Jour. Ind. Soc. Ag. Statistics
46(2), 1994 : 236–253

On Nearly πPS Sampling Scheme-II

R.R. Saxena and Padam Singh*
NCERT, New Delhi
(Received : June, 1992)

Summary

A nearly πPS sampling scheme is proposed for a sample of size n. It satisfies many requirements of a good sampling scheme. An empirical study of the stability of variance and comparison of its efficiency with some exactly πPS sampling schemes indicate that the performance of the proposed sampling scheme is satisfactory.

Key words : Bernoulli trial, Inclusion probability, probability proportional to size sampling, non-negativity of variance estimator, Randomised probability proportional to size, Systematic sampling.

Introduction

In order to make efficient use of the Horvitz Thompson estimator [7] for using unequal probability sampling scheme, the fulfillment of the following conditions is desirable.

(i) The inclusion probability of units should be exactly proportional to their size measures.

(ii) Joint inclusion probabilities of all possible pairs of units should be non-zero to ensure the estimability of the variance of the estimator.

(iii) The sampling scheme should be more efficient than the probability proportional to size (PPS) sampling scheme with replacement.

(iv) The sample size n is fixed.

Most of the existing πPS sampling schemes for the sample size n are not satisfactory. For example, Sampford's πPS sampling scheme [12] and Goodman and Kish [5] Randomised PPS systematic sampling scheme are although simple so far as selection procedure is concerned, but both have a limitation that expressions for joint inclusion probabilities are known only asymptotically. In view of the stated limitations, Sunter [15] commented that a sampling scheme with departure from condition (i) above may be acceptable provided the amount of departure is small and known. Such sampling schemes are known as nearly πPS sampling schemes. His aforesaid comments were based on the fact that size measures are not exactly proportional to the values of the study variable. These comments of Sunter led to the exploration of nearly πPS sampling schemes.

* IRMS (ICMR), New Delhi.

Singh and Kaur [13] proposed a nearly πPS sampling scheme in which initial sample is selected by PPSWR and rest of the units are selected by SRSWOR to make sample size 'n' fixed. Their sampling scheme, in addition to being simple, is highly satisfactory as compared to the well known πPS sampling schemes. The present paper also proposes a simple nearly πPS sampling scheme. The properties of the proposed sampling scheme are examined in detail.

2. *Proposed Sampling Scheme*

Consider a finite population consisting of N units. Further, let Y_m (X_m) be the value of study variable (size measure) for the mth unit (say, U_m), $m = 1, 2, \ldots, N$. We assume $X_m > 0$ for all m. The proposed sampling scheme for selecting a sample of size 'n' (say, S_u) consists of the following steps.

Step 1. Arrange the units of the population in increasing order of the size measure.

Step 2. Select a sample of size n by simple random sampling without replacement (SRSWOR).

Step 3. Arrange the selected units in increasing order of indices. Let us denote these by S_x and let S'_x be its complement.

Step 4. Perform Bernoulli trial on the first unit of S_x say U_i, with known probability of success p_i ($0 \leq p_i \leq 1$).

(a) If outcome of Bernoulli trial is success, retain U_i in S_u.

(b) Otherwise (i.e., in the case of failure), select a unit by SRSWOR from amongst those from S'_x having indices greater than i. Include this selected unit in S_u.

Step 5. Repeat step 4 for the 2nd, 3rd, ..., nth unit of S_x.

Note 1 : The sampling scheme requires a vector $P = (p_1, p_2, \ldots, p_n)$ such that $0 \leq p_i < 1$ for all $i \leq N - n$ and for the last n units $p_i = 1$.

Note 2 : It may be noted that S'_x changes after each failure in the selection procedure.

3. *Illustration*

Consider a population of 13 units with size measures as 8, 2, 3, 7, 8, 4, 3, 6, 7, 4, 5, 2 and 6 respectively. Suppose a sample of size 4 is desired to be drawn from this population.

Step 1 : Arrange the units in increasing order of X_i's. The arranged population is 2, 2, 3, 3, 4, 4, 5, 6, 6, 7, 7, 8 and 8.

Let p_i's, the probability of success of Bernoulli trial, for the units of the population be 0.40, 0.35, 0.49, 0.44, 0.57, 0.52, 0.65, 0.80, 0.76, 1.00, 1.00, 1.00 and 1.00 respectively.

Step 2 & 3 : Select a sample of 4 units by SRSWOR. Let the sample S_x after the rearrangement of units be (U_2, U_4, U_6, and U_{12}).

Step 4 :

(a) Perform Bernoulli trial on U_2 with $p_2 = 0.35$. Suppose Bernoulli trial on U_2 is success. Then U_2 is retained in S_u as per step 4(a).

(b) Perform Bernoulli trial on U_4 with $p_4 = 0.44$. Assume its outcome to be failure. Now select one unit by SRSWOR from those units of S'_x whose indices are more than 4, i.e., U_5, U_7, U_8, U_9, U_{10}, U_{11}, and U_{13}. The selected unit is, say U_{10}, then U_{10} is retained in S_u because of the step 4(b).

