6 Continuous Time Markov Chains

6.1 Pure Birth Processes

In this chapter, we present several important examples of continuous time, discrete state, and Markov processes. Specifically, we deal here with a family of random variables $\{X(t); 0 \leq t < \infty\}$ where the possible values of $X(t)$ are the nonnegative integers. We shall restrict attention to the case where $\{X(t)\}$ is a Markov process with stationary transition probabilities. Thus, the transition probability function for $t > 0$,

$$P_{ij}(t) = \Pr\{X(t+u) = j | X(u) = i\}, \qquad i, j = 0, 1, 2, \ldots,$$

is independent of $u \geq 0$.

It is usually more natural in investigating particular stochastic models based on physical phenomena to prescribe the so-called infinitesimal probabilities relating to the process and then derive from them an explicit expression for the transition probability function. For the case at hand, we will postulate the form of $P_{ij}(h)$ for h small, and, using the Markov property, we will derive a system of differential equations satisfied by $P_{ij}(t)$ for all $t > 0$. The solution of these equations under suitable boundary conditions gives $P_{ij}(t)$.

By way of introduction to the general pure birth process, we review briefly the axioms characterizing the Poisson process.

6.1.1 Postulates for the Poisson Process

The Poisson process is the prototypical pure birth process. Let us point out the relevant properties. The Poisson process is a Markov process on the nonnegative integers for which

(i) $\Pr\{X(t+h) - X(t) = 1 | X(t) = x\} = \lambda h + o(h)$ as $h \downarrow 0$
$(x = 0, 1, 2, \ldots)$.

(ii) $\Pr\{X(t+h) - X(t) = 0 | X(t) = x\} = 1 - \lambda h + o(h)$ as $h \downarrow 0$.

(iii) $X(0) = 0$.

The precise interpretation of (i) is the relationship

$$\lim_{h \to 0+} \frac{\Pr\{X(t+h) - X(t) = 1 | X(t) = x\}}{h} = \lambda.$$

The $o(h)$ symbol represents a negligible remainder term in the sense that if we divide the term by h, then the resulting value tends to zero as h tends to zero. Notice that the right side of (i) is independent of x.

An Introduction to Stochastic Modeling. DOI: 10.1016/B978-0-12-381416-6.00006-X
© 2011 Elsevier Inc. All rights reserved.

These properties are easily verified by direct computation, since the explicit formulas for all the relevant properties are available. Problem 6.1.13 calls for showing that these properties, in fact, define the Poisson process.

6.1.2 Pure Birth Process

A natural generalization of the Poisson process is to permit the chance of an event occurring at a given instant of time to depend upon the number of events that have already occurred. An example of this phenomenon is the reproduction of living organisms (and hence the name of the process), in which under certain conditions—e.g., sufficient food, no mortality, no migration—the infinitesimal probability of a birth at a given instant is proportional (directly) to the population size at that time. This example is known as the *Yule process* and will be considered in detail later.

Consider a sequence of positive numbers, $\{\lambda_k\}$. We define a pure birth process as a Markov process satisfying the following postulates:

1. $\Pr\{X(t+h)-X(t)=1|X(t)=k\}=\lambda_k h+o_{1,k}(h)(h\to 0+)$.
2. $\Pr\{X(t+h)-X(t)=0|X(t)=k\}=1-\lambda_k h+o_{2,k}(h)$. (6.1)
3. $\Pr\{X(t+h)-X(t)<0|X(t)=k\}=0\ (k\geq 0)$.
As a matter of convenience, we often add the postulate
4. $X(0)=0$.

With this postulate, $X(t)$ does not denote the population size but, rather, the number of births in the time interval $(0,t]$.

Note that the left sides of Postulates (1) and (2) are just $P_{k,k+1}(h)$ and $P_{k,k}(h)$, respectively (owing to stationarity), so that $o_{1,k}(h)$ and $o_{2,k}(h)$ do not depend upon t.

We define $P_n(t)=\Pr\{X(t)=n\}$, assuming $X(0)=0$.

By analyzing the possibilities at time t just prior to time $t+h$ (h small), we will derive a system of differential equations satisfied by $P_n(t)$ for $t\geq 0$, namely

$$\begin{aligned} P_0'(t)&=-\lambda_0 P_0(t),\\ P_n'(t)&=-\lambda_n P_n(t)+\lambda_{n-1}P_{n-1}(t) \qquad \text{for } n\geq 1, \end{aligned} \tag{6.2}$$

with initial conditions

$$P_0(0)=1, \qquad P_n(0)=0, \qquad n>0.$$

Indeed, if $h>0, n\geq 1$, then by invoking the law of total probability, the Markov property, and Postulate (3), we obtain

$$\begin{aligned} P_n(t+h)&=\sum_{k=0}^{\infty}P_k(t)\Pr\{X(t+h)=n|X(t)=k\}\\ &=\sum_{k=0}^{\infty}P_k(t)\Pr\{X(t+h)-X(t)=n-k|X(t)=k\}\\ &=\sum_{k=0}^{n}P_k(t)\Pr\{X(t+h)-X(t)=n-k|X(t)=k\}. \end{aligned}$$

Now for $k = 0, 1, \ldots, n-2$, we have

$$\begin{aligned}
&\Pr\{X(t+h) - X(t) = n-k | X(t) = k\} \\
&\quad \le \Pr\{X(t+h) - X(t) \ge 2 | X(t) = k\} \\
&\quad = o_{1,k}(h) + o_{2,k}(h),
\end{aligned}$$

or

$$\Pr\{X(t+h) - X(t) = n-k | X(t) = k\} = o_{3,n,k}(h), \qquad k = 0, \ldots, n-2.$$

Thus,

$$\begin{aligned}
P_n(t+h) = {} & P_n(t)\left[1 - \lambda_n h + o_{2,n}(h)\right] + P_{n-1}(t)\left[\lambda_{n-1}h + o_{1,n-1}(h)\right] \\
& + \sum_{k=0}^{n-2} P_k(t) o_{3,n,k}(h) k,
\end{aligned}$$

or

$$\begin{aligned}
&P_n(t+h) - P_n(t) \\
&\quad = P_n(t)\left[-\lambda_n h + o_{2,n}(h)\right] + P_{n-1}(t)\left[\lambda_{n-1}h + o_{1,n-1}(h)\right] + o_n(h),
\end{aligned} \tag{6.3}$$

where, clearly, $\lim_{h\downarrow 0} o_n(h)/h = 0$ uniformly in $t \ge 0$, since $o_n(h)$ is bounded by the finite sum $\Sigma_{k=0}^{n-2} o_{3,n,k}(h)$, which does not depend on t.

Dividing by h and passing to the limit $h \downarrow 0$, we validate the relations (6.2), where on the left side we should, to be precise, write the derivative from the right. With a little more care, however, we can derive the same relation involving the derivative from the left. In fact, from (6.3), we see at once that the $P_n(t)$ are continuous functions of t. Replacing t by $t - h$ in (6.3), dividing by h, and passing to the limit $h \downarrow 0$, we find that each $P_n(t)$ has a left derivative that also satisfies equation (6.2).

The first equation of (6.2) can be solved immediately and yields

$$P_0(t) = \exp\{-\lambda_0 t\} \qquad \text{for } t > 0. \tag{6.4}$$

Define S_k as the time between the kth and the $(k+1)$st birth, so that

$$P_n(t) = \Pr\left\{\sum_{i=0}^{n-1} S_i \le t < \sum_{i=0}^{n} S_i\right\}.$$

The random variables S_k are called the "sojourn times" between births, and

$$W_k = \sum_{i=0}^{k-1} S_i = \text{the time at which the } k\text{th birth occurs.}$$

We have already seen that $P_0(t) = \exp\{-\lambda_0 t\}$. Therefore,

$$\Pr\{S_0 \leq t\} = 1 - \Pr\{X(t) = 0\} = 1 - \exp\{-\lambda_0 t\};$$

that is S_0 has an exponential distribution with parameter λ_0. It may be deduced from Postulates (1) through (4) that $S_k, k > 0$, also has an exponential distribution with parameter λ_k and that the S_i's are mutually independent.

This description characterizes the pure birth process in terms of its sojourn times, in contrast to the infinitesimal description corresponding to (6.1).

To solve the differential equations of (6.2) recursively, introduce $Q_n(t) = e^{\lambda_n t} P_n(t)$ for $n = 0, 1, \ldots$. Then,

$$\begin{aligned} Q_n'(t) &= \lambda_n e^{\lambda_n t} P_n(t) + e^{\lambda_n t} P_n'(t) \\ &= e^{\lambda_n t} \left[\lambda_n P_n(t) + P_n'(t) \right] \\ &= e^{\lambda_n t} \lambda_{n-1} P_{n-1}(t) \qquad [\text{using (6.2)}]. \end{aligned}$$

Integrating both sides of these equations and using the boundary condition $Q_n(0) = 0$ for $n \geq 1$ gives

$$Q_n(t) = \int_0^t e^{\lambda_n x} \lambda_{n-1} P_{n-1}(x) dx,$$

or

$$P_n(t) = \lambda_{n-1} e^{-\lambda_n t} \int_0^t e^{\lambda_n x} P_{n-1}(x) dx, \qquad n = 1, 2, \ldots. \tag{6.5}$$

It is now clear that all $P_k(t) \geq 0$, but there is still a possibility that

$$\sum_{n=0}^{\infty} P_n(t) < 1.$$

To secure the validity of the process, i.e., to assure that $\Sigma_{n=0}^{\infty} P_n(t) = 1$ for all t, we must restrict the λ_k according to the following:

$$\sum_{n=0}^{\infty} P_n(t) = 1 \quad \text{if and only if} \sum_{n=0}^{\infty} \frac{1}{\lambda_n} = \infty. \tag{6.6}$$

The intuitive argument for this result is as follows: The time S_k between consecutive births is exponentially distributed with a corresponding parameter λ_k. Therefore, the quantity $\Sigma_n 1/\lambda_n$ equals the expected time before the population becomes infinite. By comparison, $1 - \Sigma_{n=0}^{\infty} P_n(t)$ is the probability that $X(t) = \infty$.

If $\Sigma_n \lambda_n^{-1} < \infty$, then the expected time for the population to become infinite is finite. It is then plausible that for all $t > 0$, the probability that $X(t) = \infty$ is positive.

When no two of the birth parameters $\lambda_0, \lambda_1, \ldots$ are equal, the integral equation (6.5) may be solved to give the explicit formula

$$\begin{aligned} P_0(t) &= \mathrm{e}^{-\lambda_0 t}, \\ P_1(t) &= \lambda_0 \left(\frac{1}{\lambda_1 - \lambda_0} \mathrm{e}^{-\lambda_0 t} + \frac{1}{\lambda_0 - \lambda_1} \mathrm{e}^{-\lambda_1 t} \right) \end{aligned} \tag{6.7}$$

and

$$\begin{aligned} P_n(t) &= \Pr\{X(t) = n | X(0) = 0\} \\ &= \lambda_0 \cdots \lambda_{n-1} \left[B_{0,n} \mathrm{e}^{-\lambda_0 t} + \cdots + B_{n,n} \mathrm{e}^{-\lambda_n t} \right] \qquad \text{for } n > 1, \end{aligned} \tag{6.8}$$

where

$$\begin{aligned} B_{0,n} &= \frac{1}{(\lambda_1 - \lambda_0) \cdots (\lambda_n - \lambda_0)}, \\ B_{k,n} &= \frac{1}{(\lambda_0 - \lambda_k) \cdots (\lambda_{k-1} - \lambda_k)(\lambda_{k+1} - \lambda_k) \cdots (\lambda_n - \lambda_k)} \\ & \qquad \text{for } 0 < k < n \end{aligned} \tag{6.9}$$

and

$$B_{n,n} = \frac{1}{(\lambda_0 - \lambda_n) \cdots (\lambda_{n-1} - \lambda_n)}.$$

Because $\lambda_j \neq \lambda_k$ when $j \neq k$ by assumption, the denominator in (6.9) does not vanish, and $B_{k,n}$ is well defined.

We will verify that $P_1(t)$, as given by (6.7), satisfies (6.5). Equation (6.4) gives $P_0(t) = \mathrm{e}^{-\lambda_0 t}$. We next substitute this in (6.5) when $n = 1$, thereby obtaining

$$\begin{aligned} P_1(t) &= \lambda_0 \mathrm{e}^{-\lambda_1 t} \int_0^t \mathrm{e}^{\lambda_1 x} \mathrm{e}^{-\lambda_0 x} \mathrm{d}x \\ &= \lambda_0 \mathrm{e}^{-\lambda_1 t} (\lambda_0 - \lambda_1)^{-1} \left[1 - \mathrm{e}^{-(\lambda_0 - \lambda_1)t} \right] \\ &= \lambda_0 \left(\frac{1}{\lambda_1 - \lambda_0} \mathrm{e}^{-\lambda_0 t} + \frac{1}{\lambda_0 - \lambda_1} \mathrm{e}^{-\lambda_1 t} \right), \end{aligned}$$

in agreement with (6.7).

The induction proof for general n involves tedious and difficult algebra. The case $n = 2$ is suggested as a problem.

6.1.3 The Yule Process

The Yule process arises in physics and biology and describes the growth of a population in which each member has a probability $\beta h + o(h)$ of giving birth to a new member during an interval of time of length $h(\beta > 0)$. Assuming independence and no interaction among members of the population, the binomial theorem gives

$$\begin{aligned}\Pr\{X(t+h) - X(t) = 1 | X(t) = n\} &= \binom{n}{1} [\beta h + o(h)] [1 - \beta h + o(h)]^{n-1} \\ &= n\beta h + o_n(h);\end{aligned}$$

for the Yule process the infinitesimal parameters are $\lambda_n = n\beta$. In words, the total population birth rate is directly proportional to the population size, the proportionality constant being the individual birth rate β. As such, the Yule process forms a stochastic analog of the deterministic population growth model represented by the differential equation $\mathrm{d}y/\mathrm{d}t = \alpha y$. In the deterministic model, the rate $\mathrm{d}y/\mathrm{d}t$ of population growth is directly proportional to population size y. In the stochastic model, the infinitesimal deterministic increase $\mathrm{d}y$ is replaced by the probability of a unit increase during the infinitesimal time interval $\mathrm{d}t$. Similar connections between deterministic rates and birth (and death) parameters arise frequently in stochastic modeling. Examples abound in this chapter.

The system of equations (6.2) in the case that $X(0) = 1$ becomes

$$P_n'(t) = -\beta [nP_n(t) - (n-1)P_{n-1}(t)], \qquad n = 1, 2, \ldots,$$

under the initial conditions

$$P_1(0) = 1, \qquad P_n(0) = 0, \qquad n = 2, 3, \ldots.$$

Its solution is

$$P_n(t) = \mathrm{e}^{-\beta t}\left(1 - \mathrm{e}^{-\beta t}\right)^{n-1}, \qquad n \geq 1, \tag{6.10}$$

as may be verified directly. We recognize (6.10) as the geometric distribution in Chapter 1, (1.26) with $p = \mathrm{e}^{-\beta t}$.

The general solution analogous to (6.8) but for pure birth processes starting from $X(0) = 1$ is

$$P_n(t) = \lambda_1 \cdots \lambda_{n-1}\left[B_{1,n}\mathrm{e}^{-\lambda_1 t} + \cdots + B_{n,n}\mathrm{e}^{-\lambda_n t}\right], \quad n > 1. \tag{6.11}$$

When $\lambda_n = \beta n$, we will show that (6.11) reduces to the solution given in (6.10) for a Yule process with parameter β. Then,

$$\begin{aligned}B_{1,n} &= \frac{1}{(\lambda_2 - \lambda_1)(\lambda_3 - \lambda_1)\cdots(\lambda_n - \lambda_1)} \\ &= \frac{1}{\beta^{n-1}(1)(2)\cdots(n-1)} \\ &= \frac{1}{\beta^{n-1}(n-1)!},\end{aligned}$$

$$
\begin{aligned}
B_{2,n} &= \frac{1}{(\lambda_1-\lambda_2)(\lambda_3-\lambda_2)\cdots(\lambda_n-\lambda_2)} \\
&= \frac{1}{\beta^{n-1}(-1)(1)(2)\cdots(n-2)} \\
&= \frac{-1}{\beta^{n-1}(n-2)!},
\end{aligned}
$$

and

$$
\begin{aligned}
\beta_{k,n} &= \frac{1}{(\lambda_1-\lambda_k)\cdots(\lambda_{k-1}-\lambda_k)(\lambda_{k+1}-\lambda_k)\cdots(\lambda_n-\lambda_k)} \\
&= \frac{(-1)^{k-1}}{\beta^{n-1}(k-1)!\,(n-k)!}.
\end{aligned}
$$

Thus, according to (6.11),

$$
\begin{aligned}
P_n(t) &= \beta^{n-1}(n-1)!\left(B_{1,n}e^{-\beta t}+\cdots+B_{n,n}e^{-n\beta t}\right) \\
&= \sum_{k=1}^{n}\frac{(n-1)!}{(k-1)!\,(n-k)!}(-1)^{k-1}e^{-k\beta t} \\
&= e^{-\beta t}\sum_{j=0}^{n-1}\frac{(n-1)!}{j!\,(n-1-j)!}\left(-e^{-\beta t}\right)^j \\
&= e^{-\beta t}\left(1-e^{-\beta t}\right)^{n-1} \qquad \text{[see Chapter 1, (1.67)]},
\end{aligned}
$$

which establishes (6.10).

Exercises

6.1.1 A pure birth process starting from $X(0)=0$ has birth parameters $\lambda_0=1$, $\lambda_1=3$, $\lambda_2=2$, and $\lambda_3=5$. Determine $P_n(t)$ for $n=0,1,2,3$.

6.1.2 A pure birth process starting from $X(0)=0$ has birth parameters $\lambda_0=1$, $\lambda_1=3$, $\lambda_2=2$, and $\lambda_3=5$. Let W_3 be the random time that it takes the process to reach state 3.

(a) Write W_3 as a sum of sojourn times and thereby deduce that the mean time is $E[W_3]=\frac{11}{6}$.

(b) Determine the mean of $W_1+W_2+W_3$.

(c) What is the variance of W_3?

6.1.3 A population of organisms evolves as follows. Each organism exists, independent of the other organisms, for an exponentially distributed length of time with parameter θ, and then splits into two new organisms, each of which exists, independent of the other organisms, for an exponentially distributed length of time with parameter θ, and then splits into two new organisms, and so on. Let $X(t)$ denote the number of organisms existing at time t. Show that $X(t)$ is a Yule process.

6.1.4 Consider an experiment in which a certain event will occur with probability αh and will not occur with probability $1-\alpha h$, where α is a fixed positive parameter and h is a small $(h < 1/\alpha)$ positive variable. Suppose that n independent trials of the experiment are carried out, and the total number of times that the event occurs is noted. Show that

(a) The probability that the event never occurs during the n trials is $1-n\alpha h+o(h)$;

(b) The probability that the event occurs exactly once is $n\alpha h+o(h)$;

(c) The probability that the event occurs twice or more is $o(h)$.

Hint: Use the binomial expansion

$$(1-\alpha h)^n = 1 - n\alpha h + \frac{n(n-1)}{2}(\alpha h)^2 - \cdots.$$

6.1.5 Using equation (6.10), calculate the mean and variance for the Yule process where $X(0)=1$.

6.1.6 Operations 1, 2, and 3 are to be performed in succession on a major piece of equipment. Operation k takes a random duration S_k that is exponentially distributed with parameter λ_k for $k=1,2,3$, and all operation times are independent. Let $X(t)$ denote the operation being performed at time t, with time $t=0$ marking the start of the first operation. Suppose that $\lambda_1=5, \lambda_2=3$, and $\lambda_3=13$. Determine

(a) $P_1(t)=\Pr\{X(t)=1\}$.

(b) $P_2(t)=\Pr\{X(t)=2\}$.

(c) $P_3(t)=\Pr\{X(t)=3\}$.

Problems

6.1.1 Let $X(t)$ be a Yule process that is observed at a random time U, where U is uniformly distributed over $[0,1)$. Show that $\Pr\{X(U)=k\}=p^k/(\beta k)$ for $k=1,2,\ldots$, with $p=1-e^{-\beta}$.

Hint: Integrate (6.10) over t between 0 and 1.

6.1.2 A Yule process with immigration has birth parameters $\lambda_k=\alpha+k\beta$ for $k=0,1,2,\ldots$. Here, α represents the rate of immigration into the population, and β represents the individual birth rate. Supposing that $X(0)=0$, determine $P_n(t)$ for $n=0,1,2,\ldots$.

6.1.3 Consider a population comprising a fixed number N of individuals. Suppose that at time $t=0$, there is exactly one *infected* individual and $N-1$ *susceptible* individuals in the population. Once infected, an individual remains in that state forever. In any short time interval of length h, *any given infected person* will transmit the disease to *any given susceptible person* with probability $\alpha h+o(h)$. (The parameter α is the *individual infection rate*.) Let $X(t)$ denote the number of infected individuals in the population at time $t\geq 0$. Then, $X(t)$ is a pure birth process on the states $0,1,\ldots,N$. Specify the birth parameters.

6.1.4 A new product (a "Home Helicopter" to solve the commuting problem) is being introduced. The sales are expected to be determined by both media (newspaper and television) advertising and word-of-mouth advertising, wherein satisfied customers tell others about the product. Assume that media advertising creates new customers according to a Poisson process of rate $\alpha = 1$ customer per month. For the word-of-mouth advertising, assume that each purchaser of a Home Helicopter will generate sales to new customers at a rate of $\theta = 2$ customers per month. Let $X(t)$ be the total number of Home Helicopter customers up to time t.

(a) Model $X(t)$ as a pure birth process by specifying the birth parameters λ_k, for $k = 0, 1, \ldots$.

(b) What is the probability that exactly two Home Helicopters are sold during the first month?

6.1.5 Let W_k be the time to the kth birth in a pure birth process starting from $X(0) = 0$. Establish the equivalence

$$\Pr\{W_1 > t, W_2 > t+s\} = P_0(t)[P_0(s) + P_1(s)].$$

From this relation together with equation (6.7), determine the joint density for W_1 and W_2, and then the joint density of $S_0 = W_1$ and $S_1 = W_2 - W_1$.

