

Unbiased Systematic Sampling Plans for the Line Intercept Method

SUSAN ANN BUTLER AND LYMAN L. MCDONALD

Abstract

Experimentors have been using cost-efficient systematically located transects in the line intercept method for some time with little support from mathematical statistics. In this paper, it is shown that for rectangular regions the usual line intercept estimators for cover, density, and other attributes are unbiased for certain systematic sampling plans. The estimators are approximately unbiased for "large" irregularly shaped study regions.

Consider line intercept sampling in a situation in which particles or items to be sampled are distributed across a study region. In general, the particles may be of any shape and size and may possess an arbitrary spatial distribution. For instance, the particles may be shrubs or patches of vegetation in a field, cells on a slide, or the projection of logs on the forest floor. It is often of interest to study certain quantitative characteristics of these particles. One might be interested in the number of particles per unit area, density, the proportion of the study region covered by particles, percentage cover, or such attributes as weight, size, yield, biomass, etc.

Consider a study region of area A and define the following parameters (usually unknown and to be estimated):

N = The total number of distinct particles in the study region.

$D = N/a =$ Density, the number of particles per unit area.

$a_i =$ The area covered by particle, $i, i=1,2,\dots,N$.

$C = \frac{\sum_{i=1}^N a_i}{A} =$ Percentage cover, the proportion of the study region covered by the particles.

$X_i =$ Any fixed, measurable attribute associated with the i^{th} particle, $i=1,2,\dots,N$.

$X = \frac{\sum_{i=1}^N x_i}{A} =$ The total quantity of the attribute per unit area.

Sample data available will include:

$K =$ The number of transects in the sample.

$L_j =$ The length of the j^{th} transect, $j=1,2,\dots,k$.

$L = \sum_{j=1}^k L_j =$ The total length of the transects.

$n =$ The number of distinct particles intersected.

$y_i =$ The length of the intersection of the i^{th} particle and a transect, $i=1,2,\dots,n$.

$w_i =$ The width of the i^{th} intersected particle relative to the transect, i.e., the distance between tangents of the particle that are parallel to the transect, $i=1,2,\dots,n$.

$x_i =$ The value of the attribute associated with the i^{th} intersected particle, $i=1,2,\dots,n$.

For illustration see Figure 1.

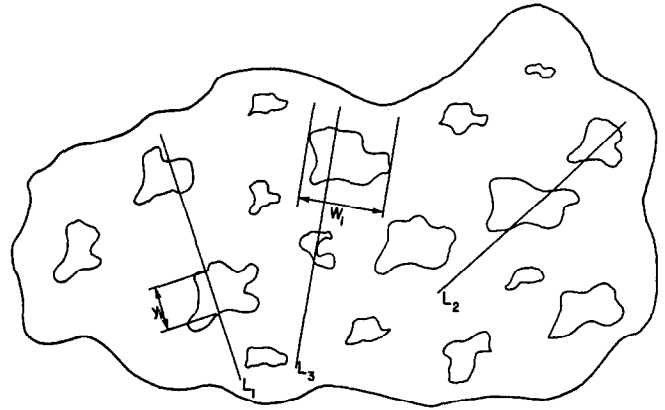


Fig. 1. Study region of area A with $N = 17$ distinct particles, $n = 6$ intersected particles, and $k = 3$ transects of equal length.

Most work to date in this area implies that data should be collected from a single randomly located transect or from a simple random sample of nonoverlapping transects of equal length. For example, Lucas and Seber (1977) derived unbiased estimators for density, D , and percentage cover, C , for any spatial distribution and a randomly located transect. DeVries (1979) derived an unbiased estimator for density for long narrow particles and a single randomly located transect, and Eberhardt (1978) derived an unbiased estimator of density using parallel but randomly located transects. McDonald (1980) showed that the Lucas and Seber estimators for density and percentage cover are unbiased for a simple random sample of unequal (random) length transects.

