

# 7

## Markov Chains

### 7.1 INTRODUCTION

Suppose that a machine with two components is inspected every hour. A given component operating at  $t = n$  hours has probability  $p$  of failing before the next inspection at  $t = n + 1$ ; a component that is not operating at  $t = n$  has probability  $r$  of being repaired at  $t = n + 1$ , regardless of how long it has been out of service. (Each repair crew works a 1-hour day and refuses to inform the next crew of any insights it may have gained.) The components are assumed to fail and be repaired independently of each other.

The situation may be summarized as follows. If  $R_n$  is the number of components in operation at  $t = n$ , then

$$\begin{aligned}
 P\{R_{n+1} = 0 \mid R_n = 0\} &= (1 - r)^2 \\
 P\{R_{n+1} = 1 \mid R_n = 0\} &= 2r(1 - r) \\
 P\{R_{n+1} = 2 \mid R_n = 0\} &= r^2 \\
 P\{R_{n+1} = 0 \mid R_n = 1\} &= p(1 - r) \\
 P\{R_{n+1} = 1 \mid R_n = 1\} &= pr + (1 - p)(1 - r) \\
 P\{R_{n+1} = 2 \mid R_n = 1\} &= (1 - p)r \\
 P\{R_{n+1} = 0 \mid R_n = 2\} &= p^2 \\
 P\{R_{n+1} = 1 \mid R_n = 2\} &= 2p(1 - p) \\
 P\{R_{n+1} = 2 \mid R_n = 2\} &= (1 - p)^2
 \end{aligned} \tag{7.1.1}$$

## 212 MARKOV CHAINS

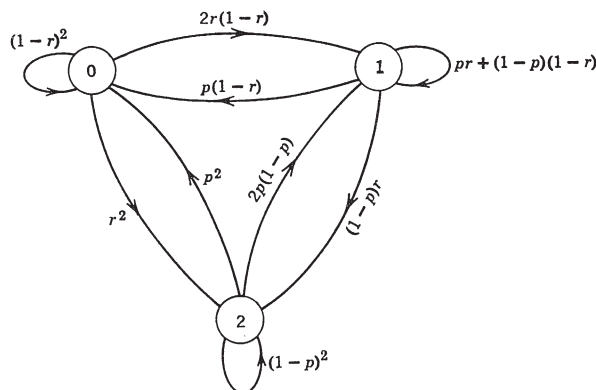


FIGURE 7.1.1 A Markov Chain.

For example, if component  $A$  is operating and component  $B$  is out of service at  $t = n$ , then in order to have one component in service at  $t = n + 1$ , either  $A$  fails and  $B$  is repaired between  $t = n$  and  $t = n + 1$ , or  $A$  does not fail and  $B$  is not repaired between  $t = n$  and  $t = n + 1$ . In order to have two components in service at  $t = n + 1$ ,  $A$  must not fail and  $B$  must be repaired. The other entries of (7.1.1) are derived similarly.

Thus, at time  $t = n$ , there are three possible states, 0, 1, and 2; to be in state  $i$  means that  $i$  components are in operation; that is,  $R_n = i$ . There are various transition probabilities  $p_{ij}$  indicating the probability of moving to state  $j$  at  $t = n + 1$ , given that we are in state  $i$  at  $t = n$ ; thus

$$p_{ij} = P\{R_{n+1} = j \mid R_n = i\}$$

(see Figure 7.1.1).

The  $p_{ij}$  may be arranged in the form of a matrix:

$$\Pi = [p_{ij}] = \begin{matrix} & \begin{matrix} 0 & 1 & 2 \end{matrix} \\ \begin{matrix} 0 \\ 1 \\ 2 \end{matrix} & \begin{bmatrix} (1-r)^2 & 2r(1-r) & r^2 \\ p(1-r) & pr + (1-p)(1-r) & (1-p)r \\ p^2 & 2p(1-p) & (1-p)^2 \end{bmatrix} \end{matrix}$$

Notice that  $\Pi$  is *stochastic*; that is, the elements are nonnegative and the row sums  $\sum_j p_{ij}$  are 1 for all  $i$ .

If  $R_n$  is the state of the process at time  $n$ , then, according to the way the problem is stated, if we know that  $R_0 = i_0, R_1 = i_1, \dots, R_n = i_n =$  (say) 2, we are in state 2 at  $t = n$ . Regardless of how we got there, once we know that  $R_n = 2$ , the probability that  $R_{n+1} = j$  is  $\binom{2}{j}(1-p)^j p^{2-j}$ ,  $j = 0, 1, 2$ . In

other words,

$$P\{R_{n+1} = i_{n+1} \mid R_0 = i_0, \dots, R_n = i_n\} = P\{R_{n+1} = i_{n+1} \mid R_n = i_n\} \quad (7.1.2)$$

This is the *Markov property*.

What we intend to show is that, given a description of a process in terms of states and transition probabilities (formally, given a stochastic matrix), we can construct in a natural way an infinite sequence of random variables satisfying (7.1.2). Assume that we are given a stochastic matrix  $\Pi = [p_{ij}]$ , where  $i$  and  $j$  range over the *state space*  $S$ , which we take to be a finite or infinite subset of the integers. Let  $p_i, i \in S$ , be a set of nonnegative numbers adding to 1 (the *initial distribution*). We specify the joint probability function of  $R_0, R_1, \dots, R_n$  as follows.

$$\begin{aligned} P\{R_0 = i_0, R_1 = i_1, \dots, R_n = i_n\} \\ = p_{i_0} p_{i_0 i_1} p_{i_1 i_2} \cdots p_{i_{n-1} i_n}, n = 1, 2, \dots, P\{R_0 = i_0\} = p_{i_0} \end{aligned} \quad (7.1.3)$$

If we sum the right side of (7.1.3) over  $i_{k+1}, \dots, i_n$ , we obtain  $p_{i_0} p_{i_0 i_1} \cdots p_{i_{k-1} i_k}$ , since  $\Pi$  is stochastic. But this coincides with the original specification of  $P\{R_0 = i_0, R_1 = i_1, \dots, R_k = i_k\}$ .

Thus the joint probability functions (7.1.3) are consistent, and therefore, by the discussion in Section 6.1, we can construct a sequence of random variables  $R_0, R_1, \dots$  such that for each  $n$  the joint probability function of  $(R_0, \dots, R_n)$  is given by (7.1.3).

Let us verify that the Markov property (7.1.2) is in fact satisfied. We have

$$\begin{aligned} P\{R_{n+1} = i_{n+1} \mid R_0 = i_0, \dots, R_n = i_n\} \\ = \frac{P\{R_0 = i_0, \dots, R_{n+1} = i_{n+1}\}}{P\{R_0 = i_0, \dots, R_n = i_n\}} \\ \quad \text{(assuming the denominator is not zero)} \\ = p_{i_n i_{n+1}} \quad \text{by (7.1.3)} \end{aligned}$$

But

$$\begin{aligned} P\{R_{n+1} = i_{n+1} \mid R_n = i_n\} &= \frac{P\{R_n = i_n, R_{n+1} = i_{n+1}\}}{P\{R_n = i_n\}} \\ &= \frac{\sum_{i_0, \dots, i_{n-1}} P\{R_0 = i_0, \dots, R_{n-1} = i_{n-1}, R_n = i_n, R_{n+1} = i_{n+1}\}}{\sum_{i_0, \dots, i_{n-1}} P\{R_0 = i_0, \dots, R_n = i_n\}} \\ &= \frac{\sum_{i_0, \dots, i_{n-1}} p_{i_0} p_{i_0 i_1} \cdots p_{i_{n-1} i_n} p_{i_n i_{n+1}}}{\sum_{i_0, \dots, i_{n-1}} p_{i_0} p_{i_0 i_1} \cdots p_{i_{n-1} i_n}} \\ &= p_{i_n i_{n+1}} \end{aligned}$$

establishing the Markov property.

## 214 MARKOV CHAINS

The sequence  $\{R_n\}$  is called the *Markov chain* corresponding to the matrix  $\Pi$  and the initial distribution  $\{p_i\}$ .  $\Pi$  is called the *transition matrix*, and the  $p_{ij}$  the *transition probabilities*, of the chain.

REMARK. The basic construction given here can be carried out if we have “nonstationary transition probabilities,” that is, if, instead of the “stationary transition probabilities”  $p_{ij}$ , we have, for each  $n = 0, 1, \dots$ , a stochastic matrix  $[{}_n p_{ij}]$ .  ${}_n p_{ij}$  is interpreted as the probability of moving to state  $j$  at time  $n + 1$  when the state at time  $n$  is  $i$ . We define  $P\{R_0 = i_0, \dots, R_n = i_n\} = p_{i_0 0} p_{i_0 i_1 1} p_{i_1 i_2 2} \cdots p_{i_{n-1} i_n n}$ . This yields a Markov process, a sequence of random variables satisfying the Markov property, with  $P\{R_n = i_n \mid R_0 = i_0, \dots, R_{n-1} = i_{n-1}\} = {}_{n-1} p_{i_{n-1} i_n}$ .

We begin the analysis of Markov chains by calculating the probability function of  $R_n$  in terms of the initial probabilities  $p_i$  and the matrix  $\Pi$ . Let  $p_j^{(n)} = P\{R_n = j\}$ ,  $n = 0, 1, \dots$ ,  $j \in S$  (thus  $p_j^{(0)} = p_j$ ). If we are in state  $i$  at time  $n - 1$ , we move to state  $j$  at time  $n$  with probability  $P\{R_n = j \mid R_{n-1} = i\} = p_{ij}$ ; thus, by the theorem of total probability,

$$p_j^{(n)} = \sum_i P\{R_{n-1} = i\} P\{R_n = j \mid R_{n-1} = i\} = \sum_i p_i^{(n-1)} p_{ij} \quad (7.1.4)$$

If  $v^{(n)} = (p_i^{(n)}, i \in S)$  is the “state distribution” or “state probability vector” at time  $n$ , then (7.1.4) may be written in matrix form as

$$v^{(n)} = v^{(n-1)} \Pi$$

Iterating this, we obtain

$$v^{(n)} = v^{(0)} \Pi^n \quad (7.1.5)$$

But suppose that we specify  $R_0 = i$ ; that is,  $p_i^{(0)} = 1$ ,  $p_j^{(0)} = 0$ ,  $i \neq j$ . Then  $v^{(0)}$  has a 1 in the  $i$ th coordinate and 0's elsewhere, so that by (7.1.5)  $v^{(n)}$  is simply row  $i$  of  $\Pi^n$ . Thus the element  $p_{ij}^{(n)}$  in row  $i$  and column  $j$  of  $\Pi^n$  is the probability that  $R_n = j$  when the initial state is  $i$ . In other words,

$$p_{ij}^{(n)} = P\{R_n = j \mid R_0 = i\} \quad (7.1.6)$$

(A slight formal quibble lies behind the phrase “in other words”; see Problem 1.) Because of (7.1.6),  $\Pi^n$  is called the  *$n$ -step transition matrix*: it follows immediately that  $\Pi^n$  is stochastic.

