Continuous-Time Markov Chains

6.1 Introduction

In this chapter we consider a class of probability models that has a wide variety of applications in the real world. The members of this class are the continuous-time analogs of the Markov chains of Chapter 4 and as such are characterized by the Markovian property that, given the present state, the future is independent of the past.

One example of a continuous-time Markov chain has already been met. This is the Poisson process of Chapter 5. For if we let the total number of arrivals by time t (that is, $N(t)$) be the state of the process at time t, then the Poisson process is a continuous-time Markov chain having states $0, 1, 2, \ldots$ that always proceeds from state n to state $n + 1$, where $n \geqslant 0$. Such a process is known as a *pure birth process* since when a transition occurs the state of the system is always increased by one. More generally, an exponential model that can go (in one transition) only from state n to either state $n - 1$ or state $n + 1$ is called a *birth and death model*. For such a model, transitions from state n to state $n + 1$ are designated as births, and those from n to $n - 1$ as deaths. Birth and death models have wide applicability in the study of biological systems and in the study of waiting line systems in which the state represents the number of customers in the system. These models will be studied extensively in this chapter.

Introduction to Probability Models, ISBN: 9780123756862

In Section 6.2 we define continuous-time Markov chains and then relate them to the discrete-time Markov chains of Chapter 4. In Section 6.3 we consider birth and death processes and in Section 6.4 we derive two sets of differential equations—the forward and backward equations—that describe the probability laws for the system. The material in Section 6.5 is concerned with determining the limiting (or long-run) probabilities connected with a continuous-time Markov chain. In Section 6.6 we consider the topic of time reversibility. We show that all birth and death processes are time reversible, and then illustrate the importance of this observation to queueing systems. In the final section we show how to "uniformize" Markov chains, a technique useful for numerical computations.

6.2 Continuous-Time Markov Chains

Suppose we have a continuous-time stochastic process $\{X(t), t \geqslant 0\}$ taking on values in the set of nonnegative integers. In analogy with the definition of a discrete-time Markov chain, given in Chapter 4, we say that the process $\{X(t), t \geqslant 0\}$ is a *continuous-time Markov chain* if for all $s, t \geqslant 0$ and nonnegative integers $i, j, x(u), 0 \leqslant u < s$

$$\begin{aligned} P\{X(t+s) = j | X(s) = i,\ X(u) = x(u),\ 0 \leqslant u < s\} \\ = P\{X(t+s) = j | X(s) = i\} \end{aligned}$$

In other words, a continuous-time Markov chain is a stochastic process having the Markovian property that the conditional distribution of the future $X(t+s)$ given the present $X(s)$ and the past $X(u)$, $0 \leqslant u < s$, depends only on the present and is independent of the past. If, in addition,

$$P\{X(t+s) = j | X(s) = i\}$$

is independent of s, then the continuous-time Markov chain is said to have stationary or homogeneous transition probabilities.

All Markov chains considered in this text will be assumed to have stationary transition probabilities.

Suppose that a continuous-time Markov chain enters state i at some time, say, time 0, and suppose that the process does not leave state i (that is, a transition does not occur) during the next ten minutes. What is the probability that the process will not leave state i during the following five minutes? Since the process is in state i at time 10 it follows, by the Markovian property, that the probability that it remains in that state during the interval $[10, 15]$ is just the (unconditional) probability that it stays in state i for at least five minutes. That is, if we let T_i denote the amount of time that the process stays in state i before making a transition into a different state, then

$$P\{T_i > 15 | T_i > 10\} = P\{T_i > 5\}$$

or, in general, by the same reasoning,

$$P\{T_i > s + t | T_i > s\} = P\{T_i > t\}$$

for all $s, t \geqslant 0$. Hence, the random variable T_i is *memoryless* and must thus (see Section 5.2.2) be *exponentially* distributed.

In fact, the preceding gives us another way of defining a continuous-time Markov chain. Namely, it is a stochastic process having the properties that each time it enters state i

(i) the amount of time it spends in that state before making a transition into a different state is exponentially distributed with mean, say, $1/v_i$, and
(ii) when the process leaves state i, it next enters state j with some probability, say, P_{ij}. Of course, the P_{ij} must satisfy

$$\begin{aligned} P_{ii} &= 0, \qquad \text{all } i \\ \sum_j P_{ij} &= 1, \qquad \text{all } i \end{aligned}$$

In other words, a continuous-time Markov chain is a stochastic process that moves from state to state in accordance with a (discrete-time) Markov chain, but is such that the amount of time it spends in each state, before proceeding to the next state, is exponentially distributed. In addition, the amount of time the process spends in state i, and the next state visited, must be independent random variables. For if the next state visited were dependent on T_i, then information as to how long the process has already been in state i would be relevant to the prediction of the next state—and this contradicts the Markovian assumption.

Example 6.1 (A Shoe Shine Shop) Consider a shoe shine establishment consisting of two chairs—chair 1 and chair 2. A customer upon arrival goes initially to chair 1 where his shoes are cleaned and polish is applied. After this is done the customer moves on to chair 2 where the polish is buffed. The service times at the two chairs are assumed to be independent random variables that are exponentially distributed with respective rates μ_1 and μ_2. Suppose that potential customers arrive in accordance with a Poisson process having rate λ, and that a potential customer will enter the system only if both chairs are empty.

The preceding model can be analyzed as a continuous-time Markov chain, but first we must decide upon an appropriate state space. Since a potential customer will enter the system only if there are no other customers present, it follows that there will always either be 0 or 1 customers in the system. However, if there is 1 customer in the system, then we would also need to know which chair he was presently in. Hence, an appropriate state space might consist of the three states 0, 1, and 2 where the states have the following interpretation:

State	*Interpretation*
0	system is empty
1	a customer is in chair 1
2	a customer is in chair 2

We leave it as an exercise for you to verify that

$$v_0 = \lambda, \quad v_1 = \mu_1, \quad v_2 = \mu_2,$$
$$P_{01} = P_{12} = P_{20} = 1$$ ■

6.3 Birth and Death Processes

Consider a system whose state at any time is represented by the number of people in the system at that time. Suppose that whenever there are n people in the system, then (i) new arrivals enter the system at an exponential rate λ_n, and (ii) people leave the system at an exponential rate μ_n. That is, whenever there are n persons in the system, then the time until the next arrival is exponentially distributed with mean $1/\lambda_n$ and is independent of the time until the next departure, which is itself exponentially distributed with mean $1/\mu_n$. Such a system is called a birth and death process. The parameters $\{\lambda_n\}_{n=0}^{\infty}$ and $\{\mu_n\}_{n=1}^{\infty}$ are called, respectively, the arrival (or birth) and departure (or death) rates.

Thus, a birth and death process is a continuous-time Markov chain with states $\{0, 1, \ldots\}$ for which transitions from state n may go only to either state $n - 1$ or state $n + 1$. The relationships between the birth and death rates and the state transition rates and probabilities are

$$\begin{aligned}
v_0 &= \lambda_0, \\
v_i &= \lambda_i + \mu_i, \qquad i > 0 \\
P_{01} &= 1, \\
P_{i,i+1} &= \frac{\lambda_i}{\lambda_i + \mu_i}, \qquad i > 0 \\
P_{i,i-1} &= \frac{\mu_i}{\lambda_i + \mu_i}, \qquad i > 0
\end{aligned}$$

The preceding follows, because if there are i in the system, then the next state will be $i + 1$ if a birth occurs before a death, and the probability that an exponential random variable with rate λ_i will occur earlier than an (independent) exponential with rate μ_i is $\lambda_i/(\lambda_i + \mu_i)$. Moreover, the time until either a birth or a death occurs is exponentially distributed with rate $\lambda_i + \mu_i$ (and so, $v_i = \lambda_i + \mu_i$).

Example 6.2 (The Poisson Process) Consider a birth and death process for which

$$\mu_n = 0, \qquad \text{for all } n \geqslant 0$$
$$\lambda_n = \lambda, \qquad \text{for all } n \geqslant 0$$

This is a process in which departures never occur, and the time between successive arrivals is exponential with mean $1/\lambda$. Hence, this is just the Poisson process. ■

A birth and death process for which $\mu_n = 0$ for all n is called a pure birth process. Another pure birth process is given by the next example.

Example 6.3 (A Birth Process with Linear Birth Rate) Consider a population whose members can give birth to new members but cannot die. If each member acts independently of the others and takes an exponentially distributed amount of time, with mean $1/\lambda$, to give birth, then if $X(t)$ is the population size at time t, then $\{X(t), t \geqslant 0\}$ is a pure birth process with $\lambda_n = n\lambda$, $n \geqslant 0$. This follows since if the population consists of n persons and each gives birth at an exponential rate λ, then the total rate at which births occur is $n\lambda$. This pure birth process is known as a *Yule process* after G. Yule, who used it in his mathematical theory of evolution. ■

Example 6.4 (A Linear Growth Model with Immigration) A model in which

$$\mu_n = n\mu, \qquad n \geqslant 1$$
$$\lambda_n = n\lambda + \theta, \qquad n \geqslant 0$$

is called a *linear growth process with immigration*. Such processes occur naturally in the study of biological reproduction and population growth. Each individual in the population is assumed to give birth at an exponential rate λ; in addition, there is an exponential rate of increase θ of the population due to an external source such as immigration. Hence, the total birth rate where there are n persons in the system is $n\lambda + \theta$. Deaths are assumed to occur at an exponential rate μ for each member of the population, so $\mu_n = n\mu$.

Let $X(t)$ denote the population size at time t. Suppose that $X(0) = i$ and let

$$M(t) = E[X(t)]$$

We will determine $M(t)$ by deriving and then solving a differential equation that it satisfies.

We start by deriving an equation for $M(t + h)$ by conditioning on $X(t)$. This yields

$$\begin{aligned} M(t + h) &= E[X(t + h)] \\ &= E[E[X(t + h)|X(t)]] \end{aligned}$$

Now, given the size of the population at time t then, ignoring events whose probability is $o(h)$, the population at time $t + h$ will either increase in size by 1 if a birth or an immigration occurs in $(t, t + h)$, or decrease by 1 if a death occurs in this interval, or remain the same if neither of these two possibilities occurs. That is, given $X(t)$,

$$X(t+h) = \begin{cases} X(t) + 1, & \text{with probability } [\theta + X(t)\lambda]h + o(h) \\ X(t) - 1, & \text{with probability } X(t)\mu h + o(h) \\ X(t), & \text{with probability } 1 - [\theta + X(t)\lambda + X(t)\mu]\, h + o(h) \end{cases}$$

Therefore,

$$E[X(t+h)|X(t)] = X(t) + [\theta + X(t)\lambda - X(t)\mu]h + o(h)$$

Taking expectations yields

$$M(t+h) = M(t) + (\lambda - \mu)M(t)h + \theta h + o(h)$$

or, equivalently,

$$\frac{M(t+h) - M(t)}{h} = (\lambda - \mu)M(t) + \theta + \frac{o(h)}{h}$$

Taking the limit as $h \to 0$ yields the differential equation

$$M'(t) = (\lambda - \mu)M(t) + \theta \tag{6.1}$$

If we now define the function $h(t)$ by

$$h(t) = (\lambda - \mu)M(t) + \theta$$

then

$$h'(t) = (\lambda - \mu)M'(t)$$

Therefore, Differential Equation (6.1) can be rewritten as

$$\frac{h'(t)}{\lambda - \mu} = h(t)$$

or

$$\frac{h'(t)}{h(t)} = \lambda - \mu$$

Integration yields

$$\log[h(t)] = (\lambda - \mu)t + c$$

or

$$h(t) = Ke^{(\lambda-\mu)t}$$

Putting this back in terms of $M(t)$ gives

$$\theta + (\lambda - \mu)M(t) = Ke^{(\lambda-\mu)t}$$

To determine the value of the constant K, we use the fact that $M(0) = i$ and evaluate the preceding at $t = 0$. This gives

$$\theta + (\lambda - \mu)i = K$$

Substituting this back in the preceding equation for $M(t)$ yields the following solution for $M(t)$:

$$M(t) = \frac{\theta}{\lambda - \mu}[e^{(\lambda-\mu)t} - 1] + ie^{(\lambda-\mu)t}$$

Note that we have implicitly assumed that $\lambda \neq \mu$. If $\lambda = \mu$, then Differential Equation (6.1) reduces to

$$M'(t) = \theta \tag{6.2}$$

Integrating (6.2) and using that $M(0) = i$ gives the solution

$$M(t) = \theta t + i$$ ■

Example 6.5 (The Queueing System $M/M/1$) Suppose that customers arrive at a single-server service station in accordance with a Poisson process having rate λ. That is, the times between successive arrivals are independent exponential random variables having mean $1/\lambda$. Upon arrival, each customer goes directly into service if the server is free; if not, then the customer joins the queue (that is, he waits in line). When the server finishes serving a customer, the customer leaves the system and the next customer in line, if there are any waiting, enters the service. The successive service times are assumed to be independent exponential random variables having mean $1/\mu$.

The preceding is known as the $M/M/1$ queueing system. The first M refers to the fact that the interarrival process is Markovian (since it is a Poisson process) and the second to the fact that the service distribution is exponential (and, hence, Markovian). The 1 refers to the fact that there is a single server.

