

5

Accounting for Observation Error in Count-Based PVAs

We have now seen repeatedly that the degree to which population growth rates fluctuate due to environmental stochasticity strongly affects population viability. Therefore, to produce a useful count-based PVA, we must accurately estimate how much of the year-to-year variation in the census¹ counts, and hence in the estimated log growth rates (i.e., the values of $\log(N_{t+1}/N_t)$), is actually caused by variation in the environment. Count-based PVAs would be simpler to perform if we could assume that all of the variation in the estimated growth rates is due to environmental stochasticity. However, we know that several factors other than the environment may cause the growth rate estimates to vary from year to year. Two such factors are demographic stochasticity and density dependence.² In the preceding chapter, we saw how to account for these two sources of variation, either by choosing an appropriate quasi-extinction threshold or by using more complex models if the data needed to parameterize them are available. Unfortunately, there is yet another factor that can introduce spurious variation into the estimated growth rates, an unwanted source of variation that we will refer to as “observation error.” Observation error is variation in the census counts (and hence the estimated population growth rates) caused by our inability to count precisely the number of individuals in the population or in the segment of the population that our census targets (e.g., adult females). Whereas demographic stochasticity and density dependence are true population processes that can have a real impact on population viability, observation error is not. Therefore, regardless of whether or not we have managed to account for demographic stochasticity and density

¹To some users, *census* refers to an exhaustive count of all members of a population. In this book, we use the term in a looser sense to mean any attempt to quantify, by any method, the size of a population. Whenever a census does not involve a completely accurate and exhaustive count, observation error is possible.

²The two other factors discussed in Chapter 4, temporal environmental autocorrelation and bonanzas/catastrophes, are truly features of the environment.

dependence, we must still consider the possible contaminating influence of observation error.

As we mentioned in Chapter 2, one approach is to use the measured variation in the population growth rate as a surrogate for the environmentally driven variation, recognizing that viability estimates produced in this way will be overly pessimistic, since observation error will inflate the measured variation. Although this route is the easiest one to follow, it would obviously be preferable to quantify the amount of observation error and remove its effects from the analysis to the greatest extent possible. The goal of this chapter is to review some of the simpler and more practical methods available to reduce the influence of observation error on viability assessments. We begin by describing some potential sources of observation error. We then describe how observation error can be reduced and how the effects on viability assessments of those errors that do occur can be lessened. We break these solutions to the problem of observation error down into those that can be employed before a census is actually conducted, during the course of the census, and after the census data have been collected. At the end of the chapter, we direct readers to references for some of the more sophisticated methods for estimating model parameters in the face of observation error, methods that lie beyond the scope of this book.

Potential Sources of Observation Error

Before describing ways to reduce its influence, we first devote a bit more attention to defining what observation error really is, and what it is not. First of all, *observation error* is not strictly synonymous with *sampling variation* (although, as we will see, sampling variation is one potential source of observation error). Sampling variation occurs whenever any statistic (such as the mean log growth rate) is estimated using data about a limited sample from a larger “population.” Here, we are using the term *population* not in the ecological sense (i.e., the set of all individuals of a given species in a prescribed area) but in the statistical sense (i.e., the set of all objects of a given type). For the purposes of conducting a count-based PVA, the statistical population we are interested in is the set of all possible growth rates that might occur in the ecological population under study. Considering that the growth rate can vary continuously over a broad range of values, the population of all possible growth rates is technically of infinite size. However, a limited set of yearly censuses can yield only a small subsample of growth rates from this infinite population. Even if we could measure the size of the ecological population, and therefore its growth rate, with perfect precision each year, *observing and counting every individual with complete accuracy*, we would still not expect to get exactly the same set of growth rates if we were to census the population annually for, say, repeated 20-year periods. Thus statistics such as the mean and variance of the log growth rates (which we use to

estimate μ and σ^2 , respectively) will differ among sets of census data due to sampling variation. The confidence limits on μ and σ^2 that we calculated in Chapter 3 used the theoretical sampling distributions of the mean and variance of a normal random variable to account for this sampling variation.

However, we are essentially *never* able to observe and count every individual in an ecological population. The population may be too large to count exhaustively, and even if it contains relatively few individuals, as do most endangered populations, they may be spread out over too large an area to search in its entirety. In these situations, the typical protocol is to *estimate* total population size at each point in time by extrapolation from a sample of the larger population. For example, we may exhaustively count the number of individuals in a set of study plots to get an estimate of mean density, which we multiply by the total occupied area to estimate total population size. Even if the count within each plot is completely accurate, our calculated mean density in a single year will be based on a limited number of plots and thus will be subject to sampling variation, which will in turn lead to observation error in the estimated population size. The important distinction to understand is that sampling variation is a consequence of limited amounts of data (even if those data are measured accurately), whereas observation error is *any* inaccuracy in estimates of population size, and hence population growth rate, whatever the cause.

Consider what effect this observation error might have on a viability assessment. As an unrealistic but informative example, let’s assume that the true population size is absolutely constant over time, but that we must estimate it by sampling a set of plots that represent a known fraction of the population’s entire geographical range. Unless the members of the population are distributed across the range in a perfectly uniform fashion each time we conduct the census (an extremely unlikely event for *any* organism), the mean numbers across all the plots, and thus the *estimates* of total population size, will vary from year to year, even though the true population size does not. That is, in some years, a relatively large fraction of the individuals in the population will happen to be located in our sampling plots at the time of the census, while in other years, few individuals will. If we now use the variance in the *estimated* log growth rates as a measure of the environmental variance σ^2 , we would obtain a positive value, and would thus conclude that the population faces some risk of quasi-extinction *even though its true size never changes*. Thus the artifactual component of the variance in the log growth rate introduced by observation error inflates our estimate of extinction risk.

The preceding paragraph assumed that the counts within sampling plots were completely accurate. However, these counts may suffer from another source of observation error: failure to determine correctly the number of individuals actually present in each plot. Survey methods will often leave some individuals undetected (e.g., bears in dense vegetation may be impossible to see in an aerial survey), leading to undercounts. Conversely, counting some

individuals more than once can lead to overcounts. If a large proportion of the sample counts taken from a population are undercounts, or a large proportion are overcounts, the mean of the samples will be a *biased* estimate of the true population mean. However, bias *per se* is not the main problem for estimating the population growth rate. For example, let's say that our census *consistently* misses 10% of the population, year after year. If the true population sizes in two successive years are 140 and 160 individuals, our estimated population sizes based on the census counts would be 126 and 144 (i.e., 90% of the true counts). However, our estimate of the population growth rate based on these population size estimates (i.e., $144/126 = 1.14 = 160/140$) would actually give us the correct answer, because a fixed proportional bias in the two population size estimates does not affect their ratio.³ The problem arises because the magnitude of the under- or over-count, and thus the degree of bias in the estimates of population size, will always vary somewhat from year to year. Imagine another pair of successive years in which the true population sizes are 140 and 160, as before. But now assume that 8% of the individuals are missed in the first year and 12% in the second year, so that now the estimated population sizes are 128.8 and 140.8. The resulting estimate of the population growth rate (1.09) is an underestimate of the true growth rate. If we switched the percentages of individuals missed in the two years, the resulting growth rate estimate (1.19) would be an overestimate of the true rate. Thus *variation* among censuses in the amount of bias in the counts will artificially inflate the estimate of the variation in the population growth rate (i.e., σ^2). Also, bias can sometimes be nonproportional (i.e., an over- or undercount by approximately the same *number* of individuals in every year); such density-dependent bias in the sample counts will translate into biased estimates of both μ and σ^2 . Variation in the number or fraction of individuals detected results from many factors, including changes from census to census in the proportion of individuals located in dense vegetation, differences in light levels or weather conditions at the times aerial surveys are conducted, or simply differences among observers in their levels of training or inherent skill at locating and distinguishing individuals of the focal species.

Thus observation errors arising either from counting mistakes or from variation among samples used to estimate population size will often introduce an unwanted source of variation in census counts. Unfortunately, describing the potential sources of observation error is far easier than designing census protocols to avoid it or developing ways to reduce its influence on parameter estimates. Still, in the remainder of this chapter, we discuss a few basic approaches that can help to reduce the unwanted influence of observation error on estimates of population viability. We use the

³We would, however, want to keep the uncensused individuals in mind when setting a quasi-extinction threshold.

word *reduce* advisedly, as we can never entirely eliminate the influence of observation error, only attempt to minimize its effects.

