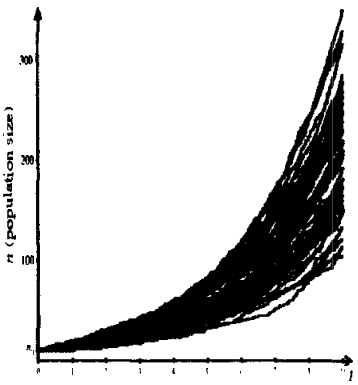

Chapter 5
Stochastic Models

5.1 Introduction

In this chapter, we consider models where outcomes are uncertain. Even though outcomes are uncertain, we can still assign probabilities to different outcomes, and then study how these probabilities change with time. An excellent reference text in this area is [3].

Many biological systems change from one state to another over time. Nerves change from quiescent to excited and back, cells change from healthy to diseased, or one plant population replaces another. While changes between states can be uncertain, probabilities of transitioning from one state to the next can be assigned nonetheless. If we know the transition probabilities between states, then we can evaluate changes in the system over time.

In preparation for the mathematical formulation of a model with uncertain outcomes, we first define the mathematical quantities that will go into the model. If S is a *sample space* (collection of all possible outcomes of an "experiment") with a probability measure, and X is a real-valued function defined over the elements of S, then X is a *random variable*. For example, if X were the fork length of a captured fish in cm, then S would be the nonnegative real numbers.

We can follow the change in a random variable as a parameter, such as time, increases. A family of random variables $\{X(t)\}$, indexed by a parameter t, is called a *stochastic process.*

We start this chapter with an example of a memoryless stochastic process, or *Markov process* (Section 5.2). We model ecosystem succession dynamics via a Markov process. Here a *Markov chain* model, describing transitions from one state to the next, can be understood with matrix theory. Next we focus on random variables (Section 5.3). We introduce probability density (Section 5.3.1) and probability mass (Section 5.3.2) as measures for sample space S, and discuss descriptive statistics (Section 5.3.3) and probability generating functions (Section 5.3.4) as means to characterize random variables. The last part of the chapter concerns applications and extensions of tools developed earlier in the chapter. We consider random motion via diffusion processes (Section 5.4), branching processes (Section 5.5), linear birth and death processes (Section 5.6), and nonlinear birth and death processes (Section 5.7). These can be used to describe animal movement (Section 5.4.2), the

extinction of family names (Section 5.3.4), the polymerase chain reaction (Section 5.5.2), population extinction (Section 5.6.2), and dynamics of the common cold (Section 5.7.1).

5.2 Markov Chains

The simplest stochastic processes are those which can be completely characterized by their current state, and where past states of a variable do not affect future outcomes. A stochastic process $\{X(t)\}$ is called a *Markov process* if it is history-independent. In the case with t being a discrete sequence $t_1, t_2, \ldots$, it is a one-step memory process, that is,

$$\begin{aligned} \Pr\{X(t_i) = x_1 | X(t_{i-1}) = x_2 \cap X(t_{i-2}) = x_3 \cap \cdots\} \\ = \Pr\{X(t_i) = x_1 | X(t_{i-1}) = x_2\}. \end{aligned} \tag{5.1}$$

Here, Pr denotes the probability associated with an event, $\cap$ means "and," and | means "given that." Markov processes are sometimes referred to as being *memoryless*; that is, the next state for the stochastic process depends only upon the current state. A *Markov chain* is a model which tracks the progression of a Markov process from time step to time step.

One example of a Markov chain involves succession in plant communities. As plant communities mature to a climax ecosystem, certain plant species replace others. In this section, we study succession in plant communities with Markov chains. We begin with a two-tree forest in Section 5.2.1. We generalize the example to formulate a Markov theory in Section 5.2.2. A large forest is considered in Section 5.2.3.

5.2.1 A Two-Tree Forest Ecosystem

By way of example, consider a population comprising red oak and hickory. At any point in space, the sample space of possible outcomes is $S = \{\text{RO}, \text{HI}\}$, where RO represents red oak and HI represents hickory. We assume that the life spans of the two trees are similar. In each generation, red oak may be replaced by itself or by hickory, and hickory may be replaced by itself or red oak. This is a Markov process, with the index t indicating the generation.

For example, suppose that when a red oak tree dies, it is equally likely to be replaced by hickory or red oak, and that when a hickory tree dies, it has probability 0.74 of being replaced by red oak and 0.26 of being replaced by hickory. These transitions can be shown in either a graphical (Figure 5.1) or a tabular (Table 5.1) format. Note that the columns of the table sum to 1.

The table format can be translated into a transition matrix. For example, the entries of Table 5.1 can be written in a transition matrix P,

$$P = \begin{pmatrix} 0.5 & 0.74 \\ 0.5 & 0.26 \end{pmatrix}.$$

To track the changes in the system over time, we define a vector $\mathbf{u}_t = (o_t, h_t)^T$ which describes the probability of red oak and of hickory at a given location in the forest after t generations. In the case of a large homogeneous forest, the same transition model would

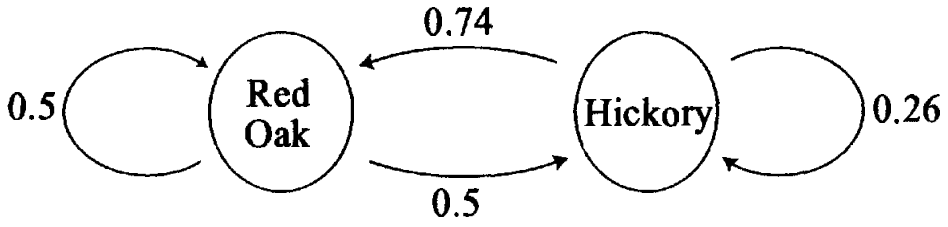

Figure 5.1. *Transitions between red oak and hickory vegetation shown as a graph.*

Table 5.1. *Transitions between red oak and hickory vegetation shown as a table.*

	Canopy	
% Saplings	Red oak	Hickory
Red oak	0.50	0.74
Hickory	0.50	0.26

apply at every point in space. Hence, o_t and h_t can be interpreted as the *proportions* of red oak and hickory in a large statistically homogeneous forest ecosystem.

If we suppose that the forest is initially 50% red oak and 50% hickory, then $\mathbf{u}_0 = (0.5, 0.5)^T$. To find $\mathbf{u}_1$, we calculate as follows:

$$\begin{aligned} o_1 &= \begin{matrix}\text{proportion}\\ \text{of red oak at}\\ \text{time 0}\end{matrix} \cdot \begin{matrix}\text{probability}\\ \text{red oak is replaced}\\ \text{by red oak}\end{matrix} + \begin{matrix}\text{proportion}\\ \text{of hickory}\\ \text{at time 0}\end{matrix} \cdot \begin{matrix}\text{probability}\\ \text{hickory is}\\ \text{replaced by red oak}\end{matrix} \\ &= (0.5) \cdot (0.5) + (0.5) \cdot (0.74) \\ &= 0.62 \end{aligned}$$

$$\begin{aligned} h_1 &= \begin{matrix}\text{proportion}\\ \text{of red oak at}\\ \text{time 0}\end{matrix} \cdot \begin{matrix}\text{probability}\\ \text{red oak is replaced}\\ \text{by hickory}\end{matrix} + \begin{matrix}\text{proportion}\\ \text{of hickory}\\ \text{at time 0}\end{matrix} \cdot \begin{matrix}\text{probability}\\ \text{hickory is}\\ \text{replaced by hickory}\end{matrix} \\ &= (0.5) \cdot (0.5) + (0.5) \cdot (0.26) \\ &= 0.38. \end{aligned}$$

In terms of the matrix formulation, we can write

$$\begin{aligned} \mathbf{u}_1 &= P\mathbf{u}_0, \\ \mathbf{u}_2 &= P\mathbf{u}_1, \\ &\vdots \\ &\text{etc.} \end{aligned}$$

If we continue this process, we observe that the forest approaches an equilibrium value. For example,

$$\mathbf{u}_4 = \begin{pmatrix} 0.596 \\ 0.404 \end{pmatrix}, \qquad \mathbf{u}_5 = \begin{pmatrix} 0.597 \\ 0.403 \end{pmatrix}.$$

The forest has reached an equilibrium $\mathbf{u}^*$ when

$$P\mathbf{u}^* = \mathbf{u}^*.$$

Here, $\mathbf{u}^*$ is an *eigenvector* corresponding to eigenvalue $\lambda = 1$. To calculate the eigenvector we solve

$$(P - I)\mathbf{u}^* = 0 \Rightarrow \begin{pmatrix} -0.5 & 0.74 \\ 0.5 & -0.74 \end{pmatrix} \begin{pmatrix} o^* \\ h^* \end{pmatrix} = 0,$$

which has solution $o^* = 0.597$, $h^* = 0.403$.

A more complex model is given below in Section 5.2.3. However, before considering this complex model, we derive the general theory.

5.2.2 Markov Chain Theory

We consider a system with n possible states for the system. Given that a transition occurs from state j, the *transition probability* p_{ij} describes the probability of the transition taking the system to state i, $1 \leq i, j, \leq n$. When the transition probabilities are entered into a *transition matrix* $P = (p_{ij})$, the matrix columns sum to 1 because a transition occurring from state j takes the system to some state i, $1 \leq i \leq n$, with probability 1. Finally, to track the probability associated with being in each state we define a *probability vector*, a vector $\mathbf{u} = (u_1, \ldots, u_n)^T$ whose nonnegative entries sum to 1.

A general Markov model for transitions then takes the form of a discrete-time dynamical system (Section 2.3),

$$\mathbf{u}_{t+1} = P\mathbf{u}_t, \qquad \mathbf{u}_0 \text{ given}, \tag{5.2}$$

where $\mathbf{u}_t$ is a probability vector and P is a transition matrix. To calculate the long-term probabilities associated with each state we can use the eigenvector of P corresponding to eigenvalue $\lambda = 1$. This result is made precise in the following theorem (see also [3]).