If we let $X(t)$ denote the number in the system at time t then $\{X(t), t \geqslant 0\}$ is a birth and death process with

$$\mu_n = \mu, \qquad n \geqslant 1$$
$$\lambda_n = \lambda, \qquad n \geqslant 0$$

■

Example 6.6 (A Multiserver Exponential Queueing System) Consider an exponential queueing system in which there are s servers available, each serving at rate μ. An entering customer first waits in line and then goes to the first free server. This is a birth and death process with parameters

$$\mu_n = \begin{cases} n\mu, & 1 \leqslant n \leqslant s \\ s\mu, & n > s \end{cases}$$
$$\lambda_n = \lambda, \qquad n \geqslant 0$$

To see why this is true, reason as follows: If there are n customers in the system, where $n \leqslant s$, then n servers will be busy. Since each of these servers works at rate μ, the total departure rate will be $n\mu$. On the other hand, if there are n customers in the system, where $n > s$, then all s of the servers will be busy, and thus the total departure rate will be $s\mu$. This is known as an $M/M/s$ queueing model. ■

Consider now a general birth and death process with birth rates $\{\lambda_n\}$ and death rates $\{\mu_n\}$, where $\mu_0 = 0$, and let T_i denote the time, starting from state i, it takes for the process to enter state $i + 1$, $i \geqslant 0$. We will recursively compute $E[T_i]$, $i \geqslant 0$, by starting with $i = 0$. Since T_0 is exponential with rate λ_0, we have

$$E[T_0] = \frac{1}{\lambda_0}$$

For $i > 0$, we condition whether the first transition takes the process into state $i - 1$ or $i + 1$. That is, let

$$I_i = \begin{cases} 1, & \text{if the first transition from } i \text{ is to } i + 1 \\ 0, & \text{if the first transition from } i \text{ is to } i - 1 \end{cases}$$

and note that

$$\begin{aligned} E[T_i | I_i = 1] &= \frac{1}{\lambda_i + \mu_i}, \\ E[T_i | I_i = 0] &= \frac{1}{\lambda_i + \mu_i} + E[T_{i-1}] + E[T_i] \end{aligned} \tag{6.3}$$

This follows since, independent of whether the first transition is from a birth or death, the time until it occurs is exponential with rate $\lambda_i + \mu_i$; if this first transition is a birth, then the population size is at $i + 1$, so no additional time is needed; whereas if it is death, then the population size becomes $i - 1$ and the additional time needed to reach $i + 1$ is equal to the time it takes to return to state i (this has mean $E[T_{i-1}]$) plus the additional time it then takes to reach $i + 1$ (this has mean $E[T_i]$). Hence, since the probability that the first transition is a birth is $\lambda_i/(\lambda_i + \mu_i)$, we see that

$$E[T_i] = \frac{1}{\lambda_i + \mu_i} + \frac{\mu_i}{\lambda_i + \mu_i}(E[T_{i-1}] + E[T_i])$$

or, equivalently,

$$E[T_i] = \frac{1}{\lambda_i} + \frac{\mu_i}{\lambda_i}E[T_{i-1}], \qquad i \geqslant 1$$

Starting with $E[T_0] = 1/\lambda_0$, the preceding yields an efficient method to successively compute $E[T_1]$, $E[T_2]$, and so on.

Suppose now that we wanted to determine the expected time to go from state i to state j where $i < j$. This can be accomplished using the preceding by noting that this quantity will equal $E[T_i] + E[T_{i+1}] + \cdots + E[T_{j-1}]$.

Example 6.7 For the birth and death process having parameters $\lambda_i \equiv \lambda$, $\mu_i \equiv \mu$,

$$\begin{aligned} E[T_i] &= \frac{1}{\lambda} + \frac{\mu}{\lambda}E[T_{i-1}] \\ &= \frac{1}{\lambda}(1 + \mu E[T_{i-1}]) \end{aligned}$$

Starting with $E[T_0] = 1/\lambda$, we see that

$$\begin{aligned} E[T_1] &= \frac{1}{\lambda}\left(1 + \frac{\mu}{\lambda}\right), \\ E[T_2] &= \frac{1}{\lambda}\left[1 + \frac{\mu}{\lambda} + \left(\frac{\mu}{\lambda}\right)^2\right] \end{aligned}$$

and, in general,

$$\begin{aligned} E[T_i] &= \frac{1}{\lambda}\left[1 + \frac{\mu}{\lambda} + \left(\frac{\mu}{\lambda}\right)^2 + \cdots + \left(\frac{\mu}{\lambda}\right)^i\right] \\ &= \frac{1 - (\mu/\lambda)^{i+1}}{\lambda - \mu}, \qquad i \geqslant 0 \end{aligned}$$

The expected time to reach state j, starting at state k, $k < j$, is

$$
\begin{aligned}
E[\text{time to go from } k \text{ to } j] &= \sum_{i=k}^{j-1} E[T_i] \\
&= \frac{j-k}{\lambda-\mu} - \frac{(\mu/\lambda)^{k+1}}{\lambda-\mu}\frac{\left[1-(\mu/\lambda)^{j-k}\right]}{1-\mu/\lambda}
\end{aligned}
$$

The foregoing assumes that $\lambda \neq \mu$. If $\lambda = \mu$, then

$$
\begin{aligned}
E[T_i] &= \frac{i+1}{\lambda}, \\
E[\text{time to go from } k \text{ to } j] &= \frac{j(j+1)-k(k+1)}{2\lambda}
\end{aligned}
$$

■

We can also compute the variance of the time to go from 0 to $i+1$ by utilizing the conditional variance formula. First note that Equation (6.3) can be written as

$$
E[T_i|I_i] = \frac{1}{\lambda_i+\mu_i} + (1-I_i)(E[T_{i-1}] + E[T_i])
$$

Thus,

$$
\begin{aligned}
\mathrm{Var}(E[T_i|I_i]) &= (E[T_{i-1}] + E[T_i])^2\,\mathrm{Var}(I_i) \\
&= (E[T_{i-1}] + E[T_i])^2\frac{\mu_i\lambda_i}{(\mu_i+\lambda_i)^2} \qquad (6.4)
\end{aligned}
$$

where $\mathrm{Var}(I_i)$ is as shown since I_i is a Bernoulli random variable with parameter $p = \lambda_i/(\lambda_i+\mu_i)$. Also, note that if we let X_i denote the time until the transition from i occurs, then

$$
\begin{aligned}
\mathrm{Var}(T_i|I_i=1) &= \mathrm{Var}(X_i|I_i=1) \\
&= \mathrm{Var}(X_i) \\
&= \frac{1}{(\lambda_i+\mu_i)^2} \qquad (6.5)
\end{aligned}
$$

where the preceding uses the fact that the time until transition is independent of the next state visited. Also,

$$
\begin{aligned}
\mathrm{Var}(T_i|I_i=0) &= \mathrm{Var}(X_i + \text{time to get back to } i + \text{time to then reach } i+1) \\
&= \mathrm{Var}(X_i) + \mathrm{Var}(T_{i-1}) + \mathrm{Var}(T_i) \qquad (6.6)
\end{aligned}
$$

where the foregoing uses the fact that the three random variables are independent. We can rewrite Equations (6.5) and (6.6) as

$$\mathrm{Var}(T_i|I_i) = \mathrm{Var}(X_i) + (1 - I_i)[\mathrm{Var}(T_{i-1}) + \mathrm{Var}(T_i)]$$

so

$$E[\mathrm{Var}(T_i|I_i)] = \frac{1}{(\mu_i + \lambda_i)^2} + \frac{\mu_i}{\mu_i + \lambda_i}[\mathrm{Var}(T_{i-1}) + \mathrm{Var}(T_i)] \tag{6.7}$$

Hence, using the conditional variance formula, which states that $\mathrm{Var}(T_i)$ is the sum of Equations (6.7) and (6.4), we obtain

$$\begin{aligned}\mathrm{Var}(T_i) =& \frac{1}{(\mu_i + \lambda_i)^2} + \frac{\mu_i}{\mu_i + \lambda_i}[\mathrm{Var}(T_{i-1}) + \mathrm{Var}(T_i)] \\ &+ \frac{\mu_i\lambda_i}{(\mu_i + \lambda_i)^2}(E[T_{i-1}] + E[T_i])^2\end{aligned}$$

or, equivalently,

$$\mathrm{Var}(T_i) = \frac{1}{\lambda_i(\lambda_i + \mu_i)} + \frac{\mu_i}{\lambda_i}\,\mathrm{Var}(T_{i-1}) + \frac{\mu_i}{\mu_i + \lambda_i}(E[T_{i-1}] + E[T_i])^2$$

Starting with $\mathrm{Var}(T_0) = 1/\lambda_0^2$ and using the former recursion to obtain the expectations, we can recursively compute $\mathrm{Var}(T_i)$. In addition, if we want the variance of the time to reach state j, starting from state k, $k < j$, then this can be expressed as the time to go from k to $k + 1$ plus the additional time to go from $k + 1$ to $k + 2$, and so on. Since, by the Markovian property, these successive random variables are independent, it follows that

$$\mathrm{Var}(\text{time to go from } k \text{ to } j) = \sum_{i=k}^{j-1} \mathrm{Var}(T_i)$$

6.4 The Transition Probability Function $P_{ij}(t)$

Let

$$P_{ij}(t) = P\{X(t + s) = j|X(s) = i\}$$

denote the probability that a process presently in state i will be in state j a time t later. These quantities are often called the *transition probabilities* of the continuous-time Markov chain.

We can explicitly determine $P_{ij}(t)$ in the case of a pure birth process having distinct birth rates. For such a process, let X_k denote the time the process spends in state k before making a transition into state $k+1$, $k \geqslant 1$. Suppose that the process is presently in state i, and let $j > i$. Then, as X_i is the time it spends in state i before moving to state $i+1$, and X_{i+1} is the time it then spends in state $i+1$ before moving to state $i+2$, and so on, it follows that $\sum_{k=i}^{j-1} X_k$ is the time it takes until the process enters state j. Now, if the process has not yet entered state j by time t, then its state at time t is smaller than j, and vice versa. That is,

$$X(t) < j \;\Leftrightarrow\; X_i + \cdots + X_{j-1} > t$$

Therefore, for $i < j$, we have for a pure birth process that

$$P\{X(t) < j | X(0) = i\} = P\left\{\sum_{k=i}^{j-1} X_k > t\right\}$$

However, since $X_i, \ldots, X_{j-1}$ are independent exponential random variables with respective rates $\lambda_i, \ldots, \lambda_{j-1}$, we obtain from the preceding and Equation (5.9), which gives the tail distribution function of $\sum_{k=i}^{j-1} X_k$, that

$$P\{X(t) < j | X(0) = i\} = \sum_{k=i}^{j-1} e^{-\lambda_k t} \prod_{r \neq k,\, r=i}^{j-1} \frac{\lambda_r}{\lambda_r - \lambda_k}$$

Replacing j by $j+1$ in the preceding gives

$$P\{X(t) < j+1 | X(0) = i\} = \sum_{k=i}^{j} e^{-\lambda_k t} \prod_{r \neq k,\, r=i}^{j} \frac{\lambda_r}{\lambda_r - \lambda_k}$$

Since

$$P\{X(t) = j | X(0) = i\} = P\{X(t) < j+1 | X(0) = i\} - P\{X(t) < j | X(0) = i\}$$

and since $P_{ii}(t) = P\{X_i > t\} = e^{-\lambda_i t}$, we have shown the following.

Proposition 6.1 For a pure birth process having $\lambda_i \neq \lambda_j$ when $i \neq j$

$$P_{ij}(t) = \sum_{k=i}^{j} e^{-\lambda_k t} \prod_{r \neq k,\, r=i}^{j} \frac{\lambda_r}{\lambda_r - \lambda_k} - \sum_{k=i}^{j-1} e^{-\lambda_k t} \prod_{r \neq k,\, r=i}^{j-1} \frac{\lambda_r}{\lambda_r - \lambda_k}, \qquad i < j$$

$$P_{ii}(t) = e^{-\lambda_i t}$$

Example 6.8 Consider the Yule process, which is a pure birth process in which each individual in the population independently gives birth at rate λ, and so $\lambda_n = n\lambda$, $n \geqslant 1$. Letting $i = 1$, we obtain from Proposition 6.1

$$\begin{aligned}
P_{1j}(t) &= \sum_{k=1}^{j} e^{-k\lambda t} \prod_{r\neq k,\, r=1}^{j} \frac{r}{r-k} - \sum_{k=1}^{j-1} e^{-k\lambda t} \prod_{r\neq k,\, r=1}^{j-1} \frac{r}{r-k} \\
&= e^{-j\lambda t} \prod_{r=1}^{j-1} \frac{r}{r-j} + \sum_{k=1}^{j-1} e^{-k\lambda t} \left(\prod_{r\neq k,\, r=1}^{j} \frac{r}{r-k} - \prod_{r\neq k,\, r=1}^{j-1} \frac{r}{r-k} \right) \\
&= e^{-j\lambda t}(-1)^{j-1} + \sum_{k=1}^{j-1} e^{-k\lambda t} \left(\frac{j}{j-k} - 1 \right) \prod_{r\neq k,\, r=1}^{j-1} \frac{r}{r-k}
\end{aligned}$$

Now,

$$\begin{aligned}
\frac{k}{j-k} \prod_{r\neq k,\, r=1}^{j-1} \frac{r}{r-k} &= \frac{(j-1)!}{(1-k)(2-k)\cdots(k-1-k)(j-k)!} \\
&= (-1)^{k-1} \binom{j-1}{k-1}
\end{aligned}$$

so

$$\begin{aligned}
P_{1j}(t) &= \sum_{k=1}^{j} \binom{j-1}{k-1} e^{-k\lambda t} (-1)^{k-1} \\
&= e^{-\lambda t} \sum_{i=0}^{j-1} \binom{j-1}{i} e^{-i\lambda t} (-1)^{i} \\
&= e^{-\lambda t}(1 - e^{-\lambda t})^{j-1}
\end{aligned}$$

Thus, starting with a single individual, the population size at time t has a geometric distribution with mean $e^{\lambda t}$. If the population starts with i individuals, then we can regard each of these individuals as starting her own independent Yule process, and so the population at time t will be the sum of i independent and identically distributed geometric random variables with parameter $e^{-\lambda t}$. But this means that the conditional distribution of $X(t)$, given that $X(0) = i$, is the same as the distribution of the number of times that a coin that lands heads on each flip with probability $e^{-\lambda t}$ must be flipped to amass a total of i heads. Hence, the population size at time t has a negative binomial distribution with parameters i and $e^{-\lambda t}$, so

$$P_{ij}(t) = \binom{j-1}{i-1} e^{-i\lambda t}(1 - e^{-\lambda t})^{j-i}, \qquad j \geqslant i \geqslant 1$$

(We could, of course, have used Proposition 6.1 to immediately obtain an equation for $P_{ij}(t)$, rather than just using it for $P_{1j}(t)$, but the algebra that would have then been needed to show the equivalence of the resulting expression to the preceding result is somewhat involved.) ∎

We shall now derive a set of differential equations that the transition probabilities $P_{ij}(t)$ satisfy in a general continuous-time Markov chain. However, first we need a definition and a pair of lemmas.

