Studies in Avian Biology No. 6:399–404, 1981.

OPTIMIZING SAMPLING FREQUENCY AND NUMBERS OF TRANSECTS AND STATIONS

CHARLES E. GATES[1]

ABSTRACT.—Five valid methods of calculating variances of average density are: (a) systematic sampling with multiple random starts; (b) systematic sampling with a single random start using either natural subunits or replication in time; (c) interpenetrating sampling; (d) direct estimation of $v(\hat{D})$; and (e) the jackknife method. Method (a) is "best," but highly impractical in many situations. Method (b) should prove very useful in those situations where the subunits are sufficiently long to provide reasonable density estimates from each subunit. Method (c) would appear useful in all situations with reasonable sample size. Methods (d) and (e) should prove useful where the subunits are so short that the individual densities are essentially meaningless. These methods are applicable for any method of determining density.

To ascertain the total length of transect needed to achieve a desired coefficient of variability, calculate $L_1 = (\mathrm{cv}_0(\hat{D}))^2 L_0/(\mathrm{cv}_1(\hat{D}))^2$ where $\mathrm{cv}_0(\cdot)$ and $\mathrm{cv}_1(\cdot)$ are the observed (in a preliminary survey) and desired c.v.'s, respectively, with LT lengths L_0 and L_1.

In optimizing the LTs with subunits (or stations) and multiple sampling dates, the larger the variance component associated with a particular source of variation the greater the number of levels of that factor required (for fixed sample size), ignoring costs. If costs are considered generalization is more difficult. Obviously, if it is much cheaper to take an additional station than to get to the transect, the effect is to tend to drive the solution to more stations per transect at the expense of transects.

The purpose of this paper is to discuss the design of sample surveys in line transect and related sampling methods. To set the stage I shall define briefly the line transect and related sampling methods, following the standardized terminology suggested by Eberhardt (1978). The *line transect* (LT) is a basic sampling method wherein an observer walks a randomly located straight line, observing the target species, whether song birds, ruffed grouse, deer, duck nests, plants or rocks. For convenience, I employ the terminology as if animals were the target species, even though the sampling method is more general. At a given sighting, the observer records one or more of the following statistics: right-angle (perpendicular) distance to the sighted individual(s), radial (sighting or flushing) distance to the sighted individual(s) and/or the sighting (flushing) angle. On the basis of these measurements and a number of assumptions (see Gates 1979), it is possible to estimate the total population in the sample area or, equivalently, the density of animals.

Closely related sampling methods include the *strip transect, line intercept* and *quadrat sampling*. A strip transect is similar to the LT except that all animals are counted within a predetermined width in which the observer is reasonably certain all animals have been seen; animals outside the strip are not counted. A quadrat is similar to the strip transect except that many small areas are censused rather than a small number of much larger strip transects. A line intercept is a line or a strip transect narrowed to the line itself. It is more commonly employed for plants and inanimate objects than for animals, although it could be used for dense populations of slow-moving animals, e.g., snails. Note that estimating densities by the line intercept and quadrat methods is considerably different from that by the line and strip transect methods. I will not discuss the former methods further and will not discuss estimation explicitly for any of the methods. I leave this discussion for others and note several recent LT reviews and announcements of general computer programs, e.g., Gates (1979) and Burnham et al. (1980).

The design of any experiment or survey is highly dependent on the variability exhibited by the variable under study. Thus computing a valid estimate of variance is a necessity. In the remainder of this paper I first discuss five ways of calculating the variance of the density estimates and consider approximations to reduce the coefficient of variation of estimated density, $s_{\hat{D}}/\hat{D}$, to a predetermined size. I then consider costs in conjunction with a more complex LT design consisting of a line with several stations or subunits, sampled over time. Data, possibly not densities, are available for each station-time period.

COMPUTING VARIANCE OF DENSITY

The principal difficulty with reducing variance of density estimates to manageable size is obtaining a LT of sufficient total length. A line or a strip transect must of necessity use a large amount of real estate, in order to minimize overlap and to assure sufficient length for estimation of the population density with precision. To achieve meaningful results for some species,

[1] Institute of Statistics, Texas A&M University, College Station, Texas 77843.

