Sheldon Ross, *Introduction to Probability models,* Ed, 7, Academic Press, 2000

4 Markov Chains

4.1. Introduction

In this chapter, we consider a stochastic process $\{X_n, n = 0, 1, 2, \ldots\}$ that takes on a finite or countable number of possible values. Unless otherwise mentioned, this set of possible values of the process will be denoted by the set of nonnegative integers $\{0, 1, 2, \ldots\}$. If $X_n = i$, then the process is said to be in state i at time n. We suppose that whenever the process is in state i, there is a fixed probability P_{ij} that it will next be in state j. That is, we suppose that

$$P\{X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \ldots, X_1 = i_1, X_0 = i_0\} = P_{ij} \tag{4.1}$$

for all states $i_0, i_1, \ldots, i_{n-1}, i, j$ and all $n \geqslant 0$. Such a stochastic process is known as a *Markov chain*. Equation (4.1) may be interpreted as stating that, for a Markov chain, the conditional distribution of any future state X_{n+1} given the past states $X_0, X_1, \ldots, X_{n-1}$ and the present state X_n, is independent of the past states and depends only on the present state.

The value P_{ij} represents the probability that the process will, when in state i, next make a transition into state j. Since probabilities are nonnegative and since the process must make a transition into some state, we have that

$$P_{ij} \geqslant 0, \qquad i, j \geqslant 0; \qquad \sum_{j=0}^{\infty} P_{ij} = 1, \qquad i = 0, 1, \ldots.$$

Let $\mathbf{P}$ denote the matrix of one-step transition probabilities P_{ij}, so that

$$\mathbf{P} = \begin{Vmatrix} P_{00} & P_{01} & P_{02} & \cdots \\ P_{10} & P_{11} & P_{12} & \cdots \\ \vdots & & & \\ P_{i0} & P_{i1} & P_{i2} & \cdots \\ \vdots & \vdots & \vdots & \end{Vmatrix}$$

C7053

Example 4.1 (Forecasting the Weather): Suppose that the chance of rain tomorrow depends on previous weather conditions only through whether or not it is raining today and not on past weather conditions. Suppose also that if it rains today, then it will rain tomorrow with probability α; and if it does not rain today, then it will rain tomorrow with probability β.

If we say that the process is in state 0 when it rains and state 1 when it does not rain, then the above is a two-state Markov chain whose transition probabilities are given by

$$\mathbf{P} = \begin{Vmatrix} \alpha & 1-\alpha \\ \beta & 1-\beta \end{Vmatrix} \quad ✦$$

Example 4.2 (A Communications System): Consider a communications system which transmits the digits 0 and 1. Each digit transmitted must pass through several stages, at each of which there is a probability p that the digit entered will be unchanged when it leaves. Letting X_n denote the digit entering the nth stage, then $\{X_n, n = 0, 1, \ldots\}$ is a two-state Markov chain having a transition probability matrix

$$\mathbf{P} = \begin{Vmatrix} p & 1-p \\ 1-p & p \end{Vmatrix} \quad ✦$$

Example 4.3 On any given day Gary is either cheerful (C), so-so (S), or glum (G). If he is cheerful today, then he will be C, S, or G tomorrow with respective probabilities 0.5, 0.4, 0.1. If he is feeling so-so today, then he will be C, S, or G tomorrow with probabilities 0.3, 0.4, 0.3. If he is glum today, then he will be C, S, or G tomorrow with probabilities 0.2, 0.3, 0.5.

Letting X_n denote Gary's mood on the nth day, then $\{X_n, n \geqslant 0\}$ is a three-state Markov chain (state $0 = C$, state $1 = S$, state $2 = G$) with transition probability matrix

$$\mathbf{P} = \begin{Vmatrix} 0.5 & 0.4 & 0.1 \\ 0.3 & 0.4 & 0.3 \\ 0.2 & 0.3 & 0.5 \end{Vmatrix} \quad ✦$$

Example 4.4 (Transforming a Process into a Markov Chain): Suppose that whether or not it rains today depends on previous weather conditions through the last two days. Specifically, suppose that if it has rained for the past two days, then it will rain tomorrow with probability 0.7; if it rained today but not yesterday, then it will rain tomorrow with probability 0.5; if it rained yesterday but not today, then it will rain tomorrow with probability 0.4; if it has not rained in the past two days, then it will rain tomorrow with probability 0.2.

