CHAPTER 3

Markov Chains and Related Stochastic Models

3.0 Introduction

Mathematical models are either deterministic or stochastic, and some settings can be represented with both deterministic and stochastic models. However, in many situations arising in the social and life sciences, there are phenomena for which stochastic models are the appropriate ones. In particular, in many circumstances the behavior of plants, animals, and people exhibits a degree of randomness that must be built into the models if predictions are to correspond with observations. There are a great variety of stochastic models—that is, sets of assumptions—that can be used to study these situations, and in this chapter we examine in detail only one rather special case. This special case, Markov chains, has proved to be widely applicable and a reasonably effective way of modeling many situations arising in the real world. Even in circumstances where detailed predictions based on the model differ from observations, we frequently gain insight into the process by studying these simple models.

We introduce the models studied in this chapter through a number of examples from a variety of settings. We then present the mathematical concepts and notation that we use in analyzing these situations. Finally, we develop parts of the theory of the models for two especially important subclasses of models. Throughout, we use the situations introduced in Section 3.1 to illustrate and apply our methods.

3.1 The Setting and Some Examples

A basic assumption we make throughout this chapter is that all situations we study have the property that we observe a system sequentially through time, and that at each observation the system can be determined to be in one of a finite number of states or to be satisfying a finite number of conditions. This is an assumption about our ability to classify circumstances or behaviors in useful ways. It is probably most effective to illustrate the notion through examples.

Animal Ranges

Consider a locale consisting of rocks, scrub brush, open meadow, and a stream (see Figure 3.1), and suppose that this locale is home for a small animal, say a marmot. We seek

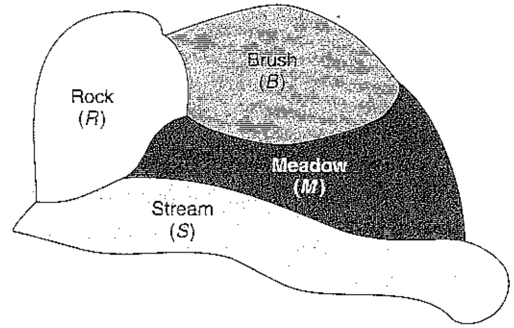


Figure 3.1

to model the movement of the marmot through time by noting its location at sequential observations and then forming a mathematical system that represents these movements in an appropriate way.

The use of a figure such as Figure 3.1 includes an assumption that the area can be partitioned meaningfully into the subareas shown, and we also assume that when an observation is made, we can determine which subarea contains the marmot. These may seem like natural and perhaps trivial assumptions, but in many experimental situations they are difficult to interpret or verify. For small animals, the task of keeping track of the creature may be a challenge, and for large animals, for which the use of tracking collars may help with the location problem, we may have the animal moving through many different subareas of its range. Also, if observations continue over an extended period of time, the nature of a specific area may change. Brush may become meadow, or during a very wet season, a part of the meadow may disappear into the stream. However, such issues are not a direct part of our current model, and we do not consider them further.

We suppose, therefore, that the location of the marmot can be tracked through time and that a sequence of observations can be represented as a sequence of locations and occupancy times. For instance, using the shorthand R, B, M, and S to denote the rocks, brush, meadow, and stream, respectively, then we might represent a particular sequence of observations by a sequence of letters (the locations) and numbers (the occupancy times):

$$R - 34.2 - B - 12.3 - R - 2.4 - B - 21.8 - M - \cdots$$

This sequence is to be interpreted as follows: When observations begin, the marmot is in the rocks, and it remains there for 34.2 minutes. It then moves to the brush, where it remains for 12.3 minutes, after which it again moves to the rocks, where it remains for 2.4 minutes, and so on. Although occupancy times play an important role in many models, we can illustrate the basic ideas of the model-building and analysis process by concentrating on the locations alone. Also, if (as is frequently the case) observations are made at discrete times, then the observational data consist solely of a sequence of locations. Depending on the criteria used to make observations, locations may or may not appear successively in the sequence. Thus the sequence of locations given above might be represented as $RBRBM \cdots$. In this representation, juxtaposition denotes the results of successive observations. If the definition of observation permits the marmot to be in the same location on successive observations, then a sequence of locations such as $RRRBMMS$ is possible. Here it might be suggestive to write the result of the seven observations as

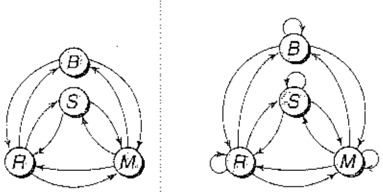
$$R \to R \to R \to B \to M \to M \to S$$

When represented using juxtaposition or a diagram as above, a sequence of observations is frequently referred to as a *sample path.*

In this example, we identify a "state of the system" with a location of the marmot. From a biological perspective, the marmot may be sleeping or resting while in the rock pile, feeding while in the meadow, and simply in transit while in the brush. In a more complex situation where we have several marmots, we need a more elaborate definition of the state of the system. We must decide, for instance, whether we are going to keep track of individual marmots. If we decide to do so, then the state corresponding to marmot number 1 being in the rocks and marmot number 2 being in the brush is different from the state corresponding to marmot number 1 being in the brush and marmot number 2 being in the rocks. If we simply keep track of how many marmots are in the various locations, then these two situations correspond to the same state.

If observations are made periodically in time, then the marmot can be in any two locations on successive observations. If the time between successive observations is short compared with the time it takes the marmot to move through a location, then it is possible for the marmot to be observed either in the same location or in adjacent locations on successive observations. Sometimes the experiment is set up so that the marmot is observed each time it moves from one location to another. In such cases, two successive observations must have the marmot in different but adjacent locations. For each assumption about how observations are made, we have a diagram similar to that shown in Figure 3.2. In Figure 3.2a the arrows represent possible moves of the marmot under the assumption that the marmot is observed only when it moves from one area to another, and in Figure 3.2b the arrows represent possible moves when it is observed whenever it moves *and* at regularly spaced times—and can therefore be in the same area on successive observations. Figure 3.2 provides two examples of a **transition diagram.** It illustrates possible transitions between states of our process. Later we will add probabilities of the transitions to the diagram.

There are natural assumptions one can make about the movements of the marmot. The likelihood of the marmot moving from the brush to the rock may depend on several of the preceding moves, or it may depend only on the immediately preceding move, or it may be independent of preceding moves. Each of these assumptions leads to a mathematical model whose predictions can be compared with observations. Here, we consider in detail only one such assumption: We assume that the likelihoods of the various possible moves



(a) (b)

Figure 3.2

for the marmot depend only on the location of the marmot, not on how much time has elapsed since we began the observations and not on the previous moves of the marmot. It is this assumption (which we will make formal soon) that distinguishes Markov chains from other stochastic processes. In circumstances where the Markov assumption is appropriate, it is customary to arrange the likelihoods or probabilities of moves in a table or matrix. We illustrate this idea with two examples.

EXAMPLE 3.1 Assume that the marmot is observed only when it moves from one subarea to another and that it is equally likely to make any move available to it. Then the probabilities of various moves are as shown in Table 3.1. ■

Table 3.1

		Location after Move			
		R	B	M	S
Location before Move	R	0	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$
	B	$\frac{1}{2}$	0	$\frac{1}{2}$	0
	M	$\frac{1}{3}$	$\frac{1}{3}$	0	$\frac{1}{3}$
	S	$\frac{1}{2}$	0	$\frac{1}{2}$	0

EXAMPLE 3.2 For this example, assume that the marmot is observed periodically and whenever it moves from one subarea to another. Assume that it is twice as likely to remain where it is as to move, and if it moves, then it is equally likely to make any move available to it. Under these assumptions, the probabilities of various moves are as shown in Table 3.2. ■

Table 3.2

		Location after Move			
		R	B	M	S
Location before Move	R	$\frac{2}{3}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$
	B	$\frac{1}{6}$	$\frac{2}{3}$	$\frac{1}{6}$	0
	M	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{2}{3}$	$\frac{1}{9}$
	S	$\frac{1}{6}$	0	$\frac{1}{6}$	$\frac{2}{3}$

The tasks of verifying the entries in these tables and finding corresponding tables in other situations are topics of the exercises.

Table 3.2 lists the probabilities for single transitions by the marmot. In that table, a transition occurs whenever a fixed period of time has elapsed or the marmot has changed areas, whichever event occurs first. We are often interested in the results of multiple transitions, and we need the probabilities of each possible result. One method of computing these probabilities is by using a simple tree diagram. For example, suppose that the marmot is in the brush at a certain time, and we want to know how likely it is to be in each of the areas

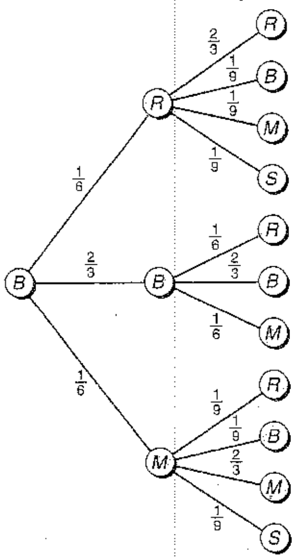


Figure 3.3

R, B, M, S after two transitions. The tree diagram in Figure 3.3 illustrates the computations needed to compute these four probabilities. We see that

$$\text{Probability}[B \to R \text{ in two transitions}] = \left(\frac{1}{6}\right)\left(\frac{2}{3}\right) + \left(\frac{2}{3}\right)\left(\frac{1}{6}\right) + \left(\frac{1}{6}\right)\left(\frac{1}{9}\right) = \frac{13}{54}$$

$$\text{Probability}[B \to B \text{ in two transitions}] = \left(\frac{1}{6}\right)\left(\frac{1}{9}\right) + \left(\frac{2}{3}\right)\left(\frac{2}{3}\right) + \left(\frac{1}{6}\right)\left(\frac{1}{9}\right) = \frac{26}{54}$$

$$\text{Probability}[B \to M \text{ in two transitions}] = \left(\frac{1}{6}\right)\left(\frac{1}{9}\right) + \left(\frac{2}{3}\right)\left(\frac{1}{6}\right) + \left(\frac{1}{6}\right)\left(\frac{2}{3}\right) = \frac{13}{54}$$

$$\text{Probability}[B \to S \text{ in two transitions}] = \left(\frac{1}{6}\right)\left(\frac{1}{9}\right) + \left(\frac{1}{6}\right)\left(\frac{1}{9}\right) = \frac{2}{54}$$

In this example it is relatively straightforward to determine the two-step transition probabilities using tree diagrams similar to Figure 3.3. In cases with more states or more steps, this method becomes unwieldy and we need an alternative.

The Effects of Group Structure on Small-Group Decision Making

In many group decision-making situations, we believe that in addition to the merits of the alternatives being considered, there are aspects of the dynamics of the group that influence the outcome. For instance, once a group of six people reaches a division of five to one in favor of some alternative, the mere fact of this division exerts some influence on the dissenting member. In this example we describe a model designed to test this conjecture in a setting in experimental psychology.

To isolate the possible influence of the group structure on decision making, we introduce an experiment designed to minimize other influences. In particular, care must be taken to ensure that the alternatives appear equally attractive and that no individual in the group assumes a leadership position. Suppose a group of people performs a sequence of trials. Each trial consists of the presentation of a stimulus—a set of alternatives—to be evaluated and a discussion of the merits of the alternatives. The discussion continues until consensus is reached. Suppose that a stimulus consists of a set of three geometrical designs that are to be evaluated according to some criteria. Each member of the group is able to convey a preference to the investigator without other members of the group knowing what that preference is, and preferences can be changed at will. The subjects are asked to express a preference as soon as the stimulus is shown to the group and then to begin discussions seeking to reach consensus. Each subject is to convey a preference change to the group right after it is conveyed to the investigator. After consensus is reached, the group is told which of the designs is "best" in terms of the criteria. The group is led to believe that there is a system behind the assignment of values to the designs in the set, but actually the designs in each set are ranked randomly. Consequently, as far as the group is concerned, each design is equally preferable. Also, techniques of selective reinforcement can be used to discourage the emergence of a group leader. For instance, selection of the best design can be manipulated so that each member of the group appears to have about the same percentage of "correct" initial selections.

Preference selections are monitored and recorded. The process continues either for a certain period of time or for a certain number of trials. In practice, it might be desirable to discard the first few trials because the subjects are becoming familiar with the experimental procedure during that period.

To make the discussion more specific, suppose that there are four individuals and that each stimulus consists of three designs. Each individual can select any of the three alternatives as the best, and consequently, there are $3^4 = 81$ possible distributions of preferences for the group. However, the experiment has been designed so that the alternatives appear equally attractive, and we ought not to distinguish among them. For example, if the alternatives are X, Y, and Z, and if the choices of the members are X, Y, Y, Y in one case and Y, X, X, X in another, then these choices should be viewed as equivalent from the standpoint of the structure of the group. Both represent a group structure in which three people vote for one alternative and a single individual votes for another. Also, there is no reason to distinguish among members of the group. That is, three votes for alternative X and one vote for Y should be considered the same regardless of which member votes for alternative Y.

It follows that the important information is the number of individuals who voted for the most popular alternative, the number who voted for the second most popular alternative, and the number who voted for the least popular alternative. That is, the relevant information is contained in a triple of integers (x, y, z), where x is the number of individuals voting for the most popular alternative, y is the number voting for the second most popular alternative, z is the number voting for the least popular alternative, and $x + y + z = 4$. The possible triples are $(4, 0, 0)$, $(3, 1, 0)$, $(2, 2, 0)$, and $(2, 1, 1)$. We refer to these triples as **group compositions** and we write them simply as xyz; thus 310 is the same as $(3, 1, 0)$.

Suppose that the group compositions are monitored continuously and every preference change is recorded. A change in group composition occurs whenever any subject changes a vote. Each preference change is equivalent to a change from one group composition to

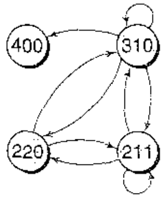


Figure 3.4

Table 3.3

		Composition after One Shift			
		400	310	220	211
Composition before Shift	310	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$
	220	0	$\frac{1}{2}$	0	$\frac{1}{2}$
	211	0	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$

another, possibly the same one. We assume that only one vote changes at a time. In the rare event that two people simultaneously indicate a change of vote, we arbitrarily select one to be changed first. It follows that preference changes by individuals are equivalent to shifts between group compositions that can be effected by the change of a single vote. For example, 211 → 220 is an admissible transition, but 211 → 400 is not. The possible shifts are shown by arrows in Figure 3.4.

Note that it is possible for a single vote to change and for a group with composition 211 to change to a group with the same composition. Likewise for a group with composition 310. Because a trial of the experiment ends when consensus is reached, there are no possible shifts from group composition 400.

The probabilities of various shifts can be conveniently summarized in a table. The entries in the table will depend, of course, on the assumptions about the voting behavior of the subjects. For example, if each subject is equally likely to change her or his vote and is equally likely to change to each of the other alternatives, then we have the information in Table 3.3. The task of verifying the entries in Table 3.3 is the topic of Exercise 5.

As a final comment in this section, we recall that an important part of a mathematical model is the assumptions. Our discussion of small-group decision making provided an example of one possible set of assumptions that can be made in that situation—of course, there are many alternatives. It is difficult (and frequently impossible) to directly test the validity of the assumptions. Instead, it is customary to compare the predictions based on the assumptions with observations. If the observations are consistent with the predictions, then one has reason to continue the study. If the observations are not consistent with the predictions, then the assumptions need to be reviewed and modified. In the following sections of this chapter, we develop a theory and techniques to make predictions, and in Chapter 4 we will develop techniques for making predictions based on simulation models. Making predictions based on assumptions and then comparing the predictions with observations are part of the cycle of model building described in Chapter 1.