Before proceeding, we note that observation error can have two distinct effects on estimates of population parameters such as μ and σ^2 . First, it can introduce *bias* into the estimated parameters. A procedure for estimating a particular parameter is biased if, when applied many times to replicate sets of data generated from a "true" model, it yields a distribution of estimates whose mean differs from the true value of the parameter. That is, biased estimators tend to under- or overestimate the true parameter value. For example, as we have just seen (and will explore in more detail below), observation error usually causes the variance of the observed log population growth rates to be upwardly biased relative to the true value of σ^2 . Bias is worrisome for PVA because it will lead to overly optimistic or overly pessimistic assessments of viability, depending on which parameter estimates are biased and in which direction. The second effect of observation error is that it can reduce the *precision* of parameter estimates. An estimation procedure that, when applied to replicate data sets produced by the same true model, yields a distribution of parameter estimates with a high variance is less precise than one that yields a low variance among the estimates, even if both produce unbiased estimates. The most general way to increase the precision of parameter estimates is to increase the amount of data used in the estimation process, but this is easier said than done for rare species. In this chapter, we will see examples of both bias and imprecision caused by observation error.

Considerations for Reducing Observation Error before a Census Is Initiated

In designing a census, care should be taken to assure that the census counts are as accurate as possible, and that the degree of accuracy changes as little as possible over the many years of a long-term census. Methods to accurately census a population are idiosyncratic to the type of organism under study, and we cannot possibly review this large field here. Instead, readers should consult one of the many taxon-specific books on population estimation (e.g., Sutherland 1996 for a range of taxa, Southwood 1978 for insects, Elzinga et al. 1998 for plants, and Thompson et al. 1998 for vertebrates, with even more specific treatments in Krebs 1989 and Wilson 1996 for mammals and Heyer 1994 for amphibians). A wealth of more general papers on the thoughtful design of monitoring plans have also been published (e.g., Vos et al. 2000 and Gibbs et al. 1999). However, a few general guidelines are worth noting here:

1. Make the first year of data collection a trial run to generate information needed to intelligently set sampling methods and effort. When possible, try several sampling methods before initiating the census or during the

- first census period to help identify the method that yields the most accurate counts. In addition, deliberately oversample (i.e., invest more search time, search more plots, and so forth) on this first census, as this is the only way to evaluate how quickly sampling variation will decrease as you increase sampling effort. Alternatively, adopt a sampling method that is known to be fairly accurate when applied to organisms and habitats that are similar to the census situation you face.
2. After fixing the census methods to be used, formalize and write-up a clear sampling protocol, and use that protocol consistently in the successive censuses. The protocol should specify that the census is to be taken in the same season each year and under similar weather conditions. In addition, make sure that all observers have received consistent training prior to the census in how to search for and recognize the study organism. To assure consistency, it helps if the same person is charged with training the observers and leading the census for as long as possible, and it helps if she or he carefully trains the succeeding leader. Furthermore, for census data to be comparable across multiple observers, it is best if highly specialized skills are not needed to accurately collect the information. These procedures should reduce variation in bias from census to census.
 3. If possible, establish a stratified,⁴ randomly selected set of census plots, transects, flight routes, or roadside sampling stations at the outset of the census, and use these same locations in subsequent censuses. This should reduce the potential for confounding spatial variation in density or in degree of detectability related to habitat complexity with temporal variability in true population size.⁵
 4. If it is not possible to count all members of the population, choose the most easily detected subset of individuals to census, or record information separately for each subset of the population, allowing separate analyses to be tried in the future. Often the most easily detected members will be the largest individuals in the population, which usually make the biggest contribution to population growth anyway.
 5. Design data collection methods and data recording protocols that allow sampling effort to be clearly understood and quantified.

⁴By *stratified*, we mean that when the focal organism inhabits two or more distinct types of habitat, we place sampling units (e.g., transects) at random within each type of habitat. We then estimate total population size as the mean density across all units in one habitat type multiplied by the area of that habitat type, and sum these products over all habitat types. A stratified sampling design can yield more accurate estimates of total population size than a strictly random design (see Greenwood 1996).

⁵One downside of using the same sampling units every year is that the counts from the same unit in successive years will not be statistically independent. Fully accounting for the covariance this introduces requires methods more sophisticated than the ones we present here.

Most of these guidelines are patently obvious, but attending to them before initiating a census might greatly reduce observation error problems later on.

Quantifying Observation Errors while a Census Is Being Conducted

The most important thing that can be done at the time of the census to help account for observation error is to quantify its magnitude. If the census is conducted by counting individuals in replicate sample units (e.g., multiple quadrats or line transects), then simply recording the separate counts in each unit, rather than only the total or the mean count, permits us to account for observation error caused by sampling variation, as we will see in the following section. If the census protocol does not call for replicate samples in every year of the census, then it would be wise to perform repeat sampling of at least a fraction of the study area in at least some years. As noted above, this repeated sampling is especially useful in the first year of a census, allowing one to determine the sampling effort needed to yield acceptably accurate data. As an example of repeated sampling, one could count the number of individuals in an area on one day, and then repeat the count in the same area at one or more times shortly thereafter. If these repeated counts are sufficiently close in time, it is probably safe to conclude that the differences among them are largely due to observation error rather than an actual change in the number of individuals in the area. (Of course for mobile species, care must be taken to account for individuals that may have been induced to leave the area due to the disturbance caused by earlier censuses.) Ideally, these repeated censuses of the same area should be performed both in years when the population size is relatively large and in years of relatively low population size, because the degree of observation error itself may depend on population size. (For example, it may be easier to miss individuals when they are scarce and hiding places are abundant.) Although this repeated sampling will add to the time and expense of performing the census, it may provide invaluable information for correcting parameter estimates for observation error. We will see how to do so in the following section.

One other thing that should be done at the time of the census deserves further mention. If the census involves simply making a total count of the number of individuals seen (i.e., if there are no obvious sampling units), the amount of effort (e.g., the number of person-days spent searching, the number of flight hours or road miles driven, or the fraction of the total area occupied by the population that is searched) should be recorded each year, and the count adjusted for effort. (For example, dividing the count by the fraction of the occupied area searched gives a measure of total population size.) If multiple observers are involved in a census it is also important to keep

separate tallies of the search time and success of each participant, in order to distinguish the effort of efficient experts from that of novices. In general, without recording sampling effort, among-year differences in observed population sizes cannot be divided into real differences versus observation error.

Correcting for Observation Errors after the Census Data Have Been Collected

Worrying about observation errors after data have been collected probably seems like shutting the stable door after the horse has escaped. However, at least some observation error is present in any data set, and hence there will always be a need for methods to reduce its influence on parameter estimates. Disentangling the variance in the estimated population growth rate that is due to environmental variation from the variance due to observation error is a complex statistical problem, and many sophisticated methods to attack this problem lie far beyond the scope of this book (see the end of the chapter). However, in this section we describe three simple approaches that can easily be applied to count-based data. As an up-front warning, we caution that all three of these methods have their limitations, which we note as we describe them. In many cases, none will work and simply proceeding without explicitly discounting for observation error (using the methods of Chapters 3 and 4) will be our only alternative.

Correcting Estimates of σ^2 for Observation Error when Counts Represent Means of Replicate Samples

As we have discussed above, the census counts we intend to use to build a count-based PVA will frequently involve means of replicate samples taken from the population. For example, to estimate the population size of an endangered plant, we might count the number of individual plants encountered in a set of quadrats or line transects, or for an animal species, we might repeatedly fly over the population in an airplane and count the number of individuals seen in each flight. Typically, we would then compute the mean of these replicate counts as a measure of the average density per sampling unit (e.g., a quadrat or a flight of a fixed duration) in each year, and then use either the means from every year (for a density-based PVA) or the means multiplied by the area occupied by the population (for an analysis based on total population size). The key point to recognize is that there is sampling variation associated with each of those means, and as we noted above, this sampling variation creates observation error. If we simply apply the methods of Chapter 3 to estimate σ^2 (i.e., if we compute the variance of the log population growth rates calculated from the sample means), the resulting value will represent an amalgamation of the true environmental variance

and the sampling variation that affects the sample means. The larger the number of samples used to calculate the means, the smaller the observation error will be, but realistic constraints of time and money will often dictate a modest number of samples (especially for technically complex sampling methods such as aerial surveys). Hence simple estimates of σ^2 will frequently be biased by observation error. In addition, observation error will reduce the precision of simple estimates of both μ and σ^2 .