It is the purpose of this paper to show that the estimators for density, percentage cover, and other attributes, are unbiased for certain systematic sampling plans. In the interest of simple notation and illustration, a single study region is addressed. However, non-homogeneous or large study regions may be stratified into smaller strata. Rectangular strata are preferred for systematic line intercept sampling. Pooling of estimates for inference toward a larger area follow the standard methods e.g., Cochran (1977).

Systematic Sampling Plans and Estimation

Case I. Rectangular Study Regions.

First consider an ideal situation in which the study region is rectangular of size L by W . Transects may be randomly located along the baseline, one side of the study region, and extended perpendicularly from the baseline across the entire region. Eberhardt (1978) derives the unbiased estimator for density.

$$\hat{D} = \frac{\sum_{i=1}^n w_i^{-1}}{L} \quad (1)$$

where w_i is as defined previously and L is the length of a randomly located transect constructed perpendicularly to the baseline.

Instead of randomly locating the transects, systematically locate parallel transects. To do so, determine the distance, d , between

Authors are with the Department of Experimental Statistics, New Mexico State University, Las Cruces, 88003; and the Department of Statistics and Zoology, University of Wyoming, Laramie 82071.

Manuscript received June 29, 1981.

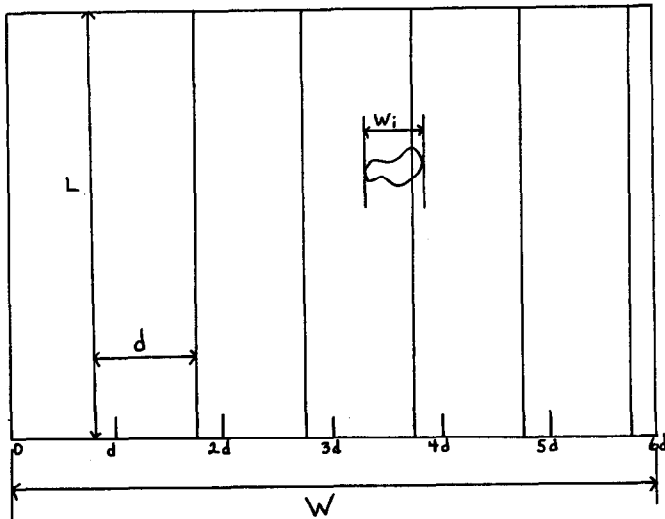


Fig. 2. Illustration of the systematic location of 6 transects in a rectangular study region. The first transect is randomly located.

transects such that d is large with respect to w_i , the width of the i^{th} particle. This will insure that no particles are intersected twice. It is also convenient to pick d such that $w/d = r$ is an integer. The first transect is then randomly located in the interval $[0, d]$. Because the transect is randomly located and every particle is in some interval, the probability that the i^{th} particle is intersected is

$$p_i = \frac{w_i}{d} \quad (2)$$

Subsequent transects are placed parallel to the first and d units apart. For illustration see Figure 2. Because the first transect is randomly located in the interval $[0, d]$ each of the other transect's starting points could be anywhere in their respective intervals. Thus, for any particular interval the probability of intersection is still as in Equation (2) and the estimator

$$\hat{D} = \frac{\sum_{i=1}^n w_i^{-1}}{\sum_{j=1}^k L_j} = \frac{\sum_{i=1}^n w_i^{-1}}{kL} \quad (3)$$

is unbiased for density. For further details of the proof, see the Appendix.

The rectangular study region naturally extends to the situation in which the region is not one large rectangle, but may be divided into r smaller rectangles. Each rectangle may be sampled with one random transect or a systematic sample of two or more transects. Density is estimated in each sub-rectangle and density for the entire study region may be estimated by the unbiased estimator.

$$\hat{D} = \frac{\sum_{j=1}^r A_j \hat{D}_j}{\sum_{j=1}^r A_j} \quad (4)$$

the weighted average of the individual estimates, where the weight A_j is the area of the j^{th} rectangle.