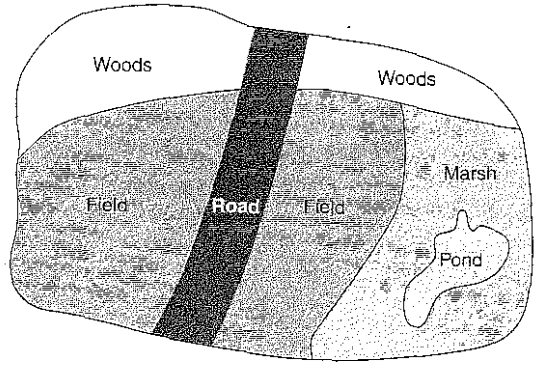


Figure 3.5

Exercises 3.1

1. In the model of a marmot's range described in this section, verify the entries in Table 3.1 under the assumptions given in Example 3.1.
2. In the model of a marmot's range described in this section, verify the entries in Table 3.2 under the assumptions given in Example 3.2.
3. A deer has as its range the area diagrammed in Figure 3.5, and its movements are observed and recorded as follows: The location of the deer is noted every hour and every time it moves from one area of its range to another. For this purpose, the woods and the field to the east of the road are distinguished from the woods and the field to the west of the road. If the deer crosses the road, then it moves only from field to field; that is, it does not move to or from the woods when crossing the road. Suppose that the probabilities of moves depend only on its current location and not on what happened prior to its last move.

   Assume that the deer is twice as likely to remain where it is as to move and that every move that does not require it to cross the road is equally likely. Also, if it moves, each option that does not require it to cross the road is three times as likely to be selected as an option that involves crossing the road. Create a table similar to Table 3.1 for this situation.
4. A marmot lives in the region diagrammed in Figure 3.6. Suppose the marmot is observed every hour and each time it moves from one area to another. Suppose that the probabilities of moves depend only on its current location, and not on what happened prior to its last move. Also suppose it is equally likely to move and to remain where it is. If it moves, the probability of its moving to an adjoining area is proportional to the number of resources available to it in that area in comparison to the resources in all adjoining areas. The areas bordering the pond have water in addition to the other resources specified. Create a table similar to Table 3.1 for this situation.
5. Consider the small-group decision-making situation described in this section in which three alternatives are presented to four individuals. If each subject is equally likely to change her or his vote, and is equally likely to change to each of the other alternatives,

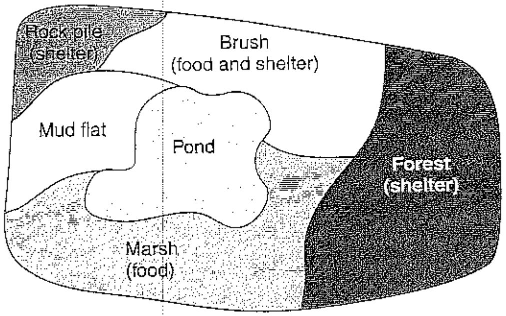


Figure 3.6

show that the probabilities of shifts between group compositions are as shown in Table 3.3.

6. Consider the small-group decision-making situation described in this section in which three alternatives are presented to four individuals. If from a specific group composition each possible shift to another group composition, possibly the same one, is equally likely, create a table similar to Table 3.3 for this case. How does the table change if only shifts to different group compositions are possible?
7. Consider a small-group decision-making situation similar to that described in this section with four alternatives presented to five individuals. What are the group compositions in this case? If each subject is equally likely to change her or his vote, and is equally likely to change to each of tbe other alternatives, find the probabilities of shifts between group compositions in this case and create a table similar to Table 3.3.
8. In the small-group decision-making experiment described in this section, define a shift toward consensus as one of the following: $211 \to 310, 220 \to 310, 310 \to 400$. Assume that a voter who can make a shift toward consensus is twice as likely to make a vote change as any other voter, and that if such a voter changes her or his vote, all changes are equally likely. Also, assume that all other voters are equally likely to change their votes and that all choices are equally likely. Create a table similar to Table 3.3 for this situation.
9. In the small-group decision-making experiment described in this section, define a shift toward consensus as one of the following: $211 \to 310, 220 \to 310, 310 \to 400$. Assume that a voter who can make a shift toward consensus is twice as likely to make a vote change as any other voter, and that if such a voter changes her or his vote, the change is twice as likely to be toward consensus as otherwise. Suppose that all other voters are equally likely to change their votes and that all vote changes are equally likely for these voters. Create a table similar to Table 3.3 for this situation.
10. Consider a small-group decision-making situation similar to the one described in this section, but with six individuals and three alternatives. Formulate a model similar to the one of this section under the following assumption: An individual who is the only person voting for an alternative is three times as likely to change her vote as an individual who is one of two or more voting for an alternative. If an individual changes her vote, she is equally likely to change to any other alternative. Create a table similar to Table 3.3 for this case.

3.2 Basic Properties of Markov Chains

With the examples and discussion of Section 3.1 as a guide, we turn to a discussion of two more general settings that include the examples of Section 3.1. We consider a system that can be in any of N possible states, and we observe the system at n successive times. The concepts of *state* and *being in a state* or *occupying a state* are taken as undefined terms. When we construct logical models of systems in specific circumstances, we must assign meanings to these terms, but in this general discussion they are left undefined. We usually refer to the states simply by the integers $1, 2, \ldots, N$.

By the very nature of a Markov chain, it is likely that for most observations, the specific state occupied by the system cannot be determined in advance, and one knows only the probabilities that it will be in the various states. Consequently, the status of a system is usually given as a state vector.

Definition 3.1 A *state vector* $\mathbf{x}$ for a Markov chain with N states is an N-vector $\mathbf{x} = [x_1 \quad x_2 \quad \cdots \quad x_N]$, where x_i is the probability that the system is in state i, $i = 1, 2, \ldots, N$. The state vector on the mth observation will be denoted by $\mathbf{x}(m)$.

As an example, to say that the state of a four-state Markov chain is specified by the state vector $[0 \quad .25 \quad .5 \quad .25]$ means that the system is not in state 1, is in state 2 with probability .25, is in state 3 with probability .5, and is in state 4 with probability .25. If the system is known to be in a specific state, say state j, then the state vector has the jth coordinate equal to 1 and the remaining coordinates equal to zero. For instance, if a four-state Markov chain is known to be in state 2, then the state vector is $[0 \quad 1 \quad 0 \quad 0]$. In a Markov chain with N states, if the system is equally likely to be in any state, then the state vector has all coordinates equal to $1/N$.

If the system is in state i on the kth observation and in state j on the $(k+1)$th observation, then we say that the system has made a *transition* from state i to state j at the kth *trial, step,* or *stage* of the process. We also say that the system has made a *move* from state i to state j.

It will be useful to work with another example, one that is somewhat simpler than those introduced in Section 3.1.

We use a setting that is familiar as a version of the classic maze of experimental psychology. The study of the behavior of mice (and rats) in mazes has been—and continues to be—important in generating and verifying hypotheses that lead to useful models for animal behavior.

EXAMPLE 3.3 Suppose that a mouse is released in the maze shown in Figure 3.7 and its behavior is observed. The illumination level in each compartment of the maze is maintained as shown in the figure. The system to be studied consists of the mouse and the maze, and we assume that the mouse is always in exactly one compartment and that it is possible to

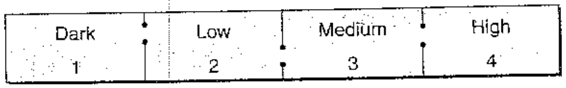


Figure 3.7

determine that compartment. The system is said to be in state i if the mouse is in compartment i, $i = 1, 2, 3, 4$. Observations are to be made, and the state of the system recorded every 2 minutes and each time the mouse moves from one compartment to another, necessarily an adjacent one. To illustrate the possible transitions, suppose that on one observation the mouse is in compartment 2. On the next observation it may be in compartment 1, 2, or 3; it cannot be in compartment 4 according to our definition of observation. ■

In the situation of Example 3.3, we assume, as we would expect, that the mouse moves in unpredictable ways, and we describe its movements in probabilistic terms. Suppose that the mouse is in state i on observation k, and we wish to determine the probability that it is in state j on observation $k + 1$. In general, we might expect this probability to depend on the states i and j, the observation k, and the history of the movements of the mouse prior to its arrival in state i on the kth observation. There are, of course, many ways in which the transition could depend on the history of the process. For instance, we could assume that the mouse has complete recall of past movements and that its transition probabilities at the kth step depend on the total prior history of its movements. Although in some circumstances such an assumption may be appropriate, in many cases it leads to very complex models without yielding any improvements in predictions. A simpler assumption would be that the transition probabilities depend only on the most recent past, say the last 1, 2, or 3 moves. We investigate here the situation wherein transition probabilities depend only on the current state, not on the prior history of the process. This includes the assumption that they do not depend on k—the number of steps for which the process has been observed.

We summarize this discussion by giving the key assumption that distinguishes Markov chains from more general stochastic processes.

■ The Markov Assumption

A **Markov chain** is a stochastic process with a finite number of states and with the property that if it is in state i on one observation, then the probability that it will be in state j on the next observation depends on states i and j (which may be the same state) and not on the observation number or on the history of the process prior to the current observation.

It will be useful to introduce the following notation and terminology.

Definition 3.2 Let p_{ij} denote the conditional probability that if the system is in state i on one observation, then it will be in state j on the next observation, $1 \le i \le N$, $1 \le j \le N$. These probabilities are called *transition probabilities*, or, more precisely, one-step transition probabilities. For each Markov chain, the $N \times N$ matrix $\mathbf{P}$ whose ij-entry is p_{ij} is called the *transition matrix* for the Markov chain.

We have

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & \cdots & p_{1N} \\ p_{21} & p_{22} & p_{23} & \cdots & p_{2N} \\ \vdots & \vdots & \vdots & & \vdots \\ p_{N1} & p_{N2} & p_{N3} & \cdots & p_{NN} \end{bmatrix}$$

as the transition matrix for a Markov chain whose transition probabilities are p_{ij}, $1 \le i \le N, 1 \le j \le N$.

Remark It is important to note that because of the Markov assumption, p_{ij} is the probability that if the system is in state i on observation k then it will be in state j on observation $k+1$, independent of k. Therefore, if one knows that (with states and observations specified) the transition probabilities do depend on the observation numbers, then the stochastic process is not a Markov chain according to our definition.

It follows from the definition of transition probability that the probability p_{ij} of making a transition from state i to state j at the kth step is the same as the probability of the system being in state j on the second observation given that it was in state i on the first observation. It is sometimes appropriate to maintain the assumption that the transition probabilities are independent of the history of the process but to permit them to depend on the time—that is, on how long the process has been observed. In this case we have a more general stochastic process usually called a nonhomogeneous Markov chain. We will not pursue this more general situation here.

Each entry in the ith row of the transition matrix is a probability, and if the system is in state i on one observation, then it must be in some state j, $1 \le j \le N$, on the next observation. Consequently, for each i we have $\sum_{j=1}^{N} p_{ij} = 1$, and the vectors $\mathbf{p}_i = [p_{i1} \quad p_{i2} \quad \cdots \quad p_{iN}]$, $i = 1, 2, \ldots, N$, are probability vectors. Each row of the transition matrix is a probability vector.

The transition matrix $\mathbf{P}$ has entries that are the probabilities of making transitions from one specified state to another in one step. There are corresponding matrices for multistep transition probabilities.

Definition 3.3 Let $\mathbf{P}(m) = [p_{ij}(m)]$ be the matrix for which the ij-entry is the probability of making a transition from state i to state j in m steps, $1 \le i \le N$, $1 \le j \le N$, $m = 2, 3, \ldots$. Clearly, $\mathbf{P}(1) = \mathbf{P}$.

Remark Note that we use the terms *step* and *move* in the same way as we use the term *transition*. It is common to talk about the probability of a transition from state i to state j in m steps, m moves, or m transitions.

EXAMPLE 3.4 Consider the situation described above in which a mouse moves in a maze with compartments illuminated at different levels, and formulate a Markov chain model under the following assumption: The mouse remains in the same compartment with probability .5, and the rest of the time it is equally likely to make any of the moves open to it.

We define the states as follows: The system is in state i if the mouse is in compartment i, $i = 1, 2, 3, 4$. Because the mouse remains in the same compartment half the time, $p_{11} = .5$, $p_{22} = .5$, $p_{33} = .5$, and $p_{44} = .5$. Next, if the mouse is in state 1 on one observation, then on the next observation it can be only in state 1 or 2. Consequently, using the assumption, we have $p_{12} = .5$. Likewise, $p_{43} = .5$. Also, if the mouse is in state 2 on one observation, then on the next observation it can be in state 1, 2, or 3. Consequently, because it is in state 2 with probability .5, it is in state 1 or 3 with probability .5, and because it is equally likely to be in either, we have $p_{21} = .25$ and $p_{23} = .25$. Similarly, $p_{32} = .25$, and $p_{34} = .25$. Finally, by the way observations are defined, p_{13}, p_{14}, p_{24}, p_{31}, p_{41}, and p_{42} all equal 0. Consequently, the transition matrix for this Markov chain is

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \\ p_{41} & p_{42} & p_{43} & p_{44} \end{bmatrix} = \begin{bmatrix} .5 & .5 & 0 & 0 \\ .25 & .5 & .25 & 0 \\ 0 & .25 & .5 & .25 \\ 0 & 0 & .5 & .5 \end{bmatrix}$$ ■

It is clear that the entries in the tables constructed in Section 3.1 are transition probabilities, and we will represent them in this way in the future.

State Vectors

A **state vector** is a probability vector that describes the status of a Markov chain at an observation, and the state vectors at two successive observations are related in a simple way. Indeed, if $\mathbf{x}(m)$ and $\mathbf{x}(m+1)$ denote the state vectors at the mth and $(m+1)$st observations, respectively, then $\mathbf{x}(m+1) = \mathbf{x}(m)\mathbf{P}$. To verify this relationship, suppose that the state vector at a specific observation is $\mathbf{x} = [x_1 \quad x_2 \quad \cdots \quad x_N]$. (Here we suppress the dependence on the observation number m for notational convenience.) Then, using the definition of p_{j1} for $j = 1, 2, \ldots, N$, we can conclude that the probability it is in state 1 on the next observation is

$$y_1 = x_1 p_{11} + x_2 p_{21} + \cdots + x_N p_{N1}$$

Thus y_1 is the dot product of $\mathbf{x}$ and the first column of $\mathbf{P}$. Likewise, this time using the definition of p_{j2} for $j = 1, 2, \ldots, N$, we can conclude that the probability it is in state 2 on the next observation is

$$y_2 = x_1 p_{12} + x_2 p_{22} + \cdots + x_N p_{N2}$$

Continuing in this fashion, we find that the probability that it is in state N on the next observation is

$$y_N = x_1 p_{1N} + x_2 p_{2N} + \cdots + x_N p_{NN}$$

That is, if the state vector at one observation is $\mathbf{x} = [x_1 \quad x_2 \quad \cdots \quad x_N]$, then the state vector at the next observation is

$$[y_1 \quad y_2 \quad \cdots \quad y_N] = \mathbf{xP}$$

Thus, with $\mathbf{x}(m)$ and $\mathbf{x}(m+1)$ as defined above, we have

$$\mathbf{x}(m+1) = \mathbf{x}(m)\mathbf{P} \tag{3.1}$$

Multistep Transitions and the Sequence of State Vectors

If the initial state vector is $\mathbf{x}$, then the state vector at the next observation is $\mathbf{xP}$, the state vector at the second observation is $(\mathbf{xP})\mathbf{P} = \mathbf{xPP} = \mathbf{xP}^2$, and so on. That is, the sequence of state vectors is

$$\mathbf{x}, \mathbf{xP}, \mathbf{xP}^2, \ldots, \mathbf{xP}^k, \ldots$$

A Markov chain is determined by the set of states, the transition matrix, and the initial state vector for the system. It is frequently useful to represent the information on the set of states and the transition probabilities in a *transition diagram*. The most common form of a transition diagram has the states represented by symbols (generally numbers or letters in small circles), an arrow directed from state i to state j if $p_{ij} > 0$, and a number near the arrow with the value of p_{ij}. The transition diagram for the model described in Example 3.3 is shown in Figure 3.8.

