tion error will decrease and then increase as model complexity increases—with respect to reliability of prediction, there is an optimal level of model complexity.

Linhart and Zucchini's approach is consistent with almost all quantitative work in this area that suggested the optimal model size is much smaller than intuition dictates. Ludwig and Walters (1985) obtained better predictions about management actions from a non-age-structured model, even when the data were derived, by simulation, from an age-structured model. That is, the "wrong" model can do better than the "right" model in prediction if parameters must be estimated. Similarly, Punt (1988) found very simple models of fisheries management, which often ignored substantial amounts of data, outperformed more complex models when parameters had to be estimated and decisions made.

When the objective is something other than prediction accuracy, the complexity of the optimal mode may be quite different. In Chapter 10, we show a fisheries example where a complex model fits the available data no better than a simpler model. However, the uncertainty in the sustainable harvest is quite low for the simple model, but high for the complex model. In this case the simple model under-represents the uncertainty, and we believe that a more complex model provides a better representation of the uncertainty.

The complexity of the optimal model will depend on the use of the model and on the data. Part of the work of the ecological detective is to iterate between alternative models, to understand their strengths and weaknesses, and to recognize that the most appropriate model will change from application to application.

Probability and Probability Models: Know Your Data

DESCRIPTIONS OF RANDOMNESS

The data we encounter in ecological settings involve different kinds of randomness. Many ecological models describe only the average, or modal, value of a parameter, but when we compare models to data, we need methods for determining the probability of individual observations, given a specific model and a value for the mean or mode of the parameter. This requires that we describe the randomness in the data. Similarly, when we build a model and want to generate a distribution of some characteristic, we first need a way to quantify the probability distribution associated with this characteristic. This involves understanding both the nature of your data and the appropriate probabilistic descriptions.

We assume that readers of this book are familiar with the normal or Gaussian distribution (the familiar "bell-shaped curve"). However, many of the distributions in nature are not normal. The purpose of this chapter is to introduce ideas about probability, describe a wide range of useful probability distributions (and consider biological processes that give rise to these distributions), and provide you with the tools you need to use these distributions in your work. We begin with advice on data and then review the concepts of probability. After that, we describe a number of different probability distributions and some of their ecological applications. We close with a description and illustration of the "Monte Carlo" method for generating data and testing models.

A modest university library will have fifty to one hundred textbooks on probability that cover the material we treat here in more detail. So why do we bother? There are two main reasons. First, we want to motivate you to be interested in other than normal distributions. Second, we want to provide enough detail so that when the distributions are used in subsequent applications, the book is self-contained. We suggest that you skim the distributional information now and return to it as needed in later chapters.

ALWAYS PLOT YOUR DATA

Ecological systems are complex. For this reason, we can hope to observe only a very small fraction of the possible variables. The largest field research programs barely scratch the surface of what could be measured. Indeed, the key questions in the design of ecological research are what experiments to perform, what to measure, and how to measure it. Whole new avenues of research have been developed based on new measurement methodologies such as radiotracking, starch gel electrophoresis, DNA fingerprinting, and individual identification of animals by natural marks.

When confronting alternative models with data, we must decide not only which models, but also which data to use. In practice we often observe more than one feature of the ecological system. For example, population surveys may be conducted in many different years, and these surveys provide the major source of information for the model. However, in some years there may be additional direct measurements of birth or death rates.

So what is the first step? Plot your data. Get to know them by using standard computer graphic routines to fit various curves (linear, polynomial, logarithmic, exponential). When there are more than two variables, plot the data in many ways and look for correlation. Think about plausible functional relationships.

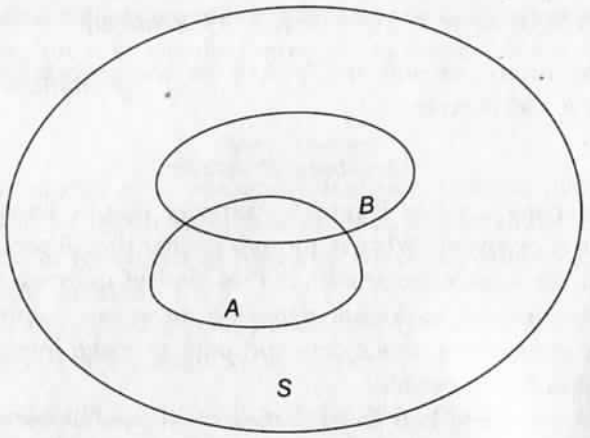

FIGURE 3.1. The probability of the event A is the area of A, however area might be defined, divided by the area of S, which is the collection of all possible outcomes of the experiment.

EXPERIMENTS, EVENTS, AND PROBABILITY

In probability theory, we are concerned with the occurrence of "events" that can be thought of as outcomes of experiments. The probability of an event A is denoted by

$$\Pr\{A\} = \text{probability that the event } A \text{ occurs.} \qquad (3.1)$$

It is helpful to think of probability in the following way. First, we imagine all the possible outcomes of the experiment and call this collection of outcomes S. A smaller collection of outcomes, A, has probability defined as the "area" of A divided by the "area" of S, with "area" suitably defined (Figure 3.1). Particular probability models give different definitions of what "area of A" really means. In any case,

$$\begin{aligned}\Pr\{A\} &= \text{probability that the event } A \text{ occurs} \\ &= (\text{area of } A)/(\text{area of } S). \qquad (3.2)\end{aligned}$$

Continuing to use this figure and the definition of probability in Equation 3.2, we see that the probability that one of two events A or B occurs is

$$\Pr\{A \text{ or } B\} = \Pr\{A\} + \Pr\{B\} - \Pr\{A \text{ and } B\}. \quad (3.3)$$

In the future, we will use $\Pr\{A,B\}$ for the probability that both A and B occur.

Conditional Probability

Referring again to Figure 3.1, suppose that we know that event A occurred. What is the probability that B occurred, given the knowledge about A? This kind of question arises all the time in ecological detection as we use models to make predictions about data and data to make inferences about different models.

If A occurred, then the collection of all possible outcomes of the experiment is no longer S, but must be A. From the definition Equation 3.2,

$$\Pr\{B \text{ occurred, given that } A \text{ occurred}\} = (\text{area common to } A \text{ and } B)/(\text{area of } A). \quad (3.4)$$

We use $\Pr\{B|A\}$ to denote the probability that B occurs given that A occurs. Dividing the numerator and denominator of the right-hand side of Equation 3.4 by the area of S and using the new notation, we have

$$\Pr\{B|A\} = \Pr\{A,B\}/\Pr\{A\}. \quad (3.5)$$

By analogy, since A and B are fully interchangeable here, we must also have

$$\Pr\{A|B\} = \Pr\{A,B\}/\Pr\{B\}. \quad (3.6)$$

We define two events as independent if knowing that one of them occurred does nothing to change our idea about the probability of the other one occurring. Thus, if A and B are independent,

$$\Pr\{A|B\} = \Pr\{A\} \quad \text{and} \quad \Pr\{B|A\} = \Pr\{B\}. \quad (3.7)$$

Using these in either Equation 3.5 or Equation 3.6, we see that for independent events

$$\Pr\{A,B\} = \Pr\{A\}\Pr\{B\}. \quad (3.8)$$

Equation 3.8 is often given as the definition of independent events, but it is actually derived from the definition based on conditioning.

Bayes' Theorem

The challenge in ecological detection (and all statistical science, for that matter) is to determine how to use the information contained in data and Bayes' theorem is a very powerful method.

From Equation 3.6, we see that $\Pr\{A,B\} = \Pr\{A|B\}\Pr\{B\}$. Using this in Equation 3.5, we have

$$\Pr\{B|A\} = \Pr\{A,B\}/\Pr\{A\} = \Pr\{A|B\}\Pr\{B\}/\Pr\{A\}. \quad (3.9)$$

The extreme left- and right-hand sides of this formula are called Bayes' theorem. It is most handy when there are a number of possible but mutually exclusive outcomes $B_1, B_2, \ldots, B_N$, one of which must occur when A occurs. The natural generalization of Equation 3.9 is to ask for the probability that B_i occurs given that A occurs (Figure 3.2). Following the reasoning that led to Equation 3.9, you should show that

$$\Pr\{B_i|A\} = \Pr\{A|B_i\}\Pr\{B_i\} \Big/ \sum_{j=1}^{N} \Pr\{A|B_j\}\Pr\{B_j\}. \quad (3.10)$$

Two hints: note that (1) the numerator on the right-hand side is the joint probability A and B_i, and (2) the denominator is the same as $\Sigma_{j=1}^{N} \Pr\{A,B_j\}$. What must be true about this expression?

We now illustrate some of the nuances of conditional probability with two examples (Bar-Hillel and Falk 1982).

Predator and Prey. Imagine a rabbit wandering through the forest. If it comes within a critical distance of a predator (e.g., a fox or coyote), there is a probability P_A that the predator will attack. In addition, suppose that the rabbit often does not observe the predator directly, but uses various

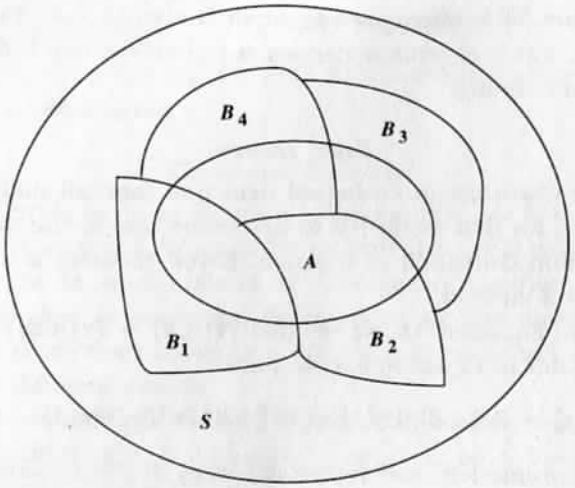

FIGURE 3.2. An illustration of Bayes' theorem for a case in which, when event A occurs, one of four other possible events $B_1, \ldots, B_4$ may also occur.

cues (e.g., scent) of the predator's presence. Assume that P_s is the probability that if the rabbit obtains such a signal, the predator is within the critical attack distance. Once the rabbit obtains such a signal, what is the probability of an attack? The answer is not $P_s P_A$, as tempting as it may seem.

In order to answer the question, we introduce events:

$$\begin{aligned} A &= \text{event of being attacked,} \\ P &= \text{event of predator present within the critical} \\ &\quad \text{attack distance,} \\ S &= \text{event of receiving the cue,} \end{aligned} \tag{3.11}$$

so that the data are

$$\begin{aligned} \Pr\{A|P\} &= P_A, \\ \Pr\{P|S\} &= P_s. \end{aligned} \tag{3.12}$$

The probability we wish to calculate is

$$\begin{aligned} \Pr\{\text{attack, given the signal}\} &= \Pr\{A|S\} \\ &= \Pr\{A,S\}/\Pr\{S\}. \end{aligned} \tag{3.13}$$

Applying Bayes' theorem,

$$\Pr\{A|S\} = \Pr\{A,S\}/\Pr\{S\} = \Pr\{S|A\}\,\Pr\{A\}/\Pr\{S\}. \tag{3.14}$$

The key piece of information in this equation is $\Pr\{S|A\}$, the probability that a signal is obtained when an attack actually does occur. This is not available from the given data and (particularly if the predator is smart) could, in fact, be 0! Thus, if the rabbit is a careless Bayesian, it may misjudge the meaning of a cue.

