United States
Department of
Agriculture

Forest Service

**Rocky Mountain
Forest and Range
Experiment Station**

Fort Collins,
Colorado 80526

**Research Paper
RM-290**

Sunter's pps Without Replacement Sampling as an Alternative to Poisson Sampling

Hans T. Schreuder, H.G. Li, and S.M. Sadooghi-Alvandi

$$V(\hat{Y}_{as}) = V[E(\hat{Y}_{as}|n_a)] + E[V(\hat{Y}_{as}|n_a)]$$
$$= V\left[\frac{n_e}{n_a} E(\hat{Y}_{us}|n_a)\right] + E\left[\frac{n_e^2}{n_a^2} V(\hat{Y}_{us}|n_a)\right]$$

$$V(\hat{Y}_{us}) = \sum_i \frac{(\sigma^2+\beta^2)\, x_i^2\, X^*}{n_t x_i} + \frac{n_t-1}{n_t} X^*\, \beta X \sum_i \gamma_i^* \beta\, x_i$$
$$- \frac{n_t-1}{n_t} X^* \sum_i \gamma_i^* (\sigma^2+\beta^2)\, x_i^2 - \beta^2 X^2$$
$$= \frac{(\sigma^2+\beta^2)\, X^* X}{n_t} + \frac{n_t-1}{n_t} \beta^2 X^* X \sum_i \gamma_i^*\, x_i$$
$$- \frac{n_t-1}{n_t} (\sigma^2+\beta^2)\, X^* \sum_i \gamma_i^*\, x_i^2 - \beta^2 X^2$$

$$V(\hat{Y}_{as}) \doteq Y2 \frac{V(n_a)}{n_e^2} + (1+3 \frac{v(n_a)}{n_e^2}) \sum_{i<j} \sum (\Pi_i \Pi_j - \Pi_{ij})$$
$$(Y_i/\Pi_i - Y_j \Pi_j)^2$$

Abstract

An alternative to Poisson sampling called Sunter sampling is introduced into the forestry literature. The probability of selecting a sample unit depends on its size and the number of previous units selected as well as on the sizes and number of units remaining in the population. This results in a less variable sample size than in Poisson sampling. Sunter sampling is more efficient than Poisson sampling where a sampling list is available prior to sampling if a slightly biased adjusted estimator similar to one in Poisson sampling is used. Approximate true variances are given for both the unadjusted and adjusted estimators. Two sample-based variance approximations provide reliable estimates of both the true and simulation variance of the adjusted estimator. Sunter sampling is not yet a practical alternative when no sampling list is available but perhaps could be an alternative to, for example, point-Poisson sampling.

USDA Forest Service
Research Paper RM-290

February 1990

Sunter's pps Without Replacement Sampling as an Alternative to Poisson Sampling

Hans T. Schreuder, Mathematical Statistician
Rocky Mountain Forest and Range Experiment Station[1]

H. G. Li, Research Associate
Statistics Department
Colorado State University

S. M. Sadooghi-Alvandi, Assistant Professor
Department of Mathematics and Statistics
Shiraz University, Iran

[1]*Headquarters is in Fort Collins, in cooperation with Colorado State University.*

Sunter's pps Without Replacement Sampling as an Alternative to Poisson Sampling

H. T. Schreuder, H. G. Li, and S. M. Sadooghi-Alvandi

Management Implications

A new sampling scheme is introduced that could be quite efficient in timber sales and other inventories. Some of these inventories currently suffer from random sample sizes which may result in inventories being more costly or less precise than planned for. The study suggests applications where this new method may be useful.

Introduction

Poisson sampling was introduced by Hajek (1957, 1964) and Grosenbaugh (1964) as a pps without replacement sampling scheme when no sampling list is available. Units are selected exactly proportional to size. Because the sample size is random, however, the sample size is frequently different from what was intended.

The purpose of this note is to discuss Sunter sampling, an alternative for which the sample size is less variable, and to compare it to Poisson sampling in applicability and efficiency. The objective is to encourage practitioners to look for practical applications for Sunter sampling. One possibility might be point-Sunter sampling as an alternative to point-Poisson sampling.

Review of Literature

Sunter (1986) gives an alternative to Poisson sampling. He starts with the premises that:

1. units are selected with probabilities, Π_i, proportional to x_i
2. the joint inclusion probabilites Π_{ij} of units indexed i and j are positive for all i, j
3. $\Pi_i\Pi_j - \Pi_{ij} > 0$ for all i, j ($i \neq j$)
4. Π_{ij} is known or can be calculated for all i,j.