(c) Perform Bernoulli trial on $U_6 = 0.52$ and let its outcome be failure. One unit is to be selected by SRSWOR from amongst (U_7, U_8, U_9, U_{11} and U_{13}). Let it be U_7. Then U_7 is retained in S_u.

(d) The unit U_{12} is retained in S_u since $p_{12} = 1.00$. Thus, the selected sample S_i is (U_2, U_{10}, U_7, U_{12}).

4. *Inclusion Probabilities*

In deriving expressions for inclusion probabilities, namely, inclusion probability (say, π_i) of U_i in S_u and the joint inclusion probability (say, π_{ij}) of (U_i, U_j) in S_u, the following results will be utilised.

Lemma (a) : Given $U_m \in S_x$, $U_i \in S'_x$ and $i > m$. The probability that $U_i \in S_u$ due to failure outcome of Bernoulli trial on U_m is $q_m/(N-n-m+v)$ where, $q_m = 1 - p_m$ and v is the position of U_m in S_x.

Lemma (b) : Further, let $(U_m, U_u) \in S_x$, $(U_i, U_j) \in S'_x$ and $j > i$, the probability that $(U_i, U_j) \subset S_u$ due to failure of Bernoulli trials on (U_m, U_u) is

$$2\{q_m/(N-n-m+v)\}\{q_u/N-n-u+h-1)\} \text{ when, } m < u < i \text{ and}$$

$$\{q_m/(N-n-m+v)\}\{q_u/N-n-u+h)\} \text{ when, } m < i \text{ and } i < u < j$$

where, v and h are the positions of U_m and U_u respectively in S_x.

Theorems 1 and 2 below present the derivation of the expressions for π_i and π_{ij} respectively.

Theorem 1 : The inclusion probability of U_i in S_u is given by

$$\pi_i = n \frac{\frac{\sum_{m=1}^{i-1} q_m}{(N-m)+p_i}}{N} \text{ where, } q_m = 1 - p_m.$$

Proof : S_u includes U_i in the following two mutually exclusive ways.

(i) S_x does not contain U_i but S_x contains at least one unit whose index is less than i and Bernoulli trials on any one such units may result into failure such that $U_i \in S_u$ as a consequence of step 4 (b).

(ii) S_x contains U_i and Bernoulli trial on U_i results into success such that $U_i \in S_u$ as a result of step 4(a).

The above two situations are discussed below :

(i) Let for a given S_x not containing U_i, there are k_i $(1 \leq k_i \leq \min.(n, i-1)$ units whose indices are less than i in S_x. We denote the set of such units by A_i. In order to get probability of $U_i \in S_u$, the result given in the lemma 'a' is used.

$$pr\{U_i \in S_u \mid S_x \not\ni U_i\} \doteq \frac{\sum_{m \in A_i} q_m}{N-n-m+v}$$

In the above expression, k_i , m and v are random variables.

Let us first take the expectation over v only. Further, it is obvious that for given v and m, there are $\binom{m-1}{v-1}\binom{N-1-m}{n-v}$ S_x's not containing U_i.

$$\text{Thus, } E\left(\frac{q_m}{N-n-m+v} \,\middle|\, m, U_i \in S'_x\right) = \frac{1}{\binom{N}{n}} \sum_{v=1}^{k_m} \frac{\binom{m-1}{v-1}\binom{N-1-m}{n-v}}{N-n-m+v} q_m$$

where, $k_m = \min.(m, n)$

The above expression simplifies to

$$\{1/\binom{N}{n}\} \sum_{v=1}^{k_m} \binom{m-1}{v-1}\binom{N-m}{n-v} \frac{q_m}{(N-m)} = (n/N)\{q_m/(N-m)\}$$

Now taking expectation over m, we get

$$\text{pr}\{U_i \in S_u \mid U_i \in S'_x\} = (n/N) \sum_{m=1}^{i-1} q_m/(N-m) \qquad (1)$$

(ii) For a given S_x containing U_i, $\text{pr}\{U_i \in S_u \mid S_x \supset U_i) = p_i$

Since, there are $\binom{N-1}{n-1}$ S_x's containing U_i, then

$$\text{pr}(U_i \in S_u \mid U_i \in S_x) = np_i/N \qquad (2)$$

Adding (1) and (2), the final expression of

$$\pi_i = n\left\{\sum_{m=1}^{i-1} q_m/(N-m) + p_i\right\}/N \qquad (3)$$

Hence the proof.

Theorem 2 : The joint inclusion probability (π_{ij}) of $(U_i, U_j) \in S_u$ for all $i < j$ is given by

$$\pi_{ij} = \frac{n(n-1)}{N(N-1)}\left[\sum_{u=1}^{i-1} \frac{q_u}{N-u-1}\left\{2\sum_{m>u}^{i-1} \frac{q_m}{N-m-1} + \sum_{m>i}^{j-1} \frac{q_m}{N-m} + p_i\right\} + p_i\left\{\sum_{m=1}^{i-1} \frac{q_m}{N-m-1} + \sum_{m>i}^{j-1} \frac{q_m}{N-m} + p_j\right\}\right]$$

Proof: The units U_i and U_j can be included together in S_u in the following four mutually exclusive ways.

