Survey Methodology, June 1993
Vol. 19, No. 1, pp. 115–121
Statistics Canada

Poisson-Poisson and Binomial-Poisson Sampling in Forestry

Z. OUYANG, H.T. SCHREUDER, T. MAX and M. WILLIAMS[1]

ABSTRACT

Binomial-Poisson and Poisson-Poisson sampling are introduced for use in forest sampling. Several estimators of the population total are discussed for these designs. Simulation comparisons of the properties of the estimators were made for three small forestry populations. A modification of the standard estimator used for Poisson sampling and a new estimator, called a modified Srivastava estimator, appear to be most efficient. The latter is unfortunately badly biased for all 3 populations.

KEY WORDS: High value timber; Volume estimation; Estimators for Poisson-Poisson sampling; Simulation comparisons; Forest sampling; Srivastava estimation.

1. INTRODUCTION

Volume estimation in forestry has been highly developed in the sense that very efficient sampling strategies are available to estimate total volume (Schreuder and Ouyang 1992). Estimating and measuring defect is often not built into these strategies since measuring defect is difficult and not economically justified in most stands. But in high value stands two-phase strategies such as Poisson-Poisson sampling may be suitable where defect is measured on trees at the second phase. To sample truck loads of logs, binomial-Poisson sampling may be a suitable sampling design.

The purpose of this article is to present the theory of binomial-Poisson and Poisson-Poisson sampling and discuss some of the properties of estimators for these designs based on simulation.

2. REVIEW OF LITERATURE

Singh and Singh (1965) developed the theory for two-phase sampling with probability proportional to size (*pps*) sampling at the second phase. Furthermore, Särndal and Swensson (1987) gave a general theory of two-phase sampling. A list of sampling units is assumed to be available at the first phase prior to sampling.

Hajek (1957) developed Poisson sampling and Grosenbaugh (1964) suggested its use for one-phase unequal probability sampling when no list is available. Poisson sampling is a scheme such that each unit in a population, say unit i, is drawn into the sample independently with probability p_i. Thus the inclusion probability of unit i is equal to p_i, and joint inclusion probability of units i and j is equal to $p_i p_j$. Binomial sampling, also often called Bernoulli sampling, is a special case of Poisson sampling when all p_i are equal.

In forest survey, Poisson sampling is often implemented as follows (Schreuder *et al.* 1968).

1. Visit the N units (say trees) in the population in any order and measure or ocularly estimate the value of a covariate x_i ($i = 1, \ldots, N$) highly correlated with the value of interest y_i ($i = 1, \ldots, N$).
2. As each x_i is observed, compare it with a random integer, δ_i, randomly selected from the range $1 \le \delta_i \le L$, where L is an integer selected prior to sampling. L is picked such that $L = X/n_e$ where X = total for the covariate in the population and n_e is the desired sample size. X is usually not known before sampling and needs to be estimated.
3. If $\delta_i \le x_i$, select the unit for the sample and measure y_i.

Implementation of this method results in a sample of size n, where $E(n) = n_e$ (if a good estimate of X was made prior to sampling). In binomial sampling all the $x_i (i = 1, \ldots, N)$ are the same (Goodman 1949).

3. SAMPLING METHODS

The United States Forest Service Region 6 (Wendall L. Jones - personal communication) uses a truck load sampling method as follows: as trucks pull up to the mill a binomial sampling technique is used to randomly select

[1] Z. Ouyang, Formerly post-doctoral fellow, Statistics Dept., Colorado State University Fort Collins, Colorado, now Research Statistician, ICI Seeds, Inc., Slater, Iowa; H.T. Schreuder, Project Leader, Multiresource Inventory Techniques Project, USDA Forest Service, Rocky Mountain Forest and Range Experiment Station, Fort Collins, Colorado; T. Max, Station Biometrician, USDA Forest Service, Pacific Northwest Experiment Station, Portland, Oregon; M. Williams, Statistician, Multiresource Inventory Techniques Project, USDA Forest Service, Rocky Mountain Forest and Range Experiment Station, Fort Collins, Colorado.

trucks to be sampled, with $p = 0.10$ say. These truck loads are measured for volume. A problem with this approach is that there are long runs of no trucks being sampled. As communicated to one of the authors, this was considered highly undesirable from a practical point of view. An alternative approach, which should decrease the frequency of long runs of no samples, and could be more efficient is to use binomial - Poisson sampling instead as follows: Apply binomial sampling with a larger p (say $p = 0.30$). The scaler visually estimates volume on the selected loads. A Poisson subsample of these loads is then selected with probability proportional to the estimated volumes and the loads selected at this phase are scaled for volume. This is binomial-Poisson sampling.

