ISSN 1946-7664. MCFNS 2012
Vol. 4, Issue 2, pp. 77–80
Mathematical and Computational Forestry & Natural-Resource Sciences

AVAILABLE ONLINE AT HTTP://MCFNS.COM

Submitted: Jul. 27, 2012
Accepted: Sep. 20, 2012
Published: Sep. 30, 2012
Last Correction: Mar. 8, 2013

SOME CURRENT SUBSAMPLING TECHNIQUES IN FORESTRY

Kim Iles

Kim Iles and Associates Ltd., Nanaimo, B.C., Canada.

Abstract. Because different tree parameters are of differing importance, and have different variability, efficiency in sampling would suggest that some of the principle variables be subsampled. One convenient way to do this is to sample different numbers of items at the same sample locations. This paper is a review of some current techniques in subsampling for measured values, especially with Variable Plot sampling, but including Fixed Plot and 3P sampling as well.

Keywords: Variable plot sampling, angle sampling, point sampling, VBAR, Big BAF, multi-point sampling.

1 Introduction

Many people consider that the beginning of unequal-probability sampling was with a famous paper by Hansen and Hurwitz (1943) about estimating with varying sampling probabilities. Foresters were among the first to actually use this kind of sampling with the adoption of Variable Plot (VP) Sampling, which was called Angle-Count Sampling by Bitterlich (1947). Although this is considered a selection which is "nonequal" for individual trees, there is a sense in which it is equal probability sampling. Although it is true that trees are chosen with different probabilities, each square foot of basal area in the stand is selected with equal probability in Variable Plot Sampling in order to establish the Volume to Basal Area Ratio or "VBAR". A Mean-of-Ratios estimator is used to combine the individual tree VBARs.

The volume of the tree is not needed in any variant of VP sampling, only the ratio of volume per square foot is required. One could just as well measure the ratio on a particular individual square foot of the tree, but the ratio is usually determined using the entire tree volume divided by the entire tree basal area. One could consider it as the average of the cluster of individual square feet that exist in each tree. Many computer programs perform the VBAR calculation explicitly for sample trees, and these programs are more flexible in adopting recent subsampling schemes. In several types of sampling, such as Critical Height Sampling (Iles, 1979) the tree volume need never be known, and an estimate of VBAR is established based on where a sample point falls within the expanded tree cross-section.

Sampled basal area is multiplied by sampled VBAR to give stand or strata volumes. Typically, both basal area and VBAR are determined at the same sample points. The two parts of this Product Estimator have different variability and different costs for their establishment, therefore suggesting different sample sizes. Traditionally, since the 1950's in the Pacific Northwest United States, this fact has been recognized by the use of "Count and Measure plots", where basal area is established on all the sample plots by counting trees; but trees are only measured on a portion of the plots, say 1/4 of them. Because counting trees is less expensive than making VBAR measurements, this provided considerable efficiency in sampling. Both the Product Estimator approach and the traditional "Double Sampling" method give the same answer when all trees are measured on some proportion of the plots and the usual ratio estimator is used. This method led to considerable angst when the measure plots were widely separated, since there was constant concern that the plots were "too far apart" and might be missing or oversampling clustered tree characteristics.

2 Subsampling issues

More recently, the method of "Big BAF" sampling solved this, by spreading the VBAR trees throughout the stand (Bell et. al., 1983). If, for instance, a 100 BAF angle gauge is used to choose VBAR trees, when a 20 BAF angle gauge prism was used to count trees, a proportion of approximately (20/100) or 1 out of 5 counted trees would be chosen as VBAR trees. Exactly the same relative selection probability would exist with either of the angle gauges, so there is no difference in the sampling properties of the trees and VBARs involved; only the number of trees differs. It is a simple subsampling scheme from the trees counted with the smaller BAF. Instead of subsampling trees and their VBARs by "clusters" with the Count-Measure approach, the trees are dispersed through the area.

Informal studies in the Pacific Northwest indicate that the efficiency of spreading out the trees gives the dispersed trees about double the efficiency for VBAR estimation that the clusters provided with the same number of trees. When tree measurement costs are high, this is a considerable advantage. The efficiency increase in hardwoods is not known, but such studies are easy to make when data is available in previous data sets where every tree on every plot was measured. Surprisingly, the Count-Measure approach is still used in many places.

