ISSN 1684-8403
Journal of Statistics
Volume 15, 2008, pp. 44-53

A Combination of Adaptive and Line Intercept Sampling Applicable in Agricultural and Environmental Studies

Azmeri Khan[1]

Abstract

An adaptive procedure is described for the line intercept sampling. Modified Hansen-Hurwitz and Horvitz-Thompson estimators are used to find estiamtors for the population mean, population density and coverage. An example is given to justify the method and to compare it with ordinary line intercept sampling.

Keywords

Adaptive Cluster Sampling, Line Intercept Sampling, Network Sampling

1. Introduction

In many studies in the field of forestry, range and harvest management, ecology etc., it will be difficult to apply usual sampling methods to identify rare characteristics in the population. For example, suppose the pest contamination of a crop is of interest. It is understood that the contamination would be rare in the wide area under cultivation of the crop, but the contamination will also be expected to be clustered, i.e., if a certain plot suffers pest attack, the neighboring plots are also likely to be attacked. Similar examples can be cited for different fields as well. In studies of salinity of the soil, similar complexity can arise. In sub-coastal zones soil salinity is a rare thing but it is of clustered nature as well.

Thompson (1990) proposed an adaptive cluster sampling (ACS) design for the populations which are rare and have cluster tendencies. In his proposed method, the study area is first divided into several geographical units. A primary sample of units is drawn from the study area using the simple random sampling (SRS)

[1] Institute of Statistical Research and Training (ISRT),
University of Dhaka, Dhaka-1000, Bangladesh
Email: azmeri@isrt.ac.bd

method. Whenever a unit satisfies a given condition (for example, possessing a minimum level of the characteristic of interest), the neighboring units of that selected unit are then added to the sample. Again some of these added units may satisfy the condition, their neighborhood units are further added to the sample, and so on. The group of units which are selected in the final sample as a result of selection of a unit in the primary sample is called a network. A unit which does not satisfy the condition but selected in the primary sample is called a network of size one. Thompson (1992) obtained an unbiased estimator of the population mean by modifying the Hansen et al. (1953) estimator. He also obtained the variance and the estimated variance of the estimator. Thompson (1991a, 1991b) extended the ACS procedure where the primary sample is selected using the systematic or the strip sampling procedure and for stratified sampling. But, for the above mentioned situations where area under investigation is large, the ACS procedure might not also be helpful since the SRS observations are usually scattered over a very big region. One of the most frequently used methods for such fields is line intercept sampling (LIS) (see, for example, Thompson, 1992), where a line is drawn along the study area and the particles (small region with evidence of salinity or pest attack) intersected by the transect line are studied (Fig. 1).

For the estimation of the total number of particles using the LIS method the width of the article intercepted by the transect line is required to be measured. The LIS procedure can also be used for estimating the coverage, i.e. how much area of the study region is covered by the particles? For coverage estimation, the length of the intersection of the particle parallel to the transect is measured. Fig. 2 shows the measurement of the width of the intercepted particle and the length of the intersection of the particle parallel to the transect. When dealing with the populations of above discussed type and if the objects are rare and have cluster tendencies, an ACS can be used where the primary sample is selected by the LIS method instead of the SRS method. In situations like the pest attack example, the SRS of plots or only a LIS procedure might not result in a substantial penetration of the pattern of the pest attack. Moreover, the coverage of the pest contamination will be very difficult to measure. A sampling procedure, applicable into these types of situations is discussed in this paper and the method is termed as adaptive line intercept sampling (ALIS) due to the nature of the method being a combination of ACS and LIS.