We shall be interested in the behavior of  $\Pi^n$  for large  $n$ . As an example, suppose that

$$\Pi = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{3}{4} & \frac{1}{4} \end{bmatrix}$$

We compute

$$\Pi^2 = \begin{bmatrix} \frac{5}{8} & \frac{3}{8} \\ \frac{9}{16} & \frac{7}{16} \end{bmatrix}, \quad \Pi^4 = \begin{bmatrix} \frac{15}{256} & \frac{102}{256} \\ \frac{153}{256} & \frac{103}{256} \end{bmatrix}$$

Thus  $p_{11}^{(4)} \sim p_{21}^{(4)}$  and  $p_{12}^{(4)} \sim p_{22}^{(4)}$ , so that the probability of being in state  $j$  at time  $t = 4$  is almost independent of the initial state. It appears as if, for large  $n$ , a “steady state” condition will be approached; the probability of being in a particular state  $j$  at  $t = n$  will be almost independent of the initial state at  $t = 0$ . Mathematically, we express this condition by saying that

$$\lim_{n \rightarrow \infty} p_{ij}^{(n)} = v_j, \quad i, j \in S \quad (7.1.7)$$

where  $v_j$  does not depend on  $i$ .

In Sections 7.4 and 7.5 we investigate the conditions under which (7.1.7) holds; it is not true for an arbitrary Markov chain.

Note that (7.1.7) is equivalent to the statement that  $\Pi^n \rightarrow$  a matrix with identical rows, the rows being  $(v_j, j \in S)$ .

► **Example 1.** Consider the simple random walk with no barriers. Then  $S =$  the integers and  $p_{i,i+1} = p, p_{i,i-1} = q = 1 - p, i = 0, \pm 1, \pm 2, \dots$

If there is an absorbing barrier at 0 (gambler’s ruin problem when the adversary has infinite capital),  $S =$  the nonnegative integers and  $p_{i,i+1} = p, p_{i,i-1} = q, i = 1, 2, \dots, p_{00} = 1$  (hence  $p_{0j} = 0, j \neq 0$ ).

If there are absorbing barriers at 0 and  $b$ , then  $S = \{0, 1, \dots, b\}$ ,  $p_{i,i+1} = p, p_{i,i-1} = q, i = 1, 2, \dots, b-1, p_{00} = p_{bb} = 1$ . ◀

► **Example 2.** Consider an infinite sequence of Bernoulli trials. Let state 1 (at  $t = n$ ) correspond to successes ( $S$ ) at  $t = n - 1$  and at  $t = n$ ; state 2 to success at  $t = n - 1$  and failure ( $F$ ) at  $t = n$ ; state 3 to failure at  $t = n - 1$  and success at  $t = n$ ; state 4 to failures at  $t = n - 1$  and at  $t = n$  (see Figure 7.1.2). We observe that  $\Pi^2$  has identical rows, the rows being

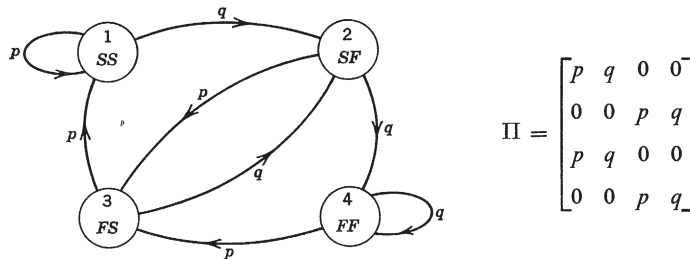


FIGURE 7.1.2

## 216 MARKOV CHAINS

$(p^2, pq, qp, q^2)$ . Hence  $\Pi^n$  has identical rows for  $n \geq 2$  (see Problem 2), so that  $\Pi^n$  approaches a limit. ◀

► **Example 3.** (A queueing example) Assume that customers are to be served at discrete times  $t = 0, 1, \dots$ , and at most one customer can be served at a given time. Say there are  $R_n$  customers before the completion of service at time  $n$ , and, in the interval  $[n, n + 1)$ ,  $N_n$  new customers arrive, where  $P\{N_n = k\} = p_k$ ,  $k = 0, 1, \dots$ . The number of customers before completion of service at time  $n + 1$  is

$$R_{n+1} = (R_n - 1)^+ + N_n$$

That is,

$$\begin{aligned} R_{n+1} &= R_n - 1 + N_n && \text{if } R_n \geq 1 \\ &= N_n && \text{if } R_n = 0 \end{aligned}$$

If the number of customers at time  $n + 1$  is  $\geq M$ , a new serving counter automatically opens and immediately serves all customers who are waiting and also those who arrive in the interval  $[n + 1, n + 2)$ ; thus  $R_{n+2} = 0$ .

The queueing process may be represented as a Markov chain with  $S =$  the nonnegative integers and transition matrix

$$\Pi = \begin{array}{c|cccccc} & 0 & 1 & 2 & 3 & \cdots & \\ \hline 0 & p_0 & p_1 & p_2 & \cdots & & \\ 1 & p_0 & p_1 & p_2 & \cdots & & \\ 2 & 0 & p_0 & p_1 & p_2 & \cdots & \\ 3 & 0 & 0 & p_0 & p_1 & p_2 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ M-1 & 0 & 0 & 0 & 0 & 0 & \cdots & p_0 & p_1 & p_2 & \cdots \\ M & 1 & 0 & \cdots & & & & & & & \\ M+1 & 1 & 0 & \cdots & & & & & & & \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{array}$$

## PROBLEMS

1. Consider a Markov chain  $\{R_n\}$  with transition matrix  $\Pi = [p_{ij}]$  and initial distribution  $\{p_i\}$ ; assume  $p_r > 0$ . Let  $\{T_n\}$  be a Markov chain with the same transition matrix, and initial distribution  $\{q_i\}$ , where  $q_r = 1$ ,  $q_j = 0$ ,  $j \neq r$ . Show that  $P\{R_n = j \mid R_0 = r\} = P\{T_n = j\} = p_{rj}^{(n)}$ ; this justifies (7.1.6).

## 7.2 STOPPING TIMES AND THE STRONG MARKOV PROPERTY 217

2. If  $\Pi$  is a transition matrix of a Markov chain, and  $\Pi^k$  has identical rows, show that  $\Pi^n$  has identical rows for all  $n \geq k$ . Similarly, if  $\Pi^k$  has a column all of whose elements are  $\geq \delta > 0$ , show that  $\Pi^n$  has this property for all  $n \geq k$ .
3. Let  $\{R_n\}$  be a Markov chain with state space  $S$ , and let  $g$  be a function from  $S$  to  $S$ . If  $g$  is one-to-one,  $\{g(R_n)\}$  is also a Markov chain (this simply amounts to relabeling the states). Give an example to show that if  $g$  is not one-to-one,  $\{g(R_n)\}$  need not have the Markov property.
4. If  $\{R_n\}$  is a Markov chain, show that

$$P\{R_{n+1} = i_1, \dots, R_{n+k} = i_k \mid R_n = i\} = p_{ii_1} p_{i_1 i_2} \cdots p_{i_{k-1} i_k}$$

## 7.2 STOPPING TIMES AND THE STRONG MARKOV PROPERTY

Let  $\{R_n\}$  be a Markov chain, and let  $T$  be the first time at which a particular state  $i$  is reached; set  $T = \infty$  if  $i$  is never reached. For example, if  $i = 3$  and  $R_0(\omega) = 4$ ,  $R_1(\omega) = 2$ ,  $R_2(\omega) = 2$ ,  $R_3(\omega) = 5$ ,  $R_4(\omega) = 3$ ,  $R_5(\omega) = 1$ ,  $R_6(\omega) = 3, \dots$ , then  $T(\omega) = 4$ . For our present purposes, the key feature of  $T$  is that if we examine  $R_0, R_1, \dots, R_k$ , we can come to a definite decision as to whether or not  $T = k$ . Formally, for each  $k = 0, 1, 2, \dots$ ,  $I_{\{T=k\}}$  can be expressed as  $g_k(R_0, R_1, \dots, R_k)$ , where  $g_k$  is a function from  $S^{k+1}$  to  $\{0, 1\}$ . A random variable  $T$ , whose possible values are the nonnegative integers together with  $\infty$ , that satisfies this condition for each  $k = 0, 1, \dots$  is said to be a *stopping time* for the chain  $\{R_n\}$ .

Now let  $T$  be the first time at which the state  $i$  is reached, as above. If we look at the sequence  $\{R_n\}$  after we arrive at  $i$ , in other words, the sequence  $R_T, R_{T+1}, \dots$ , it is reasonable to expect that we have a Markov chain with the same transition probabilities as the original chain. After all, if  $T = k$ , we are looking at the sequence  $R_k, R_{k+1}, \dots$ . However, since  $T$  is a random variable rather than a constant, there is something to be proved. We first introduce a new concept.

If  $T$  is a stopping time, an event  $A$  is said to be *prior to  $T$*  iff, whenever  $T = k$ , we can tell by examination of  $R_0, \dots, R_k$  whether or not  $A$  has occurred. Formally, for each  $k = 0, 1, \dots$ ,  $I_{A \cap \{T=k\}}$  can be expressed as  $h_k(R_0, R_1, \dots, R_k)$ , where  $h_k$  is a function from  $S^{k+1}$  to  $\{0, 1\}$ .

► **Example 1.** If  $T$  is a stopping time for the Markov chain  $\{R_n\}$ , define the random variable  $R_T$  as follows.

If  $T(\omega) = k$ , take  $R_T(\omega) = R_k(\omega)$ ,  $k = 0, 1, \dots$ .

If  $T(\omega) = \infty$ , take  $R_T(\omega) = c$ , where  $c$  is an arbitrary element not belonging to the state space  $S$ . If we like we can replace  $S$  by  $S \cup \{c\}$  and define  $p_{cc} = 1$ ,  $p_{cj} = 0$ ,  $j \in S$ .

## 218 MARKOV CHAINS

If  $A = \{R_{T-j} = i\}$ ,  $i \in S$ ,  $j = 0, 1, \dots$ , then  $A$  is prior to  $T$ . (Set  $R_{T-j} = R_0$  if  $j > T$ .)