For any pair of states i and j, let

$$q_{ij} = v_i P_{ij}$$

Since v_i is the rate at which the process makes a transition when in state i and P_{ij} is the probability that this transition is into state j, it follows that q_{ij} is the rate, when in state i, at which the process makes a transition into state j. The quantities q_{ij} are called the *instantaneous transition rates*. Since

$$v_i = \sum_j v_i P_{ij} = \sum_j q_{ij}$$

and

$$P_{ij} = \frac{q_{ij}}{v_i} = \frac{q_{ij}}{\sum_j q_{ij}}$$

it follows that specifying the instantaneous transition rates determines the parameters of the continuous-time Markov chain.

Lemma 6.2

(a) $$\lim_{h\to 0} \frac{1 - P_{ii}(h)}{h} = v_i$$

(b) $$\lim_{h\to 0} \frac{P_{ij}(h)}{h} = q_{ij} \qquad \text{when } i \neq j$$

Proof. We first note that since the amount of time until a transition occurs is exponentially distributed it follows that the probability of two or more transitions in a time h is $o(h)$. Thus, $1 - P_{ii}(h)$, the probability that a process in state i at time 0 will not be in state i at time h, equals the probability that a transition occurs within time h plus something small compared to h. Therefore,

$$1 - P_{ii}(h) = v_i h + o(h)$$

and part (a) is proven. To prove part (b), we note that $P_{ij}(h)$, the probability that the process goes from state i to state j in a time h, equals the probability that a transition occurs in this time multiplied by the probability that the transition is into state j, plus something small compared to h. That is,

$$P_{ij}(h) = hv_i P_{ij} + o(h)$$

and part (b) is proven. ■

Lemma 6.3 For all $s \geqslant 0$, $t \geqslant 0$,

$$P_{ij}(t+s) = \sum_{k=0}^{\infty} P_{ik}(t)P_{kj}(s) \tag{6.8}$$

Proof. In order for the process to go from state i to state j in time $t + s$, it must be somewhere at time t and thus

$$\begin{aligned}
P_{ij}(t+s) &= P\{X(t+s) = j|X(0) = i\} \\
&= \sum_{k=0}^{\infty} P\{X(t+s) = j, X(t) = k|X(0) = i\} \\
&= \sum_{k=0}^{\infty} P\{X(t+s) = j|X(t) = k, X(0) = i\} \cdot P\{X(t) = k|X(0) = i\} \\
&= \sum_{k=0}^{\infty} P\{X(t+s) = j|X(t) = k\} \cdot P\{X(t) = k|X(0) = i\} \\
&= \sum_{k=0}^{\infty} P_{kj}(s)P_{ik}(t)
\end{aligned}$$

and the proof is completed. ■

The set of Equations (6.8) is known as the *Chapman–Kolmogorov* equations. From Lemma 6.3, we obtain

$$\begin{aligned}
P_{ij}(h+t) - P_{ij}(t) &= \sum_{k=0}^{\infty} P_{ik}(h)P_{kj}(t) - P_{ij}(t) \\
&= \sum_{k \neq i} P_{ik}(h)P_{kj}(t) - [1 - P_{ii}(h)]P_{ij}(t)
\end{aligned}$$

and thus

$$\lim_{h \to 0} \frac{P_{ij}(t+h) - P_{ij}(t)}{h} = \lim_{h \to 0} \left\{ \sum_{k \neq i} \frac{P_{ik}(h)}{h} P_{kj}(t) - \left[\frac{1 - P_{ii}(h)}{h} \right] P_{ij}(t) \right\}$$

Now, assuming that we can interchange the limit and the summation in the preceding and applying Lemma 6.2, we obtain

$$P'_{ij}(t) = \sum_{k \neq i} q_{ik} P_{kj}(t) - v_i P_{ij}(t)$$

It turns out that this interchange can indeed be justified and, hence, we have the following theorem.

Theorem 6.1 (Kolmogorov's Backward Equations) For all states i, j, and times $t \geqslant 0$,

$$P'_{ij}(t) = \sum_{k \neq i} q_{ik} P_{kj}(t) - v_i P_{ij}(t)$$

Example 6.9 The backward equations for the pure birth process become

$$P'_{ij}(t) = \lambda_i P_{i+1,j}(t) - \lambda_i P_{ij}(t)$$

■

Example 6.10 The backward equations for the birth and death process become

$$P'_{0j}(t) = \lambda_0 P_{1j}(t) - \lambda_0 P_{0j}(t),$$
$$P'_{ij}(t) = (\lambda_i + \mu_i)\left[\frac{\lambda_i}{\lambda_i + \mu_i} P_{i+1,j}(t) + \frac{\mu_i}{\lambda_i + \mu_i} P_{i-1,j}(t)\right] - (\lambda_i + \mu_i) P_{ij}(t)$$

or equivalently,

$$P'_{0j}(t) = \lambda_0 [P_{1j}(t) - P_{0j}(t)],$$
$$P'_{ij}(t) = \lambda_i P_{i+1,j}(t) + \mu_i P_{i-1,j}(t) - (\lambda_i + \mu_i) P_{ij}(t), \qquad i > 0 \tag{6.9}$$

■

Example 6.11 (A Continuous-Time Markov Chain Consisting of Two States) Consider a machine that works for an exponential amount of time having mean $1/\lambda$ before breaking down; and suppose that it takes an exponential amount of time having mean $1/\mu$ to repair the machine. If the machine is in working condition at time 0, then what is the probability that it will be working at time $t = 10$?

To answer this question, we note that the process is a birth and death process (with state 0 meaning that the machine is working and state 1 that it is being repaired) having parameters

$$\lambda_0 = \lambda, \qquad \mu_1 = \mu,$$
$$\lambda_i = 0, \quad i \neq 0, \qquad \mu_i = 0, \quad i \neq 1$$

We shall derive the desired probability, namely, $P_{00}(10)$ by solving the set of differential equations given in Example 6.10. From Equation (6.9), we obtain

$$P'_{00}(t) = \lambda[P_{10}(t) - P_{00}(t)], \tag{6.10}$$

$$P'_{10}(t) = \mu P_{00}(t) - \mu P_{10}(t) \tag{6.11}$$

Multiplying Equation (6.10) by μ and Equation (6.11) by λ and then adding the two equations yields

$$\mu P'_{00}(t) + \lambda P'_{10}(t) = 0$$

By integrating, we obtain

$$\mu P_{00}(t) + \lambda P_{10}(t) = c$$

However, since $P_{00}(0) = 1$ and $P_{10}(0) = 0$, we obtain $c = \mu$ and hence,

$$\mu P_{00}(t) + \lambda P_{10}(t) = \mu \tag{6.12}$$

or equivalently,

$$\lambda P_{10}(t) = \mu[1 - P_{00}(t)]$$

By substituting this result in Equation (6.10), we obtain

$$\begin{aligned} P'_{00}(t) &= \mu[1 - P_{00}(t)] - \lambda P_{00}(t) \\ &= \mu - (\mu + \lambda)P_{00}(t) \end{aligned}$$

Letting

$$h(t) = P_{00}(t) - \frac{\mu}{\mu + \lambda}$$

we have

$$\begin{aligned} h'(t) &= \mu - (\mu + \lambda)\left[h(t) + \frac{\mu}{\mu + \lambda}\right] \\ &= -(\mu + \lambda)h(t) \end{aligned}$$

or

$$\frac{h'(t)}{h(t)} = -(\mu + \lambda)$$

By integrating both sides, we obtain

$$\log h(t) = -(\mu + \lambda)t + c$$

or

$$h(t) = Ke^{-(\mu+\lambda)t}$$

and thus

$$P_{00}(t) = Ke^{-(\mu+\lambda)t} + \frac{\mu}{\mu + \lambda}$$

which finally yields, by setting $t = 0$ and using the fact that $P_{00}(0) = 1$,

$$P_{00}(t) = \frac{\lambda}{\mu + \lambda}e^{-(\mu+\lambda)t} + \frac{\mu}{\mu + \lambda}$$

From Equation (6.12), this also implies that

$$P_{10}(t) = \frac{\mu}{\mu + \lambda} - \frac{\mu}{\mu + \lambda}e^{-(\mu+\lambda)t}$$

Hence, our desired probability is as follows:

$$P_{00}(10) = \frac{\lambda}{\mu + \lambda}e^{-10(\mu+\lambda)} + \frac{\mu}{\mu + \lambda}$$ ■

Another set of differential equations, different from the backward equations, may also be derived. This set of equations, known as *Kolmogorov's forward equations* is derived as follows. From the Chapman–Kolmogorov equations (Lemma 6.3), we have

$$\begin{aligned} P_{ij}(t + h) - P_{ij}(t) &= \sum_{k=0}^{\infty} P_{ik}(t)P_{kj}(h) - P_{ij}(t) \\ &= \sum_{k \neq j} P_{ik}(t)P_{kj}(h) - [1 - P_{jj}(h)]P_{ij}(t) \end{aligned}$$

and thus

$$\lim_{h \to 0} \frac{P_{ij}(t + h) - P_{ij}(t)}{h} = \lim_{h \to 0} \left\{ \sum_{k \neq j} P_{ik}(t)\frac{P_{kj}(h)}{h} - \left[\frac{1 - P_{jj}(h)}{h}\right] P_{ij}(t) \right\}$$

and, assuming that we can interchange limit with summation, we obtain from Lemma 6.2

$$P'_{ij}(t) = \sum_{k \neq j} q_{kj} P_{ik}(t) - v_j P_{ij}(t)$$

Unfortunately, we cannot always justify the interchange of limit and summation and thus the preceding is not always valid. However, they do hold in most models, including all birth and death processes and all finite state models. We thus have the following.

Theorem 6.2 (Kolmogorov's Forward Equations) Under suitable regularity conditions,

$$P'_{ij}(t) = \sum_{k \neq j} q_{kj} P_{ik}(t) - v_j P_{ij}(t) \tag{6.13}$$

We shall now solve the forward equations for the pure birth process. For this process, Equation (6.13) reduces to

$$P'_{ij}(t) = \lambda_{j-1} P_{i,j-1}(t) - \lambda_j P_{ij}(t)$$

However, by noting that $P_{ij}(t) = 0$ whenever $j < i$ (since no deaths can occur), we can rewrite the preceding equation to obtain

$$\begin{aligned} P'_{ii}(t) &= -\lambda_i P_{ii}(t), \\ P'_{ij}(t) &= \lambda_{j-1} P_{i,j-1}(t) - \lambda_j P_{ij}(t), \qquad j \geqslant i+1 \end{aligned} \tag{6.14}$$

Proposition 6.4 For a pure birth process,

$$\begin{aligned} P_{ii}(t) &= e^{-\lambda_i t}, & i \geqslant 0 \\ P_{ij}(t) &= \lambda_{j-1} e^{-\lambda_j t} \int_0^t e^{\lambda_j s} P_{i,j-1}(s)\, ds, & j \geqslant i+1 \end{aligned}$$

Proof. The fact that $P_{ii}(t) = e^{-\lambda_i t}$ follows from Equation (6.14) by integrating and using the fact that $P_{ii}(0) = 1$. To prove the corresponding result for $P_{ij}(t)$, we note by Equation (6.14) that

$$e^{\lambda_j t} \left[P'_{ij}(t) + \lambda_j P_{ij}(t) \right] = e^{\lambda_j t} \lambda_{j-1} P_{i,j-1}(t)$$

or

$$\frac{d}{dt} \left[e^{\lambda_j t} P_{ij}(t) \right] = \lambda_{j-1} e^{\lambda_j t} P_{i,j-1}(t)$$

Hence, since $P_{ij}(0) = 0$, we obtain the desired results. ■

Example 6.12 (Forward Equations for Birth and Death Process) The forward equations (Equation 6.13) for the general birth and death process become

$$\begin{aligned} P'_{i0}(t) &= \sum_{k \neq 0} q_{k0} P_{ik}(t) - \lambda_0 P_{i0}(t) \\ &= \mu_1 P_{i1}(t) - \lambda_0 P_{i0}(t) \end{aligned} \tag{6.15}$$

$$\begin{aligned} P'_{ij}(t) &= \sum_{k \neq j} q_{kj} P_{ik}(t) - (\lambda_j + \mu_j) P_{ij}(t) \\ &= \lambda_{j-1} P_{i,j-1}(t) + \mu_{j+1} P_{i,j+1}(t) - (\lambda_j + \mu_j) P_{ij}(t) \end{aligned} \tag{6.16}$$

■

6.5 Limiting Probabilities

In analogy with a basic result in discrete-time Markov chains, the probability that a continuous-time Markov chain will be in state j at time t often converges to a limiting value that is independent of the initial state. That is, if we call this value P_j, then

$$P_j \equiv \lim_{t \to \infty} P_{ij}(t)$$

where we are assuming that the limit exists and is independent of the initial state i.

