United States
Department
of Agriculture

Forest Service

Rocky Mountain
Research Station

Research Note
RMRS-RN-17

August 2004

Sampling Using a Fixed Number of Trees Per Plot

Hans T. Schreuder[1]

Abstract—The fixed number of trees sample design proposed by Jonsson and others (1992) may be dangerous in applications if a probabilistic framework of sampling is desired. The procedure can be seriously biased. Examples are given here.

Introduction

Density based sampling is often considered attractive. For example, variable radius plot (VRP) sampling is popular because it avoids the excessive sampling of small trees. We show here that the procedure of using plots in the field to obtain a fixed number of trees per plot is undesirable since it unnecessarily violates some key issues regarding probabilistic sampling. The fact that it may have small bias in some simulation studies does not prove that it is a satisfactory procedure since realistic situations can invariably be found or constructed where bias will be serious.

Review of Literature

Jonsson and others (1992) propose selecting a fixed number of trees at each primary sampling unit so that each tree has a different probability of selection. Using the unbiased Horvitz-Thompson estimator requires knowledge of the actual probabilities of selection of the sample trees. However, these can only be approximated unless the locations of all trees are known. This is clearly not practical; even then it would be cumbersome to calculate those probabilities. They approximate the probabilities of selection of the trees in a given plot by dividing the n trees selected by the local tree density. They state that the proposed estimator is then approximately unbiased if the variable of interest is spatially independent of the local tree densities.

Lessard and others (1995) compare this n-tree distance sampling to fixed and variable radius plot sampling in estimating basal area per acre. They conclude that variations of n-tree sampling, especially 3-tree sampling, are more biased and more variable than fixed and variable radius plot sampling in all stands examined. They consider n-tree sampling to be fast and cost-competitive for estimating basal area and density and have the advantage of being able to provide estimates of spatial pattern parameters because distance measurements to trees is an integral part of n-tree distance sampling. Lessard and others (2002) compared n-tree distance sampling with fixed-radius plot sampling for estimating number of trees per ha. They show that n-tree distance based sampling estimators are at least as precise as those for plot sampling generally for both random and clustered forests, provided that the fixed radius plot size is less than the ratio of (n-2) and less than the expected density where n is the number of trees included at an n-tree location. They do not break out bias of the estimators in this study.

Lynch and Rusydi (1999) compared n-tree sampling with some other sampling techniques and found that the n-tree sampling underestimated both volume and density in the plantations sampled in Indonesia.

---

[1] **Hans T. Schreuder** is a Mathematical Statistician (retired) with the Rocky Mountain Research Station in Fort Collins, CO.

The Horvitz-Thompson estimator is an unbiased estimator of Y:

$$\hat{Y}_{HT} = \sum_{i=1}^{n} y_i / \pi_i, \qquad (1)$$

where y_i is the attribute of interest and π_i is the first order inclusion probability.

The variance is given by:

$$V(\hat{Y}_{HT}) = 1/2[\sum_{i \neq j}^{N} (\pi_i \pi_j - \pi_{ij})(y_i / \pi_i - y_j / \pi_j)^2 \qquad (2)$$

or

$$V(\hat{Y}_{HT}) = 1/2[w_{ij}(y_i / \pi_i - y_j / \pi_j)^2 \qquad (2a)$$

with $w_{ij} = \pi_i \pi_j - \pi_{ij}$.

Unbiased variance estimators are:

$$v_1(\hat{Y}_{HT}) = 1/2[\sum_{i \neq j}^{n} \{(\pi_i \pi_j - \pi_{ij}) / \pi_{ij}\}(y_i / \pi_i - y_j / \pi_j)^2 \qquad (3)$$

and

$$v_2(\hat{Y}_{HT}) = [\sum_{i=1}^{n} \{(1-\pi_i)/\pi_i^2\} y_i^2 + \sum_{i \neq j}^{n} \{(\pi_{ij} - \pi_i \pi_j)/\pi_{ij}\}(y_i y_j / \pi_i \pi_j)] \qquad (4)$$

(Schreuder and others 1993).

Methods

If the estimated probabilities of selection of trees are π_i^*, $(i = 1, \ldots N)$ with N the number of trees in the population, then

with bias:

$Bias = \sum_{i=1}^{N} (1 - \pi_i / \pi_i^*) y_i$. Thus, the degree of bias of the estimator depends on how well we estimate π_i, the true probabilities of selection for the trees.