To see this, note that  $\{R_{T-j} = i\} \cap \{T = k\} = \{R_{k-j} = i\} \cap \{T = k\}$ ; examination of  $R_0, \dots, R_k$  determines the value of  $R_{k-j}$ , and also determines whether or not  $T = k$ . ◀

► **Example 2.** If  $T$  is a stopping time for the Markov chain  $\{R_n\}$ , then  $\{T = r\}$  is prior to  $T$  for all  $r = 0, 1, \dots$ . For

$$\begin{aligned} \{T = r\} \cap \{T = k\} &= \emptyset & \text{if } r \neq k \\ &= \{T = k\} & \text{if } r = k \end{aligned}$$

In either case  $I_{\{T=r\} \cap \{T=k\}}$  is a function of  $R_0, \dots, R_k$ , since  $T$  is a stopping time. ◀

**Theorem 1.** Let  $T$  be a stopping time for the Markov chain  $\{R_n\}$ . If  $A$  is prior to  $T$ , then

$$\begin{aligned} P(A \cap \{R_T = i, R_{T+1} = i_1, \dots, R_{T+k} = i_k\}) \\ = P(A \cap \{R_T = i\})p_{ii_1}p_{i_1i_2} \cdots p_{i_{k-1}i_k}(i, i_1, \dots, i_k \in S) \end{aligned}$$

PROOF. The probability of the set on the left is

$$\begin{aligned} \sum_{n=0}^{\infty} P(A \cap \{T = n, R_n = i, R_{n+1} = i_1, \dots, R_{n+k} = i_k\}) \\ = \sum_{n=0}^{\infty} P(A \cap \{T = n, R_n = i\}) \\ \times P(R_{n+1} = i_1, \dots, R_{n+k} = i_k \mid A \cap \{T = n, R_n = i\}) \end{aligned}$$

(Actually we sum only over those  $n$  for which  $P(A \cap \{T = n, R_n = i\}) > 0$ .)  
Now

$$\begin{aligned} P(R_{n+1} = i_1, \dots, R_{n+k} = i_k \mid A \cap \{T = n, R_n = i\}) \\ = P\{R_{n+1} = i_1, \dots, R_{n+k} = i_k \mid R_n = i\} \end{aligned}$$

since  $I_{A \cap \{T=n\}}$  is a function of  $R_0, R_1, \dots, R_n$  (see Problem 1)

$$= p_{ii_1}p_{i_1i_2} \cdots p_{i_{k-1}i_k}$$

(Problem 4, Section 7.1). Thus the summation becomes

$$\begin{aligned} \sum_{n=0}^{\infty} P(A \cap \{T = n, R_n = i\})p_{ii_1}p_{i_1i_2} \cdots p_{i_{k-1}i_k} \\ = P(A \cap \{R_T = i\})p_{ii_1}p_{i_1i_2} \cdots p_{i_{k-1}i_k} \end{aligned}$$

and the result follows.



## 7.2 STOPPING TIMES AND THE STRONG MARKOV PROPERTY 219

**Theorem 2 (Strong Markov Property).** Let  $T$  be a stopping time for the Markov chain  $\{R_n\}$ . Then

$$(a) \quad P\{R_{T+1} = i_1, \dots, R_{T+k} = i_k \mid R_T = i\} = p_{ii_1} p_{i_1 i_2} \cdots p_{i_{k-1} i_k} \quad \text{if} \\ P\{R_T = i\} > 0 \quad (i, i_1, \dots, i_k \in S)$$

(b) If  $A$  is prior to  $T$ , then

$$\begin{aligned} P(R_{T+1} = i_1, \dots, R_{T+k} = i_k \mid A \cap \{R_T = i\}) \\ = P\{R_{T+1} = i_1, \dots, R_{T+k} = i_k \mid R_T = i\} \\ \text{if } P(A \cap \{R_T = i\}) > 0 \quad (i, i_1, \dots, i_k \in S) \end{aligned}$$

PROOF. (a) follows from Theorem 1 by taking  $A = \Omega$ . (b) follows upon dividing the equality of Theorem 1 by  $P(A \cap \{R_T = i\})$  and using (a).

Thus the sequence  $R_T, R_{T+1}, \dots$  has essentially the same properties as the original sequence  $R_0, R_1, \dots$ .

REMARK. The strong Markov property reduces to the ordinary Markov property (7.1.2) if we set  $k = 1$ ,  $T \equiv n$ , and  $A = \{R_0 = j_0, \dots, R_{n-1} = j_{n-1}\}$ . For  $T$  is a stopping time since

$$\begin{aligned} I_{\{T=k\}} &= g_k(R_0, \dots, R_k) = 0 & \text{if } k \neq n \\ &= 1 & \text{if } k = n \end{aligned}$$

and  $A$  is prior to  $T$  since

$$\begin{aligned} I_{A \cap \{T=k\}} &= h_k(R_0, \dots, R_k) \\ &= 1 & \text{if } k = n \text{ and } R_0 = j_0, \dots, R_{n-1} = j_{n-1} \\ &= 0 & \text{otherwise} \end{aligned}$$

## PROBLEMS

1. Let  $\{R_n\}$  be a Markov chain. If  $D$  is an event whose occurrence or nonoccurrence is determined by examination of  $R_0, \dots, R_n$ , that is,  $I_D$  is a function of  $R_0, \dots, R_n$ , or, equivalently,  $D$  is of the form  $\{(R_0, \dots, R_n) \in B\}$  for some  $B \subset S^{n+1}$ , show that

$$\begin{aligned} P(R_{n+1} = i_1, \dots, R_{n+k} = i_k \mid D \cap \{R_n = i\}) \\ = P\{R_{n+1} = i_1, \dots, R_{n+k} = i_k \mid R_n = i\} \quad \text{if } P(D \cap \{R_n = i\}) > 0 \end{aligned}$$

2. If  $\{R_n\}$  is a Markov chain, show that the “reversed sequence”  $\dots R_n, R_{n-1}, R_{n-2}, \dots$  also has the Markov property.

### 7.3 CLASSIFICATION OF STATES

In this section we examine various modes of behavior of Markov chains. A key to the analysis is the following result. We consider a fixed Markov chain  $\{R_n\}$  throughout.

**Theorem 1 (First Entrance Theorem).** *Let  $f_{ii}^{(n)}$  be the probability that the first return to  $i$  will occur at time  $n$ , when the initial state is  $i$ , that is,*

$$f_{ii}^{(n)} = P\{R_n = i, R_k \neq i \text{ for } 1 \leq k \leq n-1 \mid R_0 = i\}, \quad n = 1, 2, \dots$$

*If  $i \neq j$ , let  $f_{ij}^{(n)}$  be the probability that the first visit to state  $j$  will occur at time  $n$ , when the initial state is  $i$ ; that is,*

$$f_{ij}^{(n)} = P\{R_n = j, R_k \neq j \text{ for } 1 \leq k \leq n-1 \mid R_0 = i\}, \quad n = 1, 2, \dots$$

*Then*

$$p_{ij}^{(n)} = \sum_{k=1}^n f_{ij}^{(k)} p_{ij}^{(n-k)}, \quad n = 1, 2, \dots$$

**PROOF.** Intuitively, if we are to be in state  $j$  after  $n$  steps, we must reach  $j$  for the first time at step  $k$ ,  $1 \leq k \leq n$ . After this happens, we are in state  $j$  and must be in state  $j$  again after the  $n-k$  remaining steps. For a formal proof, we use the strong Markov property. Assume that the initial state is  $i$ , and let  $T$  be the time of the first visit to  $j$  ( $T = \min \{k \geq 1 : R_k = j\}$  if  $R_k = j$  for some  $k = 1, 2, \dots$ ;  $T = \infty$  if  $R_k \neq j$  for all  $k = 1, 2, \dots$ ). Then

$$\begin{aligned} P\{R_n = j\} &= \sum_{k=1}^n P\{R_n = j, T = k\} \\ &= \sum_{k=1}^n P\{T = k, R_{T+n-k} = j\} \end{aligned}$$

But

$$P\{T = k, R_{T+n-k} = j\} = P\{T = k\} P\{R_{T+n-k} = j \mid T = k\}$$

and since  $\{T = k\} = \{T = k, R_T = j\}$ ,

$$\begin{aligned} P\{R_{T+n-k} = j \mid T = k\} &= P\{R_{T+n-k} = j \mid R_T = j, T = k\} \\ &= P\{R_{T+n-k} = j \mid R_T = j\} \end{aligned}$$

by Theorem 2b of Section 7.2

$$= p_{jj}^{(n-k)} \quad \text{by Theorem 2a of Section 7.2}$$

Since  $P\{T = k\} = f_{ij}^{(k)}$ , the result follows.

Now let

$$f_{ii} = \sum_{n=1}^{\infty} f_{ii}^{(n)}$$

$f_{ii}$  is the probability of eventual return to state  $i$  when the initial state is  $i$ .

## 7.3 CLASSIFICATION OF STATES 221

**Theorem 2.** *If the initial state is  $i$ , the probability of returning to  $i$  at least  $r$  times is  $(f_{ii})^r$ .*

**PROOF.** The result is immediate if  $r = 1$ . If it holds when  $r = m - 1$ , let  $T$  be the time of first return to  $i$ . Then, starting from  $i$ ,  $P\{\text{return to } i \text{ at least } m \text{ times}\} = \sum_{k=1}^{\infty} P\{T = k, \text{ at least } m - 1 \text{ returns after } T\}$ . But

$$\begin{aligned} & P\{T = k, \text{ at least } m - 1 \text{ returns after } T\} \\ &= P\{T = k\} P\{R_{T+1}, R_{T+2}, \dots \text{ returns to } i \text{ at least } m - 1 \text{ times} \mid T = k\} \\ &= P\{T = k\} \\ &\quad \times P\{R_{T+1}, R_{T+2}, \dots \text{ returns to } i \text{ at least } m - 1 \text{ times} \mid R_T = i, T = k\} \end{aligned}$$

By the strong Markov property this may be written as

$$\begin{aligned} & P\{T = k\} P\{R_{T+1}, R_{T+2}, \dots \text{ returns to } i \text{ at least } m - 1 \text{ times} \mid R_T = i\} \\ &= P\{T = k\} P\{R_1, R_2, \dots \text{ returns to } i \text{ at least } m - 1 \text{ times} \mid R_0 = i\} \\ &= f_{ii}^{(k)} (f_{ii})^{m-1} \quad \text{by the induction hypothesis} \end{aligned}$$

Thus the probability of returning to  $i$  at least  $m$  times is

$$\sum_{k=1}^{\infty} f_{ii}^{(k)} (f_{ii})^{m-1} = f_{ii} (f_{ii})^{m-1}$$

which is the desired result.

**COROLLARY.** Let the initial state be  $i$ . If  $f_{ii} = 1$ , the probability of returning to  $i$  infinitely often is 1. If  $f_{ii} < 1$ , the probability of returning to  $i$  infinitely often is 0.

**PROOF.** The events  $\{\text{return to } i \text{ at least } r \text{ times}\}$ ,  $r = 1, 2, \dots$  form a contracting sequence whose intersection is  $\{\text{return to } i \text{ infinitely often}\}$ . Thus the probability of returning to  $i$  infinitely often is  $\lim_{r \rightarrow \infty} (f_{ii})^r$ , and the result follows.

**DEFINITION.** If  $f_{ii} = 1$ , we say that the state  $i$  is *recurrent* or *persistent*; if  $f_{ii} < 1$ , we say that  $i$  is *transient*.

It is useful to have a criterion for recurrence in terms of the probabilities  $p_{ii}^{(n)}$ , since these numbers are often easier to handle than the  $f_{ii}^{(n)}$ .