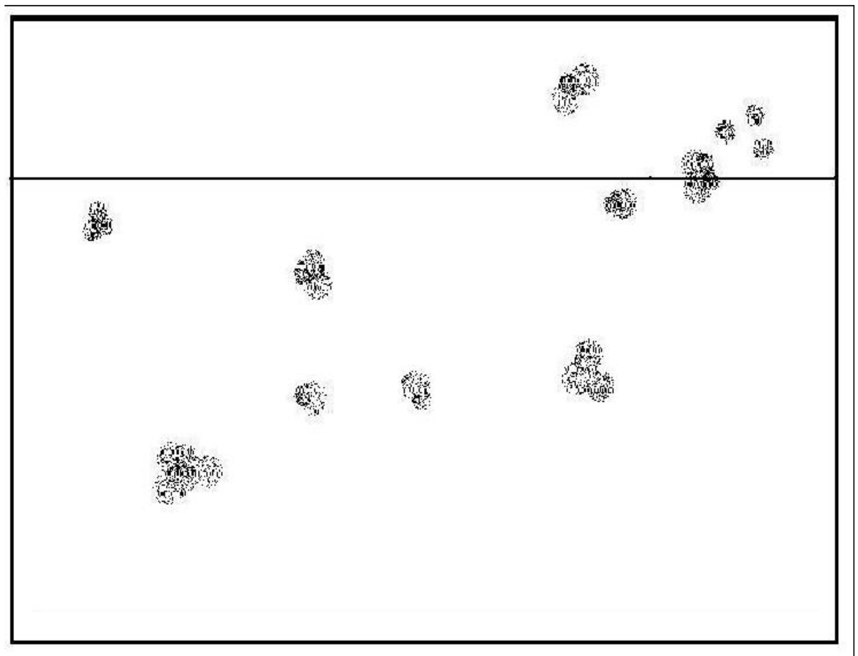

Fig. 1: A line intercept sampling

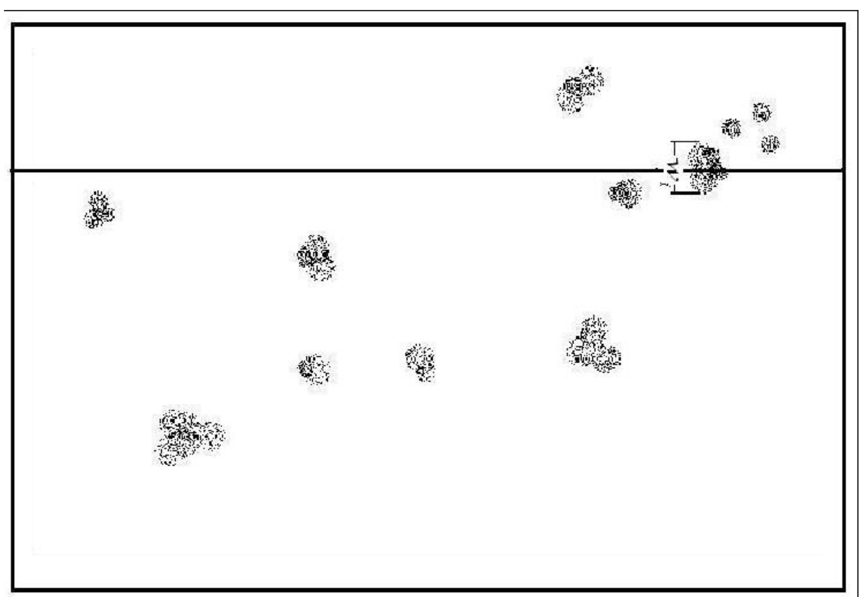

Fig. 2: The width of the intercepted particle in line intercept sampling

2. Sampling Procedure

In the simple LIS method, *n* points are randomly selected along the baseline *b* and *n* transact lines of length *l* are drawn along the study area which are perpendicular to the baseline. The particles intersected by the transect lines are observed. But if a rare and cluster population is under study, whenever a particle is intersected by the transect line it is likely that there are more particles around that intersected particle. Those neighboring particles can be added in the sample using the technique of ACS method. In the adaptive line intercept sampling (ALIS) procedure, whenever a particle is intersected by the transect line, half circles of radius r (r is arbitrarily chosen by the researcher) centering the particle are drawn on both side of the transect line, if any particle is observed within any of those half circles, the half circle is extended with radius *2r* and the selected half circles are observed. If any particle is observed, half circle is extended with radius *3r* and studied. The procedure is continued as long as any particle is observed (Fig. 3).