For high-value timber stands in the Pacific Northwest of the United States highly accurate estimates of net volume, that is, usable volume is often desired. Actually cutting down and destructively measuring sample trees is the most reliable method of determining net volume, *i.e.* total volume minus defective volume (Johnson and Hartman 1972). Poisson-Poisson sampling may be a good sampling design in this situation. The procedure is:

1. Select n_1 out of the N trees in the population by Poisson sampling, selecting the trees proportional to some estimate of gross volume, say x_1 = diameter at breast height squared (d^2). With Poisson sampling actual sample size is random, say n_1 where $E(n_1) = n_{e1}$. Ocularly estimate say x_2 = ocular net volume.
2. Select n_2 out of the n_1 sample trees proportional to x_2, by Poisson sampling. Here $E(n_2) = n_{e2}$ is the expected sample size at the second phase.

The n_2 sample trees are then cut and destructively measured for gross, net, and defective volume. To maintain maximum efficiency in both inventory and operations it is probably best to implement both sampling phases at once and mark the n_2 sample trees at inventory time. Ascertaining usable volume for these n_2 trees is done later either by a different crew or by carrying the sample trees into a sawmill to process them for actual wood products. Binomial-Poisson sampling is a special case of this. (If the second phase is implemented separately from the first phase then a list of sampling units is available to implement the second phase and some *pps* procedure with fixed sampling size should be used instead of Poisson sampling. This approach is usually inefficient because it requires two trips to the field location).

4. NOTATION

N = Population size (not known until sampling is completed).

n_e = Expected sample size in one-phase Poisson sampling.

n = Achieved sample size in one-phase Poisson sampling.

n_{e1} = Expected sample size of first phase in two-phase Poisson sampling.

n_1 = Achieved sample size of first phase in two-phase Poisson sampling.

n_{e2} = Expected sample size of second phase in two-phase Poisson sampling.

n_2 = Achieved sample size of second phase in two-phase Poisson sampling.

Y = Total usable volume in the population (to be estimated by two-phase sampling), $Y = \sum_{i=1}^{N} y_i$.

x_{1i} = Covariate value for tree i at phase 1, say tree diameter at breast height squared (D^2).

X_1 = $\sum_{i=1}^{N} x_{1i}$ (known after implementing the first phase in the entire population).

$\pi_i(P)$ = Probability of selecting tree i in one-phase Poisson sampling $(= n_e x_{1i}/X_1)$. If all the $\pi_i(P)$ are equal, this is one-phase binomial sampling.

π_{1i} = Probability of selecting tree i at phase 1 $(= n_{e1}x_{1i}/X_1)$.

x_{2i} = Covariate value for tree i at phase 2, say ocular estimate of net volume.

X_2 = Total amount of ocularly-estimated volume in the population (only obtained for the n_1 sample trees at the first phase so X_2 can only be estimated).

π_{2i} = Probability of selecting tree i at the second phase $(= n_{e2}x_{2i}/\sum_{i=1}^{n_1} x_{2i})$.

y_i = Value of interest for tree i (say net volume).

π_i = Probability of selecting tree i through both sampling phases $(= \pi_{1i}\pi_{2i})$.

π_i^* = Approximate probability of selecting tree i through both sampling phases $(= \pi_{1i}^*\pi_{2i}^*$ where $\pi_{1i}^* = n_1 x_{1i}/X_1$ and $\pi_{2i}^* = n_2 x_{2i}/\sum_{i=1}^{n_1} x_{2i})$.

5. THEORY

For Poisson sampling, the estimator

$$\hat{Y}_u = \sum_{i=1}^{n} y_i/\pi_i(P), \tag{1}$$

is unbiased but very inefficient and should be replaced by the following approximately unbiased estimator (Grosenbaugh 1964):

$$\hat{Y}_a = \left\{\frac{n_e}{n}\hat{Y}_u \quad \text{if} \quad n > 0\right\} \tag{2}$$
$$= 0 \quad \text{if} \quad n = 0.$$

The variance of $\hat{Y}_a$, as given in Brewer and Hanif (1983), is

$$V(\hat{Y}_a) = \sum_{i=1}^{W} \pi_i(P)\,[1 - \pi_i(P)] \left[\frac{y_i}{\pi_i(P)} - \frac{Y}{n_e}\right]^2 + p_0 Y^2,$$

where $p_0 = P(n = 0)$.