Theorem 5.1. *Providing some power of P has all positive entries, then for any probability vector $\mathbf{u}_0$ and model $\mathbf{u}_{t+1} = P\mathbf{u}_t$, $\mathbf{u}_t \to \mathbf{u}^*$ as $t \to \infty$, where $P\mathbf{u}^* = \mathbf{u}^*$.*

The requirement that some power of P has all positive entries (P is primitive) ensures that, given enough time-steps, one can transition from any state to any other state and hence the result is independent of the original state $\mathbf{u}_0$.

5.2.3 The Princeton Forest Ecosystem

A more complex model for successional dynamics was made for the well-studied Princeton forest ecosystem [90, 91]. Here the transitional probabilities between five dominant trees were measured in terms of which species replaced resident trees, once they died. Results were as given in Figure 5.2 and Table 5.2.

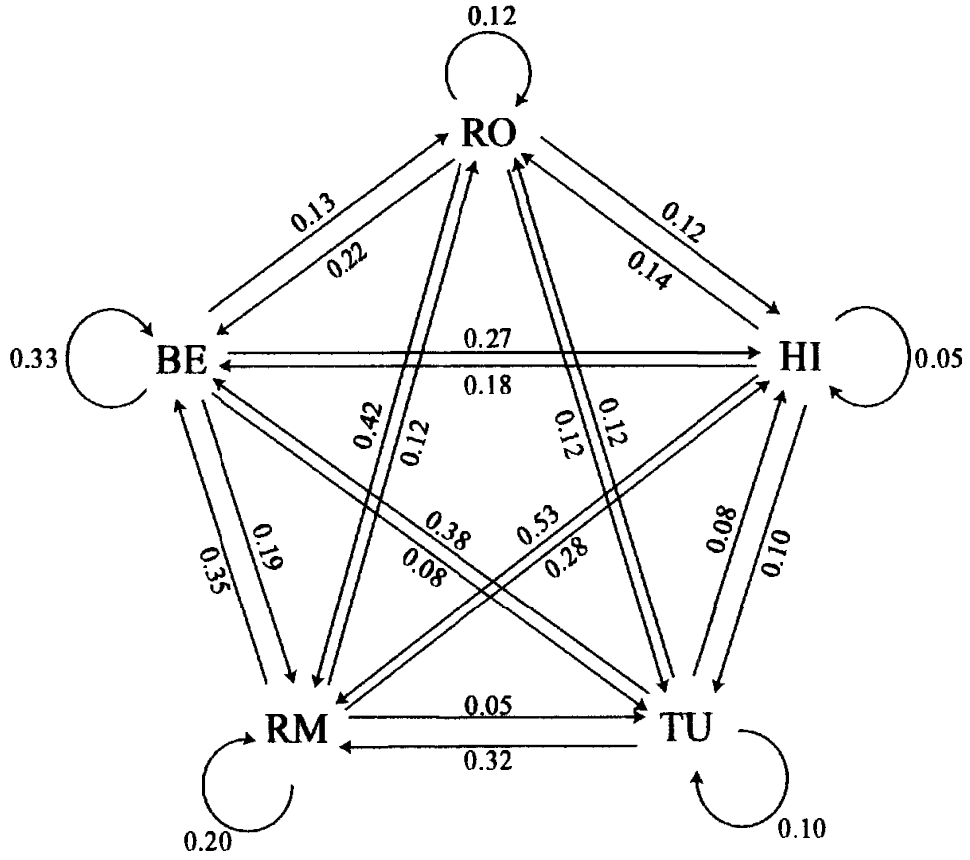

Figure 5.2. *Transitions between tree species in the Princeton forest shown as a graph. Based on* [91, 90].

Table 5.2. *Transition probabilities for the Princeton forest ecosystem shown as a table. Based on* [91, 90].

RO = Red oak
HI = Hickory
TU = Tulip tree
RM = Red maple
BE = Beech

	Canopy				
% Saplings	RO	HI	TU	RM	BE
Red oak	0.12	0.14	0.12	0.12	0.13
Hickory	0.12	0.05	0.08	0.28	0.27
Tulip tree	0.12	0.10	0.10	0.05	0.08
Red maple	0.42	0.53	0.32	0.20	0.19
Beech	0.22	0.18	0.38	0.35	0.33

For this full forest succession model, knowledge of the transitional probabilities can be translated into a prediction for the climax successional community, u^*, by applying Theorem 5.1 (see Exercise 5.8.2):

$$\mathbf{u}^* = \begin{pmatrix} 0.128 \\ 0.197 \\ 0.080 \\ 0.298 \\ 0.297 \end{pmatrix} \begin{matrix} \longleftarrow \text{RO} \\ \longleftarrow \text{HI} \\ \longleftarrow \text{TU} \\ \longleftarrow \text{RM} \\ \longleftarrow \text{BE} \end{matrix} . \tag{5.3}$$

A comparison of the prediction and the observed proportions of trees in the climax forest areas was made in [90, 91], where it was shown that the prediction and observation were closely correlated.

5.3 Working with Random Variables

In the previous section, our mathematical analysis of the forest succession involved calculating changes in probabilities associated with the random variable $X(t)$ taking on different values. In that example, $X(t)$ described the event of having a particular sort of tree at a given location in the forest. In different problems, a random variable could describe any number of things. For example, the random variable A could be used to describe the length of time a cell remains alive.

Random variables can be broken down into two main classes: continuous (the set S of values that X takes is continuous) and discrete (the set S of values that X takes is discrete). For each of these classes, we need a mathematical formalism which assigns a probability to the different values in S that the random variable can attain.

To derive more complex models for stochastic processes, we need some basic laws of probability that can be applied to random variables. We motivate discussion of the laws by a simple example which involves cell death.

Consider the length of time A for which a cell lives. A is a random variable defined on the nonnegative real numbers which describes the age of the cell. We may be interested in relating events associated with random variables, for example, the event in which the cell dies by age a_2, given that it was alive at age $a_1 < a_2$.

The *law of conditional probability* can be used to relate conditional events to other events. Specifically, if E_1 and E_2 are events which occur with nonzero probability, then

$$\Pr\{E_2|E_1\} = \frac{\Pr\{E_2 \cap E_1\}}{\Pr\{E_1\}}. \tag{5.4}$$

In the above example, if we let E_2 be the event that $A \in (a_1, a_2)$ and E_1 be the event that $A > a_1$, then the probability that the cell dies during age interval (a_1, a_2), given that it was alive at age $a_1 < a_2$, is

$$\begin{aligned}\Pr\{A \in (a_1, a_2)|A > a_1\} &= \frac{\Pr\{A \in (a_1, a_2) \cap A > a_1\}}{\Pr\{A > a_1\}} \\ &= \frac{\Pr\{A \in (a_1, a_2)\}}{\Pr\{A > a_1\}}.\end{aligned} \tag{5.5}$$

5.3.1 Probability Density

In the above instance of cell death, the age of cell death A is a continuous random variable defined over the nonnegative real numbers. Here, the age of death is described by a *probability density function* $f(a)$ such that

$$\Pr\{a_1 \leq A < a_2\} = \int_{a_1}^{a_2} f(\alpha)\, d\alpha. \tag{5.6}$$

The related *cumulative density function* describing the probability that the cell dies by age a is

$$F(a) = \Pr\{A < a\} = \int_0^a f(\alpha)\, d\alpha. \tag{5.7}$$

The fundamental theorem of calculus relates the two functions by $f(a) = F'(a)$. The actual probability density function f depends upon the model we choose for cell death.

To show how f can be derived from first principles, we consider a simple model for cell death. Suppose the age-dependent death rate is given by $\mu(a)$. Our model assumes that the probability of the cell dying in time interval $(a, a + \Delta a)$, given that it was alive at age a, is equal to $\mu(a)\Delta a + o(\Delta a)$, where $o(x)$ is the Landau symbol, designating lower order terms ($\lim_{x\to 0} o(x)/x = 0$). In terms of the random variable A, this model can be written as

$$\Pr\{A \in (a, a + \Delta a) | A > a\} = \mu(a)\Delta a + o(\Delta a). \tag{5.8}$$

Now

$$\Pr\{A > a\} = \Pr\{A \in (a, a + \Delta a)\} + \Pr\{A > a + \Delta a\}. \tag{5.9}$$

Using (5.5) and (5.8), we rewrite the last statement as follows:

$$\begin{aligned} \Pr\{A > a\} &= \Pr\{A \in (a, a + \Delta a) | A > a\} \Pr\{A > a\} + \Pr\{A > a + \Delta a\} \\ &= \Pr\{A > a\}\mu(a)\Delta a + o(\Delta a) + \Pr\{A > a + \Delta a\}, \end{aligned} \tag{5.10}$$

so that

$$\Pr\{A > a + \Delta a\} - \Pr\{A > a\} = -\Pr\{A > a\}\mu(a)\Delta a + o(\Delta a). \tag{5.11}$$

Dividing both sides of (5.11) by Δa and taking the limit as $\Delta a \to 0$ yields

$$\frac{dp}{da} = -\mu(a)p(a), \qquad p(0) = 1, \tag{5.12}$$

where $p(a) = \Pr\{A > a\}$. This differential equation for the probability of surviving to age a has solution

$$p(a) = \exp\left(-\int_0^a \mu(\tau)\, d\tau\right). \tag{5.13}$$

Returning to the definition of the cumulative density function for the age of cell death (5.7), we observe that $F(a) = 1 - p(a)$; hence the probability density for cell death is

$$f(a) = -p'(a) = \mu(a)\exp\left(-\int_0^a \mu(\tau)\, d\tau\right). \tag{5.14}$$

For the case where $\mu(a)$ is a constant, we have an exponentially distributed waiting time or cell death, $f(a) = \mu\exp(-\mu a)$.