Note that Poisson sampling satisfies all these conditions except that $\Pi_{ij} = \Pi_i\Pi_j$ for all $i \neq j$, so that $\Pi_i\Pi_j - \Pi_{ij} = 0$, for all $i \neq j$.

As Sunter (1986) indicates, under any set of positive selection probabilities, a Horvitz-Thompson estimator of the population total $Y = \sum_{i=1}^{N} Y_i$ is given by

$$\hat{Y} = \sum_{i=1}^{n_a} Y_i/\Pi_i \qquad [1]$$

where n_a is the achieved sample size, with variance

$$V(\hat{Y}) = \sum_{i=1}^{N} Y_i^2(1-\Pi_i)/\Pi_i - \sum_{i\neq j} Y_iY_j(\Pi_i\Pi_j-\Pi_{ij})/\Pi_i\Pi_j$$

and unbiased variance estimator (if $\Pi_{ij} > 0$ for all i,j)

$$v(\hat{Y}) = \sum_{i=1}^{n_a} Y_i^2(1-\Pi_i)/\Pi_i^2 - \sum_{i\neq j} Y_iY_j(\Pi_i\Pi_j-\Pi_{ij})/\Pi_{ij}\Pi_i\Pi_j \qquad [2]$$

For fixed n_a:

$$\sum_{i=1}^{N}\Pi_i = n_a,\ \sum_{j(\neq i)}^{N}\Pi_{ij} = (n_a-1)\Pi_i,\ \sum_{i\neq j}^{N}\Pi_{ij} = n_a(n_a-1)$$

$$V(\hat{Y}) = \sum_{i<j}^{N}(\Pi_i\Pi_j-\Pi_{ij})(Y_i/\Pi_i - Y_j\Pi_j)^2 \qquad [3]$$

and

$$v(\hat{Y}) = \sum_{i<j}^{n_a}(\Pi_i\Pi_j-\Pi_{ij})(Y_i/\Pi_i - Y_j/\Pi_j)^2/\Pi_{ij} \qquad [4]$$

Note that the first premise arises from the desire to make the square term in equation [3] small, which will be true if there is proportionality between the variate value and size measure. Satisfying the second and fourth premises ensures that either equation [2] or equation [4] can be computed as appropriate. The third premise implies that the variance estimate in equation [4] cannot be negative and indicates some stability in the estimate.

Stuart (1963) indicates that it is advantageous to have $\Pi_i\Pi_j - \Pi_{ij} < 0$ for the largest $(Y_i/\Pi_i - Y_j/\Pi_j)^2$ so that there are pps schemes that can be more efficient than either Poisson or Sunter sampling. But it is advantageous to be able to compute all joint probabilities of selection and to always have positive variance estimates. Sunter (1977) first proposed a list sequential sampling scheme that results in fixed sample size. The disadvantage of this scheme is that the second part of the sample is chosen by simple random sampling and that the joint probability of selection is not available between the sample units in the second part and those in the first part of the sample. Sunter then proposes a modified sequential sampling scheme as follows:

a. Select a number x^* such that for $\Pi_i = n_t x_i/Z$

$$n_t x_i \leq X_i^* = \sum_{j=1}^{N} x_j + x^* \quad (i = 1,2,\ldots,N) \qquad [5]$$

$$\text{For } x^* = 0,\ n_t x_{n_t} \leq X_{n_t}^* \text{ if } n_t = 1$$

where n_t is the target sample size. Sampling is proportional to x.

b. Select the i-th unit with conditional probability

$$P(U_i|n_i) = n_i x_i/X_i^* \qquad [6]$$

where U_i is the event that unit i is selected, and n_i denotes the number of units remaining to be selected when the i-th unit is reached.

Sampling stops when n_t units are selected or when the population of units has been exhausted. Thus, the

achieved sample size n_a is no greater than the target sample size n_t. The above is referred to as Sunter sampling in the rest of this paper.

Assume, without loss of generality, that the size measures x_i are scaled such that $\sum_{i=1}^{N} x_i = 1$, then units are selected with probability

$$\Pi_i = n_t x_i / X^* \quad (i = 1,2,\ldots,N)$$

and joint probabilities

$$\Pi_{ij} = n_t(n_t-1)x_i x_j \gamma_i^* / X^* \quad (j > i = 1,2,\ldots,N-1)$$

where $X^* = X_1^* = 1 + x^*$, $X_i^* = \sum_{j=1}^{N} x_j + x^*$,

$$\gamma_1^* = 1/X_2^*, \gamma_i^* = (1/X_{i+1}^*)(1-x_1/X_2^*) \ldots$$
$$(1-x_{i-1}/X_i^*)(i = 2,3,\ldots,N-1),$$

