OVERVIEW

TOWARDS RELIABLE BIRD SURVEYS: ACCOUNTING FOR INDIVIDUALS PRESENT BUT NOT DETECTED

William L. Thompson

U.S. Geological Survey Arkansas Cooperative Fish and Wildlife Research Unit, Department of Biological Sciences, University of Arkansas, Fayetteville, Arkansas 72701, USA

Using just the count of birds detected (per unit effort) as an index of abundance is neither scientifically sound nor reliable... It is necessary to adjust study counts by the detection probability. (Burnham 1981:325)

Counts of birds seen, heard, or captured are commonly used to elucidate avian-habitat relationships, investigate responses of avian populations to management treatments or to environmental disturbances, estimate spatial distributions of species, and monitor population trends. The point-count method, in which an observer records all birds detected within either a fixed or an unlimited distance from a point during a specified time period (Ferry and Frochot 1970, Hutto et al. 1986), is the most widely used counting method in bird population studies (Ralph et al. 1995, Rosenstock et al. 2002). Point counts and other methods that are based on observed counts to estimate abundance, such as mist netting (Karr 1981), rely on the assumption that numbers of individuals detected (e.g. seen, heard, or captured) represent a constant proportion of actual numbers present across space and time. That is, if the true number of birds within a surveyed area increases by 20% during successive samples, observed counts are assumed to increase by the same percentage. Similarly, counts in different areas during the same time period are assumed to represent the same proportion of birds present within each of those areas. The validity of this proportionality assumption has been questioned for decades (e.g. Burnham 1981) because the many factors affecting detection probabilities of individuals (i.e. probability of correctly identifying the presence of an individual) are neither constant within and among species and habitats nor constant across time (e.g. see Thompson et al. 1998; and Rosenstock et al. 2002 and references therein). Nonetheless, ornithologists continue to rely on survey methods whose results depend upon validity of the proportionality assumption for meaningful interpretation.

Violation of the proportionality assumption may lead to incorrect conclusions regarding comparative numbers, spatial distributions, trends, or habitat relationships of bird populations. For example, assume a bird population counted within an area had an average detection probability of 0.4 during an initial survey and one of 0.2 during a later survey, but actual number of individuals remained constant over both surveys (e.g. 100 birds). The ratio of observed counts, where count 1 would yield 100 × 0.4 = 40 birds and count 2 would produce 100 × 0.2 = 20 birds, would incorrectly indicate a 40/20 = 50% population decline when there was no decline. A similar example can be applied to populations in different areas or habitats (and hence detection probabilities) during a single time period. Further, relatively small differences (e.g. >9%) in detection probabilities can lead to misleading conclusions (J. R. Sauer and W. A. Link unpubl. manuscript). The proportionality assumption may only be safely ignored when either at least 95% of the individuals present are counted (White and Bennett 1996) or, if monitoring trends, the change in population numbers is large relative to degree of undercounting. Otherwise, counts will need to be properly adjusted for individuals present but not detected to avoid potential for false patterns generated by flawed data. Rosenstock et al.’s (2002) finding that 95% of sampled articles relied on unadjusted counts is all the more sobering in light of that potential for misleading results.

In this overview, I describe a general sampling framework for conducting population studies (see Skalski and Robson 1992, Thompson et al. 1998, and Morrison et al. 2001 for more details), potential sources of error associated with population estimates, and how bird
counts fit within the general study framework. I then review approaches to obtaining bird population estimates recently suggested by Bart and Earnst (2002), Nichols et al. (2000), and Rosenstock et al. (2002) that adjust for birds present but not detected and therefore produce reliable estimates when appropriately applied. My objectives are to provide a broader context within which these and other counting methods can be viewed and to discuss conditions under which these various approaches would be expected to yield reliable results.

GENERAL STUDY DESIGN FRAMEWORK

Assuming one has developed meaningful questions or objectives to be addressed, designing a population study proceeds with defining the population of interest within a specific area and time period. This is the target population, which is used in a statistical sense and hence may or may not exactly correspond to a biological population. The next step is to define the sampling frame, which is a complete listing or mapping of sampling units (e.g. plots, quadrats, and transects; Cochran 1977). An example of a sampling frame in a bird population study would be an area of interest subdivided into sampling units that are explicitly delineated. Each sampling unit may or may not contain birds. A more loosely defined frame may consist of an area whose boundary is explicitly delineated and within which sampling units are randomly placed and surveyed (Thompson et al. 1998:7–10).