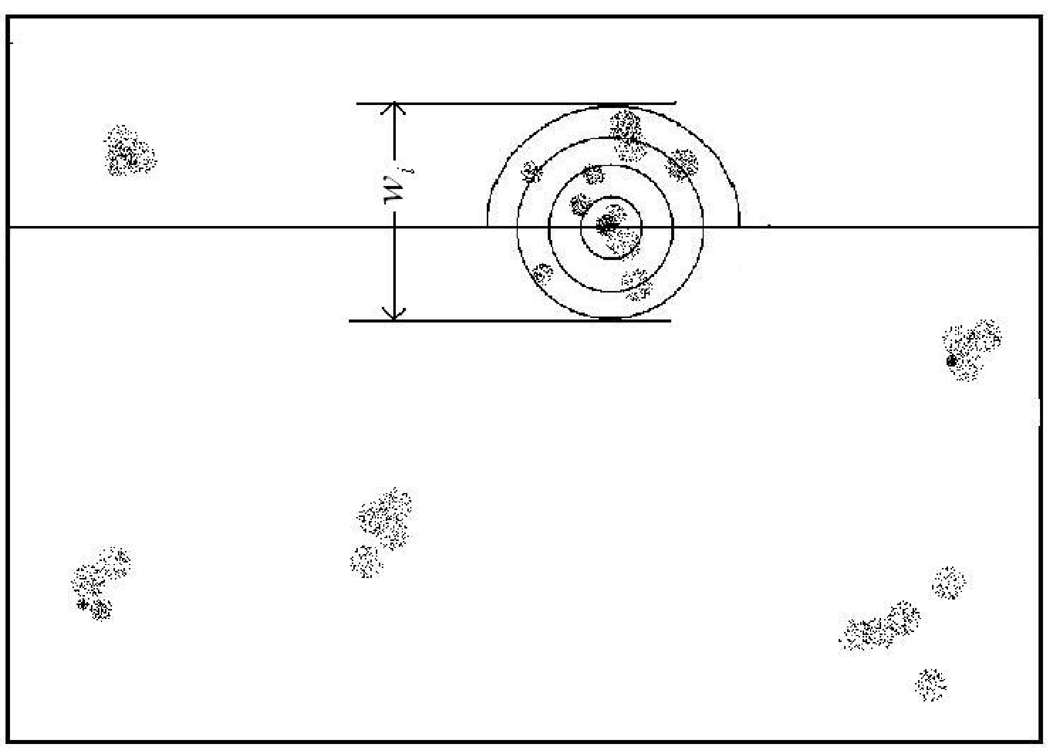

Fig. 3: The width of the intercepted network in adaptive line intercept sampling

As a result of selection of one particle in the primary sample a group of particles can be added to the sample, this group of particles is termed as a network. The particles within a network have the property that if any particle of this network is intersected by the transect line all other particles will be included in the sample.

3. Estimators using ALIS

3.1 Hansen-Hurwitz Estimator

As the ACS is dealing with networks, the selection probability of network is required. The primary sample is selected by the LIS method and hence the probability depends on the width of the selected particle and the width of the baseline (see, for example, Thompson, 1992). Suppose that there are K_i particles in the i^{th} network, y_i is the value of the variable of interest associated with i^{th} network. Let w_i be the width of the baseline from which the perpendicular lines intersect the final half circles of both sides of the transect (Fig. 4). So the probability of selecting the i^{th} network is $p_i = \frac{w_i}{b}$.

Now following the formulas of the LIS estimator of the population total τ of the variable of interest and its estimated variance are proposed as follows:

$$\hat{\tau} = \frac{1}{n}\sum_{l=1}^{n} v_l, \text{ where } v_l = \sum_{i=1}^{k}\frac{y_i}{p_i}, \tag{3.1}$$

$$v(\hat{\tau}) = \frac{1}{n}\sum_{l=1}^{n}\frac{(v_l - \hat{\tau})^2}{n-1}. \tag{3.2}$$

Suppose that the study area is A, then the population density per unit τ is $D = \frac{\tau}{A}$.

Often D is required to be estimated. An estimator of D is as follows:

$$\hat{D} = \frac{\hat{\tau}}{A} = \frac{1}{nA}\sum_{l=1}^{n}\sum_{i=1}^{k}\frac{y_i}{p_i} \ , \tag{3.3}$$

and its estimated variance is given by

$$v(\hat{D}) = \frac{1}{A^2}v(\hat{\tau}) = \frac{1}{A^2 n}\sum_{l=1}^{n}\frac{(v_l - \hat{\tau})^2}{n-1}. \tag{3.4}$$

3.2 HorvitzThompson estimator of ALIS

The population total τ can also be estimated using the Horvitz-Thompson estimator as follows:

$$\tilde{\tau} = \sum_{i=1}^{k} \frac{y_i}{\pi_i}, \tag{3.5}$$

where $\pi_i = 1-(1-p_i)^n$ is the inclusion probability of i^{th} network and

$$v(\tilde{\tau}) = \sum_{i=1}^{n}\left(\frac{1-\pi_i}{\pi_i^2}\right)y_i^2 + \sum_{i=1}^{n}\sum_{j=1}^{n}\left(\frac{\pi_{tj}-\pi_i\pi_j}{\pi_i\pi_j}\right)\frac{y_i y_j}{\pi_{ij}}, \tag{3.6}$$

where π_{tj} is the joint inclusion probability of i^{th} and j^{th} network and

$$\pi_{ij} = \pi_i + \pi_j - 1 + \left(1 - \frac{w_i + w_j - w_{ij}}{b}\right)^n,$$

with w_{tj} being the width of the baseline from which the perpendicular line intersects both the i^{th} and j^{th} networks.

