BIOMETRICS 39, 965–976
December 1983

Unbiased Estimation in Line-Intercept Sampling

Lee Kaiser[1]
Department of Mathematics, University of Auckland,
Auckland, New Zealand

SUMMARY

A theory for unbiased estimation of the total of arbitrary particle characteristics in line-intercept sampling, for transects of fixed and of random length, is presented. This theory unifies present line-intercept sampling results. Examples are given and variance estimation is discussed.

1. Introduction and Literature Review

Line-intercept sampling (LIS) is a method of sampling particles in a region whereby, roughly, a particle is sampled if a chosen line segment, called a 'transect', intersects the particle. It has the advantage over 'quadrat sampling' in that there is no need to delineate the quadrats and determine which objects are in each quadrat. Examples of the economics of LIS versus quadrat sampling can be found in Canfield (1941), Bauer (1943), Warren and Olsen (1964), and Bailey (1970). The particles may represent plants, shrubs, tree crowns, nearly stationary animals, animal dens or signs, roads, logs, forest debris, particles on a microscope slide, particles in a plane section of a rock or metal sample, etc.

In early biological applications, sampling with a transect appears to have been a purposive-sampling technique for studying how vegetation varies with changing environment, with the transect running perpendicular to the zonation (Weaver and Clements, 1929). In the study of range vegetation, Canfield (1941) incorporated random placement of the transect and, by taking the proportion of the sampled transect intercepted by the vegetation, obtained an unbiased estimator of coverage—that is, the ratio of the area covered by the vegetation to the area of the region of interest. However, he did not prove the unbiasedness of this estimator. Canfield called this method the 'line-interception method'. He also discussed such design questions as how many lines of what length are required and whether or not the area of interest should be stratified. Bauer (1943) compared transect sampling to quadrat sampling in an area of dense chapparal vegetation and in a laboratory experiment. He concluded that '. . . transect sampling deserves much wider use . . .'. McIntyre (1953) investigated the possibility of using data on intercept lengths in order to estimate not only coverage, but also density—that is, the ratio of the number of particles to the area of the region of interest. He was able to do this for populations which consisted of particles that were all magnifications of a known shape. Lucas and Seber (1977) presented and proved the unbiasedness of estimators of particle density and coverage for arbitrarily shaped and located particles when the transect is randomly placed. Their estimator of coverage is the same as that of Canfield (1941). Eberhardt (1978) reviewed three transect methods for use in ecological sampling: LIS, and two methods, 'line-transect sampling' and 'strip-transect sampling', in which the particles are points and the probability of observing a particle is a function of its perpendicular distance

---

[1] Present address: Sterling-Winthrop Research Institute, Rensselaer, New York 12144, U.S.A.

Key words: Line-intercept sampling; Line-intersect sampling; Transect; Random-length transect; Unbiased estimation.

to the transect. He also gave an example of the use of LIS in the estimation of the number of den sites in a large prairie-dog town. De Vries (1979b) showed how LIS can be used to estimate the density of animals with 'elliptical flushing regions', and obtained a different estimator from that obtained by Burnham (1979) using line-transect sampling. Seber and Pemberton (1979) investigated the use of LIS for estimating aspects of plant cuticles in rumen and fecal samples on microscope slides.

Line-intercept sampling was introduced in forestry by Warren and Olsen (1964), who coined the term 'line-intersect sampling' which is used to describe the technique in the forestry literature, as a means to estimate rapidly the volume of logging waste produced during clearfelling operations. One thing that distinguishes LIS in forestry applications is that the particles in the region are generally assumed to be needles, usually to represent log axes. Van Wagner (1968) proved the unbiasedness of Warren and Olsen's estimator under certain assumptions, one of which is that the logs are placed and oriented uniformly at random in the area. Under similar assumptions, de Vries (1973) obtained unbiased estimators of the total of arbitrary log characteristics, and considered variances and variance estimation. Again under the assumption of random particle placement, de Vries and Van Eijnsbergen (1973) presented estimators of the total of arbitrary particle characteristics allowing arbitrary particle shapes, and compared these for the case in which the particles are circular arcs. De Vries (1979) gave a good review of LIS with an emphasis on forestry applications and made suggestions for its use in other fields. This paper contains further references on LIS in the forestry literature.

Related problems may be found in the areas of petrography, quantitative metallography, and stereology, on which there is a vast literature. For references see De Hoff and Rhines (1968) and Davy and Miles (1977). A major problem in these areas is that of estimating aggregate properties of particles, usually assumed opaque, in an opaque three-dimensional medium. The usual sampling scheme is to examine a random plane section of the medium and to measure relevant aspects of the intersected particles. However, rather than measure all the particles in the plane section, a second stage of sampling may be used at this point. In this stage the problem is identical to that in LIS. Transect sampling techniques have been proposed and these will be referenced in the following sections.

The purpose of this paper is to state the problem of LIS in a sufficiently general form and to present sampling schemes and estimators which, when specialized to particular problems, will yield the unbiased estimators obtained by other investigators in this field. This is done in §2 and §3 for transects of fixed length, and in §5 for transects of random length. Variance estimation is discussed in §4.