In a similar manner, the estimators for percentage cover,

$$\hat{C} = \frac{\sum_{i=1}^n y_i}{\sum_{j=1}^k L_j} \quad (5)$$

and the total of an attribute per unit area, X ,

$$\hat{X} = \frac{\sum_{i=1}^n x_i w_i^{-1}}{\sum_{j=1}^k L_j} \quad (6)$$

can be shown to be unbiased for a systematic sample of parallel transects in rectangular study regions.

Case II. Irregular Shaped Study Regions.

Devries (1979) derived the probability that a randomly tossed needle (a well-defined longest chord of a particle) intersects a straight line in a plane using a solution to a modified version of the Buffon needle problem. Using this probability, he derived the unbiased estimator for density,

$$\hat{D} = \frac{\Pi}{2L} \sum_{i=1}^n \ell_i^{-1} \quad (7)$$

where L is the length of the transect, and ℓ_i is the length of a well-defined longest chord of the intersected particles, $i=1,2,\dots,n$. The original Buffon needle problem consists of randomly dropping needles on a grid of equidistant parallel lines in a plane. Solomon (1978) showed that the probability, p_i , of the i^{th} randomly placed needle intersecting a line is

$$p_i = \frac{2 \ell_i}{\pi d} \quad (8)$$

where d is the distance between lines. In order to avoid a needle intersecting two lines and to avoid problems with the boundary it is necessary that $\ell_i \leq d$, the center of the needle is in the plane, and the lines are long relative to ℓ_i . The derivation of p_i is based on the fact that the distribution of θ_i , the acute angle between the closest line and the needle, is uniform on the interval $[0, \pi/2]$, the distribution of m_i , the perpendicular distance from the center of the needle to the closest line, is uniform on the interval $[0, d/2]$, and m_i and θ_i are stochastically independent, see Figure 3.

In many situations, the particles are not randomly distributed, i.e., reality is not equivalent to randomly dropping needles on a grid of equidistant lines. However, suppose that the particles are fixed, as they are, and that the grid is randomly dropped on the particles. The distribution of θ_i is still uniform on the interval $[0, \pi/2]$, the distribution of m_i is uniform on the interval $[0, d/2]$, and m_i and θ_i are still stochastically independent. Thus, the probability of a line intersecting the longest chord of a particle is as in Equation (8). A systematic sample constructed as follows will be equivalent to randomly dropping a grid on the study region.

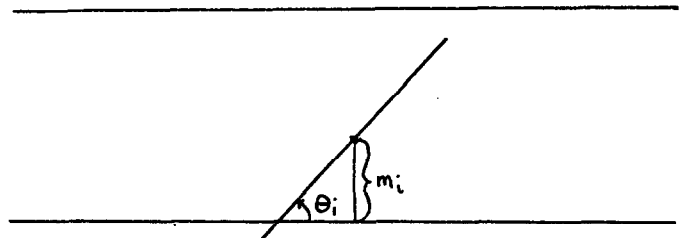


Fig. 3. The longest chord (needle) of a particle intersecting a transect. The value m_i is the perpendicular distance from the center of the chord to the closest transect. The angle θ_i is the acute angle between the chord and the closest transect.

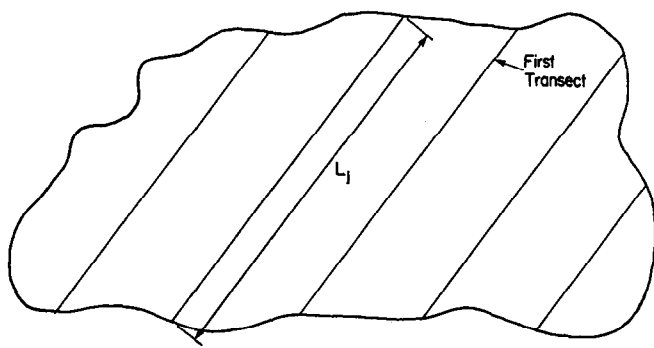


Fig. 4. A systematic sample taken by randomly locating a point, p and then randomly determining a direction, θ , for the first transect. Subsequent transects are located parallel to the first and equidistant apart.