**Theorem 3.** *The state  $i$  is recurrent iff  $\sum_{n=1}^{\infty} p_{ii}^{(n)} = \infty$ .*

**PROOF.** By the first entrance theorem,

$$p_{ii}^{(n)} = \sum_{k=1}^n f_{ii}^{(k)} p_{ii}^{(n-k)}$$

## 222 MARKOV CHAINS

so that

$$\begin{aligned}\sum_{n=1}^{\infty} p_{ii}^{(n)} &= \sum_{n=1}^{\infty} \sum_{k=1}^n f_{ii}^{(k)} p_{ii}^{(n-k)} = \sum_{k=1}^{\infty} f_{ii}^{(k)} \sum_{n=k}^{\infty} p_{ii}^{(n-k)} \\ &= f_{ii} \sum_{n=0}^{\infty} p_{ii}^{(n)}\end{aligned}$$

Thus

$$\sum_{n=1}^{\infty} p_{ii}^{(n)} = f_{ii} \left( 1 + \sum_{n=1}^{\infty} p_{ii}^{(n)} \right)$$

Hence, if

$$\sum_{n=1}^{\infty} p_{ii}^{(n)} < \infty$$

then  $f_{ii} < 1$ , so that  $i$  is transient. Now

$$\sum_{n=1}^N p_{ii}^{(n)} = \sum_{n=1}^N \sum_{k=1}^n f_{ii}^{(k)} p_{ii}^{(n-k)} = \sum_{k=1}^N f_{ii}^{(k)} \sum_{n=k}^N p_{ii}^{(n-k)} \leq \sum_{k=1}^N f_{ii}^{(k)} \sum_{r=0}^N p_{ii}^{(r)}$$

Thus

$$f_{ii} = \sum_{k=1}^{\infty} f_{ii}^{(k)} \geq \sum_{k=1}^N f_{ii}^{(k)} \geq \frac{\sum_{n=1}^N p_{ii}^{(n)}}{\sum_{r=0}^N p_{ii}^{(r)}} \rightarrow 1 \quad \text{as } N \rightarrow \infty \quad \text{if } \sum_{n=1}^{\infty} p_{ii}^{(n)} = \infty$$

Therefore  $\sum_{n=1}^{\infty} p_{ii}^{(n)} = \infty$  implies that  $f_{ii} = 1$ , so that  $i$  is recurrent.

We denote by  $f_{ij}$  the probability of ever visiting  $j$  at some future time, starting from  $i$ ; that is,

$$f_{ij} = \sum_{k=1}^{\infty} f_{ij}^{(k)}$$

**Theorem 4.** *If  $j$  is a transient state and  $i$  an arbitrary state, then*

$$\sum_{n=1}^{\infty} p_{ij}^{(n)} < \infty$$

hence

$$p_{ij}^{(n)} \rightarrow 0 \quad \text{as } n \rightarrow \infty$$

PROOF. By the first entrance theorem,

$$\begin{aligned}\sum_{n=1}^{\infty} p_{ij}^{(n)} &= \sum_{n=1}^{\infty} \sum_{k=1}^n f_{ij}^{(k)} p_{jj}^{(n-k)} = \sum_{k=1}^{\infty} f_{ij}^{(k)} \sum_{n=0}^{\infty} p_{jj}^{(n)} \\ &= f_{ij} \sum_{n=0}^{\infty} p_{jj}^{(n)}\end{aligned}$$

## 7.3 CLASSIFICATION OF STATES 223

But  $f_{ij}$ , being a probability, is  $\leq 1$ , and  $\sum_{n=0}^{\infty} p_{ij}^{(n)} < \infty$  by Theorem 3; the result follows.

**REMARK.** If  $j$  is a transient state and the initial state of the chain is  $i$ , then, by Theorem 4 above and the Borel-Cantelli lemma (Theorem 4, Section 6.6), with probability 1 the state  $j$  will be visited only finitely many times. Alternatively, we may use the argument of Theorem 2 (with initial state  $i$  and  $T =$  the time of first visit to  $j$ ) to show that the probability that  $j$  will be visited at least  $m$  times is  $f_{ij}(f_{jj})^{m-1}$ . Now  $f_{jj} < 1$  since  $j$  is transient, and thus, if we let  $m \rightarrow \infty$ , we find (as in the corollary to Theorem 2) that the probability that  $j$  will be visited infinitely often is 0.

In fact this result holds for an arbitrary initial distribution. For

$$\begin{aligned} P\{R_n = j \text{ for infinitely many } n\} \\ = \sum_i P\{R_0 = i\} P\{R_n = j \text{ for infinitely many } n \mid R_0 = i\} = 0 \end{aligned}$$

It follows that if  $B$  is a finite set of transient states, then the probability of remaining in  $B$  forever is 0. For if  $R_n \in B$  for all  $n$ , then, since  $B$  is finite, we have, for some  $j \in B$ ,  $R_n = j$  for infinitely many  $n$ ; thus

$$P\{R_n \in B \text{ for all } n\} \leq \sum_{j \in B} P\{R_n = j \text{ for infinitely many } n\} = 0$$

One of our main problems will be to classify the states of a given chain as to recurrence or nonrecurrence. The first step is to introduce an equivalence relation on the state space and show that within each equivalence class all states are of the same type.

**DEFINITION.** If  $i$  and  $j$  are distinct states, we say that  $i$  leads to  $j$  iff  $f_{ij} > 0$ ; that is, it is possible to reach  $j$ , starting from  $i$ . Equivalently,  $i$  leads to  $j$  iff  $p_{ij}^{(n)} > 0$  for some  $n \geq 1$ . By convention,  $i$  leads to itself. We say that  $i$  and  $j$  communicate iff  $i$  leads to  $j$  and  $j$  leads to  $i$ .

We define an *equivalence relation* on the state space  $S$  by taking  $i$  equivalent to  $j$  iff  $i$  and  $j$  communicate. (It is not difficult to verify that we have a legitimate equivalence relation.) The next theorem shows that *recurrence or nonrecurrence is a class property*: that is, if one state in a given equivalence class is recurrent, all states are recurrent.

**Theorem 5.** If  $i$  is recurrent and  $i$  leads to  $j$ , then  $j$  is recurrent. Furthermore,  $f_{ij} = f_{ji} = 1$ . In fact, if  $f'_{ij}$  is the probability that  $j$  will be visited infinitely often when the initial state is  $i$ , then  $f'_{ij} = f'_{ji} = 1$ .

## 224 MARKOV CHAINS

PROOF. Start in state  $i$ , and let  $T$  be the time of the first visit to  $j$ . Then

$$\begin{aligned}
 1 &= \sum_{k=1}^{\infty} P\{T = k\} + P\{T = \infty\} \\
 &= \sum_{k=1}^{\infty} P\{T = k, \text{infinitely many visits to } i \text{ after } T\} + P\{T = \infty\} \\
 &\hspace{25em} \text{since } i \text{ is recurrent} \\
 &= \sum_{k=1}^{\infty} P\{T = k\} P\{R_{T+1}, R_{T+2}, \dots \text{ visits } i \text{ infinitely often} \mid T = k, R_T = j\} \\
 &\hspace{25em} + 1 - f_{ij} \\
 &= \sum_{k=1}^{\infty} f_{ij}^{(k)} P\{R_1, R_2, \dots \text{ visits } i \text{ infinitely often} \mid R_0 = j\} + 1 - f_{ij}
 \end{aligned}$$

Thus

$$1 = f_{ij} f'_{ji} + 1 - f_{ij}, \quad \text{or} \quad f_{ij} = f_{ij} f'_{ji}$$

Since  $f_{ij} > 0$  by hypothesis,  $f'_{ji} = 1$ .

Now if  $p_{ij}^{(r)} > 0$ ,  $p_{ji}^{(s)} > 0$ , then

$$p_{jj}^{(n+r+s)} \geq p_{ji}^{(s)} p_{ii}^{(n)} p_{ij}^{(r)}$$

since one way of going from  $j$  to  $j$  in  $n + r + s$  steps is to go from  $j$  to  $i$  in  $s$  steps, from  $i$  to  $i$  in  $n$  steps, and finally from  $i$  to  $j$  in  $r$  steps. It follows from Theorem 3 that  $\sum_{n=1}^{\infty} p_{jj}^{(n)} = \infty$ ; hence  $j$  is recurrent.

Finally, we have  $j$  recurrent and  $f_{ji} > 0$ . By the above argument, with  $i$  and  $j$  interchanged,  $f'_{ij} = 1$ . Since  $f_{ij} \geq f'_{ij}$ ,  $f_{ji} \geq f'_{ji}$ , it follows that  $f_{ij} = f_{ji} = 1$  and the theorem is proved.

**Theorem 6.** *If a finite chain (i.e.,  $S$  a finite set), it is not possible for all states to be transient.*

*In particular, if every state in a finite chain can be reached from every other state, so that there is only one equivalence class (namely  $S$ ), then all states are recurrent.*

PROOF. If  $S = \{1, 2, \dots, r\}$ , then  $\sum_{j=1}^r p_{ij}^{(n)} = 1$  for all  $n$ . Let  $n \rightarrow \infty$ . By Theorem 4 and the fact that the limit of a finite sum is the sum of the limits, we have  $0 = \sum_{j=1}^r \lim_{n \rightarrow \infty} p_{ij}^{(n)} = \lim_{n \rightarrow \infty} \sum_{j=1}^r p_{ij}^{(n)} = 1$ , a contradiction.

In the case of a finite chain, it is easy to decide whether or not a given class is recurrent; we shall see how to do this in a moment.

**DEFINITION.** A nonempty subset  $C$  of the state space  $S$  is said to be *closed* iff it is not possible to leave  $C$ ; that is,  $\sum_{j \in C} p_{ij} = 1$  for all  $i \in C$ . Notice that if  $C$  is closed, then the submatrix  $[p_{ij}]$ ,  $i, j \in C$ , is stochastic; hence so is  $[p_{ij}^{(n)}]$ ,  $i, j \in C$ .

**Theorem 7.**  $C$  is closed iff for all  $i \in C$  ( $i$  leads to  $j$  implies  $j \in C$ ).

**PROOF.** Let  $C$  be closed. If  $i \in C$  and  $i$  leads to  $j$ , then  $p_{ij}^{(n)} > 0$  for some  $n$ . If  $j \notin C$ , then  $\sum_{k \in C} p_{ik}^{(n)} < 1$ , a contradiction. Conversely, if the condition is satisfied and  $C$  is not closed, then  $\sum_{j \in C} p_{ij} < 1$  for some  $i \in C$ ; hence  $p_{ij} > 0$  for some  $i \in C, j \notin C$ , a contradiction.

**Theorem 8.**

(a) Let  $C$  be a recurrent class. Then  $C$  is closed.

(b) If  $C$  is any equivalence class, no proper subset of  $C$  is closed.

(c) In a finite chain, every closed equivalence class  $C$  is recurrent.

Thus, in a finite chain, the recurrent classes are simply those classes that are closed.

**PROOF.**

(a) Let  $C$  be a recurrent class. If  $C$  is not closed, then by Theorem 7 we have some  $i \in C$  leading to a  $j \notin C$ . But by Theorem 5,  $i$  and  $j$  communicate, and so  $i$  is equivalent to  $j$ . This contradicts  $i \in C, j \notin C$ .

(b) Let  $D$  be a (nonempty) proper subset of the arbitrary equivalence class  $C$ . Pick  $i \in D$  and  $j \in C, j \notin D$ . Then  $i$  leads to  $j$ , since both states belong to the same equivalence class. Thus  $D$  cannot be closed.

(c) Consider  $C$  itself as a chain; this is possible since  $\sum_{j \in C} p_{ij} = 1, i \in C$ . (We are simply restricting the original transition matrix to  $C$ .) By Theorem 6 and the fact that recurrence is a class property,  $C$  is recurrent.

► **Example 1.** Consider the chain of Figure 7.3.1. (An arrow from  $i$  to  $j$  indicates that  $p_{ij} > 0$ .) There are three equivalence classes,  $C_1 = \{1, 2\}$ ,  $C_2 = \{3, 4, 5\}$ , and  $C_3 = \{6\}$ . By Theorem 8,  $C_2$  is recurrent and  $C_1$  and  $C_3$  are transient.

