7

CONTINUOUS-TIME MARKOV CHAINS

> Life defies our phrases, it is infinitely continuous and subtle and shaded, whilst our verbal terms are discrete, rude, and few.
>
> –William James

7.1 INTRODUCTION

In this chapter, we extend the Markov chain model to continuous time. A continuous-time process allows one to model not only the transitions between states, but also the duration of time in each state. The central Markov property continues to hold—given the present, past and future are independent.

The Markov property is a form of memorylessness. This leads to the exponential distribution. In a continuous-time Markov chain, when a state is visited, the process stays in that state for an exponentially distributed length of time before moving to a new state. If one just watches the sequence of states that are visited, ignoring the length of time spent in each state, the process looks like a discrete-time Markov chain.

One of the Markov chains introduced in this book was the three-state weather chain of Example 2.3, with state space {rain, snow, clear}. Consider a continuous-time extension. Assume that rainfall lasts, on average, 3 hours at a time. When it snows, the duration, on average, is 6 hours. And the weather stays clear, on average, for

Introduction to Stochastic Processes with R, First Edition. Robert P. Dobrow.

12 hours. Furthermore, changes in weather states are described by the stochastic transition matrix

$$
\widetilde{\boldsymbol{P}} = \begin{array}{r} \\ \text{Rain} \\ \text{Snow} \\ \text{Clear} \end{array}
\begin{array}{c} \begin{array}{ccc} \text{Rain} & \text{Snow} & \text{Clear} \end{array} \\ \begin{pmatrix} 0 & 1/2 & 1/2 \\ 3/4 & 0 & 1/4 \\ 1/4 & 3/4 & 0 \end{pmatrix} \end{array}. \tag{7.1}
$$

To elaborate, assume that it is currently snowing. Under this model, it snows for an exponential length of time with parameter $\lambda_s = 1/6$. (Remember that the parameter of an exponential distribution is the reciprocal of the mean.) Then, the weather changes to either rain or clear, with respective probabilities 3/4 and 1/4. If it switches to rain, it rains for an exponential length of time with parameter $\lambda_r = 1/3$. Then, the weather changes to either snow or clear with equal probability, and so on. Figure 7.1 gives an example of how the process unfolds over 50 hours.

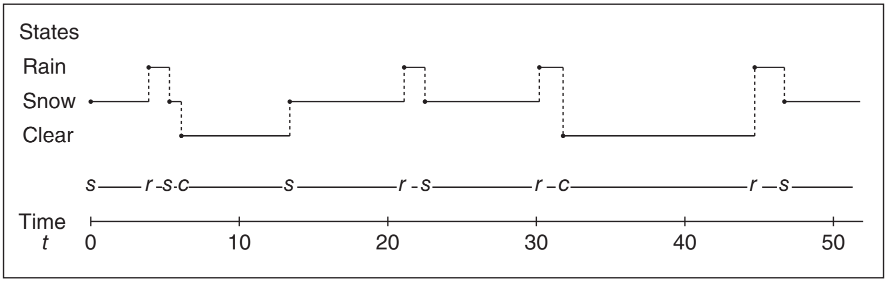

Figure 7.1 Realization of a continuous-time weather chain.

Let X_t denote the weather at time t. Then, $(X_t)_{t\geq 0}$ is a continuous-time Markov chain. The $\widetilde{\boldsymbol{P}}$ matrix, exponential time parameters $(\lambda_r, \lambda_s, \lambda_c) = (1/3, 1/6, 1/12)$, and initial distribution completely specify the process. That is, they are sufficient for computing all probabilities of the form $P\left(X_{t_1} = i_1, \ldots, X_{t_n} = i_n\right)$, for $n \geq 1$, states $i_1, \ldots, i_n$, and times $t_1, \ldots, t_n \geq 0$.

Markov Transition Function

The formal treatment of continuous-time Markov chains begins with the defining Markov property.

Markov Property

A continuous-time stochastic process $(X_t)_{t\geq 0}$ with discrete state space $\mathcal{S}$ is a *continuous-time Markov chain* if

$$P(X_{t+s} = j | X_s = i, X_u = x_u, 0 \leq u < s) = P(X_{t+s} = j | X_s = i),$$

for all $s, t, \geq 0$, $i, j, x_u \in S$, and $0 \leq u < s$.

The process is said to be *time-homogeneous* if this probability does not depend on s. That is,

$$P(X_{t+s} = j|X_s = i) = P(X_t = j|X_0 = i), \text{ for } s \geq 0. \tag{7.2}$$

The Markov chains we treat in this book are all time-homogeneous. For each $t \geq 0$, the transition probabilities in Equation (7.2) can be arranged in a matrix function $\boldsymbol{P}(t)$ called the *transition function*, with

$$P_{ij}(t) = P(X_t = j|X_0 = i).$$

Note that for the weather chain, the $\widetilde{\boldsymbol{P}}$ matrix in Equation (7.1) is *not* the transition function. The $\widetilde{\boldsymbol{P}}$ matrix gives the probabilities of moving from state to state in a discretized process in which time has been ignored.

The transition function $\boldsymbol{P}(t)$ has similar properties as that of the transition matrix for a discrete-time Markov chain. For instance, the Chapman–Kolmogorov equations hold.

Chapman–Kolmogorov Equations

For a continuous-time Markov chain $(X_t)_{t\geq 0}$ with transition function $\boldsymbol{P}(t)$,

$$\boldsymbol{P}(s+t) = \boldsymbol{P}(s)\boldsymbol{P}(t),$$

for $s, t \geq 0$. That is,

$$P_{ij}(s+t) = [P(s)P(t)]_{ij} = \sum_k P_{ik}(s)P_{kj}(t), \text{ for states } i, j, \text{ and } s, t \geq 0.$$

Proof. By conditioning on X_s,

$$\begin{aligned} P_{ij}(s+t) &= P(X_{s+t} = j|X_0 = i) \\ &= \sum_k P(X_{s+t} = j|X_s = k, X_0 = i)P(X_s = k|X_0 = i) \\ &= \sum_k P(X_{s+t} = j|X_s = k)P(X_s = k|X_0 = i) \\ &= \sum_k P(X_t = j|X_0 = k)P(X_s = k|X_0 = i) \\ &= \sum_k P_{ik}(s)P_{kj}(t) = [P(s)P(t)]_{ij}, \end{aligned}$$

where the third equality is because of the Markov property, and the fourth equality is by time-homogeneity. ■

Example 7.1 (Poisson process) A Poisson process $(N_t)_{t\geq 0}$ with parameter λ is a continuous-time Markov chain. The Markov property holds as a consequence of stationary and independent increments. For $0 \leq i \leq j$,

$$\begin{aligned}
P_{ij}(t) = P(N_{t+s} = j|N_s = i) &= \frac{P(N_{t+s} = j, N_s = i)}{P(N_s = i)} \\
&= \frac{P(N_{t+s} - N_s = j - i, N_s = i)}{P(N_s = i)} \\
&= P(N_{t+s} - N_s = j - i) \\
&= P(N_t = j - i) = \frac{e^{-\lambda t}(\lambda t)^{j-i}}{(j-i)!}.
\end{aligned}$$

The transition function is

$$\boldsymbol{P}(t) = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ \vdots \end{array} \begin{array}{c} \begin{array}{ccccc} 0 & 1 & 2 & 3 & \cdots \end{array} \\ \begin{pmatrix} e^{-\lambda t} & (\lambda t)e^{-\lambda t} & (\lambda t)^2 e^{-\lambda t}/2 & (\lambda t)^3 e^{-\lambda t}/6 & \cdots \\ 0 & e^{-\lambda t} & (\lambda t)e^{-\lambda t} & (\lambda t)^2 e^{-\lambda t}/2 & \cdots \\ 0 & 0 & e^{-\lambda t} & (\lambda t)e^{-\lambda t} & \cdots \\ 0 & 0 & 0 & e^{-\lambda t} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \end{array}.$$

■

Holding Times and Embedded Chains

By homogeneity, when a Markov chain visits state i its forward evolution from that time onward behaves the same as the process started in i at time $t = 0$. Time-homogeneity and the Markov property characterizes the distribution of the length of time that a continuous-time chain stays in state i before transitioning to a new state.

Holding Times are Exponentially Distributed

Let T_i be the *holding time* at state i, that is, the length of time that a continuous-time Markov chain started in i stays in i before transitioning to a new state. Then, T_i has an exponential distribution.

Proof. We show that T_i is memoryless. Let $s, t \geq 0$. For the chain started in i, the event $\{T_i > s\}$ is equal to the event that $\{X_u = i$, for $0 \leq u \leq s\}$. Since $\{T_i > s + t\}$ implies $\{T_i > s\}$,

$$P(T_i > s + t|X_0 = i) = P(T_i > s + t, T_i > s|X_0 = i)$$

$$
\begin{aligned}
&= P(T_i > s+t | X_0 = i, T_i > s) P(T_i > s | X_0 = i) \\
&= P(T_i > s+t | X_u = i, \quad \text{for } 0 \le u \le s) P(T_i > s | X_0 = i) \\
&= P(T_i > s+t | X_s = i) P(T_i > s | X_0 = i) \\
&= P(T_i > t | X_0 = i) P(T_i > s | X_0 = i),
\end{aligned}
$$

where the next to last equality is because of the Markov property, and the last equality is because of homogeneity. This gives that T_i is memoryless. The result follows since the exponential distribution is the only continuous distribution that is memoryless. ∎

For each i, let q_i be the parameter of the exponential distribution for the holding time T_i. We assume that $0 < q_i < \infty$. Technically, a continuous-time process can be defined where $q_i = 0$ or $+\infty$. In the former case, when i is visited the process never leaves, and i is called an *absorbing state*. In the latter case, the process leaves i immediately upon entering i. This would allow for infinitely many transitions in a finite interval. Such a process is called *explosive*.

The evolution of a continuous-time Markov chain which is neither absorbing nor explosive can be described as follows. Starting from i, the process stays in i for an exponentially distributed length of time, on average $1/q_i$ time units. Then, it hits a new state $j \neq i$, with some probability p_{ij}. The process stays in j for an exponentially distributed length of time, on average $1/q_j$ time units. It then hits a new state $k \neq j$, with probability p_{jk}, and so on.

The transition probabilities p_{ij} describe the discrete transitions from state to state. If we ignore time, and just watch state to state transitions, we see a sequence $Y_0, Y_1, \ldots$, where Y_n is the nth state visited by the continuous process. The sequence $Y_0, Y_1, \ldots$ is a discrete-time Markov chain called the *embedded chain*.

Let $\widetilde{\boldsymbol{P}}$ be the transition matrix for the embedded chain. That is, $\widetilde{P}_{ij} = p_{ij}$. Then, $\widetilde{\boldsymbol{P}}$ is a stochastic matrix whose diagonal entries are 0.

Example 7.2 (Poisson process) For a Poisson process with parameter λ, the holding time parameters are constant. That is, $q_i = \lambda$, for $i = 0, 1, 2, \ldots$ The process moves from 0 to 1 to 2, and so on. The transition matrix of the embedded chain is

$$
\widetilde{\boldsymbol{P}} = \begin{array}{c|cccccc}
 & 0 & 1 & 2 & 3 & 4 & \cdots \\
\hline
0 & 0 & 1 & 0 & 0 & 0 & \cdots \\
1 & 0 & 0 & 1 & 0 & 0 & \cdots \\
2 & 0 & 0 & 0 & 1 & 0 & \cdots \\
3 & 0 & 0 & 0 & 0 & 1 & \cdots \\
4 & 0 & 0 & 0 & 0 & 0 & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots
\end{array}.
$$

∎

Example 7.3 (Two-state chain) A two-state continuous-time Markov chain is specified by two holding time parameters as depicted in the transition graph in Figure 7.2. Edges of the graph are labeled with holding time rates, not probabilities.

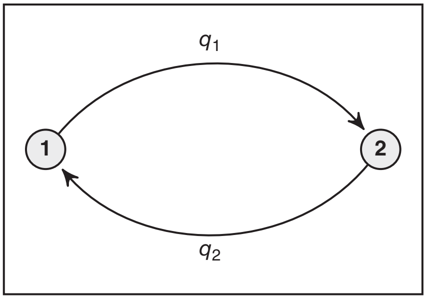

Figure 7.2 Two-state process.

The embedded chain transition matrix is

$$\widetilde{\boldsymbol{P}} = \begin{array}{c} \\ 1 \\ 2 \end{array}\!\!\begin{array}{c} \begin{array}{cc} 1 & 2 \end{array} \\ \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \end{array}.$$

■

The process stays in state 1 an exponential length of time with parameter q_1 before moving to 2. It stays in 2 an exponential length of time with parameter q_2 before moving to 1, and so on.

7.2 ALARM CLOCKS AND TRANSITION RATES

A continuous-time Markov chain can also be described by specifying *transition rates* between pairs of states. Central to this approach is the notion of the *exponential alarm clock*.

Imagine that for each state i, there are independent alarm clocks associated with each of the states that the process can visit after i. If j can be hit from i, then the alarm clock associated with (i, j) will ring after an exponentially distributed length of time with parameter q_{ij}. When the process first hits i, the clocks are started simultaneously. The first alarm that rings determines the next state to visit. If the (i, j) clock rings first and the process moves to j, a new set of exponential alarm clocks are started, with transition rates $q_{j1}, q_{j2}, \ldots$ Again, the first alarm that rings determines the next state hit, and so on.

The q_{ij} are called the *transition rates* of the continuous-time process. From the transition rates, we can obtain the holding time parameters and the embedded chain transition probabilities.

Consider the process started in i. The clocks are started, and the first one that rings determines the next transition. The time of the first alarm is the minimum of independent exponential random variables with parameters $q_{i1}, q_{i2}, \ldots$ Recall results for the exponential distribution and the minimum of independent exponentials, as given in Equations (6.1) and (6.2). The minimum has an exponential distribution with parameter $\sum_k q_{ik}$. That is, the chain stays at i for an exponentially distributed amount of time with parameter $\sum_k q_{ik}$. From the discussion of holding times, the exponential length of time that the process stays in i has parameter q_i. That is,

$$q_i = \sum_k q_{ik}.$$

The interpretation is that the rate that the process leaves state i is equal to the sum of the rates from i to each of the next states.

From i, the chain moves to j if the (i, j) clock rings first, which occurs with probability $q_{ij}/\sum_k q_{ik} = q_{ij}/q_i$. Thus, for the embedded chain transition probabilities

$$p_{ij} = \frac{q_{ij}}{\sum_k q_{ik}} = \frac{q_{ij}}{q_i}.$$

Example 7.4 The general three-state continuous-time Markov chain is described by the transition graph in Figure 7.3. In terms of the transition rates, holding time parameters are

$$(q_1, q_2, q_3) = (q_{12} + q_{13}, q_{21} + q_{23}, q_{31} + q_{32}),$$

with embedded chain transition matrix

$$\widetilde{\boldsymbol{P}} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array}\!\!\begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{pmatrix} 0 & q_{12}/q_1 & q_{13}/q_1 \\ q_{21}/q_2 & 0 & q_{23}/q_2 \\ q_{31}/q_3 & q_{32}/q_3 & 0 \end{pmatrix} \end{array}.$$

■

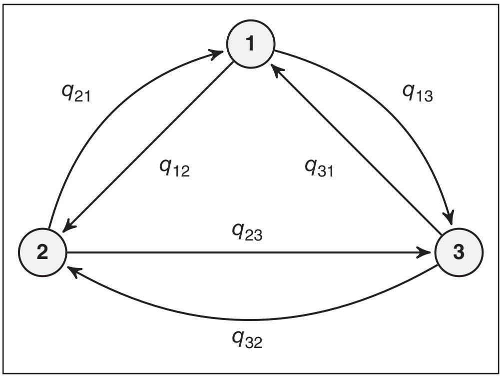

Figure 7.3 Transition rates for three-state chain.

Example 7.5 (Registration line) It is time for students to register for classes, and a line is forming at the registrar's office for those who need assistance. It takes the registrar an exponentially distributed amount of time to service each student, at the rate of one student every 5 minutes. Students arrive at the office and get in line according to a Poisson process at the rate of one student every 4 minutes. Line size is capped at 4 people. If an arriving student finds that there are already 4 people in line, then they try again later. As soon as there is at least one person in line, the registrar starts assisting the first available student. The arrival times of the students are independent of the registrar's service time.

Let X_t be the number of students in line at time t. Then, $(X_t)_{t\geq 0}$ is a continuous-time Markov chain on $\{0, 1, 2, 3, 4\}$.