(i) $(U_i, U_j) \in S'_x$

(ii) $U_i \in S_x$ but $U_j \in S_x$

(iii) $U_i \in S_x$ but $U_j \in S_x$

(iv) $(U_i, U_j) \in S_x$

(i) When $(U_i, U_j) \in S'_x$, (U_i, U_j) can be included in S_u at the step 4(b) if

(a) indices of at least two units of S_x are less than i; and

(b) index of at least one unit of S_x is less than i and in addition to it, index of atleast one unit of S_x is greater than i but less than j.

For given S_x, let $A_i \subset S_x$ has k_i units whose indices are less than i and $A_j \subset S_x$ has k_j units having indices greater than i but less than j. Using the lemma (b), the required probability for both the above cases (a) and (b) is given by

$$\text{pr}\{(U_i, U_j) \in S_u \mid S_x \not\ni (U_i, U_j)\}$$

$$= 2 \sum\sum_{(U_m, U_u) \in A_i} \frac{q_m q_u}{(N-n-m+v)(N-n-u+h-1)}$$

$$+ \sum_{U_m \in A_i} \frac{q_m}{N-n-m+v} \sum_{U_u \in A_j} \frac{q_u}{N-n-u+h}$$

For given m and u, v and h are random variables. Taking expectation over all S_x' s containing $U_m, U_u) \in A_i$, we get

$$E\left\{\frac{q_m q_u}{(N-n-m+v)(N-n-u+h)} \mid m, u\right\} = \frac{1}{\binom{N}{n}} \sum_{v=1}^{k_i} \binom{m-1}{v-1} \frac{q_m}{N-n-m+v}$$

$$\sum_{h>v}^{k_i} \binom{u-m-1}{n-v-1} \frac{q_u}{N-n-u+h-1} \binom{N-u-2}{n-h}$$

$$= n(n-1)\{q_m/(N-m-1)\}\{q_u/(N-u-1)\}/\{N(N-1)\}$$

Thus, $\text{pr}\{(U_i, U_j) \in S_u \mid (U_i, U_j) \in S'_x, (m, u) \le i\}$

$$= \frac{2n(n-1)}{N(N-1)} \sum_{m=1}^{i-1} \frac{q_m}{(N-m-1)} \sum_{u>m}^{i-1} \frac{q_u}{(N-u-1)} \tag{4}$$

Similarly, in situation (b), when $U_m \in A_i$ and $U_u \in A_j$, for given m and u

$$E\left\{\frac{q_m}{(N-n-m+v)} \frac{q_u}{(N-n-u+h)} \mid m,u\right\}$$

$$= \frac{1}{\binom{N}{n}} \sum_{v=1}^{k_i} \binom{m-1}{v-1} \frac{q_m}{(N-n-m+v)} \sum_{h>v}^{k_j} \binom{u-m-2}{h-v-1} \frac{q_u}{(N-n-u+h)} \binom{N-u-1}{n-h}$$

$$= \frac{n(n-1)}{N(N-1)} \frac{q_m}{N-m-1} \frac{q_u}{(N-u)}$$

Finally, $pr\{(U_i, U_j) \in S_u \mid (U_i, U_j) \in S'_x, m < i < u, u < j\}$

$$= \frac{n(n-1)}{N(N-1)} \sum_{m=1}^{i-1} \frac{q_m}{(n-m-1)} \sum_{u=i+1}^{j-1} \frac{q_u}{N-u} \tag{5}$$

(ii) Only one unit, say $U_m \in A_i$, causes inclusion of U_i in S_u at the step 4(b) and $U_m \in S_u$ at the step 4 (a).

Given $S_x \not\ni U_i, U_j \in S_x$, let v and h are the respective positions of U_m and U_j in S_x.

$$pr\{S_u \supset (U_i, U_j) \mid S_x \supset U_i, S_r \supset U_j, v, h\} = \sum_{U_m \in A_i} \frac{q_m}{N-n-m+v} p_j$$

Extending the logic used in part (i) of theorem 1, we get

$pr\{S_u \supset (U_i, U_j) \mid U_i \in S'_x. U_j \in S_x\}$

$$= \frac{n(n-)}{N(N-1)} p_j \sum_{m=1}^{i-1} \frac{q_m}{N-m-1} \tag{6}$$

(iii) Using the above procedure, the probability of $S_u \supset (U_i, U_j)$ when $U_i \in S_x$ but $U_j \in S'_x$ is given by

$$\frac{n(n-)}{N(N-1)} p_i \left(\sum_{m=1}^{i-1} \frac{q_m}{N-m-1} + \sum_{m=i+1}^{j-1} \frac{q_m}{N-m} \right) \tag{7}$$