Although one hopes that the sampled population is the same as the target population, often subareas within the original study area may not be accessible for a variety of reasons, such as private landowners forbidding access. Thus, these subareas are not available for surveying and hence are not part of the sampled population. Statistically based inferences cannot be validly extended to them. Other common examples are counts conducted along roads or trails so that other areas of interest have zero probability of being surveyed. Here the sampling frame would be composed only of roads or trails and their adjacent areas where birds have a nonzero probability of being sampled, but perhaps not detected, during the survey. Consequently, inferences cannot be properly made to bird populations beyond the surveyed area unless one is willing to assume areas on and adjacent to roads or trails support similar numbers of birds as those away from those features. This assumption seems questionable given that roads and trails are typically not placed randomly and factors affecting their adjacent habitat structure and composition may be expressed differently than in surrounding areas.

SOURCES OF ERROR IN POPULATION ESTIMATES

Two types of error may be associated with bird population estimates—bias and variance. Bias is a systematic error that leads to either underestimation (negative bias) or overestimation (positive bias) of the parameter of interest, such as abundance. This error may arise from non-random selection of sampling units, such as conducting counts exclusively from roads (selection bias; see Thompson et al. 1998), and from the counting process. Bias in the counting process is subdivided into response error and nonresponse error. Response error refers to a misrecording of information on a detected individual, such as misidentifying one species for another (e.g. looks or sounds similar). Another example would be misrecording data on to a data sheet. Nonresponse error arises from not detecting every individual within a surveyed area and failure to adjust count results accordingly. That is, unadjusted counts would exhibit negative bias relative to true numbers of individuals because detection probabilities would be <1. Conversely, detection probabilities would be overestimated (i.e. assumed to be 1) and hence exhibit positive bias for an incomplete count. Our ability to produce reliable population estimates frequently revolves around proper estimation of detection probability.

Variance, the second type of error associated with population estimates, is a measure of precision, which is the degree of spread in parameter estimates over repeated samples. Multiple complete counts within a specific area would have no variance if the number of individuals present did not change during the counting period. Imprecise or “noisy” counts may obscure signals in the data that, for instance, may lead to inaction by managers when a management response is warranted. Note that single point counts provide no measure of precision; even variance estimates based on multiple point
counts lack a theoretical foundation so that they are likely biased by some unknown amount.

An additional component of uncertainty associated with bird counts is variation in numbers across time and space. This form of variation is due to environmental and demographic processes affecting avian abundance and spatial distribution (Burnham et al. 1987). That is, its source is not related to counting or random selection of sampling units.

**CATEGORIZATION OF BIRD COUNTS**

Methods for counting birds or other quantities of interest, such as nests, may be described by a hierarchy of dichotomies (Fig. 1). The initial hierarchical level distinguishes complete from incomplete counts, whereas the second level separates counts based on spatial scale of application. A complete count of birds within an entire study area over a specified time period represents a true census. Ornithologists frequently misuse “census” as a synonym for “survey,” the latter referring to an incomplete count. For instance, in its common usage, “strip census” does not refer to a complete count but rather to an (unadjusted) incomplete count along a line of fixed width.

Complete counts are rarely possible in studies of mobile populations in general, and bird populations in particular, except perhaps at smaller spatial scales or within habitats where individuals or other quantities of interest are readily detectable by the counting method. Complete counts within selected portions of a study area, such as true plot or strip censuses, produce survey estimates because only a portion of the population of interest was completely counted, which then is extrapolated to the entire area (Fig. 1).

When complete counts are not possible, one obtains an incomplete count either over the entire study area or, more typically, within some portion of that area (Fig. 1). When sampling from a portion of the study area, choice of sampling units (e.g. plot, quadrat, line transect) should be based on some form of random selection, that is, each unit should have a known, nonzero probability of selection so that inferences can be extended to the entire set of units or study area. Within those selected units, observed counts can be unadjusted (e.g. fixed-radius point counts) or adjusted for incomplete detection of individuals. In the latter case, there are two forms of adjustment: ad hoc methods (e.g. Emlen line transects; Emlen 1971, 1977) and techniques whose adjustments are based on statistical theory (e.g. double-observer method, Nichols et al. 2000; distance sampling, Rosenstock et al. 2002).

**ALTERNATIVES TO UNADJUSTED COUNTS**

Because of the great potential for unadjusted counts to be confounded by factors affecting detection probabilities of individuals, one should consider alternative techniques that, when properly applied, provide unbiased estimates of abundance or density. However, even with these latter techniques, their key assumptions should be rigorously evaluated. Otherwise, one may be left with an expensive index estimate. When possible, a pilot study should precede a full study to assess feasibility of proposed counting methods, and if those methods are determined to be feasible, provide initial estimates for calculating number of samples needed to achieve precise estimates of abundance or density.