To demonstrate the potential magnitudes of the bias and imprecision imparted by observation error, we performed a simple computer simulation. We began by generating a “true” population trajectory using the density-independent model:

$$N_{t+1} = N_t \exp(\mu + \sigma x_t) \quad (5.1)$$

where x_t is drawn at random from a normal distribution with a mean of zero and a standard deviation of 1. In this model, the true mean and variance of the log population growth rate are μ and σ^2 , respectively. For each census t , we then mimicked the observation process by drawing n_t samples $N_{i,t}$ ($i = 1 \dots n_t$) from a negative binomial distribution⁶ such that the mean of a large number of samples would equal the true population density, N_t . The negative binomial distribution commonly provides a close fit to the frequency distribution of the number of individuals in replicate samples from a population (Southwood 1978). The negative binomial is a discrete, strictly nonnegative probability distribution, meaning that it can only take on integer values of zero or greater, which is appropriate given that the number of individuals observed in each sampling unit can only be a nonnegative integer. More importantly, the negative binomial distribution can fit a wide variety of sampling distributions, ranging from randomly dispersed to

⁶The negative binomial distribution arises most simply in the following way. If individuals are distributed completely at random, the counts from replicate samples will follow a Poisson probability distribution, in which the mean and variance of the samples will be equal. But if individuals are aggregated, we can think of some samples as having been drawn from Poisson distributions with high means and other samples as having been drawn from Poisson distributions with low means. More specifically, if the means of a series of Poisson random variables are themselves drawn from a gamma probability distribution, $G(x|k, k/m) = (k/m)^k x^{k-1} e^{-(k/m)x}/\Gamma(k)$

(where $\Gamma(k) = \int_0^\infty z^{k-1} e^{-z} dz$ is the gamma function), then the resulting random numbers will come from a negative binomial distribution with a mean of m and a variance of $m + m^2/k$. Hence, to generate a sample y from a negative binomial distribution, we first generate a value of x from the gamma distribution $G(x|k, k/m)$, and then generate a value of y from the Poisson distribution $P(y|x) = x^y e^{-x}/y!$ (see Fishman 1973). The MATLAB statistics toolbox also includes a function (nbirnd) for directly generating random numbers from a negative binomial distribution.

highly aggregated (Figure 5.1). A sampling distribution in which the variance is approximately equal to the mean (Figure 5.1A) implies that the location of each individual was determined at random, without regard to the locations of other individuals in the population. In contrast, an aggregated distribution, such as the one illustrated in Figure 5.1E, arises when individuals are highly clumped in space, so that most samples include few or no individuals but a few samples contain many individuals. It makes intuitive sense that when the counts vary as much among individual sampling units as they do in a highly aggregated distribution, the mean of the samples $\bar{N}_t = \frac{1}{n_t} \sum_{i=1}^{n_t} N_{t,i}$, which we use as our estimate of population density, will also vary considerably from one set of samples to another, so there will be a high potential for observation error in the estimated population size.

For our example, we used the true values $\mu = 0.0296$ (i.e., $\lambda_G = e^{\ell} = 1.03$) and $\sigma^2 = 0.01$, and we generated $n_t = 10$ samples for each of 21 annual censuses. We then used the mean densities, the \bar{N}_t 's, to estimate the log population growth rates between each pair of censuses, and computed the mean and variance of those estimated log population growth rates as our estimates of μ and σ^2 . We will label these estimates $\hat{\mu}$ and $\hat{\sigma}^2$, and refer to them as the *raw* estimates of μ and σ^2 , as they are uncorrected for observation error. We repeated this entire process 1,000 times to get a distribution of likely values of $\hat{\mu}$ and $\hat{\sigma}^2$ for each of several different levels of sampling variation. Specifically, we set the variance of the negative binomial distribution from which we drew the 10 samples in year t to equal 1, 2, 4, 8, or 16 times the true population density in that year, N_t , which was the expected mean of the samples. That is, we used sampling distributions similar to those illustrated in Figure 5.1 to mimic spatial arrangements of individuals that ranged from random to highly aggregated. Because the variance was a constant multiple of the mean, the actual variance among the samples was higher in years when true population size, and hence the mean of the samples, was high. We might well expect to see a higher variance among the samples when the mean is high than when the mean is low (although the opposite is also possible).

The distributions of the raw estimates of μ , the $\hat{\mu}$'s, are shown in Figure 5.2A. Even with relatively low variance among sample units (variance = mean), the distribution of the $\hat{\mu}$'s is quite broad (ranging from negative to positive values), meaning that it is difficult to estimate μ with precision using only 21 years of data. Nevertheless, even a high amount of observation error (i.e., high variance among samples) does not cause the mean of the $\hat{\mu}$'s to deviate appreciably from the true value of μ . In other words, the mean of the log population growth rates is a nearly unbiased estimator of the true μ . However, notice that the spread of the estimates increases with increasing among-sample variance; that is, observation error decreases the precision of the raw estimates of μ . Unfortunately, observation error has an even more

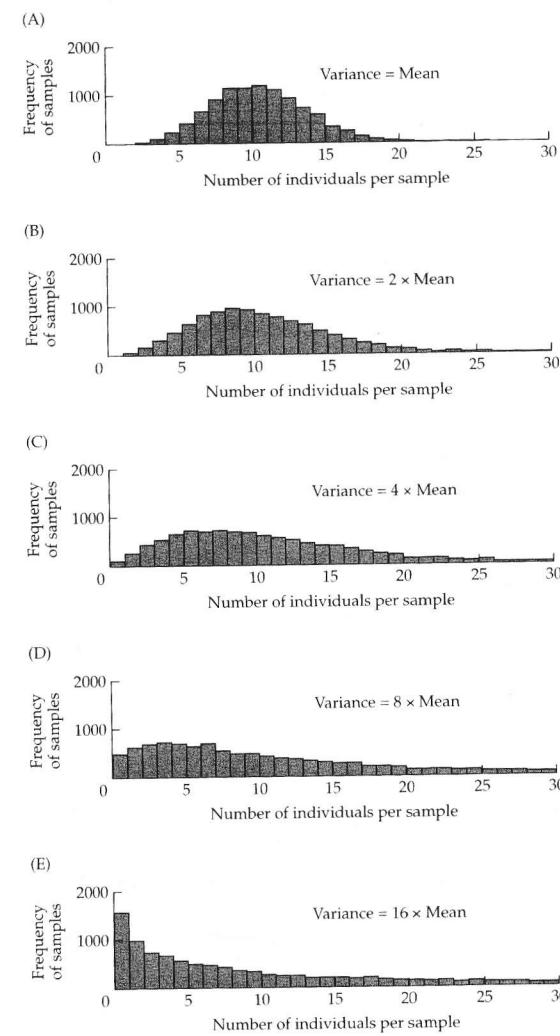
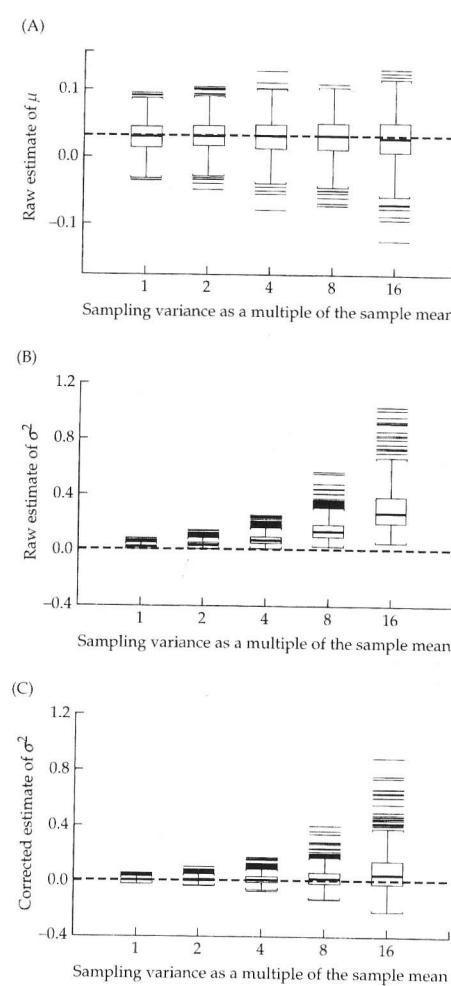


Figure 5.1 Negative binomial distributions describing the frequency of samples with differing numbers of individuals. All the distributions illustrated include 10,000 samples with a mean of 10 individuals per sample. Although the means are the same, the degree of aggregation of individuals among samples, which is reflected in the variance of the distribution as a multiple of the mean, increases from top to bottom. In Plot E (and to a lesser extent in Plots D and C), a few samples had substantially more than 30 individuals but are not shown.

