1 Markov Chains

Consider a system that can be in any one of a finite or countably infinite number of states. Let $\mathscr{S}$ denote this set of states. We can assume that $\mathscr{S}$ is a subset of the integers. The set $\mathscr{S}$ is called the *state space* of the system. Let the system be observed at the discrete moments of time $n = 0, 1, 2, \ldots$, and let X_n denote the state of the system at time n.

Since we are interested in non-deterministic systems, we think of X_n, $n \geq 0$, as random variables defined on a common probability space. Little can be said about such random variables unless some additional structure is imposed upon them.

The simplest possible structure is that of independent random variables. This would be a good model for such systems as repeated experiments in which future states of the system are independent of past and present states. In most systems that arise in practice, however, past and present states of the system influence the future states even if they do not uniquely determine them.

Many systems have the property that given the present state, the past states have no influence on the future. This property is called the *Markov property*, and systems having this property are called *Markov chains*. The Markov property is defined precisely by the requirement that

$$(1) \quad P(X_{n+1} = x_{n+1} \mid X_0 = x_0, \ldots, X_n = x_n) = P(X_{n+1} = x_{n+1} \mid X_n = x_n)$$

for every choice of the nonnegative integer n and the numbers $x_0, \ldots, x_{n+1}$, each in $\mathscr{S}$. The conditional probabilities $P(X_{n+1} = y \mid X_n = x)$ are called the *transition probabilities* of the chain. In this book we will study Markov chains having *stationary* transition probabilities, i.e., those such that $P(X_{n+1} = y \mid X_n = x)$ is independent of n. From now on, when we say that X_n, $n \geq 0$, forms a Markov chain, we mean that these random variables satisfy the Markov property and have stationary transition probabilities.

The study of such Markov chains is worthwhile from two viewpoints. First, they have a rich theory, much of which can be presented at an elementary level. Secondly, there are a large number of systems arising in practice that can be modeled by Markov chains, so the subject has many useful applications.

In order to help motivate the general results that will be discussed later, we begin by considering Markov chains having only two states.

1.1. Markov chains having two states

For an example of a Markov chain having two states, consider a machine that at the start of any particular day is either broken down or in operating condition. Assume that if the machine is broken down at the start of the nth day, the probability is p that it will be successfully repaired and in operating condition at the start of the $(n + 1)$th day. Assume also that if the machine is in operating condition at the start of the nth day, the probability is q that it will have a failure causing it to be broken down at the start of the $(n + 1)$th day. Finally, let $\pi_0(0)$ denote the probability that the machine is broken down initially, i.e., at the start of the 0th day.

Let the state 0 correspond to the machine being broken down and let the state 1 correspond to the machine being in operating condition. Let X_n be the random variable denoting the state of the machine at time n. According to the above description

$$P(X_{n+1} = 1 \mid X_n = 0) = p,$$

$$P(X_{n+1} = 0 \mid X_n = 1) = q,$$

and

$$P(X_0 = 0) = \pi_0(0).$$

Since there are only two states, 0 and 1, it follows immediately that

$$P(X_{n+1} = 0 \mid X_n = 0) = 1 - p,$$

$$P(X_{n+1} = 1 \mid X_n = 1) = 1 - q,$$

and that the probability $\pi_0(1)$ of being initially in state 1 is given by

$$\pi_0(1) = P(X_0 = 1) = 1 - \pi_0(0).$$

From this information, we can easily compute $P(X_n = 0)$ and $P(X_n = 1)$. We observe that

$$\begin{aligned}
P(X_{n+1} = 0) &= P(X_n = 0 \text{ and } X_{n+1} = 0) + P(X_n = 1 \text{ and } X_{n+1} = 0) \\
&= P(X_n = 0)P(X_{n+1} = 0 \mid X_n = 0) \\
&\quad + P(X_n = 1)P(X_{n+1} = 0 \mid X_n = 1) \\
&= (1 - p)P(X_n = 0) + qP(X_n = 1) \\
&= (1 - p)P(X_n = 0) + q(1 - P(X_n = 0)) \\
&= (1 - p - q)P(X_n = 0) + q.
\end{aligned}$$

Now $P(X_0 = 0) = \pi_0(0)$, so

$$P(X_1 = 0) = (1 - p - q)\pi_0(0) + q$$

and

$$\begin{aligned} P(X_2 = 0) &= (1 - p - q)P(X_1 = 0) + q \\ &= (1 - p - q)^2\pi_0(0) + q[1 + (1 - p - q)]. \end{aligned}$$

It is easily seen by repeating this procedure n times that

$$(2) \qquad P(X_n = 0) = (1 - p - q)^n\pi_0(0) + q \sum_{j=0}^{n-1} (1 - p - q)^j.$$

In the trivial case $p = q = 0$, it is clear that for all n

$$P(X_n = 0) = \pi_0(0) \qquad \text{and} \qquad P(X_n = 1) = \pi_0(1).$$

Suppose now that $p + q > 0$. Then by the formula for the sum of a finite geometric progression,

$$\sum_{j=0}^{n-1} (1 - p - q)^j = \frac{1 - (1 - p - q)^n}{p + q}.$$

We conclude from (2) that

$$(3) \qquad P(X_n = 0) = \frac{q}{p + q} + (1 - p - q)^n \left(\pi_0(0) - \frac{q}{p + q}\right),$$

and consequently that

$$(4) \qquad P(X_n = 1) = \frac{p}{p + q} + (1 - p - q)^n \left(\pi_0(1) - \frac{p}{p + q}\right).$$

Suppose that p and q are neither both equal to zero nor both equal to 1. Then $0 < p + q < 2$, which implies that $|1 - p - q| < 1$. In this case we can let $n \to \infty$ in (3) and (4) and conclude that

$$(5) \qquad \lim_{n\to\infty} P(X_n = 0) = \frac{q}{p + q} \qquad \text{and} \qquad \lim_{n\to\infty} P(X_n = 1) = \frac{p}{p + q}.$$

We can also obtain the probabilities $q/(p + q)$ and $p/(p + q)$ by a different approach. Suppose we want to choose $\pi_0(0)$ and $\pi_0(1)$ so that $P(X_n = 0)$ and $P(X_n = 1)$ are independent of n. It is clear from (3) and (4) that to do this we should choose

$$\pi_0(0) = \frac{q}{p + q} \qquad \text{and} \qquad \pi_0(1) = \frac{p}{p + q}.$$

Thus we see that if X_n, $n \geq 0$, starts out with the initial distribution

$$P(X_0 = 0) = \frac{q}{p + q} \qquad \text{and} \qquad P(X_0 = 1) = \frac{p}{p + q},$$

then for all n

$$P(X_n = 0) = \frac{q}{p+q} \qquad \text{and} \qquad P(X_n = 1) = \frac{p}{p+q}.$$

The description of the machine is vague because it does not really say whether X_n, $n \geq 0$, can be assumed to satisfy the Markov property. Let us suppose, however, that the Markov property does hold. We can use this added information to compute the joint distribution of $X_0, X_1, \ldots, X_n$.

For example, let $n = 2$ and let x_0, x_1, and x_2 each equal 0 or 1. Then

$$\begin{aligned} &P(X_0 = x_0, X_1 = x_1, \text{ and } X_2 = x_2) \\ &\quad = P(X_0 = x_0 \text{ and } X_1 = x_1)P(X_2 = x_2 \mid X_0 = x_0 \text{ and } X_1 = x_1) \\ &\quad = P(X_0 = x_0)P(X_1 = x_1 \mid X_0 = x_0)P(X_2 = x_2 \mid X_0 = x_0 \text{ and } X_1 = x_1). \end{aligned}$$

Now $P(X_0 = x_0)$ and $P(X_1 = x_1 \mid X_0 = x_0)$ are determined by p, q, and $\pi_0(0)$; but without the Markov property, we cannot evaluate $P(X_2 = x_2 \mid X_0 = x_0 \text{ and } X_1 = x_1)$ in terms of p, q, and $\pi_0(0)$. If the Markov property is satisfied, however, then

$$P(X_2 = x_2 \mid X_0 = x_0 \text{ and } X_1 = x_1) = P(X_2 = x_2 \mid X_1 = x_1),$$

which is determined by p and q. In this case

$$\begin{aligned} &P(X_0 = x_0, X_1 = x_1, \text{ and } X_2 = x_2) \\ &\qquad = P(X_0 = x_0)P(X_1 = x_1 \mid X_0 = x_0)P(X_2 = x_2 \mid X_1 = x_1). \end{aligned}$$

For example,

$$\begin{aligned} &P(X_0 = 0, X_1 = 1, \text{ and } X_2 = 0) \\ &\qquad = P(X_0 = 0)P(X_1 = 1 \mid X_0 = 0)P(X_2 = 0 \mid X_1 = 1) \\ &\qquad = \pi_0(0)pq. \end{aligned}$$

The reader should check the remaining entries in the following table, which gives the joint distribution of X_0, X_1, and X_2.

x_0	x_1	x_2	$P(X_0 = x_0, X_1 = x_1, \text{ and } X_2 = x_2)$
0	0	0	$\pi_0(0)(1 - p)^2$
0	0	1	$\pi_0(0)(1 - p)p$
0	1	0	$\pi_0(0)pq$
0	1	1	$\pi_0(0)p(1 - q)$
1	0	0	$(1 - \pi_0(0))q(1 - p)$
1	0	1	$(1 - \pi_0(0))qp$
1	1	0	$(1 - \pi_0(0))(1 - q)q$
1	1	1	$(1 - \pi_0(0))(1 - q)^2$

1.2. Transition function and initial distribution

Let X_n, $n \geq 0$, be a Markov chain having state space $\mathscr{S}$. (The restriction to two states is now dropped.) The function $P(x, y)$, $x \in \mathscr{S}$ and $y \in \mathscr{S}$, defined by

$$(6) \qquad P(x, y) = P(X_1 = y \mid X_0 = x), \qquad x, y \in \mathscr{S},$$

is called the *transition function* of the chain. It is such that

$$(7) \qquad P(x, y) \geq 0, \qquad x, y \in \mathscr{S},$$

and

$$(8) \qquad \sum_y P(x, y) = 1, \qquad x \in \mathscr{S}.$$

Since the Markov chain has stationary probabilities, we see that

$$(9) \qquad P(X_{n+1} = y \mid X_n = x) = P(x, y), \qquad n \geq 1.$$

It now follows from the Markov property that

$$(10) \quad P(X_{n+1} = y \mid X_0 = x_0, \ldots, X_{n-1} = x_{n-1}, X_n = x) = P(x, y).$$

In other words, if the Markov chain is in state x at time n, then no matter how it got to x, it has probability $P(x, y)$ of being in state y at the next step. For this reason the numbers $P(x, y)$ are called the *one-step transition probabilities* of the Markov chain.

The function $\pi_0(x)$, $x \in \mathscr{S}$, defined by

$$(11) \qquad \pi_0(x) = P(X_0 = x), \qquad x \in \mathscr{S},$$

is called the *initial distribution* of the chain. It is such that

$$(12) \qquad \pi_0(x) \geq 0, \qquad x \in \mathscr{S},$$

and

$$(13) \qquad \sum_x \pi_0(x) = 1.$$

The joint distribution of $X_0, \ldots, X_n$ can easily be expressed in terms of the transition function and the initial distribution. For example,

$$\begin{aligned} P(X_0 = x_0, X_1 = x_1) &= P(X_0 = x_0)P(X_1 = x_1 \mid X_0 = x_0) \\ &= \pi_0(x_0)P(x_0, x_1). \end{aligned}$$

Also,

$$\begin{aligned} &P(X_0 = x_0, X_1 = x_1, X_2 = x_2) \\ &\qquad = P(X_0 = x_0, X_1 = x_1)P(X_2 = x_2 \mid X_0 = x_0, X_1 = x_1) \\ &\qquad = \pi_0(x_0)P(x_0, x_1)P(X_2 = x_2 \mid X_0 = x_0, X_1 = x_1). \end{aligned}$$

Since X_n, $n \geq 0$, satisfies the Markov property and has stationary transition probabilities, we see that

$$\begin{aligned} P(X_2 = x_2 \mid X_0 = x_0, X_1 = x_1) &= P(X_2 = x_2 \mid X_1 = x_1) \\ &= P(X_1 = x_2 \mid X_0 = x_1) \\ &= P(x_1, x_2). \end{aligned}$$

Thus

$$P(X_0 = x_0, X_1 = x_1, X_2 = x_2) = \pi_0(x_0)P(x_0, x_1)P(x_1, x_2).$$

By induction it is easily seen that

$$(14) \qquad P(X_0 = x_0, \ldots, X_n = x_n) = \pi_0(x_0)P(x_0, x_1) \cdots P(x_{n-1}, x_n).$$

It is usually more convenient, however, to reverse the order of our definitions. We say that $P(x, y)$, $x \in \mathscr{S}$ and $y \in \mathscr{S}$, is a *transition function* if it satisfies (7) and (8), and we say that $\pi_0(x)$, $x \in \mathscr{S}$, is an *initial distribution* if it satisfies (12) and (13). It can be shown that given any transition function P and any initial distribution π_0, there is a probability space and random variables X_n, $n \geq 0$, defined on that space satisfying (14). It is not difficult to show that these random variables form a Markov chain having transition function P and initial distribution π_0.