6.1.6 A fatigue model for the growth of a crack in a discrete lattice proposes that the size of the crack evolves as a pure birth process with parameters

$$\lambda_k = (1+k)^{\rho} \qquad \text{for } k = 1, 2, \ldots.$$

The theory behind the model postulates that the growth rate of the crack is proportional to some power of the stress concentration at its ends and that this stress concentration is itself proportional to some power of $1+k$, where k is the crack length. Use the sojourn time description to deduce that the mean time for the crack to grow to infinite length is finite when $\rho > 1$ and that, therefore, the failure time of the system is a well-defined and finite-valued random variable.

6.1.7 Let λ_0, λ_1, and λ_2 be the parameters of the independent exponentially distributed random variables S_0, S_1, and S_2. Assume that no two of the parameters are equal.

(a) Verify that

$$\Pr\{S_0 > t\} = e^{-\lambda_0 t},$$

$$\Pr\{S_0 + S_1 > t\} = \frac{\lambda_1}{\lambda_1 - \lambda_0} e^{-\lambda_0 t} + \frac{\lambda_0}{\lambda_0 - \lambda_1} e^{-\lambda_1 t},$$

and evaluate in similar terms

$$\Pr\{S_0 + S_1 + S_2 > t\}.$$

(b) Verify equation (6.8) in the case that $n = 2$ by evaluating

$$P_2(t) = \Pr\{X(t) = 2\} = \Pr\{S_0 + S_1 + S_2 > t\} - \Pr\{S_0 + S_1 > t\}.$$

6.1.8 Let $N(t)$ be a pure birth process for which

$$\Pr\{\text{an event happens in}(t, t+h)|N(t) \text{ is odd}\} = \alpha h + o(h),$$
$$\Pr\{\text{an event happens in}(t, t+h)|N(t) \text{ is even}\} = \beta h + o(h),$$

where $o(h)/h \to 0$ as $h \downarrow 0$. Take $N(0) = 0$. Find the following probabilities:

$$P_0(t) = \Pr\{N(t) \text{ is even}\}; \qquad P_1(t) = \Pr\{N(t) \text{ is odd}\}.$$

Hint: Derive the differential equations

$$P_0'(t) = \alpha P_1(t) - \beta P_0(t) \quad \text{and} \quad P_1'(t) = -\alpha P_1(t) + \beta P_0(t)$$

and solve them by using $P_0(t) + P_1(t) = 1$.

6.1.9 Under the conditions of Problem 6.8, determine $E[N(t)]$.

6.1.10 Consider a pure birth process on the states $0, 1, \ldots, N$ for which $\lambda_k = (N-k)\lambda$ for $k = 0, 1, \ldots, N$. Suppose that $X(0) = 0$. Determine $P_n(t) = \Pr\{X(t) = n\}$ for $n = 0, 1$, and 2.

6.1.11 Beginning with $P_0(t) = \mathrm{e}^{-\lambda_0 t}$ and using equation (6.5), calculate $P_1(t), P_2(t)$, and $P_3(t)$ and verify that these probabilities conform with equation (6.7), assuming distinct birth parameters.

6.1.12 Verify that $P_2(t)$, as given by (6.8), satisfies (6.5) by following the calculations in the text that showed that $P_1(t)$ satisfies (6.5).

6.1.13 Using (6.5), derive $P_n(t)$ when all birth parameters are the same constant λ and show that

$$P_n(t) = \frac{(\lambda t)^n \mathrm{e}^{-\lambda t}}{n!}, \qquad n = 0, 1, \ldots.$$

Thus, the postulates of Section 6.1.1 serve to define the Poisson processes.

6.2 Pure Death Processes

Complementing the increasing pure birth process is the decreasing pure death process. It moves successively through states $N, N-1, \ldots, 2, 1$ and ultimately is absorbed in state 0 (extinction). The process is specified by the death parameters $\mu_k > 0$ for $k = 1, 2, \ldots, N$, where the sojourn time in state k is exponentially distributed with parameter μ_k, all sojourn times being independent. A typical sample path is depicted in Figure 6.1.

Alternatively, we have the infinitesimal description of a pure death process as a Markov process $X(t)$ whose state space is $0, 1, \ldots, N$ and for which

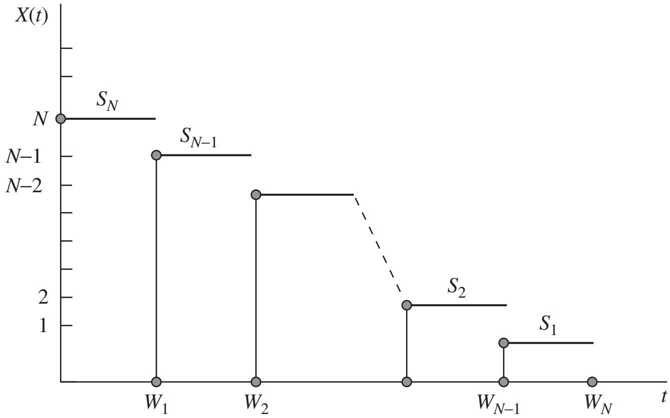

Figure 6.1 A typical sample path of a pure death process, showing the sojourn times $S_N, \ldots, S_1$ and the waiting times $W_1, W_2, \ldots, W_N$.

$$
\begin{aligned}
&\textbf{(i)}\ \Pr\{X(t+h) = k-1 | X(t) = k\} = \mu_k h + o(h), k = 1, \ldots, N; \\
&\textbf{(ii)}\ \Pr\{X(t+h) = k | X(t) = k\} = 1 - \mu_k h + o(h), k = 1, \ldots, N; \\
&\textbf{(iii)}\ \Pr\{X(t+h) > k | X(t) = k\} = 0, k = 0, 1, \ldots, N.
\end{aligned}
\tag{6.12}
$$

The parameter μ_k is the "death rate" operating or in effect while the process sojourns in state k. It is a common and useful convention to assign $\mu_0 = 0$.

When the death parameters $\mu_1, \mu_2, \ldots, \mu_N$ are distinct, i.e., $\mu_j \neq \mu_k$ if $j \neq k$, then we have the explicit transition probabilities

$$P_N(t) = \mathrm{e}^{-\mu_N t};$$

and for $n < N$,

$$
\begin{aligned}
P_n(t) &= \Pr\{X(t) = n | X(0) = N\} \\
&= \mu_{n+1}\mu_{n+2}\cdots\mu_N \left[A_{n,n}\mathrm{e}^{-\mu_n t} + \cdots + A_{N,n}\mathrm{e}^{-\mu_N t}\right],
\end{aligned}
\tag{6.13}
$$

where

$$A_{k,n} = \frac{1}{(\mu_N - \mu_k)\cdots(\mu_{k+1} - \mu_k)(\mu_{k-1} - \mu_k)\cdots(\mu_n - \mu_k)}.$$

6.2.1 The Linear Death Process

As an example, consider a pure death process in which the death rates are proportional to population size. This process, which we will call the *linear death process*, complements the Yule, or linear birth, process. The parameters are $\mu_k = k\alpha$, where α is the individual death rate in the population. Then,

$$
\begin{aligned}
A_{n,n} &= \frac{1}{(\mu_N - \mu_n)(\mu_{N-1} - \mu_n)\cdots(\mu_{n+1} - \mu_n)} \\
&= \frac{1}{\alpha^{N-n-1}(N-n)(N-n-1)\cdots(2)(1)}, \\
A_{n+1,n} &= \frac{1}{(\mu_N - \mu_{n+1})\cdots(\mu_{n+2} - \mu_{n+1})(\mu_n - \mu_{n+1})} \\
&= \frac{1}{\alpha^{N-n-1}(N-n-1)\cdots(1)(-1)}, \\
A_{k,n} &= \frac{1}{(\mu_N - \mu_k)\cdots(\mu_{k+1} - \mu_k)(\mu_{k-1} - \mu_k)\cdots(\mu_n - \mu_k)} \\
&= \frac{1}{\alpha^{N-n-1}(N-k)\cdots(1)(-1)(-2)\cdots(n-k)} \\
&= \frac{1}{\alpha^{N-n-1}(-1)^{k-n}(N-k)!\,(k-n)!}.
\end{aligned}
$$

Then,

$$
\begin{aligned}
P_n(t) &= \mu_{n+1}\mu_{n+2}\cdots\mu_N \sum_{k=n}^{N} A_{k,n}\mathrm{e}^{-\mu_k t} \\
&= \alpha^{N-n-1}\frac{N!}{n!}\sum_{k=n}^{N}\frac{\mathrm{e}^{-k\alpha t}}{\alpha^{N-n-1}(-1)^{k-n}(N-k)!\,(k-n)!} \\
&= \frac{N!}{n!}\mathrm{e}^{-n\alpha t}\sum_{j=0}^{N-n}\frac{(-1)^j\mathrm{e}^{-j\alpha t}}{(N-n-j)!\,j!} \\
&= \frac{N!}{n!\,(N-n)!}\mathrm{e}^{-n\alpha t}\left(1-\mathrm{e}^{-\alpha t}\right)^{N-n}, \quad n = 0, \ldots, N.
\end{aligned}
\tag{6.14}
$$

Let T be the time of population extinction. Formally, $T = \min\{t \geq 0; X(t) = 0\}$. Then, $T \leq t$ if and only if $X(t) = 0$, which leads to the cumulative distribution function of T via

$$
\begin{aligned}
F_T(t) &= \Pr\{T \leq t\} = \Pr\{X(t) = 0\} \\
&= P_0(t) = \left(1 - \mathrm{e}^{-\alpha t}\right)^N, \quad t \geq 0.
\end{aligned}
\tag{6.15}
$$

The linear death process can be viewed in yet another way, a way that again confirms the intimate connection between the exponential distribution and a continuous time parameter Markov chain. Consider a population consisting of N individuals, each of whose lifetimes is an independent exponentially distributed random variable with

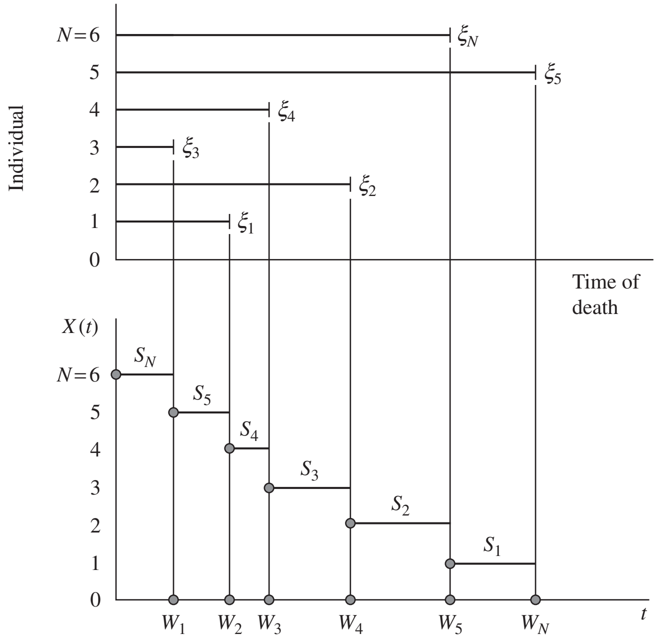

Figure 6.2 The linear death process. As depicted here, the third individual is the first to die, the first individual is the second to die, etc.

parameter α. Let $X(t)$ be the number of survivors in this population at time t. Then, $X(t)$ is the linear pure death process whose parameters are $\mu_k = k\alpha$ for $k = 0, 1, \ldots, N$. To help understand this connection, let $\xi_1, \xi_2, \ldots, \xi_N$ denote the times of death of the individuals labeled $1, 2, \ldots, N$, respectively. Figure 6.2 shows the relation between the individual lifetimes $\xi_1, \xi_2, \ldots, \xi_N$ and the death process $X(t)$.

The sojourn time in state N, denoted by S_N, equals the time of the earliest death, or $S_N = \min\{\xi_1, \ldots, \xi_N\}$. Since the lifetimes are independent and have the same exponential distribution,

$$\begin{aligned}\Pr\{S_N > t\} &= \Pr\{\min\{\xi_1, \ldots, \xi_N\} > t\}\\ &= \Pr\{\xi_1 > t, \ldots, \xi_N > t\}\\ &= [\Pr\{\xi_1 > t\}]^N\\ &= \mathrm{e}^{-N\alpha t}.\end{aligned}$$

That is, S_N has an exponential distribution with parameter $N\alpha$. Similar reasoning applies when there are k members alive in the population. The memoryless property of the exponential distribution implies that the remaining lifetime of each of these k individuals is exponentially distributed with parameter α. Then, the sojourn time S_k is the minimum of these k remaining lifetimes and hence is exponentially distributed with parameter $k\alpha$. To give one more approach in terms of transition rates, each individual in the population has a constant death rate of α in the sense that

$$\begin{aligned}\Pr\{t < \xi_1 < t+h \mid t < \xi_1\} &= \frac{\Pr\{t < \xi_1 < t+h\}}{\Pr\{t < \xi_1\}} \\ &= \frac{e^{-\alpha t} - e^{-\alpha(t+h)}}{e^{-\alpha t}} \\ &= 1 - e^{-\alpha h} \\ &= \alpha h + o(h) \quad \text{as } h \downarrow 0.\end{aligned}$$

If each of k individuals alive in the population at time t has a constant death rate of α, then the total population death rate should be $k\alpha$, directly proportional to the population size. This shortcut approach to specifying appropriate death parameters is a powerful and often-used tool of stochastic modeling. The next example furnishes another illustration of its use.

6.2.2 Cable Failure Under Static Fatigue

A cable composed of parallel fibers under tension is being designed to support a high-altitude weather balloon. With a design load of 1000 kg and a design lifetime of 100 years, how many fibers should be used in the cable?

The low-weight, high-strength fibers to be used are subject to *static fatigue*, or eventual failure when subjected to a constant load. The higher the constant load, the shorter the life, and experiments have established a linear plot on log–log axes between average failure time and load that is shown in Figure 6.3.

The relation between mean life μ_T and load l that is illustrated in Figure 6.3 takes the analytic form

$$\log_{10}\mu_T = 2 - 40\log_{10} l.$$

Were the cable to be designed on the basis of average life, to achieve the 100 year design target each fiber should carry 1 kg. Since the total load is 1000 kg, $N = \frac{1000}{1} = 1000$ fibers should be used in the cable.

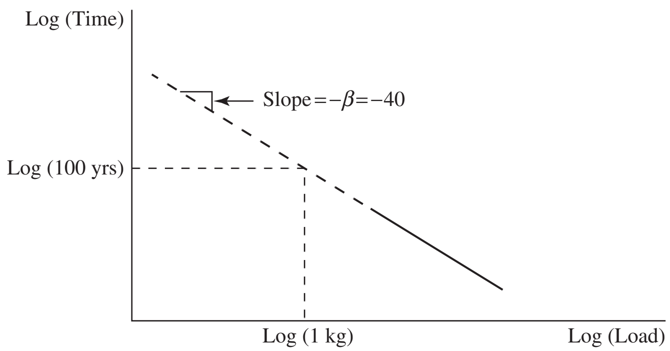

Figure 6.3 A linear relation between log mean failure time and log load.

One might suppose that this large number ($N = 1000$) of fibers would justify designing the cable based on average fiber properties. We shall see that such reasoning is dangerously wrong.

Let us suppose, however, as is the case with many modern high-performance structural materials, that there is a large amount of random scatter of individual fiber lifetimes about the mean. How does this randomness affect the design problem?

Some assumption must be made concerning the probability distribution governing individual fiber lifetimes. In practice, it is extremely difficult to gather sufficient data to determine this distribution with any degree of certainty. Most data do show, however, a significant degree of skewness, or asymmetry. Because it qualitatively matches observed data and because it leads to a pure death process model that is accessible to exhaustive analysis, we will assume that the probability distribution for the failure time T of a single fiber subjected to the time-varying tensile load $l(t)$ is given by

$$\Pr\{T \le t\} = 1 - \exp\left\{ -\int_0^t K[l(s)]\mathrm{d}s \right\}, \quad t \ge 0.$$

This distribution corresponds to a *failure rate*, or *hazard rate*, of $r(t) = K[l(t)]$ wherein a single fiber, having not failed prior to time t and carrying the load $l(t)$, will fail during the interval $(t, t + \Delta t]$ with probability

$$\Pr\{t < T \le t + \Delta t | T > t\} = K[l(t)]\Delta t + o(\Delta t).$$

The function $K[l]$, called the *breakdown rule*, expresses how changes in load affect the failure probability. We are concerned with the *power law breakdown rule* in which $K[l] = l^\beta / A$ for some positive constants A and β. Assuming power law breakdown, under a constant load $l(t) = l$, the single fiber failure time is exponentially distributed with mean $\mu_T = E[T|l] = 1/K[l] = Al^{-\beta}$. A plot of mean failure time versus load is linear on log–log axes, matching the observed properties of our fiber type. For the design problem, we have $\beta = 40$ and $A = 100$.

Now, place N of these fibers in parallel and subject the resulting bundle or cable to a total load, constant in time, of NL, where L is the nominal load per fiber. What is the probability distribution of the time at which the cable fails? Since the fibers are in parallel, this system failure time equals the failure time of the last fiber.

Under the stated assumptions governing single-fiber behavior, $X(t)$, the number of unfailed fibers in the cable at time t, evolves as a pure death process with parameters $\mu_k = kK[NL/k]$ for $k = 1, 2, \ldots, N$. Given $X(t) = k$ surviving fibers at time t and assuming that the total bundle load NL is shared equally among them, then each carries load NL/k and has a corresponding failure rate of $K[NL/k]$. As there are k such survivors in the bundle, the bundle, or system, failure rate is $\mu_k = kK[NL/k]$ as claimed.

It was mentioned earlier that the system failure time was W_N, the waiting time to the Nth fiber failure. Then, $\Pr\{W_N \le t\} = \Pr\{X(t) = 0\} = P_0(t)$, where $P_n(t)$ is given explicitly by (2.13) in terms of $\mu_1, \ldots, \mu_N$. Alternatively, we may bring to bear the sojourn time description of the pure death process and, following Figure 6.1, write

$$W_N = S_N + S_{N-1} + \cdots + S_1,$$

where $S_N, S_{N-1}, \ldots, S_1$ are independent exponentially distributed random variables and S_k has parameter $\mu_k = kK[NL/k] = k(NL/k)^\beta/A$. The mean system failure time is readily computed to be

$$\begin{aligned} E[W_N] &= E[S_N] + \cdots + E[S_1] \\ &= AL^{-\beta} \sum_{k=1}^{N} \frac{1}{k}\left(\frac{k}{N}\right)^\beta \\ &= AL^{-\beta} \sum_{k=1}^{N} \left(\frac{k}{N}\right)^{\beta-1}\left(\frac{1}{N}\right). \end{aligned} \tag{6.16}$$

The sum in the expression for $E[W_N]$ seems formidable at first glance, but a very close approximation is readily available when N is large. Figure 6.4 compares the sum to an integral.

From Figure 6.4, we see that

$$\sum_{k=1}^{N}\left(\frac{k}{N}\right)^{\beta-1}\left(\frac{1}{N}\right) \approx \int_0^1 x^{\beta-1}\mathrm{d}x = \frac{1}{\beta}.$$

Indeed, we readily obtain

$$\begin{aligned} \frac{1}{\beta} = \int_0^1 x^{\beta-1}\mathrm{d}x \le \sum_{k=1}^{N}\left(\frac{k}{N}\right)^{\beta-1}\left(\frac{1}{N}\right) &\le \int_{1/N}^{1+1/N} x^{\beta-1}\mathrm{d}x \\ &= \left(\frac{1}{\beta}\right)\left[\left(1+\frac{1}{N}\right)^\beta - \left(\frac{1}{N}\right)^\beta\right]. \end{aligned}$$

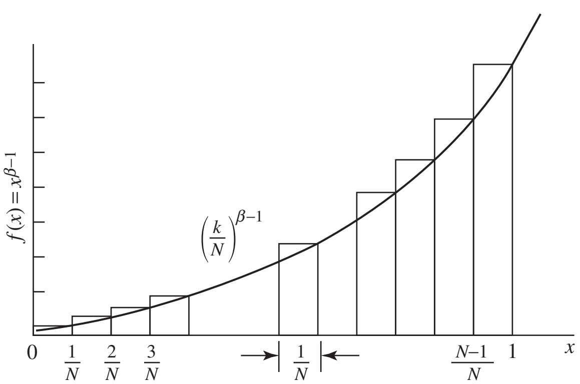

Figure 6.4 The sum $\Sigma_{k=1}^{N}(k/N)^{\beta-1}(1/N)$ is a Riemann approximation to $\int_0^1 x^{\beta-1}\mathrm{d}x = 1/\beta$.

When $N = 1000$ and $\beta = 40$, the numerical bounds are

$$\left(\frac{1}{40}\right) \leq \sum_{k=1}^{N}\left(\frac{k}{N}\right)^{\beta-1}\left(\frac{1}{N}\right) \leq \left(\frac{1}{40}\right)(1.0408),$$

which shows that the integral determines the sum to within about 4%.

Substituting $1/\beta$ for the sum in (6.16) gives the average cable life

$$E[W_N] \approx \frac{A}{\beta L^\beta}$$

to be compared with the average fiber life of

$$\mu_T = \frac{A}{L^\beta}.$$

That is, a cable lasts only about $1/\beta$ as long as an average fiber under an equivalent load. With $A = 100$, $\beta = 40$, and $N = 1000$, the designed cable would last, on the average, $100\big/\big[40(1)^{40}\big] = 2.5$ years, far short of the desired life of 100 years. The cure is to increase the number of fibers in the cable, thereby decreasing the per fiber load. Increasing the number of fibers from N to N' decreases the nominal load per fiber from L to $L' = NL/N'$. To achieve parity in fiber-cable lifetimes, we equate

$$\frac{A}{L^\beta} = \frac{A}{\beta(NL/N')^\beta},$$

or

$$N' = N\beta^{1/\beta}.$$

For the given data, this calls for $N' = 1000(40)^{1/40} = 1097$ fibers. That is, the design lifetime can be restored by increasing the number of fibers in the cable by about 10%.