A systematic sample requires that the first transect be randomly located and that subsequent transects be located parallel to the first and equidistant apart. Random location of the first transect may be accomplished by one of two procedures. In the first procedure, the first transect is located by randomly locating a point in the study region, randomly determining a direction at the point, and passing the transect through the point in the direction specified. In the second, determine a random point along a randomly located baseline and then extend the first transect perpendicularly to the baseline at that point. The transects may originate on one border and terminate on the opposite border of the study region. (See Figures 4 and 5.)

When a systematic sample is taken, the estimator for density

$$\hat{D} = \left[\frac{\pi}{2} \right] \frac{\sum_{i=1}^n e_i^{-1}}{\sum_{j=1}^k L_j} \quad (9)$$

is no longer unbiased. It has been suggested (Seber 1979) that two possible sources of bias are introduced. First, the estimator is biased by the curvature of the boundary of the study region. Thus, \hat{D} does not estimate the density for the exact study region. (See Figure 6 and the Appendix.) Seber also suggested that the random length of the transects introduces a second source of bias and discusses a Jack-Knifing technique to reduce the bias. However, McDonald (1980) showed that random lengths do not bias the estimator and hence the Jack-knifing procedure is not necessary.

If the curvature is slight, the lines are long relative to the size of the particles, the particles are inside the study region and the distance between transects is small relative to the size of the study region, the bias may be considered negligible for systematic sampling plans. The bias will be small if the combined area of the series of rectangles suggested in Figure 6 is close to the area of the study region, see the Appendix.

Lucas and Seber (1977) prove that for the case of a randomly located transect and any size, shape, orientation, or spatial distribution of particles, the estimators

$$\hat{D} = \frac{\sum_{i=1}^n w_i^{-1}}{L} \quad (10)$$

and

$$\hat{C} = \frac{\sum_{i=1}^n y_i}{L} \quad (11)$$

are unbiased for density and percentage cover respectively.

Instead of one randomly located transect or simple random sample of transects, assume that a systematic sample of k parallel unequal length transects extend from border to border across a fixed study region containing N particles of any distribution. The

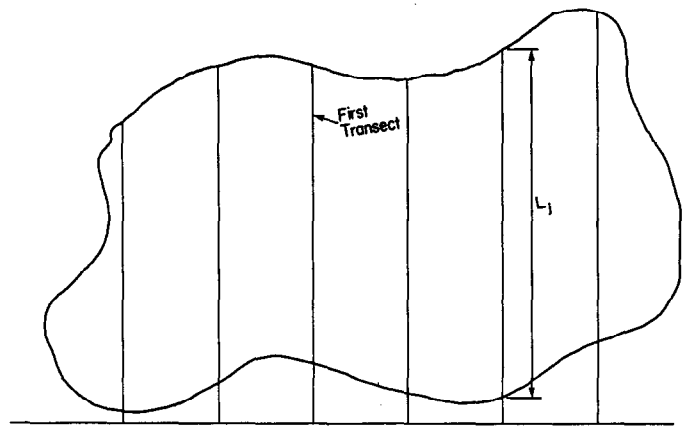


Fig. 5. A systematic sample taken by randomly locating a baseline, randomly locating the first transect along the baseline at point p , and placing subsequent transects parallel to the first and equidistant apart.

systematic sample may be constructed in one of the ways previously described. Assume the particles are "small" relative to the length of the transects and the distance between transects. (To guarantee that the probability of partial intersections or intersection of a particle by more than one transect is negligible.) As in the previous situation, the estimators of Equations (10) and (11) are potentially biased due to the curvature of the study region. Again, if the curvature is slight, the lines are long relative to the size of the particles, the particles are inside the study region, and the distance between transects is small relative to the study region, then the bias may be considered negligible for the systematic plans.