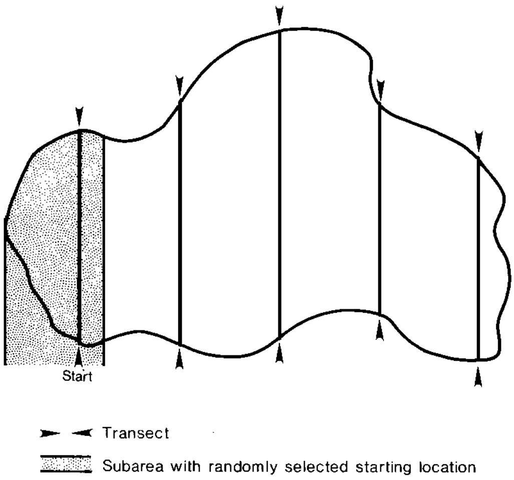

FIGURE 1. Systematic sampling with a single random start.

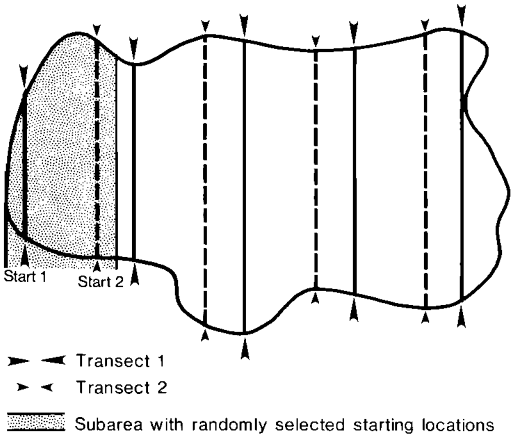

FIGURE 2. Systematic sampling with two random starts.

e.g., game birds, it may be necessary to have the line transect length 30, 40 or more km in length. If right-angle distances of only 100 meters to either side of the line are conservatively estimated (all of this is highly species-dependent, needless to say), then 6 km^2 would be utilized in a 30 km length.

Suppose the area being sampled is not sufficiently long (or wide) for 40 km of transect, e.g., an area 20 × 20 km. Then one could randomly locate in the sample area two transects of 20 km each (with restricted randomization such that there was no overlap). To ensure both that the entire area is representatively sampled and that there is no overlap, one could use *systematic sampling* (SS). For instance, one might select a random number between 1 and 10, say 5.2. This first selection determines directly the starting point; the second segment would start at 15.2 km and would be parallel to the first segment. If the SS were to be replicated in the true sense of the word, *two* random starting points would be required, say 5.2 and 7.3; thus the second portions of the transects would begin at 15.2 and 17.3 km from the base. The two techniques are called, respectively, SS with a single random start and SS with multiple random starts (Sukhatme 1954, Cochran 1977) (see Figs. 1 and 2).

Prior to discussing potential improvements in the design of a survey, a reasonably good estimate of variability of density estimate is required. Thus the estimation of variance must be discussed, which is related to the concept of replication. To the sampling purist, SS with at least two random starts would be required to have valid replication and thus valid variance estimation. My first reaction was in sympathy with this viewpoint, but on further reflection I moved away from that viewpoint. Some individuals would like to subsample without limit, dividing one large sample into more and more pieces, thus giving a large number of degrees of freedom for estimation of the variance. However, in using a line or strip transect, this "infinite" subdivision is not practical; if many subtransects were formed then most would have 0 animals sighted with a scattering of 1's, a very few 2's and so on. Such a situation would be totally impractical. To estimate density with any precision, large subtransects would have to be used. Natural subdivisions of the transects as shown in Figures 1 and 2 should be permissible. In fact, it may be necessary to clump adjacent subtransects to obtain a sufficient sample size for a reasonable estimate.

An objection of the sampling purists to using natural subdivisions or large fractions of single transects is that treatment of these subtransects as independent samples is incorrect. The theoretical difficulty is that, since these subtransects are physically close to one another, there may be large positive correlations among the dependent subtransects so that variance is underestimated. However, the situation does not concern me greatly because, unless the transects are very long, a high degree of variability will be associated with the estimation of density from each subtransect. In fact, the high variability ordinarily will swamp the positive correlation between adjacent subtransects. There is one important qualification in the use of SS that must be kept in mind. One should be certain the distance between parallel subunits does not coincide with some topographic feature, e.g., ridges.

This could prove disastrous in either estimation of density or variances.

On the other hand, if the transects are too long it is likely that heterogeneous habitat will be encountered. This introduces the topic of stratified sampling. One should stratify within each habitat type markedly influencing the density of animals. Using optimum allocation, habitats with either a greater density of the target species or increased variability will require a larger sample than otherwise. (Greater density leads to increased variance, everything else being held constant.) Similarly, habitats with reduced density or variability will require shorter transect lengths than otherwise. If we fail to stratify, then the lengths of transects in each habitat will be approximately proportional to the total area of each habitat (stratum), which will undoubtedly not be optimum. Further discussion will focus on optimizing surveys within strata or where stratification is not required.