Exercises

6.2.1 A pure death process starting from $X(0) = 3$ has death parameters $\mu_0 = 0$, $\mu_1 = 3$, $\mu_2 = 2$, and $\mu_3 = 5$. Determine $P_n(t)$ for $n = 0, 1, 2, 3$.

6.2.2 A pure death process starting from $X(0) = 3$ has death parameters $\mu_0 = 0$, $\mu_1 = 3$, $\mu_2 = 2$, and $\mu_3 = 5$. Let W_3 be the random time that it takes the process to reach state 0.

- **(a)** Write W_3 as a sum of sojourn times and thereby deduce that the mean time is $E[W_3] = \frac{31}{30}$.
- **(b)** Determine the mean of $W_1 + W_2 + W_3$.
- **(c)** What is the variance of W_3?

6.2.3 Give the transition probabilities for the pure death process described by $X(0) = 3, \mu_3 = 1, \mu_2 = 2$, and $\mu_1 = 3$.

6.2.4 Consider the linear death process (Section 6.2.1) in which $X(0) = N = 5$ and $\alpha = 2$. Determine $\Pr\{X(t) = 2\}$.

Hint: Use equation (6.14).

Problems

6.2.1 Let $X(t)$ be a pure death process starting from $X(0) = N$. Assume that the death parameters are $\mu_1, \mu_2, \ldots, \mu_N$. Let T be an independent exponentially distributed random variable with parameter θ. Show that

$$\Pr\{X(T) = 0\} = \prod_{i=1}^{N} \frac{\mu_i}{\mu_i + \theta}.$$

6.2.2 Let $X(t)$ be a pure death process with constant death rates $\mu_k = \theta$ for $k = 1, 2, \ldots, N$. If $X(0) = N$, determine $P_n(t) = \Pr\{X(t) = n\}$ for $n = 0, 1, \ldots, N$.

6.2.3 A pure death process $X(t)$ with parameters $\mu_1, \mu_2, \ldots$ starts at $X(0) = N$ and evolves until it reaches the absorbing state 0. Determine the mean area under the $X(t)$ trajectory.

Hint: This is $E[W_1 + W_2 + \cdots + W_N]$.

6.2.4 A chemical solution contains N molecules of type A and M molecules of type B. An irreversible reaction occurs between type A and B molecules in which they bond to form a new compound AB. Suppose that in any small time interval of length h, any particular unbonded A molecule will react with any particular unbonded B molecule with probability $\theta h + o(h)$, where θ is a reaction rate. Let $X(t)$ denote the number of unbonded A molecules at time t.

(a) Model $X(t)$ as a pure death process by specifying the parameters.

(b) Assume that $N < M$ so that eventually all of the A molecules become bonded. Determine the mean time until this happens.

6.2.5 Consider a cable composed of fibers following the breakdown rule $K[l] = \sinh(l) = \frac{1}{2}\left(e^{l} - e^{-l}\right)$ for $l \geq 0$. Show that the mean cable life is given by

$$E[W_N] = \sum_{k=1}^{N} \{k \sinh(NL/k)\}^{-1} = \sum_{k=1}^{N} \left\{\frac{k}{N} \sinh\left(\frac{L}{k/N}\right)\right\}^{-1} \left(\frac{1}{N}\right)$$
$$\approx \int_0^1 \{x \sinh(L/x)\}^{-1} dx.$$

6.2.6 Let T be the time to extinction in the linear death process with parameters $X(0) = N$ and α (see Section 6.2.1).

(a) Using the sojourn time viewpoint, show that

$$E[T] = \frac{1}{\alpha}\left[\frac{1}{N} + \frac{1}{N-1} + \cdots + \frac{1}{1}\right].$$

(b) Verify the result of (a) by using equation (6.15) in

$$E[T] = \int_0^\infty \Pr\{T > t\}\mathrm{d}t = \int_0^\infty [1 - F_T(t)]\mathrm{d}t.$$

Hint: Let $y = 1 - \mathrm{e}^{-\alpha t}$.

6.3 Birth and Death Processes

An obvious generalization of the pure birth and pure death processes discussed in Sections 6.1 and 6.2 is to permit $X(t)$ both to increase and to decrease. Thus, if at time t the process is in state n, it may, after a random sojourn time, move to either of the neighboring states $n+1$ or $n-1$. The resulting *birth and death process* can then be regarded as the continuous-time analog of a random walk (Chapter 3, Section 3.5.3).

Birth and death processes form a powerful tool in the kit of the stochastic modeler. The richness of the birth and death parameters facilitates modeling a variety of phenomena. At the same time, standard methods of analysis are available for determining numerous important quantities such as stationary distributions and mean first passage times. This section and later sections contain several examples of birth and death processes and illustrate how they are used to draw conclusions about phenomena in a variety of disciplines.

6.3.1 Postulates

As in the case of the pure birth processes, we assume that $X(t)$ is a Markov process on the states $0, 1, 2, \ldots$ and that its transition probabilities $P_{ij}(t)$ are stationary; that is

$$P_{ij}(t) = \Pr\{X(t+s) = j | X(s) = i\} \quad \text{for all} \quad s \geq 0.$$

In addition, we assume that the $P_{ij}(t)$ satisfy

1. $P_{i,i+1}(h) = \lambda_i h + o(h)$ as $h \downarrow 0, i \geq 0$;
2. $P_{i,i-1}(h) = \mu_i h + o(h)$ as $h \downarrow 0, i \geq 1$;
3. $P_{i,i}(h) = 1 - (\lambda_i + \mu_i)h + o(h)$ as $h \downarrow 0, i \geq 0$;
4. $P_{ij}(0) = \delta_{ij}$;
5. $\mu_0 = 0, \lambda_0 > 0, \mu_i, \lambda_i > 0, i = 1, 2, \ldots$.

The $o(h)$ in each case may depend on i. The matrix

$$\mathbf{A} = \begin{Vmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & \cdots \\ \mu_1 & -(\lambda_1+\mu_1) & \lambda_1 & 0 & \cdots \\ 0 & \mu_2 & -(\lambda_2+\mu_2) & \lambda_2 & \cdots \\ 0 & 0 & \mu_3 & -(\lambda_3+\mu_3) & \cdots \\ \vdots & \vdots & \vdots & \vdots & \end{Vmatrix} \tag{6.17}$$

is called the *infinitesimal generator* of the process. The parameters λ_i and μ_i are called, respectively, the infinitesimal birth and death rates. In Postulates (1) and (2), we are assuming that if the process starts in state i, then in a small interval of time the probabilities of the population increasing or decreasing by 1 are essentially proportional to the length of the interval.

Since the $P_{ij}(t)$ are probabilities, we have $P_{ij}(t) \geq 0$ and

$$\sum_{j=0}^{\infty} P_{ij}(t) \leq 1. \tag{6.18}$$

Using the Markov property of the process, we may also derive the so-called *Chapman–Kolmogorov equation*

$$P_{ij}(t+s) = \sum_{k=0}^{\infty} P_{ik}(t)P_{kj}(s). \tag{6.19}$$

This equation states that in order to move from state i to state j in time $t+s, X(t)$ moves to some state k in time t and then from k to j in the remaining time s. This is the continuous-time analog of formula (3.11) in Chapter 3.

Thus far, we have mentioned only the transition probabilities $P_{ij}(t)$. In order to obtain the probability that $X(t) = n$, we must specify where the process starts or more generally the probability distribution for the initial state. We then have

$$\Pr\{X(t) = n\} = \sum_{i=0}^{x} q_i P_{in}(t),$$

where

$$q_i = \Pr\{X(0) = i\}.$$

6.3.2 Sojourn Times

With the aid of the preceding assumptions, we may calculate the distribution of the random variable S_i, which is the sojourn time of $X(t)$ in state i; that is, given that the process is in state i, what is the distribution of the time S_i until it first leaves state i? If we let

$$\Pr\{S_i \geq t\} = G_i(t),$$

it follows easily by the Markov property that as $h \downarrow 0$,

$$\begin{aligned} G_i(t+h) &= G_i(t)G_i(h) = G_i(t)[P_{ii}(h) + o(h)] \\ &= G_i(t)[1 - (\lambda_i + \mu_i)h] + o(h), \end{aligned}$$

or

$$\frac{G_i(t+h) - G_i(t)}{h} = -(\lambda_i + \mu_i)G_i(t) + o(1)$$

so that

$$G_i'(t) = -(\lambda_i + \mu_i)G_i(t). \tag{6.20}$$

If we use the conditions $G_i(0) = 1$, the solution of this equation is

$$G_i(t) = \exp[-(\lambda_i + \mu_i)t];$$

that is, S_i follows an exponential distribution with mean $(\lambda_i + \mu_i)^{-1}$. The proof presented here is not quite complete, since we have used the intuitive relationship

$$G_i(h) = P_{ii}(h) + o(h)$$

without a formal proof.

According to Postulates (1) and (2), during a time duration of length h, a transition occurs from state i to $i+1$ with probability $\lambda_i h + o(h)$ and from state i to $i-1$ with probability $\mu_i h + o(h)$. It follows intuitively that, given that a transition occurs at time t, the probability that this transition is to state $i+1$ is $\lambda_i/(\lambda_i + \mu_i)$ and to state $i-1$ is $\mu_i/(\lambda_i + \mu_i)$. The rigorous demonstration of this result is beyond the scope of this book.

It leads to an important characterization of a birth and death process, however, wherein the description of the motion of $X(t)$ is as follows: The process sojourns in a given state i for a random length of time whose distribution function is an exponential distribution with parameter $(\lambda_i + \mu_i)$. When leaving state i the process enters either state $i+1$ or state $i-1$ with probabilities $\lambda_i/(\lambda_i + \mu_i)$ and $\mu_i/(\lambda_i + \mu_i)$, respectively. The motion is analogous to that of a random walk except that transitions occur at random times rather than at fixed time periods.

The traditional procedure for constructing birth and death processes is to prescribe the birth and death parameters $\{\lambda_i, \mu_i\}_{i=0}^{\infty}$ and build the path structure by utilizing the preceding description concerning the waiting times and the conditional transition probabilities of the various states. We determine realizations of the process as follows.

Suppose $X(0) = i$; the particle spends a random length of time, exponentially distributed with parameter $(\lambda_i + \mu_i)$, in state i and subsequently moves with probability $\lambda_i/(\lambda_i + \mu_i)$ to state $i + 1$ and with probability $\mu_i/(\lambda_i + \mu_i)$ to state $i - 1$. Next, the particle sojourns a random length of time in the new state and then moves to one of its neighboring states and so on. More specifically, we observe a value t_1 from the exponential distribution with parameter $(\lambda_i + \mu_i)$ that fixes the initial sojourn time in state i. Then, we toss a coin with probability of heads $p_i = \lambda_i/(\lambda_i + \mu_i)$. If heads (tails) appear, we move the particle to state $i + 1 (i - 1)$. In state $i + 1$, we observe a value t_2 from the exponential distribution with parameter $(\lambda_{i+1} + \mu_{i+1})$ that fixes the sojourn time in the seconds state visited. If the particle at the first transition enters state $i - 1$, the subsequent sojourn time t_2' is an observation from the exponential distribution with parameter $(\lambda_{i-1} + \mu_{i-1})$. After the second wait is completed, a Bernoulli trial is performed that chooses the next state to be visited, and the process continues in the same way.

A typical outcome of these sampling procedures determines a realization of the process. Its form might be, e.g.,

$$X(t) = \begin{cases} i, & \text{for } 0 < t < t_1, \\ i+1, & \text{for } t_1 < t < t_1 + t_2, \\ i, & \text{for } t_1 + t_2 < t < t_1 + t_2 + t_3, \\ \vdots & \qquad\qquad \vdots \end{cases}$$

Thus, by sampling from exponential and Bernoulli distributions appropriately, we construct typical sample paths of the process. Now, it is possible to assign to this set of paths (realizations of the process) a probability measure in a consistent way so that $P_{ij}(t)$ is determined satisfying (6.18) and (6.19). This result is rather deep, and its rigorous discussion is beyond the level of this book. The process obtained in this manner is called the minimal process associated with the infinitesimal matrix $\mathbf{A}$ defined in (6.17).

The preceding construction of the minimal process is fundamental, since the infinitesimal parameters need not determine a unique stochastic process obeying (6.18), (6.19), and Postulates 1 through 5 of Section 6.3.1. In fact, there could be several Markov processes that possess the same infinitesimal generator. Fortunately, such complications do not arise in the modeling of common phenomena. In the special case of birth and death processes for which $\lambda_0 > 0$, a sufficient condition that there exists a unique Markov process with transition probability function $P_{ij}(t)$ for which the infinitesimal relations (6.18) and (6.19) hold is that

$$\sum_{n=0}^{\infty} \frac{1}{\lambda_n \theta_n} \sum_{k=0}^{n} \theta_k = \infty, \tag{6.21}$$

where

$$\theta_0 = 1, \quad \theta_n = \frac{\lambda_0 \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \cdots \mu_n}, \quad n = 1, 2, \ldots.$$

In most practical examples of birth and death processes, the condition (6.21) is met, and the birth and death process associated with the prescribed parameters is uniquely determined.

6.3.3 Differential Equations of Birth and Death Processes

As in the case of the pure birth and pure death processes, the transition probabilities $P_{ij}(t)$ satisfy a system of differential equations known as the backward Kolmogorov differential equations. These are given by

$$\begin{aligned} P'_{0j}(t) &= -\lambda_0 P_{0j}(t) + \lambda_0 P_{1j}(t), \\ P'_{ij}(t) &= \mu_i P_{i-1,j}(t) - (\lambda_i + \mu_i) P_{ij}(t) + \lambda_i P_{i+1,j}(t), \quad i \geq 1, \end{aligned} \tag{6.22}$$

and the boundary condition $P_{ij}(0) = \delta_{ij}$.

To derive these, we have, from equation (6.19),

$$\begin{aligned} P_{ij}(t+h) &= \sum_{k=0}^{\infty} P_{ik}(h) P_{kj}(t) \\ &= P_{i,i-1}(h) P_{i-1,j}(t) + P_{i,i}(h) P_{ij}(t) + P_{i,i+1}(h) P_{i+1,j}(t) \\ &\quad + \sum\nolimits_k' P_{ik}(h) P_{kj}(t), \end{aligned} \tag{6.23}$$

where the last summation is over all $k \neq i-1, i, i+1$. Using Postulates (1), (2), and (3) of Section 6.3.1, we obtain

$$\begin{aligned} \sum\nolimits_k' P_{ik}(h) P_{kj}(t) &\leq \sum\nolimits_k' P_{ik}(h) \\ &= 1 - [P_{i,i}(h) + P_{i,i-1}(h) + P_{i,i+1}(h)] \\ &= 1 - [1 - (\lambda_i + \mu_i)h + o(h) + \mu_i h + o(h) + \lambda_i h + o(h)] \\ &= o(h) \end{aligned}$$

so that

$$P_{ij}(t+h) = \mu_i h P_{i-1,j}(t) + [1 - (\lambda_i + \mu_i)h] P_{ij}(t) + \lambda_i h P_{i+1,j}(t) + o(h).$$

Transposing the term $P_{ij}(t)$ to the left-hand side and dividing the equation by h, we obtain, after letting $h \downarrow 0$,

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

The backward equations are deduced by decomposing the time interval $(0, t+h)$, where h is positive and small, into the two periods

$$(0, h), \quad (h, t+h)$$

and examining the transition in each period separately. In this sense, the backward equations result from a "first step analysis," the first step being over the short time interval of duration h.

A different result arises from a "last step analysis," which proceeds by splitting the time interval $(0, t+h)$ into the two periods

$$(0, t), \quad (t, t+h)$$

and adapting the preceding reasoning. From this viewpoint, under more stringent conditions, we can derive a further system of differential equations

$$\begin{aligned} P'_{i0}(t) &= -\lambda_0 P_{i,0}(t) + \mu_1 P_{i,1}(t), \\ P'_{ij}(t) &= \lambda_{j-1} P_{i,j-1}(t) - (\lambda_j + \mu_j) P_{ij}(t) + \mu_{j+1} P_{i,j+1}(t), \quad j \geq 1, \end{aligned} \tag{6.24}$$

with the same initial condition $P_{ij}(0) = \delta_{ij}$. These are known as the forward Kolmogorov differential equations. To derive these equations, we interchange t and h in equation (6.23), and under stronger assumptions in addition to Postulates (1), (2), and (3), it can be shown that the last term is again $o(h)$. The remainder of the argument is the same as before. The usefulness of the differential equations will become apparent in the examples that we study in this and the next section.

A sufficient condition that (6.24) hold is that $[P_{kj}(h)]/h = o(1)$ for $k \neq j, j-1, j+1$, where the $o(1)$ term apart from tending to zero is uniformly bounded with respect to k for fixed j as $h \to 0$. In this case, it can be proved that $\sum'_k P_{ik}(t) P_{kj}(h) = o(h)$.

Example *Linear Growth with Immigration* A birth and death process is called a linear growth process if $\lambda_n = \lambda n + a$ and $\mu_n = \mu n$ with $\lambda > 0$, $\mu > 0$, and $a > 0$. Such processes occur naturally in the study of biological reproduction and population growth. If the state n describes the current population size, then the average instantaneous rate of growth is $\lambda n + a$. Similarly, the probability of the state of the process decreasing by one after the elapse of a small duration of time h is $\mu n h + o(h)$. The factor λn represents the natural growth of the population owing to its current size, while the second factor a may be interpreted as the infinitesimal rate of increase of the population due to an external source such as immigration. The component μn, which gives the mean infinitesimal death rate of the present population, possesses the obvious interpretation.

If we substitute the above values of λ_n and μ_n in (6.24), we obtain

$$\begin{aligned} P'_{i0}(t) &= -a P_{i0}(t) + \mu P_{i1}(t), \\ P'_{ij}(t) &= [\lambda(j-1) + a] P_{i,j-1}(t) - [(\lambda+\mu) j + a] P_{ij}(t) \\ &\quad + \mu(j+1) P_{i,j+1}(t), \quad j \geq 1. \end{aligned}$$

Now, if we multiply the jth equation by j and sum, it follows that the expected value

$$E[X(t)] = M(t) = \sum_{j=1}^{\infty} j P_{ij}(t)$$

satisfies the differential equation

$$M'(t) = a + (\lambda - \mu)M(t),$$

with initial condition $M(0) = i$, if $X(0) = i$. The solution of this equation is

$$M(t) = at + i \quad \text{if } \lambda = \mu$$

and

$$M(t) = \frac{a}{\lambda - \mu}\left\{e^{(\lambda-\mu)t} - 1\right\} + ie^{(\lambda-\mu)t} \quad \text{if } \lambda \neq \mu. \tag{6.25}$$

The second moment, or variance, may be calculated in a similar way. It is interesting to note that $M(t) \to \infty$ as $t \to \infty$ if $\lambda \geq \mu$, while if $\lambda < \mu$, the mean population size for large t is approximately

$$\frac{a}{\mu - \lambda}.$$

These results suggest that in the second case, wherein $\lambda < \mu$, the population stabilizes in the long run in some form of statistical equilibrium. Indeed, it can be shown that a limiting probability distribution $\{\pi_j\}$ exists for which $\lim_{t\to\infty} P_{ij}(t) = \pi_j, j = 0, 1, \ldots$. Such limiting distributions for general birth and death processes are the subject of Section 6.4.

Example *The Two-State Markov Chain* Consider a Markov chain $\{X(t)\}$ with state $\{0, 1\}$ whose infinitesimal matrix is

$$\mathbf{A} = \begin{array}{c} \\ 0 \\ 1 \end{array} \!\!\begin{array}{c} \begin{array}{cc} \;\;0 & \quad 1 \end{array} \\ \left\|\begin{array}{cc} -\alpha & \alpha \\ \beta & -\beta \end{array}\right\| \end{array}. \tag{6.26}$$

The process alternates between states 0 and 1. The sojourn times in state 0 are independent and exponentially distributed with parameter α. Those in state 1 are independent and exponentially distributed with parameter β. This is a finite-state birth and death process for which $\lambda_0 = \alpha$, $\lambda_1 = 0$, $\mu_0 = 0$, and $\mu_1 = \beta$. The first Kolmogorov forward equation in (6.24) becomes

$$P'_{00}(t) = -\alpha P_{00}(t) + \beta P_{01}(t). \tag{6.27}$$

Now, $P_{01}(t) = 1 - P_{00}(t)$, which placed in (6.27) gives

$$P'_{00}(t) = \beta - (\alpha + \beta)P_{00}(t).$$

Let $Q_{00}(t) = \mathrm{e}^{(\alpha+\beta)t} P_{00}(t)$. Then,

$$\begin{aligned}\frac{\mathrm{d}Q_{00}(t)}{\mathrm{d}t} &= \mathrm{e}^{(\alpha+\beta)t} P'_{00}(t) + (\alpha+\beta)\mathrm{e}^{(\alpha+\beta)t} P_{00}(t) \\ &= \mathrm{e}^{(\alpha+\beta)t}\left[P'_{00}(t) + (\alpha+\beta)P_{00}(t)\right] \\ &= \beta \mathrm{e}^{(\alpha+\beta)t},\end{aligned}$$

which can be integrated immediately to yield

$$\begin{aligned}Q_{00}(t) &= \beta \int \mathrm{e}^{(\alpha+\beta)t}\mathrm{d}t + C \\ &= \left(\frac{\beta}{\alpha+\beta}\right)\mathrm{e}^{(\alpha+\beta)t} + C.\end{aligned}$$

The initial condition $Q_{00}(0) = 1$ determines the constant of integration to be $C = \alpha/(\alpha+\beta)$. Thus,

$$Q_{00}(t) = \mathrm{e}^{(\alpha+\beta)t} P_{00}(t) = \left(\frac{\beta}{\alpha+\beta}\right)\mathrm{e}^{(\alpha+\beta)t} + \left(\frac{\alpha}{\alpha+\beta}\right) \tag{6.28}$$

and

$$P_{00}(t) = \frac{\beta}{\alpha+\beta} + \frac{\alpha}{\alpha+\beta}\mathrm{e}^{-(\alpha+\beta)t}. \tag{6.29a}$$

Since $P_{01}(t) = 1 - P_{00}(t)$, we have

$$P_{01}(t) = \frac{\alpha}{\alpha+\beta} - \frac{\alpha}{\alpha+\beta}\mathrm{e}^{-(\alpha+\beta)t}, \tag{6.29b}$$

and by symmetry,

$$P_{11}(t) = \frac{\alpha}{\alpha+\beta} + \frac{\beta}{\alpha+\beta}\mathrm{e}^{-(\alpha+\beta)t}, \tag{6.29c}$$

$$P_{10}(t) = \frac{\beta}{\alpha+\beta} - \frac{\beta}{\alpha+\beta}\mathrm{e}^{-(\alpha+\beta)t}. \tag{6.29d}$$

These transition probabilities assume a more succinct form if we reparametrize according to $\pi = \alpha/(\alpha+\beta)$ and $\tau = \alpha+\beta$. Then,

$$P_{00}(t) = (1-\pi) + \pi \mathrm{e}^{-\tau t}, \tag{6.30a}$$

$$P_{01}(t) = \pi - \pi \mathrm{e}^{-\tau t}, \tag{6.30b}$$

$$P_{10}(t) = (1-\pi) - (1-\pi)\mathrm{e}^{-\tau t}, \tag{6.30c}$$

and

$$P_{11}(t) = \pi + (1-\pi)\mathrm{e}^{-\tau t}. \tag{6.30d}$$

Observe that

$$\lim_{t\to\infty} P_{01}(t) = \lim_{t\to\infty} P_{11}(t) = \pi$$

so that π is the long run probability of finding the process in state 1 independently of where the process began. The long run behavior of general birth and death processes is the subject of the next section.