To derive a set of equations for the P_j, consider first the set of forward equations

$$P'_{ij}(t) = \sum_{k \neq j} q_{kj} P_{ik}(t) - v_j P_{ij}(t) \tag{6.17}$$

Now, if we let t approach ∞, then assuming that we can interchange limit and summation, we obtain

$$\begin{aligned} \lim_{t \to \infty} P'_{ij}(t) &= \lim_{t \to \infty} \left[\sum_{k \neq j} q_{kj} P_{ik}(t) - v_j P_{ij}(t) \right] \\ &= \sum_{k \neq j} q_{kj} P_k - v_j P_j \end{aligned}$$

However, as $P_{ij}(t)$ is a bounded function (being a probability it is always between 0 and 1), it follows that if $P'_{ij}(t)$ converges, then it must converge to 0 (why is this?). Hence, we must have

$$0 = \sum_{k \neq j} q_{kj} P_k - v_j P_j$$

or

$$v_j P_j = \sum_{k \neq j} q_{kj} P_k, \qquad \text{all states } j \tag{6.18}$$

The preceding set of equations, along with the equation

$$\sum_j P_j = 1 \tag{6.19}$$

can be used to solve for the limiting probabilities.

Remarks

(i) We have assumed that the limiting probabilities P_j exist. A sufficient condition for this is that

(a) all states of the Markov chain communicate in the sense that starting in state i there is a positive probability of ever being in state j, for all i, j and
(b) the Markov chain is positive recurrent in the sense that, starting in any state, the mean time to return to that state is finite

If conditions (a) and (b) hold, then the limiting probabilities will exist and satisfy Equations (6.18) and (6.19). In addition, P_j also will have the interpretation of being the long-run proportion of time that the process is in state j.

(ii) Equations (6.18) and (6.19) have a nice interpretation: In any interval $(0, t)$ the number of transitions into state j must equal to within 1 the number of transitions out of state j (why?). Hence, in the long run, the rate at which transitions into state j occur must equal the rate at which transitions out of state j occur. When the process is in state j, it leaves at rate v_j, and, as P_j is the proportion of time it is in state j, it thus follows that

$$v_j P_j = \text{rate at which the process leaves state } j$$

Similarly, when the process is in state k, it enters j at a rate q_{kj}. Hence, as P_k is the proportion of time in state k, we see that the rate at which transitions from k to j occur is just $q_{kj} P_k$; thus

$$\sum_{k \neq j} q_{kj} P_k = \text{rate at which the process enters state } j$$

So, Equation (6.18) is just a statement of the equality of the rates at which the process enters and leaves state j. Because it balances (that is, equates) these rates, Equation (6.18) is sometimes referred to as a set of "balance equations."

(iii) When the limiting probabilities P_j exist, we say that the chain is ergodic. The P_j are sometimes called stationary probabilities since it can be shown that (as in the discrete-time case) if the initial state is chosen according to the distribution $\{P_j\}$, then the probability of being in state j at time t is P_j, for all t.

Let us now determine the limiting probabilities for a birth and death process. From Equation (6.18) or equivalently, by equating the rate at which the process leaves a state with the rate at which it enters that state, we obtain

State	*Rate at which leave = rate at which enter*
0	$\lambda_0 P_0 = \mu_1 P_1$
1	$(\lambda_1 + \mu_1)P_1 = \mu_2 P_2 + \lambda_0 P_0$
2	$(\lambda_2 + \mu_2)P_2 = \mu_3 P_3 + \lambda_1 P_1$
$n, n \geqslant 1$	$(\lambda_n + \mu_n)P_n = \mu_{n+1} P_{n+1} + \lambda_{n-1} P_{n-1}$

By adding to each equation the equation preceding it, we obtain

$$
\begin{aligned}
\lambda_0 P_0 &= \mu_1 P_1, \\
\lambda_1 P_1 &= \mu_2 P_2, \\
\lambda_2 P_2 &= \mu_3 P_3, \\
&\;\vdots \\
\lambda_n P_n &= \mu_{n+1} P_{n+1}, \qquad n \geqslant 0
\end{aligned}
$$

Solving in terms of P_0 yields

$$
\begin{aligned}
P_1 &= \frac{\lambda_0}{\mu_1} P_0, \\
P_2 &= \frac{\lambda_1}{\mu_2} P_1 = \frac{\lambda_1 \lambda_0}{\mu_2 \mu_1} P_0, \\
P_3 &= \frac{\lambda_2}{\mu_3} P_2 = \frac{\lambda_2 \lambda_1 \lambda_0}{\mu_3 \mu_2 \mu_1} P_0, \\
&\;\vdots \\
P_n &= \frac{\lambda_{n-1}}{\mu_n} P_{n-1} = \frac{\lambda_{n-1} \lambda_{n-2} \cdots \lambda_1 \lambda_0}{\mu_n \mu_{n-1} \cdots \mu_2 \mu_1} P_0
\end{aligned}
$$

And by using the fact that $\sum_{n=0}^{\infty} P_n = 1$, we obtain

$$
1 = P_0 + P_0 \sum_{n=1}^{\infty} \frac{\lambda_{n-1} \cdots \lambda_1 \lambda_0}{\mu_n \cdots \mu_2 \mu_1}
$$

or

$$P_0 = \frac{1}{1 + \sum_{n=1}^{\infty} \frac{\lambda_0 \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \cdots \mu_n}}$$

and so

$$P_n = \frac{\lambda_0 \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \cdots \mu_n \left(1 + \sum_{n=1}^{\infty} \frac{\lambda_0 \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \cdots \mu_n}\right)}, \qquad n \geqslant 1 \tag{6.20}$$

The foregoing equations also show us what condition is necessary for these limiting probabilities to exist. Namely, it is necessary that

$$\sum_{n=1}^{\infty} \frac{\lambda_0 \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \cdots \mu_n} < \infty \tag{6.21}$$

This condition also may be shown to be sufficient.

In the multiserver exponential queueing system (Example 6.6), Condition (6.21) reduces to

$$\sum_{n=s+1}^{\infty} \frac{\lambda^n}{(s\mu)^n} < \infty$$

which is equivalent to $\lambda/s\mu < 1$.

For the linear growth model with immigration (Example 6.4), Condition (6.21) reduces to

$$\sum_{n=1}^{\infty} \frac{\theta(\theta + \lambda) \cdots (\theta + (n-1)\lambda)}{n! \mu^n} < \infty$$

Using the ratio test, the preceding will converge when

$$\begin{aligned} \lim_{n\to\infty} \frac{\theta(\theta + \lambda) \cdots (\theta + n\lambda)}{(n+1)! \mu^{n+1}} \frac{n! \mu^n}{\theta(\theta + \lambda) \cdots (\theta + (n-1)\lambda)} &= \lim_{n\to\infty} \frac{\theta + n\lambda}{(n+1)\mu} \\ &= \frac{\lambda}{\mu} < 1 \end{aligned}$$

That is, the condition is satisfied when $\lambda < \mu$. When $\lambda \geqslant \mu$ it is easy to show that Condition (6.21) is not satisfied.

Example 6.13 (A Machine Repair Model) Consider a job shop that consists of M machines and one serviceman. Suppose that the amount of time each machine runs before breaking down is exponentially distributed with mean $1/\lambda$, and suppose that the amount of time that it takes for the serviceman to fix a machine is exponentially distributed with mean $1/\mu$. We shall attempt to answer these questions: (a) What is the average number of machines not in use? (b) What proportion of time is each machine in use?

Solution: If we say that the system is in state n whenever n machines are not in use, then the preceding is a birth and death process having parameters

$$\mu_n = \mu \qquad n \geqslant 1$$
$$\lambda_n = \begin{cases} (M-n)\lambda, & n \leqslant M \\ 0, & n > M \end{cases}$$

This is so in the sense that a failing machine is regarded as an arrival and a fixed machine as a departure. If any machines are broken down, then since the serviceman's rate is $\mu, \mu_n = \mu$. On the other hand, if n machines are not in use, then since the $M - n$ machines in use each fail at a rate λ, it follows that $\lambda_n = (M - n)\lambda$. From Equation (6.20) we have that P_n, the probability that n machines will not be in use, is given by

$$\begin{aligned} P_0 &= \frac{1}{1 + \sum_{n=1}^{M} [M\lambda(M-1)\lambda \cdots (M-n+1)\lambda/\mu^n]} \\ &= \frac{1}{1 + \sum_{n=1}^{M} (\lambda/\mu)^n M!/(M-n)!}, \\ P_n &= \frac{(\lambda/\mu)^n M!/(M-n)!}{1 + \sum_{n=1}^{M} (\lambda/\mu)^n M!/(M-n)!}, \qquad n = 0, 1, \ldots, M \end{aligned}$$

Hence, the average number of machines not in use is given by

$$\sum_{n=0}^{M} nP_n = \frac{\sum_{n=0}^{M} n(\lambda/\mu)^n M!/(M-n)!}{1 + \sum_{n=1}^{M} (\lambda/\mu)^n M!/(M-n)!} \tag{6.22}$$

To obtain the long-run proportion of time that a given machine is working we will compute the equivalent limiting probability of the machine working. To do so, we condition on the number of machines that are not working to obtain

$$\begin{aligned} P\{\text{machine is working}\} &= \sum_{n=0}^{M} P\{\text{machine is working}|n \text{ not working}\}P_n \\ &= \sum_{n=0}^{M} \frac{M-n}{M} P_n \qquad \text{(since if } n \text{ are not working, then } M-n \text{ are working!)} \end{aligned}$$

$$= 1 - \sum_0^M \frac{nP_n}{M}$$

where $\sum_0^M nP_n$ is given by Equation (6.22). ■

Example 6.14 (The *M/M/1* Queue) In the $M/M/1$ queue $\lambda_n = \lambda$, $\mu_n = \mu$ and thus, from Equation (6.20),

$$\begin{aligned} P_n &= \frac{(\lambda/\mu)^n}{1 + \sum_{n=1}^{\infty}(\lambda/\mu)^n} \\ &= (\lambda/\mu)^n(1 - \lambda/\mu), \qquad n \geqslant 0 \end{aligned}$$

provided that $\lambda/\mu < 1$. It is intuitive that λ must be less than μ for limiting probabilities to exist. Customers arrive at rate λ and are served at rate μ, and thus if $\lambda > \mu$, then they arrive at a faster rate than they can be served and the queue size will go to infinity. The case $\lambda = \mu$ behaves much like the symmetric random walk of Section 4.3, which is null recurrent and thus has no limiting probabilities. ■

Example 6.15 Let us reconsider the shoe shine shop of Example 6.1, and determine the proportion of time the process is in each of the states 0, 1, 2. Because this is not a birth and death process (since the process can go directly from state 2 to state 0), we start with the balance equations for the limiting probabilities.

State	*Rate that the process leaves = rate that the process enters*
0	$\lambda P_0 = \mu_2 P_2$
1	$\mu_1 P_1 = \lambda P_0$
2	$\mu_2 P_2 = \mu_1 P_1$

Solving in terms of P_0 yields

$$P_2 = \frac{\lambda}{\mu_2} P_0, \qquad P_1 = \frac{\lambda}{\mu_1} P_0$$

which implies, since $P_0 + P_1 + P_2 = 1$, that

$$P_0 \left[1 + \frac{\lambda}{\mu_2} + \frac{\lambda}{\mu_1} \right] = 1$$

or

$$P_0 = \frac{\mu_1 \mu_2}{\mu_1 \mu_2 + \lambda(\mu_1 + \mu_2)}$$

and

$$P_1 = \frac{\lambda\mu_2}{\mu_1\mu_2 + \lambda(\mu_1 + \mu_2)},$$

$$P_2 = \frac{\lambda\mu_1}{\mu_1\mu_2 + \lambda(\mu_1 + \mu_2)}$$ ■

Example 6.16 Consider a set of n components along with a single repairman. Suppose that component i functions for an exponentially distributed time with rate λ_i and then fails. The time it then takes to repair component i is exponential with rate $\mu_i, i = 1, \ldots, n$. Suppose that when there is more than one failed component the repairman always works on the most recent failure. For instance, if there are at present two failed components—say, components 1 and 2 of which 1 has failed most recently—then the repairman will be working on component 1. However, if component 3 should fail before 1's repair is completed, then the repairman would stop working on component 1 and switch to component 3 (that is, a newly failed component preempts service).