Smith's Children (Bar-Hillel and Falk, 1982). Smith has two children. You meet Smith and a child who is a boy. What is the probability that the other child is also a boy?

There are two lines of reasoning about this problem. If the sexes of the children are determined independently and with equal probability, then by independence

$$\begin{aligned} &\Pr\{\text{second child is a boy} \mid \text{first child is a boy}\} \\ &\qquad = \Pr\{\text{second child is a boy}\} = 1/2. \end{aligned} \tag{3.15}$$

The second line of reasoning is the following. Before meeting the first child, the possible events in Smith's family are $\{GG,GB,BG,BB\}$, where G denotes girl and B denotes boy. The information that the child we met is a boy eliminates GG as one of the possible events, so that given this information, the possible events are $\{GB,BG,BB\}$. With this line of reasoning, if each family mix is equally likely, the probability that the second child is a boy is $1/3$.

Clearly, these two lines of reasoning cannot be correct. One approach is to forget about the problem, since "Both arguments appear reasonable and both have been used in practice. What to do about the contradiction? The easiest way out is that of a formalist, who refuses to see a problem if it is not formulated in an impeccable manner. *But problems are not solved by ignoring them.*" (Feller 1971, 12, emphasis added.)

The difficulty lies in how we use the information that one of the children is a boy. We want to find

$$\Pr\{\text{family type is } BB \mid \text{met child is a boy}\} = \Pr\{\text{family type is } BB, \text{ met child is a boy}\}/\Pr\{\text{met child is a boy}\}. \quad (3.16)$$

Allowing all four possible family types, we have:

Family type	Prior probability	Pr{meeting a boy, given family type}
BB	1/4	1
BG	1/4	1/2
GB	1/4	1/2
GG	1/4	0

Assuming independence of the met child and the family type, the joint probability of family type and meeting a boy is

$$\Pr\{\text{family type is } BB, \text{ met child is a boy}\} = (1/4) \times 1 = 1/4,$$
$$\Pr\{\text{family type is } BG, \text{ met child is a boy}\} = (1/4) \times (1/2) = 1/8,$$
$$\Pr\{\text{family type is } GB, \text{ met child is a boy}\} = (1/4) \times (1/2) = 1/8,$$
$$\Pr\{\text{family type is } GG, \text{ met child is a boy}\} = (1/4) \times 0 = 0,$$

so that we have

$$\Pr\{\text{met child is a boy}\} = 1/4 + 1/8 + 1/8 = 1/2,$$

and using this in Equation 3.16 we conclude that

$$\Pr\{\text{second child is a boy} \mid \text{met child is a boy}\} = 1/2. \quad (3.17)$$

Thus, the first line of reasoning is correct and the second is not. We encourage you to think about what was wrong with the second line of reasoning. In particular, does the fact of

meeting a boy change the probabilities for the four family types?

Random Variables, Distribution Functions, and Density Functions

A random variable Z is one that can take more than one value in which the values are determined by probabilities. If the random variable takes discrete values, we write

$$\Pr\{Z = k\} = f_k, \quad (3.18)$$

where $0 \le f_k \le 1$ and $\Sigma_k f_k = 1$. For example, Z might take the values 1,2, . . . 10, each with equal probability 0.1. Then $f_k = 0.1$ and $\Sigma_{k=1}^{z} f_k$ is the probability that $Z \le z$, which we shall denote by $F(z)$; Figure 3.3 illustrates this idea. $F(z)$ is called the cumulative distribution function. Cumulative distribution functions should have the following properties: (i) as $z \to -\infty$, $F(z) \to 0$; (ii) as $z \to \infty$, $F(z) \to 1$; (iii) $F(z)$ never decreases as z increases.

When the data are continuous variables, such as lengths, weights, or time, we cannot write the probability distributions in the same way since z can take an infinite number of values in any finite interval. In such a case, we begin with the cumulative distribution function, also indicated by $F(z)$ and which has the same interpretation,

$$F(z) = \Pr\{Z \le z\}. \quad (3.19)$$

An example of such a cumulative distribution function is

$$F(z) = \begin{cases} 0 & \text{if } z < 0, \\ 1 - e^{-rz} & \text{if } z \ge 0, \end{cases} \quad (3.20)$$

which is called the "negative exponential distribution function" (Figure 3.4).

When Z is continuous, we can no longer speak of the event "$Z = z$." Instead, we consider the chance that Z takes a value in a small neighborhood Δz of z and we can evaluate it with the following logic (we encourage you to sketch out this idea using Figure 3.4):

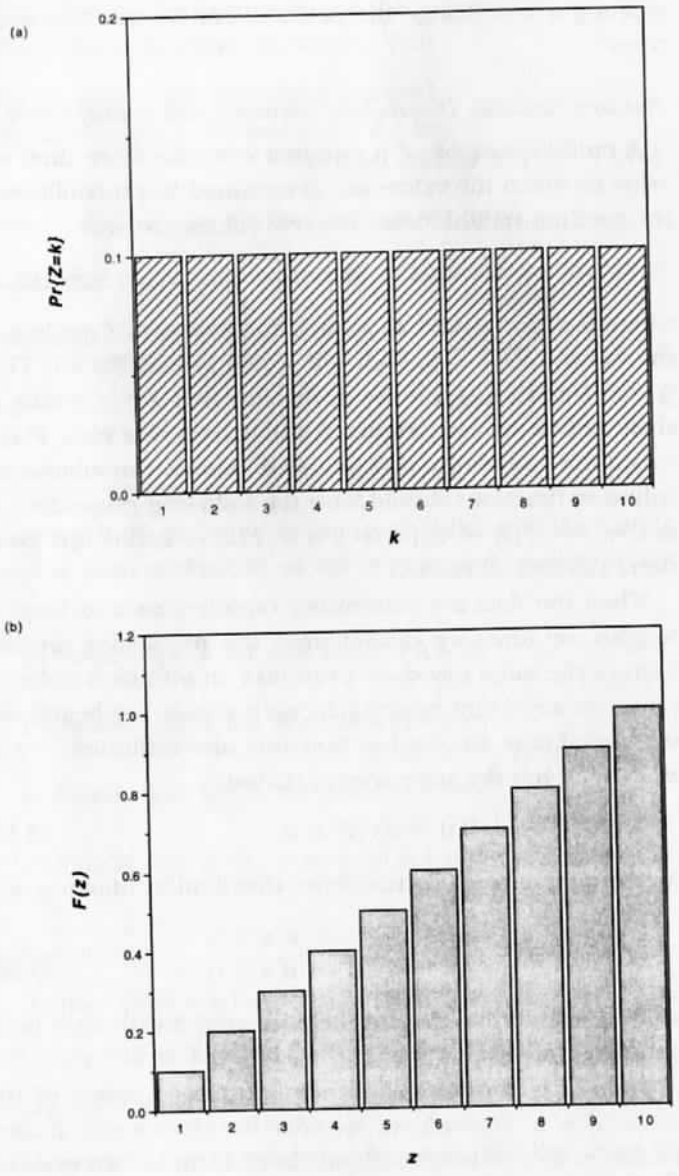

FIGURE 3.3. (a) The random variable Z may be any of 1,2, . . . , 10, each with equal probability $f_k = 0.1$. Such a random variable is said to be uniformly distributed. (b) The probability that Z is less than or equal to z, $F(z)$, is obtained by summing the f_k.

$$\begin{aligned}\Pr\{z \le Z \le z + \Delta z\} &= \Pr\{Z \le z + \Delta z\} - \Pr\{Z \le z\} \\ &= F(z + \Delta z) - F(z). \end{aligned} \tag{3.21}$$

Since Δz is assumed to be a small value, we use a Taylor expansion[1] of $F(z + \Delta z)$

$$F(z+\Delta z) = F(z) + F'(z)\Delta z + \frac{1}{2}F''(z)(\Delta z)^2 + \cdots. \tag{3.22}$$

We scoop all the terms involving high powers of dz into the single expression $o(\Delta z)$. This handy notation will be used in other places in the book. Equation 3.22 becomes

$$F(z + \Delta z) = F(z) + F'(z)\Delta z + o(\Delta z), \tag{3.23}$$

and using this in Equation 3.21,

$$\Pr\{z \le Z \le z + \Delta z\} = F'(z)\Delta z + o(\Delta z). \tag{3.24}$$

The derivative $F'(z)$ is called the probability density function and is denoted by the symbol $f(z)$. For example, a continuous distribution might be used to represent the lengths of animals in a population. When such a graph is drawn using real data, it is often a histogram, where the ordinate is the number of individuals falling in each length interval. When it is represented as a continuous curve, the appropriate label is $f(z)$, which is interpreted as the frequency distribution of outcomes. For the negative exponential distribution function, the probability density function is (Figure 3.4b)

$$f(z) = re^{-rz}. \tag{3.25}$$

These ideas of probability can be nicely illustrated by a study of predation (Box 3.2).

[1]You are going to need six facts from calculus in order to completely understand this chapter. They are given in Box 3.1.

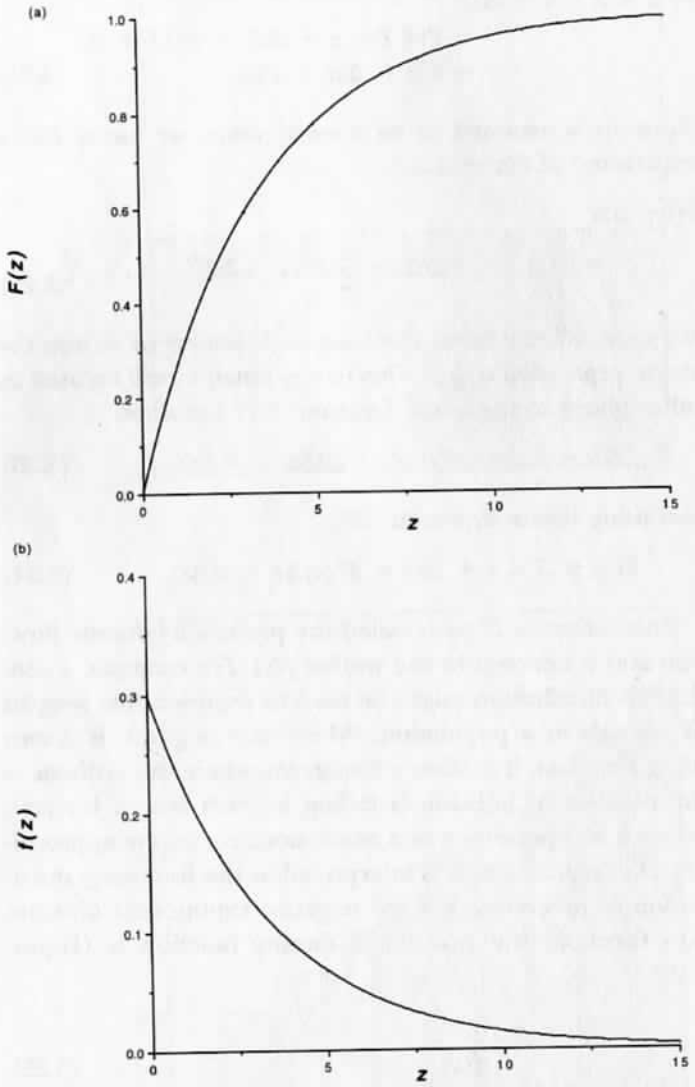

FIGURE 3.4. (a) The negative exponential distribution function $F(z) = 1 - e^{-rz}$ for $r = 0.3$. (b) The corresponding density function $f(z) = re^{-rz}$.