and x^* is a constant such that $n_t x_i \le X_i^*$ for $i = 1, 2, \ldots, N$. Note that the expected sample size, n_e, is

$$E(n_a) = n_t/X^* = n_e,$$

$$V(n_a) = \sum_{i=1}^{N} \Pi_i(1-\Pi_i) - 2 \sum_{i<j}^{N} (\Pi_i\Pi_j - \Pi_{ij}),$$

and $$P(n_a < n_t) \le n_t x^*/X^*$$

This sampling scheme reduces the variance in the sample size n_a relative to Poisson sampling by the difference of the sums of product and joint probabilities.

The selection algorithm will work for any ordering of the population, but will be easier in certain situations. An ideal ordering would be close to the requirements of Corollary 1.1 in Sunter (1986) which states: if $x_{N-n_t+1} = x_{N-n_t+2} = \ldots = x_N$, $n_t x_i < X_i = \sum_{j=1}^{N} x_j$ with $i = 1, 2, \ldots, N-n_t$ for some ordering of the population, then $x^* = 0$ will work. Then a sample of exactly n_t units will be obtained. This is rarely satisfied in practice. The next best ordering may be to arrange the population units by decreasing order of the size of x. If a sampling list is available, x^* can be computed easily.

Analytical Comparisons

The Horvitz-Thompson estimator for Sunter sampling is

$$\hat{Y}_{us} = \sum_{i=1}^{n_a} Y_i/\Pi_i = \sum_{i=1}^{n_a} \frac{Y_i}{x_i} X^*/n_t \qquad [7]$$

and can be rewritten as

$$\hat{Y}_{us} = X/n_e \sum_{i=1}^{n_a} Y_i/x_i \qquad [8]$$

where X is the population total of the auxiliary variable x.

Analogous to the (biased) adjusted estimator, $\hat{Y}_a$, suggested by Grosenbaugh (1967), a more efficient but biased estimator should be

$$\hat{Y}_{as} = \hat{Y}_{us}\, n_e/n_a$$

The variance of $\hat{Y}_{us}$ can be found in most sampling textbooks. The true variance of $\hat{Y}_{as}$ is very complicated. An approximate variance of $\hat{Y}_{as}$ similar to the one suggested by Schreuder et al. (1971) for Poisson sampling is

$$V_G(\hat{Y}_{as}) \doteq \left[\sum_{i=1}^{N} \frac{x_i}{X} \left(\frac{Y_i X}{x_i} - Y\right)^2/n_e\right] \left[1 + \frac{V(n_a)}{n_e^2}\right] \qquad [9]$$

The most suitable situation for pps sampling without replacement with the Horvitz-Thompson estimator in equation [1] is the one suggested by Van Deusen (1987):

$$Y_i = \beta x_i + e_i \qquad [10]$$

with $Ee_i = 0$ and $Ee_ie_j = \begin{cases} \sigma^2 x_i^2 & \text{for } i = j \\ 0 & \text{for } i \ne j \end{cases}$

Under this model

$$V(\hat{Y}_a) \doteq \sigma^2 X^2/n_e \left[1 + 1/n_e - \sum_{i=1}^{N} x_i^2/X^2\right] \qquad [11]$$

where $\hat{Y}_a$ is the adjusted estimator of Poisson sampling, and n_e is the expected sample size. It is shown in the appendix that under this same model

$$V(\hat{Y}_{us}) = \frac{1}{n_t}(\sigma^2 + \beta^2)X^*X + \frac{n_t-1}{n_t}\beta^2 X^* X \sum_{i=1}^{N} \gamma_i^* x_i \qquad [12]$$
$$- \frac{n_t-1}{n_t}(\sigma^2 + \beta^2)X^* \sum_{i=1}^{N} \gamma_i^* x_i^2 - \beta^2 X^2$$

and an approximate variance for $\hat{Y}_{as}$ is

$$V(\hat{Y}_{as}) \doteq \sigma^2 X^2/n_e \left[\frac{1}{n_e} - n_t(n_t-1)\left(X^2/X^* - \sum_{i=1}^{N} x_i^2/X^*/\gamma_i^*\right)\right] \qquad [13]$$

The comparison does not lead to a simple analytical comparison of which method (Poisson or Sunter) has the smallest variance because the joint probabilities of selection under Sunter sampling are analytically very complex and no simplifying assumption is obvious. The two methods are compared based on their approximate true variances and also by a simulation study for some populations. The approximate true variance of $\hat{Y}_{as}$ is given in the appendix.

