ACTA STEREOL 1984; 3/1: 3-10
ORIGINAL PAPER

THE RELATIVE EFFICIENCY OF POINT COUNT ESTIMATORS OF AREAL COVER

Pamela J. Davy* and Anthony J. Jakeman
Department of Statistics, IAS and Centre for Resource and Environmental Studies, Australian National University Canberra ACT 2600

ABSTRACT

The efficiency of linear point counts for the estimation of feature area in a two-dimensional matrix is investigated. With the line intercept method providing a reference estimator, simple formulae are derived to obtain the increase in error for point count estimators over the same length and the number of counts required to achieve the same variance. In an application to the estimation of areal cover in an arid grassland region, the formulae are seen to be a valuable aid in determining the precise spacing and number of grid points required to achieve a specified accuracy.

INTRODUCTION

It is well known that the following simple estimators of the fractional area of features in a two-dimensional matrix are unbiased:

L_L : the sum of linear transect intercept lengths divided by total transect length

P_P : the sum of point count intersections divided by total number of counts

Although the analysis extends, only point counts with uniform spacings Δ will be considered here because random point counts are usually bothersome in practice.

Of the above, it is intuitively obvious that the method based on L_L provides the lowest variance for a fixed transect length since the point count method reduces to this as the spacing Δ between counts approaches zero. However, to achieve a desired variance it may be less time consuming to take point counts over longer transect lengths than that needed for line intercept measurements. Moreover, it is practically infeasible in some cases to obtain accurate intercepts. For example, it is difficult to run a continuous measure through vegetation consisting of dense shrubs. On the other hand, it is simple enough to force the spokes of an unrimmed wheel through such vegetation and to count the number of spokes falling within and outside designated vegetation types.

The aim of this paper is to provide simple formulae that can be used to determine the following with respect to lineal intercept estimators of a given length L:

(a) the relative increases in error e_Δ incurred when point counts of different uniform spacings Δ are taken over the same length L;

(b) the number n_Δ or additional length of point counts required at specified uniform spacings Δ to achieve the same variance as L_L.

The theoretical foundations to provide such answers are given in the next section as are the required formulae. These are followed by results to corroborate them. A simple but flexible covariance function for the distribution of areal features is assumed to illustrate the results. In the final section, the usefulness of the formulae is demonstrated in an application involving the estimation of areal cover in an arid grassland area of Central Australia.

RELATIVE EFFICIENCY OF POINT COUNT ESTIMATION

Consider a second-order stationary random closed set X on the real line with centred covariance function C and lineal proportion $p = \Pr(0 \in X)$, and let $q = 1-p$. X may in turn have been obtained by stereological sampling from a planar or spatial process, but in this paper this further structure is ignored. The objective is to compare the efficiencies of the two estimators L_L and P_P of p, where L_L is the proportion of a transect of length L occupied by X, and P_P is the proportion of a grid of n points at spacing Δ occupied by X.

As shown, for example, in Stoyan (1979)

$$\mathrm{Var}(L_L) = \int_o^1 2\ (1-x)\ C(Lx)dx \tag{1}$$

and by expressing P_P as the mean of n correlated random indicator variables,

$$\mathrm{Var}(P_P) = n^{-1}\ C(0) + 2n^{-1} \sum_{i=1}^{n-1} (1 - i/n)\ C(i\Delta) \tag{2}$$

The formulae (1) and (2) are only useful in practice if the form of the covariance function is known and this is not usually the case.

When $n\Delta = L$, (2) is just a trapezoidal approximation to the integral in (1), whose associated error depends upon both the non-linearity of the integrand and the class width n^{-1}. For a quadratic integrand f(x), it is easy to show that the error involved in the trapezoidal approximation over [0,1] is

$$(f'(1) - f'(0))/12n^2 \tag{3}$$

Furthermore, the derivative of (1-x) C(Lx) is (1-x) C'(Lx) - C(Lx), which assumes values LC'(0) - C(0) and -C(L) for x = 0 and 1, respectively. Therefore, for sufficiently smooth C, it may be assumed on the basis of (1), (2) and (3) that

$$\begin{aligned}\mathrm{Var}(P_P) - \mathrm{Var}(L_L) &= (-LC'(0) + C(0) - C(L))/6n^2 + o(n^{-2}) \\ &= \Delta^2\ (-C'(0) + (C(0) - C(L))/L)/6L + o(\Delta^2)\end{aligned} \tag{4}$$

As shown in Serra (1982), $-C'(0) = \frac{1}{2} P_L$, where P_L is the expected number of boundary points per unit length. Also, if it is assumed that the range a of the process defined by

$$a = 2 \int_0^\infty C(x)\, dx/pq \tag{5}$$

is finite, it is also shown in Serra (1982) that

$$\mathrm{Var}(L_L) = apqL^{-1} + o(L^{-1}) \tag{6}$$

In practice, a is a measure of the distance at which the covariance becomes negligible and is known at least to an order of magnitude.