Figure 5.2 (A) Sampling variation does not bias the raw estimates of μ , but it does reduce their precision. (B) Sampling variation in the means of replicate samples introduces bias into raw estimates of σ^2 . (C) The corrected estimate of σ^2 calculated from Equation 5.4 reduces the observation error bias. Dashed lines indicate the true value of each parameter. The rectangular boxes contain the central 50% of the estimates, the heavy line in the center of each box is the median, the whiskers extend to the estimate closest to 1.5 times the interquartile range above and below the box, and more extreme estimates are indicated by separate lines. Estimates from 1,000 simulated data sets, each of 21 years' duration, are included.



profound influence on raw estimates of σ^2 . The distributions of those estimates are illustrated in Figure 5.2B; in all cases, the true value of σ^2 is 0.01. As for $\hat{\mu}$, $\hat{\sigma}^2$ based on only 21 years of data with 10 samples per year has relatively low precision (for example, the central 95% of the estimates lie

between 0.011 and 0.056 at the lowest sampling variance shown in Figure 5.2B). But what is even more worrisome about the raw estimates $\hat{\sigma}^2$ is that they are biased upwards, and to an increasingly large degree as the sampling variance increases. For example, for the highest sampling variance shown in Figure 5.2B, the mean of the 1000 values of $\hat{\sigma}^2$ is 0.306, which overestimates the true value of σ^2 by a factor of 30. Thus if we do not correct our estimates of σ^2 for observation errors, we will obtain unrealistically high values that will lead to correspondingly pessimistic assessments of population viability. Also notice that increasing the amount of observation error progressively degrades the precision of $\hat{\sigma}^2$, just as it does for $\hat{\mu}$.

Fortunately, when counts are based on means of replicate samples, we can do better than to simply use the means alone to estimate σ^2 , as we did in the preceding paragraph. Specifically, we can make use of information contained in the variances of each year's samples to partially correct $\hat{\sigma}^2$ for sampling variation. The logic underlying the correction is detailed in the following paragraphs.

If we consider the single interval between censuses t and $t+1$, our estimate of the log population growth rate over that interval will be $\log \lambda_t = \log(\bar{N}_{t+1} / \bar{N}_t)$. Because $\log \lambda_t$ is a function of the sample means \bar{N}_t and \bar{N}_{t+1} , we can express the variance in $\log \lambda_t$ as a function of the sampling variances of these two means. A useful approximation⁷ for the sampling variance in $\log \lambda_t$ is:

$$\text{Var}(\log \lambda_t) \approx \left(\frac{\partial \log \lambda_t}{\partial \bar{N}_t} \right)^2 \text{Var}(\bar{N}_t) + \left(\frac{\partial \log \lambda_t}{\partial \bar{N}_{t+1}} \right)^2 \text{Var}(\bar{N}_{t+1}) = \frac{\text{Var}(\bar{N}_t)}{\bar{N}_t^2} + \frac{\text{Var}(\bar{N}_{t+1})}{\bar{N}_{t+1}^2} \quad (5.2)$$

Because \bar{N}_t and \bar{N}_{t+1} are means across sampling units, then by the Central Limit Theorem their variances can be approximated by the squared standard error of the mean. That is,

⁷This approximation uses the so-called delta method, which gives an approximation for the variance of a function of independent random variables in terms of the variances of the individual random variables. Specifically, if $f(x,y)$ is a function of the random variables x and y , and if the variances of x and y are not too large, then

$$\text{Var}(f) \approx \left(\frac{\partial f}{\partial x} \right)^2 \text{Var}(x) + \left(\frac{\partial f}{\partial y} \right)^2 \text{Var}(y)$$

A slightly more accurate approximation is obtained by adding the following three terms:

$$\frac{1}{2} \left(\frac{\partial f}{\partial x} \right)^2 \left(\frac{\partial f}{\partial y} \right)^2 \text{Var}(x)\text{Var}(y) - \frac{1}{4} \left(\frac{\partial f}{\partial x} \right)^4 (\text{Var}(x))^2 - \frac{1}{4} \left(\frac{\partial f}{\partial y} \right)^4 [\text{Var}(y)]^2$$

but in practice the contribution of these terms is often negligible.

$$\text{Var}(\bar{N}_t) = s_t^2 / n_t \quad (5.3)$$

where s_t is the standard deviation of the n_t sample counts used to compute \bar{N}_t . Substituting Equation 5.3 into 5.2 thus provides a way to estimate the variance in $\log \lambda_t$ that results from sampling variation, using all of the counts in each of the separate sampling units.

We now have to use this result for one inter-census interval to adjust our estimate of σ^2 . A series of censuses performed in $q+1$ years, each comprised of replicate samples, yields q estimates of $\log \lambda$ and, using Equations 5.2 and 5.3, q estimates of the variance in $\log \lambda$. The simplest approach⁸ is to use the average of the q variances as an overall measure of the mean variance in $\log \lambda$ that is due to sampling variation in the \bar{N}_t 's:

$$\overline{\text{Var}(\log \lambda)} = \frac{1}{q} \sum_{t=1}^q \left[\frac{s_t^2}{n_t \bar{N}_t^2} + \frac{s_{t+1}^2}{n_{t+1} \bar{N}_{t+1}^2} \right] \quad (5.4)$$

Subtracting this component due to sampling variation from the raw estimate $\hat{\sigma}^2$ obtained as in Chapter 3 [i.e., as the variance of the q values of $\log(\bar{N}_{t+1} / \bar{N}_t)$] produces an estimate of the environmental variance in $\log \lambda$ corrected for the sampling variation in the \bar{N}_t 's:

$$\sigma_{\text{corr}}^2 = \hat{\sigma}^2 - \overline{\text{Var}(\log \lambda)} \quad (5.5)$$

The MATLAB code in Box 5.1 computes corrected estimates of σ^2 using both Equation 5.5 and an alternative approach that accounts for different amounts of sampling variation in different years. The sample data in Box 5.1 were generated using Equation 5.1 with $\mu = 0.0296$ and $\sigma^2 = 0.01$, and with the variance of the negative binomial samples equal to four times their expected mean (compare to Figure 5.1C). For this data set, the corrected estimate $\sigma_{\text{corr}}^2 = 0.0101$ is much closer to the true value of σ^2 than is the raw estimate $\hat{\sigma}^2 = 0.0671$.⁹

How well does Equation 5.5 reduce the observation error bias across the many simulated data sets used in Figure 5.2? Most encouragingly, for small to moderate levels of sampling variation, the bias seen in Figure 5.2B is sharply reduced by using Equation 5.5 to estimate σ^2 (Figure 5.2C). Still,

⁸An alternative is to assign different weights to years with different levels of sampling variance; see Chapter 8 for a discussion of White's (2000) method to do this. Also see Box 5.1.

⁹However, for this data set, the estimate of μ , 0.073, is rather bad. Nevertheless, over many replicate data sets, estimating μ as the mean log population growth rate yields nearly unbiased estimates of the true μ even in the presence of sampling variation, as we saw in Figure 5.2A.

BOX 5.1 MATLAB code to correct a raw estimate of σ^2 for sampling variation when census counts represent means from replicate samples.

```
% Program correct_sigma2
% Corrects the estimate of sigma^2 for sampling variation when
% census counts represent means of replicate samples

***** USER-SUPPLIED INFORMATION *****

% The matrix "samples" stores the counts from each sample;
% it has censuses (years) as its columns and individual samples
% as its rows; -1 is used as a place holder at the bottom of
% columns corresponding to censuses with fewer samples
samples=[
10 11 7 0 14 0 6 26 22 12 6 10 22 20 39 31 23 21 14 14 19;
8 4 6 13 1 14 23 10 11 11 10 8 11 27 19 18 16 15 32 38 21;
1 10 11 27 36 16 18 13 29 7 5 2 16 7 25 15 21 12 15 18 9;
7 8 1 9 17 14 9 11 24 13 4 8 12 28 8 15 21 11 30 28 15;
0 16 10 5 16 17 4 19 13 14 21 10 14 6 21 12 8 8 6 14 32;
7 6 5 4 2 15 3 10 13 9 15 22 6 33 14 23 8 14 27 28 17;
5 11 10 21 10 23 12 7 14 12 11 11 22 10 44 5 12 17 21 42 40;
2 16 15 8 6 6 12 14 15 18 9 5 4 7 20 22 14 5 16 13 30;
6 5 12 11 10 6 18 8 17 7 22 13 -1 5 15 7 8 29 25 16 23;
6 6 -1 25 10 14 19 4 16 7 19 6 -1 10 16 18 14 15 11 26 18];

% To calculate confidence limits using the method of White (see
% below), users must supply the critical values of the chi-
% squared distribution with degrees of freedom equal to the
% number of censuses minus one; see Chapter 3 of Morris and Doak,
% Quantitative Conservation Biology, for how to
% compute these values. The following are critical chi-square
% values for p=0.025 and p=0.975 with 20 degrees of freedom:
chi2crit=[34.16958143, 9.590772474];

% Enter tolerance for computing corrected estimate and its
% confidence limits using White's method; smaller tolerance means
% more accuracy
tolerance=1E-8;

*****
```

