Chapter 10

Finite-State Markov Chains

Introductory Example: Googling Markov Chains

Google means many things: it is an Internet search engine, the company that produces the search engine, and a verb meaning to search on the Internet for a piece of information. Although it may seem hard to believe, there was a time before people could “google” to find the capital of Botswana, or a recipe for deviled eggs, or other vitally important matters. Users of the Internet depend on trustworthy search engines – the amount of available information is so vast that the searcher relies on the search engine not only to find those webpages that contain the terms of the search, but also to return first those webpages most likely to be relevant to the search. Early search engines had no good way of determining which pages were more likely to be relevant. Searchers had to check the returned pages one by one, which was a tedious and frustrating process. This situation improved markedly in 1998, when search engines began to use the information contained in the hyperlinked structure of the World Wide Web to help to rank pages. Foremost among this new generation of search engines was Google, a project of two computer science graduate students at Stanford University: Sergey Brin and Lawrence Page.

Brin and Page reasoned that a webpage was important if it had hyperlinks to it from other important pages. They used the idea of random surfer: a web surfer moving from webpage to webpage merely by choosing at random which hyperlink to follow. The motion of the surfer among the webpages can be modeled using Markov chains, which were introduced in Section 4.9. The pages that this random surfer visits more often ought to be more important, and thus more relevant if their content matches the terms of a search. Although Brin and Page did not not know it at the time, they were attempting to find the steady-state vector for a particular Markov chain whose transition matrix modeled the hyperlinked structure of the web. After some important modifications to this impressively large matrix (detailed in Section 10.2), a steady-state vector can be found, and its entries can be interpreted as the amount of time a random surfer will spend at each webpage. The calculation of this steady-state vector is the basis for Google’s PageRank algorithm.

So the next time you google the capital of Botswana, know that you are using the results of this chapter to find just the right webpage.

Even though the number of webpages is huge, it is still finite. When the link structure of the World Wide Web is modeled by a Markov chain, each webpage is a state of the Markov chain. This chapter continues the study of Markov chains begun in Section 4.9, focusing on those Markov chains with a finite number of states. Section 10.1 introduces useful terminology and develops some examples of Markov chains: signal transmission models, diffusion models from physics, and random walks on various sets. Random walks on directed graphs will have particular application to the PageRank algorithm. Section 10.2 defines the steady-state vector for a Markov chain. Although all Markov chains have a steady-state vector, not all Markov chains converge to a steady-state vector. When the Markov chain converges to a steady-state vector, that vector can be interpreted as telling the amount of time the chain will spend in each state. This interpretation is necessary for the PageRank algorithm, so the conditions under which a Markov chain converges to a steady-state vector will be developed. The model for the link structure of the World Wide Web will then be modified to meet these conditions, forming what is called the Google matrix. Sections 10.3 and 10.4 discuss Markov chains that do not converge to a steady-state vector. These Markov chains can be used to model situations in which the chain eventually becomes confined to one state or a set of states. Section 10.5 introduces the fundamental matrix. This matrix can be used to calculate the expected number of steps it takes the chain to move from one state to another, as well as the probability that the chain ends up confined to a particular state. In Section 10.6, the fundamental matrix is applied to a model for run production in baseball: the number of batters in a half inning and the state in which the half inning ends will be of vital importance in calculating the expected number of runs scored.

10.1 Introduction and Examples

Recall from Section 4.9 that a **Markov chain** is a mathematical model for movement between states. A process starts in one of these states and moves from state to state. The moves between states are called **steps** or **transitions**. The terms "chain" and "process" are used interchangeably, so the chain can be said to move between states and to be "at a state" or "in a state" after a certain number of steps.

The state of the chain at any given step is not known; what is known is the probability that the chain moves from state j to state i in one step. This probability is called a **transition probability** for the Markov chain. The transition probabilities are placed in a matrix called the **transition matrix** P for the chain by entering the probability of a transition from state j to state i at the (i, j)-entry of P. So if there were m states named 1, 2, $\ldots m$, the transition matrix would be the $m \times m$ matrix

From:

$$P = \begin{array}{c} \begin{array}{ccccc} 1 & j & & m & \end{array} \\ \left[\begin{array}{ccccc} & \vdots & & & \\ & \downarrow & & & \\ & p_{ij} & \cdots & \longrightarrow & \\ & & & & \end{array}\right] \end{array} \begin{array}{l} \text{To:} \\ 1 \\ \\ i \\ m \end{array}$$

The probabilities that the chain is in each of the possible states after n steps are listed in a **state vector** $\mathbf{x}_n$. If there are m possible states the state vector would be

$$\mathbf{x}_n = \begin{bmatrix} a_1 \\ \vdots \\ a_j \\ \vdots \\ a_m \end{bmatrix} \quad \longleftarrow \text{Probability that the chain is at state } j \text{ after } n \text{ steps}$$

State vectors are **probability vectors** since their entries must sum to 1. The state vector $\mathbf{x}_0$ is called the **initial probability vector**.

Notice that the j^{th} column of P is a probability vector – its entries list the probabilities of a move from state j to the states of the Markov chain. The transition matrix is thus a **stochastic matrix** since all of its columns are probability vectors.

The state vectors for the chain are related by the equation

$$\mathbf{x}_{n+1} = P\mathbf{x}_n \tag{1}$$

for $n = 1, 2, \ldots$. Notice that Equation (1) may be used to show that

$$\mathbf{x}_n = P^n\mathbf{x}_0 \tag{2}$$

Thus any state vector $\mathbf{x}_n$ may be computed from the initial probability vector $\mathbf{x}_0$ and an appropriate power of the transition matrix P.

This chapter concerns itself with Markov chains with a finite number of states; that is, those chains for which the transition matrix P is of finite size. To use a finite-state Markov chain to model a process, the process must have the following properties, which are implied by Equations (1) and (2).

1. Since the values in the vector $\mathbf{x}_{n+1}$ depend only on the transition matrix P and on $\mathbf{x}_n$, the state of the chain before time n must have no effect on its state at time $n+1$ and beyond.

2. Since the transition matrix P does not change with time, the probability of a transition from one state to another must not depend upon how many steps the chain has taken.

Even with these restrictions, Markov chains may be used to model an amazing variety of processes. Here is a sampling.

Signal Transmission

Consider the problem of transmitting a signal along a telephone line or by radio waves. Each piece of data must pass through a multi-stage process to be transmitted, and at each stage there is a probability that a transmission error will cause the data to be corrupted. Assume that the probability of an error in transmission is not effected by transmission errors in the past and does not depend on time, and that the number of possible pieces of data is finite. The transmission process may then be modeled by a Markov chain. The object of interest is the probability that a piece of data goes through the entire multi-stage process without error. Here is an example of such a model.

EXAMPLE 1 Suppose that each bit of data is either a 0 or a 1, and at each stage there is a probability p that the bit will pass through the stage unchanged. Thus the probability is $1-p$ that the bit will be transposed. The transmission process is modeled by a Markov chain, with states 0 and 1 and transition matrix

$$\begin{array}{cc} & \text{From:} \\ & \begin{array}{cc} \scriptstyle 0 & \scriptstyle 1 \end{array} \quad \text{To:} \\ P = & \begin{bmatrix} p & 1-p \\ 1-p & p \end{bmatrix} \begin{array}{c} \scriptstyle 0 \\ \scriptstyle 1 \end{array} \end{array}$$

It is often easier to visualize the action of a Markov chain by representing its transition probabilities graphically as in Figure 1. The points are the states of the chain, and the arrows represent the transitions.

Suppose that $p = .99$. Find the probability that the signal 0 will still be a 0 after a 2-stage transmission process.

Solution Since the signal begins as 0, the probability that the chain begins at 0 is 100%, or 1; that is, the initial probability vector is

$$\mathbf{x}_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

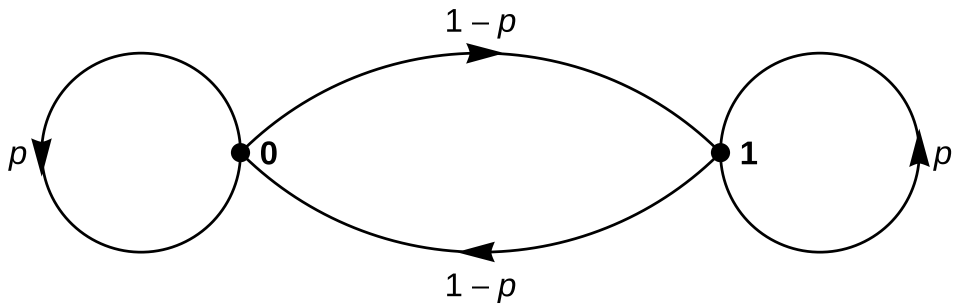


Figure 1: Transition diagram for signal transmission.

To find the probability of a two-step transition, compute

$$\mathbf{x}_2 = P^2\mathbf{x}_0 = \begin{bmatrix} .99 & .01 \\ .01 & .99 \end{bmatrix}^2 \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} .9802 & .0198 \\ .0198 & .9802 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} .9802 \\ .0198 \end{bmatrix}$$

The probability that the signal 0 will still be a 0 after the 2-stage process is thus $.9802$. Notice that this is not the same as the probability that the 0 is transmitted without error; that probability would be $(.99)^2 = .9801$. Our analysis includes the very small probability that the 0 is erroneously changed to 1 in the first step, then back to 0 in the second step of transmission. ■

Diffusion

Consider two compartments filled with different gases which are separated only by a membrane which allows molecules of each gas to pass from one container to the other. The two gases will then diffuse into each other over time, so that each container will contain some mixture of the gases. The major question of interest is what mixture of gases is in each container at a time after the containers are joined. A famous mathematical model for this process was described originally by the physicists Paul and Tatyana Ehrenfest. Since their preferred term for "container" was urn, the model is called the **Ehrenfest urn model** for diffusion.

Label the two urns A and B, and place k molecules of gas in each urn. At each time step, select one of the $2k$ molecules at random and move it from its urn to the other urn, and keep track of the number of molecules in urn A. This process can be modeled by a finite-state Markov chain: the number of molecules in urn A after $n+1$ time steps depends only on the number in urn A after n time steps, the transition probabilities do not change with time, and the number of states is finite.

EXAMPLE 2 For this example, let $k = 3$. Then the two urns contain a total of 6 molecules, and the possible states for the Markov chain are 0, 1, 2, 3, 4, 5, and 6. Notice first that if there are 0 molecules in urn A at time n, then there must be 1 molecule in urn A at time $n+1$, and if there are 6 molecules in urn A at time n, then there must be 5 molecules in urn A at time $n+1$. In terms of the transition matrix P, this means that the columns in P corresponding to states 0 and 6 are

$$\mathbf{p}_0 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \text{ and } \mathbf{p}_6 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}$$

If there are i molecules in urn A at time n, with $0 < i < 6$, then there must be either $i-1$ or $i+1$ molecules in urn A at time $n+1$. In order for a transition from i to $i-1$ molecules to occur, one of the i molecules in urn A must be selected to move; this event happens with probability $i/6$. Likewise a transition from i to $i+1$ molecules occurs when one of the $6-i$ molecules in urn B is selected, and this occurs with probability $(6-i)/6$. Allowing i to range from 1 to 5 creates the columns of P corresponding to these states, and the transition matrix for the Ehrenfest urn model with $k = 3$ is thus

$$P = \begin{array}{c} \begin{array}{ccccccc} 0 & 1 & 2 & 3 & 4 & 5 & 6 \end{array} \\ \left[\begin{array}{ccccccc} 0 & 1/6 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1/3 & 0 & 0 & 0 & 0 \\ 0 & 5/6 & 0 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 2/3 & 0 & 2/3 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 0 & 5/6 & 0 \\ 0 & 0 & 0 & 0 & 1/3 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1/6 & 0 \end{array}\right] \end{array} \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array}$$

Figure 2 shows a transition diagram of this Markov chain. Another model for diffusion will be considered in the Exercises for this section. ■

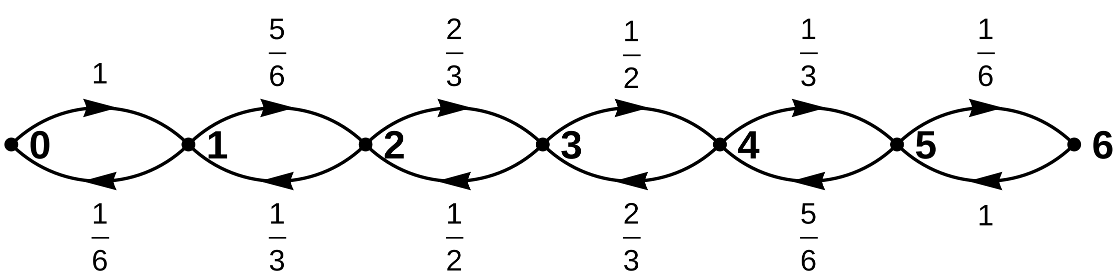


Figure 2: Transition diagram of the Ehrenfest urn model.

Random Walks on $\{1, \ldots, n\}$

Molecular motion has long been an important issue in physics. Einstein and others investigated Brownian motion, which is a mathematical model for the motion of a molecule exposed to collisions with other molecules. The analysis of Brownian motion turns out to be quite complicated, but a discrete version of Brownian motion called a **random walk** provides an introduction to this important model. Think of the states $\{1, 2, \ldots, n\}$ as lying on a line. Place a molecule at a point that is not on the end of the line. At each step the molecule moves left one unit with probability p and right one unit with probability $1 - p$. See Figure 3. The molecule thus "walks randomly" along the line. If $p = 1/2$, the walk is called **simple**, or **unbiased**. If $p \neq 1/2$, the walk is said to be **biased**.

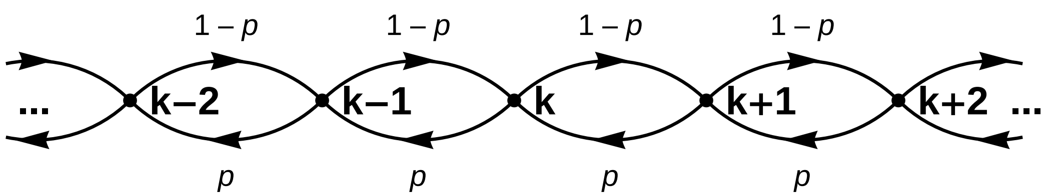


Figure 3: A graphical representation of a random walk.

The molecule must move either to the left or right at the states $2, \ldots, n-1$, but it cannot do this at the endpoints 1 and n. The molecule's possible movements at the endpoints 1 and n must be specified. One possibility is to have the molecule stay at an endpoint forever once it reaches either end of the line. This is called a **random walk with absorbing boundaries**, and the endpoints 1 and n are called **absorbing states**. Another possibility is to have the molecule bounce back one unit when an endpoint is reached. This is called a **random walk with reflecting boundaries**.

EXAMPLE 3 A random walk on $\{1, 2, 3, 4, 5\}$ with absorbing boundaries has a transition matrix of

$$
\begin{array}{cc}
 & \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\
P = & \begin{bmatrix} 1 & p & 0 & 0 & 0 \\ 0 & 0 & p & 0 & 0 \\ 0 & 1-p & 0 & p & 0 \\ 0 & 0 & 1-p & 0 & 0 \\ 0 & 0 & 0 & 1-p & 1 \end{bmatrix} \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array}
\end{array}
$$

since the molecule at state 1 has probability 1 of staying at state 1, and a molecule at state 5 has probability 1 of staying at state 5. A random walk on $\{1, 2, 3, 4, 5\}$ with reflecting boundaries has a transition matrix of

$$
\begin{array}{cc}
 & \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\
P = & \begin{bmatrix} 0 & p & 0 & 0 & 0 \\ 1 & 0 & p & 0 & 0 \\ 0 & 1-p & 0 & p & 0 \\ 0 & 0 & 1-p & 0 & 1 \\ 0 & 0 & 0 & 1-p & 0 \end{bmatrix} \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array}
\end{array}
$$

since the molecule at state 1 has probability 1 of moving to state 2, and a molecule at state 5 has probability 1 of moving to state 4. ■

In addition to their use in physics, random walks also occur in problems related to gambling and its more socially acceptable variants: the stock market and the insurance industry.

EXAMPLE 4 Consider a very simple casino game. A gambler (who still has some money left with which to gamble) flips a fair coin and calls heads or tails. If the gambler is correct, he wins a dollar; if he is wrong, he loses a dollar. Suppose that the gambler will quit the game when either he has won n dollars or has lost all of his money.

Suppose that $n = 7$ and the gambler starts with \$4. Notice that the gambler's winnings move either up or down \$1 at each move, and once the gambler's winnings reach 0 or 7, they do not change any more since the gambler has quit the game. Thus the gambler's winnings may be modeled by a random walk with absorbing boundaries and states $\{0, 1, 2, 3, 4, 5, 6, 7\}$. Since a move up or down is equally likely in this case, $p = 1/2$ and the walk is simple.

Random Walk on Graphs

It is useful to perform random walks on geometrical objects other than the one-dimensional line. For example, a **graph** is a collection of points and lines connecting some of the points. The points of a graph are called vertices, and the lines connecting the vertices are called the edges. In Figure 4, the vertices are labeled with the numbers 1 through 7.

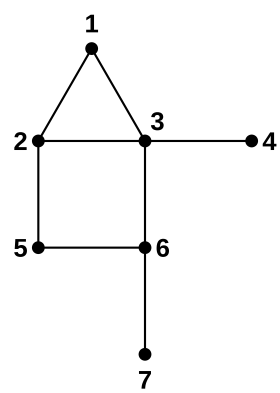


Figure 4: A graph with seven vertices.

To define simple random walk on a graph, allow the chain to move from vertex to vertex on the graph. At each step the chain is equally likely to move along any of the edges attached to the vertex. For example, if the molecule is at state 5 in Figure 4, it has probability $1/2$ of moving to state 2 and probability $1/2$ of moving to state 6. This Markov chain is called a **simple random walk on a graph**.

EXAMPLE 5 Simple random walk on the graph in Figure 4 has transition matrix

$$P = \begin{array}{c} \begin{array}{ccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{array} \\ \left[\begin{array}{ccccccc} 0 & 1/3 & 1/4 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 1/4 & 0 & 1/2 & 0 & 0 \\ 1/2 & 1/3 & 0 & 1 & 0 & 1/3 & 0 \\ 0 & 0 & 1/4 & 0 & 0 & 0 & 0 \\ 0 & 1/3 & 1/4 & 0 & 0 & 1/3 & 0 \\ 0 & 0 & 0 & 0 & 1/2 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1/3 & 0 \end{array}\right] \end{array} \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{array}$$

Find the probability that the chain in Figure 4 moves from state 6 to state 2 in exactly 3 steps.

Solution Compute

$$\mathbf{x}_3 = P^3\mathbf{x}_0 = P^3 \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} .0833 \\ .0417 \\ .4028 \\ 0 \\ .2778 \\ 0 \\ .1944 \end{bmatrix}$$

so the probability of moving from state 6 to state 2 in exactly 3 steps is $.0417$. ■

Sometimes interpreting a random process as a random walk on a graph can be useful.

EXAMPLE 6 Suppose a mouse runs through the five-room maze on the left side of Figure 5. The mouse moves to a different room at each time step. When the mouse is in a particular room, it is equally likely to choose any of the doors out of the room. Note that a Markov chain can model the motion of the mouse. Find the probability that a mouse starting in room 3 returns to that room in exactly 5 steps.

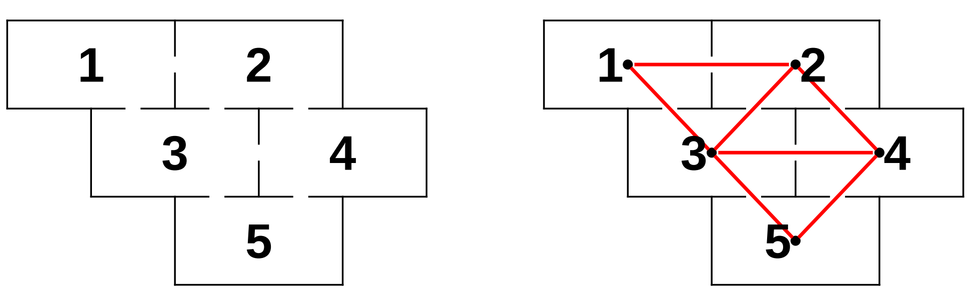


Figure 5: The five-room maze with overlaid graph.

Solution A graph is overlaid on the maze on the right side of Figure 5. Notice that the motion of the mouse is identical to simple random walk on the graph, so the transition matrix is

$$P = \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{bmatrix} 0 & 1/3 & 1/4 & 0 & 0 \\ 1/2 & 0 & 1/4 & 1/3 & 0 \\ 1/2 & 1/3 & 0 & 1/3 & 1/2 \\ 0 & 1/3 & 1/4 & 0 & 1/2 \\ 0 & 0 & 1/4 & 1/3 & 0 \end{bmatrix} \end{array} \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array}$$

and find that

$$\mathbf{x}_5 = P^5\mathbf{x}_0 = P^5 \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} .1507 \\ .2143 \\ .2701 \\ .2143 \\ .1507 \end{bmatrix}$$

Thus the probability of a return to room 3 in exactly 5 steps is $.2701$. ■

Another interesting object on which to walk randomly is a **directed graph**. A directed graph is a graph in which the vertices are not joined by lines but by arrows. See Figure 6.

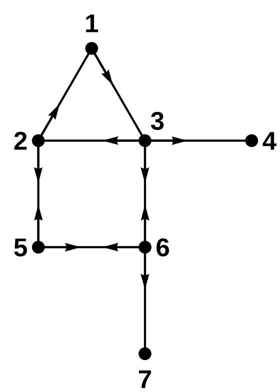


Figure 6: A directed graph with seven vertices.

To perform a simple random walk on a directed graph, allow the chain to move from vertex to vertex on the graph but only in the directions allowed by the arrows. At each step the walker is equally likely to move away from its current state along any of the arrows pointing away from the vertex. For example, if the molecule is at state 6 in Figure 6, it has probability $1/3$ of moving to state 3, state 5, and state 7.

The PageRank algorithm which Google uses to rank the importance of pages on the World Wide Web (see the Chapter Introduction) begins with a simple random walk on a directed graph. The Web is modeled as a directed graph where the vertices are the pages and an arrow is drawn from page j to page i if there is a hyperlink from page j to page i. A person surfs randomly in the following way: when the surfer gets to a page, he or she chooses a link from the page so that it is equally probable to choose any of the possible "outlinks." The surfer then follows the link to arrive at another page. The person surfing in that way is performing a simple random walk on the directed graph that is the World Wide Web.

EXAMPLE 7 Consider a set of seven pages hyperlinked by the directed graph in Figure 6. If the random surfer starts at page 5, find the probability that the surfer is at page 3 after four clicks.

Solution The transition matrix for the simple random walk on the directed graph is

$$\begin{array}{cc} & \begin{array}{ccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{array} \\ P = & \begin{bmatrix} 0 & 1/2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/3 & 0 & 1/2 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1/3 & 0 \\ 0 & 0 & 1/3 & 1 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 0 & 0 & 1/3 & 0 \\ 0 & 0 & 1/3 & 0 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1/3 & 1 \end{bmatrix} \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{array} \end{array}$$

Notice that there are no arrows coming from either state 4 or state 7. If the surfer clicks on a link to either of these pages, there is no link to click on next.[1] For that reason, the transition probabilities p_{44} and p_{77} are set equal to 1 – the chain must stay at state 4 or state 7 forever once it enters either of these states. Computing $\mathbf{x}_4$ gives

$$\mathbf{x}_4 = P^4\mathbf{x}_0 = \begin{bmatrix} .1319 \\ .0833 \\ .0880 \\ .1389 \\ .2199 \\ .0833 \\ .2546 \end{bmatrix}$$

so the probability of being at 3 after exactly 4 clicks is $.0880$. ■

States 4 and 7 are absorbing states for the Markov chain in the previous example. In technical terms they are called **dangling nodes** and are quite common on the Web – data pages in particular usually have no links leading from them. Dangling nodes will appear in the next section, where the PageRank algorithm will be explained.

As was noted in Section 4.9, the most interesting questions about Markov chains concern their long-term behavior; that is, the behavior of $\mathbf{x}_n$ as n increases. This study will occupy a large portion of this chapter. The foremost issues in our study will be whether the sequence of vectors $\{\mathbf{x}_n\}$ is converging to some limiting vector as n increases, and how to interpret this limiting vector if it exists. Convergence to a limiting vector will be addressed in the next section.

[1] Using the “Back” key is not allowed – the state of the chain before time n must have no effect on its state at time $n + 1$ and beyond.

Practice Problems

1. Fill in the missing entries in the stochastic matrix

$$P = \begin{bmatrix} .1 & * & .2 \\ * & .3 & .3 \\ .6 & .2 & * \end{bmatrix}$$

2. In the signal transmission model in Example 1, suppose that $p = .03$. Find the probability that the signal "1" will be a "0" after a 3-stage transmission process.

10.1 Exercises

In Exercises 1 and 2, determine whether P is a stochastic matrix. If P is not a stochastic matrix, explain why not.

1. a. $P = \begin{bmatrix} .3 & .4 \\ .7 & .6 \end{bmatrix}$ b. $P = \begin{bmatrix} .3 & .7 \\ .4 & .6 \end{bmatrix}$

2. a. $P = \begin{bmatrix} 1 & .5 \\ 0 & .5 \end{bmatrix}$ b. $P = \begin{bmatrix} .2 & 1.1 \\ .8 & -.1 \end{bmatrix}$

In Exercises 3 and 4, compute $\mathbf{x}_3$ in two ways: by computing $\mathbf{x}_1$ and $\mathbf{x}_2$, and by computing P^3.

3. $P = \begin{bmatrix} .6 & .5 \\ .4 & .5 \end{bmatrix}; \mathbf{x}_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$

4. $P = \begin{bmatrix} .3 & .8 \\ .7 & .2 \end{bmatrix}; \mathbf{x}_0 = \begin{bmatrix} .5 \\ .5 \end{bmatrix}$

In Exercises 5 and 6, the transition matrix P for a Markov chain with states 0 and 1 is given. Assume that in each case the chain starts in state 0 at time $n = 0$. Find the probability that the chain is in state 1 at time n.

5. $P = \begin{bmatrix} 1/3 & 3/4 \\ 2/3 & 1/4 \end{bmatrix}, n = 3$

6. $P = \begin{bmatrix} .4 & .2 \\ .6 & .8 \end{bmatrix}, n = 5$

In Exercises 7 and 8, the transition matrix P for a Markov chain with states 0, 1 and 2 is given. Assume that in each case the chain starts in state 0 at time $n = 0$. Find the probability that the chain is in state 1 at time n.

7. $P = \begin{bmatrix} 1/3 & 1/4 & 1/2 \\ 1/3 & 1/2 & 1/4 \\ 1/3 & 1/4 & 1/4 \end{bmatrix}, n = 2$

8. $P = \begin{bmatrix} .1 & .2 & .4 \\ .6 & .3 & .4 \\ .3 & .5 & .2 \end{bmatrix}, n = 3$

9. Consider a pair of Ehrenfest urns. If there are currently 3 molecules in one urn and 5 in the other, what is the probability that the exact same situation will apply after
   a. 4 selections?
   b. 5 selections?

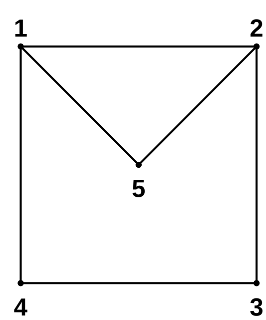
10. Consider a pair of Ehrenfest urns. If there are currently no molecules in one urn and 7 in the other, what is the probability that the exact same situation will apply after
    a. 4 selections?
    b. 5 selections?

11. Consider unbiased random walk on the set $\{1, 2, 3, 4, 5, 6\}$. What is the probability of moving from 2 to 3 in exactly 3 steps if the walk has
    a. reflecting boundaries?
    b. absorbing boundaries?

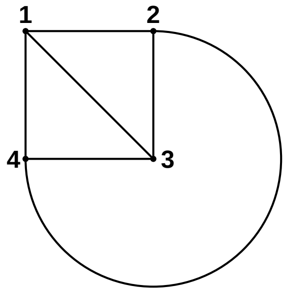
12. Consider biased random walk on the set $\{1, 2, 3, 4, 5, 6\}$ with $p = 2/3$. What is the probability of moving from 2 to 3 in exactly 3 steps if the walk has
    a. reflecting boundaries?
    b. absorbing boundaries?

In Exercises 13 and 14, find the transition matrix for the simple random walk on the given graph.

13.

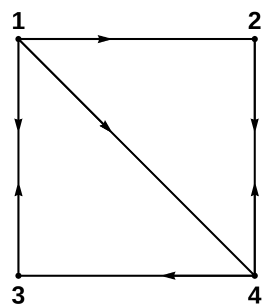


14.

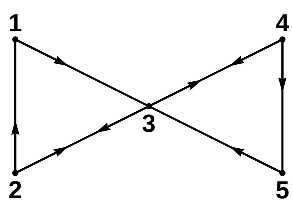


In Exercises 15 and 16, find the transition matrix for the simple random walk on the given directed graph.

15.



16.

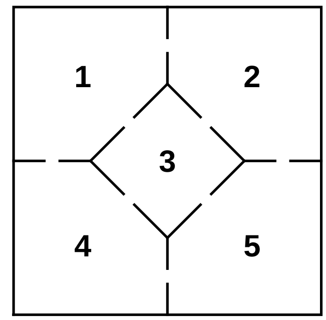


In Exercises 17 and 18, suppose a mouse wanders through the given maze. The mouse must move into a different room at each time step, and is equally likely to leave the room through any of the available doorways.