There is no foolproof method for classifying the states of an infinite chain, but in some cases an analysis can be done quickly. Consider the chain of Example 3, Section 7.1, and assume that all  $p_k > 0$ . Then every state is reachable from every other state, so that the entire state space forms a

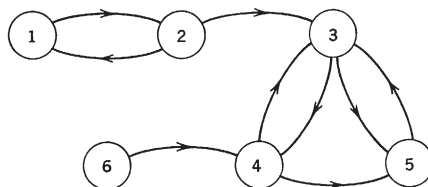


FIGURE 7.3.1 A Finite Markov Chain.

## 226 MARKOV CHAINS

single equivalence class. We claim that the class is recurrent. For assume the contrary; then all states are transient. Let 0 be the initial state; by the remark after Theorem 4, the set  $B = \{0, 1, \dots, M-1\}$  will be visited only finitely many times; that is, eventually  $R_n \geq M$  (with probability 1). But by definition of the transition matrix, if  $R_n \geq M$  then  $R_{n+1} = 0$ , a contradiction. ◀

We now describe another basic class property, that of periodicity.

If  $p_{ii}^{(n)} > 0$  for some  $n \geq 1$ , that is, if starting from  $i$  it is possible to return to  $i$ , we define the *period* of  $i$  (notation:  $d_i$ ) as the greatest common divisor of the set of positive integers  $n$  such that  $p_{ii}^{(n)} > 0$ . Equivalently, the period of  $i$  is the greatest common divisor of the set of positive integers  $n$  such that  $f_{ii}^{(n)} > 0$  (see Problem 1c).

**Theorem 9.** *If the distinct states  $i$  and  $j$  are in the same equivalence class, they have the same period.*

PROOF. Since  $i$  and  $j$  communicate, each has a period. If  $p_{ij}^{(r)} > 0$ ,  $p_{ji}^{(s)} > 0$ , then  $p_{jj}^{(n+r+s)} \geq p_{ji}^{(s)} p_{ii}^{(n)} p_{ij}^{(r)}$  (see the argument of Theorem 5). Set  $n = 0$  to obtain  $p_{jj}^{(r+s)} > 0$ , so that  $r + s$  is a multiple of  $d_j$ . Thus if  $n$  is not a multiple of  $d_j$  (so neither is  $n + r + s$ ) we have  $p_{jj}^{(n+r+s)} = 0$ ; hence  $p_{ii}^{(n)} = 0$ . But this says that if  $p_{ii}^{(n)} > 0$  then  $n$  is a multiple of  $d_j$ ; hence  $d_j \leq d_i$ . By a symmetrical argument,  $d_i \leq d_j$ .

The transitions from state to state within a closed equivalence class  $C$  of period  $d > 1$ , although random, have a certain cyclic pattern, which we now describe.

Let  $i, j \in C$ ; if  $p_{ij}^{(r)} > 0$  and  $p_{ji}^{(s)} > 0$ , let  $t$  be such that  $p_{ji}^{(t)} > 0$ . Then  $p_{ii}^{(r+t)} \geq p_{ij}^{(r)} p_{ji}^{(t)} > 0$ ; hence  $d$  divides  $r + t$ . Similarly,  $d$  divides  $s + t$ , and so  $d$  divides  $s - r$ .

Thus, if  $r = ad + b$ ,  $a$  and  $b$  integers,  $0 \leq b \leq d - 1$ , then  $s = cd + b$  for some integer  $c$ . Consequently, if  $i$  leads to  $j$  in  $n$  steps, then  $n$  is of the form  $ed + b$ , that is,

$$n \equiv b \pmod{d}$$

where the integer  $b$  depends on the states  $i$  and  $j$  but is independent of  $n$ .

Now fix  $i \in C$  and define

$$C_0 = \{j \in C: p_{ij}^{(n)} > 0 \text{ implies } n \equiv 0 \pmod{d}\}$$

$$C_1 = \{j \in C: p_{ij}^{(n)} > 0 \text{ implies } n \equiv 1 \pmod{d}\}$$

.

.

.

$$C_{d-1} = \{j \in C: p_{ij}^{(n)} > 0 \text{ implies } n \equiv d - 1 \pmod{d}\}$$



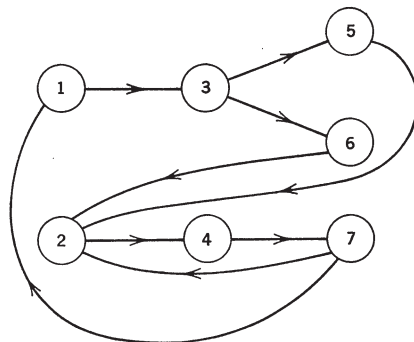


FIGURE 7.3.2 A Periodic Chain.

Then

$$C = \bigcup_{j=0}^{d-1} C_j$$

**Theorem 10.** If  $k \in C_t$  and  $p_{kj} > 0$ , then  $j \in C_{t+1}$  (with indices reduced mod  $d$ ; i.e.,  $C_d = C_0$ ,  $C_{d+1} = C_1$ , etc.). Thus, starting from  $i$ , the chain moves from  $C_0$  to  $C_1$  to  $\dots$  to  $C_{d-1}$  back to  $C_0$ , and so on. The  $C_j$  are called the cyclically moving subclasses of  $C$ .

**PROOF.** Choose an  $n$  such that  $p_{ik}^{(n)} > 0$ . Then  $n$  is of the form  $ad + t$ . Now  $p_{ij}^{(n+1)} \geq p_{ik}^{(n)} p_{kj} > 0$ ; hence  $i$  leads to  $j$ , and therefore  $j \in C$ , since  $C$  is closed. But  $n \equiv t \pmod{d}$ ; hence  $n + 1 \equiv t + 1 \pmod{d}$ , so that  $j \in C_{t+1}$ .

► **Example 2.** Consider the chain of Figure 7.3.2. Since every state leads to every other state, the entire state space forms a closed equivalence class  $C$  (necessarily recurrent by Theorem 6). We now describe an effective procedure that can be used to find the period of any finite closed equivalence class. Start with any state, say 1, and let  $C_0$  be the subclass containing 1. Then all states reachable in one step from 1 belong to  $C_1$ ; in this case  $3 \in C_1$ . All states reachable in one step from 3 belong to  $C_2$ ; in this case  $5, 6 \in C_2$ . Continue in this fashion to obtain the following table, constructed according to the rule that all states reachable in one step from at least one state in  $C_i$  belong to  $C_{i+1}$ .

$C_0$	$C_1$	$C_2$	$C_3$	$C_4$	$C_5$	$C_6$
1	3	5, 6	2	4	7	1, 2

Stop the construction when a class  $C_k$  is reached that contains a state belonging to some  $C_j$ ,  $j < k$ . Here we have  $2 \in C_3 \cap C_6$ ; hence  $C_3 = C_6$ . Also,

## 228 MARKOV CHAINS

$1 \in C_0 \cap C_6$ ; hence  $C_0 = C_3 = C_6$ . Repeat the process with  $C_0$  = the class containing 1 and 2; all states reachable in one step from either 1 or 2 belong to  $C_1$ . We obtain

$$\begin{array}{cccc} C_0 & C_1 & C_2 & C_3 \\ 1, 2 & 3, 4 & 5, 6, 7 & 1, 2 \end{array}$$

We find that  $C_0 = C_3$  (which we already knew). Since  $C_0 \cup C_1 \cup C_2$  is the entire equivalence class  $C$ , we are finished. We conclude that the period is 3, and that  $C_0 = \{1, 2\}$ ,  $C_1 = \{3, 4\}$ ,  $C_2 = \{5, 6, 7\}$ .

If  $C$  has only a finite number of states, the above process must terminate in a finite number of steps.

We have the following schematic representation of the powers of the transition matrix.

$$\begin{array}{ccc} & C_0 & C_1 & C_2 \\ \Pi = \begin{matrix} C_0 \\ C_1 \\ C_2 \end{matrix} & \begin{bmatrix} 0 & x & 0 \\ 0 & 0 & x \\ x & 0 & 0 \end{bmatrix} & & \begin{matrix} C_0 & C_1 & C_2 \\ \Pi^2 = \begin{matrix} C_0 \\ C_1 \\ C_2 \end{matrix} & \begin{bmatrix} 0 & 0 & x \\ x & 0 & 0 \\ 0 & x & 0 \end{bmatrix} \end{array}$$
  

$$\begin{array}{ccc} & C_0 & C_1 & C_2 \\ \Pi^3 = \begin{matrix} C_0 \\ C_1 \\ C_2 \end{matrix} & \begin{bmatrix} x & 0 & 0 \\ 0 & x & 0 \\ 0 & 0 & x \end{bmatrix} & & (x \text{ stands for positive element}) \end{array}$$

Notice that  $\Pi^4$  has the same form as  $\Pi$  but is not the same numerically; similarly,  $\Pi^5$  has the same form as  $\Pi^2$ ,  $\Pi^6$  has the same form as  $\Pi^3$ , and so on. ◀

► **Example 3.** Consider the simple random walk.

- If there are no barriers, the entire state space forms a closed equivalence class with period 2. We have seen that  $f_{00} = 1 - |p - q|$  [see (6.2.7)]; by symmetry,  $f_{ii} = f_{00}$  for all  $i$ . Thus if  $p = q$  the class is recurrent, and if  $p \neq q$  the class is transient.
- If there is an absorbing barrier at 0, then there are two classes,  $C = \{0\}$  and  $D = \{1, 2, \dots\}$ .  $C$  is clearly recurrent, and since  $D$  is not closed, it is transient by Theorem 8.  $C$  has period 1, and  $D$  has period 2.
- If there are absorbing barriers at 0 and  $b$ , then there are three equivalence classes,  $C = \{0\}$ ,  $D = \{1, 2, \dots, b-1\}$ ,  $E = \{b\}$ .  $C$  and  $E$  have period 1 and are recurrent;  $D$  has period 2 and is transient. ◀

**REMARK.** We have seen that if  $B$  is a finite set of transient states, the probability of remaining forever in  $B$  is 0. This is not true for an

## 7.3 CLASSIFICATION OF STATES 229

infinite set of transient states. For example, in the simple random walk with an absorbing barrier at 0, if the initial state is  $x \geq 1$  and  $p > q$ , there is a probability  $1 - (q/p)^x > 0$  of remaining forever in the transient class  $D = \{1, 2, \dots\}$  [see (6.2.6)].