For large L and small Δ, (4) and (6) can be used to obtain the approximation

$$\frac{\mathrm{Var}(P_P) - \mathrm{Var}(L_L)}{\mathrm{Var}(L_L)} \simeq \Delta^2 P_L / 12\, apq \tag{7}$$

From the conditional variance formula $\mathrm{Var}(X|Y) = \mathrm{Var}(X) - \mathrm{Var}(E(X|Y))$ and noting that $E(P_P|L_L) = L_L$, it is valid to write

$$\mathrm{Var}(P_P) - \mathrm{Var}(L_L) = \mathrm{Var}(P_P|L_L),$$

where the right-hand side is to be interpreted as the additional squared error due to grid sampling. This formula is well known (Matheron, 1971, p.68; Journel, 1978, p.67). The relative error

$$e_\Delta = (\mathrm{Var}(P_P|L_L)/\mathrm{Var}(L_L))^{1/2}$$

may be approximated by

$$\tilde{e}_\Delta = (P_L/12\, apq)^{1/2}\, \Delta \tag{8}$$

Equation (8) gives the relative increase in error due to grid sampling, which is approximately linear in Δ for small Δ.

By (6), the right-hand side of (7) may equally well be interpreted as the approximate relative increase in length of grid required to achieve the same variance as a lineal estimator of length L or, in other words, the approximate number of grid points required is given by the formula

$$\tilde{n}_{\Delta} = (1 + \Delta^2 P_L/12\ apq)L/\Delta \tag{9}$$

It can be concluded from (7) that point count estimation is most effective when P_L/a is small and pq is large. Clearly, this occurs when the covariance function tends slowly towards zero and p is not too close to 0 or 1.

RESULTS

To illustrate the theory in the preceding section, let us consider the case of an exponential covariance function (see, for example, Pielou, 1964; Matheron, 1971).

$$C(x) = pq \exp(-\lambda x)$$

with $P_L = 2pq\lambda$ and $a = 2\lambda^{-1}$. Evaluation of (1) and (2) yields

$$\mathrm{Var}(L_L) = 2pq\ (\lambda L)^{-1} \left\{1 - (\lambda L)^{-1} (1 - \exp(-\lambda L))\right\} \tag{10}$$

and

$$\mathrm{Var}(P_P) = 2pq\ \left\{n\ (1 - \exp(-\lambda\Delta))\right\}^{-1} \left\{ \tfrac{1}{2} (1 + \exp(-\lambda\Delta)) - \right.$$
$$\left. [n\ (1 - \exp(-\lambda\Delta))]^{-1} \exp(-\lambda\Delta)\ (1 - \exp(-\lambda n\Delta)) \right\} \tag{11}$$

The exact relative increase in error e_Δ for $L = n\Delta$ depends only upon the two scale invariant quantities $\lambda\Delta$ and λL, while the approximate relative error increase $\tilde{e}_\Delta$ may be written as $\lambda\Delta/\sqrt{12}$. The approximation may be confirmed analytically, and is illustrated in the following table.

TABLE 1. The first row gives the approximate relative percentage increases in error for grid sampling which are compared with the exact increases for an exponential covariance

$\lambda\Delta$		.2	.4	1	2
$100\ \tilde{e}_\Delta$		5.77	11.55	28.87	57.74
$100\ e_\Delta$	λL				
	20	6.01	12.00	29.78	58.10
	40	5.89	11.76	29.20	57.01
	100	5.82	11.62	28.86	56.37
	200	5.80	11.58	28.76	56.18