2. A Fixed-Transect-Length Sampling Scheme and Unbiased Estimators

Suppose that there is a planar region R of area A in which there are N fixed particles, $P_1, \ldots, P_N$, of any shape, not necessarily convex, but where each particle is assumed to be a connected set of points (Buck, 1965, p. 29). A fixed characteristic x_i and a variable y_i, which can depend on the transect sampled, are defined on P_i, $i = 1, \ldots, N$. We wish to estimate $\lambda_x = \sum x_i/A$. (All summations without limits indicated will refer to the sum from 1 to N). For example, x_i may be the area of P_i, and y_i the length of the intersection of P_i and the transect; or P_i may be a road of length x_i, and y_i the number of intersections which the transect makes with the road.

Let some arbitrary direction in the plane be denoted by $\theta = 0$. Particles will be sampled in the following way. A point is chosen uniformly at random in R; that is, if B is any subset of R, then the probability that B contains the point is proportional to the area of B. This point is then the midpoint of a transect of fixed length L which has direction θ generated according to an arbitrary distribution on $[0, \pi)$. The midpoint and θ are chosen independently. In most applications the transect distribution will be either degenerate in a particular direction or uniform on $[0, \pi)$; that is, $\theta \sim U(0, \pi)$. The particle P_i is said to be sampled, and x_i and y_i are measured, if the transect completely intersects P_i. In order to handle partial intercepts we follow McIntyre (1953) and Lucas and Seber (1977) in choosing one end of the transect at random and sampling partially intercepted particles on this end, ignoring partial intercepts on the other end.

Often, LIS results are derived by using a fixed transect and assuming that particles have been located and oriented on R uniformly at random. As this assumption usually will not hold in practice, it is desirable to see what results can be obtained by assuming, as is done in sampling theory, that the population is fixed and that randomness enters the problem only through the sampling scheme. Thus, in the present work, all probabilities and moments are calculated over the random placement of the transect for the fixed population.

Let $w_i(\theta)$ be the maximum perpendicular distance between lines in the direction $\theta \in [0, \pi)$ tangential to P_i, and define the indicator random variable t_i to be 1 if P_i is sampled and 0 otherwise.

Consider the particle P_i in Fig. 1. Given that the direction of the transect is θ, P_i is completely intercepted by the transect if the transect midpoint is within the unshaded region bounded by the broken lines. This region has area $Lw_i(\theta) - a_i$, where a_i is the area of P_i. If the midpoint of the transect is in either shaded region, then P_i is sampled with probability $\frac{1}{2}$. The shaded regions have total area $2a_i$. Since all the regions under discussion are entirely within R,

$$\begin{aligned} E(t_i \mid \theta) = \mathrm{pr}(t_i = 1 \mid \theta) &= \frac{Lw_i(\theta) - a_i}{A} + \tfrac{1}{2}\left(\frac{2a_i}{A}\right) \\ &= \frac{Lw_i(\theta)}{A}, \end{aligned} \tag{2.1}$$

and, taking the expectation over θ,

$$\begin{aligned} E(t_i) = \mathrm{pr}(t_i = 1) &= \frac{LE_\theta\{w_i(\theta)\}}{A} \\ &= \frac{Lc_i}{A}, \text{ say.} \end{aligned} \tag{2.2}$$

Thus, the probability that P_i is sampled is proportional to a measure of the size of P_i.

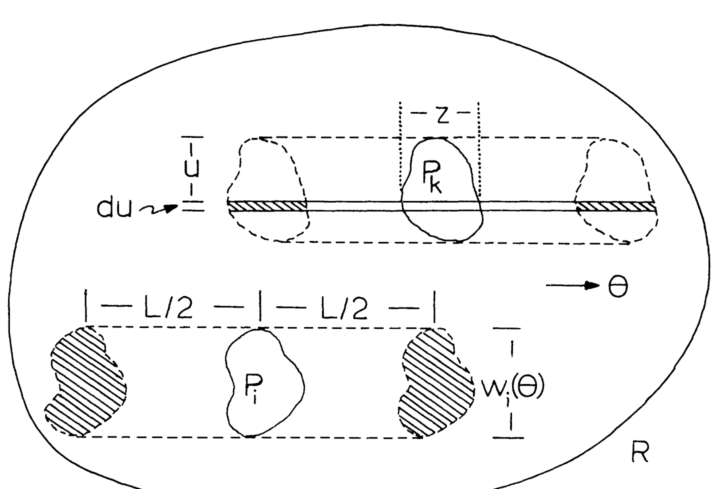

Figure 1. An illustration of the formula $\mathrm{pr}(t_i = 1 \mid \theta) = Lw_i(\theta)/A$, with $U \mid \theta, t_k = 1 \sim U(0, w_k(\theta))$.