BOX 3.1

THE CALCULUS FACTS YOU NEED FOR THIS CHAPTER

1. Definition of the derivative:

$$\frac{dF}{dx} = F'(x) = \lim_{\Delta x \to 0} \frac{F(x + \Delta x) - F(x)}{\Delta x}.$$

2. Derivative of the exponential function:

$$\frac{d}{dx} e^{kx} = ke^{kx}.$$

3. Exponential function as a limit:

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

4. Integral as a limit of a sum:

$$\int_a h(z)\,dz = \lim_{\Delta z \to 0} \sum_{z=a}^{b} h(z)\,\Delta z,$$

where the summation goes from $z = a$ to $z = b$ in steps of Δz.

5. Taylor expansion for a function of one variable:

$$F(x) = F(a) + F'(a)(x - a) + \frac{1}{2}F''(a)(x - a)^2 + \ldots,$$

where $F'(a)$ is the first derivative of $F(x)$ evaluated at $x = a$, $F''(a)$ is the second derivative of $F(x)$ evaluated at $x = a$ and "+ . . ." means terms that are higher powers of $(x - a)$, such as $(x - a)^3$, $(x - a)^4$, etc.

BOX 3.1 CONT.

Taylor expansion for a function of two variables:

$$F(x,y) = F(a,b) + \frac{\partial F(a,b)}{\partial x}(x-a) + \frac{\partial F(a,b)}{\partial y}(y-b) + \frac{1}{2}\left(\frac{\partial^2 F(a,b)}{\partial x^2}(x-a)^2 + 2\frac{\partial^2 F(a,b)}{\partial x\,\partial y}(x-a)(y-b) + \frac{\partial^2 F(a,b)}{\partial y^2}(y-b)^2 \right),$$

where $\partial F(a,b)/\partial x$ and $\partial F(a,b)/\partial y$ are the first partial derivatives of $F(x,y)$ with respect to x and y, evaluated at $x = a$ and $y = b$; $\partial^2 F(a,b)/\partial x^2$, $\partial^2 F(a,b)/\partial x\,\partial y$, and $\partial^2 F(a,b)/\partial y^2$ are the second partial derivatives with respect to x, with respect to x once and y once, and y, evaluated at $x = a$ and $y = b$.

6. The chain rule:

$$\frac{d}{dx} f(g(x)) = f'(g)g'(x).$$

Expectation, Variance, Standard Deviation, and Coefficient of Variation

We denote average, mean, or expectation by $E\{\ \}$. For a discrete random variable and for any function $g(z)$, we define the expectation by

$$E\{Z\} = \sum_z z f_z \qquad \text{or} \qquad E\{Z\} = \int z f(z)\, dz \tag{3.26}$$

for discrete and continuous random variables, respectively. (Refer again to Box 3.1, for the calculus facts regarding the relationship between sums and integrals.)

BOX 3.2

RANDOM SEARCH AND PREDATION

The rules of probability we just discussed provide some interesting insights into predation. We encourage you to work out all the details of this example, because it will help solidify the notions of probability and the notation we use.

Suppose that an organism searches for food and

$$\Pr\{\text{finding food in the next increment of time } \Delta t \mid \text{no food found thus far}\} = c\Delta t + o(\Delta t),$$

where c is a fixed constant (this will turn out to be very important) and $o(\Delta t)$ represents terms that are higher powers of Δt. We set

$$Q(t) = \Pr\{\text{not finding food in the interval } [0,t]\}$$

and note that for the animal not to find food in the interval $[0,t + \Delta t]$ it first must not find food in the interval $[0,t]$ and then not find food in the next Δt. Assuming that these are independent events (what is the biological implication of this assumption?),

$$Q(t + dt) = Q(t)[1 - c\Delta t + o(\Delta t)].$$

Subtracting $Q(t)$ from both sides we have

$$Q(t + \Delta t) - Q(t) = -cQ(t)\,\Delta t + o(\Delta t).$$

Dividing both sides by dt and letting $\Delta t \to 0$ gives the derivative of $Q_0(t)$ on the left-hand side (see Box 3.1). Since $o(\Delta t)$ denotes terms that are like $(\Delta t)^2$, etc., $o(\Delta t)/\Delta t \to 0$ as $\Delta t \to 0$. Thus, the difference equation becomes a differential equation for $Q(t)$:

BOX 3.2 CONT.

$$\frac{dQ}{dt} = -\,c\,Q(t).$$

We see that the derivative of $Q_0(t)$ is a constant times $Q(t)$. This means that $Q(t)$ must be an exponential (see Box 3.1) of the form $Q(t) = Ae^{-ct}$. Since $Q(0) = 1$ (no food is found before the start of the search for food), the constant $A = 1$. We have demonstrated that

$$\Pr\{\text{not finding food in } [0,t]\} = Q(t) = e^{-ct}.$$

Koopman (1980) derives this formula in a different way in which the biological interpretation of c becomes more apparent. Suppose that the search takes place in a "large" region of area $\mathcal{A}$ that contains the food item. Assume that W is the detection width of the searching animal, in the sense that if the food is within a distance $W/2$ of the animal, the food is discovered. If v is the speed of the searching animal, in the interval of time dt the animal covers an area $Wv\Delta t$ and detects the food with probability $Wv\Delta t/\mathcal{A}$. Envision the time interval $[0,t]$ divided into n legs of length t/n, so that $\Delta t = 1/n$. Assuming that detection on each leg is independent of previous legs gives

$$\begin{aligned}\Pr\{\text{no detection of food in } [0,t]\} &= [\Pr\{\text{no detection of food on a single leg}\}]^n \\ &= \left(1 - \frac{Wvt}{\mathcal{A}n}\right)^n.\end{aligned}$$

In the limit (see Box 3.1) that $n \to \infty$, the right-hand side of this expression becomes $e^{-Wvt/\mathcal{A}}$:

$$\Pr\{\text{no detection of food in } [0,t]\} = e^{-Wvt/\mathcal{A}},$$

BOX 3.2 CONT.

so that the interpretation of c is c = detection rate = $Wv/\mathcal{A}$, and these parameters—W, v, and $\mathcal{A}$—can be measured independently of the searching process. Because $Q(t) = e^{-ct}$ and it is only possible to take the exponential of dimensionless quantities, we conclude that the units of c (which are often denoted by $[c]$) must be 1/time. Since the units of W are length, of v are length/time and of $\mathcal{A}$ are (length)^2, we see that $Wv/\mathcal{A}$ has units of 1/time, as it should if our analysis is correct.

We shall now use notions of conditional probability to demonstrate the "memoryless property" of this model, assuming once again that c is a fixed and certain parameter. We begin with $Q(t) = e^{-ct}$ and ask: What is the probability that the animal does not find food between t and $t + s$, given that it did not find food up to time t? Applying the definition of conditional probability,

$$\Pr\{\text{no food in } (t,t+s) \mid \text{no food in } (0,t)\} = \frac{\Pr\{\text{no food in } (t,t+s) \text{ and no food in } (0,t)\}}{\Pr\{\text{no food in } (0,t)\}}.$$

Since the numerator is the same as no event in the interval from 0 to $t + s$, we have

$$\begin{aligned}\Pr\{\text{no food in } (t,t+s) \mid \text{no food in } (0,t)\} &= e^{-c(t+s)}/e^{-ct} \\ &= e^{-cs} = \Pr\{\text{no food in } (0,s)\}.\end{aligned}$$

Thus, the fact that no food was found before time t provides no information about the probability of events after time t. The predator in this model does not "learn." This is somewhat discomforting, because we expect that a failed search

BOX 3.2 CONT.

should provide information about the search rate c. But remember that we assumed c to be known and fixed. Later in this chapter, after discussing the gamma density, we will consider how failed searches may change our view of the frequency distribution of c, if we allow it to be uncertain.

These definitions generalize for any function $g(Z)$; for example, $E\{g(Z)\} = \Sigma_z\, g(z)f_z$. The generalization is very handy for computing measures of variability about the average. If we denote the average by m_1, the variance of the random variable Z is

$$\begin{aligned}\mathrm{VAR}\{Z\} &= E\{(Z-m_1)^2\} \\ &= \sum_z (z-m_1)^2 f_z \quad \text{or} \quad \int (z-m_1)^2 f(z)\, dz, \end{aligned} \tag{3.27}$$

depending on whether the random variable is discrete or continuous. The variance gives a sense of the "spread" of values of Z around the average.

Two other measures of variability of Z are the standard deviation,

$$\mathrm{SD}\{Z\} = \sqrt{\mathrm{VAR}\{Z\}} \tag{3.28}$$

and the coefficient of variation

$$\mathrm{CV}\{Z\} = \frac{\mathrm{SD}\{Z\}}{E\{Z\}}. \tag{3.29}$$

We are partial to the coefficient of variation as a measure of variation for the following reason. The standard deviation has the same units as Z, so that the coefficient of variation is a dimensionless measure of variability in which the scaling is relative to the mean. To see why this kind of scaling is important, consider the following two sequences of numbers:

A: 45, 32, 12, 23, 26, 27, 39

B: 1040, 1027, 1007, 1018, 1021, 1022, 1034

When asked which sequence is more variable, most people will say that sequence A is more variable. Sequence B is sequence A plus 995, so that the variance of these two sequences is exactly the same. However, the coefficient of variation of sequence B is much smaller than that of A. You should (i) verify that this is true by computation, and (ii) understand the reason for this being true. Some cognitive psychologists have argued that this is a matter of context: "Which series exhibits more variability? Most people answer series A. However, the statistical measure of *variance*—which indicates the amount of irregular variations from the mean of a series of numbers—is *the same* for both series. Series B is simply series A plus a constant. However, intuitive judgments of variability are usually influenced by the size or context of the series or objects. That is subjectively relative variability is more salient than variability per se" (Hogarth 1980, 44).

But when numbers have units, both the magnitude and the variability have meaning. For example, suppose that we measure the weights of five rodents and these are 0.079, 0.120, 0.085, 0.099, and 0.100 kg respectively. The average weight is 0.0966 kg, the variance is 2.018×10^{-4} kg^2 (why kg^2?), and the coefficient of variation is 0.147. If the animals were weighed in grams rather than kilograms, the average would be 96.6 g and the variance 201.84 g^2 but the coefficient of variation would remain the same at about 15%.