EXAMPLE

Estimates of areal cover and size distribution are required for incorporation into a cell model of fire spread behaviour for areas of spinifex grassland in the Uluru National Park in Central Australia (Saxon et al., 1982). Fortunately, the spinifex plant is quite circular in shape so that an Abel equation (Davy and Jakeman, in preparation) relates lineal intercept or grid point distributions to the distribution of the spinifex diameters. Depending upon the fire history of an area and hence upon the age of the spinifex population, the average diameter of a plant can vary from a fraction of a metre up to 3 metres. However, for an area to be fire prone, the average diameter must be greater than half a metre. It is important, therefore, when taking transects for size distribution determination, that uniform counts be taken at least every metre and 1/2 and 1/4 metre intervals are obviously more desirable. Since the data requirements for areal cover estimation are less stringent than those for size distribution, the transect information obtained for sizes can also be used for estimation of areal cover. Therefore, calculation of $\tilde{e}_\Delta$ and $\tilde{n}_\Delta$ can be restricted to $\Delta = 0.25$, 0.5 and 1.0 here.

Because it is expected that the model of fire behaviour needs to be accurate to about 10 per cent, it is necessary to impose a standard deviation on the areal cover estimation of around 5 per cent. Since p tends to vary around 0.5, such a standard deviation is 0.025. If the exponential covariance function is used to check the results given by (8) and (9), a value of P_L and hence λ is required in addition to p.

Consider, for example, recent samples taken through a 25 year old community of spinifex plants. For this area, $P_L = 0.58$ boundaries per unit length and $P_P = 0.53$. Using $p = 0.53$, equations (10), (11), (8) and (9) can be applied to estimate: the standard error of such an areal cover for the lineal interception method against increasing transect length; the increase in error when a grid count (of 1/4, 1/2 and 1m spacings) is used, and the number of point counts required to achieve the same variance as the lineal interception method.

Table 2 reports the results for L = 100, 200, 400, 700 and 1,000m. It shows that a transect of 700m is required to achieve a standard deviation on areal cover of around 5% (.025/.53 x 100) using lineal interception. However, with $n_1 = 780$, only an extra 80 metres of point counting at 1m intervals is required to obtain a 5% error on the point count estimate.

It is clear that as the grid spacing decreases, so does the length of point count required. For $\Delta = 1/4$ m, only an extra 5m (2820/4 - 700) is necessary. However, the workload at $\Delta = 1/4$ is relatively high when it is considered that almost 4 times (2820/780) as many wheel point counts are needed compared to using $\Delta = 1$. Therefore, it is recommended that the largest possible grid spacing be used which still constrains the size distribution data to remain accurate also to within a 5% standard error.

TABLE 2. For various transect lengths L, grid spacings $\Delta = 1.0$, 0.5 and 0.25, $P_L = 0.58$ and $p = 0.53$, the table gives with respect to lineal interception the relative increase in error for grid sampling using the approximate formula (8), the exact increase for the exponential covariance function and the approximate number of grid points required to achieve the same variance using (9)

L (metres)		100	200	400	700	1000
$\sqrt{Var(L_L)}$		.065	.046	.033	.025	.021
e_1		.335	.334	.333	.333	.333
$\tilde{e}_1$	.336					
$\tilde{n}_1$		112	223	446	780	1113
$e_{0.5}$		.169	.168	.168	.168	.168
$\tilde{e}_{0.5}$	.168					
$\tilde{n}_{0.5}$		206	412	823	1440	2057
$e_{0.25}$		.085	.084	.084	.084	.084
$\tilde{e}_{0.25}$	.084					
$\tilde{n}_{0.25}$		403	806	1612	2820	4029

ACKNOWLEDGEMENTS

The authors would like to thank Dr E.C. Saxon and Mr G.F. Griffin for providing the transect data.

REFERENCES

Journel A. Mining Geostatistics, New York: Academic Press, 1978.

Matheron G. The Theory of Regionalized Variables and its Applications, Ecole des Mines de Paris, 1971.

Pielou EC. The spatial pattern of two-phase patchworks of vegetation. Biometrics 1964; 20: 156-167.

Saxon EC, Griffin GF and Allan GE. Conservation of pyrogenic landscapes, Preschedule AS 89/2. Rangelands Research Unit, Commonwealth Scientific and Industrial Research Organisation, Alice Springs, Australia, 1982.

Serra J. Image Analysis and Mathematical Morphology, New York: Academic Press, 1982.

Stoyan D. On the accuracy of lineal analysis. Biometrical J 1979; 21: 439-449.

Accepted: 1983-11-14