For Poisson-Poisson (PP) sampling, an estimator for Y analogous to $\hat{Y}_u$ above is the unbiased estimator

$$\hat{Y}_1 = \sum_{i=1}^{n_2} y_i/\pi_i. \qquad (3)$$

This estimator can be horribly inefficient as pointed out for $\hat{Y}_u$ in Poisson sampling (Schreuder *et al.* 1968).

The variance of $\hat{Y}_1$ can be written down by using the general formulas developed by Särndal and Swensson (1987) for unbiased estimation in double sampling:

$$V(\hat{Y}_1) = \sum_{i=1}^{N} \left(\frac{1 - \pi_{1i}}{\pi_{1i}}\right) y_i^2 + E_1\left\{\sum_{i=1}^{n_1} \left(\frac{1 - \pi_{2i}}{\pi_{2i}}\right) \left(\frac{y_i}{\pi_{1i}}\right)^2\right\},$$

where E_1 denotes expectation over the first-phase sample. Since $\hat{Y}_1$ is not efficient we do not give its variance estimator. Analogous to the more efficient adjusted estimator in Poisson sampling we have the approximately unbiased estimator

$$\hat{Y}_2 = \sum_{i=1}^{n_2} y_i/\pi_i^* = \hat{Y}_1 (n_{e1}/n_1)(n_{e2}/n_2). \qquad (4)$$

The variance of $\hat{Y}_2$ is:

$$V(\hat{Y}_2) = p(\phi)\, Y^2 + \sum_{i=1}^{N} \pi_{1i}(1 - \pi_{1i}) \left(\frac{y_i}{\pi_{1i}} - \frac{Y}{n_{e1}}\right)^2$$

$$+ \sum_{s_1 \neq \phi} p_1(s_1) \left\{\sum_{i \in s_1} \pi_{2i}(1 - \pi_{2i}) \left(\frac{y_i}{\pi_{1i}\pi_{2i}} - \frac{n_1 Y}{n_{e1} n_{e2}}\right)^2\right\},$$

where s_1 denotes the first-phase sample, $p(\phi)$ is the probability of drawing an empty sample, which is equal to

$$p(\phi) = p_1(\phi) + \sum_{s_1 \neq \phi} p_1(s_1) p_2(\phi),$$

and p_1 and p_2 denote respectively the sampling design for the first-phase and the second-phase sampling design conditional on the sample drawn in the first-phase.

Usually, population size is large and the first phase sample size is also large (compared to the second phase sample size). Thus we can safely assume $p_1(\phi) \doteq 0$ (compared to $p_2(\phi)$). For example, if we draw a first phase sample with expected sample size 50 out of a population of size 500, and then we draw a second phase sample with expected sample size 20 out of the first phase sample, all by using bionomial sampling, the inclusion probability in the first phase is 0.1 and the probability to draw an empty first phase sample is $(0.9)^{500}$; but the inclusion probability in the second phase is roughly .04 and the probability to draw an empty second phase sample is $(0.6)^{50}$. Notice that $(0.9)^{500} \doteq (0.3487)^{50} << (0.6)^{50}$ Thus, in most practical applications,

$$p_1(\phi) \doteq 0.$$

A variance estimator of $\hat{Y}_2$ can hence be easily given:

$$v_1(\hat{Y}_2) = p_2(\phi)\, \hat{Y}_2^2$$

$$+ \frac{n_{e1} n_{e2}}{n_1 n_2} \sum_{i=1}^{n_2} (1 - \pi_{1i})(y_i/\pi_{1i} - \hat{Y}_2/n_{e1})^2/\pi_{2i}$$

$$+ \frac{n_{e2}}{n_2}\left[\sum_{i=1}^{n_2} (1 - \pi_{2i}) \left(\frac{y_i}{\pi_{1i}\pi_{2i}} - \frac{n_1 \hat{Y}_2}{n_{e1} n_{e2}}\right)^2\right]. \qquad (5)$$

Estimator (5) should work well in usual applications. Sometimes when ocularly estimating net volume, however, the field worker may estimate that a tree has no value but turns out to be incorrect. Thus, some x_{2i}, hence π_{2i}, will be zero (in the simulations a small value is added to those so that $\pi_{2i} > 0$). In this case, a more stable term is needed to replace the last term in (5). Notice that

$$\frac{n_{e2}}{n_2} \sum_{i=1}^{n_2} \pi_{1i}(1 - \pi_{1i})(y_i/\pi_{1i} - \hat{Y}_2/n_{e1})^2/\pi_{2i}$$

is an improved estimator of

$$\sum_{i=1}^{n_1} \pi_{1i}(1 - \pi_{1i})(y_i/\pi_{1i} - \hat{Y}_2/n_{e1})^2. \qquad (6)$$