General stochastic processes can be studied using tree diagrams, and in particular, Markov chains can be studied in that way. There is, of course, a close connection between the information usually included on tree diagrams and the information included in a transition matrix. Given the initial state of the system, either tree diagrams or multistep transition matrices can be used to determine how the process evolves over time. Also, using a one-step transition matrix, it is always possible to determine the multistep matrix by using a tree diagram, and we illustrate this in the next example.

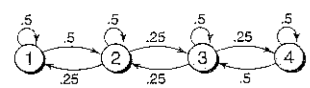


Figure 3.8

EXAMPLE 3.5 Consider the situation described in Example 3.4 and determine the 2-step transition matrix—that is, the matrix of two-step transition probabilities—by using tree diagrams.

Because the first row of the two-step transition matrix consists of the probabilities of making transitions from state 1 to states 1, 2, 3, 4 in two steps, we begin by constructing the tree diagram when the process begins in state 1. That tree diagram is shown in Figure 3.9(a). Using the information on the tree diagram, we find that

$$p_{11}(2) = .375, \quad p_{12}(2) = .5, \quad p_{13}(2) = .125, \quad p_{14}(2) = 0$$

To determine the entries in the second row of the two-step matrix, we use a tree diagram for which the process begins in state 2, as shown in Figure 3.9(b). Using the information on the tree diagram, we find that

$$p_{21}(2) = .25, \quad p_{22}(2) = .4375, \quad p_{23}(2) = .25, \quad p_{24}(2) = .0625$$

Similar arguments lead to the third and fourth rows of the two-step transition matrix. They are

$$\begin{array}{llll} p_{31}(2) = .0625, & p_{32}(2) = .25, & p_{33}(2) = .4375, & p_{34}(2) = .25 \\ p_{41}(2) = 0, & p_{42}(2) = .125, & p_{43}(2) = .5, & p_{44}(2) = .375 \end{array}$$

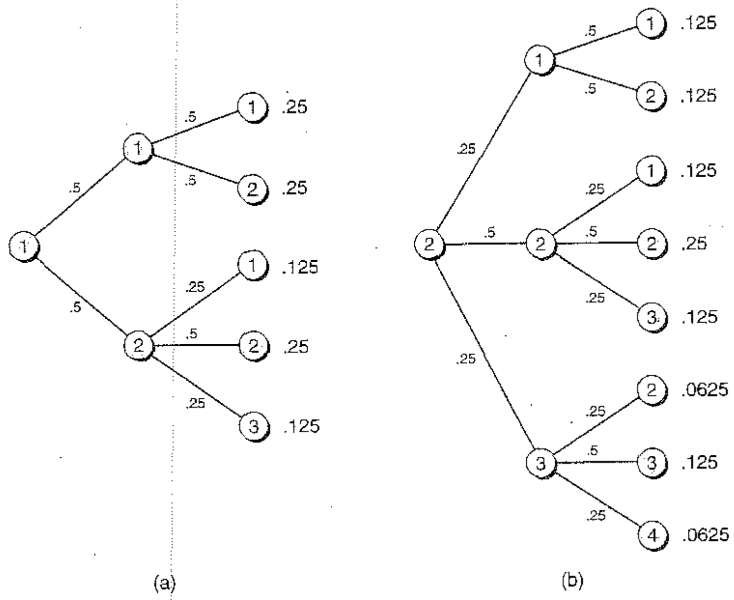


Figure 3.9

It follows that the two-step transition matrix $\mathbf{P}(2)$ is

$$\mathbf{P}(2) = \begin{bmatrix} .3750 & .5000 & .1250 & 0 \\ .2500 & .4375 & .2500 & .0625 \\ .0625 & .2500 & .4375 & .2500 \\ 0 & .1250 & .5000 & .3750 \end{bmatrix}$$ ■

The technique illustrated in Example 3.5 is a very general one, and it can be used, in particular, to construct the m-step transition matrix for any integer m. However, it is clear that for Markov chains with a large number of states or a large number of steps m, the effort involved in using this method can be prohibitive. Indeed, one of the great benefits of using Markov chains in mathematical models is that there is a much simpler way to determine multistep transition matrices once you know the one-step transition matrix.

THEOREM 3.1 Let $\mathbf{P} = [p_{ij}]$ be the transition matrix of a Markov chain. Then the ij-entry of the m-step transition matrix $\mathbf{P}(m)$ is the ij-entry of $\mathbf{P}^m$, the mth power of the one-step transition matrix.

Proof. The proof is a direct consequence of the definition of conditional probability and the Markov assumption, and it illustrates a useful approach to computing multistep transition probabilities. Indeed, the ij-entry of $\mathbf{P}(m)$ is the conditional probability that the system is in state j given that it began in state i and made m transitions. Consider each of the m-step sample paths from state i to state j as consisting of a path of length $m-1$ followed by a single step. Then, after $m-1$ steps the system must be in some state, say state k. The conditional probability that it is in state k given that it began in state i and made $m-1$ transitions is the ik-entry of the $(m-1)$-step transition matrix, $\mathbf{P}(m-1)$. Therefore, the probability that the system is in state k after $m-1$ steps and in state j after m steps is equal to $p_{ik}(m-1)p_{kj}$. Finally, using the fact that each sample path is in some state k after $m-1$ transitions, we have

$$p_{ij}(m) = \sum_{k=1}^{N} p_{ik}(m-1)p_{kj}$$

This shows that $\mathbf{P}(m) = \mathbf{P}(m-1)\mathbf{P}$. Applying the same reasoning to $\mathbf{P}(m-1)$, we find that $\mathbf{P}(m-1) = \mathbf{P}(m-2)\mathbf{P}$, and consequently $\mathbf{P}(m) = \mathbf{P}(m-2)\mathbf{P}\mathbf{P} = \mathbf{P}(m-2)\mathbf{P}^2$. Continuing the argument, we find that

$$\mathbf{P}(m) = \mathbf{P}(m-1)\mathbf{P} = \mathbf{P}(m-2)\mathbf{P}^2 = \mathbf{P}(m-3)\mathbf{P}^3 = \cdots = \mathbf{P}(1)\mathbf{P}^{(m-1)}$$

But $\mathbf{P}(1) = \mathbf{P}$, and consequently $\mathbf{P}(m) = \mathbf{P}^m$. ■

EXAMPLE 3.6 Consider the situation described in Example 3.4, and compute the two-step transition matrix $\mathbf{P}(2)$ using Theorem 3.1.

In Example 3.4 we determined the transition matrix for the Markov chain to be

$$\mathbf{P} = \begin{bmatrix} .5 & .5 & 0 & 0 \\ .25 & .5 & .25 & 0 \\ 0 & .25 & .5 & .25 \\ 0 & 0 & .5 & .5 \end{bmatrix}$$

Using Theorem 3.1, we find that the two-step transition matrix $\mathbf{P}(2)$ is

$$\mathbf{P}(2) = \begin{bmatrix} .5 & .5 & 0 & 0 \\ .25 & .5 & .25 & 0 \\ 0 & .25 & .5 & .25 \\ 0 & 0 & .5 & .5 \end{bmatrix} \begin{bmatrix} .5 & .5 & 0 & 0 \\ .25 & .5 & .25 & 0 \\ 0 & .25 & .5 & .25 \\ 0 & 0 & .5 & .5 \end{bmatrix}$$

$$= \begin{bmatrix} .3750 & .5000 & .1250 & 0 \\ .2500 & .4375 & .2500 & .0625 \\ .0625 & .2500 & .4375 & .2500 \\ 0 & .1250 & .5000 & .3750 \end{bmatrix}$$

which is the same result we obtained in Example 3.5, as it must be. ■

EXAMPLE 3.7 Consider the process of group decision making described in Section 3.1. In particular, suppose we have three alternatives and a group of four individuals. We form a Markov chain model under the assumption that each individual is equally likely to change her or his vote and is equally likely to change to each of the other alternatives.

Table 3.3 contains the probabilities for shifts from the group compositions 310, 220, and 211 to the group compositions 400, 310, 220, and 211. To make use of the Markov chain concept, we need to include the group composition 400 as a state. The situation was originally described as an experiment that ended as soon as consensus was reached—that is, as soon as the group reached composition 400. However, for the purpose of our Markov chain concept, it is useful to view 400 just as we view any other state, but with the characteristic that the system never leaves that state. This can be accomplished by setting the transition probability from the state corresponding to group composition 400 to itself equal to 1, and the probability of making a transition from state 400 to any other state equal to 0.

For this example, we define the states of the system as the group compositions, and we define state 1 as group composition 400, and states 2, 3, and 4 as group compositions 310, 220, and 211, respectively. Then, with the understanding that once the system reaches group composition 400 it does not leave it, we have the transition matrix

$$\mathbf{P} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ \frac{1}{8} & \frac{1}{8} & \frac{3}{8} & \frac{3}{8} \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{bmatrix}$$

Suppose the group is initially in state 4; that is, the group composition is 211. How many vote changes are required before the probability of the group reaching consensus first reaches .5? We answer this question by computing successive powers $\mathbf{P}^m$ of the transition matrix $\mathbf{P}$ and asking for the smallest integer m for which $p_{41}(m) \geq .5$. We have $p_{41}(k) < .5$ for $1 \leq k \leq 19$ and

$$\mathbf{P}(20) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ .5536 & .1265 & .1049 & .2150 \\ .5064 & .1399 & .1160 & .2377 \\ .4941 & .1433 & .1189 & .2436 \end{bmatrix}$$

$$\mathbf{P}(21) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ .5694 & .1220 & .1012 & .2074 \\ .5239 & .1349 & .1119 & .2293 \\ .5121 & .1383 & .1147 & .2350 \end{bmatrix}$$

from which we see that after 20 vote changes, the probability $p_{41}(20) = .494$, and after 21 vote changes, $p_{41}(21) = .512$. Thus 21 vote changes are required for the probability of the group reaching consensus to exceed .5. Recall that the system remains in state 1 once it arrives there. Consequently, the sequence of entries in the (4, 1) spot in the matrices $P(m)$ is a monotone nondecreasing sequence, and once an entry is greater than .5, all subsequent entries are also greater than .5. Of course, this answer depends heavily on the initial assumptions of the model. A different set of assumptions about the likelihood of vote changes would give a different matrix $\mathbf{P}$ and a different answer. ■

Exercises 3.2

In these exercises, forming a Markov chain model requires that you identify the states and find the transition matrix.

1. Suppose that a mouse moves in the maze shown in Figure 3.7 and that observations are made every 5 minutes and every time the mouse changes compartments. Formulate a Markov chain model under the following assumptions: The mouse remains in the same compartment 40% of the time, and if it has a choice when it moves, it moves to a darker compartment twice as often as to a lighter one.

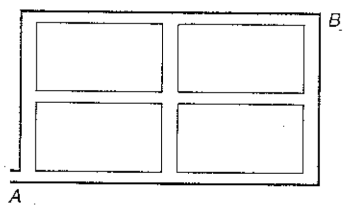


Figure 3.10

2. In the model formulated in Exercise 1, suppose the mouse is initially in the highly illuminated compartment.
   (a) Find the probability that it does not leave the highly illuminated compartment in the first five transitions.
   (b) Find the probability that it is in the same compartment after five observations.
   (c) Find the probability that it is not in the highly illuminated compartment on exactly one of the next five observations.
3. Suppose that a mouse moves in the maze shown in Figure 3.7 and that observations are made every time the mouse changes compartments. Formulate a Markov chain model under the following assumption: Whenever the mouse has a choice, it moves to a darker compartment three times as often as to a lighter one.
4. Suppose that we have the situation considered in Example 3.7. If the group is initially divided 220, how many vote changes are necessary before the probability of consensus first exceeds $\frac{1}{4}$?
5. In the group decision-making situation described in Section 3.1, define a shift toward consensus as one of the following: $211 \rightarrow 310$, $220 \rightarrow 310$, $310 \rightarrow 400$. Assume that a voter who can make a shift toward consensus is twice as likely to make a vote change as any other voter, and that if such a voter changes her or his vote, all changes are equally likely. Also, assume that all other voters are equally likely to change their votes and that all choices are equally likely.
   (a) Form a Markov chain model for this situation and find the transition matrix.
   (b) If the group is initially distributed 220, what is the most likely group composition after five vote changes? After ten vote changes?
6. An inebriated bicyclist cycles through the neighborhood shown in Figure 3.10. He begins at location A, and he traverses the streets at random. During each time interval he either rests at an intersection or pedals exactly one block.

   Suppose that at each intersection the bicyclist is three times as likely to pedal as to rest. If he pedals, he is equally likely to take any street open to him. Form a Markov chain model for this situation.
7. In the setting described in Exercise 6, suppose that the bicyclist never rests, that at any intersection he is equally likely to take any street available to him, and that once he reaches location B he stays there. Form a Markov chain model using these assumptions.
8. Consider a small-group decision-making situation similar to that described in Section 3.1 but with five individuals and three alternatives. Formulate a Markov chain

model under the following assumption: An individual who is the only person voting for an alternative is twice as likely to change her vote as a person who is one of a group of two or more voting for an alternative. If an individual changes her vote, then the probability of her changing to a particular alternative is proportional to the number of individuals voting for that alternative.

If the group initially has four individuals voting for the most popular alternative, find the probability that consensus is reached after at most three vote changes.

9. A Bloomington resident commutes to work in Indianapolis, and he encounters several traffic lights on the way to work each day. Over a period of time, the following pattern has emerged:
   - Each day the first light is green.
   - If a light is green, then the next one is always red.
   - If he encounters a green light and then a red one, then the next will be green with probability .6 and red with probability .4.
   - If he encounters two red lights in a row, then the next will be green with probability p and red with probability $1 - p$.

   Formulate a Markov chain model for this situation.

10. Consider a small-group decision-making situation similar to that described in Section 3.1 with five individuals and four alternatives. What are the group compositions in this case? Form a Markov chain model under the following assumption: Each subject is equally likely to change her or his vote and is equally likely to change to each of the other alternatives.

    If the group initially has four individuals voting for the most popular alternative, find the probability that the same holds after four vote changes.

11. A marmot lives in the region shown in Figure 3.11. Suppose that the marmot is observed every hour and each time it moves from one area to another. Formulate a Markov chain model under the following assumptions: The marmot is twice as likely to move as to remain where it is, and if it moves, the probability of its moving to a particular area is proportional to the number of resources available to it in that area in comparison to the number of resources available to it in the adjoining areas. The areas bordering the pond have water in addition to the resources specified.

    If the marmot begins in the rock pile, find the probability it is in the south meadow on the fifth observation.

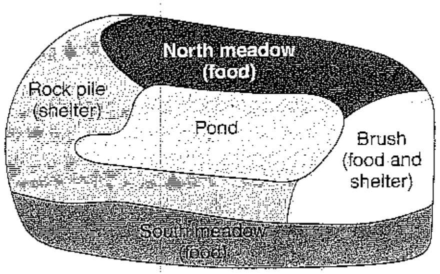


Figure 3.11

12. In the situation described in Exercise 11:
    (a) Suppose the marmot is initially in the north meadow, and find the probability that it is in the rocks after 10 transitions. Find the same probability after 20 and 40 transitions.
    (b) Suppose the marmot is initially in the brush, and determine the same probabilities as in part (a).

13. There are two coins, one fair and one biased with $\Pr[H] = .3$. A game is played by successively flipping the coins as follows:
    - The game begins with a flip of the fair coin, and the result, H or T, is noted.
    - If the result of a flip is H, then the other coin is used on the next flip, and the result is noted.
    - If the result of a flip is T, then the same coin is used on the next flip, and the result is noted.

    (a) Formulate a Markov chain model for this situation.
    (b) Find the probability that the fourth flip is a head.