**DOUBLE SAMPLING**

Double sampling (Cochran 1977) is an approach in which relatively cheap and easy incomplete counts (e.g. point counts) are conducted within a random sample of sampling units followed by more expensive and difficult complete counts within a random subsample of those selected units. Results from this subsam-
ple then are used as a correction factor for adjusting incomplete counts; that correction is typically in the form of a ratio estimator (Cochran 1977). Assuming a random sample and subsample of units, the critical assumption of double sampling is that all individuals are counted within the subsample of units. Although this method also assumes a linear relationship between incomplete and complete counts within the subsample of surveyed units, simulations based on a ratio estimator indicated relatively minor levels of bias with less than perfect correlations ($r < 1.0$; on average 90% of 95% confidence intervals contained true value; Thompson 2002). In addition, the same method for obtaining incomplete counts should be applied to all randomly sampled units (Bart and Earnst 2002).

Double sampling has mostly been applied within an incomplete aerial survey and complete ground count context for estimating wildlife populations (Jolly 1969, Eberhardt et al. 1979, Martin et al. 1979, Handel and Gill 1992). Bart and Earnst (2002) used a double-sampling scheme to estimate number of shorebird nests, which was extrapolated to numbers of pairs, in tundra habitat in Alaska. They provided cost functions and sample-size formulas for allocating effort and expenses between the two samples (also see formulas presented by Eberhardt and Simmons 1987).

Bart and Earnst (2002) discussed several advantages of double sampling compared to unadjusted counts. If key assumptions are met, the advantages of double sampling mainly relate to its low bias and therefore reliability of information it may provide for monitoring trends and evaluating bird–habitat relationships. Further, supplemental information, such as nest success, can be gathered if the parameter of interest is number of nests. There also is flexibility in choice of which unadjusted counting method to employ, as long as the same method is used for all randomly sampled units.

Double sampling offers a potentially useful alternative to unadjusted counts when all individuals or items of interest can be detected within a sampling unit. This may be the case for fixed and readily detectable features in open habitats (e.g., shorebird nests in the tundra), but is more problematic when applied to mobile populations in less open habitats (birds in forests). However, even nests may be difficult to locate in areas with heavy vegetational cover and structure. As with any method for population estimation, key assumptions underlying double sampling (e.g., complete counts in a subsample of units) should be rigorously evaluated before incorporating that approach into a population study.

**Double-Observable Approach**

Nichols et al. (2000) adapted an aerial survey method suggested by Cook and Jacobson (1979) to adjust point counts of birds for those individuals present but not detected. This double-observer approach is based on a primary observer who relays all birds he or she detects to a secondary observer. The secondary observer records those birds and any additional ones he or she detects that the primary observer missed. The two observers alternate in the primary and secondary roles. Detection rates for the two observers are calculated by species or species group (species with similar detection probabilities) and are combined with number of birds detected across sampled points to adjust observed counts.

The double-observer approach assumes that observer counts are independent and the probability of each observer recording a bird is the same regardless of primary or secondary role (Pollock and Kendall 1987). Nichols et al. (2000) discussed potential ways to meet the independence assumption, including positioning the secondary observer behind the primary observer to minimize the chance of the latter keying in on birds detected by the former. Further, the act of recording data by the secondary observer may inhibit his or her ability to detect birds, especially for abundant species. To address that problem, Nichols et al. (2000) suggested that either both observers record data or a third observer be added as the sole data recorder. Differences in detection probabilities related to distance from a given point may be addressed through use of fixed-radius plots, which would be required to extrapolate counts to a larger area anyway. Including observer, species, species group, and other relevant covariates in models of detection probabilities also may help address variability in detection probabilities. As mentioned by these authors, the double-observer approach will not work well
for species with low detection probabilities or those that occur in low numbers or both.

In addition to the double-observer approach, Nichols et al. (2000) mentioned double surveys as an alternative to traditional point counts. Double surveys require two or more independent observers to map locations of detected birds at each survey point or at each point along a surveyed transect. Then, those mapped locations are compared to generate an abundance estimate based on a mark–recapture estimator. Here, “mark–recapture” refers to comparing mapped locations of one observer (“mark”) to those mapped by other observers and treating those in common as “recaptures.” A modified Lincoln-Petersen estimator (Chapman 1951) is typically used in the case of two observers (e.g. Magnusson et al. 1978, Anthony et al. 1999). Other mark–recapture models (Otis et al. 1978) that account for factors such as variable detection probabilities among birds can be used with more observers. In this case, at least four observers should be used when feasible, although that will have to be tempered by the ability to satisfy the independence assumption. Moreover, capture–recapture models discussed by Huggins (1989, 1991) allow inclusion of covariates that may affect detection probabilities; those and other capture–recapture models are available in program MARK (White and Burnham 1999). Manly et al. (1996) offered an alternative double-count method that used logistic regression models and relevant variables affecting detection probabilities to estimate abundance. These double-survey alternatives warrant further investigation for use in counting birds.