**TERMINOLOGY.** A state (or class) is said to be *aperiodic* iff its period  $d$  is 1, *periodic* iff  $d > 1$ .

## PROBLEMS

1. (a) Let  $A$  be a (possibly infinite) set of positive integers with greatest common divisor  $d$ . Show that there is a finite subset of  $A$  with greatest common divisor  $d$ .  
 (b) If  $A$  is a nonempty set of positive integers with greatest common divisor  $d$ , and  $A$  is closed under addition, show that all sufficiently large multiples of  $d$  belong to  $A$ .  
 (c) If  $d_i$  is the period of the state  $i$ , show that  $d_i$  is the greatest common divisor of  $\{n \geq 1 : f_{ii}^{(n)} > 0\}$ .
2. A state  $i$  is said to be *essential* iff its equivalence class is closed. Show that  $i$  is essential iff, whenever  $i$  leads to  $j$ , it follows that  $j$  leads to  $i$ .
3. Prove directly (without using Theorem 5) that an equivalence class that is not closed must be transient.
4. Classify the states of the following Markov chains. [In (a) and (b) assume  $0 < p < 1$ .]  
 (a) Simple random walk with reflecting barrier at 0 ( $S = \{1, 2, \dots\}$ ,  $p_{11} = q$ ,  $p_{i,i+1} = p$  for all  $i$ ,  $p_{i,i-1} = q$ ,  $i = 2, 3, \dots$ )  
 (b) Simple random walk with reflecting barriers at 0 and  $l + 1$  ( $S = \{1, 2, \dots, l\}$ ,  $p_{11} = q$ ,  $p_{ll} = q$ ,  $p_{i,i+1} = p$ ,  $i = 1, 2, \dots, l - 1$ ,  $p_{i,i-1} = q$ ,  $i = 2, 3, \dots, l$ )  
 (c)

$$\Pi = \begin{bmatrix} .2 & .8 & 0 & 0 \\ 0 & 0 & .1 & .9 \\ 0 & 0 & .2 & .8 \\ .7 & .3 & 0 & 0 \end{bmatrix}$$

(d)

$$\Pi = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{3} & \frac{2}{3} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

## 230 MARKOV CHAINS

5. Let  $R_0, R_1, R_2, \dots$  be independent random variables, all having the same distribution function, with values in the countable set  $S$ ; assume  $P\{R_i = j\} > 0$  for all  $j \in S$ .
- (a) Show that  $\{R_n\}$  may be regarded as a Markov chain; what is  $\Pi$ ?
- (b) Classify the states of the chain.
6. Let  $i$  be a state of a Markov chain, and let

$$H(z) = \sum_{n=0}^{\infty} f_{ii}^{(n)} z^n, \quad U(z) = \sum_{n=0}^{\infty} p_{ii}^{(n)} z^n, \quad |z| \leq 1$$

[take  $f_{ii}^{(0)} = 0$ ]. Use the first entrance theorem to show that  $U(z) - 1 = H(z)U(z)$ .

## 7.4 LIMITING PROBABILITIES

In this section we investigate the limiting behavior of the  $n$ -step transition probability  $p_{ij}^{(n)}$ . The basic result is the following.

**Theorem 1.** Let  $f_1, f_2, \dots$  be a sequence of nonnegative numbers with  $\sum_{n=1}^{\infty} f_n = 1$ , such that the greatest common divisor of  $\{j: f_j > 0\}$  is 1. Set  $u_0 = 1$ ,  $u_n = \sum_{k=1}^n f_k u_{n-k}$ ,  $n = 1, 2, \dots$ . Define  $\mu = \sum_{n=1}^{\infty} n f_n$ . Then  $u_n \rightarrow 1/\mu$  as  $n \rightarrow \infty$ .

We shall apply the theorem to a Markov chain with  $f_n = f_{ii}^{(n)}$ ,  $i$  a given recurrent state with period 1; then  $u_n = p_{ii}^{(n)}$  by the first entrance theorem. Also,  $\mu = \mu_i = \sum_{n=1}^{\infty} n f_{ii}^{(n)}$ , so that if  $T$  is the time required to return to  $i$  when the initial state is  $i$ , then  $\mu_i = E(T)$ . If  $i$  is an arbitrary recurrent state of a Markov chain,  $\mu_i$  is called the *mean recurrence time* of  $i$ .

Theorem 1 states that  $p_{ii}^{(n)} \rightarrow 1/\mu_i$ ; thus, starting in  $i$ , there is a limiting probability for state  $i$ , namely, the reciprocal of the mean recurrence time. Intuitively, if  $\mu_i =$  (say) 4, then for large  $n$  we should be in state  $i$  roughly one quarter of the time, and it is reasonable to expect that  $p_{ii}^{(n)} \rightarrow 1/4$ .

**PROOF.** We first list three results from real analysis that will be needed. All numbers  $a_{kj}$ ,  $c_j$ ,  $a_j$ ,  $b_j$  are assumed real.

1. *Fatou's Lemma:* If  $|a_{kj}| \leq c_j$ ,  $k, j = 1, 2, \dots$ , and  $\sum_j c_j < \infty$ , then

$$\limsup_{k \rightarrow \infty} \sum_j a_{kj} \leq \sum_j \limsup_{k \rightarrow \infty} a_{kj}$$

and

$$\liminf_{k \rightarrow \infty} \sum_j a_{kj} \geq \sum_j \liminf_{k \rightarrow \infty} a_{kj}$$

The “lim inf” statement holds without the hypothesis that

$$|a_{kj}| \leq c_j, \sum_j c_j < \infty, \quad \text{if all } a_{kj} \geq 0$$

2. *Dominated Convergence Theorem:* If  $|a_{kj}| \leq c_j$ ,  $k, j = 1, 2, \dots$ ,  $\sum_j c_j < \infty$ , and  $\lim_{k \rightarrow \infty} a_{kj} = a_j$ ,  $j = 1, 2, \dots$ , then

$$\lim_{k \rightarrow \infty} \sum_j a_{kj} = \sum_j \lim_{k \rightarrow \infty} a_{kj} \left( = \sum_j a_j \right)$$

(The dominated convergence theorem follows from Fatou’s lemma. Alternatively, a fairly short direct proof may be given.)

$$3. \liminf_{k \rightarrow \infty} (a_k + b_k) \leq \liminf_{k \rightarrow \infty} a_k + \limsup_{k \rightarrow \infty} b_k.$$

(This follows quickly from the definitions of lim inf and lim sup.)

We now prove Theorem 1. First notice that  $0 \leq u_n \leq 1$  for all  $n$  (by induction). Define

$$r_n = \sum_{j=n+1}^{\infty} f_j, \quad n = 0, 1, \dots$$

Then

$$u_n = f_1 u_{n-1} + \dots + f_n u_0 = (r_0 - r_1) u_{n-1} + \dots + (r_{n-1} - r_n) u_0, \quad n \geq 1 \quad (7.4.1)$$

Since  $r_0 = \sum_{j=1}^{\infty} f_j = 1$ , we have  $u_n = r_0 u_n$ , and thus we may rearrange terms in (7.4.1) to obtain

$$r_0 u_n + r_1 u_{n-1} + \dots + r_n u_0 = r_0 u_{n-1} + \dots + r_{n-1} u_0, \quad n \geq 1$$

This indicates that  $\sum_{k=0}^n r_k u_{n-k}$  is independent of  $n$ ; hence

$$\sum_{k=0}^n r_k u_{n-k} = r_0 u_0 = 1, \quad n = 0, 1, \dots \quad (7.4.2)$$

[An alternative proof that  $\sum_{k=0}^n r_k u_{n-k} = 1$ : construct a Markov chain with  $f_{ii}^{(n)} = f_n$ ,  $p_{ii}^{(n)} = u_n$  (see Problem 1). Then

$$\begin{aligned} \sum_{k=0}^n r_k u_{n-k} &= \sum_{k=0}^n u_k r_{n-k} \\ &= \sum_{k=0}^n p_{ii}^{(k)} P\{T > n - k \mid R_0 = i\} \end{aligned}$$

(where  $T$  is the time required to return to  $i$  when the initial state is  $i$ )

$$\begin{aligned} &= \sum_{k=0}^n P\{R_k = i, R_{k+1} \neq i, \dots, R_n \neq i \mid R_0 = i\} \\ &= P\{R_k = i \text{ for some } k = 0, 1, \dots, n \mid R_0 = i\} \\ &= 1 \end{aligned}$$

## 232 MARKOV CHAINS

Now let  $b = \limsup_n u_n$ . Pick a subsequence  $\{u_{n_k}\}$  converging to  $b$ . Then

$$\begin{aligned} b &= \lim_k u_{n_k} = \liminf_k u_{n_k} \\ &= \liminf_k \left[ f_i u_{n_k-i} + \sum_{\substack{j=1 \\ j \neq i}}^{n_k} f_j u_{n_k-j} \right] \\ &\leq \liminf_k (f_i u_{n_k-i}) + \left( \sum_{\substack{j=1 \\ j \neq i}}^{\infty} f_j \right) b \end{aligned}$$

[We use here the fact that  $\liminf_k (a_k + b_k) \leq \liminf_k a_k + \limsup_k b_k$ . Furthermore, if we take  $u_n = 0$  for  $n < 0$ , then

$$\limsup_k \left( \sum_{\substack{j=1 \\ j \neq i}}^{\infty} f_j u_{n_k-j} \right) \leq \sum_{\substack{j=1 \\ j \neq i}}^{\infty} f_j \limsup_k (u_{n_k-j}) \leq b \sum_{\substack{j=1 \\ j \neq i}}^{\infty} f_j$$

Notice that since  $|f_j u_{n_k-j}| \leq f_j$  and  $\sum_{j=1}^{\infty} f_j < \infty$ , Fatou's lemma applies.]

Therefore

$$b \leq f_i \liminf_k u_{n_k-i} + (1 - f_i)b$$

or

$$f_i \liminf_k u_{n_k-i} \geq f_i b$$

Thus  $f_i > 0$  implies  $u_{n_k-i} \rightarrow b$  as  $k \rightarrow \infty$ .

It follows that  $u_{n_k-i} \rightarrow b$  for sufficiently large  $i$ . For if  $f_i > 0$ , we apply the above argument to the sequence  $u_{n_k-i}$ ,  $k = 1, 2, \dots$ , to show that  $f_i > 0$  implies  $u_{n_k-i-j} \rightarrow b$ . Thus if  $t = \sum_{r=1}^m a_r i_r$ , where the  $a_r$  are positive integers and  $f_{i_r} > 0$ , then  $u_{n_k-t} \rightarrow b$ . The set  $S$  of all such  $t$ 's is closed under addition and has greatest common divisor 1, since  $S$  is generated by the positive integers  $i$  for which  $f_i > 0$ . Thus (Problem 1b, Section 7.3)  $S$  contains all sufficiently large positive integers. Say  $u_{n_k-i} \rightarrow b$  for  $i \geq I$ . By (7.4.2),

$$\sum_{j=0}^{\infty} r_j u_{n_k-I-j} = 1 \quad (\text{with } u_n = 0 \text{ for } n < 0) \quad (7.4.3)$$

If  $\sum_{j=0}^{\infty} r_j < \infty$ , the dominated convergence theorem shows that we may let  $k \rightarrow \infty$  and take limits term by term in (7.4.3) to obtain  $b \sum_{j=0}^{\infty} r_j = 1$ . If  $\sum_{j=0}^{\infty} r_j = \infty$ , Fatou's lemma gives  $1 \geq b \sum_{j=0}^{\infty} r_j$ ; hence  $b = 0$ . In either case, then,

$$b = \left[ \sum_{j=0}^{\infty} r_j \right]^{-1}$$

But

$$\begin{aligned} r_0 &= f_1 + f_2 + f_3 + \cdots \\ r_1 &= f_2 + f_3 + \cdots \\ r_2 &= f_3 + \cdots \\ &\vdots \\ &\vdots \\ &\vdots \end{aligned}$$

Hence

$$\sum_{n=0}^{\infty} r_n = \sum_{n=1}^{\infty} n f_n = \mu$$

Consequently  $b = \limsup_n u_n = 1/\mu$ . By an entirely symmetric argument,  $\liminf_n u_n = 1/\mu$ , and the result follows.