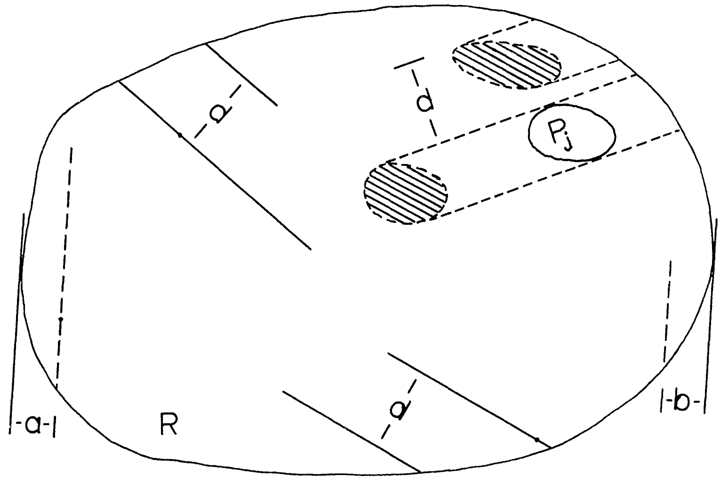

Figure 2. An illustration of a scheme for bringing into R any portion of the transect outside of R. Note that, for the dashed transect, the distances a and b between the transect and the tangent lines to R are such that $a + b = d$.

For particles within a distance $\frac{1}{2}L$ of the boundary of R, (2.1) will not hold for all θ. One way to make (2.1) hold when R is convex is to bring into R any portion of the transect which lies outside of R by continuing the transect a perpendicular distance d [$> \max\{w_i(\theta)\}$ so that no particle can be intercepted twice by the same transect] away from the portion of the transect already in R. Figure 2 illustrates such transects where the portion outside of R is brought back on the right-hand side of the portion in R. It does not matter whether the portion outside of R is brought in always on the right of the portion of the transect in R, always on the left, or on a side chosen at random. For the transect illustrated by the dashed line the sampler might rather bring the small portion of the transect in on the left and incur a small bias in the estimator. An argument identical to that given for P_i in Fig. 1, applied to P_j in Fig. 2, shows that (2.1) holds in this case also.

An important aspect of this sampling scheme is that given the direction, θ, of the transect and that $t_k = 1$, the perpendicular distance U, say, between the left-hand tangent line to P_k (when facing in the direction θ) and the transect has a uniform distribution on $(0, w_k(\theta))$. To show this, we have, using the definition of conditional probability,

$$\operatorname{pr}\{U \in (u, u + du) \mid \theta, t_k = 1\} = \frac{\operatorname{pr}\{U \in (u, u + du), t_k = 1 \mid \theta\}}{\operatorname{pr}(t_k = 1 \mid \theta)}. \tag{2.3}$$

Consider the particle P_k in Fig. 1. Let z be the length of the intercept of P_k and a transect with $U = u$. Now, given θ, U will be in $(u, u + du)$ and P_k will be sampled with probability $\frac{1}{2}$ if the transect midpoint is in either of the shaded strips, which have total area $2zdu$, or with probability 1 if the midpoint is in the unshaded strip which has area $(L - z)\, du$. Thus,

$$\begin{aligned}\operatorname{pr}\{U \in (u, u + du), t_k = 1 \mid \theta\} &= \frac{(L - z)\, du + \frac{1}{2}(2zdu)}{A} \\ &= \frac{Ldu}{A}.\end{aligned} \tag{2.4}$$

Combining (2.1) and (2.4) with (2.3) shows that

$$\operatorname{pr}\{U \in (u, u + du) \mid \theta, t_k = 1\} = \frac{du}{w_k(\theta)},$$

that is $U \mid \theta,\ t_k = 1 \sim \mathrm{U}(0,\ w_k(\theta))$. An identical argument holds for particles 'near' the boundary of R.

Now, two unbiased estimators of λ_x are

$$\hat{\lambda}_1 = \frac{1}{A} \sum \frac{t_i y_i}{\mathrm{E}(t_i y_i)} x_i \tag{2.5}$$

and

$$\hat{\lambda}_2 = \frac{1}{A} \sum \frac{t_i y_i}{\mathrm{E}(t_i y_i \mid \theta)} x_i. \tag{2.6}$$

Note that if the distribution of θ is degenerate then $\hat{\lambda}_1 = \hat{\lambda}_2$. These estimators are exemplified in §3.

3. Examples

By properly specifying the shape of the particles and properly choosing y_i and the distribution of θ in (2.5) and (2.6), all the available unbiased estimators in the LIS literature can be generated. Examples 1(a)–(c) deal with degenerate y_i and Examples 2(a)–(d) deal with nondegenerate y_i.

Example 1(*a*). Suppose that particles are of arbitrary shape and that $y_i = 1$ for all i. Then, from (2.1) and (2.2),

$$\hat{\lambda}_1 = \frac{1}{L} \sum \frac{t_i x_i}{c_i} \tag{3.1}$$

and

$$\hat{\lambda}_2 = \frac{1}{L} \sum \frac{t_i x_i}{w_i(\theta)} \tag{3.2}$$

are unbiased estimators of λ_x. The estimator $\hat{\lambda}_2$ was given by Lucas and Seber (1977) for $x_i = 1$, and by McDonald (1980) for arbitrary x_i. For a special case, $\hat{\lambda}_1$ has been used in the forestry literature: see Example 1(c).