If we let the state at time n depend only on whether or not it is raining at time n, then the above model is not a Markov chain (why not?). However, we can transform the above model into a Markov chain by saying that the state at any time is determined by the weather conditions during both that day and the previous day. In other words, we can say that the process is in

state 0 if it rained both today and yesterday,
state 1 if it rained today but not yesterday,
state 2 if it rained yesterday but not today,
state 3 if it did not rain either yesterday or today.

The preceding would then represent a four-state Markov chain having a transition probability matrix

$$\mathbf{P} = \begin{Vmatrix} 0.7 & 0 & 0.3 & 0 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 0.4 & 0 & 0.6 \\ 0 & 0.2 & 0 & 0.8 \end{Vmatrix}$$

The reader should carefully check the matrix **P**, and make sure he or she understands how it was obtained. ✦

Example 4.5 (A Random Walk Model): A Markov chain whose state space is given by the integers $i = 0, \pm1, \pm2, \ldots$ is said to be a random walk if, for some number $0 < p < 1$,

$$P_{i,i+1} = p = 1 - P_{i,i-1}, \qquad i = 0, \pm1, \ldots$$

The preceding Markov chain is called a *random walk* for we may think of it as being a model for an individual walking on a straight line who at each point of time either takes one step to the right with probability p or one step to the left with probability $1 - p$. ✦

Example 4.6 (A Gambling Model): Consider a gambler who, at each play of the game, either wins \$1 with probability p or loses \$1 with probability $1 - p$. If we suppose that our gambler quits playing either when he goes broke or he attains a fortune of \$$N$, then the gambler's fortune is a Markov chain having transition probabilities

$$P_{i,i+1} = p = 1 - P_{i,i-1}, \qquad i = 1, 2, \ldots, N-1$$
$$P_{00} = P_{NN} = 1$$

States 0 and N are called *absorbing* states since once entered they are never left. Note that the above is a finite state random walk with absorbing barriers (states 0 and N). ✦

4.2. Chapman–Kolmogorov Equations

We have already defined the one-step transition probabilities P_{ij}. We now define the n-step transition probabilities P^n_{ij} to be the probability that a process in state i will be in state j after n additional transitions. That is,

$$P^n_{ij} = P\{X_{n+m} = j \mid X_m = i\}, \qquad n \geqslant 0, i, j \geqslant 0$$

Of course $P^1_{ij} = P_{ij}$. The *Chapman–Kolmogorov equations* provide a method for computing these n-step transition probabilities. These equations are

$$P^{n+m}_{ij} = \sum_{k=0}^{\infty} P^n_{ik} P^m_{kj} \qquad \text{for all } n, m \geqslant 0, \text{ all } i, j \tag{4.2}$$

and are most easily understood by noting that $P^n_{ik} P^m_{kj}$ represents the probability that starting in i the process will go to state j in $n + m$ transitions through a path which takes it into state k at the nth transition. Hence, summing over all intermediate states k yields the probability that the process will be in state j after $n + m$ transitions. Formally, we have

$$\begin{aligned} P^{n+m}_{ij} &= P\{X_{n+m} = j \mid X_0 = i\} \\ &= \sum_{k=0}^{\infty} P\{X_{n+m} = j, X_n = k \mid X_0 = i\} \\ &= \sum_{k=0}^{\infty} P\{X_{n+m} = j \mid X_n = k, X_0 = i\} P\{X_n = k \mid X_0 = i\} \\ &= \sum_{k=0}^{\infty} P^m_{kj} P^n_{ik} \end{aligned}$$

If we let $\mathbf{P}^{(n)}$ denote the matrix of n-step transition probabilities P^n_{ij}, then Equation (4.2) asserts that

$$\mathbf{P}^{(n+m)} = \mathbf{P}^{(n)} \cdot \mathbf{P}^{(m)}$$

where the dot represents matrix multiplication.* Hence, in particular,

$$\mathbf{P}^{(2)} = \mathbf{P}^{(1+1)} = \mathbf{P} \cdot \mathbf{P} = \mathbf{P}^2$$

and by induction

$$\mathbf{P}^{(n)} = \mathbf{P}^{(n-1+1)} = \mathbf{P}^{n-1} \cdot \mathbf{P} = \mathbf{P}^n$$

That is, the n-step transition matrix may be obtained by multiplying the matrix **P** by itself n times.

