

growing season. Most populations will be affected by such variation, either directly or indirectly through its effects on interacting species (e.g., prey, predators, competitors, diseases, and so on.). When we use a sequence of censuses to estimate measures of population viability, we must account for the pervasive effect of environmental variation that can be seen in most count data. To build a general understanding of how this is done, we first expand the overview of how environmental variation affects population dynamics that we gave in Chapter 2. We next review key theoretical results that underlie the simplest count-based methods in PVA. Then, with the necessary background in place, we delve into the details of using count data to assess population viability.

Population Dynamics in a Random Environment

Let us put more flesh on the bare-bones overview of how temporal variability influences population dynamics that we gave in Chapter 2. Specifically, let us return to the simple model for discrete-time geometric population growth in a randomly varying environment:

$$N_{t+1} = \lambda_t N_t \quad (3.1)$$

where N_t is the number of individuals in the population in year t and λ_t is the population growth rate, or the amount by which population size multiplies from year t to year $t + 1$. Recall that Equation 3.1 assumes that population growth is density independent (i.e., λ_t is not affected by population size, N_t). If there is no variation in the environment from year to year, then the population growth rate λ will be constant, and only three qualitative types of population growth are possible (Figure 3.1A): geometric increase (if $\lambda > 1$), geometric decline to extinction (if $\lambda < 1$), and stasis (if λ exactly equals 1). However, by causing survival and reproduction to vary from year to year, environmental variability will cause the population growth rate, λ_t , to vary as well, and unlike the simple example in Chapter 2 in which we assumed λ_t only took on two values, in reality we expect λ_t to vary over a continuous range of values. Moreover, if the environmental fluctuations driving changes in population growth include an element of unpredictability (as factors such as rainfall and temperature are likely to do), then we will not be able to predict with certainty what the exact sequence of future population growth rates will be. As a consequence, even if we know the current population size and both the average value and the degree of variability in the population growth rate, the best we can do is to make probabilistic statements about the number of individuals the population will include at some time in the future. That is, we must view change in population size over time as a stochastic process.

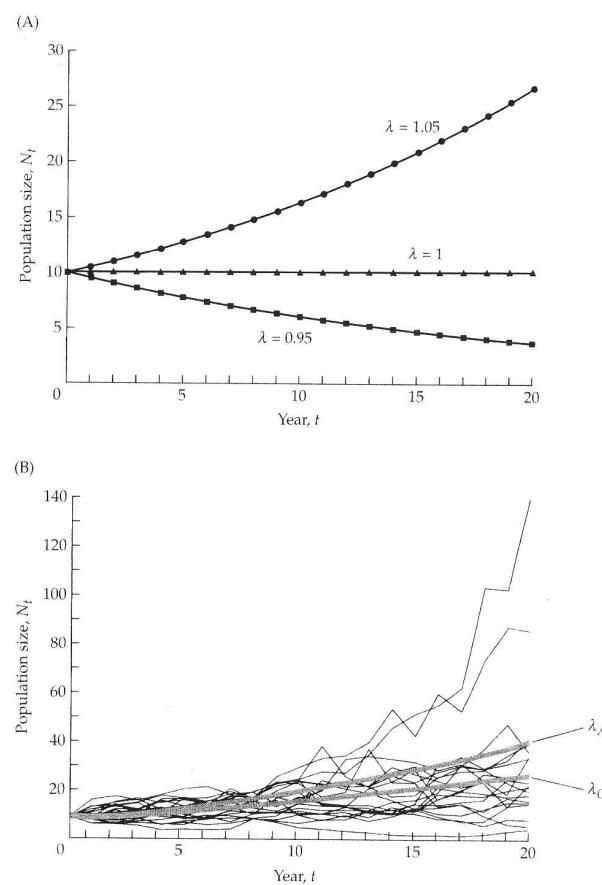


Figure 3.1 Patterns of deterministic versus stochastic growth. (A) Predicted population sizes under deterministic geometric growth with λ greater than, less than, or equal to 1. (B) Twenty realizations of stochastic population growth according to Equation 3.1. For all realizations, the population growth rate λ_t is log-normally distributed with an arithmetic mean of $\lambda_A = 1.0725$ and a variance of 0.047. The upper and lower broad gray lines show the deterministic predictions using the arithmetic and geometric mean growth rates, respectively.

To illustrate stochastic population growth, Figure 3.1B shows 20 realizations of Equation 3.1 in which the value of the population growth rate λ was drawn at random each year.¹ All realizations start at a population size of 10 individuals. Thus each realization can be viewed as a possible trajectory the population might follow given its initial size and the distribution of λ . Figure 3.1B illustrates three fundamental features of stochastic population growth. First, the realizations diverge over time, so that the farther into the future you go, the more variable the predictions about likely population sizes become. Second, the realizations do not follow very well the predicted trajectory based upon the arithmetic mean population growth rate, λ_A (this is the trajectory shown by the upper broad line in Figure 3.1B). In particular, even though in this case λ_A predicts that the population should slowly increase, a few realizations explode over the 20 years illustrated, whereas others decline. Thus extinction is possible even though the population size predicted by the arithmetic mean growth rate increases. Third, the endpoints of the 20 realizations shown are highly skewed, with a few trajectories winding up much higher than λ_A would suggest, but most ending well below this average prediction (in fact, three realizations ended lower than their starting population size). This skew is due in part to the multiplicative nature of population growth (see Equation 2.3). Because the size of the population after 20 years is proportional to the product of the population growth rates in each of those years, a long string of chance “good” years (i.e., those with high rates of population growth) can carry the population to a very high level of abundance, whereas strings of “bad” years tend to confine the population to the restricted zone between the average and zero abundance. We also saw in Chapter 2 that, because population growth is multiplicative, the geometric mean population growth rate λ_G is a better descriptor of the behavior of a typical realization. In fact, λ_G predicts the median population size at any point in the future; that is, for a large number of realizations, half will lie above and half below the population size predicted by λ_G (shown by the lower broad line in Figure 3.1B; in this particular sample of 20 realizations, more happened to fall below the median than above, but λ_G still does a better job of “splitting the difference” than does λ_A).

Skewness in the distribution of the likely future size of a population is a general feature of density-independent population growth in a stochastic environment. In fact, we can make an even more precise statement: the probability that the population will be of a certain size at a future time horizon will usually be well described by a particular skewed probability distribution,

¹Specifically, λ is lognormally distributed (that is, the log of λ follows the familiar normal distribution); the lognormal distribution is one of several distributions that can appropriately describe random λ_t values because, unlike the normal, it never

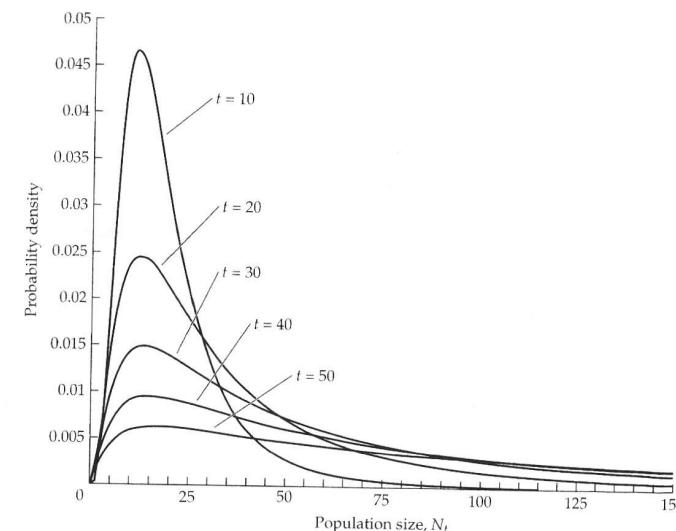


Figure 3.2 The probability that a population following the stochastic growth process illustrated in Figure 3.1B will lie within a certain size interval at a given time in the future is described by this lognormal probability density function.

the lognormal² (Lewontin and Cohen 1969). The particular lognormal distribution that would describe a large number of trajectories corresponding to the example in Figure 3.1B is shown in Figure 3.2 (compare this figure to Figures 2.4 and 2.5). The y -axis in Figure 3.2 is probability density; to obtain the actual probability that the population size lies between two particular values (the equivalent of the class boundaries in a histogram), we would calculate the integral of this probability density function between those two values of N_t . Thus population size is most likely to lie near the peak of the lognormal probability density function in Figure 3.2, but because the function is skewed to the right, the arithmetic mean population size will lie to the right of the peak. As Figure 3.2 illustrates, both the mean and the most likely population size increase over time, as does the probability density for larger population sizes. This makes sense given that most population trajectories tend to grow (Figure 3.1B).

²Thanks to the Central Limit theorem, this conclusion is true asymptotically regardless of the probability distribution from which the λ 's are drawn. See Lewontin and

If population size itself follows a lognormal distribution, then the natural logarithm of population size will be normally distributed. This follows from the definition of a lognormal random variable; if N_t is lognormally distributed, then $\log N_t = X_t$, where X_t is *normally* distributed.³ As does the lognormal distribution of N_t values (Figure 3.2), the normal distribution of $\log N_t$ values will change over time (Figure 3.3). Specifically, its mean will either increase or decrease, depending on whether separate realizations of the stochastic population growth process tend to grow (as in Figure 3.1) or decline, but its variance will strictly increase (as prediction becomes less certain over longer time intervals). In this chapter, we will use $\log N_t$, rather than the untransformed population size, because a normal distribution arises naturally from the physical process of diffusion, which we will use below to approximate the extinction process.