14. There are six balls, two red and four green, distributed between boxes labeled 1 and 2, three balls in each box. When the game begins, there are two green balls and one red ball in box 1. The game is played as follows: A ball is selected at random from each box. The ball selected from box 1 is placed in box 2, the ball selected from box 2 is placed in box 1, and the colors of the balls in each box are noted. Then two more balls are selected, and play continues.
    (a) Formulate a Markov chain model for this situation. What are your states?
    (b) Find the probability that there are exactly two red balls in box 1 after three plays of the game.

15. Joe and Jess play a game as follows: An unfair coin with $\Pr[H] = .6$ is flipped, and the result is noted. If it comes up heads, then Jess pays Joe one dollar, and if it comes up tails, then Joe pays Jess one dollar. If each player has money, then the coin is flipped again. The game ends as soon as one player has all the money. When the game begins, Joe has one dollar and Jess has three dollars.
    (a) Formulate a Markov chain model for this game.
    (b) Find the probability that Jess has all the money after not more than four flips of the coin.

16. An experiment consists of flipping an unfair coin with $\Pr[H] = .6$ repeatedly, noting the result of each flip, until there are three consecutive heads. At that point the experiment ends.
    (a) Formulate a Markov chain model for this experiment.
    (b) Find the probability that the experiment ends after exactly six flips of the coin.

3.3 Classification of Markov Chains and the Long-Range Behavior of Regular Markov Chains

Markov chains, as examples of stochastic processes, can be used to yield information on the probabilities of events, events described in terms of states or sets of states. A key tool in studying Markov chains is the multistep transition matrix. In Section 3.2 we showed that for every Markov chain, the m-step transition matrix is the mth power of the one-step transition matrix. Beyond this common behavior, Markov chains are quite diverse. The goal of this section is to illustrate some of this diversity, to provide a useful way to classify Markov chains, and to study some selected classes in detail.

EXAMPLE 3.8 Consider two Markov chains with state space $\{1, 2, 3\}$; the first has transition matrix $\mathbf{P}$ and the second has transition matrix $\mathbf{T}$.

$$\mathbf{P} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \quad \text{and} \quad \mathbf{T} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ .5 & .5 & 0 \end{bmatrix}$$

The transition diagrams for these Markov chains are shown in Figure 3.12: Figure 3.12(a) shows the transition diagram for the Markov chain with transition matrix $\mathbf{P}$, and Figure 3.12(b) shows the transition diagram for the Markov chain with transition matrix $\mathbf{T}$.

First we study the Markov chain with transition matrix $\mathbf{P}$. If the system begins in state 1, then the sequence of state vectors is $[1 \quad 0 \quad 0] \to [0 \quad 1 \quad 0] \to [0 \quad 0 \quad 1] \to [1 \quad 0 \quad 0] \to [0 \quad 1 \quad 0] \to \cdots$. That is, the system cycles repeatedly through states 1, 2, and 3 in that order. This can also be shown by examining the powers of the transition matrix $\mathbf{P}$. Indeed, the third power of $\mathbf{P}$ is the identity matrix $\mathbf{I}$, so if the system has state vector $\mathbf{x} = [x_1 \quad x_2 \quad x_3]$ on observation m, then it has the same state vector on observation $m + 3(\mathbf{x}\mathbf{P}^3 = \mathbf{x})$.

The behavior of the Markov chain with transition matrix $\mathbf{T}$ is quite different. We have

$$\mathbf{T}^2 = \begin{bmatrix} 0 & 0 & 1 \\ .5 & .5 & 0 \\ 0 & .5 & .5 \end{bmatrix} \quad \mathbf{T}^3 = \begin{bmatrix} .5 & .5 & 0 \\ 0 & .5 & .5 \\ .25 & .25 & .5 \end{bmatrix} \quad \mathbf{T}^4 = \begin{bmatrix} 0 & .5 & .5 \\ .25 & .25 & .5 \\ .25 & .5 & .25 \end{bmatrix}$$

and

$$\mathbf{T}(30) = \mathbf{T}^{30} \begin{bmatrix} .2000 & .4000 & .4000 \\ .2000 & .4000 & .4000 \\ .2000 & .4000 & .4000 \end{bmatrix}$$

In fact, $\mathbf{T}(m)$ is the same as $\mathbf{T}(30)$—at least to the accuracy shown—for all $m > 30$. (See Exercise 5 for additional information on this situation.) We see that the rows of $\mathbf{T}(30)$ are all the same. One consequence of this is that for all observations numbered 30 and beyond, the state vector of the system is $[.2000 \quad .4000 \quad .4000]$ *independent of the initial state*. Indeed, for an initial state vector $[x \quad y \quad z]$, we have

$$[x \quad y \quad z]\,\mathbf{T}(m) = [x \quad y \quad z] \begin{bmatrix} .2000 & .4000 & .4000 \\ .2000 & .4000 & .4000 \\ .2000 & .4000 & .4000 \end{bmatrix}$$

$$= [.2000x+.2000y+.2000z \quad .4000x+.4000y+.4000z \quad .4000x+.4000y+.4000z]$$
$$= [.2000(x+y+z) \quad .4000(x+y+z) \quad .4000(x+y+z)]$$
$$= [.2000 \quad .4000 \quad .4000]$$

for any probability vector $[x \quad y \quad z]$.

That is, the system "forgets" the initial state, or the early history. The same conclusion holds for any matrix that differs from $\mathbf{T}$ only in the last row and that has a third row equal to $[p \quad 1-p \quad 0]$, $0 < p < 1$. In this case, the entries in the powers will be different (they will depend on the value of p), but the conclusion will be the same: As the number m of transitions increases, the m-step transition matrix $T(m)$ approaches a matrix with entries that are all positive and with rows that are all the same. ■

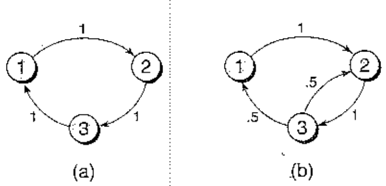


Figure 3.12

EXAMPLE 3.9 A Markov chain with state space $S = \{1, 2, 3, 4, 5\}$ has the transition diagram shown in Figure 3.13.

The transition matrix for this chain is

$$\mathbf{P} = \begin{bmatrix} 0 & 0 & .8 & .2 & 0 \\ 0 & .3 & 0 & 0 & .7 \\ .1 & .6 & .3 & 0 & 0 \\ 0 & .5 & .5 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

In this example, the probability of a direct transition from state 1 to state 2 is 0, but it is possible to go from state 1 to state 2 in more than one step, and in fact $p_{12}(2) > 0$. These facts are clear from Figure 3.13. However, we also see that it is impossible to go from state 2 to state 1 in any number of steps: $p_{21}(m) = 0$ for all values of m. It *is* possible to go from state 1 to states 3 and 4, and state 1 can be reached from states 3 and 4. Thus states 1, 3, and 4 are mutually accessible from each other. The set of all states can be partitioned using this "reachability" criterion, and we next turn to a systematic discussion of this idea. ■

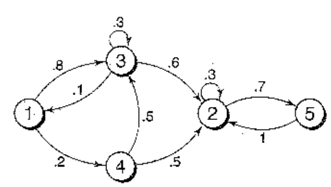


Figure 3.13

The idea introduced in Example 3.9 is helpful in classifying Markov chains, and we now show how to use it systematically. Suppose that we have a Markov chain with state space $S = \{1, 2, 3, \ldots, N\}$ and transition matrix $\mathbf{P} = (p_{ij})$. We say that state j is **accessible** from state i if there is an integer k such that $p_{ij}(k) > 0$, and we say that states i and j are **mutually accessible** if state i is accessible from state j and state j is accessible from state i. This concept can be used to partition the state space into classes of mutually accessible states. Begin with state 1, and let S_1 denote the set of all states that are mutually accessible from state 1. If $S_1 = S$, the entire set of states, then we are finished. If not, then there is a state, call it j, that is not in S_1. Let S_2 denote the set of all states that are mutually accessible with j. Continue in this way, and construct a collection of disjoint subsets of S whose union is S. Each subset consists of states that are mutually accessible, and no state belongs to more than one subset.

EXAMPLE 3.10 Applying our method of partitioning the states to Example 3.9, we have $S_1 = \{1, 3, 4\}$ and $S_2 = \{2, 5\}$. Note that $S_1 \cup S_2 = S$ and $S_1 \cap S_2 = \emptyset$. Now suppose that we form a transition matrix with the states reordered so that the states in the set S_2 are listed first and the states in the set S_1 are listed next. The order in which the states are listed within the sets S_1 and S_2 is unimportant, but the order must be the same in the rows and in the columns of the transition matrix. We have the new transition matrix

$$\begin{array}{cc} & \begin{array}{cc|ccc} \overbrace{2 \quad 5}^{S_2} & & \overbrace{1 \quad 3 \quad 4}^{S_1} \end{array} \\ \begin{array}{c} S_2\begin{cases} 2 \\ 5 \end{cases} \\ S_1\begin{cases} 1 \\ 3 \\ 4 \end{cases} \end{array} & \left[\begin{array}{cc|ccc} .3 & .7 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & .8 & .2 \\ .6 & 0 & .1 & .3 & 0 \\ .5 & 0 & 0 & .5 & 0 \end{array}\right] \end{array}$$

where the row and column labels denote the states. ■

In Example 3.10 the transition matrix has two block matrices on the main diagonal. These are shown by the lines inside the transition matrix. The 2×2 block in the upper-left corner contains the transition probabilities for transitions between states in the set $\{2, 5\}$, and the 3×3 block in the lower-right corner contains the transition probabilities for the set of states $\{1, 3, 4\}$. There is a 2×3 matrix of zeros in the upper-right corner, a consequence of the fact that no state in the set $\{1, 3, 4\}$ is accessible from a state in the set $\{2, 5\}$. The 3×2 matrix in the lower-left corner contains transition probabilities for transitions from states in the set $\{1, 3, 4\}$ to states in the set $\{2, 5\}$.

The form of the transition matrix displayed in Example 3.10 can be achieved in the general case. That is, it is always possible to relabel the classes $S_1, S_2, \ldots, S_k$ as S_1', $S_2', \ldots, S_k'$ so that the resulting transition matrix has the form

$$\begin{bmatrix} \mathbf{A}_1 & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{X} & \mathbf{A}_2 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{X} & \mathbf{X} & \mathbf{A}_3 & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & & \vdots \\ \mathbf{X} & \mathbf{X} & \mathbf{X} & \cdots & \mathbf{A}_k \end{bmatrix}$$

where each of the matrices A_j contains transition probabilities for transitions between states in the set S_j', $j = 1, 2, \ldots, k$. In Example 3.10 we have $S_1' = S_2$ and $S_2' = S_1$. Entries in the transition matrix below these blocks, the blocks denoted by X's in the transition matrix, are transition probabilities between states that are not mutually accessible. Each "0" entry represents a block of zeros. A transition matrix written in this form is said to be in **canonical form.** One of the advantages of writing a matrix in canonical form is that the matrix of m-step transition probabilities has the same form. That is, the block submatrices on the diagonal contain the m-step transition probabilities within each class, there are zeros above these diagonal blocks, and the entries below the diagonal blocks are m-step transition probabilities between states in different classes.

Definition 3.4 A Markov chain for which every two states are mutually accessible is said to be **ergodic.**

We remark that the transition matrix of an ergodic Markov chain is in canonical form no matter how the states are ordered (of course, the order must be the same in the rows and columns).

Definition 3.5 Let j be a state. Then

The **index set** of the state j, denoted by $I(j)$, is the set of all integers m such that $p_{jj}(m) > 0$.

The **period** of state j, denoted by $d(j)$, is defined to be

(i) 0 if the index $I(j)$ is empty;
(ii) the greatest common divisor of the integers in $I(j)$ if the index set is not empty.

Note that the index set of a state j consists of the set of all integers m such that there is a positive probability of making transitions from state j to state j in m steps.

EXAMPLE 3.11 The Markov chain with transition matrix **P** in Example 3.8 has index sets for states 1, 2, and 3 given by $I(1) = \{3, 6, 9, \ldots\}$, $I(2) = \{3, 6, 9, \ldots\}$, and $I(3) = \{3, 6, 9, \ldots\}$, respectively. It follows that $d(1) = 3$, $d(2) = 3$, and $d(3) = 3$.

The Markov chain with transition matrix **T** in Example 3.8 has index sets for states 1, 2, and 3 given by $I(1) = \{3, 5, 6, \ldots\}$, $I(2) = \{2, 3, 4, \ldots\}$, and $I(3) = \{2, 3, 4, \ldots\}$, respectively. It follows that $d(1) = 1$, $d(2) = 1$, and $d(3) = 1$. Note that for matrix **T** the periods of all states are the same, but the index sets are not identical. ■

EXAMPLE 3.12 Consider the Markov chain with the transition diagram given in Figure 3.14.

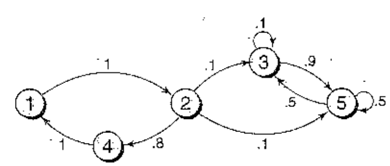


Figure 3.14

The transition matrix for this Markov chain, given in canonical form, is

$$\begin{array}{c} \\ 3 \\ 5 \\ 1 \\ 2 \\ 4 \end{array} \begin{array}{c} \begin{array}{ccccc} 3 & 5 & 1 & 2 & 4 \end{array} \\ \begin{bmatrix} .1 & .9 & 0 & 0 & 0 \\ .5 & .5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ .1 & .1 & 0 & 0 & .8 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} \end{array}$$

where the state numbers are listed to the left of the rows and above the columns. As we noted earlier (in Example 3.10), the order in which states 3 and 5 are listed is uuimportant, and the order in which states 1, 2, and 4 are listed is unimportant. However, states 3 and 5 must be listed before states 1, 2, and 4, and the states must be listed in the same order in the rows as in the columns.

The index sets for states 1, 2, 3, 4, and 5 are given by $I(1) = \{3, 6, 9, \ldots\}$, $I(2) = \{3, 6, 9, \ldots\}$, $I(4) = \{3, 6, 9, \ldots\}$, $I(3) = \{1, 2, 3, \ldots\}$, and $I(5) = \{1, 2, 3, \ldots\}$. It follows that $d(1) = 3$, $d(2) = 3$, $d(4) = 3$, $d(3) = 1$, and $d(5) = 1$. ∎

We note from these examples that in each case, the periods of all states that are mutually accessible are the same. This is a general result.

THEOREM 3.2 If states i and j are mutually accessible, then $d(i) = d(j)$.

Proof. Let m and n be integers such that $p_{ij}(m) > 0$ and $p_{ji}(n) > 0$. Then

$$p_{ii}(m+n) \geq p_{ij}(m)p_{ji}(n) > 0 \qquad \text{and}$$
$$p_{jj}(m+n) \geq p_{ji}(n)p_{ij}(m) > 0$$

and it follows that $m + n \in I(i)$ and $m + n \in I(j)$.

Next, let k be any integer in $I(j)$ and h any divisor of the elements in $I(i)$. Then $p_{ii}(m+n+k) \geq p_{ij}(m)p_{jj}(k)p_{ji}(n) > 0$, and consequently $m+n+k \in I(i)$. Therefore, h divides $m + n + k$. However, h divides $m + n$, so h must divide k. But k was any element in $I(j)$, so h must divide every element of $I(j)$, and because $d(j)$ is the greatest common divisor of elements of $I(j)$, $h \leq d(j)$. Finally, because h was an arbitrary divisor of the elements of $I(i)$, we have $d(i) \leq d(j)$.

A similar argument shows that $d(j) \leq d(i)$, and consequently $d(i) = d(j)$. ∎

Definition 3.6 A Markov chain is a **regular Markov chain** if it is ergodic and the period of each state is 1.