Example 1(*b*). Suppose that the problem is as in Example 1(a) with $\theta \sim \mathrm{U}(0, \pi)$. Let c_i^* be the circumference of the convex hull of P_i, where the convex hull of P_i is the smallest convex set which contains P_i. In this case, since each particle is assumed connected, $c_i = c_i^*/\pi$ (Kendall and Moran, 1963, p. 58) and

$$\hat{\lambda}_1 = \frac{\pi}{L} \sum \frac{t_i x_i}{c_i^*}. \tag{3.3}$$

One may measure c_i^* by measuring the length of rope, say, which wraps around P_i.

Example 1(*c*). Suppose that the problem is as in Example 1(b) with P_i a line segment, called a 'needle', of length l_i. By considering the needle to have infinitesimal width, we see that $c_i^* = 2l_i$ and therefore

$$\hat{\lambda}_1 = \frac{\pi}{2L} \sum \frac{t_i x_i}{l_i}.$$

This is an estimator often used in forestry applications of LIS where the needles represent the axes of logs (de Vries, 1979). See also Example 2(c).

Before going to the examples with nondegenerate y_i it will be convenient to develop a formula for $\mathrm{E}(t_i y_i \mid \theta)$. Conditioning on t_i given θ,

$$\mathrm{E}(t_i y_i \mid \theta) = \mathrm{E}(t_i y_i \mid \theta, t_i = 0)\mathrm{pr}(t_i = 0 \mid \theta) + \mathrm{E}(t_i y_i \mid \theta, t_i = 1)\mathrm{pr}(t_i = 1 \mid \theta).$$

Given $t_i = 0$, $t_i y_i$ is zero, and the first term is zero. Given $t_i = 1$, $t_i y_i$ equals y_i, and, from (2.1),

$$\mathrm{E}(t_i y_i \mid \theta) = \mathrm{E}(y_i \mid \theta, t_i = 1)\frac{L w_i(\theta)}{A}. \tag{3.4}$$

Example 2(*a*). Suppose that particles are of arbitrary shape and that θ has an arbitrary distribution. Suppose we are interested in estimating $\sum a_i/A$, where a_i is the area of P_i. Let y_i equal the length of the intersection of P_i with the line in R which contains the transect. Denoting by U the distance defined in §2 and using the facts that $U \mid \theta, t_i = 1 \sim \mathrm{U}(0, w_i(\theta))$ and that y_i is a function of the transect only through U, Lucas and Seber (1977) showed that $\mathrm{E}(y_i \mid \theta, t_i = 1) = a_i/w_i(\theta)$. From (3.4), $\mathrm{E}(t_i y_i \mid \theta) = La_i/A = \mathrm{E}(t_i y_i)$ and

$$\hat{\lambda}_1 = \hat{\lambda}_2 = \frac{1}{L}\sum t_i y_i. \tag{3.5}$$

Thus, only the length of intersection needs to be measured to obtain an unbiased estimator of $\sum a_i/A$.

The first occurrence of (3.5) appears to have been in Rosiwal (1898), although it was presented without probability arguments as an approximation to $\sum a_i/A$. For a description of Rosiwal's work see Chayes (1956). The unbiasedness of (3.5) was proved by Lucas and Seber (1977).

Example 2(*b*). Suppose that particles in R are the projections onto R of three-dimensional objects resting on R. Suppose we wish to estimate $\sum v_i/A$, where v_i is the volume of the ith object. Let y_i be the area of the intersection of the ith object with the plane perpendicular to R which contains the transect. By an analysis very similar to that used by Lucas and Seber for the problem in Example 2(a) it is easy to see that $\mathrm{E}(y_i \mid \theta, t_i = 1) = v_i/w_i(\theta)$. From (3.4), $\mathrm{E}(t_i y_i \mid \theta) = Lv_i/A = \mathrm{E}(t_i y_i)$ and

$$\hat{\lambda}_1 = \hat{\lambda}_2 = \frac{1}{L}\sum t_i y_i.$$

Thus, only the area of intersection needs to be measured to obtain an unbiased estimator of $\sum v_i/A$.

An example similar to Example 2(b) can be given with y_i equal to the length of the boundary of the intersection of the ith object with the plane perpendicular to R containing the transect. In this case, $\mathrm{E}(t_i y_i \mid \theta) = Ls_i/A$, where s_i is the surface area of the ith object.

Example 2(*c*). Suppose that there are N logs above R and that the particles in R are needles which are the projections onto R of the axes of the logs. Suppose that we wish to estimate $\sum v_i/A$, where v_i is the volume of the ith log. As in Example 2(b) we could let y_i be the area of the intersection of the ith log with the plane perpendicular to R which contains the transect, but this intersection would be elliptical, the area of which would be difficult to measure. It is more convenient to let y_i be the area of the intersection of (i) the plane perpendicular to the ith log axis which intersects this axis directly above the point of intersection of the transect and the ith needle and (ii) the ith log.