By using the coefficient of variation, one takes this comparison out of the realm of the subjective and into the realm of the objective, with a measure of variation that is context-free because it has no dimensions. There is a tradition in ecology, which we elaborate during the discussion of the Poisson distribution, of comparing the mean and variance of data in order to determine whether the subject of

study is "clumped" or not. This can only make sense if the random variable Z is dimensionless.

The Delta Method

When $g(z)$ is nonlinear, $E\{g(Z)\}$ is generally not equal to $g(E\{Z\})$. We encourage you to try out a numerical investigation for $g(z) = z^4$ using both the numerical data in Figure 3.3 and the negative exponential distribution of Figure 3.4. Because $E\{g(Z)\}$ may be difficult to find, an approximation commonly used is the "delta method" (Seber 1980). As before, let $m_1 = E\{Z\}$ and construct a two-term Taylor expansion $g(Z)$ around m_1:

$$g(Z) = g(m_1) + g'(m_1)\,(Z - m_1) + \frac{1}{2}g''(m_1)\,(Z - m_1)^2 + \ldots, \quad (3.30)$$

where, also as before, $g'(m_1)$ and $g''(m_1)$ denote the first and second derivatives of $g(z)$ evaluated at $z = m_1$. Taking the expectation and ignoring all the terms represented by the ellipsis "+ . . . ," we have

$$E\{g(Z)\} = E\{g(m_1)\} + E\{g'(m_1)\,(Z - m_1)\} + \frac{1}{2}E\{g''(m_1)\,(z - m_1)^2\}. \quad (3.31)$$

You should verify from the definition of expectation that for any constant c,

$$E\{c\} = c \quad \text{and} \quad E\{cg(Z)\} = cE\{g(Z)\} \quad (3.32)$$

and that

$$E\{(Z - m_1)\} = 0. \quad (3.33)$$

Since $g(m_1)$, $g'(m_1)$, and $g''(m_1)$ are constants, Equation 3.31 becomes

$$E\{g(Z)\} = g(m_1) + \frac{1}{2}g''(m_1)\,\mathrm{VAR}(Z). \quad (3.34)$$

We prefer to call this the method of "navy math," since it was commonly used by scientists in the Operations Evaluation Group (OEG) (see Tidman 1984) during World War II (Morse 1977) as a quick means of computing expectations. Those scientists and the ones who followed (Mangel 1982) are the inspiration for the part played by Kelly McGillis in *Top Gun*.

PROCESS AND OBSERVATION UNCERTAINTIES

Before discussing particular probability distributions, let us spend time thinking about how stochasticity enters into ecological models. Ecological models often begin with a description of the processes of interest (e.g., birth rates, death rates, migration rates, etc.). For this reason, these models are sometimes called "process models." Uncertainty may enter into these processes because parameters vary in unpredictable ways.

To collect data about an ecological system, we observe it, and there will usually be uncertainty associated with the observations. For instance, suppose that we model a population by

$$N_{t+1} = sN_t + b_t, \quad (3.35)$$

where N_t is the number of animals in the population at the start of period t, s is a survival probability from t to $t + 1$, and b_t is the number of new individuals added in the interval t to $t + 1$.

Uncertainty could enter in a number of different ways. For example, if birth rates fluctuate from one year to the next, we could write

$$N_{t+1} = sN_t + b_t + W_t, \quad (3.36)$$

where W_t represents "process uncertainty," "process stochasticity," "process error," or "process noise" (depending on the particular subfield of ecology, all these terms are used). We use upper case to remind ourselves that W_t is drawn from a distribution; a particular value would be denoted by w_t. In

principle, W_t could arise from a number of the distributions we describe below and could depend on population size.

Since it is likely that there is uncertainty associated with the observations, we describe the observation model as

$$N_{\text{obs},t} = N_t + V_t, \tag{3.37}$$

where $N_{\text{obs},t}$ is the observed population size at time t and the "observation uncertainty" (or any of the other terms) V_t might also depend on population size.

The process and observation models are now combined into a "full" model of the system:

$$\begin{aligned} N_{t+1} &= sN_t + b_t + W_t, \\ N_{\text{obs},t} &= N_t + V_t. \end{aligned} \tag{3.38}$$

To complete the model, we must specify the distributions of W_t and V_t and the initial population size. We shall return to this model at the end of the chapter, once the requisite skills are developed.

Since ecological detection involves comparing different models, it is useful at this point to think about other versions of the observation model.

Bias. Field methods for estimating animal abundance usually involve an unknown bias. For example, not all animals may be seen. In air surveys of marine mammals there is usually an unknown proportion of the animals below the surface. Transect counts of birds or smaller mammals almost always involve a fraction of the animals that cannot be seen from the observer's platform. To account for this effect, we might modify the observation model to

$$N_{\text{obs},t} = qN_t + V_t. \tag{3.39}$$

Here, the parameter q allows for bias of the observation system: When q is less than 1 we tend to undercount the animals, and when q is greater than 1 we tend to overcount them. As before, V_t represents the observation uncertainty. It is almost always helpful, and frequently essential, to do

experiments to determine q. However, in some instances, as in fisheries, we must estimate q from the same data that we use to estimate the parameters of the process model.

Nonlinearity. We generalize the observation model further by including a nonlinear relationship between true abundance and observed abundance:

$$N_{\text{obs},t} = q(N_t)^c + V_t. \tag{3.40}$$

When c is greater than 1 the estimated abundance rises more rapidly than real abundance, and when c is less than 1 the estimated abundance changes less than real abundance.

A Detection Threshold. There may be a minimum threshold population size below which no animals can be seen, such as species where some proportion of the population finds hiding places. In this case, the observation model becomes

$$N_{\text{obs},t} = \max\{a + q(N_t)^c + V_t, 0\}, \tag{3.41}$$

where $\max\{A,B\} = A$ if $A > B$ and $\max\{A,B\} = B$ otherwise. If $A < 0$, it represents the population density below which no animals can be seen. If $A > 0$, some animals will appear to be present even when none are present. This could be due, for example, to improper species identification.

In summary, there is always an observation process interposed between the ecological system and our notebooks. Every effort should be made to understand, calibrate, and model the observation process. Doing this is an essential component of ecological detection.

Additional Data. In some years we may have additional sources of data. For example, suppose that in one year we had also conducted a study that provided an estimate of the number of deaths, in addition to the annual survey of abundance. Our model predicts the number of deaths as

$$D_t = (1 - s)N_t, \tag{3.42}$$

where D_t is the number of deaths in year t. If we assume that the process uncertainty is entirely due to variation in births, then the observation model for deaths is

$$D_{\text{obs},t} = (1 - s)N_t + V_d, \tag{3.43}$$

where V_d is the uncertainty associated with the observation of the number of dead animals. Our model now predicts both the number of animals and the number of deaths, and when we see how well alternative models fit the data, we can compare the predictions with these observations. In later chapters, we will explore how to use multiple observations in a more rigorous framework.

However, we cannot conduct ecological detection without knowledge of the probability distributions that might describe the various kinds of uncertainty. This is what we consider next.

SOME USEFUL PROBABILITY DISTRIBUTIONS

We now provide a review of a number of probability distributions that are tools for the ecological detective. We encourage you to skim this section now and return to it as the distributions are used in subsequent chapters. *However,* whether or not you read it carefully now, you should read the next section on the Monte Carlo method.

This review is not comprehensive. Our goal is to provide enough information so that you will know how to compute Pr{data | model} and Pr{model | data}, which are the essentials for ecological detection. We provide an ecological scenario for most of the probability distributions, to help make them more concrete. Once again, we encourage you to visit the library and find a mathematics or statistics text that deals with elementary probability theory. Our favorite textbook in introductory probability is by Feller (1968).

We describe four distributions (the binomial, multinomial, Poisson, and negative binomial) in which the ran-

dom variables are discrete and observations take only integer values. The binomial distribution is commonly used in mark and recapture studies, where a discrete number of individuals are examined. The Poisson is most often used when dealing with counts of the number of plants or animals per unit time or space, or in the analysis of the number of individuals captured. When the data indicate more variability than is consistent with the Poisson distribution, the negative binomial distribution is more appropriate.

We describe four cases in which the random variable is continuous. The first is the normal or Gaussian distribution, which is the commonly used "bell-shaped curve." It has two parameters: the mean and the standard deviation. The normal distribution is commonly used because of a theorem of probability called "the central limit theorem" (Feller 1968), which asserts that, in general (and there are some ecologically important exceptions), when the sum of a large number of random variables is properly scaled (we shall describe this below), the result is approximately normally distributed. This means, for example, that binomial processes with a large number of trials can be approximated by a normally distributed random variable. The normal distribution is symmetric about the mean, which poses many problems in ecology, because this assigns positive probability to values of the random variable that are less than 0, but often the random variable itself (such as length) will have to be greater than 0.

One solution to this problem is to use the log-normal distribution, in which we replace the assumption that the random variable Z has a normal distribution with the assumption that $\log(Z)$—where log denotes the natural logarithm—has a normal distribution. This distribution has an asymmetric shape with a long tail and the property that values of the associated random variable cannot be less than zero. The chi-square distribution is also based on the normal distribution and arises in the study of the distribution of differences between predictions and data.

TABLE 3.1. Common probability distributions classified according to the nature of the trials and observations.

	Observations	
	Discrete	Continuous
Trials		
Discrete	Binomial	Normal Log-normal Gamma
Continuous	Poisson Negative binomial	—

Finally, we introduce the gamma probability distribution, which is a very flexible continuous distribution that can be used for describing a wide variety of data. It is also an essential component for some of the Bayesian analyses we conduct.

In summary, experiments can involve either discrete or continuous conditions, and the data can be either discrete or continuous (Table 3.1). An overview of these distributions is given in Table 3.2.

The Binomial Distribution

Perhaps the simplest of probability distribution is the binomial distribution with parameters N and p; which we denote by $\mathbf{B}(N,p)$. It arises, for example, in a situation in which an experiment with only two outcomes is repeated N times, and the random variable Z measures the number of times a specified outcome occurs. If p is the chance that the specified outcome occurs in an experiment, then the random variable Z takes integer values ranging from 0 to N according to the rule

$$\Pr\{Z = k\} = p(k,N) = \binom{N}{k} p^k(1-p)^{N-k}. \tag{3.44}$$

TABLE 3.2. Density, mean, and coefficient of variation of the distributions commonly used by the ecological detective.

