Journal of Official Statistics, Vol. 19, No. 1, 2003, pp. 59–70

Poisson Mixture Sampling Combined with Order Sampling

Hannu Kröger,[1] *Carl-Erik Särndal*[2]*, and Ismo Teikari*[3]

The term *Poisson Mixture* (PoMix) *sampling* refers to a family of sampling designs based on the Permanent Random Number (PRN) technique and useful for sampling highly skewed populations, such as those arising in many business surveys. Traditional Poisson πps sampling is a special case of PoMix sampling, but some PoMix designs are considerably more efficient than Poisson πps. When used with common estimators, some PoMix designs can lead to a considerably lower variance than Poisson πps. PoMix sampling gives a random sample size, regarded by some as a disadvantage. Therefore, we create in this article a family of *fixed size PoMix* designs, by using the central idea in *order sampling*: The population units are ordered by a ranking variable, and the sample consists of the n units with the smallest ranking variable values. This article reports results of a Monte Carlo simulation, where fixed size PoMix sampling is found to outperform other fixed size πps designs, and where (less surprisingly) regression and ratio estimators outperform the Horvitz-Thompson estimator. We show that the variance advantage of PoMix sampling is explained by a pronounced population skewness combined with a mildly heteroscedastic variance around the linear regression line.

Key words: Business surveys; permanent random numbers; skewed populations; fixed sample size.

1. Poisson Mixture Sampling

Poisson Mixture (PoMix) sampling, introduced in Kröger, Särndal and Teikari (1999), consists of a family of designs useful for sampling skewed populations, such as those often encountered in business surveys. Every PoMix sampling design can be viewed as a mixture of two traditional Poisson sampling schemes: Bernoulli sampling (defined as Poisson sampling with a constant inclusion probably for all units) and Poisson πps sampling (defined as Poisson sampling with inclusion probabilities strictly proportional to a size measure x_k). PoMix sampling can be carried out as Poisson sampling with a set of inclusion probabilities defined as a mixture, or a linear combination, of constant probabilities and probabilities proportional to size. An attractive feature of a well-chosen PoMix design is its ability to produce considerably more precise estimates than Poisson πps sampling, even though Poisson πps is itself highly efficient, at least in the presence of a strongly correlated measure of size.

[1] Lohjantie 16A4, SF-03100 Nummela, Finland. E-mail: hannu.kroger@eds.com
[2] 2115 Erinbrook Crescent, No. 44, Ottawa, Ontario K1B 4J5, Canada. Email: carl.sarndal@rogers.com
[3] Statistics Finland, P.O. Box 3A, FIN-00022 Statistics Finland, Finland. Email: ismo.teikari@stat.fi
Acknowledgment: The authors gratefully acknowldge the cooperation of an Associate Editor and two anonymous referees. Their comments helped improve the quality of the manuscript.

PoMix sampling is based on the Permanent Random Number (PRN) technique, which offers the survey statistician a range of possibilities for sample coordination and control of response burden. A PRN is a uniformly distributed random number attached to a population unit at birth. Important early references to PRN methodology are Brewer, Early, and Joyce (1972), and Atmer, Thulin, and Bäcklund (1975). Important new developments of PRN techniques are known under the collective term of order sampling. Some recent references are Ohlsson (1995, 1998), Saavedra (1995), Rosén (1997a, 1997b), Aires (1999, 2000), and Holmberg and Swensson (2001).

We assume that the sampling frame lists the N units of the target population $U = \{1, \ldots, k, \ldots, N\}$. To each unit is attached a PRN. Denote by y the variable of interest and by y_k its value for unit k. We wish to estimate the population y-total $Y = \sum_U y_k$ on the basis of a sample s drawn from U by a PoMix sampling design. (For any set C of units, $C \subseteq U$, $\sum_C y_k$ will be used as shorthand for $\sum_{k \in C} y_k$.) PoMix sampling produces a sample s of random size. Denote by n the expected size of s. The PoMix family of designs is indexed by a continuous parameter, B, the *Bernoulli width*, which satisfies $0 \leq B \leq f$, where $f = n/N$ is the predetermined expected sampling fraction. Poisson πps sampling is obtained for $B = 0$, and Bernoulli sampling for $B = f$. Let x_k be the positive size measure for unit k, for example, the number of persons employed in enterprise k. Define the *relative size* of k by $A_k = nx_k / \sum_U x_k$. We assume that $A_k < 1$ holds for all $k \in U$. If the specified expected size n is too large for this to hold, we set aside large units as a stratum of units selected with certainty, a ''take-all stratum,'' up to a point where the recomputed relative size is less than unity for all remaining units. In the expression $A_k = nx_k / \sum_U x_k$, U and n will then refer, respectively, to the set of the remaining units and the expected size of the remaining sample of units, drawn with inclusion probabilities strictly less than unity.