Discussion

Investigators are often reluctant to use a systematic sampling plan since spatial dependencies may be introduced into the data. Rather than avoiding systematic sampling or ignoring the dependencies, one might model the correlation between data points and use the correlation in the analysis of the data. This technique is beyond the scope of the present paper and is addressed in Butler (1981). The interested reader may also refer to Journel and Huijbregts (1978) where applications in mining and engineering are considered. Even though exact variance estimators for systematic sampling plans are not readily available, the variance estimates from the standard estimators of simple random sampling plans are recommended and are usually conservative (i.e. too large). See Mendenhall et al. (1971) and Cochran (1977). The primary problem in using systematic sampling arises if unknown cyclic patterns exist in the size, orientation, or spatial distribution of particles. The

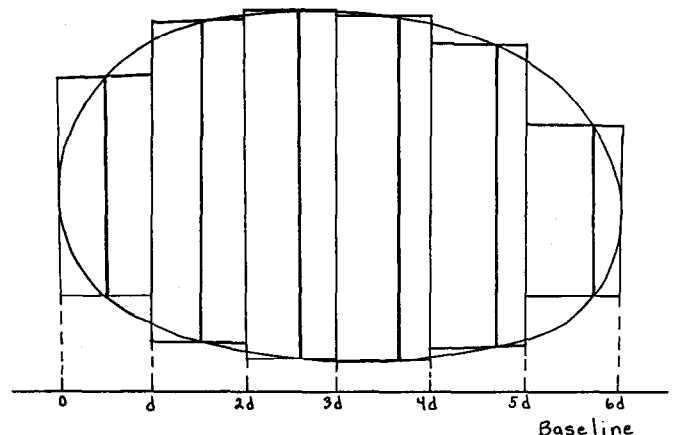


Fig. 6. Six transects systematically located in an elliptical shaped study region. If density, percent cover, or another attribute are estimated, the estimates are unbiased for the area enclosed by the rectangles and biased for the elliptical study region.

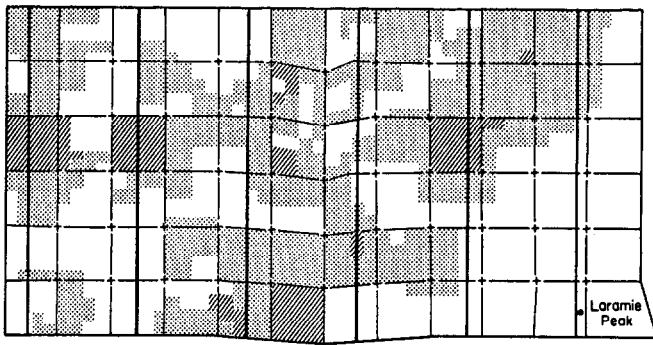


Fig. 7. Two townships in the Laramie Peak District of the Medicine Bow National Forest. The white land is federally owned, the dotted land is privately owned, and the striped land is state owned. Six transects are systematically located. The first is located .5 cm. from the left hand border and others are located parallel to the first and 2.6 cm. apart.

estimators are still unbiased but may have extremely large variances. Also the standard variance estimators (assuming simple random sampling) may not be conservative in this case. Hopefully such cyclic patterns will be recognized and used to an advantage in stratification of the study region into homogeneous subunits.

The main advantages of systematic sampling (if care is taken in the planning) are that:

- (1) It is often possible to obtain more information per unit cost than from simple random sampling, see for instance Mendenhall et al. (1971).
- (2) It is guaranteed that the sample will be spread uniformly over the study area (or subunit thereof).
- (3) The estimators used in line-intercept sampling are unbiased for rectangular study regions and nearly so for most "large" irregularly shaped study regions.

Systematic plans in line-intercept sampling have been used with little support from mathematical statistics. Experimentors have probably known that the resulting estimators are not only very cost-efficient but (at least intuitively) are also unbiased. For example, Eberhardt (1978) illustrates the use of Equation (1) with systematically located transects even though his derivation makes the assumption that a simple random sample is used. Granted that unbiasedness is not the only desirable property of estimators, its presence helps put the use of systematic line-intercept sampling on a firmer mathematical basis.