Distribution	Density	Mean	CV
Binomial	$\Pr\{Z = k\} = \binom{N}{k} p^k(1-p)^{N-k}$	Np	$\sqrt{\frac{1-p}{Np}}$
Poisson	$\Pr\{Z = k\} = \frac{e^{-rt}(rt)^k}{k!}$	rt	$\sqrt{\frac{1}{rt}}$
Negative binomial	$\Pr\{Z = s\} = \frac{\Gamma(k+s)}{\Gamma(k)s!}\left(\frac{k}{k+m}\right)^k\left(\frac{m}{k+m}\right)^s$	m	$\sqrt{\frac{1}{m}+\frac{1}{k}}$
Normal	$f(z) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(z-m)^2}{2\sigma^2}\right)$	m	$\frac{\sigma}{m}$
Gamma	$f(z) = \frac{a^n}{\Gamma(n)} e^{-az} z^{n-1}$	$\frac{n}{a}$	$\frac{1}{\sqrt{n}}$

In this equation

$$\binom{N}{k} = \frac{N!}{k!(N-k)!}.$$

You should know the following facts about the binomial distribution (Feller 1968). The mean and variance are

$$E\{Z\} = \sum_{k=0}^{N} k \Pr\{Z = k\}$$

$$= Np \quad \text{and} \quad \text{VAR}\{Z\} = Np(1-p). \tag{3.45}$$

The coefficient of variation is

$$\text{CV}\{Z\} = \sqrt{\frac{1-p}{Np}}. \tag{3.46}$$

When p is fixed, the coefficient of variation decreases as N increases. This means that the relative variability shown by Z decreases with the number of experiments conducted.

The values of the binomial probability distribution can be computed by an iterative procedure. First, note that

$$p(0,N) = (1-p)^N. \tag{3.47}$$

Then note that $p(k,N)$ and $p(k-1,N)$ can be related as follows:

$$p(k,N) = \binom{N}{k} p^k (1-p)^{N-k}$$

$$= \frac{N!}{k!(N-k)!} p^k (1-p)^{N-k}$$

$$= \frac{N![N-(k-1)]}{k(k-1)![N-(k-1)]!} p(p^{k-1})(1-p)^{N-(k-1)}(1-p)^{-1}$$

$$= \left[\frac{N-k+1}{k}\right]\left[\frac{p}{1-p}\right] p(k-1,N). \tag{3.48}$$

Equations 3.47 and 3.48 can be implemented in the following manner:

Pseudocode 3.1

Step 1. Specify p and N.

Step 2. Find $p(0,N)$ from Equation 3.47.

Step 3. For $k = 1$ to N, find $p(k,N)$ from Equation 3.48 and print out results in a form that you like.

An Ecological Scenario. Sampling for Pests. Suppose that we are sampling fruit for infestations by a pest and know that the chance that a fruit is infested is p. If N fruit are sampled, the probability that k of them are infested is given by the binomial distribution. You should use a program based on this pseudocode to predict the distribution of infested fruit if we sample 10 fruit, and p is 0.1, 0.2, or 0.3.

In most situations, we would not know p, but need to determine it by sampling fruit. How many fruit should be sampled? How do we estimate p from this sample? What confidence can we associate with this estimate? This becomes a problem in ecological detection that we discuss later.

The Multinomial Distribution

The multinomial distribution is the extension of the binomial distribution to a case with more than two possible outcomes of the experiment. For example, suppose that the fruit just described could be infested by more than one kind of pest, but there is only one species of pest per fruit. Then the data would be the number of uninfested fruit, the number infested by pest type 1, the number infested by pest type 2, etc.

Suppose that there are M possible outcomes; we then have a vector of random variables Z_i, where Z_i is the number of times the i^{th} kind of outcome occurred. Instead of Equation 3.44 we now consider

$$\begin{aligned}\Pr\{Z_1 &= k_1, Z_2 = k_2, \ldots, Z_M \\ &= k_M \text{ in } N \text{ experiments}\} \\ &= p(k_1, k_2, \ldots, k_M, N),\end{aligned} \tag{3.49}$$

which is given by

$$p(k_1, k_2, \ldots k_M, N) = \frac{N!}{k_1! k_2! \ldots k_M!} p_1^{k_1} p_2^{k_2} \ldots p_m^{k_m}. \tag{3.50}$$

We encourage you to develop a pseudocode for the multinomial distribution.

The Poisson Distribution

The binomial distribution is one for which the random variable takes discrete values in discrete experiments or trials. In the same way, the Poisson distribution (or Poisson process, to indicate that something is happening over time) is one for which the random variable takes discrete values during continuous sampling (usually area or time; we use time for definiteness). The Poisson distribution can be derived as the limit of a binomial distribution when $N \to \infty$ and $p \to 0$ in such a way that Np is constant (Feller 1968).

If $Z(t)$ has a Poisson distribution, then

$$\Pr\{Z(t) = k\} = \frac{e^{-rt}(rt)^k}{k!}. \tag{3.51}$$

Here r is called the "rate parameter" of the Poisson distribution. You should know the following facts about the Poisson distribution (Feller 1968).

The mean and variance are

$$E\{Z(t)\} = rt \tag{3.52}$$

and

$$\text{VAR}\{Z\} = rt, \tag{3.53}$$

so that the coefficient of variation is

$$\text{CV}\{Z\} = \sqrt{\frac{1}{rt}}. \tag{3.54}$$

Thus, for r fixed, the coefficient of variation decreases as t increases.

The Poisson distribution can be derived from assumptions about what happens in a very small (infinitesimal) amount of time (Feller 1968). Suppose that Δt is a very short time interval. We assume that either nothing happens in this time interval or one event happens, and that the probabilities are

$$\begin{aligned}\Pr\{\text{no event in } \Delta t\} &= e^{-r\Delta t}, \\ \Pr\{\text{exactly one event in } \Delta t\} &= 1 - e^{-r\Delta t}.\end{aligned} \tag{3.55}$$

In probability textbooks one usually finds this written as $\Pr\{\text{more than one event in } \Delta t\} = o(\Delta t)$, where $o(\Delta t)$ is the notation that we introduced earlier denoting terms that are high powers of Δt. Since $e^x = 1 + x + x^2/2 + \ldots$,

$$\begin{aligned}\Pr\{\text{no events in } \Delta t\} &= 1 - r\Delta t + o(\Delta t), \\ \Pr\{\text{one event in } \Delta t\} &= r\Delta t + o(\Delta t).\end{aligned} \tag{3.56}$$

We strongly recommend using Equation 3.55 whenever numerical computation is done, because Equation 3.56 is only an approximation, whereas Equation 3.55 is fundamentally true. For example, regardless of the value of Δt, Equation 3.56 can lead to probabilities that are bigger than 1 or less than 0 if r is big enough; this does not happen with Equation 3.55.

The mean and variance of the Poisson process are equal. Also, note from Equation 3.55 that the chance of an event in the next bit of time depends only on the time interval and not on any history or current state of the system. We saw this previously with the discussion of random search. Thus, there is a tendency to think of the Poisson distribution as representing "randomness." Since the mean and variance are equal, the tradition evolved in ecology to con-

sider the ratio of the variance of the data to the mean of the data. If this is about 1, then the data are considered to be random, and if the ratio is considerably bigger than 1, then the data are considered to be clumped. Such reasoning only works for special kinds of data, because for this to make sense at all, the data must be dimensionless so that the variance-to-mean ratio has no units.

As with the binomial distribution, it is empowering to be able to compute the terms of the Poisson distribution yourself. This can be done by an iterative procedure. Once again, we begin by setting $p(0,t) = e^{-rt}$. Successive terms are then computed by recognizing that

$$p(k,t) = \frac{e^{-rt}(rt)^k}{k!} = \frac{rt}{k}\frac{e^{-rt}(rt)^{k-1}}{(k-1)!} = \left(\frac{rt}{k}\right) p(k-1,t). \tag{3.57}$$

Before we describe the pseudocode, note the following. Unlike the binomial distribution (which has exactly N terms), the Poisson distribution has no limit on the number of terms. Thus, when computing it, you must introduce a cutoff (close to 1), so that when the sum of terms exceeds that cutoff, the computation stops. A pseudocode for this computation is:

Pseudocode 3.2

1. Specify r, t, and the cutoff.
2. Set $p(0,t) = e^{-rt}$. Set sum = $p(0,t)$.
3. Cycle over values of $k \geq 1$ and find $p(k,t)$ from Equation 3.57. Replace sum by sum + $p(k,t)$.
   If the sum is less than the cutoff, return to step 2; otherwise go to step 4.
4. Print out results as you desire.

The Normal or Gaussian Distribution

The two distributions considered thus far involve a random variable Z that takes discrete values. The usual example of a random variable taking continuous values is the normal or Gaussian random variable. We will use the notation $\mathcal{N}(m,\sigma^2)$ to denote a random variable X that is normally distributed with mean m and variance σ^2. We use the symbol X, rather than Z, to remind you that these are names of random variables. As long as you remember that they have specific meanings and biological interpretations, there will be no problem.

We need the following facts about the normal distribution. The distribution function $F(x)$ is

$$F(x) = \Pr\{X \leq x\} = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{x} \exp\left(\frac{-(s-m)^2}{2\sigma^2}\right) ds. \tag{3.58}$$

In this expression, the integration variable s takes all values between $s = -\infty$ and $s = x$. Since it must be true that $\Pr\{-\infty < X < \infty\} = 1$,

$$\frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \exp\left(\frac{-(s-m)^2}{2\sigma^2}\right) ds = 1, \tag{3.59}$$

which means that

$$\int_{-\infty}^{\infty} \exp\left(\frac{-(s-m)^2}{2\sigma^2}\right) ds = \sqrt{2\pi\sigma^2}. \tag{3.60}$$

This is a handy trick for evaluating complicated integrals that are associated with probability functions, and we will use it later.

The normal density function $f(x)$ is

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(x-m)^2}{2\sigma^2}\right). \tag{3.61}$$

The function $f(x)$ is the familiar "bell-shaped curve." Plot it, if it is not completely familiar; vary m and σ to see how they affect the shape.

If X is $\mathcal{N}(m,\sigma^2)$, then the transformed variable $Y = (X - m)/\sigma$ is normally distributed $\mathcal{N}(0,1)$. The distribution function of Y is given the symbol $P_N(y)$:

$$P_N(y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} \exp\left(-\frac{s^2}{2}\right) ds, \tag{3.62}$$

and once again the integration variable ranges from $s = -\infty$ to $s = y$. This function is especially useful. Note that $P_N(0) = 1/2$ and that if $y < 0$, then $P_N(y) = 1 - P_N(|y|)$.

To find $P_N(y)$ one can compute the value of the integral numerically, but a number of excellent algebraic approximations exist (Abramowitz and Stegun 1965, 932), and we recommend their use. If $y \geq 0$, the following approximation is accurate to 10^{-5}:

$$P_N(y) = 1 - \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right)[a_1 t + a_2 t^2 + a_3 t^3], \tag{3.63}$$

where $t = 1/(1 + py)$, and the constants are $p = 0.332\ 67$, $a_1 = 0.436\ 183\ 6$, $a_2 = -0.120\ 167\ 6$, and $a_3 = 0.937\ 298\ 0$.

It often happens that we want to invert the normal distribution function. That is, we wish to find a value y_p such that $P_N(y_p) = p$, where the value of p is specified. There exist nice algebraic methods for this inversion as well (Abramowitz and Stegun 1965, 933). If $0.5 \leq p \leq 1$, then the following approximation is accurate to 4.5×10^{-4}:

$$y_p = t - \frac{c_0 + c_1 t + c_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3}, \tag{3.64}$$

where $t = \sqrt{\log(1/p^2)}$, $c_0 = 2.515\ 517$, $c_1 = 0.802\ 853$, $c_2 = 0.010\ 328$, $d_1 = 1.432\ 788$, $d_2 = 0.189\ 269$, and d_3

$= 0.001\ 308$. If $p < 0.5$, then we find the value of y_{1-p} according to the same formula and then set $y_p = -y_{1-p}$.