As originally presented in Kröger, Särndal, and Teikari (1999), PoMix is a random sample size design defined as follows. Fix a value f for the desired sampling rate and a value B for the Bernoulli width; $0 \leq B \leq f$. The choice of B is discussed later. For $k = 1, \ldots, N$, define and compute

$$\pi_k = \Pr(k \in s) = B + (1 - Q)A_k \tag{1.1}$$

where $Q = B/f$. Then carry out N independent Bernoulli experiments, one for each unit, giving the kth unit the probability $\pi_k = B + (1 - Q)A_k$ of ''success'' (= selection), $k = 1, \ldots, N$. The resulting sample size is then random with the expected value $\sum_U \pi_k = n$, for any value of $B \in [0, f]$. It also follows that when B is chosen as distinctly larger than zero, no unit will end up having an inclusion probability extremely close to zero, with the advantage that excessively large weights $1/\pi_k$ are avoided. The term ''Poisson Mixture'' is appropriate considering that the inclusion probability (1.1) can be written alternatively as the linear combination $\pi_k = Q\pi_k^{BE} + (1 - Q)\pi_k^{\pi\text{ps}}$, where $Q = B/f$ is the mixing proportion, $\pi_k^{BE} = f$ for all $k \in U$ as in Bernoulli sampling (which is a special case of Poisson sampling), and $\pi_k^{\pi\text{ps}} = A_k = nx_k / \sum_U x_k$ as in Poisson πps.

2. Questions Arising About PoMix Sampling

Kröger, Särndal, and Teikari (1999) studied the performance of several ratio type estimators under PoMix sampling and found that if the PoMix parameter B in (1.1) is fixed at a

value distinctly larger than zero, there is often a considerable variance reduction compared to Poisson πps ($B = 0$). We shall refer to this as the *variance advantage* of PoMix sampling. For the quite skewed Finnish business survey data used in the study, ratio type estimators gave a variance advantage of the order of 50%, relative to $B = 0$, realized for a value of B roughly equal to $0.3f$. The simulation also included the Horvitz-Thompson (HT) estimator, which is of more limited interest, because its variance is considerably larger than that of the ratio estimators, confirming the rule stating that in order to profit fully from the available auxiliary information, one should use it both at the sampling stage (in the sampling design) and at the estimation stage (in the estimator formula).

It is well known that a random sample size, which is a feature of the PoMix sampling just described, can substantially increase the variance of the HT estimator, compared to when this estimator is used with a fixed sample size design. It is also known that there is essentially no such penalty for the Generalized Regression (GREG) family of estimators, which includes the ratio estimators. Nevertheless, some users prefer a sampling procedure guaranteeing a fixed sample size, even though the ultimate sample size is still unpredictable because of the nonresponse that will almost certainly affect the survey. The study in Kröger, Särndal, and Teikari (1999) generated some new issues. The recent work on ordered sampling designs by Ohlsson (1995, 1998), Rosén (1997a, 1997b) and others suggests that it is possible to construct a fixed size variety of PoMix sampling starting from the PoMix inclusion probabilities given by (1.1), or, putting it differently, to create an order sampling design giving approximately the target inclusion probabilities (1.1). The following questions arise:

(i) For a fixed size variety of PoMix sampling, will the variance advantage found for random size PoMix be preserved? Will GREG estimators continue to improve on the HT estimator if the latter is no longer handicapped by a random sample size?

(ii) If we use a regression estimator (which presupposes a nonzero intercept) rather than a ratio estimator (which assumes a regression through the origin), will the variance advantage of PoMix sampling persist? Put differently, if available information about the population size, N, is also incorporated into the estimator formula, will the advantage persist? This is a valid question because it could be argued that if all available information, including N, is exploited at the estimation stage, then PoMix sampling might not offer any variance advantage.

(iii) What is the effect of a pronounced population skewness on the efficiency of PoMix sampling? Is it true that its variance advantage is greater for highly skewed populations? One may suspect this to be the case because a high skewness may cause a high incidence of very small units having extremely large sampling weights under Poisson πps. A remedy would be to impose a lower bound on the inclusion probabilities, as is done in PoMix sampling when B is distinctly larger than 0.

Several possibilities may exist for imposing a fixed size requirement on PoMix sampling. The device proposed in Section 4 brings in the principal idea behind *order sampling*, as presented in Rosén (1997a, b) and Ohlsson (1995). We combine PoMix sampling (to get the variance advantage) with order sampling (to get a fixed sample size). One can safely predict that the fixed size feature will improve the efficiency of the HT estimator

while leaving essentially unchanged the efficiency of a GREG estimator. With this point of departure, our objective is to throw light on Questions (i) to (iii). The article is arranged as follows: Section 3 reviews the main ideas of order sampling. Section 4 introduces fixed size PoMix sampling. Section 5 presents the estimators used in our simulation, carried out on six artificially generated populations described in Section 6. Simulation results and answers to Questions (i) to (iii) are given in the concluding Section 7.

3. Order Sampling

The central idea in order sampling is to compute a ranking variable for each population unit and then to let the sample be defined by the n smallest ranking variable values. By definition, the sample size is then the same for all possible samples. The term order sampling comes from Rosén (1997a, 1997b); an important special case had been considered by Ohlsson (1995). Order sampling is defined by the following three-step algorithm, referred to also in later sections:

(i) Compute the ranking variable value for unit k, denoted ξ_k, $k = 1, \ldots, N$;
(ii) Sort the N units by the size of ξ_k, from the smallest to the largest;
(iii) Define the sample to consist of the first n units in the sorted list.