If there is no one in line, then the size of the line increases to 1 when a student arrives. If there are 4 people in line, then the number decreases to 3 when the registrar finishes assisting the student they are meeting with. If there are 1, 2, or 3 students in line, then the line size can either decrease or increase by 1. If a student arrives at the registrar's office before the registrar has finished serving the student being helped, then the line increases by 1. If the registrar finishes serving the student being helped before another student arrives, the line decreases by 1.

Imagine that when there is 1 person in line two exponential alarm clocks are started—one for student arrivals, with rate 1/4, the other for the registrar's service time, with rate 1/5. If the arrival time clock rings first, the line increases by one. If the service clock rings first, the line decreases by one. The same dynamics hold if there are 2 or 3 people in line. The process is described by the transition graph in Figure 7.4.

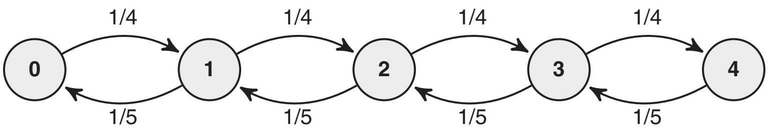

Figure 7.4 Transition rate graph for registration line Markov chain.

The holding time parameters are

$$(q_0, q_1, q_2, q_3, q_4) = \left(\frac{1}{4}, \frac{9}{20}, \frac{9}{20}, \frac{9}{20}, \frac{1}{5}\right),$$

with embedded chain transition matrix

$$\widetilde{\boldsymbol{P}} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{ccccc} \;\;0\;\; & \;\;1\;\; & \;\;2\;\; & \;\;3\;\; & \;\;4\;\; \end{array} \\ \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 4/9 & 0 & 5/9 & 0 & 0 \\ 0 & 4/9 & 0 & 5/9 & 0 \\ 0 & 0 & 4/9 & 0 & 5/9 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} \end{array}.$$

Since students arrive in line at a faster rate than the registrar's service time, the line tends to grow over time. See a realization of the process on $[0, 60]$ in Figure 7.5. ■

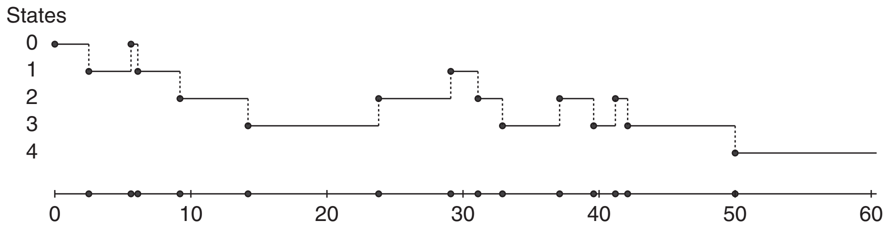

Figure 7.5 Realization of the size of the line at the registrar's office.

7.3 INFINITESIMAL GENERATOR

In continuous time, transition rates play a fundamental role when working with Markov chains. Since the derivative of a function describes its rate of change, it is not surprising that the derivative of the transition function $\boldsymbol{P}'(t)$ is most important.

Assume that $(X_t)_{t\geq 0}$ is a continuous-time Markov chain with transition function $\boldsymbol{P}(t)$. Assume the transition function is differentiable. Note that

$$P_{ij}(0) = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{if } i \neq j. \end{cases}$$

If $X_t = i$, then the instantaneous transition rate of hitting $j \neq i$ is

$$\begin{aligned}
\lim_{h\to 0^+} \frac{E(\text{Number of transitions to } j \text{ in } (t, t+h])}{h} &= \lim_{h\to 0^+} \frac{P(X_{t+h} = j | X_t = i)}{h} \\
&= \lim_{h\to 0^+} \frac{P(X_h = j | X_0 = i)}{h} \\
&= \lim_{h\to 0^+} \frac{P_{ij}(h)}{h} \\
&= \lim_{h\to 0^+} \frac{P_{ij}(h) - P_{ij}(0)}{h} \\
&= P'_{ij}(0).
\end{aligned}$$

The first equality is because for h sufficiently small, the number of transitions to j in $(t, t+h]$ is either 0 or 1. Let $\boldsymbol{Q} = \boldsymbol{P}'(0)$. The off-diagonal entries of $\boldsymbol{Q}$ are the instantaneous transition rates, which are the transition rates q_{ij} introduced in the last section. That is, $Q_{ij} = q_{ij}$, for $i \neq j$. In the language of infinitesimals, if $X_t = i$, then the chance that $X_{t+dt} = j$ is $q_{ij}\,dt$.

The diagonal entries of $\boldsymbol{Q}$ are

$$Q_{ii} = P'_{ii}(0) = \lim_{h\to 0^+} \frac{P_{ii}(h) - P_{ii}(0)}{h} \lim_{h\to 0^+} \frac{P_{ii}(h) - 1}{h}$$

$$= \lim_{h\to 0^+} -\frac{\sum_{j\neq i} P_{ij}(h)}{h} = -\sum_{j\neq i} \lim_{h\to 0^+} \frac{P_{ij}(h)}{h}$$
$$= -\sum_{j\neq i} Q_{ij} = -\sum_{j\neq i} q_{ij} = -q_i.$$

The $\boldsymbol{Q}$ matrix is called the *generator* or *infinitesimal generator*. It is the most important matrix for continuous-time Markov chains. Here, we derived $\boldsymbol{Q}$ from the transition function $\boldsymbol{P}(t)$. However, in a modeling context one typically starts with $\boldsymbol{Q}$, identifying the transition rates q_{ij} based on the qualitative and quantitative dynamics of the process. The transition function and related quantities are derived from $\boldsymbol{Q}$.

Clearly, the generator is not a stochastic matrix. Diagonal entries are negative, entries can be greater than 1, and rows sum to 0.

Example 7.6 The infinitesimal generator matrix for the registration line chain of Example 7.5 is

$$\boldsymbol{Q} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{ccccc} 0 & 1 & 2 & 3 & 4 \end{array} \\ \begin{pmatrix} -1/4 & 1/4 & 0 & 0 & 0 \\ 1/5 & -9/20 & 1/4 & 0 & 0 \\ 0 & 1/5 & -9/20 & 1/4 & 0 \\ 0 & 0 & 1/5 & -9/20 & 1/4 \\ 0 & 0 & 0 & 1/5 & -1/5 \end{pmatrix} \end{array}.$$

■

Example 7.7 The generator for a Poisson process with parameter λ is

$$\boldsymbol{Q} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ \vdots \end{array} \begin{array}{c} \begin{array}{cccccc} 0 & 1 & 2 & 3 & 4 & \cdots \end{array} \\ \begin{pmatrix} -\lambda & \lambda & 0 & 0 & 0 & \cdots \\ 0 & -\lambda & \lambda & 0 & 0 & \cdots \\ 0 & 0 & -\lambda & \lambda & 0 & \cdots \\ 0 & 0 & 0 & -\lambda & \lambda & \cdots \\ 0 & 0 & 0 & 0 & -\lambda & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \end{array}.$$

■

For a continuous-time Markov chain, the transition probabilities of the embedded chain can be derived from first principles from the transition function and generator matrix. From i, the probability that if a transition occurs at time t the process moves to a different state $j \neq i$ is

$$\lim_{h\to 0^+} P(X_{t+h} = j | X_t = i, X_{t+h} \neq i) = \lim_{h\to 0^+} P(X_h = j | X_0 = i, X_h \neq i)$$
$$= \lim_{h\to 0^+} \frac{P(X_h = j, X_h \neq i, X_0 = i)}{P(X_h \neq i, X_0 = i)}$$

$$= \lim_{h\to 0^+} \frac{P(X_h = j|X_0 = i)}{P(X_h \neq i|X_0 = i)}$$
$$= \lim_{h\to 0^+} \frac{P_{ij}(h)/h}{1 - P_{ii}(h)/h}$$
$$= \frac{q_{ij}}{q_i},$$

which is independent of t. This gives the transition probability p_{ij} of the embedded chain. It also establishes the relationship between instantaneous transition rates, holding time parameters, and the embedded chain transition probabilities.

Instantaneous Rates, Holding Times, Transition Probabilities

For a continuous-time Markov chain, let q_{ij}, q_i, and p_{ij} be defined as above. For $i \neq j$,

$$q_{ij} = q_i p_{ij}.$$

For discrete-time Markov chains, there is no generator matrix and the probabilistic properties of the stochastic process are captured by the transition matrix $\boldsymbol{P}$. For continuous-time Markov chains the generator matrix $\boldsymbol{Q}$ gives a complete description of the dynamics of the process. The distribution of any finite subset of the X_t, and all probabilistic quantities of the stochastic process, can, in principle, be obtained from the infinitesimal generator and the initial distribution.

Forward, Backward Equations

The transition function $\boldsymbol{P}(t)$ can be computed from the generator $\boldsymbol{Q}$ by solving a system of differential equations.

Kolmogorov Forward, Backward Equations

A continuous-time Markov chain with transition function $\boldsymbol{P}(t)$ and infinitesimal generator $\boldsymbol{Q}$ satisfies the *forward equation*

$$\boldsymbol{P}'(t) = \boldsymbol{P}(t)\boldsymbol{Q} \tag{7.3}$$

and the *backward equation*

$$\boldsymbol{P}'(t) = \boldsymbol{Q}\boldsymbol{P}(t). \tag{7.4}$$

Equivalently, for all states i and j,

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

and

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

Proof. The equations are a consequence of the Chapman–Kolmogorov property. For the forward equation, for $h, t \geq 0$,

$$\begin{aligned}\frac{\boldsymbol{P}(t+h) - \boldsymbol{P}(t)}{h} &= \frac{\boldsymbol{P}(t)\boldsymbol{P}(h) - \boldsymbol{P}(t)}{h} \\ &= \boldsymbol{P}(t)\left(\frac{\boldsymbol{P}(h) - \boldsymbol{I}}{h}\right) \\ &= \boldsymbol{P}(t)\left(\frac{\boldsymbol{P}(h) - \boldsymbol{P}(0)}{h}\right).\end{aligned}$$

Taking limits as $h \to 0^+$ gives $\boldsymbol{P}'(t) = \boldsymbol{P}(t)\boldsymbol{Q}$. The backward equation is derived similarly, starting with $\boldsymbol{P}(t+h) = \boldsymbol{P}(h)\boldsymbol{P}(t)$. ■

Example 7.8 (Poisson process) The transition probabilities for the Poisson process with parameter λ,

$$P_{ij}(t) = \frac{e^{-\lambda t}(\lambda t)^{j-i}}{(j-i)!}, \text{ for } j \geq i,$$

were derived in Example 7.1. They satisfy the Kolmogorov forward equations

$$\begin{aligned}P'_{ii}(t) &= -\lambda P_{ii}(t), \\ P'_{ij}(t) &= -\lambda P_{ij}(t) + \lambda P_{i,j-1}, \text{ for } j = i+1, i+2, \ldots,\end{aligned}$$

and the backward equations

$$\begin{aligned}P'_{ii}(t) &= -\lambda P_{ii}(t), \\ P'_{ij}(t) &= -\lambda P_{ij}(t) + \lambda P_{j+1,i}, \text{ for } j = i+1, i+2, \ldots\end{aligned}$$

■

Example 7.9 (Two-state process) For a continuous-time process with generator

$$\boldsymbol{Q} = \begin{array}{c c} & \begin{array}{cc} 1 & 2 \end{array} \\ \begin{array}{c} 1 \\ 2 \end{array} & \begin{pmatrix} -\lambda & \lambda \\ \mu & -\mu \end{pmatrix} \end{array}$$

the forward equations give

$$\begin{aligned} P'_{11}(t) &= -P_{11}(t)q_1 + P_{12}(t)q_{21} \\ &= -\lambda P_{11}(t) + (1 - P_{11}(t))\mu \\ &= \mu - (\lambda + \mu)P_{11}(t), \end{aligned}$$

using the fact that the first row of the matrix $\boldsymbol{P}(t)$ sums to 1. The solution to the linear differential equation is

$$P_{11}(t) = \frac{\mu}{\lambda + \mu} + \frac{\lambda}{\lambda + \mu} e^{-(\lambda+\mu)t}.$$

Also,

$$P'_{22}(t) = -P_{22}(t)q_2 + P_{21}(t)q_{12} = \lambda - (\lambda + \mu)P_{22}(t),$$

with solution

$$P_{22}(t) = \frac{\lambda}{\lambda + \mu} + \frac{\mu}{\lambda + \mu} e^{-(\lambda+\mu)t}.$$

The transition function is

$$\boldsymbol{P}(t) = \frac{1}{\lambda + \mu} \begin{array}{c} \begin{array}{cc} \quad 1 \quad & \quad\quad 2 \end{array} \\ \begin{pmatrix} \mu + \lambda e^{-(\lambda+\mu)t} & \lambda - \lambda e^{-(\lambda+\mu)t} \\ \mu - \mu e^{-(\lambda+\mu)t} & \lambda + \mu e^{-(\lambda+\mu)t} \end{pmatrix} \end{array}. \tag{7.5}$$

■

Matrix Exponential

The Kolmogorov backward equation $\boldsymbol{P}'(t) = \boldsymbol{QP}(t)$ is a matrix equation, which bears a striking resemblance to the nonmatrix differential equation $p'(t) = qp(t)$, where p is a differentiable function and q is a constant. If $p(0) = 1$, the latter has the unique solution

$$p(t) = e^{tq}, \text{ for } t \geq 0.$$

If you are not familiar with solutions to matrix differential equations it might be tempting to try to solve the backward equation by analogy, and write

$$\boldsymbol{P}(t) = e^{t\boldsymbol{Q}}, \text{ for } t \geq 0,$$

since $\boldsymbol{P}(0) = \boldsymbol{I}$. Remarkably, this is exactly correct, as long as $e^{t\boldsymbol{Q}}$ is defined properly.

Matrix Exponential

Let $\boldsymbol{A}$ be a $k \times k$ matrix. The *matrix exponential* $e^{\boldsymbol{A}}$ is the $k \times k$ matrix

$$e^{\boldsymbol{A}} = \sum_{n=0}^{\infty} \frac{1}{n!}\boldsymbol{A}^n = \boldsymbol{I} + \boldsymbol{A} + \frac{1}{2}\boldsymbol{A}^2 + \frac{1}{6}\boldsymbol{A}^3 + \cdots .$$

The matrix exponential is the matrix version of the exponential function and reduces to the ordinary exponential function e^x when $\boldsymbol{A}$ is a 1×1 matrix. The matrix $e^{\boldsymbol{A}}$ is well-defined as its defining series converges for all square matrices $\boldsymbol{A}$.

The matrix exponential satisfies many familiar properties of the exponential function. These include

1. $e^{\boldsymbol{0}} = \boldsymbol{I}$.
2. $e^{\boldsymbol{A}} e^{-\boldsymbol{A}} = \boldsymbol{I}$.
3. $e^{(s+t)\boldsymbol{A}} = e^{s\boldsymbol{A}} e^{t\boldsymbol{A}}$.
4. If $\boldsymbol{AB} = \boldsymbol{BA}$, then $e^{\boldsymbol{A}+\boldsymbol{B}} = e^{\boldsymbol{A}} e^{\boldsymbol{B}} = e^{\boldsymbol{B}} e^{\boldsymbol{A}}$.
5. $\frac{d}{dt} e^{t\boldsymbol{A}} = \boldsymbol{A} e^{t\boldsymbol{A}} = e^{t\boldsymbol{A}} \boldsymbol{A}$.

For a continuous-time Markov chain with generator $\boldsymbol{Q}$, the matrix exponential $e^{t\boldsymbol{Q}}$ is the unique solution to the forward and backward equations. Letting $\boldsymbol{P}(t) = e^{t\boldsymbol{Q}}$ gives

$$\boldsymbol{P}'(t) = \frac{d}{dt} e^{t\boldsymbol{Q}} = \boldsymbol{Q} e^{t\boldsymbol{Q}} = e^{t\boldsymbol{Q}} \boldsymbol{Q} = \boldsymbol{P}(t)\boldsymbol{Q} = \boldsymbol{Q}\boldsymbol{P}(t).$$

Transition Function and Generator

For a continuous-time Markov chain with transition function $\boldsymbol{P}(t)$ and infinitesimal generator $\boldsymbol{Q}$,

$$\boldsymbol{P}(t) = e^{t\boldsymbol{Q}} = \sum_{n=0}^{\infty} \frac{1}{n!} (t\boldsymbol{Q})^n = \boldsymbol{I} + t\boldsymbol{Q} + \frac{t^2}{2}\boldsymbol{Q}^2 + \frac{t^3}{6}\boldsymbol{Q}^3 + \cdots \tag{7.6}$$

Computing the matrix exponential is often numerically challenging. Finding accurate and efficient algorithms is still a topic of current research. Furthermore, the transition function is difficult to obtain in closed form for all but the most specialized models. For applied problems, numerical approximation methods are often needed. The `R` package `expm` contains the function `expm(mat)` for computing the matrix exponential of a numerical matrix `mat`.