The reader may be bothered by the possibility that some of the conditional probabilities we have discussed may not be well defined. For example, the left side of (1) is not well defined if

$$P(X_0 = x_0, \ldots, X_n = x_n) = 0.$$

This difficulty is easily resolved. Equations (7), (8), (12), and (13) defining the transition functions and the initial distributions are well defined, and Equation (14) describing the joint distribution of $X_0, \ldots, X_n$ is well defined. It is not hard to show that if (14) holds, then (1), (6), (9), and (10) hold whenever the conditional probabilities in the respective equations are well defined. The same qualification holds for other equations involving conditional probabilities that will be obtained later.

It will soon be apparent that the transition function of a Markov chain plays a much greater role in describing its properties than does the initial distribution. For this reason it is customary to study simultaneously all Markov chains having a given transition function. In fact we adhere to the usual convention that by "a Markov chain having transition function P," we really mean the family of all Markov chains having that transition function.

1.3. Examples

In this section we will briefly describe several interesting examples of Markov chains. These examples will be further developed in the sequel.

Example 1. Random walk. Let $\xi_1, \xi_2, \ldots$ be independent integer-valued random variables having common density f. Let X_0 be an integer-valued random variable that is independent of the ξ_i's and set $X_n = X_0 + \xi_1 + \cdots + \xi_n$. The sequence X_n, $n \geq 0$, is called a *random walk*. It is a Markov chain whose state space is the integers and whose transition function is given by

$$P(x, y) = f(y - x).$$

To verify this, let π_0 denote the distribution of X_0. Then

$$\begin{aligned} P(X_0 = x_0, \ldots, X_n = x_n) & \\ &= P(X_0 = x_0, \xi_1 = x_1 - x_0, \ldots, \xi_n = x_n - x_{n-1}) \\ &= P(X_0 = x_0)P(\xi_1 = x_1 - x_0) \cdots P(\xi_n = x_n - x_{n-1}) \\ &= \pi_0(x_0) f(x_1 - x_0) \cdots f(x_n - x_{n-1}) \\ &= \pi_0(x_0) P(x_0, x_1) \cdots P(x_{n-1}, x_n), \end{aligned}$$

and thus (14) holds.

Suppose a "particle" moves along the integers according to this Markov chain. Whenever the particle is in x, regardless of how it got there, it jumps to state y with probability $f(y - x)$.

As a special case, consider a *simple random walk* in which $f(1) = p$, $f(-1) = q$, and $f(0) = r$, where p, q, and r are nonnegative and sum to one. The transition function is given by

$$P(x, y) = \begin{cases} p, & y = x + 1, \\ q, & y = x - 1, \\ r, & y = x, \\ 0, & \text{elsewhere.} \end{cases}$$

Let a particle undergo such a random walk. If the particle is in state x at a given observation, then by the next observation it will have jumped to state $x + 1$ with probability p and to state $x - 1$ with probability q; with probability r it will still be in state x.

Example 2. Ehrenfest chain. The following is a simple model of the exchange of heat or of gas molecules between two isolated bodies. Suppose we have two boxes, labeled 1 and 2, and d balls labeled $1, 2, \ldots, d$. Initially some of these balls are in box 1 and the remainder are in box 2. An integer is selected at random from $1, 2, \ldots, d$, and the ball labeled by that integer is removed from its box and placed in the opposite box. This procedure is repeated indefinitely with the selections being independent from trial to trial. Let X_n denote the number of balls in box 1 after the nth trial. Then X_n, $n \geq 0$, is a Markov chain on $\mathscr{S} = \{0, 1, 2, \ldots, d\}$.

The transition function of this Markov chain is easily computed. Suppose that there are x balls in box 1 at time n. Then with probability x/d the ball drawn on the $(n + 1)$th trial will be from box 1 and will be transferred to box 2. In this case there will be $x - 1$ balls in box 1 at time $n + 1$. Similarly, with probability $(d - x)/d$ the ball drawn on the $(n + 1)$th trial will be from box 2 and will be transferred to box 1, resulting in $x + 1$ balls in box 1 at time $n + 1$. Thus the transition function of this Markov chain is given by

$$P(x, y) = \begin{cases} \dfrac{x}{d}, & y = x - 1, \\ 1 - \dfrac{x}{d}, & y = x + 1, \\ 0, & \text{elsewhere.} \end{cases}$$

Note that the Ehrenfest chain can in one transition only go from state x to $x - 1$ or $x + 1$ with positive probability.

A state a of a Markov chain is called an *absorbing state* if $P(a, a) = 1$ or, equivalently, if $P(a, y) = 0$ for $y \neq a$. The next example uses this definition.

Example 3. Gambler's ruin chain. Suppose a gambler starts out with a certain initial capital in dollars and makes a series of one dollar bets against the house. Assume that he has respective probabilities p and $q = 1 - p$ of winning and losing each bet, and that if his capital ever reaches zero, he is ruined and his capital remains zero thereafter. Let X_n, $n \geq 0$, denote the gambler's capital at time n. This is a Markov chain in which 0 is an absorbing state, and for $x \geq 1$

$$P(x, y) = \begin{cases} q, & y = x - 1, \\ p, & y = x + 1, \\ 0, & \text{elsewhere.} \end{cases} \tag{15}$$

Such a chain is called a *gambler's ruin chain* on $\mathscr{S} = \{0, 1, 2, \ldots\}$. We can modify this model by supposing that if the capital of the gambler increases to d dollars he quits playing. In this case 0 and d are both absorbing states, and (15) holds for $x = 1, \ldots, d - 1$.

For an alternative interpretation of the latter chain, we can assume that two gamblers are making a series of one dollar bets against each other and that between them they have a total capital of d dollars. Suppose the first gambler has probability p of winning any given bet, and the second gambler has probability $q = 1 - p$ of winning. The two gamblers play until one of them goes broke. Let X_n denote the capital of the first gambler at time n. Then X_n, $n \geq 0$, is a gambler's ruin chain on $\{0, 1, \ldots, d\}$.

Example 4. Birth and death chain. Consider a Markov chain either on $\mathscr{S} = \{0, 1, 2, \ldots\}$ or on $\mathscr{S} = \{0, 1, \ldots, d\}$ such that starting from x the chain will be at $x - 1$, x, or $x + 1$ after one step. The transition function of such a chain is given by

$$P(x, y) = \begin{cases} q_x, & y = x - 1, \\ r_x, & y = x, \\ p_x, & y = x + 1, \\ 0, & \text{elsewhere}, \end{cases}$$

where p_x, q_x, and r_x are nonnegative numbers such that $p_x + q_x + r_x = 1$. The Ehrenfest chain and the two versions of the gambler's ruin chain are examples of *birth and death chains*. The phrase "birth and death" stems from applications in which the state of the chain is the population of some living system. In these applications a transition from state x to state $x + 1$ corresponds to a "birth," while a transition from state x to state $x - 1$ corresponds to a "death."

In Chapter 3 we will study *birth and death processes*. These processes are similar to birth and death chains, except that jumps are allowed to occur at arbitrary times instead of just at integer times. In most applications, the models discussed in Chapter 3 are more realistic than those obtainable by using birth and death chains.

Example 5. Queuing chain. Consider a service facility such as a checkout counter at a supermarket. People arrive at the facility at various times and are eventually served. Those customers that have arrived at the facility but have not yet been served form a waiting line or queue. There are a variety of models to describe such systems. We will consider here only one very simple and somewhat artificial model; others will be discussed in Chapter 3.

Let time be measured in convenient periods, say in minutes. Suppose that if there are any customers waiting for service at the beginning of any given period, exactly one customer will be served during that period, and that if there are no customers waiting for service at the beginning of a period, none will be served during that period. Let ξ_n denote the number of new customers arriving during the nth period. We assume that $\xi_1, \xi_2, \ldots$ are independent nonnegative integer-valued random variables having common density f.

Let X_0 denote the number of customers present initially, and for $n \geq 1$, let X_n denote the number of customers present at the end of the nth period. If $X_n = 0$, then $X_{n+1} = \xi_{n+1}$; and if $X_n \geq 1$, then $X_{n+1} = X_n + \xi_{n+1} - 1$. It follows without difficulty from the assumptions on ξ_n, $n \geq 1$, that X_n, $n \geq 0$, is a Markov chain whose state space is the nonnegative integers and whose transition function P is given by

$$P(0, y) = f(y)$$

and

$$P(x, y) = f(y - x + 1), \qquad x \geq 1.$$

Example 6. Branching chain. Consider particles such as neutrons or bacteria that can generate new particles of the same type. The initial set of objects is referred to as belonging to the 0th generation. Particles generated from the nth generation are said to belong to the $(n + 1)$th generation. Let X_n, $n \geq 0$, denote the number of particles in the nth generation.

Nothing in this description requires that the various particles in a generation give rise to new particles simultaneously. Indeed at a given time, particles from several generations may coexist.

A typical situation is illustrated in Figure 1: one initial particle gives rise to two particles. Thus $X_0 = 1$ and $X_1 = 2$. One of the particles in the first generation gives rise to three particles and the other gives rise to one particle, so that $X_2 = 4$. We see from Figure 1 that $X_3 = 2$. Since neither of the particles in the third generation gives rise to new particles, we conclude that $X_4 = 0$ and consequently that $X_n = 0$ for all $n \geq 4$. In other words, the progeny of the initial particle in the zeroth generation become extinct after three generations.

Figure 1

In order to model this system as a Markov chain, we suppose that each particle gives rise to ξ particles in the next generation, where ξ is a nonnegative integer-valued random variable having density f. We suppose that the number of offspring of the various particles in the various generations are chosen independently according to the density f.

Under these assumptions X_n, $n \geq 0$, forms a Markov chain whose state space is the nonnegative integers. State 0 is an absorbing state. For if there are no particles in a given generation, there will not be any particles in the next generation either. For $x \geq 1$

$$P(x, y) = P(\xi_1 + \cdots + \xi_x = y),$$

where $\xi_1, \ldots, \xi_x$ are independent random variables having common density f. In particular, $P(1, y) = f(y)$, $y \geq 0$.

If a particle gives rise to $\xi = 0$ particles, the interpretation is that the particle dies or disappears. Suppose a particle gives rise to ξ particles, which in turn give rise to other particles; but after some number of generations, all descendants of the initial particle have died or disappeared (see Figure 1). We describe such an event by saying that the descendants of the original particle eventually become *extinct*. An interesting problem involving branching chains is to compute the probability ρ of eventual extinction for a branching chain starting with a single particle or, equivalently, the probability that a branching chain starting at state 1 will eventually be absorbed at state 0. Once we determine ρ, we can easily find the probability that in a branching chain starting with x particles the descendants of each of the original particles eventually become extinct. Indeed, since the particles are assumed to act independently in giving rise to new particles, the desired probability is just ρ^x.

The branching chain was used originally to determine the probability that the male line of a given person would eventually become extinct. For this purpose only male children would be included in the various generations.

Example 7. Consider a gene composed of d subunits, where d is some positive integer and each subunit is either normal or mutant in form. Consider a cell with a gene composed of m mutant subunits and $d - m$ normal subunits. Before the cell divides into two daughter cells, the gene duplicates. The corresponding gene of one of the daughter cells is composed of d units chosen at random from the $2m$ mutant subunits and the $2(d - m)$ normal subunits. Suppose we follow a fixed line of descent from a given gene. Let X_0 be the number of mutant subunits initially present, and let X_n, $n \geq 1$, be the number present in the nth descendant gene. Then X_n, $n \geq 0$, is a Markov chain on $\mathscr{S} = \{0, 1, 2, \ldots, d\}$ and

$$P(x, y) = \frac{\binom{2x}{y}\binom{2d - 2x}{d - y}}{\binom{2d}{d}}.$$

States 0 and d are absorbing states for this chain.

1.4. Computations with transition functions

Let X_n, $n \geq 0$, be a Markov chain on $\mathscr{S}$ having transition function P. In this section we will show how various conditional probabilities can be expressed in terms of P. We will also define the n-step transition function of the Markov chain.

We begin with the formula

$$(16) \qquad P(X_{n+1} = x_{n+1}, \ldots, X_{n+m} = x_{n+m} \mid X_0 = x_0, \ldots, X_n = x_n) \\ = P(x_n, x_{n+1}) \cdots P(x_{n+m-1}, x_{n+m}).$$

To prove (16) we write the left side of this equation as

$$\frac{P(X_0 = x_0, \ldots, X_{n+m} = x_{n+m})}{P(X_0 = x_0, \ldots, X_n = x_n)}.$$

By (14) this ratio equals

$$\frac{\pi_0(x_0)P(x_0, x_1) \cdots P(x_{n+m-1}, x_{n+m})}{\pi_0(x_0)P(x_0, x_1) \cdots P(x_{n-1}, x_n)},$$

which reduces to the right side of (16).