If it is not feasible to replicate over space, it may be feasible to replicate over time. This is commonly done in LT sampling. Obviously, the time frame must be short enough so that significant mortality or recruitment could not have occurred, and ambient conditions should be similar. If density has changed, then an average density will be estimated with increased variability due to change in density.

However, transects need not be partitioned into either natural or artificial units to estimate variances of mean density. A legitimate sampling method for estimating sampling variances with one true replication is called *interpenetrating sampling* (Cochran 1977) and is closely related to the statistical jackknife method (to be described later). In interpenetrating sampling, the data are randomly sampled after collection. Suppose each sighting is randomly assigned to one of k subsamples. The density is then estimated from each subsample, where the number of observations will be a random variable. For b = 4 the LT length will be 25% of its former value. The variance is then determined from the densities of the individual groups, $\hat{D}_1, \hat{D}_2, \ldots, \hat{D}_k$, and is an unbiased estimate of $V(\hat{D})$ provided there is no correlation between the errors of measurement of any two sampling units in different groups. This condition would appear to be met in transect sampling. The disadvantage to the procedure is that if two individuals calculate the variance, even with the same number of subsamples, they will not obtain exactly the same answer. The method is not unique in that sense. Interpenetrating sampling is illustrated in Table 1. In the original population there were 40 sightings. Each sighting was randomly and independently assigned to one of four subsamples with the resulting subsample sizes, n, estimated intercepts, $\hat{f}(0)$, and densities shown in Table 1. The density, $\hat{D}$, was calculated assuming length of the line transect $L = 100$ km and distances recorded to the nearest meter.

TABLE 1
ILLUSTRATION OF INTERPENETRATING SAMPLING

Subsample	Number sightings, n	Intercept, $\hat{f}(0)$	$\hat{D}$
1	12	.301	72.2
2	12	.256	61.4
3	7	.177	24.8
4	9	.172	31.0
Total	40	.244	48.8

Another way (the "direct" method) of determining the variance of estimated density is to consider the general LT density estimator

$$\hat{D} = cn\hat{f}(0)$$

where c is the constant, $1/(2L)$. The variance of $\hat{D}$ may be written

$$V(\hat{D}) = c^2V[n\hat{f}(0)].$$

The expression in brackets is a product of variables. Using known information on the variance of a product of variables and that n and $\hat{f}(0)$ will be uncorrelated or very close to it yields

$$V(\hat{D}) = c^2[V(n)E^2\hat{f}(0) + V(\hat{f}(0))E^2(n)],$$

where $V(n)$ and $V(\hat{f}(0))$ are the variances of n and $\hat{f}(0)$, and $E(\hat{f}(0))$ and $E(n)$ the expected values of $\hat{f}(0)$ and n, respectively. If n is binomial, then $V(n) = NPQ$, $E(n) = NP$, where P is the probability of flushing an animal given that it is in the transect and $Q = 1 - P$. However, unless the animals truly flush independently of one another, it is unlikely that n will be binomially distributed. (It is more likely that n follows a negative binomial distribution.) Thus $V(n)$ and $V(\hat{f}(0))$ could be estimated empirically from natural subunits of a transect, although there seems to be no advantage in doing that over calculating the empirical variance of $\hat{D}$ from the $\hat{D}_i$ (as done in the interpenetrating sampling procedure).

Burnham et al. (1980) observe that for their recommended estimators, e.g., the Fourier series, the variance of $\hat{f}(0)$ is readily obtainable. Thus if one of those estimators is used, the only problem is in the calculation of $v(n)$. This quantity may always be calculated by empirical methods if natural subunits of a LT are available. If not, the binomial, the Poisson or negative binomial approximation to $v(n)$ would have to be used, depending on the user's best appraisal.

TABLE 2
ILLUSTRATION OF THE JACKKNIFE ESTIMATION OF D_J AND $v(D_J)$[a]

i	n_i	l_i	$n - i$	$L - l_i$	$\hat{D}_i$	$\hat{D}^{(i)}$
1	14	3	121	25	101.00	84.66
2	20	4	115	21	101.75	160.56
3	43	9	92	19	98.66	100.48
4	18	3	117	25	100.33	90.22
5	23	5	112	23	95.80	115.12
6	17	4	118	24	100.25	93.25

[a] Adapted from Burnham et al. (1980); $R = 6$, $L = 28$, $n = 135$, $D = 99.25$.