Illustration

Suppose that it is of interest to estimate the proportions of federal, state and private lands in the Laramie Peak District of the Medicine Bow National Forest. To illustrate the procedure, consider Figure 7 which consists of two of the many townships in the region. The different land owners are denoted as white for federal,

Table 1. The total length of intersections for each of 6 systematically located transects are given for each of the 3 land types. The distance between transects is $d = 2.6$ cm. and the first is located .5 cm. from the left hand border. The total length of intersection for all transects are given for each land type as well as the total number of particles intersected, n .

Transect	Land type			State (cm)
	L_j	Private (cm)	Federal (cm)	
1	7.6	3.5	2.9	1.2
2	7.6	0.6	5.6	1.4
3	7.6	3.2	3.7	0.7
4	7.6	3.9	3.1	0.6
5	7.6	2.7	3.6	1.3
6	7.6	2.9	4.7	0.0
Totals	49.6	16.8	23.6	5.2
		($n=16$)	($n=15$)	($n=5$)

dotted for private, and striped for state. Six transects are systematically located in the following manner. The base of the region is divided into six intervals each of width 2.6 cm. The first transect is randomly located in the first interval. Its starting point is randomly determined to be .5 cm from the left hand border. All measurements are made in centimeters and the results are tabulated in Table I. Percentage cover for each land type is estimated by

$$\hat{C} = \frac{\sum_{i=1}^n y_i}{6L}$$

Let \hat{C}_f , \hat{C}_p , \hat{C}_s denote the estimates for percentage cover of federal, private, and state lands respectively. The following estimates are computed.

$$\hat{C}_f = .52 \quad \hat{C}_p = .37 \quad \hat{C}_s = .11$$

Therefore it is estimated that 52% of the two townships is federally owned, 37% privately owned, and 11% owned by the state. The actual proportions for these two townships are 50% federally owned, 42% privately owned, and 8% owned by the state. Although no variance estimators exist for systematic sampling plans. The following variances are computed using the standard variance estimators for simple random sampling:

$$S_r^2 = .0183 \quad S_p^2 = .0233 \quad S_s^2 = .00496$$

Literature Cited

- Butler, S.A. 1981. Some aspects of line intercept sampling. Ph.D. thesis. Laramie, Wyo.
- Cochran, W.G. 1977. Sampling techniques. John Wiley and sons, Inc., New York.
- DeVries, P.G. 1979. Line intersect sampling—statistical theory, applications, and suggestions for extended use in ecological inventory. In: Sampling Biological Populations, Vol. 5 of Statistical Ecology, p. 1-70, R.M. Cormack, G.P. Patil and D.S. Robson (eds.), International Co-operative Publishing House, Fairland, Md.
- Eberhardt, L.L. 1978. Transect methods for population studies. J. Wildl. Manage. 42:1-31.
- Journel, A.G., and Ch. J. Huijbregts. 1978. Mining geostatistics, Academic Press, New York.
- Lucas, H.A., and G.F. Seber. 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. J. Wildl. Manage. 44:530-533.
- Mendenhall, W., L. Ott, and R.L. Schaeffer. 1971. Elementary survey sampling. Duxbury Press. Belmont, Calif.
- Seber, G.A.F. 1979. Transects of random length. In: Sampling Biological Populations, Vol. 5 of Statistical Ecology, p. 182-192, R.M. Cormack, G.P. Patil and D.S. Robson (eds.), international Co-operative Publishing House, Fairland, Md.
- Solomon, H. 1978. Geometric probability. Society for Industrial and Applied Mathematics. Penn.

Appendix

Eberhardt (1978) argues that the estimator for density,

$$\hat{D} = \frac{\sum_{i=1}^n w_i^{-1}}{\sum_{j=1}^k L_j} \quad (A.1)$$

is unbiased for a simple random sample of parallel transects constructed perpendicularly to one side of a rectangular study region. The estimator in Equation (A.1) is also unbiased if a systematic sample of parallel transects is constructed rather than a simple random sample.

Proof:

Because the first transect is randomly located in some interval, [0,d], and every particle is in some such interval, the probability that the i^{th} particle is intersected is $p_i = w_i/d$. Where w_i is the width of the particle and d is the distance between transects.