To ensure that the estimator does not become too large when one or more π_{2i} are close to zero, we use the following estimator

$$\left\{\left[\sum_{i=1}^{n_2} \pi_{1i}(1 - \pi_{1i})\left(\frac{y_i}{\pi_{1i}} - \frac{\hat{Y}_2}{n_{e1}}\right)^2\right]\Big/ \left[\sum_{i=1}^{n_2} \pi_{2i}\right]\right\} n_{e2}. \qquad (7)$$

If we consider x_{2i} as the auxiliary characteristic of $\pi_{1i}(1 - \pi_{1i})\ (y_i/\pi_{1i} - \hat{Y}_2/n_{e1})^2$, then (7) is a ratio estimator of (6), since $\pi_{2i} \propto x_{2i}$ for $i = 1, \ldots, n_1$. But since x_{2i} is not necessarily approximately proportional to $\pi_{1i}\ (1 - \pi_{1i})\ (y_i/\pi_{1i} - \hat{Y}_2/n_{e1})^2$, (7) may not be a very efficient estimator of (6). The advantage of using (7) is that $\sum_{i=1}^{n_2} \pi_{2i}$ will not be close to zero, so that (7) will be stable.

This leads to the following variance estimator:

$$v_2(\hat{Y}_2) = p_2(\phi)\ \hat{Y}_2^2 + \frac{n_{e1}n_{e2}}{n_1}\left[\sum_{i=1}^{n_2} (1 - \pi_{1i})\left(\frac{y_i}{\pi_{1i}} - \frac{\hat{Y}_2}{n_{e1}}\right)^2\right]\Big/ \sum_{i=1}^{n_2} \pi_{2i} + \frac{n_{e2}}{n_2}\sum_{i=1}^{n_2} (1 - \pi_{2i})\left(\frac{y_i}{\pi_{1i}\pi_{2i}} - \frac{n_1\hat{Y}_2}{n_{e1}n_{e2}}\right)^2, \qquad (8)$$

which is less affected by small probabilities than (5) and hence is more stable. We will use (8) instead of (5) as a variance estimator of $\hat{Y}_2$.

Let E_1 denote the expectation with respect to the first phase and E_2 denote the expectation with respect to the second phase. Since n_2 is the actual sample size and $E\ n_2 = E_1E_2n_2 = E_1n_{e2}$, the adjusted estimator in *PP* sampling should be $E_1n_{e2}/n_2\ \hat{Y}_1$. But the quantity E_1n_{e2} is not available and is replaced by n_{e2} to obtain the following estimator:

$$\hat{Y}_3 = \frac{n_{e2}}{n_2}\hat{Y}_1. \qquad (9)$$

$\hat{Y}_3$ should also have very small bias and the variance of $\hat{Y}_3$ is

$$V(\hat{Y}_3) = p(\phi)Y^2 + \sum_{i=1}^{N} \frac{1 - \pi_{1i}}{\pi_{1i}}y_i^2 + \sum_{s\neq\phi} p_1(s_1)\left\{\sum_{i\in S_1} \pi_{2i}(1 - \pi_{2i})\ \left(\frac{y_i}{\pi_{i1}\pi_{2i}} - \frac{Y}{n_{e2}}\right)^2\right\}.$$

A variance estimator of $\hat{Y}_3$ is

$$v(\hat{Y}_3) = p_2\ (\phi)\hat{Y}_3^2 + \frac{n_{e2}}{n_2}\left[\sum_{i=1}^{n_2} \pi_{2i}(1 - \pi_{1i})\left(\frac{y_i}{\pi_{1i}\pi_{2i}}\right)^2\right] + \sum_{i=1}^{n_2} (1 - \pi_{2i})\left(\frac{y_i}{\pi_{1i}\pi_{2i}} - \frac{\hat{Y}_3}{n_{e2}}\right)^2. \qquad (10)$$

Another possible estimator is based on the idea that we first want an efficient estimator of the first-phase information. This is accomplished by an analogous estimator to $\hat{Y}_a$ in eq. (2):

$$\hat{Y}_a(2) = \sum_{i=1}^{n_2} (y_i/\pi_{2i})n_{e2}/n_2 \quad \text{if} \quad n_2 > 0.$$