It is instructive to examine alternative methods of expressing the direct variance of $\hat{D}$ (replacing $E(n)$ and $E\{\hat{f}(0)\}$ by n and $\hat{f}(0)$, respectively):

$$v(\hat{D}) = \hat{D}^2\left[\frac{v(n)}{n^2} + \frac{v(\hat{f}(0))}{\hat{f}^2(0)}\right]$$

from which it follows that

$$cv(\hat{D}) = cv(\hat{n}) + cv(\hat{f}(0)),$$

where $cv(\cdot)$ and $v(\cdot)$ are the sample coefficient of variation and variance, respectively.

A special case of the direct method of calculating variance is to calculate the theoretical variance directly. For example, Gates et al. (1968) give

$$v(\hat{D}) = \frac{n}{(A\hat{P})^2}\left[\hat{Q} + \frac{n}{n - 2}\right]$$

where $\hat{P} = 2L/A\hat{\lambda}$, and A is the area of the study site. However, it is dangerous to use such variances, as they depend heavily on two assumptions—exponentiality of right angle sighting distances in this case—and on the strict independence of sightings. The failure of the assumptions will cause the estimated variance to underestimate the true variance by an unknown amount.

The fifth method for estimating variance of density is the jackknife method. The technique is illustrated by Burnham et al. (1980), whose Table 4 we modify and present here as Table 2. Basically, the method requires a series of natural subunits. The set of data from each subunit is omitted, one at a time, with the density estimated from the remaining data. These densities are called pseudovalues, $\hat{D}^{(i)}$, and are used to calculate the average density and ultimately $v(\hat{D})$:

$$\hat{D}^{(i)} = \frac{L\hat{D} - (L - l_i)\hat{D}_i}{l_i}$$

where l_i is the length of the ith subunit and $\hat{D}_i$ its density. Then

$$\hat{D}_J = \frac{1}{L}\sum_{i=1}^{R} l_i\hat{D}^{(i)}$$

and

$$v(\hat{D}_J) = \sum_{1}^{R} \frac{l_i(\hat{D}^{(i)} - \hat{D}_J)^2}{L(R - 1)},$$

where R is the number of subunits. For the data illustrated in Table 2, $\hat{D}_J = 107.85$ with $v(\hat{D}_J) = 130.60$. Thus 95% confidence intervals, using the t statistic with five degrees of freedom are 78.5 to 137.23. The chief disadvantage of this procedure is that computations are fairly heavy with a desk calculator. They are admirably adapted to the computer, however.

LENGTH OF LINE TRANSECT NEEDED

Given now that some legitimate estimate of sampling variance of density is computable, how can we improve our sampling in the next iteration? Gates et al. (1968) gave a procedure for estimating the length of line transect needed to reduce the ratio of $v(\hat{N})/\hat{N}$ to some predetermined value R for their parametric estimator. The difficulty with their expression is that it is highly dependent on the exponentiality of the right angle flushing distances.

A more general criterion would be to make the reasonable assumption that the product of LT length and the squares of the coefficients of

TABLE 3
MEAN SQUARE EXPECTATIONS FOR MULTIPLE STATIONS PER TRANSECT, SAMPLED AT VARIOUS TIME INTERVALS

Source of variation	Degrees of freedom	Mean square	Expected mean square
Transects	$t - 1$	M_t	$\sigma^2{}_e + s\sigma^2{}_{tw} + w\sigma^2{}_{s(t)} + ws\sigma^2{}_t$
Stations (T)	$t(s - 1)$	$M_{s(t)}$	$\sigma^2{}_e + w\sigma^2{}_{s(t)}$
Times	$w - 1$	M_w	$\sigma^2{}_e + s\sigma^2{}_{tw} + st\sigma^2{}_w$
Times × tran.	$(w - 1)(t - 1)$	M_{tw}	$\sigma^2{}_e + s\sigma^2{}_{tw}$
Residual	$t(s - 1)(w - 1)$	M_e	$\sigma^2{}_e$

variation (cv) of observed densities are proportional at different lengths:

$$L_0(cv_0(\hat{D}))^2 = L_1(cv_1(\hat{D}))^2$$

where $cv_0(\hat{D})$, L_0, $cv_1(\hat{D})$ and L_1 represent, respectively, the observed cv in a survey of a similar species in a similar habitat or small preliminary survey of length L_0 and the desired cv in the final survey with total length L_1. Solving for L_1, we have

$$L_1 = \frac{(cv_0(\hat{D}))^2 L_0}{(cv_1(\hat{D}))^2}.$$

This result is identical to that found by Burnham et al. (1980:35). Thus if a small survey is run with $cv = 0.3$ and $L = 3$ km, but a cv of 0.1 is desired, $L_1 = (0.3)^2 4/(0.1)^2 = 36$ km.