17. The mouse is placed in room 2 of the maze below.

   a. Construct a transition matrix and an initial probability vector for the mouse's travels.

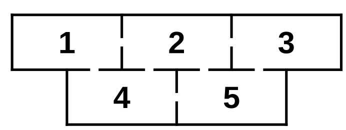
   b. What are the probabilities that the mouse is in each of the rooms after 3 moves?



18. The mouse is placed in room 3 of the maze below.

   a. Construct a transition matrix and an initial probability vector for the mouse's travels.

   b. What are the probabilities that the mouse is in each of the rooms after 4 moves?

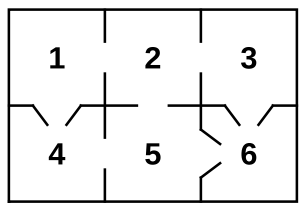


In Exercises 19 and 20, suppose a mouse wanders through the given maze some of whose doors are "one-way": they are just large enough for the mouse to squeeze through in only one direction. The mouse still must move into a different room at each time step if possible. When faced with accessible openings into two or more rooms, the mouse chooses them with equal probability.

19. The mouse is placed in room 1 of the maze below.

   a. Construct a transition matrix and an initial probability vector for the mouse's travels.

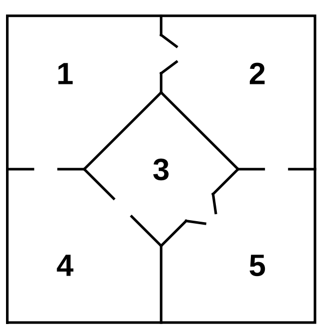
b. What are the probabilities that the mouse is in each of the rooms after 4 moves?



20. The mouse is placed in room 1 of the maze below.

a. Construct a transition matrix and an initial probability vector for the mouse's travels.

b. What are the probabilities that the mouse is in each of the rooms after 3 moves?



In Exercises 21 and 22, mark each statement True or False. Justify each answer.

21. a. The columns of a transition matrix for a Markov chain must sum to 1.

b. The transition matrix P may change over time.

c. The (i, j)-entry in a transition matrix P gives the probability of a move from state j to state i.

22. a. The rows of a transition matrix for a Markov chain must sum to 1.

b. If $\{\mathbf{x}_n\}$ is a Markov chain, then $\mathbf{x}_{n+1}$ must depend only on the transition matrix and $\mathbf{x}_n$.

c. The (i, j)-entry in P^3 gives the probability of a move from state i to state j in exactly three time steps.

23. The weather in Charlotte, North Carolina can be classified as either sunny, cloudy, or rainy on a given day. Climate data from 2003[2] reveal the following facts:

- If a day is sunny, then the next day is sunny with probability .65, cloudy with probability .1, and rainy with probability .25.
- If a day is cloudy, then the next day is sunny with probability .25, cloudy with probability .25, and rainy with probability .5.
- If a day is rainy, then the next day is sunny with probability .25, cloudy with probability .15, and rainy with probability .60.

Suppose it is cloudy on Monday. Use a Markov chain to find the probabilities of the different kinds of possible weather on Friday.

24. Suppose that whether it rains in Charlotte tomorrow depends on the weather conditions for today and yesterday. Climate data in 2003 [1] show that

- If it rained yesterday and today, then it will rain tomorrow with probability .58.

[2]http://www.wunderground.com/history/airport/KCLT/2003/1/1/MonthlyHistory.html

- If it rained yesterday but not today, then it will rain tomorrow with probability .29.
- If it rained today but not yesterday, then it will rain tomorrow with probability .47.
- If it did not rain yesterday and today, then it will rain tomorrow with probability .31.

Even though the weather depends on the last two days in this case, we can create a Markov chain model using the states

1 it rained yesterday and today
2 it rained yesterday but not today
3 it rained today but not yesterday
4 it did not rain yesterday and today

So, for example, the probability of a transition from state 1 to state 1 is $.58$, and the transition from state 1 to state 3 is 0.

a. Complete the creation of the transition matrix for this Markov chain.

b. If it rains on Tuesday and is clear on Wednesday, what is the probability of no rain on the next weekend?

25. Consider a set of four webpages hyperlinked by the directed graph in Exercise 15. If a random surfer starts at page 1, what is the probability that the surfer is at each of the pages after 3 clicks?

26. Consider a set of five webpages hyperlinked by the directed graph in Exercise 16. If a random surfer starts at page 2, what is the probability that the surfer is at each of the pages after 4 clicks?

27. Consider a model for signal transmission where data is sent as two-bit bytes. Then there are four possible bytes 00, 01, 10, and 11 which are the states of the Markov chain. At each stage there is a probability p that each bit will pass through the stage unchanged.

    a. Construct the transition matrix for the model.

    b. Suppose that $p = .99$. Find the probability that the signal "01" will still be "01" after a three-stage transmission.

28. Consider a model for signal transmission where data is sent as three-bit bytes. Construct the transition matrix for the model.

29. Another version of the Ehrenfest model for diffusion starts with k molecules of gas in each urn. One of the $2k$ molecules is picked at random just as in the Ehrenfest model in the text. The chosen molecule is then moved to the other urn with a fixed probability p and is placed back in its urn with probability $1-p$. (Note that the Ehrenfest model in the text is this model with $p = 1$.)

    a. Let $k = 3$. Find the transition matrix for this model.

    b. Let $k = 3$ and $p = 1/2$. If there are currently no balls in Urn A, what is the probability that there will be 3 balls in Urn A after 5 selections?

30. Another model for diffusion is called the **Bernoulli-LaPlace** model. Two urns (Urn A and Urn B) contain a total of $2k$ molecules. In this case, k of the molecules are of one type (called Type I molecules) and k are of another type (Type II molecules). In addition, k molecules must be in each urn all times. At each time step, a pair of molecules is selected, one from Urn A and one from Urn B, and these molecules change urns. Let the Markov chain model

the number of Type I molecules in Urn A (which is also the number of Type II molecules in Urn B).

a. Suppose that there are j Type I molecules in Urn A with $0 < j < k$. Explain why the probability of a transition to $j-1$ Type I molecules in Urn A is $(j/k)^2$, and why the probability of a transition to $j+1$ Type I molecules in Urn A is $((k-j)/k)^2$.

b. Let $k = 5$. Use the result in part a. to set up the transition matrix for the Markov chain which models the number of Type I molecules in Urn A.

c. Let $k = 5$ and begin with all Type I molecules in Urn A. What is the distribution of Type I molecules after 3 time steps?

31. To win a game in tennis, one player must score four points and must also score at least two points more than his or her opponent. Thus if the two players have scored an equal number of points (which is called "deuce" in tennis jargon), one player must then score two points in a row to win the game. Suppose that players A and B are playing a game of tennis which is at deuce. If A wins the next point, it is called "advantage A" while if B wins the point it is "advantage B." If the game is at advantage A and player A wins the next point, then player A wins the game. If player B wins the point at advantage A the game is back at deuce.

a. Suppose that the probability that A wins any point is p. Model the progress of a tennis game starting at deuce using a Markov chain with the five states

1 deuce
2 advantage A
3 advantage B
4 A wins the game
5 B wins the game

Find the transition matrix for this Markov chain.

b. Let $p = .6$. Find the probability that the game is at "advantage B" after three points starting at deuce.

32. Volleyball uses two different scoring systems in which a team must win by at least two points. In both systems, a *rally* begins with a serve by one of the teams and ends when the ball goes out of play, touches the floor, or a player commits a fault. The team that wins the rally gets to serve for the next rally. Games are played to 15, 25 or 30 points.

a. In *rally point scoring* the team that wins a rally is awarded a point no matter which team served for the rally. Assume that team A has probability p of winning a rally for which it serves, and that team B has probability q of winning a rally for which it serves. Model the progress of a volleyball game using a Markov chain with the six states

1 tied – A serving
2 tied – B serving
3 A ahead by 1 point – A serving
4 B ahead by 1 point – B serving
5 A wins the game
6 B wins the game

Find the transition matrix for this Markov chain.

b. Suppose that team A and team B are tied 15-15 in a 15-point game and that team A is serving. Let $p = q = .6$. Find the probability that the game is not finished after three rallies.

c. In *side out scoring* the team that wins a rally is awarded a point only when it served for the rally. Assume that team A has probability p of winning a rally for which it serves, and that team B has probability q of winning a rally for which it serves. Model the progress of a volleyball game using a Markov chain with the eight states

1	tied – A serving
2	tied – B serving
3	A ahead by 1 point – A serving
4	A ahead by 1 point – B serving
5	B ahead by 1 point – A serving
6	B ahead by 1 point – B serving
7	A wins the game
8	B wins the game

Find the transition matrix for this Markov chain.

d. Suppose that team A and team B are tied 15-15 in a 15-point game and that team A is serving. Let $p = q = .6$. Find the probability that the game is not finished after three rallies.

33. Suppose that P is a stochastic matrix all of whose entries are greater than or equal to p. Show that all of the entries in P^n are greater than or equal to p for $n = 1, 2, \ldots$.

Solutions to Practice Problems

1. Since a stochastic matrix must have columns that sum to 1,

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

2. The transition matrix for the model is

$$P = \begin{bmatrix} .97 & .03 \\ .03 & .97 \end{bmatrix}$$

Since the signal begins as "1", the initial probability vector is

$$\mathbf{x}_0 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

To find the probability of a three-step transition, compute

$$\mathbf{x}_2 = P^3\mathbf{x}_0 = \begin{bmatrix} .9153 & .0847 \\ .0847 & .9153 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} .0847 \\ .9153 \end{bmatrix}$$

The probability of a change to "0" is thus $.0847$.

10.2 The Steady-State Vector and Google's PageRank

As was seen in Section 4.9, the most interesting aspect of a Markov chain is its long-range behavior: the behavior of $\mathbf{x}_n$ as n increases without bound. In many cases, the sequence of vectors $\{\mathbf{x}_n\}$ is converging to a vector which is called the **steady-state vector** for the Markov chain. This section will review how to compute the steady-state vector of a Markov chain, explain how to interpret this vector if it exists, and will offer an expanded version of Theorem 18 in Section 4.9, which describes the circumstances under which $\{\mathbf{x}_n\}$ converges to a steady-state vector. This Theorem will be applied to the Markov chain model used for the World Wide Web in the previous section and will show how the PageRank method for ordering the importance of webpages is derived.

Steady-State Vectors

In many cases, the Markov chain $\mathbf{x}_n$ and the matrix P^n change very little for large values of n.

EXAMPLE 1 To begin, recall Example 3 from Section 4.9. That example concerned a Markov chain with transition matrix $P = \begin{bmatrix} .5 & .2 & .3 \\ .3 & .8 & .3 \\ .2 & 0 & .4 \end{bmatrix}$ and initial probability vector $\mathbf{x}_0 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$. The vectors $\mathbf{x}_n$ were seen to be converging to the vector $\mathbf{q} = \begin{bmatrix} .3 \\ .6 \\ .1 \end{bmatrix}$. This result may be written as $\lim_{n\to\infty} \mathbf{x}_n = \mathbf{q}$. Increasing powers of the transition matrix P may also be computed, giving:

$$P^2 = \begin{bmatrix} .3700 & .2600 & .3300 \\ .4500 & .7000 & .4500 \\ .1800 & .0400 & .2200 \end{bmatrix} \qquad P^3 = \begin{bmatrix} .3290 & .2820 & .3210 \\ .5250 & .6500 & .5250 \\ .1460 & .0680 & .1540 \end{bmatrix}$$

$$P^4 = \begin{bmatrix} .3133 & .2914 & .3117 \\ .5625 & .6250 & .5625 \\ .1242 & .0836 & .1258 \end{bmatrix} \qquad P^5 = \begin{bmatrix} .3064 & .2958 & .3061 \\ .5813 & .6125 & .5813 \\ .1123 & .0917 & .1127 \end{bmatrix}$$

$$P^{10} = \begin{bmatrix} .3002 & .2999 & .3002 \\ .5994 & .6004 & .5994 \\ .1004 & .0997 & .1004 \end{bmatrix} \qquad P^{15} = \begin{bmatrix} .3000 & .3000 & .3000 \\ .6000 & .6000 & .6000 \\ .1000 & .1000 & .1000 \end{bmatrix}$$

so the sequence of matrices $\{P^n\}$ also seems to be converging to a matrix as n increases, and this matrix has the unusual property that all of its columns equal $\mathbf{q}$. The example also showed that $P\mathbf{q} = \mathbf{q}$. This equation forms the definition for the steady-state vector, and is a straightforward way to calculate it.

DEFINITION If P is a stochastic matrix, then a **steady-state vector** (or **equilibrium vector** or **invariant probability vector**) for P is a probability vector $\mathbf{q}$ such that

$$P\mathbf{q} = \mathbf{q}$$

Exercises 36 and 37 will show that every stochastic matrix P has a steady-state vector $\mathbf{q}$. Notice that 1 must be an eigenvalue of any stochastic matrix, and the steady-state vector is a probability vector which is also an eigenvector of P associated with the eigenvalue 1.

Although the definition of the steady-state vector makes the calculation of $\mathbf{q}$ straightforward, it has a major drawback: there are Markov chains which have a steady-state vector $\mathbf{q}$ but for which $\lim_{n\to\infty} \mathbf{x}_n \neq \mathbf{q}$: the definition is not sufficient for $\mathbf{x}_n$ to converge. Examples 3-5 below will show different ways in which $\mathbf{x}_n$ can fail to converge – later in this section the conditions under which $\lim_{n\to\infty} \mathbf{x}_n = \mathbf{q}$ will be restated. For now, consider what $\mathbf{q}$ means when $\lim_{n\to\infty} \mathbf{x}_n = \mathbf{q}$, as it does in the example above. When $\lim_{n\to\infty} \mathbf{x}_n = \mathbf{q}$ there are two ways to interpret this vector.

- Since $\mathbf{x}_n$ is approximately equal to $\mathbf{q}$ for large n,the entries in $\mathbf{q}$ approximate the probability that the chain is in each state after n time steps. Thus in the above example, no matter the value of the initial probability vector, after many steps the probability that the chain is in state 1 is approximately $q_1 = .3$. Likewise the probability that the chain is in state 2 in the distant future is approximately $q_2 = .6$, and the probability that the chain is in state 3 in the distant future is approximately $q_3 = .1$. So the entries in $\mathbf{q}$ give **long-run probabilities**.

- When N is large, $\mathbf{q}$ approximates $\mathbf{x}_n$ for almost all values of $n \leq N$. Thus the entries in $\mathbf{q}$ approximate the proportion of time steps that the chain spends in each state. In the above example, the chain will end up spending $.3$ of the time steps in state 1, $.6$ of the time steps in state 2, and $.1$ of the time steps in state 2. So the entries in $\mathbf{q}$ give the proportion of the time steps spent in each state, which are called the **occupation times** for each state.

EXAMPLE 2 For an application of computing $\mathbf{q}$, consider the rat-in-the-maze example (Example 6, Section 10.1). In this example, the position of a rat in a five-room maze is modeled by a Markov chain with states $\{1, 2, 3, 4, 5\}$ and transition matrix

$$P = \begin{array}{c} \begin{array}{ccccc} \scriptstyle 1 & \scriptstyle 2 & \scriptstyle 3 & \scriptstyle 4 & \scriptstyle 5 \end{array} \\ \begin{bmatrix} 0 & 1/3 & 1/4 & 0 & 0 \\ 1/2 & 0 & 1/4 & 1/3 & 0 \\ 1/2 & 1/3 & 0 & 1/3 & 1/2 \\ 0 & 1/3 & 1/4 & 0 & 1/2 \\ 0 & 0 & 1/4 & 1/3 & 0 \end{bmatrix} \end{array} \begin{array}{c} \\ \scriptstyle 1 \\ \scriptstyle 2 \\ \scriptstyle 3 \\ \scriptstyle 4 \\ \scriptstyle 5 \end{array}$$

The steady-state vector may be computed by solving the system $P\mathbf{q} = \mathbf{q}$, which is equivalent to the homogeneous system $(P - I)\mathbf{q} = \mathbf{0}$. Row reduction gives

$$\begin{bmatrix} -1 & 1/3 & 1/4 & 0 & 0 & 0 \\ 1/2 & -1 & 1/4 & 1/3 & 0 & 0 \\ 1/2 & 1/3 & -1 & 1/3 & 1/2 & 0 \\ 0 & 1/3 & 1/4 & -1 & 1/2 & 0 \\ 0 & 0 & 1/4 & 1/3 & -1 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 & -3/2 & 0 \\ 0 & 0 & 1 & 0 & -2 & 0 \\ 0 & 0 & 0 & 1 & -3/2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

so a general solution is

$$q_5 \begin{bmatrix} 1 \\ 3/2 \\ 2 \\ 3/2 \\ 1 \end{bmatrix}$$

Letting q_5 be the reciprocal of the sum of the entries in the vector gives the steady-state vector

$$\mathbf{q} = \frac{1}{7}\begin{bmatrix} 1 \\ 3/2 \\ 2 \\ 3/2 \\ 1 \end{bmatrix} = \begin{bmatrix} 1/7 \\ 3/14 \\ 2/7 \\ 3/14 \\ 1/7 \end{bmatrix} \approx \begin{bmatrix} .142857 \\ .214286 \\ .285714 \\ .214286 \\ .142857 \end{bmatrix}$$

There are again two interpretations for $\mathbf{q}$: long-run probabilities and occupation times. After many moves, the probability that the rat will be in room 1 at a given time is approximately 1/7 no matter how where the rat began its journey. Put another way, the rat is expected to be in room 1 for 1/7 (or about 14.3%) of the time.

Again notice that taking high powers of the transition matrix P gives matrices whose columns are converging to $\mathbf{q}$; for example,

$$P^{10} = \begin{bmatrix} .144169 & .141561 & .142613 & .144153 & .142034 \\ .212342 & .216649 & .214286 & .211922 & .216230 \\ .285226 & .285714 & .286203 & .285714 & .285226 \\ .216230 & .211922 & .214286 & .216649 & .212342 \\ .142034 & .144153 & .142613 & .141561 & .144169 \end{bmatrix}$$

The columns of P^{10} are very nearly equal to each other, and each column is also nearly equal to $\mathbf{q}$.
■

Interpreting the Steady-State Vector

As noted above, every stochastic matrix will have a steady-state vector, but in some cases steady-state vectors cannot be interpreted as vectors of long-run probabilities or of occupation times. The following examples show some difficulties.

EXAMPLE 3 Consider an unbiased random walk on $\{1, 2, 3, 4, 5\}$ with absorbing boundaries. The transition matrix is

$$P = \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{bmatrix} 1 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 1 \end{bmatrix} \end{array} \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array}$$

Notice that only two long-term possibilities exist for this chain – it must end up in state 0 or state 4. Thus the probability that the chain is in states 1, 2 or 3 becomes smaller and smaller as n increases, as P^n illustrates:

$$P^{20} = \begin{bmatrix} 1 & .74951 & .49951 & .24951 & 0 \\ 0 & .00049 & 0 & .00049 & 0 \\ 0 & 0 & .00098 & 0 & 0 \\ 0 & .00049 & 0 & .00049 & 0 \\ 0 & .24951 & .49951 & .74951 & 1 \end{bmatrix}, P^{30} = \begin{bmatrix} 1 & .749985 & .499985 & .249985 & 0 \\ 0 & .000015 & 0 & .000015 & 0 \\ 0 & 0 & .000030 & 0 & 0 \\ 0 & .000015 & 0 & .000015 & 0 \\ 0 & .249985 & .499985 & .749985 & 1 \end{bmatrix}$$

It seems that P^n converges to the matrix

$$\begin{bmatrix} 1 & .75 & .5 & .25 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & .25 & .5 & .75 & 1 \end{bmatrix}$$

as n increases. But the columns of this matrix are not equal; the probability of ending up either at 1 or at 5 depends on where the chain begins. Although the chain has steady-state vectors, they cannot be interpreted as in Example 1. Exercise 23 confirms that if $0 \leq q \leq 1$ the vector

$$\begin{bmatrix} q \\ 0 \\ 0 \\ 0 \\ 1-q \end{bmatrix}$$

is a steady-state vector for P. This matrix then has an infinite number of possible steady-state vectors, which shows in another way that $\mathbf{x}_n$ cannot be expected to have convergent behavior which does not depend on $\mathbf{x}_0$. ■

EXAMPLE 4 Consider an unbiased random walk on $\{1, 2, 3, 4, 5\}$ with reflecting boundaries. The transition matrix is

$$\begin{array}{cc} & \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ P = & \begin{bmatrix} 0 & 1/2 & 0 & 0 & 0 \\ 1 & 0 & 1/2 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 1/2 & 0 & 1 \\ 0 & 0 & 0 & 1/2 & 0 \end{bmatrix} \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \end{array}$$

If the chain $\mathbf{x}_n$ starts at state 1, notice that it can return to 1 only when n is even, while the chain can be at state 2 only when n is odd. In fact, the chain must be at an even-numbered site when n is odd and at an odd-numbered site when n is even. If the chain were to start at state 2, however, this situation would be reversed: the chain must be at an odd-numbered site when n is odd and at an even-numbered site when n is even. Therefore, P^n cannot converge to a unique matrix since P^n looks very different depending on whether n is even or odd, as shown:

$$P^{20} = \begin{bmatrix} .2505 & 0 & .2500 & 0 & .2495 \\ 0 & .5005 & 0 & .4995 & 0 \\ .5000 & 0 & .5000 & 0 & .5000 \\ 0 & .4995 & 0 & .5005 & 0 \\ .2495 & 0 & .2500 & 0 & .2505 \end{bmatrix}, P^{21} = \begin{bmatrix} 0 & .2502 & 0 & .2498 & 0 \\ .5005 & 0 & .5000 & 0 & .4995 \\ 0 & .5000 & 0 & .5000 & 0 \\ .4995 & 0 & .5000 & 0 & .5005 \\ 0 & .2498 & 0 & .2502 & 0 \end{bmatrix}$$

Even though P^n does not converge to a unique matrix, P does have a steady-state vector. In fact,

$$\begin{bmatrix} 1/8 \\ 1/4 \\ 1/4 \\ 1/4 \\ 1/8 \end{bmatrix}$$

is a steady-state vector for P (see Exercise 32). This vector **can** be interpreted as giving long-run probabilities and occupation times in a sense that will be made precise in Section 10.4. ■

EXAMPLE 5 Consider a Markov chain on $\{1, 2, 3, 4, 5\}$ with transition matrix

$$P = \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{bmatrix} 1/4 & 1/3 & 1/2 & 0 & 0 \\ 1/4 & 1/3 & 1/4 & 0 & 0 \\ 1/2 & 1/3 & 1/4 & 0 & 0 \\ 0 & 0 & 0 & 1/3 & 3/4 \\ 0 & 0 & 0 & 2/3 & 1/4 \end{bmatrix} \end{array} \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array}$$

If this Markov chain begins in states 1, 2, or 3, then it must always be at one of those states. Likewise if the chain starts at states 4 or 5, then it must always be at one of those states. The chain splits into two separate chains, each with its own steady-state vector. In this case P^n converges to a matrix whose columns are not equal. The vectors

$$\begin{bmatrix} 4/11 \\ 3/11 \\ 4/11 \\ 0 \\ 0 \end{bmatrix} \text{ and } \begin{bmatrix} 0 \\ 0 \\ 0 \\ 9/17 \\ 8/17 \end{bmatrix}$$

both satisfy the definition of steady-state vector (Exercise 33). The first vector gives the limiting probabilities if the chain starts at states 1, 2, or 3, and the second does the same for the states 4 and 5. ■

Regular Matrices

Examples 1 and 2 show that in some cases a Markov chain $\mathbf{x}_n$ with transition matrix P has a steady-state vector $\mathbf{q}$ for which

$$\lim_{n\to\infty} P^n = \begin{bmatrix} \mathbf{q} & \mathbf{q} & \cdots & \mathbf{q} \end{bmatrix}$$

In these cases, $\mathbf{q}$ can be interpreted as a vector of long-run probabilities or occupation times for the chain. These probabilities or occupation times do not depend on the initial probability vector; that is, for any probability vector $\mathbf{x}_0$,

$$\lim_{n\to\infty} P^n\mathbf{x}_0 = \lim_{n\to\infty} \mathbf{x}_n = \mathbf{q}$$

Notice also that $\mathbf{q}$ is the only probability vector which is also an eigenvector of P associated with the eigenvalue 1.

Examples 3, 4, and 5 do not have such a steady-state vector $\mathbf{q}$. In Examples 3 and 5 the steady-state vector is not unique; in all three examples the matrix P^n does not converge to a matrix with equal columns as n increases. The goal is then to find some property of the transition matrix P that leads to these different behaviors, and to show that this property causes the differences in behavior.

A little calculation shows that in Examples 3, 4, and 5, every matrix of the form P^k has some zero entries. In Examples 1 and 2, however, some power of P has all positive entries. As was mentioned in Section 4.9, this is exactly the property that is needed.

DEFINITION A stochastic matrix P is **regular** if some power P^k contains only strictly positive entries.

Since the matrix P^k contains the probabilities of a k-step move from one state to another, a Markov chain with a regular transition matrix has the property that, for some k, it is possible to move from any state to any other in exactly k steps. The following theorem expands upon the content of Theorem 18 in Section 4.9. One idea must be defined before the theorem is presented. The limit of a sequence of $m \times n$ matrices is the $m \times n$ matrix (if one exists) whose (i, j) entry is the limit of the (i, j) entries in the sequence of matrices. With that understanding, here is the theorem.

THEOREM 1 If P is a regular $m \times m$ transition matrix with $m \geq 2$, then the following statements are all true.

(a) There is a stochastic matrix Π such that $\lim_{n\to\infty} P^n = \Pi$.

(b) Each column of Π is the same probability vector $\mathbf{q}$.

(c) For any initial probability vector $\mathbf{x}_0$, $\lim_{n\to\infty} P^n\mathbf{x}_0 = \mathbf{q}$.

(d) The vector $\mathbf{q}$ is the unique probability vector which is an eigenvector of P associated with the eigenvalue 1.

(e) All eigenvalues λ of P other than 1 have $|\lambda| < 1$.

A proof of Theorem 1 is given in Appendix 1. Theorem 1 is a special case of the Perron-Frobenius Theorem, which is used in applications of linear algebra to economics, graph theory, and systems analysis. Theorem 1 shows that a Markov chain with a regular transition matrix has the properties found in Examples 1 and 2. For example, since the transition matrix P in Example 1 is regular, Theorem 1 justifies the conclusion that P^n converges to a stochastic matrix all of whose columns equal $\mathbf{q} = \begin{bmatrix} .3 \\ .6 \\ .1 \end{bmatrix}$, as numerical evidence seemed to indicate.

PageRank and the Google Matrix

In Section 1, the notion of a simple random walk on a graph was defined. The World Wide Web can be modeled as a directed graph, with the vertices representing the webpages and the arrows representing the links between webpages. Let P be the huge transition matrix for this Markov chain. If the matrix P were regular, then Theorem 1 would show that there is a steady-state vector $\mathbf{q}$ for the chain, and that the entries in $\mathbf{q}$ can be interpreted as occupation times for each state. In terms of the model, the entries in $\mathbf{q}$ would tell what fraction of the random surfer's time was spent at each webpage. The founders of Google, Sergey Brin and Lawrence Page, reasoned that "important" pages had links coming from other "important" pages. Thus the random surfer would spend more time at more important pages and less time at less important pages. But the amount of time spent at each page is just the occupation time for each state in the Markov chain. This observation is the basis for PageRank, which is the model that Google uses to rank the importance of all webpages it catalogs:

> The importance of a webpage may be measured by the relative size of the corresponding entry in the steady-state vector $\mathbf{q}$ for an appropriately chosen Markov chain.

Unfortunately, simple random walk on the directed graph model for the Web is not the appropriate Markov chain, because the matrix P is not regular. Thus Theorem 1 will not apply. For example, consider the seven-page Web modeled in Section 10.1 using the directed graph in Figure 1. The transition matrix is

$$P = \begin{array}{c} \begin{array}{ccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{array} \\ \begin{bmatrix} 0 & 1/2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/3 & 0 & 1/2 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1/3 & 0 \\ 0 & 0 & 1/3 & 1 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 0 & 0 & 1/3 & 0 \\ 0 & 0 & 1/3 & 0 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1/3 & 1 \end{bmatrix} \end{array} \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{array}$$

Pages 4 and 7 are dangling nodes, and so are absorbing states for the chain. Just as in Example 3, the presence of absorbing states implies that the state vectors $\mathbf{x}_n$ do not approach a unique limit as $n \to \infty$. To handle dangling nodes, an adjustment is made to P:

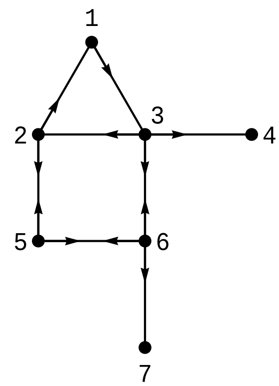


Figure 1: A seven-page Web.