Until now, we have talked about the growth of N_t . How do we characterize the growth of $\log N_t$? Above and in Chapter 2, we argued that the best predictor of whether N_t will increase or decrease over the long term is λ_G , the geometric mean of the population growth rates each year. On the scale of log population size, the log of λ_G is the most natural way to express population growth. The geometric mean of λ is the value that would give the same average annual population growth rate as is observed over a long sequence of stochastically varying growth rates. That is, since $N_{t+1} = (\lambda_t \lambda_{t-1} \lambda_{t-2} \dots \lambda_2 \lambda_1 \lambda_0)N_0$ (see Equation 2.3), λ_G is defined as

$$(\lambda_G)^t = \lambda_t \lambda_{t-1} \lambda_{t-2} \dots \lambda_2 \lambda_1 \lambda_0 \quad \text{or} \quad \lambda_G = (\lambda_t \lambda_{t-1} \lambda_{t-2} \dots \lambda_2 \lambda_1 \lambda_0)^{1/t} \quad (3.2)$$

Converting this formula for λ_G to the log scale yields a new measure of “average” population growth:⁴

$$\mu = \log \lambda_G \approx \frac{\log \lambda_t + \log \lambda_{t-1} + \log \lambda_{t-2} + \dots + \log \lambda_2 + \log \lambda_1 + \log \lambda_0}{t} \quad (3.3)$$

As Equation 3.3 states, the correct measure of stochastic population growth on a log scale, μ , is equal to the log of λ_G or, equivalently, to the arithmetic mean of the $\log \lambda_t$ values. Thus if we can estimate the value of μ (which we will see how to do below), then we can immediately determine the geometric mean population growth rate, and thus whether the population will tend to grow or decline. In particular, Equation 3.3 predicts that if $\mu > 0$, then $\lambda_G > 1$, and most population trajectories will grow, whereas if $\mu < 0$, then $\lambda_G < 1$, and most trajectories will decline.

³In this book, when we write “log” we always mean the natural logarithm (following mathematical convention).

⁴Strictly speaking, the rightmost expression in Equation 3.3 approaches $\log \lambda_G$ only

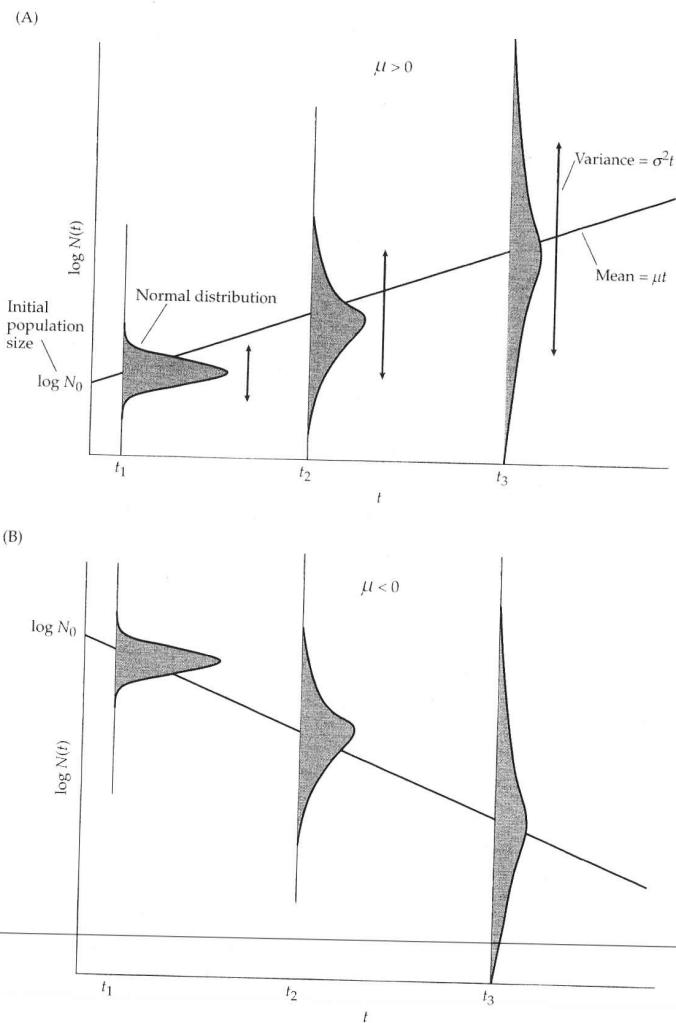


Figure 3.3 Normal distributions of the log population size, when the parameter μ is (A) positive or (B) negative.

Because of this relationship between μ and λ_G , the parameter μ is of direct interest as a metric of population viability, telling us the direction in which the population will tend to move over time. But remember that even if μ is positive, there is still some chance that the population will fall to low levels, even to zero, as Figures 3.1B and 3.3A show. Thus in addition to knowing the population's general tendency to grow or decline, we would also like to predict the full probability distribution for different population sizes, in particular so that we can calculate the probability that the population will fall below a specified quasi-extinction threshold at or before a specified future time horizon. To fully characterize the changing normal distribution of log population size (Figure 3.3), we will need two parameters. The parameter μ not only equals the log of λ_G ; it also represents the rate at which the mean of the normal distribution of log population size changes over time (that is, at time t , the mean of $\log N_t$ equals μt ; see Figure 3.3). But we must also define a second parameter, called σ^2 (sigma-squared), that represents the rate at which the variance of the distribution increases over time (that is, at time t , the variance of $\log N_t$ equals $\sigma^2 t$; Figure 3.3). Whereas μ is approximated by the (arithmetic) mean of the log population growth rates (i.e., the mean of the $\log \lambda_t$ values; see Equation 3.3), σ^2 is approximated by the variance of the $\log \lambda_t$'s, as we will see below. Our use of the symbol σ^2 to represent the variance of $\log \lambda$ follows Lande and Orzack 1988 and Dennis et al. 1991, the two most important sources for the material presented in this chapter. Throughout the book, we use σ^2 strictly to represent the variance in the log population growth rate that is caused by environmental stochasticity.

Together, μ and σ^2 fully describe the normal probability distribution of future log population sizes. Specifically, a positive value of μ indicates an environment in which most realizations tend to grow, whereas a negative μ indicates that declining realizations predominate. The more the population growth rate λ varies from year to year as a result of environmental stochasticity, the greater will be the value of σ^2 and the greater the range of possible population sizes in the future. These two measures of the dynamics of log population size are used extensively in the calculation of extinction times. Because their verbal definitions are somewhat tortuous ("the mean and variance of the log population growth rate"), we will usually refer to them by their symbols. Nonetheless, to really understand how simple count-based PVAs work, you will need to fix the meanings of μ and σ^2 in your mind.

The Relationship Between the Probability of Extinction and the Parameters μ and σ^2

Because μ and σ^2 describe the changing probability that the log population will be within a given range (Figure 3.3), it makes intuitive sense that

if we know their values (as well as the current population size and the quasi-extinction threshold), we can calculate the probability of quasi-extinction at any future time. In this section, we review the theoretical underpinnings behind the calculation of extinction probabilities using μ and σ^2 . This section is more mathematically challenging than the rest of the chapter, and you may want to skip it. If you do, you will still be able to understand and use the following sections on how to do simple count-based PVAs, provided that you accept on faith the formulas relating μ and σ^2 to extinction risk. However, it is important to recognize the basic patterns in extinction times as well as the assumptions underlying the theory, so we urge you to read the last three paragraphs of this section even if you skip the rest.

For PVA purposes, we consider a population to be "extinct" (either literally so or at least in need of a radically different management approach) if it ever hits a quasi-extinction threshold. (In Chapter 2 we discussed several factors that determine the choice of a reasonable quasi-extinction threshold.) Thus to measure extinction risk, we want to calculate the probability that the population hits the threshold *at any point during the interval between the present time and a specified future time horizon*. We begin by calculating the probability that extinction first occurs in one small segment of that time interval, and then we sum the probabilities over all segments. This calculation is performed by viewing the changing normal distribution of log population size in Figure 3.3 as though it were a cloud of particles undergoing diffusion with drift. To envision this process, think of a large number of small beads released at a single point near the bottom of an infinitely deep river, with height above the river bottom standing in for log population size. The path of each bead represents a single realization of "population" growth. As time evolves, the cloud of beads will move downriver, but its vertical position will change as well, moving upward or downward depending on whether the individual beads tend to float or sink, and spreading out due to turbulent flow of the water. The rate of floating or sinking is represented by μ , and the rate of turbulent diffusion by σ^2 . The river bottom represents the quasi-extinction threshold; we will assume that any bead that hits the bottom will adhere to it. (In mathematical terms, the quasi-extinction threshold is a so-called absorbing lower boundary.) Even if the individual beads tend to float, some of them will be pushed to the bottom by turbulence, just as some population trajectories will hit low levels even as most trajectories tend to grow (Figure 3.3A), and the more turbulent is the flow of the water (i.e., the larger is σ^2), the quicker some beads will be pushed to the bottom. Treating the change in log population size (or in the vertical position of a bead) as a process of diffusion with drift allows the rate at which realizations first hit the lower boundary to be calculated. (For full details on this diffusion approximation, see Cox and Miller 1965, pages 208–222, and Lande and Orzack 1988; for a clearly presented alternative derivation of Equation 3.4, see Whitmore and Seshadri 1987.) Using a diffusion approximation, the

probability density for first hitting the quasi-extinction threshold (that is, going “extinct”) in a small segment of time beginning at t is given by the so-called inverse Gaussian distribution:⁵

$$g(t|\mu, \sigma^2, d) = \frac{d}{\sqrt{2\pi\sigma^2 t^3}} \exp\left[\frac{-(d + \mu t)^2}{2\sigma^2 t}\right] \quad (3.4)$$

where $d = \log N_c - \log N_x$ is the difference between the log of the current population size N_c and the log of the extinction threshold, N_x . (The notation $g(t|\mu, \sigma^2, d)$ means that the value of g at time t depends on the values of μ , σ^2 , and d .) Notice that the extinction probability depends only on the difference (on the log scale) between the current and threshold population sizes, not on their actual values.

Figure 3.4 illustrates how the shape of the inverse Gaussian distribution (Equation 3.4) is affected by the values of μ and σ^2 . The most likely time to hit the threshold is the peak of the distribution. The position of the peak is shifted toward shorter times as the value of σ^2 increases. (Compare curves with the same value of μ but different values of σ^2 .) This is because greater variation drives populations down to the threshold more rapidly, as we would expect from the analogy with diffusing beads. Figure 3.4A shows that more negative values of μ lead to shorter quasi-extinction times. (Compare the two curves with $\sigma^2 = 0.04$.) This also makes sense; if the mean of the normal distribution of log population size declines more rapidly, quasi-extinction should occur earlier. If μ is positive, the probability of quasi-extinction occurring in each time segment is smaller than if μ is negative. (Compare Figure 3.4 A and B, and note the different scales of the y-axes.) The more positive μ is, the smaller is the total area under the curve (which represents the ultimate probability of quasi-extinction, as we will see below). However, notice that as μ becomes more positive, the peak of the distribution (i.e., the most likely quasi-extinction time) actually *decreases*. (Compare curves for $\mu = 0.01$ and $\mu = 0.03$ with the same value of σ^2 in Figure 3.4B.) The reason for this somewhat counterintuitive result (as explained by Lande and Orzack 1988 and Dennis et al. 1991) is that when μ is strongly positive, most realizations tend to increase rapidly to high numbers, from which they

⁵Strictly speaking, Equation 3.4 is a “proper” probability distribution (i.e., a function that integrates to 1) only if $\mu < 0$. This merely reflects the fact that when $\mu > 0$, not all realizations hit the threshold, so the ultimate probability of extinction is less than 1 (see Equation 3.7). Equation 3.4 can be converted into a proper probability distribution for the time to hit the threshold *given that the threshold is reached eventually* by replacing the numerator of the exponential term by $-(d - |\mu|t)^2$. However, the “improper” distribution is actually more useful to us, as it reflects the full, unconditional probabilities of extinction. See Lande and Orzack 1988.