Specific situations:

Assume that we have a grid sample of n locations imposed on the population of interest and that we have a consistent bias in the sense that in the top and bottom half of the population, respectively, the π_i (i=1...N) are over- and underestimated by 10 percent. Then we have:

$$E(\hat{Y}_{HT}) = \sum_{i=1}^{N} (y_i / \pi_i^*) \pi_i = \sum_{i=1}^{N_1} \{y_i / (1.1 \pi_i)\} \pi_i$$

$$+ \sum_{i=N_1+1}^{N_2} \{y_i / (0.9 \pi_i)\} \pi_i = \sum_{i=1}^{N_1} y_i / 1.1 + \sum_{i=N_1+1}^{N_2} y_i / 0.9.$$

Hence the bias is $0.1 \left\{ \sum_{i=N_1=1}^{N_2} y_i / 0.9 - \sum_{i=1}^{N_1} y_i / 1.1 \right\}$. Clearly this bias can be quite serious.

It is easy to construct realistic cases where estimating π_i by π_i^* can be seriously in error. See figure 1 for an example.

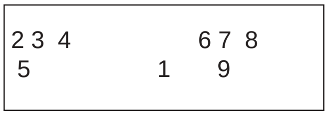


Figure 1—An illustration of a possible scenario with numbered tree locations, where probabilities of selection will be seriously underestimated.

The Horvitz-Thompson estimator $\hat{Y}_{HT}$ is an unbiased estimator of Y:

$$\hat{Y}_{HT} = \sum_{i=1}^{n} y_i / \pi_i, \tag{1}$$

where y_i is the attribute of interest and π_i is the first order inclusion probability.

The variance is given by:

$$V(\hat{Y}_{HT}) = 1/2[\sum_{i \neq j}^{N} (\pi_i\pi_j - \pi_{ij})(y_i/\pi_i - y_j/\pi_j)^2 \tag{2}$$

or

$$V(\hat{Y}_{HT}) = 1/2[w_{ij}(y_i/\pi_i - y_j/\pi_j)^2 \tag{2a}$$

with $w_{ij} = \pi_i\pi_j - \pi_{ij}$.

Unbiased variance estimators are:

$$v_1(\hat{Y}_{HT}) = 1/2[\sum_{i \neq j}^{n} \{(\pi_i\pi_j - \pi_{ij})/\pi_{ij}\}(y_i/\pi_i - y_j/\pi_j)^2 \tag{3}$$

and

$$v_2(\hat{Y}_{HT}) = [\sum_{i=1}^{n} \{(1-\pi_i)/\pi_i^2\} y_i^2 + \sum_{i \neq j}^{n} \{(\pi_{ij} - \pi_i\pi_j)/\pi_{ij}\}(y_iy_j/\pi_i\pi_j)] \tag{4}$$

(Schreuder and others 1993).

Methods

If the estimated probabilities of selection of trees are π_i^*, $(i = 1,\ldots N)$ with N the number of trees in the population, then

$E(\hat{Y}_{HT}) = E\sum_{i=1}^{n} y_i/\pi_i^* = \sum_{i=1}^{N}(y_i/\pi_i^*)\pi_i$, with bias:

$Bias = \sum_{i=1}^{N}(1-\pi_i/\pi_i^*)y_i$. Thus, the degree of bias of the estimator depends on how well we estimate π_i, the true probabilities of selection for the trees.

Specific situations:

Assume that we have a grid sample of n locations imposed on the population of interest and that we have a consistent bias in the sense that in the top and bottom half of the population, respectively, the π_i (i=1…N) are over- and underestimated by 10 percent. Then we have:

$$E(\hat{Y}_{HT}) = \sum_{i=1}^{N}(y_i/\pi_i^*)\pi_i = \sum_{i=1}^{N_1}\{y_i/(1.1\pi_i)\}\pi_i$$

$$+ \sum_{i=N_1+1}^{N_2}\{y_i/(0.9\pi_i)\}\pi_i = \sum_{i=1}^{N_1} y_i/1.1 + \sum_{i=N_1+1}^{N_2} y_i/0.9.$$

Hence the bias is $0.1\left\{\sum_{i=N_1=1}^{N_2} y_i/0.9 - \sum_{i=1}^{N_1} y_i/1.1\right\}$. Clearly this bias can be quite serious.

It is easy to construct realistic cases where estimating π_i by π_i^* can be seriously in error. See figure 1 for an example.

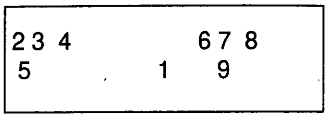


Figure 1—An illustration of a possible scenario with numbered tree locations, where probabilities of selection will be seriously underestimated.