`R`: Computing the Transition Function

For the registration line Markov chain of Example 7.5, we find the transition function $\boldsymbol{P}(t)$ for $t = 2.5$.

```
# matrixexp.R
> install.packages("expm")
> library(expm)
> Q <- matrix(c(-1/4,1/4,0,0,0,1/5,-9/20,1/4,0,0,0,
```

```
+ 1/5,-9/20,1/4,0,0,0,1/5,-9/20,1/4,0,0,0,1/5,-1/5),
+ nrow=5,byrow=T)
> Q
      0     1     2     3     4
0 -0.25  0.25  0.00  0.00  0.00
1  0.20 -0.45  0.25  0.00  0.00
2  0.00  0.20 -0.45  0.25  0.00
3  0.00  0.00  0.20 -0.45  0.25
4  0.00  0.00  0.00  0.20 -0.20
> P  <- function(t) expm(t*Q)
> P(2.5)
      0     1     2     3     4
0 0.610 0.290 0.081 0.016 0.003
1 0.232 0.443 0.238 0.071 0.017
2 0.052 0.190 0.435 0.238 0.085
3 0.008 0.045 0.191 0.446 0.310
4 0.001 0.008 0.054 0.248 0.688
```

Diagonalization*

If the $\boldsymbol{Q}$ matrix is diagonalizable, then so is $e^{t\boldsymbol{Q}}$, and the transition function can be expressed in terms of the eigenvalues and eigenvectors of $\boldsymbol{Q}$. Write $\boldsymbol{Q} = \boldsymbol{SDS}^{-1}$, where

$$\boldsymbol{D} = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_k \end{pmatrix}$$

is a diagonal matrix whose diagonal entries are the eigenvalues of $\boldsymbol{Q}$, and $\boldsymbol{S}$ is an invertible matrix whose columns are the corresponding eigenvectors. This gives

$$\begin{aligned} e^{t\boldsymbol{Q}} &= \sum_{n=0}^{\infty} \frac{1}{n!}(t\boldsymbol{Q})^n = \sum_{n=0}^{\infty} \frac{t^n}{n!}(\boldsymbol{SDS}^{-1})^n \\ &= \sum_{n=0}^{\infty} \frac{t^n}{n!}\boldsymbol{SD}^n\boldsymbol{S}^{-1} = \boldsymbol{S}\left(\sum_{n=0}^{\infty} \frac{t^n}{n!}\boldsymbol{D}^n\right)\boldsymbol{S}^{-1} \\ &= \boldsymbol{S}e^{t\boldsymbol{D}}\boldsymbol{S}^{-1}, \end{aligned}$$

where

$$e^{t\boldsymbol{D}} = \sum_{n=0}^{\infty} \frac{t^n}{n!}\boldsymbol{D}^n = \sum_{n=0}^{\infty} \frac{t^n}{n!} \begin{pmatrix} \lambda_1^n & 0 & \cdots & 0 \\ 0 & \lambda_2^n & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_k^n \end{pmatrix}$$

$$= \begin{pmatrix} \sum_{n=0}^{\infty} (t\lambda_1)^n/n! & 0 & \cdots & 0 \\ 0 & \sum_{n=0}^{\infty} (t\lambda_2)^n/n! & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sum_{n=0}^{\infty} (t\lambda_k)^n/n! \end{pmatrix}$$

$$= \begin{pmatrix} e^{t\lambda_1} & 0 & \cdots & 0 \\ 0 & e^{t\lambda_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & e^{t\lambda_k} \end{pmatrix}.$$

Example 7.10 For the two-state chain, the generator

$$\boldsymbol{Q} = \begin{pmatrix} -\lambda & \lambda \\ \mu & -\mu \end{pmatrix}$$

is diagonalizable with eigenvalues 0 and $-(\lambda + \mu)$, and corresponding eigenvectors $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $\begin{pmatrix} -\lambda \\ \mu \end{pmatrix}$. This gives

$$\boldsymbol{Q} = \boldsymbol{SDS}^{-1} = \begin{pmatrix} 1 & -\lambda \\ 1 & \mu \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & -(\lambda+\mu) \end{pmatrix} \begin{pmatrix} \mu/(\lambda+\mu) & \lambda/(\lambda+\mu) \\ -1/(\lambda+\mu) & 1/(\lambda+\mu) \end{pmatrix}.$$

The transition function is

$$\begin{aligned} \boldsymbol{P}(t) &= e^{t\boldsymbol{Q}} = \boldsymbol{S}e^{t\boldsymbol{D}}\boldsymbol{S}^{-1} \\ &= \begin{pmatrix} 1 & -\lambda \\ 1 & \mu \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & e^{-t(\lambda+\mu)} \end{pmatrix} \begin{pmatrix} \mu/(\lambda+\mu) & \lambda/(\lambda+\mu) \\ -1/(\lambda+\mu) & 1/(\lambda+\mu) \end{pmatrix} \\ &= \frac{1}{\lambda+\mu} \begin{pmatrix} \mu + \lambda e^{-t(\lambda+\mu)} & \lambda - \lambda e^{-t(\lambda+\mu)} \\ \mu - \mu e^{-t(\lambda+\mu)} & \lambda + \mu e^{-t(\lambda+\mu)} \end{pmatrix}. \end{aligned}$$

This result was also shown in Example 7.9 as the solution to the Kolmogorov forward equations. ■

Example 7.11 (DNA evolution) Continuous-time Markov chains are used to study the evolution of DNA sequences. Numerous models have been proposed for the evolutionary changes on the genome as a result of mutation. Such models are often specified in terms of transition rates between base nucleotides adenine, guanine, cytosine, and thymine at a fixed chromosome location.

The *Jukes–Cantor* model assumes that all transition rates are the same. The infinitesimal generator, with parameter $r > 0$, is

$$\boldsymbol{Q} = \begin{array}{c} \\ a \\ g \\ c \\ t \end{array} \begin{array}{c} \begin{array}{cccc} a & g & c & t \end{array} \\ \begin{pmatrix} -3r & r & r & r \\ r & -3r & r & r \\ r & r & -3r & r \\ r & r & r & -3r \end{pmatrix} \end{array}.$$

The generator is diagonalizable with linearly independent eigenvectors

$$\begin{pmatrix}-1\\0\\0\\1\end{pmatrix}, \begin{pmatrix}-1\\0\\1\\0\end{pmatrix}, \begin{pmatrix}-1\\1\\0\\0\end{pmatrix}, \begin{pmatrix}1\\1\\1\\1\end{pmatrix},$$

corresponding to eigenvalues $-4r, -4r, -4r$, and 0. This gives

$$\begin{aligned}
\boldsymbol{P}(t) &= e^{t\boldsymbol{Q}} = \boldsymbol{S}e^{t\boldsymbol{D}}\boldsymbol{S}^{-1}\\
&= \begin{pmatrix}-1 & -1 & -1 & 1\\ 0 & 0 & 1 & 1\\ 0 & 1 & 0 & 1\\ 1 & 0 & 0 & 1\end{pmatrix}\begin{pmatrix}e^{-4rt} & 0 & 0 & 0\\ 0 & e^{-4rt} & 0 & 0\\ 0 & 0 & e^{-4rt} & 0\\ 0 & 0 & 0 & 1\end{pmatrix}\begin{pmatrix}-\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & \frac{3}{4}\\ -\frac{1}{4} & -\frac{1}{4} & \frac{3}{4} & -\frac{1}{4}\\ -\frac{1}{4} & \frac{3}{4} & -\frac{1}{4} & -\frac{1}{4}\\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4}\end{pmatrix}\\
&= \frac{1}{4}\begin{pmatrix}1+3e^{-4rt} & 1-e^{-4rt} & 1-e^{-4rt} & 1-e^{-4rt}\\ 1-e^{-4rt} & 1+3e^{-4rt} & 1-e^{-4rt} & 1-e^{-4rt}\\ 1-e^{-4rt} & 1-e^{-4rt} & 1+3e^{-4rt} & 1-e^{-4rt}\\ 1-e^{-4rt} & 1-e^{-4rt} & 1-e^{-4rt} & 1+3e^{-4rt}\end{pmatrix}.
\end{aligned}$$

The Jukes–Cantor model does not distinguish between base types. Nucleotides a and g are purines, c and t are pyrimidines. Changes from purine to purine or pyrimidine to pyrimidine are called *transitions*. Changes from purine to pyrimidine or vice versa are called *transversions*.

The *Kimura* model, which includes two parameters r and s, distinguishes between transitions and transversions. The generator is

$$\boldsymbol{Q} = \begin{array}{c c} & \begin{array}{cccc} a & g & c & t \end{array}\\ \begin{array}{c} a\\ g\\ c\\ t \end{array} & \begin{pmatrix} -(r+2s) & r & s & s\\ r & -(r+2s) & s & s\\ s & s & -(r+2s) & r\\ s & s & r & -(r+2s) \end{pmatrix} \end{array}.$$

The matrix is diagonalizable. The respective matrices of eigenvalues and eigenvectors are

$$\boldsymbol{D} = \begin{pmatrix}-2(r+s) & 0 & 0 & 0\\ 0 & -2(r+s) & 0 & 0\\ 0 & 0 & -4s & 0\\ 0 & 0 & 0 & 0\end{pmatrix} \quad \text{and} \quad \boldsymbol{S} = \begin{pmatrix}-1 & 0 & -1 & 1\\ 0 & -1 & 1 & 1\\ 1 & 0 & -1 & 1\\ 0 & 1 & 1 & 1\end{pmatrix}.$$

The transition function $\boldsymbol{P}(t) = e^{t\boldsymbol{Q}} = \boldsymbol{S}e^{t\boldsymbol{D}}\boldsymbol{S}^{-1}$ is

$$\boldsymbol{P}_{xy}(t) = \begin{cases} \left(1 + e^{-4st} - 2e^{-2(r+s)t}\right)/4, & \text{if } xy \in \{ag, ga, ct, tc\}, \\ \left(1 - e^{-4st}\right)/4, & \text{if } xy \in \{ac, at, gc, gt, ca, cg, ta, tg\}, \\ \left(1 + e^{-4st} + 2e^{-2(r+s)t}\right)/4, & \text{if } xy \in \{aa, gg, cc, tt\}. \end{cases}$$

Estimating Mutation Rate and Evolutionary Distance

Continuous-time Markov models are used by biologists and geneticists to estimate the evolutionary distance between species, as well as the related mutation rate. The following application is based on Durrett (2002).

A statistic for estimating evolutionary distance from DNA sequences is the number of locations in the sequences that differ. Given two DNA strands and a common nucleotide site, consider the probability q that the nucleotides in the two strands are identical given that the most recent common ancestor occurred t time units ago. Assume at that time that the nucleotide at the given site was a. Then, the probability that the two sequences are identical, by the Jukes–Cantor model, is

$$\begin{aligned} q &= (P_{aa}(t))^2 + (P_{ag}(t))^2 + (P_{ac}(t))^2 + (P_{at}(t))^2 \\ &= \left(\frac{1}{4} + \frac{3}{4}e^{-4rt}\right)^2 + 3\left(\frac{1}{4} - \frac{1}{4}e^{-4rt}\right)^2 \\ &= \frac{1}{4} + \frac{3}{4}e^{-8rt}, \end{aligned}$$

which is independent of the starting nucleotide. Thus the probability p that the two sites are different is

$$p = 1 - q = \frac{3}{4}\left(1 - e^{-8rt}\right).$$

Solving for rt gives

$$rt = -\frac{1}{8}\ln\left(1 - \frac{4p}{3}\right).$$

In the Jukes–Cantor model, nucleotides change at the rate of $3r$ transitions per time unit. In two DNA strands we expect $2(3rt) = 6rt$ substitutions over t years. Let K denote the number of substitutions that occur over t years. Then,

$$E(K) = 6rt = -\frac{3}{4}\ln\left(1 - \frac{4p}{3}\right).$$

To estimate the actual number of substitutions that occurred since the most recent common ancestor t time units ago, take

$$\widehat{K} = -\frac{3}{4}\ln\left(1 - \frac{4\widehat{p}}{3}\right),$$

where $\widehat{p}$ is the observed fraction of differences between two sequences.

The number of nucleotide substitutions per site between two sequences since their separation from the last common ancestor is called the *evolutionary distance* between the sequences. It is an important quantity for estimating the rate of evolution and the divergence time between species.

Durrett compares the sequence of mitochondrial RNA for the somatotropin gene (a growth hormone) for rats and humans. The observed proportion of differences between the two sequences at a fixed site in the gene is $\hat{p} = 0.366$. Hence, the estimate of the number of substitutions at that site that have occurred since divergence of the two species is

$$\hat{K} = -\frac{3}{4}\ln\left(1 - \frac{4(0.366)}{3}\right) = 0.502.$$

Using the fact that rats and humans diverged about 80 million years ago, and choosing 0.502 as the number of substitutions per nucleotide, the mutation rate at that position is estimated as $0.502/(8 \times 10^7) = 6.275 \times 10^{-9}$ per year. ■

Example 7.12 (Using symbolic software) Symbolic software systems, such as *Mathematica* and *Maple*, work in exact integer arithmetic, and can be used to find the matrix exponential when the generator matrix contains symbolic parameters. *Wolfram Alpha*, which is freely available on the web, has the command **MatrixExp** for computing the matrix exponential.

To find the transition function for the Jukes–Cantor model from the previous example, we type the following command from our web browser

$$> \quad \textbf{MatrixExp}[t\{\{-3r, r, r, r\}, \{r, -3r, r, r\}, \{r, r, -3r, r\}, \{r, r, r, -3r\}\}],$$

which gives the output shown in Figure 7.6. ■

7.4 LONG-TERM BEHAVIOR

For continuous-time Markov chains, limiting and stationary distributions are defined similarly as for discrete time.

Limiting Distribution

A probability distribution $\boldsymbol{\pi}$ is the *limiting distribution* of a continuous-time Markov chain if for all states i and j,

$$\lim_{t\to\infty} \boldsymbol{P}_{ij}(t) = \pi_j.$$

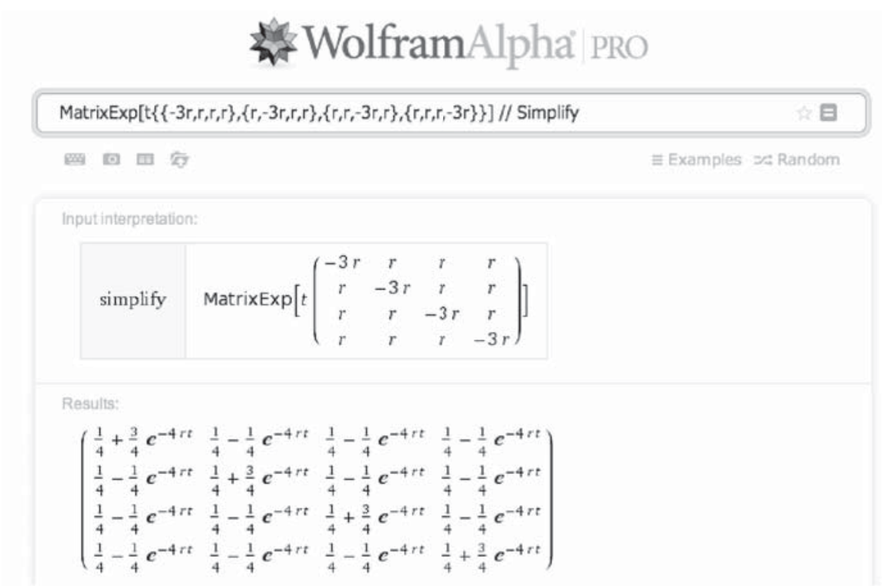

Figure 7.6 Computing the matrix exponential with *WolframAlpha*. *Source:* See Wolfram Alpha LLC (2015).

Stationary Distribution

A probability distribution $\boldsymbol{\pi}$ is a *stationary distribution* if

$$\boldsymbol{\pi} = \boldsymbol{\pi}\boldsymbol{P}(t), \text{ for } t \geq 0.$$

That is, for all states j,

$$\pi_j = \sum_i \pi_i P_{ij}(t), \text{ for } t \geq 0.$$

As in the discrete case, the limiting distribution, if it exists, is a stationary distribution. However, the converse is not necessarily true and depends on the class structure of the chain. For characterizing the states of a continuous-time Markov chain, notions of accessibility, communication, and irreducibility are defined as in the discrete case. A continuous-time Markov chain is *irreducible* if for all i and j, $P_{ij}(t) > 0$ for some $t > 0$.

In one regard, the classification of states is easier in continuous time since periodicity is not an issue. All states are essentially aperiodic, a consequence of the following lemma.