ADJUSTMENT 1: If the surfer reaches a dangling node, the surfer will pick any page in the Web with equal probability and will move to that page. In terms of the transition matrix P, if state j is an absorbing state, replace column j of P with the vector

$$\begin{bmatrix} 1/n \\ 1/n \\ \vdots \\ 1/n \end{bmatrix}$$

where n is the number of rows (and columns in P).
In the seven-page example, the transition matrix is now

$$P_* = \begin{array}{c} \begin{array}{ccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{array} \\ \begin{bmatrix} 0 & 1/2 & 0 & 1/7 & 0 & 0 & 1/7 \\ 0 & 0 & 1/3 & 1/7 & 1/2 & 0 & 1/7 \\ 1 & 0 & 0 & 1/7 & 0 & 1/3 & 1/7 \\ 0 & 0 & 1/3 & 1/7 & 0 & 0 & 1/7 \\ 0 & 1/2 & 0 & 1/7 & 0 & 1/3 & 1/7 \\ 0 & 0 & 1/3 & 1/7 & 1/2 & 0 & 1/7 \\ 0 & 0 & 0 & 1/7 & 0 & 1/3 & 1/7 \end{bmatrix} \end{array} \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{array}$$

Yet even this adjustment is not sufficient to ensure that the transition matrix is regular: while dangling nodes are no longer possible, it is still possible to have "cycles" of pages. If page j linked only to page i and page i linked only to page j, a random surfer entering either page would be condemned to spend eternity linking from page i to page j and back again. Thus the columns of P_*^k corresponding to these pages would always have zeros in them, and the transition matrix P_* would not be regular. Another adjustment is needed.

ADJUSTMENT 2: Let p be a number between 0 and 1. Assume the surfer is now at page j. With probability p the surfer will pick from among all possible links from the page j with equal probability and will move to that page. With probability $1 - p$ the surfer will pick *any* page in the

Web with equal probability and will move to that page. In terms of the transition matrix P_*, the new transition matrix will be

$$G = pP_* + (1-p)K$$

where $K$is an $n \times n$ matrix all of whose columns are[3]

$$\begin{bmatrix} 1/n \\ 1/n \\ \vdots \\ 1/n \end{bmatrix}$$

The matrix G is called the **Google matrix**, and G is now a regular matrix since all entries in $G^1 = G$ are positive. Although any value of p between 0 and 1 is allowed, Google is said to use a value of $p = .85$ for their PageRank calculations. In the seven-page Web example, the Google matrix is thus

$$G = .85 \begin{bmatrix} 0 & 1/2 & 0 & 1/7 & 0 & 0 & 1/7 \\ 0 & 0 & 1/3 & 1/7 & 1/2 & 0 & 1/7 \\ 1 & 0 & 0 & 1/7 & 0 & 1/3 & 1/7 \\ 0 & 0 & 1/3 & 1/7 & 0 & 0 & 1/7 \\ 0 & 1/2 & 0 & 1/7 & 0 & 1/3 & 1/7 \\ 0 & 0 & 1/3 & 1/7 & 1/2 & 0 & 1/7 \\ 0 & 0 & 0 & 1/7 & 0 & 1/3 & 1/7 \end{bmatrix} + .15 \begin{bmatrix} 1/7 & 1/7 & 1/7 & 1/7 & 1/7 & 1/7 & 1/7 \\ 1/7 & 1/7 & 1/7 & 1/7 & 1/7 & 1/7 & 1/7 \\ 1/7 & 1/7 & 1/7 & 1/7 & 1/7 & 1/7 & 1/7 \\ 1/7 & 1/7 & 1/7 & 1/7 & 1/7 & 1/7 & 1/7 \\ 1/7 & 1/7 & 1/7 & 1/7 & 1/7 & 1/7 & 1/7 \\ 1/7 & 1/7 & 1/7 & 1/7 & 1/7 & 1/7 & 1/7 \\ 1/7 & 1/7 & 1/7 & 1/7 & 1/7 & 1/7 & 1/7 \end{bmatrix}$$

$$= \begin{bmatrix} .021429 & .446429 & .021429 & .142857 & .021429 & .021429 & .142857 \\ .021429 & .021429 & .304762 & .142857 & .446429 & .021429 & .142857 \\ .871429 & .021429 & .021429 & .142857 & .021429 & .304762 & .142857 \\ .021429 & .021429 & .304762 & .142857 & .021429 & .021429 & .142857 \\ .021429 & .446429 & .021429 & .142857 & .021429 & .304762 & .142857 \\ .021429 & .021429 & .304762 & .142857 & .446429 & .021429 & .142857 \\ .021429 & .021429 & .021429 & .142857 & .021429 & .304762 & .142857 \end{bmatrix}$$

It is now possible to find the steady-state vector by the methods of this section:

$$\mathbf{q} = \begin{bmatrix} .116293 \\ .168567 \\ .191263 \\ .098844 \\ .164054 \\ .168567 \\ .092413 \end{bmatrix}$$

so the most important page according to PageRank is page 3 which has the largest entry in $\mathbf{q}$. The complete ranking is 3, 2 and 6, 5, 1, 4, and 7.

[3]PageRank really uses a K which has all its columns equal to a probability vector $\mathbf{v}$ which could be linked to an individual searcher of group of searchers. This modification also makes it easier to police the Web for websites attempting to generate Web traffic. See *Google's PageRank and Beyond: the Science of Search Engine Rankings* by Amy N. Langville and Carl D. Meyer (Princeton: Princeton University Press, 2006) for more information.

NUMERICAL NOTE

The computation of $\mathbf{q}$ is not a trivial task, since the Google matrix has over 8 billion rows and columns. Google uses a version of the power method introduced in Section 5.8 to compute $\mathbf{q}$. While the power method was used in that section to estimate the eigenvalues of a matrix, it can also be used to provide estimates for eigenvectors. Since $\mathbf{q}$ is an eigenvector of G corresponding to the eigenvalue 1, the power method applies. It turns out that only between 50 and 100 iterations of the method are needed to get the vector $\mathbf{q}$ to the accuracy that Google needs for its rankings. It still takes days for Google to compute a new $\mathbf{q}$, which it does every month.

Practice Problem

1. Consider the Markov chain on $\{1, 2, 3\}$ with transition matrix

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

   a. Show that P is a regular matrix.

   b. Find the steady-state vector for this Markov chain.

   c. What fraction of the time does this chain spend in state 2? Explain your answer.

10.2 Exercises

In Exercises 1 and 2, consider a Markov chain on $\{1, 2\}$ with the given transition matrix P. In each exercise, use two methods to find the probability that, in the long run, the chain is in state 1. First, raise P to a high power. Then directly compute the steady-state vector.

1. $P = \begin{bmatrix} .2 & .4 \\ .8 & .6 \end{bmatrix}$

2. $P = \begin{bmatrix} .95 & .05 \\ .05 & .95 \end{bmatrix}$

In Exercises 3 and 4, consider a Markov chain on $\{1, 2, 3\}$ with the given transition matrix P. In each exercise, use two methods to find the probability that, in the long run, the chain is in state 1. First, raise P to a high power. Then directly compute the steady-state vector.

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

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

In Exercises 5 and 6, find the matrix to which P^n converges as n increases.

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

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

In Exercises 7 and 8, determine whether the given matrix is regular. Explain your answer.

7. $P = \begin{bmatrix} 1/3 & 0 & 1/2 \\ 1/3 & 1/2 & 1/2 \\ 1/3 & 1/2 & 0 \end{bmatrix}$

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

9. Consider a pair of Ehrenfest urns with a total of 8 molecules divided between them.

   a. Find the transition matrix for the Markov chain which models the number of molecules in Urn A, and show that this matrix is not regular.

   b. Assuming that the steady-state vector may be interpreted as occupation times for this Markov chain, in what state will this chain spend the most steps?

10. Consider a pair of Ehrenfest urns with a total of 7 molecules divided between them.

    a. Find the transition matrix for the Markov chain which models the number of molecules in Urn A, and show that this matrix is not regular.

    b. Assuming that the steady-state vector may be interpreted as occupation times for this Markov chain, in what state will this chain spend the most steps?

11. Consider unbiased random walk with reflecting boundaries on $\{1, 2, 3, 4, 5, 6\}$.

    a. Find the transition matrix for the Markov chain and show that this matrix is not regular.

    b. Assuming that the steady-state vector may be interpreted as occupation times for this Markov chain, in what state will this chain spend the most steps?

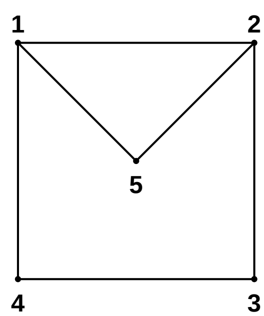
12. Consider biased random walk with reflecting boundaries on $\{1, 2, 3, 4, 5, 6\}$ with $p = 2/3$.

    a. Find the transition matrix for the Markov chain and show that this matrix is not regular.

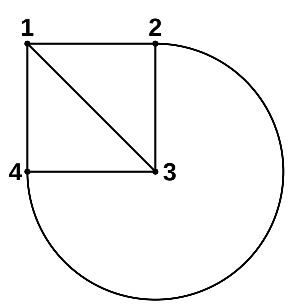
    b. Assuming that the steady-state vector may be interpreted as occupation times for this Markov chain, in what state will this chain spend the most steps?

In Exercises 13 and 14, consider a simple random walk on the graph given. In the long run, what fraction of the time is the walk at the various states?

13.

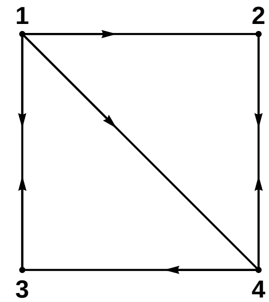


14.

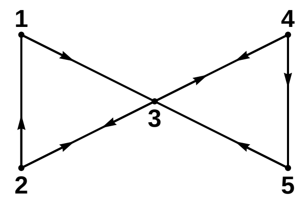


In Exercises 15 and 16, consider a simple random walk on the directed graph given. In the long run, what fraction of the time is the walk at the various states?

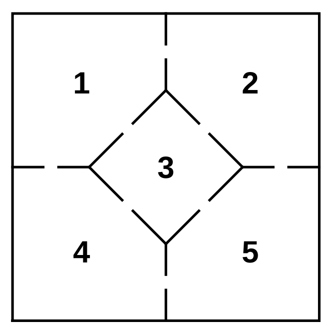
15.



16.

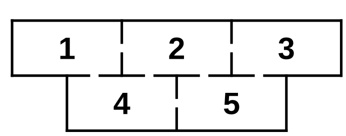


17. Consider the mouse in the following maze from Section 1, Exercise 17.



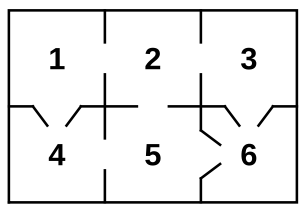
The mouse must move into a different room at each time step, and is equally likely to leave the room through any of the available doorways. If you go away from the maze for a while, what is the probability that the mouse is in room 3 when you return?

18. Consider the mouse in the following maze from Section 1, Exercise 18.



What fraction of time does it spend in room 3?

19. Consider the mouse in the following maze that includes “one-way” doors from Section 1, Exercise 19.

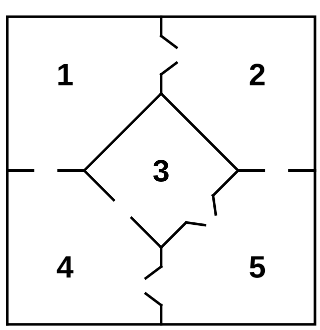


Show that

$$\mathbf{q} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

is a steady-state vector for the associated Markov chain, and interpret this result in terms of the mouse’s travels through the maze.

20. Consider the mouse in the following maze that includes “one-way” doors.



What fraction of time does it spend in each of the rooms in the maze?

In Exercises 21 and 22, mark each statement True or False. Justify each answer.

21. a. Every stochastic matrix has a steady-state vector.

    b. If its transition matrix is regular, then the steady-state vector gives information on long-run probabilities of the Markov chain.

    c. If $\lambda = 1$ is an eigenvalue of a matrix P, then P is regular.

22. a. Every stochastic matrix is regular.

    b. If P is a regular stochastic matrix, then P^n approaches a matrix with equal columns as n increases.

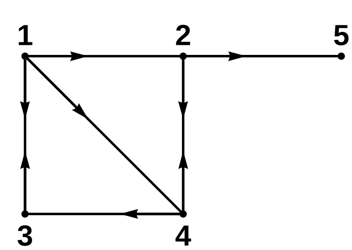
    c. If $\lim_{n\to\infty} \mathbf{x}_n = \mathbf{q}$, then the entries in $\mathbf{q}$ may be interpreted as occupation times.

23. Suppose that the weather in Charlotte is modeled using the Markov chain in Section 1, Exercise 17. Over the course of a year, about how many days in Charlotte are sunny, cloudy, and rainy according to the model?

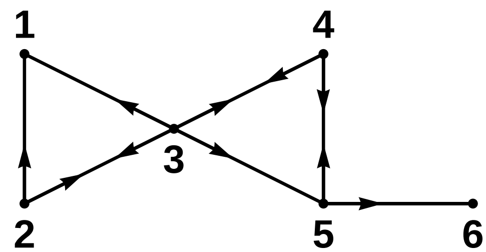
24. Suppose that the weather in Charlotte is modeled using the Markov chain in Section 1, Exercise 18. Over the course of a year, about how many days in Charlotte are rainy according to the model?

In Exercises 25 and 26, consider a set of webpages hyperlinked by the given directed graph. Find the Google matrix for each graph and compute the PageRank of each page in the set.

25.



26.



27. A genetic trait is often governed by a pair of genes, one inherited from each parent. The genes may be of two types, often labelled A and a. An individual then may have three different pairs: AA, Aa (which is the same as aA), or aa. In many cases the AA and Aa individuals cannot be otherwise distinguished – in these cases gene A is *dominant* and gene a is *recessive*. Likewise an AA individual is called *dominant* and an aa individual is called *recessive*. An Aa individual is called a *hybrid*.

   a. Show that if a dominant individual is mated with a hybrid, the probability of an offspring being dominant is $1/2$ and the probability of an offspring being a hybrid is $1/2$.

   b. Show that if a recessive individual is mated with a hybrid, the probability of an offspring being recessive is $1/2$ and the probability of an offspring being a hybrid is $1/2$.

   c. Show that if a hybrid individual is mated with a hybrid, the probability of an offspring being dominant is $1/4$, the probability of an offspring being recessive is $1/4$, and the probability of an offspring being a hybrid is $1/2$.

28. Consider beginning with an individual of known type and mating it with a hybrid, then mating an offspring of this mating with a hybrid, and so on. At each step an offspring is mated with a hybrid. The type of the offspring can be modelled by a Markov chain with states AA, Aa, and aa.

   a. Find the transition matrix for this Markov chain.

   b. If the mating process of the previous Exercise is continued for a extended period of time, what percent of the offspring are of each type?

29. Consider the variation of the Ehrenfest urn model of diffusion studied in Section 1, Exercise 29, where one of the $2k$ molecules is chosen at random and is then moved between the urns with a fixed probability p.

   a. Let $k = 3$ and suppose that $p = 1/2$. Show that the transition matrix for the Markov chain that models the number of molecules in Urn A is regular.

   b. Let $k = 3$ and suppose that $p = 1/2$. In what state will this chain spend the most steps, and what fraction of the steps will the chain spend at this state?

   c. Does the answer to part b. change if a different value of p with $0 < p < 1$ is used?

30. Consider the Bernoulli-Laplace of diffusion studied in Section 1, Exercise 30.

   a. Let $k = 5$ and show that the transition matrix for the Markov chain that models the number of Type I molecules in Urn A is regular.

   b. Let $k = 5$. In what state will this chain spend the most steps, and what fraction of the steps will the chain spend at this state?

31. Let $0 \leq q \leq 1$. Show that $\begin{bmatrix} q \\ 0 \\ 0 \\ 0 \\ 1-q \end{bmatrix}$ is a steady-state vector for the Markov chain in Example 3.

32. Consider the Markov chain in Example 4.

    a. Show that $\begin{bmatrix} 1/8 \\ 1/4 \\ 1/4 \\ 1/4 \\ 1/8 \end{bmatrix}$ is a steady-state vector for this Markov chain.

    b. Compute the average of the entries in P^{20} and P^{21} given in Example 4. What do you find?

33. Show that $\begin{bmatrix} 4/11 \\ 3/11 \\ 4/11 \\ 0 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 0 \\ 0 \\ 9/17 \\ 8/17 \end{bmatrix}$ are steady-state vectors for the Markov chain in Example 5. If the chain is equally likely to begin in each of the states, what is the probability of being in state 1 after many steps?

34. Let $0 \leq p, q \leq 1$, and define

$$P = \begin{bmatrix} p & 1-q \\ 1-p & q \end{bmatrix}$$

    a. Show that 1 and $p+q-1$ are eigenvalues of P.

    b. By Theorem 1, for what values of p and q will P fail to be regular?

    c. Find a steady-state vector for P.

35. Let $0 \leq p, q \leq 1$, and define

$$P = \begin{bmatrix} p & q & 1-p-q \\ q & 1-p-q & p \\ 1-p-q & p & q \end{bmatrix}$$

    a. For what values of p and q is P a regular stochastic matrix?

    b. Given that P is regular, find a steady-state vector for P.

36. Let A be an $m \times m$ non-negative matrix, $\mathbf{x}$ be in $\mathbb{R}^m$, and $\mathbf{y} = A\mathbf{x}$. Show that

$$|y_1| + \ldots + |y_m| \leq |x_1| + \ldots + |x_m|$$

with equality holding if and only if all of the nonzero entries in $\mathbf{x}$ have the same sign.

37. Show that every stochastic matrix has a steady-state vector using the following steps.

    a. Let P be a stochastic matrix. By Exercise 30 in Section 4.9, $\lambda = 1$ is an eigenvalue for P. Let $\mathbf{v}$ be an eigenvector of P associated with $\lambda = 1$. Use Exercise 36 to conclude that the nonzero entries in $\mathbf{v}$ must have the same sign.

    b. Show how to produce a steady-state vector for P from $\mathbf{v}$.

38. Consider simple random walk on a finite connected graph. (A graph is connected if it is possible to move from any vertex of the graph to any other along the edges of the graph).

    a. Explain why this Markov chain must have a regular transition matrix.

    b. Use the results of Exercises 13 and 14 to hypothesize a formula for the steady-state vector for such a Markov chain.

39. By Theorem 1 (e) all eigenvalues λ of a regular matrix other than 1 have the property that $|\lambda| < 1$; that is, the eigenvalue 1 is a *strictly dominant eigenvalue*. Suppose that P is an $n \times n$ regular matrix with

eigenvalues $\lambda_1 = 1, \ldots, \lambda_n$ ordered so that $|\lambda_1| > |\lambda_2| \geq |\lambda_3| \geq \ldots \geq |\lambda_n|$. Suppose that $\mathbf{x}_0$ is a linear combination of eigenvectors of P.

a. Use Equation (2) in Section 5.8 to derive an expression for $\mathbf{x}_k = P^k\mathbf{x}_0$.

b. Use the result of part (a) to derive an expression for $\mathbf{x}_k - \mathbf{q}$, and explain how the value of $|\lambda_2|$ effects the speed with which $\{\mathbf{x}_k\}$ converges to $\mathbf{q}$.

Solutions to Practice Problem

1. a. Since

$$P^2 = \begin{bmatrix} 1/4 & 1/4 & 1/2 \\ 1/2 & 1/4 & 1/4 \\ 1/4 & 1/2 & 1/4 \end{bmatrix}$$

P is regular by the definition with $k = 2$.

b. Solve the equation $P\mathbf{q} = \mathbf{q}$, which may be re-written as $(P - I)\mathbf{q} = \mathbf{0}$. Since

$$P - I = \begin{bmatrix} -1/2 & 0 & 1/2 \\ 1/2 & -1/2 & 0 \\ 0 & 1/2 & -1/2 \end{bmatrix}$$

Row reducing the augmented matrix gives

$$\begin{bmatrix} -1/2 & 0 & 1/2 & 0 \\ 1/2 & -1/2 & 0 & 0 \\ 0 & 1/2 & -1/2 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

so the general solution is $q_3 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$. Since $\mathbf{q}$ must be a probability vector, set $q_3 = 1/(1+1+1) = 1/3$ and compute that

$$\mathbf{q} = \frac{1}{3}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \end{bmatrix}$$

c. The chain will spend $1/3$ of its time in state 2 since the entry in $\mathbf{q}$ corresponding to state 2 is $1/3$, and we can interpret the entries as occupation times.

10.3 Communication Classes

Section 10.2 showed that if the transition matrix for a Markov chain is regular, then $\mathbf{x}_n$ converges to a unique steady-state vector for any choice of initial probability vector. That is, $\lim_{n\to\infty} \mathbf{x}_n = \mathbf{q}$, where $\mathbf{q}$ is the unique steady-state vector for the Markov chain. Examples 3, 4, and 5 of Section 10.2 illustrated that, even though every Markov chain has a steady-state vector, not every Markov chain has the property that $\lim_{n\to\infty} \mathbf{x}_n = \mathbf{q}$. The goal of the next two sections is to study these examples further, and to show that Examples 3, 4, and 5 of Section 10.2 describe all the ways in which Markov chains fail to converge to a steady-state vector. The first step is to study which states of the Markov chain can be reached from other states of the chain.

Communicating States

Suppose that state j and state i are two states of a Markov chain. If the state j can be reached from the state i in a finite number of steps and the state i can be reached from the state j in a finite number of steps, then the states j and i are said to **communicate**. If P is the transition matrix for the chain, then the entries in P^k give the probabilities of going from one state to another in k steps:

$$\begin{array}{cc} \text{From:} & \\ \begin{array}{cccc} 1 & j & \quad & m \end{array} & \text{To:} \\ P^k = \left[\begin{array}{cccc} & \vdots & & \\ & \downarrow & & \\ & p_{ij} & \cdots & \rightarrow \\ & & & \end{array}\right] & \begin{array}{c} 1 \\ \\ i \\ m \end{array} \end{array}$$

and powers of P can be used to make the following definition.

DEFINITION Let i and j be two states of a Markov chain with transition matrix P. Then state i **communicates** with state j if there exist nonnegative integers m and n such that the (j, i) entry of P^m and the (i, j) entry of P^n are both strictly positive. That is, state i communicates with state j if it is possible to go from state i to state j in m steps and from state j to state i in n steps.

This definition implies three properties that will allow the states of a Markov chain to be placed into groups called **communication classes**. First, the definition allows the integers m and n to be zero, in which case the (i, i) entry of $P^0 = I$ is 1, which is positive. This insures that every state communicates with itself. Because both (i, j) and (j, i) are included in the definition, it follows that if state i communicates with state j then state j communicates with state i. Finally, you will show in Exercise 36 that if state i communicates with state j and state j communicates with state k, then state i communicates with state k. These three properties are called respectively the **reflexive**, **symmetric**, and **transitive** properties:

(a) (Reflexive Property) Each state communicates with itself.

(b) (Symmetric Property) If state i communicates with state j, then state j communicates with state i.

(c) (Transitive Property) If state i communicates with state j, and state j communicates with state k, then state i communicates with state k.

A relation with these three properties is called an **equivalence relation**. The communication relation is an equivalence relation on the state space for the Markov chain. Using the properties listed above simplifies determining which states communicate.

EXAMPLE 1 Consider an unbiased random walk with absorbing boundaries on $\{1, 2, 3, 4, 5\}$. Find which states communicate.

Solution The transition matrix P is given below along with the transition diagram for this Markov chain:

$$P = \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{bmatrix} 1 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 1 \end{bmatrix} \end{array} \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array}$$

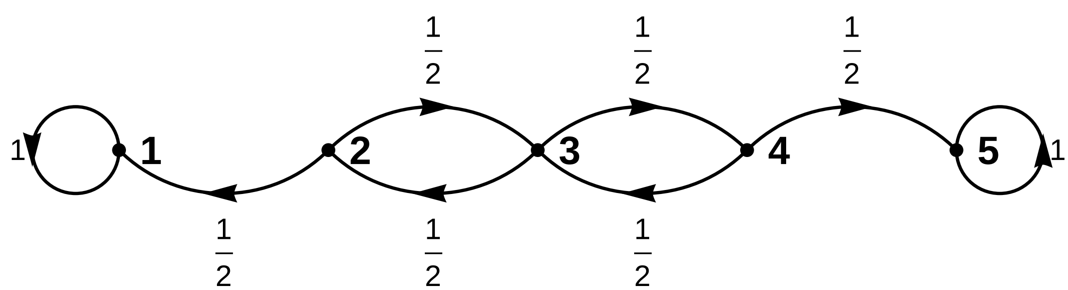


Figure 1: Unbiased random walk with absorbing boundaries.

First note by the reflexive property each state communicates with itself. It is clear from the diagram that states 2, 3 and 4 communicate with each other. The same conclusion may be reached using the definition by finding that the $(2, 3)$, $(3, 2)$, $(3, 4)$, and $(4, 3)$ entries in P are positive, thus states 2 and 3 communicate, as do states 3 and 4. States 2 and 4 must also communicate by the transitive property. Now consider state 1 and state 5. If the chain starts in state 1, it cannot move to any state other than itself. Thus it is not possible to go from state 1 to any other state in any number of steps, and state 1 does not communicate with any other state. Likewise state 5 does not communicate with any other state. In summary,

State 1 communicates with state 1.
State 2 communicates with state 2, state 3, and state 4.
State 3 communicates with state 2, state 3, and state 4.
State 4 communicates with state 2, state 3, and state 4.
State 5 communicates with state 5.

Notice that even though the states 1 and 5 do not communicate with states 2, 3 and 4, it is possible to go **from** these states either state 1 or state 5 in a finite number of steps: this is clear from the diagram, or by confirming that the appropriate entries in P, P^2, or P^3 are positive. ■

In Example 1 the state space $\{1, 2, 3, 4, 5\}$ can now be divided into the classes $\{1\}$, $\{2, 3, 4\}$, and $\{5\}$. The states in each of these classes communicate only with the other members of their class. This division of the state space occurs because the communication relation is an equivalence relation. The communication relation **partitions** the state space into **communication classes**. Each state in a Markov chain communicates only with the members of its communication class. For the Markov chain in Example 1, the communication classes are $\{1\}$, $\{2, 3, 4\}$, and $\{5\}$.

EXAMPLE 2 Consider an unbiased random walk with reflecting boundaries on $\{1, 2, 3, 4, 5\}$. Find the communication classes for this Markov chain.

Solution The transition matrix P for this chain, as well as P^2, P^3, and P^4, are shown below.

$$P = \begin{array}{cc} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} & \\ \begin{bmatrix} 0 & 1/2 & 0 & 0 & 0 \\ 1 & 0 & 1/2 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 1/2 & 0 & 1 \\ 0 & 0 & 0 & 1/2 & 0 \end{bmatrix} & \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \end{array}, \quad P^2 = \begin{array}{cc} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} & \\ \begin{bmatrix} 1/2 & 0 & 1/4 & 0 & 0 \\ 0 & 3/4 & 0 & 1/4 & 0 \\ 1/2 & 0 & 1/2 & 0 & 1/2 \\ 0 & 1/4 & 0 & 3/4 & 0 \\ 0 & 0 & 1/4 & 0 & 1/2 \end{bmatrix} & \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \end{array}$$

$$P^3 = \begin{array}{cc} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} & \\ \begin{bmatrix} 0 & 3/8 & 0 & 1/8 & 0 \\ 3/4 & 0 & 1/2 & 0 & 1/4 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 1/4 & 0 & 1/2 & 0 & 3/4 \\ 0 & 1/8 & 0 & 3/8 & 0 \end{bmatrix} & \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \end{array}, \quad P^4 = \begin{array}{cc} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} & \\ \begin{bmatrix} 3/8 & 0 & 1/4 & 0 & 1/8 \\ 0 & 5/8 & 0 & 3/8 & 0 \\ 1/2 & 0 & 1/2 & 0 & 1/2 \\ 0 & 3/8 & 0 & 5/8 & 0 \\ 1/8 & 0 & 1/4 & 0 & 3/8 \end{bmatrix} & \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \end{array}$$

The transition diagram for this Markov chain is given in Figure 2.

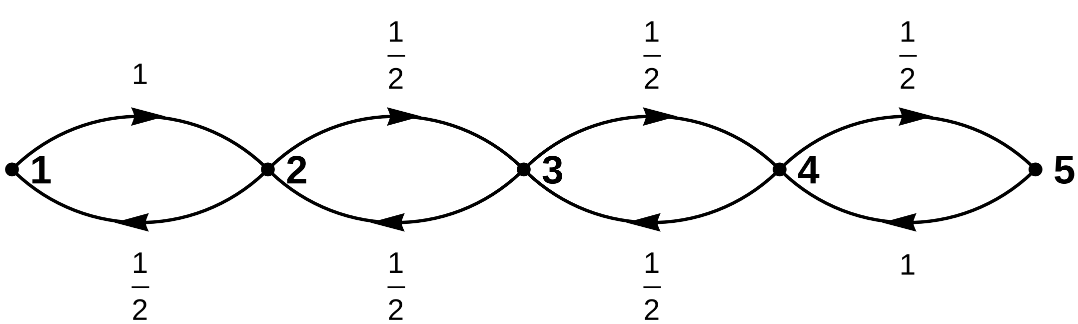


Figure 2: Unbiased random walk with reflecting boundaries.

Notice that the (i, j) entry in at least one of these matrices is positive for any choice of i and j. Thus every state is reachable from any other state in 4 steps or fewer, and every state communicates with every state. There is only one communication class: $\{1, 2, 3, 4, 5\}$. ■

EXAMPLE 3 Consider the Markov chain given in Example 5 of Section 10.2. Find the communication classes for this Markov chain.