5.3.2 Probability Mass

When the random variable X is discrete, the probabilities associated with each outcome in $S = \{x_0, x_1, x_2, \ldots\}$ are given by a *probability mass function* with values

$$p_n = \Pr\{X = x_n\}. \tag{5.15}$$

The simplest discrete random variable is a *Bernoulli* random variable. This random variable X has only two possible outcomes: 0 (failure) or 1 (success). For example, it could be used to describe whether a given cell is alive or dead at some fixed age a. Here, we associate success with the outcome of being alive at the fixed age a. The sample space is $S = \{0, 1\}$, and the probabilities associated with each outcome are $p_0 = \Pr\{X = 0\} = 1 - p$ and $p_1 = \Pr\{X = 1\} = p$, where $p = \Pr\{\text{success}\}$. For the above cell death model, $p = p(a)$, given by (5.13), describes the probability of being alive (success) at age a. A simpler example is given by the outcome from flipping a coin, which yields $X = 0$ (tails) or $X = 1$ (heads). Here, X is distributed as a Bernoulli random variable and, if the coin is fair, $p = 0.5$.

We now consider the case where there are many identical cells or, alternatively, many coin flips. The sum of m independent identically distributed Bernoulli random variables is a *binomial* random variable. The probability of n successes in m trials is

$$p_n = \text{Bin}(n; m, p) = \binom{m}{n} p^n (1-p)^{m-n}, \qquad n = 0, 1, 2, \ldots, m. \tag{5.16}$$

For example, the probability of seven heads in ten coin flips is $p_7 = \text{Bin}(7; 10, 0.5) = 0.117$. Note that the binomial theorem ensures that the probabilities sum to 1:

$$\sum_{n=0}^{m} p_n = \sum_{n=0}^{m} \binom{m}{n} p^n (1-p)^{m-n} = (p + (1-p))^m = 1. \tag{5.17}$$

If we consider a population of m independent, identical cells, and denote

$$X_i(a) = \begin{cases} 1 & \text{if cell } i \text{ is alive at age } a, \\ 0 & \text{if cell } i \text{ is dead at age } a, \end{cases} \qquad i = 1, \ldots, m, \tag{5.18}$$

then we can count the number of cells alive at age a as

$$Y(a) = \sum_{i=1}^{m} X_i(a). \tag{5.19}$$

At any fixed age a, $Y(a)$ is the sum of m independent, identical cells. Hence $Y(a)$ is a stochastic process described by a binomial random variable, with $p_n = \text{Bin}(n; m, p(a))$ and $p(a)$ as given in (5.13).

If each trial is an independent identically distributed Bernoulli random variable, then the probability of the kth success occurring on the nth trial is governed by the negative binomial distribution

$$\begin{aligned} p_n &= \text{NB}(n; k, p) \\ &= \Pr\{k-1 \text{ successes on the first } n-1 \text{ trials}\} \cdot \Pr\{\text{success}\} \\ &= \binom{n-1}{k-1} p^{k-1} (1-p)^{n-k} \cdot p \\ &= \binom{n-1}{k-1} p^k (1-p)^{n-k}, \qquad n = k, k+1, k+2, \ldots. \end{aligned} \tag{5.20}$$

For example, the probability of the seventh head occurring on the tenth coin flip is $p_7 = \mathrm{NB}(7; 10, 0.5) = 0.082$. Returning to the example of cell death, the probability of finding the kth living cell on the nth cell checked is $p_n = \mathrm{NB}(n; k, p(a))$.

5.3.3 Descriptive Statistics

When we use stochastic processes to describe the uncertain behavior of biological models, it is convenient to have summary statistics to describe the qualitative features. The most commonly used measures are the mean, which describes average values, and the variance, which describes variability about the mean.

Given a random variable X, whose sample space is the natural numbers and whose probabilities are $p_n = \Pr\{X = n\}$, $n = 0, 1, 2, \ldots$, and any real-valued function of X, denoted by ϕ, the *expected value* of $\phi(X)$ is

$$E(\phi(X)) = \sum_{n=0}^{\infty} \phi(n) p_n. \tag{5.21}$$

For some functions ϕ, the sum will not converge. When this is true, we say that $E(\phi(X))$ does not exist.

We note that expectation is a linear operator, so that for any functions ϕ and ψ, the expectation of a linear combination of ϕ and ψ is the linear combination of the expectations: $E(a\phi(X) + b\psi(X)) = aE(\phi(X)) + bE(\psi(X))$ for all real numbers a and b.

When $\phi = X^m$, the *mth moment* M_m is

$$M_m = E(X^m) = \sum_{n=0}^{\infty} n^m p_n. \tag{5.22}$$

The first moment M_1 is also referred to as the *expected value of* X, $E(X)$, or *mean of* X, μ. The variance is the expected value of the squared deviations about the mean,

$$\mathrm{var}(X) = \sigma^2 = E((X - \mu)^2) = \sum_{n=0}^{\infty} (n - \mu)^2 p_n. \tag{5.23}$$

Using the linear operator property of expectation, we note that the variance can be rewritten as $\mathrm{var}(X) = \sigma^2 = E(X^2 - 2\mu X + \mu^2) = E(X^2) - 2\mu E(X) + \mu^2 = M_2 - M_1^2$. This is sometimes referred to as the computational form of the variance.

While the variance gives a measure of squared deviations about the mean, we may be interested in a measure for the typical spread about the mean. This is the standard deviation, $\sigma = \sqrt{\sigma^2}$. The *coefficient of variation*, a scaled measure of the spread, is c.v.$= \sigma/\mu$.

We now consider some examples taken from the distributions discussed in earlier sections. The first and second moments of a Bernoulli random variable are both given by p, and hence the variance is $M_2 - M_1^2 = p - p^2 = p(1 - p)$, which is highest for values of p intermediate between 0 and 1.

To calculate the mean of the binomial distribution, we must evaluate

$$\begin{aligned} M_1 = E(X) &= \sum_{n=0}^{m} np_n = \sum_{n=0}^{m} n \binom{m}{n} p^n (1-p)^{m-n} \\ &= \sum_{n=1}^{m} \frac{m!}{(n-1)!(m-n)!} p^n (1-p)^{m-n} \\ &= mp \sum_{n=1}^{m} \binom{m-1}{n-1} p^{(n-1)} (1-p)^{m-n} \\ &= mp. \end{aligned} \tag{5.24}$$

A similar calculation gives

$$M_2 - M_1 = E(X(X-1)) = m(m-1)p^2. \tag{5.25}$$

Hence the variance is $\sigma^2 = M_2 - M_1^2 = E(X(X-1)) + M_1 - M_1^2 = mp(1-p)$.

Note that the mean and variance of the binomial distribution are simply m times the mean and variance for the Bernoulli distribution. In general, if $Y = \sum X_i$, then $E(Y) = \sum E(X_i)$, and if the X_i's are independent ($E(X_i X_j) = E(X_i)E(X_j)$), then $\text{var}(Y) = \sum \text{var}(X_i)$. The proof of this is left as an exercise (Exercise 5.8.3).

The expected values for continuous random variables are defined in an analogous way to those for discrete random variables. Given a random variable X, whose sample space is the nonnegative real numbers and whose probability density function is $f(x)$, the expectation of $\phi(X)$ is

$$E(\phi(X)) = \int_0^\infty \phi(x) f(x)\, dx, \tag{5.26}$$

providing it exists. As with the previous discussion of discrete random variables, the mean and variance can be calculated from the moments of X.

In our cell death example, we may be interested in the mean and variance in the age of death A for a cell where A is exponentially distributed. Here, integration by parts yields

$$\begin{aligned} M_1 = E(A) &= \int_0^\infty a\mu \exp(-\mu a)\, da = \mu^{-1}, \\ M_2 = E(A^2) &= \int_0^\infty a^2 \mu \exp(-\mu a)\, da = 2\mu^{-2}, \end{aligned} \tag{5.27}$$

and hence A has a mean of μ^{-1} and the computational formula for the variance yields $\text{var}(A) = M_2 - M_1^2 = \mu^{-2}$.

5.3.4 The Generating Function

One of the workhorses in stochastic processes is the generating function. Given any discrete random variable Y that assumes values in the natural numbers n with probability p_n, the *generating function* is defined as

$$g(s) = \sum_{n=0}^{\infty} s^n p_n, \qquad 0 \le s \le 1. \tag{5.28}$$

Formally, we may write

$$g(s) = E(s^Y) \tag{5.29}$$

to denote the generating function. Note that the power series in (5.28) converges for all $0 \leq s \leq 1$ and is an increasing function of s, with $g(0) = p_0$, $g(1) = 1$, and hence $p_0 \leq g(s) \leq 1$.

All the information about the random variable Y is contained within its generating function $g(s)$. To observe this, note that Taylor's theorem permits us to expand a function in terms of its derivatives at zero. Thus, by calculating derivatives, we regain the probabilities p_n associated with the random variable as follows:

$$p_n = \frac{1}{n!} \frac{d^n g}{ds^n}\bigg|_{s=0}. \tag{5.30}$$

The generating function allows us to compute the mean as

$$E(Y) = \sum_{n=0}^{\infty} np_n = g'(1). \tag{5.31}$$

The variance takes only slightly more effort. We can write

$$g''(1) = \sum_{n=0}^{\infty} n(n-1)p_n, \tag{5.32}$$

yielding

$$\text{var}(Y) = g''(1) + g'(1) - g'^2(1). \tag{5.33}$$

In other words, the generating function allows us to compute all the probabilities and statistics we need in a straightforward way. As we will show in several cases, it may be easier to compute the generating function $g(s)$ for a discrete random variable than to compute the values of p_n directly.

By way of example, we consider the Bernoulli random variable of Section 5.3.2. Recall that the random variable Y has only two possible outcomes: 0 with probability $1 - p$, and 1 with probability p. The generating function is simply $g(s) = (1 - p) + sp$. The derivatives, evaluated at $s = 1$, are $g'(1) = p$ and $g''(1) = 0$. Hence formula (5.31) yields the mean for the Bernoulli random variable as p, and formula (5.33) yields the variance as $p(1 - p)$. These formulae were derived, using different methods, in Section 5.3.3.