Lemma 7.1. *If $P_{ij}(t) > 0$, for some $t > 0$, then $P_{ij}(t) > 0$, for all $t > 0$.*

The result is intuitive since if $P_{ij}(t) > 0$ for some t, then there exists a path from i to j in the embedded chain, and for any time s there is positive probability of reaching j from i in s time units. We forego the complete proof, but show the result for forward time. Assume that $P_{ij}(t) > 0$ for some t. Then, for $s \geq 0$,

$$P_{ij}(t+s) = \sum_k P_{ik}(t)P_{kj}(s) \geq P_{ij}(t)P_{jj}(s) \geq P_{ij}(t)e^{-q_j s} > 0.$$

For the penultimate inequality, $e^{-q_j s}$ is the probability that there is no transition from j by time s. The latter event implies that the process started at j is at j at time s, whose probability is $P_{jj}(s)$.

A finite-state continuous-time Markov chain is irreducible if all the holding time parameters are positive. On the contrary, if $q_i = 0$ for some i, then i is an *absorbing state*. If we assume that all the holding time parameters are finite, then there are two possibilities: (i) the process is irreducible, all states communicate, and $P_{ij}(t) > 0$, for $t > 0$ and all i, j or (ii) the process contains one or more absorbing states.

The following fundamental limit theorem is given without proof. Note the analogies with the discrete-time results, for example, Theorems 3.6 and 3.8.

Fundamental Limit Theorem

Theorem 7.2. *Let $(X_t)_{t\geq 0}$ be a finite, irreducible, continuous-time Markov chain with transition function $\boldsymbol{P}(t)$. Then, there exists a unique stationary distribution $\boldsymbol{\pi}$, which is the limiting distribution. That is, for all j,*

$$\lim_{t\to\infty} P_{ij}(t) = \pi_j, \textit{for all initial } i.$$

Equivalently,

$$\lim_{t\to\infty} \boldsymbol{P}(t) = \boldsymbol{\Pi},$$

where $\boldsymbol{\Pi}$ is a matrix all of whose rows are equal to $\boldsymbol{\pi}$.

Example 7.13 (Two-state process) For the continuous-time Markov chain on two states,

$$\lim_{t\to\infty} \boldsymbol{P}(t) = \lim_{t\to\infty} \frac{1}{\lambda+\mu}\begin{pmatrix} \mu + \lambda e^{-(\lambda+\mu)t} & \lambda - \lambda e^{-(\lambda+\mu)t} \\ \mu - \mu e^{-(\lambda+\mu)t} & \lambda + \mu e^{-(\lambda+\mu)t} \end{pmatrix} = \frac{1}{\lambda+\mu}\begin{pmatrix} \mu & \lambda \\ \mu & \lambda \end{pmatrix}.$$

The stationary distribution is

$$\boldsymbol{\pi} = \left(\frac{\mu}{\lambda+\mu}, \frac{\lambda}{\lambda+\mu}\right).$$

■

The following result links the stationary distribution with the generator.

Stationary Distribution and Generator Matrix

A probability distribution $\boldsymbol{\pi}$ is a stationary distribution of a continuous-time Markov chain with generator $\boldsymbol{Q}$ if and only if

$$\boldsymbol{\pi Q} = \mathbf{0}.$$

That is,

$$\sum_i \pi_i Q_{ij} = 0, \text{ for all } j.$$

Proof. Assume that $\boldsymbol{\pi} = \boldsymbol{\pi P}(t)$, for all $t \geq 0$. Take the derivative of both sides of the equation at $t = 0$ to get $\mathbf{0} = \boldsymbol{\pi P}'(0) = \boldsymbol{\pi Q}$. Conversely, assume that $\boldsymbol{\pi Q} = \mathbf{0}$. Then,

$$\mathbf{0} = \boldsymbol{\pi Q P}(t) = \boldsymbol{\pi P}'(t), \text{ for } t \geq 0,$$

by the Kolmogorov backward equation. Since the derivative is equal to $\mathbf{0}$, $\boldsymbol{\pi P}(t)$ is a constant. In particular, $\boldsymbol{P}(0) = \boldsymbol{I}$, and thus $\boldsymbol{\pi P}(t) = \boldsymbol{\pi P}(0) = \boldsymbol{\pi}$, for $t \geq 0$. ■

The stationary probability π_j can be interpreted as the long-term proportion of time that the chain spends in state j. This is analogous to the discrete-time case in which the stationary probability represents the long-term fraction of transitions that the chain visits a given state.

Example 7.14 (Eat, play, sleep) Jesse is a newborn baby who is always in one of three states: eat, play, and sleep. He eats on average for 30 minutes at a time; plays on average for 1 hour; and sleeps for about 3 hours. After eating, there is a 50–50 chance he will sleep or play. After playing, there is a 50–50 chance he will eat or sleep. And after sleeping, he always plays. Jesse's life is governed by a continuous-time Markov chain. What proportion of the day does Jesse sleep?

Solution The holding time parameters for the three-state chain (in hour units) are $(q_e, q_p, q_s) = (2, 1, 1/3)$. The embedded chain transition probabilities are

$$\widetilde{\boldsymbol{P}} = \begin{array}{r} \\ \text{Eat} \\ \text{Play} \\ \text{Sleep} \end{array} \begin{array}{c} \begin{array}{ccc} \text{Eat} & \text{Play} & \text{Sleep} \end{array} \\ \begin{pmatrix} 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \end{pmatrix} \end{array}.$$

With $q_{ij} = q_i p_{ij}$, the generator matrix is

$$Q = \begin{array}{c} \\ \text{Eat} \\ \text{Play} \\ \text{Sleep} \end{array} \begin{array}{c} \begin{array}{ccc} \text{Eat} & \text{Play} & \text{Sleep} \end{array} \\ \begin{pmatrix} -2 & 1 & 1 \\ 1/2 & -1 & 1/2 \\ 0 & 1/3 & -1/3 \end{pmatrix} \end{array}.$$

The linear system $\boldsymbol{\pi Q} = \mathbf{0}$ gives

$$\begin{aligned} -2\pi_e + (1/2)\pi_p \qquad\qquad &= 0, \\ \pi_e - \qquad \pi_p + (1/3)\pi_s &= 0, \\ \pi_e + (1/2)\pi_p - (1/3)\pi_s &= 0, \end{aligned}$$

with solution

$$\boldsymbol{\pi} = (\pi_e, \pi_p, \pi_s) = \left(\frac{1}{14}, \frac{4}{14}, \frac{9}{14}\right) = (0.071, 0.286, 0.643).$$

Jesses spends almost two-thirds of his day sleeping.

Using R, we compute the transition function, and then take $\boldsymbol{P}(t)$ for large t (in this case $t = 100$) to find the approximate limiting distribution. ■

R: Eat, Play, Sleep

```
> install.packages("expm")
> library(expm)
> Q = matrix(c(-2,1,1,1/2,-1,1/2,0,1/3,-1/3),
+  nrow=3,byrow=T)
> P <- function(t) {expm(t*Q)}
> P(100)
          Eat    Play   Sleep
Eat   0.07143 0.28571 0.64286
Play  0.07143 0.28571 0.64286
Sleep 0.07143 0.28571 0.64286
```

Example 7.15 (DNA evolution) The Jukes–Cantor and Kimura models, introduced in Example 7.11, have uniform limiting distribution $\boldsymbol{\pi} = (1/4, 1/4, 1/4, 1/4)$. An objection to these models is that nucleotides are not distributed uniformly on the genome. For instance, frequencies for human DNA are approximately

a	g	c	t
0.292	0.207	0.207	0.292

The four-parameter *Felsenstein* model, introduced in 1981, has generator matrix

$$\boldsymbol{Q} = \begin{array}{cc} & \begin{array}{cccc} a & g & c & t \end{array} \\ \begin{array}{c} a \\ g \\ c \\ t \end{array} & \begin{pmatrix} -\alpha(1-p_a) & \alpha p_g & \alpha p_c & \alpha p_t \\ \alpha p_a & \alpha(1-p_g) & \alpha p_c & \alpha p_t \\ \alpha p_a & \alpha p_g & -\alpha(1-p_c) & \alpha p_t \\ \alpha p_a & \alpha p_g & \alpha p_c & -\alpha(1-p_t) \end{pmatrix} \end{array},$$

where $p_a + p_g + p_c + p_t = 1$ and $\alpha > 0$. One checks that $\boldsymbol{\pi} = (p_a, p_g, p_c, p_t)$ satisfies $\boldsymbol{\pi Q} = \boldsymbol{0}$ and is the stationary distribution. The transition function, which we do not derive, is

$$\boldsymbol{P}_{ij}(t) = \begin{cases} (1-e^{-\alpha t})p_j, & \text{if } i \neq j, \\ e^{-\alpha t} + (1-e^{-\alpha t})p_j, & \text{if } i = j. \end{cases}$$

■

Absorbing States

An absorbing Markov chain is one in which there is at least one absorbing state. Let $(X_t)_{t\geq 0}$ be an absorbing continuous-time Markov chain on $\{1, \ldots, k\}$. For simplicity, assume the chain has one absorbing state a. As in the discrete case, the nonabsorbing states are *transient*. There is positive probability that the chain, started in a transient state, gets absorbed and never returns to that state. For transient state i, we derive an expression for a_i, the expected time until absorption.

Let T denote the set of transient states. Write the generator in canonical block matrix form

$$\boldsymbol{Q} = \begin{array}{cc} & \begin{array}{cc} a & T \end{array} \\ \begin{array}{c} a \\ T \end{array} & \begin{pmatrix} 0 & \boldsymbol{0} \\ * & \boldsymbol{V} \end{pmatrix} \end{array},$$

where $\boldsymbol{V}$ is a $(k-1) \times (k-1)$ matrix.

Mean Time Until Absorption

Theorem 7.3. *For an absorbing continuous-time Markov chain, define a square matrix $\boldsymbol{F}$ on the set of transient states, where F_{ij} is the expected time, for the chain started in i, that the process is in j until absorption. Then,*

$$\boldsymbol{F} = -\boldsymbol{V}^{-1}.$$

For the chain started in i, the mean time until absorption is,

$$a_i = \sum_j F_{ij}.$$

The matrix $\boldsymbol{F}$ is called the fundamental matrix.

Proof. The proof is based on conditioning on the first transition. For the chain started in i, consider the mean time F_{ij} spent in j before absorption.

Assume that $j \neq i$ and that the process first moves to $k \neq i$. The probability of moving from i to k is q_{ik}/q_i. If $k = a$ is the absorbing state, the time spent in j is 0. Otherwise, the mean time in j until absorption is F_{kj}. This gives

$$\begin{aligned} F_{ij} &= \sum_{\substack{k\in T\\ k\neq i}} \frac{q_{ik}}{q_i} F_{kj} = \frac{1}{q_i} \sum_{\substack{k\in T\\ k\neq i}} V_{ik} F_{kj} \\ &= \frac{1}{q_i}[(VF)_{ij} - V_{ii}F_{ij}] = \frac{1}{q_i}[(VF)_{ij} + q_i F_{ij}] \\ &= \frac{1}{q_i}(VF)_{ij} + F_{ij}. \end{aligned}$$

Thus $(VF)_{ij} = 0$. We have used that $V_{ij} = Q_{ij}$, for $i, j \in T$.

Assume that $j = i$. The mean time spent in i before transitioning to a new state is $1/q_i$. Conditioning on the next state,

$$\begin{aligned} F_{ii} &= \frac{q_{ia}}{q_i}\left(\frac{1}{q_i}\right) + \sum_{\substack{k\in T\\ k\neq i}} \frac{q_{ik}}{q_i}\left(\frac{1}{q_i} + F_{ki}\right) \\ &= \frac{q_{ia}}{q_i^2} + \frac{1}{q_i^2}\sum_{\substack{k\in T\\ k\neq i}} q_{ik} + \frac{1}{q_i}\sum_{\substack{k\in T\\ k\neq i}} q_{ik}F_{ki} \\ &= \frac{-1}{q_i^2} Q_{ii} + \frac{1}{q_i}[(VF)_{ii} - V_{ii}F_{ii}] \\ &= \frac{1}{q_i} + \frac{1}{q_i}(VF)_{ii} + F_{ii}. \end{aligned}$$

Hence, $(VF)_{ii} = -1$.

In summary,

$$(VF)_{ij} = \begin{cases} -1, & \text{if } i = j, \\ 0, & \text{if } i \neq j. \end{cases}$$

That is, $\boldsymbol{VF} = -\boldsymbol{I}$. and $\boldsymbol{F} = -\boldsymbol{V}^{-1}$.

For the chain started in i, the mean time until absorption is the sum of the mean times in each transient state j until absorption, which gives the final result. ∎

Example 7.16 (Multistate models) Multistate Markov models are used in medicine to model the course of diseases. A patient may advance into, or recover from, successively more severe stages of a disease until some terminal state. Each stage represents a state of an absorbing continuous-time Markov chain.

Bartolomeo et al. (2011) develops such a model to study the progression of liver disease among patients diagnosed with cirrhosis of the liver. The general form of the infinitesimal generator matrix for their three-parameter model is

$$\boldsymbol{Q} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{ccc} \quad 1 \quad & \quad 2 \quad & 3 \end{array} \\ \begin{pmatrix} -(q_{12}+q_{13}) & q_{12} & q_{13} \\ 0 & -q_{23} & q_{23} \\ 0 & 0 & 0 \end{pmatrix} \end{array},$$

where state 1 represents cirrhosis, state 2 denotes liver cancer (hepatocellular carcinoma), and state 3 is death. The fundamental matrix is

$$\boldsymbol{F} = -\begin{pmatrix} -(q_{12}+q_{13}) & q_{12} \\ 0 & -q_{23} \end{pmatrix}^{-1} = \begin{array}{c} \begin{array}{cc} \quad 1 \quad & \quad 2 \end{array} \\ \begin{pmatrix} 1/\left(q_{12}+q_{13}\right) & q_{12}/(q_{23}(q_{12}+q_{13})) \\ 0 & 1/q_{23} \end{pmatrix} \end{array},$$

with mean absorption times

$$a_1 = \frac{1}{q_{12}+q_{13}} + \frac{q_{12}}{q_{23}(q_{12}+q_{13})} \quad \text{and} \quad a_2 = \frac{1}{q_{23}}.$$

From a sample of 1,925 patients diagnosed with cirrhosis, and data on the number of months at each stage of the disease, the authors estimate the parameters of the model for subgroups depending on age, gender, and other variables. Their mean parameter estimates are $\widehat{q_{12}} = 0.0151$, $\widehat{q_{13}} = 0.0071$, and $\widehat{q_{23}} = 0.0284$.

Plots of the transition probabilities between states are shown in Figure 7.7. The fundamental matrix is estimated to be

$$\widehat{\boldsymbol{F}} = \begin{array}{c} \\ 1 \\ 2 \end{array} \begin{array}{c} \begin{array}{cc} \quad 1 \quad & \quad 2 \end{array} \\ \begin{pmatrix} 45.05 & 23.95 \\ 0.00 & 35.21 \end{pmatrix} \end{array}.$$

From the fundamental matrix, the estimated mean time to death for patients with liver cirrhosis is $45.05 + 23.95 = 69$ months. See the foregoing R code for a simulation of the mean time to death.

R: Simulation of Time to Absorption

```
# absorption.R
> trials <- 100000
> simlist <- numeric(trials)
> init <- 1   # initial state of liver cirrhosis
> for (i in 1:trials) {
+ state <- init
+ t <- 0
```

```
+ while (TRUE) {
+ if (state == 1) { q12 <- rexp(1,0.0151)
+   q13 <- rexp(1,0.0071) }
+   if (q12 < q13) {t <-t + q12
+   state <- 2}
+   else {t <- t + q13
+   break}
+ if (state == 2) {q23 <- rexp(1,0.0284)
+   t <- t + q23
+   break}
+ }
+ simlist[i] <- t  }
> mean(simlist)
[1] 69.01561
```

■

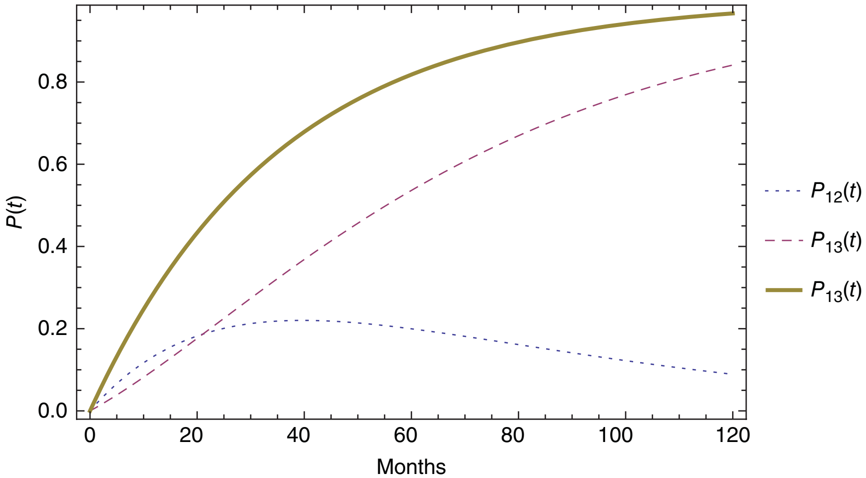

Figure 7.7 Estimated transition probabilities for stages of liver cirrhosis. *Source:* Bartolomeo et al. (2011).