(iv) To include (U_i, U_j) in S_u when $(U_i, U_j) \in S_x$, the Bernoulli trial on both the units should result into success.

$$pr\{S_u)(U_i, U_j) \mid (U_i, U_j) \in S_x\} = \frac{n(n-1)}{N(N-1)} p_i p_j \tag{8}$$

Adding (4) to (8), we finally get

$$\pi_{ij} = \frac{n(n-1)}{N(N-1)}\left[\sum_{m=1}^{i-1}\frac{q_m}{N-m-1}\left\{2\sum_{u>m}^{i-1}\frac{q_u}{N-u-1}+\sum_{u>i}^{j-1}\frac{q_u}{N-u}+p_j\right\}\right.$$
$$\left.+\,p_i\left\{\sum_{m=1}^{i-1}\frac{q_m}{N-m-1}+\sum_{m>i}^{j-1}\frac{q_m}{N-m}+p_j\right\}\right] \qquad (9)$$

5. *Limitation of the sampling scheme*

(*a*) One of the important property of a good sampling scheme is that n is fixed. At times, replacement of $U_m \in S_x$ at the step 4(b) is not possible when $m > N-n$ because S'_x may not contain even a single unit with index greater than m. To avoid this situation, it is necessary that the last 'n' units of the population are not subjected to Bernoulli trial, i.e., $p_m = 1$ for $m > N-n$.

(*b*) Further, to make efficient use of Horvitz Thompson estimator (loc. cit), π_i should be proportional to size measure (X_i). To satisfy this characteristic, values of p_i's in terms of π_i's are found below by using (3).

$$\pi_i = \frac{nX_i}{X} = \frac{n}{N}\left\{\sum_{m=1}^{i-1}\frac{q_m}{N-m}+p_i\right\} \text{ where, } X = \sum_{m=1}^{N} X_m \text{ and } \bar{X} = \frac{X}{N}$$

or, $$p_i = \frac{X_i}{\bar{X}} - \sum_{m=1}^{i=1}\frac{q_m}{N-m} \quad \text{and}$$

$$p_{i-1} = \frac{X_{i-1}}{\bar{X}} - \sum_{m=1}^{i-2}\frac{q_m}{N-m}$$

Taking the difference, the following recurrence relationship is obtained.

$$p_i = \frac{N-i+2}{N-i+1}\,p_{i-1} + \left(\frac{X_i}{\bar{X}} - \frac{X_{i-1}}{\bar{X}}\right) - \frac{1}{N-i+1}$$

Knowing, $p_1 = X_1/\bar{X}$, the above relationship is simplified to

or, $$p_i = \frac{X_i}{\bar{X}} - \sum_{m=1}^{i-1}\frac{(\bar{X}-X_m)}{\bar{X}(N-i+1)} \qquad (10)$$

For p_i's to be in the range of 0 to 1,

$$\sum_{m=1}^{i-1} \frac{(\overline{X}-X_m)}{N-i+1} < X_i \leq \overline{X} + \sum_{m=1}^{i-1} \frac{(\overline{X}-X_m}{N-i+1} \qquad (11)$$

The above condition should hold good for the first N–n units of the population.

At times, this condition may not hold true in a given situation because of variability in X_i's specially when N is small. Such units of a population not fulfilling the condition at (11) can be shifted to the position beyond N–n units and assigned $p_i = 1$.

This shifting of units may cause further violation of the condition at (11) for some larger units. Such units, if any, should also be treated in the stated way. Thus, it may happen that p_m is assigned value 1 for the $r \geq n$ units, i.e., the last n units and also for some (r–n) units violating the stated condition.

6. *Characteristics of the Sampling Scheme*

The proposed sampling scheme has the following desirable properties :

(i) The sample size n is fixed

(ii) π_i's are exactly proportional to their respective size measures for the first N–r units ($r \geq n$) and for the remaining units it is in proportion to the average size measure of the last r units.

(iii) $\pi_{ij} > 0$ for all pairs of units ensuring unbiased estimation of the variance.

(iv) Estimators of variance due to Horvitz and Thompson (loc. cit.) and Yates and Grundy (1953) will always be positive. This is proved in Theorem 3.

(v) Efficiency of the proposed sampling scheme depends upon the variability in the last r units because inclusion probabilities of these units are not proportional to respective sizes. Its performance is, however, empirically compared with some standard unequal probability sampling schemes in the next section.

Theorem 3 : For the proposed sampling scheme, the Yates-Grundy form of the variance estimator takes non-negative values always.