As we mentioned earlier, according to the central limit theorem (CLT), appropriately normalized sums of random variables have a distribution function that approaches the normal distribution. Suppose $\{Z_k\}$ is a sequence of independent random variables with $m_k = E\{Z_k\}$ and $\sigma_k^2 = \text{VAR}\{Z_k\}$, and set

$$S_n = \sum_{k=X}^{n} Z_k,$$

$$m_n = \sum_{k=1}^{n} m_k,$$

$$s_n^2 = \sum_{k=1}^{n} \sigma_k^2. \tag{3.65}$$

According to the CLT, the variable $Z = (S_n - m_n)/s_n$ is approximately normally distributed with mean 0 and variance 1. If the Z_k have the same distribution with common mean m and variance σ^2, then the $\mathcal{N}(0,1)$ random variable is $(S_n - nm)/\sigma\sqrt{n}$. We shall use the central limit theorem in the next section to motivate the log-normal distribution, and in the next chapter for the determination of the observation effort when monitoring the incidental catch of seabirds in a fishery.

The Log-Normal Distribution

To understand the log-normal distribution, imagine a population of initial size N_0 during a nonbreeding season. We expect the number of individuals alive at some later day t, N_t, to be the product of N_0 and the daily survival probabilities $\{s_i\}$, where s_i is the probability that an individual survives from day i to day $i + 1$. Thus

$$N_t = N_0 s_0 s_1 \cdots s_{t-2} s_{t-1}. \tag{3.66}$$

Taking logarithms of both sides gives

$$\log(N_t) = \log(N_0) + \log(s_0) + \log(s_1) + \cdots + \log(s_{t-1}). \tag{3.67}$$

If the daily survival probabilities are random variables, then using the central limit theorem, we assume that an appropriate normally distributed random variable Y can be constructed from the sum $\log(s_0) + \log(s_1) + \cdots + \log(s_{t-1})$. We then say that $Z = e^Y$ has a log-normal distribution, and we can rewrite Equation 3.66 as

$$N_t = N_0 e^Y = N_0 Z \tag{3.68}$$

One advantage of the log-normal distribution is that a normal random variable takes values between $-\infty$ and ∞, but many ecological variables are typically positive. The log-normal random variable takes only positive values. In addition, the log-normal distribution has a long tail, which is common to ecological data.

We will now explore some properties of the log-normally distributed random variable $Z = e^Y$, where we assume that Y is $\mathcal{N}(0,\sigma^2)$. We begin with the distribution function

$$F(z) = \Pr\{Z \le z\} = \Pr\{e^Y \le z\} = \Pr\{Y \le \log(z)\}. \tag{3.69}$$

Since we know that Y is normally distributed with mean 0 and variance σ^2,

$$F(z) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\log(z)} \exp\left(-\frac{s^2}{2\sigma^2}\right) ds. \tag{3.70}$$

The density function is found by taking the derivative of $F(z)$, and using the chain rule when evaluating the derivative of the integral,

$$f(z) = F'(z) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\log z)^2}{2\sigma^2}\right) \frac{1}{z}. \tag{3.71}$$

Thus, although Y has a normal density function, the density of Z is skewed, and given by Equation 3.71.

Finally, let us evaluate the mean of the random variable Z. Before doing the calculation, we can try to develop some intuition. The mean of the random variable Y is 0, and Y takes positive and negative values. However, $Z = e^Y$ will only take positive values, so that we expect the mean of Z to be larger than 0. We shall now demonstrate this. We start with

$$E\{Z\} = \int_0^\infty z f(z)\, dz = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^y \exp\left(-\frac{y^2}{2\sigma^2}\right) dy, \tag{3.72}$$

which is justified by noting that, as z varies from 0 to ∞ with density $f(z)$ given by Equation 3.71, y varies from $-\infty$ to ∞ with the standard normal density. Bringing the two exponential terms together gives

$$E\{Z\} = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \exp\left(-\frac{y^2}{2\sigma^2} + y\right) dy. \tag{3.73}$$

We now complete the square in the exponent according to

$$\frac{y^2}{2\sigma^2} - y = \frac{1}{2\sigma^2}\left[y^2 - 2\sigma^2 y\right] = \frac{1}{2\sigma^2}\left[(y-\sigma^2)^2 - \sigma^4\right], \tag{3.74}$$

so that the expected value of Z becomes

$$E\{Z\} = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\sigma^2}\left[(y-\sigma^2)^2 - \sigma^4\right]\right) dy$$

$$= \exp\left(\frac{\sigma^2}{2}\right) \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\sigma^2}(y-\sigma^2)^2\right) dy. \tag{3.75}$$

The integrand in Equation 3.75 is a normal density with mean and variance σ^2, but the range of integration is over all values of y; hence the integral must be equal to 1, and we obtain

$$E\{Z\} = \exp\left(\frac{\sigma^2}{2}\right). \tag{3.76}$$

We thus see that the mean of Z is indeed greater than 0, and that the random variable $Z \exp(-\sigma^2/2)$ will have a mean equal to 1. We will use the log-normal distribution extensively in the case studies of fisheries management.

The Chi-Square Distribution

Another random variable connected to the normal distribution arises as follows. Suppose that the response Z to a control variable X is

$$Z = X + Y, \tag{3.77}$$

where Y is normally distributed with mean 0 and variance 1. The squared deviation between the prediction and the independent variable is then

$$(Z - X)^2 = Y^2, \tag{3.78}$$

and is called the chi-square random variable. If we had n independent variables $\{X_i\}$ and responses $\{Z_i\}$, then the total squared deviation would be

$$\sum_{i=1}^{n} (Z_i - X_i)^2 = \sum_{i=1}^{n} Y_i^2, \tag{3.79}$$

which is called the chi-square random variable with n degrees of freedom and is given the symbol χ_n^2.

The Gamma Distribution

The gamma distribution also takes non-negative values, can have a long tail, and is very useful in Bayesian analysis.

A random variable Z follows a gamma density with parameters a and n if the probability density function is

$$f(z) = \frac{a^n}{\Gamma(n)} e^{-az} z^{n-1}. \tag{3.80}$$

In this equation, $\Gamma(n)$ is read "gamma of n" and is described in Box 3.3.

If you don't worry about these things, just think of $a^n/\Gamma(n)$ as a normalization constant to ensure that $f(z)$ defined in Equation 3.80 is a true probability density (i.e., its integral is 1); the gamma function plays the same role that $\binom{N}{k}$ plays in the binomial distribution. That is, since $\Pr\{0 \leq Z < \infty\}$,

$$\int_0^\infty \frac{a^n}{\Gamma(n)} e^{-az} z^{n-1}\, dz = 1. \tag{3.81}$$

Since $a^n/\Gamma(n)$ is a constant, it can be brought out of the integral sign:

$$\frac{a^n}{\Gamma(n)} \int_0^\infty e^{-az} z^{n-1}\, dz = 1$$

$$\text{or} \qquad \int_0^\infty e^{-az} z^{n-1}\, dz = \frac{\Gamma(n)}{a^n}. \tag{3.82}$$

We now consider some properties of the gamma density, Equation 3.80. To begin, note that if $n = 1$, since $\Gamma(1) = 0! = 1$, $f(z) = e^{-az}$, which is the exponential density.

When $n < 1$, as $z \to 0$, $z^{n-1} \to \infty$, so that $f(z) \to \infty$. When $n > 1$, z^{n-1} will approach 0 as $z \to 0$, so that $f(0) = 0$ and the gamma density has a peak (Figure 3.5) because $e^{-az} \to 0$ as z increases. Thus, the single parameter n controls the wide-ranging shape of this density.

BOX 3.3

AN ASIDE ON THE GAMMA FUNCTION

Most readers will feel comfortable with the more common special functions such as $\log(x)$, e^x, or $\sin(x)$ and $\cos(x)$. These relatively simple transcendental functions (i) are encountered frequently, (ii) often have simple physical interpretations, (iii) are well tabulated, and (iv) have simple power series and limiting behaviors, such as $x^n/e^x \to 0$ as $x \to \infty$ for any n, or $\lim_{x \to 0} (\sin x)/x = 1$.

The gamma function shares many of the same qualities. A good source book is by Abramowitz and Stegun (1965). The gamma function $\Gamma(n)$ arises in classical applied mathematics, and is defined by the integral

$$\Gamma(n) = \int_0^\infty e^{-t} t^{n-1} \, dt.$$

Integrating by parts gives

$$\Gamma(n+1) = \int_0^\infty e^{-t} t^n \, dt = -e^{-t} t^n \Big|_0^\infty + \int_0^\infty e^{-t} n t^{n-1} \, dt = n\Gamma(n),$$

so that we conclude that

$$\Gamma(n+1) = n\Gamma(n).$$

This recurrence formula is similar to the one for factorials in which $n! = n(n-1)!$ For integer values of n, $\Gamma(n+1) = n!$ The general recurrence holds for all values of n, however, not just integer ones.

BOX 3.3 CONT.

Abramowitz and Stegun (1965, 256) show that $\Gamma(n)$ can be calculated from the formula

$$\Gamma(n) = \frac{1}{\sum_{k=1}^{\infty} c_k n^k}.$$

The best way to use this formula is to write $n = n_I + n_F$, where n_I is an integer and n_F is a fraction with $0 < n_F < 1$. First compute $\Gamma(n_F)$ from the power series and then use the recurrence relationship. For example, $\Gamma(3.7) = 2.7\Gamma(2.7) = (2.7)(1.7)\Gamma(1.7) = (2.7)(1.7)(0.7)\Gamma(0.7)$. The first nineteen of the c_k are (Abramowitz and Stegun 1965, 256):

k	c_k					
1	1.000	000	000	000	000	0
2	0.577	215	664	901	532	9
3	−0.655	878	071	520	253	8
4	−0.042	002	635	034	095	2
5	0.166	538	611	382	291	5
6	−0.042	197	734	555	544	3
7	−0.009	621	971	527	877	0
8	0.007	218	943	246	663	0
9	−0.001	165	167	591	859	1
10	−0.000	215	241	674	114	9
11	0.000	128	050	282	388	2
12	−0.000	020	134	854	780	7
13	−0.000	001	250	493	482	1
14	0.000	001	133	027	232	0
15	−0.000	000	205	633	841	7
16	0.000	000	006	116	095	0
17	0.000	000	005	002	007	5
18	−0.000	000	001	181	274	6
19	0.000	000	000	104	342	7