* If **A** is an $N \times M$ matrix whose element in the ith row and jth column is a_{ij} and **B** is a $M \times K$ matrix whose element in the ith row and jth column is b_{ij}, then $\mathbf{A} \cdot \mathbf{B}$ is defined to be the $N \times K$ matrix whose element in the ith row and jth column is $\sum_{k=1}^{M} a_{ik} b_{kj}$.

Example 4.7 Consider Example 4.1 in which the weather is considered as a two-state Markov chain. If $\alpha = 0.7$ and $\beta = 0.4$, then calculate the probability that it will rain four days from today given that it is raining today.

Solution: The one-step transition probability matrix is given by

$$\mathbf{P} = \begin{Vmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{Vmatrix}$$

Hence,

$$\mathbf{P}^{(2)} = \mathbf{P}^2 = \begin{Vmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{Vmatrix} \cdot \begin{Vmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{Vmatrix}$$
$$= \begin{Vmatrix} 0.61 & 0.39 \\ 0.52 & 0.48 \end{Vmatrix},$$
$$\mathbf{P}^{(4)} = (\mathbf{P}^2)^2 = \begin{Vmatrix} 0.61 & 0.39 \\ 0.52 & 0.48 \end{Vmatrix} \cdot \begin{Vmatrix} 0.61 & 0.39 \\ 0.52 & 0.48 \end{Vmatrix}$$
$$= \begin{Vmatrix} 0.5749 & 0.4251 \\ 0.5668 & 0.4332 \end{Vmatrix}$$

and the desired probability P^4_{00} equals 0.5749. ✦

Example 4.8 Consider Example 4.4. Given that it rained on Monday and Tuesday, what is the probability that it will rain on Thursday?

Solution: The two-step transition matrix is given by

$$\mathbf{P}^{(2)} = \mathbf{P}^2 = \begin{Vmatrix} 0.7 & 0 & 0.3 & 0 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 0.4 & 0 & 0.6 \\ 0 & 0.2 & 0 & 0.8 \end{Vmatrix} \cdot \begin{Vmatrix} 0.7 & 0 & 0.3 & 0 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 0.4 & 0 & 0.6 \\ 0 & 0.2 & 0 & 0.8 \end{Vmatrix}$$
$$= \begin{Vmatrix} 0.49 & 0.12 & 0.21 & 0.18 \\ 0.35 & 0.20 & 0.15 & 0.30 \\ 0.20 & 0.12 & 0.20 & 0.48 \\ 0.10 & 0.16 & 0.10 & 0.64 \end{Vmatrix}$$

Since rain on Thursday is equivalent to the process being in either state 0 or state 1 on Thursday, the desired probability is given by $P^2_{00} + P^2_{01} = 0.49 + 0.12 = 0.61$. ✦

So far, all of the probabilities we have considered are conditional probabilities. For instance, P^n_{ij} is the probability that the state at time n is j *given* that the initial state at time 0 is i. If the unconditional distribution of the state at time n is desired, it is necessary to specify the probability distribution of the initial state. Let us denote this by

$$\alpha_i \equiv P\{X_0 = i\}, \qquad i \geqslant 0 \left(\sum_{i=0}^{\infty} \alpha_i = 1 \right)$$

All unconditional probabilities may be computed by conditioning on the initial state. That is,

$$P\{X_n = j\} = \sum_{i=0}^{\infty} P\{X_n = j \mid X_0 = i\} P\{X_0 = i\}$$
$$= \sum_{i=0}^{\infty} P^n_{ij} \alpha_i$$

For instance, if $\alpha_0 = 0.4$, $\alpha_1 = 0.6$, in Example 4.7, then the (unconditional) probability that it will rain four days after we begin keeping weather records is

$$P\{X_4 = 0\} = 0.4P^4_{00} + 0.6P^4_{10}$$
$$= (0.4)(0.5749) + (0.6)(0.5668)$$
$$= 0.5700$$

4.3. Classification of States

State j is said to be *accessible* from state i if $P^n_{ij} > 0$ for some $n \geqslant 0$. Note that this implies that state j is accessible from state i if and only if, starting in i, it is possible that the process will ever enter state j. This is true since if j is not accessible from i, then

$$P\{\text{ever enter } j \mid \text{start in } i\} = P\left\{ \bigcup_{n=0}^{\infty} \{X_n = j\} \mid X_0 = i \right\}$$
$$\leqslant \sum_{n=0}^{\infty} P\{X_n = j \mid X_0 = i\}$$
$$= \sum_{n=0}^{\infty} P^n_{ij}$$
$$= 0$$

Two states i and j that are accessible to each other are said to *communicate*, and we write $i \leftrightarrow j$.

