Thailand Statistician
July 2011; 9(2) : 143-150
http://statassoc.or.th
Contributed paper

A Generalized πps Sampling Scheme of Two Units

Lokanath Sahoo* [a] and Shishir Nayak [b]

[a] Department of Statistics, Utkal University, Bhubaneswar 751004, India

[b] Department of Statistics, Dhenkanal College, Dhenkanal, India

* Author for correspondence; e-mail: lnsahoostatuu@rediffmail.com

Received: 13 March 2011

Accepted: 17 May 2011

Abstract

This paper proposes a generalized method of πps sampling of two units for estimating a finite population total. The novel feature of the method is that not only it retains its πps properties but also flexible in the sense of leading to many other πps sampling schemes.

Keywords: auxiliary variable, inclusion probability, joint inclusion probability, unequal probability sampling.

1. Introduction and Principles of πps Sampling Scheme

Let y_i be the value of the study variable y for the $i-$ th unit of a finite population $U, i = 1,2,..., N.$ Assume that a sample s of n distinct units is drawn from U according to some unequal probability sampling without replacement scheme with π_i as the inclusion probability of $i-$th unit and π_{ij} as the joint inclusion probability of $i-$ th and $j-$ th units. For estimating population total $Y = \sum_{i=1}^{N} y_i$, Horvitz and

Thompson [1] considered an estimator (called as HT estimator), defined by $t_{HT} = \sum_{i \in s} \frac{y_i}{\pi_i}$. It is true that the HT method of estimation works effectively if π_i is exactly proportional to y_i. But, since the $y-$values are unknown at the sampling stage, sampling schemes which ensure $\pi_i \propto x_i$ are usually employed in practice, where x_i (>0) is the size measure of the $i-$th unit of U and such schemes are called πps sampling schemes. As the πps schemes are operated in combination with HT estimator, they must satisfy some desirable features viz., (i) $\pi_i = np_i$, where $p_i = \frac{x_i}{X}$ is the initial probability of selection of $i-$th unit such that $X = \sum_{i=1}^{N} x_i$, (ii) $\sum_{i=1}^{N} \pi_i = n$, (iii) $\sum_{i \neq j}^{N} \pi_{ij} = (n-1)\pi_i$, (iv) $\sum_i \sum_{j<i} \pi_{ij} = \frac{1}{2}n(n-1)$, and (v) $\pi_i \pi_j \leq \pi_{ij}$, for all $i \neq j$, in order to make Sen [2], and Yates and Grundy [3] unbiased estimator of $Var(t_{HT})$ given by

$$v(t_{HT}) = \frac{1}{2} \sum_{i \neq j \in s} \frac{\pi_i \pi_j - \pi_{ij}}{\pi_{ij}} \left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j} \right)^2, \qquad (1.1)$$

non-negative.

Many concentrated efforts have been made in the literature to develop πps schemes. Brewer and Hanif [4] and Chaudhuri and Vos [5] have made elaborate discussions on a number of such methods. But a majority of them are restricted to $n = 2$ only as the calculation of π_{ij} becomes cumbersome when $n > 2$. However, πps schemes with $n = 2$ are useful in stratified sampling, where stratification is sufficiently ‘deep’ (cf, Chaudhuri and Vos [5, p.148]).

Our present paper makes an attempt to develop a generalized sampling scheme for achieving πps requirements and providing an unbiased and non-negative Sen-Yates-Grundy variance estimator. A number of πps schemes are identified as particular cases of this generalized scheme. Although the scheme can be applicable for $n > 2$, we are confined to $n = 2$ only in order to avoid complexity in deriving expression for π_{ij}.