To analyze the preceding as a continuous-time Markov chain, the state must represent the set of failed components in the order of failure. That is, the state will be $i_1, i_2, \ldots, i_k$ if $i_1, i_2, \ldots, i_k$ are the k failed components (all the other $n-k$ being functional) with i_1 having been the most recent failure (and is thus presently being repaired), i_2 the second most recent, and so on. Because there are $k!$ possible orderings for a fixed set of k failed components and $\binom{n}{k}$ choices of that set, it follows that there are

$$\sum_{k=0}^{n} \binom{n}{k} k! = \sum_{k=0}^{n} \frac{n!}{(n-k)!} = n! \sum_{i=0}^{n} \frac{1}{i!}$$

possible states.

The balance equations for the limiting probabilities are as follows:

$$\left(\mu_{i_1} + \sum_{\substack{i \neq i_j \\ j=1,\ldots,k}} \lambda_i\right) P(i_1, \ldots, i_k) = \sum_{\substack{i \neq i_j \\ j=1,\ldots,k}} P(i, i_1, \ldots, i_k)\mu_i + P(i_2, \ldots, i_k)\lambda_{i_1},$$

$$\sum_{i=1}^{n} \lambda_i P(\phi) = \sum_{i=1}^{n} P(i)\mu_i \tag{6.23}$$

where ϕ is the state when all components are working. The preceding equations follow because state $i_1, \ldots, i_k$ can be left either by a failure of any of the additional components or by a repair completion of component i_1. Also, that state can be entered either by a repair completion of component i when the state is $i, i_1, \ldots, i_k$ or by a failure of component i_1 when the state is $i_2, \ldots, i_k$.

However, if we take

$$P(i_1, \ldots, i_k) = \frac{\lambda_{i_1} \lambda_{i_2} \cdots \lambda_{i_k}}{\mu_{i_1} \mu_{i_2} \cdots \mu_{i_k}} P(\phi) \tag{6.24}$$

then it is easily seen that Equations (6.23) are satisfied. Hence, by uniqueness these must be the limiting probabilities with $P(\phi)$ determined to make their sum equal 1. That is,

$$P(\phi) = \left[1 + \sum_{i_1, \ldots, i_k} \frac{\lambda_{i_1} \cdots \lambda_{i_k}}{\mu_{i_1} \cdots \mu_{i_k}}\right]^{-1}$$

As an illustration, suppose $n = 2$ and so there are five states ϕ, 1, 2, 12, 21. Then from the preceding we would have

$$P(\phi) = \left[1 + \frac{\lambda_1}{\mu_1} + \frac{\lambda_2}{\mu_2} + \frac{2\lambda_1\lambda_2}{\mu_1\mu_2}\right]^{-1},$$

$$P(1) = \frac{\lambda_1}{\mu_1} P(\phi),$$

$$P(2) = \frac{\lambda_2}{\mu_2} P(\phi),$$

$$P(1, 2) = P(2, 1) = \frac{\lambda_1\lambda_2}{\mu_1\mu_2} P(\phi)$$

It is interesting to note, using Equation (6.24), that given the set of failed components, each of the possible orderings of these components is equally likely. ■

6.6 Time Reversibility

Consider a continuous-time Markov chain that is ergodic and let us consider the limiting probabilities P_i from a different point of view than previously. If we consider the sequence of states visited, ignoring the amount of time spent in each state during a visit, then this sequence constitutes a discrete-time Markov chain with transition probabilities P_{ij}. Let us assume that this discrete-time Markov chain, called the embedded chain, is ergodic and denote by π_i its limiting probabilities. That is, the π_i are the unique solution of

$$\pi_i = \sum_j \pi_j P_{ji}, \qquad \text{all} \quad i$$

$$\sum_i \pi_i = 1$$

Now, since π_i represents the proportion of transitions that take the process into state i, and because $1/v_i$ is the mean time spent in state i during a visit, it seems intuitive that P_i, the proportion of time in state i, should be a weighted average of the π_i where π_i is weighted proportionately to $1/v_i$. That is, it is intuitive that

$$P_i = \frac{\pi_i/v_i}{\sum_j \pi_j/v_j} \tag{6.25}$$

To check the preceding, recall that the limiting probabilities P_i must satisfy

$$v_i P_i = \sum_{j \neq i} P_j q_{ji}, \qquad \text{all } i$$

or equivalently, since $P_{ii} = 0$

$$v_i P_i = \sum_j P_j v_j P_{ji}, \qquad \text{all } i$$

Hence, for the P_is to be given by Equation (6.25), the following would be necessary:

$$\pi_i = \sum_j \pi_j P_{ji}, \qquad \text{all } i$$

But this, of course, follows since it is in fact the defining equation for the π_is.

Suppose now that the continuous-time Markov chain has been in operation for a long time, and suppose that starting at some (large) time T we trace the process going backward in time. To determine the probability structure of this reversed process, we first note that given we are in state i at some time—say, t—the probability that we have been in this state for an amount of time greater than s is just $e^{-v_i s}$. This is so, since

$$\begin{aligned}
&P\{\text{process is in state } i \text{ throughout } [t-s,t] | X(t) = i\} \\
&\quad = \frac{P\{\text{process is in state } i \text{ throughout } [t-s,t]\}}{P\{X(t)=i\}} \\
&\quad = \frac{P\{X(t-s)=i\}e^{-v_i s}}{P\{X(t)=i\}} \\
&\quad = e^{-v_i s}
\end{aligned}$$

since for t large $P\{X(t-s) = i\} = P\{X(t) = i\} = P_i$.

In other words, going backward in time, the amount of time the process spends in state i is also exponentially distributed with rate v_i. In addition, as was shown in Section 4.8, the sequence of states visited by the reversed process constitutes a discrete-time Markov chain with transition probabilities Q_{ij} given by

$$Q_{ij} = \frac{\pi_j P_{ji}}{\pi_i}$$

Hence, we see from the preceding that the reversed process is a continuous-time Markov chain with the same transition rates as the forward-time process and with one-stage transition probabilities Q_{ij}. Therefore, the continuous-time Markov chain will be *time reversible*, in the sense that the process reversed in time has the same probabilistic structure as the original process, if the embedded chain is time reversible. That is, if

$$\pi_i P_{ij} = \pi_j P_{ji}, \qquad \text{for all } i, j$$

Now, using the fact that $P_i = (\pi_i/v_i) \big/ \left(\sum_j \pi_j/v_j\right)$, we see that the preceding condition is equivalent to

$$P_i q_{ij} = P_j q_{ji}, \qquad \text{for all } i, j \tag{6.26}$$

Since P_i is the proportion of time in state i and q_{ij} is the rate when in state i that the process goes to j, the condition of time reversibility is that *the rate at which the process goes directly from state i to state j is equal to the rate at which it goes directly from j to i*. It should be noted that this is exactly the same condition needed for an ergodic discrete-time Markov chain to be time reversible (see Section 4.8).

An application of the preceding condition for time reversibility yields the following proposition concerning birth and death processes.

Proposition 6.5 An ergodic birth and death process is time reversible.

Proof. We must show that the rate at which a birth and death process goes from state i to state $i + 1$ is equal to the rate at which it goes from $i + 1$ to i. In any length of time t the number of transitions from i to $i + 1$ must equal to within 1 the number from $i + 1$ to i (since between each transition from i to $i + 1$ the process must return to i, and this can only occur through $i + 1$, and vice versa). Hence, as the number of such transitions goes to infinity as $t \to \infty$, it follows that the rate of transitions from i to $i + 1$ equals the rate from $i + 1$ to i. ■

Proposition 6.5 can be used to prove the important result that the output process of an $M/M/s$ queue is a Poisson process. We state this as a corollary.

Corollary 6.6 Consider an $M/M/s$ queue in which customers arrive in accordance with a Poisson process having rate λ and are served by any one of s servers—each having an exponentially distributed service time with rate μ. If $\lambda < s\mu$, then the output process of customers departing is, after the process has been in operation for a long time, a Poisson process with rate λ.

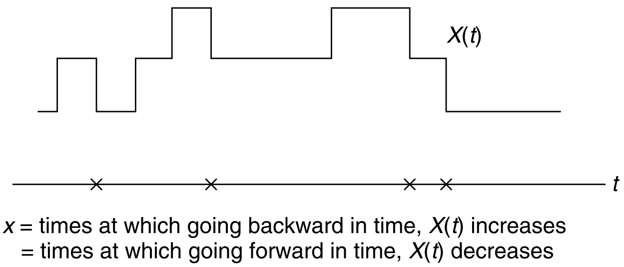

Figure 6.1 The number in the system.

Proof. Let $X(t)$ denote the number of customers in the system at time t. Since the $M/M/s$ process is a birth and death process, it follows from Proposition 6.5 that $\{X(t), t \geqslant 0\}$ is time reversible. Going forward in time, the time points at which $X(t)$ increases by 1 constitute a Poisson process since these are just the arrival times of customers. Hence, by time reversibility the time points at which $X(t)$ increases by 1 when we go backward in time also constitute a Poisson process. But these latter points are exactly the points of time when customers depart (see Figure 6.1). Hence, the departure times constitute a Poisson process with rate λ. ■

Example 6.17 Consider a first come first serve $M/M/1$ queue, with arrival rate λ and service rate μ, where $\lambda < \mu$, that is in steady state. Given that customer C spends a total of t time units in the system, what is the conditional distribution of the number of others that were present when C arrived?

Solution: Suppose that C arrived at time s and departed at time $s + t$. Because the system is first come first served, the number that were in the system when C arrived is equal to the number of departures of other customers that occur after time s and before time $s + t$, which is equal to the number of arrivals in the reversed process in that interval of time. Now, in the reversed process C would have arrived at time $s + t$ and departed at time s. Because the reversed process is also an $M/M/1$ queueing system, the number of arrivals during that interval of length t is Poisson distributed with mean λt. (For a more direct argument for this result, see Section 8.3.1.) ■

We have shown that a process is time reversible if and only if

$$P_i q_{ij} = P_j q_{ji} \qquad \text{for all } i \neq j$$

Analogous to the result for discrete-time Markov chains, if we can find a probability vector $\mathbf{P}$ that satisfies the preceding then the Markov chain is time reversible and the P_is are the long-run probabilities. That is, we have the following proposition.

Proposition 6.7 If for some set $\{P_i\}$

$$\sum_i P_i = 1, \qquad P_i \geqslant 0$$

and

$$P_i q_{ij} = P_j q_{ji} \qquad \text{for all } \; i \neq j \tag{6.27}$$

then the continuous-time Markov chain is time reversible and P_i represents the limiting probability of being in state i.

Proof. For fixed i we obtain upon summing Equation (6.27) over all $j : j \neq i$

$$\sum_{j \neq i} P_i q_{ij} = \sum_{j \neq i} P_j q_{ji}$$

or, since $\sum_{j \neq i} q_{ij} = v_i$,

$$v_i P_i = \sum_{j \neq i} P_j q_{ji}$$

Hence, the P_is satisfy the balance equations and thus represent the limiting probabilities. Because Equation (6.27) holds, the chain is time reversible. ■

Example 6.18 Consider a set of n machines and a single repair facility to service them. Suppose that when machine $i, i = 1, \ldots, n$, goes down it requires an exponentially distributed amount of work with rate μ_i to get it back up. The repair facility divides its efforts equally among all down components in the sense that whenever there are k down machines $1 \leqslant k \leqslant n$ each receives work at a rate of $1/k$ per unit time. Finally, suppose that each time machine i goes back up it remains up for an exponentially distributed time with rate λ_i.

The preceding can be analyzed as a continuous-time Markov chain having 2^n states where the state at any time corresponds to the set of machines that are down at that time. Thus, for instance, the state will be $(i_1, i_2, \ldots, i_k)$ when machines $i_1, \ldots, i_k$ are down and all the others are up. The instantaneous transition rates are as follows:

$$q_{(i_1,\ldots,i_{k-1}),(i_1,\ldots,i_k)} = \lambda_{i_k},$$
$$q_{(i_1,\ldots,i_k),(i_1,\ldots,i_{k-1})} = \mu_{i_k}/k$$

where $i_1, \ldots, i_k$ are all distinct. This follows since the failure rate of machine i_k is always λ_{i_k} and the repair rate of machine i_k when there are k failed machines is μ_{i_k}/k.