Solution The transition matrix for this Markov chain is

$$P = \begin{array}{c} \\ \\ \\ \\ \\ \end{array}\begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{bmatrix} 1/4 & 1/3 & 1/2 & 0 & 0 \\ 1/4 & 1/3 & 1/4 & 0 & 0 \\ 1/2 & 1/3 & 1/4 & 0 & 0 \\ 0 & 0 & 0 & 1/3 & 3/4 \\ 0 & 0 & 0 & 2/3 & 1/4 \end{bmatrix} \end{array} \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array}$$

and a transition diagram is

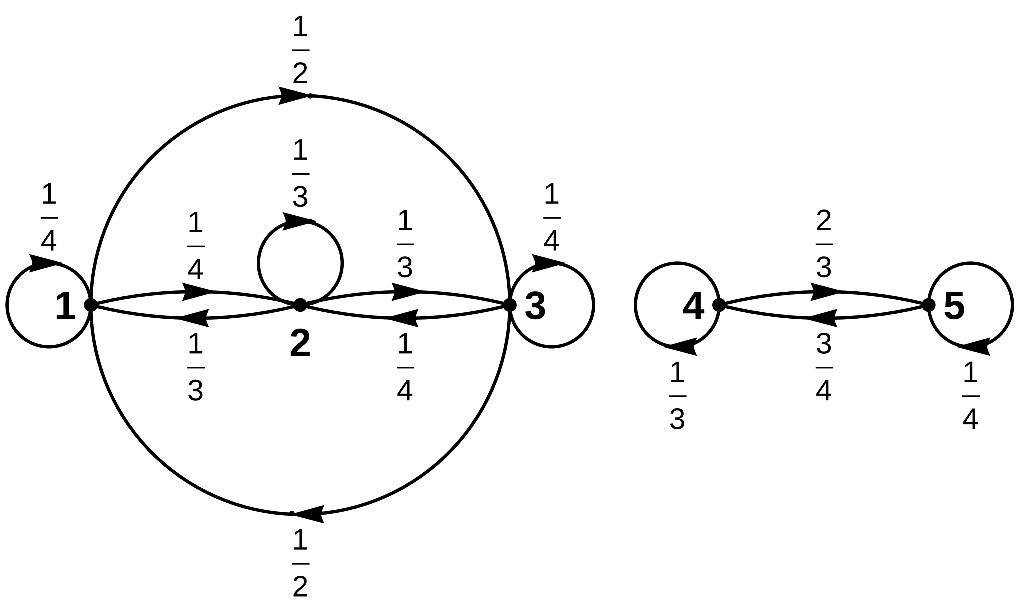


Figure 3: Transition diagram for Example 3.

It is impossible to move from any of the states 1, 2, or 3 to either of the states 4 or 5, so these states must be in separate communication classes. In addition, state 1, state 2, and state 3 communicate; state 4 and state 5 also communicate. Thus the communication classes for this Markov chain are $\{1, 2, 3\}$ and $\{4, 5\}$. ■

The Markov chains in Examples 1 and 3 have more than one communication class, while the Markov chain in Example 2 has only one communication class. This distinction leads to the following definitions.

DEFINITION A Markov chain with only one communication class is **irreducible**. A Markov chain with more than one communication class is **reducible**.

Thus the Markov chains in Examples 1 and 3 are reducible, while the Markov chain in Example 2 is irreducible. Irreducible Markov chains and regular transition matrices are connected by the following theorem.

THEOREM 2 If a Markov chain has a regular transition matrix, then it is irreducible.

Proof Suppose that P is a regular transition matrix for a Markov chain. Then by definition, there is a k such that P^k is a positive matrix. That is, for any states i and j, the (i, j) and (j, i) elements in P^k are strictly positive. Thus there is a positive probability of moving from i to j and from j to i in exactly k steps, and so i and j communicate with each other. Since i and j are any states and must be in the same communication class, there can be only one communication class for the chain, so the Markov chain must be irreducible. ■

Example 2 shows that the converse of Theorem 2 is not true, because the Markov chain in this example is irreducible, but its transition matrix is not regular.

EXAMPLE 4 Consider the Markov chain whose transition diagram is given in Figure 4. Determine whether this Markov chain is reducible or irreducible.

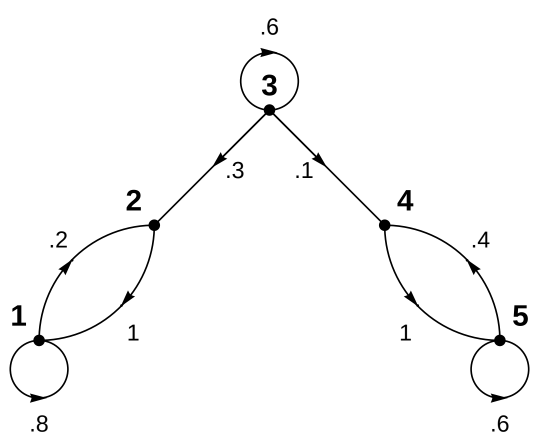


Figure 4: Transition diagram for Example 4.

Solution The diagram shows that states 1 and 2 communicate, as do states 4 and 5. Notice that states 1 and 2 cannot communicate with states 3, 4, or 5 since the probability of moving from state 2 to state 3 is 0. Likewise states 4 and 5 cannot communicate with states 1, 2, or 3 since the probability of moving from state 4 to state 3 is 0. Finally, state 3 cannot communicate with any state other than itself since it is impossible to return to state 3 from any other state. Thus the communication classes for this Markov chain are $\{1, 2\}$, $\{3\}$, and $\{4, 5\}$. Since there is more than one communication class, this Markov chain is reducible. ■

Mean Return Times

Let $\mathbf{q}$ be the steady-state vector for an irreducible Markov chain. It can be shown using advanced methods in probability theory that the entries in $\mathbf{q}$ may be interpreted as occupation times; that is, q_i is the fraction of time steps that the chain will spend at state i. For example, consider a Markov chain on $\{1, 2, 3\}$ with steady-state vector $\mathbf{q} = \begin{bmatrix} .2 \\ .5 \\ .3 \end{bmatrix}$. In the long run the chain will spend about half of its steps in state 2. If the chain is currently at state 2, it should take about $2 = 1/.5$ steps to return to state 2. Likewise since the chain spends about $1/5$ of its time in state 1, it should visit state 1 once every 5 steps.

Given a Markov chain and states i and j, a quantity of considerable interest is the number of steps n_{ij} that it will take for the system to first visit state i given that it is started in state j. The value of n_{ij} cannot be known – it could be any positive integer depending on how the Markov chain evolves. Such a quantity is known as a **random variable**. Since n_{ij} is unknowable, the **expected value** of n_{ij} is studied instead. The expected value of a random variable functions as a type of average value of the random variable. The following definition will be used in subsequent sections.

DEFINITION The **expected value of a random variable** X which takes on the values $x_1, x_2, \ldots$ is

$$E[X] = x_1 P(X = x_1) + x_2 P(X = x_2) + \cdots = \sum_{k=1}^{\infty} x_k P(X = x_k) \tag{1}$$

where $P(X = x_k)$ denotes the probability that the random variable X equals the value x_k.

Now let $t_{ii} = E[n_{ii}]$ be the expected value of n_{ii}, which is the expected number of steps it will take for the system to return to state i given that it starts in state i. Unfortunately, Equation 1 will not be helpful at this point. Instead proceeding intuitively, the system should spend 1 step at state i for each t_{ii} steps on average. It seems reasonable to say that the system will, over the long run, spend about $1/t_{ii}$ of the time at state i. But that quantity is q_i, so the expected time steps needed to return, or **mean return time** to a state i, is the reciprocal of q_i. This informal argument may be made rigorous using methods from probability theory; see Appendix 2 for a complete proof.

THEOREM 3 Consider an irreducible Markov chain with a finite state space, let n_{ij} be the number of steps until the chain first visits state i given that the chain starts in state j, and let $t_{ii} = E[n_{ii}]$. Then

$$t_{ii} = \frac{1}{q_i} \tag{2}$$

where q_i is the entry in the steady-state vector $\mathbf{q}$ corresponding to state i.

The above example matches Equation 2: $t_{11} = 1/.2 = 5$, $t_{22} = 1/.5 = 2$, and $t_{33} = 1/.3 = 10/3$. Recall that the mean return time is a expected value, so the fact that t_{33} is not an integer ought not be troubling. Section 10.5 will include a discussion of $t_{ij} = E[n_{ij}]$ where $i \neq j$.

Practice Problem

1. Consider the Markov chain on $\{1, 2, 3, 4\}$ with transition matrix

$$P = \begin{bmatrix} 1/4 & 1/3 & 1/2 & 0 \\ 0 & 1/3 & 0 & 1/3 \\ 3/4 & 0 & 1/2 & 1/3 \\ 0 & 1/3 & 0 & 1/3 \end{bmatrix}$$

   Determine the communication classes for this chain.

10.3 Exercises

In Exercises 1-6, consider a Markov chain with state space with $\{1, 2, \dots, n\}$ and the given transition matrix. Find the communication classes for each Markov chain, and state whether the Markov chain is reducible or irreducible.

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

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

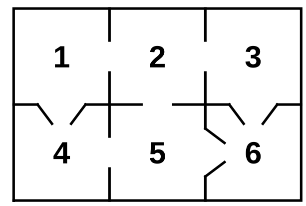
3. $\begin{bmatrix} 1 & 1/2 & 1/2 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1/2 \end{bmatrix}$

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

5. $\begin{bmatrix} 0 & 0 & .4 & 0 & .8 & 0 \\ 0 & 0 & 0 & .7 & 0 & .5 \\ .3 & 0 & 0 & 0 & .2 & 0 \\ 0 & .1 & 0 & 0 & 0 & .5 \\ .7 & 0 & .6 & 0 & 0 & 0 \\ 0 & .9 & 0 & .3 & 0 & 0 \end{bmatrix}$

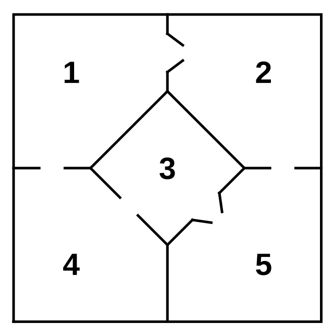
6. $\begin{bmatrix} 0 & 1/3 & 0 & 2/3 & 1/2 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 & 1/3 & 0 \\ 0 & 2/3 & 0 & 1/3 & 0 & 0 & 2/5 \\ 1/2 & 0 & 1/2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 3/5 \\ 0 & 0 & 0 & 0 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2/3 & 0 \end{bmatrix}$

7. Consider the mouse in the following maze from Section 1, Exercise 19.



Find the communication classes for the Markov chain that models the mouse's travels through this maze. Is this Markov chain reducible or irreducible?

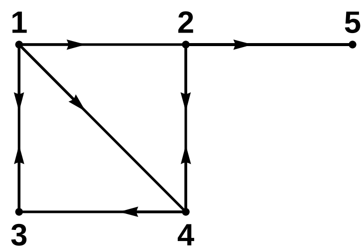
8. Consider the mouse in the following maze from Section 1, Exercise 20.



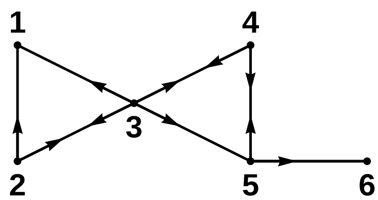
Find the communication classes for the Markov chain that models the mouse's travels through this maze. Is this Markov chain reducible or irreducible?

In Exercises 9 and 10, consider the set of webpages hyperlinked by the given directed graph. Find the communication classes for the Markov chain that models a random surfer's progress through this set of webpages.

9.



10.

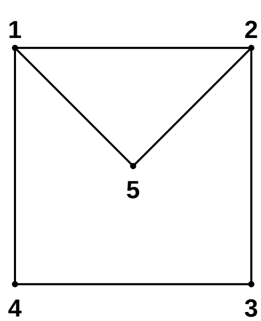


11. Consider unbiased random walk with reflecting boundaries on $\{1, 2, 3, 4, 5, 6\}$. Find the communication classes for this Markov chain and determine whether it is reducible or irreducible.

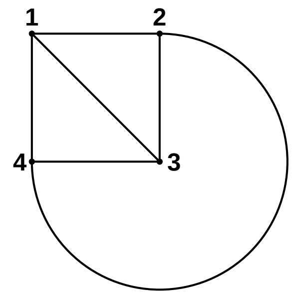
12. Consider unbiased random walk with absorbing boundaries on $\{1, 2, 3, 4, 5, 6\}$. Find the communication classes for this Markov chain and determine whether it is reducible or irreducible.

In Exercises 13 and 14, consider a simple random walk on the graph given. Show that the Markov chain is irreducible and calculate the mean return times for each state.

13.

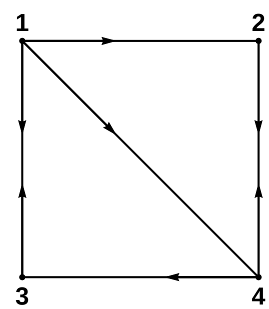


14.

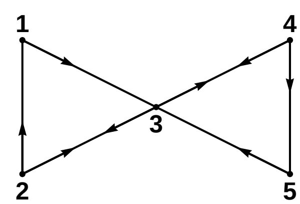


In Exercises 15 and 16, consider a simple random walk on the directed graph given. Show that the Markov chain is irreducible and calculate the mean return times for each state.

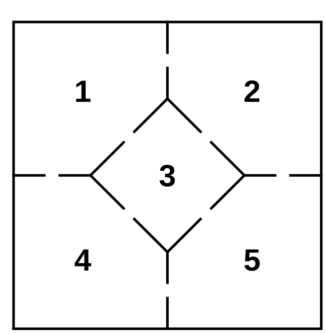
15.



16.

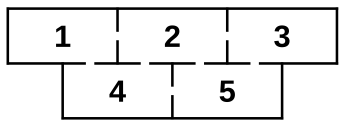


17. Consider the mouse in the following maze from Section 1, Exercise 17.



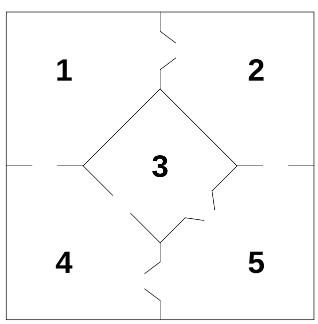
If the mouse starts in room 3, how long on average will it take the mouse to return to room 3?

18. Consider the mouse in the following maze from Section 1, Exercise 18.



If the mouse starts in room 2, how long on average will it take the mouse to return to room 2?

In Exercises 19 and 20, consider the mouse in the following maze from Section 2, Exercise 20.



19. If the mouse starts in room 1, how long on average will it take the mouse to return to room 1?

20. If the mouse starts in room 4, how long on average will it take the mouse to return to room 4?

In Exercises 21 and 22, mark each statement True or False. Justify each answer.

21. a. If it is possible to go from state i to state j in n steps, where $n \geq 0$, then states i and j communicate with each other.

b. If a Markov chain is reducible, then it cannot have a regular transition matrix.

c. The entries in the steady-state vector are the mean return times for each state.

22. a. An irreducible Markov chain must have a regular transition matrix.

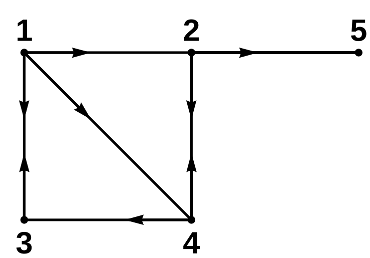
b. If the (i, j) and (j, i) entries in P^k are positive for some k, then the states i and j communicate with each other.

c. If state i communicates with state j and state j communicates with state k, then state i communicates with state k.

23. Suppose that the weather in Charlotte is modeled using the Markov chain in Section 1, Exercise 23. About how many days elapse in Charlotte between rainy days?

24. Suppose that the weather in Charlotte is modeled using the Markov chain in Section 1, Exercise 24. About how many days elapse in Charlotte between consecutive rainy days?

25. The following set of web pages hyperlinked by the directed graph was studied in Section 2, Exercise 25.

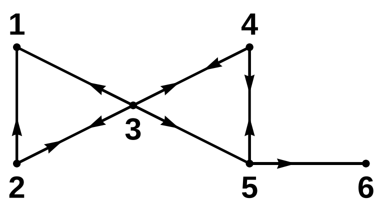


Consider randomly surfing on this set of web pages using the Google matrix as the transition matrix.

a. Show that this Markov chain is irreducible.

b. Suppose the surfer starts at page 1. How many mouse clicks on average must the surfer make to get back to page 1?

26. The following set of web pages hyperlinked by the directed graph that was studied in Section 2, Exercise 26.



Repeat Exercise 25 for this set of web pages.

27. Consider the pair of Ehrenfest urns studied in Section 2, Exercise 9. Suppose that there are now 4 molecules in Urn A. How many steps on average will be needed until there are again 4 molecules in Urn A?

28. Consider the pair of Ehrenfest urns studied in Section 2, Exercise 10. Suppose that Urn A is now empty. How many steps on average will be needed until Urn A is again empty?

29. A variation of the Ehrenfest model of diffusion studied in Section 2, Exercise 29. Consider this model with $k = 3$ and $p = 1/2$ and suppose that there are now 3 molecules in Urn A. How many draws on average will be needed until there are again 3 molecules in Urn A?

30. Consider the Bernoulli-Laplace model of diffusion studied in Section 2, Exercise 30. Let $k = 5$. Suppose that all of the Type I molecules are now in Urn A. How many draws on average will be needed until all of the Type I molecules are again in Urn A?

31. A Markov chain model for scoring a tennis game was studied in Section 1, Exercise 31. What are the communication classes for this Markov chain?

32. A Markov chain model for the rally point method for scoring a volleyball game was studied in Section 1, Exercise 32. What are the communication classes for this Markov chain?

In Exercises 33 and 34, consider the Markov chain on $\{1, 2, 3, 4, 5\}$ with transition matrix

$$P = \begin{bmatrix} 0 & 0 & 0 & 1/2 & 1 \\ 1/3 & 0 & 0 & 0 & 0 \\ 2/3 & 0 & 0 & 0 & 0 \\ 0 & 2/5 & 1/5 & 1/2 & 0 \\ 0 & 3/5 & 4/5 & 0 & 0 \end{bmatrix}$$

33. Show that this Markov chain is irreducible.

34. Suppose the chain starts in state 1. What is the expected number of steps until it is in state 1 again?

35. How does the presence of dangling nodes in a set of hyperlinked webpages affect the communication classes of the associated Markov chain?

36. Show that the communication relation is transitive. **Hint**: Show that the (i, k)-entry of P^{n+m} must be greater than or equal to the product of the (i, j)-entry of P^m and the (j, k)-entry of P^n.

Solution to Practice Problem

1. First note that states 1 and 3 communicate with each other, as do states 2 and 4. However, there is no way to proceed from either 1 or 3 to either 2 or 4, so the communication classes are $\{1, 3\}$ and $\{2, 4\}$.

10.4 Classification of States and Periodicity

The communication classes of a Markov chain have important properties which help determine whether the state vectors converge to a unique steady-state vector. These properties are studied in this section, and it will be shown that Examples 3, 4 and 5 in Section 10.2 are examples of all the ways that the state vectors of a Markov chain can fail to converge to a unique steady-state vector.

Recurrent and Transient States

One way to describe the communication classes is to determine whether it is possible for the Markov chain to leave the class once it has entered it.

DEFINITION Let C be a communication class of states for a Markov chain, and let j be a state in C. If there is a state i not in C and $k > 0$ such that the (i, j) entry in P^k is positive, then the class C is called a **transient class**, and each state in C is a **transient state**. If a communication class is not transient, it is called a **recurrent class** and each state in the class is a **recurrent state**.

Suppose that C is a transient class. Notice that once the system moves from C to a another communication class D, the system can never return to C. This is true because D cannot contain a state i from which it is possible to move to a state in C. If D did contain such a state i, then the transitive property of the communication relation would imply that every state in C communicates with every state in D. This is impossible.

EXAMPLE 1 Consider the Markov chain on $\{1, 2, 3, 4, 5\}$ studied in Example 4 of Section 10.3. Its transition diagram is given in Figure 1. Determine whether each of the communication classes is transient or recurrent.

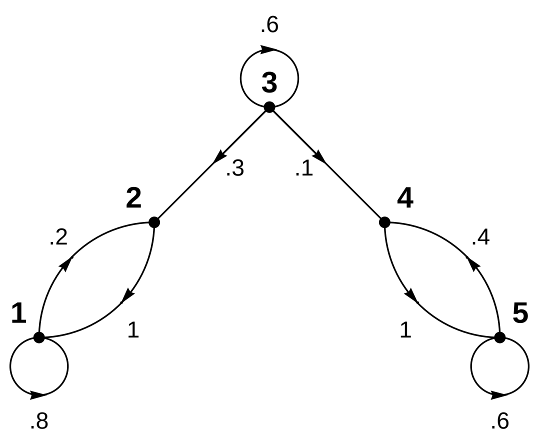


Figure 1: Transition diagram for Example 1.

Solution The communication classes were found to be $\{1, 2\}$, $\{3\}$, and $\{4, 5\}$. First consider the class $\{3\}$. There is a positive probability of a transition from state 3 to state 2, so applying the definition with $k = 1$ shows that $\{3\}$ is a transient class. Now consider $\{1, 2\}$. The probability of a one-step transition from either state 1 or 2 to any of states 3, 4, or 5 is zero, and this is also true for any number of steps. If the system starts in state 1 or 2, it will always stay in state 1 or 2. The class $\{1, 2\}$ is thus a recurrent class. A similar argument shows that $\{4, 5\}$ is also a recurrent class. ■

EXAMPLE 2 Consider the random walk with reflecting boundaries studied in Example 2 in Section 10.3. Determine whether each of the communication classes is transient or recurrent.

Solution This Markov chain is irreducible: the single communication class for this chain is $\{1, 2, 3, 4, 5\}$. By the definition, this class cannot be transient. Thus the communication class must be recurrent. ■

The result of the preceding example may be generalized to any irreducible Markov chain.

REMARK All states of an irreducible Markov chain are recurrent.

Suppose that a reducible Markov chain has two transient classes C_1 and C_2 and no recurrent classes. Since C_1 is transient, there must be a state in C_2 which can be reached from a state in C_1. Since C_2 is transient, there must be a state in C_1 which can be reached from C_2. Thus all states in C_1 and C_2 communicate, which is impossible. Thus the Markov chain must have at least one recurrent class. This argument can be generalized to refer to any reducible Markov chain with any number of transient classes, which along the previous remark proves the following.

REMARK Every Markov chain must have at least one recurrent class.

EXAMPLE 3 Consider the Markov chain studied in Example 3 of Section 10.3. Determine whether each of the communication classes is transient or recurrent.

Solution The transition matrix for this Markov chain is

$$P = \begin{array}{c} \begin{matrix} 1 & \;\;2 & \;\;3 & \;\;4 & \;\;5 \end{matrix} \\ \begin{bmatrix} 1/4 & 1/3 & 1/2 & 0 & 0 \\ 1/4 & 1/3 & 1/4 & 0 & 0 \\ 1/2 & 1/3 & 1/4 & 0 & 0 \\ 0 & 0 & 0 & 1/3 & 3/4 \\ 0 & 0 & 0 & 2/3 & 1/4 \end{bmatrix} \end{array} \begin{matrix} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{matrix}$$

and the two communication classes are $\{1, 2, 3\}$ and $\{4, 5\}$. The matrix P may be written as the partitioned matrix $P = \begin{bmatrix} P_1 & O \\ O & P_2 \end{bmatrix}$, where

$$P_1 = \begin{array}{c} \begin{matrix} 1 & \;\;2 & \;\;3 \end{matrix} \\ \begin{bmatrix} 1/4 & 1/3 & 1/2 \\ 1/4 & 1/3 & 1/4 \\ 1/2 & 1/3 & 1/4 \end{bmatrix} \end{array} \begin{matrix} \\ 1 \\ 2 \\ 3 \end{matrix} \text{ and } P_2 = \begin{array}{c} \begin{matrix} 4 & \;\;5 \end{matrix} \\ \begin{bmatrix} 1/3 & 3/4 \\ 2/3 & 1/4 \end{bmatrix} \end{array} \begin{matrix} \\ 4 \\ 5 \end{matrix}$$

and O is an appropriately sized zero matrix. Using block multiplication,

$$P^k = \begin{bmatrix} P_1^k & O \\ O & P_2^k \end{bmatrix}$$

for all $k > 0$. Thus if state j is in one class and state i is in the other, the (i, j) and (j, i) entries of P^k are zero for all $k > 0$. Thus both classes of this Markov chain must be recurrent. ■

EXAMPLE 4 Consider altering the previous example slightly to get a Markov chain with transition matrix

$$\begin{array}{cc} & \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ P = & \begin{bmatrix} 1/4 & 1/3 & 1/2 & 0 & 0 \\ 1/4 & 1/3 & 1/4 & 0 & 0 \\ 1/2 & 1/3 & 1/4 & 0 & 1/4 \\ 0 & 0 & 0 & 1/3 & 1/2 \\ 0 & 0 & 0 & 2/3 & 1/4 \end{bmatrix} \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \end{array}$$

and transition diagram given in Figure 2. Determine whether each of the communication classes is transient or recurrent.

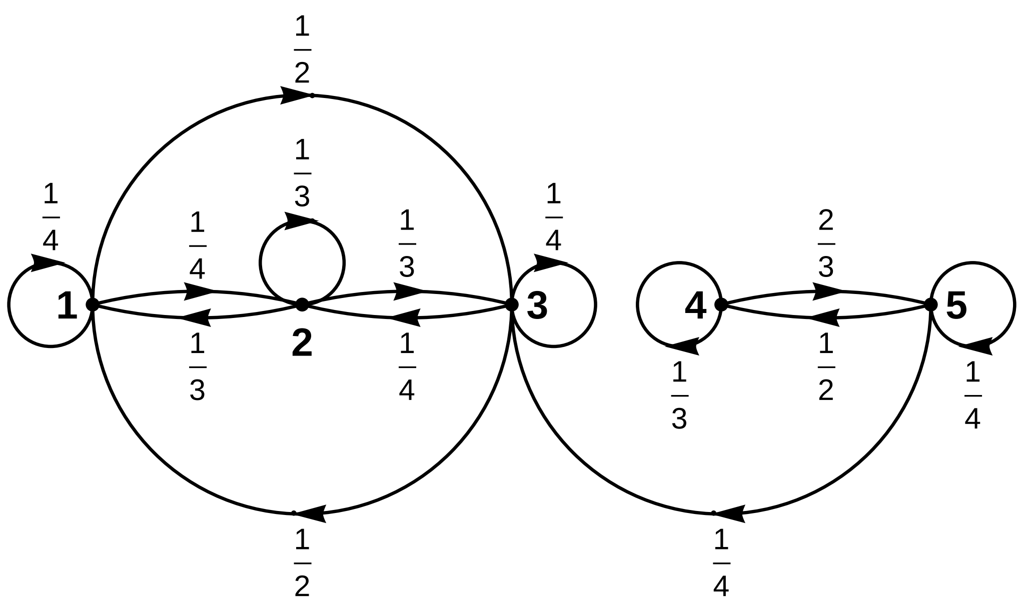


Figure 2: Transition diagram for Example 4.

Solution The communication classes are still $\{1, 2, 3\}$ and $\{4, 5\}$. Now the $(5, 3)$ entry is not zero, so $\{4, 5\}$ is a transient class. By the above remark the chain must have at least one recurrent class, so $\{1, 2, 3\}$ must be that recurrent class. This result may also be proven using partitioned matrices.

Let $P = \begin{bmatrix} P_1 & S \\ O & Q \end{bmatrix}$, where P_1 is as in the previous example,

$$Q = \begin{array}{c} \\ \\ \end{array}\begin{array}{cc} 4 & 5 \\ \end{array}$$

$$Q = \begin{array}{cc|c} 4 & 5 & \\ \hline 1/3 & 1/2 & 4 \\ 2/3 & 1/4 & 5 \end{array} \quad \text{and } S = \begin{array}{cc|c} 4 & 5 & \\ \hline 0 & 0 & 1 \\ 0 & 0 & 2 \\ 0 & 1/4 & 3 \end{array}$$

The submatrix P_1 is a transition matrix in its own right: it describes transitions within the recurrent class $\{1, 2, 3\}$. The matrix S records the probabilities of transitions from the transient class $\{4, 5\}$ into the recurrent class $\{1, 2, 3\}$. The matrix Q records the probabilities of transitions within the transient class $\{4, 5\}$. Block multiplication (see Section 2.4) now gives

$$P^k = \begin{bmatrix} P_1^k & S_k \\ O & Q^k \end{bmatrix}$$

for some non-zero matrix S_k. Since the lower left block is O for all matrices P^k, it is impossible to leave the class $\{1, 2, 3\}$ after entering it, and $\{1, 2, 3\}$ is a recurrent class. ■

In Examples 3 and 4, the states were ordered so that the members of each class were grouped together. In Example 4, the recurrent classes were listed first followed by the transient classes. This ordering was convenient, as it allowed for the use of partitioned matrices to determine the recurrent and transient classes. It is also possible to use block multiplication to compute powers of the transition matrix P if the states are reordered in the manner done in Examples 3 and 4: the states in each communication class are consecutive, and if there are any transient classes, the recurrent classes are listed first, followed by the transient classes. A transition matrix with its states thus reordered is be said to be in **canonical form**. To see how this reordering works, consider the following example.