Two sample-based variance approximations suggest themselves for $V(\hat{Y}_{as})$. The first one is the Yates-Grundy variance estimate which assumes a fixed sample size, i.e., $v(\hat{Y}_{YG}) = v(\hat{Y})$ in equation [4] and the second one is with the replacement variance estimator corrected for the random sample size

$$v_G(\hat{Y}_{as}) = \frac{X^2}{n_e} \sum_{i=1}^{n_a} \left(\frac{Y_i}{x_i} - \sum_{i=1}^{n_a} \frac{Y_i}{n_e x_i}\right)^2/(n_e-1) \qquad [14]$$

Populations and Simulations

To develop an understanding of gains and losses in efficiency and bias due to the validity of error structure

for postulated linear models, the loblolly pine population described in Schreuder and Thomas (1985) was used as a starting point. This population consists of 1,801 loblolly pine trees with cubic foot volume as the dependent variable (y) and tree basal diameter squared times total height (x) as the auxiliary variable. This data set is basically a subset of a larger data set of 5,500 trees described in McClure et al. (1983), selected such that the diameter distribution of these trees is similar to that for loblolly pine in the Piedmont of the Southeast, a highly skewed distribution with many small and few large trees.

From this data set (N = 1,801) the following parameters were computed. The mean-of-ratios regression parameter is

$$\beta_{mr} = \sum_{i=1}^{N} Y_i / N x_i$$

$$\text{with variance } \sigma^2_{mr} = \sum_{i=1}^{N} [(Y_i - \beta_{mr} x_i)/x_i]^2 / N$$

The ratio-of-means regression parameter is

$$\beta_{rm} = \sum_{i=1}^{N} Y_i / \sum_{i=1}^{N} x_i$$

$$\text{with variance } \sigma^2_{rm} = \sum_{i=1}^{N} [(Y_i - \beta_{rm} x_i) / \sqrt{x_i}]^2 / N$$

The simple linear regression parameter is:

$$\beta_r = \sum_{i=1}^{N} (Y_i - \bar{Y})(x_i - \bar{X}) / \sum_{i=1}^{N} (x_i - \bar{X})^2$$

with intercept parameter $\alpha_r = \bar{Y} - \beta_r \bar{X}$. Variances used with these two parameters were

$$\sigma^2_{r1} = \sum_{i=1}^{N} [(Y_i - \alpha_r - \beta_r x_i)/x_i]^2 / N \text{ and}$$

$$\sigma^2_{r2} = \sum_{i=1}^{N} [(Y_i - \alpha_r - \beta_r x_i)/\sqrt{x_i}]^2 / N$$

Then populations 1–10 were generated in pairs (replicates) using the same x's as in the original population and with the following Y's.

For populations 1 and 2:
$Y_i - e_i = \beta_{mr} x_i$ where $e_i \sim N(0, \sigma^2_{mr} x_i^2)$
For populations 3 and 4:
$Y_i - e_i = \beta_{rm} x_i$ where $e_i \sim N(0, \sigma^2_{rm} x_i)$
For populations 5 and 6:
$Y_i - e_i = \alpha_r + \beta_r x_i$ where $e_i \sim N(0, \sigma^2_{r1} x_i^2)$
For populations 7 and 8:
$Y_i - e_i = \alpha_r + \beta_r x_i$ where $e_i \sim N(0, \sigma^2_{r2} x_i)$
and for populations 9 and 10:
$Y_i - e_i = \alpha_r + \beta_r x_i$ where $e_i \sim N(0, \sigma^2_{r2} x_i^{2.1})$

To save computer time the same random numbers were used for populations 1, 3, 5, 7, and 9 and similarly for populations 2, 4, 6, 8, and 10. Population 11 is the original loblolly pine population. All populations are the same size. Parameter values are $\beta_{mr} = 0.001836$, $\beta_{rm} = 0.002016$, $\alpha_r = -1.146$, and $\beta_r = 0.002209$.

Another population data set, the Pine Creek diameter data set, for which 2,748 tree diameters (D) were measured in 1967 and 1979, was also used. The size of the intercept is close to the size of the slope for the relationship between D^2 in 1979 and 1967. The plotted relationship seems to indicate a very strong linear relationship with variability increasing with increase in diameter class size.

For the Pine Creek data set, populations 1–10 were generated as above for the loblolly pine but with parameter values $\beta_{mr} = 1.323$, $\beta_{rm} = 1.272$, $\alpha_r = 2.75$, and $\beta_r = 1.110$.