Proof : Estimator of the variance due to Yates and Grundy [16] for the horvitz Thompson (loc. cit.) estimator is given by

$$\sum_{i=1}^{n} \sum_{j>i}^{n} \left(\frac{\pi_i \pi_j - \pi_{ij}}{\pi_{ij}}\right)\left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j}\right)^2$$

Evidently, contribution of each pair of unit of the population will be positive if and only if $\pi_i \pi_j - \pi_{ij} > 0$ for all i and j (j > i). To evaluate the inequality, the expression of π_{ij} at (9) is written as

$$\frac{N(N-1)}{n(n-1)} \pi_{ij} = -\sum_{m=1}^{i-1} \left(\frac{q_m}{N-m-1} \right)^2 + \left(\sum_{m=1}^{i-1} \frac{q_m}{N-m-1} + p_i \right) \left(\sum_{m=1}^{i-1} \frac{q_m}{N-m-1} + \sum_{m=i+1}^{j-1} \frac{q_m}{N-m} + p_i \right)$$

Using (3), the above equation reduces to

$$\frac{N(N-1)}{n(n-1)} \pi_{ij} = -\sum_{m=1}^{i-1} \left\{ Q_m \left(q_m + \frac{q_m}{N-m} + Q_m \right) \right\} + \left(\sum_{m=1}^{i-1} + \frac{N}{n} \pi_i \right) \left(\sum_{m=1}^{i-1} Q_m - \frac{q_i}{N-1} + \frac{N}{n} \pi_j \right) \quad (12)$$

where $Q_m = \dfrac{q_m}{(N-m)(N-m-1)}$

Substituting (12) and after rearranging terms, we obtain

$$\pi_i \pi_j - \pi_{ij} = \frac{n(n-1)}{N(N-1)} \sum_{m=1}^{i-1} \left[Q_m \left\{ q_m + \frac{q_m}{N-m} - \sum_{u=1}^{m-1} Q_u + \frac{q_i}{N-i} - \sum_{u=1}^{m-1} Q_u \right\} \right]$$
$$+ \frac{\pi_j}{N-1} \left\{ \frac{N-n}{n} \pi_i - (n-1) \sum_{m=1}^{i-1} Q_m \right\} + \frac{n-1}{N-1} \pi_i \left(\frac{q_i}{N-i} - \sum_{m=1}^{i-1} Q_m \right) \quad (13)$$

The r.h.s. of (13) > 0 if

(a) $\dfrac{q_m}{N-1} - \sum_{u=1}^{m-1} Q_u > 0$ for all m, and

(b) $\dfrac{N-n}{n} \pi_i - (n-1) \sum_{m=1}^{i-1} Q_m > 0$

To evaluate (a) and (b), express (10) in the following form

$$q_i = \pi'_i + \sum_{m=1}^{i-1} \frac{\pi'_i}{N-i+1} \text{ where, } \pi'_i = 1 - N\pi_i/n \quad (14)$$

Now, $$\sum_{m=1}^{i-1} Q_m = \sum_{m=1}^{i-1} q_m / \{(N-m-1)(N-m)\}$$

Substituting q_m from 14, we get

$$= \sum_{m=1}^{i-1} \frac{\pi'_m}{(N-m)(N-m-1)} + \sum_{m=1}^{i-1} \frac{1}{(N-m-1)(N-m)(N-m+1)} \sum_{u=1}^{m-1} \pi'_u$$

$$= \frac{1}{2}\left[\sum_{m=1}^{i-1} \pi'_m \left\{\frac{1}{(N-m-1)(N-m)} + \frac{1}{(N-i)(N-i+1)}\right\}\right] \quad \ldots (15)$$

Using (14) and (15), the l.h.s. of condition at (a) above transforms to

$$\frac{1}{N-m}\pi'_m + \sum_{u=1}^{m-1} \frac{\pi'_u}{(N-m)(N-m+1)} - 1/2 \sum_{u=1}^{m-1} \pi'_u \left\{\frac{1}{(N-u-1)(N-u)} + \frac{1}{(N-m)(N-m+1)}\right\}$$

$$= \frac{1}{N-m}\pi'_m + \frac{1}{2}\sum_{u=1}^{m-1} \pi'_u \left\{\frac{1}{(N-m)(N-m+1)} - \frac{1}{(N-u)(N-u-1)}\right\} > 0$$

(b) $$\frac{N-n}{n}\pi_i - (n-1)\sum_{m=1}^{i-1} Q_m$$

$$= (1-\frac{n}{N})(1-q_i) + \frac{1}{2}\left[\sum \pi'_m \left\{\frac{2(1-n/N)}{N-i+1} - \frac{n-1}{(N-i)(N-i+1)} - \frac{n-1}{(N-m)(N-m-1)}\right\}\right]$$

$$= (1-\frac{n}{N})(1-q_i) + \frac{1}{2}\left[\sum_{m=1}^{i-1} \pi'_m \left\{\frac{2(1-\frac{n}{N})(N-i) - n + 1}{(N-i)(N-i+1)} - \frac{n-1}{(N-m)(N-m-1)}\right\}\right] > 0$$

because $\frac{1}{(N-i)(N-i+1)} > \frac{1}{(N-m)(N-m-1)}$ for all m.

As (a) and (b) are individually greater than 0, $\pi_i \pi_j - \pi_{ij} \geq 0$

Hence, the proof.