Exercises

6.3.1 Particles are emitted by a radioactive substance according to a Poisson process of rate λ. Each particle exists for an exponentially distributed length of time, independent of the other particles, before disappearing. Let $X(t)$ denote the number of particles alive at time t. Argue that $X(t)$ is a birth and death process and determine the parameters.

6.3.2 Patients arrive at a hospital emergency room according to a Poisson process of rate λ. The patients are treated by a single doctor on a first come, first served basis. The doctor treats patients more quickly when the number of patients waiting is higher. An industrial engineering time study suggests that the mean patient treatment time when there are k patients in the system is of the form $m_k = \alpha - \beta k/(k+1)$, where α and β are constants with $\alpha > \beta > 0$. Let $N(t)$ be the number of patients in the system at time t (waiting and being treated). Argue that $N(t)$ might be modeled as a birth and death process with parameters $\lambda_k = \lambda$ for $k = 0, 1, \ldots$ and $\mu_k = k/m_k$ for $k = 0, 1, \ldots$. State explicitly any necessary assumptions.

6.3.3 Let $\{V(t)\}$ be the two-state Markov chain whose transition probabilities are given by (6.30a–d). Suppose that the initial distribution is $(1-\pi, \pi)$. That is, assume that $\Pr\{V(0) = 0\} = 1-\pi$ and $\Pr\{V(0) = 1\} = \pi$. In this case, show that $\Pr\{V(t) = 1\} = \pi$ for all times $t > 0$.

Problems

6.3.1 Let $\xi_n, n = 0, 1, \ldots,$ be a two-state Markov chain with transition probability matrix

$$\mathbf{P} = \begin{array}{c c} & \begin{array}{cc} 0 & 1 \end{array} \\ \begin{array}{c} 0 \\ 1 \end{array} & \left\| \begin{array}{cc} 0 & 1 \\ 1-\alpha & \alpha \end{array} \right\| \end{array}.$$

Let $\{N(t); t \geq 0\}$ be a Poisson process with parameter λ. Show that

$$X(t) = \xi_{N(t)}, \quad t \geq 0,$$

is a two-state birth and death process and determine the parameters λ_0 and μ_1 in terms of α and λ.

6.3.2 Collards were planted equally spaced in a single row in order to provide an experimental setup for observing the chaotic movements of the flea beetle (*Phyllotreta cruciferae*). A beetle at position k in the row remains on that plant for a random length of time having mean m_k (which varies with the "quality" of the plant) and then is equally likely to move right $(k+1)$ or left $(k-1)$. Model the position of the beetle at time t as a birth and death process having parameters $\lambda_k = \mu_k = 1/(2m_k)$ for $k = 1, 2, \ldots, N-1$, where the plants are numbered $0, 1, \ldots, N$. What assumptions might be plausible at the ends 0 and N?

6.3.3 Let $\{V(t)\}$ be the two-state Markov chain whose transition probabilities are given by (6.30a–d). Suppose that the initial distribution is $(1-\pi, \pi)$. That is, assume that $\Pr\{V(0) = 0\} = 1 - \pi$ and $\Pr\{V(0) = 1\} = \pi$. For $0 < s < t$, show that

$$E[V(s)V(t)] = \pi - \pi P_{10}(t-s),$$

whence

$$\mathrm{Cov}[V(s), V(t)] = \pi(1-\pi)\mathrm{e}^{-(\alpha+\beta)|t-s|}.$$

6.3.4 *A Stop-and-Go Traveler* The velocity $V(t)$ of a stop-and-go traveler is described by the two-state Markov chain whose transition probabilities are given by (6.30a–d). The distance traveled in time t is the integral of the velocity:

$$S(t) = \int_0^t V(u)\mathrm{d}u.$$

Assuming that the velocity at time $t = 0$ is $V(0) = 0$, determine the mean of $S(t)$. Take for granted the interchange of integral and expectation in

$$E[S(t)] = \int_0^t E[V(u)]\mathrm{d}u.$$

6.4 The Limiting Behavior of Birth and Death Processes

For a general birth and death process that has no absorbing states, it can be proved that the limits

$$\lim_{t\to\infty} P_{ij}(t) = \pi_j \geq 0 \tag{6.31}$$

exist and are independent of the initial state i. It may happen that $\pi_j = 0$ for all states j. When the limits π_j are strictly positive, however, and satisfy

$$\sum_{j=0}^{\infty} \pi_j = 1, \tag{6.32}$$

they form a probability distribution that is called, naturally enough, the *limiting distribution* of the process. The limiting distribution is also a *stationary distribution* in that

$$\pi_j = \sum_{i=0}^{\infty} \pi_i P_{ij}(t), \tag{6.33}$$

which tells us that if the process starts in state i with probability π_i, then at any time t it will be in state i with the same probability π_i. The proof of (6.33) follows from (6.19) and (6.31) if we let $t \to \infty$ and use the fact that $\Sigma_{i=0}^{\infty} \pi_i = 1$.

The general importance of birth and death processes as models derives in large part from the availability of standard formulas for determining if a limiting distribution exists and what its values are when it does. These formulas follow from the Kolmogorov forward equations (6.24) that were derived in Section 6.3.3:

$$\begin{aligned} P'_{i,0}(t) &= -\lambda_0 P_{i,0}(t) + \mu_1 P_{i,1}(t), \\ P'_{i,j}(t) &= \lambda_{j-1} P_{i,j-1}(t) - (\lambda_j + \mu_j) P_{ij}(t) + \mu_{j+1} P_{i,j+1}(t), \quad j \geq 1, \end{aligned} \tag{6.34}$$

with the initial condition $P_{iJ}(0) = \delta_{ij}$. Now pass to the limit as $t \to \infty$ in (6.34) and observe first that the limit of the right side of (6.34) exists according to (6.31). Therefore, the limit of the left side, the derivatives $P'_{ij}(t)$, exists as well. Since the probabilities are converging to a constant, the limit of these derivatives must be zero. In summary, passing to the limit in (6.34) produces

$$\begin{aligned} 0 &= -\lambda_0 \pi_0 + \mu_1 \pi_1, \\ 0 &= \lambda_{j-1} \pi_{j-1} - (\lambda_j + \mu_j) \pi_j + \mu_{j+1} \pi_{j+1}, \quad j \geq 1. \end{aligned} \tag{6.35}$$

The solution to (6.35) is obtained by induction. Letting

$$\theta_0 = 1 \quad \text{and} \quad \theta_j = \frac{\lambda_0 \lambda_1 \cdots \lambda_{j-1}}{\mu_1 \mu_2 \cdots \mu_J} \quad \text{for } j \geq 1, \tag{6.36}$$

we have $\pi_1 = \lambda_0 \pi_0 / \mu_1 = \theta_1 \pi_0$. Then, assuming that $\pi_k = \theta_k \pi_0$ for $k = 1, \ldots, j$, we obtain

$$\begin{aligned} \mu_{j+1} \pi_{j+1} &= (\lambda_j + \mu_j) \theta_j \pi_0 - \lambda_{j-1} \theta_{j-1} \pi_0 \\ &= \lambda_j \theta_j \pi_0 + (\mu_j \theta_j - \lambda_{j-1} \theta_{j-1}) \pi_0 \\ &= \lambda_j \theta_j \pi_0, \end{aligned}$$

and finally

$$\pi_{j+1} = \theta_{j+1}\pi_0.$$

In order that the sequence $\{\pi_j\}$ define a distribution, we must have $\Sigma_j \pi_j = 1$. If $\Sigma\theta_j < \infty$, then we may sum the following,

$$\begin{aligned}
\pi_0 &= \theta_0\pi_0\\
\pi_1 &= \theta_1\pi_0\\
\pi_2 &= \theta_2\pi_0\\
\vdots\ \ & \quad \vdots\\
\hline
1 &= (\Sigma\theta_k)\pi_0
\end{aligned}$$

to see that $\pi_0 = 1/\Sigma_{k=0}^{\infty}\theta_k$, and then

$$\pi_j = \theta_j\pi_0 = \frac{\theta_j}{\Sigma_{k=0}^{\infty}\theta_k} \quad \text{for } j = 0, 1, \ldots. \tag{6.37}$$

If $\Sigma\theta_k = \infty$, then necessarily $\pi_0 = 0$, and then $\pi_j = \theta_j\pi_0 = 0$ for all j, and there is no limiting distribution ($\lim_{t\to\infty} P_{ij}(t) = 0$ for all j).

Example *Linear Growth with Immigration* As described in the example at the end of Section 6.3.3, this process has birth parameters $\lambda_n = a + \lambda n$ and death parameters $\mu_n = \mu n$ for $n = 0, 1, \ldots,$ where $\lambda > 0$ is the individual birth rate, $a > 0$ is the rate of immigration into the population, and $\mu > 0$ is the individual death rate.

Suppose $\lambda < \mu$. It was shown in Section 6.3.3 that the population mean $M(t)$ converges to $a/(\mu - \lambda)$ as $t \to \infty$. Here, we will determine the limiting distribution of the process under the same condition $\lambda < \mu$.

Then, $\theta_0 = 1$, $\theta_1 = a/\mu$, $\theta_2 = a(a+\lambda)/[\mu(2\mu)]$, $\theta_3 = a(a+\lambda)(a+2\lambda)/[\mu(2\mu)(3\mu)]$, and, in general,

$$\begin{aligned}
\theta_k &= \frac{a(a+\lambda)\cdots[a+(k-1)\lambda]}{\mu^k(k)!}\\
&= \frac{(a/\lambda)[(a/\lambda)+1]\cdots[(a/\lambda)+k-1]}{k!}\left(\frac{\lambda}{\mu}\right)^k\\
&= \binom{(a/\lambda)+k-1}{k}\left(\frac{\lambda}{\mu}\right)^k.
\end{aligned}$$

Now, use the infinite binomial formula (Chapter 1, equation (1.71)),

$$(1-x)^{-N} = \sum_{k=0}^{\infty}\binom{N+k-1}{k}x^k \quad \text{for} |x| < 1,$$

to determine that

$$\sum_{k=0}^{\infty}\theta_k = \sum_{k=0}^{\infty}\binom{(a/\lambda)+k-1}{k}\left(\frac{\lambda}{\mu}\right)^k = \left(1-\frac{\lambda}{\mu}\right)^{-(a/\lambda)}$$

when $\lambda < \mu$. Thus, $\pi_0 = (1-\lambda/\mu)^{a/\lambda}$, and

$$\pi_k = \left(\frac{\lambda}{\mu}\right)^k \frac{(a/\lambda)[(a/\lambda)+1]\cdots[(a/\lambda)+k-1]}{k!}(1-\lambda/\mu)^{a/\lambda} \quad \text{for } k > 1.$$

Example *Repairman Models* A system is composed of N machines, of which at most $M \le N$ can be operating at any one time. The rest are "spares." When a machine is operating, it operates a random length of time until failure. Suppose this failure time is exponentially distributed with parameter μ.

When a machine fails, it undergoes repair. At most R machines can be "in repair" at any one time. The repair time is exponentially distributed with parameter λ. Thus, a machine can be in any of four states: (1) operating; (2) "up," but not operating, i.e., a spare; (3) in repair; and (4) waiting for repair. There is a total of N machines in the system. At most M can be operating. At most R can be in repair.

The action is diagrammed in Figure 6.5.

Let $X(t)$ be the number of machines "up" at time t, either operating or spare. Then, (we assume) the number operating is $\min\{X(t), M\}$, and the number of spares is $\max\{0, X(t) - M\}$. Let $Y(t) = N - X(t)$ be the number of machines "down." Then, the number in repair is $\min\{Y(t), R\}$, and the number waiting for repair is $\max\{0, Y(t) - R\}$.

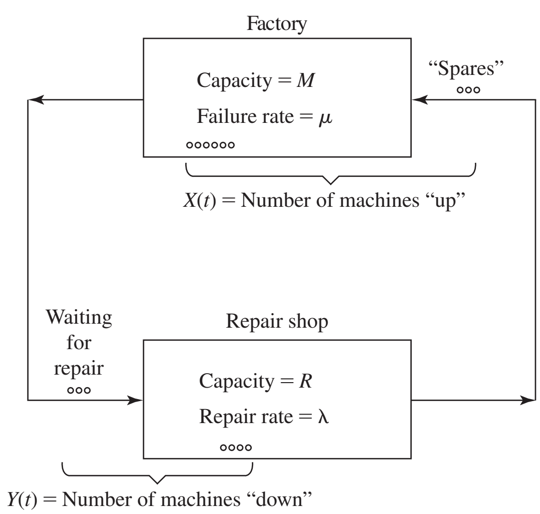

Figure 6.5 Repairman model.

The foregoing formulas permit us to determine the number of machines in any category, once $X(t)$ is known.

Then, $X(t)$ is a finite-state birth and death process[1] with parameters

$$\begin{aligned}\lambda_n &= \lambda \times \min\{N-n, R\}\\ &= \begin{cases}\lambda R & \text{for } n = 0, 1, \ldots, N-R,\\ \lambda(N-n) & \text{for } n = N-R+1, \ldots, N,\end{cases}\end{aligned}$$

and

$$\mu_n = \mu \times \min\{n, M\} = \begin{cases}\mu n & \text{for } n = 0, 1, \ldots, M,\\ \mu M & \text{for } n = M+1, \ldots, N.\end{cases}$$

It is now a routine task to determine the limiting probability distribution for any values of λ, μ, N, M, and R. (See Problems 6.4.1 and 6.4.7.) In terms of the limiting probabilities $\pi_0, \pi_1, \ldots, \pi_N$, some quantities of interest are the following:

$$\begin{aligned}\textit{Average Machines Operating} &= \pi_1 + 2\pi_2 + \cdots + M\pi_M\\ &\quad + M(\pi_{M+1} + \cdots + \pi_N);\\ \textit{Long Run Utilization} &= \frac{\text{Average Machines Operating}}{\text{Capacity}}\\ &= \frac{\pi_1 + 2\pi_2 + \cdots + M\pi_M}{M}\\ &\quad + (\pi_{M+1} + \cdots + \pi_N);\\ \textit{Average Idle Repair Capacity} &= 1\pi_{N-R+1} + 2\pi_{N-R+2} + \cdots + R\pi_N.\end{aligned}$$

These and other similar quantities can be used to evaluate the desirability of adding additional repair capability, additional spare machines, and other possible improvements.

The stationary distribution assumes quite simple forms in certain special cases. For example, consider the special case in which $M = N = R$. The situation arises, for instance, when each machine's operator becomes its repairman upon its failure. Then, $\lambda_n = \lambda(N-n)$ and $\mu_n = \mu n$ for $n = 0, 1, \ldots, N$, and following (6.36), we determine $\theta_0 = 1, \theta_1 = \lambda N/\mu, \theta_2 = (\lambda N)\lambda(N-1)/\mu(2\mu)$, and, in general,

$$\theta_k = \frac{N(N-1)\cdots(N-k+1)}{(1)(2)\cdots(k)}\left(\frac{\lambda}{\mu}\right)^k = \binom{N}{k}\left(\frac{\lambda}{\mu}\right)^k.$$

[1] The definition of birth and death processes was given for an infinite number of states. The adjustments in the definitions and analyses for the case of a finite number of states are straightforward and even simpler than the original definitions and are left to the reader.

The binomial formula $(1+x)^N = \Sigma_{k=0}^{N}\binom{N}{k}x^k$ applies to yield

$$\sum_{k=0}^{N}\theta_k = \sum_{k=0}^{N}\binom{N}{k}\left(\frac{\lambda}{\mu}\right)^k = \left(1+\frac{\lambda}{\mu}\right)^N.$$

Thus, $\pi_0 = [1+(\lambda/\mu)]^{-N} = [\mu/(\lambda+\mu)]^N$, and

$$\begin{aligned}\pi_k &= \binom{N}{k}\left(\frac{\lambda}{\mu}\right)^k [\mu/(\lambda+\mu)]^N \\ &= \binom{N}{k}\left(\frac{\lambda}{\lambda+\mu}\right)^k\left(\frac{\mu}{\lambda+\mu}\right)^{N-k}.\end{aligned} \tag{6.38}$$

We recognize (6.38) as the familiar binomial distribution.

Example *Logistic Process* Suppose we consider a population whose size $X(t)$ ranges between two fixed integers N and $M (N < M)$ for all $t \geq 0$. We assume that the birth and death rates per individual at time t are given by

$$\lambda = \alpha(M - X(t)) \quad \text{and} \quad \mu = \beta(X(t) - N)$$

and that the individual members of the population act independently of each other. The resulting birth and death rates for the population then become

$$\lambda_n = \alpha n(M-n) \quad \text{and} \quad \mu_n = \beta n(n-N).$$

To see this, we observe that if the population size $X(t)$ is n, then each of the n individuals has an infinitesimal birth rate λ so that $\lambda_u = \alpha n(M-n)$. The same rationale applies in the interpretation of the μ_n.

Under such conditions, one would expect the process to fluctuate between the two constants N and M, since, e.g., if $X(t)$ is near M, the death rate is high and the birth rate is low, and then $X(t)$ will tend toward N. Ultimately, the process should display stationary fluctuations between the two limits N and M.

The stationary distribution in this case is

$$\pi_{N+m} = \frac{c}{N+m}\binom{M-N}{m}\left(\frac{\alpha}{\beta}\right)^m, \quad m = 0, 1, 2, \ldots, M-N,$$

where c is an appropriate constant determined so that $\Sigma_m \pi_{N+m} = 1$. To see this, we observe that

$$\begin{aligned}\theta_{N+m} &= \frac{\lambda_N\lambda_{N+1}\cdots\lambda_{N+m-1}}{\mu_{N+1}\mu_{N+2}\cdots\mu_{N+m}} \\ &= \frac{\alpha^m N(N+1)\cdots(N+m-1)(M-N)\cdots(M-N-m+1)}{\beta^m(N+1)\cdots(N+m)m!} \\ &= \frac{N}{N+m}\binom{M-N}{m}\left(\frac{\alpha}{\beta}\right)^m.\end{aligned}$$

Example *Some Genetic Models* Consider a population consisting of N individuals who are either of gene type **a** or gene type **A**. The state of the process $X(t)$ represents the number of **a** individuals at time t. We assume that the probability that any individual dies and is replaced by another during the time interval $(t, t+h)$ is $\lambda h + o(h)$ independent of the values of $X(t)$ and that the probability of two or more changes occurring in a time interval h is $o(h)$.

The changes in the population structure are affected as follows. An individual is to be replaced by another chosen randomly from the population; that is, if $X(t) = j$, then an **a**-type is selected to be replaced with probability j/N and an **A**-type with probability $1 - j/N$. We refer to this stage as death. Next, birth takes place by the following rule. Another selection is made randomly from the population to determine the type of the new individual replacing the one who died. The model introduces mutation pressures that admit the possibility that the type of the new individual may be altered upon birth. Specifically, let γ_1 denote the probability that an **a**-type mutates to an **A**-type, and let γ_2 denote the probability of an **A**-type mutating to an **a**-type.

The probability that the new individual added to the population is of type **a** is

$$\frac{j}{N}(1-\gamma_1) + \left(1 - \frac{j}{N}\right)\gamma_2. \tag{6.39}$$

We deduce this formula as follows: The probability that we select an **a**-type and that no mutation occurs is $(j/N)(1-\gamma_1)$. Moreover, the final type may be an **a**-type if we select an **A**-type that subsequently mutates into an **a**-type. The probability of this contingency is $(1 - j/N)\gamma_2$. The combination of these two possibilities gives (6.39).

We assert that the conditional probability that $X(t+) - X(t) = 1$ when a change of state occurs is

$$\left(1 - \frac{j}{N}\right)\left[\frac{j}{N}(1-\gamma_1) + \left(1 - \frac{j}{N}\right)\gamma_2\right], \quad \text{where } X(t) = j. \tag{6.40}$$

In fact, the **a**-type population size can increase only if an **A**-type dies (is replaced). This probability is $1 - (j/N)$. The second factor is the probability that the new individual is of type **a** as in (6.39).

In a similar way, we find that the conditional probability that $X(t+) - X(t) = -1$ when a change of state occurs is

$$\frac{j}{N}\left[\left(1 - \frac{j}{N}\right)(1-\gamma_2) + \frac{j}{N}\gamma_1\right], \quad \text{where } X(t) = j.$$

The number of type **a** individuals in the population is thus a birth and death process with a finite number of states and infinitesimal birth and death rates

$$\lambda_j = \lambda\left(1 - \frac{j}{N}\right)\left[\frac{j}{N}(1-\gamma_1) + \left(1 - \frac{j}{N}\right)\gamma_2\right]N$$

and

$$\mu_j = \lambda \frac{j}{N}\left[\frac{j}{N}\gamma_1 + \left(1 - \frac{j}{N}\right)(1-\gamma_2)\right]N, \quad 0 \le j \le N.$$

Although these parameters seem rather complicated, it is interesting to see what happens to the stationary measure $\{\pi_k\}_{k=0}^{N}$ if we let the population size $N \to \infty$ and the probabilities of mutation per individual γ_1 and γ_2 tend to zero in such a way that $\gamma_1 N \to \kappa_1$ and $\gamma_2 N \to \kappa_2$, where $0 < \kappa_1,\ \kappa_2 < \infty$. At the same time, we shall transform the state of the process to the interval $[0, 1]$ by defining new states j/N, i.e., the fraction of **a**-types in the population. To examine the stationary density at a fixed fraction x, where $0 < x < 1$, we shall evaluate π_k as $k \to \infty$ in such a way that $k = [xN]$, where $[xN]$ is the greatest integer less than or equal to xN.

