4

Markov Processes

Chapter 2 gave a detailed treatment of (discrete-time) Markov chains. The key Markov property was that the future was independent of the past given the present. A Markov process is simply the continuous analogue to a chain, namely, the Markov property is maintained and time is measured continuously. For example, suppose we are keeping track of patients within a large hospital, and the state of our process is the specific department housing the patient. If we record the patient's location every day, the resulting process will be a Markov chain if the Markov property holds. If, however, we record the patient's location continuously throughout the day, then the resulting process will be called a Markov process if the Markov property holds. Thus, instead of transitions only occurring at times 1, 2, 3, etc., transitions may occur at times 2.54 or 7.01.

Our major reason for covering Markov processes is that they play a key role in queueing theory, and queueing theory has applications in many areas. (In particular, any system in which there may be a buildup of waiting lines is a candidate for a queueing model.) Therefore, in this chapter, we give a cursory treatment of Markov processes in preparation for the next chapter dealing with queues.

4.1 BASIC DEFINITIONS

The definition for a Markov process is similar to that for a Markov chain except the Markov property must be shown to hold for all future times instead of just for one step (since "one step" has no meaning in continuous time).

DEFINITION 4.1 *The process $Y = \{Y_t; t \geq 0\}$ with finite state space E is a* MARKOV PROCESS *if the following holds for all $j \in E$ and $t, s \geq 0$*

$$\Pr\{Y_{t+s} = j \mid Y_u; u \leq t\} = \Pr\{Y_{t+s} = j \mid Y_t\}.$$

Furthermore, the Markov process is said to have STATIONARY *transitions if*

$$\Pr\{Y_{t+s} = j \mid Y_t = i\} = \Pr\{Y_s = j \mid Y_0 = i\}.$$

These equations are completely analogous to those in Definition 2.3. Think of the present time as being time t. The left-hand side of the first equation in Definition 4.1 indicates that a prediction of the future s time units from now is desired given all the past history up to and including the current time t. The right-hand side of the equation indicates that the prediction is the same if the only information available is the state of the system at the present time. (Note that an implication of this definition is that it is a waste of resources to keep historical records of a process if it is Markovian.) The second equation of the definition is the time homogenous condition, which indicates that the probability law governing the process does not change during the life of the process.

EXAMPLE 4.1 A salesman lives in town a and is responsible for towns a, b, and c. The amount of time he spends in each town is random. After some study it has been determined that the amount of consecutive time spent in any one town follows an exponentially distributed random variable [see Eq. (1.15)] with the mean time depending on the town. In his hometown, a mean time of 2 weeks is spent, in town b a mean time of 1 week is spent, and in town c a mean time of 1.5 weeks is spent. When he leaves town a, he flips a coin to determine to which town he goes next; when he leaves either town b or c, he flips two coins so that there is a 75% chance of returning to a and a 25% chance of going to the other town. Let Y_t be a random variable denoting the town that the salesman is in at time t. The process $Y = \{Y_t; t \geq 0\}$ is a Markov process due to the lack of memory inherent in the exponential distribution (see Exercise 1.23).

The construction of the Markov process in Example 4.1 is instructive in that it contains the major characterizations for any finite state Markov process. A Markov process remains in each state for an exponential length of time and then when it jumps, it jumps according to a Markov chain (see Figure 4.1). To describe this characterization mathematically, we introduce some additional notation that should be familiar due to its similarity to the previous chapter. We first let $T_0 = 0$ and $X_0 = Y_0$. Then, we denote the successive jump times of the process by $T_1, T_2, \cdots$, and denote the state of the process immediately after the jumps as $X_1, X_2, \cdots$. To be more explicit, for $n = 0, 1, \cdots$,

$$T_{n+1} = \min\{t > T_n : Y_t \neq Y_{T_n}\}, \quad \text{and}$$
$$X_n = Y_{T_n}.$$

The time between jumps (namely, $T_{n+1} - T_n$) is called the *sojourn time*. The Markov property implies that the only information needed for predicting

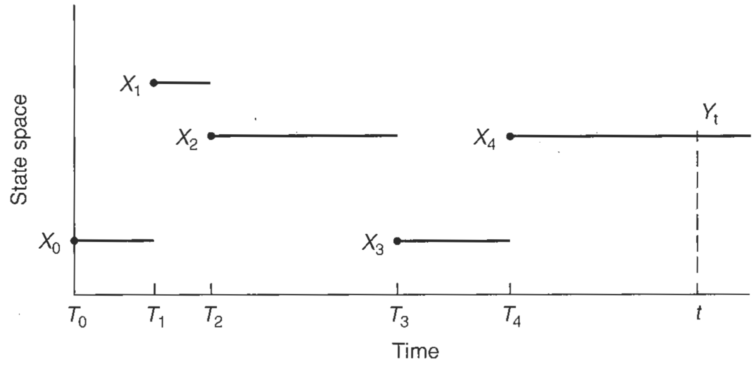


FIGURE 4.1 Typical realization for a Markov process.

the sojourn time is the current state. Or, to say it mathematically,

$$\Pr\{T_{n+1} - T_n \le t \mid X_0, \cdots, X_n, T_0, \cdots, T_n\} = \Pr\{T_{n+1} - T_n \le t \mid X_n\}.$$

The Markov property also implies that if we focus our attention only on the process $X_0, X_1, \cdots$, then we would see a Markov chain. This Markov chain is called the *imbedded Markov chain.*

DEFINITION 4.2 *Let $Y = \{Y_t; t \ge 0\}$ be a Markov process with finite state space E and with jump times denoted by $T_0, T_1, \cdots$, and the imbedded process at the jump times denoted by $X_0, X_1, \cdots$. Then there is a collection of scalars, $\lambda(i)$ for $i \in E$, called the* MEAN SOJOURN RATES, *and a Markov matrix, **P**, called the* IMBEDDED MARKOV MATRIX, *that satisfy the following:*

$$\Pr\{T_{n+1} - T_n \le t \mid X_n = i\} = 1 - e^{-\lambda(i)t},$$

$$\Pr\{X_{n+1} = j \mid X_n = i\} = P(i,j),$$

*where each $\lambda(i)$ is nonnegative and the diagonal elements of **P** are zero.*