The transition matrix **P** of Example 3.8 is not the transition matrix of a regular Markov chain because the period of each state is 3, but the transition matrix **T** is the transition matrix of a regular Markov chain. Indeed, for the Markov chain with matrix **T**, inspection of the transition diagram in Figure 3.12(b) shows that every two states are mutually accessible. Also, the index set for state 1 includes the integers 3 and 5, and consequently the period of state 1 is 1. As we will show later (in Theorem 3.4), the Markov chain with transition matrix **T** has the property that for every initial state vector $\mathbf{x}_0$, the state vector after m

transitions, $\mathbf{x}(m)$, tends to the state vector $[.2 \quad .4 \quad .4]$. This convergence of $\{\mathbf{x}(m)\}$ is a general property of regular Markov chains, and it is part of the content of Theorem 3.4.

Before turning to the main result of this section, Theorem 3.4, we remark that the definition of a regular Markov chain given in Definition 3.6 is equivalent to a condition on the powers of the transition matrix. This equivalent condition is frequently taken as the definition of a regular Markov chain. We state here a result that says our definition of *regular* implies the condition, and we provide a proof of the result in the Appendix to this chapter. The fact that the condition implies our definitiou of *regular* is the topic of Exercise 6.

THEOREM 3.3 If **P** is the transition matrix of a regular Markov chain, then there is an integer r such that $\mathbf{P}^n$ has only positive entries for all integers $n > r$. ∎

Much of the usefulness of regular Markov chains as models rests on the fact that in the long run, the state vectors tend to a limit: $\lim_{m\to\infty} \mathbf{x}(m)$ exists. The limit is independent of the initial state, it has all positive coordinates, and it can be determined by solving a system of linear equations. These results are the content of Theorem 3.4, and the proofs are provided in the Chapter Appendix.

THEOREM 3.4 Let **P** be the transition matrix of a regular Markov chain. Then

(i) The limit $\lim_{m\to\infty} \mathbf{P}^m$ exists and is a matrix **H** all of whose rows are the same vector **s** (called the steady-state vector for **P**). The coordinates in **s** are all positive.
(ii) The vector **s** is a probability vector that satisfies the equation $\mathbf{s} = \mathbf{sP}$.
(iii) If **x** is any probability vector that satisfies the equation $\mathbf{x} = \mathbf{xP}$, then $\mathbf{x} = \mathbf{s}$. ∎

We note that the convergence of the m-step transition matrices $\mathbf{P}(m)$ to a limit whose rows are all the same means that the m-step state vectors $\mathbf{x}(m)$ converge to a limit and that the limit is the common row of the limit of the transition matrices. The limit of the state vectors is independent of the initial state vector. This result can be interpreted as meaning that as the number of transitions increases, the system "forgets" the initial state. The long-term behavior of the state vector in a regular Markov chain does not depend on the initial state. The same conclusion need not hold for other ergodic Markov chains.

EXAMPLE 3.13 Consider the mouse moving in a maze as described in Examples 3.3 and 3.4, and determine the probability that the mouse will be in the dark compartment in the long run.

When phrased in this way, the problem asks for the coordinate of the limiting state vector (if the limit exists) corresponding to the dark compartment. First, the Markov chain is regular—that is, it is ergodic and each state is of period 1—so the state vectors tend to a limit. Next, the limit vector can be obtained by finding a probability vector **x** that satisfies the system of equations $\mathbf{x} = \mathbf{xP}$, where **P** is the transition matrix of Example 3.3:

$$\mathbf{P} = \begin{bmatrix} .5 & .5 & 0 & 0 \\ .25 & .5 & .25 & 0 \\ 0 & .25 & .5 & .25 \\ 0 & 0 & .5 & .5 \end{bmatrix}.$$

We write the system $\mathbf{x} = \mathbf{xP}$ as $\mathbf{x}(\mathbf{I} - \mathbf{P}) = 0$; if we set $\mathbf{x} = [x_1 \quad x_2 \quad x_3 \quad x_4]$, the system of equations becomes

$$(3.2) \qquad [x_1 \quad x_2 \quad x_3 \quad x_4] \begin{bmatrix} .5 & -.5 & 0 & 0 \\ -.25 & .5 & -.25 & 0 \\ 0 & -.25 & .5 & -.25 \\ 0 & 0 & -.5 & .5 \end{bmatrix} = [0 \quad 0 \quad 0 \quad 0]$$

The system of equations (3.2) has infinitely many solutions. However, we are interested only in solutions that are probability vectors, and adding that requirement, namely

$$(3.3) \qquad x_1 + x_2 + x_3 + x_4 = 1$$

gives a unique solution.

The system of equations that consists of the four equations (3.2) and Equation (3.3) consists of five equations in four variables, and a unique solution is determined by any three of the equations of (3.2) together with Eqation (3.3). Equation (3.3) must be retained, and we can select any three of the four equations in (3.2). We choose the first three equations of (3.2) and Equation (3.3). We have

$$\begin{aligned} x_1 + \quad x_2 + \quad x_3 + \quad x_4 &= 1 \\ .5x_1 - .25x_2 \qquad\qquad &= 0 \\ -.5x_1 + .5x_2 - .25x_3 \qquad &= 0 \\ - .25x_2 + .5x_3 - .25x_4 &= 0 \end{aligned}$$

Solving this system, we find

$$x_1 = \frac{1}{6}, \quad x_2 = \frac{1}{3}, \quad x_3 = \frac{1}{3}, \quad x_4 = \frac{1}{6}$$

Using this information, we can answer the original question. The mouse will be in the dark compartment, compartment 1, about $\frac{1}{6}$ of the time in the long run. ■

Remark The system of equations $\mathbf{x} = \mathbf{xP}$, or $\mathbf{x}(\mathbf{I} - \mathbf{P}) = 0$, has the unknown vector $\mathbf{x}$ on the left and the matrix on the right. This is a departure from the common notation for systems of equations, and it arises from the way we define transition probabilities. It is common for systems of linear equations to be written with the variable $\mathbf{x}$ on the right—that is, in the form $\mathbf{Ax} = \mathbf{b}$, where $\mathbf{A}$ is the coefficient matrix for the system.

Exercises 3.3

1. Transition matrices for ten Markov chains are shown below. In each case, write the transition matrix in canonical form and find the period of each state. If the chain is

regular, find the steady-state vector. Which chains are ergodic but not regular?

a. $\begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & .8 & 0 & 0 & .2 \\ .4 & 0 & 0 & 0 & .6 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$ **b.** $\begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & .4 & 0 & .6 & 0 \\ .2 & .3 & 0 & .5 & 0 \\ 0 & .5 & 0 & .1 & .4 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$ **c.** $\begin{bmatrix} .2 & 0 & .8 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & .8 & 0 & 0 & .2 \\ 0 & 0 & .4 & 0 & .6 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$

d. $\begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ .5 & 0 & 0 & .5 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}$ **e.** $\begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & .4 & 0 & .6 \\ .2 & 0 & 0 & .8 & 0 \end{bmatrix}$ **f.** $\begin{bmatrix} 0 & 0 & .2 & 0 & .8 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$

g. $\begin{bmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & .5 & 0 & .5 & 0 \\ 0 & 0 & .8 & 0 & .2 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix}$ **h.** $\begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & .5 & 0 & 0 & .5 \\ .8 & 0 & 0 & .2 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix}$ **i.** $\begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ .8 & 0 & 0 & .2 & 0 \\ 0 & 0 & 0 & .6 & .4 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & .8 & 0 & 0 & .2 \end{bmatrix}$

j. $\begin{bmatrix} .5 & 0 & .5 & 0 & 0 \\ .7 & 0 & .1 & 0 & .2 \\ .5 & 0 & .1 & .4 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix}$

2. A transition matrix for a Markov chain is shown below. Write this matrix in canonical form.

$$\begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ .6 & .2 & 0 & .2 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & .2 & .5 & .3 & 0 \\ 0 & 0 & 0 & .1 & .6 & 0 & 0 & .3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

3. In the setting described in Exercise 11 of Section 3.2, find the long-range probability of the marmot being in each of the areas.
4. In the setting described in Exercise 13 of Section 3.2, find the long-range probability that the flip is made with the biased coin.
5. Show that if $\mathbf{P}$ is the transition matrix for a Markov chain and $\mathbf{H}$ is a matrix with all rows equal to the same probability vector $\mathbf{w}$, then $\mathbf{PH}$ is a matrix all of whose rows are the vector $\mathbf{w}$.

6. Show that if **P** is the transition matrix of a Markov chain for which there is an integer r such that $\mathbf{P}^r$ has only positive entries, then the Markov chain is ergodic and of period 1. That is, the Markov chain is regular according to Definition 3.6.
7. Let **P** be the transition matrix for a Markov chain, and suppose there is an integer r such that all entries of $\mathbf{P}^r$ are greater than $h > 0$. Show that for all integers $m > r$, the entries of $\mathbf{P}^m$ are also greater than h.
8. Amy is studying the feeding habits of a certain bird. She observes that the bird always comes the first day she makes food available. After that, however, whenever food is available the pattern of feeding is as follows:
   - If the bird feeds one day, then it never feeds the next day.
   - If the bird feeds on day $n-1$ and does not feed on day n, then it feeds on day $n+1$ with probability .75 and does not feed on day $n+1$ with probability .25.
   - If the bird feeds neither on day $n-1$ nor on day n, then it feeds on day $n+1$ with probability .85 and does not feed on day $n+1$ with probability .15.

   (a) Formulate a Markov chain model for this situation.
   (b) In the long run, on what fraction of the days does the bird feed?
9. In the situation of Exercise 8, suppose the first two parts of the feeding pattern remain the same but the third part is replaced by the following:
   - If the bird feeds neither on day $n-1$ nor on day n, then it feeds on day $n+1$ with probability p and does not feed on day $n+1$ with probability $1-p$.

   (a) Formulate a Markov chain for this situation.
   (b) After an extended period of feeding, the probability that the bird feeds on a particular day is a function of p. Find this function and graph it for $0 < p < 1$.
10. Suppose that a mouse moves in the maze shown in Figure 3.7 and that observations are made every 5 minutes and every time the mouse moves from one compartment to another. Assume that the mouse remains where it is with probability .4 and that whenever it has a choice, it is three times as likely to move to a darker compartment as to a lighter one. In the long run, what is the probability that it is in the compartment with low illumination?
11. A computer consultant allocates her time in one-week blocks among two employers and vacations. She is very well paid by employer A, but she dislikes the work. She enjoys working for employer B, but the pay is poor. She always takes a week of vacation when she shifts from one employer to the other, and she never takes more than one week of vacation at a time. If she is on vacation, then she selects an employer at random, and A is selected with probability .6. If this is her first week working for employer A, then she will take a vacation next week with probability .2, and if she has worked for employer A for two weeks (or more), then she will take a vacation next week with probability .5. If this is her first week working for employer B, then she will take a vacation next week with probability .1, and if she has worked for employer B for two weeks (or more), then she will take a vacation next week with probability .3. Suppose she starts by taking a vacation and then beginning to work for employer A.

    (a) Formulate a Markov chain model for this situation and find the transition matrix.
    (b) In the long run, how much time does she spend on vacation?

3.4 Absorbing Chains and Applications to Ergodic Chains

In Section 3.3 we introduced a classification of Markov chains, and we considered the special case of chains (ergodic and regular Markov chains) with the property that any two states are mutually accessible—that is, the states form a single equivalence class. Those chains that have at least one pair of states that are not mutually accessible are more complex, and the behavior of systems modeled with such chains shows a great deal of variety. Rather than conducting a systematic study of these more general chains, we turn to another special case: Markov chains in which one or more of the equivalence classes of states consist of a single state. As we shall see, these chains are also very useful as models.

Definition 3.7 A state i of a Markov chain is an **absorbing state** if $p_{ii} = 1$. A Markov chain is said to be an **absorbing Markov chain** if

1. There is at least one absorbing state, and
2. For each nonabsorbing state j, there is an absorbing state k and a number m of steps such that the probability of making a transition from j to k in m steps is positive—that is, $p_{jk}(m) > 0$.

The definition of an absorbing state means that state i is absorbing if the ith row of the transition matrix has a 1 in the ith column, and then, necessarily all other entries in the ith row are 0. Entries in the ith row and the ith column are said to be on the *main diagonal* of the transition matrix. It is important to note that a state j with a 1 in the jth row of the transition matrix in a position other than on the main diagonal is not an absorbing state.

Condition 2 of Definition 3.7 can be stated as follows: For each nonabsorbing state, it is possible to make a transition to some absorbing state in some number of steps. It is not necessary that each absorbing state be accessible from each nonabsorbing state, only that *some* absorbing state be accessible.

It is conventional to write the canonical form of the transition matrix for an absorbing Markov chain with the absorbing states listed first.

EXAMPLE 3.14 Let **P** and **T** be the transition matrices for Markov chains with five states.

$$\mathbf{P} = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & .8 & 0 & 0 & .2 \\ .4 & 0 & 0 & 0 & .6 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \qquad \mathbf{T} = \begin{bmatrix} .5 & 0 & .3 & 0 & .2 \\ .6 & .2 & 0 & .2 & 0 \\ .3 & 0 & .6 & 0 & .1 \\ 0 & 0 & 0 & 1 & 0 \\ .6 & 0 & .4 & 0 & 0 \end{bmatrix}$$

For the matrix **P**, state 2 is an absorbing state and all other states are nonabsorbing. Note that state 1 is nonabsorbing even though the first row has one entry equal to 1 and all other entries equal to zero; the entry 1 is not on the main diagonal. A similar comment holds for state 5. The absorbing state can be reached in one step from state 3 and in two or more steps from each of the other nonabsorbing states. Therefore, both conditions are satisfied and **P** is the transition matrix of an absorbing Markov chain. A canonical form for the matrix **P**, with states as shown, is

$$\begin{array}{c|c} & \begin{array}{ccccc} 2 & 1 & 3 & 4 & 5 \end{array} \\ \begin{array}{c} 2 \\ 1 \\ 3 \\ 4 \\ 5 \end{array} & \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ .8 & 0 & 0 & 0 & .2 \\ 0 & .4 & 0 & 0 & .6 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \end{array}$$

For the transition matrix **T**, state 4 is an absorbing state and all other states are nonabsorbing. It is possible to reach state 4 in a single step from state 2. However, there is a single absorbing state, and it is not possible to reach this state from state 1, 3, or 5. Consequently, **T** is not the transition matrix of an absorbing Markov chain. A canonical form for the matrix **T**, with states listed as shown, is

$$\begin{array}{c|c} & \begin{array}{ccccc} 4 & 1 & 3 & 5 & 2 \end{array} \\ \begin{array}{c} 4 \\ 1 \\ 3 \\ 5 \\ 2 \end{array} & \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & .5 & .3 & .2 & 0 \\ 0 & .3 & .6 & .1 & 0 \\ 0 & .6 & .4 & 0 & 0 \\ .2 & .6 & 0 & 0 & .2 \end{bmatrix} \end{array}$$

Note that the equivalence class of states $\{1, 3, 5\}$ has the property that once the system enters the class, it never leaves it. ■

Our convention is that the canonical form of the transition matrix of an absorbing Markov chain has states ordered so that absorbing states are listed first. That is, if there are N states and k of them are absorbing, then we suppose that the states have been relabeled so that states $1, 2, \ldots, k$ are absorbing and states $k+1, k+2, \ldots, N$ are nonabsorbing. With this convention, the transition matrix **P** has the form

$$\mathbf{P} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{R} & \mathbf{Q} \end{bmatrix} \tag{3.4}$$

where **I** is a $k \times k$ identity matrix whose row and column labels correspond to absorbing states; **0** is a matrix with all entries equal to zero; **R** is a $(N-k) \times k$ matrix in which the row labels correspond to nonabsorbing states and the column labels correspond to absorbing states; the entries of **R** give the probabilities of direct transitions from nonabsorbing states to absorbing states; and **Q** is an $(N-k) \times (N-k)$ matrix whose entries give the probabilities of transitions between nonabsorbing states. As usual, the states must be listed in the same order in the rows and in the columns. When we refer to a transition matrix of an absorbing Markov chain being written in canonical form, we mean the form (3.4).