We now apply Theorem 1 to gain complete information about the limiting behavior of the  $n$ -step transition probability  $p_{ij}^{(n)}$ . A recurrent state  $j$  is said to be *positive* iff its mean recurrence time  $\mu_j$  is  $< \infty$ , *null* iff  $\mu_j = \infty$ .

**Theorem 2.**

(a) *If the state  $j$  is transient then  $\sum_{n=1}^{\infty} p_{ij}^{(n)} < \infty$  for all  $i$ , hence*

$$p_{ij}^{(n)} \rightarrow 0 \quad \text{as } n \rightarrow \infty$$

PROOF. This is Theorem 4 of Section 7.3.

(b) *If  $j$  is recurrent and aperiodic, and  $i$  belongs to the same equivalence class as  $j$ , then  $p_{ij}^{(n)} \rightarrow 1/\mu_j$ . Furthermore,  $\mu_j$  is finite iff  $\mu_i$  is finite. If  $i$  belongs to a different class, then  $p_{ij}^{(n)} \rightarrow f_{ij}/\mu_j$ .*

PROOF.

$$p_{ij}^{(n)} = \sum_{k=1}^{\infty} f_{ij}^{(k)} p_{jj}^{(n-k)}$$

by the first entrance theorem [take  $p_{jj}^{(r)} = 0$ ,  $r < 0$ ]. By the dominated convergence theorem, we may take limits term by term as  $n \rightarrow \infty$ ; since  $p_{jj}^{(n-k)} \rightarrow 1/\mu_j$  by Theorem 1, we have

$$p_{ij}^{(n)} \rightarrow \left( \sum_{k=1}^{\infty} f_{ij}^{(k)} \right) \frac{1}{\mu_j} = \frac{f_{ij}}{\mu_j}$$

If  $i$  and  $j$  belong to the same recurrent class,  $f_{ij} = 1$ .

Now assume that  $\mu_j$  is finite. If  $p_{ij}^{(r)}, p_{ji}^{(s)} > 0$ , then  $p_{ii}^{(n+r+s)} \geq p_{ij}^{(r)} p_{jj}^{(n)} p_{ji}^{(s)}$ ; this is bounded away from 0 for large  $n$ , since  $p_{jj}^{(n)} \rightarrow 1/\mu_j > 0$ . But if  $\mu_i = \infty$ , then  $p_{ii}^{(n)} \rightarrow 0$  as  $n \rightarrow \infty$ , a contradiction. This proves (b).

## 234 MARKOV CHAINS

- (c) Let  $j$  be recurrent with period  $d > 1$ . Let  $i$  be in the same class as  $j$ , with  $i \in$  the cyclically moving subclass  $C_r$ ,  $j \in C_{r+a}$ . Then  $p_{ij}^{(nd+a)} \rightarrow d/\mu_j$ . Also,  $\mu_j$  is finite iff  $\mu_i$  is finite, so that the property of being recurrent positive (or recurrent null) is a class property.

PROOF. First assume  $a = 0$ . Then  $j$  is recurrent and aperiodic relative to the chain with transition matrix  $\Pi^d$ . (If  $A$  has greatest common divisor  $d$ , then the greatest common divisor of  $\{x/d: x \in A\}$  is 1.) By (b),

$$p_{ij}^{(nd)} \rightarrow \frac{1}{\sum_{k=1}^{\infty} k f_{jj}^{(kd)}} = \frac{d}{\sum_{k=1}^{\infty} k d f_{jj}^{(kd)}} = \frac{d}{\mu_j}$$

Now, having established the result for  $a = r$ , assume  $a = r + 1$  and write

$$p_{ij}^{(nd+r+1)} = \sum_k p_{ik} p_{kj}^{(nd+r)} \rightarrow \sum_k p_{ik} \frac{d}{\mu_j} = \frac{d}{\mu_j}$$

as asserted.

The argument that  $\mu_j$  is finite iff  $\mu_i$  is finite is the same as in (b), with  $nd$  replacing  $n$ .

- (d) If  $j$  is recurrent with period  $d > 1$ , and  $i$  is an arbitrary state, then

$$p_{ij}^{(nd+a)} \rightarrow \left[ \sum_{k=0}^{\infty} f_{ij}^{(kd+a)} \right] \frac{d}{\mu_j}, \quad a = 0, 1, \dots, d-1$$

The expression in brackets is the probability of reaching  $j$  from  $i$  in a number of steps that is  $\equiv a \pmod{d}$ . Thus, if  $j$  is recurrent null,  $p_{ij}^{(n)} \rightarrow 0$  as  $n \rightarrow \infty$  for all  $i$ .

PROOF.

$$p_{ij}^{(nd+a)} = \sum_{k=1}^{nd+a} f_{ij}^{(k)} p_{jj}^{(nd+a-k)}, \quad a = 0, 1, \dots, d-1$$

Since  $j$  has period  $d$ ,  $p_{jj}^{(nd+a-k)} = 0$  unless  $k - a$  is of the form  $rd$  (necessarily  $r \leq n$ ); hence

$$p_{ij}^{(nd+a)} = \sum_{r=0}^n f_{ij}^{(rd+a)} p_{jj}^{((n-r)d)}$$

Let  $n \rightarrow \infty$  and use (c) to finish the proof.

- (e) A finite chain has no recurrent null states.



## 7.4 LIMITING PROBABILITIES 235

PROOF. Let  $C$  be a finite recurrent null class, say,  $C = \{1, 2, \dots, r\}$ . Then

$$\sum_{j=1}^r p_{ij}^{(n)} = 1, \quad i \in C$$

Let  $n \rightarrow \infty$ ; by (d) we obtain  $0 = 1$ , a contradiction.

## PROBLEMS

- With  $f_n$  and  $u_n$  as in Theorem 1, show how to construct a Markov chain with a state  $i$  such that  $f_{ii}^{(n)} = f_n$  and  $p_{ii}^{(n)} = u_n$  for all  $n$ .
- (The renewal theorem) Let  $T_1, T_2, \dots$  be independent random variables, all with the same distribution function, taking values on the positive integers. (Think of the  $T_i$  as waiting times for customers to arrive, or as lifetimes of a succession of products such as light bulbs. If  $T_1 + \dots + T_n = x$ , bulb  $n$  has burned out at time  $x$ , and the light must be renewed by placing bulb  $n + 1$  in position.) Assume that the greatest common divisor of  $\{x: P\{T_k = x\} > 0\}$  is  $d$  and let  $G(n) = \sum_{k=1}^{\infty} P\{T_1 + \dots + T_k = n\}$ ,  $n = 1, 2, \dots$ . If  $\mu = E(T_i)$ , show that  $\lim_{n \rightarrow \infty} G(nd) = d/\mu$ ; interpret the result intuitively.
- Show that in any Markov chain,  $(1/n) \sum_{k=1}^n p_{ij}^{(k)}$  approaches a limit as  $n \rightarrow \infty$ , namely,  $f_{ij}/\mu_j$ . (Define  $\mu_j = \infty$  if  $j$  is transient.) HINT:

$$\frac{1}{n} \sum_{k=1}^n p_{ij}^{(k)} = \sum_{r=1}^d \frac{1}{n} \sum_{\substack{k=1 \\ k \equiv r \pmod{d}}}^n p_{ij}^{(k)}, \quad d = \text{period of } j$$

- Let  $V_{ij}$  be the number of visits to the state  $j$ , starting at  $i$ . (If  $i = j$ ,  $t = 0$  counts as a visit.)
  - Show that  $E(V_{ij}) = \sum_{n=0}^{\infty} p_{ij}^{(n)}$ . Thus  $i$  is recurrent iff  $E(V_{ii}) = \infty$ , and if  $j$  is transient,  $E(V_{ij}) < \infty$  for all  $i$ .
  - Let  $C$  be a transient class,  $N_{ij} = E(V_{ij})$ ,  $i, j \in C$ . Show that

$$\begin{aligned} N_{ij} &= \delta_{ij} + \sum_{k \in C} p_{ik} N_{kj} & (\delta_{ij} &= 1, i = j \\ & & = 0, i \neq j) \end{aligned}$$

In matrix form,  $N = I + QN$ ,  $Q = \Pi$  restricted to  $C$ .

- Show that  $(I - Q)N = N(I - Q) = I$  so that  $N = (I - Q)^{-1}$  in the case of a finite chain (the inverse of an infinite matrix need not be unique).

REMARK. (a) implies that in the gambler's ruin problem with finite capital, the average duration of the game is finite. For if the initial capital is  $i$  and  $D$  is the duration of the game, then  $D = \sum_{j=1}^{b-1} V_{ij}$ , so that  $E(D) < \infty$ .

## 7.5 STATIONARY AND STEADY-STATE DISTRIBUTIONS

A *stationary distribution* for a Markov chain with state space  $S$  is a set of numbers  $v_i, i \in S$ , such that  $v_i \geq 0$ ,  $\sum_{i \in S} v_i = 1$ , and

$$\sum_{i \in S} v_i p_{ij} = v_j, \quad j \in S$$

Thus, if  $V = (v_i, i \in S)$ , then  $V\Pi = V$ . By induction,  $V\Pi^n = V\Pi(\Pi^{n-1}) = V\Pi^{n-1} = \cdots = V\Pi = V$ , so that  $V\Pi^n = V$  for all  $n = 0, 1, \dots$ . Therefore, if the initial state distribution is  $V$ , the state distribution at all future times is still  $V$ . Furthermore, since

$$\begin{aligned} P\{R_n = i, R_{n+1} = i_1, \dots, R_{n+k} = i_k\} \\ &= P\{R_n = i\} p_{ii_1} p_{i_1 i_2} \cdots p_{i_{k-1} i_k} \quad (\text{Problem 4, Section 7.1}) \\ &= v_i p_{ii_1} p_{i_1 i_2} \cdots p_{i_{k-1} i_k} \end{aligned}$$

the sequence  $\{R_n\}$  is stationary; that is, the joint probability function of  $R_n, R_{n+1}, \dots, R_{n+k}$  does not depend on  $n$ .

Stationary distributions are closely related to limiting probabilities. The main result is the following.

**Theorem 1.** Consider a Markov chain with transition matrix  $[p_{ij}]$ . Assume

$$\lim_{n \rightarrow \infty} p_{ij}^{(n)} = q_j$$

for all states  $i, j$  (where  $q_j$  does not depend on  $i$ ). Then

- (a)  $\sum_{j \in S} q_j \leq 1$  and  $\sum_{i \in S} q_i p_{ij} = q_j, j \in S$ .
- (b) Either all  $q_j = 0$ , or else  $\sum_{j \in S} q_j = 1$ .
- (c) If all  $q_j = 0$ , there is no stationary distribution. If  $\sum_{j \in S} q_j = 1$ , then  $\{q_j\}$  is the unique stationary distribution.