In this example for n=5, tree 1 will be selected for any sample location falling in the rectangle. Hence π_i =1 and π_i^* will never be close to 1 for this tree for any sample falling in the rectangle.

The following illustrates another problem with fixed number of trees per plot sampling: Assume that a plot location falls in part (a) and another in part (b) (figure 2). For n=5, plot size for (a) will obviously be much smaller than for (b) and measuring plot size for (b) would take time.

Sample location (a):

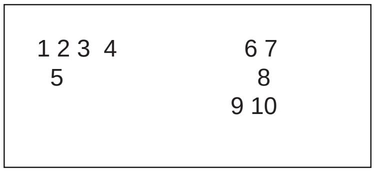
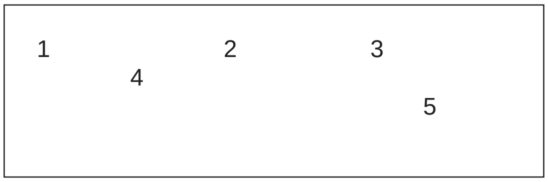


Figure 2—An illustration of a possible scenario with widely varying plot sizes.

Sample location (b):



What are we gaining by this approach relative to a fixed area plot sample? Clearly in the latter case we would have several samples with 1 or even 0 trees in it. Certainly it is easier to obtain the information required for the fixed area plots. Are we gaining in information with the fixed number of tree plots? It is not clear that we are unless we are talking about variables keyed to specific trees requiring a large or guaranteed number of sample trees. If distances to trees are required, they could easily be measured with either the fixed or variable radius plots.

Conclusions

Using designs with fixed number of tree plot sampling can be dangerous. The method is biased and can be hard to implement in practice. In addition it is not fulfilling a need that cannot be met by traditional sampling designs.

Acknowledgments

The author would like to thank Drs. Tom Lynch and Veronica Lessard for their valuable comments.

Literature Cited

Jonsson, B.; Holm, S.; Kallur, H. 1992. A forest inventory method based on density-adapted circular plot size. Scandinavian Journal of Forest Research. 7: 405-421.

Lessard, V. C.; Reed, D. D.; Drummer, T. D. 1995. N-tree distance sampling compared to fixed radius plot and variable radius point sampling in forest inventory estimation of basal area per acre. In proceedings Monte Verita Conference on Forest Survey Designs; "Simplicity versus Efficiency" and Assessment of Non-Timber Resources. M. Kohl and P. Brassel, Eds. May 2-7, 1994. Monte Verita Ascona, pp 81-90. Swiss Federal Institute of Technology. Section of Forest Inventory and Planning. Zurich, Switzerland.

Lessard, V. C.; Reed, D. D.; Drummer, T. D. 2002. Precision of density estimates from fixed-radius plots compared to N-tree distance sampling. Forest Science. 48: 1-6.

Lynch, T. B.; Rusydi, R. 1999. Distance sampling for forest inventory in Indonesian teak plantations. Forest Ecology and Management 113:215-221.

Schreuder, H. T.; Gregoire, T. G.; Wood, G. B. 1993. Sampling methods for multiresource inventories. New York: John Wiley and Sons. 446 p.

RMRS
ROCKY MOUNTAIN RESEARCH STATION

The Rocky Mountain Research Station develops scientific information and technology to improve management, protection, and use of the forests and rangelands. Research is designed to meet the needs of the National Forest managers, Federal and State agencies, public and private organizations, academic institutions, industry, and individuals.

Studies accelerate solutions to problems involving ecosystems, range, forests, water, recreation, fire, resource inventory, land reclamation, community sustainability, forest engineering technology, multiple use economics, wildlife and fish habitat, and forest insects and diseases. Studies are conducted cooperatively, and applications may be found worldwide.

Research Locations

Flagstaff, Arizona
Fort Collins, Colorado*
Boise, Idaho
Moscow, Idaho
Bozeman, Montana
Missoula, Montana
Lincoln, Nebraska
Reno, Nevada
Albuquerque, New Mexico
Rapid City, South Dakota
Logan, Utah
Ogden, Utah
Provo, Utah
Laramie, Wyoming

*Station Headquarters, Natural Resources Research Center, 2150 Centre Avenue, Building A, Fort Collins, CO 80526.

You may order additional copies of this publication by sending your mailing information in label form through one of the following media. Please specify the publication title and series number.

Fort Collins Service Center

Telephone	(970) 498-1392
FAX	(970) 498-1396
E-mail	rschneider@fs.fed.us
Web site	http://www.fs.fed.us/rm
Mailing address	Publications Distribution Rocky Mountain Research Station 240 West Prospect Road Fort Collins, CO 80526