One could easily imagine many other ways to choose from the trees initially counted by the 20 BAF, such as randomly choosing every 5^{th} tree with the 20 BAF prism; but the larger BAF selection was easy to use and understand, and auditing function for this approach was simple. Not all subsamples are appropriate for all computations, of course, but the use of subsampling is not difficult to do correctly in this case. If you cared about the average distance from all the 20 BAF trees to the plot center, you would use a wedge of the original plot to subsample trees. The essential idea is just to sample the smaller BAF tree selection by distributing them more equally across the area sampled. Any process which spreads the VBAR measurements more equally has the increased efficiency of systematic sampling, and is more comfortable for users who are concerned about clusters measured at too great an interval. The Big BAF method is simple, and is quickly spreading among contractors and others who are not heavily constrained by traditional and standardized procedures or by compilation programs that owe their structure to the Fixed Plots of 5 decades ago.

3 Historical Development

Apparently, the first use of the term "VBAR" was due to J. R. Dilworth in the 1950's (John Bell, personal communication). Recently, I discovered a letter from 1949 from Lew Grosenbaugh to his colleague Mr. Wheeler, which mentioned not only the use of the VBAR method, but clearly suggested the subsampling approach. Grosenbaugh certainly anticipated the use of a VBAR ratio, as well as the issue of subsampling for it. When the Big BAF method was published in 1983, an extensive poling of specialists in the field at that time produced no knowledge of this simple technique. I recently found an easily overlooked comment in Grosenbaugh's 1950 paper "*Plotless Timber Estimates – New, Fast, Easy* that suggested the possible use of subsampling with a second BAF. Grosenbaugh never seems to have actively supported a subsampling procedure of this sort. I suspect that this is because of the issue of the variance formulas involved.

4 Estimators and their variance

For forest sampling done today in many parts of the world, the volume of the stand is determined by a Product Estimator where the stand basal area is multiplied by the ratio estimate for volume from measured plots. When all trees are measured on some plots the usual calculation of volume is exactly the same as the use of a Double Sampling procedure where the plot basal areas on each "count plot" are multiplied by the ratio-of-means estimate for VBAR from measured plots. The volume estimates of the Product Estimator vs. the Double-Sampling estimator are identical, even though the mathematics seem different. This is just an illusion of the seemingly different algebra.

The variance estimators, however, are not the same. Neither of them are correct for variance, because we are dealing with systematic sampling in virtually all practical applications. In the early days of Variable Plot Sampling, the fact that the Double Sampling equation was exact for a random sample and available in traditional textbooks was very important to some people, and using anything else invited contention. Grosenbaugh was probably just avoiding this friction by avoiding the Big BAF method.

The Product Estimator assumes that the basal area estimate and the VBAR estimate are statistically independent. With measurements taken at the sample points, this is not strictly the case. There is only a trivial correlation between the estimates of VBAR and basal area on sample points in typical forest sampling. The conifer gross volume VBAR, and in secondgrowth timber the net VBAR, are very consistent and highly correlated with tree height. Most modern forest sampling is done with strata that are strongly determined by tree height, therefore the VBAR is fairly consistent and has virtually no significant correlation to plot basal area estimates.

This being the case, the product estimator of variance is quite good, and will be highly correlated to the double sampling variance estimator when they can be compared. In practical situations, there is no way to know which is more accurate, and there is little difference between them. For anyone wanting to take the small correlation into account, the paper by Goodman (1960) and his subsequent papers handle this correlation explicitly. In forestry, the simple Standard Error in percent known as "Bruce's Formula" is widely used.

$$SE\%_{combined} = \sqrt{(SE\%_{basal\,area})^2 + (SE\%_{VBAR})^2}$$

The reason this matters is that if you are subsampling with the Big BAF method, then the traditional double sampling formula does not apply and cannot be used. The product formula for the mean is unbiased (except for the trivial bias in the ratio estimators used by both methods), and the traditional product formula for variance is quite reasonable; certainly it ignores a smaller affect than is ignored by treating the samples as random selections. Compilation programs that have always used the product formula for the average and variance with calculated VBARs for sample trees have no trouble processing Big BAF samples.

Using a larger prism to select only a few trees for boring to determine growth has been suggested before, such as in the textbook by Husch et al (1982, page 287). This was done to automatically weight the measurement by the basal area of the trees in the stand as well as to reduce the number of trees bored. Other random selections normally require weighting procedures for the measurements. If several trees are initially selected, but only one is measured at random, then the results of that measurement should be weighted by the number of trees from which it was selected. Since boring trees is difficult, and often requires tedious adjustments for missing sections, a subsample is clearly desirable. Arbitrary selections, such as "the first tree from North", have all the biases that are present with nonsample data.