Stationary Distribution of Embedded Chain

For a continuous-time Markov chain, the stationary distribution $\boldsymbol{\pi}$ is not the same as the stationary distribution of the embedded chain, which we denote as $\boldsymbol{\psi}$. Consider a three-state process with generator

$$\boldsymbol{Q} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{pmatrix} -2 & 1 & 1 \\ 2 & -4 & 2 \\ 4 & 4 & -8 \end{pmatrix} \end{array}.$$

The unique solution of $\pi Q = \mathbf{0}$ is $\pi = (4/7, 2/7, 1/7)$, which is the stationary distribution of the continuous-time chain. However, the embedded chain has transition matrix

$$\widetilde{P} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array}\!\!\begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{pmatrix} 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{pmatrix} \end{array},$$

with uniform stationary distribution $\psi = (1/3, 1/3, 1/3)$.

Note the difference in interpretation of the two distributions. The stationary probability π_j is the long-term proportion of *time* that the process spends in state j. On the other hand, the embedded chain stationary probability ψ_j is the long-term proportion of transitions that the process makes into state j.

Each stationary distribution can be derived from the other. From $\pi Q = \mathbf{0}$, we have that

$$\pi_j q_j = \sum_{i \neq j} \pi_i q_{ij}, \text{ for all } j. \tag{7.7}$$

Define $\widetilde{\psi}_j = \pi_j q_j$. Then, Equation (7.7) gives

$$\widetilde{\psi}_j = \sum_{i \neq j} \widetilde{\psi}_i p_{ij}.$$

Thus $\widetilde{\psi}$ satisfies $\widetilde{\psi}\widetilde{P} = \widetilde{\psi}$. To obtain a stationary distribution for the embedded chain, normalize $\widetilde{\psi}$. Let

$$\psi_j = \frac{\widetilde{\psi}_j}{\sum_k \widetilde{\psi_k}} = \frac{\pi_j q_j}{\sum_k \pi_k q_k}.$$

Then, $\psi = \psi\widetilde{P}$, and ψ is the stationary distribution of the embedded Markov chain.

Having derived ψ from π, we can also derive π from ψ. Since $\psi_j = C\pi_j q_j$, where C is an appropriate normalizing constant, we have that $\pi_j = \psi_j/(Cq_j)$. This gives,

$$\pi_j = \frac{\psi_j/q_j}{\sum_k \psi_k/q_k}, \text{ for all } j.$$

Example 7.17 The embedded chain of a continuous-time process has transition matrix

$$\widetilde{P} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array}\!\!\begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{pmatrix} 0 & 1 & 0 \\ 1/3 & 0 & 2/3 \\ 0 & 1 & 0 \end{pmatrix} \end{array}.$$

Assume that the process stays at state 1 on average 5 minutes before moving to 2. From state 2, it stays on average 2 minutes before moving to a new state. From 3, it stays on average 4 minutes before transitioning to 2. Find the stationary distribution of the continuous-time chain.

Solution Solve $\boldsymbol{\psi}\widetilde{\boldsymbol{P}} = \boldsymbol{\psi}$ to get $\boldsymbol{\psi} = (1/6, 1/2, 1/3)$. Holding time parameters are $(q_1, q_2, q_3) = (1/5, 1/2, 1/4)$. The stationary distribution $\boldsymbol{\pi}$ is proportional to

$$(\psi_1/q_1, \psi_2/q_2, \psi_3/q_3) = (5/6, 2/2, 4/3),$$

which gives $\boldsymbol{\pi} = (5/19, 6/19, 8/19)$. ■

Global Balance

Assume that $\boldsymbol{\pi}$ is the stationary distribution of a continuous-time Markov chain. From $\boldsymbol{\pi Q} = \mathbf{0}$, we have that

$$\sum_{i \neq j} \pi_i q_{ij} = \pi_j q_j, \text{ for all } j. \tag{7.8}$$

The holding time parameter q_j is the transition rate from j. Since π_j is the long-term proportion of time the process visits j, the right-hand side of Equation (7.8) is the long-term rate that the process leaves j. Also, $\pi_i q_{ij}$ is the long-term rate of transitioning from i to j. Thus, the left-hand side of Equation (7.8) is the long-term rate that the process enters j.

Equations (7.8) are known as the *global balance equations*. They say that in stationarity the rates in and out of any state are the same.

Example 7.18 In Example 7.17, we have that

$$(\pi_1 q_1, \pi_2 q_2, \pi_3 q_3) = (1/19, 3/19, 2/19).$$

By global balance, in the long-term, every 19 minutes the process sees one transition to and from state 1, three transitions to and from state 2, and two transitions to and from state 3.

We illustrate one of these facts by simulating the number of transitions to and from state 2 for the first 19 minutes of the stationary continuous-time chain.

R: Simulating Global Balance

```
# globalbalance.R
> trials <- 100000
> simin2 <- simout2 <- numeric(trials)
> for (i in 1:trials) {
+ state <- sample(1:3,1, prob=c(5/19,6/19,8/19))
+ t <- 0  # time
+ in2  <- 0  # counter for transitions to 2
+ out2 <- 0  # counter for transitions from 2
+ while (t < 19) {
```

```
+ if (state == 1) {t <- t+rexp(1,1/5)
+   if (t > 19) {break}
+   state <- 2
+   in2 <- in2 + 1 }
+ if (state == 3) {t <- t + rexp(1,1/4)
+   if (t > 19) {break}
+   state <- 2
+   in2 <- in2 + 1}
+ if (state ==2) { r1 <- rexp(1, (1/2)*(1/3))
+   r3 <- rexp(1, (1/2)*(2/3))
+   if (r1 < r3) { t <- t + r1
+   if (t > 19) {break}
+   out2 <- out2 + 1
+   state <- 1} else { t <-  t + r3
+   if (t > 19) {break}
+   out2 <- out2 + 1
+   state <- 3}}
+ }
+ simin2[i] <- in2
+ simout2[i] <- out2}
> mean(simin2)   # mean transitions to 2
[1] 2.99994
> mean(simout2)  # mean transitions from 2
[1] 3.00264
```

■

7.5 TIME REVERSIBILITY

Intuitively, a continuous-time Markov chain is *time reversible* if the process in forward time is indistinguishable from the process in reversed time. A consequence is that for all states i and j, the long-term *forward* transition rate from i to j is equal to the long-term *backward* rate from j to i. This is a stronger condition than global balance, which says that the long-term rate from a given state is equal to the long-term rate into that state.

Time Reversibility

A continuous-time Markov chain with generator $\boldsymbol{Q}$ and unique stationary distribution $\boldsymbol{\pi}$ is said to be *time reversible* if

$$\pi_i q_{ij} = \pi_j q_{ji}, \text{ for all } i,j. \tag{7.9}$$

Equations (7.9) give the *local balance*, or *detailed balance*, equations. They say that the long-term transition rate from i to j is equal to the long-term transition rate from j to i.

The local balance equations are a property of reversible chains, which can be used to find the stationary distribution. If a probability distribution λ satisfies

$$\lambda_i q_{ij} = \lambda_j q_{ji}, \text{ for all } i, j,$$

then the continuous-time chain is time reversible and $\boldsymbol{\pi} = \lambda$ is the unique stationary distribution of the chain. Summing over i gives

$$\sum_i \lambda_i q_{ij} = \lambda_j \sum_i q_{ji} = 0.$$

That is, $\boldsymbol{\lambda Q} = \mathbf{0}$.

Example 7.19 The general, nine-parameter, time reversible model for DNA substitutions has generator matrix

$$\boldsymbol{Q} = \begin{array}{c} \\ a \\ g \\ c \\ t \end{array} \begin{array}{c} \begin{array}{cccc} a & g & c & t \end{array} \\ \begin{pmatrix} - & \alpha p_g & \beta p_c & \gamma p_t \\ \alpha p_a & - & \delta p_c & \epsilon p_t \\ \beta p_a & \delta p_g & - & \eta p_t \\ \gamma p_a & \epsilon p_g & \eta p_c & - \end{pmatrix} \end{array},$$

where $p_a + p_g + p_c + p_t = 1$, and $\alpha, \beta, \gamma, \delta, \epsilon, \eta > 0$. Diagonal entries are chosen so that rows sum to 0. One checks that that $p_x q_{xy} = q_{yx} p_y$ for $x, y \in \{a, g, c, t\}$. It follows that the unique stationary distribution is $\boldsymbol{\pi} = (p_a, p_g, p_c, p_t)$. ■

A continuous-time Markov chain is time reversible if and only if its embedded discrete-time chain is time reversible. From Equation (7.9), we obtain that

$$\pi_i q_i p_{ij} = \pi_j q_j p_{ji}, \text{ for all } i, j.$$

Dividing both sides by the normalizing constant $\sum_k \pi_k q_k$ gives

$$\psi_i p_{ij} = \psi_j p_{ji}, \text{ for all } i, j.$$

That is, $\boldsymbol{\psi}$ satisfies the detailed balance equations for the embedded chain. In stationarity, the frequency of transitions from i to j is the same as that from j to i.

Tree Theorem

A *tree* is an undirected graph in which any pair of vertices is connected by exactly one path. Equivalently, a tree is a connected graph that has no cycles. Figure 7.8 shows three examples of trees.

Consider an undirected transition graph for a continuous-time Markov chain where vertices i and j are joined by an edge if $q_{ij} > 0$ or $q_{ji} > 0$. The next theorem, from

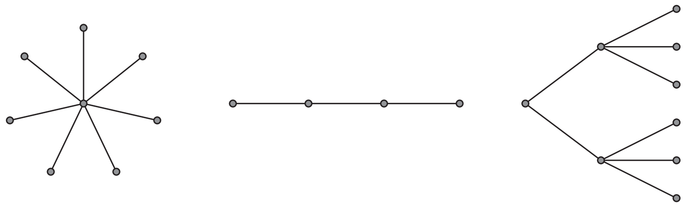

Figure 7.8 Trees.

Kelly (1994), gives a sufficient condition for time reversibility based on the structure of the graph.

Markov Processes on Trees are Time Reversible

Theorem 7.4. *Assume that the transition graph of an irreducible continuous-time Markov chain is a tree. Then, the process is time reversible.*

The result is shown by means of the following lemma, which is of independent interest. It states that the long-term transition rate out of any nonempty subset of states is equal to the long-term transition rate into that set.

Lemma 7.5. *Let $\mathcal{S}$ denote the state space of a continuous-time Markov chain with stationary distribution $\boldsymbol{\pi}$ and generator $\boldsymbol{Q}$. For any nonempty subset of the state space $A \subseteq \mathcal{S}$,*

$$\sum_{i \in A} \sum_{j \notin A} \pi_i q_{ij} = \sum_{i \in A} \sum_{j \notin A} \pi_j q_{ji}. \tag{7.10}$$

Proof of Lemma 7.5. The left-hand side of Equation (7.10) is equal to

$$\sum_{i \in A} \sum_{j \in S} \pi_i q_{ij} - \sum_{i \in A} \sum_{j \in A} \pi_i q_{ij} = -\sum_{i \in A} \sum_{j \in A} \pi_i q_{ij},$$

since the rows of $\boldsymbol{Q}$ sum to 0. The right-hand side of Equation (7.10) is equal to

$$\sum_{i \in A} \sum_{j \in S} \pi_j q_{ji} - \sum_{i \in A} \sum_{j \in A} \pi_j q_{ji} = -\sum_{i \in A} \sum_{j \in A} \pi_i q_{ij},$$

since $\boldsymbol{\pi}$ is the stationary distribution and $\boldsymbol{\pi Q} = \mathbf{0}$. ∎

Proof of Theorem 7.4. Assume that the transition graph is a tree. We show that

$$\pi_i q_{ij} = \pi_j q_{ji}, \text{ for all } i, j.$$

If there is no edge between i and j in the transition graph, then $q_{ij} = q_{ji} = 0$. Assume that there is an edge between i and j. A tree has the property, since it contains no cycles, that removal of any edge cuts the graph into two disconnected components. Let A denote the set of states connected to i, after removing the edge between i and j. Since the only vertex connected to i in the original graph, which is not in S, is j,

$$\sum_{k\in A}\sum_{l\notin A} \pi_k q_{kl} = \pi_i q_{ij}$$

and

$$\sum_{k\in A}\sum_{l\notin A} \pi_l q_{lk} = \pi_j q_{ji}.$$

By Lemma 7.5, the result is proven. ■

Theorem 7.4 gives a sufficient condition for a continuous-time Markov chain to be time reversible. However, the condition is not necessary. Following is an example of a reversible Markov chain whose undirected transition graph is not a tree.

Example 7.20 The continuous-time Markov chain with symmetric generator matrix

$$\boldsymbol{Q} = \begin{pmatrix} -2 & 1 & 1 \\ 1 & -2 & 1 \\ 1 & 1 & -2 \end{pmatrix}$$

has stationary distribution $\boldsymbol{\pi} = (1/3, 1/3, 1/3)$. One checks that the local balance equations are satisfied and the process is time reversible. However, the undirected transition graph is a triangle. ■

Birth-and-Death Process

Birth-and-death processes form a large class of time reversible, continuous-time Markov chains, which arise in many applications. For these processes, transitions only occur to neighboring states. *Births* occur from i to $i + 1$ at the rate λ_i. *Deaths* occur from i to $i - 1$ at the rate μ_i.

The generator matrix for the general birth-and-death process on $\{0, 1, \dots\}$ is

$$\boldsymbol{Q} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ \vdots \end{array} \begin{array}{c} \begin{array}{ccccc} 0 & 1 & 2 & 3 & \cdots \end{array} \\ \begin{pmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & \cdots \\ \mu_1 & -(\lambda_1+\mu_1) & \lambda_1 & 0 & \cdots \\ 0 & \mu_2 & -(\lambda_2+\mu_2) & \lambda_2 & \cdots \\ 0 & 0 & \mu_3 & -(\lambda_3+\mu_3) & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \end{array}.$$

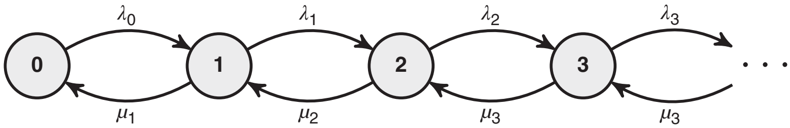

Figure 7.9 Birth-and-death process.

See Figure 7.9 for the directed transition graph. The undirected graph is a path, which is a tree. By Theorem 7.4, the chain is reversible.

The local balance equations for a birth-and-death process are

$$\pi_i \lambda_i = \pi_{i+1}\mu_{i+1}, \text{ for } i = 0, 1, \ldots$$

Solving for the stationary distribution,

$$\pi_1 = \pi_0 \frac{\lambda_0}{\mu_1},$$

$$\pi_2 = \pi_1 \frac{\lambda_1}{\mu_2} = \pi_0 \frac{\lambda_0 \lambda_1}{\mu_1 \mu_2},$$

and so on, giving

$$\pi_k = \pi_0 \frac{\lambda_0 \cdots \lambda_{k-1}}{\mu_1 \cdots \mu_k} = \pi_0 \prod_{i=1}^{k} \frac{\lambda_{i-1}}{\mu_i}, \text{ for } k = 0, 1, \ldots, \tag{7.11}$$

where we use the convention, for $k = 0$, that an empty product is equal to 1. For the components of $\boldsymbol{\pi}$ to sum to 1, we need

$$1 = \sum_{k=0}^{\infty} \pi_k = \pi_0 \sum_{k=0}^{\infty} \prod_{i=1}^{k} \frac{\lambda_{i-1}}{\mu_i}. \tag{7.12}$$

A necessary and sufficient condition for the stationary distribution to exist is that

$$\sum_{k=0}^{\infty} \prod_{i=1}^{k} \frac{\lambda_{i-1}}{\mu_i} < \infty,$$

in which case we can solve for π_0 in Equation (7.12).