7. *Empirical Illustration*

Performance of the (i) proposed sampling scheme was empirically examined vis-a-vis (ii) Hartley and Rao [6] sampling (HRS); (iii) Sampford's [12] sampling scheme (SS) and (iv) Probability proportional to size with replacement (PPS) sampling scheme. Of the 18 populations considered for this purpose, six have been generated (appendix) and the remaining 12 are taken from those available in the literature. The first three populations are similar to those considered by Cochran [3] for which correlation between X and Y/X (r (X, Y/X)) is an important consideration. For populations 1, 2 and 3, the value of r(X, Y/X) = 1, – 1 and 0 respectively. Besides size (N) of population varying from 13 to 34 for all populations, the value of r, i.e. number of last units for which $p_i = 1$, has been computed for each population which indicated that in only population 14, the condition (11) is violated by 2 units. coefficient of variation of X(CV(X)) and of Y/X(CV(Y/X)) may supply useful information in explaining performance of the proposed sampling scheme.

Correlations r(X, Y) and r(X, Y/X) have also been computed for this purpose. The percentage of variance of X of last r units (V(U)) to the total variance (V(T)) is also considered an important parameter in the explanation of various aspects of the sampling scheme. All these characteristics of the populations are presented in Table 1.

(a) Efficiency of the proposed sampling scheme

Relative efficiency of the proposed, HRS and SS sampling schemes with respect to PPS is presented in Table 2 along with the variance of PPS. In order to study the performance of the proposed scheme, two sets of relative efficiencies are computed for each population. The first set of relative efficiency given in table 2 does not disturb the X_i's. But in the second set, X_i's of the last r units is replaced by $\sum_{m=N-r+1}^{N} X_m/r$.

This set of values of relative efficiency compared with the first set indicated improvement of efficiency of the proposed sampling scheme substantially in the cases of populations 1, 12, 13, 14, 15 and 17.

Although there is no definite indicator, it may be observed that CV(Y/X) for these populations range between 13.64 to 17.03. The following discussions pertain to the first set of relative efficiency.

For populations 1 to 3, simulated for r(X,Y/X), the variability in the last four units is 24%. It is interesting to observe that for the population 2, the proposed sampling scheme, with r(x, Y/X) = –1.0, is almost as efficient as those of HRS and SS schemes. For population 1, (r(X, Y/X) = 1.0), the proposed

Table 1. Characteristics of populations considered for empirical study.

Sr. No.	N	r	CV(X)	(CV(Y/X)	r(X, Y)	r(X, Y/X)	V(U)/V(T) x 100	Population Description
1	13	4	29.65	17.01	0.99	0.99	24.09	Generated Cochran [3] r(x, y/x) = 1
2	13	4	29.65	17.01	0.94	–0.99	24.09	r(X, Y/X) = 0
3	13	4	29.65	10.06	0.94	0.00	24.09	R(X, Y/X) = –1
4	15	4	9.43	2.76	0.96	0.00	30.62	Generated g = 0
5	15	4	9.43	10.62	0.67	0.00	30.62	Generated g = 1
6	15	4	9.43	37.76	0.22	0.00	30.62	Generated g = 2
7	20	4	18.82	41.60	–0.19	–0.45	43.82	Murthy [9] p. 178 Sl. No. 1 to 20
8	20	4	16.82	19.40	0.78	0.26	23.00	Murthy [9] p. 128 Sl. No. 1 to 20
9	20	4	17.50	22.70	0.63	0.10	35.24	Popu. (class II) Chandrasekhar [1] Padamanabha [10]
10	20	4	13.50	18.07	0.64	0.04	19.59	Popu. (class III) do
11	20	4	42.64	15.09	0.87	0.30	76.99	Horvitz & Thompson [7] p. 682 1 to 20
12	20	4	38.89	17.03	0.87	–0.09	72.30	Des Raj [4] p. 283 Sl. No. 1 to 20
13	14	4	41.76	16.65	0.93	–0.48	64.26	Rao [11] p. 207 Sl. No. 1 to 14
14	14	6	116.21	72.11	0.97	0.55	152.55	Kish [8] p. 42 Sl. No. 1 to 14
15	14	4	67.73	13.64	0.99	0.16	44.09	Sukhatme [4] p. 183 Sl. No. 1 to 14
16	30	4	81.40	38.73	0.96	–0.32	72.21	Cochran [2] p. 113 Sl. No. 20 to 49
17	30	4	46.57	45.92	0.63	–0.28	77.36	Sukhatme [14] p. 279 Sl. No. 1 to 30
18	34	4	75.91	19.85	0.92	–0.13	39.48	Sukhatme [14] p. 183 Sl. No. 1 to 34

(*Note* : V(U) is variance of X of last r units and V(T) is the variance of all units. V(U)/V(T) × 100 is the variability in the last r units.)