EXAMPLE 5 The Markov chain in Example 1 has transition matrix

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

and its communication classes are $\{1, 2\}$, $\{3\}$, and $\{4, 5\}$. To place the matrix in canonical form, reorder the classes $\{1, 2\}$, $\{4, 5\}$, and $\{3\}$; that is, rearrange the states in the order 1, 2, 4, 5, 3. To perform this rearrangement, first rearrange the columns, which produces the matrix

$$\begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{bmatrix} .8 & 1 & 0 & 0 & 0 \\ .2 & 0 & .3 & 0 & 0 \\ 0 & 0 & .6 & 0 & 0 \\ 0 & 0 & .1 & 0 & .4 \\ 0 & 0 & 0 & 1 & .6 \end{bmatrix} \end{array} \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \xrightarrow[\text{columns}]{\text{rearrange}} \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 4 & 5 & 3 \end{array} \\ \begin{bmatrix} .8 & 1 & 0 & 0 & 0 \\ .2 & 0 & 0 & 0 & .3 \\ 0 & 0 & 0 & 0 & .6 \\ 0 & 0 & 0 & .4 & .1 \\ 0 & 0 & 1 & .6 & 0 \end{bmatrix} \end{array} \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array}$$

Now rearranging the rows produces the transition matrix in canonical form:

$$\begin{array}{cc} \begin{array}{ccccc} 1 & 2 & 4 & 5 & 3 \end{array} & \\ \begin{bmatrix} .8 & 1 & 0 & 0 & 0 \\ .2 & 0 & 0 & 0 & .3 \\ 0 & 0 & 0 & 0 & .6 \\ 0 & 0 & 0 & .4 & .1 \\ 0 & 0 & 1 & .6 & 0 \end{bmatrix} & \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \end{array} \xrightarrow[\text{rows}]{\text{rearrange}} \begin{array}{cc} \begin{array}{ccccc} 1 & 2 & 4 & 5 & 3 \end{array} & \\ \begin{bmatrix} .8 & 1 & 0 & 0 & 0 \\ .2 & 0 & 0 & 0 & .3 \\ 0 & 0 & 0 & .4 & .1 \\ 0 & 0 & 1 & .6 & 0 \\ 0 & 0 & 0 & 0 & .6 \end{bmatrix} & \begin{array}{c} 1 \\ 2 \\ 4 \\ 5 \\ 3 \end{array} \end{array}$$

The transition matrix may be divided as follows:

$$P = \begin{array}{cc} \begin{array}{ccccc} 1 & 2 & 4 & 5 & 3 \end{array} & \\ \left[\begin{array}{cc|ccc} .8 & 1 & 0 & 0 & 0 \\ .2 & 0 & 0 & 0 & .3 \\ \hline 0 & 0 & 0 & .4 & .1 \\ 0 & 0 & 1 & .6 & 0 \\ 0 & 0 & 0 & 0 & .6 \end{array}\right] & \begin{array}{c} 1 \\ 2 \\ 4 \\ 5 \\ 3 \end{array} \end{array} = \left[\begin{array}{c|c} P_1 & S \\ \hline O & Q \end{array}\right]$$

In general, suppose that P is the transition matrix for a reducible Markov chain with r recurrent classes and one or more transient classes. A canonical form of P is

$$P = \left[\begin{array}{ccc|c} P_1 & \cdots & O & \\ \vdots & \ddots & \vdots & S \\ O & \cdots & P_r & \\ \hline & O & & Q \end{array}\right]$$

Here P_i is the transition matrix for the i^{th} recurrent class, O is an appropriately sized zero matrix, Q records transitions within the transient classes, and S contains the probabilities of transitions from the transient classes to the recurrent classes. Since P is a partitioned matrix, it is relatively easy to take powers of it using block multiplication:

$$P^k = \left[\begin{array}{ccc|c} P_1^k & \cdots & O & \\ \vdots & \ddots & \vdots & S_k \\ O & \cdots & P_r^k & \\ \hline & O & & Q^k \end{array}\right]$$

for some matrix S_k. The matrices Q, S, and S_k help to answer questions about the long-term behavior of the Markov chain which are addressed in Section 10.5.

Periodicity

A final way of classifying states is to examine at what times it is possible for the system to return to the state in which it begins. Consider the following simple example.

EXAMPLE 6 A Markov chain on $\{1,2,3\}$ has transition matrix

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

The transition diagram is quite straightforward:

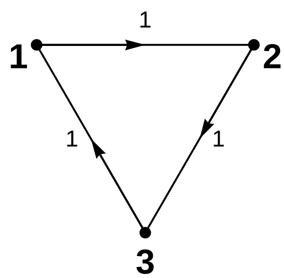


Figure 3: Transition diagram for Example 6.

The system must return to its starting point in three steps and every time the number of steps is a multiple of three.

EXAMPLE 7 A Markov chain on $\{1,2,3,4\}$ has transition matrix

$$P = \begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{bmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 1 \\ 0 & 1/2 & 0 & 0 \end{bmatrix} \end{array} \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array}$$

and transition diagram

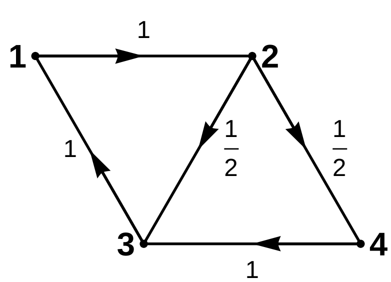


Figure 4: Transition diagram for Example 7.

If the system starts in states 1, 2, or 3, the system may return to its starting point in three steps or in four steps, and may return every time the number of steps is $3a + 4b$ for some non-negative integers a and b. It can be shown that every positive integer greater than 5 may be written in that form, so if the system starts in states 1, 2, or 3, it may also return to its starting point at any number of steps greater than 5. If the system starts in state 4, the system may return to its starting point in four steps or in seven steps, and a similar argument shows that the system may also return to its starting point at any number of steps greater than 17.

EXAMPLE 8 The unbiased random walk on $\{1, 2, 3, 4, 5\}$ with reflecting boundaries has transition diagram

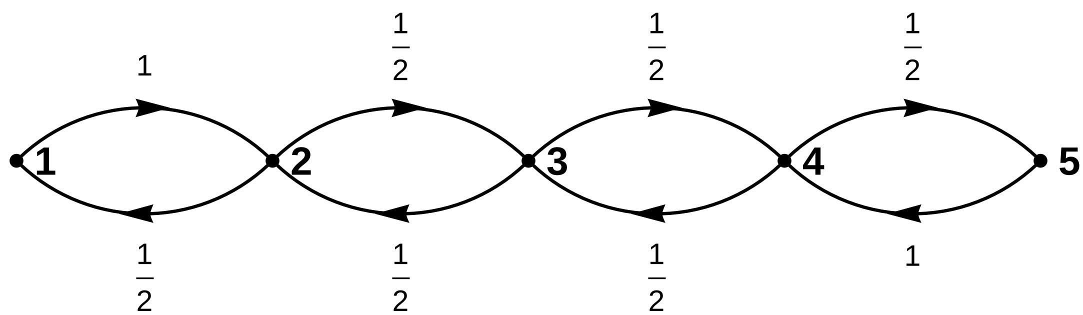


Figure 5: Unbiased random walk with reflecting boundaries.

From this diagram, one can see that it will always take an even number of steps for the system to return to the state in which it started.

In Examples 6 and 8, the time steps at which the system may return to its initial site may be are multiples of a number d: $d = 3$ for Example 6, while $d = 2$ for Example 8. This number d is called the *period* of the state, and is defined as follows.

DEFINITION The **period** d of a state i of a Markov chain is the greatest common divisor of all time steps n such that the probability that the Markov chain started at i visits i at time step n is strictly positive.

Using a careful analysis of the set of states visited by the Markov chain, it may be shown that the period of each state in a given communication class is the same, so the period is a property of communication classes. See Appendix 2 for a proof of this fact, which leads to the following definition.

DEFINITION The **period** of a communication class C is the period of each state in C. If a Markov chain is irreducible, then the period of the chain is the period of its single communication class. If the period of every communication class (and thus every state) is $d = 1$, then the Markov chain is **aperiodic**.

The reason that the greatest common divisor appears in the definition is to allow a period to be assigned to all states of all Markov chains. In Example 7, the system may return to its starting state after any sufficiently large number of steps, so the period of each state is $d = 1$. That is, the Markov chain in Example 7 is aperiodic. Notice that this chain does not exhibit periodic behavior, so the term aperiodic is quite apt. Using the definition confirms that the period of the Markov chain in Example 6 is $d = 3$, while the period of the Markov chain in Example 8 is $d = 2$. The next theorem describes the transition matrix of an irreducible and aperiodic Markov chain.

THEOREM 4 Let P be the transition matrix for an irreducible, aperiodic Markov chain. Then P is a regular matrix.

Proof Let P be an $n \times n$ transition matrix for an irreducible, aperiodic Markov chain. To show that P is regular, find a power P^k of P must be found for which every entry is strictly positive. Let $1 \leq i, j \leq n$. Since the Markov chain is irreducible, there must be a number a which depends on i and j such that the (i, j)-element in P^a is strictly positive. Since the Markov chain is aperiodic, there is a number b which depends on i such that the (i, i)-element in P^m is strictly positive for all $m \geq b$. Now note that since $P^{a+m} = P^a P^m$, the (i, j)-element in P^{a+m} must be greater than the product of the (i, j)-element in P^a and the (i, i)-element in P^m. Thus the (i, j)-element in P^{a+m} must be strictly positive for all $m \geq b$. Now let k be the maximum over all pairs (i, j) of the quantity $a + b$. This maximum exists because the state space is finite, and the (i, j)-element of P^k must be strictly positive for all pairs (i, j). Thus every entry of P^k is strictly positive, and P is a regular matrix. ■

So, if P is the transition matrix for an irreducible, aperiodic Markov chain, then P must be regular and Theorem 1 must apply to P. Thus there is a steady-state vector $\mathbf{q}$ for which

$$\lim_{n\to\infty} P^n \mathbf{x}_0 = \mathbf{q}$$

for any choice of initial probability vector $\mathbf{x}_0$. What can be said about the steady-state vector $\mathbf{q}$ if an irreducible Markov chain has period $d > 1$? The following result is proven in more advanced texts in probability theory.

THEOREM 5 Let P be the transition matrix for an irreducible Markov chain with period $d > 1$, and let $\mathbf{q}$ be the steady-state vector for the Markov chain. Then for any initial probability vector $\mathbf{x}_0$,

$$\lim_{n\to\infty} \frac{1}{d}\left(P^{n+1} + \cdots + P^{n+d}\right)\mathbf{x}_0 = \mathbf{q}$$

Theorem 5 says that in the case of an irreducible Markov chain with period $d > 1$, the vector $\mathbf{q}$ is the limit of the average of the probability vectors $P^{n+1}\mathbf{x}_0, P^{n+2}\mathbf{x}_0, \ldots, P^{n+d}\mathbf{x}_0$. When a Markov chain is irreducible with period $d > 1$, the vector $\mathbf{q}$ may still be interpreted as a vector of occupation times.

EXAMPLE 9 The period of the irreducible Markov chain in Example 8 is $d = 2$, so the Markov chain has period $d > 1$. Let n be an even integer. Taking high powers of the transition matrix P shows that

$$P^n \longrightarrow \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{bmatrix} 1/4 & 0 & 1/4 & 0 & 1/4 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 1/2 & 0 & 1/2 & 0 & 1/2 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 1/4 & 0 & 1/4 & 0 & 1/4 \end{bmatrix} \end{array} \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \text{ and } P^{n+1} \longrightarrow \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{bmatrix} 0 & 1/4 & 0 & 1/4 & 0 \\ 1/2 & 0 & 1/2 & 0 & 1/2 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 1/2 & 0 & 1/2 & 0 & 1/2 \\ 0 & 1/4 & 0 & 1/4 & 0 \end{bmatrix} \end{array} \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array}$$

So for any initial probability vector $\mathbf{x}_0$,

$$\lim_{n\to\infty} \frac{1}{2}\left(P^n + P^{n+1}\right)\mathbf{x}_0 = \begin{bmatrix} 1/8 & 1/8 & 1/8 & 1/8 & 1/8 \\ 1/4 & 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 & 1/4 \\ 1/8 & 1/8 & 1/8 & 1/8 & 1/8 \end{bmatrix} \mathbf{x}_0 = \begin{bmatrix} 1/8 \\ 1/4 \\ 1/4 \\ 1/4 \\ 1/8 \end{bmatrix}$$

But this vector was the steady-state vector for this Markov chain calculated in Exercise 32 of Section 10.2. Theorem 5 is thus confirmed in this case.

The steady-state vector for a *reducible* Markov chain will be discussed in detail in the next section.

Practice Problem

1. Consider the Markov chain on $\{1, 2, 3, 4\}$ with transition matrix

$$P = \begin{bmatrix} 1/4 & 1/3 & 1/2 & 0 \\ 0 & 1/3 & 0 & 1/3 \\ 3/4 & 0 & 1/2 & 1/3 \\ 0 & 1/3 & 0 & 1/3 \end{bmatrix}$$

Identify the communication classes of the chain as either recurrent or transient, and reorder the states to produce a matrix in canonical form.

10.4 Exercises

In Exercises 1-6, consider a Markov chain with state space with $\{1, 2, \ldots, n\}$ and the given transition matrix. Identify the communication classes for each Markov chain as recurrent or transient, and find the period of each communication class.

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

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

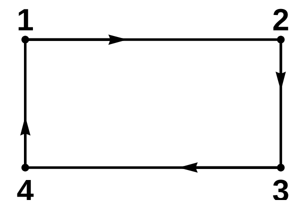
3. $\begin{bmatrix} 1 & 1/2 & 1/2 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1/2 \end{bmatrix}$

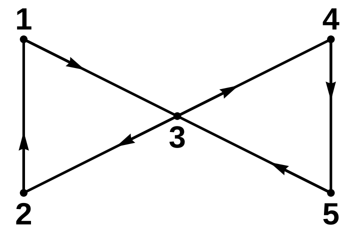
4. $\begin{bmatrix} 0 & 0 & 0 & 1 & 1 \\ 1/3 & 0 & 0 & 0 & 0 \\ 2/3 & 0 & 0 & 0 & 0 \\ 0 & 1/4 & 2/3 & 0 & 0 \\ 0 & 3/4 & 1/3 & 0 & 0 \end{bmatrix}$

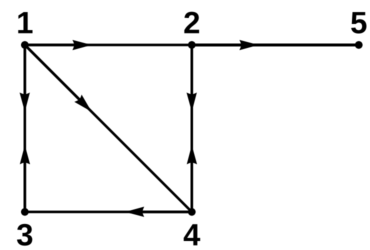
5. $\begin{bmatrix} 0 & 0 & .4 & 0 & .8 & 0 \\ 0 & 0 & 0 & .7 & 0 & .5 \\ .3 & 0 & 0 & 0 & .2 & 0 \\ 0 & .1 & 0 & 0 & 0 & .5 \\ .7 & 0 & .6 & 0 & 0 & 0 \\ 0 & .9 & 0 & .3 & 0 & 0 \end{bmatrix}$

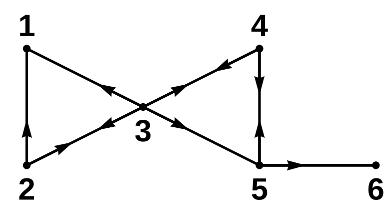
6. $\begin{bmatrix} 0 & 1/3 & 0 & 2/3 & 1/2 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 & 1/3 & 0 \\ 0 & 2/3 & 0 & 1/3 & 0 & 0 & 2/5 \\ 1/2 & 0 & 1/2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 3/5 \\ 0 & 0 & 0 & 0 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2/3 & 0 \end{bmatrix}$

In Exercises 7-10, consider a simple random walk on the given directed graph. Identify the communication classes as of this Markov chain as recurrent or transient, and find the period of each communication class.

7. 

8. 

9. 

10. 

11. Reorder the states in the Markov chain in Exercise 1 to produce a transition matrix in canonical form.

12. Reorder the states in the Markov chain in Exercise 2 to produce a transition matrix in canonical form.

13. Reorder the states in the Markov chain in Exercise 3 to produce a transition matrix in canonical form.

14. Reorder the states in the Markov chain in Exercise 4 to produce a transition matrix in canonical form.

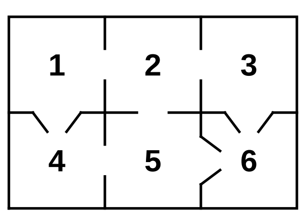
15. Reorder the states in the Markov chain in Exercise 5 to produce a transition matrix in canonical form.

16. Reorder the states in the Markov chain in Exercise 6 to produce a transition matrix in canonical form.

17. Find the transition matrix for the Markov chain in Exercise 9 and reorder the states to produce a transition matrix in canonical form.

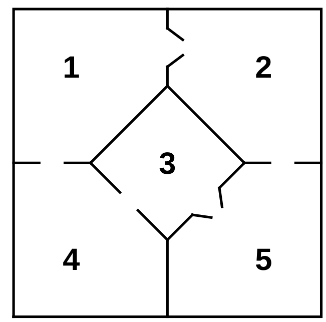
18. Find the transition matrix for the Markov chain in Exercise 10 and reorder the states to produce a transition matrix in canonical form.

19. Consider the mouse in the following maze from Section 1, Exercise 19.



   a. Identify the communication classes of this Markov chain as recurrent or transient.
   b. Find the period of each communication class.
   c. Find the transition matrix for the Markov chain and reorder the states to produce a transition matrix in canonical form.

20. Consider the mouse in the following maze from Section 1, Exercise 20.



   a. Identify the communication classes of this Markov chain as recurrent or transient.
   b. Find the period of each communication class.
   c. Find the transition matrix for the Markov chain and reorder the states to produce a transition matrix in canonical form.

In Exercises 21-22, mark each statement True or False. Justify each answer.

21. a. If two states i and j are both recurrent, then they must belong to the same communication class.
   b. All of the states in an irreducible Markov chain are recurrent.
   c. Every Markov chain must have at least one transient class.

22. a. If state i is recurrent and state i communicates with state j, then state j is also recurrent.
   b. If two states of a Markov chain have different periods, then the Markov chain is reducible.
   c. Every Markov chain must have exactly one recurrent class.

23. Confirm Theorem 5 for the Markov chain in Exercise 7 by taking powers of the transition matrix (see Example 9).

24. Confirm Theorem 5 for the Markov chain in Exercise 8 by taking high powers of the transition matrix (see Example 9).

25. Consider the Markov chain on $\{1, 2, 3\}$ with transition matrix

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

a. Explain why this Markov chain is irreducible and has period 2.

b. Find a steady-state vector $\mathbf{q}$ for this Markov chain.

c. Find an invertible matrix A and a diagonal matrix D such that $P = ADA^{-1}$. (See Section 5.3.)

d. Use the result from part (c) to show that P^n may be written as

$$\begin{bmatrix} 1/4 & 1/4 & 1/4 \\ 1/2 & 1/2 & 1/2 \\ 1/4 & 1/4 & 1/4 \end{bmatrix} + (-1)^n \begin{bmatrix} 1/4 & -1/4 & 1/4 \\ -1/2 & 1/2 & -1/2 \\ 1/4 & -1/4 & 1/4 \end{bmatrix}$$

e. Use the result from d. to confirm Theorem 5 for P.

26. Follow the plan of Exercise 25 to confirm Theorem 5 for the Markov chain with transition matrix

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

where $0 < p < 1$.

27. Confirm Theorem 5 for the Markov chain in Example 6.

28. Matrix multiplication can be used to find the canonical form of a transition matrix. Consider the matrix P in Example 5 and the matrix

$$E = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

Notice that the rows of E are the rows of the identity matrix in the order 1,2,4,5,3.

a. Compute EP and explain what has happened to the matrix P.

b. Compute PE^T and explain what has happened to the matrix P.

c. Compute EPE^T and explain what has happened to the matrix P.

29. Let A be an $n \times n$ matrix and let E be a $n \times n$ matrix resulting from permuting the rows of I_n, the $n \times n$ identity matrix. The matrix E is called a **permutation matrix**.

a. Show that EA is the matrix A with its rows permuted in exactly the same order that the rows of I_n were permuted to form E. **Hint:** Any permutation of rows can be written as a sequence of swaps of pairs of rows.

b. Apply the result of part a. to A^T to show that AE^T is the matrix A with its columns permuted in exactly the same order that the rows of I_n were permuted to form E.

c. Explain why EAE^T is the matrix A with its rows and columns permuted in exactly the same order that the rows of I_n were permuted to form E.

d. In the process of finding the canonical form of a transition matrix, does it matter whether the rows of the matrix or the columns of the matrix are permuted first? Why or why not?

Solution to Practice Problem

1. First note that states 1 and 3 communicate with each other, as do states 2 and 4. However, there is no way to proceed from either 1 or 3 to either 2 or 4, so the communication classes are $\{1,3\}$ and $\{2,4\}$. Since the chain stays in the class $\{1,3\}$ after it enters this class, the class $\{1,3\}$ is recurrent. Likewise, there is a positive probability of leaving the class $\{2,4\}$ at any time, so the class $\{2,4\}$ is transient. One ordering of the states that produces a canonical form is 1,3,2,4: the corresponding transition matrix is

$$P \xrightarrow[\text{columns}]{\text{rearrange}} \begin{array}{c} \begin{array}{cccc} 1 & 3 & 2 & 4 \end{array} \\ \begin{bmatrix} 1/4 & 1/2 & 1/3 & 0 \\ 0 & 0 & 1/3 & 1/3 \\ 3/4 & 1/2 & 0 & 1/3 \\ 0 & 0 & 1/3 & 1/3 \end{bmatrix} \end{array} \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \xrightarrow[\text{rows}]{\text{rearrange}} \begin{array}{c} \begin{array}{cccc} 1 & 3 & 2 & 4 \end{array} \\ \begin{bmatrix} 1/4 & 1/2 & 1/3 & 0 \\ 3/4 & 1/2 & 0 & 1/3 \\ 0 & 0 & 1/3 & 1/3 \\ 0 & 0 & 1/3 & 1/3 \end{bmatrix} \end{array} \begin{array}{c} \\ 1 \\ 3 \\ 2 \\ 4 \end{array}$$

10.5 The Fundamental Matrix

The return time for a state in an irreducible Markov chain was defined in Section 10.3 to be the expected number of steps needed for the system to return its starting state. This section studies the expected number of steps needed for a system to pass from one state to another state, which is called a transit time. Another quantity of interest is the probability that the system visits one state before it visits another. It is perhaps surprising that discussing these issues for irreducible Markov chains begins by working with reducible Markov chains, particularly those with transient states.

The Fundamental Matrix and Transient States

The first goal is to compute the expected number of visits the system makes to a state i given that the system starts in state j, where j is a transient state. Suppose that a Markov chain has at least one transient state. Its transition matrix may be written in canonical form as

$$P = \left[\begin{array}{c|c} R & S \\ \hline O & Q \end{array}\right]$$

Since at least one state is transient, S has at least one non-zero entry. In order for P to be a stochastic matrix at least one of the columns of Q must sum to less than one. The matrix Q is called a **substochastic matrix**. It can be shown that

$$\lim_{k\to\infty} Q^k = O$$

for any substochastic matrix Q. This fact implies that if the system is started in a transient class, it must eventually make a transition to a recurrent class and never visit any state outside that recurrent class again. The system is thus eventually **absorbed** by some recurrent class.

Now let j and i be transient states, and suppose that the Markov chain starts at state j. Let v_{ij} be the number of visits the system makes to state i before the absorption into a recurrent class. The goal is to calculate $E[v_{ij}]$, which is the expected value of v_{ij}. To do so a special kind of random variable called an **indicator random variable** is useful. An **indicator random variable** I is a random variable which is 1 if an event happens and is 0 if the event does not happen. Symbolically,

$$I = \begin{cases} 0 & \text{if the event does not happen} \\ 1 & \text{if the event happens} \end{cases}$$

The expected value of an indicator random variable may be easily calculated:

$$E[I] = 0 \cdot P(I=0) + 1 \cdot P(I=1) = P(I=1) = P(\text{event happens}) \tag{1}$$

Returning to the discussion of the number of visits to state i starting at state j, let I_k be the indicator random variable for the event "the system visits state i at step k." Then

$$v_{ij} = I_0 + I_1 + I_2 + \ldots = \sum_{k=0}^{\infty} I_k$$

since a visit to state i at a particular time will cause 1 to be added to the running total of visits kept in v_{ij}. Using Equation 1, the expected value of v_{ij} is

$$E[v_{ij}] = E\left[\sum_{k=0}^{\infty} I_k\right] = \sum_{k=0}^{\infty} E[I_k] = \sum_{k=0}^{\infty} P(I_k = 1) = \sum_{k=0}^{\infty} P(\text{visit to } i \text{ at step } k)$$

But $P(\text{visit to } i \text{ at step } k)$ is just the (i, j) entry in the matrix Q^k, so

$$E[v_{ij}] = \sum_{k=0}^{\infty} (Q^k)_{ij}$$

Thus the expected number of times that the system visits state i starting at state j is the (i, j)-entry in the matrix

$$I + Q + Q^2 + Q^3 + \ldots = \sum_{k=0}^{\infty} Q^k$$

Using the argument given in Section 2.6 (p.154-155),

$$I + Q + Q^2 + Q^3 + \ldots = (I - Q)^{-1}$$

The matrix $(I - Q)^{-1}$ is called the **fundamental matrix** of the Markov chain and is denoted by M. The interpretation of the entries in M is given in the following theorem.

THEOREM 6 Let j and i be transient states of a Markov chain, and let Q be that portion of the transition matrix which governs movement between transient states.

a. When the chain starts at a transient state j, the (i, j) entry of $M = (I - Q)^{-1}$ is the expected number of visits to the transient state i before absorption into a recurrent class.

b. When the chain starts at a transient state j, the sum of the entries in column j of $M = (I - Q)^{-1}$ is the expected number of time steps until absorption.

An alternative proof of Theorem 6 is given in Appendix 2.

EXAMPLE 1 Consider an unbiased random walk on $\{1, 2, 3, 4, 5\}$ with absorbing boundaries. If the system starts in state 3, ind the expected number of visits to state 2 before absorption. Also find the expected number of steps until absorption starting at the states 2, 3, and 4.

Solution Placing the states in the order 1, 5, 2, 3, 4 produces a transition matrix in canonical form:

$$\begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{bmatrix} 1 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 1 \end{bmatrix} \end{array} \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \xrightarrow[\text{columns}]{\text{rearrange}} \begin{array}{c} \begin{array}{ccccc} 1 & 5 & 2 & 3 & 4 \end{array} \\ \begin{bmatrix} 1 & 0 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 1/2 & 0 \\ 0 & 1 & 0 & 0 & 1/2 \end{bmatrix} \end{array} \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \xrightarrow[\text{rows}]{\text{rearrange}}$$

$$\begin{array}{c c}
\begin{array}{ccccc} 1 & 5 & 2 & 3 & 4 \end{array} & \\
\left[\begin{array}{cc|ccc} 1 & 0 & 1/2 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1/2 \\ \hline 0 & 0 & 0 & 1/2 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 1/2 & 0 \end{array}\right] & \begin{array}{c} 1 \\ 5 \\ 2 \\ 3 \\ 4 \end{array}
\end{array}$$

The matrix Q and the fundamental matrix $M = (I - Q)^{-1}$ are

$$Q = \begin{array}{c c}
\begin{array}{ccc} 2 & 3 & 4 \end{array} & \\
\begin{bmatrix} 0 & 1/2 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 1/2 & 0 \end{bmatrix} & \begin{array}{c} 2 \\ 3 \\ 4 \end{array}
\end{array}
\text{ and } M = \begin{array}{c c}
\begin{array}{ccc} 2 & 3 & 4 \end{array} & \\
\begin{bmatrix} 3/2 & 1 & 1/2 \\ 1 & 2 & 1 \\ 1/2 & 1 & 3/2 \end{bmatrix} & \begin{array}{c} 2 \\ 3 \\ 4 \end{array}
\end{array}$$

Starting at state 3, the expected number of visits to state 2 until absorption is the entry of M whose row corresponds to state 2 and whose column corresponds to state 3. This value is 1, so the chain will visit state 2 once on the average before being absorbed. The sum of the columns of M corresponding to states 2 and 4 is three, so the expected number of steps until absorption is three if starting at either state 2 or state 4. Likewise the expected number of steps until absorption starting at state 3 is four. ■

Transit Times

Consider the problem of calculating the expected number of steps t_{ji} needed to travel from state j to state i in an irreducible Markov chain. If the states i and j are the same state, the value t_{jj} is the expected return time to state j found in Section 10.4. The value t_{ji} will be called the **transit time** (or **mean first passage time**) from state j to state i. Surprisingly, the insight into transient states provided by Theorem 6 is exactly what is needed to calculate t_{ji}.

Finding the transit time of a Markov chain from state j to state i begins by changing the transition matrix P for the chain. First reorder the states so that state i comes first. The new matrix has the form

$$\left[\begin{array}{c|c} p_{ii} & S \\ \hline X & Q \end{array}\right]$$

for some matrices S, X, and Q. Next change the first column of the matrix from $\begin{bmatrix} p_{ii} \\ X \end{bmatrix}$ to $\begin{bmatrix} 1 \\ O \end{bmatrix}$, where O is a zero vector of appropriate size. In terms of probabilities, it is now impossible to leave state i after entering it. State i is now an absorbing state for the Markov chain, and the transition matrix now has the form

$$\left[\begin{array}{c|c} 1 & S \\ \hline O & Q \end{array}\right]$$

The expected number of steps t_{ji} that it takes to reach state i after starting at state j may be calculated using Theorem 6(b): it will be the sum of the column of M corresponding to state j.

EXAMPLE 2 Consider an unbiased random walk on $\{1, 2, 3, 4, 5\}$ with reflecting boundaries. Find the expected number of steps t_{j4} required to get to state 4 starting at any state $j \neq 4$ of the chain.