One consequence of writing the transition matrix in this form is that the multistep transition matrices have a particularly simple form—for example,

$$\mathbf{P}(2) = \mathbf{P}^2 = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{R}_2 & \mathbf{Q}^2 \end{bmatrix}$$

where $\mathbf{R}_2 = \mathbf{R} + \mathbf{QR}$. In general, for any integer m, $m \geq 2$,

$$\mathbf{P}(m) = \mathbf{P}^m = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{R}_m & \mathbf{Q}^m \end{bmatrix} \tag{3.5}$$

where $\mathbf{R}_m$ can be computed successively as

$$\mathbf{R}_m = \mathbf{R} + \mathbf{Q}\mathbf{R}_{m-1}, \qquad \mathbf{R}_1 = \mathbf{R} \tag{3.6}$$

or as

$$\mathbf{R}_m = \mathbf{R}_{m-1} + \mathbf{Q}^{m-1}\mathbf{R}, \qquad \mathbf{R}_1 = \mathbf{R} \tag{3.7}$$

Both of these equations for $\mathbf{R}_m$ will be useful, as we shall see. The matrix $\mathbf{P}(m)$ contains the m-step transition probabilities, and therefore the entries of $\mathbf{Q}^m$ are the m-step transition probabilities from one nonabsorbing state to another, and the entries of $\mathbf{R}_m$ are the m-step transition probabilities from nonabsorbing states to absorbing states.

We note for emphasis that the state labels in the transition matrix in canonical form may differ from the original state labels. In situations where questions about the original setting are to be answered, care must be used in keeping track of the changes in state labels.

Several properties of the matrices $\mathbf{Q}^m$ and $\mathbf{R}_m$ are useful and will enable us to develop techniques for answering many interesting questions. We begin with an examination of the matrix **Q**.

Suppose we have an absorbing Markov chain with a transition matrix **P** in canonical form, and suppose the system begins in the nonabsorbing state i. If j is another nonabsorbing state, $i \neq j$, then the probability that the system is in state j on the first (subsequent) observation is q_{ij}. The probability that it is in state j on the second observation is $q_{ij}(2)$, and so on. Let $E(i, j; m)$ be the expected number of times the system is in state j given that it started in state i and continued for m transitions. We develop an expression for $E(i, j; m)$.

For fixed states i and j, define a random variable X_k by

$$X_k = \begin{cases} 1 & \text{if the system began in state } i \text{ and is in state } j \text{ after the } k\text{th transition} \\ 0 & \text{if the system began in state } i \text{ and is not in state } j \text{ after the } k\text{th transition} \end{cases}$$

It follows from the definition of the expected value of a random variable that $E[X_k] = q_{ij}(k)$, for $k = 1, 2, \ldots, m$, and

$$\begin{aligned} E(i, j; m) &= E[X_1] + E[X_2] + \cdots + E[X_m] \\ &= q_{ij}(1) + q_{ij}(2) + \cdots + q_{ij}(m) \end{aligned} \tag{3.8}$$

If states i and j are the same, then the expression has an additional term as a result of the fact that the system began in state j:

$$E(j, j; m) = 1 + q_{jj}(1) + q_{jj}(2) + \cdots + q_{jj}(m) \tag{3.9}$$

Because i and j could be any nonabsorbing states, we have shown that the ij-entry in the matrix

$$\mathbf{I} + \mathbf{Q} + \mathbf{Q}^2 + \cdots + \mathbf{Q}^m \tag{3.10}$$

is the expected number of times the system is in state j given that it started in state i and made m transitions. Note that the matrix $\mathbf{I}$ in (3.10) has the same dimensions as $\mathbf{Q}$.

If the system begins in a nonabsorbing state, there is a positive probability that it will reach some absorbing state, and after doing so, the system remains there for all subsequent observations. In fact, as the number of observations increases, the probability of finding the system in a nonabsorbing state becomes arbitrarily small, and the probability that the system is in an absorbing state approaches 1. Indeed, the probability that the system is in a nonabsorbing state after m transitions becomes small sufficiently fast (as m increases) that the series

$$\mathbf{I}+\mathbf{Q}+\mathbf{Q}^2+\mathbf{Q}^3+\cdots \tag{3.11}$$

converges. The convergence of the series (3.11), together with the meanings of the partial sums (3.10) as given in (3.8) and (3.9), gives a highly useful result.

THEOREM 3.5 If the transition matrix of an absorbing Markov chain is written in canonical form as

$$\mathbf{P}=\begin{bmatrix}\mathbf{I} & \mathbf{0}\\ \mathbf{R} & \mathbf{Q}\end{bmatrix}$$

then the matrix $\mathbf{I}-\mathbf{Q}$ has an inverse, and the series $\mathbf{I}+\mathbf{Q}+\mathbf{Q}^2+\mathbf{Q}^3+\cdots$ converges to the inverse of $\mathbf{I}-\mathbf{Q}$.

The matrix $\mathbf{N}=(\mathbf{I}-\mathbf{Q})^{-1}$ is called the **fundamental matrix** of the Markov chain, and the ij-entry in $\mathbf{N}$ is the expected number of visits to nonabsorbing state j given that the system began in nonabsorbing state i and continued until an absorbing state was reached. The sum of the entries in the ith row of $\mathbf{N}$ is the expected number of transitions before an absorbing state is reached. The state labels are those of the transition matrix in canonical form. ∎

The details of arguments justifying this theorem are included in the Chapter Appendix.

EXAMPLE 3.15 An absorbing Markov chain has the transition matrix

$$\begin{bmatrix}0 & 1 & 0 & 0 & 0\\ .2 & 0 & .1 & .1 & .6\\ 0 & 0 & 1 & 0 & 0\\ 0 & 0 & 0 & 1 & 0\\ .2 & .2 & .3 & .1 & .2\end{bmatrix}$$

(a) If the system begins in state 2, find the expected number of visits to state 5 before an absorbing state is reached.
(b) If the system begins in state 2, find the expected number of transitions before an absorbing state is reached.

We begin by writing the transition matrix in canonical form.

$$\begin{array}{c c}
 & \begin{array}{ccccc}3 & 4 & 1 & 2 & 5\end{array}\\
\begin{array}{c}3\\4\\1\\2\\5\end{array} & \begin{bmatrix}1 & 0 & 0 & 0 & 0\\ 0 & 1 & 0 & 0 & 0\\ 0 & 0 & 0 & 1 & 0\\ .1 & .1 & .2 & 0 & .6\\ .3 & .1 & .2 & .2 & .2\end{bmatrix}
\end{array}$$

Because the question is posed in terms of the original state labels, we retain those labels above and to the left of the new matrix. The matrices $\mathbf{Q}$, $\mathbf{I}-\mathbf{Q}$, and $\mathbf{N}$ are

$$\mathbf{Q}=\begin{bmatrix}0 & 1 & 0\\ .2 & 0 & .6\\ .2 & .2 & .2\end{bmatrix}\quad \mathbf{I}-\mathbf{Q}=\begin{bmatrix}1 & -1 & 0\\ -.2 & 1 & -.6\\ -.2 & -.2 & .8\end{bmatrix}\quad \mathbf{N}=\begin{bmatrix}1.7 & 2 & 1.5\\ .7 & 2 & 1.5\\ .6 & 1 & 2\end{bmatrix}$$

Recall that the state labels for the first, second, and third rows of $\mathbf{Q}$, $\mathbf{I}-\mathbf{Q}$, and $\mathbf{N}$, are 1, 2, and 5, respectively.

Using the fundamental matrix $\mathbf{N}$, we can answer the questions. The answer to question (a) is the entry in the second row and third column of $\mathbf{N}$: 1.5. If the system begins in state 2, then the expected number of visits to state 5 before an absorbing state is reached is 1.5.

The answer to question (b) is the sum of the entries in the second row of $\mathbf{N}$: $.7+2+1.5=4.2$. If the system begins in state 2, then the expected number of transitions before an absorbing state is reached is 4.2. Of course, it may reacb an absorbing state in one transition: There is a positive probability that it moves directly from state 2 to state 3 (or 4). However, it may also take more than one transition, and we now know that the expected number is 4.2. ∎

EXAMPLE 3.16 Consider the small-group decision-making situation described in Section 3.1, and suppose we have a group of four individuals and three alternatives. Define a vote change as a "shift toward consensus" if it results in one of the group composition shifts: $310 \to 400$, $220 \to 310$, $211 \to 310$. Assume that an individual who can effect a shift toward consensus is twice as likely to change a vote as one who cannot. Also assume that the probability of changing a vote to another alternative is proportional to the number of individuals voting for that alternative. Suppose the group initially has group composition 211. Find the expected number of vote changes before consensus is reached.

Denote the group compositions 400, 310, 220, 211 as states 1, 2, 3, and 4, respectively. Then a Markov chain model for this situation under these assumptions has transition matrix (see Exercise 6)

$$\begin{bmatrix}1 & 0 & 0 & 0\\ \frac{2}{5} & 0 & \frac{3}{5} & 0\\ 0 & 1 & 0 & 0\\ 0 & \frac{4}{9} & \frac{2}{9} & \frac{3}{9}\end{bmatrix}$$

This is a transition matrix for an absorbing chain, and the matrix is already written in canonical form. The matrix $\mathbf{Q}$ and the fundamental matrix $\mathbf{N}$ are

$$\mathbf{Q} = \begin{bmatrix} 0 & \frac{3}{5} & 0 \\ 1 & 0 & 0 \\ \frac{4}{9} & \frac{2}{9} & \frac{3}{9} \end{bmatrix} \qquad \mathbf{N} = \begin{bmatrix} \frac{5}{2} & \frac{3}{2} & 0 \\ \frac{5}{2} & \frac{5}{2} & 0 \\ \frac{15}{6} & \frac{11}{6} & \frac{9}{6} \end{bmatrix}$$

The third row of $\mathbf{N}$ is associated with state 4, group composition 211. Therefore, the expected number of transitions before reaching state 1 is $\frac{15}{6} + \frac{11}{6} + \frac{9}{6} = \frac{35}{6}$. ■

The matrix $\mathbf{N}$ can also be used to determine the probabilities of absorption in the various absorbing states. To determine how, we return to Equations (3.6) and (3.7),

$$\mathbf{R}_m = \mathbf{R}_{m-1} + \mathbf{Q}^{m-1}\mathbf{R} \qquad \text{and} \qquad \mathbf{R}_m = \mathbf{R} + \mathbf{Q}\mathbf{R}_{m-1},\ \mathbf{R}_1 = \mathbf{R}$$

Suppose that the system begins in the ith nonabsorbing state. The probability that it is in the jth absorbing state after the first transition is r_{ij}. The probability that it is in the jth absorbing state after the second transition is the ij-entry of the matrix $\mathbf{R}_2$, and in general, the probability that it is in the jth absorbing state after the mth transition is the ij-entry of the matrix $\mathbf{R}_m$.

Next, from the expression $\mathbf{R}_m = \mathbf{R}_{m-1} + \mathbf{Q}^{m-1}\mathbf{R}$ we see that $\mathbf{R}_m \geq \mathbf{R}_{m-1}$, where the symbol $\geq$ means entrywise ineqnality. That is, each of the seqnences of numbers obtained by fixing i and j and taking the ij-entry in the matrix $\mathbf{R}_m$ as the mth entry in the sequence is a monotone nondecreasing sequence of numbers. Moreover, each of these numbers must be less than or equal to 1 because the rows of the m-step transition matrix are probability vectors. Therefore, each of the sequences of numbers converges, and so the sequence of matrices $\{\mathbf{R}_m\}$ converges. Define the matrix $\mathbf{A}$ to be the limit of the sequence $\{\mathbf{R}_m\}$:

$$\mathbf{A} = \lim_{m\to\infty} \mathbf{R}_m$$

It follows from the discussion of the meaning of the entries in $\mathbf{R}_m$ that the ij-entry of the matrix $\mathbf{A}$ has the following interpretation:

> If an absorbing Markov chain is initially in state i, then the probability that it is absorbed in nonabsorbing state j is the ij-entry of the matrix $\mathbf{A}$.

Here as elsewhere in the discussion, it is important to remember that the references to states i and j refer to the states of the matrix in canonical form, and references to the original state labels must be translated into the new state labels.

The definition of the matrix $\mathbf{A}$ given above is as a limit—not particularly well suited for computation—and it is useful to have an alternative means of determining $\mathbf{A}$. There is an expression for the matrix $\mathbf{A}$ that involves only the matrices $\mathbf{R}$ and $\mathbf{N}$. To determine the expression, we recall that $\mathbf{R}_m = \mathbf{R} + \mathbf{Q}\mathbf{R}_{m-1}$, and if we take the limit of both sides as $m \to \infty$, we have

$$\mathbf{A} = \mathbf{R} + \mathbf{Q}\mathbf{A}$$

From this we have

$$\mathbf{A} - \mathbf{Q}\mathbf{A} = (\mathbf{I} - \mathbf{Q})\mathbf{A} = \mathbf{R}$$

Finally, because the inverse of $\mathbf{I} - \mathbf{Q}$ exists and is equal to $\mathbf{N}$, if we multiply both sides of the expression $(\mathbf{I} - \mathbf{Q})\mathbf{A} = \mathbf{R}$ on the left by $\mathbf{N}$, we have $\mathbf{A} = \mathbf{N}\mathbf{R}$. We summarize this result as a theorem.

THEOREM 3.6 Suppose that the transition matrix of an absorbing Markov chain is written in canonical form as

$$\mathbf{P} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{R} & \mathbf{Q} \end{bmatrix}$$

and that $\mathbf{N}$ is the fundamental matrix for $\mathbf{P}$. Then the entry in the ith row and jth column of the matrix $\mathbf{A} = \mathbf{N}\mathbf{R}$ is the probability that the system is absorbed in the jth absorbing state given that it began in the ith nonabsorbing state. ■

EXAMPLE 3.17 An absorbing Markov chain has the following transition matrix.

$$\begin{bmatrix} 0 & 0 & .5 & 0 & .5 \\ 0 & 1 & 0 & 0 & 0 \\ .2 & .1 & .2 & .1 & .4 \\ 0 & 0 & 0 & 1 & 0 \\ .1 & .1 & 0 & .6 & .2 \end{bmatrix}$$

If the system is initially in state 3, find the probability that it is absorbed in state 4.

We rewrite the transition matrix with the states listed in the order 2, 4, 1, 3, 5. Then the matrices $\mathbf{Q}$, $\mathbf{R}$, and $\mathbf{N}$ are

$$\mathbf{Q} = \begin{bmatrix} 0 & .5 & .5 \\ .2 & .2 & .4 \\ .1 & 0 & .2 \end{bmatrix} \qquad \mathbf{R} = \begin{bmatrix} 0 & 0 \\ .1 & .1 \\ .1 & .6 \end{bmatrix} \qquad \mathbf{N} = \begin{bmatrix} 1.28 & .80 & 1.20 \\ .40 & 1.50 & 1.00 \\ .16 & .10 & 1.40 \end{bmatrix}$$

Therefore, the matrix $\mathbf{A} = \mathbf{N}\mathbf{R}$ is

$$\mathbf{A} = \begin{bmatrix} .20 & .80 \\ .25 & .75 \\ .15 & .85 \end{bmatrix}$$

To complete the example, we need to identify the entry of $\mathbf{A}$ that gives the probability of absorption in state 4 given a start in state 3. The states have been relabeled in the order 2, 4, 1, 3, 5. Therefore, state 3 corresponds to the second row of the matrix $\mathbf{A}$, and state 4 corresponds to the second column. It follows that the desired probability is .75. ■

Applications of Absorbing Chains to Ergodic Chains

One of the common uses of the ideas and techniques introduced here for absorbing chains is to determine useful information about ergodic chains. In many situations we can do so by constructing an absorbing Markov chain that is based on the ergodic chain and the information desired.