Note that any state communicates with itself since, by definition,

$$P_{ii}^0 = P\{X_0 = i \mid X_0 = i\} = 1$$

The relation of communication satisfies the following three properties:

(i) State i communicates with state i, all $i \geqslant 0$.
(ii) If state i communicates with state j, then state j communicates with state i.
(iii) If state i communicates with state j, and state j communicates with state k, then state i communicates with state k.

Properties (i) and (ii) follow immediately from the definition of communication. To prove (iii) suppose that i communicates with j, and j communicates with k. Thus, there exist integers n and m such that $P_{ij}^n > 0$, $P_{jk}^m > 0$. Now by the Chapman–Kolmogorov equations, we have that

$$P_{ik}^{n+m} = \sum_{r=0}^{\infty} P_{ir}^n P_{rk}^m \geqslant P_{ij}^n P_{jk}^m > 0$$

Hence, state k is accessible from state i. Similarly, we can show that state i is accessible from state k. Hence, states i and k communicate.

Two states that communicate are said to be in the same *class*. It is an easy consequence of (i), (ii), and (iii) that any two classes of states are either identical or disjoint. In other words, the concept of communication divides the state space up into a number of separate classes. The Markov chain is said to be *irreducible* if there is only one class, that is, if all states communicate with each other.

Example 4.9 Consider the Markov chain consisting of the three states 0, 1, 2 and having transition probability matrix

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

It is easy to verify that this Markov chain is irreducible. For example, it is possible to go from state 0 to state 2 since

$$0 \to 1 \to 2$$

That is, one way of getting from state 0 to state 2 is to go from state 0 to state 1 (with probability $\frac{1}{2}$) and then go from state 1 to state 2 (with probability $\frac{1}{4}$). ✦

Example 4.10 Consider a Markov chain consisting of the four states 0, 1, 2, 3 and having transition probability matrix

$$\mathbf{P} = \begin{Vmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ 0 & 0 & 0 & 1 \end{Vmatrix}$$

The class of this Markov chain are $\{0, 1\}$, $\{2\}$, and $\{3\}$. Note that while state 0 (or 1) is accessible from state 2, the reverse is not true. Since state 3 is an absorbing state, that is, $P_{33} = 1$, no other state is accessible from it. ✦

For any state i we let f_i denote the probability that, starting in state i, the process will ever reenter state i. State i is said to be *recurrent* if $f_i = 1$ and *transient* if $f_i < 1$.

Suppose that the process starts in state i and i is recurrent. Hence, with probability 1, the process will eventually reenter state i. However, by the definition of a Markov chain, it follows that the process will be starting over again when it reenters state i and, therefore, state i will eventually be visited again. Continual repetition of this argument leads to the conclusion that *if state i is recurrent then, starting in state i, the process will reenter state i again and again and again—in fact, infinitely often.*

mp relies on this

On the other hand, suppose that state i is transient. Hence, each time the process enters state i there will be a positive probability, namely, $1 - f_i$, that it will never again enter that state. Therefore, starting in state i, the probability that the process will be in state i for exactly n time periods equals $f_i^{n-1}(1 - f_i)$, $n \geqslant 1$. In other words, *if state i is transient then, starting in state i, the number of time periods that the process will be in state i has a geometric distribution with finite mean* $1/(1 - f_i)$.

From the preceding two paragraphs, it follows that *state i is recurrent if and only if, starting in state i, the expected number of time periods that the process is in state i is infinite.* But, letting

$$I_n = \begin{cases} 1, & \text{if } X_n = i \\ 0, & \text{if } X_n \neq i \end{cases}$$

we have that $\sum_{n=0}^{\infty} I_n$ represents the number of periods that the process is in state i. Also,

$$\begin{aligned} E\left[\sum_{n=0}^{\infty} I_n \,\middle|\, X_0 = i\right] &= \sum_{n=0}^{\infty} E[I_n \mid X_0 = i] \\ &= \sum_{n=0}^{\infty} P\{X_n = i \mid X_0 = i\} \\ &= \sum_{n=0}^{\infty} P_{ii}^n \end{aligned}$$

We have thus proven the following.