This estimator can be expanded to estimate Y by dividing the first-phase sample by its probability of selection and we obtain

$$\hat{Y}_4 = \left[\hat{Y}_a(2)\Big/\left\{\prod_{i\in s}^{n_1} p_{1i}\prod_{j\notin s}^{N} (1 - p_{1j})\right\}\right]\Big/2^{N-1}, \qquad (11)$$

where $i\in s$ indicates that unit i is in the sample, $j\notin s$ indicates that j is not in the sample, $p_{1i} = n_{e1}x_{1i}/X_1$, and 2^{N-1} is the number of all samples.

The variance of $\hat{Y}_4$ is

$$V(\hat{Y}_4) \doteq (2^{-2(N-1)})\left[\sum_{s_1\neq\phi} T(s_1)^2/p_1(s_1)\right] - Y^2 + (2^{-2(N-1)})\sum_{s_1\neq\phi}\left\{\sum_{i\in s_1} \pi_{2i}(1 - \pi_{2i}) \left[\frac{y_i}{\pi_{2i}} - \frac{1}{n_{e2}}T(s_1)\right]^2 + p_2(\phi)T(s_1)^2\right\}/p_1(s_1),$$

where $T(s_1)$ is the total of y over s_1. It can be easily derived by using the formula

$$V(\hat{Y}_4) = V_1E_2(\hat{Y}_4) + E_1V_2(\hat{Y}_4),$$

and the variance given for $\hat{Y}_a$.

This estimator is expected to be highly unstable. A possible improvement is to condition the estimator on the actual sample size obtained, *i.e.*,

$$\hat{Y}_5 = \left[\hat{Y}_a(2) \Big/ \left\{ \frac{\left[\prod_{i\in s}^{n_1} \pi_{1i} \prod_{j\notin s} (1 - \pi_{1j}) \right]}{P_1(n_1)} \cdot \binom{N-1}{n_1 - 1} \right\} \right], \quad (12)$$

where $P_1(n_1)$ is the probability of drawing a first phase sample of size n_1.

To compute this probability, let I_i be the random variable which is 1 if unit i is in the sample and 0 otherwise. Hence $n_1 = \sum_{i=1}^{N} I_i$, and

$$E(n_1) = n_{e1}, \ \mathrm{Var}(n_1) = \sum_{i=1}^{N} \pi_{1i}(1 - \pi_{1i}) = d.$$

If

$$r = \frac{n_0 - n_e}{\sqrt{d}},$$

$$\phi(r) = (2\pi)^{-1/2} \exp[-\frac{1}{2}r^2],$$

$$f_m(r) = \left[\frac{1}{\sqrt{d}}\right] \phi(r) \left[1 + \sum_{j=1}^{m} p_j(r)\right],$$

where $P_j(r)$ are Edgeworth polynomials. Then

$P_1(n_1) \doteq f_{m1}(r)$ and specifically, for $m = 2$

$$P_1(n_1) \doteq f_2(r) = \left[\frac{1}{\sqrt{d}}\right] \phi(r) \left[1 + \frac{1 - 2\bar{\bar{\pi}}}{6\sqrt{d}} (r^3 - 3r) + \frac{1}{4!} \frac{1 - \overline{6\pi(1-\pi)}}{d} (r^4 - 6r^2 + 3) + \frac{10}{6!} \frac{(1 - 2\bar{\bar{\pi}})^2}{d} (r^6 - 15r^4 + 45r^2 - 15)\right],$$

where

$$\bar{\bar{\pi}} = \frac{\sum_{i=1}^{N} \pi_i^2(1 - \pi_i)}{\sum_{i=1}^{N} \pi_i(1 - \pi_i)}, \ \overline{\pi(1 - \pi)} = \frac{\sum_{i=1}^{N} \pi_{1i}^2(1 - \pi_{1i})^2}{\sum_{i=1}^{N} \pi_{1i}(1 - \pi_{1i})}$$

(Hájek 1981).

$\hat{Y}_4$ and $\hat{Y}_5$ are only given for completeness. They are not considered further since both are unstable.