It is convenient to rewrite (16) as

$$(17) \quad P(X_{n+1} = y_1, \ldots, X_{n+m} = y_m \mid X_0 = x_0, \ldots, X_{n-1} = x_{n-1}, X_n = x) \\ = P(x, y_1)P(y_1, y_2) \cdots P(y_{m-1}, y_m).$$

Let $A_0, \ldots, A_{n-1}$ be subsets of $\mathscr{S}$. It follows from (17) and Exercise 4(a) that

$$(18) \quad P(X_{n+1} = y_1, \ldots, X_{n+m} = y_m \mid X_0 \in A_0, \ldots, X_{n-1} \in A_{n-1}, X_n = x) \\ = P(x, y_1)P(y_1, y_2) \cdots P(y_{m-1}, y_m).$$

Let $B_1, \ldots, B_m$ be subsets of $\mathscr{S}$. It follows from (18) and Exercise 4(b) that

$$(19) \quad P(X_{n+1} \in B_1, \ldots, X_{n+m} \in B_m \mid X_0 \in A_0, \ldots, X_{n-1} \in A_{n-1}, X_n = x) \\ = \sum_{y_1 \in B_1} \cdots \sum_{y_m \in B_m} P(x, y_1)P(y_1, y_2) \cdots P(y_{m-1}, y_m).$$

The *m-step transition function* $P^m(x, y)$, which gives the probability of going from x to y in m steps, is defined by

$$(20)\qquad P^m(x, y) = \sum_{y_1} \cdots \sum_{y_{m-1}} P(x, y_1)P(y_1, y_2) \cdots P(y_{m-2}, y_{m-1})P(y_{m-1}, y)$$

for $m \geq 2$, by $P^1(x, y) = P(x, y)$, and by

$$P^0(x, y) = \begin{cases} 1, & x = y, \\ 0, & \text{elsewhere.} \end{cases}$$

We see by setting $B_1 = \cdots = B_{m-1} = \mathscr{S}$ and $B_m = \{y\}$ in (19) that

$$(21)\qquad P(X_{n+m} = y \mid X_0 \in A_0, \ldots, X_{n-1} \in A_{n-1}, X_n = x) = P^m(x, y).$$

In particular, by setting $A_0 = \cdots = A_{n-1} = \mathscr{S}$, we see that

$$(22)\qquad P(X_{n+m} = y \mid X_n = x) = P^m(x, y).$$

It also follows from (21) that

$$(23)\qquad P(X_{n+m} = y \mid X_0 = x, X_n = z) = P^m(z, y).$$

Since (see Exercise 4(c))

$$\begin{aligned} P^{n+m}(x, y) &= P(X_{n+m} = y \mid X_0 = x) \\ &= \sum_z P(X_n = z \mid X_0 = x)P(X_{n+m} = y \mid X_0 = x, X_n = z) \\ &= \sum_z P^n(x, z)P(X_{n+m} = y \mid X_0 = x, X_n = z), \end{aligned}$$

we conclude from (23) that

$$(24)\qquad P^{n+m}(x, y) = \sum_z P^n(x, z)P^m(z, y).$$

For Markov chains having a finite number of states, (24) allows us to think of P^n as the nth power of the matrix P, an idea we will pursue in Section 1.4.2.

Let π_0 be an initial distribution for the Markov chain. Since

$$\begin{aligned} P(X_n = y) &= \sum_x P(X_0 = x, X_n = y) \\ &= \sum_x P(X_0 = x)P(X_n = y \mid X_0 = x), \end{aligned}$$

we see that

$$(25)\qquad P(X_n = y) = \sum_x \pi_0(x)P^n(x, y).$$

This formula allows us to compute the distribution of X_n in terms of the initial distribution π_0 and the n-step transition function P^n.

For an alternative method of computing the distribution of X_n, observe that

$$P(X_{n+1} = y) = \sum_x P(X_n = x, X_{n+1} = y)$$
$$= \sum_x P(X_n = x)P(X_{n+1} = y \mid X_n = x),$$

so that

$$P(X_{n+1} = y) = \sum_x P(X_n = x)P(x, y). \tag{26}$$

If we know the distribution of X_0, we can use (26) to find the distribution of X_1. Then, knowing the distribution of X_1, we can use (26) to find the distribution of X_2. Similarly, we can find the distribution of X_n by applying (26) n times.

We will use the notation $P_x(\ )$ to denote probabilities of various events defined in terms of a Markov chain starting at x. Thus

$$P_x(X_1 \neq a, X_2 \neq a, X_3 = a)$$

denotes the probability that a Markov chain starting at x is in a state a at time 3 but not at time 1 or at time 2. In terms of this notation, (19) can be rewritten as

$$P(X_{n+1} \in B_1, \ldots, X_{n+m} \in B_m \mid X_0 \in A_0, \ldots, X_{n-1} \in A_{n-1}, X_n = x) = P_x(X_1 \in B_1, \ldots, X_m \in B_m). \tag{27}$$

1.4.1. Hitting times. Let A be a subset of $\mathscr{S}$. The *hitting time* T_A of A is defined by

$$T_A = \min(n > 0 : X_n \in A)$$

if $X_n \in A$ for some $n > 0$, and by $T_A = \infty$ if $X_n \notin A$ for all $n > 0$. In other words, T_A is the first positive time the Markov chain is in (*hits*) A. Hitting times play an important role in the theory of Markov chains. In this book we will be interested mainly in hitting times of sets consisting of a single point. We denote the hitting time of a point $a \in \mathscr{S}$ by T_a rather than by the more cumbersome notation $T_{\{a\}}$.

An important equation involving hitting times is given by

$$P^n(x, y) = \sum_{m=1}^{n} P_x(T_y = m)P^{n-m}(y, y), \qquad n \geq 1. \tag{28}$$

In order to verify (28) we note that the events $\{T_y = m, X_n = y\}$, $1 \leq m \leq n$, are disjoint and that

$$\{X_n = y\} = \bigcup_{m=1}^{n} \{T_y = m, X_n = y\}.$$

We have in effect decomposed the event $\{X_n = y\}$ according to the hitting time of y. We see from this decomposition that

$$\begin{aligned} P^n(x, y) &= P_x(X_n = y) \\ &= \sum_{m=1}^{n} P_x(T_y = m, X_n = y) \\ &= \sum_{m=1}^{n} P_x(T_y = m)P(X_n = y \mid X_0 = x, T_y = m) \\ &= \sum_{m=1}^{n} P_x(T_y = m)P(X_n = y \mid X_0 = x, X_1 \neq y, \ldots, \\ &\qquad\qquad\qquad\qquad\qquad X_{m-1} \neq y, X_m = y) \\ &= \sum_{m=1}^{n} P_x(T_y = m)P^{n-m}(y, y), \end{aligned}$$

and hence that (28) holds.

Example 8. Show that if a is an absorbing state, then $P^n(x, a) = P_x(T_a \leq n)$, $n \geq 1$.

If a is an absorbing state, then $P^{n-m}(a, a) = 1$ for $1 \leq m \leq n$, and hence (28) implies that

$$\begin{aligned} P^n(x, a) &= \sum_{m=1}^{n} P_x(T_a = m)P^{n-m}(a, a) \\ &= \sum_{m=1}^{n} P_x(T_a = m) = P_x(T_a \leq n). \end{aligned}$$

Observe that

$$P_x(T_y = 1) = P_x(X_1 = y) = P(x, y)$$

and that

$$P_x(T_y = 2) = \sum_{z \neq y} P_x(X_1 = z, X_2 = y) = \sum_{z \neq y} P(x, z)P(z, y).$$

For higher values of n the probabilities $P_x(T_y = n)$ can be found by using the formula

$$(29) \qquad P_x(T_y = n + 1) = \sum_{z \neq y} P(x, z)P_z(T_y = n), \qquad n \geq 1.$$

This formula is a consequence of (27), but it should also be directly obvious. For in order to go from x to y for the first time at time $n + 1$, it is necessary to go to some state $z \neq y$ at the first step and then go from z to y for the first time at the end of n additional steps.

1.4.2. Transition matrix. Suppose now that the state space $\mathscr{S}$ is finite, say $\mathscr{S} = \{0, 1, \ldots, d\}$. In this case we can think of P as the *transition matrix* having $d + 1$ rows and columns given by

$$\begin{array}{c} \\ 0 \\ \vdots \\ d \end{array} \begin{array}{c} \begin{array}{ccc} 0 & \cdots & d \end{array} \\ \begin{bmatrix} P(0, 0) & \cdots & P(0, d) \\ \vdots & & \vdots \\ P(d, 0) & \cdots & P(d, d) \end{bmatrix} \end{array}.$$

For example, the transition matrix of the gambler's ruin chain on $\{0, 1, 2, 3\}$ is

$$\begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 0 \\ q & 0 & p & 0 \\ 0 & q & 0 & p \\ 0 & 0 & 0 & 1 \end{bmatrix} \end{array}.$$

Similarly, we can regard P^n as an *n-step transition matrix*. Formula (24) with $m = n = 1$ becomes

$$P^2(x, y) = \sum_z P(x, z)P(z, y).$$

Recalling the definition of ordinary matrix multiplication, we observe that the two-step transition matrix P^2 is the product of the matrix P with itself. More generally, by setting $m = 1$ in (24) we see that

$$P^{n+1}(x, y) = \sum_z P^n(x, z)P(z, y). \tag{30}$$

It follows from (30) by induction that the n-step transition matrix P^n is the nth power of P.

An initial distribution π_0 can be thought of as a $(d + 1)$-dimensional row vector

$$\pi_0 = (\pi_0(0), \ldots, \pi_0(d)).$$

If we let π_n denote the $(d + 1)$-dimensional row vector

$$\pi_n = (P(X_n = 0), \ldots, P(X_n = d)),$$

then (25) and (26) can be written respectively as

$$\pi_n = \pi_0 P^n$$

and

$$\pi_{n+1} = \pi_n P.$$

The two-state Markov chain discussed in Section 1.1 is one of the few examples where P^n can be found very easily.

Example 9. Consider the two-state Markov chain having one-step transition matrix

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

where $p + q > 0$. Find P^n.

In order to find $P^n(0, 0) = P_0(X_n = 0)$, we set $\pi_0(0) = 1$ in (3) and obtain

$$P^n(0, 0) = \frac{q}{p + q} + (1 - p - q)^n \frac{p}{p + q}.$$

In order to find $P^n(0, 1) = P_0(X_n = 1)$, we set $\pi_0(1) = 0$ in (4) and obtain

$$P^n(0, 1) = \frac{p}{p + q} - (1 - p - q)^n \frac{p}{p + q}.$$

Similarly, we conclude that

$$P(^n1, 0) = \frac{q}{p + q} - (1 - p - q)^n \frac{q}{p + q}$$

and

$$P^n(1, 1) = \frac{p}{p + q} + (1 - p - q)^n \frac{q}{p + q}.$$

It follows that

$$P^n = \frac{1}{p + q} \begin{bmatrix} q & p \\ q & p \end{bmatrix} + \frac{(1 - p - q)^n}{p + q} \begin{bmatrix} p & -p \\ -q & q \end{bmatrix}.$$

1.5. Transient and recurrent states

Let X_n, $n \geq 0$, be a Markov chain having state space $\mathscr{S}$ and transition function P. Set

$$\rho_{xy} = P_x(T_y < \infty).$$

Then ρ_{xy} denotes the probability that a Markov chain starting at x will be in state y at some positive time. In particular, ρ_{yy} denotes the probability that a Markov chain starting at y will ever return to y. A state y is called *recurrent* if $\rho_{yy} = 1$ and *transient* if $\rho_{yy} < 1$. If y is a recurrent state, a Markov chain starting at y returns to y with probability one. If y is a transient state, a Markov chain starting at y has positive probability $1 - \rho_{yy}$ of never returning to y. If y is an absorbing state, then $P_y(T_y = 1) =$

$P(y, y) = 1$ and hence $\rho_{yy} = 1$; thus an absorbing state is necessarily recurrent.