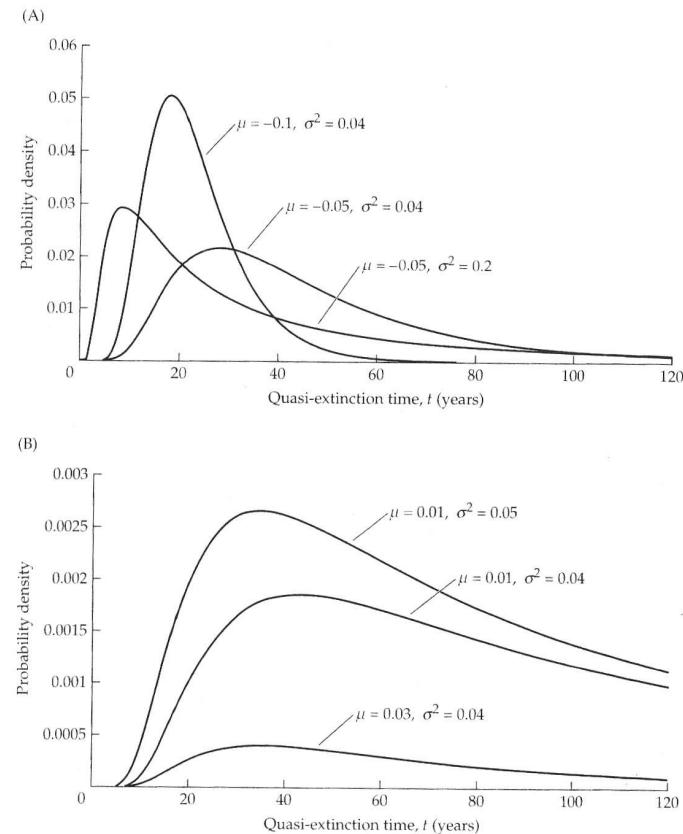


Figure 3.4 The inverse Gaussian distribution (Equation 3.3) with (A) $\mu < 0$ and (B) $\mu > 0$. Here, the current population size and the quasi-extinction threshold are $N_c = 10$ and $N_x = 1$, respectively.

are unlikely to ever hit the quasi-extinction threshold. It is only early in the growth process, when population size is still low, that a chance string of bad years can push the population down to the threshold. So, extinction is unlikely to happen, but if it does, it will occur rapidly.

The inverse Gaussian distribution gives the probability that quasi-extinction occurs in a very small time segment. To calculate the probability that the threshold is reached at any time between the present ($t = 0$) and a future time of interest ($t = T$), we integrate Equation 3.4 from $t = 0$ to $t = T$ (that is, we sum up the probabilities for each small interval of time). The result is the cumulative distribution function (or CDF) for the time to quasi-extinction,

$$G(T | d, \mu, \sigma^2) = \Phi\left(\frac{-d - \mu T}{\sqrt{\sigma^2 T}}\right) + \exp(-2\mu d/\sigma^2) \Phi\left(\frac{-d + \mu T}{\sqrt{\sigma^2 T}}\right) \quad (3.5)$$

where $\Phi(z)$ is the standard normal cumulative distribution function

$$\Phi(z) = \left(1/\sqrt{2\pi}\right) \int_{-\infty}^z \exp(-y^2/2) dy \quad (3.6)$$

$\Phi(z)$ is simply the area from $-\infty$ to z under a normal curve with a mean of zero and a variance of 1. Values of $\Phi(z)$ are tabulated in most basic probability reference books (e.g., Abramowitz and Stegun 1964), and can be calculated using built-in functions in mathematical software packages such as MATLAB and even spreadsheet programs such as Excel.

Figure 3.5 illustrates how the probability of hitting the threshold according to Equation 3.5 increases as the time horizon T is moved farther into the future. If μ is substantially negative, the probability of extinction increases rapidly with time, quickly reaching a value near 1. If the time horizon is not far into the future, increasing the value of σ^2 increases the probability that extinction will have occurred before the horizon is reached (once again, greater year-to-year variability leads to higher extinction risk). For positive μ , the probability of extinction increases slowly with time, and never reaches a value of 1. The probability of ultimate extinction (that is, at a time horizon of infinity) can be calculated by taking the integral of the inverse Gaussian distribution from $t = 0$ to $t = \infty$:

$$G(\infty | d, \mu, \sigma^2) = \begin{cases} 1 & \text{if } \mu \leq 0 \\ \exp(-2\mu d/\sigma^2) & \text{if } \mu > 0 \end{cases} \quad (3.7)$$

Thus if μ is zero or negative, ultimate extinction is certain, and increasing σ^2 has no effect on the probability that the threshold is eventually reached. In contrast, if μ is positive, ultimate extinction is not a certainty, and increasing σ^2 increases the probability of ultimate extinction.

The rest of this chapter is devoted to the calculation of extinction risk estimates using count data, but even in the absence of any data, knowledge of

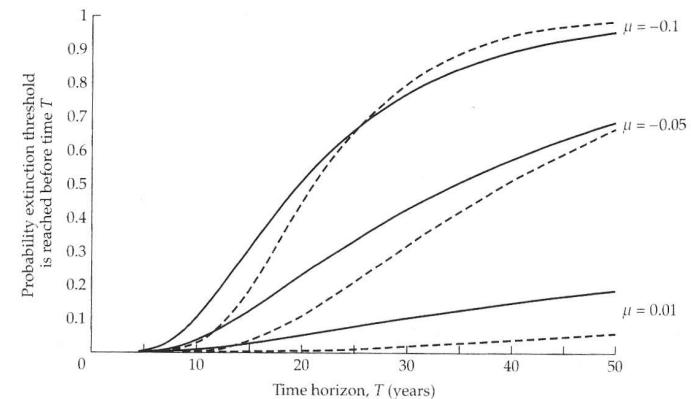


Figure 3.5 Cumulative distribution function (Equation 3.4) for the time to reach an extinction threshold of $N_x = 1$ from an initial population of $N_c = 10$; $\sigma^2 = 0.04$ (dashed lines) or 0.08 (solid lines).

how the CDF is affected by its underlying parameters can help us to make useful *qualitative* assessments of relative viability for a species of critical concern, especially if we can use natural history information to make inferences about the local environment or about the life history of the species in question. For example, we will frequently be able to make an educated guess that the environment one population experiences is likely to be more variable than another's in ways that will affect population growth. Such differences in σ^2 influence the extinction time CDF even when its other determinants (μ and the starting and threshold population sizes) are fixed (Figure 3.5). Similarly, some species will have life history features (e.g., long-lived adults) that buffer their populations against year-to-year environmental variation, while others do not. With a low σ^2 , values of μ that are only slightly positive are sufficient to minimize extinction risks, whereas with high variance, much greater mean growth is needed. Thus we can state that, all else being equal, the greater the environmentally driven fluctuations in population growth rate, the greater will be the risk of extinction, especially for short-term time horizons, a statement that, though qualitative, nonetheless provides some useful guidance.

Before going on, it is also important to recognize that three key assumptions underlie the use of a diffusion approximation to derive the key Equations 3.4, 3.5, and 3.7. First, we assume that the environmental perturbations affecting the population growth rate are small to moderate, just as gas mole-

cules typically diffuse through a series of small jumps.⁶ In other words, we assume that large catastrophes or bonanzas (Chapter 2) do not occur frequently enough to be important. Second, we assume that the changes in population size are independent between one time interval and the next; that is, we assume that strings of good or bad years occur no more frequently than would be expected by chance. Third, we assume that the values of μ and σ^2 do not change over time, either in response to changes in population density or because of temporal trends in environmental conditions. We show how to test for violations of these assumptions in the following section, and how to deal with some of these violations (should they occur) in Chapter 4.

Using Count Data to Estimate the Population Growth Parameters μ and σ^2 : An Illustration Using the Yellowstone Grizzly Bear Census

We argued in Chapter 2 that the extinction time cumulative distribution function or CDF, given by Equation 3.5 and illustrated in Figure 3.5, is perhaps the single most useful metric of a population's risk of extinction. To calculate it, we need only four quantities: the current population size N_i , the extinction threshold $N_{X'}$, and the values of μ and σ^2 . We turn now to methods for estimating μ and σ^2 from a series of census counts.

Let us assume that we have conducted a total of $q + 1$ annual censuses of a population at times $t_0, t_1, t_2, \dots, t_q$, having obtained the census counts $N_0, N_1, N_2, \dots, N_q$. The censuses need not have been conducted in consecutive years, but in a seasonal environment, the censuses should have been performed at the same time of year. Although μ and σ^2 can be used to describe future population size distributions (as shown in Figure 3.3), they are more directly measures of the mean and variance of the *change* in log population sizes. Thus it is not the raw census counts themselves that we want to use to estimate μ and σ^2 , but the amount by which the natural logarithms of the counts change over each of the inter-census intervals. For example, over the time interval of length $(t_{i+1} - t_i)$ years between censuses i and $i + 1$, the logs of the counts change by an amount

$$\log N_{i+1} - \log N_i = \log(N_{i+1}/N_i) = \log \lambda_i \quad (3.8)$$

⁶Technically, diffusion models also assume that the state variable (in this case population size) is continuous, whereas in reality, a population can only contain an integer number of individuals. The assumption of small environmental perturbations is therefore tantamount to assuming that the *change* in numbers over a short time interval *relative* to the total number of individuals in the population is small, an assumption that is likely to break down at small population sizes (Ludwig 1996a). Using a moderate quasi-extinction threshold, as we advocated in Chapter 2, ameliorates to some extent the effects of this assumption.