Just as we earlier calculated the variance for a sum of independent random variables, we can also calculate the generating function for a sum of independent random variables. Whereas the variances add, the generating functions multiply. Specifically, if X and Y are independent random variables with probability mass functions g_n and h_n, and generating functions $g(s)$ and $h(s)$, respectively, then $X + Y$ has generating function $g(s)h(s)$. To show this, we first observe that

$$\Pr\{X + Y = n\} = \sum_{k=1}^{n} g_k h_{n-k}. \tag{5.34}$$

Hence the generating function for $X + Y$ is

$$\sum_{n=0}^{\infty}\left(\sum_{k=0}^{n} g_k h_{n-k}\right) s^n = \sum_{k=0}^{\infty}\sum_{n=k}^{\infty} g_k h_{n-k} s^n = \sum_{k=0}^{\infty} g_k s^k \sum_{n=k}^{\infty} h_{n-k} s^{n-k} = g(s)h(s). \quad (5.35)$$

Note that the calculation is made possible by the careful switching of the limits for the double sum after the first "equals" sign of the above line. You may want to check this to make sure you agree with it. This result can be extended in a straightforward way to the sum of m random variables. The generating function for the sum is the m-fold product of the individual generating functions. We use this result when analyzing branching processes in Section 5.5.

We now extend the previous example of using generating functions to calculate the mean and variance of a Bernoulli random variable. Recall that the sum of m independent Bernoulli random variables is the binomial random variable Bin$(n; m, p)$ (equation (5.16)). Hence the generating function for Bin$(n; m, p)$ is simply the m-fold product of the Bernoulli random variable generating function, $(1 - p + sp)$, with itself, $g(s) = (1 - p + sp)^m$. This generating function $g(s)$ has derivatives $g'(1) = mp$ and $g''(1) = m(m-1)p^2$, and so formulae (5.31) and (5.33) yield mp and $mp(1-p)$ for the mean and variance, respectively. These formulae were derived, using different methods, in Section 5.3.3.

In the remainder of the chapter, we apply the methods from this section to develop and analyze a series of stochastic models. We motivate each model with a biological problem, but the mathematical tools that are brought to bear on the problem have general application to the analysis of stochastic processes.

5.4 Diffusion Processes

Most living organisms move in space. Given that we have some information about how an organism moves over short time scales, can we determine where it is likely to be over long time scales? If movement rules are simple, mathematical models can be used to translate the movement rules into equations. As we will show in this section, analysis of the resulting equations yields a probability density function that can be used to track the changing location of the animal over time.

We consider an individual executing a random walk in one-dimensional space. At each time step, the individual jumps to either the right or the left, and its new position is determined by its current position plus a random increment to the left or right. This is another example of a Markov process, because the current location plus the random increment is sufficient to determine the next position. The precise path taken to get to the current location plays no role in determining future positions.

In the next section, we calculate the probability mass function for the location of the individual after a given number of time steps. We show that, after a sufficiently large number of time steps, the probability mass function can be approximated by a Gaussian probability density function.

time step	1	2	3	4	5	6	7	8	9	10
	L	*R*	*L*	*R*	*L*	*L*	*L*	*L*	*R*	*L*
	11	12	13	14	15	16	17	18	19	20
	R	*L*	*R*	*L*	*L*	*L*	*R*	*L*	*L*	*L*

Figure 5.3. *Jumps to the left and right, for* 20 *time steps, are generated by flipping a fair coin* 20 *times.*

5.4.1 Random Motion in One Dimension

Suppose an individual, released at $x = 0$, moves back and forth randomly along a line in fixed steps $\lambda = \Delta x$ at fixed time intervals $\tau = \Delta t$, and that this movement is unbiased (equal probability of moving right and left). After k time steps (time $= k\tau$), the individual is anywhere from $x = -k\lambda$ to $x = k\lambda$.

We describe the stochastic process with $p_n(k)$, the probability the individual reaches n space steps to the right ($x = n\lambda$) after k time steps ($t = k\tau$). Suppose that to reach $n\lambda$, the individual has moved a steps to the right and b steps to the left.

$$n = a - b, \quad k = a + b,$$

implying that $k + n = 2a$ is even. Thus k odd implies n odd, and k even implies n even. After n time steps, only every other point ("evens" or "odds") can be occupied.

For example, consider the case where, after $k = 20$ time steps, the jumps to the left L and right R are given by the sequence shown in Figure 5.3. These were generated by flipping a fair coin 20 times, and writing R for heads and L for tails. Here, $a = 6$, $b = 14$, $k = 20$, $n = -8$.

Other orderings of $a = 6$ jumps to the right and $b = 14$ jumps to the left in the $k = 20$ time steps would describe other paths that would still lead to location $n = -8$ after $k = 20$ time steps. The number of possible ways to insert $a = 6$ R's into a sequence of length 20 is given by the following combinatorial expression:

$$\frac{20 \cdot 19 \cdot 18 \cdot 17 \cdot 16 \cdot 15}{6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} = \frac{20!}{14!\,6!} = \binom{20}{6}.$$

In general, the number of possible paths that an individual can take to reach $x = n\lambda$ in k time steps (time $= k\tau$) is given by

$$\frac{k!}{a!\,b!} = \frac{k!}{a!\,(k-a)!} = \binom{k}{a},$$

with $\binom{k}{a}$ being the number of possible combinations of a moves to the right in k time steps (and $a = \frac{k+n}{2}$). The expression $\binom{k}{a}$ is referred to as the *binomial coefficient.* It features in both the binomial distribution, (5.16), and the *binomial expansion*

$$(x + y)^k = \sum_{a=0}^{k} \binom{k}{a} x^{k-a} y^a.$$

The total number of possible k-step paths is 2^k, so

$$p_n(k) = \frac{\text{no. of possible paths to reach } x = n\lambda \text{ in } k \text{ time steps}}{\text{no. of possible paths in } k \text{ time steps}} = \frac{1}{2^k}\binom{k}{a}.$$

This is a binomial distribution, (5.16), where the probability of success is $\frac{1}{2}$. Note that for $n + k$ even,

$$\sum_{n=-k}^{k} p_n(k) = \sum_{a=0}^{k} \binom{k}{a} \left(\frac{1}{2}\right)^{k-a} \left(\frac{1}{2}\right)^{a} = \left(\frac{1}{2} + \frac{1}{2}\right)^{k} = 1,$$

where we used the facts that $n = -k$ implies $a = 0$ and $n = k$ implies $a = k$.

Providing k is sufficiently large (i.e., a sufficiently large number of time steps have taken place), the binomial distribution can be approximated with a normal (Gaussian) distribution with variance k [3],

$$p_n(k) \approx \left[\frac{2}{\pi k}\right]^{\frac{1}{2}} \exp\left[\frac{-n^2}{2k}\right].$$

We can translate this result in terms of continuous time and space by recalling that $n = \frac{x}{\lambda}$ and $k = \frac{t}{\tau}$. What happens when $\lambda, \tau \to 0$, but $x = n\lambda$ and $t = k\tau$ are finite? The relevant quantity is

$$p(x,t) = \lim_{\substack{\lambda \to 0 \\ \tau \to 0}} \frac{p_{(x/\lambda)}(t/\tau)}{2\lambda} \approx \left[\frac{\tau}{2(\lambda)^2} \frac{1}{\pi t}\right]^{\frac{1}{2}} \exp\left[\frac{-\tau}{2(\lambda)^2} \frac{x^2}{t}\right]$$

$$= \left[\frac{1}{4\pi Dt}\right]^{\frac{1}{2}} \exp\left[\frac{-x^2}{4Dt}\right] \tag{5.36}$$

if we assume

$$D = \lim_{\substack{\lambda \to 0 \\ \tau \to 0}} \frac{(\lambda)^2}{2\tau} \neq 0. \tag{5.37}$$

This assumption implies that individuals move very quickly ($\lambda/\tau \to \infty$), but switch direction very frequently ($\tau \to 0$). This limit is often referred to as the *parabolic limit* because, as we will see below, it can be connected to the parabolic diffusion PDE. Since the speeds and times can be scaled by changing the units used to measure them, we can say, equivalently, that the parabolic limit is valid when we are describing movement over small spatial scales ("small" units for space) and large time scales ("large" units for time), relative to the characteristic space and time steps taken by the individual.

Note that (5.36) is the fundamental solution of the diffusion equation as shown in Section 4.3.2. The coefficient D is called the *diffusion coefficient*. It has dimensions length^2/time and measures how efficiently individuals disperse from high to low density. For example, hemoglobin in blood has diffusion coefficient $D \approx 10^{-7}$ cm^2/sec, whereas oxygen in blood has diffusion coefficient $D \approx 10^{-5}$ cm^2/sec.

In the next section, we consider the limiting case of small time and space steps and large velocity. In this case, the probability density function, describing the location of the individual, satisfies a parabolic PDE called the diffusion equation.

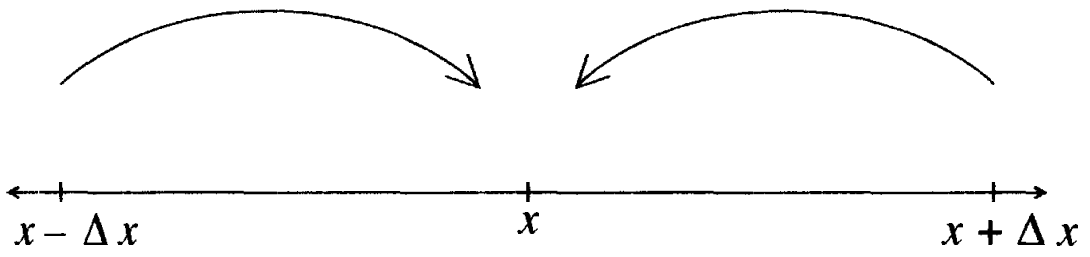

Figure 5.4. *Movement on the lattice giving rise to the diffusion equation. Here an individual arriving at x can come from locations $x - \Delta x$ and $x + \Delta x$, where $\Delta x = \lambda$ in equation* (5.38).