Keeping these relations in mind, we write

$$\lambda_j = \frac{\lambda(N-j)}{N}(1-\gamma_1-\gamma_2)j\left(1+\frac{a}{j}\right), \quad \text{where } a = \frac{N\gamma_2}{1-\gamma_1-\gamma_2},$$

and

$$\mu_j = \frac{\lambda(N-j)}{N}(1-\gamma_1-\gamma_2)j\left(1+\frac{b}{N-j}\right), \quad \text{where } b = \frac{N\gamma_1}{1-\gamma_1-\gamma_2}.$$

Then,

$$\begin{aligned}\log\theta_k &= \sum_{j=0}^{k-1}\log\lambda_j - \sum_{j=1}^{k}\log\mu_j \\ &= \sum_{j=1}^{k-1}\log\left(1+\frac{a}{j}\right) - \sum_{j=1}^{k-1}\log\left(1+\frac{b}{N-j}\right) + \log Na \\ &\quad - \log(N-k)k\left(1+\frac{b}{N-k}\right).\end{aligned}$$

Now, using the expansion

$$\log(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \cdots, \quad |x| < 1,$$

it is possible to write

$$\sum_{j=1}^{k-1}\log\left(1+\frac{a}{j}\right) = a\sum_{j=1}^{k-1}\frac{1}{j} + c_k,$$

where c_k approaches a finite limit as $k \to \infty$. Therefore, using the relation

$$\sum_{j=1}^{k-1} \frac{1}{j} \sim \log k \quad \text{as } k \to \infty,$$

we have

$$\sum_{j=1}^{k-1} \log\left(1 + \frac{a}{j}\right) \sim \log k^a + c_k \quad \text{as } k \to \infty.$$

In a similar way, we obtain

$$\sum_{j=1}^{k-1} \log\left(1 + \frac{b}{N-j}\right) \sim \log \frac{N^b}{(N-k)^b} + d_k \quad \text{as } k \to \infty,$$

where d_k approaches a finite limit as $k \to \infty$. Using the above relations, we have

$$\log \theta_k \sim \log\left(C_k \frac{k^a (N-k)^b Na}{N^b (N-k)k}\right) \quad \text{as } k \to \infty, \tag{6.41}$$

where $\log C_k = c_k + d_k$, which approaches a limit, say C, as $k \to \infty$. Notice that $a \to \kappa_2$ and $b \to \kappa_1$ as $N \to \infty$. Since $k = [Nx]$, we have, for $N \to \infty$,

$$\theta_k \sim C\kappa_2 N^{\kappa_2 - 1} x^{\kappa_2 - 1}(1-x)^{\kappa_1 - 1}.$$

Now, from (6.41), we have

$$\theta_k \sim aC_k k^{a-1}\left(1 - \frac{k}{N}\right)^{b-1}.$$

Therefore,

$$\frac{1}{N^a} \sum_{k=0}^{N-1} \theta_k \sim \frac{a}{N} \sum_{k=0}^{N-1} C_k \left(\frac{k}{N}\right)^{a-1} \left(1 - \frac{k}{N}\right)^{b-1}.$$

Since $C_k \to C$ as k tends to ∞, we recognize the right side as a Riemann sum approximation of

$$\kappa_2 C \int_0^1 x^{\kappa_2 - 1}(1-x)^{\kappa_1 - 1} \mathrm{d}x.$$

Thus,

$$\sum_{i=0}^{N}\theta_i \sim N^{\kappa_2}\kappa_2 C\int\limits_0^1 x^{\kappa_2-1}(1-x)^{\kappa_1-1}\mathrm{d}x$$

so that the resulting density on [0, 1] is

$$\frac{\theta_k}{\Sigma\theta_i}\sim\frac{1}{N}\frac{x^{\kappa_2-1}(1-x)^{\kappa_1-1}}{\int_0^1 x^{\kappa_2-1}(1-x)^{\kappa_1-1}\mathrm{d}x}=\frac{x^{\kappa_2-1}(1-x)^{\kappa_1-1}\mathrm{d}x}{\int_0^1 x^{\kappa_2-1}(1-x)^{\kappa_1-1}\mathrm{d}x},$$

since $\mathrm{d}x \sim 1/N$. This is a beta distribution with parameters κ_1 and κ_2.

Exercises

6.4.1 In a birth and death process with birth parameters $\lambda_n = \lambda$ for $n = 0, 1, \ldots$ and death parameters $\mu_n = \mu n$ for $n = 0, 1, \ldots$, we have

$$P_{0j}(t)=\frac{(\lambda p)^j \mathrm{e}^{-\lambda p}}{j!},$$

where

$$p=\frac{1}{\mu}\left[1-\mathrm{e}^{-\mu t}\right].$$

Verify that these transition probabilities satisfy the forward equations (6.34), with $i = 0$.

6.4.2 Let $X(t)$ be a birth and death process where the possible states are $0, 1, \ldots, N$, and the birth and death parameters are, respectively, $\lambda_n = \alpha(N-n)$ and $\mu_n = \beta n$. Determine the stationary distribution.

6.4.3 Determine the stationary distribution for a birth and death process having infinitesimal parameters $\lambda_n = \alpha(n+1)$ and $\mu_n = \beta n^2$ for $n = 0, 1, \ldots$, where $0 < \alpha < \beta$.

6.4.4 Consider two machines, operating simultaneously and independently, where both machines have an exponentially distributed time to failure with mean $1/\mu$ (μ is the failure rate). There is a single repair facility, and the repair times are exponentially distributed with rate λ.

(a) In the long run, what is the probability that no machines are operating?

(b) How does your answer in (a) change if at most one machine can operate, and thus be subject to failure, at any time?

6.4.5 Consider the birth and death parameters $\lambda_n = \theta < 1$ and $\mu_n = n/(n+1)$ for $n = 0, 1, \ldots$. Determine the stationary distribution.

6.4.6 A birth and death process has parameters $\lambda_n = \lambda$ and $\mu_n = n\mu$, for $n = 0, 1, \ldots$. Determine the stationary distribution.

Problems

6.4.1 For the repairman model of the second example of this section, suppose that $M = N = 5, R = 1, \lambda = 2$, and $\mu = 1$. Using the limiting distribution for the system, determine

(a) The average number of machines operating.

(b) The equipment utilization.

(c) The average idle repair capacity.

How do these system performance measures change if a second repairman is added?

6.4.2 Determine the stationary distribution, when it exists, for a birth and death process having constant parameters $\lambda_n = \lambda$ for $n = 0, 1, \ldots$ and $\mu_n = \mu$ for $n = 1, 2, \ldots$.

6.4.3 A factory has five machines and a single repairman. The operating time until failure of a machine is an exponentially distributed random variable with parameter (rate) 0.20 per hour. The repair time of a failed machine is an exponentially distributed random variable with parameter (rate) 0.50 per hour. Up to five machines may be operating at any given time, their failures being independent of one another, but at most one machine may be in repair at any time. In the long run, what fraction of time is the repairman idle?

6.4.4 This problem considers a continuous time Markov chain model for the changing pattern of relationships among members in a group. The group has four members: a, b, c, and d. Each pair of the group may or may not have a certain relationship with each other. If they have the relationship, we say that they are *linked*. For example, being linked may mean that the two members are communicating with each other. The following graph illustrates links between a and b, between a and c, and between b and d:

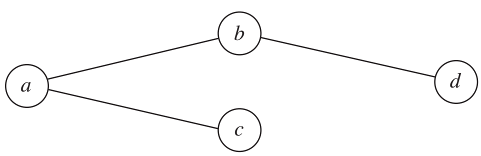

Figure 6.6

Suppose that any pair of unlinked individuals will become linked in a small time interval of length h with probability $\alpha h + o(h)$. Any pair of linked individuals will lose their link in a small time interval of length h with probability $\beta h + o(h)$. Let $X(t)$ denote the number of linked pairs of individuals in the group at time t. Then, $X(t)$ is a birth and death process.

(a) Specify the birth and death parameters λ_k and μ_k for $k = 0, 1, \ldots$.

(b) Determine the stationary distribution for the process.

6.4.5 A chemical solution contains N molecules of type A and an equal number of molecules of type B. A reversible reaction occurs between type A and B molecules in which they bond to form a new compound AB. Suppose that in any small time interval of length h, any particular unbonded A molecule will react with any particular unbonded B molecule with probability $\alpha h + o(h)$, where α is a reaction rate of formation. Suppose also that in any small time interval of length h, any particular AB molecule disassociates into its A and B constituents with probability $\beta h + o(h)$, where β is a reaction rate of dissolution. Let $X(t)$ denote the number of AB molecules at time t. Model $X(t)$ as a birth and death process by specifying the parameters.

6.4.6 A time-shared computer system has three terminals that are attached to a central processing unit (CPU) that can simultaneously handle at most two active users. If a person logs on and requests service when two other users are active, then the request is held in a buffer until it can receive service. Let $X(t)$ be the total number of requests that are either active or in the buffer at time t. Suppose that $X(t)$ is a birth and death process with parameters

$$\lambda_k = \begin{cases} \lambda & \text{for } k = 0, 1, 2, \\ 0 & \text{for } k \geq 3 \end{cases}$$

and

$$\mu_k = \begin{cases} k\mu & \text{for } k = 0, 1, 2, \\ 2\mu & \text{for } k = 3. \end{cases}$$

Determine the long run probability that the computer is fully loaded.

6.4.7 A system consists of three machines and two repairmen. At most two machines can operate at any time. The amount of time that an operating machine works before breaking down is exponentially distributed with mean 5. The amount of time that it takes a single repairman to fix a machine is exponentially distributed with mean 4. Only one repairman can work on a failed machine at any given time. Let $X(t)$ be the number of machines in operating condition at time t.

(a) Calculate the long run probability distribution for $X(t)$.

(b) If an operating machine produces 100 units of output per hour, what is the long run output per hour of the system?

6.4.8 A birth and death process has parameters

$$\lambda_k = \alpha(k+1) \quad \text{for } k = 0, 1, 2, \ldots,$$

and

$$\mu_k = \beta(k+1) \quad \text{for } k = 1, 2, \ldots.$$

Assuming that $\alpha < \beta$, determine the limiting distribution of the process. Simplify your answer as much as possible.

6.5 Birth and Death Processes with Absorbing States

Birth and death processes in which $\lambda_0 = 0$ arise frequently and are correspondingly important. For these processes, the zero state is an absorbing state. A central example is the linear-growth birth and death process without immigration (cf. Section 6.3.3). In this case, $\lambda_n = n\lambda$ and $\mu_n = n\mu$. Since growth of the population results exclusively from the existing population, it is clear that when the population size becomes zero it remains zero thereafter; that is, 0 is an absorbing state.

6.5.1 Probability of Absorption into State 0

It is of interest to compute the probability of absorption into state 0 starting from state $i(i \geq 1)$. This is not, a priori, a certain event, since conceivably the particle (i.e., state variable) may wander forever among the states $(1, 2, \ldots)$ or possibly drift to infinity.

Let $u_i(i = 1, 2, \ldots)$ denote the probability of absorption into state 0 from the initial state i. We can write a recursion formula for u_i by considering the possible states after the first transition. We know that the first transition entails the movements

$$
\begin{aligned}
i &\to i+1 \qquad \text{with probability } \frac{\lambda_i}{\mu_i + \lambda_i}, \\
i &\to i-1 \qquad \text{with probability } \frac{\mu_i}{\mu_i + \lambda_i}.
\end{aligned}
$$

Invoking the familiar first step analysis, we directly obtain

$$
u_i = \frac{\lambda_i}{\mu_i + \lambda_i} u_{i+1} + \frac{\mu_i}{\mu_i + \lambda_i} u_{i-1}, \quad i \geq 1, \tag{6.42}
$$

where $u_0 = 1$.

Another method for deriving (6.42) is to consider the "embedded random walk" associated with a given birth and death process. Specifically, we examine the birth and death process only at the transition times. The discrete time Markov chain generated in this manner is denoted by $\{Y_n\}_{n=0}^{\infty}$, where $Y_0 = X_0$ is the initial state and $Y_n(n \geq 1)$ is the state at the nth transition. Obviously, the transition probability matrix has the form

$$
\mathbf{P} = \begin{Vmatrix}
1 & 0 & 0 & 0 & \cdots \\
q_1 & 0 & p_1 & 0 & \cdots \\
0 & q_2 & 0 & p_2 & \cdots \\
\vdots & \vdots & & &
\end{Vmatrix},
$$

where

$$
p_i = \frac{\lambda_i}{\lambda_i + \mu_i} = 1 - q_i \quad \text{for } i \geq 1.
$$

The probability of absorption into state 0 for the embedded random walk is the same as for the birth and death processes, since both processes execute the same transitions. A closely related problem (gambler's ruin) for a random walk was examined in Chapter 3, Section 3.6.1.

We turn to the task of solving (6.42) subject to the conditions $u_0 = 1$ and $0 \le u_i \le 1 (i \ge 1)$. Rewriting (6.42), we have

$$(u_{i+1} - u_i) = \frac{\mu_i}{\lambda_i}(u_i - u_{i-1}), \quad i \ge 1.$$

Defining $v_i = u_{i+1} - u_i$, we obtain

$$v_i = \frac{\mu_i}{\lambda_i} v_{i-1}, \quad i \ge 1.$$

Iteration of the last relation yields the formula $v_i = \rho_i v_0$, where

$$\rho_0 = 1 \quad \text{and} \quad \rho_i = \frac{\mu_1 \mu_2 \cdots \mu_i}{\lambda_1 \lambda_2 \cdots \lambda_i} \quad \text{for } i \ge 1;$$

and with $u_{i+1} - u_i = v_i$,

$$u_{i+1} - u_i = v_i = \rho_i v_0 = \rho_i(u_1 - u_0) = \rho_i(u_1 - 1) \quad \text{for } i \ge 1.$$

Summing these last equations from $i = 1$ to $i = m - 1$, we have

$$u_m - u_1 = (u_1 - 1)\sum_{i=1}^{m-1} \rho_i, \quad m > 1. \tag{6.43}$$

Since u_m by its very meaning is bounded by 1, we see that if

$$\sum_{i=1}^{\infty} \rho_i = \infty, \tag{6.44}$$

then necessarily $u_1 = 1$ and $u_m = 1$ for all $m \ge 2$. In other words, if (6.44) holds, then ultimate absorption into state 0 is certain from any initial state.

Suppose $0 < u_1 < 1$; then, of course,

$$\sum_{i=1}^{\infty} \rho_i < \infty.$$

Obviously, u_m is decreasing in m, since passing from state m to state 0 requires entering the intermediate states in the intervening time. Furthermore, it can be shown that $u_m \to 0$ as $m \to \infty$. Now, letting $m \to \infty$ in (6.43) permits us to solve for u_1; thus,

$$u_1 = \frac{\Sigma_{i=1}^{\infty} \rho_i}{1 + \Sigma_{i=1}^{\infty} \rho_i};$$

and then from (6.43), we obtain

$$u_m = \frac{\Sigma_{i=m}^{\infty}\rho_i}{1+\Sigma_{i=1}^{\infty}\rho_i}, \quad m \geq 1.$$

6.5.2 Mean Time Until Absorption

Consider the problem of determining the mean time until absorption, starting from state m.

We assume that condition (6.44) holds so that absorption is certain. Notice that we cannot reduce our problem to a consideration of the embedded random walk, since the actual time spent in each state is relevant for the calculation of the mean absorption time.

Let w_i be the mean absorption time starting from state i (this could be infinite). Considering the possible states following the first transition, instituting a first step analysis, and recalling the fact that the mean waiting time in state i is $(\lambda_i + \mu_i)^{-1}$ (it is actually exponentially distributed with parameter $\lambda_i + \mu_i$), we deduce the recursion relation

$$w_i = \frac{1}{\lambda_i+\mu_i} + \frac{\lambda_i}{\lambda_i+\mu_i}w_{i+1} + \frac{\mu_i}{\lambda_i+\mu_i}w_{i-1}, \quad i \geq 1, \tag{6.45}$$

where $w_0 = 0$. Letting $z_i = w_i - w_{i+1}$ and rearranging (6.45) leads to

$$z_i = \frac{1}{\lambda_i} + \frac{\mu_i}{\lambda_i}z_{i-1}, \quad i \geq 1. \tag{6.46}$$

Iterating this relation gives

$$\begin{aligned}
z_1 &= \frac{1}{\lambda_1} + \frac{\mu_1}{\lambda_1}z_0,\\
z_2 &= \frac{1}{\lambda_2} + \frac{\mu_2}{\lambda_2}z_1 = \frac{1}{\lambda_2} + \frac{\mu_2}{\lambda_2\lambda_1} + \frac{\mu_2\mu_1}{\lambda_2\lambda_1}z_0,\\
z_3 &= \frac{1}{\lambda_3} + \frac{\mu_3}{\lambda_3\lambda_2} + \frac{\mu_3\mu_2}{\lambda_3\lambda_2\lambda_1} + \frac{\mu_3\mu_2\mu_1}{\lambda_3\lambda_2\lambda_1}z_0,
\end{aligned}$$

and finally

$$z_m = \sum_{i=1}^{m}\frac{1}{\lambda_i}\prod_{j=i+1}^{m}\frac{\mu_j}{\lambda_j} + \left(\prod_{j=1}^{m}\frac{\mu_j}{\lambda_j}\right)z_0.$$

(The product $\Pi_{m+1}^{m}\mu_j/\lambda_j$ is interpreted as 1.) Using the notation

$$\rho_0 = 1 \quad \text{and} \quad \rho_i = \frac{\mu_1\mu_2\cdots\mu_i}{\lambda_1\lambda_2\cdots\lambda_i}, \quad i \geq 1,$$

the expression for z_m becomes

$$z_m = \sum_{i=1}^{m} \frac{1}{\lambda_i} \frac{\rho_m}{\rho_i} + \rho_m z_0,$$

or, since $z_m = w_m - w_{m+1}$ and $z_0 = w_0 - w_1 = -w_1$, then

$$\frac{1}{\rho_m}(w_m - w_{m+1}) = \sum_{i=1}^{m} \frac{1}{\lambda_i \rho_i} - w_1. \tag{6.47}$$

If $\Sigma_{i=1}^{\infty}(1/\lambda_i\rho_i) = \infty$, then inspection of (6.47) reveals that necessarily $w_1 = \infty$. Indeed, it is probabilistically evident that $w_m < w_{m+1}$ for all m, and this property would be violated for m large if we assume to the contrary that w_1 is finite.

Now, suppose $\Sigma_{i=1}^{\infty}(1/\lambda_i\rho_i) < \infty$; then, letting $m \to \infty$ in (6.47) gives

$$w_1 = \sum_{i=1}^{\infty} \frac{1}{\lambda_i \rho_i} - \lim_{m\to\infty} \frac{1}{\rho_m}(w_m - w_{m+1}).$$

It is more involved but still possible to prove that

$$\lim_{m\to\infty} \frac{1}{\rho_m}(w_m - w_{m+1}) = 0,$$

and then,

$$w_1 = \sum_{i=1}^{\infty} \frac{1}{\lambda_i \rho_i}.$$

We summarize the discussion of this section in the following theorem:

Theorem 6.1. *Consider a birth and death process with birth and death parameters λ_n and $\mu_n, n \geq 1$, where $\lambda_0 = 0$ so that 0 is an absorbing state.*
The probability of absorption into state 0 from the initial state m is

$$u_m = \begin{cases} \dfrac{\Sigma_{i=m}^{\infty}\rho_i}{1+\Sigma_{i=1}^{\infty}\rho_i} & \text{if } \sum\limits_{i=1}^{\infty} \rho_i < \infty, \\ 1 & \text{if } \sum\limits_{i=1}^{\infty} \rho_i = \infty. \end{cases} \tag{6.48}$$

The mean time to absorption is

$$w_m = \begin{cases} \infty & \text{if } \sum\limits_{i=1}^{\infty} \frac{1}{\lambda_i\rho_i} = \infty, \\ \sum\limits_{i=1}^{\infty} \frac{1}{\lambda_i\rho_i} + \sum\limits_{k=1}^{m-1} \rho_k \sum\limits_{j=k+1}^{\infty} \frac{1}{\lambda_j\rho_j} & \text{if } \sum\limits_{i=1}^{\infty} \frac{1}{\lambda_i\rho_i} < \infty, \end{cases} \tag{6.49}$$

where $\rho_0 = 1$ and $\rho_i = (\mu_1\mu_2\cdots\mu_i)/(\lambda_1\lambda_2\cdots\lambda_i)$.

Example *Population Processes* Consider the linear growth birth and death process without immigration (cf. Section 6.3.3) for which $\mu_n = n\mu$ and $\lambda_n = n\lambda, n = 0, 1, \ldots$. During a short time interval of length h, a *single individual* in the population dies with probability $\mu h + o(h)$ and gives birth to a new individual with probability $\lambda h + o(h)$, and thus, $\mu > 0$ and $\lambda > 0$ represent the *individual* death and birth rates, respectively.

Substitution of $a = 0$ and $i = m$ in equation (6.25) determines the mean population size at time t for a population starting with $X(0) = m$ individuals. This mean population size is $M(t) = m\mathrm{e}^{(\lambda-\mu)t}$, exhibiting exponential growth or decay depending on whether $\lambda > \mu$ or $\lambda < \mu$.

Let us now examine the extinction phenomenon and determine the probability that the population eventually dies out. This phenomenon corresponds to absorption in state 0 for the birth and death process.