In Equation 3.8, $\lambda_i = N_{i+1}/N_i$ is the population growth rate between census i and census $i + 1$, emphasizing the fact that it is the logs of the population growth rates that we will use to estimate μ and σ^2 . The counts from $q + 1$ censuses allow us to calculate log population growth rates for q time intervals, although as noted above, these intervals need not be of the same length.

An example of count-based data that we will analyze extensively in this chapter comes from the ongoing census of adult female grizzly bears (*Ursus arctos*) in the Greater Yellowstone ecosystem. (For an earlier analysis of the same population, see Dennis et al. 1991.) This population is currently designated as threatened by the United States Fish and Wildlife Service, and is completely isolated from other grizzly bear populations. Each year, bear biologists count the number of unique female bears with cubs (offspring in their first year of life) in the entire Yellowstone population. Because the litters of one to three cubs remain closely associated with their mothers, females with cubs are the most conspicuous, and therefore most reliably censused, element of the population. Censuses were originally performed by observing bears at park garbage dumps, but they have been conducted by aerial survey since the closure of the dumps in 1970–1971. The counts of females with cubs are used to estimate the total number of adult females in the population. Specifically, the number of adult females in year t is estimated as the sum of the observed number of females with cubs in years t , $t+1$, and $t+2$ (Eberhardt et al. 1986). The logic underlying this estimate is that the interval between litters produced by the same mother is at least three years, so that females with cubs observed in years $t+1$ and $t+2$ could not have been the same individuals that were observed with cubs in year t . Yet, if they are observed in later years, they must have been alive in year t (although they may not have been *adults* in year t , which introduces some error). The estimated numbers of adult females from 39 annual censuses of the Yellowstone population, beginning in 1959, are shown in Figure 3.6 and listed in Table 3.1. (These data can also be obtained at the Interagency Grizzly Bear Study Team's website: www.nrmrc.usgs.gov/research/igbst/pub.htm.)

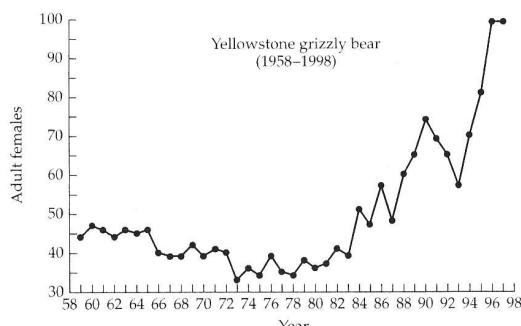
We now review two methods to estimate μ and σ^2 , illustrating both methods with the grizzly bear data in Table 3.1.

Estimating μ and σ^2 as the mean and variance of $\log(N_{i+1}/N_i)$

If the censuses were conducted at regular yearly intervals, as was the Yellowstone grizzly bear census, the simplest method of estimating μ and σ^2 is to calculate the arithmetic mean and the sample variance of the $\log(N_{i+1}/N_i)$ values, given respectively by

$$\hat{\mu} = \frac{1}{q} \sum_{i=0}^{q-1} \log(N_{i+1}/N_i) \quad \text{and} \quad \hat{\sigma}^2 = \frac{1}{q-1} \sum_{i=0}^{q-1} (\log(N_{i+1}/N_i) - \hat{\mu})^2 \quad (3.9)$$

Figure 3.6 Census counts of adult female grizzly bears in the Greater Yellowstone population (data from Eberhardt et al. 1986 and Haroldson 1999).



(where the hats over $\hat{\mu}$ and $\hat{\sigma}^2$ indicate that they are estimates). Notice that in calculating $\hat{\sigma}^2$, we divide by $q-1$ instead of q , thus using the typical formula to obtain an unbiased estimate of variance from a sample of data (Dennis et al. 1991⁷). The values of $\hat{\mu}$ and $\hat{\sigma}^2$ are easily obtained by applying the functions AVERAGE and VAR (respectively) in Microsoft Excel to the log (N_{i+1}/N_i) values, or by using similar routines in any statistical package.

The estimates $\hat{\mu}$ and $\hat{\sigma}^2$ for the Yellowstone grizzly bear calculated using Equations 3.9 are given in Table 3.1. The value of $\hat{\mu}$ is positive, reflecting a general upward trend in the number of adult female bears over the sampling period (Figure 3.6). The value of $\hat{\sigma}^2$ is about 60% of the value of $\hat{\mu}$; recall that $\hat{\sigma}^2$ is a measure of the year-to-year variability in the census counts.

Estimating μ and σ^2 by Linear Regression

If the censuses were not taken at equal time intervals, we should not use Equations 3.9 to estimate μ and σ^2 , because they do not account for the fact that both the mean and variance in population change should be larger for pairs of censuses that are separated by longer intervals of time.⁸ In such cases, we can estimate the parameters using a linear regression method proposed by Dennis et al. (1991). In fact, the linear regression method, although somewhat more complicated to execute, has advantages over Equations 3.9 even when the censuses occurred every year. In particular, the output produced by most widely available regression packages allows us to easily

⁷We use the symbol $\hat{\sigma}^2$ to represent the unbiased estimate of σ^2 . Dennis et al. (1991) use $\hat{\sigma}^2$.

⁸An alternate formula that does account for the length of the time interval is presented in Dennis et al. 1991 (see their Equations 24 & 25), but as it does not have the advantages of their regression approach, we do not present it here.

TABLE 3.1 Estimated number of adult female grizzly bears in the Greater Yellowstone population^a

Census, <i>i</i>	Year, <i>t_i</i>	Adult females, <i>N_i</i>	$\log(N_{i+1}/N_i)$
1	1959	44	0.0660
2	1960	47	-0.0215
3	1961	46	-0.0445
4	1962	44	0.0445
5	1963	46	-0.0220
6	1964	45	0.0220
7	1965	46	-0.1398
8	1966	40	-0.0253
9	1967	39	0.0000
10	1968	39	0.0741
11	1969	42	-0.0741
12	1970	39	0.0500
13	1971	41	-0.0247
14	1972	40	-0.1924
15	1973	33	0.0870
16	1974	36	-0.0572
17	1975	34	0.1372
18	1976	39	-0.1082
19	1977	35	-0.0290
20	1978	34	0.1112
21	1979	38	-0.0541
22	1980	36	0.0274
23	1981	37	0.1027
24	1982	41	-0.0500
25	1983	39	0.2683
26	1984	51	-0.0817
27	1985	47	0.1929
28	1986	57	-0.1719
29	1987	48	0.2231
30	1988	60	0.0800
31	1989	65	0.1297
32	1990	74	-0.0700
33	1991	69	-0.0597
34	1992	65	-0.1313
35	1993	57	0.2054
36	1994	70	0.1460
37	1995	81	0.2007
38	1996	99	0.0000
39	1997	99	0.0000

Mean: $\hat{\mu} = 0.02134$
Sample variance: $\hat{\sigma}^2 = 0.01305$

^aData from Eberhardt et al. 1986 and Haroldson 1999.

place confidence limits on the parameters and to test assumptions of the underlying PVA model, as we now demonstrate.

REGRESSION PROCEDURE The basic idea of this method is to regress the log population growth rate over a time interval against the amount of time elapsed. However, the most straightforward way to do this does not conform to one of the assumptions of standard linear regression, namely that the variance in the dependent variable (log population growth, in this case) is constant over different values of the independent variable (time elapsed). As we noted above, the variance of log population size increases with time (see Figure 3.3), which means that the variance in population *change* over a time interval is also dependent on the time elapsed. In particular, if two censuses are separated by $t_{i+1} - t_i$ years, then $\log(N_{i+1}/N_i)$ has a variance of $\sigma^2(t_{i+1} - t_i)$. To make our regression conform to the assumption of equal variances, we need to transform the rate of population change (and also the time elapsed) to get rid of this time dependence. This is accomplished by dividing both the log population growth rate and the time elapsed by $\sqrt{t_{i+1} - t_i}$, which makes the variance in the transformed population change variable equal to σ^2 for any time interval.

Thus, to perform the linear regression, we first calculate a transformation of the length of each time interval to use as the independent variable:

$$x_i = \sqrt{t_{i+1} - t_i} \quad (3.10a)$$

We then calculate a transformed variable of population change, the dependent variable, as:

$$y_i = \log(N_{i+1}/N_i)/\sqrt{t_{i+1} - t_i} = \log(N_{i+1}/N_i)/x_i \quad (3.10b)$$

For example, if we have entered the values of t_i and N_i from Table 3.1 in the first 39 rows of Columns A and B in an Excel worksheet, we can perform the first 39 transformations in Equations 3.10a,b by entering the formulas =SQRT(A2-A1) in cell C1 and =LN(B2/B1)/C1 in cell D1 and then “filling down” the subsequent 38 rows of Columns C and D. Notice that when all adjacent censuses are one year apart, as in the grizzly bear data set, all the x values will equal 1, and y_i will equal $\log(N_{i+1}/N_i)$.

The final step is to perform a linear regression of the y_i 's against the x_i 's, *forcing the regression intercept to be zero*. (In some statistics packages, such as Excel's regression routine, this is accomplished by choosing “no constant” in the regression options.) By fixing the regression intercept at zero, we are enforcing the rule that there can be no change in population size if no time has elapsed. The slope of the regression is an estimate of μ and the regression's error mean square estimates σ^2 (Figure 3.7). For example, the following SAS command applied to the transformed grizzly bear data generates

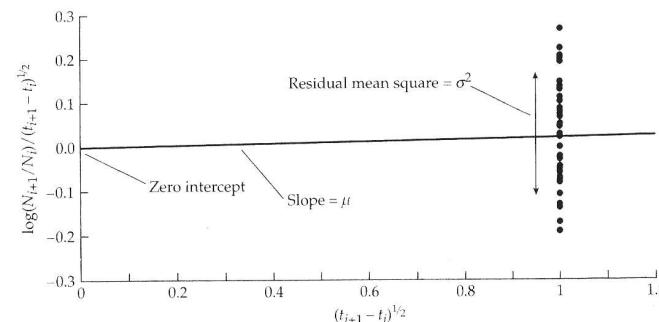


Figure 3.7 Estimating μ and σ^2 for the adult female population of Yellowstone grizzly bears using linear regression.

the output in Box 3.1:

```
proc reg;
  model y=x / noint dw influence;
run;
```

(3.11)

The command `noint` induces a regression with no intercept, and the commands `dw` and `influence` instruct SAS to print regression diagnostics of which we will make use below. In Box 3.1, the Parameter Estimate for the Variable x (0.021337) is the regression slope, which is our estimate of μ , and the second entry (0.01305) in the column Mean Square in the Analysis of Variance table is the error mean square, which is our estimate of σ^2 . The output in Box 3.1 is equivalent to that produced by virtually all good statistics packages, although terminology may differ. (For example, the error mean square is sometimes labeled “residual mean square”.) Notice that the regression yields the same estimates for μ and σ^2 that we obtained in Table 3.1. However, the regression output also provides additional useful information, as we now review.