5.4.2 Diffusion Equation

An alternate approach to the random walk derives the diffusion PDE for a probability density function describing the location of the individual directly from a *master equation*. Let $X(t)$ be a stochastic process describing the location of an individual at time t, which is released at location $x = 0$ at time $= 0$ ($X(0) = 0$). We define a time-dependent probability density function $p(x, t)$ and small space interval λ, such that $p(x, t)\lambda =$ probability an individual released at $x = 0$ and time $= 0$ is between x and $x + \lambda$ at time t.

Assume the random walk is unbiased. The *master equation* which describes movement on the lattice of points situated a distance λ apart is

$$p(x, t + \tau) = \frac{1}{2}\, p(x - \lambda, t) + \frac{1}{2}\, p(x + \lambda, t) \tag{5.38}$$

(see Figure 5.4).

Expanding in Taylor series gives

$$\begin{aligned} & p(x,t) + \tau \frac{\partial p}{\partial t}(x,t) + \frac{(\tau)^2}{2}\frac{\partial^2 p}{\partial t^2}(x,t) + \text{h.o.t.} \\ & = \frac{1}{2}\left\{ p(x,t) - \lambda \frac{\partial p}{\partial x}(x,t) + \frac{(\lambda)^2}{2}\frac{\partial^2 p}{\partial x^2}(x,t) + \text{h.o.t.} \right. \\ & \quad \left. + \; p(x,t) + \lambda \frac{\partial p}{\partial x}(x,t) + \frac{(\lambda)^2}{2}\frac{\partial^2 p}{\partial x^2}(x,t) + \text{h.o.t.} \right\}, \end{aligned}$$

where h.o.t. indicates higher-order terms in the Taylor series. The above equation can be simplified to yield

$$\frac{\partial p}{\partial t} + \frac{\tau}{2}\frac{\partial^2 p}{\partial t^2} = \frac{(\lambda)^2}{2\tau}\frac{\partial^2 p}{\partial x^2}.$$

Taking the limit as $\lambda, \tau \to 0$ so that $\frac{(\lambda)^2}{2\tau} \to D$ yields the diffusion equation

$$\frac{\partial p}{\partial t} = D\frac{\partial^2 p}{\partial x^2}. \tag{5.39}$$

The solution to (5.39), corresponding to the point release of an individual, is a Gaussian centered about zero, with variance $2Dt$:

$$p(x,t) = \left[\frac{1}{4\pi Dt}\right]^{\frac{1}{2}} \exp\left[\frac{-x^2}{4Dt}\right], \tag{5.40}$$

as we found in (5.36), and also earlier in (4.16). This solution can be verified by substitution and by noting that as $t \to 0$, $p(x,t) \to \delta(x)$ (see Exercise 4.5.2). Because (5.40) is an even function, the first moment $M_1 = E(X)$ is equal to zero (see (5.26)), and hence the variance is the second moment. We can derive a differential equation for the second moment,

$$M_2(t) = \int_{-\infty}^{\infty} x^2 p(x,t)\, dx,$$

by using (5.39) and integrating by parts,

$$\dot{M}_2 = \int_{-\infty}^{\infty} x^2 p_t\, dx = \int_{-\infty}^{\infty} x^2 D p_{xx}\, dx = x^2 D p_x\Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} 2xDp_x\, dx$$

$$= -2xp\Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} 2Dp\, dx = 2D\int_{-\infty}^{\infty} p(x,t)\, dx = 2D.$$

Solving this differential equation subject to the initial condition $M_2(0) = 0$ gives

$$M_2 = 2Dt, \tag{5.41}$$

which implies that the second moment M_2 grows linearly with time at rate $2D$. This linear growth in the second moment describes increasing uncertainty as to the location of the individual as time progresses (see Figure 4.6).

5.5 Branching Processes

A branching process is a stochastic process that describes a reproducing population. The random variable is the number of individuals in each generation, where it is assumed that the behavior of the younger generation is (stochastically) independent from the older generation. The simplest and most important example of a branching process is the Galton–Watson process, which is discussed in Section 5.5.1; an application to a polymerase chain reaction is given in Section 5.5.2.

5.5.1 Galton–Watson Process

Reverend H.W. Watson and Francis Galton [159] were interested in the extinction of family names. In an age where family name inheritance was restricted to males, it was possible that a run of "bad luck" would result in a family name going extinct even if, on average, a man were to have more than one son. Watson and Galton supposed that $p_0, p_1, p_2, \ldots$ were the probabilities that a man has 0, 1, 2, ... sons, and determined the probability that the direct male line (i.e., family name arising from that original man) was extinct after r generations.

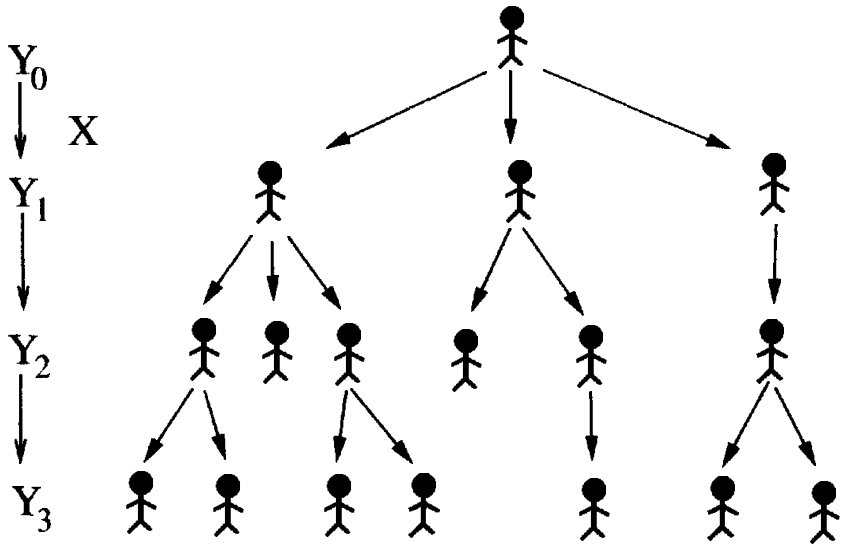

Figure 5.5. *The Galton–Watson process.*

More generally, they determined the probability associated with any specified number of direct male descendants in a given generation. Whereas rules for inheritance of family names have changed since the times of Galton and Watson, the mathematics behind the calculations remains relevant today and can be applied to a variety of interesting stochastic problems.

The Galton–Watson process is a Markov process in discrete time. Here, Y_t denotes the number of parents at time step t, $t = 0, 1, 2, \ldots$. Initially, there is a single parent ($Y_0 = 1$), who has a random number of offspring X with $\Pr(X = n) = p_n$. After a single generation, $Y_1 = X$. In the next generation, each of the Y_1 offspring becomes a parent and has a random number of offspring, and so on (Figure 5.5). Here the conditional distribution of Y_{t+1}, given that $Y_t = m$, is the sum of m independent variables, each with the same distribution as X. This is referred to as a branching process. An excellent discussion of branching processes is given in [102], and we follow the approach given there.

Our investigation into Watson and Galton's question starts by calculating the change in the generating function for the number of parents, Y_t, from one time step to the next. By knowing the original number of parents, Y_0, and by tracking the change in generation from time step to time step, we can calculate the generating function for the number of parents at every time step. As discussed in Section 5.3.4 and shown in (5.30), the generating function will then give us the probability mass function for the number of parents in each time step. If we are interested in the likelihood of extinction of the family name by generation t, then this can be calculated as the probability mass associated with $Y_t = 0$.

We suppose that the generating function for the number of offspring, X, from a parent is given by $g(s)$, and that there are Y_t identical parents at time step t. For example, if each parent had 0, 1, or 2 offspring with probability 1/3, then $p_0 = p_1 = p_2 = 1/3$, $p_n = 0$ for $n > 2$, and $g(s) = 1/3 + s/3 + s^2/3$.

The parents in generation $t + 1$ are the offspring from generation t. The number of offspring arising from each of the Y_t parents in generation t is assumed to be independent from the number of offspring arising from the other $Y_t - 1$ parents in generation t. Hence

the number of parents in generation $t+1$ is given by the sum

$$\sum_{i=1}^{Y_t} X_i, \tag{5.42}$$

where the X_i's are independent, identically distributed random variables, each with generating function $g(s)$.

We know that the generating function for the sum of m independent, identically distributed random variables, each with generating function $g(s)$, is simply $(g(s))^m$ (see Section 5.3.4). Therefore, if we knew the number of parents in generation t to be $Y_t = m$, then the generating function for the number of parents in the next generation would be $(g(s))^m$.

However, our calculation is complicated by the fact that the number of parents in generation t, Y_t, is a random variable. So as to keep track of this random variable, we denote the probability mass for the random variable Y_t as $q_{tm} = \Pr\{Y_t = m\}$ and the generating function for Y_t as $h_t(s)$. Thus the generating function for the number of parents in generation $t+1$ is $(g(s))^m$, conditioned upon the different values that m can attain in generation t,

$$h_{t+1}(s) = \sum_{m=0}^{\infty} \Pr\{Y_t = m\}(g(s))^m = \sum_{m=0}^{\infty} q_{tm}(g(s))^m = h_t(g(s)) = h_t \circ g(s), \tag{5.43}$$

where $\circ$ indicates functional composition.

At $t = 0$, there is a single parent ($Y_0 = 1$) so that $h_0(s) = s$. Applying (5.43), we observe that $h_1(s) = g(s)$. Continuing to the next generation, we have $h_2(s) = g \circ g(s)$. Returning to our above example where each parent has 0, 1, or 2 offspring with probability 1/3, we calculate $h_0(s) = s$, $h_1(s) = g(s) = 1/3 + s/3 + s^2/3$, $h_2(s) = g \circ g(s) = 1/3 + (1/3+s/3+s^2/3)/3+(1/3+s/3+s^2/3)^2/3 = 13/27+5s/27+6s^2/27+2s^3/27+s^4/27$, and so forth. Using either the definition of the generating function (5.28) or equation (5.30), we can deduce that, after two generations, the probabilities of having 0, 1, 2, 3, or 4 parents are 13/27, 5/27, 6/27, 2/27, and 1/27, respectively. Hence the probability of extinction after two generations is 13/27.