PROOF.

$$\sum_j q_j = \sum_j \lim_n p_{ij}^{(n)} \leq \liminf_n \sum_j p_{ij}^{(n)}$$

by Fatou's lemma; hence

$$\sum_j q_j \leq 1$$

Now

$$\sum_i q_i p_{ij} = \sum_i (\lim_n p_{ki}^{(n)}) p_{ij} \leq \liminf_n \sum_i p_{ki}^{(n)} p_{ij} = \liminf_n p_{kj}^{(n+1)} = q_j$$

## 7.5 STATIONARY AND STEADY-STATE DISTRIBUTIONS 237

But if  $\sum_i q_i p_{ij_0} < q_{j_0}$  for some  $j_0$ , then

$$\sum_j q_j > \sum_j \sum_i q_i p_{ij} = \sum_i q_i \sum_j p_{ij} = \sum_i q_i$$

which is a contradiction. This proves (a).

Now if  $Q = (q_i, i \in S)$ , then by (a),  $Q\Pi = Q$ ; hence, by induction,  $Q\Pi^n = Q$ , that is,  $\sum_i q_i p_{ij}^{(n)} = q_j$ . Thus

$$q_j = \lim_n \sum_i q_i p_{ij}^{(n)} = \sum_i q_i \lim_n p_{ij}^{(n)}$$

by the dominated convergence theorem. Hence  $q_j = (\sum_i q_i)q_j$ , proving (b).

Finally, if  $\{v_i\}$  is a stationary distribution, then  $\sum_i v_i p_{ij}^{(n)} = v_j$ . Let  $n \rightarrow \infty$  to obtain  $\sum_i v_i q_j = v_j$ , so that  $q_j = v_j$ . Consequently, if a stationary distribution exists, it is unique and coincides with  $\{q_j\}$ . Therefore no stationary distribution can exist if all  $q_j = 0$ ; if  $\sum_j q_j = 1$ , then, by (a),  $\{q_j\}$  is stationary and the result is established.

The numbers  $v_i, i \in S$ , are said to form a *steady-state distribution* iff  $\lim_{n \rightarrow \infty} p_{ij}^{(n)} = v_j$  for all  $i, j \in S$ , and  $\sum_{j \in S} v_j = 1$ . Thus we require that limiting probabilities exist (independent of the initial state) and form a probability distribution.

In the case of a finite chain, a set of limiting probabilities that are independent of the initial state must form a steady-state distribution, that is, the case in which all  $q_j = 0$  cannot occur in Theorem 1. For  $\sum_{j \in S} p_{ij}^{(n)} = 1$  for all  $i \in S$ ; let  $n \rightarrow \infty$  to obtain, since  $S$  is finite,  $\sum_{j \in S} q_j = 1$ . If the chain is infinite, this result is no longer valid. For example, if all states are transient, then  $p_{ij}^{(n)} \rightarrow 0$  for all  $i, j$ .

If  $\{q_j\}$  is a steady-state distribution,  $\{q_j\}$  is the unique stationary distribution, by Theorem 1. However, a chain can have a unique stationary distribution without having a steady-state distribution, in fact without having limiting probabilities. We give examples later in the section.

We shall establish conditions under which a steady-state distribution exists after we discuss the existence and uniqueness of stationary distributions.

Let  $N$  be the number of positive recurrent classes.

**CASE 1.**  $N = 0$ . Then all states are transient or recurrent null. Hence  $p_{ij}^{(n)} \rightarrow 0$  for all  $i, j$  by Theorem 2 of Section 7.4, so that, by Theorem 1 of this section, there is no stationary distribution.

**CASE 2.**  $N = 1$ . Let  $C$  be the unique positive recurrent class. If  $C$  is aperiodic, then, by Theorem 2 of Section 7.4,  $p_{ij}^{(n)} \rightarrow 1/\mu_j, i, j \in C$ . If  $j \notin C$ , then  $j$  is transient or recurrent null, so that  $p_{ij}^{(n)} \rightarrow 0$  for all  $i$ . By Theorem 1,

## 238 MARKOV CHAINS

if we assign  $v_j = 1/\mu_j$ ,  $j \in C$ ,  $v_j = 0$ ,  $j \notin C$ , then  $\{v_j\}$  is the unique stationary distribution, and  $p_{ij}^{(n)} \rightarrow v_j$  for all  $i, j$ .

Now assume  $C$  periodic, with period  $d > 1$ . Let  $D$  be a cyclically moving subclass of  $C$ . The states of  $D$  are recurrent and aperiodic relative to the transition matrix  $\Pi^d$ . By Theorem 2 of Section 7.4,  $p_{ij}^{(nd)} \rightarrow d/\mu_j$ ,  $i, j \in D$ ; hence  $\{d/\mu_j, j \in D\}$  is the unique stationary distribution for  $D$  relative to  $\Pi^d$  (in particular,  $\sum_{j \in D} 1/\mu_j = 1/d$ ). It follows that  $v_j = 1/\mu_j$ ,  $j \in C$ ,  $v_j = 0$ ,  $j \notin C$ , gives the unique stationary distribution for the original chain (see Problem 1).

CASE 3.  $N \geq 2$ . There is a unique stationary distribution for each positive recurrent class, hence uncountably many stationary distributions for the original chain. For if  $V_1\Pi = V_1$ ,  $V_2\Pi = V_2$ , then, if  $a_1, a_2 \geq 0$ ,  $a_1 + a_2 = 1$ , we have

$$(a_1V_1 + a_2V_2)\Pi = a_1V_1 + a_2V_2$$

In summary, *there is a unique stationary distribution if and only if there is exactly one positive recurrent class.*

Finally, we have the basic theorem concerning steady-state distributions.

**Theorem 2.**

- (a) *If there is a steady-state distribution, there is exactly one positive recurrent class  $C$ , and this class is aperiodic; also,  $f_{ij} = 1$  for all  $j \in C$  and all  $i \in S$ .*
- (b) *Conversely, if there is exactly one positive recurrent class  $C$ , which is aperiodic, and, in addition,  $f_{ij} = 1$  for all  $j \in C$  and all  $i \in S$ , then a steady-state distribution exists.*

**PROOF.**

- (a) Let  $\{v_j\}$  be a steady-state distribution. By Theorem 1,  $\{v_j\}$  is the unique stationary distribution; hence there must be exactly one positive recurrent class  $C$ . Suppose that  $C$  has period  $d > 1$ , and let  $i \in$  a cyclically moving subclass  $C_0, j \in C_1$ . Then  $p_{ij}^{(nd+1)} \rightarrow d/\mu_j$  by Theorem 2 of Section 7.4, and  $p_{ij}^{(nd)} = 0$  for all  $n$ . Since  $d/\mu_j > 0$ ,  $p_{ij}^{(n)}$  has no limit as  $n \rightarrow \infty$ , contradicting the hypothesis. If  $j \in C$  and  $i \in S$ , then by Theorem 2(b) of Section 7.4,  $p_{ij}^{(n)} \rightarrow f_{ij}/\mu_j$ , hence  $v_j = f_{ij}/\mu_j$ . Since  $v_j$  does not depend on  $i$ , we have  $f_{ij} = f_{jj} = 1$ .
- (b) By Theorem 2(b) of Section 7.4,

$$\begin{aligned} p_{ij}^{(n)} &\rightarrow \frac{f_{ij}}{\mu_j} && \text{for all } i, \text{ if } j \in C \\ &\rightarrow 0 && \text{for all } i \text{ if } j \notin C \end{aligned}$$

## 7.5 STATIONARY AND STEADY-STATE DISTRIBUTIONS 239

since in this case  $j$  is transient or recurrent null. Therefore, if  $f_{ij} = 1$  for all  $i \in S$  and  $j \in C$ , the limit  $v_j$  is independent of  $i$ . Since  $C$  is positive,  $v_j > 0$  for  $j \in C$ ; hence, by Theorem 1,  $\sum_j v_j = 1$  and the result follows.

[Note that if a steady state distribution exists, there are no recurrent null classes (or closed transient classes). For if  $D$  is such a class and  $i \in D$ , then since  $D$  is closed,  $f_{ij} = 0$  for all  $j \in C$ , a contradiction. Thus in Theorem 2, the statement “there is exactly one positive recurrent class, which is aperiodic” may be replaced by “there is exactly one recurrent class, which is positive and aperiodic”.]

**COROLLARY.** Consider a finite chain.

- (a) A steady-state distribution exists iff there is exactly one closed equivalence class  $C$ , and  $C$  is aperiodic.
- (b) There is a unique stationary distribution iff there is exactly one closed equivalence class.

**PROOF.** The result follows from Theorem 2, with the aid of Theorem 8c of Section 7.3, Theorem 2e of Section 7.4, and the fact that if  $B$  is a finite set of transient states, the probability of remaining forever in  $B$  is 0 (see the remark after Theorem 4 of Section 7.3).

[It is not difficult to verify that a finite chain has at least one closed equivalence class. Thus a finite chain always has at least one stationary distribution.]

**REMARK.** Consider a finite chain with exactly one closed equivalence class, which is periodic. Then, by the above corollary, there is a unique stationary distribution but no steady-state distribution, in fact no limiting probabilities (see the argument of Theorem 2a). For example, consider the chain with transition matrix

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

The unique stationary distribution is  $(1/2, 1/2)$ , but

$$\Pi^n = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad n \text{ even}$$

$$\Pi^n = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad n \text{ odd}$$

and therefore  $\Pi^n$  does not approach a limit.

Usually the easiest way to find a steady-state distribution  $\{v_j\}$ , if it exists,

## 240 MARKOV CHAINS

is to use the fact that a steady-state distribution must be the unique stationary distribution. Thus we solve the equations

$$\sum_{i \in S} v_i p_{ij} = v_j, \quad j \in S$$

under the conditions that all  $v_i \geq 0$  and  $\sum_{i \in S} v_i = 1$ .

## PROBLEMS

1. Show that if there is a single positive recurrent class  $C$ , then  $\{1/\mu_j, j \in C\}$ , with probability 0 assigned to states outside  $C$ , gives the unique stationary distribution for the chain. HINT:  $p_{ii}^{(nd)} = \sum_{k \in C} p_{ik}^{(nd-1)} p_{ki}$ ,  $i \in C$ . Use Fatou's lemma to show that  $1/\mu_i \geq \sum_{k \in C} (1/\mu_k) p_{ki}$ . Then use the fact that  $\sum_{j \in C} 1/\mu_j = 1$ .
2. (a) If, for some  $N$ ,  $\Pi^N$  has a column bounded away from 0, that is, if for some  $j_0$  and some  $\delta > 0$  we have  $p_{ij_0}^{(N)} \geq \delta > 0$  for all  $i$ , show that there is exactly one recurrent class (namely, the class of  $j_0$ ); this class is positive and aperiodic.  
(b) In the case of a finite chain, show that a steady-state distribution exists iff  $\Pi^N$  has a positive column for some  $N$ .
3. Classify the states of the following Markov chains. Discuss the limiting behavior of the transition probabilities and the existence of steady-state and stationary distributions.
  1. Simple random walk with no barriers.
  2. Simple random walk with absorbing barrier at 0.
  3. Simple random walk with absorbing barriers at 0 and  $b$ .
  4. Simple random walk with reflecting barrier at 0.
  5. Simple random walk with reflecting barriers at 0 and  $l + 1$ .
  6. The chain of Example 2, Section 7.1.
  7. The chain of Problem 4c, Section 7.3.
  8. The chain of Problem 4d, Section 7.3.
  9. A sequence of independent random variables (Problem 5, Section 7.3).
  10. The chain with transition matrix

$$\Pi = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{matrix} & \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{3}{4} & \frac{1}{4} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \end{matrix}$$