The ranking variable value for unit k is a function of its PRN and its size measure. Several ranking variables are possible. Ohlsson (1995) used the ranking variable $\xi_k = \xi_{1k}$, where $\xi_{1k} = r_k/A_k$; he termed the procedure Sequential Poisson sampling. Rosén (1997a, b) proposed the ranking variable $\xi_k = \xi_{2k}$, where $\xi_{2k} = [r_k/(1 - r_k)][A_k/(1 - A_k)]^{-1}$ and termed the resulting procedure Pareto πps sampling. He was led to it by minimizing an approximation to the variance of the HT estimator. Consequently we expect the ranking variables ξ_{2k} to give a smaller variance for the HT estimator than the ξ_{1k}. (This result cannot be assumed to hold for a GREG estimator, if different from the HT estimator.) For the HT estimator, Rosén (1997b) compared Pareto πps with Sequential Poisson in a variety of cases, showing that the variance reduction is insignificant for a sampling fraction of 0.1; for a sampling fraction of 0.3, it ranged up to a few percent, but remained modest at usually less than 3%.

The quantity A_k appearing in the expressions for ξ_{1k} and ξ_{2k} is now called the *target inclusion probability* for unit k. It differs somewhat from the realized inclusion probability. But Rosén (1997a, b) showed that for various order sampling schemes, including Pareto πps and Sequential Poisson, the realized inclusion probabilities π_k are very close to the targeted ones, that is, $\pi_k \approx A_k$ with excellent approximation. Our simulations, carried out under the extensions of the technique spelled out in Section 4 below, leave no reason to believe otherwise. The closeness of the actual inclusion probabilities of Pareto πps to their target values has been studied, for example, in Aires (1999), for small and moderate sample sizes.

4. Combining PoMix Sampling and Order Sampling

In this section we define *fixed size PoMix sampling*, using the ranking idea in order sampling. We define a new ranking variable ξ_k as a function of the PRN, r_k, and of the modified relative size defined by $A_{\mathrm{mod},k} = B + (1 - Q)A_k$. Recall that $A_{\mathrm{mod},k}$ is the *exact*

inclusion probability of k under PoMix sampling. For fixed size PoMix, now to be defined, $A_{\mathrm{mod},k}$ becomes the *target* inclusion probability of k.

Fixed size PoMix sampling (Sequential Poisson variety) is defined as follows: Step (i) of the three-step order sampling algorithm in Section 3 is carried out with the ranking variable values $\xi_k = \xi_{1\,\mathrm{mod},k}$, $k = 1, \ldots, N$; where $\xi_{1\,\mathrm{mod},k} = r_k/A_{\mathrm{mod},k}$. Steps (ii) and (iii) of the algorithm are as before. This scheme was used for the simulations reported in Section 6. There are two special cases of interest: (1) Sequential Poisson sampling is obtained for $B = 0$; (2) Simple Random Sampling Without Replacement (SRS) is obtained for $B = f$.

We can obtain a fixed size PoMix sampling (Pareto πps variety) by a corresponding modification of the ranking variable for Pareto πps. The ranking variable is then defined as $\xi_k = \xi_{2\,\mathrm{mod},k}$, where $\xi_{2\,\mathrm{mod},k} = [r_k/(1 - r_k)][A_{\mathrm{mod},k}/(1 - A_{\mathrm{mod},k})]^{-1}$. This alternative was also studied in our simulations, but since there were no appreciable gains in efficiency – and sometimes small losses – the results for this option are not shown in Section 7. The conclusion is not surprising, considering that the advantage that Pareto πps may have for the HT estimator is not automatically transferable to a GREG estimator. Pareto πps is the special case obtained for $B = 0$.

Fixed size PoMix can also be carried out with an arbitrary starting point D in the unit interval. Then, in the definitions of $\xi_{1\,\mathrm{mod},k}$ and $\xi_{2\,\mathrm{mod},k}$, replace r_k by $r_k^o = r_k - D$ if $D < r_k \leq 1$ and by $r_k^o = r_k - D + 1$ if $0 \leq r_k \leq D$. The rest of the algorithm is as before.

5. Estimators Included in the Simulation Study

Our simulation included the Horvitz-Thompson estimator,

$$\hat{Y}_{HT} = \sum_s a_k y_k \tag{5.1}$$

with $a_k = 1/\pi_k$. Its main function is to provide a benchmark with which to compare the better alternatives belonging to the GREG family of estimators, which can be written in the general form

$$\hat{Y}_{GREG} = \sum_s w_k y_k \tag{5.2}$$

where $w_k = a_k g_k$, with $g_k = 1 + \lambda_s' \mathbf{x}_k/c_k$ where $\lambda_s' = (\mathbf{X} - \hat{\mathbf{X}}_{HT})'(\sum_s a_k \mathbf{x}_k \mathbf{x}_k'/c_k)^{-1}$ and $\hat{\mathbf{X}}_{HT} = \sum_s a_k \mathbf{x}_k$, where $\mathbf{x}_k$ is the vector of auxiliary variables whose population total $\mathbf{X} = \sum_U \mathbf{x}_k$ is assumed known. The c_k are specified constants. A standard choice is $c_k = 1$ for all k; other choices have only a mild effect on the variance of $\hat{Y}_{GREG}$. Equivalently, (5.2) can be written in the ''regression form,'' that is, we have $\hat{Y}_{GREG} = \hat{Y}_{HT} + (\mathbf{X} - \hat{\mathbf{X}}_{HT})'\mathbf{b}$, where the term $(\mathbf{X} - \hat{\mathbf{X}}_{HT})'\mathbf{b}$ with $\mathbf{b} = (\sum_s a_k \mathbf{x}_k \mathbf{x}_k'/c_k)^{-1} \sum_s a_k \mathbf{x}_k y_k/c_k$ represents a regression adjustment to the HT estimator. As a result, $\hat{Y}_{GREG}$ has a considerably smaller variance than $\hat{Y}_{HT}$ whenever the regression of y on $\mathbf{x}$ is strong.