With each new generation, we iterate with the generating function to obtain

$$h_t(s) = h_{t-1} \circ g(s) = g^t(s), \tag{5.44}$$

a t-fold composition of the generating function $g(s)$. While the generating function $h_t(s)$ may not have such a simple form as given in the example above, (5.44) gives us a straightforward method for calculating $h_t(s)$. From this generating function, we can use the methods of Section 5.3.4 to calculate the probability mass function for the number of parents at each time step.

The expected number of parents in the tth generation is

$$E(Y_t) = h_t'(1) = g'(h_{t-1}(1))h_{t-1}'(1) = g'(1)h_{t-1}'(1) = R_0 E(Y_{t-1}), \tag{5.45}$$

where $R_0 = g'(1) = E(Y_1)$, and hence $E(Y_t) = R_0^t$, so the expectation grows geometrically with *reproduction ratio* R_0.

Calculation of the variance from (5.33) by similar methods is a little more involved and yields

$$\text{var}(Y_t) = \begin{cases} \frac{(R_0^t-1)R_0^t}{R_0(R_0-1)}\sigma^2, & R_0 \neq 1, \\ \sigma^2 t, & R_0 = 1, \end{cases} \tag{5.46}$$

where σ^2 is the variance of the generating function $g(s)$ (see Exercise 5.8.7). Higher moments also can be found in a similar manner.

To determine the chance of a lineage going extinct, we can use the cobwebbing method of Section 2.2. The probability of being extinct in generation t, x_t, is $x_t = h_t(0)$. Equation (5.44) tells us that x_t satisfies

$$x_{t+1} = g(x_t), \tag{5.47}$$

with initial condition $x_0 = 0$. Recall that $R_0 = g'(1)$, and hence there are two generic behaviors, depending upon whether $R_0 < 1$ (subcritical case; see Figure 5.6, top panel) or $R_0 > 1$ (supercritical case; see Figure 5.6, bottom panel). In the subcritical case, eventual extinction is inevitable. In the supercritical case, $x_t \to x^*$, the unique root to $g(x) = x$, and eventual extinction is possible, but not inevitable.

5.5.2 Polymerase Chain Reaction

Polymerase chain reaction (PCR) is a standard technique of molecular biology in which a small amount of nucleic acid (DNA or RNA) taken from a probe is multiplied so that it can be detected. This method is the first step in DNA fingerprinting, preceding the sequencing of the amplified nucleic acid. However, most investigations are interested not only in determining the sequencing of the nucleic acid, but also in calculating the amount of the sequenced strands present in the original probe. Quantitative PCR allows the user to calculate the starting amounts of the nucleic acid template by analyzing the amount of DNA produced during each cycle of PCR. The technique relies on the fluorescence of a reporter molecule that increases as product accumulates with each cycle of amplification.

The strings of nucleic acid encoding the DNA or RNA are incubated with a mixture of primers and nucleotides. This mixture allows the strings to replicate. String replication is stochastic, with the amplification factor (probability of doubling, or p_d) ranging between 0.6 and 0.8. Each string is assumed to be independent of the others. Hence this can be described as a Galton–Watson process with $p_0 = 1 - p_d$, $p_2 = p_d$ (given), and $p_n = 0$, $n \neq 0, 2$. The generating function associated with the PCR process is $g(s) = (1 - p_d) + p_d s^2$, and therefore the reproductive ratio is $R_0 = g'(1) = 2p_d$.

Using Y_t to denote the number of copies of a given strand at time step t, we now consider the question of estimating the number of strands in the original probe, Y_0, given estimates for the number of strands in two successive probes $\tilde{Y}_t$, and $\tilde{Y}_{t+1}$ for t large (typically $10 \leq t \leq 20$). Here, the tilde indicates that we do not know the precise number of strands Y_t and Y_{t+1}, only estimates based on fluorescence levels.

In the Galton–Watson process (above), we assumed a single parent at time $t = 0$. If we modify the analysis to allow for Y_0 parents at time $t = 0$, the equation for the expected

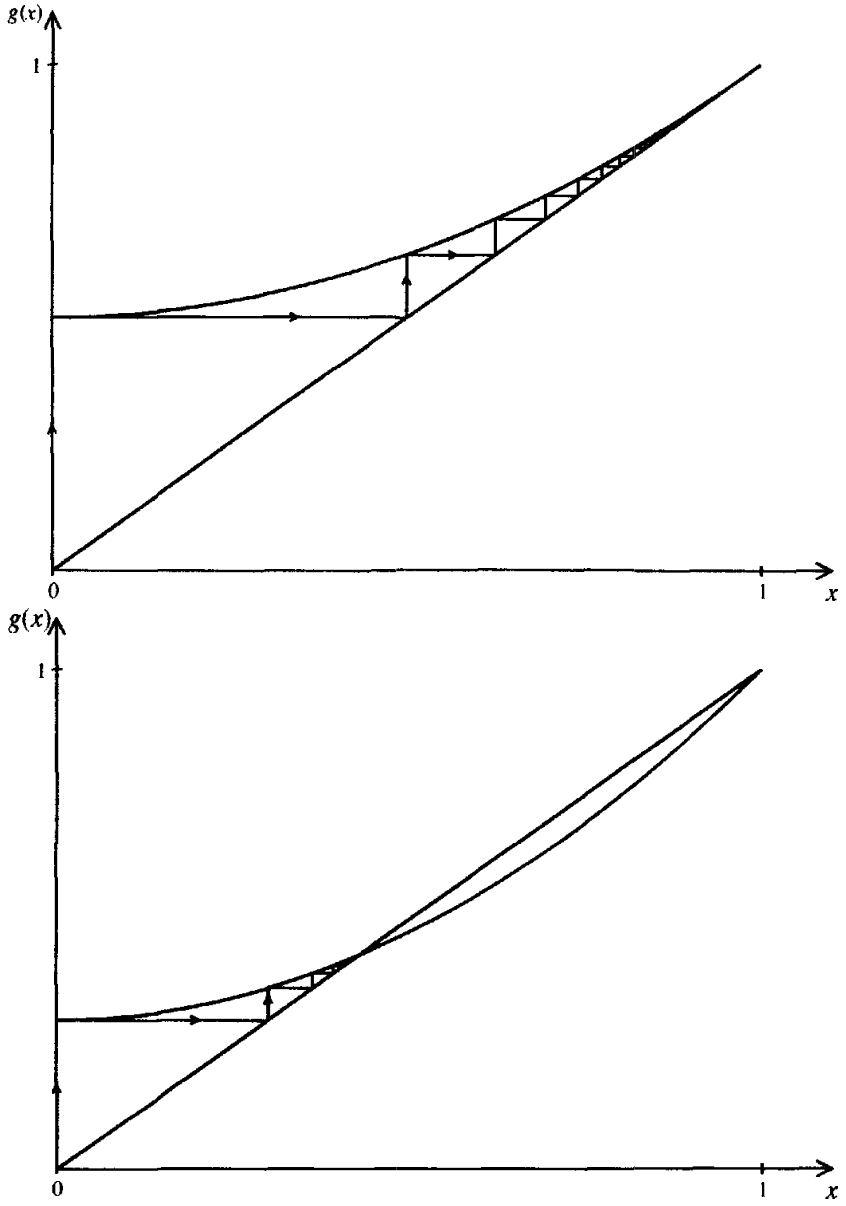

Figure 5.6. *Sketch of the extinction probability for the subcritical case (top) and the supercritical case (bottom). Cobwebbing tells us how the probability that the lineage goes extinct changes with time (equation* (5.47)*). x_t indicates the probability of being extinct after t generations. The initial condition is $x_0 = 0$. Top: In the subcritical case ($R_0 < 1$), extinction is certain. Here, cobwebbing shows $x_t \to 1$. Bottom: In the supercritical case ($R_0 > 1$), extinction is possible, but not inevitable. Here, cobwebbing shows $x_t \to x^* \neq 1$.*

number of strands after t generations becomes $E(Y_t) = R_0^t Y_0$ and so $R_0 = E(Y_{t+1})/E(Y_t)$. We define the estimator for the reproductive ratio, $\hat{R}_0$, as follows:

$$\hat{R}_0 = \frac{\tilde{Y}_{t+1}}{\tilde{Y}_t}. \tag{5.48}$$

Knowing R_0 and t, we find $Y_0 = E(Y_t)/R_0^t$. This yields a simple estimator for the original

number of strands, namely,

$$\hat{Y}_0 = \frac{\tilde{Y}_t}{\hat{R}_0^t} = \tilde{Y}_t \left(\frac{\tilde{Y}_{t+1}}{\tilde{Y}_t} \right)^{-t} = \frac{\tilde{Y}_t^{t+1}}{\tilde{Y}_{t+1}^t}. \tag{5.49}$$

Of course, each time the PCR experiment is repeated, this estimator will give slightly different values for $\hat{Y}_0$. This is because the $\tilde{Y}_t$ values will vary between replicates. It is possible to analyze the variance of the estimator (5.49) by using simulations and other methods and to calculate confidence intervals, although we do not pursue this here.

5.6 Linear Birth and Death Process

Populations are subject to two primary types of stochasticity. *Environmental stochasticity* refers to variation and uncertainty in the environmental conditions in which a population finds itself. These conditions include effects of temperature, rainfall, competition from other species, and so forth. *Demographic stochasticity* refers to variation and uncertainty arising from the unpredictable behavior of the individuals that make up a population. It is relevant when population sizes are small (e.g., fewer than 25). Here, populations with a positive net growth rate can still go extinct due to a "run of bad luck," where insufficient individuals reproduce before they die. In this section, we consider how to model demographic stochasticity in continuous time using a linear birth and death model. Here it is assumed that individuals act independently from one another, so there are no nonlinear interaction terms in the equations. We derive formulae for the mean and variance of a population undergoing stochastic birth/death, and calculate the probability of extinction. We start by considering a pure birth process in Section 5.6.1 and include death in Section 5.6.2.