**Distance Sampling**

Distance sampling methods (Burnham et al. 1980, Buckland et al. 1993) use distances or distance categories from a point or line to a detected individual as a basis for a detectability correction via a detection function. Recorded distances or distance categories then are modeled using various model forms of detection functions; the best-fitting model is used to generate density estimates.

Rosenstock et al. (2002) described distance sampling within the context of bird counts, especially with respect to the key assumptions underlying this collection of methods (line and point transects [variable circular-plots]). That is, distance sampling methods assume that all birds on the transect line or point are detected, birds do not move in response to the observer prior to detection, and distances or distance categories are accurately recorded. Of those three, the first two are more difficult to satisfy. Although there are approaches for accounting for incomplete detections on a line or point (Alpizar-Jara and Pollock 1997, Borchers et al. 1998) and for responsive movements (Palka and Hammond 2001), those were developed mainly for marine or aerial surveys. Thus, they may be difficult to apply to bird counts. Nonetheless, Rosenstock et al. (2002) described various approaches that may be helpful for meeting those assumptions. More generally, they emphasized the need for accounting for counting bias and for presenting valid estimates of precision with abundance or density estimates.

Studies by DeSante (1981, 1986) often are cited as evidence that distance sampling techniques, and variable circular-plots (Ramsey and Scott 1979) in particular, do not perform well in field tests. Buckland et al. (1993) discussed design problems in those studies that would result in poor estimates of density; their comments also are pertinent to recent field tests by Tarvin et al. (1998) and Jones et al. (2000). None of those studies used a known population of birds in an area that was large enough to have adequate numbers of detections to properly fit distance data. Intensive nest and territory searches are usually subject to error, except perhaps when applied to small areas. The only published study to date that truly used a known density of birds to evaluate distance sampling was Nelson and Fancy (1999). They released 41 radio-marked birds into unoccupied forest habitat and reported that this approach provided unbiased density estimates (even with small sample sizes; see modifications for less abundant species in Ramsey et al. 1987 and Fancy 1997).

Distance sampling potentially offers a rigorous approach to obtaining valid density estimates. However, key assumptions of that approach should be evaluated, especially those of complete detection of birds on a line or point and no responsive movement by birds prior to detection.
DISCUSSION

A primary goal of a population study is to obtain population estimates with low (or no) bias and high precision in a cost-effective and logistically feasible manner. Common methods of surveying birds may fail to meet that goal because of either lack of adjustment or lack of proper adjustment of results to account for birds present but not detected (Nichols 2000, Bart and Earnst 2002, Rosenstock et al. 2002). Bias associated with unadjusted counts is exacerbated if one chooses survey locations or sampling units based on nonrandom selection (e.g., counts conducted along roads or trails). Further, note that on average, a larger number of samples or counts will increase precision, but it will not decrease bias in large populations (Yates 1981).

Standardization of counting protocols often is suggested as a remedy for relieving effects of bias inherent in unadjusted counts. However, standardization will not remove counting biases, although it certainly can maximize detection rates and improve precision. That is, the vast number of factors potentially affecting detection probabilities (see Thompson et al. 1998 and Rosenstock et al. 2002 and references therein), and therefore bias, in unadjusted counts are far too complex and variable in field studies to control for via standardization. Moreover, some factors cannot be standardized, such as bird density and breeding status (Bart and Schoultz 1984).

There appears to be a common misconception among biologists (and ornithologists specifically) that count data can lack statistical rigor if one is only interested in trends. If counts are biased or imprecise, placing them in a time series does not magically transform them into something meaningful. However, unadjusted counts may be useful if they are reasonably precise and their bias is small relative to the magnitude of population change. Thus, for large changes in an initially abundant population, unadjusted counts may suffice. However, even in this case, it would typically take a large change before we took notice, that is, our actions would be reactive rather than proactive. Unadjusted counts will definitely not suffice for species of concern, that is, those already occurring in low or greatly reduced numbers.

Reliable count data are a necessity for valid conclusions. One should not adopt a counting technique simply because it is commonly used. Suitability of a technique is based on reliability of its results, not on majority vote. Bart and Earnst (2002), Nichols et al. (2000), and Rosenstock et al. (2002) suggested useful alternatives to unadjusted counts that, when appropriately applied, will provide reliable estimates of avian abundance and density. However, more emphasis should be placed on validating these and other counting methods by investigating possible violations of key assumptions within the context of a given study (e.g., using radio-telemetry to investigate possible response movements by birds in distance sampling) and by comparing observed population estimates with a known population of birds to serve as a benchmark (e.g., detection rates of a marked subpopulation treated as true population). In general, we should strive for a higher standard regarding quality of information obtained in wildlife population studies (Anderson 2001) because it is the resource that ultimately suffers if unreliable data are used in setting conservation priorities and making management decisions.

LITERATURE CITED