Solution The transition matrix for this Markov chain is

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

First reorder the states to list state 4 first, then convert state 4 to an absorbing state.

$$\begin{array}{cc} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} & \\ \begin{bmatrix} 0 & 1/2 & 0 & 0 & 0 \\ 1 & 0 & 1/2 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 1/2 & 0 & 1 \\ 0 & 0 & 0 & 1/2 & 0 \end{bmatrix} & \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \end{array} \xrightarrow[\text{columns}]{\text{rearrange}} \begin{array}{cc} \begin{array}{ccccc} 4 & 1 & 2 & 3 & 5 \end{array} & \\ \begin{bmatrix} 0 & 0 & 1/2 & 0 & 0 \\ 0 & 1 & 0 & 1/2 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 1 \\ 1/2 & 0 & 0 & 0 & 0 \end{bmatrix} & \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \end{array} \xrightarrow[\text{rows}]{\text{rearrange}}$$

$$\begin{array}{cc} \begin{array}{ccccc} 4 & 1 & 2 & 3 & 5 \end{array} & \\ \begin{bmatrix} 0 & 0 & 0 & 1/2 & 1 \\ 0 & 0 & 1/2 & 0 & 0 \\ 0 & 1 & 0 & 1/2 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 \\ 1/2 & 0 & 0 & 0 & 0 \end{bmatrix} & \begin{array}{c} 4 \\ 1 \\ 2 \\ 3 \\ 5 \end{array} \end{array} \xrightarrow[\text{state 4}]{\text{convert}} \begin{array}{cc} \begin{array}{ccccc} 4 & 1 & 2 & 3 & 5 \end{array} & \\ \left[\begin{array}{c|cccc} 1 & 0 & 0 & 1/2 & 1 \\ \hline 0 & 0 & 1/2 & 0 & 0 \\ 0 & 1 & 0 & 1/2 & 0 \\ 0 & 0 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right] & \begin{array}{c} 4 \\ 1 \\ 2 \\ 3 \\ 5 \end{array} \end{array}$$

The matrix Q and the fundamental matrix $M = (I - Q)^{-1}$ are

$$Q = \begin{array}{cc} \begin{array}{cccc} 1 & 2 & 3 & 5 \end{array} & \\ \begin{bmatrix} 0 & 1/2 & 0 & 0 \\ 1 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} & \begin{array}{c} 1 \\ 2 \\ 3 \\ 5 \end{array} \end{array} \text{ and } M = \begin{array}{cc} \begin{array}{cccc} 1 & 2 & 3 & 5 \end{array} & \\ \begin{bmatrix} 3 & 2 & 1 & 0 \\ 4 & 4 & 2 & 0 \\ 2 & 2 & 2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} & \begin{array}{c} 1 \\ 2 \\ 3 \\ 5 \end{array} \end{array}$$

Summing the columns of M gives $t_{14} = 9$, $t_{24} = 8$, $t_{34} = 5$, and $t_{54} = 1$. ■

Absorption Probabilities

Suppose that a Markov chain has more than one recurrent class and at least one transient state j. If the chain starts at state j, then the chain will eventually be absorbed into one of the recurrent classes; the probability that the chain is absorbed into a particular recurrent class is called the **absorption probability** for that recurrent class. The fundamental matrix is used in calculating the absorption probabilities.

To calculate the absorption probabilities, begin by changing the transition matrix for the Markov chain. First write all recurrent classes as single states i with $p_{ii} = 1$; that is, each recurrent class coalesces into an absorbing state. (Exercises 37 and 38 explore the information that the absorption probabilities give for recurrent classes with more than one state.) A canonical form for this altered transition matrix is

$$P = \left[\begin{array}{c|c} I & S \\ \hline O & Q \end{array}\right]$$

where the identity matrix describes the lack of movement between the absorbing states.

Let j be a transient state and i be an absorbing state for the changed Markov chain; to find the probability that the chain starting at j is eventually absorbed by i, consider the (i, j) entry in the matrix P^k. This entry is the probability that a system which starts at state j is at state i after k steps. Since i is an absorbing state, in order for the system to be at state i, the system must have been absorbed by state i at some step at or before the k^{th} step. Thus the probability that the system has been absorbed by state i at or before the k^{th} step is just the (i, j)-entry in the matrix P^k, and the probability that the chain starting at j is eventually absorbed by i is the (i, j) entry in $\lim\limits_{k\to\infty} P^k$. Computing P^k using rules for multiplying partitioned matrices (see Section 2.4) gives

$$P^2 = \left[\begin{array}{c|c} I & S + SQ \\ \hline O & Q^2 \end{array}\right], P^3 = \left[\begin{array}{c|c} I & S + SQ + SQ^2 \\ \hline O & Q^3 \end{array}\right],$$

and it may be proved by induction (Exercise 28) that

$$P^k = \left[\begin{array}{c|c} I & S_k \\ \hline O & Q^k \end{array}\right], \text{ where } S_k = S + SQ + SQ^2 + \ldots + SQ^{k-1} = S\left(I + Q + Q^2 + \ldots + Q^{k-1}\right)$$

Since j is a transient state and i is an absorbing state, only the entries in S_k need be considered. The probability that the chain starting at j is eventually absorbed by i may thus be found by investigating the matrix

$$A = \lim_{k\to\infty} S_k = \lim_{k\to\infty} S(I + Q + Q^2 + \ldots + Q^{k-1}) = \lim_{k\to\infty} S(I + Q + Q^2 + \ldots) = SM$$

where M is the fundamental matrix for the Markov chain with coalesced recurrent classes. The (j, i) entry in A is the probability that the chain starting at j is eventually absorbed by i. The following Theorem summarizes these ideas; an alternative proof is given in Appendix 2.

THEOREM 7 Suppose that the recurrent classes of a Markov chain are all absorbing states. Let j be a transient state and i be an absorbing state of this chain. Then the probability that the Markov chain starting at state j is eventually absorbed by state i is the (i, j)-element of the matrix $A = SM$, where M is the fundamental matrix of the Markov chain and S is that portion of the transition matrix that governs movement from transient states to absorbing states.

EXAMPLE 3 Consider the unbiased random walk on $\{1, 2, 3, 4, 5\}$ with absorbing boundaries studied in Example 1. Find the probability that the chain is absorbed into state 1 given that the chain starts at state 4.

Solution: Placing the states in the order $\{1, 5, 2, 3, 4\}$, gives the canonical form of the transition matrix:

$$
\begin{array}{c c}
\begin{array}{ccccc} 1 & 5 & 2 & 3 & 4 \end{array} & \\
\left[\begin{array}{cc|ccc}
1 & 0 & 1/2 & 0 & 0 \\
0 & 1 & 0 & 0 & 1/2 \\
\hline
0 & 0 & 0 & 1/2 & 0 \\
0 & 0 & 1/2 & 0 & 1/2 \\
0 & 0 & 0 & 1/2 & 0
\end{array}\right] & \begin{array}{c} 1 \\ 5 \\ 2 \\ 3 \\ 4 \end{array}
\end{array}
$$

The matrix Q and the fundamental matrix $M = (I - Q)^{-1}$ are

$$
Q = \begin{array}{cc}
\begin{array}{ccc} 2 & 3 & 4 \end{array} & \\
\begin{bmatrix} 0 & 1/2 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 1/2 & 0 \end{bmatrix} & \begin{array}{c} 2 \\ 3 \\ 4 \end{array}
\end{array}
\text{ and } M = \begin{array}{cc}
\begin{array}{ccc} 2 & 3 & 4 \end{array} & \\
\begin{bmatrix} 3/2 & 1 & 1/2 \\ 1 & 2 & 1 \\ 1/2 & 1 & 3/2 \end{bmatrix} & \begin{array}{c} 2 \\ 3 \\ 4 \end{array}
\end{array}
$$

so

$$
A = SM = \begin{bmatrix} 1/2 & 0 & 0 \\ 0 & 0 & 1/2 \end{bmatrix} \begin{bmatrix} 3/2 & 1 & 1/2 \\ 1 & 2 & 1 \\ 1/2 & 1 & 3/2 \end{bmatrix} = \begin{array}{cc}
\begin{array}{ccc} 2 & 3 & 4 \end{array} & \\
\begin{bmatrix} 3/4 & 1/2 & 1/4 \\ 1/4 & 1/2 & 3/4 \end{bmatrix} & \begin{array}{c} 1 \\ 5 \end{array}
\end{array}
$$

The columns of A correspond to the transient states 2, 3, and 4 in that order, while the rows correspond to the absorbing states 1 and 5. The probability the chain started at state 4 is absorbed at state 1 is $1/4$.

Absorption probabilities may be used to compute the probability that a system modeled by an irreducible Markov chain visits one state before another.

EXAMPLE 4 Consider a simple random walk on the graph in Figure 1. What is the probability that a walker starting at state 1 visits state 4 before visiting state 7?

Solution Changing state 4 and state 7 to absorbing states and then computing the absorption probabilities starting at state 1 will answer this question. Begin by reordering the states as 4, 7, 1, 2, 3, 5, 6 and rewrite states 4 and 7 as absorbing states:

$$
\begin{array}{cc}
\begin{array}{ccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{array} & \\
\begin{bmatrix}
0 & 1/3 & 1/4 & 0 & 0 & 0 & 0 \\
1/2 & 0 & 1/4 & 0 & 1/2 & 0 & 0 \\
1/2 & 1/3 & 0 & 1 & 0 & 1/3 & 0 \\
0 & 0 & 1/4 & 0 & 0 & 0 & 0 \\
0 & 1/3 & 0 & 0 & 0 & 1/3 & 0 \\
0 & 0 & 1/4 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 1/2 & 1/3 & 0
\end{bmatrix} & \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{array}
\end{array}
\xrightarrow[\text{columns}]{\text{rearrange}}
\begin{array}{cc}
\begin{array}{ccccccc} 4 & 7 & 1 & 2 & 3 & 5 & 6 \end{array} & \\
\begin{bmatrix}
0 & 0 & 0 & 1/3 & 1/4 & 0 & 0 \\
0 & 0 & 1/2 & 0 & 1/4 & 1/2 & 0 \\
1 & 0 & 1/2 & 1/3 & 0 & 0 & 1/3 \\
0 & 0 & 0 & 0 & 1/4 & 0 & 0 \\
0 & 0 & 0 & 1/3 & 0 & 0 & 1/3 \\
0 & 1 & 0 & 0 & 1/4 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1/2 & 1/3
\end{bmatrix} & \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{array}
\end{array}
\xrightarrow[\text{rows}]{\text{rearrange}}
$$

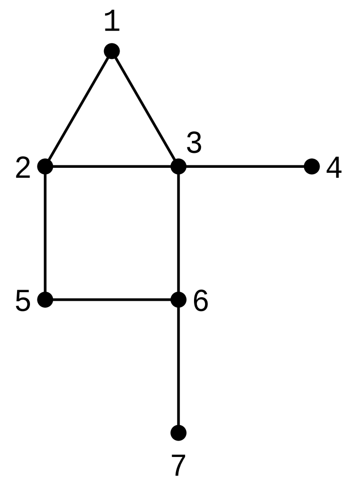


Figure 1: The graph for Example 4.

$$\begin{array}{cc} \begin{array}{ccccccc} 4 & 7 & 1 & 2 & 3 & 5 & 6 \end{array} & \\ \begin{bmatrix} 0 & 0 & 0 & 0 & 1/4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1/2 & 1/3 \\ 0 & 0 & 0 & 1/3 & 1/4 & 0 & 0 \\ 0 & 0 & 1/2 & 0 & 1/4 & 1/2 & 0 \\ 1 & 0 & 1/2 & 1/3 & 0 & 0 & 1/3 \\ 0 & 0 & 0 & 1/3 & 0 & 0 & 1/3 \\ 0 & 1 & 0 & 0 & 1/4 & 0 & 0 \end{bmatrix} & \begin{array}{c} 4 \\ 7 \\ 1 \\ 2 \\ 3 \\ 5 \\ 6 \end{array} \end{array} \xrightarrow[\text{states 4 and 7}]{\text{convert}} \begin{array}{cc} \begin{array}{ccccccc} 4 & 7 & 1 & 2 & 3 & 5 & 6 \end{array} & \\ \begin{bmatrix} 1 & 0 & 0 & 0 & 1/4 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1/2 & 1/3 \\ 0 & 0 & 0 & 1/3 & 1/4 & 0 & 0 \\ 0 & 0 & 1/2 & 0 & 1/4 & 1/2 & 0 \\ 0 & 0 & 1/2 & 1/3 & 0 & 0 & 1/3 \\ 0 & 0 & 0 & 1/3 & 0 & 0 & 1/3 \\ 0 & 0 & 0 & 0 & 1/4 & 0 & 0 \end{bmatrix} & \begin{array}{c} 4 \\ 7 \\ 1 \\ 2 \\ 3 \\ 5 \\ 6 \end{array} \end{array}$$

The resulting transition matrix is $\left[\begin{array}{c|c} I & S \\ \hline O & Q \end{array}\right]$, with

$$S = \begin{array}{cc} \begin{array}{ccccc} 1 & 2 & 3 & 5 & 6 \end{array} & \\ \begin{bmatrix} 0 & 0 & 1/4 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 1/3 \end{bmatrix} & \begin{array}{c} 4 \\ 7 \end{array} \end{array} \quad \text{and } Q = \begin{array}{cc} \begin{array}{ccccc} 1 & 2 & 3 & 5 & 6 \end{array} & \\ \begin{bmatrix} 0 & 1/3 & 1/4 & 0 & 0 \\ 1/2 & 0 & 1/4 & 1/2 & 0 \\ 1/2 & 1/3 & 0 & 0 & 1/3 \\ 0 & 1/3 & 0 & 0 & 1/3 \\ 0 & 0 & 1/4 & 0 & 0 \end{bmatrix} & \begin{array}{c} 1 \\ 2 \\ 3 \\ 5 \\ 6 \end{array} \end{array}$$

so

$$M = (I - Q)^{-1} = \begin{array}{cc} \begin{array}{ccccc} 1 & 2 & 3 & 5 & 6 \end{array} & \\ \begin{bmatrix} 12/5 & 8/5 & 6/5 & 6/5 & 4/5 \\ 12/5 & 31/10 & 17/10 & 11/5 & 13/10 \\ 12/5 & 34/15 & 38/15 & 28/15 & 22/15 \\ 6/5 & 22/15 & 14/15 & 34/15 & 16/15 \\ 6/5 & 13/10 & 11/10 & 8/5 & 19/10 \end{bmatrix} & \begin{array}{c} 1 \\ 2 \\ 3 \\ 5 \\ 6 \end{array} \end{array}$$

and

$$A = SM = \begin{array}{c} \\ \\ \end{array}\begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 5 & 6 \end{array} \\ \begin{bmatrix} 3/5 & 17/30 & 19/30 & 7/15 & 11/30 \\ 2/5 & 13/30 & 11/30 & 8/15 & 19/30 \end{bmatrix} \end{array} \begin{array}{c} \\ 4 \\ 7 \end{array}$$

Since the first column of A corresponds to state 1 and the rows correspond to 4 and 7 respectively, the probability of visiting 4 before visiting 7 is $3/5$. ■

A mathematical model that uses Theorems 6 and 7 appears in Section 10.6.

Practice Problems

1. Consider a Markov chain on $\{1, 2, 3, 4\}$ with transition matrix

$$P = \begin{bmatrix} 1 & 1/2 & 0 & 0 \\ 0 & 1/6 & 1/2 & 0 \\ 0 & 1/3 & 1/6 & 0 \\ 0 & 0 & 1/3 & 1 \end{bmatrix}$$

   a. If the Markov chain starts at state 2, find the expected number of steps before the chain is absorbed.
   b. If the Markov chain starts at state 2, find the probability that the chain is absorbed at state 1.

2. Consider a Markov chain on $\{1, 2, 3, 4\}$ with transition matrix

$$P = \begin{bmatrix} 2/3 & 1/2 & 0 & 0 \\ 1/3 & 1/6 & 1/2 & 0 \\ 0 & 1/3 & 1/6 & 1/2 \\ 0 & 0 & 1/3 & 1/2 \end{bmatrix}$$

   a. If the Markov chain starts at state 2, find the expected number of steps required to reach state 4.
   b. If the Markov chain starts at state 2, find the probability that state 1 is reached before state 4.

10.5 Exercises

In Exercises 1-3, find the fundamental matrix of the Markov chain with the given transition matrix. Assume that the state space in each case is $\{1, 2, \ldots, n\}$. If reordering of states in necessary, list the order in which the states have been reordered.

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

2. $$\begin{bmatrix} 1 & 0 & 0 & 1/4 & 1/5 \\ 0 & 1 & 0 & 1/8 & 1/10 \\ 0 & 0 & 1 & 1/8 & 1/5 \\ 0 & 0 & 0 & 1/4 & 3/10 \\ 0 & 0 & 0 & 1/4 & 1/5 \end{bmatrix}$$

3. $$\begin{bmatrix} 1/5 & 0 & 1/10 & 0 & 1/5 \\ 1/5 & 1 & 1/5 & 0 & 1/5 \\ 1/5 & 0 & 1/5 & 0 & 1/4 \\ 1/5 & 0 & 1/4 & 1 & 1/10 \\ 1/5 & 0 & 1/4 & 0 & 1/4 \end{bmatrix}$$

In Exercises 4-6, find the matrix $A = \lim_{n\to\infty} S_n$ for the Markov chain with the given transition matrix. Assume that the state space in each case is $\{1, 2, \ldots, n\}$. If reordering of states in necessary, list the order in which the states have been reordered.

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

5. $$\begin{bmatrix} 1 & 0 & 0 & 1/4 & 1/5 \\ 0 & 1 & 0 & 1/8 & 1/10 \\ 0 & 0 & 1 & 1/8 & 1/5 \\ 0 & 0 & 0 & 1/4 & 3/10 \\ 0 & 0 & 0 & 1/4 & 1/5 \end{bmatrix}$$

6. $$\begin{bmatrix} 1/5 & 0 & 1/10 & 0 & 1/5 \\ 1/5 & 1 & 1/5 & 0 & 1/5 \\ 1/5 & 0 & 1/5 & 0 & 1/4 \\ 1/5 & 0 & 1/4 & 1 & 1/10 \\ 1/5 & 0 & 1/4 & 0 & 1/4 \end{bmatrix}$$

7. Suppose that the Markov chain in Exercise 1 starts at state 3. How many visits will the chain make to state 4 on the average before absorption?

8. Suppose that the Markov chain in Exercise 2 starts at state 4. How many steps will the chain take on the average before absorption?

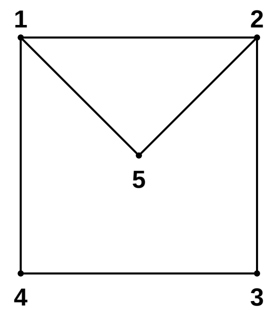
9. Suppose that the Markov chain in Exercise 3 starts at state 1. How many steps will the chain take on the average before absorption?

10. Suppose that the Markov chain in Exercise 4 starts at state 3. What is the probability that the chain is absorbed at state 1?

11. Suppose that the Markov chain in Exercise 5 starts at state 4. Find the probabilities that the chain is absorbed at states 1, 2, and 3.

12. Suppose that the Markov chain in Exercise 6 starts at state 5. Find the probabilities that the chain is absorbed at states 2 and 4.

13. Consider simple random walk on the following graph.

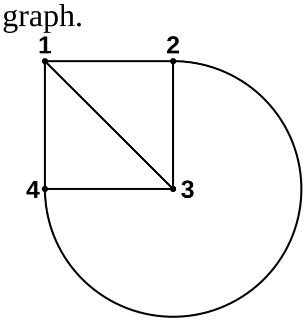


a. Suppose that the walker begins in state 5. What is the expected number of visits to state 2 before the walker visits state 1?

b. Suppose again that the walker begins in state 5. What is the expected number of steps until the walker reaches state 1?

c. Now suppose that the walker starts in state 1. What is the probability that the walker reaches state 5 before reaching state 2?

14. Consider simple random walk on the following graph.

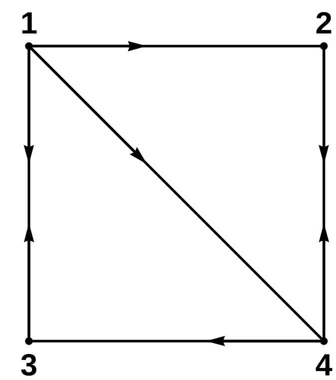


a. Suppose that the walker begins in state 3. What is the expected number of visits to state 2 before the walker visits state 1?

b. Suppose again that the walker begins in state 3. What is the expected number of steps until the walker reaches state 1?

c. Now suppose that the walker starts in state 1. What is the probability that the walker reaches state 3 before reaching state 2?

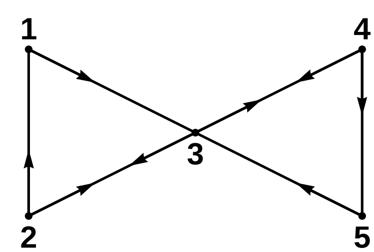
15. Consider simple random walk on the following directed graph. Suppose that the walker starts at state 1.



a. How many visits does the walker expect to make to state 2 before visiting state 3?

b. How many steps does the walker expect to take before visiting state 3?

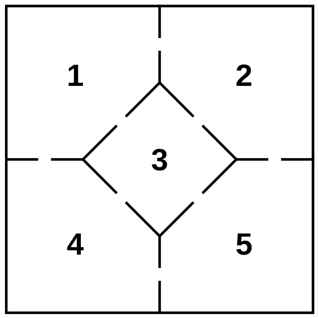
16. Consider simple random walk on the following directed graph. Suppose that the walker starts at state 4.



a. How many visits does the walker expect to make to state 3 before visiting state 2?

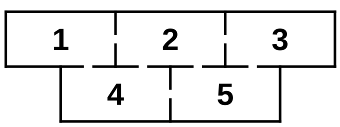
b. How many steps does the walker expect to take before visiting state 2?

17. Consider the mouse in the following maze from Section 1, Exercise 17.



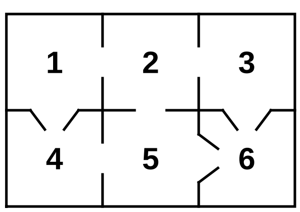
If the mouse starts in room 2, what is the probability that the mouse visits room 3 before visiting room 4?

18. Consider the mouse in the following maze from Section 1, Exercise 18.



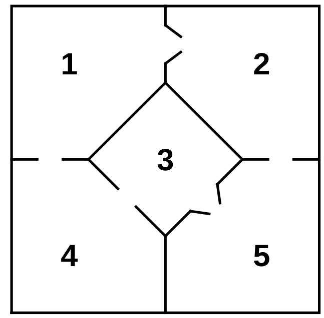
If the mouse starts in room 1, what is the probability that the mouse visits room 3 before visiting room 4?

19. Consider the mouse in the following maze from Section 1, Exercise 19.



If the mouse starts in room 1, how many steps on the average will it take the mouse to get to room 6?

20. Consider the mouse in the following maze from Section 1, Exercise 20.



If the mouse starts in room 1, how many steps on the average will it take the mouse to get to room 5?

In Exercises 21-22, mark each statement True or False. Justify each answer.

21. a. The (i, j)-element in the fundamental matrix M is the expected number of visits to the transient state j prior to absorption, starting at the transient state i.

b. The (j, i)-element in the fundamental matrix gives the expected number of visits to state i starting at state j prior to absorption.

c. The probability that the Markov chain starting at state i is eventually absorbed by state j is the (j, i)-element of the matrix $A = SM$, where M is the fundamental matrix of the Markov chain and S is that portion of the transition matrix that governs movement from transient states to absorbing states.

22. a. The sum of the column j of the fundamental matrix M is the expected number of time steps until absorption.

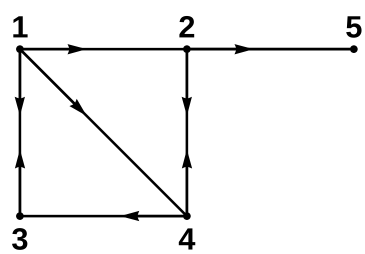
b. Transit times may be computed directly from the entries in the transition matrix.

c. If A is a $m \times m$ substochastic matrix, then the entries in A^n approach 0 as n increases.

23. Suppose that the weather in Charlotte is modeled using the Markov chain in Section 1, Exercise 23. If it is sunny today, what is the probability that the weather is cloudy before it is rainy?

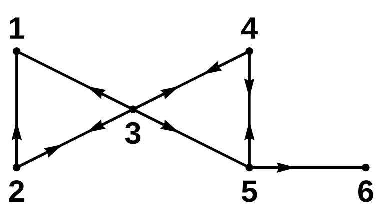
24. Suppose that the weather in Charlotte is modeled using the Markov chain in Section 1, Exercise 24. If it rained yesterday and today, how many days on the average will it take before there are two consecutive days with no rain?

25. Consider a set of webpages hyperlinked by the given directed graph that was studied in Section 2, Exercise 25.



If a random surfer starts on page 1, how many mouse clicks on the average will the surfer make before becoming stuck at a dangling node?

26. Consider a set of webpages hyperlinked by the given directed graph that was studied in Section 2, Exercise 26.



If a random surfer starts on page 3, what is the probability that the surfer will eventually become stuck on page 1, which is a dangling node?

Exercises 27-30 concern the Markov chain model for scoring a tennis match described in Section 1, Exercise 31. Suppose that Player A and player B are playing a tennis match, that the probability that player A wins any point is $p = .6$, and that the game is currently at "deuce."

27. How many more points will the tennis game be expected to last?

28. Find the probability that player A wins the game.

29. Repeat Exercise 27 if the game is

    a. currently at "advantage A".
    b. currently at "advantage B".

30. Repeat Exercise 28 if the game is

    a. currently at "advantage A".
    b. currently at "advantage B".

Exercises 31-36 concern the two Markov chain models for scoring volleyball games described in Section 1, Exercise 32. Suppose that teams A and player B are playing a 15-point volleyball game which is tied 15-15 with team A serving. Suppose that the probability p that team A wins any rally for which it serves is $p = .7$, and the probability q that team B wins any rally for which it serves is $q = .6$.

31. Suppose that rally point scoring is being used. How many more rallies will the volleyball game be expected to last?

32. Suppose that rally point scoring is being used. Find the probability that team A wins the game.

33. Suppose that side out scoring is being used. How many more rallies will the volleyball game be expected to last?

34. Suppose that side out scoring is being used. Find the probability that team A wins the game.

35. Rally point scoring was introduced to make volleyball matches take less time. Considering the results of Exercises 31 and 33, does using rally point scoring really lead to fewer rallies being played?

36. Since $p = .7$ and $q = .6$, it seems that team A is the dominant team. Does it really matter which scoring system is chosen? Should the manager of each team have a preference?

37. Consider a Markov chain on $\{1, 2, 3, 4, 5\}$ with transition matrix

$$P = \begin{bmatrix} 1/4 & 1/2 & 1/3 & 0 & 1/4 \\ 3/4 & 1/2 & 0 & 1/3 & 1/4 \\ 0 & 0 & 0 & 1/3 & 0 \\ 0 & 0 & 1/3 & 0 & 0 \\ 0 & 0 & 1/3 & 1/3 & 1/2 \end{bmatrix}$$

Find $\lim_{n\to\infty} P^n$ by the following steps.

a. What are the recurrent and transient classes for this chain?

b. Find the limiting matrix for each recurrent class.

c. Determine the long range probabilities for the Markov chain starting from each transient state.

d. Use the results of (b) and (c) to find $\lim_{n\to\infty} P^n$.

e. Confirm your answer in (d) by taking P to a high power.

38. Consider a Markov chain on $\{1, 2, 3, 4, 5, 6\}$ with transition matrix

$$P = \begin{bmatrix} 1/3 & 1/2 & 0 & 0 & 1/2 & 0 \\ 2/3 & 1/2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/4 & 2/3 & 0 & 1/2 \\ 0 & 0 & 3/4 & 1/3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/4 & 1/4 \\ 0 & 0 & 0 & 0 & 1/4 & 1/4 \end{bmatrix}$$

Find $\lim_{n\to\infty} P^n$ by the following steps.

a. What are the recurrent and transient classes for this chain?

b. Find the limiting matrix for each recurrent class.

c. Find the absorption probabilities from each transient state into each recurrent class.

d. Use the results of (b) and (c) to find $\lim_{n\to\infty} P^n$.

e. Confirm your answer in (d) by taking P to a high power.

39. Show that if $P = \left[\begin{array}{c|c} I & S \\ \hline O & Q \end{array}\right]$, then $P^n = \left[\begin{array}{c|c} I & S_n \\ \hline O & Q^n \end{array}\right]$, where $S_n = S + SQ + SQ^2 + \ldots + SQ^{n-1} = S\,(I + Q + Q^2 + \ldots + Q^{n-1})$.