Let l_i equal the length of the ith log, δ_i the angle of the ith log axis with respect to R, and γ_i the angle of the ith needle with respect to $\theta = 0$. Thus, the length of the ith needle is $l_i \cos \delta_i$.

Given θ and that the ith needle has been intercepted, the distance, r, from one end of the ith log axis along the axis to the sampling plane has a uniform distribution on $(0, l_i)$. Now, y_i

is a function of r and

$$\mathrm{E}(y_i \mid \theta, t_i = 1) = \frac{1}{l_i} \int_0^{l_i} y_i(r)\, dr = \frac{v_i}{l_i}.$$

Here, $w_i(\theta) = l_i \cos \delta_i \sin |\theta - \gamma_i|$ and, from (3.4),

$$\begin{aligned}\mathrm{E}(t_i y_i \mid \theta) &= L v_i w_i(\theta)/(A l_i) \\ &= (L v_i / A) \cos \delta_i \sin |\theta - \gamma_i|.\end{aligned}$$

Thus

$$\hat{\lambda}_2 = \frac{1}{L} \sum \frac{t_i y_i}{\cos \delta_i \sin |\theta - \gamma_i|}. \tag{3.6}$$

When $\theta \sim \mathrm{U}(0, \pi)$, then, as in Example 1(c), $\mathrm{E}\{w_i(\theta)\} = (2l_i/\pi)\cos \delta_i$ and

$$\hat{\lambda}_1 = \frac{\pi}{2L} \sum \frac{t_i y_i}{\cos \delta_i}. \tag{3.7}$$

For the case $\delta_i \equiv 0$, (3.7) is often used in forestry applications of LIS. Van Wagner (1968) pointed out that if (3.7) is used, with $\delta_i \equiv 0$, a small bias will be incurred if the logs are moderately tilted. For example, if all the δ_i are less than 25°, the relative bias of (3.7) with $\delta_i \equiv 0$ is less than 10%. Brown and Roussopoulos (1974) gave a table of factors by which to multiply $(\pi/2L) \sum t_i y_i$ in an effort to correct the bias due to object tilt when LIS is used to estimate the volume of forest debris.

Assume that all log axes are parallel with R. Van Wagner (1968) derived the estimator (3.7) using a fixed transect and assuming that logs are distributed uniformly at random on R with individual logs uniformly oriented. He investigated the bias (over placement of the population of logs for the fixed transect) when the logs tend to be oriented in a particular direction and suggested using three transects at 60° angles to each other in order to guarantee a small bias.

When the transect is located uniformly at random in R with $\theta \sim \mathrm{U}(0, \pi)$, we see that (3.7) is unbiased, over placement of the transect, for any population of logs, whether oriented or not. However, if the logs do tend to be oriented (3.7) will have a large variance. In this case it would probably be preferable to use (3.6) with θ degenerate at an angle perpendicular to the general orientation of the logs. How this scheme compares with the above scheme of three transects at 60° angles needs investigation.

The problem in Example 2(c) was stated with logs as objects. In general, the estimators of volume/area in Example 2(c) may be preferable to the estimator of volume/area in Example 2(b) when objects are such that the areas of plane sections in a particular direction relative to an object are easier to measure than the areas of plane sections in arbitrary directions.

Example 2(*d*). Suppose that particles in R are arbitrarily shaped, with the length of the boundary of P_i being l_i. Define y_i to be the number of intersections of the boundary of P_i with the line in R which contains the transect. Also, let $K_i(\theta)$ be the total projection of the boundary of P_i onto a line in R perpendicular to θ, where each point of the projection is counted as many times as there are points on the boundary which project onto it (Kendall and Moran, 1963, p. 59; de Vries, 1979).

Now, $\mathrm{E}(y_i \mid \theta, t_i = 1) = K_i(\theta)/w_i(\theta)$. When $\theta \sim \mathrm{U}(0, \pi)$, then (Kendall and Moran, 1963, p. 59) $\mathrm{E}_\theta\{K_i(\theta)\} = 2l_i/\pi$, and therefore

$$\hat{\lambda}_1 = \frac{\pi}{2L} \sum \frac{t_i y_i}{l_i} x_i. \tag{3.8}$$

This, except for differences in notation, is equation (8) of de Vries and Van Eijnsbergen (1973). Smith and Guttman (1953) gave (3.8) with $x_i = l_i$ with particular reference to the estimation of the surface area of objects in a three-dimensional medium.

De Vries (1979, p. 44) stated that 'We have no expression for [pr(t_i = 1)] unless the [particle's] shape can be exactly specified (e.g., a circular arc). Consequently we have no generally valid total-estimator [$\hat{\lambda}_x$] here. *It is advised not to use the practice of considering multiple intersections just as one intersection in LIS'* [his italics]. However, the practice of considering multiple intersections just as one intersection is simply the practice of using $y_i = 1$ for all i in (2.5).