The population density per unit can also be estimated as

$$\tilde{D} = \frac{\tilde{\tau}}{A} = \frac{1}{A}\sum_{i=1}^{k}\frac{y_i}{\pi_i} \quad , \tag{3.7}$$

and with its estimated variance as

$$v(\tilde{D}) = \frac{1}{A^2}v(\hat{\tau}) = \frac{1}{A^2}\left[\sum_{i=1}^{n}\left(\frac{1-\pi_i}{\pi_i^2}\right)y_t^2 + \sum_{i=1}^{n}\sum_{j=1}^{n}\left(\frac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j}\right)\frac{y_t y_j}{\pi_{ij}}\right] \tag{3.8}$$

3.3 Coverage Estimation using ALIS

Coverage can also be estimated by ALIS. Let L_i be the length of i^{th} transect, n be the number of randomly placed transect, m be the number of particles intersected by the transect lines, M be the number of particles in the study area A and a_k be the area covered by the particle k $(k = 1, 2, ..., M)$. To measure the lengths of the intersections of every particles with the transects, lines parallel to the transect are drawn from random points of the particles. For each particle, the length of the intersection of the j^{th} particle in i^{th} network intersected by the transect is measured, let this length be denoted by x_{ij} .

The average length of the intersections for the i^{th} network can be obtained as follows:

$$\bar{x}_i = \frac{\sum_{j=1}^{K_i} x_{ij}}{K_i}. \tag{3.9}$$

The formula for estimator of the coverage $C = \sum_{k=1}^{M} \frac{a_k}{A}$ can be given as follows:

$$\overline{C} = \frac{\sum_{j=1}^{m} \bar{x}_j}{\sum_{l=1}^{n} L_l}. \tag{3.10}$$

4. An Example

To illustrate how one might use the estimators given in Section 2, simulated data by Diggle (1983) is used. He described a conditional poisson cluster process, where 100 elements are randomly distributed among 25 parents. This data set is considered as our population in this example as it satisfies the assumptions of the ACS procedure being applicable when the population is rare and has cluster tendency. The ALIS procedure is applied to this population to obtain an estimate of the population total and the population density. Here our study area A is 196 sq units and the length of the baseline is $b = 14$ and population total τ is 100 with a density of 0.51 per sq unit.

Adaptive Line Intercept Sampling

To perform ALIS, at first two transects of lengths l_1 and l_2 are drawn from two random points of the baseline b, where $l_1 = l_2 = 14$ unit. First transect line intersects one element and second transect intersects two elements of the population (Fig. 4).

For the first transect, two half circles of radius 0.4 unit are drawn on both sides of the transact centering the point of the intersection of the element and the transect and both half circles are observed. Elements are found on both of the half circles, so the half circles are extended to radius $2r$ i.e. 0.8 unit and observed. But now no element is found on the left half circle, so the half circle is extended only on the right side of radius $3r$ i.e. 1.2 unit and the area of the half circle is observed. The procedure is continued as long as an element is observed. Finally, as a result of adaptation, a cluster of elements is found and the width of the half circles is measured and it is found that $w_1 = 3.8$ unit.

Similar procedure is followed for the second transect and two clusters are found and the width of the half circles are measured and it is found that $w_2 = 0.9$ unit and $w_3 = 1.7$ unit. Now the probabilities of the selection of the networks are calculated using (3.1) as follows:

$$p_1 = \frac{3.8}{14} = 0.271,\ p_2 = \frac{0.9}{14} = 0.064,\ p_3 = \frac{1.7}{14} = 0.121$$

and the total number of elements is estimated using equation (3.2) as follows :

$$\hat{\tau} = \frac{1}{2}(\nu_1 + \nu_2) = \frac{1}{2}\left(\frac{18}{0.271} + \left(\frac{3}{0.064} + \frac{6}{0.121}\right)\right) = 83.33 \cong 83,$$

and the estimated variance of the above estimate is calculated as follows:

$$\mathrm{v}(\hat{\tau}) = 555.445.$$

This provides a two-standard deviation limit of

$$\left(83 - 2\times\sqrt{555},\ 83 - 2\times\sqrt{555}\right) = (60,\ 107).$$

Now the density of the population is also calculated using equation (3.3) as follows:

$$\hat{D} = \frac{83.33}{196} = 0.43\ ,$$

with estimated variance $v(\hat{D}) = 0.015$.