Let $1_y(z)$, $z \in \mathscr{S}$, denote the indicator function of the set $\{y\}$ defined by

$$1_y(z) = \begin{cases} 1, & z = y, \\ 0, & z \neq y. \end{cases}$$

Let $N(y)$ denote the number of times $n \geq 1$ that the chain is in state y. Since $1_y(X_n) = 1$ if the chain is in state y at time n and $1_y(X_n) = 0$ otherwise, we see that

$$N(y) = \sum_{n=1}^{\infty} 1_y(X_n). \tag{31}$$

The event $\{N(y) \geq 1\}$ is the same as the event $\{T_y < \infty\}$. Thus

$$P_x(N(y) \geq 1) = P_x(T_y < \infty) = \rho_{xy}.$$

Let m and n be positive integers. By (27), the probability that a Markov chain starting at x first visits y at time m and next visits y n units of time later is $P_x(T_y = m)P_y(T_y = n)$. Thus

$$\begin{aligned} P_x(N(y) \geq 2) &= \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} P_x(T_y = m)P_y(T_y = n) \\ &= \left(\sum_{m=1}^{\infty} P_x(T_y = m)\right)\left(\sum_{n=1}^{\infty} P_y(T_y = n)\right) \\ &= \rho_{xy}\rho_{yy}. \end{aligned}$$

Similarly we conclude that

$$P_x(N(y) \geq m) = \rho_{xy}\rho_{yy}^{m-1}, \qquad m \geq 1. \tag{32}$$

Since

$$P_x(N(y) = m) = P_x(N(y) \geq m) - P_x(N(y) \geq m + 1),$$

it follows from (32) that

$$P_x(N(y) = m) = \rho_{xy}\rho_{yy}^{m-1}(1 - \rho_{yy}), \qquad m \geq 1. \tag{33}$$

Also

$$P_x(N(y) = 0) = 1 - P_x(N(y) \geq 1),$$

so that

$$P_x(N(y) = 0) = 1 - \rho_{xy}. \tag{34}$$

These formulas are intuitively obvious. To see why (33) should be true, for example, observe that a chain starting at x visits state y exactly m times if and only if it visits y for a first time, returns to y $m - 1$ additional times, and then never again returns to y.

We use the notation $E_x(\ )$ to denote expectations of random variables defined in terms of a Markov chain starting at x. For example,

$$E_x(1_y(X_n)) = P_x(X_n = y) = P^n(x, y). \tag{35}$$

It follows from (31) and (35) that

$$\begin{aligned} E_x(N(y)) &= E_x\left(\sum_{n=1}^{\infty} 1_y(X_n)\right) \\ &= \sum_{n=1}^{\infty} E_x(1_y(X_n)) \\ &= \sum_{n=1}^{\infty} P^n(x, y). \end{aligned}$$

Set

$$G(x, y) = E_x(N(y)) = \sum_{n=1}^{\infty} P^n(x, y).$$

Then $G(x, y)$ denotes the expected number of visits to y for a Markov chain starting at x.

Theorem 1 (i) *Let y be a transient state. Then*

$$P_x(N(y) < \infty) = 1$$

and

$$G(x, y) = \frac{\rho_{xy}}{1 - \rho_{yy}}, \qquad x \in \mathscr{S}, \tag{36}$$

which is finite for all $x \in \mathscr{S}$.

(ii) *Let y be a recurrent state. Then $P_y(N(y) = \infty) = 1$ and $G(y, y) = \infty$. Also*

$$P_x(N(y) = \infty) = P_x(T_y < \infty) = \rho_{xy}, \qquad x \in \mathscr{S}. \tag{37}$$

If $\rho_{xy} = 0$, then $G(x, y) = 0$, while if $\rho_{xy} > 0$, then $G(x, y) = \infty$.

This theorem describes the fundamental difference between a transient state and a recurrent state. If y is a transient state, then no matter where the Markov chain starts, it makes only a finite number of visits to y and the expected number of visits to y is finite. Suppose instead that y is a recurrent state. Then if the Markov chain starts at y, it returns to y infinitely often. If the chain starts at some other state x, it may be impossible for it to ever hit y. If it is possible, however, and the chain does visit y at least once, then it does so infinitely often.

Proof. Let y be a transient state. Since $0 \le \rho_{yy} < 1$, it follows from (32) that

$$P_x(N(y) = \infty) = \lim_{m\to\infty} P_x(N(y) \ge m) = \lim_{m\to\infty} \rho_{xy}\rho_{yy}^{m-1} = 0.$$

By (33)

$$\begin{aligned} G(x, y) &= E_x(N(y)) \\ &= \sum_{m=1}^{\infty} mP_x(N(y) = m) \\ &= \sum_{m=1}^{\infty} m\rho_{xy}\rho_{yy}^{m-1}(1 - \rho_{yy}). \end{aligned}$$

Substituting $t = \rho_{yy}$ in the power series

$$\sum_{m=1}^{\infty} mt^{m-1} = \frac{1}{(1-t)^2},$$

we conclude that

$$G(x, y) = \frac{\rho_{xy}}{1 - \rho_{yy}} < \infty.$$

This completes the proof of (i).

Now let y be recurrent. Then $\rho_{yy} = 1$ and it follows from (32) that

$$\begin{aligned} P_x(N(y) = \infty) &= \lim_{m\to\infty} P_x(N(y) \ge m) \\ &= \lim_{m\to\infty} \rho_{xy} = \rho_{xy}. \end{aligned}$$

In particular, $P_y(N(y) = \infty) = 1$. If a nonnegative random variable has positive probability of being infinite, its expectation is infinite. Thus

$$G(y, y) = E_y(N(y)) = \infty.$$

If $\rho_{xy} = 0$, then $P_x(T_y = m) = 0$ for all finite positive integers m, so (28) implies that $P^n(x, y) = 0$, $n \ge 1$; thus $G(x, y) = 0$ in this case. If $\rho_{xy} > 0$, then $P_x(N(y) = \infty) = \rho_{xy} > 0$ and hence

$$G(x, y) = E_x(N(y)) = \infty.$$

This completes the proof of Theorem 1. ∎

Let y be a transient state. Since

$$\sum_{n=1}^{\infty} P^n(x, y) = G(x, y) < \infty, \qquad x \in \mathscr{S},$$

we see that

(38) $$\lim_{n\to\infty} P^n(x, y) = 0, \qquad x \in \mathscr{S}.$$

A Markov chain is called a *transient chain* if all of its states are transient and a *recurrent chain* if all of its states are recurrent. It is easy to see that a Markov chain having a finite state space must have at least one recurrent state and hence cannot possibly be a transient chain. For if $\mathscr{S}$ is finite and all states are transient, then by (38)

$$\begin{aligned} 0 &= \sum_{y \in \mathscr{S}} \lim_{n \to \infty} P^n(x, y) \\ &= \lim_{n \to \infty} \sum_{y \in \mathscr{S}} P^n(x, y) \\ &= \lim_{n \to \infty} P_x(X_n \in \mathscr{S}) \\ &= \lim_{n \to \infty} 1 = 1, \end{aligned}$$

which is a contradiction.

1.6. Decomposition of the state space

Let x and y be two not necessarily distinct states. We say that x *leads to* y if $\rho_{xy} > 0$. It is left as an exercise for the reader to show that x leads to y if and only if $P^n(x, y) > 0$ for some positive integer n. It is also left to the reader to show that if x leads to y and y leads to z, then x leads to z.

Theorem 2 *Let x be a recurrent state and suppose that x leads to y. Then y is recurrent and $\rho_{xy} = \rho_{yx} = 1$.*

Proof. We assume that $y \neq x$, for otherwise there is nothing to prove. Since

$$P_x(T_y < \infty) = \rho_{xy} > 0,$$

we see that $P_x(T_y = n) > 0$ for some positive integer n. Let n_0 be the least such positive integer, i.e., set

$$n_0 = \min(n \geq 1 \colon P_x(T_y = n) > 0). \tag{39}$$

It follows easily from (39) and (28) that $P^{n_0}(x, y) > 0$ and

$$P^m(x, y) = 0, \qquad 1 \leq m < n_0. \tag{40}$$

Since $P^{n_0}(x, y) > 0$, we can find states $y_1, \ldots, y_{n_0 - 1}$ such that

$$P_x(X_1 = y_1, \ldots, X_{n_0 - 1} = y_{n_0 - 1}, X_{n_0} = y) = P(x, y_1) \cdots P(y_{n_0 - 1}, y) > 0.$$

None of the states $y_1, \ldots, y_{n_0 - 1}$ equals x or y; for if one of them did equal x or y, it would be possible to go from x to y with positive probability in fewer than n_0 steps, in contradiction to (40).

We will now show that $\rho_{yx} = 1$. Suppose on the contrary that $\rho_{yx} < 1$. Then a Markov chain starting at y has positive probability $1 - \rho_{yx}$ of never hitting x. More to the point, a Markov chain starting at x has the positive probability

$$P(x, y_1) \cdots P(y_{n_0-1}, y)(1 - \rho_{yx})$$

of visiting the states $y_1, \ldots, y_{n_0-1}, y$ successively in the first n_0 times and never returning to x after time n_0. But if this happens, the Markov chain never returns to x at any time $n \geq 1$, so we have contradicted the assumption that x is a recurrent state.

Since $\rho_{yx} = 1$, there is a positive integer n_1 such that $P^{n_1}(y, x) > 0$. Now

$$\begin{aligned} P^{n_1+n+n_0}(y, y) &= P_y(X_{n_1+n+n_0} = y) \\ &\geq P_y(X_{n_1} = x, X_{n_1+n} = x, X_{n_1+n+n_0} = y) \\ &= P^{n_1}(y, x)P^n(x, x)P^{n_0}(x, y). \end{aligned}$$

Hence

$$\begin{aligned} G(y, y) &\geq \sum_{n=n_1+1+n_0}^{\infty} P^n(y, y) \\ &= \sum_{n=1}^{\infty} P^{n_1+n+n_0}(y, y) \\ &\geq P^{n_1}(y, x)P^{n_0}(x, y) \sum_{n=1}^{\infty} P^n(x, x) \\ &= P^{n_1}(y, x)P^{n_0}(x, y)G(x, x) = +\infty, \end{aligned}$$

from which it follows that y is also a recurrent state.

Since y is recurrent and y leads to x, we see from the part of the theorem that has already been verified that $\rho_{xy} = 1$. This completes the proof. ∎

A nonempty set C of states is said to be *closed* if no state inside of C leads to any state outside of C, i.e., if

(41) $$\rho_{xy} = 0, \qquad x \in C \text{ and } y \notin C.$$

Equivalently (see Exercise 16), C is closed if and only if

(42) $$P^n(x, y) = 0, \qquad x \in C,\ y \notin C, \text{ and } n \geq 1.$$

Actually, even from the weaker condition

(43) $$P(x, y) = 0, \qquad x \in C \text{ and } y \notin C,$$

we can prove that C is closed. For if (43) holds, then for $x \in C$ and $y \notin C$

$$\begin{aligned} P^2(x, y) &= \sum_{z \in \mathcal{S}} P(x, z)P(z, y) \\ &= \sum_{z \in C} P(x, z)P(z, y) = 0, \end{aligned}$$

and (42) follows by induction. If C is closed, then a Markov chain starting in C will, with probability one, stay in C for all time. If a is an absorbing state, then $\{a\}$ is closed.

A closed set C is called *irreducible* if x leads to y for all choices of x and y in C. It follows from Theorem 2 that if C is an irreducible closed set, then either every state in C is recurrent or every state in C is transient. The next result is an immediate consequence of Theorems 1 and 2.

Corollary 1 *Let C be an irreducible closed set of recurrent states. Then $\rho_{xy} = 1$, $P_x(N(y) = \infty) = 1$, and $G(x, y) = \infty$ for all choices of x and y in C.*

An *irreducible Markov chain* is a chain whose state space is irreducible, that is, a chain in which every state leads back to itself and also to every other state. Such a Markov chain is necessarily either a transient chain or a recurrent chain. Corollary 1 implies, in particular, that an irreducible recurrent Markov chain visits every state infinitely often with probability one.

We saw in Section 1.5 that if $\mathscr{S}$ is finite, it contains at least one recurrent state. The same argument shows that any finite closed set of states contains at least one recurrent state. Now let C be a finite irreducible closed set. We have seen that either every state in C is transient or every state in C is recurrent, and that C has at least one recurrent state. It follows that every state in C is recurrent. We summarize this result:

Theorem 3 *Let C be a finite irreducible closed set of states. Then every state in C is recurrent.*

Consider a Markov chain having a finite number of states. Theorem 3 implies that if the chain is irreducible it must be recurrent. If the chain is not irreducible, we can use Theorems 2 and 3 to determine which states are recurrent and which are transient.

Example 10. Consider a Markov chain having the transition matrix

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

Determine which states are recurrent and which states are transient.

As a first step in studying this Markov chain, we determine by inspection which states lead to which other states. This can be indicated in matrix form as

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

The x, y element of this matrix is $+$ or 0 according as ρ_{xy} is positive or zero, i.e., according as x does or does not lead to y. Of course, if $P(x, y) > 0$, then $\rho_{xy} > 0$. The converse is certainly not true in general. For example, $P(2, 0) = 0$; but

$$P^2(2, 0) = P(2, 1)P(1, 0) = \tfrac{1}{5} \cdot \tfrac{1}{4} = \tfrac{1}{20} > 0,$$

so that $\rho_{20} > 0$.

State 0 is an absorbing state, and hence also a recurrent state. We see clearly from the matrix of $+$'s and 0's that $\{3, 4, 5\}$ is an irreducible closed set. Theorem 3 now implies that 3, 4, and 5 are recurrent states. States 1 and 2 both lead to 0, but neither can be reached from 0. We see from Theorem 2 that 1 and 2 must both be transient states. In summary, states 1 and 2 are transient, and states 0, 3, 4, and 5 are recurrent.