An alternative to $\hat{Y}_4$ and $\hat{Y}_5$ is to correct $\hat{Y}_a(2)$ using an expansion factor based on the information for covariate x_1. These estimators are sensible if $\hat{Y}_a(2)/\sum_{i=1}^{n_1} x_{1i}$ is an approximately unbiased estimator of $R = Y/X_1$ which is true for binomial-Poisson (BP) but not for PP sampling. This fact is verified by simulation, but the reason why approximate unbiasedness holds for binomial-Poisson is that $\hat{Y}_a(2)/\sum_{i=1}^{n_1} x_{1i}$ under binomial sampling is similar to the ratio estimator under simple random sampling. Hence the following estimator is only appropriate for *BP* sampling.

$$\hat{Y}_6 = X_1 \left[\hat{Y}_a(2) \Big/ \sum_{i=1}^{n_1} x_{1i} \right]. \quad (13)$$

The variance of $\hat{Y}_6$ is

$$V(\hat{Y}_6) = \frac{N^2}{n_{e1}^2} \sum_{i=1}^{N} (y_i - Rx_i)^2 \pi_{1i} + E_1 \left\{ \frac{x_i}{n_1 \bar{x}_1 s_1} \left[\sum_{i=1}^{n_1} \pi_{2i}(1 - \pi_{2i}) \left(\frac{y_i}{\pi_{2i}} - \frac{n\bar{y}_{s1}}{n_{e2}} \right)^2 + p_2(\phi)\, n_1^2 \bar{y}_{s1}^2 \right] \right\}.$$

Another promising estimator is based on Srivastava's (1985) proposed unbiased estimator $\hat{Y}_{sr1}$ based on the sample weight function concept. Srivastava and Ouyang (1992) developed a structure for the sample weight in order that $\hat{Y}_{sr1}$ has zero variance at some points of the parameter space $\{y_1, \ldots, y_N\}$. The sample weight function can use any information other than that given in a sample. Examples of this kind of information have been given in Srivastava and Ouyang (1992) and Ouyang and Schreuder (1992). If the information can be formulated as a model

$$y_i = \alpha + \beta x_i + e_i, \ i = 1, \ldots, N, \quad (14)$$

then the so called "generalized ratio estimator approximation" (Ouyang *et al.* (1992)) can be used which gives the following estimator of the population total:

$$\hat{Y}_7 = \left[\frac{\hat{Y}_1}{\sum_{i=1}^{n_2} y_i^* / (\pi_{1i}\pi_{2i})} \right] Y^*, \qquad (15)$$

with $\hat{\alpha}$ and $\hat{\beta}$ weighted regression coefficients, and y_i^* calculated by $y_i^* = \hat{\alpha} + \hat{\beta}x_{1i}$ and $Y^* = \sum_{i=1}^{N} y_i^*$.

Note that $\hat{Y}_7$ is dependent on the model assumption.

6. SIMULATIONS

Simulation samples with first-and second-phase samples of expected sizes 50 and 20 in Poisson-Poisson and binomial-Poisson sampling were each drawn from three populations. Two populations were high-value fir, cedar and pine trees. Population 1, called BLM1 (Data from unpublished report "Comparison of volume estimates made by several timber measurement methods in western Oregon" by G. B. Hartman. Feb., 1971. Bureau of Land Management, Portland, Oregon), contained 331 trees and population 2, called BLM2, included 510 trees (Johnson and Hartman 1972). Measured variables on each tree were: net volume scaled (nvs), net volume dendrometered (nvd), and diameter at breast height (d). Here nvs $(= y)$ is the variable of interest, $x_1 = d^2$ is used in the first phase of PP sampling and x_2 = nvd is the more expensively but presumably additionally useful covariate obtained at the second level of PP sampling; 200,000 simulations were performed. Ideally, one would like the first- and second-level covariates to be relatively uncorrelated yet both highly correlated with y. These would be d^2 or nvd at the first phase and some measure of defect at the second-phase. Unfortunately, to do this in a satisfactory manner requires separating trees into a class where the field worker is comfortable estimating defect and another class for which he does not. This was not done for the available data. In BP sampling trees were selected with equal probabilities at the first phase and proportional to x_2 at the second phase. Population 3, a mapped data set, called Surinam, was also used since it was cleaner than the other populations in terms of having available more sensible variables for Poisson-Poisson sampling. The population consists of a 60-ha mapped Surinam forest for which only species and diameters were recorded (Schreuder *et al.* 1987). Tree heights and standing tree volumes for other species were superimposed on these trees as described in Schreuder *et al.* (1992). The resulting population consists of 5,525 trees for which tree diameter (d), height (h) and volume (v) were available. This yielded covariates $x_1 = h_h^2$ and x_2 = standing gross tree volume for PP sampling. For BP sampling x_2 was used at the second phase. Board foot volume (y) was also added to the data set. Included are 10 trees for which d^2h is large ($\geq$ 60,000) but bd. ft. volume is essentially zero; 10,000 simulations were performed for the Surinam data. Results for BLM1, BLM2, and Surinam are given in Tables 1, 2 and 3 respectively.