As with the loblolly pine, the same random numbers were used for populations 1, 3, 5, 7, and 9 and for 2, 4, 6, 8, and 10, respectively. Population 11 is the original Pine Creek diameter population. All 11 populations are the same size.

Using Poisson sampling and Sunter sampling, 2,000 samples of intended size 20 were drawn from each population. In Sunter sampling, the order of visiting the units is very important. First units were arranged in decreasing size of the auxiliary variable with $x^* = 0.12\%$ of X for the loblolly pine and $x^* = 0.37\%$ of X for the Pine Creek. This is a good strategy for minimizing the value of x^* so as to guarantee $\Pi_i = nx_i/X^* \leq 1$ for all i (i=1, ..., N). The expected sample size is 19.98 for the loblolly pine ($x^* = 0.12\%X$), and 19.93 for the Pine Creek ($x^* = 0.37\%X$).

To evaluate other orderings of the population, the population units were randomly ordered twice. For these orderings $x^* = 4.27\%X$ and $x^* = 5.23\%X$ were required for these two situations for the loblolly pine data set and $x^* = 1.05\%X$ and $x^* = 1.26\%X$ for the Pine Creek data sets. The expected sample sizes are 19.15 ($x^* = 4.27\%X$) and 19.02 ($x^* = 5.23\%X$), respectively, for the two loblolly pine data sets, 19.79 ($x^* = 1.05\%X$) and 19.75 ($x^* = 1.26\%X$), respectively, for the two Pine Creek data sets.

Computational and Simulation Results

In all cases in the simulation study, the estimation biases of $\hat{Y}_{us}$, $\hat{Y}_{as}$, and $\hat{Y}_a$ were negligible and are not shown. In table 1, the expected variance under the model $\hat{Y}_{us}(V(\hat{Y}_{us}))$ is shown in equation [12], the approximate true variance for $\hat{Y}_{as}(V(\hat{Y}_{as}))$ in equation [13], the approximate true variance for $\hat{Y}_a(V(\hat{Y}_a))$ in equation [11], simulation variances for $\hat{Y}_{as}$ and $\hat{Y}_a$ based on 2,000 simulations and average sample-based variance estimates for $\hat{Y}_{as}$ based on fixed-sample size ($v(\hat{Y}_{YG})$) in equation [14], and based on with-replacement sampling ($v_G(\hat{Y}_{as})$) in equation [14] for populations 1 and 2 of each data set (only these two populations were generated using equation [4]). Based on the theoretical variances it is clear that $\hat{Y}_{as}$ is more efficient than $\hat{Y}_{us}$ for Sunter sampling and also more efficient than $\hat{Y}_a$ for Poisson sampling. The simulation variances confirm that $\hat{Y}_{as}$ for Sunter sampling is more efficient than $\hat{Y}_a$ for Poisson sampling.

Either $v(\hat{Y}_{YG})$ or $v_G(\hat{Y}_{as})$ seem to give reliable variance estimates of the true variance of $\hat{Y}_{as}$ (expressed either by the approximate variance $V(\hat{Y}_{as})$ or the simulation vari-

Table 1.—Comparison of theoretical, iterated, and average sample variances for Sunter and Poisson sampling for populations 1 and 2 of both data sets with x sorted in decreasing order.

Data set	Method	x^*	Estimator	Variance	Population 1	Population 2
Loblolly	Sunter	0.12%X	$\hat{Y}_{us}$	$V(\hat{Y}_{us})/10^6$	0.89	0.85
	Sunter	0.12%X	$\hat{Y}_{as}$	$V(\hat{Y}_{as})/10^6$	0.79	0.78
	Poisson	____	$\hat{Y}_a$	$V(\hat{Y}_a)/10^6$	0.85	0.80
	Sunter	0.12%X	$\hat{Y}_{as}$	Iterated variance/10^6	0.74	0.78
	Poisson	____	$\hat{Y}_a$	Iterated variance/10^6	0.82	0.79
	Sunter	0.12%X	$\hat{Y}_a$	$v(\hat{Y}_G)/10^6$	0.78	0.79
	Sunter	0.12%X	$\hat{Y}_a$	$v_G(\hat{Y}_{as})/10^6$	0.80	0.76
Pine Creek	Sunter	0.37%X	$\hat{Y}_{as}$	$V(\hat{Y}_{us})/10^7$	1.06	1.07
	Sunter	0.37%X	$\hat{Y}_{as}$	$V(\hat{Y}_{as})/10^7$	0.82	0.81
	Poisson	____	$\hat{Y}_a$	$V(\hat{Y}_a)/10^7$	0.89	0.85
	Sunter	0.37%X	$\hat{Y}_{as}$	Iterated variance/10^7	0.85	0.85
	Poisson	____	$\hat{Y}_a$	Iterated variance/10^7	0.88	0.83
	Sunter	0.37%X	$\hat{Y}_a$	$v(\hat{Y}_G)/10^7$	0.82	0.89
	Sunter	0.37%X	$\hat{Y}_a$	$v_G(\hat{Y}_{as})/10^7$	0.85	0.82

ance). Neither sample-based variance estimator is clearly better.