2. Suggested Sampling Scheme

For each i, define $h_i = p_i^{\delta} \Big/ \sum_{j=1}^{N} p_j^{d}$ and a collection of revised probabilities $\{P_1, P_2, ..., P_N\}$ for the N population units where P_i is defined by

$$P_i = \frac{(2p_i - \lambda h_i)(1 - h_i)}{(1 - 2h_i)}, \ i = 1,2,...,N, \qquad (2.1)$$

such that δ and λ are constants. Here δ is pre-determined whereas λ is determined in order to fulfill the basic requirement $\sum_{i=1}^{N} P_i = 1$ which finally, after a considerable simplification provides

$$\lambda = \sum_{i=1}^{N} \frac{p_i}{1 - 2h_i} \Big/ \sum_{i=1}^{N} \frac{h_i(1 - h_i)}{1 - 2h_i}.$$

The suggested generalized sampling scheme for $n = 2$ [$S(\delta)$, say] consists of the following steps:

Step I: Draw the first unit, say i, with revised probability P_i and without replacement

Step II: Draw the second unit, say j, from the remaining $(N - 1)$ units with conditional probability

$$P_{j/i} = \frac{h_j}{1 - h_i}. \qquad (2.2)$$

Remark 2.1 **:** The suggested method obviously requires that the revised probability P_i should be a non-negative quantity. Hence, a sufficient condition for applicability of the scheme is that $h_i \leq \min\left(\frac{1}{2}, \frac{2p_i}{\lambda}\right) \forall i$. But, this is not surprising as many popular πps methods considered in the survey sampling literature are dependent on this type of restrictions (cf, Brewer [6], Sampford [7], Durbin [8]).

3. Inclusion Probabilities and Properties of $S(\delta)$

We note that,

$$\pi_i = P_i + \sum_{j \neq i} P_j \frac{h_i}{1-h_j}$$

$$= 2p_i - h_i\left[\lambda\left(1+\sum_{j=1}^{N}\frac{h_j}{1-h_j}\right) - 2\sum_{j=1}^{N}\frac{p_j}{1-2h_j}\right].$$

Then, substituting value of λ we get, on simplification, that

$$\pi_i = 2p_i. \tag{3.1}$$

The second order inclusion probabilities are given by

$$\pi_{ij} = P_i P_{j/i} + P_j P_{i/j}$$

$$= \frac{(2p_i - \lambda h_i)h_j}{(1-2h_i)} + \frac{(2p_j - \lambda h_j)h_i}{(1-2h_j)}. \tag{3.2}$$

Now we establish the following properties in respect of $S(\delta)$:

(a) $$\sum_{i=1}^{N}\pi_i = 2\sum_{i=1}^{N} p_i = 2$$

(b) $$\sum_{j\neq i}^{N}\pi_{ij} = \frac{2p_i - \lambda h_i}{(1-2h_i)}\sum_{j\neq i}^{N} h_j + h_i\sum_{j\neq i}^{N}\frac{2p_j - \lambda h_j}{(1-2h_j)}$$

$$= 2p_i - h_i\left[\lambda\left(1+\sum_{j=1}^{N}\frac{h_j}{1-h_j}\right) - 2\sum_{j=1}^{N}\frac{p_j}{1-2h_j}\right]$$

$$= 2p_i$$

$$= \pi_i$$

(c) $$\sum_{i=1}^{N}\sum_{j<i}\pi_{ij} = \sum_{i=1}^{N}\left(\frac{1}{2}\sum_{i\neq j}^{N}\pi_{ij}\right) = \frac{1}{2}\sum_{i=1}^{N}\pi_i = 1$$

(d) After a considerable simplification, for any arbitrary i and j, we have

$$\pi_i \pi_j - \pi_{ij} = \frac{(2p_i - \lambda h_i)(2p_j - \lambda h_j)}{(1-2h_i)(1-2h_j)} \left(\sum_{k>2} h_k \right)^2$$

$$+ h_i h_j \left[\sum_{k>2} \frac{2p_k - \lambda h_k}{1-2h_k} \right]^2 + \pi_{ij} \sum_{k>2} \frac{(2p_k - \lambda h_k) h_k}{(1-2h_k)}$$

$$\geq 0,$$

implying that under $S(\delta)$, $v(t_{HT}) \geq 0$.

4. Some Special Cases of $S(\delta)$

We now consider some noteworthy specific cases of $S(\delta)$. But, it is very clear that the domain of $S(\delta)$ is not restricted only to these cases. Some more such schemes can also come out for other choices of δ.