Let $\mathscr{S}_T$ denote the collection of transient states in $\mathscr{S}$, and let $\mathscr{S}_R$ denote the collection of recurrent states in $\mathscr{S}$. In Example 10, $\mathscr{S}_T = \{1, 2\}$ and $\mathscr{S}_R = \{0, 3, 4, 5\}$. The set $\mathscr{S}_R$ can be decomposed into the disjoint irreducible closed sets $C_1 = \{0\}$ and $C_2 = \{3, 4, 5\}$. The next theorem shows that such a decomposition is always possible whenever $\mathscr{S}_R$ is nonempty.

Theorem 4 *Suppose that the set $\mathscr{S}_R$ of recurrent states is nonempty. Then $\mathscr{S}_R$ is the union of a finite or countably infinite number of disjoint irreducible closed sets $C_1, C_2, \ldots$.*

Proof. Choose $x \in \mathscr{S}_R$ and let C be the set of all states y in $\mathscr{S}_R$ such that x leads to y. Since x is recurrent, $\rho_{xx} = 1$ and hence $x \in C$. We will now verify that C is an irreducible closed set. Suppose that y is in C and y leads to z. Since y is recurrent, it follows from Theorem 2 that z is recurrent. Since x leads to y and y leads to z, we conclude that x leads to z. Thus z is in C. This shows that C is closed. Suppose that y and z are both in C. Since x is recurrent and x leads to y, it follows from

Theorem 2 that y leads to x. Since y leads to x and x leads to z, we conclude that y leads to z. This shows that C is irreducible.

To complete the proof of the theorem, we need only show that if C and D are two irreducible closed subsets of $\mathscr{S}_R$, they are either disjoint or identical. Suppose they are not disjoint and let x be in both C and D. Choose y in C. Now x leads to y, since x is in C and C is irreducible. Since D is closed, x is in D, and x leads to y, we conclude that y is in D. Thus every state in C is also in D. Similarly every state in D is also in C, so that C and D are identical. ∎

We can use our decomposition of the state space of a Markov chain to understand the behavior of such a system. If the Markov chain starts out in one of the irreducible closed sets C_i of recurrent states, it stays in C_i forever and, with probability one, visits every state in C_i infinitely often. If the Markov chain starts out in the set of transient states $\mathscr{S}_T$, it either stays in $\mathscr{S}_T$ forever or, at some time, enters one of the sets C_i and stays there from that time on, again visiting every state in that C_i infinitely often.

1.6.1 Absorption probabilities. Let C be one of the irreducible closed sets of recurrent states, and let $\rho_C(x) = P_x(T_C < \infty)$ be the probability that a Markov chain starting at x eventually hits C. Since the chain remains permanently in C once it hits that set, we call $\rho_C(x)$ the probability that a chain starting at x is *absorbed* by the set C. Clearly $\rho_C(x) = 1$, $x \in C$, and $\rho_C(x) = 0$ if x is a recurrent state not in C. It is not so clear how to compute $\rho_C(x)$ for $x \in \mathscr{S}_T$, the set of transient states.

If there are only a finite number of transient states, and in particular if $\mathscr{S}$ itself is finite, it is always possible to compute $\rho_C(x)$, $x \in \mathscr{S}_T$, by solving a system of linear equations in which there are as many equations as unknowns, i.e., members of $\mathscr{S}_T$. To understand why this is the case, observe that if $x \in \mathscr{S}_T$, a chain starting at x can enter C only by entering C at time 1 or by being in $\mathscr{S}_T$ at time 1 and entering C at some future time. The former event has probability $\sum_{y \in C} P(x, y)$ and the latter event has probability $\sum_{y \in \mathscr{S}_T} P(x, y)\rho_C(y)$. Thus

$$(44) \qquad \rho_C(x) = \sum_{y \in C} P(x, y) + \sum_{y \in \mathscr{S}_T} P(x, y)\rho_C(y), \qquad x \in \mathscr{S}_T.$$

Equation (44) holds whether $\mathscr{S}_T$ is finite or infinite, but it is far from clear how to solve (44) for the unknowns $\rho_C(x)$, $x \in \mathscr{S}_T$, when $\mathscr{S}_T$ is infinite. An additional difficulty is that if $\mathscr{S}_T$ is infinite, then (44) need not have a unique solution. Fortunately this difficulty does not arise if $\mathscr{S}_T$ is finite.

Theorem 5 *Suppose the set* $\mathscr{S}_T$ *of transient states is finite and let* C *be an irreducible closed set of recurrent states. Then the system of equations*

$$f(x) = \sum_{y \in C} P(x, y) + \sum_{y \in \mathscr{S}_T} P(x, y)f(y), \qquad x \in \mathscr{S}_T, \tag{45}$$

has the unique solution

$$f(x) = \rho_C(x), \qquad x \in \mathscr{S}_T. \tag{46}$$

Proof. If (45) holds, then

$$f(y) = \sum_{z \in C} P(y, z) + \sum_{z \in \mathscr{S}_T} P(y, z)f(z), \qquad y \in \mathscr{S}_T.$$

Substituting this into (45) we find that

$$f(x) = \sum_{y \in C} P(x, y) + \sum_{y \in \mathscr{S}_T} \sum_{z \in C} P(x, y)P(y, z) + \sum_{y \in \mathscr{S}_T} \sum_{z \in \mathscr{S}_T} P(x, y)P(y, z)f(z).$$

The sum of the first two terms is just $P_x(T_C \le 2)$, and the third term reduces to $\sum_{z \in \mathscr{S}_T} P^2(x, z)f(z)$, which is the same as $\sum_{y \in \mathscr{S}_T} P^2(x, y)f(y)$. Thus

$$f(x) = P_x(T_C \le 2) + \sum_{y \in \mathscr{S}_T} P^2(x, y)f(y).$$

By repeating this argument indefinitely or by using induction, we conclude that for all positive integers n

$$f(x) = P_x(T_C \le n) + \sum_{y \in \mathscr{S}_T} P^n(x, y)f(y), \qquad x \in \mathscr{S}_T. \tag{47}$$

Since each $y \in \mathscr{S}_T$ is transient, it follows from (38) that

$$\lim_{n \to \infty} P^n(x, y) = 0, \qquad x \in \mathscr{S} \quad \text{and} \quad y \in \mathscr{S}_T. \tag{48}$$

According to the assumptions of the theorem, $\mathscr{S}_T$ is a finite set. It therefore follows from (48) that the sum in (47) approaches zero as $n \to \infty$. Consequently for $x \in \mathscr{S}_T$

$$f(x) = \lim_{n \to \infty} P_x(T_C \le n) = P_x(T_C < \infty) = \rho_C(x),$$

as desired. ∎

Example 11. Consider the Markov chain discussed in Example 10. Find

$$\rho_{10} = \rho_{\{0\}}(1) \qquad \text{and} \qquad \rho_{20} = \rho_{\{0\}}(2).$$

From (44) and the transition matrix in Example 10, we see that ρ_{10} and ρ_{20} are determined by the equations

$$\rho_{10} = \tfrac{1}{4} + \tfrac{1}{2}\rho_{10} + \tfrac{1}{4}\rho_{20}$$

and

$$\rho_{20} = \tfrac{1}{5}\rho_{10} + \tfrac{2}{5}\rho_{20}.$$

Solving these equations we find that $\rho_{10} = \frac{3}{5}$ and $\rho_{20} = \frac{1}{5}$.

By similar methods we conclude that $\rho_{\{3,4,5\}}(1) = \frac{2}{5}$ and $\rho_{\{3,4,5\}}(2) = \frac{4}{5}$. Alternatively, we can obtain these probabilities by subtracting $\rho_{\{0\}}(1)$ and $\rho_{\{0\}}(2)$ from 1, since if there are only a finite number of transient states,

$$(49) \qquad \sum_i \rho_{C_i}(x) = 1, \qquad x \in \mathscr{S}_T.$$

To verify (49) we note that for $x \in \mathscr{S}_T$

$$\sum_i \rho_{C_i}(x) = \sum_i P_x(T_{C_i} < \infty) = P_x(T_{\mathscr{S}_R} < \infty).$$

Since there are only a finite number of transient states and each transient state is visited only finitely many times, the probability $P_x(T_{\mathscr{S}_R} < \infty)$ that a recurrent state will eventually be hit is 1, so (49) holds.

Once a Markov chain starting at a transient state x enters an irreducible closed set C of recurrent states, it visits every state in C. Thus

$$(50) \qquad \rho_{xy} = \rho_C(x), \qquad x \in \mathscr{S}_T \text{ and } y \in C.$$

It follows from (50) that in our previous example

$$\rho_{13} = \rho_{14} = \rho_{15} = \rho_{\{3,4,5\}}(1) = \tfrac{2}{5}$$

and

$$\rho_{23} = \rho_{24} = \rho_{25} = \rho_{\{3,4,5\}}(2) = \tfrac{4}{5}.$$

1.6.2. Martingales. Consider a Markov chain having state space $\{0, \ldots, d\}$ and transition function P such that

$$(51) \qquad \sum_{y=0}^{d} yP(x, y) = x, \qquad x = 0, \ldots, d.$$

Now

$$\begin{aligned} E[X_{n+1} \mid X_0 &= x_0, \ldots, X_{n-1} = x_{n-1}, X_n = x] \\ &= \sum_{y=0}^{d} yP[X_{n+1} = y \mid X_0 = x_0, \ldots, X_{n-1} = x_{n-1}, X_n = x] \\ &= \sum_{y=0}^{d} yP(x, y) \end{aligned}$$

by the Markov property. We conclude from (51) that

$$(52) \qquad E[X_{n+1} \mid X_0 = x_0, \ldots, X_{n-1} = x_{n-1}, X_n = x] = x,$$

i.e., that the expected value of X_{n+1} given the past and present values of $X_0, \ldots, X_n$ equals the present value of X_n. A sequence of random variables having this property is called a *martingale*. Martingales, which need not be Markov chains, play a very important role in modern probability theory. They arose first in connection with gambling. If X_n denotes the capital of a gambler after time n and if all bets are "fair," that is, if they result in zero expected gain to the gambler, then X_n, $n \geq 0$, forms a martingale. Gamblers were naturally interested in finding some betting strategy, such as increasing their bets until they win, that would give them a net expected gain after making a series of fair bets. That this has been shown to be mathematically impossible does not seem to have deterred them from their quest.

It follows from (51) that

$$\sum_{y=0}^{d} yP(0, y) = 0,$$

and hence that $P(0, 1) = \cdots = P(0, d) = 0$. Thus 0 is necessarily an absorbing state. It follows similarly that d is an absorbing state. Consider now a Markov chain satisfying (51) and having no absorbing states other than 0 and d. It is left as an exercise for the reader to show that under these conditions the states $1, \ldots, d - 1$ each lead to state 0, and hence each is a transient state. If the Markov chain starts at x, it will eventually enter one of the two absorbing states 0 and d and remain there permanently.

It follows from Example 8 that

$$\begin{aligned} E_x(X_n) &= \sum_{y=0}^{d} yP_x(X_n = y) \\ &= \sum_{y=0}^{d} yP^n(x, y) \\ &= \sum_{y=1}^{d-1} yP^n(x, y) + dP^n(x, d) \\ &= \sum_{y=1}^{d-1} yP^n(x, y) + dP_x(T_d \leq n). \end{aligned}$$

Since states $1, 2, \ldots, d - 1$ are transient, we see that $P^n(x, y) \to 0$ as $n \to \infty$ for $y = 1, 2, \ldots, d - 1$. Consequently,

$$\lim_{n \to \infty} E_x(X_n) = dP_x(T_d < \infty) = d\rho_{xd}.$$

On the other hand, it follows from (51) (see Exercise 13(a)) that $EX_n = EX_{n-1} = \cdots = EX_0$ and hence that $E_x(X_n) = x$. Thus

$$\lim_{n \to \infty} E_x(X_n) = x.$$

By equating the two values of this limit, we conclude that

$$\rho_{xd} = \frac{x}{d}, \qquad x = 0, \ldots, d. \tag{53}$$

Since $\rho_{x0} + \rho_{xd} = 1$, it follows from (53) that

$$\rho_{x0} = 1 - \frac{x}{d}, \qquad x = 0, \ldots, d.$$

Of course, once (53) is conjectured, it is easily proved directly from Theorem 5. We need only verify that for $x = 1, \ldots, d - 1$,

$$\frac{x}{d} = P(x, d) + \sum_{y=1}^{d-1} \frac{y}{d} P(x, y). \tag{54}$$

Clearly (54) follows from (51).

The genetics chain introduced in Example 7 satisfies (51) as does a gambler's ruin chain on $\{0, 1, \ldots, d\}$ having transition matrix of the form

$$\begin{bmatrix} 1 & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & & & & \cdot \\ \cdot & \frac{1}{2} & 0 & \frac{1}{2} & & & \cdot \\ \cdot & & & & & & \cdot \\ \cdot & & & & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & \cdot & \cdot & \cdot & \cdot & 0 & 1 \end{bmatrix}.$$

Suppose two gamblers make a series of one dollar bets until one of them goes broke, and suppose that each gambler has probability $\frac{1}{2}$ of winning any given bet. If the first gambler has an initial capital of x dollars and the second gambler has an initial capital of $d - x$ dollars, then the second gambler has probability $\rho_{xd} = x/d$ of going broke and the first gambler has probability $1 - (x/d)$ of going broke.