5.6.1 Pure Birth Process

To start, we ignore death and consider a *pure birth* process where individuals give birth at rate b. In the absence of demographic stochasticity, the underlying model equation would be

$$\frac{dn}{dt} = bn, \qquad n(0) = n_0, \tag{5.50}$$

where $n(t)$ is the number of individuals in the population at time t, and n_0 is the number of individuals at time $t = 0$. This equation has the solution $n(t) = n_0 \exp(bt)$.

We define the underlying stochastic process by

$$N(t) = \text{number of individuals at time } t \text{ (random variable)},$$
$$p_n(t) = \Pr\{N(t) = n\}, \; n = 0, 1, 2, \ldots.$$

We assume that the birth event is a *Poisson process*, namely, that the probability of the event occurring in a short period of time τ is proportional to τ, and the probability of two events occurring during the short period of time is $o(\tau)$.

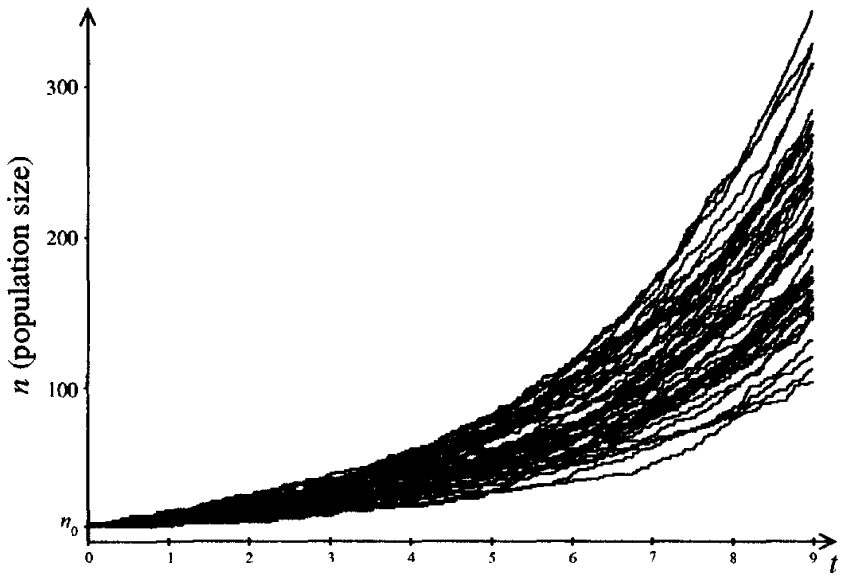

Figure 5.7. *Stochastic simulation of the pure birth process given in* (5.51). *Parameters are* $b = 1/3$ *and* $n_0 = 10$. *Fifty different trajectories are given. They differ only in the seed for the random number generator used.*

For one individual, we have

$$\begin{aligned}\Pr\{1 \text{ birth in } [t, t+\tau)\} &= b\tau + o(\tau),\\ \Pr\{>1 \text{ birth in } [t, t+\tau)\} &= o(\tau),\\ \Pr\{0 \text{ births in } [t, t+\tau)\} &= 1 - b\tau + o(\tau).\end{aligned}$$

For n individuals, we have

$$\begin{aligned}\Pr\{1 \text{ birth in } [t, t+\tau)\} &= nb\tau(1-b\tau)^{n-1} = nb\tau + o(\tau),\\ \Pr\{m \text{ births in } [t, t+\tau)\} &= \binom{n}{m}(b\tau)^m(1-b\tau)^{n-m} = o(\tau),\ 1 < m \le n,\\ \Pr\{0 \text{ births in } [t, t+\tau)\} &= 1 - nb\tau + o(\tau).\end{aligned} \tag{5.51}$$

Figure 5.7 shows a stochastic simulation of this process.

To translate this stochastic process into a differential equation, we require a *master equation* that relates probabilities at different time steps,

$$\begin{aligned}p_n(t+\tau) &= p_{n-1}(t)\cdot \Pr\{1 \text{ birth in } [t, t+\tau)\} + p_n(t)\cdot \Pr\{0 \text{ births in } [t, t+\tau)\}\\ &= p_{n-1}(t)(n-1)b\tau + p_n(t)(1-bn\tau) + o(\tau).\end{aligned}$$

After rearranging, we can write

$$\frac{p_n(t+\tau) - p_n(t)}{\tau} = b\{(n-1)p_{n-1}(t) - np_n(t)\}.$$

As $\tau \to 0$, we obtain an *infinite* system of ODEs,

$$\frac{d}{dt}p_n(t) = b\{(n-1)p_{n-1}(t) - np_n(t)\}, \quad n = n_0, n_0+1, n_0+2, n_0+3, \ldots, \quad p_{n_0-1} = 0, \tag{5.52}$$

with initial data, describing n_0 individuals present at time $t = 0$, as follows:

$$p_n(0) = \begin{cases} 1 & \text{if } n = n_0, \\ 0 & \text{otherwise.} \end{cases} \tag{5.53}$$

It is possible to solve this system exactly (see [102] or Exercise 5.8.9). However, we will focus on calculating the mean M_1 and variance $\sigma^2 = M_2 - M_1^2$ of $N(t)$ (see (5.21) and (5.23)), by first deriving differential equations for M_1 and $M_2 - M_1$:

$$\begin{aligned} \frac{dM_1}{dt} = \sum_{n=1}^{\infty} n\dot{p}_n &= \sum_{n=1}^{\infty} bn\{(n-1)p_{n-1} - np_n\} \\ &= b\sum_{n=0}^{\infty}\{(n+1)np_n - n^2p_n\} \\ &= b\sum_{n=1}^{\infty} np_n \\ &= bM_1. \end{aligned} \tag{5.54}$$

Together with the initial condition $M_1(0) = n_0$, equation (5.54) has the solution

$$M_1(t) = n_0 e^{bt},$$

with the same solution as given by the linear deterministic model (5.50).

To find a differential equation for the variance σ^2 we consider first $M_2 - M_1 = E(X(X-1))$ (5.25) and obtain

$$\begin{aligned} \frac{d(M_2 - M_1)}{dt} = \frac{d}{dt}\sum_{n=1}^{\infty} n(n-1)p_n &= b\sum_{n=1}^{\infty} n(n-1)\{(n-1)p_{n-1} - np_n\} \\ &= b\sum_{n=1}^{\infty}\{(n+1)n^2p_n - (n-1)n^2p_n\} \\ &= 2b\sum_{n=1}^{\infty} n^2p_n. \end{aligned}$$

Hence $\frac{d}{dt}\{M_2 - M_1\} = 2bM_2$. Using (5.54), this can be rewritten as

$$\frac{dM_2}{dt} = 2bM_2 + bM_1. \tag{5.55}$$

Thus, the differential equation for the variance is

$$\begin{aligned} \frac{d\sigma^2}{dt} = \frac{d}{dt}(M_2 - M_1^2) &= \frac{dM_2}{dt} - 2M_1\frac{dM_1}{dt} \\ &= \frac{dM_2}{dt} - 2bM_1^2 \\ &= 2bM_2 + bM_1 - 2bM_1^2 \\ &= 2b\sigma^2 + bM_1. \end{aligned}$$

The initial condition (5.53) yields an initial variance of zero, $\sigma^2(0) = 0$. The above equation, with $\sigma^2(0) = 0$ and $M_1(0) = n_0$, can be solved with an integrating factor to yield

$$\sigma^2(t) = n_0 e^{bt}\left(e^{bt} - 1\right), \tag{5.56}$$

which implies that the variance increases exponentially for large time t. For example, for the simulation shown in Figure 5.7, the values $n_0 = 10$ and $b = 1/3$ yield a final variance of $\sigma^2(9) = 3833$ and standard deviation of $\sigma(9) = 61.9$ by the final time $t = 9$.

In the pure birth process, the probability of the population going extinct is zero, because there is no death included in the model. To include death, we consider a simple birth and death process in the next section.

5.6.2 Birth and Death Process

In the previous section, we neglected death of individuals, but could derive a simple system that can be solved explicitly (see Exercise 5.8.9) and whose mean and variance can be calculated in a straightforward way. Death cannot be ignored in realistic biological models. When it is included, the model becomes only slightly more complex, but the analysis of the model becomes considerably more challenging.

When we extend the analysis from the previous section to a population of individuals that give birth at a rate b and die at a rate d, the transitions for n individuals in a time step of length τ become

$$\begin{aligned} \Pr\{1 \text{ birth in } [t, t+\tau)\} &= nb\,\tau + o(\tau), \\ \Pr\{1 \text{ death in } [t, t+\tau)\} &= nd\,\tau + o(\tau), \\ \Pr\{\text{no change in } [t, t+\tau)\} &= 1 - n(b+d)\tau + o(\tau), \end{aligned} \tag{5.57}$$

and the probability of having more than one birth or death in the time step is $o(\tau)$. Figure 5.8 shows these transitions diagrammatically, and Figure 5.9 shows a stochastic simulation of the process.

The master equation for this process is

$$\begin{aligned} p_n(t+\tau) = {} & (n-1)\,b\tau\; p_{n-1}(t) \\ & + (n+1)\,d\tau\; p_{n+1}(t) + (1 - n\tau\;(b+d))\,p_n(t) + o(\tau), \end{aligned}$$

and the corresponding differential equation is

$$\frac{dp_n}{dt} = (n-1)\,b\;p_{n-1} + (n+1)\,d\;p_{n+1} - (b+d)\,n\;p_n, \quad n = 0, 1, 2, \ldots, \quad p_{-1} = 0, \tag{5.58}$$

with initial condition

$$p_n(0) = \begin{cases} 1 & \text{if } n = n_0, \\ 0 & \text{otherwise.} \end{cases}$$

For the simple birth process, $p_n(t)$ depends only upon $p_n(t)$ and the preceding $p_{n-1}(t)$. In the birth-death process, $p_n(t)$ depends not only on $p_n(t)$ and $p_{n-1}(t)$, but also on the as yet unknown $p_{n+1}(t)$.