Table 2. Relative efficiency of sampling schemes with resspect to PPS

Popu No.	Relative efficiency (Undisturbed X_i)				Relative efficiency (For last r units $X_i = \Sigma X_m/r$			
	Variance PPS	Proposed	HRS	SS	Variance PPS	Proposed	HRS	SS
1	6620.87	112.83	132.20	133.18	7644.92	130.28	133.87	134.68
2	6620.87	130.24	132.20	133.18	6406.00	126.03	130.54	131.45
3	1843.62	150.84	138.30	139.29	1629.60	133.33	134.81	134.85
4	37.84	127.45	126.79	126.79	37.78	127.25	126.84	126.84
5	565.48	134.70	127.11	127.11	534.42	127.30	126.90	126.90
6	8212.50	130.43	127.50	127.50	8049.63	127.85	127.41	127.41
7	52702.00	120.11	118.62	118.66	51659.40	117.73	118.24	118.27
8	29352712.00	116.11	115.55	115.56	291297.50	115.42	115.46	115.46
9	6308124.00	121.31	120.71	120.71	6283382.00	120.83	120.67	120.67
10	779144.00	120.65	118.47	118.47	766296.00	118.66	118.45	118.45
11	1181.36	149.19	133.33	133.59	1007.46	127.23	126.46	126.56
12	1623.33	118.31	127.56	127.57	1770.43	129.04	127.24	127.26
13	20019.00	77.12	126.69	126.98	369913.00	142.51	139.46	140.28
14	1623.33	17.94	124.65	125.05	11029.73	138.23	106.37	104.18
15	34948.00	57.79	151.46	152.54	108653.00	179.68	170.60	170.85
16	$2.27*10^5$	107.32	107.35	107.55	$2.27*10^5$	107.38	106.55	106.60
17	$4.37*10^7$	108.63	112.55	112.57	$4.56*10^7$	113.32	112.97	112.99

sampling scheme is less efficient than the two whereas it performs better than these for population 3 having r(X, Y/X) = 0.0.

The populations 4, 5 and 6, generated under finite population g- model for g = 0, 1 and 2 respectively, have about 30% variability in the last four units. The performance of the proposed sampling scheme is slightly better than HRS and SS schemes.

The next 6 populations each with size 20, have been taken from different published sources. The variability of the last four units ranges from 19% to 72%. The r(X, Y) varies for –0.19 to 0.87. The relative efficiency of the proposed sampling scheme for these populations is almost the same as that of the two sampling schemes. However, the proposed sampling scheme for population 11 performs substantially better than the HRS and SS schemes.

Populations 13 to 15, each of size 14, have also been taken from the published sources. The percentage of variability in the last r units varies from 45% to 150%. The performance of the proposed sampling scheme is worst for these populations. These populations are positively skewed and X is strongly correlated with Y, (r(X, Y) > .87). For such populations, the proposed sampling scheme should be used withe care. It is important to mention that for population 14, the rearrangement of two units was necessary to avoid $p_i < 0$ or $p_i > 1$ due to condition at (11). It resulted into assigning inclusion probabilities of smaller and larger units equal. Consequently, efficiency of the proposed sampling scheme dropped drastically.

The remaining three populations of size 30 or more have also been taken from published work. The ratio of variability of last four units is over 70% except for population 18. The relative efficiency of the proposed sampling for population 16 is the same as that of the other two. For population 17, the efficiency of the proposed sampling scheme is slightly less than HRS and SS schemes. But for population 18 it is much higher than the efficiency of the other two sampling schemes. This trend in relative efficiency may be attributed to the high variability in the last four units and also to the high correlation between Y and X.

It may be mentioned that population 15 of size 14 is part of population 18 having size as 34. It has been attempted to examine the loss in efficiency vis-a-vis population size. It is evident, as expected, that for not highly skewed populations, specially when population size (N) is large, the proposed sampling scheme is as good as other πPS sampling schemes and sometimes even better. For skewed populations with small N, the proposed sampling scheme should be used with caution.

(b) Stability and non-negativity of the variance estimator

The stability of the variance estimator depends upon the values of the ratio of $\pi_{ij}/\pi_i \pi_j$. The indicator for this is taken as min $(\pi_{ij}/\pi_i \pi_j)$, and presented in Table 3. It is observed from the values of min $(\pi_{ij}/\pi_i \pi_j)$, that the proposed sampling scheme is as good as HRS and SS schemes except for population 14.

Table 3. Stability and non-negativity of the variance estimator

Popn Sr. No.	minimum $\frac{\pi_{ij}}{\pi_i \pi_j}$			maximum $\frac{\pi_{ij}}{\pi_i \pi_j}$		
	Proposed	HRS	SS	Proposed	HRS	SS
1	0.76	0.75	0.75	0.88	0.87	0.87
2	0.75	0.75	0.76	0.86	0.87	0.87
3	0.76	0.75	0.75	0.88	0.87	0.87
4	0.79	0.79	0.79	0.83	0.82	0.82
5	0.79	0.79	0.79	0.83	0.82	0.82
6	0.79	0.79	0.79	0.83	0.82	0.82
7	0.77	0.76	0.76	0.84	0.82	0.82
8	0.78	0.77	0.77	0.82	0.81	0.81
9	0.78	0.77	0.77	0.84	0.82	0.82
10	0.78	0.78	0.78	0.82	0.81	0.81
11	0.75	0.74	0.74	0.88	0.87	0.87
12	0.74	0.74	0.74	0.86	0.85	0.85
13	0.74	0.74	0.73	0.90	0.91	0.90
14	0.36	0.67	0.63	0.84	1.03	1.01
15	0.61	0.69	0.66	0.92	0.94	0.94
16	0.72	0.72	0.72	0.90	0.90	0.89
17	0.70	0.73	0.73	0.85	0.83	0.83
18	0.66	0.72	0.71	0.87	0.86	0.85