BOX 5.1 (continued)

```

options=optimset('TolX',tolerance);
          % set tolerance for the
          % function fzero below

% The vector n stores the number of samples for each census
n=sum(samples>=0);

% q=number of censuses minus 1
q=size(samples,2) - 1;

% Compute means and variances of the samples from each census
for t=1:q+1
    samplest=samples(1:n(t),t);
    Nbar(t)=mean(samplest);
    Vs(t)=var(samplest);
end;

% Compute raw estimates of mu and sigma^2 using the conventional * method
for t=1:q
    loglam(t)=log(Nbar(t+1)/Nbar(t));
end;
muest=mean(loglam);
disp('Raw estimate of mu:');
disp(muest);
s2raw=var(loglam);
disp('Raw estimate of sigma^2:');
disp(s2raw);

% Compute the component of total variance due to sampling
% variation
for t=1:q
    SampleVar(t)=Vs(t)/(n(t)*Nbar(t)^2) + ...
    Vs(t+1)/(n(t+1)*Nbar(t+1)^2);
end;
MeanVar=sum(SampleVar)/q; % mean sampling variation across years

% Compute the simple corrected estimate of sigma^2
s2corr=s2raw-MeanVar;
disp('Sigma^2 with simple correction for sampling variation');
disp(s2corr);

```

BOX 5.1 (continued)

```

% Compute the corrected estimate of sigma^2 using the method of
% White, which weights by the sum of the environmental and
% sampling variance in each year; see White 2000.
Dev2=(loglam-muest).^2;
s2corrwhite=fzero(inline('sum(Dev2./(s2+SampleVar))-q+1',...
    's2','Dev2','SampleVar','q'),s2raw,options,Dev2,SampleVar,q);
disp('Sigma^2 with White''s correction for sampling variation');
disp(s2corrwhite);

% Compute confidence interval for White's version of the
% corrected sigma^2;
% WARNING: Users may need to try different values of 'start'
% below to get reasonable values for the confidence limits, as
% there may be more than one root to the equation being solved by
% the function fzero.

start=0.001;
s2lower=fzero(inline('sum(Dev2./(s2+SampleVar))-chi2', 's2', ...
    'Dev2','SampleVar', 'chi2'),...
    start,options,Dev2,SampleVar,chi2crit(1));

start=0.1;
s2upper=fzero(inline('sum(Dev2./(s2+SampleVar))-chi2', 's2', ...
    'Dev2','SampleVar', 'chi2'),...
    start,options,Dev2,SampleVar,chi2crit(2));

CI=[s2lower s2upper];
disp('Confidence interval for White''s correction:');
disp(CI);

```

¹ Note that this program provides an alternative to using results based on diffusion approximations in Chapter 3. That is, rather than assume the log population growth rate is normally distributed and that large environmental perturbations do not occur, users can easily modify this program to choose population growth rates *only* from among those actually observed. Specifically, set "probout" to 1 and put all the observed population growth rates (*not* log population growth rates) in "outliers". (Of course, other modifications such as eliminating the "if rand>probout" statement would make the program faster.) Also see the program "randdraw" in Box 2.1.

five caveats must be kept in mind. First, when the variance among samples is large, σ_{corr}^2 reduces, but does not eliminate, the observation error bias. For example, when the sampling variance was 16 times greater than the mean, the average of the 1,000 corrected estimates was 0.076, substantially

greater than the true value, although much closer to it than the mean of the uncorrected estimates (compare Figures 5.2B and C). As the sampling variance becomes large, the approximation used in Equation 5.2 breaks down. Second, even though it reduces the bias, the corrected estimate does not reduce the imprecision that observation error introduces into the raw estimates. (Note the similar ranges of the raw and corrected estimates in Figures 5.2 B and C). When the sampling variance is high, a large number of samples are needed to achieve precise estimates of σ^2 (Figure 5.3). Note that while increasing the sample size also reduces the bias in the raw estimates, the cor-

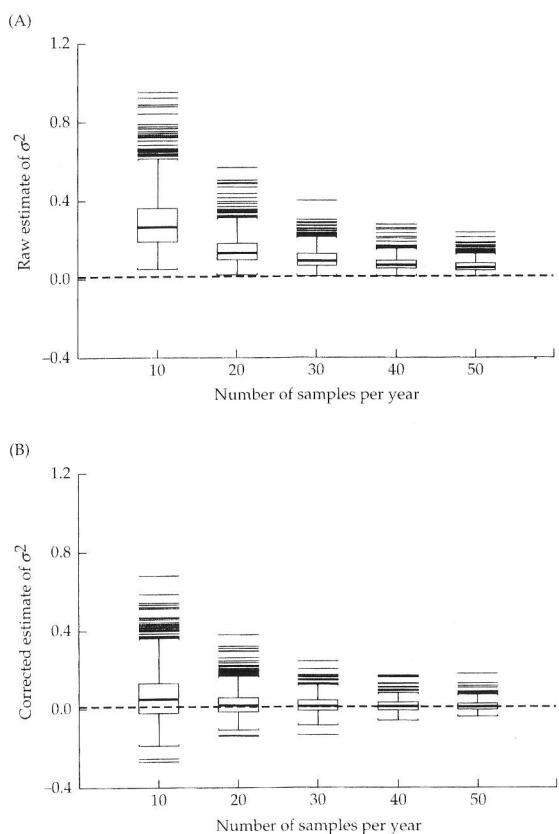


Figure 5.3 Increasing the number of samples taken per year decreases the imprecision in both the raw (A) and corrected (B) estimates of σ^2 . Here, the true value of σ^2 is 0.01 (dashed lines), and the sampling variance is 16 times greater than the mean. Symbols as in Figure 5.2.

rected estimate is always less biased than the raw estimate (means of the estimates in Figure 5.3 for sample sizes of 10 through 50 were 0.0711 versus 0.2971, 0.0281 versus 0.1485, 0.0213 versus 0.1026, 0.0157 versus 0.0774, and 0.0146 versus 0.0643 for the corrected versus raw estimates, respectively). Because it will rarely be possible to estimate μ and σ^2 with as much precision as we would desire, it is important to compute confidence intervals for our estimates.¹⁰ Third, the model used to generate the data sets in Figures 5.2 and 5.3 was density-independent (i.e., Equation 5.1). The method used in Box 5.1 may be less successful at correcting for sampling variation when the data reflect the influence of density dependence. Fourth, the approximation underlying Equation 5.2 assumes that the mean of the samples, \bar{N}_t , is an unbiased estimate of the true mean population density in year t . When undercounting and overcounting cause \bar{N}_t to be a biased estimate of the true density, σ^2_{corr} will be less successful at removing the influence of observation error. For example, let's assume that counting errors cause the number of individuals observed in each plot to vary uniformly among years between 0% and 200% of the numbers actually present (that is, in some years all the counts are over-counts and in other years all are under-counts). Relative to the case when \bar{N}_t is an unbiased estimate, this level of counting error increases σ^2_{corr} by approximately 20% when the variance of the samples equals their mean, but it increases σ^2_{corr} by only 3% when the variance is 16 times greater than the mean (results not shown). Taking steps to assure that the counts are as accurate as possible is the best way to reduce this bias. Finally, when the sampling variance is high, Equation 5.5 can produce negative values of σ^2_{corr} (Figures 5.2C and 5.3). Because a true variance cannot be negative, we would have little choice in such cases but to conclude that high sampling variation makes it impossible to rule out the possibility that the true value of σ^2 is actually zero. By running a pair of PVAs with σ^2 set both at zero and at the raw estimate, we could at least delimit the optimistic and pessimistic bounds of reality, respectively. Despite these caveats, when the variance among samples is modest, density dependence is weak, and the sample counts are reasonably accurate, Equation 5.5 presents a useful tool for obtaining better estimates of σ^2 in the face of sampling variation.