Proposition 4.1 State i is

$$\text{recurrent if } \sum_{n=1}^{\infty} P_{ii}^n = \infty,$$

$$\text{transient if } \sum_{n=1}^{\infty} P_{ii}^n < \infty$$

The argument leading to the preceding proposition is doubly important because it also shows that a transient state will only be visited a finite number of times (hence the name transient). This leads to the conclusion that in a finite-state Markov chain not all states can be transient. To see this, suppose the states are $0, 1, \ldots, M$ and suppose that they are all transient. Then after a finite amount of time (say, after time T_0) state 0 will never be visited, and after a time (say, T_1) state 1 will never be visited, and after a time (say, T_2) state 2 will never be visited, etc. Thus, after a finite time $T = \max\{T_0, T_1, \ldots, T_M\}$ no states will be visited. But as the process must be in some state after time T we arrive at a contradiction, which shows that at least one of the states must be recurrent.

Another use of Proposition 4.1 is that it enables us to show that recurrence is a class property.

Corollary 4.2 If state i is recurrent, and state i communicates with state j, then state j is recurrent.

Proof To prove this we first note that, since state i communicates with state j, there exist integers k and m such that $P_{ij}^k > 0$, $P_{ji}^m > 0$. Now, for any integer n

$$P_{jj}^{m+n+k} \geqslant P_{ji}^m P_{ii}^n P_{ij}^k$$

This follows since the left side of the above is the probability of going from j to j in $m+n+k$ steps, while the right side is the probability of going from j to j in $m+n+k$ steps via a path that goes from j to i in m steps, then from i to i in an additional n steps, then from i to j in an additional k steps.

From the preceding we obtain, by summing over n, that

$$\sum_{n=1}^{\infty} P_{jj}^{m+n+k} \geqslant P_{ji}^m P_{ij}^k \sum_{n=1}^{\infty} P_{ii}^n = \infty$$

since $P_{ji}^m P_{ij}^k > 0$, and $\sum_{n=1}^{\infty} P_{ii}^n$ is infinite since state i is recurrent. Thus, by Proposition 4.1 it follows that state j is also recurrent. ✦

Remarks (i) Corollary 4.2 also implies that transience is a class property. For if state i is transient and communicates with state j, then state j must also be transient. For if j were recurrent then, by Corollary 4.2, i would also be recurrent and hence could not be transient.

(ii) Corollary 4.2 along with our previous result that not all states in a finite Markov chain can be transient leads to the conclusion that all states of a finite irreducible Markov chain are recurrent.

Example 4.11 Let the Markov chain consisting of the states 0, 1, 2, 3 have the transition probability matrix

$$\mathbf{P} = \begin{Vmatrix} 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{Vmatrix}$$

Determine which states are transient and which are recurrent.

Solution: It is a simple matter to check that all states communicate and hence, since this is a finite chain, all states must be recurrent. ✦

Example 4.12 Consider the Markov chain having states 0, 1, 2, 3, 4 and

$$\mathbf{P} = \begin{Vmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{4} & \frac{1}{4} & 0 & 0 & \frac{1}{2} \end{Vmatrix}$$

Determine the recurrent state.

Solution: This chain consists of the three classes $\{0, 1\}$, $\{2, 3\}$, and $\{4\}$. The first two classes are recurrent and the third transient. ✦

Example 4.13 (A Random Walk): Consider a Markov chain whose state space consists of the integers $i = 0, \pm 1, \pm 2, \ldots$, and have transition probabilities given by

$$P_{i,i+1} = p = 1 - P_{i,i-1}, \qquad i = 0, \pm 1, \pm 2, \ldots$$

where $0 < p < 1$. In other words, on each transition the process either moves one step to the right (with probability p) or one step to the left (with

Substitution of the preceding into Equation (4.6) yields

$$E\left[\sum_{k=0}^{\infty} I_k\right] \leqslant \sum_{k=0}^{\infty} \frac{a_0 kp(1-p)^{k-1}}{1 - a_0[1 - kp(1-p)^{k-1}] - a_1(1-p)^k}$$
$$< \infty$$

where the convergence follows by noting that when k is large the denominator of the expression in the preceding sum converges to $1 - a_0$ and so the convergence or divergence of the sum is determined by whether or not the sum of the terms in the numerator converge and $\sum_{k=0}^{\infty} k(1-p)^{k-1} < \infty$.