Table 1

Simulation results for BLM1 (N = 331) population. 200,000 simulations were performed using $x_1 = D^2$ and x_2 = nvd as covariates*

Estimator	Bias		SE		EASE	
	BP	PP	BP	PP	BP	PP
$\hat{Y}_1$	0.021	0.011	42.495	53.228		
$\hat{Y}_2$	−0.045	−0.770	37.272	48.219	97.787	97.806
$\hat{Y}_3$	−0.050	−0.777	39.819	49.349	97.492	96.763
$\hat{Y}_6$	0.012		39.992			
$\hat{Y}_7$	−0.036	3.650	18.881	21.885		

Table 2

Simulation results for BLM2 (N = 510) population. 200,000 simulations were performed using $x_1 = D^2$ and x_2 = nvd as covariates*

Estimator	Bias		SE		EASE	
	BP	PP	BP	PP	BP	PP
$\hat{Y}_1$	0.146	0.059	95.708	62.500		
$\hat{Y}_2$	0.055	−0.424	90.247	55.876	100.325	98.583
$\hat{Y}_3$	0.050	−0.411	91.259	58.701	99.779	98.679
$\hat{Y}_6$	0.146		94.100			
$\hat{Y}_7$	0.486	4.391	26.788	19.855		

Table 3

Simulation results for Surinam (N = 5,525) population. 10,000 simulations were performed using $x_1 = D^2$ and x_2 = ocular estimate of net volume*

Estimator	Bias		SE		EASE	
	BP	PP	BP	PP	BP	PP
$\hat{Y}_1$	0.764	0.364	25.709	25.924		
$\hat{Y}_2$	0.290	−0.402	15.636	10.845	97.492	97.37
$\hat{Y}_3$	0.019	−0.463	20.989	17.886	100.364	98.945
$\hat{Y}_6$	1.013		20.822			
$\hat{Y}_7$	2.277	2.426	22.428	17.397		

* All tables give bias and standard error (SE) expressed as a percentage of the population net volume. The estimated average standard error (EASE) is expressed as a percentage of the simulation standard error. Expected sample sizes are n_{e1} = 50 and n_{e2} = 20 for both binomial-Poisson (BP) and Poisson-Poisson (PP) sampling.

7. RESULTS AND DISCUSSION

For PP sampling $\hat{Y}_2$ is the most efficient estimator of the three ($\hat{Y}_1$, $\hat{Y}_2$, and $\hat{Y}_3$) relatively assumption-free estimators for BLM1 and BLM2; $\hat{Y}_3$ is slightly less efficient than $\hat{Y}_2$. Note that $\hat{Y}_7$ is even more efficient than $\hat{Y}_2$ but $\hat{Y}_7$ has a serious bias in some cases. The variance estimators for $\hat{Y}_2$ and $\hat{Y}_3$, $v(\hat{Y}_2)$ and $v(\hat{Y}_3)$, in eq. (8) and (10) are approximately unbiased.

For BP sampling, $\hat{Y}_7$ has negligible bias and the smallest standard error of all the estimators. $\hat{Y}_2$ is considerably less efficient than $\hat{Y}_7$ for BLM1 and BLM2 but more efficient than the other estimators. The variance estimators for both $\hat{Y}_2$ and $\hat{Y}_3$ are approximately unbiased.

Note for BLM1, BP sampling is always more efficient than PP sampling whereas for BLM2 PP sampling is more efficient with $\hat{Y}_1$, $\hat{Y}_2$ and $\hat{Y}_3$. This is because x_2 is not the logical variable to measure after the effect of x_1 is removed. Unfortunately a better variable to assess defect was not available for these data. For BLM1 x_2 did not but for BLM2 it did improve estimation.