In the simulations reported in Sections 6 and 7, we used a unidimensional, continuous x-variable. The $\mathbf{x}_k$-vector is therefore either $\mathbf{x}_k = (1, x_k)'$, corresponding to a regression with intercept, or $\mathbf{x}_k = x_k$, corresponding to a regression without intercept. In the first case, with $\mathbf{x}_k = (1, x_k)'$ and $c_k = 1$ for all k, (5.2) gives the regression estimator

$$\hat{Y}_{REG} = N\tilde{y}_s + (X - N\tilde{x}_s)b \tag{5.3}$$

with $X = \sum_U x_k$ and

$$b = \frac{\sum_s a_k(y_k - \tilde{y}_s)(x_k - \tilde{x}_s)}{\sum_s a_k(x_k - \tilde{x}_s)^2}; \quad \tilde{x}_s = \frac{\sum_s a_k x_k}{\sum_s a_k}; \quad \tilde{y}_s = \frac{\sum_s a_k y_k}{\sum_s a_k}$$

In the second case, letting $\mathbf{x}_k = c_k = x_k$, (5.2) gives the ratio estimator

$$\hat{Y}_{RA} = X \frac{\sum_s a_k y_k}{\sum_s a_k x_k} \tag{5.4}$$

In the literature one can find support for other c_k-weights than those just mentioned (see, for example, the discussion in Brewer 1999 and Särndal 1996). The latter reference recommends $c_k = 1/(1 - a_k)$; the estimators with these weights (and $\mathbf{x}_k = (1, x_k)'$ or $\mathbf{x}_k = x_k$ as earlier) were also tried in the simulation, but led to no significant variance reduction compared to (5.3) and (5.4).

6. Description of the Simulation

Two factors that affect the variance advantage of PoMix sampling and need to be studied are: (i) the skewness of the population, and (ii) the residual variance pattern around the regression line of y on x. We examined the three estimators, $\hat{Y}_{REG}$, $\hat{Y}_{RA}$ and $\hat{Y}_{HT}$, under fixed size PoMix sampling from six artificially created populations representing different conditions of skewness and residual variance. Each population is of size $N = 1{,}000$ and is generated as follows. First, create 1,000 x_k-values according to a specified distribution shape. Then, for each given x_k, generate y_k in such a way that the regression of y on x is linear through the origin, $k = 1, \ldots, 1{,}000$. Two different Weibull distributions were used, resulting in two sets of x_k-values. Three different values were used for the parameter, denoted p, that determines the heteroscedastic residual variance pattern around a linear regression of y on x. This gave a total of $2 \times 3 = 6$ populations, each consisting of 1,000 pairs (x_k, y_k).

The Weibull distribution function with parameters $\alpha > 0$, $\gamma > 0$ is given by $F(x) = 1 - \exp(-\gamma x^{\alpha})$ for $x > 0$. We worked with the cases $\alpha = 1, \gamma = 2$ (an exponential distribution, thus rather skewed) and $\alpha = \frac{1}{2}, \gamma = 2$ (more highly skewed). The expected value of x equals $\frac{1}{2}$ in both cases; the skewness equals 2.0 for the exponential and 6.6 for the more highly skewed distribution. (The measure of skewness is defined as the third central moment divided by (variance)$^{3/2}$.) For each of these two cases, we used the inverse of the distribution function to obtain 1,000 ''systematically spaced'' x_k-values, as $x_k = [-\ln(1 - P_k)/\gamma]^{1/\alpha}$ where $P_k = (k - 0.5)/N$; $k = 1, \ldots, N = 1{,}000$.

Next, given x_k, we created the corresponding value $y_k, k = 1, \ldots, 1{,}000$, as a realization of the Gamma (a, b) distribution with the density $f(y) = [\Gamma(b)a^b]^{-1} y^{b-1} \exp(-y/a)$ for $y > 0$, where a and b are specified so that y_k conditionally on x_k has the expected value βx_k and the variance $\sigma^2 x_k^p$, for a specified exponent p, and suitably chosen values of β and σ^2. For unit k, this was realized letting $a = (\sigma^2/\beta)x_k^{p-1}$; $b = (\beta^2/\sigma^2)x_k^{2-p}$, because this choice gives the desired properties $\mathrm{E}(y_k|x_k) = ab = \beta x_k$ and $\mathrm{Var}(y_k|x_k) = a^2 b = \sigma^2 x_k^p$. We used the three values $p = 1$, $p = 1.5$ and $p = 2$. We fixed $\beta = 2$ as a common theoretical regression slope for all six populations, whereas the value of σ^2 was adjusted so that the theoretical coefficient of correlation between x and y is always 0.90. We thus obtained six sets of points (x_k, y_k), $k = 1, \ldots, 1{,}000$, which share the following

characteristics: The mean of x is roughly $\frac{1}{2}$, the mean of y is roughly 1, the computed regression line of y on x is approximately $y = 2x$, the correlation coefficient computed on the 1,000 realized points (x_k, y_k) is close to 0.90. The six sets differ in regard to such aspects as the variance of y and the heteroscedastic variance pattern around the regression line.