Table 2 shows a very consistent though small advantage in efficiency for $\hat{Y}_{as}$ (Sunter sampling) relative to $\hat{Y}_a$ (Poisson sampling) for almost all of the 11 populations in each data set. This is true even though the expected sample size for Sunter sampling is somewhat smaller especially for the x^* values much larger than 0. There is little difference in the results for the three different x^* values even though the expected sample size for Sunter sampling is less with increasing x^*. There is no obvious change in gain in Sunter sampling with randomly ordered populations over Poisson sampling for different variance functions or presence or absence of an intercept in the y–x relationship. Estimator $\hat{Y}_{us}$ of Sunter sampling could be very inefficient especially when the x^* values are much larger than 0. As could be expected, there seems to be some correlation in the results of populations 1, 3, 5, 7, and 9 and those of populations 2, 4, 6, 8, and 10.

Practical Considerations

If a sampling list is available prior to sampling, there is no question that $\hat{Y}_{as}$ for Sunter sampling should be preferred over $\hat{Y}_a$ for Poisson sampling. By arranging the population in decreasing order of x, x^* can usually be taken to be close to zero so that the attained sample size

Table 2.—Ratio of square root, iterated mean square errors of Poisson over Sunter sampling ratio estimators for the loblolly pine-based and Pine Creek diameter-based populations for x sorted from largest to smallest, and for two randomly arranged sequences with 2,000 iterations.

Data Set	x^*	Estimator	Population $\alpha = 0$ 1	2	3	4	$\alpha = 0$ 5	6	7	8	9	10	11
Loblolly	0.12%X	$\hat{Y}_{as}$	1.00	1.03	1.02	1.03	1.02	1.02	1.03	1.02	1.02	1.02	1.01
		$\hat{Y}_{us}$	0.98	1.02	0.99	1.01	1.01	1.01	1.01	1.01	1.01	1.01	0.99
	4.27%X	$\hat{Y}_{as}$	0.99	1.03	0.97	1.03	1.01	1.01	0.99	1.02	1.01	1.01	1.01
		$\hat{Y}_{us}$	0.69	0.73	0.60	0.63	0.72	0.74	0.67	0.69	0.74	0.76	0.63
	5.23%X	$\hat{Y}_{as}$	1.01	1.01	1.02	1.01	1.04	1.01	1.04	1.00	1.03	1.01	0.98
		$\hat{Y}_{us}$	0.70	0.67	0.61	0.58	0.72	0.70	0.67	0.63	0.74	0.70	0.58
Pine Creek	0.37%X	$\hat{Y}_{as}$	1.00	1.03	1.00	1.02	1.01	1.03	1.00	1.02	1.01	1.03	1.03
		$\hat{Y}_{us}$	0.89	0.90	0.88	0.87	0.90	0.91	0.88	0.88	0.90	0.91	0.87
	1.05%X	$\hat{Y}_{as}$	1.03	0.99	1.03	0.99	1.03	0.99	1.03	0.99	1.03	0.99	1.03
		$\hat{Y}_{us}$	0.75	0.70	0.72	0.67	0.76	0.71	0.73	0.70	0.77	0.72	0.72
	1.26%X	$\hat{Y}_{as}$	1.07	1.03	1.06	1.02	1.09	1.04	1.08	1.02	1.09	1.03	1.04
		$\hat{Y}_{us}$	0.73	0.72	0.70	0.69	0.76	0.74	0.72	0.70	0.77	0.74	0.67

will be identical to or nearly identical to the desired sample size. The simulation results show that $\hat{Y}_{as}$ for Sunter sampling will be more efficient.

When no sampling list is available, for example, in timber sales for which Poisson sampling was originally suggested in forestry (Grosenbaugh 1964), it is unlikely that Sunter sampling is a practical alternative. Implementation of this method requires the ability to compute the probability of selecting a unit which depends on the number selected and the number and sizes of units remaining to be selected in the population, as indicated by equations [5] and [6]. This would likely be too cumbersome in practice, particularly in light of the small gain in efficiency this seems to yield as indicated in table 1 for the x^* values considered.