Solutions to Practice Problems

1. a. Since 1 and 4 are absorbing states, reordering the states as $\{1, 4, 2, 3\}$ produces the canonical form

$$\begin{array}{cc} & \begin{array}{cccc} 1 & 4 & 2 & 3 \end{array} \\ P = & \begin{bmatrix} 1 & 0 & 1/2 & 0 \\ 0 & 1 & 0 & 1/3 \\ 0 & 0 & 1/6 & 1/2 \\ 0 & 0 & 1/3 & 1/6 \end{bmatrix} \end{array} \begin{array}{c} 1 \\ 4 \\ 2 \\ 3 \end{array}$$

So

$$\begin{array}{cc} & \begin{array}{cc} 2 & 3 \end{array} \\ Q = & \begin{bmatrix} 1/6 & 1/2 \\ 1/3 & 1/6 \end{bmatrix} \end{array} \begin{array}{c} 2 \\ 3 \end{array} \text{ and } \begin{array}{cc} & \begin{array}{cc} 2 & 3 \end{array} \\ M = & \begin{bmatrix} 30/19 & 18/19 \\ 12/19 & 30/19 \end{bmatrix} \end{array} \begin{array}{c} 2 \\ 3 \end{array}$$

The expected number of steps needed starting at state 2 before the chain is absorbed is the sum of the entries in the column of M corrsponding to state 2, which is

$$\frac{30}{19} + \frac{12}{19} = \frac{42}{19}$$

b. Using the canonical form of the transition matrix, we see that

$$\begin{array}{cc} & \begin{array}{cc} 2 & 3 \end{array} \\ S = & \begin{bmatrix} 1/2 & 0 \\ 0 & 1/3 \end{bmatrix} \end{array} \begin{array}{c} 1 \\ 4 \end{array} \text{ and } \begin{array}{cc} & \begin{array}{cc} 2 & 3 \end{array} \\ A = SM = & \begin{bmatrix} 15/19 & 9/19 \\ 4/19 & 10/19 \end{bmatrix} \end{array} \begin{array}{c} 1 \\ 4 \end{array}$$

The probability that the chain is absorbed at state 1 given that the Markov chain starts at state 2 is the entry in A whose row corresponds to state 1 and whose column corresponds to state 2; this entry is $15/19$.

2. a. Reorder the states as $\{4, 1, 2, 3\}$ and make state 4 into an absorbing state and to produce the canonical form

$$\begin{array}{cc} & \begin{array}{cccc} 4 & 1 & 2 & 3 \end{array} \\ P = & \begin{bmatrix} 1 & 0 & 0 & 1/3 \\ 0 & 2/3 & 1/2 & 0 \\ 0 & 1/3 & 1/6 & 1/2 \\ 0 & 0 & 1/3 & 1/6 \end{bmatrix} \end{array} \begin{array}{c} 4 \\ 1 \\ 2 \\ 3 \end{array}$$

So

$$\begin{array}{cc} & \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ Q = & \begin{bmatrix} 2/3 & 1/2 & 0 \\ 1/3 & 1/6 & 1/2 \\ 0 & 1/3 & 1/6 \end{bmatrix} \end{array} \begin{array}{c} 1 \\ 2 \\ 3 \end{array} \text{ and } \begin{array}{cc} & \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ M = & \begin{bmatrix} 14.25 & 11.25 & 6.75 \\ 7.50 & 7.50 & 4.50 \\ 3.00 & 3.00 & 3.00 \end{bmatrix} \end{array} \begin{array}{c} 1 \\ 2 \\ 3 \end{array}$$

The expected number of steps required to reach 4, starting at state 2, is the sum of the entries in the column of M corrsponding to state 2, which is

$$11.25 + 7.50 + 3.00 = 21.75$$

b. Make states 1 and 4 into absorbing states and reorder the states as $\{1, 4, 2, 3\}$ to produce the canonical form

$$P = \begin{array}{c} \\ \\ \\ \\ \\ \end{array}\begin{array}{cc} \begin{array}{cccc} 1 & 4 & 2 & 3 \end{array} & \\ \begin{bmatrix} 1 & 0 & 1/2 & 0 \\ 0 & 1 & 0 & 1/3 \\ 0 & 0 & 1/6 & 1/2 \\ 0 & 0 & 1/3 & 1/6 \end{bmatrix} & \begin{array}{c} 1 \\ 4 \\ 2 \\ 3 \end{array} \end{array}$$

So

$$Q = \begin{array}{cc} \begin{array}{cc} 2 & 3 \end{array} & \\ \begin{bmatrix} 1/6 & 1/2 \\ 1/3 & 1/6 \end{bmatrix} & \begin{array}{c} 2 \\ 3 \end{array} \end{array}, \; M = \begin{array}{cc} \begin{array}{cc} 2 & 3 \end{array} & \\ \begin{bmatrix} 30/19 & 18/19 \\ 12/19 & 30/19 \end{bmatrix} & \begin{array}{c} 2 \\ 3 \end{array} \end{array}$$

$$S = \begin{array}{cc} \begin{array}{cc} 2 & 3 \end{array} & \\ \begin{bmatrix} 1/2 & 0 \\ 0 & 1/3 \end{bmatrix} & \begin{array}{c} 1 \\ 4 \end{array} \end{array} \text{ and } A = SM = \begin{array}{cc} \begin{array}{cc} 2 & 3 \end{array} & \\ \begin{bmatrix} 15/19 & 9/19 \\ 4/19 & 10/19 \end{bmatrix} & \begin{array}{c} 1 \\ 4 \end{array} \end{array}$$

Thus the probability that, starting at state 2, state 1 is reached before state 4 is the entry in A whose row corresponds to state 1 and whose column corresponds to state 2; this entry is $15/19$.

10.6 Markov Chains and Baseball Statistics

Markov chains are used to model a wide variety of systems. The examples and exercises in this chapter have shown how Markov chains may be used to model situations in topic one, topic two, and topic three. The final example to be explored is a model for how runners proceed around the bases in baseball. This model leads to useful measures of expected run production both for a team and for individual players.

Baseball Modeled by a Markov Chain

Many baseball fans carefully study the statistics of their favorite teams. The teams themselves use baseball statistics for individual players to determine strategy during games, and to make hiring decisions.[4] This section shows how a Markov chain is used to predict the number of earned runs a team will score and to compare the offensive abilities of different players. Some exercises suggest how to use Markov chains to investigate matters of baseball strategy, such as deciding whether to attempt a sacrifice or a steal.

The Markov chain in this section provides a way to analyze how runs are scored during half of one inning of a baseball game. The states of the chain are the various configurations of runners on base and the number of outs. See Table 10.1.

The first state in the left column of Table 10.1 ("no bases occupied, 0 outs") is the initial state of the chain, when the baseball half-inning begins (that is, when one team becomes the team "at bat"). The four states in the right column describe the various ways the half-inning can end (when the third out occurs and the teams trade places). Physically, the half-inning is completed when the third out occurs. Mathematically, the Markov chain continues in one of the four "final" states. (The model only applies to a game in which each half-inning is completed.) So, each of these four states is an absorbing state of the chain. The other 24 states are transient states, because whenever an out is made, the states with fewer outs can never occur again.

The Markov chain moves from state to state because of the actions of the batters. The transition probabilities of the chain are the probabilities of possible outcomes of a batter's action. For a Markov chain, the transition probabilities must remain the same from batter to batter, so the model does not allow for variations between batters. This assumption means that each batter for a team hits as an "average batter" for the team.[5]

The model also assumes that only the batter determines how the runners move around the bases. This means that stolen bases, wild pitches, and passed balls are not considered. Also, errors by the players in the field are not allowed, so the model only calculates earned runs – runs that are scored without the benefit of fielding errors. Finally, the model considers only seven possible outcomes at the plate: a single (arriving safely at first base and stopping there), a double (arriving safely at second base), a triple (arriving safely at third base), a home run, a walk (advancing to first

[4] The use of statistical analysis in baseball is called **sabermetrics** as a tribute to SABR, the Society for American Baseball Research. An overview of sabermetrics can be found at `http://en.wikipedia.org/wiki/Sabermetrics`

[5] This unrealistic assumption can be overcome by using a more complicated model which uses different transition matrices for each batter. Nevertheless, the model presented here can lead to useful information about the team. Later in the section, the model will be used to evaluate individual players.

Bases Occupied	Outs	State
None	0	0:0
First	0	1:0
Second	0	2:0
Third	0	3:0
First and Second	0	12:0
First and Third	0	13:0
Second and Third	0	23:0
First, Second, and Third	0	123:0
None	1	0:1
First	1	1:1
Second	1	2:1
Third	1	3:1
First and Second	1	12:1
First and Third	1	13:1
Second and Third	1	23:1
First, Second, and Third	1	123:1
None	2	0:2
First	2	1:2
Second	2	2:2
Third	2	3:2
First and Second	2	12:2
First and Third	2	13:2
Second and Third	2	23:2
First, Second, and Third	2	123:2

Left on Base	Outs	State
0	3	0:3
1	3	1:3
2	3	2:3
3	3	3:3

Table 10.1: The 28 States of a Baseball Markov Chain

base without hitting the ball), a hit batsman (a pitched ball hits the batter, and the batter advances to first base), and an "out." Thus, the model allows no double or triple plays, no sacrifices, and no sacrifice flies. However, Markov chain models can be constructed that include some of these excluded events.[6]

Constructing the Transition Matrix

The 28×28 transition matrix for the Markov chain has the canonical form

$$P = \begin{bmatrix} I_4 & S \\ O & Q \end{bmatrix} \tag{1}$$

where I_4 is the 4×4 identity matrix (because the only recurrent states are the four absorbing states, one of which is entered when the third out occurs), S is a 4×24 matrix, and Q is a 24×24

[6]Other models use "play-by-play" data. The numbers of transitions between states are counted and scaled to produce a transition matrix. For these models it does not matter *how* the runners advance, merely that they do.

substochastic matrix. The columns of S and Q correspond to the transient states, in the order shown in Table 10.1. The entries in S describe the transitions from the 24 transient states (with 0, 1, or 2 outs) to the absorbing states (with 3 outs). Note that the only way to enter an absorbing state is to come from a state with 2 outs. Let p_O denote the probability that the batter makes an out. Then S may be written in block form, with three 4×8 blocks, as

$$S = \begin{bmatrix} O & O & X \end{bmatrix}, \quad \text{where} \quad X = \begin{array}{c c} \begin{array}{cccccccc} \scriptstyle 0:2 & \scriptstyle 1:2 & \scriptstyle 2:2 & \scriptstyle 3:2 & \scriptstyle 12:2 & \scriptstyle 13:2 & \scriptstyle 23:2 & \scriptstyle 123:2 \end{array} & \\ \begin{bmatrix} p_O & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & p_O & p_O & p_O & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & p_O & p_O & p_O & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & p_O \end{bmatrix} & \begin{array}{c} \scriptstyle 0:3 \\ \scriptstyle 1:3 \\ \scriptstyle 2:3 \\ \scriptstyle 3:3 \end{array} \end{array} \tag{2}$$

The matrix X describes the transitions from the transient states with 2 outs to the absorbing states with 3 outs. (For example, columns 2, 3, and 4 of X list the probabilities that the batter makes the third out when one runner is on one of the three bases. The substochastic matrix Q has the following block form, with 8×8 blocks,

$$Q = \begin{array}{c c} \begin{array}{ccc} \scriptstyle 0 & \scriptstyle 1 & \scriptstyle 2 \end{array} & \\ \begin{bmatrix} A & O & O \\ B & A & O \\ O & B & A \end{bmatrix} & \begin{array}{c} \scriptstyle 0 \\ \scriptstyle 1 \\ \scriptstyle 2 \end{array} \end{array} \tag{3}$$

The labels on the rows and columns of Q represent the number of outs. The four zero blocks in Q reflect the facts that the number of outs cannot go from 1 to 0, from 2 to 0 or 1, or from 0 directly to 2 in one step. The matrix A describes how the various base configurations change when the number of outs does not change.

The entries in A and B depend on how the batter's action at the plate affects any runners that may already be on base. The Markov chain model presented here makes the assumptions shown in Table 10.2. The exercises consider some alternate assumptions.

The entries in the 8×8 matrices A and B are constructed from the probabilities of the six batting events in Table 10.2. Denote these probabilities by p_W, p_1, p_2, p_3, p_H, and p_O, respectively. The notation p_O was introduced earlier during the construction of the matrix S.

The 8×8 matrix B involves the change of state when the number of outs increases. In this case, the configuration of runners on the bases does not change (see Table 10.2). So

$$B = p_O I$$

where I is the 8×8 identity matrix.[7]

The matrix A concerns the situations in which the batter does not make an out and either succeeds in reaching one of the bases or hits a home run. The construction of A is discussed in Example 1 below and in the exercises. The labels on the rows and columns of A correspond to the states in Table 10.2. Here k is the fixed number of outs: either 0, 1 or 2.

[7]A batter can make an out in three ways – by striking out, by hitting a fly ball that is caught, or by hitting a ground ball that is thrown to first base before the batter arrives. When the second or third case occurs, a runner on a base sometimes can advance one base, but may also make an out and be removed from the bases. Table 10.2 excludes these possibilities. However, see Exercise xx.

Batting Event	Outcome
Walk or Hit Batsman	The batter advances to first base. A runner on first base advances to second base. A runner on second base advances to third base only if first base was also occupied. A runner on third base scores only if first base and second base were also occupied.
Single	The batter advances to first base. A runner on first base advances to second base. A runner on third base scores. A runner on second advances to third base half of the time and scores half of the time.
Double	The batter advances to second base. A runner on first base advances to third base. A runner on second base scores. A runner on third base scores.
Triple	The batter advances to third base. A runner on first base scores. A runner on second base scores. A runner on third base scores.
Home Run	The batter scores. A runner on first base scores. A runner on second base scores. A runner on third base scores.
Out	No runners advance. The number of outs increases by one.

Table 10.2: Assumptions about Advancing Runners

$$A = \begin{array}{cc} \begin{matrix} \text{0:k} & \text{1:k} & \text{2:k} & \text{3:k} & \text{12:k} & \text{13:k} & \text{23:k} & \text{123:k} \end{matrix} & \\ \begin{bmatrix} p_H & p_H & p_H & p_H & p_H & p_H & p_H & p_H \\ p_W + p_1 & 0 & .5p_1 & p_1 & 0 & 0 & .5p_1 & 0 \\ p_2 & 0 & p_2 & p_2 & 0 & 0 & p_2 & 0 \\ p_3 & p_3 & p_3 & p_3 & p_3 & p_3 & p_3 & p_3 \\ 0 & p_W + p_1 & p_W & 0 & .5p_1 & p_1 & 0 & .5p_1 \\ 0 & 0 & .5p_1 & p_W & 0 & 0 & .5p_1 & 0 \\ 0 & p_2 & 0 & 0 & p_2 & p_2 & 0 & p_2 \\ 0 & 0 & 0 & 0 & p_W + .5p_1 & p_W & p_W & p_W + .5p_1 \end{bmatrix} & \begin{matrix} \text{0:k} \\ \text{1:k} \\ \text{2:k} \\ \text{3:k} \\ \text{12:k} \\ \text{13:k} \\ \text{23:k} \\ \text{123:k} \end{matrix} \end{array}$$

The analysis in Example 1 below requires two facts from probability theory. If an event can occur in two mutually exclusive ways, with probabilities p_1 and p_2, then the probability of the event is $p_1 + p_2$. The probability that two independent events both occur is the product of the separate probabilities for each event.

EXAMPLE 1

a. Justify the transition probabilities for the initial state "no bases occupied."

b. Justify the transition probabilities for the initial state "second base occupied."

Solution a. For the first column of A, either the batter advances to one of the bases or hits a home run. So the probability that the bases remain unoccupied is p_H. The batter advances to first base when the batter either walks (or is hit by a pitch), or hits a single. Since the desired outcome can be reached in two different ways, the probability of success is the sum of the two probabilities, namely, $p_W + p_1$. The probabilities of the batter advancing to second base or third base are, respectively, p_2 and p_3. All other outcomes are impossible, because there can be at most one runner on base after one batter when the starting state has no runners on base.

b. This concerns the third column of A. The initial state is 2:k (a runner on second base, k outs). For entry $(1,3)$ of A, the probability of a transition "to state 0:k" is required. Suppose that only second base is occupied and the batter does not make an out. Only a home run will empty the bases, so the $(1,3)$-entry is p_H.

Entry $(2,3)$: ("to state 1:k") To leave a player only on first base, the batter must get to first base and the player on second base must reach home plate successfully.[8] From Table 10.2, the probability of reaching home plate successfully from second base is $.5$. Now, assume that these two events are independent, because only the actions of the batter (and Table 10.2) influence the outcome. In this case, the probability of both events happening at the same time is the product of these two probabilities, so the $(2,3)$-entry is $.5p_1$.

Entry $(3,3)$: ("to state 2:k") To leave only one player on second base, the batter must reach second base (a "double") and the runner on second base must score. The second condition, however, is automatically satisfied because of the assumption in Table 10.2. So the probability of success in this case is p_2. This is the $(3,3)$-entry.

Entry $(4,3)$: ("to state 3:k") An argument similar to that for the $(3,3)$-entry gives that the $(4,3)$-entry is p_3.

Entry $(5,3)$: ("to state 12:k") To leave players on first base and second base, the batter must get to first base and the player on second base must remain there. However, from Table 10.2, if the batter hits a single, the runner on second base will at least get to third base. So, the only way for the desired outcome to occur is for the batter to get a walk or be hit by a pitch. The $(5,3)$-entry is thus p_W.

Entry $(6,3)$: ("to state 13:k") This concerns the batter getting to first base and the runner on second base advancing to third base. That can happen only if the batter hits a single, with probability p_1, and the runner on second base stops at third base, which happens with probability $.5$ (by Table 10.2). Since both events are required, the $(6,3)$-entry is the product $.5p1$.

[8]The only other way to make the player on second "disappear" would be for the player to be tagged "out", but the model does not permit outs for runners on the bases.

Entry $(7, 3)$: ("to state 23:k") To leave players on second base and third base, the batter must hit a double and the runner on second base must advance only to base 3. Table 10.2 rules this out – when the batter hits a double, the runner on second base scores. Thus the $(7, 3)$-entry is zero.

Entry $(8, 3)$: The starting state has just one runner on base. The next state cannot have three runners on base, so the $(8, 3)$-entry is zero. ■

EXAMPLE 2 Batting statistics are often displayed as in Table 10.3. Use the data from Table 10.3 to obtain the transition probabilities for the 2002 Atlanta Braves.

Walks	Hit Batsmen	Singles	Doubles	Triples	Home Runs	Outs
558	54	959	280	25	164	4067

Table 10.3: Atlanta Braves Batting Statistics – 2002 Season

Solution The sum of the entries in Table 10.3 is 6107. This is the total number of Atlanta Braves players who came to bat during the 2002 baseball season. From the first two columns, there are 612 walks or hit batsmen. So, $p_W = 612/6107 = .1002$. Of the 6107 times a player came to bat, a player hit a single 959 times, so $p_1 = 959/6107 = .1570$. Similar calculations provide $p_2 = .0458$, $p_3 = .0041$, $p_H = .0269$, and $p_O = .660$. These values are placed in the matrices shown above to produce the transition matrix for the Markov chain.[9] ■

Applying the Model

Now that the data for the stochastic matrix is available, Theorems 5 and 6 from Section 10.5 can provide information about how many earned runs to expect from the Atlanta Braves during a typical game. The goal is to calculate how many earned runs the Braves will score on average in each half-inning. First, observe that since three batters must make an out to finish one half-inning, the number of runs scored in that half-inning is given by

$$[\#\text{of runs}] = [\#\text{of batters}] - [\#\text{of runners left on base}] - 3 \tag{4}$$

If R is the number of runs scored in the half-inning, B is the number of batters, and L is the number left on base, Equation (4) becomes

$$R = B - L - 3 \tag{5}$$

The quantity of interest is $E[R]$, the expected number of earned runs scored. Properties of expected value give that

$$E[R] = E[B] - E[L] - 3 \tag{6}$$

Each batter moves the Markov chain ahead one step. So, the expected number of batters in a half-inning $E[B]$ is the expected number of steps to absorption (at the third out) when the chain begins

[9] The 28×28 transition matrix is available at `www.laylinalgebra.com`

at the initial state “0 bases occupied, 0 outs”. This initial state corresponds to the fifth column of the transition matrix

$$P = \begin{bmatrix} I_4 & S \\ O & Q \end{bmatrix}$$

In baseball terms, Theorem 5 shows that:

> The expected number of players that bat in one half-inning is the sum of the entries in column 1 of the fundamental matrix $M = (I - Q)^{-1}$.

Thus $E[B]$ may be computed. The other quantity needed in (6) above is $E[L]$, the expected number of batters left on base in a typical half-inning. This is given by the following sum:

$$E[L] = 0 \cdot P(L = 0) + 1 \cdot P(L = 1) + 2 \cdot P(L = 2) + 3 \cdot P(L = 3) \tag{7}$$

Theorem 6 can provide this information because the recurrent classes for the chain are just the four absorbing states (at the end of the half-inning). The probabilities needed in (7) are the probabilities of absorption into the four final states of the half-inning given that the initial state of the system is “0 bases occupied, 0 outs”. So the desired probabilities are in column 1 of the matrix SM, where M is the fundamental matrix of the chain and $S = \begin{bmatrix} O & O & X \end{bmatrix}$ as in (2). The probabilities can be used to calculate $E[L]$ using (7), and thus to find $E[R]$.

EXAMPLE 3 When the Atlanta Braves data from Example 2 is used to construct the transition matrix (not shown here), it turns out that the sum of the first column of the fundamental matrix M is 4.5048, and the first column of the matrix SM is

$$\begin{bmatrix} .3520 \\ .3309 \\ .2365 \\ .0805 \end{bmatrix}$$

Compute the number of earned runs the Braves can expect to score per inning based on their performance in 2002. How many earned runs does the model predict for the entire season, if the Braves play $1443\ 2/3$ innings, as they did in 2002?

Solution The first column of SM shows that, for example, the probability that the Braves left no runners on base is $.3520$. The expected number of runners left on base is

$$E[L] = 0(.3520) + 1(.3309) + 2(.2365) + 3(.0805) = 1.0454$$

The expected number of batters is $E[B] = 4.5048$, the sum of the first column of M. From equation (6), the expected number of earned runs $E[R]$ is

$$E[R] = E[B] - E[L] - 3 = 4.5048 - 1.0454 - 3 = .4594$$

The Markov chain model predicts that the Braves should average $.4594$ earned runs per inning. In $1443\ 2/3$ innings, the total number of earned runs expected is

$$.4594 \times 1443.67 = 663.22$$

The actual number of earned runs for the Braves in 2002 was 636, so the model's error is 27.22 runs, or about 4.3%. ■

Mathematical models are used by some major league teams to compare the offensive profiles of single players. To analyze a player using the Markov chain model, use the player's batting statistics instead of a team's statistics. Compute the expected number of earned runs that a team of such players would score in an inning. This number is generally multiplied by 9 to give what has been termed an "offensive earned run average."

EXAMPLE 4 Table 10.4 shows the career batting statistics for Jose Oquendo, who played for the New York Mets and St. Louis Cardinals in the 1980's and 1990's. Compute his offensive earned run average.

Walks	Hit Batsmen	Singles	Doubles	Triples	Home Runs	Outs
448	5	679	104	24	14	2381

Table 10.4: Jose Oquendo Batting Statistics

Solution Construct the transition matrix from this data as described in Example 2, and then compute M and SM. The sum of the first column of M is 4.6052, so a team entirely composed of Jose Oquendos would come to bat an average of 4.6052 times per inning. That is, $E[B] = 4.6052$. The first column of SM is

$$\begin{bmatrix} .2844 \\ .3161 \\ .2725 \\ .1270 \end{bmatrix}$$

so the expected number of runners left on base is

$$E[L] = 0(.2844) + 1(.3161) + 2(.2725) + 3(.1270) = 1.2421$$

From equation (6), the expected number of earned runs is

$$E[R] = E[B] - E[L] - 3 = 4.6052 - 1.2421 - 3 = .3631$$

The offensive earned run average for Jose Oquendo is $.3631 \times 9 = 3.2679$. This compares to an offensive earned run average of about 10 for teams composed of the greatest hitters in baseball history. See the Exercises. ■

Practice Problems

1. Let A be the matrix just before Example 1. Explain why entry $(3, 6)$ is zero.

2. Explain why entry $(6, 3)$ of A is $.5p_1$.

10.6 Exercises

In Exercises 1-6, justify the transition probabilities for the given initial states. See Example 1.

1. first base occupied

2. third base occupied

3. first and second bases occupied

4. first and third bases occupied

5. second and third bases occupied

6. first, second, and third bases occupied

7. Major League batting statistics for the 2006 season are shown in Table 10.5. Compute the transition probabilities for this data as was done in Example 2, and find the matrix A for this data.

8. Find the complete transition matrix for the model using the Major League data in Table 10.5.

9. It can be shown that the sum of the first column of M for the 2006 Major League data is 4.53933, and that the first column of SM for the 2006 Major League data is

$$\begin{bmatrix} .34973 \\ .33414 \\ .23820 \\ .07793 \end{bmatrix}$$

Find the expected number of earned runs per inning in a Major League game in 2006.

10. The number of innings batted in the Major Leagues in the 2006 season was 43,257, and the number of earned runs scored was 21,722. What is the total number of earned runs scored for the season predicted by the model, and how does it compare with the actual number of earned runs scored?

11. Batting statistics for three of the greatest batters in Major League history are shown in Table 10.6. Compute the transition probabilities for this data for each player.

12. The sums of the first columns of M for the player data in Table 10.6 and the first columns of SM for the player data in Table 10.6 is given in Table 10.7. Find and compare the offensive earned run averages of these players. Which batter does the model say was the best of these three?

13. Consider the second columns of the matrices M and SM, which correspond to the "Runner on first, none out" state.

   a. What information does the sum of the second column of M give?

   b. What value can you calculate using the second column of SM?

   c. What would the calculation of expected runs scored using the data from the second columns mean?

Exercises 14-18 show how the model for run production in the text can be used to determine baseball strategy. Suppose that you are managing a baseball team and have access to the matrices M and SM for your team.

14. The sum of the column of M corresponding to the "Runner on first, none out" state is 4.53933, and the column of SM corresponding to the "Runner on first, none out" state is

$$\begin{bmatrix} .06107 \\ .35881 \\ .41638 \\ .16374 \end{bmatrix}$$

Your team now has a runner on first and no outs. How many earned runs do you expect your team to score this inning?

15. The sum of the column of M corresponding to the "Runner on second, none out" state is 4.53933, and the column of SM corresponding to the "Runner on second, none out" state is

$$\begin{bmatrix} .06107 \\ .47084 \\ .34791 \\ .12018 \end{bmatrix}$$

How many earned runs do you expect your team to score if there is a runner on second and no outs?

16. The sum of the column of M corresponding to the "Bases empty, one out" state is 3.02622, and the column of SM corresponding to the "Bases empty, one out" state is

$$\begin{bmatrix} .48513 \\ .31279 \\ .16060 \\ .04148 \end{bmatrix}$$

How many earned runs do you expect your team to score if the bases are empty and one out?

17. Suppose that a runner for your team is on first base with no outs. You have to decide whether to tell the baserunner to attempt to steal second base. If the steal is successful, there will be a runner on second base and no outs. If the runner is caught stealing, the bases will be empty and there will be one out. Suppose further that the baserunner has a probability of $p = .8$ of stealing successfully. Does attempting a steal in this circumstance increase or decrease the number of earned runs your team will score this innning?

18. In the previous Exercise, let p be the probability that the baserunner steals second successfully. For which values of p would you as manager call for an attempted steal?

Walks	Hit Batsmen	Singles	Doubles	Triples	Home Runs	Outs
15847	1817	29600	9135	952	5386	122268

Table 10.5: Major League Batting Statistics – 2006 Season

Name	Walks	Hit Batsmen	Singles	Doubles	Triples	Home Runs	Outs
Barry Bonds[10]	2426	103	1443	587	77	734	6666
Babe Ruth	2062	43	1517	506	136	714	5526
Ted Williams	2021	39	1537	525	71	521	5052

Table 10.6: Batting Statistics for Leading Batters

	Sum of First Column of M	First Column of SM
Barry Bonds	5.41674	$\begin{bmatrix} .283348 \\ .294212 \\ .258310 \\ .164131 \end{bmatrix}$
Babe Ruth	5.70250	$\begin{bmatrix} .268150 \\ .295908 \\ .268120 \\ .167822 \end{bmatrix}$
Ted Williams	5.79929	$\begin{bmatrix} .233655 \\ .276714 \\ .290207 \\ .199425 \end{bmatrix}$

Table 10.7: Model Information for Batting Statistics

[10]Barry Bonds' statistics are complete through the 2006 season.

Appendix 1: Proof of Theorem 1

Here is a restatement of Theorem 1, which will be proven in this appendix:

THEOREM 1 If P is a regular $m \times m$ transition matrix with $m \geq 2$, then the following statements are all true.

(a) There is a stochastic matrix Π such that $\lim_{n\to\infty} P^n = \Pi$.

(b) Each column of Π is the same probability vector $\mathbf{q}$.

(c) For any initial probability vector $\mathbf{x}_0$, $\lim_{n\to\infty} P^n\mathbf{x}_0 = \mathbf{q}$.

(d) The vector $\mathbf{q}$ is the unique probability vector which is an eigenvector of P associated with the eigenvalue 1.

(e) All eigenvalues λ of P other than 1 have $|\lambda| < 1$.

The proof of Theorem 1 requires creating an order relation for vectors, and begins with the consideration of matrices whose entries are strictly positive or non-negative.

DEFINITION If $\mathbf{x}$ and $\mathbf{y}$ are in $\mathbb{R}^m$, then

a. $\mathbf{x} > \mathbf{y}$ if $x_i > y_i$ for $i = 1, 2, \ldots, m$.

b. $\mathbf{x} < \mathbf{y}$ if $x_i < y_i$ for $i = 1, 2, \ldots, m$.

c. $\mathbf{x} \geq \mathbf{y}$ if $x_i \geq y_i$ for $i = 1, 2, \ldots, m$.

d. $\mathbf{x} \leq \mathbf{y}$ if $x_i \leq y_i$ for $i = 1, 2, \ldots, m$.