Hence, the time reversible equations from (6.27) are

$$P(i_1,\ldots,i_k)\mu_{i_k}/k = P(i_1,\ldots,i_{k-1})\lambda_{i_k}$$

or

$$\begin{aligned}
P(i_1,\ldots,i_k) &= \frac{k\lambda_{i_k}}{\mu_{i_k}}P(i_1,\ldots,i_{k-1}) \\
&= \frac{k\lambda_{i_k}}{\mu_{i_k}}\frac{(k-1)\lambda_{i_{k-1}}}{\mu_{i_{k-1}}}P(i_1,\ldots,i_{k-2}) \qquad \text{upon iterating} \\
&= \\
&\;\vdots \\
&= k!\prod_{j=1}^{k}(\lambda_{i_j}/\mu_{i_j})P(\phi)
\end{aligned}$$

where ϕ is the state in which all components are working. Because

$$P(\phi) + \sum P(i_1,\ldots,i_k) = 1$$

we see that

$$P(\phi) = \left[1 + \sum_{i_1,\ldots,i_k} k!\prod_{j=1}^{k}(\lambda_{i_j}/\mu_{i_j})\right]^{-1} \tag{6.28}$$

where the preceding sum is over all the $2^n - 1$ nonempty subsets $\{i_1,\ldots,i_k\}$ of $\{1,2,\ldots,n\}$. Hence, as the time reversible equations are satisfied for this choice of probability vector it follows from Proposition 6.7 that the chain is time reversible and

$$P(i_1,\ldots,i_k) = k!\prod_{j=1}^{k}(\lambda_{i_j}/\mu_{i_j})P(\phi)$$

with $P(\phi)$ being given by (6.28).

For instance, suppose there are two machines. Then, from the preceding we would have

$$P(\phi) = \frac{1}{1 + \lambda_1/\mu_1 + \lambda_2/\mu_2 + 2\lambda_1\lambda_2/\mu_1\mu_2},$$

$$P(1) = \frac{\lambda_1/\mu_1}{1 + \lambda_1/\mu_1 + \lambda_2/\mu_2 + 2\lambda_1\lambda_2/\mu_1\mu_2},$$
$$P(2) = \frac{\lambda_2/\mu_2}{1 + \lambda_1/\mu_1 + \lambda_2/\mu_2 + 2\lambda_1\lambda_2/\mu_1\mu_2},$$
$$P(1,2) = \frac{2\lambda_1\lambda_2}{\mu_1\mu_2[1 + \lambda_1/\mu_1 + \lambda_2/\mu_2 + 2\lambda_1\lambda_2/\mu_1\mu_2]}$$ ■

Consider a continuous-time Markov chain whose state space is S. We say that the Markov chain is truncated to the set $A \subset S$ if q_{ij} is changed to 0 for all $i \in A$, $j \notin A$. That is, transitions out of the class A are no longer allowed, whereas ones in A continue at the same rates as before. A useful result is that if the chain is time reversible, then so is the truncated one.

Proposition 6.8 A time reversible chain with limiting probabilities $P_j, j \in S$ that is truncated to the set $A \subset S$ and remains irreducible is also time reversible and has limiting probabilities P_j^A given by

$$P_j^A = \frac{P_j}{\sum_{i \in A} P_i}, \qquad j \in A$$

Proof. By Proposition 6.7 we need to show that, with P_j^A as given,

$$P_i^A q_{ij} = P_j^A q_{ji} \qquad \text{for } \ i \in A, \ j \in A$$

or, equivalently,

$$P_i q_{ij} = P_j q_{ji} \qquad \text{for } i \in A, \ j \in A$$

But this follows since the original chain is, by assumption, time reversible. ■

Example 6.19 Consider an $M/M/1$ queue in which arrivals finding N in the system do not enter. This finite capacity system can be regarded as a truncation of the $M/M/1$ queue to the set of states $A = \{0, 1, \ldots, N\}$. Since the number in the system in the $M/M/1$ queue is time reversible and has limiting probabilities $P_j = (\lambda/\mu)^j(1 - \lambda/\mu)$ it follows from Proposition 6.8 that the finite capacity model is also time reversible and has limiting probabilities given by

$$P_j = \frac{(\lambda/\mu)^j}{\sum_{i=0}^{N} (\lambda/\mu)^i}, \qquad j = 0, 1, \ldots, N$$ ■

Another useful result is given by the following proposition, whose proof is left as an exercise.

Proposition 6.9 If $\{X_i(t),\ t \geqslant 0\}$ are, for $i = 1, \ldots, n$, independent time reversible continuous-time Markov chains, then the vector process $\{(X_i(t), \ldots, X_n(t)),\ t \geqslant 0\}$ is also a time reversible continuous-time Markov chain.

Example 6.20 Consider an n-component system where component $i, i = 1, \ldots, n$, functions for an exponential time with rate λ_i and then fails; upon failure, repair begins on component i, with the repair taking an exponentially distributed time with rate μ_i. Once repaired, a component is as good as new. The components act independently except that when there is only one working component the system is temporarily shut down until a repair has been completed; it then starts up again with two working components.

(a) What proportion of time is the system shut down?
(b) What is the (limiting) averaging number of components that are being repaired?

Solution: Consider first the system without the restriction that it is shut down when a single component is working. Letting $X_i(t),\ i = 1, \ldots, n$, equal 1 if component i is working at time t and 0 if it failed, then $\{X_i(t),\ t \geqslant 0\}$, $i = 1, \ldots, n$, are independent birth and death processes. Because a birth and death process is time reversible, it follows from Proposition 6.9 that the process $\{(X_1(t), \ldots, X_n(t)),\ t \geqslant 0\}$ is also time reversible. Now, with

$$P_i(j) = \lim_{t\to\infty} P\{X_i(t) = j\}, \qquad j = 0, 1$$

we have

$$P_i(1) = \frac{\mu_i}{\mu_i + \lambda_i}, \qquad P_i(0) = \frac{\lambda_i}{\mu_i + \lambda_i}$$

Also, with

$$P(j_1, \ldots, j_n) = \lim_{t\to\infty} P\{X_i(t) = j_i, i = 1, \ldots, n\}$$

it follows, by independence, that

$$P(j_1, \ldots, j_n) = \prod_{i=1}^{n} P_i(j_i), \qquad j_i = 0, 1,\ i = 1, \ldots, n$$

Now, note that shutting down the system when only one component is working is equivalent to truncating the preceding unconstrained system to the set consisting of all states except the one having all components down. Therefore, with P_T denoting a probability for the truncated system, we have from

Proposition 6.8 that

$$P_T(j_1, \ldots, j_n) = \frac{P(j_1, \ldots, j_n)}{1 - C}, \qquad \sum_{i=1}^{n} j_i > 0$$

where

$$C = P(0, \ldots, 0) = \prod_{j=1}^{n} \lambda_j/(\mu_j + \lambda_j)$$

Hence, letting $(\mathbf{0}, 1_i) = (0, \ldots, 0, 1, 0, \ldots, 0)$ be the n vector of zeroes and ones whose single 1 is in the ith place, we have

$$\begin{aligned} P_T(\text{system is shut down}) &= \sum_{i=1}^{n} P_T(\mathbf{0}, 1_i) \\ &= \frac{1}{1 - C} \sum_{i=1}^{n} \left(\frac{\mu_i}{\mu_i + \lambda_i}\right) \prod_{j \neq i} \left(\frac{\lambda_j}{\mu_j + \lambda_j}\right) \\ &= \frac{C \sum_{i=1}^{n} \mu_i/\lambda_i}{1 - C} \end{aligned}$$

Let R denote the number of components being repaired. Then with I_i equal to 1 if component i is being repaired and 0 otherwise, we have for the unconstrained (nontruncated) system that

$$E[R] = E\left[\sum_{i=1}^{n} I_i\right] = \sum_{i=1}^{n} P_i(0) = \sum_{i=1}^{n} \lambda_i/(\mu_i + \lambda_i)$$

But, in addition,

$$\begin{aligned} E[R] &= E[R|\text{all components are in repair}]C \\ &\quad + E[R|\text{not all components are in repair}](1 - C) \\ &= nC + E_T[R](1 - C) \end{aligned}$$

implying that

$$E_T[R] = \frac{\sum_{i=1}^{n} \lambda_i/(\mu_i + \lambda_i) - nC}{1 - C} \qquad \blacksquare$$

6.7 Uniformization

Consider a continuous-time Markov chain in which the mean time spent in a state is the same for all states. That is, suppose that $v_i = v$, for all states i. In this case since the amount of time spent in each state during a visit is exponentially distributed with rate v, it follows that if we let $N(t)$ denote the number of state transitions by time t, then $\{N(t), t \geqslant 0\}$ will be a Poisson process with rate v.

To compute the transition probabilities $P_{ij}(t)$, we can condition on $N(t)$:

$$
\begin{aligned}
P_{ij}(t) &= P\{X(t) = j | X(0) = i\} \\
&= \sum_{n=0}^{\infty} P\{X(t) = j | X(0) = i,\ N(t) = n\} P\{N(t) = n | X(0) = i\} \\
&= \sum_{n=0}^{\infty} P\{X(t) = j | X(0) = i,\ N(t) = n\} e^{-vt} \frac{(vt)^n}{n!}
\end{aligned}
$$

Now, the fact that there have been n transitions by time t tells us something about the amount of time spent in each of the first n states visited, but since the distribution of time spent in each state is the same for all states, it follows that knowing that $N(t) = n$ gives us no information about which states were visited. Hence,

$$
P\{X(t) = j | X(0) = i,\ N(t) = n\} = P_{ij}^n
$$

where P_{ij}^n is just the n-stage transition probability associated with the discrete-time Markov chain with transition probabilities P_{ij}; and so when $v_i \equiv v$

$$
P_{ij}(t) = \sum_{n=0}^{\infty} P_{ij}^n e^{-vt} \frac{(vt)^n}{n!} \tag{6.29}
$$

Equation (6.29) is often useful from a computational point of view since it enables us to approximate $P_{ij}(t)$ by taking a partial sum and then computing (by matrix multiplication of the transition probability matrix) the relevant n stage probabilities P_{ij}^n.

Whereas the applicability of Equation (6.29) would appear to be quite limited since it supposes that $v_i \equiv v$, it turns out that most Markov chains can be put in that form by the trick of allowing fictitious transitions from a state to itself. To see how this works, consider any Markov chain for which the v_i are bounded, and let v be any number such that

$$
v_i \leqslant v, \qquad \text{for all } i \tag{6.30}
$$

When in state i, the process actually leaves at rate v_i; but this is equivalent to supposing that transitions occur at rate v, but only the fraction v_i/v of transitions are real ones (and thus real transitions occur at rate v_i) and the remaining fraction $1 - v_i/v$ are fictitious transitions that leave the process in state i. In other words, any Markov chain satisfying Condition (6.30) can be thought of as being a process that spends an exponential amount of time with rate v in state i and then makes a transition to j with probability P^*_{ij}, where

$$P^*_{ij} = \begin{cases} 1 - \frac{v_i}{v}, & j = i \\ \frac{v_i}{v} P_{ij}, & j \neq i \end{cases} \tag{6.31}$$

Hence, from Equation (6.29) we have that the transition probabilities can be computed by

$$P_{ij}(t) = \sum_{n=0}^{\infty} P^{*n}_{ij} e^{-vt} \frac{(vt)^n}{n!}$$

where P^*_{ij} are the n-stage transition probabilities corresponding to Equation (6.31). This technique of uniformizing the rate in which a transition occurs from each state by introducing transitions from a state to itself is known as *uniformization*.

Example 6.21 Let us reconsider Example 6.11, which models the workings of a machine—either on or off—as a two-state continuous-time Markov chain with

$$\begin{aligned} P_{01} &= P_{10} = 1, \\ v_0 = \lambda, &\qquad v_1 = \mu \end{aligned}$$

Letting $v = \lambda + \mu$, the uniformized version of the preceding is to consider it a continuous-time Markov chain with

$$\begin{aligned} P_{00} &= \frac{\mu}{\lambda + \mu} = 1 - P_{01}, \\ P_{10} &= \frac{\mu}{\lambda + \mu} = 1 - P_{11}, \\ v_i &= \lambda + \mu, \qquad i = 1, 2 \end{aligned}$$

As $P_{00} = P_{10}$, it follows that the probability of a transition into state 0 is equal to $\mu/(\lambda + \mu)$ no matter what the present state. Because a similar result is true for state 1, it follows that the n-stage transition probabilities are given by

$$\begin{aligned} P^n_{i0} &= \frac{\mu}{\lambda + \mu}, \qquad n \geqslant 1,\ i = 0, 1 \\ P^n_{i1} &= \frac{\lambda}{\lambda + \mu}, \qquad n \geqslant 1,\ i = 0, 1 \end{aligned}$$

Hence,

$$\begin{aligned}
P_{00}(t) &= \sum_{n=0}^{\infty} P_{00}^{n} e^{-(\lambda+\mu)t} \frac{[(\lambda+\mu)t]^n}{n!} \\
&= e^{-(\lambda+\mu)t} + \sum_{n=1}^{\infty} \left(\frac{\mu}{\lambda+\mu}\right) e^{-(\lambda+\mu)t} \frac{[(\lambda+\mu)t]^n}{n!} \\
&= e^{-(\lambda+\mu)t} + [1 - e^{-(\lambda+\mu)t}] \frac{\mu}{\lambda+\mu} \\
&= \frac{\mu}{\lambda+\mu} + \frac{\lambda}{\lambda+\mu} e^{-(\lambda+\mu)t}
\end{aligned}$$

Similarly,

$$\begin{aligned}
P_{11}(t) &= \sum_{n=0}^{\infty} P_{11}^{n} e^{-(\lambda+\mu)t} \frac{[(\lambda+\mu)t]^n}{n!} \\
&= e^{-(\lambda+\mu)t} + [1 - e^{-(\lambda+\mu)t}] \frac{\lambda}{\lambda+\mu} \\
&= \frac{\lambda}{\lambda+\mu} + \frac{\mu}{\lambda+\mu} e^{-(\lambda+\mu)t}
\end{aligned}$$

The remaining probabilities are

$$\begin{aligned}
P_{01}(t) &= 1 - P_{00}(t) = \frac{\lambda}{\lambda+\mu}[1 - e^{-(\lambda+\mu)t}], \\
P_{10}(t) &= 1 - P_{11}(t) = \frac{\mu}{\lambda+\mu}[1 - e^{-(\lambda+\mu)t}]
\end{aligned}$$ ■

Example 6.22 Consider the two-state chain of Example 6.20 and suppose that the initial state is state 0. Let $O(t)$ denote the total amount of time that the process is in state 0 during the interval $(0, t)$. The random variable $O(t)$ is often called the *occupation time*. We will now compute its mean.