5 Other sampling methods

Subsampling for tree characteristics can be done on Fixed Plots as well. Simply dividing a circular plot into x parts and choosing one part randomly will obviously correctly choose approximately 1/x of the total number of trees on the entire plots. Using a "Split Plot" in this way is fundamentally different than enlarging or shrinking a plot until a certain number of trees is obtained, which can cause bias. In most cases, the divided parts do not need to be of the same size, and can be divided for convenience or to give approximately the same number of trees in the subsampled sections.

Using a constant size plot that is smaller than some standard larger plot, for example when a large plot is used to estimate numbers of trees and a smaller plot is used to get volume per tree, is correct as long as edge affects are accounted for. The recent Walkthrough method by Ducey et. al. (2004) continues to be the most reasonable correction for edge affects, and is very nearly unbiased except for very small or contorted situations. In these cases, the more tedious methods called the "Toss-Back" method (Iles, 2004, page 641) or the newer "Through and Fro estimator" method of Flewelling and Strunk (2012) seem to be the only general unbiased alternatives.

With 3P sampling, a subsampling approach can be easily implemented. The equally probable systematic selection of some of the 3P trees (or a random selection of 1/x) might well be chosen for a more expensive assessment. Actual felling of the trees, such as has been used in the Pacific Northwest, is often too much effort for all the sample trees, but might be efficient with a small subset. Felling some of the trees seems to be the only way to approach many of the small biases from studies or assumptions that simply cannot be quite correct.

The rate of subsampling can be complicated to compute exactly, especially for small sample sizes, but for practical purposes the ratio of the Coefficient of Variation of the variables is a very close guideline. If the variability in 3 stages in a Product Estimator for Variable Plot Sampling is : 60% for Basal Area, 30% for VBAR, and 5% for correction of VBAR using felled trees, then sampling in the ratio of 60:30:5 during the three stages would be indicated. Given that the variability is never known before the sample, this ratio is probably as good a guideline as can be obtained.

6 Conclusion

Sampling all trees on all plots is normally very inefficient. There are simple techniques for selecting subsample trees, and for spreading these selections systematically or semi-systematically throughout the initial stage of sampling. These techniques apply to Fixed, Variable or 3P methods, and are easily accommodated by many compiling programs. Where the compiling programs are limiting, they should be changed. There are no problems with estimation of the mean, and I would argue that variance calculation by Bruce's method is more than adequate for practical purposes.

Acknowledgments

The author would like to thank three anonymous reviewers for their comments on the paper.

References

Bell, J.F., K. Iles, D. Marshall. 1983, Balancing the Ratio of TreeCountOnly sample points and VBAR Measurements in Variable Plot Sampling. Proceedings: Renewable Resource Inventories for Monitoring Changes and Trends. Corvallis, Oregon.

Bitterlich, W. 1947. The angle count method. Allgemeine Forst und Jagdzeitung 58:94–96.

Ducey, M.J., J.H. Gove, T.H. Valentine. 2004. A walk-through solution to the boundary overlap problem. Forest Science. 50(4): 427–435.

Flewelling, J.W., J.L. Strunk. 2012. The Walk Through and Fro Estimator for Edge Bias Avoidance. Forest Science. http://dx.doi.org/10.5849/forsci.10-135. (In Press). Published online on Apr. 26, 2012.

Goodman, L.A. 1960. On the Exact Variance of Products Journal of the American Statistical Association. 55(292), pp. 708–713.

Grosenbaugh, L.R. 1952. Plotless Timber Estimates – New, Fast, Easy. Journal of Forestry, 50(1), pp. 32–37.

Hansen, M.M., W.N. Hurwitz. 1943. On the theory of sampling from finite populations. Annals of Mathematical Statistics. 14:333–362.

Husch, B., C. Miller, T. Beers. 1982. Forest Inventory, John Wiley and Sons, Third Edition. New York, NY. 402 p.

Iles, K. 1979. Some Techniques to Generalize the use of Variable Plot and Line Intersect Sampling. Forest Resource Inventories Workshop, Colorado State University.

Iles, K. 2003. A Sampler of Inventory Topics. Kim Iles & Associates Ltd. ISBN 0-9732198-0-7, pp. 869.