We drew 10,000 repeated fixed size PoMix samples, for each of a number of different values of $B \in [0, f]$. For the exponential case ($\alpha = 1$), the size of each sample was $n = 100$ out of $N = 1{,}000$, for a sampling fraction of $f = 0.10$. For the more skewed case ($\alpha = \frac{1}{2}$), the sample size was $n = 87$ out of $N = 987$, for a sampling fraction of $f = 87/987 = 0.088$, because 13 out of the original 1,000 units were set aside as a take-all stratum, following the procedure presented in Section 1.

We examined both the Sequential Poisson and the Pareto πps variety of fixed size PoMix, but results are reported in Section 7 only for the former variety, for reasons already mentioned. For each realized sample we computed the three estimators $\hat{Y}_{REG}$, $\hat{Y}_{RA}$ and $\hat{Y}_{HT}$, as given by (5.3), (5.4), and (5.1) with $a_k = a_k^*$ where $a_k^* = 1/A_{\mathrm{mod},k}$. That is, we trust the contention that the realized inclusion probabilities are sufficiently close to the targeted ones, $A_{\mathrm{mod},k}$, so that no appreciable bias will affect the estimates; this was in fact confirmed by our simulation. We computed various Monte Carlo performance measures for each estimator, $\hat{Y}_{REG}$, $\hat{Y}_{RA}$ and $\hat{Y}_{HT}$. Letting $\hat{Y}$ denote one of these, we computed $MCE\hat{Y}$, $MCV\hat{Y}$, that is, the Monte Carlo expectation and the Monte Carlo variance, defined as the mean and the variance, respectively, of the 10,000 realized values of $\hat{Y}$. The Monte Carlo measure of the relative bias is then $(MCE\hat{Y} - Y)/Y$. As expected, this quantity was always very near zero, so results on the relative bias are omitted in the tables which follow.

Although variance estimation is not the primary objective in this article, we also computed an estimated variance $\hat{V}(\hat{Y})$ for each of the 10,000 estimates, as well as the Monte Carlo expectation, $MCE\hat{V}(\hat{Y})$, defined as the mean of the 10,000 realized values of the variance estimator $\hat{V}(\hat{Y})$ associated with $\hat{Y}$. We constructed the variance estimator from the idea that the first and second order inclusion probabilities under fixed size PoMix are in close approximation to their exactly known counterparts under random size PoMix. That is, we acted as if $\pi_k = A_{\mathrm{mod},k} = B + (1 - Q)A_k$ and $\pi_{k\ell} = \pi_k \pi_\ell$ for all $k \neq \ell$. The variance estimator for $\hat{Y} = \hat{Y}_{REG}$ and $\hat{Y} = \hat{Y}_{RA}$ is then of the simple form

$$\hat{V}(\hat{Y}) = \sum_s a_k^*(a_k^* - 1)(g_k e_k)^2 \tag{6.1}$$

where $a_k^* = 1/A_{\mathrm{mod},k}$ and the weights g_k, given in (5.2), and the regression residuals $e_k = y_k - \mathbf{x}_k' \mathbf{b}$ are specific to each of the two estimators. Despite the simple construction, this variance estimator worked well in most cases. For $\hat{Y}_{HT}$, we used the expression

$$\hat{V}(\hat{Y}_{HT}) = \frac{1}{n(n-1)} \sum_s (1 - A_{\mathrm{mod},k}) \left(\frac{n y_k}{A_{\mathrm{mod},k}} - \hat{Y}_{HT} \right)^2 \tag{6.2}$$

For each sample, we also computed the confidence interval $\hat{Y} \pm 1.96\sqrt{\hat{V}(\hat{Y})}$, targeted for an approximate 95% confidence level. The realized coverage rate (the percentage of the 10,000 confidence intervals that contain the true total Y) was close to the desired 95%

rate in most cases, and for this reason, results on coverage rates are omitted in the tables which follow.

The 10,000 repetitions were realized as follows: First, assign PRN's (that is, 1,000 independent realizations of the $Unif(0, 1)$ random variable) to the 1,000 units, then, using these PRN's, realize 100 fixed size PoMix samples (each sample replaced), then assign new PRN's to the 1,000 units, realize 100 more samples, and so on until 100 PRN assignments have been realized, each with 100 drawn samples. The total number of repetitions is thus $100 \times 100 = 10{,}000$. The rationale for reassigning the PRN's is to create smoother conditions for the Monte Carlo experiment. In the simulation, each new sample selection was set in motion by a new start value, D, drawn at random in the unit interval and used for computing the N ranking variable values as described at the end of Section 4.

7. Results and Conclusions

Our comments in this section on the simulation results make use of the following terms in regard to the six generated populations: Those with $p = 1$ and $p = 1.5$ are called ''moderately heteroscedastic''; those with $p = 2$ ''markedly heteroscedastic''; those with $\alpha = \frac{1}{2}$ ''highly skewed''; and those with $\alpha = 1$ ''moderately skewed.'' By ''the variance advantage of PoMix'' for a given estimator ($\hat{Y}_{REG}$ or $\hat{Y}_{RA}$ or $\hat{Y}_{HT}$), we mean the difference, if positive, between that estimator's variance under PoMix with $B = 0$ and the variance of the same estimator under PoMix sampling with a value of B at or in the neighbourhood of the value that yields the smallest variance, in this case 0.02 to 0.05. The ''relative variance advantage'' is the difference divided by the variance realized for $B = 0$. ''Better than'' means ''has lower variance than.''