EXAMPLE 4.2 Consider again Example 4.1. The imbedded Markov chain described by the above definition is

$$\boldsymbol{P} = \begin{matrix} a \\ b \\ c \end{matrix} \begin{bmatrix} 0 & 0.50 & 0.50 \\ 0.75 & 0 & 0.25 \\ 0.75 & 0.25 & 0 \end{bmatrix},$$

and the mean sojourn rates are $\lambda(a) = \frac{1}{2}$, $\lambda(b) = 1$, and $\lambda(c) = \frac{2}{3}$. (Note that mean rates are always the reciprocal of mean times.) ■

To avoid pathological cases, we shall adopt the policy in this chapter that the mean time spent in each state be finite and nonzero. The notions of state classification can now be carried over from Chapter 2.

DEFINITION 4.3 *Let $Y = \{Y_t; t \geq 0\}$ be a Markov process with finite state space E such that $0 < \lambda(i) < \infty$ for all $i \in E$. A state is called* RECURRENT *or* TRANSIENT *according to whether it is recurrent or transient in the imbedded Markov chain of the process. A set of states is* IRREDUCIBLE *for the process if it is irreducible for the imbedded chain.*

4.2 STEADY-STATE PROPERTIES

There is a close relationship between the steady-state probabilities for the Markov process and the steady-state probabilities of the imbedded chain. Specifically, the Markov process probabilities are obtained by the weighted average of the imbedded Markov chain probabilities according to each state's sojourn times. To show the relationship mathematically, consider $Y = \{Y_t; t \geq 0\}$ to be a Markov process with an irreducible, recurrent state space. Let $X = \{X_n; n = 0, 1, \cdots\}$ be the imbedded Markov chain, with $\boldsymbol{\pi}$ its steady-state probabilities; namely,

$$\pi(j) = \lim_{n \to \infty} \Pr\{X_n = j \mid X_0 = i\}.$$

(In other words, $\boldsymbol{\pi P} = \boldsymbol{\pi}$, where $\boldsymbol{P}$ is the Markov matrix of the imbedded chain.) The steady-state probabilities for the Markov process are denoted by the vector $\boldsymbol{p}$ where

$$p(j) = \lim_{t \to \infty} \Pr\{Y_t = j \mid Y_0 = i\}. \tag{4.1}$$

Note that the limiting values are independent of the initial state since the state space is assumed to be irreducible and recurrent. The relationship between the vectors $\boldsymbol{\pi}$ and $\boldsymbol{p}$ is given by

$$p(j) = \frac{\pi(j)/\lambda(j)}{\sum_{k \in E} \pi(k)/\lambda(k)} \tag{4.2}$$

where E is the state space and $\lambda(j)$ is the mean sojourn rate for state j.

The use of Eq. (4.2) is easily illustrated with Example 4.2 since the steady-state probabilities $\boldsymbol{\pi}$ were calculated in Chapter 2. Using these previous results (page 56), we have that the imbedded chain has steady-state probabilities given by $\boldsymbol{\pi} = (\frac{3}{7}, \frac{2}{7}, \frac{2}{7})$. Combining these with the mean sojourn rates in Example 4.2 and with Eq. (4.2) yields

$$p(a) = \frac{6}{11}, \quad p(b) = \frac{2}{11}, \quad \text{and} \quad p(c) = \frac{3}{11}.$$

The salesman of the example thus spends 54.55% of his time in town a, 18.18% of his time in town b, and 27.27% of his time in town c.

The information contained in the imbedded Markov matrix and the mean sojourn rates can be combined into one matrix, which is called the *generator matrix* of the Markov process.

DEFINITION 4.4 *Let $Y = \{Y_t; t \geq 0\}$ be a Markov process with an imbedded Markov matrix $\boldsymbol{P}$ and a mean sojourn rate of $\lambda(i)$ for each state i in the state space. The* GENERATOR MATRIX, $\boldsymbol{G}$, *for the Markov process is given by*

$$G(i,j) = \begin{cases} -\lambda(i) & \text{for } i = j \\ \lambda(i)P(i,j) & \text{for } i \neq j \end{cases}.$$

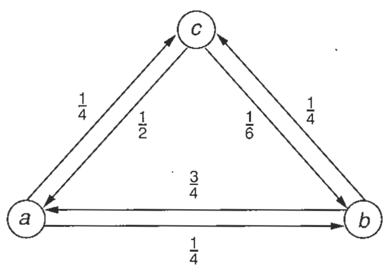
Generator matrices are extremely important in the application of Markov processes. A generator matrix for a Markov process has two properties: (1) Each row sum is zero and (2) the off-diagonal elements are nonnegative. These properties can be seen from Definition 4.4 by remembering that the imbedded Markov matrix is nonnegative with row sums of one and has zeros on the diagonals. The physical interpretation of $\boldsymbol{G}$ is that $G(i,j)$ is the *rate* at which the process goes from state i to state j.

EXAMPLE 4.3 We again return to our initial example to illustrate the generator matrix. Using the imbedded Markov chain and sojourn times given in Example 4.2, the generator matrix for that Markov process is

$$\boldsymbol{G} = \begin{matrix} a \\ b \\ c \end{matrix} \begin{bmatrix} -\frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{3}{4} & -1 & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{6} & -\frac{2}{3} \end{bmatrix}.$$

Thus, Definition 4.4 indicates that if the transition matrix for the imbedded Markov chain and the mean sojourn rates are known, the generator matrix can be computed. State diagrams for Markov processes are very similar to diagrams for Markov chains, except that the arrows representing transition probabilities represent transition rates (see Figure 4.2).

FIGURE 4.2 State diagram for the Markov process of Example 4.3.