Consider an ergodic chain with transition matrix $\mathbf{P}$. By the definition of an ergodic chain, for any states i and j, there is an integer m such that the probability of making a transition from state i to state j in m steps is positive. What is the expected number of transitions to first reach state j given a start in state i? To answer the question, we construct an absorbing chain with the same states, with transition matrix $\mathbf{P}'$ that has the same rows as $\mathbf{P}$ with the exception of the jth row. The jth row of $\mathbf{P}'$ has a 1 on the main diagonal and zeros in all other entries. That is, we have replaced state j by an absorbing state. Transition probabilities among all other states are as in the original chain, and, in particular, transition probabilities into the new state j are the same as those into the old state j. The absorbing chain behaves as follows:

- If the system begins in state j, then it remains there
- If the system begins in state $i, i \neq j$, then it proceeds just as in the original ergodic system until it reaches state j for the first time. Once in state j, it remains there.

Because the original system was ergodic, it is possible to reach state j from every other state, and the new process satisfies the conditions for an absorbing Markov chain.

Next, write the transition matrix for the new chain in canonical form and find the fundamental matrix $\mathbf{N}$. By Theorem 3.5, the entries in $\mathbf{N}$ give the expected number of times the system is in each nonabsorbing state prior to reaching the absorbing state (there is a single absorbing state in this case). Interpreting this in terms of the original process, we see that these numbers give the expected number of times the system is in each state $i, i \neq j$, before it first reaches state j.

EXAMPLE 3.18 A regular Markov chain has the transition matrix

$$\mathbf{P} = \begin{bmatrix} .25 & .25 & .5 & 0 \\ 0 & .25 & .5 & .25 \\ .25 & .25 & .5 & 0 \\ .25 & 0 & .5 & .25 \end{bmatrix}$$

If the system is initially in state 2, find (a) the expected number of visits to state 3 before it first reaches state 4, and (b) the expected number of transitions before it first reaches state 4.

Because this is an ergodic Markov chain (every regular chain is ergodic), we can answer the question by constructing an absorbing chain. We are interested in what happens before the system first reaches state 4, and we construct an absorbing chain by replacing state 4 with an absorbing state. The transition matrix for the new absorbing chain is

$$\begin{bmatrix} .25 & .25 & .5 & 0 \\ 0 & .25 & .5 & .25 \\ .25 & .25 & .5 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Writing this matrix in canonical form yields

$$\begin{array}{c} \\ 4 \\ 2 \\ 1 \\ 3 \end{array} \begin{array}{c} \begin{array}{cccc} 4 & 2 & 1 & 3 \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 0 \\ .25 & .25 & 0 & .5 \\ 0 & .25 & .25 & .5 \\ 0 & .25 & .25 & .5 \end{bmatrix} \end{array}$$

where the original state labels are as shown. The matrix $\mathbf{Q}$ and the fundamental matrix $\mathbf{N}$ are

$$\mathbf{Q} = \begin{bmatrix} .25 & 0 & .5 \\ .25 & .25 & .5 \\ .25 & .25 & .5 \end{bmatrix} \qquad \mathbf{N} = \begin{bmatrix} 4 & 2 & 6 \\ 4 & 4 & 8 \\ 4 & 3 & 9 \end{bmatrix}$$

The rows and columns of the matrix $\mathbf{N}$ correspond to states 2, 1, 3, in that order. Consequently, we conclude that if the system begins in state 2, then the expected number of visits to state 3 before it reaches an absorbing state is 6. Thus, in the original setting the expected number of visits to state 3 before first reaching state 4 is 6.

Also, if the absorbing chain begins in state 2, then the expected number of transitions before it is absorbed is $4 + 2 + 6 = 12$. Consequently, in the original setting the expected number of transitions before the system first reaches state 4 is 12. ■

EXAMPLE 3.19 Consider the situation described in Example 3.3 of Section 3.2, in which a mouse moves in a maze with compartments illuminated at different levels. Assume that half the time the mouse is in the same compartment on successive observations, and half the time it moves from its starting compartment to an adjacent compartment between observations. If it moves, then it is equally likely to move to any compartment open to it. If the mouse begins in compartment 1 (see Figure 3.7), find the expected number of transitions before it first reaches compartment 4.

The transition matrix for this system, determined in Example 3.4 of Section 3.2, is

$$\mathbf{P} = \begin{bmatrix} .5 & .5 & 0 & 0 \\ .25 & .5 & .25 & 0 \\ 0 & .25 & .5 & .25 \\ 0 & 0 & .5 & .5 \end{bmatrix}$$

Because the task is to find the number of transitions before the system reaches state 4, we construct an absorbing Markov chain with state 4 replaced by an absorbing state. The transition matrix for the new chain is

$$\mathbf{P}' = \begin{bmatrix} .5 & .5 & 0 & 0 \\ .25 & .5 & .25 & 0 \\ 0 & .25 & .5 & .25 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

We write the new transition matrix in canonical form, and we label the rows and columns with the original state labels to retain that information. We have

$$\mathbf{P} = \begin{array}{c} \\ 4 \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{cccc} 4 & 1 & 2 & 3 \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & .5 & .5 & 0 \\ 0 & .25 & .5 & .25 \\ .25 & 0 & .25 & .5 \end{bmatrix} \end{array}$$

The fundamental matrix for this chain is

$$\mathbf{N} = \begin{bmatrix} 3 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 3 \end{bmatrix}$$

The system began in state 1, and state 1 corresponds to the first row of this fundamental matrix. The sum of the entries in the first row is 6, and consequently, the expected number of transitions before the system first reaches state 4, given that it begins in state 1, is 6. ■

In Examples 3.18 and 3.19 we have used the technique of creating a new Markov chain with an absorbing state to determine the expected number of transitions before the system first reaches a specified state. The same technique can be used to find the probability that the system visits state i before state j. For example, given an ergodic Markov chain and a specified starting state (different from states i and j), we can find the probability that state i is visited before state j by creating a new Markov chain with states i and j absorbing—as in Examples 3.18 and 3.19—and then using the associated matrix $\mathbf{A}$ as in Example 3.17.

Exercises 3.4

1. Transition matrices for six Markov chains are shown below. In each case determine whether the Markov chain is absorbing, and write the transition matrix in canonical form. If the chain is absorbing, find the fundamental matrix $\mathbf{N}$.

a. $\begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & .5 & 0 & .5 \\ 0 & 0 & 1 & 0 & 0 \\ .8 & 0 & .2 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix}$ **b.** $\begin{bmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & .1 & .3 & .1 & .5 \\ 0 & 0 & 1 & 0 & 0 \\ .4 & 0 & 0 & .2 & .4 \\ 0 & 0 & 0 & .8 & .2 \end{bmatrix}$

c. $\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & .5 & .5 \\ 0 & 0 & 0 & 0 & 1 \\ .5 & .5 & 0 & 0 & 0 \\ 0 & .5 & .5 & 0 & 0 \end{bmatrix}$ **d.** $\begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & .4 & .6 \\ 0 & 0 & 0 & 1 & 0 \\ .5 & 0 & .5 & 0 & 0 \end{bmatrix}$

e. $\begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & .4 & 0 & .6 & 0 \\ .1 & 0 & .1 & 0 & .8 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & .5 & 0 & 0 & .5 \end{bmatrix}$ **f.** $\begin{bmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ .2 & 0 & 0 & 0 & .8 \end{bmatrix}$

2. The transition diagram for a Markov chain is shown in Figure 3.15.
   (a) Find the transition matrix and write it in canonical form.
   (b) If the system is initially in state 2, find the expected number of transitions before it first reaches an absorbing state.

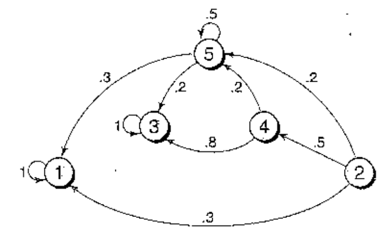


Figure 3.15

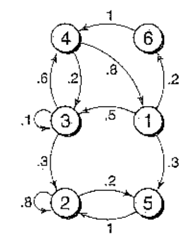


Figure 3.16

3. A Markov chain has the transition diagram shown in Figure 3.16. Describe the behavior of the Markov chain in as much detail as you can, including (to the extent possible) information on the long-run behavior of the state vector. To what extent does the long-run behavior depend on the initial state vector?

4. A transition matrix for a Markov chain is shown below.

$$\begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ .5 & .2 & 0 & .3 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & .3 & .3 & .4 & 0 \\ 0 & 0 & 0 & .2 & .5 & 0 & 0 & .3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

   (a) Write this matrix in canonical form.
   (b) Find the fundamental matrix for your canonical form.
   (c) If the process is initially in state 2, find the expected number of transitions before it first reaches an absorbing state.
   (d) If the process begins in state 8, find the probability that it is absorbed in state 4.
   (e) If the process begins in state 8, find the probability that it is in state 3 before it is absorbed.
   (f) If the process is initially in state 3, find the expected number of visits to state 1 before it is absorbed.

5. A transition matrix for a Markov chain is shown below.

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

(a) Write this matrix in canonical form.
(b) Find the fundamental matrix for your canonical form.
(c) If the process is initially in state 5, find the expected number of transitions before it is absorbed.
(d) If the process begins in state 2, find the probability that it is absorbed in state 9.
(e) If the process begins in state 2, find the probability that it is in state 8 before it is absorbed.
(f) If the process is initially in state 1, find the expected number of visits to state 5 before it is absorbed.

6. Consider the small-group decision-making situation described in Section 3.1, and suppose we have a group of four individuals and three alternatives. Define a vote change as a "shift toward consensus" if it results in one of the group composition shifts: $310 \to 400$, $220 \to 310$, $211 \to 310$. Assume that an individual who can effect a shift toward consensus is twice as likely to change a vote than one who cannot. Also assume that the probability of changing a vote to another alternative is proportional to the number of individuals voting for that alternative. Suppose that the group initially has group composition 211. Formulate a Markov chain model for this situation, find the transition matrix, and find the fundamental matrix.

7. Consider a small-group decision-making situation similar to that described in Section 3.1 with six individuals and three alternatives. Formulate a Markov chain model under the following assumption: An individual who is the only person voting for an alternative is twice as likely to change her vote as a person who is one of a group of two or more voting for an alternative. If an individual changes her vote, then the probability of changing to an alternative is proportional to the number of individuals voting for that alternative.

(a) If the group is initially divided 3, 2, and 1, find the expected number of vote changes before consensus is reached.
(b) If the group is initially divided 2, 2, and 2, find the probability that at some point it is divided 3, 3, and 0 before consensus is reached.

8. From one academic year to another, each student at Gigantic State University moves on to the next class, flunks out, or remains in the same class with probabilities p, q, and r, respectively. A student is said to be in state 1 if graduated, 2 if flunked out, 3 if a senior, 4 if a junior, 5 if a sophomore, and 6 if a freshman. Formulate a Markov chain model for this situation, and then answer the following:

(a) Find the transition matrix.
(b) If $p = .7$, $q = .2$, and $r = .1$, find the fundamental matrix.
(c) With p, q, and r as in (b), how long does a student need to be in college before there is a better than even chance of his or her graduating?

9. A marmot lives in the region shown in Figure 3.6. Suppose the marmot is observed every hour and each time it moves from one area to another. Also, suppose that it will be observed in the same location on successive observations with probability .2. If it moves, the probability of its moving to an area is proportional to the number of resources available to it in that area in comparison to the number of resources available to it in the adjoining areas. The areas bordering the pond have water in addition to the other resources specified, and the marmot is never actually in the pond.

(a) Formulate a Markov chain model for this situation.
(b) If the marmot begins at its den in the rock pile, find the expected number of observations before it first reaches the marsh.
(c) If the marmot begins at its den in the rock pile, find the probability that it reaches the forest without going through the marsh.

10. Joe and Jess play a game as follows: An unfair coin with $\Pr[H] = .6$ is flipped and the result is noted. If it comes up heads, then Jess pays Joe one dollar, and if it comes up tails, then Joe pays Jess one dollar. If each player has money, then the coin is flipped again. The game ends as soon as one player has all the money. When the game begins, Joe has one dollar and Jess has three dollars. (See also Exercise 15 of Section 3.2.)

(a) Find the expected number of plays before the game ends.
(b) Find the probability that Jess wins the game.

11. There are six balls, three red and three green, distributed between boxes labeled 1 and 2, with three balls in each box. A game is played as follows: A ball is selected at random from each box. The ball selected from box 1 is placed in box 2, the ball selected from box 2 is placed in box 1, and the colors of the balls in each box are noted. Then two more balls are selected, and play continues. When the game begins there are two green balls and one red ball in box 2. Find the probability that all balls in box 1 are green before all balls in box 2 are green. (See also Exercise 14 of Section 3.2.)

12. An unfair coin with $\Pr[H] = .4$ is flipped repeatedly, and the result of each flip is noted. Find the expected number of flips before there are three consecutive heads for the first time.

13. Each morning Reba decides how to get from her apartment to the university. She can walk, bike, take the bus, or ride with a friend who also has an 8:00 a.m. class. It is too much work to plan in advance, so each morning she makes a random choice subject to the following conditions:

- If she walks on one day, then the next day she walks with probability .4; all other choices are equally likely.
- If she rides the bus one day, then the next day she is twice as likely to ride the bus as not; all other choices are equally likely.

- If she rides with a friend one day, then she does not ride with a friend next day; all other choices are equally likely.
- If she rides her bike one day, then she always walks the next day.

If she rides with a friend on Monday, find the expected number of days before she first rides the bus.

Chapter Appendix: Mathematical Details

The primary focus of this book is on model building. However, there are times when a slightly deeper look into the mathematical aspects of the models yields interesting and useful information. Thus, in this Appendix we consider in greater detail the following topics that were introduced but not explored in Chapter 3.

1. The relation between Markov chains that are ergodic and of period 1 and Markov chains with a transition matrix $\mathbf{P}$ for which some power $\mathbf{P}^r$ has only positive entries.
2. The long-run behavior of the powers of the transition matrix of a regular Markov chain. This result can be used to give information on the long-run behavior of the state vector for a regular Markov chain.
3. The behavior of the powers $\mathbf{Q}^m$, where $\mathbf{Q}$ is the matrix of transition probabilities among the nonabsorbing states in an absorbing Markov chain, and the invertibility of the matrix $(\mathbf{I}-\mathbf{Q})$.

It is the details of the second and third topics that are most likely to be useful, and we begin with them. (The details of the proof of Theorem 3.3 are included primarily for completeness, although they do provide some insights into the structure of Markov chains with period 1.)

Proof of Theorem 3.4 The hypothesis of the theorem is that we have a Markov chain with a transition matrix $\mathbf{P}$ and an integer r such that $\mathbf{P}^r$ has only positive entries. We show that there is a probability vector $\mathbf{s}$ such that

$$\lim_{r\to\infty} \mathbf{P}^r = \begin{bmatrix} \mathbf{s} \\ \mathbf{s} \\ \vdots \\ \mathbf{s} \\ \mathbf{s} \end{bmatrix}. \tag{3.12}$$

To this end, let r be an integer such that all entries in $\mathbf{P}^r$ are positive—say all are greater than $h > 0$. It follows that all entries of $\mathbf{P}^m$ are greater than h for all $m \geq r$ (Exercise 7 of Section 3.3). Consider the first column of $\mathbf{P}^m$ and denote it by $\mathbf{p}^1(m)$. Set

$$s(m) = \min\{p_{i1}(m),\ i = 1, 2, \ldots, N\} \quad \text{and}$$
$$t(m) = \max\{p_{i1}(m),\ i = 1, 2, \ldots, N\}$$

Thus $s(m)$ is the smallest probability that the system is in state 1 on the mth observation given a start in state $1, 2, \ldots, N$, and $t(m)$ is the largest such probability. The following is a useful fact about Markov chains, and it will be helpful in verifying our result.