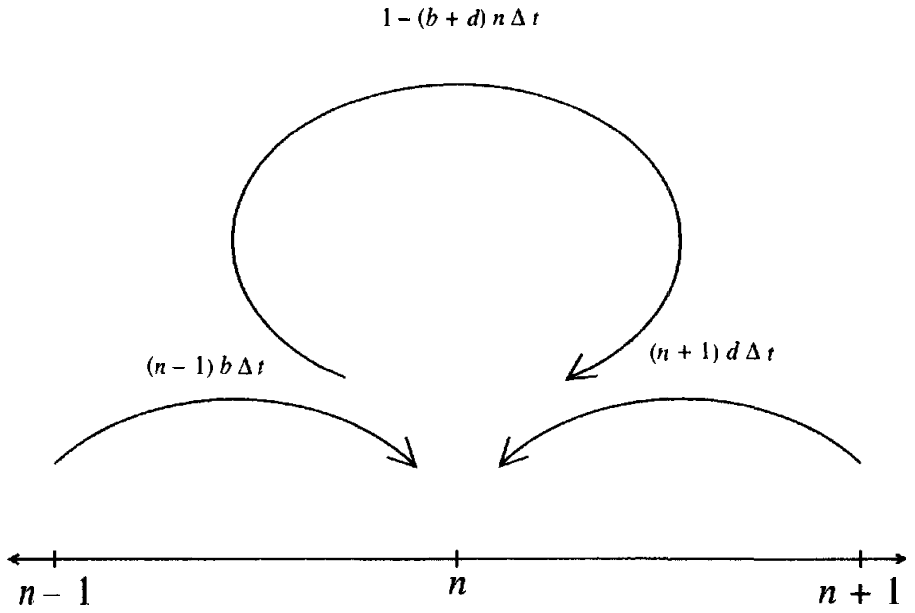

Figure 5.8. *The linear birth-death process can be described by transitions between the natural numbers. Here, in a small time step τ, a population of size n can arise because of single birth in a population of size $n-1$, because of single death in a population of size $n+1$, or because of no change in the population size. These transitions are shown in* (5.57).

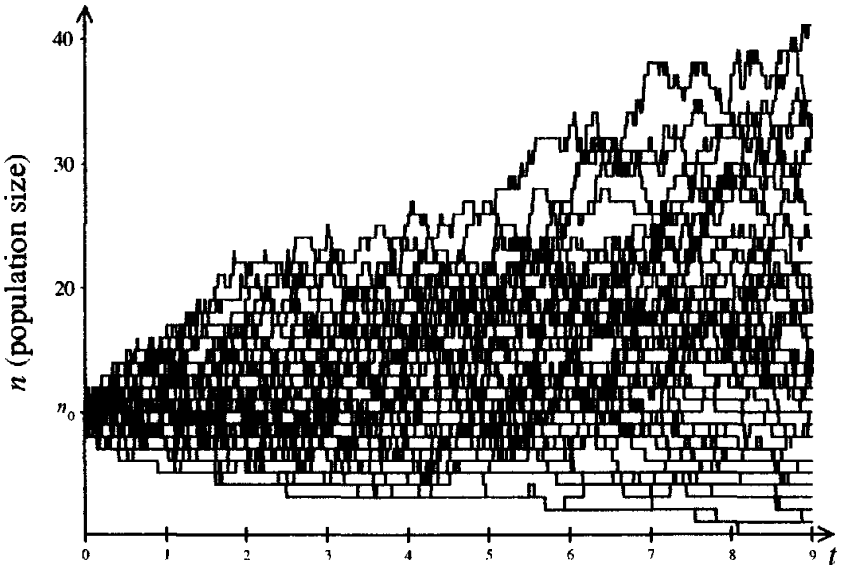

Figure 5.9. *Stochastic simulation of the birth and death process given in* (5.57). *Parameters are $b = 1/3$, $d = 0.3$, and $n_0 = 10$. Fifty different trajectories are given. They differ only in the seed for the random number generator used.*

To get around this problem, we solve for all p_n's at once using the probability generating function,

$$g(s,t) = \sum_{n=0}^{\infty} s^n p_n(t), \qquad g(s,0) = s^{n_0}.$$

Here, the original definition of the generating function (5.28) has been extended to allow for time-dependency so that the function g depends upon t as well as s. The probability of extinction at time t can be calculated directly from g:

$$p_0(t) = \Pr\{\text{extinction by time } t\} = g(0, t). \tag{5.59}$$

The formulae for probabilities, mean, and variance follow from (5.30), (5.31), and (5.33):

$$p_n(t) = \frac{1}{n!} \frac{\partial^n g}{\partial s^n}\bigg|_{s=0}, \tag{5.60}$$

$$g_s(1, t) = \sum_{n=0}^{\infty} n\ p_n(t) = E(N(t)) = M_1(t), \tag{5.61}$$

$$g_{ss}(1, t) = \sum_{n=0}^{\infty} n(n-1)\ p_n(t) = M_2(t) - M_1(t), \tag{5.62}$$

$$\begin{aligned} \operatorname{var}(N(t)) &= M_2(t) - M_1(t)^2 \\ &= \left[\ g_{ss} + g_s - g_s^2\ \right]_{s=1}. \end{aligned} \tag{5.63}$$

To calculate $g(s, t)$, we first derive a PDE satisfied by g, as follows:

$$\begin{aligned} \frac{\partial g}{\partial t} &= \sum_{n=0}^{\infty} s^n\ \dot{p}_n(t) \\ &= b \sum_{n=0}^{\infty} (n-1)\ s^n\ p_{n-1}(t) + d \sum_{n=0}^{\infty} (n+1)\ s^n\ p_{n+1}(t) - (b+d) \sum_{n=0}^{\infty} n\ s^n\ p_n(t) \\ &= b \sum_{k=1}^{\infty} k\ s^{k+1}\ p_k(t) + d \sum_{k=1}^{\infty} k\ s^{k-1}\ p_k(t) - (b+d) \sum_{k=1}^{\infty} k\ s^k\ p_k(t) \\ &\quad \text{(where } k = n-1,\ k = n+1, \text{ and } k = n, \text{ respectively, for the three sums)} \\ &= (bs - d)(s-1) \sum_{k=1}^{\infty} k\ s^{k-1}\ p_k(t) = (bs - d)(s-1) \frac{\partial g}{\partial s}. \end{aligned}$$

Thus, the generating function satisfies a first-order PDE,

$$\frac{\partial g}{\partial t} - (bs - d)(s-1) \frac{\partial g}{\partial s} = 0, \qquad g(s, 0) = s^{n_0}. \tag{5.64}$$

The solution to this equation can be found using the *method of characteristics*. While this is a very useful method for solving first-order PDEs, it is beyond the scope of this book. Details of the method applied to this equation are given in [102]. The interested reader is encouraged to look up the solution method from [102]. Alternatively, the solution,

$$g = \left[\frac{d - c(s)\ e^{-t(b-d)}}{b - c(s)\ e^{-t(b-d)}} \right]^{n_0}, \quad c(s) = \frac{bs - d}{s - 1}, \quad b \neq d,$$

can be verified by substituting directly into (5.64). The case $b = d$ is covered in detail in [102].

Substitution of the solution $g(s,t)$ into the formula for the mean yields

$$E(N(t)) = g_s(1) = n_0\, e^{(b-d)t} = n_0\, e^{rt}, \qquad r = b-d,$$

as predicted by the deterministic model. The variance is

$$\text{var}(N(t)) = \left[\, g_{ss} + g_s - g_s^2 \,\right]_{s=1} = \frac{n_0\,(b+d)}{r}\, e^{rt}\,(e^{rt}-1).$$

Growth or decay of the variance over time depends upon the sign of $r = b-d$: the variance grows if $b > d$ and decays if $b < d$. The variance also scales with the birth and death rates. It increases with b and d, even if r is held constant.

The probability of extinction is

$$p_0(t) = g(0,t) = \left[\frac{d(1-e^{-rt})}{b-de^{-rt}}\right]^{n_0}.$$

For example, for the simulation shown in Figure 5.9, values of $b = 1/3$, $d = 3/10$, and $n_0 = 10$ give $p_0(9) = 0.028$. Thus, approximately 1 out of the 50 simulations should have gone extinct by time 9. If we are interested in whether the population eventually goes extinct, we must consider two cases, namely, $b > d$ ($r > 0$) and $b < d$ ($r < 0$). If $b > d$ ($r > 0$), then

$$\lim_{t\to\infty} p_0(t) = \left[\frac{d}{b}\right]^{n_0},$$

which implies that the probability of extinction is greater than zero, even though the birth rate is greater than the death rate. For example, for the simulation shown in Figure 5.9, the probability of eventual extinction is $(d/b)^{n_0} = 0.34$, so approximately 17 out of the 50 simulations in Figure 5.9 should eventually go extinct. On the other hand, if $b < d$ ($r < 0$), then

$$\lim_{t\to\infty} p_0(t) = 1,$$

which implies certain extinction.

5.7 Nonlinear Birth-Death Process

In this section, we develop techniques for dealing with a nonlinear birth-death process. We already know *linear* birth-death processes, where individuals act independently of each other. For many systems in biology, this hypothesis is not completely appropriate: a growing population eventually reaches the limits of the carrying capacity of the ecosystem (individuals start to compete for resources); an infection cannot grow exponentially (eventually all susceptibles are infected). If the correlations between individuals are strong, then a linear model is no longer appropriate and the dependence between the individuals must be taken into account.

In the following two sections, we model the common cold in households as a nonlinear birth-death process. Development of the model is given in Section 5.7.1, and analysis of the model as an embedded discrete-time Markov process is given in Section 5.7.2.