It should also be clear that not only can the transition rate matrix be obtained from the imbedded Markov matrix and mean sojourn rates, but the reverse is also true. From the generator matrix, the absolute value of the diagonal elements gives the mean sojourn rates. The transition matrix for the imbedded chain is then obtained by dividing the off-diagonals element by the absolute value of that row's diagonal element. The diagonal elements of the transition matrix for the imbedded Markov chain are zero. For example, suppose a Markov process with state space $\{a, b, c\}$ has a generator matrix given by

$$\boldsymbol{G} = \begin{matrix} a \\ b \\ c \end{matrix} \begin{bmatrix} -2 & 2 & 0 \\ 2 & -4 & 2 \\ 1 & 4 & -5 \end{bmatrix}.$$

The mean sojourn rates are thus $\lambda_a = 2$, $\lambda_b = 4$, and $\lambda_c = 5$, and the transition matrix for the imbedded Markov chain is given by

$$\boldsymbol{P} = \begin{matrix} a \\ b \\ c \end{matrix} \begin{bmatrix} 0 & 1 & 0 \\ 0.5 & 0 & 0.5 \\ 0.2 & 0.8 & 0 \end{bmatrix}.$$

▶ *Suggestion: Do Exercises 4.1 and 4.2a.*

It is customary when describing a Markov process to obtain the generator matrix directly and never specifically look at the imbedded Markov matrix. The usual technique is to obtain the off-diagonal elements first and then let the diagonal element equal the negative of the sum of the off-diagonal elements on the row. The generator is used to obtain the steady-state probabilities directly as follows.

Property 4.5 Let $Y = \{Y_t; t \geq 0\}$ be a Markov process with an irreducible, recurrent state space E and with generator matrix $\boldsymbol{G}$. Furthermore, let $\boldsymbol{p}$ be a vector of steady-state probabilities defined by Eq. (4.1). Then $\boldsymbol{p}$ is the solution to

$$\boldsymbol{pG} = \boldsymbol{0}$$

$$\sum_{j \in E} p(j) = 1.$$

Obviously, this equation is very similar to the analogous equation in Property 2.16. The difference is in the right-hand side of the first equation. The easy way to remember which equation to use is that the multiplier for the right-hand side is the same as the row sums. Since Markov matrices have row sums of one, the right-hand side is *one* times $\boldsymbol{\pi}$; since the generator matrix has row sums of zero, the right-hand side is *zero* times $\boldsymbol{p}$.

EXAMPLE 4.4 A small garage operates a tractor repair service in a busy farming community. Because the garage is small, only two tractors can be kept there at any one time and the owner is the only repairman. If one tractor is being repaired and one is waiting, then any further customers who request service are turned away. The time between the arrival of customers is an exponential random variable and an average of four customers arrive each day. Repair time varies considerably and it is also an exponential random variable. Based on previous records, the owner repairs an average of five tractors per day if kept busy. The arrival and repair of tractors can be described as a Markov process with state space $\{0, 1, 2\}$ where a state represents the number of tractors in the garage. Instead of determining the Markov matrix for the imbedded chain, the generator matrix can be calculated directly. As in Markov chain problems, drawing a state diagram is helpful. State diagrams for Markov processes show transition rates instead of transition probabilities. Figure 4.3 shows the diagram with the transition rates. The rate of going from state 0 to state 1 and from state 1 to state 2 is four per day since that is the arrival rate of customers to the system. The rate of going from state 2 to state 1 and from state 1 to state 0 is five per day because that is the repair rate. Transition rates are only to adjacent states because both repairs and arrivals only occur one at a time. From the diagram, the generator matrix is obtained as

$$\boldsymbol{G} = \begin{matrix} 0 \\ 1 \\ 2 \end{matrix} \begin{bmatrix} -4 & 4 & 0 \\ 5 & -9 & 4 \\ 0 & 5 & -5 \end{bmatrix}.$$

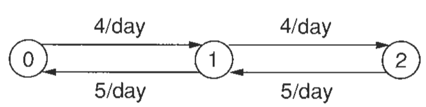
The steady-state probabilities are thus obtained by solving

$$\begin{aligned} -4p_0 + 5p_1 \qquad &= 0, \\ 4p_0 - 9p_1 + 5p_2 &= 0, \\ 4p_1 - 5p_2 &= 0, \\ p_0 + p_1 + p_2 &= 1, \end{aligned}$$

which yields (after deleting one of the first three equations) $p_0 = \frac{25}{61}$, $p_1 = \frac{20}{61}$, and $p_2 = \frac{16}{61}$. As with Markov chains, there will always be one redundant equation. The steady-state equations can be interpreted to indicate that the owner is idle 41% of the time and 26% of the time the shop is full.

An assumption was made in Example 4.4 regarding exponential random variables. The length of time that the process is in state 1 is a function of

FIGURE 4.3 State diagram for the Markov process of Example 4.4.



two exponentially distributed random variables. In effect, the sojourn time in state 1 depends on a "race" between the next arrival and the completion of the repair. The assumption is that the minimum of two exponential random variables is also exponentially distributed. Actually, that is not too difficult to prove, so we leave it as an exercise.