This provides a two-standard deviation limit for the density as:

$$\left(0.43 - 2\times\sqrt{0.015},\ 0.43 - 2\times\sqrt{0.015}\right) = (0.31,\ 0.55).$$

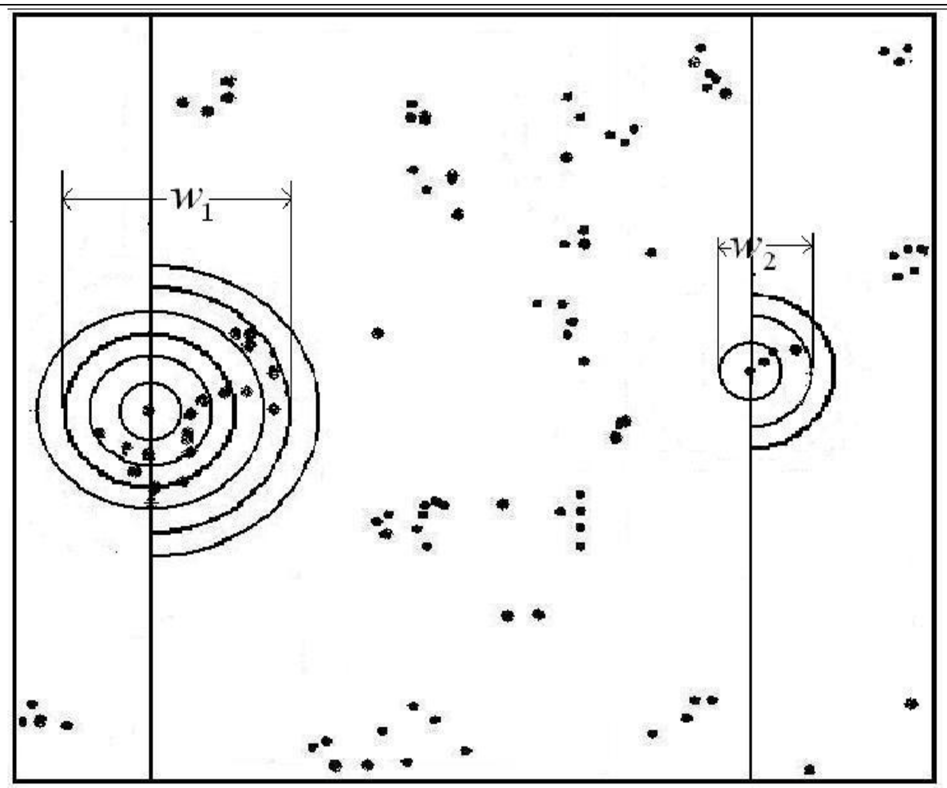

Fig. 4: An example of adaptive line intercept sampling

5. Conclusions

From the sample obtained for the example, we can see that the point estimate of the population total possesses enough accuracy (83 as an estimate of 100), also the two standard deviation limit for the population total contains the population total which is indicative of the fact that the estimator and its variance both are of applicable level. Similar features are observed while estimating the density. The ALIS method is seen to be an practically applicable method which can be advantageously applicable in situation where the variable under interest is rare and has a cluster tendency. Especially in agricultural and environmental studies where the geographic or spatial dispersion, density, coverage of a characteristic are of importance, the ALIS method can be an appropriate choice.

References

1. Diggle, P. J. (1983). *Statistical Analysis of Spatial Point Patterns*, Academic Press, New York.
2. Hansen, M. H., Hurwitz, W. N. and Madow, W. G. (1953). *Sample Survey Methods*, Wiley Publications, New York.
3. Thompson, S. K. (1990). Adaptive cluster sampling. *Journal of the American Statistical Association*, **85**, 1050-1059.
4. Thompson, S.K. (1991a). Adaptive cluster sampling: Designs with primary and secondary units, *Biometrics*, **47**, 1103-1115.
5. Thompson, S.K. (1991b). Stratified adaptive cluster sampling, *Biometrika*, **78**, 389-397.
6. Thompson, S. K. (1992). *Sampling*, John Wiley & Sons, New York.