For both PP and BP sampling, using population Surinam, $\hat{Y}_2$ is again the most efficient estimator of the three ($\hat{Y}_1$, $\hat{Y}_2$ and $\hat{Y}_3$) relatively assumption-free estimators. $\hat{Y}_3$ is considerably less efficient than $\hat{Y}_2$. $\hat{Y}_7$ is less efficient than $\hat{Y}_2$ and is substantially more biased for this population. $v(\hat{Y}_2)$ and $v(\hat{Y}_3)$ seems to be a approximately unbiased variance estimators for $\hat{Y}_2$ and $\hat{Y}_3$. For this population PP sampling is more efficient than BP sampling with $\hat{Y}_2$ showing that in this case both $x_1 = d^2h$ and x_2 = standing gross total volume are useful in sampling.

Actually, it is not surprising to see $\hat{Y}_2$ is the most efficient estimator, since it uses the most amount of information at both the design and estimation stages. Estimator $\hat{Y}_7$ tends to be even more efficient in terms of mean squared error, but with larger bias. This is because $\hat{Y}_7$ is based on the model given in equation (14). If the model is correct, $\hat{Y}_7$ should be preferred over $\hat{Y}_2$, since $\hat{Y}_7$ incorporates even more information from the population. But otherwise, $\hat{Y}_2$ should be preferred. $\hat{Y}_7$ is not recommended if model (14) is not justified.

8. RECOMMENDATIONS

1. Both Poisson-Poisson and binomial-Poisson sampling are useful in practical forest sampling. With either procedure, estimator $\hat{Y}_2$ should be used. This estimator, with negligible bias and high efficiency, is analogous to the adjusted estimator $\hat{Y}_a$ used in Poisson sampling and has a reliable variance estimator.
2. Estimator $\hat{Y}_7$ is considerably more efficient than $\hat{Y}_2$ for 2 populations but should not be used in preference to $\hat{Y}_2$ until it has been more fully investigated in additional studies. $\hat{Y}_7$ tends to be seriously biased in these simulations.

ACKNOWLEDGEMENT

We appreciate valuable comments by a referee.

REFERENCES

BREWER, K.R.W., and HANIF, M. (1983). *Sampling with Unequal Probabilities*. New York: Springer-Verlag.

GOODMAN, L.A. (1949). On the estimation of the number of classes in a population. *Annals Mathematical Statistics*, 20, 572-579.

GROSENBAUGH, L.R. (1964). Some suggestions for better sample-tree measurement. *Society of American Foresters. Proceedings*, 36-42.

HAJAK, J. (1957). Some contributions to the theory of probability sampling. *Bulletin of the International Statistical Institute*, 36,127-133.

JOHNSON, F.A., and HARTMAN, G.B. (1972). Fall, buck and scale cruising. *Journal of Forestry*, 566-568.

OUYANG, Z. (1990). Investigation of some estimators and strategies in sampling proposed by Srivastava. PhD thesis. Colorado State University Fort Collins, CO, 83.

OUYANG, Z., and SCHREUDER, H.T. (1992). Srivastava estimation in forestry. Submitted to *Forest Science*.

OUYANG, Z., SRIVASTAVA, J.N., and SCHREUDER, H.T. (1992). A general ratio estimator and its application in model based inference. *Annals. Institute of Statistical Mathematics*, (in press).

SÄRNDAL, C.-E., and SWENSSON, B. (1987). A general view of estimation for two phases of selection with applications to two-phase sampling and nonresponse. *International Statistical Review*, 55, 279-294.

SCHREUDER, H.T., SEDRANSK, J., and WARE, K.D. (1968). Sampling and some alternatives, I. *Forest Science*, 14, 429–454.

SCHREUDER, H.T., BANYARD, S.C., and BRINK, G.E. (1987). Comparison of three sampling methods in estimating stand parameters for a tropical forest. *Forest Ecology and Management*, 21, 119-128.

SCHREUDER, H.T., and OUYANG, Z. (1992). Optimal sampling strategies for weighted linear regression estimation. *Canadian Journal of Forest Research*, 22, 239-247.

SCHREUDER, H.T., OUYANG, Z., and WILLIAMS, M. (1992). Point-Poisson, point-pps, and modified point-pps sampling: Efficiency and variance estimation. *Canadian Journal of Forest Research*, (in press).

SINGH, D., and SINGH, B.D. (1965). Some contributions to two-phase sampling. *Australian Journal of Statistics*, 7, 45-47.

SRIVASTAVA, J.N. (1985). On a general theory of sampling, using experimental design. Concepts I: Estimation. *Bulletin of the International Statistical Institute*, 51, 1-16.

SRIVASTAVA, J.N., and OUYANG, Z. (1992). Studies on a general estimator in sampling, utilizing extraveous information through a sampling weight function. *Journal of Statistical Planning and Inference*, 31, 177-196.