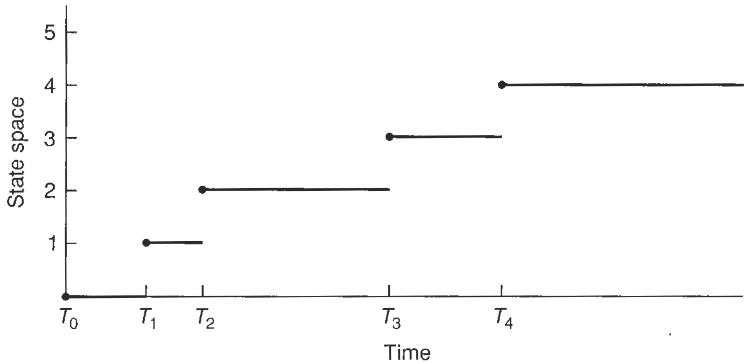
EXAMPLE 4.5 **Poisson Process.** We are interested in counting the number of vehicles that go through a tollbooth. The arriving vehicles form a Poisson process (page 18) with a mean rate of 120 cars per hour. Since a Poisson process has exponential interarrival times, a Poisson process is also a Markov process with a state space equal to the nonnegative integers. A possible realization of the process is given in Figure 4.4, and the generator matrix is infinite dimensioned given by

$$G = \begin{array}{c} 0 \\ 1 \\ 2 \\ \vdots \end{array} \begin{bmatrix} -120 & 120 & 0 & 0 & \cdots \\ 0 & -120 & 120 & 0 & \cdots \\ 0 & 0 & -120 & 120 & \cdots \\ \vdots & & & \ddots & \end{bmatrix}.$$

Assume that we are actually interested in the cumulative amount of tolls collected. The toll is a function of the type of vehicle going through the tollbooth. Fifty percent of the vehicles pay 25 cents, one-third of the vehicles pay 75 cents, and one-sixth of the vehicles pay $1. If we let the state space be the cumulative number of quarters collected, the process of interest is called a compound Poisson process. A typical realization is given in Figure 4.5 and the generator matrix is given by

$$G = \begin{array}{c} 0 \\ 1 \\ 2 \\ \vdots \end{array} \begin{bmatrix} -120 & 60 & 0 & 40 & 20 & 0 & 0 & \cdots \\ 0 & -120 & 60 & 0 & 40 & 20 & 0 & \cdots \\ 0 & 0 & -120 & 60 & 0 & 40 & 20 & \cdots \\ \vdots & & & \ddots & & & & \end{bmatrix}.$$

FIGURE 4.4 Typical realization for a Poisson process.



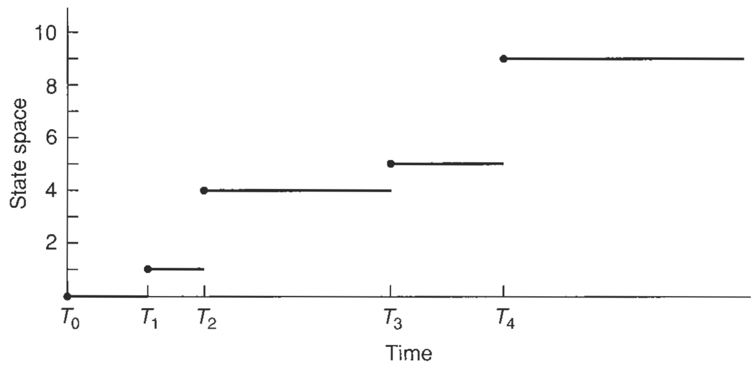


FIGURE 4.5 Typical realization for a compound Poisson process.

Notice that the only difference between the Poisson process and the compound Poisson process is that the Poisson process always increases by jump sizes of exactly one unit, whereas the compound Poisson process increases by an amount equal to a random variable. Furthermore, the sequence of random variables governing the size of the successive jumps is an independent, identically distributed sequence. ■

▶ *Suggestion: Do Exercises 4.2b, 4.3, 4.6, 4.7, 4.9, 4.13, and 4.14.*

4.3 REVENUES AND COSTS

Markov processes often represent revenue-producing situations, so the steady-state probabilities are used to determine a long-run average return. Suppose that whenever the Markov process is in state i, a profit is produced at a rate of $f(i)$. Thus, $\boldsymbol{f}$ is a vector of profit rates. The long-run profit per unit time is then simply the (vector) product of the steady-state probabilities times $\boldsymbol{f}$.

Property 4.6 Let $Y = \{Y_t; t \geq 0\}$ be a Markov process with an irreducible, recurrent state space E and a profit rate vector denoted by $\boldsymbol{f}$ [i.e., $f(i)$ is the rate that profit is accumulated whenever the process is in state i]. Furthermore, let $\boldsymbol{p}$ denote the steady-state probabilities as defined by Eq. (4.1). Then, the long-run profit per unit time is given by

$$\lim_{t\to\infty} \frac{1}{t} E\left[\int_0^t f(Y_s)\, ds\right] = \sum_{j\in E} p(j)f(j)$$

independent of the initial state.

▶ *Suggestion: Do Exercises 4.2c and 4.4a.*

The left-hand side of the preceding equation is simply the mathematical way of writing the long-run average accumulated profit. It is sometimes important to include revenue produced at transition times so that the total profit is the sum of the accumulation of profit resulting from the sojourn times and the jumps. In other words, we would not only have the profit rate vector $\boldsymbol{f}$ but also a matrix that indicates a profit obtained at each jump.

Property 4.7 Let $Y = \{Y_t; t \geq 0\}$ be a Markov process with an irreducible, recurrent state space E, a profit rate vector denoted by $\boldsymbol{f}$, and a matrix of jump profits denoted by $\boldsymbol{h}$ [i.e., $h(i, j)$ is the profit obtained whenever the process jumps from state i to state j]. Furthermore, let $\boldsymbol{p}$ denote the steady-state probabilities as defined by Eq. (4.1). Then, the long-run profit per unit time is given by

$$\lim_{t\to\infty} \frac{1}{t} E\left[\int_0^t f(Y_s)\, ds + \sum_{s\leq t} h(Y_{s-}, Y_s)\right] =$$

$$\sum_{i\in E} p(i)\left[f(i) + \sum_{k\in E} G(i, k)h(i, k)\right]$$

where $h(i, i) = 0$ and $\boldsymbol{G}$ is the generator matrix of the process.