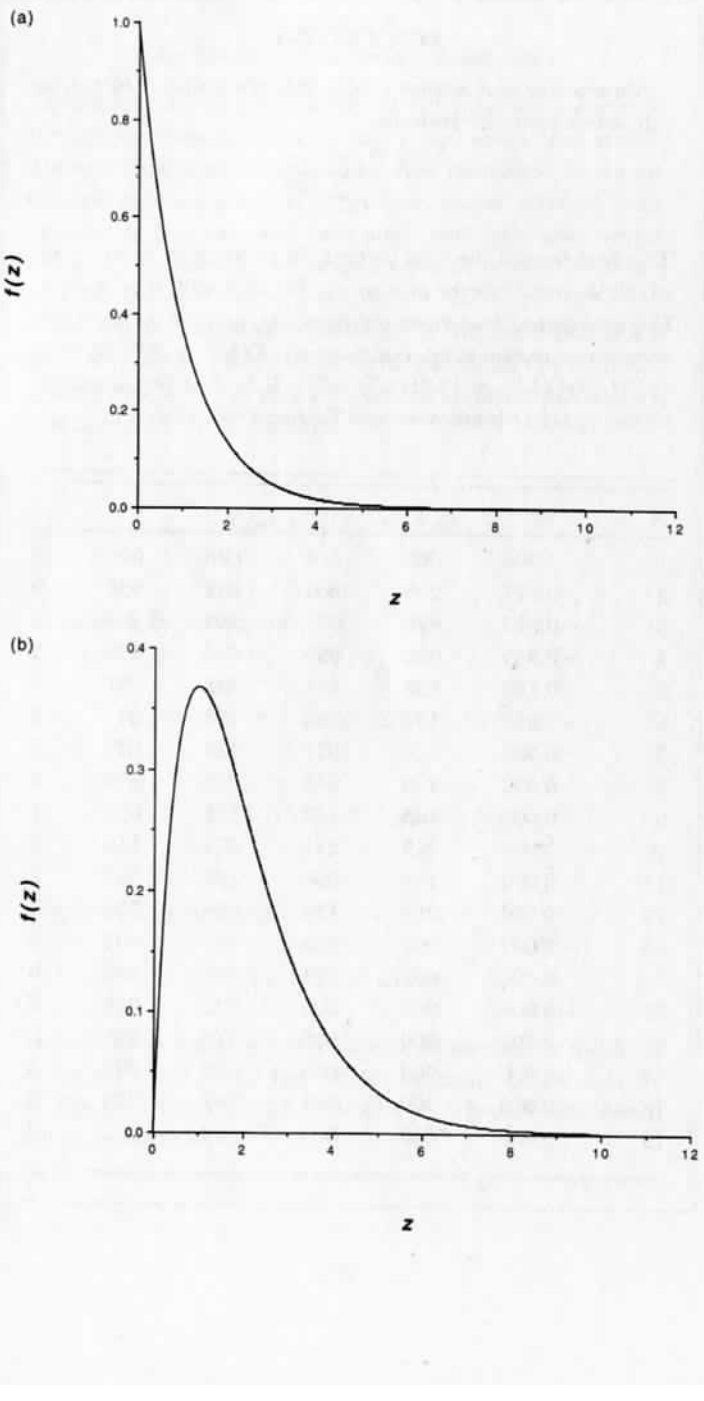

FIGURE 3.5. The gamma probability density $f(z)$ for three values of the parameters n and a. Note the range of shapes that is possible for this density. (a) $n = 1$, $a = 1$; (b) $n = 2$, $a = 1$; (c) $n = 2$, $a = 0.5$.

The expected value of Z is

$$E\{Z\} = \int_0^\infty z f(z)\, dz = \int_0^\infty \frac{a^n}{\Gamma(n)} e^{-az} z^n\, dz. \tag{3.83}$$

Using a modification of Equation 3.82 gives

$$E\{Z\} = \frac{\Gamma(n+1)}{a^{n+1}} \frac{a^n}{\Gamma(n)} = \frac{n}{a}. \tag{3.84}$$

The mean of the gamma density is the ratio of the parameters. The most likely value (i.e., the "mode") of the gamma density is found by setting the derivative of $f(z)$ equal to 0 and solving for $z^* = (n - 1)/a$, so that the most likely value of the gamma density occurs at a value smaller than the mean, and therefore the density has a long tail.

We encourage you to find the second moment using the same method, and then show that the coefficient of variation is

$$\mathrm{CV}\{Z\} = \frac{1}{\sqrt{n}}, \tag{3.85}$$

so that the single parameter n also controls the coefficient of variation.

Ecological Scenario: The Return to Random Search. We now return to the discussion of random search by predators (Box 3.2). Recall that we concluded there that

$$Q(t) = \Pr\{\text{no food is found between 0 and } t\} = e^{-ct}, \tag{3.86}$$

and that this function has the memoryless property that unsuccessful search up to time t provides no information about the chance of success after that time.

Previously, we assumed that c was a fixed constant. Let us now suppose, however, that c has a frequency distribution. For example, the search rate might vary across seasons, across spatial locations as the predator searches, or across individual prey items. In that case, Equation 3.86 is reinterpreted as the conditional probability of not finding food, given the value of c. Assume that c has a gamma density. Then the joint probability of not finding food and the value of c is

$$\Pr\{\text{no food is found between 0 and } t \text{ and the search parameter takes the value } c\} = e^{-ct}\frac{a^n}{\Gamma(n)}e^{-ac}c^{n-1}. \tag{3.87}$$

Consequently, the probability of not finding food is

$$\Pr\{\text{no food is found between 0 and } t\} = \int_0^\infty \Pr\left\{\begin{array}{l}\text{no food is found between 0 and } t\\ \text{and the search parameter takes}\\ \text{the value } c\end{array}\right\} dc$$

$$= \int_0^\infty \frac{a^n}{\Gamma(n)}e^{-(a+t)c}c^{n-1}\,dc = \left(\frac{a}{a+t}\right)^n. \tag{3.88}$$

You should verify the integral, once again by using logic similar to that in Equation 3.84. You should also verify, following the same calculation as in Box 3.2, that the distribution in Equation 3.88 does not have the memoryless property. We return to this example once more, when we discuss Bayesian analysis, because if c has a distribution of values, the predator can learn from its failed search and learning changes the frequency distribution. The precise way that this is done requires the methodology introduced in Chapter 9.

The Negative Binomial Distribution

The negative binomial distribution arises in two ways, and both are relevant to the ecological detective. First, imagine a sequence of independent experiments, each of which has probability p of succeeding. We are interested in the number of experiments needed before s successes occur. In particular, we ask for the probability that the s^{th} success occurs on trial $Z = u + s$, where u is the number of unsuccessful experiments, so that $u = 0, 1, 2, \ldots$. The s^{th} success can happen on trial $u + s$ only if there are $s - 1$ successes in the first $u + s - 1$ experiments and a success on the $(u + s)^{\text{th}}$ experiment. The probability of the latter event is p and the probability of the former is given by the binomial distribution

$$\binom{u+s-1}{s-1}p^{s-1}(1-p)^{u+s-1-(s-1)} = \binom{u+s-1}{u}p^{s-1}(1-p)^u. \tag{3.89}$$

Multiplying this expression by p, we obtain

$$\Pr\{s^{\text{th}} \text{ success occurs on trial } u+s\}$$

$$= \binom{u+s-1}{u} p^s(1-p)^u. \tag{3.90}$$

This is the first form of the negative binomial distribution. Here the parameters are u and p with the possible values $u > 0$ and $0 < p < 1$.

The second form of the negative binomial distribution arises when we consider a Poisson process in which the rate parameter has a probability distribution. In that case, we can interpret Equation 3.51 as a conditional probability:

$$\Pr\{Z(t) = s \mid \text{parameter} = r\} = \frac{e^{-rt}(rt)^s}{s!}. \tag{3.91}$$

Now assume that r has a gamma density with parameters n and a, so that the expected value of r is n/a. The unconditional distribution of $Z(t)$ is found by integrating the product of the conditional distribution Equation 3.91 and the gamma density, since this product is the $\Pr\{Z(t) = s$ and the parameter $= r\}$, over all possible values of r:

$$\Pr\{Z(t) = s\} = \int_0^\infty \frac{e^{-rt}(rt)^s}{s!} \frac{a^n}{\Gamma(n)} e^{-ar} r^{n-1} \, dr. \tag{3.92}$$

Taking everything that is constant out of the integral gives

$$\Pr\{Z(t) = s\} = \frac{t^s a^n}{s!\Gamma(n)} \int_0^\infty e^{-(t+a)r} r^{s+n-1} \, dr. \tag{3.93}$$

Computing the integral as before,

$$\int_0^\infty e^{-(t+a)r} r^{s+n-1} \, dr = \frac{\Gamma(n+s)}{(a+t)^{n+s}}, \tag{3.94}$$

so that

$$\Pr\{Z(t) = s\} = \frac{t^s a^n}{s!\Gamma(n)} \frac{\Gamma(n+s)}{(a+t)^{n+s}}$$

$$= \frac{\Gamma(n+s)}{\Gamma(n)} \frac{t^s}{s!} \frac{a^n}{(a+t)^{s+n}}$$

$$= \frac{\Gamma(n+s)}{\Gamma(n)s!} \left(\frac{t}{a+t}\right)^s \left(\frac{a}{a+t}\right)^n. \tag{3.95}$$

If we set $p = a/(a+t)$, then Equation 3.95 can be rewritten as

$$\Pr\{Z(t) = s\} = \binom{n+s-1}{s} p^n(1-p)^s, \tag{3.96}$$

which is analogous to Equation 3.90 with n replacing u. The difference is that we now allow any value of n, whereas in Equation 3.90 the understanding is implicitly that u is at least 1.

The mean of the negative binomial distribution is

$$E\{Z(t)\} = \frac{n(1-p)}{p} = \frac{n}{a} t = m(t) \tag{3.97}$$

and the variance is

$$\text{VAR}\{Z(t)\} = m(t) + \frac{m(t)^2}{n}. \tag{3.98}$$

Unlike the case of the Poisson distribution, in which the variance and mean are equal, the variance of the negative binomial distribution will always be larger than the mean. Hence, n is often called the "overdispersion" parameter. We can see this more clearly by considering the coefficients of variation. For the Poisson distribution,

$$\text{CV}_{\text{Poisson}}\{Z(t)\} = \frac{1}{\sqrt{rt}}, \tag{3.99}$$

whereas for the negative binomial distribution,

$$\text{CV}_{\text{NB}}\{Z(t)\} = \sqrt{\frac{1}{m(t)} + \frac{1}{n}}. \tag{3.100}$$

From Equation 3.97 we see that as $t \to \infty$, $m(t) \to \infty$. Although the CV of the Poisson distribution goes to 0 as $t \to \infty$, the CV of the negative binomial distribution approaches a constant. Note that as n increases, CV_{NB} approaches $CV_{Poisson}$. This can be shown more precisely: that as $n \to \infty$, the negative binomial distribution becomes more and more Poisson-like.

A form of the negative binomial distribution commonly encountered in ecological texts (e.g., Southwood 1966), and one that we find handy to use, is

$$\Pr\{Z(t) = s\} = \frac{\Gamma(k + s)}{\Gamma(k)s!}\left(1 + \frac{m}{k}\right)^{-k}\left(\frac{m}{m + k}\right)^{s}, \quad (3.101)$$

where k and m are parameters. Using Equation 3.95, setting $m(t) = (n/a)\ t$, and doing some algebra shows that

$$\Pr\{Z(t) = s\} = \frac{\Gamma(n + s)}{\Gamma(n)s!}\left(\frac{m(t)}{n + m(t)}\right)^{s}\left(\frac{n}{n + m(t)}\right)^{n}. \quad (3.102)$$

Comparing Equations 3.101 and 3.102, we see that $m(t)$ and m have exactly the same interpretation as the mean, and that k and n have exactly the same interpretation as the overdispersion parameter.