Perhaps point-Sunter sampling might be a practical alternative to point-Poisson sampling. Point-Poisson sampling (Grosenbaugh 1971, 1979; Schreuder et al. 1984) is a highly efficient sampling strategy where trees at each point are first selected proportional to basal area (point sampling) and then a subsample of trees is selected proportional to tree height by Poisson sampling for volume measurement. In some cases, it may be desirable to select one or more sample trees per sample point. This is not guaranteed by point-Poisson sampling but might be guaranteed by implementation of some point-Sunter sampling scheme (where the subsample of trees at each point is selected by Sunter sampling). For each sample point, a list of trees could be obtained so that Sunter sampling could be implemented at those points where sample trees are desired. This would probably require a programmed computer to indicate sample trees. The results of the simulations offer some encouragement that this may also be more efficient than point-Poisson sampling.

Literature Cited

Grosenbaugh, L. R. 1964. Some suggestions for better sample-tree measurements. Society of American Foresters (1963, Boston, MA): Proceedings: 36–42.

Grosenbaugh, L. R. 1967. The gains from sample-tree selection with unequal probabilities. Journal of Forestry. 65: 203–206.

Grosenbaugh, L. R. 1971. STX1-11-71 for dendrometry or multistage 3P samples. Publication FS-277. Washington, DC: U.S. Department of Agriculture, Forest Service. 63 p.

Grosenbaugh, L. R. 1979. 3P sampling theory, examples, and rationale. Tech. Note 331. Denver, CO: U.S. Department of the Interior, Bureau of Land Management. 18 p.

Hajek, J. 1957. Some contributions to the theory of probability sampling. Bulletin International Statistical Institute. 36: 127–133.

Hajek, J. 1964. Asymptotic theory of rejective sampling with varying probabilities from a finite population. Annals of Mathematical Statistics. 35: 1491–1523.

McClure, J. P.; Schreuder, H. T.; Wilson, R. L. 1983. A comparison of volume table equations for white oak and loblolly pine. Res. Pap. SE-240. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southeastern Forest Experiment Station. 8 p.

Schreuder, H. T.; Brink, G. E.; Wilson, R. L. 1984. Alternative estimators for point-Poisson sampling. Forest Science. 30(3): 803–812.

Schreuder, H. T.; Sedransk, J.; Ware, K. D.; Hamilton, D. A. 1971. 3-P sampling and some alternatives II. Forest Science. 17: 103–118.

Schreuder, H. T.; Thomas, C. E. 1985. Efficient sampling techniques for timber sale surveys and inventory updates. Forest Science. 31(4): 857–866.

Stuart, A. 1963. Some remarks on sampling with unequal probabilities. Bulletin International Statistical Institute. 40(2): 773–779.

Sunter, A. 1977. List sequential sampling with equal or unequal probabilities without replacement. Applied Statistics. 26: 261–268.

Sunter, A. 1986. Solutions to the problem of unequal probability sampling without replacement. International Statistical Review. 54(1): 33–50.

Van Deusen, P. C. 1987. 3-P sampling and design versus model-based estimates. Canadian Journal of Forest Research. 17: 115–117.

Appendix

It is known that the variance for the Horvitz-Thompson estimator is

$$V(\hat{Y}_{us}) = \sum_{i}^{N} \frac{1-\Pi_i}{\Pi_i} Y_i^2 + \sum_{i\neq j}^{N}\sum^{N} \frac{\Pi_{ij}-\Pi_i\,\Pi_j}{\Pi_i\,\Pi_j} Y_i\,Y_j$$

$$= \sum_{i}^{N} \frac{Y_i^2}{\Pi_i} + \sum_{i\neq j}^{N}\sum^{N} \frac{\Pi_{ij}}{\Pi_i\,\Pi_j} Y_i\,Y_j - (\sum_i Y_i)^2$$

Since

$$\Pi_i = n_t\,X_i/X^* \text{ and } \Pi_{ij} = n_t(n_t-1)\,x_i\,x_j\,\gamma_i^*/X^*$$

for Sunter's sampling, we have

$$V(\hat{Y}_{us}) = \sum_{i}^{N} \frac{Y_i^2 X^*}{n_t x_i} + \sum_{i\neq j}^{N}\sum^{N} \frac{n_t-1}{n_t} \gamma_i^*\,X^* Y_i Y_j - (\sum_i^N Y_i)^2$$

$$= \sum_{i}^{N} \frac{Y_i^2 X^*}{n_t x_i} + \frac{n_t-1}{n_t} X^* (\sum_i^N \gamma_i^* Y_i Y - \sum_i^N \gamma_i^* Y_i^2) - Y^2$$

where $Y = \sum_i Y_i$.