Stationary Distribution for Birth-and-Death Process

For a birth-and-death process with birth rates λ_i and death rates μ_i, for $i = 1, 2, \ldots,$ assume that

$$\sum_{k=0}^{\infty} \prod_{i=1}^{k} \frac{\lambda_{i-1}}{\mu_i} < \infty.$$

Then, the unique stationary distribution $\boldsymbol{\pi}$ is

$$\pi_k = \pi_0 \prod_{i=1}^{k} \frac{\lambda_{i-1}}{\mu_i}, \text{ for } k = 1, 2, \ldots,$$

where

$$\pi_0 = \left(\sum_{k=0}^{\infty} \prod_{i=1}^{k} \frac{\lambda_{i-1}}{\mu_i} \right)^{-1}.$$

Example 7.21 (Continuous-time random walk) Consider a continuous-time version of simple random walk on $\{0, 1, 2, \ldots\}$ with reflecting boundaries. From 0, the walk moves to 1 after an exponentially distributed length of time with rate λ. From $i > 0$, transitions to the left occur at the rate μ, and transitions to the right occur at the rate λ. Find the stationary distribution.

Solution The walk is a birth-and-death process with constant birth rate $\lambda_i = \lambda$ and death rate $\mu_i = \mu$. For the stationary distribution,

$$\pi_k = \pi_0 \prod_{i=1}^{k} \frac{\lambda}{\mu} = \pi_0 \left(\frac{\lambda}{\mu} \right)^k, \text{ for } k = 0, 1, \ldots,$$

with

$$\pi_0 = \left(\sum_{k=0}^{\infty} \left(\frac{\lambda}{\mu} \right)^k \right)^{-1} = 1 - \frac{\lambda}{\mu},$$

provided that $\lambda < \mu$. In that case,

$$\pi_k = (1 - \lambda/\mu)(\lambda/\mu)^k, \text{ for } k = 0, 1, \ldots$$

For $\lambda < \mu$, the stationary distribution is the geometric distribution on $\{0, 1, \ldots\}$ with parameter $1 - \lambda/\mu$. ∎

Example 7.22 (Yule process) The Yule process arises in biology to describe the growth of a population where each individual gives birth to an offspring at a constant rate λ independently of other individuals. Let X_t denote the size of the population at time t. If $X_t = i$, then a new individual is born when one of the i members of the population gives birth, which occurs at rate $i\lambda$. A Yule process is a birth-and-death process with birth rate $\lambda_i = i\lambda$ and death rate $\mu_i = 0$. In a Yule process, all states are transient and no limiting distribution exists.

Assume that the initial size of the population is 1. The Yule process satisfies the Kolmogorov forward equations

$$P'_{1,j}(t) = -\lambda j P_{1,j}(t) + \lambda(j-1)P_{1,j-1}(t), \text{ for } j = 1, 2, \dots,$$

with

$$P_{1,1}(0) = 1 \quad \text{and} \quad P_{1,j} = 0, \text{ for } j \geq 2.$$

The solution to the system of differential equations, which may be verified directly, is

$$P_{1,j}(t) = e^{-\lambda t}(1 - e^{-\lambda t})^{j-1}, \text{ for } j = 1, 2 \dots,$$

which is a geometric distribution with parameter $e^{-\lambda t}$.

For the process started with i individuals, the transition function is

$$P_{ij}(t) = \binom{j-1}{i-1} e^{-\lambda i t}(1 - e^{-\lambda t})^{j-i}, \text{ for } j \geq i, \tag{7.13}$$

which is a negative binomial distribution. See Exercise 7.16. ■

Example 7.23 An academic support center has N tutors who help students with their homework. Students arrive at the center according to a Poisson process at rate λ. Each tutor takes an exponential length of time to work with students. Tutors' service times and student arrival times are independent. If all the tutors are busy when a student arrives at the center, the student will leave. Let X_t denote the number of tutors who are busy at time t. Find the stationary distribution.

Solution Assume that $i < N$ tutors are busy at time t. The number of busy tutors will increase by one if a student arrives, which occurs at rate λ. That is, $q_{i,i+1} = \lambda$, for $i = 0, \dots, N-1$.

On the other hand, if $i > 0$ tutors are busy, the number of busy tutors decreases by one if one of the i busy tutors finishes with their student. Each tutor's service time is exponentially distributed with parameter μ. Thus, the first time one of the i tutors is free is the minimum of i exponential random variables with parameter μ, which has an exponential distribution with parameter $i\mu$. This gives $q_{i,i-1} = i\mu$, for $i = 1, \dots, N$.

The process is a finite birth-and-death process with constant birth rate and linear death rate. The generator for $N = 4$ tutors is

$$\boldsymbol{Q} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{ccccc} 0 & \quad 1 & \qquad 2 & \qquad 3 & \qquad 4 \end{array} \\ \begin{pmatrix} -\lambda & \lambda & 0 & 0 & 0 \\ \mu & -(\lambda+\mu) & \lambda & 0 & 0 \\ 0 & 2\mu & -(\lambda+2\mu) & \lambda & 0 \\ 0 & 0 & 3\mu & -(\lambda+3\mu) & \lambda \\ 0 & 0 & 0 & 4\mu & -4\mu \end{pmatrix} \end{array}.$$

For general N, the stationary distribution is

$$\pi_k = \pi_0 \prod_{i=1}^{k} \frac{\lambda}{i\mu} = \pi_0 \frac{(\lambda/\mu)^k}{k!}, \text{ for } k = 0, 1, \dots, N,$$

with

$$\pi_0 = \left(\sum_{k=0}^{N} \prod_{i=1}^{k} \frac{\lambda}{i\mu} \right)^{-1} = \left(\sum_{k=0}^{N} \frac{(\lambda/\mu)^k}{k!} \right)^{-1}.$$

The distribution is a truncated Poisson distribution on $\{0, 1, \dots, N\}$ with parameter λ/μ. For large N, the distribution is an approximate Poisson distribution, and the long-term expected number of busy tutors is λ/μ. ■

Common birth-and-death processes are listed in Table 7.1.

TABLE 7.1 Types of Birth-and-Death Processes

Type	Birth Rate	Death Rate
Pure birth	λ_i	$\mu_i = 0$
Poisson process	$\lambda_i = \lambda$	$\mu_i = 0$
Pure death	$\lambda_i = 0$	μ_i
Linear process	$\lambda_i = i\lambda, i > 0$	$\mu_i = i\mu$
Yule process	$\lambda_i = \lambda i, i, \lambda > 0$	$\mu_i = 0$
Linear with immigration	$\lambda_i = i\lambda + \alpha, i, \alpha > 0$	$\mu_i = i\mu$

7.6 QUEUEING THEORY

Queueing theory is the study of waiting lines, or queues. In the terminology of queueing theory, *customers* arrive at a facility for *service*. If the service is not immediate, they wait for service, and leave the system when the service is complete. The framework is very general and could describe a diner waiting to be seated at a restaurant, a computer program waiting to be run, and a machine waiting to be repaired.

The general queueing model can be quite broad with many parameters, which describe things such as the distribution of arrival times, the distribution of service times, the number of servers, the capacity of the system, the stages of service, and how customers line up to be served.

A standard notation of the form A/B/*n* is used to describe a queueing model, where A denotes the arrival time distribution, B the service time distribution, and *n* the number of servers.

The M/M/1 queue is a basic model. The M stands for Markov or memoryless. In this model, both arrival and service times have exponential distributions, and there is one server. The M/M/1 queue is a birth-and-death process with constant birth and death rates.

The queueing models we explore in this section are continuous-time Markov chains where X_t denotes the number of customers in the system at time t.

A central result in queueing theory is Little's formula, which is deceptively simple and remarkably diverse.

Little's Formula

In a queueing system, let L denote the long-term average number of customers in the system, λ the rate of arrivals, and W the long-term average time that a customer is in the system. Then,

$$L = \lambda W.$$

The power of Little's formula is that it applies to a very broadly defined queueing system. We will not prove the formula but justify it intuitively with a *proof by picture*. The argument is based on Gross et al. (2008).

Consider a realization of a queueing system between the time when a customer first enters the system and when the system is next empty. Assume that four customers enter the system in the time interval $[0, t]$ with arrival and departure times

Customer	Arrival Time	Departure Time
1	t_1	t_5
2	t_2	t_4
3	t_3	t_7
4	t_6	t

See Figure 7.10. Let A be the area of the shaded region. Little's formula is obtained by computing A in two ways.

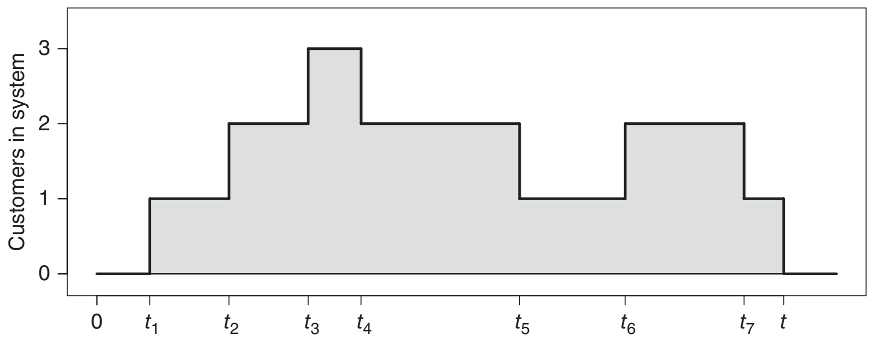

Figure 7.10 Little's formula is obtained by finding the area of the shaded region two ways.

The average length of time W that a customer spends in the system is

$$\begin{aligned} W &= \frac{(t_5 - t_1) + (t_4 - t_2) + (t_7 - t_3) + (t - t_6)}{4} \\ &= \frac{(t - t_1) + (t_5 - t_2) + (t_7 - t_6) + (t_4 - t_3)}{\text{Number of customers in } [0, t]} \\ &= \frac{A}{\text{Number of customers in } [0, t]}. \end{aligned} \tag{7.14}$$

In addition, the average number of customers L in the system is

$$\begin{aligned} L &= \frac{1}{t}[1(t_2 - t_1) + 2(t_3 - t_2) + 3(t_4 - t_3) + 2(t_5 - t_4) \\ &\quad +1(t_6 - t_5) + 2(t_7 - t_6) + 1(t - t_7)] \\ &= \frac{A}{t}. \end{aligned} \tag{7.15}$$

Equations (7.14) and (7.15) give

$$L = \frac{A}{t} = \frac{W \times \text{Number of customers in } [0, t]}{t}.$$

For large t, this gives $L = \lambda W$.

Example 7.24 (At the carwash) Cars arrive at a drive-through carwash according to a Poisson process at the rate of nine customers per hour. The time to wash a car has an exponential distribution with mean 5 minutes. Many questions can be asked.

1. How many cars, on average, are at the carwash?
2. How long, on average, is a customer at the carwash?
3. How long, on average, does a customer wait to be served?
4. What is the expected number of cars waiting to be served?

Solution Let X_t denote the number of cars in the system at hour t. The process is an M/M/1 queueing system, which is a birth-and-death process with constant birth and death rates. The arrival rate is $\lambda = 9$. Since the average time to wash a car is one-twelfth of an hour, the service rate is $\mu = 12$. The limiting distribution probabilities (see Example 7.21) are

$$\pi_k = \left(1 - \frac{\lambda}{\mu}\right)\left(\frac{\lambda}{\mu}\right)^k = \left(\frac{1}{4}\right)\left(\frac{3}{4}\right)^k, \text{ for } k = 0, 1, \dots,$$

which is a geometric distribution with parameter $p = 1 - \lambda/\mu = 1/4$.

1. The long-term expected number of cars at the carwash is the mean of the geometric distribution. A geometric distribution with parameter p, which takes values on $\{0, 1, \ldots\}$, has expectation $(1-p)/p$. The desired expectation is

$$\frac{\lambda/\mu}{1-\lambda/\mu} = \frac{\lambda}{\mu-\lambda} = 3.$$

On average, there will be three cars at the carwash. In the notation of Little's formula, $L = 3$.

2. By Little's formula, the long-term average time W that a customer is at the carwash is

$$W = \frac{L}{\lambda} = \frac{\lambda/(\mu-\lambda)}{\lambda} = \frac{1}{\mu-\lambda} = \frac{1}{3}.$$

A customer will be at the carwash, on average, for 20 minutes.

3. Let W_q denote the long-term average time a customer spends in the queue waiting to be served. Let W_s be the average time it takes for a car to be washed. Then, $W = W_q + W_s$. The service time for a car to be washed is the mean of an exponential distribution with parameter μ. Thus, $W_s = 1/\mu$. This gives

$$W_q = W - W_s = \frac{1}{\mu-\lambda} - \frac{1}{\mu} = \frac{\lambda}{\mu(\lambda-\mu)} = \frac{1}{4}.$$

A customer waits, on average, 15 minutes to be served.

4. We can consider the process restricted to just the queue as its own queueing system. Little's formula applies with $L_q = \lambda W_q$, where L_q denotes the long-term average number of cars waiting to be served. This gives

$$L_q = \lambda W_q = \frac{\lambda^2}{\mu(\lambda-\mu)} = \frac{9}{4}.$$

On average, there are 2.25 cars in the queue. ■

M/M/c Queue

An M/M/c queue has c servers. Consider the dynamics of the process. If there are $0 < k \le c$ customers then k servers are busy. The number of customers will decrease by one, the first time one of the servers completes their service. The time until that happens is the minimum of k independent exponential random variables, which has an exponential distribution with parameter $k\mu$. If there are more than c customers in the system, then the time until the first service time is complete is the minimum of c independent exponential random variables, and thus is exponentially distributed with parameter $c\mu$.

The M/M/c queue is a birth-and-death process with parameters $\lambda_i = \lambda$, for all i, and

$$\mu_i = \begin{cases} i\mu, & \text{for } i = 1, \ldots, c, \\ c\mu, & \text{for } i = c+1, c+2, \ldots \end{cases}$$

We have that

$$
\begin{aligned}
\sum_{k=0}^{\infty}\prod_{i=1}^{k}\frac{\lambda_i}{\mu_i} &= \sum_{k=0}^{c-1}\prod_{i=1}^{k}\frac{\lambda}{i\mu} + \sum_{k=c}^{\infty}\left(\prod_{i=1}^{c}\frac{\lambda}{i\mu}\right)\left(\prod_{i=c+1}^{k}\frac{\lambda}{c\mu}\right)\\
&= \sum_{k=0}^{c-1}\left(\frac{\lambda}{\mu}\right)^k\frac{1}{k!} + \frac{1}{c!}\sum_{k=c}^{\infty}\left(\frac{\lambda}{\mu}\right)^k\left(\frac{1}{c}\right)^{k-c}\\
&= \sum_{k=0}^{c-1}\left(\frac{\lambda}{\mu}\right)^k\frac{1}{k!} + \frac{(\lambda/\mu)^c}{c!}\sum_{k=c}^{\infty}\left(\frac{\lambda}{c\mu}\right)^{k-c}.
\end{aligned}
$$

The infinite sum converges for $0 < \lambda < c\mu$, in which case the stationary distribution $\boldsymbol{\pi}$ exists, with

$$
\pi_0 = \left(\sum_{k=0}^{c-1}\left(\frac{\lambda}{\mu}\right)^k\frac{1}{k!} + \frac{(\lambda/\mu)^c}{c!}\left(\frac{1}{1-\lambda/c\mu}\right)\right)^{-1}.
$$

The stationary probabilities are

$$
\pi_k = \begin{cases} \dfrac{\pi_0}{k!}\left(\dfrac{\lambda}{\mu}\right)^k, & \text{for } 0 \le k < c,\\[2ex] \dfrac{\pi_0}{c^{k-c}c!}\left(\dfrac{\lambda}{\mu}\right)^k, & \text{for } k \ge c. \end{cases}
$$

Example 7.25 (At the hair salon) A hair salon has five chairs. Customers arrive at the salon at the rate of 6 per hour. The hair stylists each take, on average, half an hour to service a customer, independent of arrival times.

1. Jill, the owner, wants to know the long-term probability that no customers are in the salon.
2. Danny, a potential customer, wants to know the average waiting time for a haircut.
3. Leslie, a hair stylist, wants to know the long-term expected number of customers in the salon.

Solution The system is an M/M/5 queue with $\lambda = 6$, $c = 5$, and $\mu = 2$.