USING REGRESSION OUTPUT TO CONSTRUCT CONFIDENCE INTERVALS FOR μ AND σ^2 Ideally, we want more than just the single best estimates of μ and σ^2 . Because these will only be *estimates* based on the limited number of transitions in the data set, we would also like to know how much confidence we can place in them. Confidence intervals provide an upper and lower value between which the true value of each parameter is likely to lie. Confidence limits help us to place the best estimates of each parameter in context. For

BOX 3.1 Output produced by SAS commands given in Equation 3.11¹.

```

Model: MODEL1
NOTE: No intercept in model. R-square is redefined.
Dependent Variable: Y

Analysis of Variance

Source          DF      Sum of Squares      Mean Square      F Value      Prob>F
Model            1        0.01730       0.01730       1.325       0.2570
Error           37       0.48295       0.01305
U Total          38       0.50025

Root MSE      0.11425      R-square      0.0346
Dep Mean      0.02134      Adj R-sq      0.0085
C.V.          535.45064

Parameter Estimates

Variable        Parameter      Standard      T for H0:
                Estimate       Error      Parameter=0      Prob > |T|
X                 1        0.021337    0.01853351     1.151       0.2570

Durbin-Watson D      2.570
(For Number of Obs.)      38
1st Order Autocorrelation      -0.288

```

¹Output referred to in the text is in bold.

example, if $\hat{\mu}$, the best estimate of μ , is positive, we would predict that most population trajectories will grow (as in Figure 3.3A). However, if the lower limit of the confidence interval for μ is negative, we cannot, based on the available data, rule out the possibility that the population will actually tend to decline over the long term.

In a proper linear regression, the estimate of the slope (i.e., $\hat{\mu}$, the estimate of μ) is normally distributed around its true value. Using this fact, many statistics packages will provide 95% confidence limits or standard errors for the regression slope. If yours does not, or if you want to calculate different confidence limits (say, 90 or 99% limits) it is easy to do so using the following formula:

$$\left(\hat{\mu} - t_{\alpha/2} \text{SE}(\hat{\mu}), \quad \hat{\mu} + t_{\alpha/2} \text{SE}(\hat{\mu}) \right) \quad (3.12)$$

(Dennis et al. 1991). Here, $t_{\alpha, q-1}$ is the critical value of the two-tailed Student's t distribution with a significance level of α and $q-1$ degrees of freedom. For example, if we want to place a 95% confidence interval around μ for the Yellowstone grizzly bear, we would set α equal to 0.05 and q equal to 38 (the number of transitions on which the regression was performed; Table 3.1), and then obtain the value of $t_{0.05, 37}$. We can look up this value in

BOX 3.1 (continued)

Obs	Residual	Rstudent	Hat	Diag	Cov	X
			H	Ratio	Dffits	Dfbetas
1	0.0447	0.3916	0.0263	1.0511	0.0644	0.0644
2	-0.0428	-0.3755	0.0263	1.0514	-0.0617	-0.0617
3	-0.0658	-0.5787	0.0263	1.0458	-0.0951	-0.0951
4	0.0232	0.2028	0.0263	1.0544	0.0333	0.0333
5	-0.0433	-0.3799	0.0263	1.0513	-0.0625	-0.0625
6	0.000663	0.0058	0.0263	1.0556	0.0010	0.0010
7	-0.1611	-1.4505	0.0263	0.9973	-0.2385	-0.2385
8	-0.0466	-0.4090	0.0263	1.0507	-0.0672	-0.0672
9	-0.0213	-0.1868	0.0263	1.0545	-0.0307	-0.0307
10	0.0528	0.4630	0.0263	1.0493	0.0761	0.0761
11	-0.0954	-0.8432	0.0263	1.0351	-0.1386	-0.1386
12	0.0287	0.2510	0.0263	1.0537	0.0413	0.0413
13	-0.0460	-0.4037	0.0263	1.0508	-0.0664	-0.0664
14	-0.2137	-1.9682	0.0263	0.9530	-0.3236	-0.3236
15	0.0657	0.5772	0.0263	1.0459	0.0949	0.0949
16	-0.0785	-0.6917	0.0263	1.0417	-0.1137	-0.1137
17	0.1159	1.0286	0.0263	1.0254	0.1691	0.1691
18	-0.1295	-1.1542	0.0263	1.0179	-0.1897	-0.1897
19	-0.0503	-0.4416	0.0263	1.0499	-0.0726	-0.0726
20	0.0899	0.7931	0.0263	1.0374	0.1304	0.1304
21	-0.0754	-0.6641	0.0263	1.0428	-0.1092	-0.1092
22	0.00606	0.0531	0.0263	1.0555	0.0087	0.0087
23	0.0814	0.7170	0.0263	1.0407	0.1179	0.1179
24	-0.0713	-0.6276	0.0263	1.0441	-0.1032	-0.1032
25	0.2470	2.3163	0.0263	0.9186	0.3808	0.3808
26	-0.1030	-0.9119	0.0263	1.0317	-0.1499	-0.1499
27	0.1716	1.5504	0.0263	0.9895	0.2549	0.2549
28	-0.1932	-1.7622	0.0263	0.9717	-0.2897	-0.2897
29	0.2018	1.8471	0.0263	0.9642	0.3037	0.3037
30	0.0587	0.5152	0.0263	1.0478	0.0847	0.0847
31	0.1084	0.9602	0.0263	1.0292	0.1579	0.1579
32	-0.0913	-0.8064	0.0263	1.0368	-0.1326	-0.1326
33	-0.0810	-0.7140	0.0263	1.0408	-0.1174	-0.1174
34	-0.1526	-1.3699	0.0263	1.0033	-0.2252	-0.2252
35	0.1841	1.6718	0.0263	0.9795	0.2748	0.2748
36	0.1247	1.1092	0.0263	1.0207	0.1824	0.1824
37	0.1794	1.6260	0.0263	0.9833	0.2673	0.2673
38	-0.0213	-0.1868	0.0263	1.0545	-0.0307	-0.0307

most standard statistics texts or compute it with the aid of mathematical software such as MATLAB⁹; even many spreadsheet programs can calculate it. For example, entering the formula =TINV(0.05, 37) into a cell in a Microsoft Excel worksheet yields the value $t_{0.05, 37} = 2.0262$. In Expression 3.12, SAS supplies $SE(\hat{\mu})$ is the standard error of $\hat{\mu}$, the estimated regression slope. SAS supplies it in the regression output (Box 3.1) in the column labeled Standard Error; it is 0.0185. The standard error can also be calculated directly from the expression $\sqrt{\hat{\sigma}^2/t_q}$, where $\hat{\sigma}^2$ is the regression error mean square and t_q is the duration of the census in years (Dennis et al. 1991); for the Yellowstone grizzly bear data in Table 3.1, the census spans $t_q = 38$ years, from 1959 to 1997. Substituting the values of $\hat{\mu}$, $t_{0.05, 37}$, and $SE(\hat{\mu})$ that we have just calculated into Expression 3.12 gives us a 95% confidence interval of $(-0.01621, 0.05889)$ for the parameter μ . The way to view this confidence interval is to imagine repeatedly sampling at random 38 log growth rates from this population; the confidence intervals computed using Equation 3.12 would include the true value of μ 95% of the time. The fact that the confidence interval we have calculated ranges from negative to positive values indicates that, despite strong growth of the population in recent years (Figure 3.6) and the relatively long-term nature of the census, we still cannot rule out the possibility that the population will decline. Notice that the regression output in Box 3.1 also provides a test of the null hypothesis that the slope of the regression is zero. The relatively large probability associated with this test (0.2570) also tells us that we cannot say definitively that the population is growing, even though $\hat{\mu}$ is positive.

The results in the preceding paragraph highlight an important conservation lesson: We should be very skeptical about single estimates of viability metrics that are presented without an associated measure of uncertainty, such as a confidence interval. The case of the Yellowstone grizzly bear is particularly apropos. Recent years have witnessed intense pressure to remove the Yellowstone grizzly bear population from the Endangered Species list; much of that pressure is motivated by a desire to show that the Endangered Species Act functions effectively to bring about recovery of threatened and endangered species and populations. However, our analysis shows that it would be dangerous to conclude from the data in Table 3.1 that, because the best estimate of μ is slightly positive, the Yellowstone grizzly bear population is growing and can be safely delisted. The lower confidence limit on μ says that such a conclusion would be premature.

To know how much confidence we can place in extinction probabilities based on the estimated values of μ and σ^2 , we also need to calculate a confidence interval for σ^2 . Such an interval can be constructed using the chi-

⁹Specifically, the MATLAB function `tiny` will compute the inverse t statistic, but users must purchase the Statistics Toolbox to use `tiny`.

square distribution (Dennis et al. 1991). For example, if we want to place a 95% confidence interval around σ^2 for the Yellowstone grizzly bear, for which we have data on $q = 38$ transitions, we first obtain the 2.5th and 97.5th percentiles of the chi-square distribution with $q - 1 = 37$ degrees of freedom: call them $\chi^2_{0.025, 37}$ and $\chi^2_{0.975, 37}$. Once again, these can be looked up in a statistics reference or calculated with the aid of MATLAB¹⁰ or a spreadsheet program; the Excel formulae =CHIINV(0.025, 37) and =CHIINV(0.975, 37) yield the values $\chi^2_{0.025, 37} = 55.6680$ and $\chi^2_{0.975, 37} = 22.1056$. We substitute these values into the following general expression for a 95% confidence interval for σ^2 based on $q - 1$ degrees of freedom:

$$\left((q-1) \hat{\sigma}^2 / \chi^2_{0.025, q-1}, (q-1) \hat{\sigma}^2 / \chi^2_{0.975, q-1} \right) \quad (3.13)$$

(Dennis et al. 1991). Note that we can use this same approach to calculate confidence intervals with other degrees of coverage. For example, to calculate a 90% rather than a 95% confidence interval, we would substitute the values of $\chi^2_{0.05, q-1}$ and $\chi^2_{0.95, q-1}$ into the first and second terms in Equation 3.13, respectively.