When $\lambda_n = n\lambda$ and $\mu_n = n\mu$, a direct calculation yields $\rho_i = (\mu/\lambda)^i$, and then,

$$\sum_{i=m}^{\infty} \rho_i = \sum_{i=m}^{\infty} (\mu/\lambda)^i = \begin{cases} \dfrac{(\mu/\lambda)^m}{1-(\mu/\lambda)} & \text{when } \lambda > \mu, \\ \infty & \text{when } \lambda \le \mu. \end{cases}$$

From Theorem 6.1, the probability of eventual extinction starting with m individuals is

$$\Pr\{\text{Extinction}|X(0) = m\} = \begin{cases} (\mu/\lambda)^m & \text{when } \lambda > \mu, \\ 1 & \text{when } \lambda \le \mu. \end{cases} \tag{6.50}$$

When $\lambda = \mu$, the process is sure to vanish eventually. Yet, in this case, the mean population size remains constant at the initial population level. Similar situations where mean values do not adequately describe population behavior frequently arise when stochastic elements are present.

We turn attention to the mean time to extinction assuming that extinction is certain, i.e., when $\lambda \le \mu$. For a population starting with a single individual, then, from (6.49) with $m = 1$, we determine this mean time to be

$$\begin{aligned} \sum_{i=1}^{\infty} \frac{1}{\lambda_i \rho_i} &= \frac{1}{\lambda} \sum_{i=1}^{\infty} \frac{1}{i} \left(\frac{\lambda}{\mu}\right)^i \\ &= \frac{1}{\lambda} \sum_{i=1}^{\infty} \int_0^{(\lambda/\mu)} x^{i-1} \mathrm{d}x \\ &= \frac{1}{\lambda} \int_0^{(\lambda/\mu)} \sum_{i=1}^{\infty} x^{i-1} \mathrm{d}x \\ &= \frac{1}{\lambda} \int_0^{(\lambda/\mu)} \frac{\mathrm{d}x}{(1-x)} \end{aligned} \tag{6.51}$$

$$= -\frac{1}{\lambda}\ln(1-x)\Big|_0^{(\lambda/\mu)}$$

$$= \begin{cases} \dfrac{1}{\lambda}\ln\left(\dfrac{\mu}{\mu-\lambda}\right) & \text{when } \mu > \lambda, \\ \infty & \text{when } \mu = \lambda. \end{cases}$$

When the birth rate λ exceeds the death rate μ, a linear growth birth and death process can, with strictly positive probability, grow without limit. In contrast, many natural populations exhibit density-dependent behavior wherein the individual birth rates decrease or the individual death rates increase or both changes occur as the population grows. These changes are ascribed to factors including limited food supplies, increased predation, crowding, and limited nesting sites. Accordingly, we introduce a notion of environmental *carrying capacity* K, an upper bound that the population size cannot exceed.

Since all individuals have a chance of dying, with a finite carrying capacity, all populations will eventually become extinct. Our measure of population fitness will be the mean time to extinction, and it is of interest to population ecologists studying colonization phenomena to examine how the capacity K, the birth rate λ, and the death rate μ affect this mean population lifetime.

The model should have the properties of exponential growth (on the average) for small populations, as well as the ceiling K beyond which the population cannot grow. There are several ways of approaching the population size K and staying there at equilibrium. Since all such models give more or less the same qualitative results, we stipulate the simplest model, in which the birth parameters are

$$\lambda_n = \begin{cases} n\lambda & \text{for } n = 0, 1, \ldots, K-1, \\ 0 & \text{for } n > K. \end{cases}$$

Theorem 6.1 yields w_1, the mean time to population extinction starting with a single individual, as given by

$$w_1 = \sum_{i=1}^{\infty} \frac{1}{\lambda_i \rho_i} = \sum_{i=1}^{\infty} \frac{\lambda_1\lambda_2\cdots\lambda_{i-1}}{\mu_1\mu_2\cdots\mu_i} = \frac{1}{\mu}\sum_{i=1}^{K}\frac{1}{i}\left(\frac{\lambda}{\mu}\right)^{i-1}. \tag{6.52}$$

Equation (6.52) isolates the distinct factors influencing the mean time to population extinction. The first factor is $1/\mu$, the mean lifetime of an individual, since μ is the individual death rate. Thus, the sum in (6.52) represents the mean *generations*, or mean lifespans, to population extinction, a dimensionless quantity that we denote by

$$M_{\mathrm{g}} = \mu w_1 = \sum_{i=1}^{K} \frac{1}{i}\theta^{i-1}, \quad \text{where } \theta = \frac{\lambda}{\mu}. \tag{6.53}$$

Next, we examine the influence of the birth–death, or reproduction, ratio $\theta = \lambda/\mu$, and the carrying capacity K on the mean time to extinction. Since λ represents the individual birth rate and $1/\mu$ is the mean lifetime of a single member in the population, we may interpret the reproduction ratio $\theta = \lambda(1/\mu)$ as the mean number of offspring of an arbitrary individual in the population. Accordingly, we might expect significantly different behavior when $\theta < 1$ as opposed to when $\theta > 1$, and this is indeed the case. A carrying capacity of $K = 100$ is small. When K is on the order of 100 or more, we have the following accurate approximations, their derivations being sketched in Exercises 6.5.1 and 6.5.2:

$$M_g \approx \begin{cases} \dfrac{1}{\theta} \ln\left(\dfrac{1}{1-\theta}\right) & \text{for } \theta < 1, \\ 0.5772157 + \ln K & \text{for } \theta = 1, \\ \dfrac{1}{K}\left(\dfrac{\theta^K}{\theta - 1}\right) & \text{for } \theta > 1. \end{cases} \tag{6.54}$$

The contrast between $\theta < 1$ and $\theta > 1$ is vivid. When $\theta < 1$, the mean generations to extinction M_g is almost independent of carrying capacity K and approaches the asymptotic value $\theta^{-1} \ln(1-\theta)^{-1}$ quite rapidly. When $\theta > 1$, the mean generations to extinction M_g grows exponentially in K. Some calculations based on (6.54) are given in Table 6.1.

Example *Sterile Male Insect Control* The screwworm fly, a cattle pest in warm climates, was eliminated from the southeastern United States by the release into the environment of sterilized adult male screwworm flies. When these males, artificially sterilized by radiation, mate with native females, there are no offspring, and in this manner, part of the reproductive capacity of the natural population is nullified by their presence. If the sterile males are sufficiently plentiful so as to cause even a small decline in the population level, then this decline accelerates in succeeding generations even if the number of sterile males is maintained at approximately the same level because the ratio of sterile to fertile males will increase as the natural population drops. Because of this compounding effect, if the sterile male control method works at all, it works to such an extent as to drive the native population to extinction in the area in which it is applied.

Recently, a multibillion-dollar effort involving the sterile male technique has been proposed for the control of the cotton boll weevil. In this instance, it was felt that

Table 6.1 Mean generations to extinction for a population starting with a single parent and where θ is the reproduction rate and K is the environmental capacity

K	$\theta = 0.8$	$\theta = 1$	$\theta = 1.2$
10	1.96	2.88	3.10
100	2.01	5.18	4.14×10^4
1000	2.01	7.48	7.59×10^{76}

pretreatment with a pesticide could reduce the natural population size to a level such that the sterile male technique would become effective. Let us examine this assumption, first with a deterministic model and then in a stochastic setting.

For both models, we suppose that sexes are present in equal numbers, that sterile and fertile males are equally competitive, and that a constant number S of sterile males is present in each generation. In the deterministic case, if N_0 fertile males are in the parent generation and the N_0 fertile females choose mates equally likely from the entire male population, then the fraction $N_0/(N_0+S)$ of these matings will be with fertile males and will produce offspring. Letting θ denote the number of male offspring that results from a fertile mating, we calculate the size N of the next generation according to

$$N_1 = \theta N_0 \left(\frac{N_0}{N_0+S} \right). \tag{6.55}$$

For a numerical example, suppose that there are $N_0 = 100$ fertile males (and an equal number of fertile females) in the parent generation of the native population, and that $S = 100$ sterile male insects are released. If $\theta = 4$, meaning that a fertile mating produces four males (and four females) for the succeeding generation, then the number of both sexes in the first generation is

$$N_1 = 4(100)\left(\frac{100}{100+100} \right) = 200;$$

the population has increased, and the sterile male control method has failed.

On the other hand, if a pesticide can be used to reduce the initial population size to $N_0 = 20$, or 20% of its former level, and $S = 100$ sterile males are released, then

$$N_1 = 4(20)\left(\frac{20}{20+100} \right) = 13.33,$$

and the population is declining. The succeeding population sizes are given in Table 6.2, above. With the pretreatment, the population becomes extinct by the fourth generation.

Often deterministic or average value models will adequately describe the evolution of large populations. But extinction is a small population phenomenon, and even in the presence of significant long-term trends, small populations are strongly influenced by

Table 6.2 The trend of an insect population subject to sterile male releases

Generation	Number of Insects Natural Population	Number of Sterile Insects	Ratio Sterile to Fertile	Number of Progeny
Parent	20	100	5:1	13.33
F_1	13.33	100	7.5:1	6.27
F_2	6.27	100	16:1	1.48
F_3	1.48	100	67.5:1	0.09
F_4	0.09	100	1156:1	—

the chance fluctuations that determine which of extinction or recolonization will occur. This fact motivates us to examine a stochastic model of the evolution of a population in the presence of sterile males. The factors in our model are

λ, the individual birth rate;
μ, the individual death rate;
$\theta = \lambda/\mu$, the mean offspring per individual;
K, the carrying capacity of the environment;
S, the constant number of sterile males in the population;
m, the initial population size.

We assume that both sexes are present in equal numbers in the natural population and that $X(t)$, the number of either sex present at time t, evolves as a birth and death process with parameters

$$\lambda_n = \begin{cases} \lambda n\left(\dfrac{n}{n+S}\right) & \text{if } 0 \le n < K, \\ 0 & \text{for } n \ge K, \end{cases}$$

and (6.56)

$$\mu_n = \mu n \quad \text{for } n = 0, 1, \ldots.$$

This is the colonization model of the *Population Processes* example, modified in analogy with (6.55) by including in the birth rate the factor $n/(n+S)$ to represent the probability that a given mating will be fertile.

To calculate the mean time to extinction w_m as given in (6.49), we first use (6.56) to determine

$$\rho_k = \frac{\mu_1\mu_2\cdots\mu_k}{\lambda_1\lambda_2\cdots\lambda_k} = \left(\frac{\mu}{\lambda}\right)^k \frac{(k+S)!}{k!\,S!} \quad \text{for } k = 1, \ldots, K-1,$$

$$\rho_0 = 1, \quad \text{and} \quad \rho_K = \infty, \text{ or } 1/\rho_K = 0,$$

and then substitute these expressions for ρ_k into (6.49) to obtain

$$\begin{aligned} w_m &= \sum_{j=1}^{K} \frac{1}{\lambda_j\rho_j} + \sum_{k=1}^{m-1} \rho_k \sum_{j=k+1}^{K} \frac{1}{\lambda_j\rho_j} \\ &= \sum_{k=0}^{m-1} \rho_k \sum_{j=k+1}^{K} \frac{1}{\lambda_j\rho_j} = \sum_{k=0}^{m-1} \rho_k \sum_{j=k+1}^{K} \frac{1}{\mu_j\rho_{j-1}} \\ &= \frac{1}{\mu}\left\{\sum_{k=0}^{m-1}\sum_{j=k}^{K-1} \frac{1}{j+1}\theta^{j-k}\frac{j!\,(S+k)!}{k!\,(S+j)!}\right\}. \end{aligned} \tag{6.57}$$

Because of the factorials, equation (6.57) presents numerical difficulties when direct computations are attempted. A simple iterative scheme works to provide accurate and effective computation, however. We let

$$\alpha_k = \sum_{j=k}^{K-1} \frac{1}{j+1} \theta^{j-k} \frac{j!\,(S+k)!}{k!\,(S+j)!}$$

so that $w_m = (\alpha_0 + \cdots + \alpha_{m-1})/\mu$. But, it is easily verified that

$$\alpha_{k-1} = \frac{1}{k} + \theta\left(\frac{k}{S+k}\right)\alpha_k.$$

Beginning with $\alpha_K = 0$, one successively computes $\alpha_{K-1}, \alpha_{K-2}, \ldots, \alpha_0$, and then $w_m = (\alpha_0 + \cdots + \alpha_{m-1})/\mu$.

Using this method, we have computed the mean generations to extinction in the stochastic model for comparison with the deterministic model as given in Table 6.3. Table 6.3 lists the mean generations to extinction for various initial population sizes m when $K = S = 100$, $\lambda = 4$, and $\mu = 1$ so that $\theta = 4$. Instead of the four generations to extinction as predicted by the deterministic model when $m = 20$, we now estimate that the population will persist for over 8 billion generations!

What is the explanation for the dramatic difference between the predictions of the deterministic model and the predictions of the stochastic model? The stochastic model allows the small but positive probability that the population will not die out but will recolonize and return to a higher level near the environmental capacity K and then persist for an enormous length of time.

While both models are qualitative, the practical implications cannot be dismissed. In any large-scale control effort, a wide range of habitats and microenvironments is bound to be encountered. The stochastic model suggests the likely possibility that some subpopulation in some pocket might persist and later recolonize the entire area.

Table 6.3 The mean lifespans to extinction in a birth and death model of a population containing a constant number $S = 100$ of sterile males

Initial Population Size	Mean Lifespans to Extinction
20	8,101,227,748
10	4,306,531
5	3,822
4	566
3	65
2	6.3
1	1.2

A sterile male program that depends on a pretreatment with an insecticide for its success is chancy at best.

Exercises

6.5.1 Assuming $\theta < 1$, verify the following steps in the approximation to M_{g}, the mean generation to extinction as given in (6.53):

$$
\begin{aligned}
M_{\mathrm{g}} &= \sum_{i=1}^{K} \frac{1}{i}\theta^{i-1} = \theta^{-1}\sum_{i=1}^{K}\int_{0}^{\theta} x^{i-1}\mathrm{d}x \\
&= \theta^{-1}\int_{0}^{\theta}\frac{1-x^K}{1-x}\mathrm{d}x = \theta^{-1}\int_{0}^{\theta}\frac{\mathrm{d}x}{1-x} - \theta^{-1}\int_{0}^{\theta}\frac{x^K}{1-x}\mathrm{d}x \\
&= \frac{1}{\theta}\ln\frac{1}{1-\theta} - \theta^{-1}\int_{0}^{\theta} x^K\left(1+x+x^2+\cdots\right)\mathrm{d}x \\
&= \frac{1}{\theta}\ln\frac{1}{1-\theta} - \frac{1}{\theta}\left(\frac{\theta^{K+1}}{K+1}+\frac{\theta^{K+2}}{K+2}+\cdots\right) \\
&= \frac{1}{\theta}\ln\frac{1}{1-\theta} - \frac{\theta^K}{K+1}\left(1+\frac{K+1}{K+2}\theta+\frac{K+1}{K+3}\theta^2+\cdots\right) \\
&\approx \frac{1}{\theta}\ln\frac{1}{1-\theta} - \frac{\theta^K}{(K+1)(1-\theta)}.
\end{aligned}
$$

6.5.2 Assume that $\theta > 1$ and verify the following steps in the approximation to M_{g}, the mean generation to extinction as given in (6.53):

$$
\begin{aligned}
M_{\mathrm{g}} &= \sum_{i=1}^{K}\frac{1}{i}\theta^{i-1} = \theta^K\sum_{i=1}^{K}\frac{1}{i}\theta^{K-i+1} \\
&= \theta^K\sum_{j=1}^{K}\frac{1}{K-j+1}\left(\frac{1}{\theta}\right)^j \\
&= \frac{\theta^{K-1}}{K}\left[1+\frac{K}{K-1}\left(\frac{1}{\theta}\right)+\frac{K}{K-2}\left(\frac{1}{\theta}\right)^2+\cdots+\frac{K}{1}\left(\frac{1}{\theta}\right)^{K-1}\right] \\
&\approx \frac{\theta^{K-1}}{K}\left[\frac{1}{1-(1/\theta)}\right] = \frac{\theta^K}{K(\theta-1)}.
\end{aligned}
$$

Problems

6.5.1 Consider the sterile male control model as described in the example entitled "*Sterile Male Insect Control*" and let u_m be the probability that the population becomes extinct before growing to size K starting with $X(0) = m$ individuals. Show that

$$u_m = \frac{\Sigma_{i=m}^{K-1} \rho_i}{\Sigma_{i=0}^{K-1} \rho_i} \quad \text{for } m = 1, \ldots, K,$$

where

$$\rho_i = \theta^{-i} \frac{(S+i)!}{i!}.$$

6.5.2 Consider a birth and death process on the states $0, 1, \ldots, 5$ with parameters

$$\begin{aligned}
&\lambda_0 = \mu_0 = \lambda_5 = \mu_5 = 0, \\
&\lambda_1 = 1, \qquad \lambda_2 = 2, \qquad \lambda_3 = 3, \quad \lambda_4 = 4, \\
&\mu_1 = 4, \qquad \mu_2 = 3, \qquad \mu_3 = 2, \quad \mu = 1.
\end{aligned}$$

Note that 0 and 5 are absorbing states. Suppose the process begins in state $X(0) = 2$.

(a) What is the probability of eventual absorption in state 0?

(b) What is the mean time to absorption?

6.6 Finite-State Continuous Time Markov Chains

A continuous time Markov chain $X(t)(t > 0)$ is a Markov process on the states $0, 1, 2, \ldots$. We assume as usual that the transition probabilities are stationary; that is,

$$P_{ij}(t) = \Pr\{X(t+s) = j | X(s) = i\}. \tag{6.58}$$

In this section, we consider only the case where the state spaces S is finite, labeled as $\{0, 1, 2, \ldots, N\}$.

The Markov property asserts that $P_{ij}(t)$ satisfies

(a) $P_{ij}(t) \geq 0$,

(b) $\sum_{j=0}^{N} P_{ij}(t) = 1, \quad i, j = 0, 1, \ldots, N$, and

(c) $P_{ik}(s+t) = \sum_{j=0}^{N} P_{ij}(s) P_{jk}(t) \quad \text{for } t, s \geq 0 \quad$ (Chapman–Kolmogorov relation),

and we postulate in addition that

(d) $\lim_{t \to 0+} P_{ij}(t) = \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases}$

If $\mathbf{P}(t)$ denotes the matrix $\|P_{ij}(t)\|_{i,j=0}^{N}$, then property (c) can be written compactly in matrix notation as

$$\mathbf{P}(t+s) = \mathbf{P}(t)\mathbf{P}(s), \quad t,s \geq 0. \tag{6.59}$$

Property (d) asserts that $\mathbf{P}(t)$ is continuous at $t = 0$, since the fact $\mathbf{P}(0) = \mathbf{I}$ ($=$ identity matrix) is implied by (6.59). It follows simply from (6.59) that $\mathbf{P}(t)$ is continuous for all $t > 0$. In fact, if $s = h > 0$ in (6.59), then because of (d), we have

$$\lim_{h\to 0+} \mathbf{P}(t+h) = \mathbf{P}(t) \lim_{h\to 0+} \mathbf{P}(h) = \mathbf{P}(t)\mathbf{I} = \mathbf{P}(t). \tag{6.60}$$

On the other hand, for $t > 0$ and $0 < h < t$, we write (6.59) in the form

$$\mathbf{P}(t) = \mathbf{P}(t-h)\mathbf{P}(h). \tag{6.61}$$

But $\mathbf{P}(h)$ is near the identity when h is sufficiently small, and so $\mathbf{P}(h)^{-1}$ [the inverse of $\mathbf{P}(h)$] exists and also approaches the identity $\mathbf{I}$. Therefore,

$$\mathbf{P}(t) = \mathbf{P}(t) \lim_{h\to 0+} (\mathbf{P}(h))^{-1} = \lim_{h\to 0+} \mathbf{P}(t-h). \tag{6.62}$$

The limit relations (6.60) and (6.62) together show that $\mathbf{P}(t)$ is continuous.

Actually, $\mathbf{P}(t)$ is not only continuous but also differentiable in that the limits

$$\begin{aligned} &\lim_{h\to 0+} \frac{1 - P_{ii}(h)}{h} = q_i, \\ &\quad \lim_{h\to 0+} \frac{P_{ij}(h)}{h} = q_{ij}, \quad i \neq j, \end{aligned} \tag{6.63}$$

exist, where $0 \leq q_{ij} < \infty (i \neq j)$ end $0 \leq q_i < \infty$. Starting with the relation

$$1 - P_{ii}(h) = \sum_{j=0, j\neq i}^{N} P_{ij}(h),$$

dividing by h, and letting h decrease to zero yields directly the relation

$$q_i = \sum_{j=0, j\neq i}^{N} q_{ij}.$$

The rates q_i and q_{ij} furnish an infinitesimal description of the process with

$$\begin{aligned} &\Pr\{X(t+h) = j | X(t) = i\} = q_{ij}h + o(h) \quad \text{for } i \neq j, \\ &\Pr\{X(t+h) = i | X(t) = i\} = 1 - q_i h + o(h). \end{aligned}$$

In contrast to the infinitesimal description, the sojourn description of the process proceeds as follows: Starting in state i, the process sojourns there for a duration that is exponentially distributed with parameter q_i. The process then jumps to state $j \neq i$ with probability $p_{ij} = q_{ij}/q_i$; the sojourn time in state j is exponentially distributed with parameter q_j, and so on. The sequence of states visited by the process, denoted by $\xi_0, \xi_1, \ldots$, is a Markov chain with discrete parameter, called the *embedded Markov chain*. Conditioned on the state sequence $\xi_0, \xi_1, \ldots$, the successive sojourn times $S_0, S_1, \ldots$ are independent exponentially distributed random variables with parameters $q_{\xi_0}, q_{\xi_1}, \ldots$, respectively.

Assuming that (6.63) has been verified, we now derive an explicit expression for $P_{ij}(t)$ in terms of the infinitesimal matrix

$$\mathbf{A} = \begin{Vmatrix} -q_0 & q_{01} & \cdots & q_{0N} \\ q_{10} & -q_1 & & q_{1N} \\ \vdots & & & \\ q_{N0} & q_{N1} & \cdots & -q_N \end{Vmatrix}.$$

The limit relations (6.63) can be expressed concisely in matrix form:

$$\lim_{h \to 0+} \frac{\mathbf{P}(h) - \mathbf{I}}{h} = \mathbf{A}, \tag{6.64}$$

which shows that $\mathbf{A}$ is the matrix derivative of $\mathbf{P}(t)$ at $t = 0$. Formally, $\mathbf{A} = \mathbf{P}'(0)$.