1. The long-term probability that no customers are in the salon is

$$
\pi_0 = \left(\sum_{k=0}^{4}\frac{3^k}{k!} + \frac{3^5}{5!}\left(\frac{1}{1-6/10}\right)\right)^{-1} = \frac{16}{343} = 0.0466.
$$

2. The long-time average waiting time in the queue, in the notation of Little's formula, is $W_q = L_q/\lambda$. To find L_q, the expected number of customers in the queue, observe that there are k people in the queue if and only if there are $k + c$ customers in the system. This gives

$$\begin{aligned}
L_q &= \sum_{k=c}^{\infty}(k-c)\pi_k = \sum_{k=c}^{\infty}(k-c)\frac{\pi_0}{c^{k-c}c!}\left(\frac{\lambda}{\mu}\right)^k \\
&= \frac{\pi_0}{c!}\left(\frac{\lambda}{\mu}\right)^c \sum_{k=c}^{\infty}(k-c)\frac{1}{c^{k-c}}\left(\frac{\lambda}{\mu}\right)^{k-c} \\
&= \frac{\pi_0}{c!}\left(\frac{\lambda}{\mu}\right)^c \sum_{k=0}^{\infty}k\left(\frac{\lambda}{c\mu}\right)^k \\
&= \frac{\pi_0}{c!}\left(\frac{\lambda}{\mu}\right)^c \frac{\lambda}{c\mu}\left(\frac{1}{1-\lambda/c\mu}\right)^2 \\
&= \frac{\pi_0 3^5}{5!}\left(\frac{6}{10}\right)\left(\frac{1}{1-6/10}\right)^2 \\
&= 0.35423.
\end{aligned}$$

By Little's formula, the expected waiting time in the queue is

$$W_q = \frac{L_q}{\lambda} = \frac{0.35423}{6} = 0.059,$$

or about 3.6 minutes.

3. The long-term expected waiting time in the system is

$$W = W_q + W_s = W_q + \frac{1}{\mu} = 0.059 + 0.5 = 0.559,$$

or about 33.54 minutes. The expected number of customers in the system, by Little's formula, is

$$L = \lambda W = 6(0.559) = 3.354.$$

Note that the last result can also be obtained by finding the expectation with respect to the stationary distribution

$$L = \sum_{k=0}^{\infty} k\pi_k.$$

■

7.7 POISSON SUBORDINATION

The times when transitions occur for a continuous-time Markov chain are exponentially distributed with holding time rates $q_1, q_2, \ldots$ In this section, we present an interesting representation of a continuous-time Markov chain in which all holding time rates are the same, but where transitions from a state to itself are allowed. The representation is remarkably useful for simulation and numerical computation.

Consider a finite-state, irreducible, discrete-time Markov chain $Y_0, Y_1, \ldots$ with transition matrix $\boldsymbol{R}$. Let $(N_t)_{t\geq 0}$ be a Poisson process with parameter λ, which is independent of the Markov chain. Define a continuous-time process $(X_t)_{t\geq 0}$ by $X_t = Y_{N_t}$. That is, transitions for the X_t process occur at the arrival times of the Poisson process. From state i, the process holds an exponentially distributed amount of time with parameter λ and then transitions to j with probability R_{ij}.

The process $(X_t)_{t\geq 0}$ is a continuous-time Markov chain whose transition function $\boldsymbol{P}(t)$ has a surprisingly simple form. By conditioning on N_t,

$$\begin{aligned}
P_{ij}(t) &= P(X_t = j|X_0 = i)\\
&= \sum_{k=0}^{\infty} P(X_t = j|N_t = k, X_0 = i)P(N_t = k|X_0 = i)\\
&= \sum_{k=0}^{\infty} P(Y_k = j|N_t = k, X_0 = i)P(N_t = k)\\
&= \sum_{k=0}^{\infty} P(Y_k = j|Y_0 = i)P(N_t = k)\\
&= \sum_{k=0}^{\infty} R_{ij}^k \frac{e^{-\lambda t}(\lambda t)^k}{k!}.
\end{aligned}$$

We say that the $(X_t)_{t\geq 0}$ Markov chain is *subordinated to a Poisson process*.

Not only can we construct a continuous-time Markov chain from a discrete-time chain and a Poisson process, but conversely many continuous-time Markov chains can be represented as a chain subordinated to a Poisson process.

Consider a continuous-time Markov chain with generator $\boldsymbol{Q}$ and holding time parameters $q_1, q_2, \ldots$ Assume the parameters are uniformly bounded. That is, there exists a constant λ such that $q_i \leq \lambda$, for all i. This will always be the case if the chain is finite, and we can take $\lambda = \max_i q_i$. Let

$$\boldsymbol{R} = \frac{1}{\lambda}\boldsymbol{Q} + \boldsymbol{I}.$$

The matrix $\boldsymbol{R}$ is a stochastic matrix. Entries are non-negative and rows sum to 1. The transition function can be given in terms of $\boldsymbol{R}$, as

$$\begin{aligned}
\boldsymbol{P}(t) = e^{t\boldsymbol{Q}} &= e^{-\lambda t}e^{t\boldsymbol{Q}}e^{\lambda t} = e^{-\lambda t}e^{t(\boldsymbol{Q}+\lambda\boldsymbol{I})}\\
&= e^{-\lambda t}\sum_{k=0}^{\infty}\frac{1}{k!}t^k(\boldsymbol{Q}+\lambda\boldsymbol{I})^k\\
&= \sum_{k=0}^{\infty}\left(\frac{1}{\lambda}\boldsymbol{Q}+\boldsymbol{I}\right)^k\frac{e^{-\lambda t}(\lambda t)^k}{k!}
\end{aligned}$$

$$= \sum_{k=0}^{\infty} \boldsymbol{R}^k \frac{e^{-\lambda t}(\lambda t)^k}{k!}.$$

The continuous-time chain is represented as a Markov chain subordinated to a Poisson process. Other names for Poisson subordination are *randomization* and *uniformization*.

Note that the $\boldsymbol{R}$ matrix is *not* the matrix of the embedded Markov chain. The entries of the embedded Markov matrix are

$$\widetilde{P}_{ij} = \begin{cases} q_{ij}/q_i, & \text{for } i \neq j, \\ 0, & \text{for } i = j, \end{cases}$$

while the entries of the $\boldsymbol{R}$ matrix are

$$R_{ij} = \begin{cases} q_{ij}/\lambda, & \text{for } i \neq j, \\ 1 - q_i/\lambda, & \text{for } i = j. \end{cases}$$

Poisson subordination can be described as follows. From a given state i, wait an exponential length of time with rate λ. Then, flip a coin whose heads probability is q_i/λ. If heads, transition to a new state according to the $\boldsymbol{R}$ matrix. If tails, stay at i and repeat. Thus, holding time parameters are constant, and transitions, or pseudo-transitions, from a state to itself are allowed.

To illustrate, consider a continuous-time Markov chain on $\{a, b, c\}$ with generator

$$\boldsymbol{Q} = \begin{array}{c} \\ a \\ b \\ c \end{array} \!\!\begin{array}{c} \begin{array}{ccc} a & b & c \end{array} \\ \begin{pmatrix} -2 & 1 & 1 \\ 1 & -3 & 2 \\ 5 & 1 & -6 \end{pmatrix} \end{array}.$$

Choose $\lambda = \max\{2, 3, 6\} = 6$. Then,

$$\boldsymbol{R} = \frac{1}{\lambda}\boldsymbol{Q} + \boldsymbol{I} = \frac{1}{6}\begin{pmatrix} -2 & 1 & 1 \\ 1 & -3 & 2 \\ 5 & 1 & -6 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 2/3 & 1/6 & 1/6 \\ 1/6 & 1/2 & 1/3 \\ 5/6 & 1/6 & 0 \end{pmatrix}.$$

See Figure 7.11 for a comparison of the dynamics of the original chain with the subordinated process.

Long-Term Behavior, Simulation, Computation

For a Markov chain subordinated to a Poisson process, the discrete $\boldsymbol{R}$-chain has the same stationary distribution as the original chain. We have that

$$\boldsymbol{\pi Q} = \boldsymbol{\pi}\lambda(\boldsymbol{R} - \boldsymbol{I}) = \lambda\boldsymbol{\pi R} - \lambda\boldsymbol{\pi}.$$

Thus, $\boldsymbol{\pi Q} = \boldsymbol{0}$ if and only if $\boldsymbol{\pi R} = \boldsymbol{\pi}$.

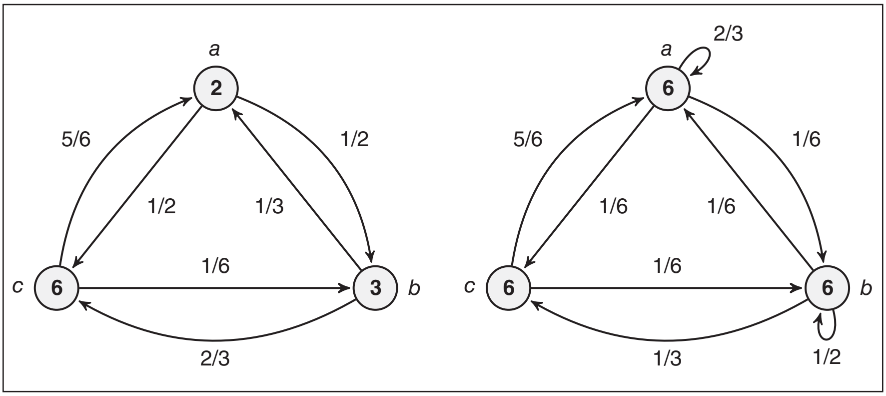

Figure 7.11 Both graphs describe the same Markov chain. Nodes are labeled with holding time parameters, edges are labeled with transition probabilities. The graph on the right shows the chain subordinated to a Poisson process. Holding time parameters are constant, and transitions to the same state are allowed.

Example 7.26 Consider a Markov chain $(X_t)_{t\geq 0}$ with generator

$$\boldsymbol{Q} = \begin{pmatrix} -2 & 1 & 1 \\ 1 & -3 & 2 \\ 0 & 1 & -1 \end{pmatrix}.$$

Letting $\lambda = \max\{q_1, q_2, q_3\} = 3$ gives

$$\boldsymbol{R} = \frac{1}{3}\boldsymbol{Q} + \boldsymbol{I} = \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 0 & 2/3 \\ 0 & 1/3 & 2/3 \end{pmatrix}.$$

Observe that $\boldsymbol{\pi} = (1/8, 2/8, 5/8)$ is the stationary distribution for both the original chain and the discrete-time $\boldsymbol{R}$-chain. ■

Poisson subordination leads to an efficient method to simulate a continuous-time Markov chain, since transitions rates are constant and thus not dependent on the current state. Here, we simulate the distribution of $X_{1.5}$ for the Markov chain in Example 7.26.

R: Simulation of the Distribution of $X_{1.5}$

```
# Psubordination.R
> Q
   1  2  3
1 -2  1  1
2  1 -3  2
3  0  1 -1
```

```
> lambda <- 3
> R <- (1/lambda)*Q+diag(3)
> R
          1         2         3
1 0.3333333 0.3333333 0.3333333
2 0.3333333 0.0000000 0.6666667
3 0.0000000 0.3333333 0.6666667
> trials <- 100000
> simlist <- numeric(trials)
> for (i in 1:trials) {
 +  s <- 0  # time
 +  state <- 1
 +  newstate <- 1
 +  while(s < 1.5) {
 +    state <- newstate
 +    s <- s+rexp(1,lambda)
 +    newstate <- sample(1:3,1,prob=r[state,]) }
 +    simlist[i] <- state
 +         }
> table(simlist)/trials
simlist
      1       2       3
0.16274 0.24958 0.58768
> expm(1.5*Q)[1,]  #   Compare with exact values
      1     2     3
1 0.163 0.249 0.588
```

Another benefit of Poisson subordination is that it gives a numerically stable method for computing a Markov transition function, as compared with the matrix exponential. For large matrices, the matrix exponential is notoriously difficult to compute. The infinite sum $\sum_{k=0}^{\infty}(t\boldsymbol{Q})^k/k!$ converges slowly. The matrix $\boldsymbol{Q}$ contains positive and negative numbers, some of which may be greater than 1, and thus $\boldsymbol{Q}^k$ may contain large positive and negative numbers, a source of numerical instability.

On the other hand, with Poisson subordination the transition function

$$P_{ij}(t) = \sum_{k=0}^{\infty} R_{ij}^k \frac{e^{-\lambda t}(\lambda t)^k}{k!} \tag{7.16}$$

can be computed numerically by truncating the infinite sum to a desired level of accuracy. Consider the approximation

$$\widehat{P_{ij}(t)} = \sum_{k=0}^{N} R_{ij}^k \frac{e^{-\lambda t}(\lambda t)^k}{k!}, \tag{7.17}$$

for some N. The absolute error of the approximation is

$$\begin{aligned}
|P_{ij}(t) - \widehat{P_{ij}(t)}| &= \left| \sum_{k=0}^{\infty} R_{ij}^k \frac{e^{-\lambda t}(\lambda t)^k}{k!} - \sum_{k=0}^{N} R_{ij}^k \frac{e^{-\lambda t}(\lambda t)^k}{k!} \right| \\
&= \sum_{k=N+1}^{\infty} R_{ij}^k \frac{e^{-\lambda t}(\lambda t)^k}{k!} \\
&\leq \sum_{k=N+1}^{\infty} \frac{e^{-\lambda t}(\lambda t)^k}{k!} \\
&= P(Y > N),
\end{aligned}$$

where Y is a Poisson random variable with parameter λt. The inequality holds since $\boldsymbol{R}^k$ is a stochastic matrix, all of whose entries are between 0 and 1.

To obtain an absolute error in the approximation of at most ϵ, choose N such that $P(Y > N) \leq \epsilon$.

The following example is small enough so that exact calculations are possible and numerical approximations are not necessary. However, it illustrates the general method.

Example 7.27 Consider a birth-and-death process on $\{0, \ldots, 5\}$ with constant death rate $\mu_i = 2$ and linear birth rates $\lambda_i = i$, for $i = 1, \ldots, 4$. The generator is

$$\boldsymbol{Q} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{cccccc} 0 & 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{pmatrix} -1 & 1 & 0 & 0 & 0 & 0 \\ 2 & -4 & 2 & 0 & 0 & 0 \\ 0 & 2 & -5 & 3 & 0 & 0 \\ 0 & 0 & 2 & -6 & 4 & 0 \\ 0 & 0 & 0 & 2 & -7 & 5 \\ 0 & 0 & 0 & 0 & 2 & -2 \end{pmatrix} \end{array}.$$

Find $\boldsymbol{P}(1.5)$ to within four-digit accuracy.

Solution Let $\lambda = 7$, and set

$$\boldsymbol{R} = \frac{1}{\lambda}\boldsymbol{Q} + \boldsymbol{I} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{cccccc} 0 & 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{pmatrix} 6/7 & 1/7 & 0 & 0 & 0 & 0 \\ 2/7 & 3/7 & 2/7 & 0 & 0 & 0 \\ 0 & 2/7 & 2/7 & 3/7 & 0 & 0 \\ 0 & 0 & 2/7 & 1/7 & 4/7 & 0 \\ 0 & 0 & 0 & 2/7 & 0 & 5/7 \\ 0 & 0 & 0 & 0 & 2/7 & 5/7 \end{pmatrix} \end{array}.$$

Choose N such that $P(Y > N) < 0.5 \times 10^{-4}$, where Y is a Poisson random variable with parameter $7 \times 1.5 = 10.5$. In R, we find that

```
> 1-ppois(24,10.5)
[1] 9.933169e-05
> 1-ppois(25,10.5)
[1] 3.921504e-05
```

Thus, truncate the infinite series at $N = 25$. This gives

$$\widehat{\boldsymbol{P}(1.5)} = \sum_{k=0}^{25} \boldsymbol{R}^k \frac{e^{-\lambda(1.5)}((1.5)t)^k}{k!}$$
$$= \begin{pmatrix} 0.50706 & 0.18466 & 0.09985 & 0.06535 & 0.05600 & 0.08703 \\ 0.36932 & 0.15278 & 0.10063 & 0.08480 & 0.09944 & 0.19299 \\ 0.19970 & 0.10063 & 0.08741 & 0.10025 & 0.15741 & 0.35457 \\ 0.08713 & 0.05653 & 0.06683 & 0.10241 & 0.19880 & 0.48826 \\ 0.03734 & 0.03315 & 0.05247 & 0.09940 & 0.21813 & 0.55948 \\ 0.02321 & 0.02573 & 0.04728 & 0.09765 & 0.22379 & 0.58230 \end{pmatrix}.$$

For this example, we can compare the numerical approximation with the matrix exponential computation $\boldsymbol{P}(1.5) = e^{1.5\boldsymbol{Q}}$, which can be found with software. The transition function, to five decimal places, is

$$\boldsymbol{P}(1.5) = \begin{pmatrix} 0.50708 & 0.18467 & 0.09985 & 0.06535 & 0.05601 & 0.08704 \\ 0.36933 & 0.15278 & 0.10064 & 0.08480 & 0.09945 & 0.19300 \\ 0.19970 & 0.10064 & 0.08741 & 0.10025 & 0.15742 & 0.35458 \\ 0.08714 & 0.05653 & 0.06683 & 0.10241 & 0.19881 & 0.48828 \\ 0.03734 & 0.03315 & 0.05247 & 0.09940 & 0.21814 & 0.55950 \\ 0.02321 & 0.02573 & 0.04728 & 0.09766 & 0.2238 & 0.58232 \end{pmatrix}.$$

By inspection, we see that our approximation is accurate to the desired level of accuracy.