For the Yellowstone grizzly bear, Equation 3.13 gives us the confidence interval (0.00867, 0.02184) for σ^2 . Thus a true value of nearly 70% higher than the best estimate of $\sigma^2 = 0.01305$ is consistent with the available data (as is a much smaller value). We will return to Equations 3.12 and 3.13 when we discuss how to incorporate parameter uncertainty into estimates of extinction probabilities.

USING REGRESSION DIAGNOSTICS TO TEST FOR TEMPORAL AUTOCORRELATION IN THE POPULATION GROWTH RATE Another advantage of estimating the parameters μ and σ^2 by regression is that we can test the assumption of the diffusion approximation that the environmental conditions (and thus the log population growth rates) are uncorrelated from one inter-census interval to the next (that is, whether a particular interval was good or bad for birth or death is independent of whether preceding or succeeding intervals were good or bad). One test makes use of the Durbin-Watson d statistic (invoked by the `dw` option in SAS), which measures the strength of autocorrelation in the regression residuals (and we are interpreting the residuals as being the result of environmental perturbations). A residual is simply an observed value of y (i.e., the value of $\log(N_{i+1}/N_i)/\sqrt{t_{i+1}-t_i}$ for census i) minus the value predicted by the regression equation (which equals $\mu\sqrt{t_{i+1}-t_i}$). The two-tailed Durbin-Watson test evaluates the null hypothesis that all of the serial autocorrelations of the residuals are zero (Draper and Smith 1981). To

¹⁰Specifically, you can use the function `chi2inv` if you have purchased MATLAB's Statistics Toolbox.

perform the test, we compare both d and $4 - d$ to upper and lower critical values d_L and d_U . Specifically, if $d < d_L$ or $4 - d < d_U$, we conclude that the residuals show significant autocorrelation; if $d > d_U$ and $4 - d > d_U$, we conclude there is no significant autocorrelation; and otherwise, the test is inconclusive. Values of d_L and d_U for a significance level of $\alpha = 0.05$ and different numbers of data points in the regression, q , are given in Table 3.2; for other values of q , d_L and d_U can be calculated by linear interpolation, or by referring to Table 3.3 in Draper and Smith 1981. For the Yellowstone grizzly bear, $q = 38$, so that $d_L = 1.33$ and $d_U = 1.44$. The calculated value of d is 2.57 (Box 3.1). Because $d > d_U$ but $4 - d$ nearly equals d_U , this value of d leads to a test result that is right on the border between inconclusive and a finding of no significant autocorrelation. We conclude that there may be some autocorrelation in the residuals, but that it is probably weak. Another sign that the autocorrelation in the residuals is at best weak is the fact that the first-order autocorrelation of the residuals¹¹ ($r = -0.288$ with 37 degrees of freedom; see Box 3.1) has a small but nonsignificant p value ($p = 0.084$; this p value is calculated by most statistical packages and can be looked up in tables of two-tailed significance levels of the Pearson correlation coefficient, given in most basic statistics references). Given the possibility that successive log population growth rates may be slightly correlated (in this case negatively), a thorough analysis of the data should include an examination of the possible effect of the observed level of autocorrelation on the estimated viability of the grizzly bear population, using methods we describe in Chapter 4.

USING REGRESSION DIAGNOSTICS TO TEST FOR OUTLIERS We can also use the regression output to evaluate whether any particular transitions are outliers or are having a disproportionate influence on the parameter estimates. There are a variety of measures of the importance of a data point on the results of a regression. These include Cook's distance measure, leverage, and studentized residuals; different statistics packages provide different combinations of these measures. For example, the "influence" option in `proc reg` of SAS produces an output table with two useful measures of the effect of each data point on the regression results (Box 3.1). The column labeled `Rstudent` gives the studentized residual for each (x,y) pair in the regression. Data points with studentized residuals greater than 2 are suspected outliers (SAS 1990). `Dffits` is a statistic that measures the influence each data point has on the regression parameter estimates; for a linear regression with no intercept, a

¹¹This autocorrelation can be computed easily as follows. Place the residuals for years 1 to $q - 1$ in one column of a spreadsheet, and the residuals for years 2 to q in the adjacent column. Then compute the Pearson correlation coefficient of the two columns of numbers. SAS performs this calculation for you when you specify the `dw` option in `proc reg` (see Box 3.1).

TABLE 3.2 Critical values for a two-tailed test of the Durbin-Watson statistic d at the $\alpha = 0.05$ significance level (from Table 3.3 in Draper and Smith 1981). q is the number of data points in a linear regression with one independent variable

q	d_U	d_L
15	0.95	1.23
20	1.08	1.28
25	1.18	1.34
30	1.25	1.38
35	1.30	1.42
40	1.35	1.45
45	1.39	1.48
50	1.42	1.50

value of `Dffits` greater than $2\sqrt{1/q}$ (where q is the number of data points, or transitions) suggests high influence (SAS 1990). For the grizzly bear data set, the critical value of `Dffits` is $2\sqrt{1/38} = 0.324$. By both the studentized residual and `Dffits` criteria, the 25th transition in the data set, corresponding to the 1983–1984 censuses, is flagged as an unusually large transition with a high influence on the estimate of μ . This transition represents a large proportional increase, in which the counts grew by more than 50% in a single year (Figure 3.6). Dennis et al. (1991) identified this same transition as an outlier in their analysis of the 1959–1987 census data.

If an outlier or influential observation is identified in your data set, you must decide whether to keep that data point or remove it from the regression. To make this decision, we recommend first examining why that point may be unusual. If the cause is "nonbiological," it might be sensible to delete the data point. For example, if unusual transitions are associated with years in which methodological problems with the census were known to have occurred, or if they correspond to a change from one method of censusing the population to another, they may represent errors in the counting process rather than periods of unusually high or low population growth. As we want our parameter estimates to include real environmental variation but to exclude observation error as much as possible (see Chapter 5), deleting such points would be justified. In the case of the Yellowstone grizzly bear, we are not aware of any such problems that occurred in the 1983 or 1984 censuses; hence we see no reason to delete the 25th transition as being unduly influenced by observation error.

On the other hand, there may be real biological reasons why a particular transition is an outlier. These reasons can be classified as one-time human impacts and extreme environmental conditions; the way we should treat

these two types of outliers differs. If an outlier can be pegged to a one-time human impact that we do not expect to recur, that outlier should be omitted for the purposes of predicting the future state of the population. A hypothetical example of such an impact on the Yellowstone grizzly bear would be effects of closing the park garbage dumps (at which the bears had become accustomed to scavenge for food). Because this event is unlikely to occur again, any unusual transitions associated with it should not be allowed to influence our predictions for the future. The unusually high 1983–1984 transition, however, is not likely to be related to the closure of the dumps, which occurred in 1970–1971.

Finally, outliers could represent the consequences of extreme environmental conditions, either catastrophes or bonanzas (Chapter 2). Recall that the diffusion approximation used to derive an expression for the cumulative probability of extinction (Equation 3.4) assumed that such large fluctuations in the population growth rate do not occur. Strictly speaking, then, we should not use Equation 3.5 to calculate an extinction probability if our estimates of μ and σ^2 have been influenced by catastrophes or bonanzas. Instead, we must resort to computer simulation to calculate extinction probabilities, as we describe in Chapter 4.

In the case of the Yellowstone grizzly bear, the only suspected outlier is an unusually large *increase* in population size. Large increases inflate the estimated values of both μ and σ^2 . As a higher value of μ reduces the probability of extinction but a higher value of σ^2 increases it (Figure 3.5), it is not clear how these two effects of the outlier on the extinction probability play out. Because of this, and because no environmental factors that could have caused the 1983–1984 transition to be a bonanza have been identified, we have chosen to retain this transition when estimating μ and σ^2 . However, readers are encouraged to explore how the probability of extinction changes if we use parameter estimates ($\hat{\mu} = 0.01467$ and $\hat{\sigma}^2 = 0.01167$) obtained after deleting the 25th transition.

As the foregoing discussion indicates, there is no general rule regarding when to omit outliers, and the decision to do so must be decided on a case-by-case basis. Our point here is simply that the regression approach makes this decision possible by identifying outliers in the first place. When outliers are omitted, it is important that you state this fact explicitly in any reports you prepare on your PVA results, and that you carefully document why those points were omitted.

USING REGRESSION DIAGNOSTICS TO TEST FOR PARAMETER CHANGES A final advantage to the linear regression approach is that it provides a way to ask whether μ or σ^2 differ significantly in distinct segments of the time series. For example, for the Yellowstone grizzly bear, we might ask whether the parameter estimates differ before versus after the closure of the garbage dumps in 1970–1971, or before versus after the 1988 Yellowstone fires.

To test for changes in μ and σ^2 before versus after a pivotal census (e.g., 1988, the year of the fires), we recommend that you perform linear regressions in the following order. The first step is to test for differences in σ^2 before and after the pivotal census. To do this, divide the transitions into those leading up to and including the pivotal census versus those from the pivotal census onward. Transform each data set using Equations 3.10a,b. Next, perform separate linear regressions with zero intercept on these two transformed data sets (e.g., using two repetitions of SAS Command 3.9 above) to obtain two estimates of σ^2 ; call them $\hat{\sigma}_b^2$ and $\hat{\sigma}_a^2$ (the subscripts here stand for “before” and “after”). For example, if we divide the Yellowstone grizzly bear censuses into “before-fire” (1959–1988) and “after-fire” (1988–1997) periods, we obtain the estimates $\hat{\sigma}_b^2 = 0.01229$ and $\hat{\sigma}_a^2 = 0.01561$ from the error mean squares of the separate regressions. There are 29 and 9 transitions in these two data sets, respectively. We can perform a two-tailed test for a significant difference between two variances (in this case $\hat{\sigma}_b^2$ and $\hat{\sigma}_a^2$) by calculating their ratio with the larger variance in the numerator and then computing its probability from an *F* distribution with the appropriate numerator and denominator degrees of freedom (Snedecor and Cochran 1980). For the grizzly bear data, the ratio $\hat{\sigma}_a^2 / \hat{\sigma}_b^2 = 1.270$ has 9 – 1 = 8 numerator and 29 – 1 = 28 denominator degrees of freedom. The probability of observing a ratio this large or larger if the two data sets have the same true value of σ^2 is 0.298 (obtained from an *F* table available in any statistics reference or by using the Excel formula =FDIST(1.270,8,28)). As this probability is larger than 0.025,¹² we conclude that there is no significant difference in the degree of variability in the grizzly bear population growth rate before versus after the 1988 Yellowstone fires.