COST EFFECTIVE SAMPLING OF LTs WITH STATIONS

Next consider a more complex sampling plan wherein the observer has stations (stops or subunits) on the transect and may be interested in sampling on more than one occasion. How may he allocate his resources in some useful way? I shall make the assumption that the average of the variable being measured (not necessarily density) does not change markedly over time (if it does, then the problem degenerates to considering the optimal sampling within dates). Assume that the researcher has t transects, each with s stations (subunits) and samples on w occasions. The random model for the situation described is

$$y_{ijk} = \mu + t_i + s_{ij} + w_k + (tw)_{ik} + \epsilon_{ijk}$$

where y_{ijk} = observed value (e.g., density or calls per three minute time period), t_i = transect effect, s_{ij} = station (subunit) within transect effect, w_k = time effect, $(tw)_{ik}$ = transect by time interaction effect, ϵ_{ijk} = random residual.

The analysis of variance appropriate to this completely random model is shown in Table 3.

The mean square expectations do not provide a criterion per se. One possible criterion for improving the sampling procedure would be to minimize the variance of a transect mean. The variance of a transect mean, $V(\bar{T})$, is the expected mean square for the transect without the $\sigma^2{}_t$ term, divided by the number of observations per transect, viz., sw. For fixed product sw, the minimization of $V(\bar{T})$ depends on the relative sizes of estimates of $\sigma^2{}_{tw}$ and $\sigma^2{}_{s(t)}$ as the relative size of $\sigma^2{}_e$ is immaterial. If $\sigma^2{}_{tw}$ is much larger than $\sigma^2{}_{s(t)}$ then the transect should be sampled more often at the expense of sampling more stations. Conversely, if $\sigma^2{}_{s(t)}$ is much larger than $\sigma^2{}_{tw}$ then more stations should be sampled at the expense of repeated sampling. If those two variance components are about the same size, then $s = w$ approximately. However, this is not a good criterion, as the number of transects is not considered and the cost of sampling is ignored. (One could optimize t and s by considering $(V(\bar{W}))$, variance of a time mean, but then no information is given on w.) It is undoubtedly more expensive to sample additional times than to sample additional stations.

Two common concepts involving costs in sampling invoke two different alternatives: (a) minimize cost subject to fixed variance or (b) minimize variance subject to fixed cost. Gates et al. (1975), with a model similar to the ANOVA model shown above, suggested specifically minimizing the variance of the overall mean subject to fixed cost. Consider a cost function such as

$$C = tc_t + wtc_w + wtsc_s$$

where t, w and s are defined as above and c_t = cost of establishing and maintaining a transect, c_w = average cost of traveling to a transect, and c_s = cost per station once the observer reaches the transect. The formal function for minimizing the overall variance subject to fixed cost, e.g., is

$$V(\bar{y}_{...}) + \lambda(C - tc_t - wtc_w - wstc_s)$$

where λ is a Lagrangian multiplier (Lindgren 1962:216–227) and

$$V(\bar{y}_{...}) = \frac{\sigma^2{}_e}{wls} + \frac{\sigma^2{}_{tw}}{tw} + \frac{\sigma^2{}_w}{w} + \frac{\sigma^2{}_{s(t)}}{st} + \frac{\sigma^2{}_t}{t}.$$

The minimization of this function requires the simultaneous solution of four non-linear equations in four unknowns (obtained by differentiating the previous expression with respect to s, t, w and λ, respectively). We need not show these but simply note that the equations cannot be solved directly, due to their non-linear nature, but must be solved by iteration. The procedure assumes that the variance components are known and treats s, t, w and λ as variables.

Gates et al. (1975) used $c_s = 0.3775$, $c_w = 5.442$ and $C_t = 1.00$ in a Mourning Dove (*Zenaida macroura*) survey in Texas, and concluded on the basis of analyzing several variables that the optimal design would be a very large number of transects, 8–13 stations per transect and one sampling time. When the number of sampling times was constrained to four, the optimum numbers of transects and stations/transect were about 170 and 5, respectively (vs. the original 91 transects and 20 stations/transect). Eventually 135 randomly-located transects with 15 stations each were established.

Modifications of the above technique would permit optimizing the number of transects and stations at a single sampling time or optimizing the number of transects and times for one station per transect. In the above development, w (or s) would be replaced by one and the number of non-linear equations would be reduced to three. The solutions would be a simplified version of the more general case.

A novel use of the procedure outlined would be to optimize the number of subunits for a lengthy transect for future similar work. Currently, it is not clear whether to have a small number of subunits with relatively small variance each or a large number of subunits to give more degrees of freedom for confidence limits but relatively large variances.