1.7. Birth and death chains

For an irreducible Markov chain either every state is recurrent or every state is transient, so that an irreducible Markov chain is either a recurrent chain or a transient chain. An irreducible Markov chain having only finitely many states is necessarily recurrent. It is generally difficult to decide whether an irreducible chain having infinitely many states is recurrent or transient. We are able to do so, however, for the birth and death chain.

Consider a birth and death chain on the nonnegative integers or on the finite set $\{0, \ldots, d\}$. In the former case we set $d = \infty$. The transition function is of the form

$$P(x, y) = \begin{cases} q_x, & y = x - 1, \\ r_x, & y = x, \\ p_x, & y = x + 1, \end{cases}$$

where $p_x + q_x + r_x = 1$ for $x \in \mathscr{S}$, $q_0 = 0$, and $p_d = 0$ if $d < \infty$. We assume additionally that p_x and q_x are positive for $0 < x < d$.

For a and b in $\mathscr{S}$ such that $a < b$, set

$$u(x) = P_x(T_a < T_b), \qquad a < x < b,$$

and set $u(a) = 1$ and $u(b) = 0$. If the birth and death chain starts at y, then in one step it goes to $y - 1$, y, or $y + 1$ with respective probabilities q_y, r_y, or p_y. It follows that

(55) $$u(y) = q_y u(y - 1) + r_y u(y) + p_y u(y + 1), \qquad a < y < b.$$

Since $r_y = 1 - p_y - q_y$, we can rewrite (55) as

(56) $$u(y + 1) - u(y) = \frac{q_y}{p_y}(u(y) - u(y - 1)), \qquad a < y < b.$$

Set $\gamma_0 = 1$ and

(57) $$\gamma_y = \frac{q_1 \cdots q_y}{p_1 \cdots p_y}, \qquad 0 < y < d.$$

From (56) we see that

$$u(y + 1) - u(y) = \frac{\gamma_y}{\gamma_{y-1}}(u(y) - u(y - 1)), \qquad a < y < b,$$

from which it follows that

$$\begin{aligned} u(y + 1) - u(y) &= \frac{\gamma_{a+1}}{\gamma_a} \cdots \frac{\gamma_y}{\gamma_{y-1}}(u(a + 1) - u(a)) \\ &= \frac{\gamma_y}{\gamma_a}(u(a + 1) - u(a)). \end{aligned}$$

Consequently,

(58) $$u(y) - u(y + 1) = \frac{\gamma_y}{\gamma_a}(u(a) - u(a + 1)), \qquad a \le y < b.$$

Summing (58) on $y = a, \ldots, b - 1$ and recalling that $u(a) = 1$ and $u(b) = 0$, we conclude that

$$\frac{u(a) - u(a + 1)}{\gamma_a} = \frac{1}{\sum_{y=a}^{b-1} \gamma_y}.$$

Thus (58) becomes

$$u(y) - u(y+1) = \frac{\gamma_y}{\sum_{y=a}^{b-1} \gamma_y}, \qquad a \le y < b.$$

Summing this equation on $y = x, \ldots, b-1$ and again using the formula $u(b) = 0$, we obtain

$$u(x) = \frac{\sum_{y=x}^{b-1} \gamma_y}{\sum_{y=a}^{b-1} \gamma_y}, \qquad a < x < b.$$

It now follows from the definition of $u(x)$ that

$$P_x(T_a < T_b) = \frac{\sum_{y=x}^{b-1} \gamma_y}{\sum_{y=a}^{b-1} \gamma_y}, \qquad a < x < b. \tag{59}$$

By subtracting both sides of (59) from 1, we see that

$$P_x(T_b < T_a) = \frac{\sum_{y=a}^{x-1} \gamma_y}{\sum_{y=a}^{b-1} \gamma_y}, \qquad a < x < b. \tag{60}$$

Example 12. A gambler playing roulette makes a series of one dollar bets. He has respective probabilities 9/19 and 10/19 of winning and losing each bet. The gambler decides to quit playing as soon as his net winnings reach 25 dollars or his net losses reach 10 dollars.

(a) Find the probability that when he quits playing he will have won 25 dollars.

(b) Find his expected loss.

The problem fits into our scheme if we let X_n denote the capital of the gambler at time n with $X_0 = 10$. Then X_n, $n \ge 0$, forms a birth and death chain on $\{0, 1, \ldots, 35\}$ with birth and death rates

$$p_x = 9/19, \qquad 0 < x < 35,$$

and

$$q_x = 10/19, \qquad 0 < x < 35.$$

States 0 and 35 are absorbing states. Formula (60) is applicable with $a = 0$, $x = 10$, and $b = 35$. We conclude that

$$\gamma_y = (10/9)^y, \qquad 0 \le y \le 34,$$

and hence that

$$P_{10}(T_{35} < T_0) = \frac{\sum_{y=0}^{9} (10/9)^y}{\sum_{y=0}^{34} (10/9)^y} = \frac{(10/9)^{10} - 1}{(10/9)^{35} - 1} = .047.$$

Thus the gambler has probability .047 of winning 25 dollars. His expected loss in dollars is $10 - 35(.047)$, which equals \$8.36.

In the remainder of this section we consider a birth and death chain on the nonnegative integers which is irreducible, i.e., such that $p_x > 0$ for $x \geq 0$ and $q_x > 0$ for $x \geq 1$. We will determine when such a chain is recurrent and when it is transient.

As a special case of (59),

$$P_1(T_0 < T_n) = 1 - \frac{1}{\sum_{y=0}^{n-1} \gamma_y}, \qquad n > 1. \tag{61}$$

Consider now a birth and death chain starting in state 1. Since the birth and death chain can move at most one step to the right at a time (considering the transition from state to state as movement along the real number line),

$$1 \leq T_2 < T_3 < \cdots. \tag{62}$$

It follows from (62) that $\{T_0 < T_n\}$, $n > 1$, forms a nondecreasing sequence of events. We conclude from Theorem 1 of Chapter 1 of Volume I[1] that

$$\lim_{n \to \infty} P_1(T_0 < T_n) = P_1(T_0 < T_n \text{ for some } n > 1). \tag{63}$$

Equation (62) implies that $T_n \geq n$ and thus $T_n \to \infty$ as $n \to \infty$; hence the event $\{T_0 < T_n \text{ for some } n > 1\}$ occurs if and only if the event $\{T_0 < \infty\}$ occurs. We can therefore rewrite (63) as

$$\lim_{n \to \infty} P_1(T_0 < T_n) = P_1(T_0 < \infty). \tag{64}$$

It follows from (61) and (64) that

$$P_1(T_0 < \infty) = 1 - \frac{1}{\sum_{y=0}^{\infty} \gamma_y}. \tag{65}$$

We are now in position to show that the birth and death chain is recurrent if and only if

$$\sum_{y=0}^{\infty} \gamma_y = \infty. \tag{66}$$

If the birth and death chain is recurrent, then $P_1(T_0 < \infty) = 1$ and (66) follows from (65). To obtain the converse, we observe that $P(0, y) = 0$ for $y \geq 2$, and hence

$$P_0(T_0 < \infty) = P(0, 0) + P(0, 1)P_1(T_0 < \infty). \tag{67}$$

---

[1] Paul G. Hoel, Sidney C. Port, and Charles J. Stone, *Introduction to Probability Theory* (Boston: Houghton Mifflin Co., 1971), p. 13.

Suppose (66) holds. Then by (65)

$$P_1(T_0 < \infty) = 1.$$

From this and (67) we conclude that

$$P_0(T_0 < \infty) = P(0, 0) + P(0, 1) = 1.$$

Thus 0 is a recurrent state, and since the chain is assumed to be irreducible, it must be a recurrent chain.

In summary, we have shown that an irreducible birth and death chain on $\{0, 1, 2, \ldots\}$ is recurrent if and only if

$$\sum_{x=1}^{\infty} \frac{q_1 \cdots q_x}{p_1 \cdots p_x} = \infty. \tag{68}$$

Example 13. Consider the birth and death chain on $\{0, 1, 2, \ldots\}$ defined by

$$p_x = \frac{x + 2}{2(x + 1)} \quad \text{and} \quad q_x = \frac{x}{2(x + 1)}, \qquad x \geq 0.$$

Determine whether this chain is recurrent or transient.

Since

$$\frac{q_x}{p_x} = \frac{x}{x + 2},$$

it follows that

$$\begin{aligned} \gamma_x = \frac{q_1 \cdots q_x}{p_1 \cdots p_x} &= \frac{1 \cdot 2 \cdots x}{3 \cdot 4 \cdots (x + 2)} \\ &= \frac{2}{(x + 1)(x + 2)} = 2\left(\frac{1}{x + 1} - \frac{1}{x + 2}\right). \end{aligned}$$

Thus

$$\begin{aligned} \sum_{x=1}^{\infty} \gamma_x &= 2 \sum_{x=1}^{\infty} \left(\frac{1}{x + 1} - \frac{1}{x + 2}\right) \\ &= 2(\tfrac{1}{2} - \tfrac{1}{3} + \tfrac{1}{3} - \tfrac{1}{4} + \tfrac{1}{4} - \tfrac{1}{5} + \cdots) \\ &= 2 \cdot \tfrac{1}{2} = 1. \end{aligned}$$

We conclude that the chain is transient.

1.8. Branching and queuing chains

In this section we will describe which branching chains are certain of extinction and which are not. We will also describe which queuing chains are transient and which are recurrent. The proofs of these results are somewhat complicated and will be given in the appendix to this chapter. These proofs can be skipped with no loss of continuity. It is interesting to note that the proofs of the results for the branching chain and the queuing chain are very similar, whereas the results themselves appear quite dissimilar.

1.8.1. Branching chain. Consider the branching chain introduced in Example 6. The extinction probability ρ of the chain is the probability that the descendants of a given particle eventually become extinct. Clearly

$$\rho = \rho_{10} = P_1(T_0 < \infty).$$

Suppose there are x particles present initially. Since the numbers of offspring of these particles in the various generations are chosen independently of each other, the probability ρ_{x0} that the descendants of each of the x particles eventually become extinct is just the xth power of the probability that the descendants of any one particle eventually become extinct. In other words,

$$\rho_{x0} = \rho^x, \qquad x = 1, 2, \ldots. \tag{69}$$

Recall from Example 6 that a particle gives rise to ξ particles in the next generation, where ξ is a random variable having density f. If $f(1) = 1$, the branching chain is degenerate in that every state is an absorbing state. Thus we suppose that $f(1) < 1$. Then state 0 is an absorbing state. It is left as an exercise for the reader to show that every state other than 0 is transient. From this it follows that, with probability one, the branching chain is either absorbed at 0 or approaches $+\infty$. We conclude from (69) that

$$P_x(\lim_{n\to\infty} X_n = \infty) = 1 - \rho^x, \qquad x = 1, 2, \ldots.$$

Clearly it is worthwhile to determine ρ or at least to determine when $\rho = 1$ and when $\rho < 1$. This can be done using arguments based upon the formula

$$\Phi(\rho) = \rho, \tag{70}$$

where Φ is the probability generating function of f, defined by

$$\Phi(t) = f(0) + \sum_{y=1}^{\infty} f(y)t^y, \qquad 0 \le t \le 1.$$

To verify (70) we observe that (see Exercise 9(b))

$$\begin{aligned}\rho = \rho_{10} &= P(1, 0) + \sum_{y=1}^{\infty} P(1, y)\rho_{y0} \\ &= P(1, 0) + \sum_{y=1}^{\infty} P(1, y)\rho^y \\ &= f(0) + \sum_{y=1}^{\infty} f(y)\rho^y \\ &= \Phi(\rho).\end{aligned}$$

Let μ denote the expected number of offspring of any given particle. Suppose $\mu \leq 1$. Then the equation $\Phi(t) = t$ has no roots in $[0, 1)$ (under our assumption that $f(1) < 1$), and hence $\rho = 1$. Thus *ultimate extinction is certain if* $\mu \leq 1$ *and* $f(1) < 1$.

Suppose instead that $\mu > 1$. Then the equation $\Phi(t) = t$ has a unique root ρ_0 in $[0, 1)$, and hence ρ equals either ρ_0 or 1. Actually ρ always equals ρ_0. Consequently, *if* $\mu > 1$ *the probability of ultimate extinction is less than one.*

The proofs of these results will be given in the appendix. The results themselves are intuitively very reasonable. If $\mu < 1$, then on the average each particle gives rise to fewer than one new particle, so we would expect the population to die out eventually. If $\mu > 1$, then on the average each particle gives rise to more than one new particle. In this case we would expect that the population has positive probability of growing rapidly, indeed geometrically fast, as time goes on. The case $\mu = 1$ is borderline; but since $\rho = 1$ when $\mu < 1$, it is plausible by "continuity" that $\rho = 1$ also when $\mu = 1$.