Let

$$Z_i = \begin{cases} 1.0 & \text{if the } i^{\text{th}} \text{ particle is intersected with} \\ & \text{probability } w_i/d. \\ 0.0 & \text{if the } i^{\text{th}} \text{ particle is not intersected} \\ & \text{probability } 1-w_i/d. \end{cases}$$

Then

$$\begin{aligned} E[\hat{D}] &= E \left[\frac{\sum_{i=1}^n w_i^{-1}}{\sum_{j=1}^k L_j} \right] \\ E[\hat{D}] &= E \left[\frac{\sum_{i=1}^N z_i w_i^{-1}}{\sum_{j=1}^k L_j} \right] \\ &= \frac{\sum_{i=1}^N E[z_i] w_i^{-1}}{\sum_{j=1}^k L_j} \\ &= \frac{\sum_{i=1}^N w_i w_i^{-1}}{\sum_{j=1}^k L_j d} \\ &= \frac{\sum_{i=1}^N 1}{A} = \frac{N}{A} = D \end{aligned}$$

If a systematic sample is taken over an irregularly shaped study region such that the transects run from border to border, the estimator for density

$$\hat{D} = \frac{\pi}{2} \left[\frac{\sum_{i=1}^n \ell_i^{-1}}{\sum_{j=1}^k L_j} \right] \quad (\text{A.2})$$

is no longer unbiased. (ℓ_i is the length of a well defined longest chord of the intersected particle.)

Proof:

Consider the expected value of \hat{D} given $L_j, j=1,2,\dots,k$. Recall that DeVries (1973) showed that the probability that the longest chord of the i^{th} particle is intersected is $p_i = 2\ell_i/\pi d$. Under the assumption of a systematic sample of parallel transects this probability does not change.

Let

$$z_i = \begin{cases} 1.0 & \text{with probability } p_i = \ell_i/\pi d \text{ if the} \\ & \text{longest chord of the } i^{\text{th}} \text{ particle is} \\ & \text{intersected.} \\ 0.0 & \text{with probability } (1-p_i) \text{ if the longest} \\ & \text{chord of the } i^{\text{th}} \text{ particle is not inter-} \\ & \text{sected.} \end{cases}$$

Then

$$\begin{aligned} E[\hat{D}|L_j] &= E \left[\frac{\frac{\pi}{2} \sum_{i=1}^n \ell_i^{-1}}{\sum_{j=1}^k L_j} \right] \\ &= E \left[\frac{\frac{\pi}{2} \sum_{i=1}^N z_i \ell_i^{-1}}{\sum_{j=1}^k L_j} \right] \\ &= \frac{\pi}{2} \left[\frac{\sum_{i=1}^N E[z_i] \ell_i^{-1}}{\sum_{j=1}^k L_j} \right] \\ &= \frac{\pi}{2} \left[\frac{\sum_{i=1}^N \frac{2 \ell_i}{\pi d} \ell_i^{-1}}{\sum_{j=1}^k L_j} \right] \\ &= \frac{\sum_{i=1}^N 1}{\sum_{j=1}^k L_j d} \\ &= \frac{N}{\sum_{j=1}^k (\text{area of the } j^{\text{th}} \text{ rectangle, } L_j \text{ by } d)} \end{aligned}$$

If the curvature is slight, the lines are long relative to the size of the particles, the particles are inside the study region, and the distance between transects is small then clearly $E[\hat{D}|L_j] \approx (N/A) = D$.

If we may assume that the bias due to curvature is negligible, the Lucas and Seber estimators of Equations (1) and (11), are unbiased for a systematic sample of k parallel transects.

Proof:

Consider the i^{th} particle and its closest transect. Closest is defined to be the minimum perpendicular distance between the center of the particles and the transect.