The simulation results for $MCV\hat{Y}$ and $MCE\hat{V}(\hat{Y})$ are shown in Table 1 (for the highly skewed population, $\alpha = \frac{1}{2}$, and the three values of p), and in Table 2 (for the moderately skewed, exponential population, $\alpha = 1$, and the three values of p). The results in both tables refer to Sequential Poisson PoMix sampling for different Bernoulli widths B between 0 and f, where $f = 87/987 = 0.088$ (for $\alpha = \frac{1}{2}$), and $f = 0.10$ (for $\alpha = 1$). Here, $B = 0$ represents Sequential Poisson sampling (the pure variety); and $B = f$ represents SRS. Our results for Pareto πps PoMix are not reported, because the differences compared to the Sequential Poisson variety were without consequence. As for the Monte Carlo error in this simulation limited to 10,000 repetitions, examination of subsets of the results suggested that the last decimal in the tables may not be ''safe'' for the Monte Carlo variance $MCV\hat{Y}$; on the other hand, the Monte Carlo expected variance estimate, $MCE\hat{V}(\hat{Y})$, showed more stability, and the tabulated last decimal is likely to be correct. We now comment on the results for the estimators, $\hat{Y}_{REG}$, $\hat{Y}_{RA}$ and $\hat{Y}_{HT}$.

7.1. *The variance advantage of PoMix for $\hat{Y}_{REG}$ and $\hat{Y}_{RA}$*

The first part of Question (i) of Section 2 is answered in the affirmative. For both estimators, the variance advantage is considerable for the moderately heteroscedastic populations ($p = 1$ and $p = 1.5$), but nonexistent, as expected, for the markedly heteroscedastic population ($p = 2$); see an analysis in Kröger, Särndal and Teikari (1999). Further, the variance advantage is greater for the highly skewed population ($\alpha = \frac{1}{2}$)

Table 1. Monte Carlo variance of point estimator, $MCV\hat{Y}$; Monte Carlo expectation of variance estimator, $MCE\hat{V}(\hat{Y})$. All entries in thousands. Upper table: $\alpha = \frac{1}{2}$, $p = 1$; Middle Table: $\alpha = \frac{1}{2}$; $p = 1.5$; Lower table: $\alpha = \frac{1}{2}$; $p = 2$

Width B	$MCV\hat{Y}$			$MCE\hat{V}(\hat{Y})$		
	$\hat{Y}_{REG}$	$\hat{Y}_{RA}$	$\hat{Y}_{HT}$	$\hat{Y}_{REG}$	$\hat{Y}_{RA}$	$\hat{Y}_{HT}$
0.000	10.95	9.90	9.90	9.30	9.78	9.91
0.010	6.24	6.53	6.70	6.40	6.37	6.65
0.020	5.20	5.57	6.14	5.84	5.82	6.48
0.030	5.18	5.54	6.95	5.55	5.52	6.67
0.040	5.19	5.56	7.51	5.50	5.47	7.25
0.050	5.22	5.62	8.38	5.55	5.56	8.22
0.060	5.52	5.93	9.78	5.66	5.67	9.64
0.070	6.07	6.53	11.81	6.13	6.16	12.18
0.080	6.60	6.98	16.55	6.78	6.78	16.77
0.088	8.36	8.62	24.86	8.01	8.02	25.33

Width B	$MCV\hat{Y}$			$MCE\hat{V}(\hat{Y})$		
	$\hat{Y}_{REG}$	$\hat{Y}_{RA}$	$\hat{Y}_{HT}$	$\hat{Y}_{REG}$	$\hat{Y}_{RA}$	$\hat{Y}_{HT}$
0.000	2.79	2.11	2.11	2.83	2.83	2.24
0.010	2.05	2.06	2.26	2.06	2.06	2.26
0.020	1.87	1.91	2.51	2.06	2.06	2.63
0.030	1.87	1.91	2.97	2.12	2.12	3.16
0.040	2.15	2.19	3.80	2.25	2.25	3.89
0.050	2.32	2.38	4.55	2.45	2.45	4.91
0.060	2.98	3.00	6.73	2.74	2.74	6.41
0.070	3.40	3.36	8.54	3.22	3.22	8.81
0.080	4.49	4.32	13.83	4.03	4.03	13.20
0.088	6.45	5.91	21.11	5.21	5.21	21.17

Width B	$MCV\hat{Y}$			$MCE\hat{V}(\hat{Y})$		
	$\hat{Y}_{REG}$	$\hat{Y}_{RA}$	$\hat{Y}_{HT}$	$\hat{Y}_{REG}$	$\hat{Y}_{RA}$	$\hat{Y}_{HT}$
0.000	1.03	0.91	0.91	1.07	0.97	0.98
0.010	1.10	1.10	1.44	1.04	1.03	1.38
0.020	1.15	1.15	1.92	1.15	1.15	1.99
0.030	1.38	1.37	2.90	1.29	1.29	2.77
0.040	1.50	1.47	3.96	1.48	1.48	3.81
0.050	1.70	1.66	5.18	1.74	1.74	5.21
0.060	2.27	2.18	7.35	2.09	2.09	7.27
0.070	2.99	2.80	9.75	2.68	2.66	10.63
0.080	4.81	4.29	16.62	3.63	3.61	16.80
0.088	6.92	5.87	28.64	5.01	5.03	28.52

than for the less skewed population ($\alpha = 1$). When a variance advantage exists, it is greatest for B-values in the range 0.02 to 0.05. To illustrate, for $\alpha = \frac{1}{2}$, $p = 1$ the relative variance advantage is between 40% and 50% for both estimators, while for $\alpha = 1$, $p = 1$, it drops to about 20%.