We can find the terms of the negative binomial distribution using an iterative procedure similar to the one used for the binomial and Poisson distributions. For purposes of commonality with most ecological texts, we adopt Equation 3.102, rewriting it with $Z(t) = Z$, $m(t) = m$, and $n = k$, so that

$$\Pr\{Z = s\} = \frac{\Gamma(k + s)}{\Gamma(k)s!}\left(\frac{m}{k + m}\right)^{s}\left(1 + \frac{m}{k}\right)^{-k}, \quad (3.103)$$

and note that the last term is the same as $[k/(k + m)]^k$ so that

$$\Pr\{Z = s\} = p(s,m,k) = \frac{\Gamma(k + s)}{\Gamma(k)s!}\left(\frac{m}{k + m}\right)^{s}\left(\frac{k}{k + m}\right)^{k}. \quad (3.104)$$

From this equation, we see that $p(0,m,k) = [k/k + m)]^k$ and that additional terms can be computed according to

$$p(s,m,k) = \frac{s + k - 1}{s}\frac{m}{k + m}p(s - 1,m,k) \quad \text{for } s = 1, 2, \ldots. \quad (3.105)$$

The iteration result, Equation 3.105, is derived in the same way as the iteration results for the binomial and Poisson distributions were derived. We encourage you to derive it and write out the pseudocode. We shall now use it.

THE MONTE CARLO METHOD

In order to confront models with data, we must estimate parameters in the models from the data and then choose one description of nature over another. Because we usually do not know the true mechanisms and processes in the natural world, we never know if the parameters that we estimate are indeed "true" or if the model that is picked is "correct." One way to increase our confidence in the methods we use is to test models and methods on sets of data in which we know exactly what is happening, i.e., where we create the data and thus know the true situation exactly. A useful method for generating such data is called the Monte Carlo method or the method of stochastic simulation (Ripley 1987).

The Monte Carlo method uses random-number generators for the construction of data. Virtually all microcomputer languages have built-in random-number generators, and these are, for almost all of our purposes, sufficient. The usual problem with such generators is that they are only quasi-random and have a periodic cycling in the generation

of the numbers. These days, however, the periods are of the order of 2^{30}, so that the difficulties are minor. The random number generators usually provide a value U that is uniformly distributed between 0 and 1. Thus, the distribution function for U is

$$F(u) = \begin{cases} u & \text{if } 0 \le u \le 1, \\ 0 & \text{otherwise,} \end{cases} \tag{3.106}$$

and the density is $f(u) = F'(u) = 1$.

To construct a random variable Z that is uniformly distributed on the interval $[A,B]$, we pick U and set

$$Z = A + (B - A)U. \tag{3.107}$$

Since the smallest value that U takes is 0, the smallest value that Z takes is A; similarly, the largest value of $Z = B$, corresponding to $U = 1$.

Typically, the command U = RND in a computer program will generate a uniformly distributed random variable (but check the manual for your software). We now describe methods for generating random variables with other distributions.

Binomial, Poisson, or Negative Binomial Random Variables

These three distributions have the common feature that the random variable Z takes integer values. We shall illustrate the method for generating individual random variables from a specific distribution using the binomial distribution, and leave the Poisson and negative binomial distributions to you.

For the binomial distribution, the probability $p(k,N)$ of obtaining exactly k successes in N experiments is given by Equation 3.44. If $p(k,N)$ is summed from $k = 0$ to $k = N$, the sum is 1. The value of k associated with a particular value of $U = u$, called k_u, is chosen so that

$$\sum_{k=0}^{k_u} p(k,N) \le U$$

and

$$\sum_{k=0}^{k_u+1} p(k,N) > U. \tag{3.108}$$

A pseudocode that implements this idea is:

Pseudocode 3.3

1. Specify parameters N and p. Choose a uniformly distributed random number U. Set $k = 0$ and SUM = 0.
2. Compute $p(k,N)$ from Equation 3.44.
3. Replace SUM by SUM + $p(k,N)$.
4. If SUM $\ge U$, then the current value of k is the number of successes in this single experiment. Otherwise, replace k by $k + 1$ and return to step 2.

Normal Random Variables

To generate normally distributed random variables, we recommend the use of the Box-Mueller scheme (Press et al. 1986, 202). Choose two uniformly distributed random numbers U_1 and U_2 and set

$$Z_1 = \sqrt{-2\log(U_1)}\cos(2\pi U_2),$$
$$Z_2 = \sqrt{-2\log(U_1)}\sin(2\pi U_2). \tag{3.109}$$

Then Z_1 and Z_2 are normally distributed random variables with mean 0 and variance 1. To make these variables normally distributed with mean m and variance σ^2, replace Z_i by $m + \sigma Z_i$. We leave writing a pseudocode to you.

Gamma Random Variables

Gamma random variables are more difficult to generate. Press et al. (1986, 204 ff.) describe one method ("the rejection method") that interested readers may wish to consult. In general, some form of integration of the probability density is needed.

An Ecological Scenario: The Simple Population Model with Process and Observation Uncertainty

We return to the model Equation 3.38. In order to generate data with this model, assumptions about W_t, V_t, and the other parameters are required. For example, we might assume that the process and observation uncertainties are normally distributed with mean 0 and standard deviations σ_W and σ_V, respectively, but that the initial population size N_0 is known exactly. As a demonstration of the importance of understanding observation and process uncertainty, and to demonstrate the Monte Carlo technique, we now perform some simple computer experiments based on the following pseudocode. A pseudocode for this model with process and observation uncertainties is:

Pseudocode 3.4

1. Specify s, b, σ_N, σ_W, σ_V, and N_0.
2. Begin a loop over 50 time steps.
3. Calculate N_{t+1} and $N_{\text{obs},t}$ from Equation 3.38.
4. Print or graph results as desired.
5. Exit after 50 time steps.

We chose $s = 0.8$, $b = 20$, and $N_0 = 50$.

To begin, we can ask how process uncertainty affects the relationship between N_t and N_{t+1}. If we allow for process uncertainty ($\sigma_W = 10$), but no observation uncertainty ($\sigma_V = 0$), the observed values are "scattered" about the true value (Figure 3.6) but will be centered on it. A standard

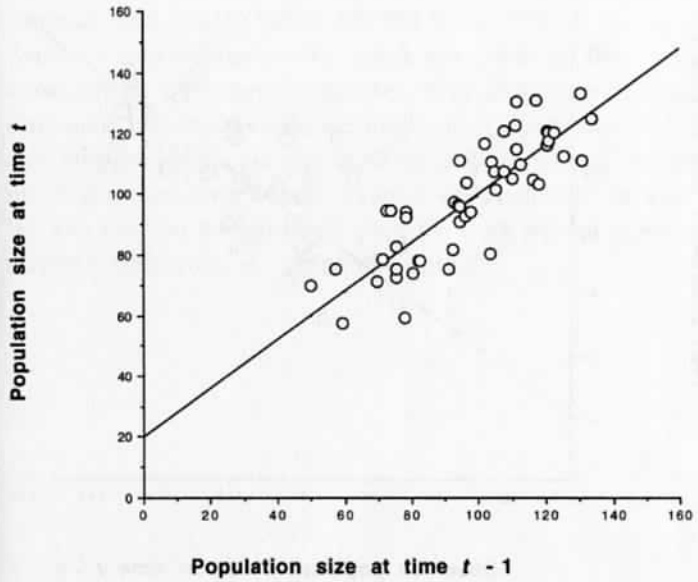

FIGURE 3.6. One Monte Carlo realization of fifty data points drawn with process uncertainty but no observation uncertainty. The solid line represents the true model (deterministic relationship).

linear regression fit to the data gives $y = 20.01 + 0.808x$ with $r^2 = 0.723$. Thus, both the birth rate (the constant in the regression) and the survival (the slope of the regression) are accurately determined.

If we now add observation uncertainty, by setting $\sigma_V = 10$, and use the same sequence of random numbers to generate the data, we obtain an apparent "relationship" (Figure 3.7) that is weaker than in the case without observation uncertainty. In this case, the regression is $y = 32.47 + 0.684x$ with $r^2 = 0.481$. Thus, we overestimate the birth rate, underestimate survival, and explain only about half as much of the variation as before. What happened? By adding variability in observations, it now appears that there are some very small population sizes and some very large ones, even

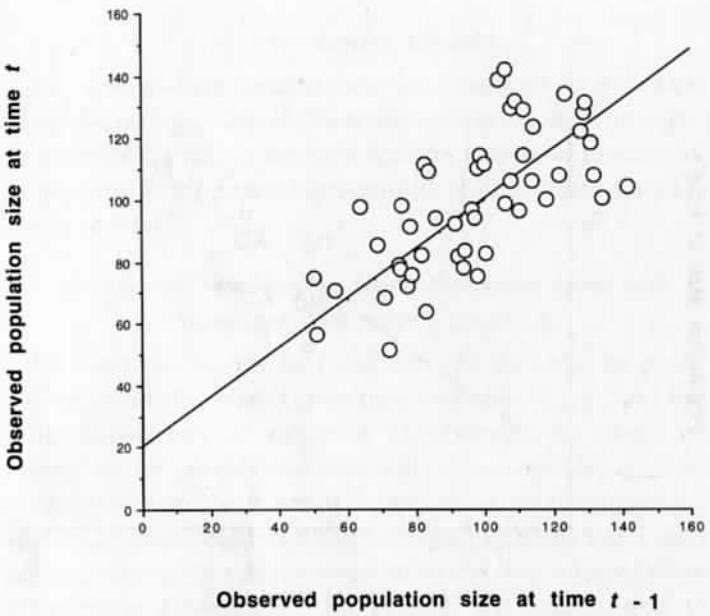

FIGURE 3.7. One Monte Carlo realization of fifty data points drawn with process and observation uncertainty. Once again, the solid line represents the true model (deterministic relationship).

though the true population size has not changed. The net effect is that the population in the next time period, N_{t+1}, appears to depend less on N_t. This is not due to a weakening of the density dependence. Rather, it is caused by the additional source of uncertainty in the model. The job of the ecological detective is to sort out such differences and then arrive at the best description of nature possible.

Bootstrap Data Sets

Another use of the Monte Carlo method is to generate "replicate" sets of data from one actual set of data. This is often called a "bootstrap" data set (Efron and Tibshirani 1991, 1993). We do it by resampling the data set with replacement. For example, in the discussion of coefficient of

variation, we described a data set of masses of rodents. The original data (in g) were {79,120,85,99,100}. A bootstrap data set is constructed by randomly picking five "new" masses from the original data set, with replacement. Thus, one such bootstrap replicate might be {79,120,85,100,100} and another might be {99,75,99,120,85}. We could use this method to generate a large number of "replicate" data sets. We will use the bootstrap method for both model selection and the evaluation of confidence limits.