Example 14. Suppose that every man in a certain society has exactly three children, which independently have probability one-half of being a boy and one-half of being a girl. Suppose also that the number of males in the nth generation forms a branching chain. Find the probability that the male line of a given man eventually becomes extinct.

The density f of the number of male children of a given man is the binomial density with parameters $n = 3$ and $p = \frac{1}{2}$. Thus $f(0) = \frac{1}{8}$, $f(1) = \frac{3}{8}$, $f(2) = \frac{3}{8}$, $f(3) = \frac{1}{8}$, and $f(x) = 0$ for $x \geq 4$. The mean number of male offspring is $\mu = \frac{3}{2}$. Since $\mu > 1$, the extinction probability ρ is the root of the equation

$$\frac{1}{8} + \frac{3}{8}t + \frac{3}{8}t^2 + \frac{1}{8}t^3 = t$$

lying in $[0, 1)$. We can rewrite this equation as

$$t^3 + 3t^2 - 5t + 1 = 0,$$

or equivalently as

$$(t - 1)(t^2 + 4t - 1) = 0.$$

This equation has three roots, namely, 1, $-\sqrt{5} - 2$, and $\sqrt{5} - 2$. Consequently, $\rho = \sqrt{5} - 2$.

1.8.2. Queuing chain. Consider the queuing chain introduced in Example 5. Let $\xi_1, \xi_2, \ldots$ and μ be as in that example. In this section we will indicate when the queuing chain is recurrent and when it is transient.

Let μ denote the expected number of customers arriving in unit time. Suppose first that $\mu > 1$. Since at most one person is served at a time and on the average more than one new customer enters the queue at a time, it would appear that as time goes on more and more people will be waiting for service and that the queue length will approach infinity. This is indeed the case, so that *if $\mu > 1$ the queuing chain is transient.*

In discussing the case $\mu \leq 1$, we will assume that the chain is irreducible (see Exercises 37 and 38 for necessary and sufficient conditions for irreducibility and for results when the queuing chain is not irreducible). Suppose first that $\mu < 1$. Then on the average fewer than one new customer will enter the queue in unit time. Since one customer is served whenever the queue is nonempty, we would expect that, regardless of the initial length of the queue, it will become empty at some future time. This is indeed the case and, in particular, 0 is a recurrent state. The case $\mu = 1$ is borderline, but again it turns out that 0 is a recurrent state. Thus *if $\mu \leq 1$ and the queuing chain is irreducible, it is recurrent.*

The proof of these results will be given in the appendix.

APPENDIX

1.9. Proof of results for the branching and queuing chains

In this section we will verify the results discussed in Section 1.8. To do so we need the following.

Theorem 6 *Let Φ be the probability generating function of a nonnegative integer-valued random variable ξ and set $\mu = E\xi$ (with $\mu = +\infty$ if ξ does not have finite expectation). If $\mu \leq 1$ and $P(\xi = 1) < 1$, the equation*

$$\Phi(t) = t \tag{71}$$

has no roots in $[0, 1)$. If $\mu > 1$, then (71) has a unique root ρ_0 in $[0, 1)$.

Graphs of $\Phi(t)$, $0 \le t \le 1$, in three typical cases corresponding to $\mu < 1$, $\mu = 1$, and $\mu > 1$ are shown in Figure 2. The fact that μ is the left-hand derivative of $\Phi(t)$ at $t = 1$ plays a fundamental role in the proof of Theorem 6.

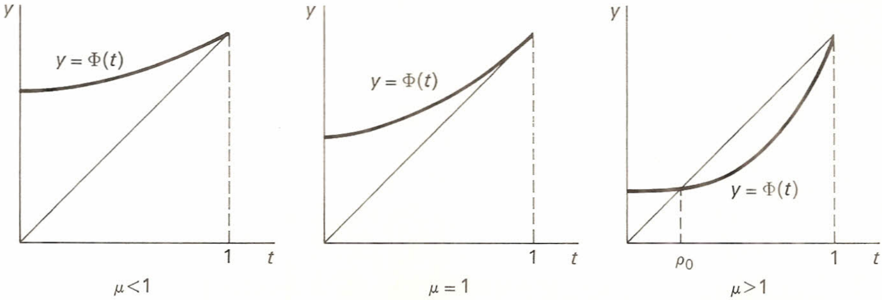

Figure 2

Proof. Let f denote the density of ξ. Then

$$\Phi(t) = f(0) + f(1)t + f(2)t^2 + \cdots$$

and

$$\Phi'(t) = f(1) + 2f(2)t + 3f(3)t^2 + \cdots.$$

Thus $\Phi(0) = f(0)$, $\Phi(1) = 1$, and

$$\lim_{t \to 1} \Phi'(t) = f(1) + 2f(2) + 3f(3) + \cdots = \mu.$$

Suppose first that $\mu < 1$. Then

$$\lim_{t \to 1} \Phi'(t) < 1.$$

Since $\Phi'(t)$ is nondecreasing in t, $0 \le t < 1$, we conclude that $\Phi'(t) < 1$ for $0 \le t < 1$. Suppose next that $\mu = 1$ and $f(1) = P(\xi = 1) < 1$. Then $f(n) > 0$ for some $n \ge 2$ (otherwise $f(0) > 0$, which implies that $\mu < 1$, a contradiction). Therefore $\Phi'(t)$ is strictly increasing in t, $0 \le t < 1$. Since

$$\lim_{t \to 1} \Phi'(t) = 1,$$

we again conclude that $\Phi'(t) < 1$ for $0 \le t < 1$.

Suppose now that $\mu \le 1$ and $P(\xi = 1) < 1$. We have shown that $\Phi'(t) < 1$ for $0 \le t < 1$. Thus

$$\frac{d}{dt}(\Phi(t) - t) < 0, \qquad 0 \le t < 1,$$

and hence $\Phi(t) - t$ is strictly decreasing on $[0, 1]$. Since $\Phi(1) - 1 = 0$, we see that $\Phi(t) - t > 0$, $0 \le t < 1$, and hence that (71) has no roots on $[0, 1)$. This proves the first part of the theorem.

Suppose next that $\mu > 1$. Then

$$\lim_{t \to 1} \Phi'(t) > 1,$$

so by the continuity of Φ' there is a number t_0 such that $0 < t_0 < 1$ and $\Phi'(t) > 1$ for $t_0 < t < 1$. It follows from the mean value theorem that

$$\frac{\Phi(1) - \Phi(t_0)}{1 - t_0} > 1.$$

Since $\Phi(1) = 1$, we conclude that $\Phi(t_0) - t_0 < 0$. Now $\Phi(t) - t$ is continuous in t and nonnegative at $t = 0$, so by the intermediate value theorem it must have a zero ρ_0 on $[0, t_0)$. Thus (71) has a root ρ_0 in $[0, 1)$. We will complete the proof of the theorem by showing that there is only one such root.

Suppose that $0 \le \rho_0 < \rho_1 < 1$, $\Phi(\rho_0) = \rho_0$, and $\Phi(\rho_1) = \rho_1$. Then the function $\Phi(t) - t$ vanishes at ρ_0, ρ_1, and 1; hence by Rolle's theorem its first derivative has at least two roots in (0, 1). By another application of Rolle's theorem its second derivative $\Phi''(t)$ has at least one root in (0, 1). But if $\mu > 1$, then at least one of the numbers $f(2), f(3), \ldots$ is strictly positive, and hence

$$\Phi''(t) = 2f(2) + 3 \cdot 2f(3)t + \cdots$$

has no roots in (0, 1). This contradiction shows that $\Phi(t) = t$ has a unique root in $[0, 1)$. ∎

1.9.1. Branching chain. Using Theorem 6 we see that the results for $\mu \le 1$ follow as indicated in Section 1.8.1.

Suppose $\mu > 1$. It follows from Theorem 6 that ρ equals ρ_0 or 1, where ρ_0 is the unique root of the equation $\Phi(t) = t$ in $[0, 1)$. We will show that ρ always equals ρ_0.

First we observe that since the initial particles act independently in giving rise to their offspring, the probability $P_y(T_0 \le n)$ that the descendants of each of the $y \ge 1$ particles become extinct by time n is given by

$$P_y(T_0 \le n) = (P_1(T_0 \le n))^y.$$

Consequently for $n \ge 0$ by Exercise 9(a)

$$\begin{aligned} P_1(T_0 \le n + 1) &= P(1, 0) + \sum_{y=1}^{\infty} P(1, y)P_y(T_0 \le n) \\ &= P(1, 0) + \sum_{y=1}^{\infty} P(1, y)(P_1(T_0 \le n))^y \\ &= f(0) + \sum_{y=1}^{\infty} f(y)(P_1(T_0 \le n))^y, \end{aligned}$$

and hence

(72) $$P_1(T_0 \le n + 1) = \Phi(P_1(T_0 \le n)), \qquad n \ge 0.$$

We will use (72) to prove by induction that

(73) $$P_1(T_0 \le n) \le \rho_0, \qquad n \ge 0.$$

Now

$$P_1(T_0 \le 0) = 0 \le \rho_0,$$

so that (73) is true for $n = 0$. Suppose that (73) holds for a given value of n. Since $\Phi(t)$ is increasing in t, we conclude from (72) that

$$P_1(T_0 \le n + 1) = \Phi(P_1(T_0 \le n)) \le \Phi(\rho_0) = \rho_0,$$

and thus (73) holds for the next value of n. By induction (73) is true for all $n \ge 0$.

By letting $n \to \infty$ in (73) we see that

$$\rho = P_1(T_0 < \infty) = \lim_{n\to\infty} P_1(T_0 \le n) \le \rho_0.$$

Since ρ is one of the two numbers ρ_0 or 1, it must be the number ρ_0.

1.9.2. Queuing chain. We will now verify the results of Section 1.8.2. Let ξ_n denote the number of customers arriving during the nth time period. Then $\xi_1, \xi_2, \ldots$ are independent random variables having common density f, mean μ, and probability generating function Φ.

It follows from Exercise 9(b) and the identity $P(0, z) \equiv P(1, z)$, valid for a queuing chain, that $\rho_{00} = \rho_{10}$. We will show that the number $\rho = \rho_{00} = \rho_{10}$ satisfies the equation

(74) $$\Phi(\rho) = \rho.$$

If 0 is a recurrent state, $\rho = 1$ and (74) follows immediately from the fact that $\Phi(1) = 1$. To verify (74) in general, we observe first that by Exercise 9(b)

$$\rho_{00} = P(0, 0) + \sum_{y=1}^{\infty} P(0, y)\rho_{y0},$$

i.e., that

(75) $$\rho = f(0) + \sum_{y=1}^{\infty} f(y)\rho_{y0}.$$

In order to compute ρ_{y0}, $y = 1, 2, \ldots$, we consider a queuing chain starting at the positive integer y. For $n = 1, 2, \ldots$, the event $\{T_{y-1} = n\}$ occurs if and only if

$$\begin{aligned} n &= \min\,(m > 0: y + (\xi_1 - 1) + \cdots + (\xi_m - 1) = y - 1) \\ &= \min\,(m > 0: \xi_1 + \cdots + \xi_m = m - 1), \end{aligned}$$

that is, if and only if n is the smallest positive integer m such that the number of new customers entering the queue by time m is one less than the number served by time m. Thus $P_y(T_{y-1} = n)$ is independent of y, and consequently $\rho_{y,y-1} = P_y(T_{y-1} < \infty)$ is independent of y for $y = 1, 2, \ldots$. Since $\rho_{10} = \rho$, we see that

$$\rho_{y,y-1} = \rho_{y-1,y-2} = \cdots = \rho_{10} = \rho.$$

Now the queuing chain can go at most one step to the left at a time, so in order to go from state $y > 0$ to state 0 it must pass through all the intervening states $y - 1, \ldots, 1$. By applying the Markov property we can conclude (see Exercise 39) that

$$\rho_{y0} = \rho_{y,y-1}\rho_{y-1,y-2}\cdots\rho_{10} = \rho^y. \tag{76}$$

It follows from (75) and (76) that

$$\rho = f(0) + \sum_{y=1}^{\infty} f(y)\rho^y = \Phi(\rho),$$

so that (74) holds.

Using (74) and Theorem 6 it is easy to see that if $\mu \leq 1$ and the queuing chain is irreducible, then the chain is recurrent. For ρ satisfies (74) and by Theorem 6 this equation has no roots in $[0, 1)$ (observe that $P(\xi_1 = 1) < 1$ if the queuing chain is irreducible). We conclude that $\rho = 1$. Since $\rho_{00} = \rho$, state 0 is recurrent, and thus since the chain is irreducible, all states are recurrent.

Suppose now that $\mu > 1$. Again ρ satisfies (74) which, by Theorem 6, has a unique root ρ_0 in $[0, 1)$. Thus ρ equals either ρ_0 or 1. We will prove that $\rho = \rho_0$.