With the aid of (6.64) and referring to (6.59), we have

$$\frac{\mathbf{P}(t+h) - \mathbf{P}(t)}{h} = \frac{\mathbf{P}(t)[\mathbf{P}(h) - \mathbf{I}]}{h} = \frac{\mathbf{P}(h) - \mathbf{I}}{h}\mathbf{P}(t). \tag{6.65}$$

The limit on the right exists, and this leads to the matrix differential equation

$$\mathbf{P}'(t) = \mathbf{P}(t)\mathbf{A} = \mathbf{A}\mathbf{P}(t), \tag{6.66}$$

where $\mathbf{P}'(t)$ denotes the matrix whose elements are $P'_{ij}(t) = \mathrm{d}P_{ij}(t)/\mathrm{d}t$. The existence of $P'_{ij}(t)$ is an obvious consequence of (6.64) and (6.65). The differential equations (6.66) under the initial condition $\mathbf{P}(0) = \mathbf{I}$ can be solved by standard methods to yield the formula

$$\mathbf{P}(t) = \mathrm{e}^{\mathbf{A}t} = \mathbf{I} + \sum_{n=1}^{\infty} \frac{\mathbf{A}^n t^n}{n!}. \tag{6.67}$$

Example *The Two-State Markov Chain* Consider a Markov chain $\{X(t)\}$ with states $\{0, 1\}$ whose infinitesimal matrix is

$$\mathbf{A} = \begin{array}{c c} & \begin{array}{cc} 0 & 1 \end{array} \\ \begin{array}{c} 0 \\ 1 \end{array} & \begin{Vmatrix} -\alpha & \alpha \\ \beta & -\beta \end{Vmatrix} \end{array}.$$

The process alternates between states 0 and 1. The sojourn times in state 0 are independent and exponentially distributed with parameter α. Those in state 1 are independent and exponentially distributed with parameter β. We carry out the matrix multiplication

$$\begin{Vmatrix}-\alpha & \alpha \\ \beta & -\beta\end{Vmatrix} \times \begin{Vmatrix}-\alpha & \alpha \\ \beta & -\beta\end{Vmatrix} = \begin{Vmatrix}\alpha^2+\alpha\beta & -\alpha^2-\alpha\beta \\ -\beta^2-\alpha\beta & \beta^2+\alpha\beta\end{Vmatrix}$$
$$= -(\alpha+\beta)\begin{Vmatrix}-\alpha & \alpha \\ \beta & -\beta\end{Vmatrix}$$

to see that $\mathbf{A}^2 = -(\alpha+\beta)\mathbf{A}$. Repeated multiplication by $\mathbf{A}$ then yields

$$\mathbf{A}^n = [-(\alpha+\beta)]^{n-1}\mathbf{A},$$

which when inserted into (6.67) simplifies the sum according to

$$\begin{aligned}\mathbf{P}(t) &= \mathbf{I} - \frac{1}{\alpha+\beta}\sum_{n=1}^{\infty}\frac{[-(\alpha+\beta)t]^n}{n!}\mathrm{A} \\ &= \mathbf{I} - \frac{1}{\alpha+\beta}\left[\mathrm{e}^{-(\alpha+\beta)t}-1\right]\mathbf{A} \\ &= \mathbf{I} + \frac{1}{\alpha+\beta}\mathbf{A} - \frac{1}{\alpha+\beta}\mathbf{A}\mathrm{e}^{-(\alpha+\beta)t}.\end{aligned}$$

And with $\pi = \alpha/(\alpha+\beta)$ and $\tau = \alpha+\beta$,

$$\mathbf{P}(t) = \begin{Vmatrix}1-\pi & \pi \\ 1-\pi & \pi\end{Vmatrix} + \begin{Vmatrix}\pi & -\pi \\ -(1-\pi) & (1-\pi)\end{Vmatrix}\mathrm{e}^{-\tau t},$$

which is the matrix expression for equations (6.30a–d).

Returning to the general Markov chain on states $\{0, 1, \ldots, N\}$, when the chain is irreducible (all states communicate), then $P_{ij}(t) > 0$ for $i, j = 0, 1, \ldots, N$ and $\lim_{t\to\infty} P_{ij}(t) = \pi_j > 0$ exists independently of the initial state i. The limiting distribution may be found by passing to the limit in (6.66), noting that $\lim_{t\to\infty}\mathbf{P}'(t) = 0$. The resulting equations for $\boldsymbol{\pi} = (\pi_0, \pi_1, \ldots, \pi_N)$ are

$$0 = \boldsymbol{\pi}\mathbf{A} = (\pi_0, \pi_1, \ldots, \pi_N)\begin{Vmatrix}-q_0 & q_{01} & \cdots & q_{0N} \\ q_{10} & -q_1 & \cdots & q_{1N} \\ \vdots & \vdots & \ldots & \vdots \\ q_{N0} & q_{N1} & & -q_N\end{Vmatrix},$$

which is the same as

$$\pi_j q_j = \sum_{i\neq j}\pi_i q_{ij}, \quad j = 0, 1, \ldots, N. \tag{6.68}$$

Equation (6.68) together with

$$\pi_0 + \pi_1 + \cdots + \pi_N = 1 \tag{6.69}$$

determines the limiting distribution.

Equation (6.68) has a mass balance interpretation that aids us in understanding it. The left side $\pi_j q_j$ represents the long run rate at which particles executing the Markov process leave state j. This rate must equal the long run rate at which particles arrive at state j if equilibrium is to be maintained. Such arriving particles must come from some state $i \neq j$, and a particle moves from state $i \neq j$ to state j at rate q_{ij}. Therefore, the right side $\Sigma_{i \neq j} \pi_i q_{ij}$ represents the total rate of arriving particles.

Example *Industrial Mobility and the Peter Principle* Let us suppose that a draftsman position at a large engineering firm can be occupied by a worker at any of three levels: $T =$ Trainee, $J =$ Junior draftsman, and $S =$ Senior draftsman. Let $X(t)$ denote the level of the person in the position at time t, and suppose that $X(t)$ evolves as a Markov chain whose infinitesimal matrix is

$$\mathbf{A} = \begin{array}{c} \\ T \\ J \\ S \end{array} \begin{array}{c} \begin{array}{ccc} T & J & S \end{array} \\ \left\| \begin{array}{ccc} -a_T & a_T & 0 \\ a_{JT} & -a_J & a_{JS} \\ a_S & 0 & -a_S \end{array} \right\| \end{array}.$$

Thus, a Trainee stays at that rank for an exponentially distributed time having parameter a_T and then becomes a Junior draftsman. A Junior draftsman stays at that level for an exponentially distributed length of time having parameter $a_J = a_{JT} + a_{JS}$. Then, the Junior draftsman leaves the position and is replaced by a Trainee with probability a_{JT}/a_J or is promoted to a Senior draftsman with probability a_{JS}/a_J and so on.

Alternatively, we may describe the model by specifying the movements during short time intervals according to

$$\begin{aligned}
&\Pr\{X(t+h) = J|X(t) = T\} = a_T h + o(h),\\
&\Pr\{X(t+h) = T|X(t) = J\} = a_{JT} h + o(h),\\
&\Pr\{X(t+h) = S|X(t) = J\} = a_{JS} h + o(h),\\
&\Pr\{X(t+h) = T|X(t) = S\} = a_S h + o(h),
\end{aligned}$$

and

$$\Pr\{X(t+h) = i|X(t) = i\} = 1 - a_i h + o(h) \quad \text{for } i = T, J, S.$$

The equations for the equilibrium distribution (π_T, π_J, π_S) are, according to (6.68),

$$\begin{aligned}
a_T\pi_T &= \qquad\qquad a_{JT}\pi_J + a_S\pi_S,\\
a_J\pi_J &= a_T\pi_T,\\
a_S\pi_S &= \qquad\qquad a_{JS}\pi_J,\\
1 &= \pi_T + \qquad \pi_J + \quad \pi_S,
\end{aligned}$$

and the solution is

$$\pi_T = \frac{a_S a_J}{a_S a_J + a_S a_T + a_T a_{JS}},$$
$$\pi_J = \frac{a_S a_T}{a_S a_J + a_S a_T + a_T a_{JS}},$$
$$\pi_S = \frac{a_T a_{JS}}{a_S a_J + a_S a_T + a_T a_{JS}}.$$

Let us consider a numerical example for comparison with an alternative model to be developed later. We suppose that the mean times in the three states are

State	**Mean Time**
T	0.1
J	0.2
S	1.0

and that a Junior draftsman leaves and is replaced by a Trainee with probability $\frac{2}{5}$ and is promoted to a Senior draftsman with probability $\frac{3}{5}$. These suppositions lead to the prescription $a_T = 10, a_{JT} = 2, a_{JS} = 3$, and $a_S = 1$. The equilibrium probabilities are

$$\pi_T = \frac{1(5)}{1(5) + 1(10) + 10(3)} = \frac{5}{45} = 0.11,$$
$$\pi_J = \frac{10}{45} = 0.22,$$
$$\pi_S = \frac{30}{45} = 0.67.$$

But the duration that people spend in any given position is not exponentially distributed in general. A bimodal distribution is often observed in which many people leave rather quickly, while others persist for a substantial time. A possible explanation for this phenomenon is found in the "Peter Principle,"[2] which asserts that a worker is promoted until finally reaching a position in which he or she is incompetent. When this happens, the worker stays in that job until retirement. Let us modify the industrial mobility model to accommodate the Peter Principle by considering two types of Junior draftsmen, *Competent* and *Incompetent*. We suppose that a fraction p of Trainees are Competent and $q = 1 - p$ are Incompetent. We assume that a competent Junior draftsman stays at that level for an exponentially distributed duration with parameter a_C and then is promoted to Senior draftsman. Finally, an incompetent Junior draftsman stays in that position until retirement, an exponentially distributed sojourn with parameter

[2] Laurence, J.P., & Hull, R. (1969). The Peter Principle. Cutchogue, NY: Buccanear Books.

a_I, and then he or she is replaced by a Trainee. The relevant infinitesimal matrix is given by

$$\mathbf{A} = \begin{array}{c} \\ T \\ I \\ C \\ S \end{array} \begin{array}{c} \begin{array}{cccc} \quad T & \quad I & \quad C & \quad S \end{array} \\ \left\| \begin{array}{cccc} -a_T & qa_T & pa_T & \\ a_{\mathrm{I}} & -a_{\mathrm{I}} & & \\ & & -a_C & a_C \\ a_S & & & -a_S \end{array} \right\| \end{array}.$$

The duration in the Junior draftsman position now follows a probability law that is a mixture of exponential densities. To compare this model with the previous model, suppose that $p = \frac{3}{5}, q = \frac{2}{5}, a_I = 2.86$, and $a_C = 10$. These numbers were chosen so as to make the mean duration as a Junior draftsman,

$$p\left(\frac{1}{a_C}\right) + q\left(\frac{1}{a_I}\right) = \left(\frac{3}{5}\right)(0.10) + \left(\frac{2}{5}\right)(0.35) = 0.20,$$

the same as in the previous calculations. The probability density of this duration is

$$f(t) = \frac{3}{5}(10)\mathrm{e}^{-10t} + \frac{2}{5}(2.86)\mathrm{e}^{-2.86t} \quad \text{for } t \geq 0.$$

This density is plotted in Figure 6.7, for comparison with the exponential density $g(t) = 5\mathrm{e}^{-5t}$, which has the same mean. The bimodal tendency is indicated in that $f(t) > g(t)$ when t is near zero and when t is very large.

With the numbers as given and $a_T = 10$ and $a_S = 1$ as before, the stationary distribution $(\pi_T, \pi_I, \pi_C, \pi_S)$ is found by solving

$$\begin{aligned} 10\pi_T &= \qquad 2.86\pi_I, \qquad\qquad 1\pi_S, \\ 2.86\pi_I &= 4\pi_T, \\ 10\pi_C &= 6\pi_T, \\ 1\pi_S &= \qquad\qquad\qquad 10\pi_C, \\ 1 &= \pi_T + \qquad \pi_I + \ \ \pi_C + \pi_S. \end{aligned}$$

The solution is

$$\begin{aligned} \pi_T &= 0.111, \qquad & \pi_I &= 0.155, \\ \pi_S &= 0.667, \qquad & \pi_C &= 0.067. \end{aligned}$$

Let us make two observations before leaving this example. First, the limiting probabilities π_T, π_S, and $\pi_J = \pi_I + \pi_C$ agree between the two models. This is a common occurrence in stochastic modeling, wherein the limiting behavior of a process is rather insensitive to certain details of the model and depends only on the first moments, or means. When this happens, the model assumptions can be chosen for their mathematical convenience with no loss.

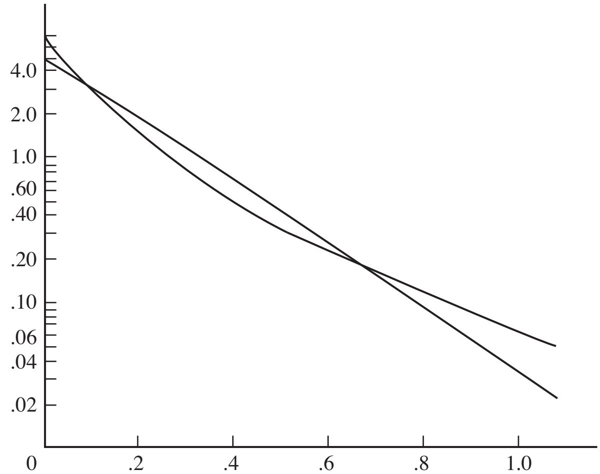

Figure 6.7 The exponential density (straight line) versus the mixed exponential density (curved line). Both distributions have the same mean. A logarithmic scale was used to accentuate the differences.

The second observation is specific to the Peter Principle. We have assumed that $p = \frac{3}{5}$ of Trainees are competent Junior draftsmen and only $q = \frac{2}{5}$ are Incompetent. Yet in the long run, a Junior draftsman is found to be Incompetent with probability $\pi_I/(\pi_I + \pi_C) = 0.155/(0.155 + 0.067) = 0.70!$

Example *Redundancy and the Burn-In Phenomenon* An airlines reservation system has two computers, one online and one backup. The operating computer fails after an exponentially distributed duration having parameter μ and is replaced by the standby. There is one repair facility, and repair times are exponentially distributed with parameter λ. Let $X(t)$ be the number of computers in operating condition at time t. Then, $X(t)$ is a Markov chain whose infinitesimal matrix is

$$\mathbf{A} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \end{array} \begin{array}{c} \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ \left\| \begin{array}{ccc} -\lambda & \lambda & 0 \\ \mu & -(\lambda+\mu) & \lambda \\ 0 & \mu & -\mu \end{array} \right\| \end{array}.$$

The stationary distribution (π_0, π_1, π_2) satisfies

$$\begin{aligned} \lambda\pi_0 &= \mu\pi_1, \\ (\lambda+\mu)\pi_1 &= \lambda\pi_0, \quad +\mu\pi_2, \\ \mu\pi_2 &= \lambda\pi_1, \\ 1 &= \pi_0 + \pi_1 + \pi_2, \end{aligned}$$

and the solution is

$$\pi_0 = \frac{1}{1+(\lambda/\mu)+(\lambda/\mu)^2},$$
$$\pi_1 = \frac{\lambda/\mu}{1+(\lambda/\mu)+(\lambda/\mu)^2},$$
$$\pi_2 = \frac{(\lambda/\mu)^2}{1+(\lambda/\mu)+(\lambda/\mu)^2}.$$

The availability, or probability that at least one computer is operating, is $1-\pi_0 = \pi_1+\pi_2$.

Often, in practice, the assumption of exponentially distributed operating times is not realistic because of the so-called *burn-in phenomenon*. This idea is best explained in terms of the *hazard* rate $r(t)$ associated with a probability density function $f(t)$ of a nonnegative failure time T. Recall that $r(t)\Delta t$ measures the conditional probability that the item fails in the next time interval $(t, t+\Delta t)$ given that it has survived up to time t, and therefore, we have

$$r(t) = \frac{f(t)}{1-F(t)} \quad \text{for } t \geq 0,$$

where $F(t)$ is the cumulative distribution function associated with the probability density function $f(t)$.

A constant hazard rate $r(t) = \lambda$ for all t corresponds to the exponential density function $f(t) = \lambda e^{-\lambda t}$ for $t \geq 0$. The burn-in phenomenon is described by a hazard rate that is initially high and then decays to a constant level, where it persists, possibly later to rise again (aging). It corresponds to a situation in which a newly manufactured or newly repaired item has a significant probability of failing early in its use. If the item survives this test period, however, it then operates in an exponential or memoryless manner. The early failures might correspond to incorrect manufacture or faulty repair, or might be a property of the materials used.

Anyone familiar with automobile repairs has experienced the burn-in phenomenon.

One of many possible ways to model the burn-in phenomenon is to use a mixture of exponential densities

$$f(t) = p\alpha e^{-\alpha t} + q\beta e^{-\beta t}, \quad t \geq 0, \tag{6.70}$$

where $0 < p = 1-q < 1$ and α, β are positive. The density function for which $p = 0.1, \alpha = 10, q = 0.9$, and $\beta = 0.909\cdots = 1/1.1$ has mean one. Its hazard rate is plotted in Figure 6.8, where the higher initial burn-in level is evident.

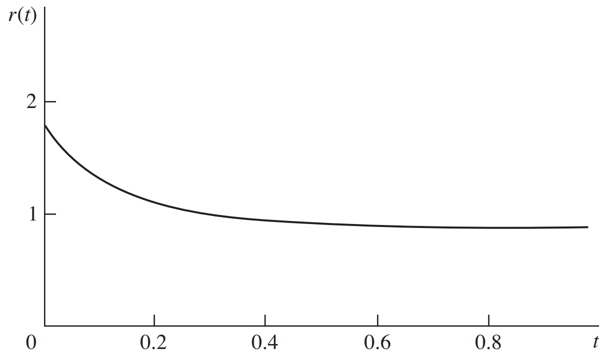

Figure 6.8 The hazard rate corresponding to the density given in (6.70). The higher hazard rate at the initial t values represents the burn-in phenomenon.

We may incorporate the burn-in phenomenon corresponding to the mixed exponential density (6.70) by expanding the state space according to the following table:

Notation	State
0	Both computers down
1_A	One operating computer, current up time has parameter α
1_B	One operating computer, current up time has parameter β
2_A	Two operating computers, current up time has parameter α
2_B	Two operating computers, current up time has parameter β

Equation (6.70) corresponds to a probability p that a computer beginning operation will have an exponentially distributed up time with parameter α and a probability q that the parameter is β. Accordingly, we have the infinitesimal matrix

$$\mathbf{A} = \begin{array}{c} \\ 0 \\ 1_A \\ 1_B \\ 2_A \\ 2_B \end{array} \begin{array}{c} \begin{array}{ccccc} 0 & 1_A & 1_B & 2_A & 2_B \end{array} \\ \left\| \begin{array}{ccccc} -\lambda & p\lambda & q\lambda & & \\ \alpha & -(\lambda+\alpha) & & \lambda & \\ \beta & & -(\lambda+\beta) & & \lambda \\ & p\alpha & q\alpha & -\alpha & \\ & p\beta & q\beta & & -\beta \end{array} \right\| \end{array}.$$

The stationary distribution can be determined in the usual way by applying (6.68).

Exercises

6.6.1 A certain type component has two states: $0 = \text{OFF}$ and $1 = \text{OPERATING}$. In state 0, the process remains there a random length of time, which is exponentially distributed with parameter α, and then moves to state 1. The time in state 1 is exponentially distributed with parameter β, after which the process returns to state 0.

The *system* has two of these components, A and B, with distinct parameters:

Component	Operating Failure Rate	Repair Rate
A	β_A	α_A
B	β_B	α_B

In order for the *system* to operate, at least one of components A and B must be operating (a parallel system). Assume that the component stochastic processes are independent of one another. Determine the long run probability that the system is operating by

(a) Considering each component separately as a two-state Markov chain and using their statistical independence;

(b) Considering the system as a four-state Markov chain and solving equations (6.68).

6.6.2 Let $X_1(t)$ and $X_2(t)$ be independent two-state Markov chains having the same infinitesimal matrix

$$\mathbf{A} = \begin{array}{c} \\ 0 \\ 1 \end{array} \begin{array}{c} \begin{array}{cc} 0 & 1 \end{array} \\ \left\| \begin{array}{cc} -\lambda & \lambda \\ \mu & -\mu \end{array} \right\| \end{array}.$$

Argue that $Z(t) = X_1(t) + X_2(t)$ is a Markov chain on the state space $S = \{0, 1, 2\}$ and determine the transition probability matrix $\mathbf{P}(t)$ for $Z(t)$.

Problems

6.6.1 Let $Y_n, n = 0, 1, \ldots,$ be a discrete time Markov chain with transition probabilities $\mathbf{P} = \|P_{ij}\|$, and let $\{N(t); t \geq 0\}$ be an independent Poisson process of rate λ. Argue that the compound process

$$X(t) = Y_{N(t)}, \quad t \geq 0,$$

is a Markov chain in continuous time and determine its infinitesimal parameters.

6.6.2 A certain type component has two states: $0 =$ OFF and $1 =$ OPERATING. In state 0, the process remains there a random length of time, which is exponentially distributed with parameter α, and then moves to state 1. The time in state 1 is exponentially distributed with parameter β, after which the process returns to state 0.

The system has three of these components, A, B, and C, with distinct parameters:

Component	Operating Failure Rate	Repair Rate
A	β_A	α_A
B	β_B	α_B
C	β_C	α_C

In order for the *system* to operate, component A must be operating, and at least one of components B and C must be operating. In the long run, what fraction of time does the system operate? Assume that the component stochastic processes are independent of one another.

6.6.3 Let $X_1(t), X_2(t), \ldots, X_N(t)$ be independent two-state Markov chains having the same infinitesimal matrix

$$\mathbf{A} = \begin{array}{c} \\ 0 \\ 1 \end{array} \begin{array}{c} \begin{array}{cc} 0 & \quad 1 \end{array} \\ \left\| \begin{array}{cc} -\lambda & \lambda \\ \mu & -\mu \end{array} \right\| \end{array}.$$

Determine the infinitesimal matrix for the Markov chain $Z(t) = X_1(t) + \cdots + X_N(t)$.