From (2.2) we do have an expression for pr(t_i = 1) and when $\theta \sim U(0, \pi)$, from Example 1(b), pr(t_i = 1) = $Lc_i^*/(\pi A)$. Thus, in this case an alternative to (3.8) is (3.3). The choice between these two estimators depends on the ease of measurement of $y_i x_i/l_i$ as opposed to the measurement of x_i/c_i^*, and on the variances of the estimators. Intuitively, (3.8) will have a larger variance than (3.3) if $y_i x_i/l_i$, given $t_i = 1$, has a positive variance. Notice that if the particles are convex, then $c_i^* = l_i$, and (3.8) and (3.3) are equal since $y_i \equiv 2$.

The unbiased estimators of McIntyre (1953) can also be derived but they will not be presented here because they require the restrictive assumption that the particles are magnifications of a known shape.

4. Variance and Variance Estimation

By the formula for the variance of a linear combination of random variables, and taking $\hat{\lambda}_1$ from (2.5), we have

$$\operatorname{var}(\hat{\lambda}_1) = \frac{1}{A^2}\sum \frac{\operatorname{var}(t_i y_i)}{\{E(t_i y_i)\}^2} + \frac{1}{A^2}\sum_{i \neq j} \frac{\operatorname{cov}(t_i y_i, t_j y_j)}{E(t_i y_i)E(t_j y_j)}. \tag{4.1}$$

The variance terms are tractable, being independent of the locations of $P_1, \ldots, P_N$ on R. The difficulty is that cov($t_i y_i$, $t_j y_j$) depends on the relative locations of P_i and P_j. Even in simple examples the variance formula remains intractable and has proved to be of no help in either the estimation of variance or the design of an LIS survey. Similar comments apply to the variance of $\hat{\lambda}_2$, from (2.6).

For the case of a population of needles located and oriented uniformly at random in R, de Vries (1973) obtained a variance formula (over placement of the needles for a fixed transect) for the estimator in Example 1(c), and showed how this formula might be used for the estimation of variance from single-line data and in the design of an LIS survey, giving due caution on its use with a fixed population. This variance formula is essentially (4.1) with the covariance terms set to zero. It is planned, in a further paper, to give conditions on a population under which one might expect to be able to ignore covariance terms and obtain a reasonable approximation to the true variance (over placement of the transect for the fixed population) in order to aid in the design of general LIS surveys.

In order to be able to estimate the variance of the estimator of λ_x, it is desirable to sample $m > 1$ transects of length L, chosen independently according to the sampling scheme. The jth transect yields an estimate, $\hat{\lambda}_j$, of λ_x. As these $\hat{\lambda}_j$ are independent and identically distributed, they are probably best pooled by taking their arithmetic mean. The sample variance of the $\hat{\lambda}_j$ can be used to estimate the variance of the mean. The Central Limit Theorem applies here, so valid confidence intervals may be obtained for large m. For moderate m it is probably reasonable to use the usual confidence intervals based on the t distribution. In most practical situations, to achieve reasonable values of the coefficient of variation of the estimator, multiple transects will be called for. Pickford and Hazard (1978) give an illustration of this point even for very 'homogeneous' populations.

5. Unbiased Estimation with Random Transect Length

It will often be convenient to sample a transect which runs entirely across the area of interest. Seber (1979) suggested one such sampling scheme in which the estimators (3.2) and (3.5) are biased, and suggested using the jackknife to reduce the bias and provide an estimate of variance. In this section we present a sampling scheme with the transect running entirely across R, for which the estimators (2.5) and (2.6), with $\mathrm{E}(t_i y_i)$ and $\mathrm{E}(t_i y_i \mid \theta)$ evaluated as if the sampling scheme of §2 with fixed L were being used, are unbiased.

This sampling scheme is as follows. Choose a point uniformly at random in R and run a transect through this point in the direction θ generated independently with arbitrary distribution. The transect will run entirely across R, and its length, L, will therefore be a random variable. This kind of sampling scheme was considered by Miles and Davy (1976) for problems in stereology. However, the scheme to be described here is not contained in the work of Miles and Davy since they restricted themselves to taking measurements in the sample line and so did not consider estimators like those in Examples 1(a) and 2(c).

Let $W(\theta)$ be the maximum perpendicular distance between lines tangent to R in the direction θ. A transect may be uniquely indentified by its direction θ and by its perpendicular distance, q, from the left-hand (when facing in the direction θ) tangent line to R in the direction θ; see Fig. 3. Thus, the set of all transects is given by $\{(\theta, q)\colon 0 \leqslant \theta < \pi, 0 < q < W(\theta)\} = D$, say. What is the probability distribution on D generated by the above sampling scheme? Let $L(\theta, q)$ denote the length of the transect with (direction, distance) pair equal to (θ, q). For a given θ the transect will have q in the interval $(q, q + dq)$ if the random point chosen in R is in a strip of area $L(\theta, q)\, dq$ about the (θ, q) transect. Thus, the density function of q given θ is $f(q \mid \theta) = L(\theta, q)/A$ for $0 < q < W(\theta)$ and is zero otherwise.

This gives an alternative method of generating a transect according to the present sampling scheme: that is, generate $\theta \in [0, \pi)$ and then generate a quantity q with $0 < q < W(\theta)$ from the density $L(\theta, q)/A$. Although it is easier to generate a transect for use in practice according to the original description of this sampling scheme, the alternative characterization is easier to use to prove the unbiasedness of the estimators.