If we let

$$I(s) = \begin{cases} 1, & \text{if } X(s) = 0 \\ 0, & \text{if } X(s) = 1 \end{cases}$$

then we can represent the occupation time by

$$O(t) = \int_0^t I(s)\, ds$$

Taking expectations and using the fact that we can take the expectation inside the integral sign (since an integral is basically a sum), we obtain

$$\begin{aligned}
E[O(t)] &= \int_0^t E[I(s)]\,ds \\
&= \int_0^t P\{X(s)=0\}\,ds \\
&= \int_0^t P_{00}(s)\,ds \\
&= \frac{\mu}{\lambda+\mu}t + \frac{\lambda}{(\lambda+\mu)^2}\{1-e^{-(\lambda+\mu)t}\}
\end{aligned}$$

where the final equality follows by integrating

$$P_{00}(s) = \frac{\mu}{\lambda+\mu} + \frac{\lambda}{\lambda+\mu}e^{-(\lambda+\mu)s}$$

(For another derivation of $E[O(t)]$, see Exercise 38.) ■

6.8 Computing the Transition Probabilities

For any pair of states i and j, let

$$r_{ij} = \begin{cases} q_{ij}, & \text{if } i \neq j \\ -v_i, & \text{if } i = j \end{cases}$$

Using this notation, we can rewrite the Kolmogorov backward equations

$$P'_{ij}(t) = \sum_{k\neq i} q_{ik}P_{kj}(t) - v_i P_{ij}(t)$$

and the forward equations

$$P'_{ij}(t) = \sum_{k\neq j} q_{kj}P_{ik}(t) - v_j P_{ij}(t)$$

as follows:

$$\begin{aligned}
P'_{ij}(t) &= \textstyle\sum_k r_{ik}P_{kj}(t) \qquad &\text{(backward)} \\
P'_{ij}(t) &= \textstyle\sum_k r_{kj}P_{ik}(t) \qquad &\text{(forward)}
\end{aligned}$$

This representation is especially revealing when we introduce matrix notation. Define the matrices $\mathbf{R}$ and $\mathbf{P}(t)$, $\mathbf{P}'(t)$ by letting the element in row i, column j of these matrices be, respectively, r_{ij}, $P_{ij}(t)$, and $P'_{ij}(t)$. Since the backward equations say that the element in row i, column j of the matrix $\mathbf{P}'(t)$ can be obtained by multiplying the ith row of the matrix $\mathbf{R}$ by the jth column of the matrix $\mathbf{P}(t)$, it is equivalent to the matrix equation

$$\mathbf{P}'(t) = \mathbf{R}\mathbf{P}(t) \tag{6.32}$$

Similarly, the forward equations can be written as

$$\mathbf{P}'(t) = \mathbf{P}(t)\mathbf{R} \tag{6.33}$$

Now, just as the solution of the scalar differential equation

$$f'(t) = cf(t)$$

(or, equivalent, $f'(t) = f(t)c$) is

$$f(t) = f(0)e^{ct}$$

it can be shown that the solution of the matrix differential Equations (6.32) and (6.33) is given by

$$\mathbf{P}(t) = \mathbf{P}(0)e^{\mathbf{R}t}$$

Since $\mathbf{P}(0) = \mathbf{I}$ (the identity matrix), this yields that

$$\mathbf{P}(t) = e^{\mathbf{R}t} \tag{6.34}$$

where the matrix $e^{\mathbf{R}t}$ is defined by

$$e^{\mathbf{R}t} = \sum_{n=0}^{\infty} \mathbf{R}^n \frac{t^n}{n!} \tag{6.35}$$

with $\mathbf{R}^n$ being the (matrix) multiplication of $\mathbf{R}$ by itself n times.

The direct use of Equation (6.35) to compute $\mathbf{P}(t)$ turns out to be very inefficient for two reasons. First, since the matrix $\mathbf{R}$ contains both positive and negative elements (remember the off-diagonal elements are the q_{ij} while the ith diagonal element is $-v_i$), there is the problem of computer round-off error when we compute the powers of $\mathbf{R}$. Second, we often have to compute many of the terms in the infinite sum (6.35) to arrive at a good approximation. However, there are certain indirect ways that we can utilize the relation in (6.34) to efficiently approximate the matrix $\mathbf{P}(t)$. We now present two of these methods.

Approximation Method 1 Rather than using Equation (6.35) to compute $e^{\mathbf{R}t}$, we can use the matrix equivalent of the identity

$$e^x = \lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n$$

which states that

$$e^{\mathbf{R}t} = \lim_{n\to\infty}\left(\mathbf{I}+\mathbf{R}\frac{t}{n}\right)^n$$

Thus, if we let n be a power of 2, say, $n = 2^k$, then we can approximate $\mathbf{P}(t)$ by raising the matrix $\mathbf{M} = \mathbf{I} + \mathbf{R}t/n$ to the nth power, which can be accomplished by k matrix multiplications (by first multiplying $\mathbf{M}$ by itself to obtain $\mathbf{M}^2$ and then multiplying that by itself to obtain $\mathbf{M}^4$ and so on). In addition, since only the diagonal elements of $\mathbf{R}$ are negative (and the diagonal elements of the identity matrix $\mathbf{I}$ are equal to 1), by choosing n large enough we can guarantee that the matrix $\mathbf{I} + \mathbf{R}t/n$ has all nonnegative elements.

Approximation Method 2 A second approach to approximating $e^{\mathbf{R}t}$ uses the identity

$$\begin{aligned} e^{-\mathbf{R}t} &= \lim_{n\to\infty}\left(\mathbf{I}-\mathbf{R}\frac{t}{n}\right)^n \\ &\approx \left(\mathbf{I}-\mathbf{R}\frac{t}{n}\right)^n \qquad \text{for } n \text{ large} \end{aligned}$$

and thus

$$\begin{aligned} \mathbf{P}(t) = e^{\mathbf{R}t} &\approx \left(\mathbf{I}-\mathbf{R}\frac{t}{n}\right)^{-n} \\ &= \left[\left(\mathbf{I}-\mathbf{R}\frac{t}{n}\right)^{-1}\right]^n \end{aligned}$$

Hence, if we again choose n to be a large power of 2, say, $n = 2^k$, we can approximate $\mathbf{P}(t)$ by first computing the inverse of the matrix $\mathbf{I} - \mathbf{R}t/n$ and then raising that matrix to the nth power (by utilizing k matrix multiplications). It can be shown that the matrix $(\mathbf{I} - \mathbf{R}t/n)^{-1}$ will have only nonnegative elements.

Remark Both of the preceding computational approaches for approximating $\mathbf{P}(t)$ have probabilistic interpretations (see Exercises 41 and 42).

Exercises

1. A population of organisms consists of both male and female members. In a small colony any particular male is likely to mate with any particular female in any time interval of length h, with probability $\lambda h + o(h)$. Each mating immediately produces one offspring, equally likely to be male or female. Let $N_1(t)$ and $N_2(t)$ denote the number of males and females in the population at t. Derive the parameters of the continuous-time Markov chain $\{N_1(t), N_2(t)\}$, i.e., the v_i, P_{ij} of Section 6.2.

*2. Suppose that a one-celled organism can be in one of two states—either A or B. An individual in state A will change to state B at an exponential rate α; an individual in state B divides into two new individuals of type A at an exponential rate β. Define an appropriate continuous-time Markov chain for a population of such organisms and determine the appropriate parameters for this model.

3. Consider two machines that are maintained by a single repairman. Machine i functions for an exponential time with rate μ_i before breaking down, $i = 1, 2$. The repair times (for either machine) are exponential with rate μ. Can we analyze this as a birth and death process? If so, what are the parameters? If not, how can we analyze it?

*4. Potential customers arrive at a single-server station in accordance with a Poisson process with rate λ. However, if the arrival finds n customers already in the station, then he will enter the system with probability α_n. Assuming an exponential service rate μ, set this up as a birth and death process and determine the birth and death rates.

5. There are N individuals in a population, some of whom have a certain infection that spreads as follows. Contacts between two members of this population occur in accordance with a Poisson process having rate λ. When a contact occurs, it is equally likely to involve any of the $\binom{N}{2}$ pairs of individuals in the population. If a contact involves an infected and a noninfected individual, then with probability p the noninfected individual becomes infected. Once infected, an individual remains infected throughout. Let $X(t)$ denote the number of infected members of the population at time t.

   (a) Is $\{X(t), t \geqslant 0\}$ a continuous-time Markov chain?
   (b) Specify its type.
   (c) Starting with a single infected individual, what is the expected time until all members are infected?

6. Consider a birth and death process with birth rates $\lambda_i = (i + 1)\lambda, i \geqslant 0$, and death rates $\mu_i = i\mu, i \geqslant 0$.

   (a) Determine the expected time to go from state 0 to state 4.
   (b) Determine the expected time to go from state 2 to state 5.
   (c) Determine the variances in parts (a) and (b).

*7. Individuals join a club in accordance with a Poisson process with rate λ. Each new member must pass through k consecutive stages to become a full member of the club. The time it takes to pass through each stage is exponentially distributed with rate

μ. Let $N_i(t)$ denote the number of club members at time t who have passed through exactly i stages, $i = 1, \ldots, k-1$. Also, let $\mathbf{N}(t) = (N_1(t), N_2(t), \ldots, N_{k-1}(t))$.

(a) Is $\{\mathbf{N}(t), t \geqslant 0\}$ a continuous-time Markov chain?

(b) If so, give the infinitesimal transition rates. That is, for any state $\mathbf{n} = (n_1, \ldots, n_{k-1})$ give the possible next states along with their infinitesimal rates.

8. Consider two machines, both of which have an exponential lifetime with mean $1/\lambda$. There is a single repairman that can service machines at an exponential rate μ. Set up the Kolmogorov backward equations; you need not solve them.

9. The birth and death process with parameters $\lambda_n = 0$ and $\mu_n = \mu, n > 0$ is called a pure death process. Find $P_{ij}(t)$.

10. Consider two machines. Machine i operates for an exponential time with rate λ_i and then fails; its repair time is exponential with rate $\mu_i, i = 1, 2$. The machines act independently of each other. Define a four-state continuous-time Markov chain that jointly describes the condition of the two machines. Use the assumed independence to compute the transition probabilities for this chain and then verify that these transition probabilities satisfy the forward and backward equations.

*11. Consider a Yule process starting with a single individual—that is, suppose $X(0) = 1$. Let T_i denote the time it takes the process to go from a population of size i to one of size $i+1$.

(a) Argue that $T_i, i = 1, \ldots, j$, are independent exponentials with respective rates $i\lambda$.

(b) Let $X_1, \ldots, X_j$ denote independent exponential random variables each having rate λ, and interpret X_i as the lifetime of component i. Argue that $\max(X_1, \ldots, X_j)$ can be expressed as

$$\max(X_1, \ldots, X_j) = \varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_j$$

where $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_j$ are independent exponentials with respective rates $j\lambda$, $(j-1)\lambda, \ldots, \lambda$.

Hint: Interpret ε_i as the time between the $i-1$ and the ith failure.

(c) Using (a) and (b) argue that

$$P\{T_1 + \cdots + T_j \leqslant t\} = (1 - e^{-\lambda t})^j$$

(d) Use (c) to obtain

$$P_{1j}(t) = (1 - e^{-\lambda t})^{j-1} - (1 - e^{-\lambda t})^j = e^{-\lambda t}(1 - e^{-\lambda t})^{j-1}$$

and hence, given $X(0) = 1$, $X(t)$ has a geometric distribution with parameter $p = e^{-\lambda t}$.

(e) Now conclude that

$$P_{ij}(t) = \binom{j-1}{i-1} e^{-\lambda t i}(1 - e^{-\lambda t})^{j-i}$$

12. Each individual in a biological population is assumed to give birth at an exponential rate λ, and to die at an exponential rate μ. In addition, there is an exponential rate

of increase θ due to immigration. However, immigration is not allowed when the population size is N or larger.
(a) Set this up as a birth and death model.
(b) If $N = 3, 1 = \theta = \lambda, \mu = 2$, determine the proportion of time that immigration is restricted.

13. A small barbershop, operated by a single barber, has room for at most two customers. Potential customers arrive at a Poisson rate of three per hour, and the successive service times are independent exponential random variables with mean $\frac{1}{4}$ hour.
(a) What is the average number of customers in the shop?
(b) What is the proportion of potential customers that enter the shop?
(c) If the barber could work twice as fast, how much more business would he do?