Hence, $E[\sum_{k=0}^{\infty} I_k] < \infty$, which implies that $\sum_{k=0}^{\infty} I_k < \infty$ with probability 1 (for if there was a positive probability that $\sum_{k=0}^{\infty} I_k$ could be ∞, then its mean would be ∞). Hence, with probability 1, there will be only a finite number of states that are initially departed via a successful transmission; or equivalently, there will be some finite integer N such that whenever there are N or more messages in the system, there will never again be a successful transmission. From this (and the fact that such higher states will eventually be reached—why?) it follows that, with probability 1, there will only be a finite number of successful transmissions. ✦

Remark For a (slightly less than rigorous) probabilistic proof of Stirling's approximation, let $X_1 X_2, \ldots$ be independent Poisson random variables each having mean 1. Let $S_n = \sum_{i=1}^{n} X_i$, and note that both the mean and variance of S_n are equal to n. Now,

$$\begin{aligned} P\{S_n = n\} &= P\{n - 1 < S_n \leqslant n\} \\ &= P\{-1/\sqrt{n} < (S_n - n)/\sqrt{n} \leqslant 0 \\ &\approx \int_{-1/\sqrt{n}}^{0} (2\pi)^{-1/2} e^{-x^2/2}\, dx \qquad \text{when } n \text{ is large, by the central limit theorem} \\ &\approx (2\pi)^{-1/2}(1/\sqrt{n}) \\ &= (2\pi n)^{-1/2} \end{aligned}$$

But S_n is Poisson with mean n, and so

$$P\{S_n = n\} = \frac{e^{-n} n^n}{n!}$$

Hence, for n large

$$\frac{e^{-n} n^n}{n!} \approx (2\pi n)^{-1/2}$$

or, equivalently

$$n! \approx n^{n+1/2} e^{-n} \sqrt{2\pi}$$

which is Stirling's approximation.

4.4. Limiting Probabilities

In Example 4.7, we calculated $\mathbf{P}^{(4)}$ for a two-state Markov chain; it turned out to be

$$\mathbf{P}^{(4)} = \begin{Vmatrix} 0.5749 & 0.4251 \\ 0.5668 & 0.4332 \end{Vmatrix}$$

From this it follows that $\mathbf{P}^{(8)} = \mathbf{P}^{(4)} \cdot \mathbf{P}^{(4)}$ is given (to three significant places) by

$$\mathbf{P}^{(8)} = \begin{Vmatrix} 0.572 & 0.428 \\ 0.570 & 0.430 \end{Vmatrix}$$

Note that the matrix $\mathbf{P}^{(8)}$ is almost identical to the matrix $\mathbf{P}^{(4)}$, and secondly, that each of the rows of $\mathbf{P}^{(8)}$ has almost identical entries. In fact it seems that P_{ij}^n is converging to some value (as $n \to \infty$) which is the same for all i. In other words, there seems to exist a limiting probability that the process will be in state j after a large number of transitions, and this value is independent of the initial state.

To make the above heuristics more precise, two additional properties of the states of a Markov chain need to be considered. State i is said to have *period* d if $P_{ii}^n = 0$ whenever n is not divisible by d, and d is the largest integer with this property. For instance, starting in i, it may be possible for the process to enter state i only at the times $2, 4, 6, 8, \ldots$, in which case state i has period 2. A state with period 1 is said to be *aperiodic*. It can be shown that periodicity is a class property. That is, if state i has period d, and states i and j communicate, then state j also has period d.

If state i is recurrent, then it is said to be *positive recurrent* if, starting in i, the expected time until the process returns to state i is finite. It can be shown that positive recurrence is a class property. While there exist recurrent states that are not positive recurrent,* it can be shown that *in a finite-state Markov chain all recurrent states are positive recurrent*. Positive recurrent, aperiodic states are called *ergodic*.

We are now ready for the following important theorem which we state without proof.

* Such states are called *null recurrent*.