Table 2. Monte Carlo variance of point estimator, $MCV\hat{Y}$; Monte Carlo expectation of variance estimator, $MCE\hat{V}(\hat{Y})$. All entries in thousands. Upper table: $\alpha = 1$; $p = 1$; Middle table: $\alpha = 1$; $p = 1.5$; Lower table: $\alpha = 1$; $p = 2$

Width B	$MCV\hat{Y}$			$MCE\hat{V}(\hat{Y})$		
	$\hat{Y}_{REG}$	$\hat{Y}_{RA}$	$\hat{Y}_{HT}$	$\hat{Y}_{REG}$	$\hat{Y}_{RA}$	$\hat{Y}_{HT}$
0.00	1.81	2.09	2.09	2.19	2.06	2.08
0.01	1.54	1.83	1.93	1.89	1.84	1.97
0.02	1.51	1.75	2.15	1.78	1.74	2.11
0.03	1.51	1.74	2.44	1.72	1.70	2.39
0.04	1.47	1.68	2.77	1.69	1.68	2.79
0.05	1.45	1.63	3.14	1.68	1.68	3.30
0.06	1.53	1.67	4.00	1.71	1.71	3.99
0.07	1.65	1.76	5.04	1.75	1.75	4.89
0.08	1.72	1.83	6.02	1.81	1.82	6.10
0.09	1.88	1.95	7.58	1.91	1.92	7.89
0.10	2.05	2.03	11.66	2.08	2.08	10.74

Width B	$MCV\hat{Y}$			$MCE\hat{V}(\hat{Y})$		
	$\hat{Y}_{REG}$	$\hat{Y}_{RA}$	$\hat{Y}_{HT}$	$\hat{Y}_{REG}$	$\hat{Y}_{RA}$	$\hat{Y}_{HT}$
0.00	1.24	1.33	1.33	1.42	1.32	1.34
0.01	1.16	1.26	1.41	1.31	1.27	1.43
0.02	1.19	1.28	1.73	1.30	1.27	1.70
0.03	1.26	1.33	2.10	1.30	1.28	2.06
0.04	1.33	1.39	2.60	1.33	1.32	2.53
0.05	1.33	1.37	3.06	1.38	1.37	3.12
0.06	1.42	1.44	3.87	1.45	1.43	3.86
0.07	1.52	1.51	4.74	1.53	1.52	4.82
0.08	1.63	1.60	6.02	1.64	1.63	6.13
0.09	1.91	1.79	8.03	1.79	1.77	7.96
0.10	2.19	2.00	11.58	2.01	1.98	10.91

Width B	$MCV\hat{Y}$			$MCE\hat{V}(\hat{Y})$		
	$\hat{Y}_{REG}$	$\hat{Y}_{RA}$	$\hat{Y}_{HT}$	$\hat{Y}_{REG}$	$\hat{Y}_{RA}$	$\hat{Y}_{HT}$
0.00	0.98	0.95	0.95	1.04	0.94	0.95
0.01	0.97	0.96	1.08	1.00	0.95	1.10
0.02	0.96	0.94	1.36	1.02	0.99	1.40
0.03	1.13	1.09	1.83	1.06	1.03	1.80
0.04	1.20	1.12	2.40	1.11	1.09	2.31
0.05	1.23	1.12	2.82	1.19	1.16	2.95
0.06	1.36	1.22	3.81	1.28	1.25	3.77
0.07	1.54	1.36	4.85	1.39	1.36	4.84
0.08	1.83	1.54	6.20	1.54	1.52	6.31
0.09	2.17	1.75	8.48	1.74	1.74	8.48
0.10	2.65	2.07	12.77	2.01	2.06	11.96

7.2. The variance advantage of PoMix for $\hat{Y}_{HT}$

For $\hat{Y}_{HT}$ a clear variance advantage of fixed size PoMix exists for the least heteroscedastic case ($p = 1$). It is roughly 40% for $\alpha = \frac{1}{2}$, $p = 1$, and roughly 8% for $\alpha = 1$, $p = 1$. But from the outset, $\hat{Y}_{HT}$ is of more limited interest than $\hat{Y}_{REG}$ and $\hat{Y}_{RA}$, because it can be taken for granted that a use of the available auxiliary information both in the sampling design and at the estimation stage (as done in $\hat{Y}_{REG}$ and $\hat{Y}_{RA}$) is at least as good as limiting its use to the design stage only (as done in $\hat{Y}_{HT}$). This is confirmed by our finding for all six populations that when B is distinctly larger than 0, the variance of $\hat{Y}_{HT}$ exceeds that of $\hat{Y}_{REG}$ and $\hat{Y}_{RA}$, sometimes considerably. The second part of Question (i) of Section 2 is thereby answered in the affirmative.