> The sequences $\{s(m)\}$ and $\{t(m)\}$ are monotone increasing and monotone decreasing, respectively. That is, $s(m+1) \geq s(m)$ and $t(m+1) \leq t(m)$ for $m = 1, 2, 3, \ldots$.

To confirm this monotonicity, we denote by $\mathbf{p}_i$ the ith row of the matrix $\mathbf{P}$, and we note that

$$\begin{aligned} s(m+1) &= \min\{p_{i1}(m+1),\ i = 1, 2, \ldots, N\} \\ &= \min\{\mathbf{p}_i \cdot \mathbf{p}^1(m),\ i = 1, 2, \ldots, N\} \end{aligned}$$

Also, since $\mathbf{p}_i$ is a probability vector,

$$\mathbf{p}_i \cdot \mathbf{p}^1(m) = \sum_{j=1}^{N} p_{ij}p_{j1}(m) \geq s(m)\sum_{j=1}^{N} p_{ij} = s(m)$$

and consequently the minimum of these inner products—which is $s(m+1)$—must also be greater than or equal to $s(m)$. The proof that $t(m+1) \leq t(m)$ is similar.

Returning to the proof of the Theorem, we have $s(m) \leq t(m) \leq t(1)$ for all m, and consequently $\{s(m)\}$ is an increasing sequence of real numbers that is bounded above. By a fundamental property of real numbers, the limit $\lim_{m\to\infty} s(m)$ exists, and we denote the limit by s. Also, the sequence $\{t(m)\}$ is monotone decreasing and $t(m) \geq s(m) \geq s(1)$, and consequently $\lim_{m\to\infty} t(m) = t$ exists. Because $s(m) \leq t(m)$ for all m, we conclude that $s \leq t$. If $s = t$, then we have shown that the entries in the first column of $\{\mathbf{P}^r\}$ all converge to a common limit as m increases. We show that the remaining possibility, $s < t$, is impossible. In this case, the sequences $\{s(m)\}$ and $\{t(m)\}$ and the limits s and t are as shown in Figure 3.17.

Now set $d = t - s$, and note that we are assuming $d > 0$. We complete the proof by showing that $s(m)$ increases by more than a fixed amount—an amount depending on h and d—in each sequence of r transitions. In fact, we show that if

1. $p_{i1}(m) \geq h$ for $1 \leq i \leq N$ and $m \geq r$,
2. $s(m) \leq s < t \leq t(m)$ for all m, and
3. $t - s = d$,

then

$$s(m+r) \geq s(m) + hd \qquad \text{for } m = 1, 2, 3 \ldots$$

To verify this, we let k be an index for which $s(m+r) = p_{k1}(m+r)$. We let $\mathbf{u}$ be an N-vector all of whose coordinates are 1. Then for $m = 1, 2, \ldots,$ we have

$$\begin{aligned} s(m+r) &= \mathbf{p}_k(r) \cdot \mathbf{p}^1(m) \\ &= \mathbf{p}_k(r) \cdot [\mathbf{p}^1(m) - s(m)\mathbf{u}] + s(m)(\mathbf{p}_k(r) \cdot \mathbf{u}) \\ &\geq s(m) + \max\{p_{kj}(r)[p_{j1}(m) - s(m)],\ j = 1, 2, \ldots\} \\ &\geq s(m) + hd \end{aligned}$$

The second inequality results from the fact that at least one of the terms $p_{j1}(m) - s(m)$ is as large as d and all of the $p_{kj}(r)$ are at least as large as h. To see that $s < t$ is

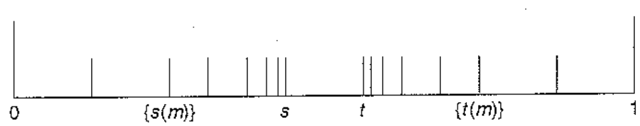


Figure 3.17

impossible, we argue as follows: Because $s(m) \to s$ as $m \to \infty$, there is an integer m_0 such that $s(m) > s - hd/2$ for all $m > m_0$. However, the above argument shows that $s(m+r) \geq s(m) + hd \geq s + hd/2$ for all $m > m_0$. But this is impossible by the meaning of s.

The proof that all entries in the first column of $\{\mathbf{P}^r\}$ tend to the same limit is complete. Also, because $s(r) > 0$, that limit must be positive. The same proof can be applied to entries in the second, ..., Nth columns of $\{\mathbf{P}^r\}$. The limit for each column will, in general, depend on the column index. The verification of Equation (3.12) is complete, and this is part (i) of Theorem 3.4.

For part (ii) of Theorem 3.4, we let $\mathbf{H}$ denote the matrix $\lim_{m\to\infty} \mathbf{P}^m$, a matrix all of whose rows are the same vector $\mathbf{s}$. We write $\mathbf{P}^{m+1} = \mathbf{P}^m\mathbf{P}$. Next, let m tend to infinity in this expression, and use Equation (3.12). Then $\mathbf{H} = \mathbf{HP}$, and the first row of each side of this expression gives $\mathbf{s} = \mathbf{sP}$.

For part (iii) of Theorem 3.4, let $\mathbf{x}$ be a probability vector that satisfies $\mathbf{x} = \mathbf{xP}$. If we multiply each side of this equation on the right by the matrix $\mathbf{P}$, we have $\mathbf{xP} = \mathbf{xPP} = \mathbf{xP}^2$, and because $\mathbf{x} = \mathbf{xP}$, we have $\mathbf{x} = \mathbf{xP}^2$. Continuing in this way, we have $\mathbf{x} = \mathbf{xP}^m$ for $m = 2, 3, 4, \ldots$. Letting m tend to infinity and using Equation (3.12), we have

$$\mathbf{x} = \mathbf{x}\begin{bmatrix} \mathbf{s} \\ \mathbf{s} \\ \vdots \\ \mathbf{s} \\ \mathbf{s} \end{bmatrix}$$

Finally, because $\mathbf{x}$ is a probability vector, the sum of the coordinates in $\mathbf{x}$ is 1, and consequently the right-hand side of the expression just above is $\mathbf{s}$. ∎

Proof of Theorem 3.5 We show first that if $\mathbf{P}$ is the transition matrix of an absorbing Markov chain written in canonical form,

$$\mathbf{P} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{R} & \mathbf{Q} \end{bmatrix}$$

then the matrix $\mathbf{I} - \mathbf{Q}$ is invertible. We do so by showing that the infinite series in expression (3.11)

$$\mathbf{I} + \mathbf{Q} + \mathbf{Q}^2 + \mathbf{Q}^3 + \cdots$$

converges and that the matrix $\mathbf{N}$, defined as the sum of the series, is the inverse of $\mathbf{I} - \mathbf{Q}$.

First we show that $\lim_{m\to\infty} \mathbf{Q}^m = \mathbf{0}$. Begin by noting that the multistep transition matrices have a particularly simple form. Indeed, as shown in Equation (3.5), we have

$$\mathbf{P}^m = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{R}_m & \mathbf{Q}^m \end{bmatrix}$$

where the entries in $\mathbf{R}_m$ are the multistep transition probabilities from nonabsorbing states to absorbing states, and the entries in $\mathbf{Q}^m$ are multistep transition probabilities within the set of nonabsorbiug states. Because there is a positive probability of reaching an absorbing state from each nonabsorbing state, for each nonabsorbing state i there is an integer m depending

on i such that the row in $\mathbf{R}_m$ corresponding to state i has a positive entry. Also, because the columns of $\mathbf{R}_m$ correspond to absorbing states, once an entry in $\mathbf{R}_m$ is positive, it remains positive for all larger integers m. For each state i, let $m(i)$ be an integer m determined as above, and let $M = \max\{m(i) \text{ for all nonabsorbing states } i\}$. Then each row of $\mathbf{R}_M$ has at least one positive entry. Let h be the smallest positive entry in $\mathbf{R}_M$. Then each row in $\mathbf{Q}^M$ has coordinates whose sum is no larger than $1 - h < 1$, and, in particular, each entry is no larger than $1 - h$.

It follows that each entry in $\mathbf{Q}^M\mathbf{Q}^M = \mathbf{Q}^{2M}$ is no larger than $(1-h)^2$, and, in general, each entry in $\mathbf{Q}^{kM}$ is no larger than $(1-h)^k$. Next, consider the sequence $\{\mathbf{Q}^m\}$. For each column in $\mathbf{Q}^m$, the sequence of largest entries forms a monotone decreasing sequence, and consequently the entries in each column of $\mathbf{Q}^m$ tend to zero, and therefore $\lim_{m\to\infty} \mathbf{Q}^m = \mathbf{0}$.

The estimates on the size of the entries in $\mathbf{Q}^m$ derived just above enable us to show that the series $\mathbf{I} + \mathbf{Q} + \mathbf{Q}^2 + \mathbf{Q}^3 + \cdots$ converges. Indeed, for each position in the sum of matrices (3.11), the associated series of numbers consists of M consecutive entries each less than $1 - h$ (which is less than 1), followed by another M consecutive entries each less than $(1-h)^2$, followed by another M consecutive entries each less than $(1-h)^3, \ldots$; and the resulting series of numbers converges. The entries in each position in the series $\mathbf{I} + \mathbf{Q} + \mathbf{Q}^2 + \mathbf{Q}^3 + \cdots$ converge, and therefore the infinite series of matrices converge. Denote the sum by $\mathbf{N}$. Finally,

$$(\mathbf{I} - \mathbf{Q})(\mathbf{I} + \mathbf{Q} + \mathbf{Q}^2 + \mathbf{Q}^3 + \cdots + \mathbf{Q}^m) = \mathbf{I} - \mathbf{Q}^{m+1}$$

Because $\lim_{m\to\infty} \mathbf{Q}^m = \mathbf{0}$, it follows that

$$\lim_{m\to\infty} (\mathbf{I} - \mathbf{Q})(\mathbf{I} + \mathbf{Q} + \mathbf{Q}^2 + \mathbf{Q}^3 + \cdots + \mathbf{Q}^m) = \lim_{m\to\infty} (\mathbf{I} - \mathbf{Q}^{m+1}) = \mathbf{I}$$

and therefore,

$$(\mathbf{I} - \mathbf{Q}) \lim_{m\to\infty} (\mathbf{I} + \mathbf{Q} + \mathbf{Q}^2 + \mathbf{Q}^3 + \cdots + \mathbf{Q}^m) = (\mathbf{I} - \mathbf{Q})\mathbf{N} = \mathbf{I}$$

A similar argument shows that $\mathbf{N}(\mathbf{I} - \mathbf{Q}) = \mathbf{I}$. Therefore, $(\mathbf{I} - \mathbf{Q})$ is invertible and $(\mathbf{I} - \mathbf{Q})^{-1} = \mathbf{N}$. ∎

Proof of Theorem 3.3 The goal is to show that if $\mathbf{P}$ is the transition matrix of an ergodic Markov chain with period 1, then there is a power of $\mathbf{P}$ all of whose entries are strictly positive. The proof is based on a result from elementary number theory concerning the greatest common divisor of a set of integers:

> If d is the greatest common divisor of a set of integers $\{n_1, n_2, n_3, \ldots, n_k\}$, then there are integers $x_1, x_2, x_3, \ldots, x_k$ such that

$$d = n_1x_1 + n_2x_2 + \cdots + n_kx_k \tag{3.13}$$

Suppose that the integers n_j are labeled so that all the positive x_j occur before any negative ones. Then d can be written as the sum $N_1 - N_2$, where N_1 is the portion of the sum (3.13) that includes all positive x_j, and $-N_2$ is the portion of the sum that includes all negative x_j. If the greatest common divisor is 1, then $1 = N_1 - N_2$, with N_1 and N_2 defined as above.

Set $N = N_2^2$. Any integer $n \geq N$ can be written as $n = N + k = N_2^2 + k$, with k a nonnegative integer. We write $k = aN_2 + b$, with $0 \leq b < N_2$ and a equal to the integer j

that satisfies $jN_2 \le k < (j+1)N_2$. With a and b defined in this way, we have

$$n = N_2^2 + k = N_2^2 + aN_2 + b = N_2^2 + aN_2 + b(N_1 - N_2) = (N_2 + a - b)N_2 + bN_1$$

which gives us a representation of n as a linear combination of N_1 and N_2 with *positive* coefficients. Thus, for any integer $n > N$, we have a representation of n as a linear combination of the integers $\{n_1, n_2, n_3, \ldots, n_k\}$ with positive coefficients.

Now, for a specific state i with period 1, let $\{n_1, n_2, n_3, \ldots, n_k\}$ be integers such that $p_{ii}(n_j) > 0$ for $j = 1, 2, \ldots, k$, and the greatest common divisor of $\{n_1, n_2, n_3, \ldots, n_k\}$ is 1. Then for every sufficiently large n, there are positive integers y_j, $j = 1, 2, \ldots, k$, for which we have

$$p_{ii}(n) = p_{ii}(y_1 n_1 + y_2 n_2 + \cdots + y_k n_k) \ge \prod_{j=1}^{k} p_{ii}(y_j n_j) \ge \prod_{j=1}^{k} p_{ii}(n_j)^{y_j} > 0$$

It follows that for any states i and j and any integer m such that $p_{ij}(m) > 0$, we have $p_{ij}(m+n) \ge p_{ii}(n) p_{ij}(m) > 0$ for all sufficiently large integers n.

We now have enough information to complete the proof of Theorem 3.3. For every pair of states i and j, there is an integer m depending on i and j and denoted by $m(i, j)$ such that $p_{ij}(m(i, j)) > 0$. By the result just above, $p_{ij}(m(i, j) + n) > 0$ for all sufficiently large n. Let M be the maximum of $m(i, j)$ over all pairs i and j, let N be the maximum of the integers N_2^2 (there is a value of N_2 for each state), and define $M^* = M + N$. Then

$$p_{ij}(M^*) = p_{ij}(m(i, j) + M^* - m(i, j)) \ge p_{ij}(m(i, j)) p_{ii}(M^* - m(i, j)).$$

Now $p_{ij}(m(i, j)) > 0$ and $M^* - m(i, j) = N + [M - m(i, j)] > N$, and consequently both factors on the right-hand side are positive. This shows that $p_{ij}(M^*) > 0$ for all states i and j. ∎

CHAPTER 4

Simulation Models

4.0 Introduction

In this chapter we discuss in more detail the computer implementation of models used to study activities or processes by imitating or simulating their actual behavior. We introduced the broad ideas and provided examples in Section 2.6. In this chapter we continue the discussion in greater depth. Our view, however, is that simulation is just one of a number of modeling techniques, and therefore we consider only the basic aspects.

Simulation models are widely used and their popularity is increasing. Many systems of current interest are large and complex, and simulation models may be a very effective way to gain insight. Indeed, software that facilitates simulation has become increasingly common, and the cost of computing continues to decline. Consequently, simulation models have become a very attractive option as an aid to understanding complex systems. Because the development of simulation models for complex systems is a time-consuming task, a number of special-purpose simulation languages have been created to reduce the effort. Often these languages are designed to handle a relatively restricted class of situations, but to do so in a convenient and efficient way. Our goal is to illustrate the fundamental concepts, so we use only widely available computer software: the scientific software package MAPLE and the spreadsheet EXCEL; there are many alternatives to both.

4.1 The Simulation Process

The simulation process is intended to help us understand the behavior of a system by using a computer to imitate its behavior or certain aspects of its behavior. Although the term *simulation* is sometimes used to refer to the use of a computer in models for completely deterministic situations—that is, models in which a specific set of inputs always yields the same set of outputs—we will use the term *numerical model* or *computational model* for such situations. We will reserve the term *simulation model* for situations in which some of the quantities or aspects of the system being studied are described in probabilistic terms. In such situations, the output data are themselves random, and consequently we can obtain only estimates of the actual behavior of the system. For instance, some of the population models studied in Section 2.2 are numerical models in the sense that the model is deterministic