Theorem 4.1 For an irreducible ergodic Markov chain $\lim_{n\to\infty} P_{ij}^n$ exists and is independent of i. Furthermore, letting

$$\pi_j = \lim_{n\to\infty} P_{ij}^n, \qquad j \geqslant 0$$

then π_j is the unique nonnegative solution of

$$\begin{aligned} \pi_j &= \sum_{i=0}^{\infty} \pi_i P_{ij}, \qquad j \geqslant 0 \\ \sum_{j=0}^{\infty} \pi_j &= 1 \end{aligned} \tag{4.7}$$

Remarks (i) Given that $\pi_j = \lim_{n\to\infty} P_{ij}^n$ exists and is independent of the initial state i, it is not difficult to (heuristically) see that the πs must satisfy Equation (4.7). Let us derive an expression for $P\{X_{n+1} = j\}$ by conditioning on the state at time n. That is,

$$\begin{aligned} P\{X_{n+1} = j\} &= \sum_{i=0}^{\infty} P\{X_{n+1} = j \mid X_n = i\} P\{X_n = i\} \\ &= \sum_{i=0}^{\infty} P_{ij} P\{X_n = i\} \end{aligned}$$

Letting $n \to \infty$, and assuming that we can bring the limit inside the summation, leads to

$$\pi_j = \sum_{i=0}^{\infty} P_{ij} \pi_i$$

(ii) It can be shown that π_j, the limiting probability that the process will be in state j at time n, also equals the long-run proportion of time that the process will be in state j.

(iii) In the irreducible, positive recurrent, *periodic* case we still have that the π_j, $j \geqslant 0$, are the unique nonnegative solution of

$$\pi_j = \sum_i \pi_i P_{ij},$$

$$\sum_j \pi_j = 1$$

But now π_j must be interpreted as the long-run proportion of time that the Markov chain is in state j.

Example 4.15 Consider Example 4.1, in which we assume that if it rains today, then it will rain tomorrow with probability α; and if it does not rain

today, then it will rain tomorrow with probability β. If we say that the state is 0 when it rains and 1 when it does not rain, then by Equation (4.7) the limiting probabilities π_0 and π_1 are given by

$$\begin{aligned} \pi_0 &= \alpha\pi_0 + \beta\pi_1, \\ \pi_1 &= (1-\alpha)\pi_0 + (1-\beta)\pi_1, \\ \pi_0 + \pi_1 &= 1 \end{aligned}$$

which yields that

$$\pi_0 = \frac{\beta}{1+\beta-\alpha}, \qquad \pi_1 = \frac{1-\alpha}{1+\beta-\alpha}$$

For example if $\alpha = 0.7$ and $\beta = 0.4$, then the limiting probability of rain is $\pi_0 = \frac{4}{7} = 0.571$. ✦

Example 4.16 Consider Example 4.3 in which the mood of an individual is considered as a three-state Markov chain having a transition probability matrix

$$\mathbf{P} = \begin{Vmatrix} 0.5 & 0.4 & 0.1 \\ 0.3 & 0.4 & 0.3 \\ 0.2 & 0.3 & 0.5 \end{Vmatrix}$$

In the long run, what proportion of time is the process in each of the three states?

Solution: The limiting probabilities π_i, $i = 0, 1, 2$, are obtained by solving the set of equations in Equation (4.1). In this case these equations are

$$\begin{aligned} \pi_0 &= 0.5\pi_0 + 0.3\pi_1 + 0.2\pi_2, \\ \pi_1 &= 0.4\pi_0 + 0.4\pi_1 + 0.3\pi_2, \\ \pi_2 &= 0.1\pi_0 + 0.3\pi_1 + 0.5\pi_2, \\ \pi_0 + \pi_1 + \pi_2 &= 1 \end{aligned}$$

Solving yields

$$\pi_0 = \tfrac{21}{62}, \qquad \pi_1 = \tfrac{23}{62}, \qquad \pi_2 = \tfrac{18}{62} \quad ✦$$

Example 4.17 (A Model of Class Mobility): A problem of interest to sociologists is to determine the proportion of society that has an upper- or lower-class occupation. One possible mathematical model would be to assume that transitions between social classes of the successive generations in a family can be regarded as transitions of a Markov chain. That is, we assume that the occupation of a child depends only on his or her parent's occupation. Let us suppose that such a model is appropriate and that the transition probability matrix is given by

$$\mathbf{P} = \begin{Vmatrix} 0.45 & 0.48 & 0.07 \\ 0.05 & 0.70 & 0.25 \\ 0.01 & 0.50 & 0.49 \end{Vmatrix} \tag{4.8}$$

That is, for instance, we suppose that the child of a middle-class worker will attain an upper-, middle, or lower-class occupation with respective probabilities 0.05, 0.70, 0.25.