To this end we first observe that by Exercise 9(a)

$$P_1(T_0 \leq n + 1) = P(1, 0) + \sum_{y=1}^{\infty} P(1, y)P_y(T_0 \leq n),$$

which can be rewritten as

$$P_1(T_0 \leq n + 1) = f(0) + \sum_{y=1}^{\infty} f(y)P_y(T_0 \leq n). \tag{77}$$

We claim next that

$$P_y(T_0 \leq n) \leq (P_1(T_0 \leq n))^y, \qquad y \geq 1 \text{ and } n \geq 0. \tag{78}$$

To verify (78) observe that if a queuing chain starting at y reaches 0 in n or fewer steps, it must reach $y - 1$ in n or fewer steps, go from $y - 1$ to $y - 2$ in n or fewer steps, etc. By applying the Markov property we can conclude (see Exercise 39) that

$$P_y(T_0 \leq n) \leq P_y(T_{y-1} \leq n)P_{y-1}(T_{y-2} \leq n)\cdots P_1(T_0 \leq n). \tag{79}$$

Since

$$P_z(T_{z-1} \le n) = P_1(T_0 \le n), \qquad 1 \le z \le y,$$

(78) is valid.

It follows from (77) and (78) that

$$P_1(T_0 \le n + 1) \le f(0) + \sum_{y=1}^{\infty} f(y)(P_1(T_0 \le n))^y,$$

i.e., that

$$(80) \qquad P_1(T_0 \le n + 1) \le \Phi(P_1(T_0 \le n)), \qquad n \ge 0.$$

This in turn implies that

$$(81) \qquad P_1(T_0 \le n) \le \rho_0, \qquad n \ge 0,$$

by a proof that is almost identical to the proof that (72) implies (73) (the slight changes needed are left as an exercise for the reader). Just as in the proof of the corresponding result for the branching chain, we see by letting $n \to \infty$ in (81) that $\rho \le \rho_0$ and hence that $\rho = \rho_0$.

We have shown that if $\mu > 1$, then $\rho_{00} = \rho < 1$, and hence 0 is a transient state. It follows that if $\mu > 1$ and the chain is irreducible, then all states are transient. If $\mu > 1$ and the queuing chain is not irreducible, then case (d) of Exercise 38 holds (why?), and it is left to the reader to show that again all states are transient.

Exercises

1 Let X_n, $n \ge 0$, be the two-state Markov chain. Find
(a) $P(X_1 = 0 \mid X_0 = 0 \text{ and } X_2 = 0)$,
(b) $P(X_1 \ne X_2)$.

2 Suppose we have two boxes and $2d$ balls, of which d are black and d are red. Initially, d of the balls are placed in box 1, and the remainder of the balls are placed in box 2. At each trial a ball is chosen at random from each of the boxes, and the two balls are put back in the opposite boxes. Let X_0 denote the number of black balls initially in box 1 and, for $n \ge 1$, let X_n denote the number of black balls in box 1 after the nth trial. Find the transition function of the Markov chain X_n, $n \ge 0$.

3 Let the queuing chain be modified by supposing that if there are one or more customers waiting to be served at the start of a period, there is probability p that one customer will be served during that period and probability $1 - p$ that no customers will be served during that period. Find the transition function for this modified queuing chain.

4 Consider a probability space $(\Omega, \mathscr{A}, P)$ and assume that the various sets mentioned below are all in $\mathscr{A}$.

(a) Show that if D_i are disjoint and $P(C \mid D_i) = p$ independently of i, then $P(C \mid \bigcup_i D_i) = p$.

(b) Show that if C_i are disjoint, then $P(\bigcup_i C_i \mid D) = \sum_i P(C_i \mid D)$.

(c) Show that if E_i are disjoint and $\bigcup_i E_i = \Omega$, then

$$P(C \mid D) = \sum_i P(E_i \mid D)P(C \mid E_i \cap D).$$

(d) Show that if C_i are disjoint and $P(A \mid C_i) = P(B \mid C_i)$ for all i, then $P(A \mid \bigcup_i C_i) = P(B \mid \bigcup_i C_i)$.

5 Let X_n, $n \geq 0$, be the two-state Markov chain.

(a) Find $P_0(T_0 = n)$.

(b) Find $P_0(T_1 = n)$.

6 Let X_n, $n \geq 0$, be the Ehrenfest chain and suppose that X_0 has a binomial distribution with parameters d and $1/2$, i.e.,

$$P(X_0 = x) = \frac{\binom{d}{x}}{2^d}, \qquad x = 0, \ldots, d.$$

Find the distribution of X_1.

7 Let X_n, $n \geq 0$, be a Markov chain. Show that

$$P(X_0 = x_0 \mid X_1 = x_1, \ldots, X_n = x_n) = P(X_0 = x_0 \mid X_1 = x_1).$$

8 Let x and y be distinct states of a Markov chain having $d < \infty$ states and suppose that x leads to y. Let n_0 be the smallest positive integer such that $P^{n_0}(x, y) > 0$ and let $x_1, \ldots, x_{n_0 - 1}$ be states such that

$$P(x, x_1)P(x_1, x_2) \cdots P(x_{n_0 - 2}, x_{n_0 - 1})P(x_{n_0 - 1}, y) > 0.$$

(a) Show that $x, x_1, \ldots, x_{n_0 - 1}, y$ are distinct states.

(b) Use (a) to show that $n_0 \leq d - 1$.

(c) Conclude that $P_x(T_y \leq d - 1) > 0$.

9 Use (29) to verify the following identities:

(a) $P_x(T_y \leq n + 1) = P(x, y) + \sum_{z \neq y} P(x, z)P_z(T_y \leq n), \qquad n \geq 0;$

(b) $\rho_{xy} = P(x, y) + \sum_{z \neq y} P(x, z)\rho_{zy}.$

10 Consider the Ehrenfest chain with $d = 3$.

(a) Find $P_x(T_0 = n)$ for $x \in \mathscr{S}$ and $1 \leq n \leq 3$.

(b) Find P, P^2, and P^3.

(c) Let π_0 be the uniform distribution $\pi_0 = (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$. Find π_1, π_2, and π_3.

11 Consider the genetics chain from Example 7 with $d = 3$.
(a) Find the transition matrices P and P^2.
(b) If $\pi_0 = (0, \frac{1}{2}, \frac{1}{2}, 0)$, find π_1 and π_2.
(c) Find $P_x(T_{\{0,3\}} = n)$, $x \in \mathscr{S}$, for $n = 1$ and $n = 2$.

12 Consider the Markov chain having state space $\{0, 1, 2\}$ and transition matrix

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

(a) Find P^2.
(b) Show that $P^4 = P^2$.
(c) Find P^n, $n \geq 1$.

13 Let X_n, $n \geq 0$, be a Markov chain whose state space $\mathscr{S}$ is a subset of $\{0, 1, 2, \ldots\}$ and whose transition function P is such that

$$\sum_y yP(x, y) = Ax + B, \qquad x \in \mathscr{S},$$

for some constants A and B.
(a) Show that $EX_{n+1} = AEX_n + B$.
(b) Show that if $A \neq 1$, then

$$EX_n = \frac{B}{1 - A} + A^n\left(EX_0 - \frac{B}{1 - A}\right).$$

14 Let X_n, $n \geq 0$, be the Ehrenfest chain on $\{0, 1, \ldots, d\}$. Show that the assumption of Exercise 13 holds and use that exercise to compute $E_x(X_n)$.

15 Let y be a transient state. Use (36) to show that for all x

$$\sum_{n=0}^{\infty} P^n(x, y) \leq \sum_{n=0}^{\infty} P^n(y, y).$$

16 Show that $\rho_{xy} > 0$ if and only if $P^n(x, y) > 0$ for some positive integer n.

17 Show that if x leads to y and y leads to z, then x leads to z.

18 Consider a Markov chain on the nonnegative integers such that, starting from x, the chain goes to state $x + 1$ with probability p, $0 < p < 1$, and goes to state 0 with probability $1 - p$.
(a) Show that this chain is irreducible.
(b) Find $P_0(T_0 = n)$, $n \geq 1$.
(c) Show that the chain is recurrent.

19 Consider a Markov chain having state space $\{0, 1, \ldots, 6\}$ and transition matrix

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

(a) Determine which states are transient and which states are recurrent.
(b) Find ρ_{0y}, $y = 0, \ldots, 6$.

20 Consider the Markov chain on $\{0, 1, \ldots, 5\}$ having transition matrix

$$\begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array}\begin{array}{c} \begin{array}{cccccc} 0 & 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 \\ \frac{1}{3} & \frac{2}{3} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{8} & 0 & \frac{7}{8} & 0 \\ \frac{1}{4} & \frac{1}{4} & 0 & 0 & \frac{1}{4} & \frac{1}{4} \\ 0 & 0 & \frac{3}{4} & 0 & \frac{1}{4} & 0 \\ 0 & \frac{1}{5} & 0 & \frac{1}{5} & \frac{1}{5} & \frac{2}{5} \end{bmatrix} \end{array}.$$

(a) Determine which states are transient and which are recurrent.
(b) Find $\rho_{\{0,1\}}(x)$, $x = 0, \ldots, 5$.

21 Consider a Markov chain on $\{0, 1, \ldots, d\}$ satisfying (51) and having no absorbing states other than 0 and d. Show that the states $1, \ldots, d - 1$ each lead to 0, and hence that each is a transient state.

22 Show that the genetics chain introduced in Example 7 satisfies Equation (51).

23 A certain Markov chain that arises in genetics has states $0, 1, \ldots, 2d$ and transition function

$$P(x, y) = \binom{2d}{y}\left(\frac{x}{2d}\right)^y\left(1 - \frac{x}{2d}\right)^{2d-y}.$$

Find $\rho_{\{0\}}(x)$, $0 < x < 2d$.

24 Consider a gambler's ruin chain on $\{0, 1, \ldots, d\}$. Find

$$P_x(T_0 < T_d), \qquad 0 < x < d.$$

25 A gambler playing roulette makes a series of one dollar bets. He has respective probabilities 9/19 and 10/19 of winning and losing each bet. The gambler decides to quit playing as soon as he either is one dollar ahead or has lost his initial capital of \$1000.

(a) Find the probability that when he quits playing he will have lost \$1000.
(b) Find his expected loss.

26 Consider a birth and death chain on the nonnegative integers such that $p_x > 0$ and $q_x > 0$ for $x \geq 1$.
(a) Show that if $\sum_{y=0}^{\infty} \gamma_y = \infty$, then $\rho_{x0} = 1$, $x \geq 1$.
(b) Show that if $\sum_{y=0}^{\infty} \gamma_y < \infty$, then

$$\rho_{x0} = \frac{\sum_{y=x}^{\infty} \gamma_y}{\sum_{y=0}^{\infty} \gamma_y}, \qquad x \geq 1.$$

27 Consider a gambler's ruin chain on $\{0, 1, 2, \ldots\}$.
(a) Show that if $q \geq p$, then $\rho_{x0} = 1$, $x \geq 1$.
(b) Show that if $q < p$, then $\rho_{x0} = (q/p)^x$, $x \geq 1$.
Hint: Use Exercise 26.

28 Consider an irreducible birth and death chain on the nonnegative integers. Show that if $p_x \leq q_x$ for $x \geq 1$, the chain is recurrent.

29 Consider an irreducible birth and death chain on the nonnegative integers such that

$$\frac{q_x}{p_x} = \left(\frac{x}{x+1}\right)^2, \qquad x \geq 1.$$

(a) Show that this chain is transient.
(b) Find ρ_{x0}, $x \geq 1$. *Hint:* Use Exercise 26 and the formula $\sum_{y=1}^{\infty} 1/y^2 = \pi^2/6$.

30 Consider the birth and death chain in Example 13.
(a) Compute $P_x(T_a < T_b)$ for $a < x < b$.
(b) Compute ρ_{x0}, $x > 0$.

31 Consider a branching chain such that $f(1) < 1$. Show that every state other than 0 is transient.

32 Consider the branching chain described in Example 14. If a given man has two boys and one girl, what is the probability that his male line will continue forever?

33 Consider a branching chain with $f(0) = f(3) = 1/2$. Find the probability ρ of extinction.

34 Consider a branching chain with $f(x) = p(1 - p)^x$, $x \geq 0$, where $0 < p < 1$. Show that $\rho = 1$ if $p \geq 1/2$ and that $\rho = p/(1 - p)$ if $p < 1/2$.

35 Let X_n, $n \geq 0$, be a branching chain. Show that $E_x(X_n) = x\mu^n$. *Hint:* See Exercise 13.

36 Let X_n, $n \geq 0$, be a branching chain and suppose that the associated random variable ξ has finite variance σ^2.
(a) Show that

$$E[X_{n+1}^2 \mid X_n = x] = x\sigma^2 + x^2\mu^2.$$

(b) Use Exercise 35 to show that

$$E_x(X_{n+1}^2) = x\mu^n\sigma^2 + \mu^2 E_x(X_n^2).$$

Hint: Use the formula $EY = \sum_x P(X = x)E[Y \mid X = x]$.