In order to study non-negativity of the variance estimator, values of max $(\pi_{ij}/\pi_i \pi_j)$ have been tabulated in Table 3. The proposed sampling scheme satisfies desirable condition, max $(\pi_{ij}/\pi_i \pi_j) < 1$, of the non-negativity of variance estimator for all the 18 populations. On the contrary, for the other πPS sampling schemes, the stated condition is violated in the case of 14th population.

REFERENCES

[1] Chandrasekhar, A., 1971. Provisional population totals, paper 1, series 1, *Census of India*, Govt. of India, 5.

[2] Cochran, W.G., 1953. Sampling technique, First edition, *John Wiley and Sons, London*, 113.

[3] Cochran, W.G., 1977. Sampling technique, Third edition, *Wiley Eastern Limited, India*, 113.

[4] Des Raj, 1965. Variance estimation in randomised systematic sampling with probability proportional to size, *J. Amer. Statist. Assoc.*, **60**, 278–84.

[5] Goodman R. and Kish, L., 1950. Controlled selection — a technique in probability sampling, *J. Amer. Statist. Assoc.*, **45**, 350–72.

[6] Hartley, H.O. and Rao, J.N.K., 1962. Sampling with unequal probabilities and without replacement, *Ann. Math. Statist.*, **33**, 350–374.

[7] Horvitz, D.G. and Thompson, D.J., 1952. A generalization of sampling without replacement from a finite universe, *J. Amer. Statist. Assoc.*, **47**, 663–685.

[8] Kish, L., 1965. Survey sampling, John Wiley & Sons Inc., New York., 42.

[9] Murthy, M.N., 1967. Smapling theory and methods, *Statistical Publishing Society, Calcutta, India*, 178–228.

[10] Padmanabha, P., 1981. Provisional population totals, paper 2, series A *Census of India, Govt. of India*, 140–143.

[11] Rao, J.N.K., 1967. On three procedures of unequal probability sampling without replacement, *J. Amer. Statist. Assoc.*, **58**, 202–215.

[12] Sampford, M.R., 1967. On sampling without replacement with unequal probability of selection, *Biometrika*, **54**, 499–513.

[13] Singh, P. and Kaur, S., 1992. On nearly πPS sampling scheme, *J. Ind. Soc. Agri. Statist.*, **44** (2), 106–121.

[14] Sukhatme, P.V., 1954. Sampling theory of Surveys, *Ames*. 183–279.

[15] Sunter, A.B., 1977. List sequential sampling with equal or unequal probabilities without replacement, *Appl. Statist.* **25**, 261–268.

[16] Yates, F. and Grundy, P.M., 1953. Selection without replacement from within strata with probability proportionate to size, *J. Roy. Statist. Soc.*, **B15**, 253–261.

APPENDIX

Generated populations which fulfill the conditions considered by Cochran (1977).

A. Populations satisfying corr. (Y/X, X)					B. Populations satisfying finite population g-model				
Unit No	Size measure	Study Variable (Y) for populations			Unit No.	Size measure	Study Variable (Y) for populations		
	(X)	(1)	(2)	(3)		(Y)	(4)	(5)	(6)
1	4	32.0	56.0	32.0	1	13	25.0	22.3	13.0
2	4	34.0	54.0	34.0	2	13	26.0	26.0	26.0
3	4	45.0	55.0	45.0	3	13	27.0	29.7	39.0
4	5	47.0	62.0	47.0	4	14	27.0	24.3	14.0
5	5	60.0	72.0	60.0	5	14	28.0	28.0	28.0
6	6	63.0	69.0	63.0	6	14	29.0	31.7	42.0
7	7	77.0	77.0	77.0	7	15	29.0	26.1	45.0
8	8	92.0	84.0	84.0	8	15	30.0	30.0	30.0
9	8	96.0	80.0	80.0	9	15	31.0	33.9	45.0
10	9	112.0	85.5	85.6	10	16	31.0	26.0	16.0
11	9	114.0	81.0	81.0	11	16	32.0	32.0	32.0
12	10	135.0	85.0	85.0	12	16	33.0	36.0	48.0
13	10	140.0	80.0	80.0	13	17	33.0	29.0	17.0
					14	17	34.0	34.0	34.0
					15	17	35.0	38.1	51.0