14. Potential customers arrive at a full-service, one-pump gas station at a Poisson rate of 20 cars per hour. However, customers will only enter the station for gas if there are no more than two cars (including the one currently being attended to) at the pump. Suppose the amount of time required to service a car is exponentially distributed with a mean of five minutes.
(a) What fraction of the attendant's time will be spent servicing cars?
(b) What fraction of potential customers are lost?

15. A service center consists of two servers, each working at an exponential rate of two services per hour. If customers arrive at a Poisson rate of three per hour, then, assuming a system capacity of at most three customers,
(a) what fraction of potential customers enter the system?
(b) what would the value of part (a) be if there was only a single server, and his rate was twice as fast (that is, $\mu = 4$)?

*16. The following problem arises in molecular biology. The surface of a bacterium consists of several sites at which foreign molecules—some acceptable and some not—become attached. We consider a particular site and assume that molecules arrive at the site according to a Poisson process with parameter λ. Among these molecules a proportion α is acceptable. Unacceptable molecules stay at the site for a length of time that is exponentially distributed with parameter μ_1, whereas an acceptable molecule remains at the site for an exponential time with rate μ_2. An arriving molecule will become attached only if the site is free of other molecules. What percentage of time is the site occupied with an acceptable (unacceptable) molecule?

17. Each time a machine is repaired it remains up for an exponentially distributed time with rate λ. It then fails, and its failure is either of two types. If it is a type 1 failure, then the time to repair the machine is exponential with rate μ_1; if it is a type 2 failure, then the repair time is exponential with rate μ_2. Each failure is, independently of the time it took the machine to fail, a type 1 failure with probability p and a type 2 failure with probability $1 - p$. What proportion of time is the machine down due to a type 1 failure? What proportion of time is it down due to a type 2 failure? What proportion of time is it up?

18. After being repaired, a machine functions for an exponential time with rate λ and then fails. Upon failure, a repair process begins. The repair process proceeds sequentially through k distinct phases. First a phase 1 repair must be performed, then a

phase 2, and so on. The times to complete these phases are independent, with phase i taking an exponential time with rate μ_i, $i = 1, \ldots, k$.

(a) What proportion of time is the machine undergoing a phase i repair?
(b) What proportion of time is the machine working?

*19. A single repairperson looks after both machines 1 and 2. Each time it is repaired, machine i stays up for an exponential time with rate λ_i, $i = 1, 2$. When machine i fails, it requires an exponentially distributed amount of work with rate μ_i to complete its repair. The repairperson will always service machine 1 when it is down. For instance, if machine 1 fails while 2 is being repaired, then the repairperson will immediately stop work on machine 2 and start on 1. What proportion of time is machine 2 down?

20. There are two machines, one of which is used as a spare. A working machine will function for an exponential time with rate λ and will then fail. Upon failure, it is immediately replaced by the other machine if that one is in working order, and it goes to the repair facility. The repair facility consists of a single person who takes an exponential time with rate μ to repair a failed machine. At the repair facility, the newly failed machine enters service if the repairperson is free. If the repairperson is busy, it waits until the other machine is fixed; at that time, the newly repaired machine is put in service and repair begins on the other one. Starting with both machines in working condition, find

(a) the expected value and
(b) the variance of the time until both are in the repair facility.
(c) In the long run, what proportion of time is there a working machine?

21. Suppose that when both machines are down in Exercise 20 a second repairperson is called in to work on the newly failed one. Suppose all repair times remain exponential with rate μ. Now find the proportion of time at least one machine is working, and compare your answer with the one obtained in Exercise 20.

22. Customers arrive at a single-server queue in accordance with a Poisson process having rate λ. However, an arrival that finds n customers already in the system will only join the system with probability $1/(n + 1)$. That is, with probability $n/(n + 1)$ such an arrival will not join the system. Show that the limiting distribution of the number of customers in the system is Poisson with mean λ/μ.

23. A job shop consists of three machines and two repairmen. The amount of time a machine works before breaking down is exponentially distributed with mean 10. If the amount of time it takes a single repairman to fix a machine is exponentially distributed with mean 8, then

(a) what is the average number of machines not in use?
(b) what proportion of time are both repairmen busy?

*24. Consider a taxi station where taxis and customers arrive in accordance with Poisson processes with respective rates of one and two per minute. A taxi will wait no matter how many other taxis are present. However, an arriving customer that does not find a taxi waiting leaves. Find

(a) the average number of taxis waiting, and
(b) the proportion of arriving customers that get taxis.

25. Customers arrive at a service station, manned by a single server who serves at an exponential rate μ_1, at a Poisson rate λ. After completion of service the customer

then joins a second system where the server serves at an exponential rate μ_2. Such a system is called a *tandem* or *sequential* queueing system. Assuming that $\lambda < \mu_i$, $i = 1, 2$, determine the limiting probabilities.

Hint: Try a solution of the form $P_{n,m} = C\alpha^n \beta^m$, and determine C, α, β.

26. Consider an ergodic $M/M/s$ queue in steady state (that is, after a long time) and argue that the number presently in the system is independent of the sequence of past departure times. That is, for instance, knowing that there have been departures 2, 3, 5, and 10 time units ago does not affect the distribution of the number presently in the system.

27. In the $M/M/s$ queue if you allow the service rate to depend on the number in the system (but in such a way so that it is ergodic), what can you say about the output process? What can you say when the service rate μ remains unchanged but $\lambda > s\mu$?

*28. If $\{X(t)\}$ and $\{Y(t)\}$ are independent continuous-time Markov chains, both of which are time reversible, show that the process $\{X(t), Y(t)\}$ is also a time reversible Markov chain.

29. Consider a set of n machines and a single repair facility to service these machines. Suppose that when machine i, $i = 1, \ldots, n$, fails it requires an exponentially distributed amount of work with rate μ_i to repair it. The repair facility divides its efforts equally among all failed machines in the sense that whenever there are k failed machines each one receives work at a rate of $1/k$ per unit time. If there are a total of r working machines, including machine i, then i fails at an instantaneous rate λ_i/r.
    (a) Define an appropriate state space so as to be able to analyze the preceding system as a continuous-time Markov chain.
    (b) Give the instantaneous transition rates (that is, give the q_{ij}).
    (c) Write the time reversibility equations.
    (d) Find the limiting probabilities and show that the process is time reversible.

30. Consider a graph with nodes $1, 2, \ldots, n$ and the $\binom{n}{2}$ arcs (i, j), $i \neq j$, $i, j, = 1, \ldots, n$. (See Section 3.6.2 for appropriate definitions.) Suppose that a particle moves along this graph as follows: Events occur along the arcs (i, j) according to independent Poisson processes with rates λ_{ij}. An event along arc (i, j) causes that arc to become excited. If the particle is at node i at the moment that (i, j) becomes excited, it instantaneously moves to node j, $i, j = 1, \ldots, n$. Let P_j denote the proportion of time that the particle is at node j. Show that

$$P_j = \frac{1}{n}$$

Hint: Use time reversibility.

31. A total of N customers move about among r servers in the following manner. When a customer is served by server i, he then goes over to server j, $j \neq i$, with probability $1/(r-1)$. If the server he goes to is free, then the customer enters service; otherwise he joins the queue. The service times are all independent, with the service times at server i being exponential with rate μ, $i = 1, \ldots, r$. Let the state at any time be the vector $(n_1, \ldots, n_r)$, where n_i is the number of customers presently at server i, $i = 1, \ldots, r$, $\sum_i n_i = N$.

(a) Argue that if $X(t)$ is the state at time t, then $\{X(t),\ t \geqslant 0\}$ is a continuous-time Markov chain.
(b) Give the infinitesimal rates of this chain.
(c) Show that this chain is time reversible, and find the limiting probabilities.

32. Customers arrive at a two-server station in accordance with a Poisson process having rate λ. Upon arriving, they join a single queue. Whenever a server completes a service, the person first in line enters service. The service times of server i are exponential with rate μ_i, $i = 1, 2$, where $\mu_1 + \mu_2 > \lambda$. An arrival finding both servers free is equally likely to go to either one. Define an appropriate continuous-time Markov chain for this model, show it is time reversible, and find the limiting probabilities.

*33. Consider two $M/M/1$ queues with respective parameters λ_i, μ_i, $i = 1, 2$. Suppose they share a common waiting room that can hold at most three customers. That is, whenever an arrival finds her server busy and three customers in the waiting room, she goes away. Find the limiting probability that there will be n queue 1 customers and m queue 2 customers in the system.

Hint: Use the results of Exercise 28 together with the concept of truncation.

34. Four workers share an office that contains four telephones. At any time, each worker is either "working" or "on the phone." Each "working" period of worker i lasts for an exponentially distributed time with rate λ_i, and each "on the phone" period lasts for an exponentially distributed time with rate μ_i, $i = 1, 2, 3, 4$.
(a) What proportion of time are all workers "working"?
Let $X_i(t)$ equal 1 if worker i is working at time t, and let it be 0 otherwise. Let $\mathbf{X}(t) = (X_1(t), X_2(t), X_3(t), X_4(t))$.
(b) Argue that $\{\mathbf{X}(t),\ t \geqslant 0\}$ is a continuous-time Markov chain and give its infinitesimal rates.
(c) Is $\{\mathbf{X}(t)\}$ time reversible? Why or why not?
Suppose now that one of the phones has broken down. Suppose that a worker who is about to use a phone but finds them all being used begins a new "working" period.
(d) What proportion of time are all workers "working"?

35. Consider a time reversible continuous-time Markov chain having infinitesimal transition rates q_{ij} and limiting probabilities $\{P_i\}$. Let A denote a set of states for this chain, and consider a new continuous-time Markov chain with transition rates q^*_{ij} given by

$$q^*_{ij} = \begin{cases} cq_{ij}, & \text{if } i \in A,\ j \notin A \\ q_{ij}, & \text{otherwise} \end{cases}$$

where c is an arbitrary positive number. Show that this chain remains time reversible, and find its limiting probabilities.

36. Consider a system of n components such that the working times of component i, $i = 1, \ldots, n$, are exponentially distributed with rate λ_i. When a component fails, however, the repair rate of component i depends on how many other components are down. Specifically, suppose that the instantaneous repair rate of component i, $i = 1, \ldots, n$, when there are a total of k failed components, is $\alpha^k \mu_i$.

(a) Explain how we can analyze the preceding as a continuous-time Markov chain. Define the states and give the parameters of the chain.
(b) Show that, in steady state, the chain is time reversible and compute the limiting probabilities.

37. For the continuous-time Markov chain of Exercise 3 present a uniformized version.
38. In Example 6.20, we computed $m(t) = E[O(t)]$, the expected occupation time in state 0 by time t for the two-state continuous-time Markov chain starting in state 0. Another way of obtaining this quantity is by deriving a differential equation for it.
(a) Show that

$$m(t+h) = m(t) + P_{00}(t)h + o(h)$$

(b) Show that

$$m'(t) = \frac{\mu}{\lambda+\mu} + \frac{\lambda}{\lambda+\mu}e^{-(\lambda+\mu)t}$$

(c) Solve for $m(t)$.
39. Let $O(t)$ be the occupation time for state 0 in the two-state continuous-time Markov chain. Find $E[O(t)|X(0) = 1]$.
40. Consider the two-state continuous-time Markov chain. Starting in state 0, find $\mathrm{Cov}[X(s), X(t)]$.
41. Let Y denote an exponential random variable with rate λ that is independent of the continuous-time Markov chain $\{X(t)\}$ and let

$$\bar{P}_{ij} = P\{X(Y) = j|X(0) = i\}$$

(a) Show that

$$\bar{P}_{ij} = \frac{1}{v_i + \lambda}\sum_k q_{ik}\bar{P}_{kj} + \frac{\lambda}{v_i+\lambda}\delta_{ij}$$

where δ_{ij} is 1 when $i = j$ and 0 when $i \neq j$.
(b) Show that the solution of the preceding set of equations is given by

$$\bar{\mathbf{P}} = (\mathbf{I} - \mathbf{R}/\lambda)^{-1}$$

where $\bar{\mathbf{P}}$ is the matrix of elements $\bar{P}_{ij}$, $\mathbf{I}$ is the identity matrix, and $\mathbf{R}$ the matrix specified in Section 6.8.
(c) Suppose now that $Y_1, \ldots, Y_n$ are independent exponentials with rate λ that are independent of $\{X(t)\}$. Show that

$$P\{X(Y_1 + \cdots + Y_n) = j|X(0) = i\}$$

is equal to the element in row i, column j of the matrix $\bar{\mathbf{P}}^n$.
(d) Explain the relationship of the preceding to Approximation 2 of Section 6.8.

*42. (a) Show that Approximation 1 of Section 6.8 is equivalent to uniformizing the continuous-time Markov chain with a value v such that $vt = n$ and then approximating $P_{ij}(t)$ by P_{ij}^{*n}.

(b) Explain why the preceding should make a good approximation.

Hint: What is the standard deviation of a Poisson random variable with mean n?

References

[1] D. R. Cox and H. D. Miller, "The Theory of Stochastic Processes," Methuen, London, 1965.

[2] A. W. Drake, "Fundamentals of Applied Probability Theory," McGraw-Hill, New York, 1967.

[3] S. Karlin and H. Taylor, "A First Course in Stochastic Processes," Second Edition, Academic Press, New York, 1975.

[4] E. Parzen, "Stochastic Processes," Holden-Day, San Francisco, California, 1962.

[5] S. Ross, "Stochastic Processes," Second Edition, John Wiley, New York, 1996.