Using Repeated Censuses of the Same Area to Discount for Observation Error

We have to modify the approach taken in the preceding section to fit another commonly used method for censusing populations. Assume that in most

¹⁰An approximate confidence interval for σ^2_{corr} can be calculated using Equation 3.13 with σ^2_{corr} in place of $\hat{\sigma}^2$. However, we should recognize that the resulting interval is likely to be too narrow, as the range of possible values of σ^2_{corr} increases as the degree of sampling variation increases (see Figure 5.2C). For an alternative approach, see White (2000). Both Boxes 5.1 and 8.1 use White's method, which is further explained under *Discounting sampling variation* in Chapter 8.

years, we estimate population size by obtaining a single count of all individuals in an area that represents a known fraction of the total range of the population and by extrapolating that count to the population as a whole. (That is, if in year t we observe c_t individuals in an area that is a fraction b_t of the entire range, where $b_t \leq 1$, the estimate of total population size is $C_t = c_t/b_t$.) However, in some years, we repeatedly census a certain area (which may be smaller than the total area censused that year) over a short period of time in order to quantify the magnitude of observation error. We will assume that the variation among these repeated censuses is due to counting errors; it is not sampling variation because we are not censusing separate plots as we did in the preceding section. Because we do not have a set of replicate samples taken every year, we cannot simply use the sampling variances associated with the means of the samples (as we did in Equations 5.2 through 5.5) to correct C_t for observation error. However, we can try to extrapolate the observation error we saw in the repeated censuses to the whole-population scale, as follows. If V_t is the variance of the counts from the area repeatedly sampled in year t , a_t is the size of the area occupied by the entire population as a multiple of the area repeatedly sampled in year t (where $a_t \geq 1$), and C_t is the estimate of the total population size in year t , then the variance in C_t that is due to observation error is $\text{Var}(C_t) \approx a_t^2 V_t$. If $\log(C_{t+1}/C_t)$ is the estimate of the log population growth rate in year t , then replacing \bar{N}_t with C_t in Equation 5.3 yields

$$\text{Var}(\log \lambda_t) = \frac{\text{Var}(C_t)}{C_t^2} + \frac{\text{Var}(C_{t+1})}{C_{t+1}^2} = \frac{a_t^2 V_t}{C_t^2} + \frac{a_{t+1}^2 V_{t+1}}{C_{t+1}^2} \quad (5.6)$$

Averaging the values of $\text{Var}(\log \lambda_t)$ across all years of the census and subtracting the average from the raw estimate $\hat{\sigma}^2$ computed using the total counts [i.e., the variance of the $\log(C_{t+1}/C_t)$'s] gives an estimate of σ^2 corrected for observation error. If repeated sampling was performed only once, the simplest option is to associate the single estimate of the observation error variance with each year's total count (i.e., use the single value of V_t for all t 's in Equation 5.6).¹¹ However, if repeated sampling was performed in years with different population sizes, we can regress the separate variance estimates $\text{Var}(C_t)$ against the total counts in those years, and we can then use the resulting regression equation to predict the likely variances of the total counts from years in which repeated sampling was not performed.

As for the previous method, several caveats apply to Equation 5.6. First, it is based on the assumption that the $\text{Var}(C_t)$'s are not too large, and so it may not compensate completely for high levels of observation error. Second, as

¹¹Alternatively, if we believe that counting errors are more likely in a larger population, we could set the observation error variance to be a constant multiple of the estimated total population size.

for N_t values in Equation 5.3, Equation 5.6 assumes that the C_t 's are unbiased estimates of the true population size. This assumption may be correct when the observation error is due to sampling variation, the sampling is appropriately randomized, and the counts are fairly accurate, but it is not strictly true when the observation error results only from counting errors. Thus we can expect to be less successful at discounting for observation error in the present case than we were in the preceding section.

A Method to Reduce the Impact of Fluctuations in Population Structure and Observation Errors without Replicate Samples

In many cases, we will completely lack replicate counts taken during a single census. For example, every census may have taken only a total count, or the published source of the data may report only the means, not the separate counts in each sample. In such cases, we cannot use the methods described in the preceding sections. However, for a certain type of species and census, an alternative method to reduce the influence of observation error has recently been proposed by Holmes (2001). The method is most appropriate for species that have a short and well-known lifespan, with individuals that reproduce only once, and for which the census targets only a subset of the population (e.g., breeding females). For reasons we explain in detail below, we do not advocate the use of this method in other cases, although future development of this new technique may make it more widely applicable. The essence of the Holmes method is to use not the raw counts, but *running sums* of a predetermined number of contiguous counts, to estimate μ and σ^2 . For example, if we are using a running sum with a length of three censuses, we would add the counts from the first, second, and third censuses, the second, third and fourth censuses, the third, fourth, and fifth censuses, and so on. With a total of n censuses and a running sum of length L , we would have $n - L + 1$ such running sums. These running sums are then used in a modification of the regression method of Dennis et al. (1991) to estimate μ and σ^2 .

There are two distinct arguments for using running sums instead of single counts. First, when the counts include only a subset of a population, variation in the counts can be caused by fluctuations in the structure of the population (i.e., the relative proportions in different age, size, or developmental classes), as well as variation in environmental conditions. Ideally, we want our estimate of σ^2 to reflect only environmentally driven fluctuations in total population size, but the number of individuals in a particular class may vary more or less than does the number of individuals in the population as a whole, so that an estimate of σ^2 based on a subset of individuals may misrepresent the population-wide σ^2 . Holmes (2001) argued that a running sum of the observed number of individuals in one class will more accurately represent the size of the population as a whole. To see why, consider a population of salmon in which individuals spawn at age 4 or 5 and then die. We census the population by counting adults as they return to

spawning streams, but we know that the population also includes pre-breeding individuals that remain at sea and so are not counted. However, if we had additional demographic information, we could use the five most recent spawner counts to calculate the *total* population size at the present time. At the current census, the population will include, in addition to the observed spawners, one-year-olds that were produced by last year's spawners, two-year-olds that were produced by the spawners two years previously, and so on. Thus if S_t is the number of spawners in year t , then the total size of the population in year t is

$$T_t = S_t + f_{t-1}S_{t-1} + f_{t-2}S_{t-2} + f_{t-3}S_{t-3} + f_{t-4}S_{t-4} \quad (5.7)$$

where f_t is the number of offspring each spawner produces in year t and $s_{x,t}$ is the fraction of offspring born in year t that survive to age x and do not return to spawn. We know that the number of terms in the sum should be no larger than 5, because we know that individuals cannot live longer than 5 years. In addition, because we know that individual salmon spawn only once, we know that the spawners observed in a given year will not include any individuals observed to spawn in previous (or subsequent) years, so there is no double counting in Equation 5.7. According to Equation 5.7, the total size of the population at the current census is a *weighted* sum of the five most recent spawner counts, where the weights are determined by the values of the f_t 's and the $s_{x,t}$'s, which change over time due to environmental variation. Unfortunately, in most cases, we do not know these values and so cannot compute the proper weights. Holmes proposed that in this situation, we use the *unweighted* sum of the five most recent counts as a surrogate for total population size, to at least partially reduce the potential for confounding fluctuations in population structure with environmental stochasticity affecting total population size.

The second reason to use running sums is that they can partially counteract observation errors. Counting errors will sometimes cause a single count to be an overestimate and sometimes an underestimate of the true population size. This is one reason why observation error artificially inflates the estimate of σ^2 . However, the odds are small that over a set of contiguous censuses, *all* the counts will be overestimates or all of them underestimates. Rather, an overestimate will frequently be followed by an underestimate, and vice versa. That is, over a set of censuses, the positive and negative errors will tend to cancel each other out, so that the sum of the counts may be closer to the sum of the true population sizes than any one count is to the true population size in a single year. However, summing adjacent counts also removes some of the *real* year-to-year variation in population size, so that the variability of running sums may *underestimate* the annual variability of population size, especially if observation errors are actually small and the length of the running sum is large. The goal is to choose a running sum length

that reduces observation error effects without leading to an underestimate of true population variability. The principal "rub" of the Holmes method is that there is no clear way to determine *a priori* what is the optimal running sum length, as we discuss in detail below. But first, we describe the method in its entirety.

As we noted above, the Holmes method employs a modified version of the regression method of Dennis et al. (1991) described in Chapter 3. In the Dennis et al. method, we regress the transformed log population growth rate between two successive censuses against the square root of the time elapsed between them. Instead, using the Holmes method we first match up all possible pairs of running sums separated by different elapsed times of $\tau = 1, 2, \dots, \tau_{max}$ years. We then compute the log population growth rate over that time interval as $\log(R_{t+\tau}/R_t)$, where R_t is the running sum beginning in year t . Next, we compute the means and variances of the $\log(R_{t+\tau}/R_t)$ values for each value of τ .¹² Finally, we separately regress these means and variances against τ . The slope of a linear regression of the means against τ is an estimate of μ , and the slope of a linear regression of the variances against τ is an estimate of σ^2 . A major difference between the regression method of Dennis et al. and that of Holmes is that in the former, the regression intercept should always be zero, whereas in the latter it is a fitted parameter. (In the case of the regression to estimate σ^2 , the intercept represents the nonprocess error variance, that is, observation error plus "errors" due to fluctuations in population structure; see Holmes and Fagan, in press.) The reasoning behind these regressions is that the means and variances of $\log(R_{t+\tau}/R_t)$ are μ and σ^2 multiplied by the time elapsed between samples, τ (see Chapter 3 if you don't remember this idea). Thus, the slopes of these regressions are in fact measurements of μ and σ^2 . Box 5.2 lists MATLAB code defining the function `dennisholmes` that estimates μ and σ^2 using Holmes' running-sum modification of the Dennis et al. regression method. Users must pass three things to the function: the name of a vector containing the census counts, the length of the running sum, and the maximum time lag in the linear regressions.¹³

To illustrate the ability of the Holmes method to reduce biases in σ^2 caused by observation error, we generated 1,000 simulated data sets from the stochastic density-independent model (Equation 5.1) with known values of μ and σ^2 , as above. We then multiplied the "true" counts in each year by a

¹²Although the Dennis et al. method can easily handle them, missing censuses would cause the lengths of the running sums to differ when using the Holmes method. However, if there are relatively few missing census values, linear interpolation can be used to replace missing census counts.