In order to distinguish between them, moments of random variables with respect to the

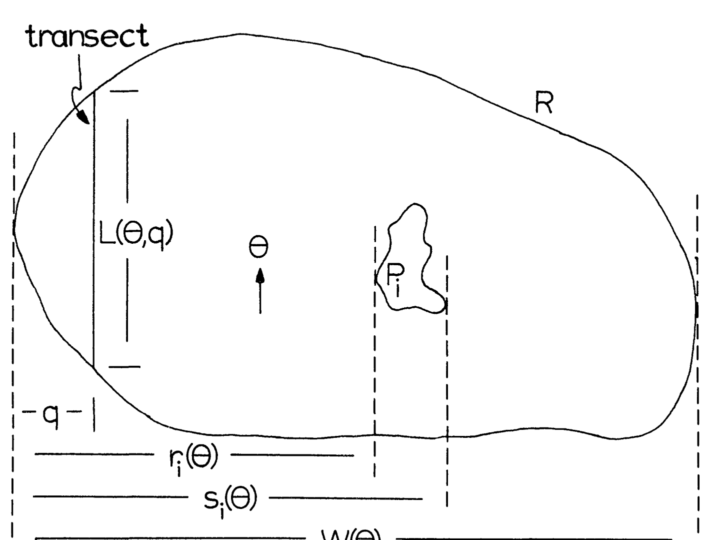

Figure 3. An illustration of some quantities needed in the proof of the unbiasedness of the estimators in the random-length sampling scheme.

sampling scheme of this section will be starred and moments with respect to the sampling scheme of §2 will be unstarred.

The statistic $\hat{\lambda}_2$ of (2.6) is unbiased if

$$\mathrm{E}^*\left\{\frac{t_i y_i}{\mathrm{E}(t_i y_i \mid \theta)} \mid \theta\right\} = 1. \tag{5.1}$$

Make explicit the dependence of y_i and t_i on the sampled transect by writing them as $y_i(\theta, q)$ and $t_i(\theta, q)$. Define $r_i(\theta)$ to be the perpendicular distance between the left-hand tangent to R in the direction θ and the left-hand tangent to P_i in the direction θ, and define $s_i(\theta)$ to be the perpendicular distance between the left-hand tangent to R in the direction θ and the right-hand tangent to P_i in the direction θ; see Fig. 3. From (3.4),

$$\mathrm{E}(t_i y_i \mid \theta) = L w_i(\theta) \mathrm{E}(y_i \mid \theta, t_i = 1)/A.$$

By the uniform distribution, given θ and $t_i = 1$, of the distance U defined in §2 when the fixed-length sampling scheme is used,

$$\begin{aligned} w_i(\theta)\mathrm{E}(y_i \mid \theta, t_i = 1) &= \int_0^{w_i(\theta)} y_i\{\theta, r_i(\theta) + u\}\, du \\ &= I(\theta), \text{ say.} \end{aligned} \tag{5.2}$$

Thus, (5.1) is true if $\mathrm{E}^*(At_i y_i/L \mid \theta) = I(\theta)$ but

$$\begin{aligned} \mathrm{E}^*\left(\frac{At_i y_i}{L} \mid \theta\right) &= \int_0^{W(\theta)} \left\{\frac{At_i(\theta, q) y_i(\theta, q)}{L(\theta, q)}\right\}\left\{\frac{L(\theta, q)}{A}\right\} dq \\ &= \int_{\{q: t_i(\theta, q) = 1\}} y_i(\theta, q) dq \\ &= \int_{r_i(\theta)}^{s_i(\theta)} y_i(\theta, q)\, dq = I(\theta), \end{aligned} \tag{5.3}$$

and $\hat{\lambda}_2$ is unbiased.

That $\mathrm{E}^*(\hat{\lambda}_1)$ of (2.5) equals $\sum x_i/A$ follows by taking expectations over θ of (5.2) and (5.3).

In all of the Examples in §3, y_i is a function of the transect only through its orientation, θ, and the distance U defined at the end of §2. Thus, the estimators presented in §3 are unbiased under either the sampling scheme of §2 or that of this section.

To estimate the variance of the estimator of λ_x, it is necessary to sample independently $m > 1$ transects according to the sampling scheme. As the m estimators obtained from these transects are independent and identically distributed it seems reasonable to pool them by taking their arithmetic mean and estimating the variance of the mean by using the sample variance among these m estimators.

The sampling scheme here differs from that of Seber (1979) in that Seber set $f(q \mid \theta) = 1/W(\theta)$ for $0 < q < W(\theta)$ and zero otherwise, while here the distribution of q given θ is 'weighted by the length of the transect'.

6. Conclusion

The examples in §3 illustrate that, for the derivation of the estimators (2.6) and (2.5) in a particular problem, for use with either the fixed- or random-transect-length sampling schemes, it is necessary to obtain expressions for $\mathrm{E}(y_i \mid \theta,\ t_i = 1)$ and $\mathrm{E}_\theta\{w_i(\theta)\mathrm{E}(y_i \mid \theta,\ t_i = 1)\}$, respectively.