The summation on the left-hand side of the preceding equation may appear different from what you are used to seeing. The notation Y_{s-} indicates the left-hand limit of the process; that is, $Y_{s-} = \lim_{t\to s^-} Y_t$ where t approaches s from the left, i.e., $t < s$. Since Markov processes are always right-continuous, the only times for which $Y_{s-} \neq Y_s$ is when s is a jump time. (When we say that the Markov process is right-continuous, we mean that $Y_s = \lim_{t\to s^+} Y_t$ where the limit is such that t approaches s from the right, i.e., $t > s$.) Because the diagonal elements of $\boldsymbol{h}$ are zero, the only times at which the summation includes nonzero terms are at jump times.

EXAMPLE 4.6 Returning to Example 4.1 for illustration, assume that the revenue possible from each town varies. Whenever the salesman is working town a, his profit comes in at a rate of \$80 per day. In town b, his profit is \$100 per day. In town c, his profit is \$125 per day. There is also a cost associated with changing towns. This cost is estimated at 25 cents per mile and it is 50 miles from a to b, 65 miles from a to c, and 80 miles from b to c. The functions (using 5-day weeks) are

$$\boldsymbol{f} = (400, 500, 625)^T,$$

$$\boldsymbol{h} = \begin{bmatrix} 0 & -12.50 & -16.25 \\ -12.50 & 0 & -20.00 \\ -16.25 & -20.00 & 0 \end{bmatrix}.$$

Therefore, using

$$G = \begin{bmatrix} -\frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{3}{4} & -1 & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{6} & -\frac{2}{3} \end{bmatrix}$$

and

$$p(a) = \frac{6}{11}, \quad p(b) = \frac{2}{11}, \quad p(c) = \frac{3}{11},$$

the long-run weekly average profit is

$$\frac{6}{11} \times (400 - 7.19) + \frac{2}{11} \times (500 - 14.38) + \frac{3}{11} \times (625 - 11.46) = 469.88.$$

▶ *Suggestion: Do Exercises 4.5 and 4.8.*

It is sometimes important to consider the time-value of the profit (or cost). In the discrete case, a discount factor of α was used. We often think in terms of a rate of return or an interest rate. If i were the rate of return, then the discount factor used for Markov chains is related to i according to the formula $\alpha = 1/(1+i)$; thus, the present value of one dollar obtained one period from the present is equal to α. For a continuous-time problem, we let β denote the discount rate, which in this case would be the same as the (nominal) rate of return except that we assume that compounding occurs continuously. Thus, the present value of one dollar obtained one period from the present equals $e^{-\beta}$. When the time-value of money is important, it is difficult to include the jump time profits but the function $\boldsymbol{f}$ is easily utilized according to the following property.

Property 4.8 Let $Y = \{Y_t; t \geq 0\}$ be a Markov process with a generator matrix $\boldsymbol{G}$, a profit rate vector $\boldsymbol{f}$, and a discount factor of β. Then, the present value of the total discounted profit (over an infinite planning horizon) is given by

$$E\left[\int_0^\infty e^{-\beta s} f(Y_s)\, ds \mid Y_0 = i\right] = ((\beta \boldsymbol{I} - \boldsymbol{G})^{-1}\boldsymbol{f})(i).$$

EXAMPLE 4.7 Assume that the salesman of Example 4.1 wants to determine the present value of the total revenue using a discount rate of 5%, where the revenue rate function is given by $\boldsymbol{f} = (400, 500, 625)^T$. The present value of all future revenue is thus given by

$$\begin{bmatrix} \frac{11}{20} & -\frac{1}{4} & -\frac{1}{4} \\ -\frac{3}{4} & \frac{21}{20} & -\frac{1}{4} \\ -\frac{1}{2} & -\frac{1}{6} & \frac{43}{60} \end{bmatrix}^{-1} \begin{bmatrix} 400 \\ 500 \\ 625 \end{bmatrix} = \begin{bmatrix} 11.31 & 3.51 & 5.17 \\ 10.54 & 4.28 & 5.17 \\ 10.34 & 3.45 & 6.21 \end{bmatrix} \begin{bmatrix} 400 \\ 500 \\ 625 \end{bmatrix} = \begin{bmatrix} 9515 \\ 9592 \\ 9741 \end{bmatrix}.$$

Note that the total discounted revenue depends on the initial state. Thus, if at time 0 the salesman is in city a, then the present value of his total revenue is \$9515; whereas, if he started in city c, the present value of his total revenue would be \$9741.

▶ *Suggestion: Do Exercise 4.4b,c.*

4.4 TIME-DEPENDENT PROBABILITIES

A thorough treatment of the time-dependent probabilities for Markov processes is beyond the scope of this book; however, we do present a brief summary in this section for those students who may want to continue studying random processes and would like to know something of what the future might hold. For other students, this section can be skipped with no loss of continuity.

As before, the Markov process will be represented by $\{Y_t; t \geq 0\}$ with finite state space E and generator matrix $\boldsymbol{G}$. The basic relationship that probabilities for a Markov process must satisfy are analogous to the relationship that the transition probabilities of a Markov chain satisfy. As you recall from Section 2.2, the following holds for Markov chains:

$$\begin{aligned}
P\{X_{n+m} = j \mid X_0 = i\} \\
&= \sum_{k \in E} \Pr\{X_n = k \mid X_0 = i\} \times \Pr\{X_{n+m} = j \mid X_n = k\} \\
&= \sum_{k \in E} \Pr\{X_n = k \mid X_0 = i\} \times \Pr\{X_m = j \mid X_0 = k\}.
\end{aligned}$$

The second equality is a result of the stationary property for Markov chains and is called the Chapman-Kolmogorov equation for Markov chains. The Chapman-Kolmogorov equation for Markov processes can be expressed similarly as

$$\begin{aligned}
P\{Y_{t+s} = j \mid Y_0 = i\} \\
&= \sum_{k \in E} \Pr\{Y_t = k \mid Y_0 = i\} \times \Pr\{Y_{t+s} = j \mid Y_t = k\} \\
&= \sum_{k \in E} \Pr\{Y_t = k \mid Y_0 = i\} \times \Pr\{Y_s = j \mid Y_0 = k\},
\end{aligned}$$

for $t, s \geq 0$ and $i, j \in E$. In other words, the Chapman-Kolmogorov property indicates that the probability of going from state i to state j in $t + s$ time units is the sum of the product of probabilities of going from state i to an arbitrary state k in t time units and going from state k to state j in s time units (see Figure 4.6).