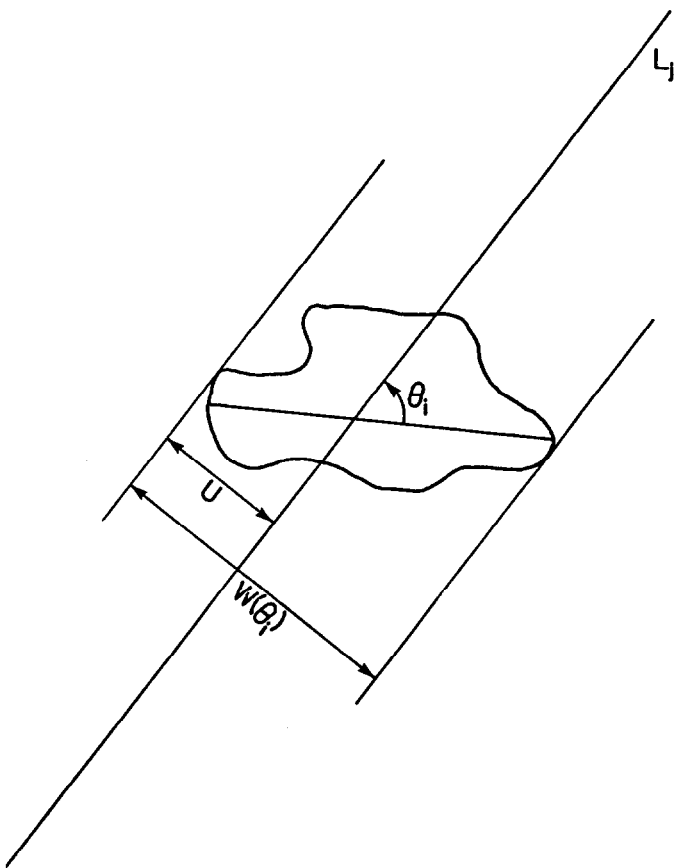


Figure A.1. The i^{th} particle intersected by the j^{th} transect. The distance, $w(\theta_i)$, is the perpendicular distance between two tangents to the particle constructed parallel to the transect. The distance, u , is the perpendicular distance between the transect and the upper tangent. The angle, θ_i , is the acute angle between a well-defined longest chord and the transect.

Define:

$\theta_i =$ The acute angle between a well-defined chord, say the longest chord, of the i^{th} particle and its closest transect, $i=1,2,\dots,n$.

$w(\theta_i) =$ The perpendicular distance between two tangents of the i^{th} particle which are parallel to the transect, a function of the angle θ and the shape of the particle.

$u_i =$ The perpendicular distance between the transect and the "upper" tangent. See Figure A.1.

Since the first transect in a systematic sample of transects is randomly positioned, the angle θ_i , for any particle and its closest transect, may take on any value in the interval $[0, \pi]$. That is θ_i is uniformly distributed on the interval $[0, \pi]$. And give θ_i , the variable u_i may take on any value in the interval $[0, w(\theta_i)]$, because the first transect is randomly located. Thus for any particle and its closest transect, u_i is uniformly distributed on the interval $[0, w(\theta_i)]$. Now we have the situation for θ_i and u_i as described by Lucas and Seber (1977) and the probability of intersection is the same. Thus, if the boundary effects are negligible, the estimators for density and percent cover,

$$\hat{D} = \frac{\sum_{i=1}^n w_i^{-1}}{\sum_{j=1}^k L_j} \quad (\text{A.3})$$

and

$$\hat{C} = \frac{\sum_{i=1}^n y_i}{\sum_{j=1}^k L_j} \quad (\text{A.4})$$

are approximately unbiased for systematic sampling.

Membership in the Society for Range Management. . .

- is open to those engaged in or interested in the study, management, or use of range ecosystems and the intelligent use of all range resources
- includes research scientists, ranchers, governmental agency administrators and technical personnel, teachers, students, and people from the business community
- provides members with two publications—one oriented to research (*Journal of Range Management*) and the other oriented to practical resource management (*Rangelands*)
- offers opportunities for face-to-face exchange of ideas at local, national, and international meetings of the Society.

Dues vary according to type of membership and geographical section. For application forms and additional information, contact the:

Executive Secretary
Society for Range Management
2760 West Fifth Avenue
Denver, Colorado 80204
(303) 571-0174