6.6.4 A system consists of two units, both of which may operate simultaneously, and a single repair facility. The probability that an operating system will fail in a short time interval of length Δt is $\mu(\Delta t) + o(\Delta t)$. Repair times are exponentially distributed, but the parameter depends on whether the failure was *regular* or *severe*. The fraction of regular failures is p, and the corresponding exponential parameter is α. The fraction of severe failure is $q = 1 - p$, and the exponential parameter is $\beta < \alpha$.

Model the system as a continuous time Markov chain by taking as states the pairs (x, y), where $x = 0, 1, 2$ is the number of units operating and $y = 0, 1, 2$ is the number of units undergoing repair for a severe failure. The possible states are $(2, 0), (1, 0), (1, 1), (0, 0), (0, 1)$, and $(0, 2)$. Specify the infinitesimal matrix $\mathbf{A}$. Assume that the units enter the repair shop on a first come, first served basis.

6.7 A Poisson Process with a Markov Intensity[3]

Consider "points" scattered in some manner along the semi-infinite interval $[0, \infty)$, and for an interval of the form $I = (a, b]$, with $0 \leq a < b < \infty$, let $N(I)$ count the number of "points" in the interval I. Then $N(I)$, as I ranges over the half-open intervals $I = (a, b]$, is a point process. (See Chapter 5, Section 5.5 for generalizations to higher dimensions.) Suppose that, conditional on a given intensity, $N(I)$ is a nonhomogeneous Poisson process, but where the intensity function $\{\lambda(t), t \geq 0\}$ is itself a stochastic process. Such point processes were introduced in Chapter 5, Section 5.1.4, where they were called *Cox processes* in honor of their discoverer. While Cox processes are sufficiently general to describe a plethora of phenomena, they remain simple enough to permit explicit calculation, at least in some instances. As an illustration, we will derive the probability of no points in an interval for a Cox process in which the intensity function is a two-state Markov chain in continuous time. The Cox process alternates between being "ON" and "OFF." When the underlying intensity is "ON," points occur according to a Poisson process of constant intensity λ. When the

[3] Starred sections contain material of a more specialized or advanced nature.

underlying process is “OFF,” no points occur. We will call this basic model a $(0, \lambda)$ *Cox process* to distinguish it from a later generalization. The $(0, \lambda)$ Cox process might describe bursts of rainfall in a locale that alternates between dry spells and wet ones, or arrivals to a queue from a supplier that randomly shuts down. Rather straightforward extensions of the techniques that we will now use in this simple case can be adapted to cover more complex models and computations, as we will subsequently show.

We assume that the intensity process $\{\lambda(t); t \geq 0\}$ is a two-state Markov chain in continuous time for which

$$\Pr\{\lambda(t+h) = \lambda | \lambda(t) = 0\} = \alpha h + o(h), \tag{6.71}$$

and

$$\Pr\{\lambda(t+h) = 0 | \lambda(t) = \lambda\} = \beta h + o(h). \tag{6.72}$$

This intensity is merely the constant λ times the two-state birth and death process of Section 6.3. As may be seen by allowing $t \to \infty$ in (6.29), such a process has the limiting distribution $\Pr\{\lambda(\infty) = 0\} = \beta/(\alpha + \beta)$ and $\Pr\{\lambda(\infty) = \lambda\} = \alpha/(\alpha + \beta)$. We will assume that the intensity process begins with this limiting distribution, or explicitly, that $\Pr\{\lambda(0) = 0\} = \beta/(\alpha + \beta)$ and $\Pr\{\lambda(0) = \lambda\} = \alpha/(\alpha + \beta)$. With this assumption, the intensity process is stationary in the sense that $\Pr\{\lambda(t) = 0\} = \beta/(\alpha + \beta)$ and $\Pr\{\lambda(t) = \lambda\} = \alpha/(\alpha + \beta)$ for all $t \geq 0$. This stationarity carries over to the Cox process $N(I)$ to imply that $\Pr\{N((0, t]) = k\} = \Pr\{N((s, s+t]) = k\}$ for all $s, t \geq 0$ and $k = 0, 1, \ldots$. We are interested in determining

$$f(t; \lambda) = \Pr\{N((0, t]) = 0\}.$$

Let

$$\Lambda(t) = \int_0^t \lambda(s)\mathrm{d}s \tag{6.73}$$

and note the conditional Poisson probability

$$\Pr\{N((0, t]) = 0 | \lambda(s) \quad \text{for } s \leq t\} = \mathrm{e}^{-\Lambda(t)}$$

so that upon removing the conditioning via the law of total probability, we obtain

$$f(t; \lambda) = E\left[\mathrm{e}^{-\Lambda(t)}\right] = f_0(t) + f_1(t), \tag{6.74}$$

where

$$f_0(t) = \Pr\{N((0, t]) = 0 \quad \text{and } \lambda(t) = 0\}, \tag{6.75}$$

and

$$f_1(t) = \Pr\{N((0,t]) = 0 \quad \text{and } \lambda(t) = \lambda\}. \tag{6.76}$$

Using an infinitesimal "last step analysis" similar to that used to derive the Kolmogorov forward equations, we will derive a pair of first-order linear differential equations for $f_0(t)$ and $f_1(t)$. To this end, by analyzing the possibilities at time t and using the law of total probability, we begin with

$$\begin{aligned} f_0(t+h) &= f_0(t)\Pr\{N((t,t+h]) = 0|\lambda(t) = 0\}\Pr\{\lambda(t+h) = 0|\lambda(t) = 0\} \\ &\quad + f_1(t)\Pr\{N((t,t+h]) = 0|\lambda(t) = \lambda\}\Pr\{\lambda(t+h) = 0|\lambda(t) = \lambda\} \\ &= f_0(t)[1-\alpha h + o(h)] + f_1(t)\mathrm{e}^{-\lambda h}\beta h + o(h), \end{aligned}$$

and

$$\begin{aligned} f_1(t+h) &= f_1(t)\Pr\{N((t,t+h]) = 0|\lambda(t) = \lambda\}\Pr\{\lambda(t+h) = \lambda|\lambda(t) = \lambda\} \\ &\quad + f_0(t)\Pr\{N((t,t+h]) = 0|\lambda(t) = 0\}\Pr\{\lambda(t+h) = \lambda|\lambda(t) = 0\} \\ &= f_1(t)\mathrm{e}^{-\lambda h}[1-\beta h + o(h)] + f_0(t)\alpha h + o(h). \end{aligned}$$

We rearrange the terms and use $\mathrm{e}^{-\lambda h} = 1 - \lambda h + o(h)$ to get

$$f_0(t+h) - f_0(t) = -\alpha f_0(t)h + \beta f_1(t)h + o(h)$$

and

$$f_1(t+h) - f_1(t) = -(\beta+\lambda)f_1(t)h + \alpha f_0(t)h + o(h),$$

which, after dividing by h and letting h tend to zero, become the differential equations

$$\frac{\mathrm{d}f_0(t)}{\mathrm{d}t} = -\alpha f_0(t) + \beta f_1(t) \tag{6.77}$$

and

$$\frac{\mathrm{d}f_1(t)}{\mathrm{d}t} = -(\beta+\lambda)f_1(t) + \alpha f_0(t). \tag{6.78}$$

The initial conditions are

$$f_0(0) = \Pr\{\lambda(0) = 0\} = \beta/(\alpha+\beta) \tag{6.79}$$

and

$$f_1(0) = \Pr\{\lambda(0) = \lambda\} = \alpha/(\alpha+\beta). \tag{6.80}$$

Such coupled first-order linear differential equations are readily solved. In our case, after carrying out the solution and simplifying the result, the answer is $\Pr\{N((0,t]) = 0\} = f_0(t) + f_1(t) = f(t;\lambda)$, where

$$f(t;\lambda) = c_+ \exp\{-\mu_+ t\} + c_- \exp\{-\mu_- t\} \tag{6.81}$$

with

$$\mu_\pm = \frac{1}{2}\left\{(\lambda+\alpha+\beta) \pm \sqrt{(\lambda+\alpha+\beta)^2 - 4\alpha\lambda}\right\}, \tag{6.82}$$

$$c_+ = \frac{[\alpha\lambda/(\alpha+\beta)] - \mu_-}{\mu_+ - \mu_-} \tag{6.83}$$

and

$$c_- = \frac{\mu_+ - [\alpha\lambda/(\alpha+\beta)]}{\mu_+ - \mu_-}. \tag{6.84}$$

A Generalization Let N be a Cox process driven by a 0–1 Markov chain $\lambda(t)$, but now suppose that when the intensity process is in state 0, the Cox process is, conditionally, a Poisson process of rate λ_0, and when the intensity process is in state 1, then the Cox process is a Poisson process of rate λ_1. The earlier Cox process had $\lambda_0 = 0$ and $\lambda_1 = \lambda$. Without loss of generality, we assume $0 < \lambda_0 < \lambda_1$.

In order to evaluate $\Pr\{N((0,t]) = 0\}$, we write N as the sum $N = N_1 + N_2$ of two independent processes, where N_1 is a Poisson process of constant rate λ_0 and N_2 is a $(0,\lambda)$ Cox process with $\lambda = \lambda_1 - \lambda_0$. Then, N is zero if and only if both N_1 and N_2 are zero, whence

$$\begin{aligned}\Pr\{N((0,t]) = 0\} &= \Pr\{N_1((0,t]) = 0\} \cdot \Pr\{N_2((0,t]) = 0\} \\ &= e^{-\lambda_0 t} f(t, \lambda_1 - \lambda_0).\end{aligned} \tag{6.85}$$

Example The tensile strength $S(t)$ of a single fiber of length t is often assumed to follow a Weibull distribution of the form

$$\Pr\{S(t) > x\} = \exp\left\{-t\sigma x^\delta\right\}, \quad \text{for } x > 0, \tag{6.86}$$

where δ and σ are positive material constants. The explicit appearance of the length t in the exponent is an expression of a weakest-link size effect, in which the fiber strength is viewed as the minimum strength of independent sections. This theory suggests that the survivorship probability of strength for a fiber of length t should satisfy the relation

$$\Pr\{S(t) > x\} = [\Pr\{S(1) > x\}]^t, \quad t > 0. \tag{6.87}$$

The Weibull distribution is the only type of distribution that is concentrated on $0 \le x < \infty$ and satisfies (6.87).

However, a fiber under stress may fail from a surface flaw such as a notch or scratch, or from an internal flaw such as a void or inclusion. Where the diameter d of the fiber varies along its length, the relative magnitude of these two types of flaws will also vary, since the surface of the fiber is proportional to d, while the volume is proportional to d^2. As a simple generalization, suppose that the two types of flaws alternate and that the changes from one flaw type to the other follow a two-state Markov chain along the continuous length of the fiber. Further, suppose that a fiber of constant type i flaw, for $i = 0, 1$, will support the load x with probability

$$\Pr\{S(t) > x\} = \exp\left\{-t\sigma_i x^{\delta_i}\right\}, \quad x > 0,$$

where σ_i and δ_i are positive constants.

We can evaluate the survivorship probability for the fiber having Markov varying flaw types by bringing in an appropriate (λ_0, λ_1) Cox process. For a fixed $x > 0$, suppose that flaws that are weaker than x will occur along a fiber of constant flaw type i according to a Poisson process of rate $\lambda_i(x) = \sigma_i x^{\delta_i}$, for $i = 0, 1$. A fiber of length t and having Markov varying flaw types will carry a load of x if and only if there are no flaws weaker than x along the fiber. Accordingly, for the random flaw type fiber, using (6.85), we have

$$\Pr\{S(t) > x\} = e^{-\lambda_0(x)t} f(t; \lambda_1(x) - \lambda_0(x)). \tag{6.88}$$

Equation (6.88) may be evaluated numerically under a variety of assumptions for comparison with observed fiber tensile strengths. Where fibers having two flaw types are tested at several lengths, (6.88) may be used to extrapolate and predict strengths at lengths not measured.

It is sometimes more meaningful to reparametrize according to $\pi = \alpha/(\alpha + \beta)$ and $\tau = \alpha + \beta$. Here, π is the long run fraction of fiber length for which the applicable flaw distribution is of type 1, and $1 - \pi$ is the similar fraction of type 0 flaw behavior. The second parameter τ is a measure of the rapidity with which the flaw types alternate. In particular, when $\tau = 0$, the diameter or flaw type remains in whichever state it began, and the survivor probability reduces to the mixture

$$\Pr\{S(t) > x\} = \pi e^{-\lambda_1(x)t} + (1 - \pi)e^{-\lambda_0(x)t}. \tag{6.89}$$

On the other hand, at $\tau = \infty$, the flaw type process alternates instantly, and the survivor probability simplifies to

$$\Pr\{S(t) > x\} = \exp\{-t[\pi\lambda_1(x) + (1 - \pi)\lambda_0(x)]\}. \tag{6.90}$$

The probability distribution for $N((0, t])$ Let $\Lambda(t)$ be the cumulative intensity for a Cox process and suppose that we have evaluated

$$g(t; \theta) = E\left[e^{-(1-\theta)\Lambda(t)}\right], \quad 0 < \theta < 1.$$

For a $(0,\lambda)$ Cox process, for instance, $g(t,\theta) = f(t;(1-\theta)\lambda)$, where f is defined in (6.81). Upon expanding as a power series in θ, according to

$$\begin{aligned} g(t;\theta) &= E\left[\mathrm{e}^{-\Lambda(t)}\sum_{k=0}^{\infty}\frac{\Lambda(t)^k}{k!}\theta^k\right] \\ &= \sum_{k=0}^{\infty}E\left[\mathrm{e}^{-\Lambda(t)}\frac{\Lambda(t)^k}{k!}\right]\theta^k \\ &= \sum_{k=0}^{\infty}\Pr\{N((0,t]) = k\}\theta^k, \end{aligned}$$

we see that the coefficient of θ^k in the power series is $\Pr\{N((0,t]) = k\}$. In principle then, the probability distribution for the points in an interval in a Cox process can be determined in any particular instance.

Exercises

6.7.1 Suppose that a $(0,\lambda)$ Cox process has $\alpha = \beta = 1$ and $\lambda = 2$. Show that $\mu_\pm = 2 \pm \sqrt{2}$ and $c_- = \frac{1}{4}(2+\sqrt{2}) = 1 - c_+$, whence

$$\Pr\{N((0,t]) = 0\} = \mathrm{e}^{-2t}\left[\cosh(\sqrt{2}t) + \frac{\sqrt{2}}{2}\sinh(\sqrt{2}t)\right].$$

6.7.2 Suppose that a $(0,\lambda)$ Cox process has $\alpha = \beta = 1$ and $\lambda = 2$. Show that

$$f_0(t) = \frac{1+\sqrt{2}}{4}\mathrm{e}^{-(2-\sqrt{2})t} + \frac{1-\sqrt{2}}{4}\mathrm{e}^{-(2+\sqrt{2})t}$$

and

$$f_1(t) = \frac{1}{4}\mathrm{e}^{-(2-\sqrt{2})t} + \frac{1}{4}\mathrm{e}^{-(2+\sqrt{2})t}$$

satisfy the differential equations (6.77) and (6.78) with the initial conditions (6.79) and (6.80).

Problems

6.7.1 Consider a stationary Cox process driven by a two-state Markov chain. Let $\pi = \alpha/(\alpha+\beta)$ be the probability that the process begins in state λ.

- **(a)** By using the transition probabilities given in (6.30a–d), show that $\Pr\{\lambda(t) = \lambda\} = \pi$ for all $t > 0$.
- **(b)** Show that $E[N((0,t])] = \pi\lambda t$ for all $t > 0$.

6.7.2 The *excess life* $\gamma(t)$ in a point process is the random length of the duration from time t until the next event. Show that the cumulative distribution function for the excess life in a Cox process is given by $\Pr\{\gamma(t) \leq x\} = 1 - \Pr\{N((t, t+x]) = 0\}$.

6.7.3 Let T be the time to the first event in a stationary $(0, \lambda)$ Cox process. Find the probability density function $\phi(t)$ for T. Show that when $\alpha = \beta = 1$ and $\lambda = 2$, this density function simplifies to $\phi(t) = \exp\{-2t\}\cosh(\sqrt{2}t)$.

6.7.4 Let T be the time to the first event in a stationary $(0, \lambda)$ Cox process. Find the expected value $E[T]$. Show that $E[T] = \frac{3}{2}$ when $\alpha = \beta = 1$ and $\lambda = 2$. What is the average duration between events in this process?

6.7.5 Determine the conditional probability of no points in the interval $(t, t+s]$, given that there are no points in the interval $(0, t]$ for a stationary Cox process driven by a two-state Markov chain. Establish the limit

$$\lim_{t\to\infty} \Pr\{N((t, t+s]) = 0 | N((0, t]) = 0\} = \mathrm{e}^{-\mu - s}, \quad s > 0.$$

6.7.6 Show that the Laplace transform

$$\phi(s; \lambda) = \int_0^\infty \mathrm{e}^{-st} f(t; \lambda) \mathrm{d}t$$

is given by

$$\phi(s; \lambda) = \frac{s + (1-\pi)\lambda + \tau}{s^2 + (\tau + \lambda)s + \pi\tau\lambda},$$

where $\tau = \alpha + \beta$ and $\pi = \alpha/(\alpha + \beta)$. Evaluate the limit (a) as $\tau \to \infty$, and (b) as $\tau \to 0$.

6.7.7 Consider a $(0, \lambda)$ stationary Cox process with $\alpha = \beta = 1$ and $\lambda = 2$. Show that $g(t; \theta) = f(t; (1-\theta)\lambda)$ is given by

$$g(t; \theta) = \mathrm{e}^{-(2-\theta)t}\left\{\cosh(Rt) + \frac{1}{R}\sinh(Rt)\right\},$$

where

$$R = \sqrt{\theta^2 - 2\theta + 2}.$$

Use this to evaluate $\Pr\{N((0, 1]) = 1\}$.

6.7.8 Consider a stationary $(0, \lambda)$ Cox process. A long duration during which no events were observed would suggest that the intensity process is in state 0. Show that

$$\Pr\{\lambda(t) = 0 | N((0, t]) = 0\} = \frac{f_0(t)}{f(t)},$$

where $f_0(t)$ is defined in (6.75).

6.7.9 Show that

$$f_0(t) = a_+ \mathrm{e}^{-\mu_+ t} + a_- \mathrm{e}^{-\mu_- t}$$

and

$$f_1(t) = b_+ \mathrm{e}^{-\mu_+ t} + b_- \mathrm{e}^{-\mu_- t}$$

with

$$a_\pm = \frac{1}{2}(1-\pi)\left[1 \mp \frac{(\alpha+\beta+\lambda)}{R}\right], R = \sqrt{(\alpha+\beta+\lambda)^2 - 4\alpha\lambda}$$

$$b_\pm = \frac{1}{2}\pi\left[1 \pm \frac{(\lambda-\alpha-\beta)}{R}\right]$$

satisfy the differential equations (6.77) and (6.78) subject to the initial conditions (6.79) and (6.80).

6.7.10 Consider a stationary $(0, \lambda)$ Cox process.

(a) Show that

$$\Pr\{N((0,h]) > 0, N((h,h+t]) = 0\} = f(t;\lambda) - f(t+h;\lambda),$$

whence

$$\Pr\{N((h,h+t]) = 0 | N((0,h]) > 0\} = \frac{f(t;\lambda) - f(t+h;\lambda)}{1 - f(h;\lambda)}.$$

(b) Establish the limit

$$\lim_{h\to 0} \Pr\{N((h,h+t]) = 0 | N((0,h]) > 0\} = \frac{f'(t;\lambda)}{f'(0;\lambda)},$$

where

$$f'(t;\lambda) = \frac{\mathrm{d}f(t;\lambda)}{\mathrm{d}t}.$$

(c) We interpret the limit in (b) as the conditional probability

$$\Pr\{N((0,t]) = 0 | \text{Event occurs at time } 0\}.$$

Show that

$$\Pr\{N((0,t]) = 0 | \text{Event at time } 0\} = p_+ \mathrm{e}^{-\mu_+ t} + p_- \mathrm{e}^{-\mu_- t},$$

where

$$p_+ = \frac{c_+\mu_+}{c_+\mu_+ + c_-\mu_-}, \quad p_- = \frac{c_-\mu_-}{c_+\mu_+ + c_-\mu_-}.$$

(d) Let τ be the time to the first event in $(0, \infty)$ in a stationary $(0, \lambda)$ Cox process with $\alpha = \beta = 1$ and $\lambda = 2$. Show that

$$E[\tau|\text{Event at time } 0] = 1.$$

Why does this differ from the result in Problem 6.7.4?

6.7.11 *A Stop-and-Go Traveler* The velocity $V(t)$ of a stop-and-go traveler is described by a two-state Markov chain. The successive durations in which the traveler is stopped are independent and exponentially distributed with parameter α, and they alternate with independent exponentially distributed sojourns, parameter β, during which the traveler moves at unit speed. Take the stationary case in which $\Pr\{V(0) = 1\} = \pi = \alpha/(\alpha + \beta)$. The distance traveled in time t is the integral of the velocity:

$$S(t) = \int_0^t V(u)\mathrm{d}u.$$

Show that

$$E\left[\mathrm{e}^{-\theta S(t)}\right] = f(t; \theta), \quad \theta \text{ real}.$$

(This is the Laplace transform of the probability density function of $S(t)$.)

6.7.12 Let τ be the time of the first event in a $(0, \lambda)$ Cox process. Let the 0 and λ states represent "OFF" and "ON," respectively.

(a) Show that the total duration in the $(0, \tau]$ interval that the system is ON is exponentially distributed with parameter λ and does not depend on α, β, or the starting state.

(b) Assume that the process begins in the OFF state. Show that the total duration in the $(0, \tau]$ interval that the system is OFF has the same distribution as

$$\sum_{k=0}^{N(\vartheta)} \eta_k,$$

where ζ is exponentially distributed with parameter $\lambda, N(t)$ is a Poisson process with parameter β, and $\eta_0, \eta_1, \ldots$ are independent and exponentially distributed with parameter α.