Had the σ^2 values differed, we would use the two separate regressions to make different estimates of both μ and σ^2 and conduct any analysis of future trends using only the more recent set of estimates. (However, we would not if, for some reason, we thought that the earlier population dynamics might be a better indicator of what will happen in the future, for example if a threat to the population that was present only during the later time interval has now been removed.) In this case, however, we have determined that σ^2 does not differ significantly before versus after the fires, but we still need to test for differences in μ while allowing only a single estimate of σ^2 that applies to the entire census. We can do this by creating two *x* variables (call them x_1 and x_2) and performing a multiple linear regression of *y* on x_1 and x_2 , once again forcing the regression to have an intercept of zero (Dennis et al. 1991). The new variable x_1 will have the values of x_i calculated from Equation 3.10a for all years before the fires and zero for all years after the fires. In contrast,

¹²A probability of 0.025 corresponds to an overall significance level of $\alpha = 0.05$ for a two-tailed *F* test (Snedecor and Cochran 1980).

variable x_2 will be zero for years before the fires and have values calculated from Equation 3.10a for all years after the fires. The y values should be computed from Equation 3.10b as before. The following SAS command performs this multiple regression of y on x_1 and x_2 , yielding the output in Box 3.2:

```
proc reg;
  model y=x1 x2 / noint;
run;
```

(3.14)

The parameter estimates associated with variables x_1 and x_2 give us $\hat{\mu}_b = 0.01069$ and $\hat{\mu}_a = 0.05564$ for the periods before and after the fires, respectively. The single estimate of σ^2 for the entire census, once again obtained from the error mean square, is $\hat{\sigma}_m^2 = 0.01303$ (the subscript m here indicates that this estimate was obtained from a multiple regression; it is not necessarily the same as the estimate obtained from a single regression, as in Box 3.1). As the estimates of a regression slope are normally distributed, we can use a two-sample t test to ask if $\hat{\mu}_b$ and $\hat{\mu}_a$ are significantly different (Dennis et al. 1991). Specifically, we compute the statistic

$$T_{q-2} = (\hat{\mu}_a - \hat{\mu}_b) / \sqrt{\hat{\sigma}_m^2[(1/j) + 1/(q-j)]} \quad (3.15)$$

where q is the total number of transitions in the data set and j is the number of transitions leading up to the pivotal census. For the Yellowstone

BOX 3.2 Output produced by SAS commands given in Equation 3.14.

Model: MODEL1
NOTE: No intercept in model. R-square is redefined.
Dependent Variable: Y
Analysis of Variance
Source DF Sum of Squares Mean Square F Value Prob>F
Model 2 0.03118 0.01559 1.197 0.3140
Error 36 0.46907 0.01303
U Total 38 0.50025
Root MSE 0.11415 R-square 0.0623
Dep Mean 0.02134 Adj R-sq 0.0102
C.V. 514.97870
Parameter Estimates
Parameter Standard T for H0:
Variable DF Estimate Error Parameter=0 Prob > T
X1 1 0.010690 0.0219667 0.504 0.6171
X2 1 0.055644 0.03804919 1.462 0.1523

grizzly bear, $q = 38$ and $j = 29$ (as above), and we substitute these and the values of $\hat{\mu}_b$, $\hat{\mu}_a$, and $\hat{\sigma}_m^2$ obtained above into Equation 3.15, yielding $T_{36} = 1.032$. The probability of obtaining this value from a t distribution with 36 degrees of freedom is 0.154 (which can be obtained using Excel formula =TDIST(1.032,36,1)). Once again, this probability does not support the hypothesis that μ differs before and after the fires. If it did, we would use the results of this regression to obtain separate estimates of $\hat{\mu}_b$ and $\hat{\mu}_a$ (with one estimate of $\hat{\sigma}_m^2$) and then decide which estimate of μ to use in viability calculations ($\hat{\mu}_a$ if we have reason to believe that post-fire conditions will continue, and $\hat{\mu}_b$ if we suspect that pre-fire conditions have returned).

Thus we have tested the assumption of the diffusion approximation that the parameters μ and σ^2 do not change, and found it to be reasonable (at least in reference to the periods before and after the 1988 fires). We now have justification for estimating μ and σ^2 with a single regression, as in Box 3.1. Once again, if any of the results were significant, we would be left in the position of having to decide, based on ancillary information, which set of parameter estimates to use. As an exercise, readers are encouraged to follow the procedure laid out above to test whether μ and σ^2 differ before versus after the dumps were closed in 1970–1971.

One caution about testing for parameter changes deserves mention. When we divide an already short data set into even smaller time intervals, we may be left with little statistical power to detect changes in parameter values. In fact, we may not even have sufficient power to detect simple trends in population size within each time interval. Hence it may be valuable to supplement the analysis described here with a power analysis for detecting population trends, using methods described in Gibbs (2000).

Using Estimates of μ and σ^2 to Calculate Probability of Extinction

Having estimated μ and σ^2 and made use of regression diagnostics to test some of the assumptions of the diffusion approximation, we are now ready to use Equation 3.5 to construct the extinction time cumulative distribution function (CDF), which we argued in Chapter 2 was the single most informative metric of a population's viability. Once we have obtained the estimates $\hat{\mu}$ and $\hat{\sigma}^2$ and determined a suitable quasi-extinction threshold, it is straightforward to use Equation 3.5 to calculate the probability that the population starting at the current size will have hit the threshold prior to each of a set of future times. The MATLAB code in Box 3.3 defines a function called `extcdf` that calculates the extinction time CDF. This function uses a second function listed in Box 3.3, `stdnormcdf`, which calculates the standard normal cumulative distribution function given by Equation 3.6. Later, we will use three other functions defined in Box 3.3 to compute confidence limits for the extinction time CDF. To make these functions accessible to MATLAB pro-

BOX 3.3 MATLAB code defining five functions used in calculating the extinction time cumulative distribution function and its confidence limits.

File extcdf.m

```
function G=extcdf(mu,sig2,d,tmax);
% extcdf(mu,sig2,d,tmax) calculates the unconditional extinction
% time cumulative distribution function from t=0 to t=tmax for
% mean and variance parameters mu and sig2 and log distance from
% the quasi-extinction threshold d;
% Modified from Lande and Orzack, Proc. Nat. Acad. Sci. USA 85:
% 7418-7421 (1988), equation 11.
% REQUIRES THE FILE stdnormcdf.m

for t=1:tmax
    G(t)=stdnormcdf((-d-mu*t)/sqrt(sig2*t)) + ...
        exp(-2*mu*d/sig2)*stdnormcdf((-d+mu*t)/sqrt(sig2*t));
end;
```

File stdnormcdf.m

```
function phi=stdnormcdf(z)
% stdnormcdf(z) calculates the standard normal cumulative
% distribution function, using the built-in MATLAB error function
% erf;
% See Abramowitz and Stegun, 1964, Handbook of Mathematical
% Functions, Dover, New York, equation 7.1.22
```

```
phi=0.5*(1 + (erf(z/sqrt(2))));
```

File gammarrv.m

```
function gamma=gammarrv(alpha,beta,n);
% gammarrv(alpha, beta,n) generates a vector of n random numbers
% from a Gamma(alpha,beta) distribution:
% f(x)=exp(-x/beta)*x^(alpha-1)/[Gamma(alpha)*beta^alpha]
% if x>=0;
% f(x)=0 if x<0
% where Gamma(z) is the gamma function.
% Mean(x)=alpha*beta; Var(x)=alpha*beta^2
% See G.S. Fishman, 1973, Concepts and Methods in Discrete Event
% Digital Simulation, Wiley, New York, pp. 208-209.
```

BOX 3.3 (continued)

```
gamma=[];
for i=1:n
    X=0;
    k=floor(alpha);
    g=alpha-k;
    if k>0
        X=-log(prod(rand(1,k)));
    end;
    if g==0
        gamma=[gamma beta*X];
    else
        a=g;
        b=1-g;
        y=1;
        z=1;
        while y+z>1
            y=rand^(1/a);
            z=rand^(1/b);
        end;
        Y=y/(y+z); % Y is a Beta(g,1-g) random variable
        Z=-log(rand); % Z is a Gamma(1,1) random variable
        gamma=[gamma beta*(X+Y*Z)];
    end;
end;
```

File betarv.m

```
function beta=betarv(m,v,n);
% betarv(m,v,n) generates a row vector of length n, the elements
% of which are Beta random variables with mean m and variance v;
% See G.S. Fishman, 1973, Concepts and Methods in Discrete Event
% Digital Simulation, Wiley, New York, pp. 204-208.
% REQUIRES THE FILE gammarrv.m

if v==0
    beta=m*ones(1,n);
elseif v>=m*(1-m),
    fprintf(1,'Variance of Beta too large given the mean');
    pause;
```

BOX 3.3 (continued)

```

else
    a=m*(m*(1-m)/v - 1);
    b=(1-m)*(m*(1-m)/v - 1);
    beta=[];
    for i=1:n
        k1=floor(a);
        k2=floor(b);
        if k1==0 & k2==0
            Y=1;
            Z=1;
            while Y+Z>1
                Y=rand^(1/a);
                Z=rand^(1/b);
            end;
        else
            Y=gammarv(a,1,1);
            Z=gammarv(b,1,1);
        end;
        beta=[beta Y/(Y+Z)];
    end;
end;
```

File chi2rv.m

```

function x=chi2rv(df);
% chi2rv(df) generates a chi-squared random number with df
% degrees of freedom;
% See G.S. Fishman, 1973, Concepts and Methods in Discrete Event
% Digital Simulation, Wiley, New York, p. 213.
% REQUIRES THE FILE betarv.m

if mod(df,2)==0
    x=-2*log(prod(rand(1,df/2)));
else
    k=(df/2)-0.5;
    v=sum(-log(rand(1,k)));
    y=betarv(0.5,0.125,1);
    z=-log(rand);
    x=2*(v+y*z);
end;
```

grams, place them in separate files with a ".m" extension (e.g., extcdf.m) and place the name of the folder containing the files in MATLAB's path settings. The function extcdf returns a column array with tmax rows containing the probabilities that the extinction threshold will have been reached by each future time. In addition to tmax, calls to the function extcdf must provide three other arguments. These are the estimated values of μ and σ^2 and d , the difference between the log of the current population size and the log of the quasi-extinction threshold. Recall that it is only this difference, rather than the actual values of the population size and the threshold, that determines the extinction probability (Equations 3.4 and 3.5).