Before giving a closed-form expression for the time-dependent probabilities, it is necessary to first discuss the exponentiation of a matrix. Recall that

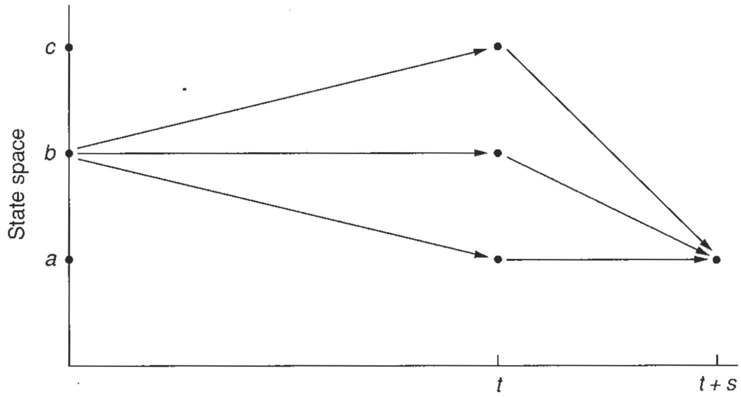


FIGURE 4.6 Graphical representation for the Chapman-Kolmogorov equations for a three-state Markov process.

for scalars, the following power series holds for all values of a:

$$e^a = \sum_{n=0}^{\infty} \frac{a^n}{n!}.$$

This same relationship holds for matrices and becomes the definition of the exponentiation of a matrix; namely,

$$e^{\boldsymbol{A}} = \sum_{n=0}^{\infty} \frac{\boldsymbol{A}^n}{n!}. \quad (4.3)$$

Note that $e^{\boldsymbol{A}}(i,j)$ refers to the $i-j$ element of the matrix $e^{\boldsymbol{A}}$; thus, in general, $e^{\boldsymbol{A}}(i,j) \neq e^{\boldsymbol{A}(i,j)}$.

Using the preceding definition and the Chapman-Kolmogorov equation, it is possible to derive the following property:

Property 4.9 Let $Y = \{Y_t; t \geq 0\}$ be a Markov process with state space E and generator matrix $\boldsymbol{G}$, then for $i, j \in E$ and $t \geq 0$

$$\Pr\{Y_t = j \mid Y_0 = i\} = e^{t\boldsymbol{G}}(i,j),$$

where $t\boldsymbol{G}$ is the matrix formed by multiplying each element of the generator matrix by the scalar t.

EXAMPLE 4.8 We are interested in an electronic device that is either "on standby" or "in use." The on-standby periods are exponentially distributed with a mean of 20 seconds, and the in-use periods are exponentially distributed with a mean of 12 seconds. Because the exponential assumption is satisfied, a Markov process $\{Y_t; t \geq 0\}$ with state space $E = \{0, 1\}$ can be used to model the electronic device as it alternates between being on standby and in use. State 0 denotes on standby, and state 1 denotes in use. The generator matrix is given by

$$\boldsymbol{G} = \begin{matrix} 0 \\ 1 \end{matrix}\begin{bmatrix} -3 & 3 \\ 5 & -5 \end{bmatrix},$$

where the time unit is minutes. Using Property 4.5, it is easy to show that the long-run probability of being in use is $\frac{3}{8}$. However, we are interested in the time-dependent probabilities. If the generator can be written in diagonal form, then it is not difficult to apply Property 4.9. In other words, if it is possible to find a matrix $\boldsymbol{Q}$ such that

$$\boldsymbol{G} = \boldsymbol{Q}\boldsymbol{D}\boldsymbol{Q}^{-1},$$

where $\boldsymbol{D}$ is a diagonal matrix, then the following holds

$$e^{t\boldsymbol{G}} = \boldsymbol{Q}e^{t\boldsymbol{D}}\boldsymbol{Q}^{-1}.$$

Furthermore, $e^{t\boldsymbol{D}}$ is a diagonal matrix whenever $\boldsymbol{D}$ is diagonal, in which case $e^{t\boldsymbol{D}}(i,i) = e^{tD(i,i)}$. For the generator matrix given, we have

$$\begin{bmatrix} -3 & 3 \\ 5 & -5 \end{bmatrix} = \begin{bmatrix} 1 & 3 \\ 1 & -5 \end{bmatrix} \cdot \begin{bmatrix} 0 & 0 \\ 0 & -8 \end{bmatrix} \cdot \begin{bmatrix} 0.625 & 0.375 \\ 0.125 & -0.125 \end{bmatrix};$$

therefore,

$$\begin{aligned} e^{\boldsymbol{G}t} &= \begin{bmatrix} 1 & 3 \\ 1 & -5 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ 0 & e^{-8t} \end{bmatrix} \cdot \begin{bmatrix} 0.625 & 0.375 \\ 0.125 & -0.125 \end{bmatrix} \\ &= \begin{bmatrix} 0.625 + 0.375e^{-8t} & 0.375 - 0.375e^{-8t} \\ 0.625 - 0.625e^{-8t} & 0.375 + 0.625e^{-8t} \end{bmatrix}. \end{aligned}$$

The probability that the electronic device is in use at time t given that it started in use at time 0 is thus given by

$$\Pr\{Y_t = 1 \mid Y_0 = 1\} = 0.375 + 0.625e^{-8t}.$$

(See Figure 4.7 for a graphical representation of the time-dependent probabilities.)