The real problems in LIS are in the area of design. While it is possible to design an LIS survey based on a pilot sample of transects, it would be desirable to have approximations to the variances of (2.5) and (2.6) which might be used with prior knowledge of the population to help in the design of an LIS survey.

Acknowledgements

I wish to thank Professor George Seber for suggesting work on this problem and for helpful discussions. Also, I greatly appreciate the detailed comments made by the referees.

Résumé

On présente une théorie pour une estimation sans biais du nombre total d'individus de nature quelconque dans un échantillonnage sur un segment, pour des transects de longueur fixée on aléatoire. Cette théorie synthétise les résultats connus sue l'échantillonnage sur un segment. Des exemples sont donnés et la variance d'estimation est discutée.

References

Bailey, G. R. (1970). A simplified method of sampling logging residue. *Forestry Chronicle* **46,** 288–294.

Bauer, H. L. (1943). The statistical analysis of chaparral and other plant communities by means of transect samples. *Ecology* **24,** 45–60.

Brown, J. K. and Roussopoulos, P. J. (1974). Eliminating biases in the planar intersect method for estimating volumes of small fuels. *Forest Science* **20,** 350–356.

Buck, R. C. (1965). *Advanced Calculus.* New York: Wiley.

Burnham, K. P. (1979). A parametric generalization of the Hayne estimator for line transect sampling. *Biometrics* **35,** 587–595.

Canfield, R. H. (1941). Application of the line intercept method in sampling range vegetation. *Journal of Forestry* **39,** 388–394.

Chayes, F. (1956). *Petrographic Modal Analysis, an Elementary Statistical Appraisal.* New York: Wiley.

Davy, P. J. and Miles, R. E. (1977). Sampling theory for opaque spatial specimens. *Journal of the Royal Statistical Society, Series B* **39,** 56–65.

De Hoff, R. T. and Rhines, F. N. (1968). *Quantitative Microscopy.* New York: McGraw-Hill.

de Vries, P. G. (1973). A general theory on line intersect sampling with application to logging residue inventory. *Mededelingen Landbouwhogeschool Wageningen* **73**(11), 1–23.

de Vries, P. G. (1979a). Line intersect sampling: statistical theory, applications, and suggestions for extended use in ecological inventory. In *Sampling Biological Populations,* R. M. Cormack, G. P. Patil and D. S. Robson (eds), 1–70. Fairland, Maryland: International Co-operative Publishing House.

de Vries, P. G. (1979b). A generalization of the Hayne-type estimator as an application of line intercept sampling. *Biometrics* **35,** 743–748.

de Vries, P. G. and Van Eijnsbergen, A. C. (1973). Line intersect sampling over populations of arbitrarily shaped elements with special reference to forestry. *Mededelingen Landbouwhogeschool Wageningen* **73**(19), 1–21.

Eberhardt, L. L. (1978). Transect methods for population studies. *Journal of Wildlife Management* **42,** 1–31.

Kendall, M. G. and Moran, P. A. P. (1963). *Geometrical Probability.* London: Griffin.

Lucas, H. A. and Seber, G. A. F. (1977). Estimating coverage and particle density using the line intercept method. *Biometrika* **64,** 618–622.

McDonald, L. L. (1980). Line-intercept sampling for attributes other than coverage and density. *Journal of Wildlife Management* **44,** 530–533.

McIntyre, G. A. (1953). Estimation of plant density using line transects. *Journal of Ecology* **41,** 319–330.

Miles, R. E. and Davy, P. J. (1976). Precise and general conditions for the validity of a comprehensive set of stereological fundamental formulae. *Journal of Microscopy* **107,** 211–226.

Pickford, S. G. and Hazard, J. W. (1978). Simulation studies on line intersect sampling of forest residue. *Forest Science* **24,** 469–483.

Rosiwal, A. (1898). On geometric rock analysis: a simple surface measurement to determine the quantitative content of the mineral constituents of a stony aggregate. *Verhandlungen der Koenig-*

liches Kristliches Reichsanstalt Wien **5-6,** 143–175.

Seber, G. A. F. (1979). Transects of random length. In *Sampling Biological Populations*, R. M. Cormack, G. P. Patil and D. S. Robson (eds), 183–192. Fairland, Maryland: International Co-operative Publishing House.

Seber, G. A. F. and Pemberton, J. R. (1979). The line intercept method for studying plant cuticles from rumen and fecal samples. *Journal of Wildlife Management* **43,** 916–925.

Smith, C. S. and Guttman, L. (1953). Measurement of internal boundaries in three-dimensional structures by random sectioning. *Journal of Metals* **5,** 81–87.

Van Wagner, C. E. (1968). The line intersect method in forest fuel sampling. *Forest Science* **14,** 20–26.

Warren, W. G. and Olsen, P. F. (1964). A line intersect technique for assessing logging waste. *Forest Science* **13,** 267–276.

Weaver, J. E. and Clements, F. E. (1929). *Plant Ecology.* New York: McGraw-Hill.

Received July 1981; *revised April and November* 1982