The limiting probabilities π_i, thus satisfy

$$\begin{aligned}\pi_0 &= 0.45\pi_0 + 0.05\pi_1 + 0.01\pi_2,\\ \pi_1 &= 0.48\pi_0 + 0.70\pi_1 + 0.50\pi_2,\\ \pi_2 &= 0.07\pi_0 + 0.25\pi_1 + 0.49\pi_2,\\ \pi_0 + \pi_1 + \pi_2 &= 1\end{aligned}$$

Hence,

$$\pi_0 = 0.07, \qquad \pi_1 = 0.62, \qquad \pi_2 = 0.31$$

In other words, a society in which social mobility between classes can be described by a Markov chain with transition probability matrix given by Equation (4.8) has, in the long run, 7 percent of its people in upper-class jobs, 62 percent of its people in middle-class jobs, and 31 percent in lower-class jobs. ✦

Example 4.18 (The Hardy–Weinberg Law and a Markov Chain in Genetics): Consider a large population of individuals each of whom possesses a particular pair of genes, of which each individual gene is classified as being of type A or type a. Assume that the proportions of individuals whose gene pairs are AA, aa, or Aa are respectively p_0, q_0, and r_0 ($p_0 + q_0 + r_0 = 1$). When two individuals mate, each contributes one of his or her genes, chosen at random, to the resultant offspring. Assuming that the mating occurs at random, in that each individual is equally likely to mate with any other individual, we are interested in determining the proportions of individuals in the next generation whose genes are AA, aa, or Aa. Calling these proportions p, q, and r, they are easily obtained by focusing attention on an individual of the next generation and then determining the probabilities for the gene pair of that individual.

To begin, note that randomly choosing a parent and then randomly choosing one of its genes is equivalent to just randomly choosing a gene from the total gene population. By conditioning on the gene pair of the parent, we see that a randomly chosen gene will be type A with probability

$$\begin{aligned}P\{A\} &= P\{A\,|\,AA\}p_0 + P\{A\,|\,aa\}q_0 + P\{A\,|\,Aa\}r_0\\ &= p_0 + r_0/2\end{aligned}$$

Similarly, it will be type a with probability

$$P\{a\} = q_0 + r_0/2$$

Thus, under random mating a randomly chosen member of the next generation will be type AA with probability p, where

$$p = P\{A\}P\{A\} = (p_0 + r_0/2)^2$$

Similarly, the randomly chosen member will be type aa with probability

$$q = P\{a\}P\{a\} = (q_0 + r_0/2)^2$$

and will be type Aa with probability

$$r = 2P\{A\}P\{a\} = 2(p_0 + r_0/2)(q_0 + r_0/2)$$

Since each member of the next generation will independently be of each of the three gene types with probabilities p, q, r, it follows that the percentages of the members of the next generation that are of type AA, aa, or Aa are respectively p, q, and r.

If we now consider the total gene pool of this next generation, then $p+r/2$, the fraction of its genes that are A, will be unchanged from the previous generation. This follows either by arguing that the total gene pool has not changed from generation to generation or by the following simple algebra:

$$\begin{aligned}p + r/2 &= (p_0 + r_0/2)^2 + (p_0 + r_0/2)(q_0 + r_0/2)\\ &= (p_0 + r_0/2)[p_0 + r_0/2 + q_0 + r_0/2]\\ &= p_0 + r_0/2 \qquad \text{since } p_0 + r_0 + q_0 = 1\\ &= P\{A\}\end{aligned} \tag{4.9}$$

Thus, the fractions of the gene pool that are A and a are the same as in the initial generation. From this it follows that, under random mating, in all successive generations after the initial one the percentages of the population having gene pairs AA, aa, and Aa will remain fixed at the values p, q, and r. This is known as the *Hardy–Weinberg law*.

Suppose now that the gene pair population has stabilized in the percentages p, q, r, and let us follow the genetic history of a single individual and