We argued previously that we should not place much faith in single estimates of μ and σ^2 because these parameters will not be estimated with high precision, given the limited amount of data typically available for threatened and endangered species and populations. For the same reason, we should not place much faith in the best estimate of the probability of extinction, which we will call \hat{G} (see Equation 3.5), that is based only on our best parameter estimates μ and σ^2 , without accounting for the uncertainty in these estimates. Thus we need a way to translate uncertainty in the estimates of μ and σ^2 into uncertainty in the probability of extinction, G . Although we cannot write down an expression for the confidence interval of G akin to Expressions 3.12 and 3.13, we can use a computer-based method known as a parametric bootstrap¹³ to approximate the confidence interval for G . Specifically, because we know the probability distributions that govern the parameter estimates,¹⁴ we can have a computer draw values of μ and σ^2 from the appropriate distributions. If both of those estimates lie within their respective confidence limits, given by Expressions 3.12 and 3.13, then we use them to calculate \hat{G} . By repeating this process many times, we obtain a range of values of \hat{G} , all of which will lie within the confidence interval for G . The extreme values of \hat{G} define the upper and lower boundaries of the confidence interval. The number of bootstrap samples (i.e., the number of values of \hat{G} we calculate) should be relatively large, so that we are reasonably sure to see the extreme values that define the confidence limits.

¹³An alternative bootstrapping approach, the nonparametric bootstrap, involves repeatedly sampling from the original data (i.e., constructing a set of q values randomly chosen (with replacement) from the observed log population growth rates, estimating μ and σ^2 and calculating G , repeating this process many times, and identifying (for example) the 2.5th and 97.5th percentiles of the resulting distribution of G values as the 95% confidence limits. For a discussion of the relative merits of the parametric and nonparametric bootstrap in a PVA context, see Ellner and Fieberg (2002).

¹⁴Specifically, $\hat{\mu}$ is normally distributed and $\hat{\sigma}^2$ has the distribution of a chi-square random variable multiplied by $\hat{\sigma}^2/q$. Furthermore, these estimates are independent, so their values can be obtained from separate random draws from the two distributions. For details, see Dennis et al. 1991.

The MATLAB code in Box 3.4 performs the bootstrap procedure described in the preceding paragraph. This program uses the function `extcdf` and the four other functions defined in Box 3.3. The user-defined parameters in Box 3.4 correspond to the values we have estimated for the Yellowstone grizzly bear population. Specifically, we use $N_c = 99$ (the number in the 1997 census; Table 3.1) as the current population size, a quasi-extinction threshold N_x of 20 reproductive females, following the guidelines we gave in Chapter 2 (remember that N_c and N_x together determine d), and estimates of μ and σ^2 from Box 3.1. The results from one run of the program with 500 bootstrap samples are shown in Figure 3.8. The best estimate of the probability that the population will decline from 99 to 20 female bears is quite low for short times into the future, reaching a value of 0.0018 at 50 years. (Note the logarithmic scale of the y -axis in Figure 3.8.) However, notice that the confidence interval for the probability of quasi-extinction widens rapidly as time increases; by 50 years, the 95% confidence interval

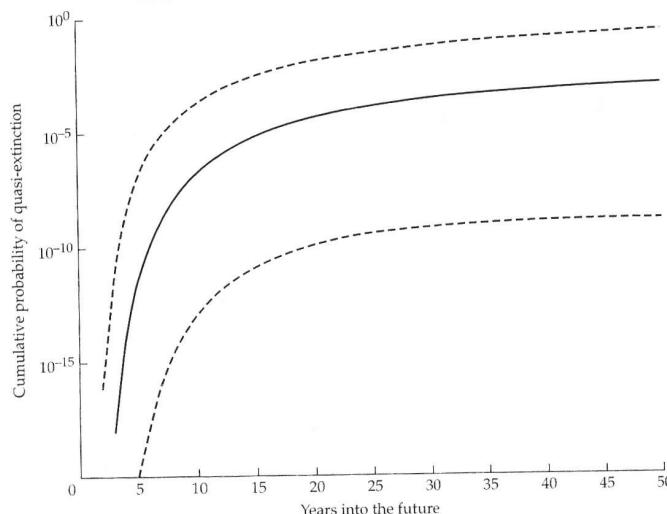


Figure 3.8 Extinction time cumulative distribution function for the Yellowstone grizzly bear population. Solid line is based on the best estimates of μ and σ^2 ; dashed lines delineate an approximate 95% confidence interval determined by a bootstrap (see Box 3.4). The x-axis shows the time required for a population of 99 female bears to decline to 20.

BOX 3.4 MATLAB code to calculate extinction probabilities and bootstrap confidence intervals¹.

```
% Program ExtProb
% ExtProb calculates the probability of extinction with
% bootstrap confidence intervals for a density-independent
% model using a diffusion approximation;
% See Dennis et al. 1991, Ecological Monographs 61: 115-143;
% REQUIRES THE FILES extcdf.m, chi2rv.m, stdnormcdf.m,
% gammargv.m, and betargv.m

%*****%
% Change the following user-defined parameters:
%*****%
mu=0.02134; % Enter the estimated value of mu
sig2=0.01305; % Enter the estimated value of sigma^2
CI_mu=[-.01621,.05889]; % Enter confidence interval for mu
CI_sig2=[.00867,.02184]; % Enter confidence interval for sigma^2
q=38; % Enter the number of transitions in the data set
tq=38; % Enter the length of the census (in years)
Nc=99; % Enter the current population size
Ne=20; % Enter the quasi-extinction threshold
tmax=50; % Enter latest time to calculate extinction
% probability
Nboot=500; % Number of bootstrap samples for calculating
% confidence intervals for extinction probabilities
%*****%
d=log(Nc/Ne); % calculate log distance to quasi-extinction
% threshold
SEmu=sqrt(sig2/tq); % calculate standard error of mu
Glo=ones(1,tmax); % initialize array to store lower bootstrap
% confidence limits for extinction
% probabilities
Gup=zeros(1,tmax); % initialize array to store upper bootstrap
% confidence limits for extinction
% probabilities
rand('state',sum(100*clock)); % seed the random number generator

% Calculate the extinction time cdf for the best parameter
% estimates
```

¹The code in Box 3.4 uses functions defined in Box 3.3.

BOX 3.4 (continued)

```

Gbest=extcdf(mu,sig2,d,tmax);

% Calculate bootstrap confidence limits for extinction
%   probabilities

for i=1:Nboot,

    % Generate a random mu within its confidence interval
    murnd=inf;
    while murnd<CI_mu(1)|murnd>CI_mu(2)
        murnd=mu+SEmu*randn;
    end;

    % Generate a random sigma^2 within its confidence interval
    sig2rnd=inf;
    while sig2rnd<CI_sig2(1)|sig2rnd>CI_sig2(2)
        sig2rnd=sig2*chi2rv(q-1)/(q-1);
    end;

    % Calculate extinction probabilities given murnd and sig2rnd
    G=extcdf(murnd,sig2rnd,d,tmax);

    % Store extreme values
    for t=1:tmax
        if G(t)<Glo(t)
            Glo(t)=G(t);
        end;
        if G(t)>Gup(t)
            Gup(t)=G(t);
        end;
    end;
end; %for i=1:Nboot

% Plot G (log-scale) vs t, using a solid line for the best
% estimate of G and dotted lines for the confidence limits
t=[1:tmax];
semilogy(t,Gup,'k:',t,Gbest,'k-',t,Glo,'k:')
axis([0,tmax,0,1])
xlabel('Years into the future')
ylabel('Cumulative probability of quasi-extinction')

```

for the probability of hitting the threshold ranges from 6.09×10^{-9} to 0.255. Thus even though the best estimate is that this population won't decline dramatically over the short term, we cannot say with certainty that the population has at least a 95% chance of remaining above 20 females over a period as short as the next five decades (and many authors have called for such high levels of safety over much longer periods). Thus due to the uncertainty in the parameter estimates $\hat{\mu}$ and $\hat{\sigma}^2$, we cannot predict the extinction probability very accurately for very far into the future, a caution that several authors have made (Ludwig 1999, Fieberg and Ellner 2000). The fact that, given the available information, we cannot make very precise statements about the extinction risk the Yellowstone grizzly bear population faces several decades from now further argues the need for caution when we consider delisting this population. Indeed, for a decision as momentous as delisting, it makes sense to use the more pessimistic bounds on our estimates of μ and σ^2 in order to act in a precautionary fashion. In this way, we can use uncertainty in extinction risk estimates to reach more valid conclusions about our understanding of population viability (Gerber et al. 1999). Wide confidence limits on extinction risk estimates realistically reflect our uncertainty given imperfect information, and quantifying this uncertainty is a useful product of a viability analysis.

Using Extinction Time Estimates for Conservation Analysis and Planning

We now give two examples that show how the extinction time CDF can be used to inform decisions about the viability of single populations or about which of several populations should receive the highest priority for acquisition or management.

Perhaps the most valuable use of the CDF is to make comparisons between the relative viabilities of two or more populations. Ideally, we would have a series of counts from each population. For example, Figure 3.9A,B shows the number of adult birds during the breeding season in two populations of the federally listed red-cockaded woodpecker, one in central Florida and the other in North Carolina. Applying the methods outlined above yields the CDFs in Figure 3.9C. Both because it has a more negative estimate for μ (-0.083 vs. -0.011) and a smaller current size, the central Florida population has a much greater probability of extinction at any future time than does the North Carolina population. This information could be very useful in deciding the population on which to focus management attention.

Often we will not have independent census data from each population about which we must make conservation decisions. However, if we have a single count of the number of individuals of a particular species in one population, we can use count data from multiple censuses of the same species at