¹³For example, if the vector `counts` contains the census counts, the command `[muest, s2est]=dennisholmes(counts, 3, 5)` will place the estimates of μ and σ^2 into the variables named `muest` and `s2est`, respectively, using a running sum of 3 and a maximum time lag of 5.

BOX 5.2 MATLAB code defining the function "dennisholmes", which estimates μ and σ^2 using the method of Holmes (2001).

```

function [mu,sigma2]=dennisholmes(RawCounts,L,taumax)
% dennisholmes(RawCounts,L,taumax) computes mu and sigma2
% using the Dennis-Holmes method;
% RawCounts=vector of census counts
% L=length of running sum
% taumax=maximum time lag in regressions;
% See Holmes, Proceeding of the National Academy of
% Sciences USA 98: 5072-5077 (2001).

NumCounts=length(RawCounts);

% First, compute running sums

RunSum=[];
for i=1:NumCounts-L+1
    RunSum=[RunSum sum(RawCounts(i:i+L-1))];
end;

% Second, compute means and variances of the log population
% growth rate at each time lag, tau

taus=[];      % array to store time lags
means=[];      % array to store means
vars=[];       % array to store variances
for tau=1:taumax
    % For each time lag,
    R=[];
    for i=1:NumCounts-L+1-tau
        % compute log growth
        R=[R log(RunSum(i+tau)/RunSum(i))]; % rates R using running
    end;                                % sums, then compute
    taus=[taus; tau];
    means=[means; mean(R)];             % their means and
    vars=[vars; var(R)];               % variances
end;

% Third, compute regression slopes of means and vars. vs. tau

% Step 1: center tau, means, and vars by subtracting
% their respective means
taus=taus-mean(taus);

```

BOX 5.2 (continued)

```

means=means-mean(means);
vars=vars-mean(vars);

% Step 2: with centered dependent and independent variables,
% the right-hand sides of the following two lines
% give the slopes of linear regressions of the means and
% variances of the log growth rates on tau, which are
% our estimates of mu and sigma^2
mu=(taus'*means)/(taus'*taus);
sigma2=(taus'*vars)/(taus'*taus);

```

lognormally distributed random variable to mimic observation errors. The observation error distribution had a mean of 1 and a variance of either 0.01 or 0.05. For each corrupted data set, we then estimated μ and σ^2 by the standard method (i.e., as the mean and variance of the observed log population growth rates, using the counts contaminated with observation error) and by the Holmes method, using the function `dennisholmes`. Both methods produce unbiased estimates of μ , although with 20 years of data the estimates range widely around the true value (Figure 5.4A,C). However, the standard method yields estimates of σ^2 that are strongly upwardly biased, especially when the observation error variance is large (Figure 5.4B,D). For example, with an observation error variance of 0.05, the standard method gives a median estimate of σ^2 that is 10 times larger than its true value. Applying the Holmes method to the same data sets using a running sum length of 3 yields values for σ^2 that are again broadly distributed,¹⁴ but that on average slightly underestimate the true value when the observation error variance is 0.01 and are nearly unbiased when the observation error variance is 0.05 (Figure 5.4B,D).

Thus the Holmes method does provide a way to reduce bias in estimates of σ^2 caused by observation error. However, as we noted above, its ability to do so without *underestimating* the true σ^2 hinges on making the correct choice.

¹⁴Because of this broad distribution, it is important to calculate a confidence interval for the running sum estimate of σ^2 so that we know how much faith to put in it. Holmes and Fagan (in press) show that an approximate $100(1 - \alpha)\%$ confidence interval for an estimate $\hat{\sigma}^2$ obtained by the Holmes method is given by $[df \hat{\sigma}^2 / \chi_{\alpha,df}^2, df \hat{\sigma}^2 / \chi_{1-\alpha,df}^2]$, where $\chi_{\alpha,df}^2$ is the 100α th percentile of the chi-square distribution with degrees of freedom given by $df = 0.333 + 0.212n - 0.387L$, where n is the number of census counts and L is the running sum length.

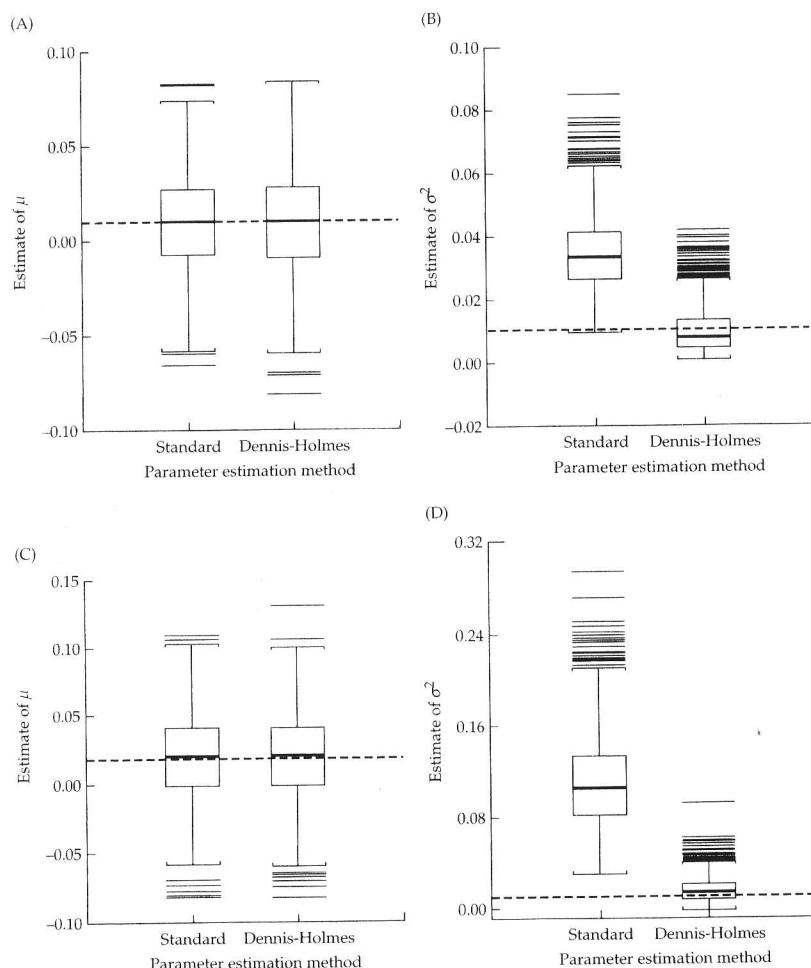


Figure 5.4 Demonstration of the effectiveness of the Dennis-Holmes method compared to the standard method of estimating μ and σ^2 . In all cases the true value of both μ and σ^2 is 0.01 (dashed lines). The observation error variance was 0.01 (A,B) or 0.05 (C,D). Distributions of estimates for 1,000 simulated data sets, each of 21 years' duration, are shown. Symbols are as in Figure 5.2. The running sum length was 3, and the maximum time lag used in the regressions was 4.

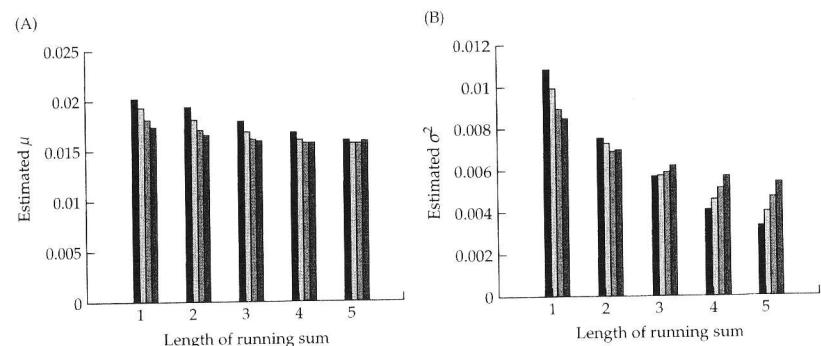
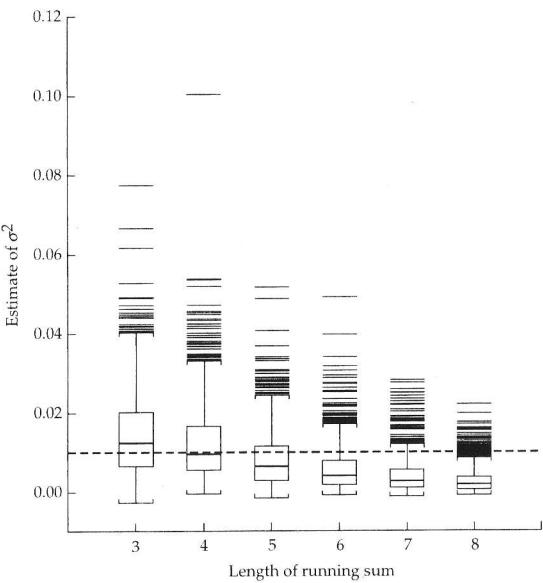