We would like to know the expected amount of time that the electronic device is in use during its first 1 minute of operation given that it was in use at time 0. Mathematically, the quantity to be calculated is given by

$$E\left[\int_0^1 Y_s \, ds \mid Y_0 = 1\right] = \int_0^1 \Pr\{Y_s = 1 \mid Y_0 = 1\} \, ds.$$

Because Y_s is 1 if the device is in use at time s and zero otherwise, the left-hand side of the above expression is the expected length of time that

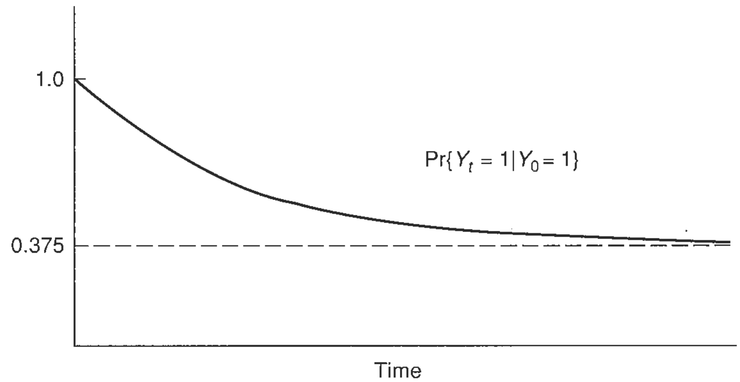


FIGURE 4.7 Time-dependent probabilities for Example 4.8.

the process is in use during its first minute of operation. The equality holds because the expected value of a Bernoulli random variable equals the probability that the random variable equals 1. (We ignore the technical details involved in interchanging the expectation operator and the integral.) Combining the preceding two equations, we have

$$E\left[\int_0^1 Y_s\, ds \mid Y_0 = 1\right] = \int_0^1 \Pr\{Y_s = 1 \mid Y_0 = 1\}\, ds$$

$$= \int_0^1 0.375 + 0.625e^{-8s}\, ds$$

$$= 0.375 + \frac{0.625}{8}(1 - e^{-8}) = 0.453 \text{ minutes}.$$

It is often useful to determine the total cost (or profit) of an operation modeled by a Markov process. In order to express the expected cost over a time interval, let $C_{(0,t)}$ be a random variable denoting the total cost incurred by the Markov process over the interval $(0, t)$. Then, for $i \in E$,

$$E[C_{(0,t)} \mid Y_0 = i] = \sum_{k \in E} f(k) \int_0^t \Pr\{Y_s = k \mid Y_0 = i\}\, ds,$$

where $f(k)$ is the cost rate incurred by the process while in state k.

For example, suppose that the electronic device costs the company 5¢ every minute it is on standby and it costs 35¢ every minute the device is in use. We may then be interested in estimating the total cost of operation for the first 2 minutes of utilizing the device given that the device was on standby when the process started; thus

$$\begin{aligned} E[C_{(0,2)} \mid Y_0 = 0] &= 0.05 \int_0^2 \Pr\{Y_s = 0 \mid Y_0 = 0\}\, ds \\ &\quad + 0.35 \int_0^2 \Pr\{Y_s = 1 \mid Y_0 = 0\}\, ds \\ &= 0.05 \times \left[0.625 \times 2 + \frac{0.375}{8}(1 - e^{-8\times 2})\right] \\ &\quad + 0.35 \times \left[0.375 \times 2 - \frac{0.375}{8}(1 - e^{-8\times 2})\right] \\ &= 0.31. \end{aligned}$$

▶ *Suggestion: Do Exercises 4.10–4.12.*

4.5 EXERCISES

4.1 The following matrix is a generator for a Markov process. Complete its entries.

$$\boldsymbol{G} = \begin{bmatrix} - & 3 & 0 & 7 \\ 5 & -12 & 4 & - \\ 1 & - & -6 & 5 \\ 3 & 2 & 9 & - \end{bmatrix}$$

4.2 Let Y be a Markov process with state space $\{a, b, c, d\}$ and an imbedded Markov chain having a Markov matrix given by

$$\boldsymbol{P} = \begin{bmatrix} 0.0 & 0.1 & 0.2 & 0.7 \\ 0.0 & 0.0 & 0.4 & 0.6 \\ 0.8 & 0.1 & 0.0 & 0.1 \\ 1.0 & 0.0 & 0.0 & 0.0 \end{bmatrix}.$$

The mean sojourn times in states a, b, c, and d are 2, 5, 0.5, and 1, respectively.

(a) Give the generator matrix for this Markov process.

(b) What is $\lim_{t\to\infty} \Pr\{Y_t = a\}$?

(c) Let $\boldsymbol{r} = (10, 25, 30, 50)^T$ be a reward vector and determine $\lim_{t\to\infty} \frac{1}{t} E[\int_0^t r(Y_t)\, dt]$.

4.3 Let T and S be exponentially distributed random variables with means $1/a$ and $1/b$, respectively. Define the random variable $U = \min\{T, S\}$. Justify the relationship $\Pr\{U > u\} = \Pr\{T > u\} \times \Pr\{S > u\}$ and derive the distribution function for U.

4.4 A revenue-producing system can be in one of four states: high income, medium income, low income, costs. The movement of the system among the states is according to a Markov process Y with state space $E = \{h, m, l, c\}$ and with a generator matrix given by

$$G = \begin{array}{c} h \\ m \\ l \\ c \end{array}\begin{bmatrix} -0.2 & 0.1 & 0.1 & 0.0 \\ 0.0 & -0.4 & 0.3 & 0.1 \\ 0.0 & 0.0 & -0.5 & 0.5 \\ 1.5 & 0.0 & 0.0 & -1.5 \end{bmatrix}.$$

While the system is in state h, m, l, or c it produces a profit at a rate of \$500, \$250, \$100, or –\$600 per time unit. The company would like to reduce the time spent in the fourth state and has determined that by doubling the cost (i.e., from \$600 to \$1200) incurred while in that state the mean time spent in the state can be cut in half. Is the additional expense worthwhile?

(a) Use the long-run average profit for the criterion.