We assume that x_i and γ_i are constants, and under equation [10]

$$E(Y_i) = \beta\,x_i \qquad E(Y_i^2) = (\sigma^2+\beta^2)\,x_i^2$$

Hence

$$V(\hat{Y}_{us}) = \sum_i \frac{(\sigma^2+\beta^2)\,x_i^2\,X^*}{n_t x_i} + \frac{n_t-1}{n_t} X^*\,\beta X \sum_i \gamma_i^* \beta\,x_i$$

$$- \frac{n_t-1}{n_t} X^* \sum_i \gamma_i^* (\sigma^2+\beta^2)\,x_i^2 - \beta^2\,X^2$$

$$= \frac{(\sigma^2+\beta^2)\,X^* X}{n_t} + \frac{n_t-1}{n_t} \beta^2 X^* X \sum_i \gamma_i^*\,x_i$$

$$- \frac{n_t-1}{n_t} (\sigma^2+\beta^2)\,X^* \sum_i \gamma_i^*\,x_i^2 - \beta^2\,X^2$$

To derive the variance of $\hat{Y}_{as}$, the conditional variance is used as follows:

$$V(\hat{Y}_{as}) = V[E(\hat{Y}_{as}|n_a)] + E[V(\hat{Y}_{as}|n_a)]$$

$$= V\left[\frac{n_e}{n_a} E(\hat{Y}_{us}|n_a)\right] + E\left[\frac{n_e^2}{n_a^2} V(\hat{Y}_{us}|n_a)\right]$$

Hence the relationship $\hat{Y}_{as} = \hat{Y}_{us}\,n_e/n_a$ is used. When n_a is fixed, we have

$$E(\hat{Y}_{us}|n_a) = Y$$

and

$$V(\hat{Y}_{us}|n_a) = \sum_{i<j}\sum (\Pi_i\Pi_j - \Pi_{ij})\,(Y_i/\Pi_i - Y_j/\Pi_j)^2$$

The above two quantities on the right hand side of the equations do not depend on n_a. So by Taylor series approximation,we have

$$V(\frac{n_e}{n_a}) = n_e^2 v\,(\frac{1}{n_a}) \doteq n_e^2 \frac{v(n_a)}{n_e^4}$$

and

$$E(\frac{n_\ell^2}{n_a^2}) \doteq n_e^2\,(\frac{1}{n_e^2} + 3\,\frac{V(n_a)}{n_e^4}),$$

where

$$v(n_a) = \Sigma\,\Pi_i\,(1-\Pi_i) - 2 \sum_{i<j}\sum (\Pi_i\Pi_j - \Pi_{ij})$$

Hence

$$V(\hat{Y}_{as}) \doteq Y^2 \frac{V(n_a)}{n_e^2} + (1+3\,\frac{v(n_a)}{n_e^2}) \sum_{i<j}\sum (\Pi_i\Pi_j - \Pi_{ij})\,(Y_i/\Pi_i - Y_j\Pi_j)^2$$

Schreuder, H. T.; Li, H. G.; Sadooghi-Alvandi, S. M. 1990. Sunter's pps without replacement sampling as an alternative to Poisson sampling. Res. Pap. RM-290. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Forest and Range Experiment Station. 6 p.

Sunter sampling is an unequal probability sampling procedure where the probability of selecting a sample unit depends on its size, the number of previous units selected, and the sizes and number of units remaining in the population. This method is compared to Poisson sampling.

Keywords: pps sampling, random sample size, joint probabilities, simulations, point-Poisson sampling

Rocky Mountains

Southwest

Great Plains

U.S. Department of Agriculture
Forest Service

Rocky Mountain Forest and Range Experiment Station

The Rocky Mountain Station is one of eight regional experiment stations, plus the Forest Products Laboratory and the Washington Office Staff, that make up the Forest Service research organization.

RESEARCH FOCUS

Research programs at the Rocky Mountain Station are coordinated with area universities and with other institutions. Many studies are conducted on a cooperative basis to accelerate solutions to problems involving range, water, wildlife and fish habitat, human and community development, timber, recreation, protection, and multiresource evaluation.

RESEARCH LOCATIONS

Research Work Units of the Rocky Mountain Station are operated in cooperation with universities in the following cities:

Albuquerque, New Mexico
Flagstaff, Arizona
Fort Collins, Colorado*
Laramie, Wyoming
Lincoln, Nebraska
Rapid City, South Dakota
Tempe, Arizona

*Station Headquarters: 240 W. Prospect Rd., Fort Collins, CO 80526