Figure 5.5 Estimates of μ and σ^2 for the Yellowstone grizzly bear using the Holmes method are sensitive to the choice of L (length of the running sum) and τ_{max} (maximum time lag in the regression). Within each group of bars with the same running sum length, τ_{max} increases from 3 (left) to 6 (right).

es for both τ_{max} , the maximum time lag in the regressions, and more importantly L , the length of the running sum. Both of these choices, but particularly the second, can strongly influence the resulting parameter estimates. For example, we applied the Holmes method to the census counts for the Yellowstone grizzly bear population (see the data in Table 3.1), varying the maximum time lag and running sum length (Figure 5.5). The estimate of μ is relatively insensitive to the choice of τ_{max} and L (Figure 5.5A), but there is a nearly threefold range of variation in the estimates of σ^2 across the values of τ_{max} and L (Figure 5.5B). Most notable is the fact that the longer the running sum, the smaller the estimate of σ^2 will be. Of course, we don't know what the "true" value of σ^2 is for the Yellowstone grizzly bear population (indeed, that is why we are trying to estimate it), so there is no clear way to determine which value of L yields the least-biased estimate. Using the simulated data sets above, where we do know the "true" value of σ^2 , it becomes clear that choosing to compute too long a running sum will cause the resulting values to sharply underestimate the true value, thus producing overly optimistic viability assessments (Figure 5.6). Remember that this effect arises because calculating running sums averages out not only observation error, but also real variation in numbers due to environmental variability.

Thus we are left with a conundrum as to how to best determine the running sum length. It is clear from the logic underlying Equation 5.7 that if our goal is to minimize the effect of fluctuations in population structure on our estimate of σ^2 , we should choose a running sum length that is close to the lifespan of the organism, hence approximating the total population size. Unfortunately, for long-lived organisms and with a census of limited dura-

Figure 5.6 Using too long a running sum in the Holmes method can lead to severe underestimates of σ^2 . The true value of σ^2 is indicated by the dashed line. Observation error variance = 0.05 and $\mu = 0.01$. Each distribution shows estimates of σ^2 for 1,000 random data sets. Other symbols are as in Figure 5.2.



tion, it will be difficult to compute such long running sums and still have enough of them to estimate parameters accurately (for example, for grizzly bears, whose potential lifespan exceeds 30 years, even the 39 year census of the Yellowstone population would yield only 10 running sums of length 30 years). If our goal is to minimize observation error bias rather than the effects of fluctuations in population structure, then there are no clear guidelines as to how long the running sums should be. This question is currently the focus of active research. Until more precise guidelines emerge, we make two provisional recommendations:

1. Use the Holmes method only for organisms that live for only a few (but more than one) years and reproduce once (i.e., semelparous animals or monocarpic plants), and do not set the running sum length to be any longer than the known lifespan. If our goal in using the running sum is to approximate the total population size at the present time, adding more terms to the running sum is superfluous, and merely risks downwardly biasing the estimate of σ^2 .

2. Repeat the analysis using a running sum length L that varies from 1 to the known lifespan, and see how much the predicted population viability is affected. The true value of σ^2 is likely to lie somewhere between the value obtained with $L = 1$ (which is likely to be an overestimate due to fluctuations in population structure and observation errors) and the value obtained with L equal to the lifespan (which is likely to be downwardly biased due to excessive averaging).

There is another reason to keep the running sums relatively short. Using short running sums, we are more likely to overestimate rather than underestimate the value of σ^2 . The resulting somewhat pessimistic estimates of population viability might cause us to take action to manage a population sooner than we otherwise might, a sensible outcome if we are taking a precautionary approach. More generally, if observation errors or fluctuations in population structures are thought to be small relative to real variability, we recommend not using the Holmes method, as it is likely to underestimate σ^2 and thus overestimate viability. Finally, if your data seem appropriate for the Holmes method, we suggest that before proceeding with any analysis you look for emerging studies that explore the methodology and tradeoffs of this still very new approach (e.g., Holmes and Fagan, in press).

A Directory to More Advanced Methods for Estimating Parameters in the Face of Observation Error

In this chapter, we have described a few simple methods to reduce the influence of observation error and to discount it when it occurs. More powerful but also more complex methods exist for estimating population parameters in the face of observation error. In closing the chapter, we direct interested readers to references on two such methods.

The first approach we briefly describe is the so-called total least-squares method (Ludwig and Walters 1981, Ludwig et al. 1988, Ludwig 1999). To use it, one assumes that the observed log population size in year t equals the log of the true population size in year t plus a deviation due to observation error. In turn, the true log-population size in year $t + 1$ represents the size predicted by the deterministic part of the population model given the true population size in year t plus a deviation due to environmental stochasticity. One then writes a sum of squares term as a weighted sum of the squared deviations due to observation errors and the squared deviations due to environmental stochasticity. Then one uses numerical methods to find values for the model parameters as well as the observation error deviations (which are treated as unknown parameters) that minimize this sum of squares. To implement this method, one must first obtain an estimate of the observation error variance, for example as the variance of replicate sam-

ples (see Carpenter et al. 1994), or by repeatedly censusing the same area, as described above. Alternatively, one can use a prior estimate of the ratio of the observation error variance and the environmental variance (Ludwig and Walters 1981, Ludwig et al. 1988, Ludwig 1999). If one lacks an estimate of the observation error variance and simply guesses at the ratio of the two variances, biased estimates of the model parameters and associated extinction risk metrics will result if the guess is not reasonably accurate (see Figure 1 in Ludwig 1999). For the many count-based data sets for which no information about the observation error variance is available, the method of total least squares will be impossible to apply.

The second, more sophisticated approach, state-space modeling, does not require prior information about the observation error variance. A state-space model has two submodels, a population model that describes the dynamics of the “true” but unobserved population size in a stochastic environment and an observation model that describes how the observed population size is related to the true population size given possible sources of observation error. The data are a series of observed population sizes (i.e., census counts). To estimate parameters, one begins by writing a likelihood function to describe the probability of obtaining the observed population sizes given a particular set of values of the parameters in the population and observation submodels (including the environmental variance and the observation error variance). One then searches for parameter values that maximize the likelihood function.¹⁵ In general, the likelihood function involves a complex product of integrals. Thus, calculating the likelihood for a set of parameter values requires either direct numerical integration (de Valpine and Hastings 2002) or stochastic integration schemes such those based on Markov Chain Monte Carlo techniques (Carlin et al. 1992, Bjørnstad et al. 1999). At present, implementing these methods is a highly technical endeavor. However, as they become more generally used in ecology over the coming years, we expect they will become widely applied tools for accounting for observation error in population viability analyses.

¹⁵See Chapter 4 for a general description of likelihood approaches to model fitting.

6

Demographic PVAs: Using Demographic Data to Build Stochastic Projection Matrix Models

The methods we have discussed in previous chapters assume that all individuals are identical, so that the models only need to consider the total number of individuals. But for many organisms, particularly the long-lived ones that are often of special conservation concern, individuals differ in important ways that affect their current and future contributions to population growth. For example, larger individuals often have a greater chance of surviving and a higher reproductive rate than smaller individuals. As a result, two populations that differ in the proportion of larger versus smaller individuals may have very different viabilities, even if they are of the same total size and experience the same sequence of environmental conditions. Populations in which individuals differ in their contributions to population growth are known as *structured populations*. The principal tool for assessing the viability of structured populations is the *population projection matrix* model, which divides the population into discrete classes and tracks the contribution of individuals in each class at one census to all classes in the following census. Different variables describing the “state” of an individual, such as size, age, or stage (e.g., seeds, seedlings, and adults for plants, or larvae, juveniles, and adults for fish or invertebrates), are most appropriate for different organisms as ways to classify individuals. As was true of the models for unstructured populations in Chapters 3, 4, and 5, projection matrix models must account not only for mean contributions of one stage to each of the others, but also for variability in these rates if they are to yield accurate assessments of population fate.

Projection matrix models have advantages and disadvantages relative to the simpler count-based models covered in the previous chapters. As we just indicated, one advantage is that structured models provide a more accurate