By contrast, if we truncate the infinite sum in the matrix exponential to 25 terms we get

$$e^{1.5\boldsymbol{Q}} \approx \sum_{k=0}^{25} (1.5\boldsymbol{Q})^k / k!$$
$$= \begin{pmatrix} -4.06 & 22.01 & -69.10 & 170.51 & -284.73 & 166.39 \\ 44.01 & -208.29 & 661.43 & -1629.30 & 2723.19 & -1590.05 \\ -138.20 & 661.43 & -2099.20 & 5174.13 & -8648.70 & 5051.54 \\ 227.34 & -1086.20 & 3449.42 & -8503.51 & 14216.90 & -8302.93 \\ -189.82 & 907.73 & -2882.90 & 7108.44 & -11884.80 & 6942.40 \\ 44.37 & -212.01 & 673.54 & -1660.59 & 2776.96 & -1621.28 \end{pmatrix},$$

which is not close to converging. In fact, it takes almost twice as many terms in this case before reaching the desired level of accuracy. ■

EXERCISES

7.1 A continuous-time Markov chain has generator matrix

$$\boldsymbol{Q} = \begin{array}{c} \\ a \\ b \\ c \end{array}\begin{array}{c} \begin{array}{ccc} a & b & c \end{array} \\ \begin{pmatrix} -1 & 1 & 0 \\ 1 & -2 & 1 \\ 2 & 2 & -4 \end{pmatrix} \end{array}.$$

Exhibit (i) the transition matrix of the embedded Markov chain and (ii) the holding time parameter for each state.

7.2 A Markov chain on $\{1, 2, 3, 4\}$ has nonzero transition rates

$$q_{12} = q_{23} = q_{31} = q_{41} = 1 \quad \text{and} \quad q_{14} = q_{32} = q_{34} = q_{43} = 2.$$

(a) Exhibit the (i) generator, (ii) holding time parameters, and (iii) transition matrix for the embedded Markov chain.

(b) If the chain is at state 1, how long on average will it take before moving to a new state?

(c) If the chain is at state 3, how long on average will it take before moving to state 4?

(d) Over the long term, what proportion of visits will be to state 2?

7.3 A three-state Markov chain has distinct holding time parameters a, b, and c. From each state, the process is equally likely to transition to the other two states. Exhibit the generator matrix and find the stationary distribution.

7.4 During lunch hour, customers arrive at a fast-food restaurant at the rate of 120 customers per hour. The restaurant has one line, with three workers taking food orders at independent service stations. Each worker takes an exponentially distributed amount of time—on average 1 minute—to service a customer. Let X_t denote the number of customers in the restaurant (in line and being serviced) at time t. The process $(X_t)_{t\geq 0}$ is a continuous-time Markov chain. Exhibit the generator matrix.

7.5 For the fast-food restaurant chain of the previous exercise, assume that customers turn away from the store if all three service stations are busy. Let Y_t denote the number of service stations busy at time t. Then, $(Y_t)_{t\geq 0}$ is a continuous-time Markov chain. Exhibit the generator matrix.

7.6 A Markov chain $(X_t)_{t\geq 0}$ on $\{1, 2, 3, 4\}$ has generator matrix

$$\boldsymbol{Q} = \begin{pmatrix} -2 & 1 & 1 & 0 \\ 1 & -3 & 1 & 1 \\ 2 & 2 & -4 & 0 \\ 1 & 2 & 3 & 6 \end{pmatrix}.$$

Use technology as needed for the following:

(a) Find the long-term proportion of time that the chain visits state 1.

(b) For the chain started in state 2, find the long-term probability that the chain visits state 3.

(c) Find $P(X_1 = 3|X_0 = 1)$.

(d) Find $P(X_5 = 1, X_2 = 4|X_1 = 3)$.

7.7 A Markov chain has generator matrix

$$\boldsymbol{Q} = \begin{pmatrix} -1 & 1 & 0 \\ 0 & -2 & 2 \\ 3 & 0 & -3 \end{pmatrix}.$$

(a) Exhibit the Kolmogorov backward equations.

(b) Find the transition function by diagonalizing the generator and finding the matrix exponential.

(c) Find the transition function using a symbolic software system such as *Wolfram Alpha*.

7.8 Consider the Jukes–Cantor model for DNA nucleotide substitution in Example 7.11. Find the transition function $\boldsymbol{P}(t)$ by solving the backward, or forward, equations.

7.9 Consider the Felsenstein DNA model of Example 7.15.

(a) Show that $\boldsymbol{\pi} = (p_a, p_c, p_c, p_t)$ is the stationary distribution.

(b) For $\alpha = 2$ and the parameter values given in the example, find the probability that an a nucleotide mutates to a c nucleotide in 1.5 time units.

7.10 Unlike the Felsenstein DNA model, introduced in Example 7.15, the Hasegawa, Kishino, Yao model distinguishes between transitions and transversions. The generator matrix is

$$\boldsymbol{Q} = \begin{array}{c} \\ a \\ g \\ c \\ t \end{array} \begin{array}{c} \begin{array}{cccc} \quad a \quad & \quad g \quad & \quad c \quad & \quad t \quad \end{array} \\ \begin{pmatrix} -\alpha p_g - \beta p_r & \alpha p_g & \beta p_c & \beta p_t \\ \alpha p_a & -\alpha p_a - \beta p_r & \beta p_c & \beta p_t \\ \beta p_a & \beta p_g & -\alpha p_t - \beta p_s & \alpha p_t \\ \beta p_a & \beta p_g & \alpha p_c & -\alpha p_c - \beta p_s \end{pmatrix} \end{array},$$

where $p_r = p_c + p_t$, $p_s = p_a + p_g$, $p_a + p_g + p_c + p_t = 1$, and $\alpha, \beta > 0$. Show that $\boldsymbol{\pi} = (p_a, p_g, p_c, p_t)$ is the stationary distribution of the chain.

7.11 Let A be a square matrix. Show that

$$\frac{d}{dt} e^{tA} = \boldsymbol{A} e^{tA} = e^{tA} \boldsymbol{A}.$$

7.12 The following result from linear algebra relates the determinant and trace of a matrix $\boldsymbol{A}$:

$$\det e^{\boldsymbol{A}} = e^{\operatorname{tr} \boldsymbol{A}}.$$

Prove this for the case that $\boldsymbol{A}$ is diagonalizable.

7.13 Assume that $\boldsymbol{\pi}$ is the limiting distribution of a continuous-time chain. Show that $\boldsymbol{\pi}$ is a stationary distribution. (*Hint:* start with the forward equation.)

7.14 For the Markov chain with transition rate graph shown in Figure 7.12, find

(a) the generator matrix,

(b) the stationary distribution of the continuous-time Markov chain,

(c) the transition matrix of the embedded chain,

(d) the stationary distribution of the embedded chain.

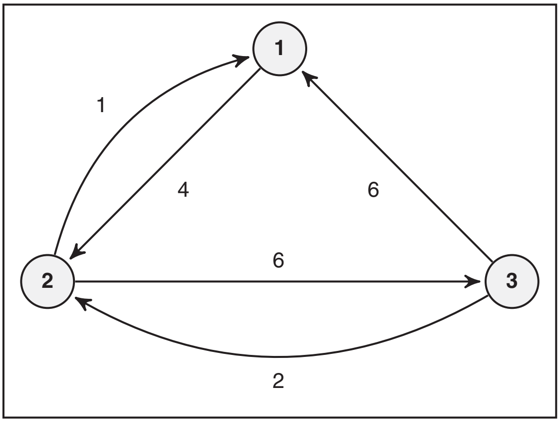

Figure 7.12

7.15 Let $(N_t)_{t\geq 0}$ be a Poisson process with parameter $\lambda = 1$. Define the process $X_t = N_t \bmod 4$, for $t \geq 0$. Then, $(X_t)_{t\geq 0}$ is a continuous-time Markov chain on $\{0, 1, 2, 3\}$.

(a) Exhibit the generator matrix.

(b) Use a symbolic software system such as *Wolfram Alpha* to find the transition function $\boldsymbol{P}(t)$.

7.16 For a Yule process started with i individuals, derive the transition probabilities

$$P_{ij}(t) = \binom{j-1}{i-1} e^{-\lambda i t}(1 - e^{-\lambda t})^{j-i}, \text{ for } j \geq i.$$

Hint: Relate the process to a Yule process started with one individual.

7.17 Each individual in a population gives birth to an offspring at the rate of 1.5 per unit time independently of other individuals. If the population starts with 4 individuals, find the mean and variance of the size of the population at time $t = 8$.

7.18 For a general birth-and-death process with birth rates λ_i and death rates μ_i, let T_i denote the time, from state i, for the process to hit state $i + 1$.

(a) Show that

$$E[T_i] = \frac{1}{\lambda_i} + \frac{\mu_i}{\lambda_i} E[T_{i-1}], \text{ for } i = 1, 2, \ldots$$

Hint: Condition on whether the first transition is a birth or a death.

(b) Solve $E[T_i]$ for the case $\lambda_i = \lambda$ and $\mu_i = \mu$, for all i.

7.19 Taxis arrive at a taxi stand according to a Poisson process with parameter λ. Customers arrive, independently of taxis, at the rate μ. If there are no taxis when a customer arrives at the stand they will leave. Assume that $\lambda < \mu$. What is the long-term probability that an arriving customer gets a taxi?

7.20 The M/M/∞ queue has infinitely many servers. Show that the limiting distribution is Poisson and find the mean number of customers in the system.

7.21 For an M/M/∞ queue with $\lambda = \mu = 1$, find the mean time until state 4 is hit for the process started in state 1.

7.22 Tom is taking an online exam in which there are four questions of increasing difficulty. Tom needs to complete each question before moving on to the next question. It takes on average 10, 20, 30, and 40 minutes, respectively, to answer each question. The times spent for each question are independent and exponentially distributed.

(a) Find the probability that after 45 minutes Tom has completed the exam.

(b) Find the probability that after 45 minutes Tom is still on the third question.

7.23 A facility has four machines, with two repair workers to maintain them. Individual machines fail on average every 10 hours. It takes an individual repair worker on average 4 hours to fix a machine. Repair and failure times are independent and exponentially distributed.

(a) Find the generator matrix.

(b) In the long term, how many machines are typically operational?

(c) If all four machines are initially working, find the probability that only two machines are working after 5 hours.

7.24 Customers arrive at a busy food truck according to a Poisson process with parameter λ. If there are i people already in line, the customer will join the line with probability $1/(i + 1)$. Assume that the chef at the truck takes, on average, α minutes to process an order.

(a) Find the long-term average number of people in line.

(b) Find the long-term probability that there are at least two people in line.

7.25 Over the long term, a continuous-time Markov chain on $\{1,2,3,4\}$ makes 10% of its transitions to 1 and 30% each to 2, 3, and 4, respectively. From state i, it stays on average i minutes before moving to a new state, for $i = 1,2,3,4$.

(a) Find the stationary distribution of the embedded discrete-time chain.

(b) Find the stationary distribution of the continuous-time chain.

7.26 A facility has three machines and three mechanics. Machines break down at the rate of one per 24 hours. Breakdown times are exponentially distributed. The time it takes a mechanic to fix a machine is exponentially distributed with mean 6 hours. Only one mechanic can work on a failed machine at any given time. Let X_t be the number of machines working at time t. Find the long-term probability that all machines are working.

7.27 Recall the discrete-time Ehrenfest dog–flea model of Example 3.7. In the continuous-time version, there are N fleas distributed between two dogs. Fleas jump from one dog to another independently at rate λ. Let X_t denote the number of fleas on the first dog.

(a) Show that the process is a birth-and-death process. Give the birth and death rates.

(b) Find the stationary distribution.

(c) Assume that fleas jump at the rate of 2 per minute. If there are 10 fleas on Cooper and no fleas on Lisa, how long, on average, will it take for Lisa to get 4 fleas?

7.28 A linear birth-and-death process with immigration (see Table 7.1) has parameters $\lambda = 3$, $\mu = 4$, and $\alpha = 2$. Find the stationary distribution.

7.29 Consider an absorbing, continuous-time Markov chain with possibly more than one absorbing states.

(a) Argue that the continuous-time chain is absorbed in state a if and only if the embedded discrete-time chain is absorbed in state a.

(b) Let

$$\boldsymbol{Q} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & -3 & 2 & 0 & 0 \\ 0 & 2 & -4 & 2 & 0 \\ 0 & 0 & 2 & -5 & 3 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \end{array}$$

be the generator matrix for a continuous-time Markov chain. For the chain started in state 2, find the probability that the chain is absorbed in state 5.

7.30 Cars arrive at a toll booth according to a Poisson process at the rate of two cars per minute. The time taken by the attendant to collect the toll is exponentially distributed with mean 20 seconds.

(a) Find the long-term mean number of cars in line at the toll booth.

(b) Find the long-term probability that there are more than three cars at the toll booth.

7.31 See the last exercise. Assume that the toll booth has two attendants on duty. Find the long-term probability there are no cars at the toll booth.

7.32 Calls come in to a computer help center at the rate of 15 calls per hour. There are three tech support workers on duty, and the times they take to provide assistance are exponentially distributed with a mean of 10 minutes.

(a) Find the average number of callers waiting to be helped.

(b) Find the average amount of time that a caller spends waiting.

7.33 Consider an M/M/1 queue. Assume that the arrival and service time rates are both increased by a factor of k. What effect does this have on

(a) the long-term expected number of customers in the system?

(b) the long-term expected time that a customer is in the system?

7.34 You are a frequent customer at a coffee shop, where you typically wait 3 minutes to be served. Furthermore, on average you spend \$4 per visit. Over many months you estimate that on a typical day there are 20 customers in the shop, which is open from 6 a.m. to 10 p.m. Estimate the shop's total revenue per day.

7.35 Consider a continuous-time Markov chain with generator

$$\boldsymbol{Q} = \begin{pmatrix} -1 & 0 & 1 \\ 1 & -2 & 1 \\ 1 & 3 & -4 \end{pmatrix}.$$

Represent the process as a Markov chain subordinated to a Poisson process. Exhibit the transition function $\boldsymbol{P}(t)$ in terms of $\boldsymbol{R}$.

7.36 A discrete-time Markov chain has transition matrix

$$\boldsymbol{P} = \begin{pmatrix} p & q \\ p & q \end{pmatrix},$$

for $p + q = 1$. Extend the process to continuous time by allowing transitions to occur at the points of a Poisson process with parameter λ. Find the transition function $\boldsymbol{P}(t)$.

7.37 R: Consider a continuous-time Markov chain with generator

$$\boldsymbol{Q} = \begin{pmatrix} -4 & 1 & 2 & 1 \\ 2 & -3 & 0 & 1 \\ 3 & 3 & -9 & 3 \\ 4 & 2 & 0 & -6 \end{pmatrix}.$$

(a) To use Poisson subordination to estimate $\boldsymbol{P}(0.8)$ to three significant digits, how many terms of the series are needed?

(b) Use (a) and compute $\boldsymbol{P}(0.8)$.

(c) Use `R`'s matrix exponential command to find $\boldsymbol{P}(0.8)$ and check that your result in (a) is accurate to three digits.

7.38 `R`: Simulate Tom's online exam in Exercise 7.22 and estimate the probabilities in that problem.

7.39 `R`: Simulate an M/M/∞ queue and verify the result of Exercise 7.20, choosing your own values for λ and μ.

7.40 `R`: A multistate Markov model for the progression of HIV infection is developed in Hendricks et al. (1996). A sample of 467 HIV-positive men were monitored between 1984 and 1993. A 7-state absorbing chain is developed where stages 1–6 represent a range of values of an immunological marker for HIV, and state 7 corresponds to AIDS. The researchers give the following estimates for the monthly transition rates of the model.

λ_{12}	λ_{21}	λ_{23}	λ_{32}	λ_{34}	λ_{43}	λ_{45}	λ_{47}	λ_{54}	λ_{56}	λ_{57}	λ_{65}	λ_{67}
0.055	0.008	0.060	0.008	0.039	0.008	0.033	0.006	0.009	0.029	0.007	0.002	0.042

(a) Find the mean time to develop AIDS from each HIV state.

(b) Starting from the first stage of HIV infection, estimate the probability of developing AIDS within k years, for $k = 5, 10, 15, 20$.