7.3. Comparing $\hat{Y}_{REG}$ with $\hat{Y}_{RA}$

Question (ii) in Section 2 receives an affirmative answer from our study in that the variance advantage for $\hat{Y}_{RA}$ is also present in $\hat{Y}_{REG}$. As described in Section 6, the populations were constructed to have a linear regression of y on x with a zero intercept term, which supports an a priori belief that $\hat{Y}_{REG}$ should not realize a smaller variance than $\hat{Y}_{RA}$, and that instead $\hat{Y}_{REG}$ may incur a slightly higher variance than $\hat{Y}_{RA}$ because of ''overfitting.'' Nevertheless, our results for the moderately heteroscedastic populations ($p = 1$ and $p = 1.5$) show a small but clear variance advantage for $\hat{Y}_{REG}$ as compared to $\hat{Y}_R$ for the B-values 0.02 to 0.05 of particular interest. The variance advantage is around 7% for $\alpha = \frac{1}{2}$, $p = 1$. However, for $\alpha = \frac{1}{2}$ and $B = 0$, the ''normal expectation'' holds that $\hat{Y}_{RA}$ has the smaller variance. The Table 1 entries of 10.95 (upper table) and 2.79 (middle table) may seem unduly large, but we believe that they are correct to within reasonable Monte Carlo error limits. Likewise, for values of B in the upper range (when the sampling approaches a selection with equal inclusion probabilities), $\hat{Y}_{RA}$ retains a certain advantage over $\hat{Y}_{REG}$, for all six populations but one. Our result is noteworthy in that it shows that it is possible for $\hat{Y}_{REG}$ (which is based on the idea of a nonzero intercept) to have a distinctly smaller variance than $\hat{Y}_{RA}$ (based on a nonzero intercept) even when there is no intercept in the population data. This supports the idea that since $\hat{Y}_{REG}$ incorporates additional auxiliary information (namely, the population size N), it should have lower variance than $\hat{Y}_{RA}$ at least under certain conditions.

7.4. The effect on the results of the skewness

Question (iii) in Section 2 asked about the effect of high skewness on the variance advantage of PoMix. We note that the population characteristics that seem to enhance the variance advantage are a high skewness ($\alpha = \frac{1}{2}$) and a moderate heteroscedasticity ($p = 1.5$ and, to an even greater extent, $p = 1$). A strikingly large variance advantage is obtained for the highly skewed case $\alpha = \frac{1}{2}$. When the population contains many units with x-values just slightly larger than 0, then their large sampling weights under Sequential Poisson sampling ($B = 0$) will cause a large and erratic contribution to the estimate of the population total. This effect is mitigated by taking B larger than zero.

7.5. The variance estimator

The proposed variance estimators (6.1) and (6.2) work satisfactorily. That is, $MCV\hat{Y}$ and $MCE\hat{V}(\hat{Y})$ are close in most cases; a more noticeable discrepancy is evident only for the case of $\hat{Y}_{REG}$ with $\alpha = \frac{1}{2}$, $p = 1$.

7.6. Concluding comment

This article has left unanswered the question of an optimal choice of B. Based on the experience available, the use of a B-value in the range 0.2f to 0.5f is recommended.

8. References

Aires, N. (1999). Algorithms to Find Exact Inclusion Probabilities for Conditional Poisson Sampling and Pareto pps Sampling Designs. Methodology and Computing in Applied Probability, 54, 459–473.

Aires, N. (2000). Comparisons Between Conditional Poisson Sampling and Pareto πps Sampling Designs. Journal of Statistical Planning and Inference, 82, 133–147.

Atmer, J., Thulin, G., and Bäcklund, S. (1975). Coordination of Samples with the JALES Technique. Statistisk tidskrift, 13, 443–450.

Brewer, K.R.W. (1999). Cosmetic Calibration with Unequal Probability Sampling. Survey Methodology, 25, 205–212.

Brewer, K.R.W., Early, L.J., and Joyce, S.F. (1972). Selecting Several Samples from a Single Population. Australian Journal of Statistics, 14, 231–239.

Holmberg, A. and Swensson, B. (2001). On Pareto πps Sampling: Reflections on Unequal Probability Sampling Strategies. Theory of Stochastic Processes, 7, 142–155.

Kröger, H., Särndal, C.E., and Teikari, I. (1999). Poisson Mixture Sampling: A Family of Designs for Coordinated Selection Using Permanent Random Numbers. Survey Methodology, 25, 3–11.

Ohlsson, E. (1995). Coordination of Samples Using Permanent Random Numbers. In Business Survey Methods, B.G. Cox, D.A. Binder, B.N. Chinnappa, A. Christianson, M.J. Colledge, and P.S. Kott (eds.) New York: Wiley, 153–169.

Ohlsson, E. (1998). Sequential Poisson Sampling. Journal of Official Statistics, 14, 149–162.

Rosén, B. (1997a). Asymptotic Theory for Order Sampling. Journal of Statistical Planning and Inference, 62, 135–158.

Rosén, B. (1997b). On Sampling with Probability Proportional to Size. Journal of Statistical Planning and Inference, 62, 159–191.

Saavedra, P. (1995). Fixed Sample Size PPS Approximations with a Permanent Random Number. Proceedings of the American Statistical Association, Section on Survey Research Methods, 697–700.

Särndal, C.E. (1996). Efficient Estimators with Simple Variance in Unequal Probability Sampling. Journal of the American Statistical Association, 91, 1289–1300.

Received September 2000
Revised June 2002