DEFINITION An $m \times n$ matrix A is **positive** if all its entries are positive. An $m \times n$ matrix A is **non-negative** if it has no negative entries.

Notice that all stochastic matrices are non-negative. Exercise 27 in Section 10.2 shows that multiplication of vectors by a positive matrix preserves order.

$$\text{If } A \text{ is a positive matrix and } \mathbf{x} > \mathbf{y}, \text{ then } A\mathbf{x} > A\mathbf{y}. \tag{1}$$

$$\text{If } A \text{ is a positive matrix and } \mathbf{x} \geq \mathbf{y}, \text{ then } A\mathbf{x} \geq A\mathbf{y}. \tag{2}$$

In addition, multiplication by non-negative matrices "almost" preserves order in the following sense.

$$\text{If } A \text{ is a non-negative matrix and } \mathbf{x} \geq \mathbf{y}, \text{ then } A\mathbf{x} \geq A\mathbf{y}. \tag{3}$$

The first step toward proving Theorem 1 is a lemma which shows how the transpose of a stochastic matrix acts on a vector.

LEMMA 1 Let P be an $m \times m$ stochastic matrix, and let ϵ be the smallest entry in P. Let $\mathbf{a}$ be in $\mathbb{R}^m$; let M_a be the largest entry in $\mathbf{a}$ and let m_a be the smallest entry in $\mathbf{a}$. Likewise let $\mathbf{b} = P^T\mathbf{a}$, M_b be the largest entry in $\mathbf{b}$ and m_b be the smallest entry in $\mathbf{b}$. Then $m_a \leq m_b \leq M_b \leq M_a$ and

$$M_b - m_b \leq (1 - 2\epsilon)(M_a - m_a)$$

Proof Create a new vector $\mathbf{c}$ from $\mathbf{a}$ by replacing every entry of $\mathbf{a}$ by M_a except for one occurrence of m_a. Suppose that this single m_a entry lies in the i^{th} row of $\mathbf{c}$. Then $\mathbf{c} \geq \mathbf{a}$. If the columns of P^T are labeled $\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_m$, we have

$$\begin{aligned} P^T\mathbf{c} &= \sum_{k=1}^{m} c_k \mathbf{q}_k \\ &= \sum_{k=1}^{m} M_a \mathbf{q}_k - M_a \mathbf{q}_i + m_a \mathbf{q}_i \end{aligned}$$

Since P is a stochastic matrix, each row of P^T sums to 1. If we let $\mathbf{u}$ be the vector in $\mathbb{R}^m$ consisting of all 1's, then $\sum_{k=1}^{m} M_a \mathbf{q}_k = M_a \sum_{k=1}^{m} \mathbf{q}_k = M_a \mathbf{u}$, and

$$\sum_{k=1}^{m} M_a \mathbf{q}_k - M_a \mathbf{q}_i + m_a \mathbf{q}_i = M_a \mathbf{u} - (M_a - m_a)\mathbf{q}_i$$

Since each entry in P (and thus P^T) is greater than or equal to ϵ, $\mathbf{q}_i \geq \epsilon \mathbf{u}$, and

$$M_a \mathbf{u} - (M_a - m_a)\mathbf{q}_i \leq M_a \mathbf{u} - \epsilon(M_a - m_a)\mathbf{u} = (M_a - \epsilon(M_a - m_a))\mathbf{u}$$

So

$$P^T\mathbf{c} \leq (M_a - \epsilon(M_a - m_a))\mathbf{u}$$

But since $\mathbf{a} \geq \mathbf{c}$ and P^T is positive, Equation 2 gives

$$\mathbf{b} = P^T\mathbf{a} \leq P^T\mathbf{c} \leq (M_a - \epsilon(M_a - m_a))\mathbf{u}$$

Thus each entry in $\mathbf{b}$ is less than or equal to $M_a - \epsilon(M_a - m_a)$. In particular,

$$M_b \leq M_a - \epsilon(M_a - m_a) \tag{4}$$

So $M_b \leq M_a$. If we now examine the vector $-\mathbf{a}$, we find that the largest entry in $-\mathbf{a}$ is $-m_a$, the smallest is $-M_a$, and similar results hold for $-\mathbf{b} = P^T(-\mathbf{a})$. Applying Equation 4 to this situation gives

$$-m_b \leq -m_a - \epsilon(-m_a + M_a) \tag{5}$$

so $m_b \geq m_a$. Adding Equations 4 and 5 together gives

$$\begin{aligned} M_b - m_b &\leq M_a - m_a - 2\epsilon(M_a - m_a) \\ &= (1 - 2\epsilon)(M_a - m_a) \end{aligned}$$

■

Proof of Theorem 1 First assume that the transition matrix P is a **positive** stochastic matrix. As above, let $\epsilon > 0$ be the smallest entry in P. Consider the vector $\mathbf{e}_j$ where $1 \leq j \leq m$. Let M_n and m_n be the largest and smallest entries in the vector $(P^T)^n\mathbf{e}_j$. Since $(P^T)^n\mathbf{e}_j = P^T(P^T)^{n-1}\mathbf{e}_j$, Theorem 2 gives

$$M_n - m_n \leq (1-2\epsilon)(M_{n-1} - m_{n-1}) \tag{6}$$

Hence by induction it may be shown that

$$M_n - m_n \leq (1-2\epsilon)^n(M_0 - m_0) = (1-2\epsilon)^n$$

Since $m \geq 2$, $0 < \epsilon \leq 1/2$. Thus $0 \leq 1 - 2\epsilon < 1$, and $\lim\limits_{n\to\infty} M_n - m_n = 0$. Therefore the entries in the vector $(P^T)^n\mathbf{e}_j$ approach the same value, say q_j, as n increases. Notice that since the entries in P^T are between 0 and 1, the entries in $(P^T)^n\mathbf{e}_j$ must also be between 0 and 1, and so q_j must also lie between 0 and 1. Now $(P^T)^n\mathbf{e}_j$ is the j^{th} column of $(P^T)^n$, which is the j^{th} row of P^n. Therefore P^n approaches a matrix all of whose rows are constant vectors, which is another way of saying the columns of P^n approach the same vector $\mathbf{q}$:

$$\lim_{n\to\infty} P^n = \Pi = \begin{bmatrix} \mathbf{q} & \mathbf{q} & \cdots & \mathbf{q} \end{bmatrix} = \begin{bmatrix} q_1 & q_1 & \cdots & q_1 \\ q_2 & q_2 & \cdots & q_2 \\ \vdots & \vdots & \ddots & \vdots \\ q_m & q_m & \cdots & q_m \end{bmatrix}$$

So Theorem 1(a) is true if P is a positive matrix. Suppose now P is regular but not positive; since P is regular, there is a power P^k of P that is positive. We need to show that $\lim\limits_{n\to\infty} M_n - m_n = 0$; the remainder of the proof follows exactly as above. No matter the value of n, there is always a multiple of k, say rk, with $rk < n \leq r(k+1)$. By the proof above $\lim\limits_{r\to\infty} M_{rk} - m_{rk} = 0$. But Equation 6 applies equally well to non-negative matrices, so $0 \leq M_n - m_n \leq M_{rk} - m_{rk}$, and $\lim\limits_{n\to\infty} M_n - m_n = 0$, proving part (a) of Theorem 1.

To prove part (b), it suffices to show that $\mathbf{q}$ is a probability vector. To see this note that since $(P^T)^n$ has row sums equal to 1 for any n, $(P^T)^n\mathbf{u} = \mathbf{u}$. Since $\lim\limits_{n\to\infty}(P^T)^n = \Pi^T$, it must be the case that $\Pi^T\mathbf{u} = \mathbf{u}$. Thus the rows of Π^T, and so also the columns of Π, must sum to 1 and $\mathbf{q}$ is a probability vector.

The proof of part (c) follows from the definition of matrix multiplication and the fact that P^n approaches Π by part (a). Let $\mathbf{x}_0$ be any probability vector. Then

$$\begin{aligned} \lim_{n\to\infty} P^n\mathbf{x}_0 &= \lim_{n\to\infty} P^n(x_1\mathbf{e}_1 + \ldots + x_m\mathbf{e}_m) \\ &= x_1(\lim_{n\to\infty} P^n\mathbf{e}_1) + \ldots + x_m(\lim_{n\to\infty} P^n\mathbf{e}_m) \\ &= x_1(\Pi\mathbf{e}_1) + \ldots + x_m(\Pi\mathbf{e}_m) = x_1\mathbf{q} + \ldots + x_m\mathbf{q} \\ &= (x_1 + \ldots + x_m)\mathbf{q} = \mathbf{q} \end{aligned}$$

since $\mathbf{x}_0$ is a probability vector.

To show part (d), we calculate $P\Pi$. First note that $\lim_{n\to\infty} P^{n+1} = \Pi$. But since $P^{n+1} = PP^n$, and $\lim_{n\to\infty} P^n = \Pi$, $\lim_{n\to\infty} P^{n+1} = P\Pi$. Thus $P\Pi = \Pi$, and any column of this matrix equation gives $P\mathbf{q} = \mathbf{q}$. Thus $\mathbf{q}$ is a probability vector that is also an eigenvector for P associated with the eigenvalue $\lambda = 1$. To show that this vector is unique, let $\mathbf{v}$ be any eigenvector for P associated with the eigenvalue $\lambda = 1$, which is also a probability vector. Then $P\mathbf{v} = \mathbf{v}$, and $P^n\mathbf{v} = \mathbf{v}$ for any n. But by part (c), $\lim_{n\to\infty} P^n\mathbf{v} = \mathbf{q}$, which can only happen if $\mathbf{v} = \mathbf{q}$. Thus $\mathbf{q}$ is unique. Note that this part of the proof has also shown that the eigenspace associated with the eigenvalue $\lambda = 1$ has dimension 1 (Exercise 29).

To prove part (e), let $\lambda \neq 1$ be an eigenvalue of P, and let $\mathbf{w}$ be an associated eigenvector. Assume that $\sum_{k=1}^{m} w_k \neq 0$. Without loss of generality, we may additionally assume that $\sum_{k=1}^{m} w_k = 1$ (as Exercise 30 proves). Then $P\mathbf{w} = \lambda\mathbf{w}$, so $P^n\mathbf{w} = \lambda^n\mathbf{w}$ for any n. By part (c), $\lim_{n\to\infty} P^n\mathbf{w} = \mathbf{q}$. Thus

$$\lim_{n\to\infty} \lambda^n\mathbf{w} = \mathbf{q} \tag{7}$$

Notice that Equation 6 can be true only if $\lambda = 1$. If $|\lambda| \geq 1$ and $\lambda \neq 1$, the left side of Equation 6 diverges; if $|\lambda| < 1$, the left side of Equation 7 must converge to $\mathbf{0} \neq \mathbf{q}$. This contradicts our assumption, so it must be the case that $\sum_{k=1}^{m} w_k = 0$. By part (a), $\lim_{n\to\infty} P^n\mathbf{w} = \Pi\mathbf{w}$. Since

$$\begin{aligned} \Pi\mathbf{w} &= \begin{bmatrix} \mathbf{q} & \mathbf{q} & \cdots & \mathbf{q} \end{bmatrix} \mathbf{w} \\ &= w_1\mathbf{q} + w_2\mathbf{q} + \cdots + w_m\mathbf{q} \\ &= (w_1 + w_2 + \cdots + w_m)\mathbf{q} = 0\mathbf{q} = \mathbf{0} \end{aligned}$$

then $\lim_{n\to\infty} P^n\mathbf{w} = \mathbf{0}$. Since $P^n\mathbf{w} = \lambda^n\mathbf{w}$ and $\mathbf{w} \neq \mathbf{0}$, $\lim_{n\to\infty} \lambda^n = 0$, and $|\lambda| < 1$. ■

Appendix 2: Probability

The purpose of this appendix is to provide some information from probability theory that can be used to develop a formal definition of a Markov chain and to prove some results from Chapter 10.

Probability

DEFINITION For each event E of the sample space S, the **probability** of E (denoted $P(E)$) is a number that has the following three properties:

a) $0 \leq P(E) \leq 1$

b) $P(S) = 1$

c) For any sequence of mutually exclusive events $E_1, E_2, \ldots$

$$P\left(\bigcup_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty} P(E_n)$$

Properties of Probability

1. $P(\emptyset) = 0$
2. $P(E^c) = 1 - P(E)$
3. $P(E \cup F) = P(E) + P(F) - P(E \cap F)$
4. If E and F are mutually exclusive events, $P(E \cup F) = P(E) + P(F)$

DEFINITION The **conditional probability** of E given F (denoted $P(E|F)$ is the probability that E occurs given that F has occurred is

$$P(E|F) = \frac{P(E \cap F)}{P(F)}$$

Law of Total Probability Let $F_1, F_2, \ldots$ be a sequence of mutually exclusive events for which

$$\bigcup_{n=1}^{\infty} F_n = S.$$

Then for any event E in the sample space S,

$$P(E) = \sum_{n=1}^{\infty} P(E|F_n)P(F_n)$$

Random Variables and Expectation

DEFINITION A random variable is a real-valued function defined on the sample space S. A **discrete random variable** is a random variable that takes on at most a countable number of possible values.

Only discrete random variables will be considered in this text; random variables that take on an uncountably infinite set of values are considered in advanced courses in probability theory. In Section 10.3 the expected value of a discrete random variable was defined. The expected value of a discrete random variable may also be defined using a function called the *probability mass function*.

DEFINITION The **probability mass function** p of a discrete random variable X is the real-valued function defined by $p(a) = P(X = a)$.

DEFINITION The **expected value** of a discrete random variable X is

$$E[X] = \sum_x xp(x)$$

where the sum is taken over all x with $p(x) > 0$.

Notice that if the random variable takes on the values $x_1, x_2, \ldots$ with positive probability, then the expected value of the random variable is

$$\sum_x xp(x) = x_1 P(X = x_1) + x_2 P(X = x_2) + \ldots$$

which matches the definition of expected value given in Section 10.3. Using the definition above it is straightforward to show that expected value has the following properties.

Properties of Expected Value For any real constant k and any discrete random variables X and Y,

1. $E[kX] = kE[X]$
2. $E[X + k] = E[X] + k$
3. $E[X + Y] = E[X] + E[Y]$
4. If f is a real-valued function, then $f(X)$ is a discrete random variable, and $E[f(X)] = \sum_x f(x)p(x)$, where the sum is taken over all x with $p(x) > 0$.

Just as probabilities can be affected by whether an event occurs, so can expected values.

DEFINITION Let X be a discrete random variable and let F be an event in the sample space S. Then the **conditional expected value** of X given F is

$$E[X|F] = \sum_x xP(X = x|F)$$

where the sum is taken over all x with $p(x) > 0$.

There is a law of total probability for expected value that will be used to prove a result from Chapter 10. Its statement and its proof follow.

Law of Total Probability for Expected Value Let $F_1, F_2, \ldots$ be a sequence of mutually exclusive events for which

$$\bigcup_{n=1}^{\infty} F_n = S.$$

Then for any discrete random variable X,

$$E[X] = \sum_{n=1}^{\infty} E[X|F_n]P(F_n)$$

Proof Let $F_1, F_2, \ldots$ be a sequence of mutually exclusive events for which $\bigcup_{n=1}^{\infty} F_n = S$, and let X be a discrete random variable. Then using the definition of expected value and the law of total probability,

$$\begin{aligned}
E[X] &= \sum_x xp(x) \\
&= \sum_x xP(X = x) \\
&= \sum_x x \sum_{n=1}^{\infty} P(X = x|F_n)P(F_n) \\
&= \sum_{n=1}^{\infty} P(F_n) \sum_x xP(X = x|F_n) \\
&= \sum_{n=1}^{\infty} E[X|F_n]P(F_n)
\end{aligned}$$

Markov chains

In Section 4.9, a Markov chain was defined as sequence of vectors. In order to understand Markov chains from a probabilistic standpoint, it is better to define a Markov chain as a sequence of random variables. To begin, consider any collection of random variables. This is called a stochastic process.

DEFINITION A **stochastic process** $\{X_n : n \in T\}$ is a collection of random variables.

NOTES:

1. The set T is called the **index set** for the stochastic process. The only set T that need be considered for this appendix is $T = \{0, 1, 2, 3, \ldots\}$, so the stochastic process can be described as the sequence of random variables $\{X_0, X_1, X_2, \ldots\}$. When $T = \{0, 1, 2, 3, \ldots\}$, the index is often identified with time and the stochastic process is called a discrete-time stochastic process. The random variable X_k is understood to be the stochastic process at time k.

2. It is assumed that the random variables in a stochastic process have a common range. This range is called the **state space** for the stochastic process. The state spaces in Chapter 10 are all finite, so the random variables X_k are all discrete random variables. If $X_k = i$, we will say that i is the **state** of the process at time k, or that the process is in state i at time (or step) k.

3. Notice that a stochastic process can be used to model movement between the states in the state space. For some element ω in the sample space S, the sequence $\{X_0(\omega), X_1(\omega), \ldots\}$ will be a sequence of states in the state space – a sequence that will potentially be different for each element in S. Usually the dependence on the sample space is ignored and the stochastic process is treated as a sequence of states, and the process is said to move (or transition) between those states as time proceeds.

4. Since a stochastic process is a sequence of random variables, the actual state that the process occupies at any given time cannot be known. The goal therefore is to find the probability that the process is in a particular state at a particular time. This amounts to finding the probability mass function of each random variable X_k in the sequence that is the stochastic process.

5. When a discrete-time stochastic process and the state space is finite, the probability mass function of each random variable X_k can be expressed as a probability vector $\mathbf{x}_k$. These probability vectors were used to define a Markov chain in Section 4.9.

In order for a discrete-time stochastic process $\{X_0, X_1, X_2, \ldots\}$ to be a Markov chain, the state of the process at time $n+1$ can depend only on the state of the process at time n. This is in contrast with a more general stochastic process, whose state at time n could depend on the entire history of the process. In terms of conditional probability, this property is

$$P(X_{n+1} = i | X_0 = j_0, X_1 = j_1, \ldots, X_n = j) = P(X_{n+1} = i | X_n = j)$$

The probability on the right side of this equation is called the transition probability from state j to state i. In general, this transition probability can change depending on the time n. This is not the case for Markov chains considered in this chapter: the transition probabilities do not change with time, so the transition probability from state j to state i is

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

A Markov chain with constant transition probabilities is called a **time-homogeneous** Markov chain. Its definition is thus

DEFINITION A time-homogeneous **Markov chain** is a discrete-time stochastic process whose transition probabilities satisfy

$$P(X_{n+1} = i | X_0 = j_0, X_1 = j_1, \ldots, X_n = j\} = P\{X_{n+1} = i | X_n = j\} = p_{ij}$$

for all times n and for all states i and j.

Using this definition it is clear that, if the number of states is finite, then a transition matrix can be constructed that has the properties assumed in Section 10.1.

Proofs of Theorems

Mean Return Times

Theorem 3 in Section 10.3 connected the steady-state vector for a Markov chain with the mean return time to a state of the chain. Here is a statement of this Theorem and a proof that relies on the law of total probability for expected value.

THEOREM 3 Let X_n, $n = 1, 2, \ldots$ be an irreducible Markov chain with finite state space S. Let n_{ij} be the number of steps until the chain first visits state i given that the chain starts in state j, and let $t_{ii} = E[n_{ii}]$. Then

$$t_{ii} = \frac{1}{q_i}$$

where q_i is the entry in the steady-state vector $\mathbf{q}$ corresponding to state i.

Proof To find an expression for t_{jj}, first produce an equation involving t_{ij} by considering the first step of the chain X_1. There are two possibilities: either $X_1 = i$ or $X_1 = k \neq i$. If $X_1 = i$, then it took exactly one step to visit state i and

$$E[n_{ij}|X_1 = i] = 1$$

If $X_1 = k \neq i$ then the chain will take one step to reach state k and then the expected number of steps the chain will make to first visit state i will be $E[n_{ik}] = t_{ik}$. Thus

$$E[n_{ij}|X_1 = k \neq i] = 1 + t_{ik}$$

By the law of total probability for expected value

$$\begin{aligned}
t_{ij} &= E[n_{ij}] \\
&= \sum_{k\in S} E[n_{ij}|X_1 = k]P(X_1 = k) \\
&= E[n_{ij}|X_1 = i]P(X_1 = i) + \sum_{k\neq i} E[n_{ij}|X_1 = k]P(X_1 = k) \\
&= 1 \cdot p_{ij} + \sum_{k\neq i}(1 + t_{ik})p_{kj} \\
&= p_{ij} + \sum_{k\neq i} p_{kj} + \sum_{k\neq i} t_{ik}p_{kj} \\
&= 1 + \sum_{k\neq i} t_{ik}p_{kj} \\
&= 1 + \sum_{k\in S} t_{ik}p_{kj} - t_{ii}p_{ij}
\end{aligned}$$

Let T be the matrix whose (i, j)-element is t_{ij} and let D be the diagonal matrix whose diagonal entries are t_{ii}. Then the final equality above may be written as

$$T_{ij} = 1 + (TP)_{ij} - (DP)_{ij} \tag{1}$$

If U is an appropriately sized matrix of ones, (1) can be written in matrix form as

$$T = U + TP - DP = U + (T - D)P \tag{2}$$

Multiplying each side of (2) by the steady-state vector $\mathbf{q}$ and recalling that $P\mathbf{q} = \mathbf{q}$ gives

$$T\mathbf{q} = U\mathbf{q} + (T - D)P\mathbf{q} = U\mathbf{q} + (T - D)\mathbf{q} = U\mathbf{q} + T\mathbf{q} - D\mathbf{q}$$

so

$$U\mathbf{q} = D\mathbf{q} \tag{3}$$

Consider the entries in each of the vectors in (3). Since U is a matrix of all ones,

$$U\mathbf{q} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} q_1 \\ q_2 \\ \vdots \\ q_n \end{bmatrix} = \begin{bmatrix} \sum_{k=1}^{n} q_k \\ \sum_{k=1}^{n} q_k \\ \vdots \\ \sum_{k=1}^{n} q_k \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

since $\mathbf{q}$ is a probability vector. Likewise

$$D\mathbf{q} = \begin{bmatrix} t_{11} & 0 & \cdots & 0 \\ 0 & t_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & t_{nn} \end{bmatrix} \begin{bmatrix} q_1 \\ q_2 \\ \vdots \\ q_n \end{bmatrix} = \begin{bmatrix} t_{11}q_1 \\ t_{22}q_2 \\ \vdots \\ t_{nn}q_n \end{bmatrix}$$

Equating corresponding entries in $U\mathbf{q}$ and $D\mathbf{q}$ gives $t_{ii}q_i = 1$, or

$$t_{ii} = \frac{1}{q_i}$$

■

Periodicity as a Class Property

In Section 10.4 it was stated that if two states belong to the same communication class then their periods must be equal. A proof of this result follows.

THEOREM Let i and j be two states of a Markov chain that are in the same communication class. Then the periods of i and j are equal.

Proof Suppose that i and j are in the same communication class for the Markov chain X, that state i has period d_i and that state j has period d_j. To simplify the exposition of the proof, the notation $(a^r)_{ij}$ will be used to refer to the (i, j) entry in the matrix A^r. Since i and j are in the same communication class, there exist positive integers m and n such that the $(p^m)_{ji} > 0$ and $(p^n)_{ij} > 0$. Let k be a positive integer such that $(p^k)_{jj} > 0$. In fact, $(p^{lk})_{jj} > 0$ for all integers $l > 1$. Now $(p^{n+lk+m})_{ii} > (p^n)_{ij}(p^{lk})_{jj}(p^m)_{ji} > 0$ for all integers $l > 1$, since a loop from state i to state i in $n + lk + m$ steps may be created in many ways, but one way is to proceed from state i to state j in n steps, then to loop from state j to state j l times using a loop of k steps each time, and then to return to state i in m steps. Since d_i is the period of state i, d_i must divide $n + lk + m$ for all integers $l > 1$. So d_i divides $n + k + m$ and $n + 2k + m$, and so divides $(n + 2k + m) - (n + k + m) = k$. Thus d_i is a common divisor of the set of all time steps k such that $(p^k)_{jj} > 0$. Since d_j is the *greatest* common divisor of the set of all time steps k such that $(p^k)_{jj} > 0$, $d_i \leq d_j$. A similar argument shows that $d_i \geq d_j$, so $d_i = d_j$. ■

The Fundamental Matrix

In Section 10.5, the number of visits v_{ij} to a transient state i that a Markov chain makes starting at the transient state j was studied. Specifically, the expected value $E[v_{ij}]$ was computed, and the fundamental matrix was defined as the matrix whose (i, j)-element is $m_{ij} = E[v_{ij}]$. The following theorem restates Theorem 6 from Section 10.5 in an equivalent form and provides a proof that relies on the law of total probability for expected value.

THEOREM 6 Let j and i be transient states of a Markov chain, let Q be that portion of the transition matrix which governs movement between transient states. Let v_{ij} be the number of visits that the chain will make to state i given that the chain starts in state j, and let $m_{ij} = E[v_{ij}]$. Then the matrix M whose (i, j)-element is m_{ij} satisfies the equation

$$M = (I - Q)^{-1}$$

Proof We produce an equation involving m_{ij} by conditioning on the first step of the chain X_1. We consider two cases: $i \neq j$ and $i = j$. First assume that $i \neq j$ and suppose that $X_1 = k$. Then we see that

$$E[v_{ij}|X_1 = k] = E[v_{ik}] \tag{4}$$

if $i \neq j$. Now assume that $i = j$. Then the previous analysis is valid, but we must add one visit to i since the chain was at state i at time 0. Thus

$$E[v_{ii}|X_1 = k] = 1 + E[v_{ik}] \tag{5}$$

We may combine equations (4) and (5) by introducing the following symbol, called the *Kronecker delta*:

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

Notice that δ_{ij} is the (i, j)-element in the identity matrix I. We can write equations (4) and (5) as

$$E[v_{ij}|X_1 = k] = \delta_{ij} + E[v_{ik}]$$

Thus by the law of total probability for expected value

$$\begin{aligned} m_{ij} &= E[v_{ij}] \\ &= \sum_{k \in S} E[v_{ij}|X_1 = k]P(X_1 = k) \end{aligned}$$

$$\begin{aligned}
&= \sum_{k \in S} (\delta_{ij} + E[v_{ik}]) P(X_1 = k) \\
&= \delta_{ij} \sum_{k \in S} P(X_1 = k) + \sum_{k \in S} E[v_{ik}] P(X_1 = k) \\
&= \delta_{ij} + \sum_{k \in S} E[v_{ik}] P(X_1 = k)
\end{aligned}$$

Now note that if k is a recurrent state, then $E[v_{ik}] = 0$. We thus only need to sum over transient states of the chain:

$$\begin{aligned}
m_{ij} &= \delta_{ij} + \sum_{k \text{transient}} E[v_{ik}] P(X_1 = k) \\
&= \delta_{ij} + \sum_{k \text{transient}} m_{ik} q_{kj}
\end{aligned}$$

since j and k are transient states and Q is defined in the statement of the Theorem. We may write the final equality above as

$$m_{ij} = I_{ij} + (MQ)_{ij}$$

or in matrix form as

$$M = I + MQ \tag{6}$$

We may rewrite (6) as

$$M - MQ = M(I - Q) = I$$

so $(I - Q)$ is invertible by the Invertible Matrix Theorem, and $M = (I - Q)^{-1}$. ■

Absorption Probabilities

In Section 10.5, the probability that the chain was absorbed into a particular absorbing state was studied. The Markov chain was assumed to have only transient and absorbing states, j is a transient state and i is an absorbing state of the chain. The probability a_{ij} that the chain is absorbed at state i given that the chain starts at state j was calculated, and it was shown that the matrix A whose (i, j)-element is a_{ij} satisfies $A = SM$, where M is the fundamental matrix and S is that portion of the transition matrix that governs movement from transient states to absorbing states. The following theorem restates this result, which was Theorem 7 in Section 10.5. An alternative proof of this result is given that relies on the law of total probability.

THEOREM 7 Consider a Markov chain with finite state space whose states are either absorbing or transient. Suppose that j be a transient state and that i is an absorbing state of the chain, and let a_{ij} be the probability that chain is absorbed at state i given that the chain starts in state j. Let A be the matrix whose (i, j)-element is a_{ij}. Then $A = SM$, where S and M are defined above.

Proof We again consider on the first step of the chain X_1. Let $X_1 = k$. There are three possibilities: k could be a transient state, k could be i, or k could be an absorbing state unequal to i. If k is transient, then

$$P\{\text{absorption at } i | X_1 = k\} = a_{ik}$$

If $k = i$, then

$$P\{\text{absorption at } i | X_1 = k\} = 1$$

while if k is absorbing state other than i,

$$P\{\text{absorption at } i | X_1 = k\} = 0$$

By the law of total probability,

$$\begin{aligned}
a_{ij} &= P\{\text{absorption at } i\} \\
&= \sum_k P\{\text{absorption at } i | X_1 = k\} P\{X_1 = k\} \\
&= 1 \cdot P\{X_1 = i\} + \sum_{k \text{ transient}} P\{\text{absorption at } i | X_1 = k\} P\{X_1 = k\} \\
&= p_{ij} + \sum_{k \text{ transient}} a_{ik} p_{kj}
\end{aligned}$$

Since j is transient and i is absorbing, $p_{ij} = s_{ij}$. Since in the final sum j and k are both transient, $p_{kj} = q_{kj}$. Thus the final equality may be written as

$$\begin{aligned}
a_{ij} &= s_{ij} + \sum_{k \text{ transient}} a_{ik} q_{kj} \\
&= s_{ij} + (AQ)_{ij}
\end{aligned}$$

or in matrix form as

$$A = S + AQ$$

This equation may be solved for A find that $A = SM$. ■