(b) Use a total discounted-cost criterion assuming a discount rate of 10%.

(c) Use a total discounted-cost criterion assuming the company uses a 25% annual rate of return, and the time step for this problem is assumed to be in weeks.

4.5 Let Y be a Markov process with state space $\{a, b, c, d\}$ and generator matrix given by

$$G = \begin{bmatrix} -5 & 4 & 0 & 1 \\ 6 & -10 & 4 & 0 \\ 0 & 0 & -1 & 1 \\ 2 & 0 & 0 & -2 \end{bmatrix}.$$

Costs are incurred at a rate of \$100, \$300, \$500, and \$1000 per time unit while the process is in states a, b, c, and d, respectively. Furthermore, each time a jump is made from state d to state a an additional cost of \$5000 is incurred. (All other jumps do not incur additional costs.) For a maintenance cost of \$400 per time unit, all cost rates can be cut in half and the "jump" cost can be eliminated. Based on long-run averages, is the maintenance cost worthwhile?

4.6 A small gas station has one pump and room for a total of three cars (one at the pump and two waiting). The time between car arrivals to the station is an exponential random variable with the average arrival rate of ten cars per hour. The time each car spends in front of the pump is an exponential random variable with a mean of 5 minutes (i.e., a mean rate of 12 per hour). If there are three cars in the station and another car arrives, the newly arrived car keeps going and never enters the station.

(a) Model the gas station as a Markov process Y, where Y_t denotes the number of cars in the station at time t. Give its generator matrix.

(b) What is the long-run probability that the station is empty?

(c) What is the long-run expected number of cars in the station?

4.7 A small convenience store has a room for only five people inside. Cars arrive at the store randomly, with the interarrival times between cars being an exponential random variable with a mean of ten cars arriving each hour. The number of people within each car is a random variable, N, where $\Pr\{N = 1\} = 0.1$, $\Pr\{N = 2\} = 0.7$, and $\Pr\{N = 3\} = 0.2$. People from the cars come into the store and stay in the store an exponential length of time. The mean length of stay in the store is 10 minutes and each person acts independent of all other people, leaving the store singly and waiting in their cars for the others. If a car arrives and the store is too full for everyone in the car to enter the store, the car will leave and nobody from that car will enter the store. Model the store as a Markov process Y, where Y_t denotes the number of individuals in the store at time t. Give its generator matrix.

4.8 A certain piece of electronic equipment has two components. The time until failure for component A is described by an exponential distribution function with a mean time of 100 hours. Component B has a mean life until failure of 200 hours and is also described by an exponential distribution. When one component fails, the equipment is turned off and maintenance is performed. The time to fix the component is exponentially distributed with a mean time of 5 hours if it was A that failed and 4 hours if it was B that failed. Let Y be a Markov process with state space $E = \{w, a, b\}$, where state w denotes that the equipment is working, a denotes that component A has failed, and b denotes that component B has failed.

(a) Give the generator for Y.

(b) What is the long-run probability that the equipment is working?

(c) An outside contractor does the repair work on the components when a failure occurs and charges \$100 per hour for time plus travel expenses, which is an additional \$500 for each visit. The company has determined that they can hire and train their own repairperson. If they have their own employee for the repair work, it will cost the company \$40 per hour while the machine is running as well as when it is down. Ignoring the initial training cost and the possibility that an employee who is hired for repair work can do other things while the machine is running, is it economically worthwhile to hire and train their own person?

4.9 An electronic component works as follows: Electric impulses arrive at the component with exponentially distributed interarrival times such that the mean arrival rate of impulses is 90 per hour. An impulse is "stored" until the third impulse arrives, then the component "fires" and enters a "recovery" phase. If an impulse arrives while the component is in the recovery phase, it is ignored. The length of time for which the component remains in the recovery phase is an exponential random variable with a mean time of 1 minute. After the recovery phase is over, the cycle is repeated; that is, the third arriving impulse will instantaneously fire the component.

(a) Give the generator matrix for a Markov process model of the dynamics of this electronic component.

(b) What is the long-run probability that the component is in the recovery phase?

(c) How many times would you expect the component to fire each hour?

4.10 Consider the electronic device that is either in the on-standby or in-use state as described in Example 4.8. Find the following quantities.

(a) The expected cost incurred during the time interval between the third and fourth minutes, given that at time zero the device was in use.

(b) The expected cost incurred during the time interval between the third and fourth minutes, given that at time zero the device was in use with probability 0.8 and on standby with probability 0.2.

4.11 Let $\boldsymbol{\mu}$ be a vector of initial probabilities for a Markov process Y, and let $\boldsymbol{f}$ denote a cost rate vector associated with Y. Write a general expression for the expected cost incurred by the Markov process during the time interval $(t, t + s)$.

4.12 Let $\boldsymbol{f}$ denote a profit rate vector associated with a Markov process Y, and let β denote a discount rate. Write a general expression for the profit returned by the process during the interval $(t, t + s)$ given that the process was in state i at time 0.

4.13 Simulate the Markov process defined by the following generator matrix:

$$\boldsymbol{G} = \begin{matrix} a \\ b \\ c \end{matrix}\begin{bmatrix} -1.0 & 0.3 & 0.7 \\ 0.1 & -0.2 & 0.1 \\ 0.3 & 0.2 & -0.5 \end{bmatrix}.$$

(Example 3.6 may provide a guide for simulating Markov processes.) From the simulation, estimate the probability that the process will be in state b at time 10 given that the process started in state a.

4.14 Write a computer program to determine the long-run probability of being in state a for the Markov process of Problem 4.13. (Remember that the long-run probability for a particular state can be estimated by the fraction of time that the process spends in that state. The estimate can also be improved slightly if the initial conditions are ignored. In other words,

$$p(j) \approx \left(\int_{t_0}^{t} I(Y_s, j)\, ds \right) \Big/ (t - t_0),$$

where t_0 is small and t is large. Part of your problem will be to determine quantitative values for "small" and "large.")