Case I **:** Let $\delta = 0$, then $h_i = \frac{1}{N}$, $\lambda = \frac{N}{N-1}$, $P_i = \frac{2(N-1)p_i - 1}{N-2}$ and $P_{j/i} = \frac{1}{N-1}$. Hence, we have $\pi_i = 2p_i$ and $\pi_{ij} = \frac{2}{N-2}\left[(p_i + p_j) - \frac{1}{N-1} \right]$ which are identical to those for the Midzuno's [9] sampling scheme for $n = 2$ (see also Horvitz and Thompson [1]).

Case II **:** Let $\delta = 1$, then $h_i = p_i$, $\lambda = \sum_{i=1}^{N} \frac{p_i}{1-2p_i} \bigg/ \sum_{i=1}^{N} \frac{p_i(1-p_i)}{1-2p_i}$, $P_i = \frac{p_i(1-p_i)}{1-2p_i} \bigg/ \sum_{i=1}^{N} \frac{p_i(1-p_i)}{1-2p_i}$ and $P_{j/i} = \frac{p_j}{1-p_i}$. Thus, we obtain $\pi_i = 2p_i$ and $\pi_{ij} = 2p_i p_j \left(\frac{1}{1-2p_i} + \frac{1}{1-2p_j} \right) \bigg/ \left(1 + \sum_{j=1}^{N} \frac{p_j}{1-2p_j} \right)$. This πps scheme is due to Brewer [6], Durbin [8] and Rao [10].

Case III **:** When $\delta = -1,\ h_i = \dfrac{1}{Tp_i}$ where $T = \sum_{i=1}^{N} \dfrac{1}{p_i}$. This case leads to a πps design with revised probability of selecting $i-$ th unit being given by $P_i = \dfrac{(2Tp_i^2 - \lambda)(Tp_i - 1)}{Tp_i(Tp_i - 2)}$, where $\lambda = T\sum_{i=1}^{N} \dfrac{p_i^2}{Tp_i - 2} \Big/ \dfrac{Tp_i - 1}{p_i(Tp_i - 2)}$ such that

$P_{j/i} = \dfrac{p_i}{(Tp_j - 1)p_j}$, $\pi_i = 2p_i$ and $\pi_{ij} = \dfrac{2Tp_i^2 - \lambda}{T(Tp_i - 2)p_j} + \dfrac{2Tp_j^2 - \lambda}{T(Tp_j - 2)p_i}$. This design was considered by Senapati et al. [11].

Case IV **:** When $\delta = 2,$ we see that $h_i = \dfrac{p_i^2}{Z}$, where $Z = \sum_{i=1}^{N} p_i^2$. This case gives rise to a new πps design for which $\gamma = \sum_{i=1}^{N} \dfrac{Zp_i}{Z - 2p_i^2} \Big/ \sum_{i=1}^{N} \dfrac{p_i^2(Z - p_i^2)}{Z(Z - 2p_i^2)}$,

$P_i = \dfrac{(2Zp_i - \gamma p_i^2)(Z - p_i^2)}{Z(Z - 2p_i^2)}$, $P_{j/i} = \dfrac{p_j^2}{Z - p_i^2}$, $\pi_i = 2p_i$ and

$\pi_{ij} = \dfrac{(2Zp_i - \gamma p_i^2)p_j^2}{Z(Z - 2p_i^2)} + \dfrac{(2Zp_j - \gamma p_j^2)p_i^2}{Z(Z - 2p_j^2)}$. This design was developed by Sahoo et al. [12]. In a similar manner, on considering $\delta = -2,$ we can generate another new sampling scheme as a member of $S(\delta)$.

4. Optimum Value of δ

The specific cases of $S(\delta)$ considered in the earlier section are restricted to some integral values of δ only; although fractional values can be considered for this purpose. But, the selection of δ restricts the operation of $S(\delta)$ because $P_i > 0$ for a finite range of δ only depending on the configurations of $x-$ and $y-$values for U. Analytically, it is not possible to trace out an optimum value of δ for which the scheme attains the maximum precision. However, we computed the relative efficiency (RE) of $S(\delta)$ compared to the probability proportional to size with replacement scheme for different values of δ using data on a number of natural and artificial populations available in the survey sampling literature. From these computed values, we noticed that RE is either a concave or a convex function of δ attaining a minimum or maximum value for a value of $\delta \in [1,2]$. We further computed this performance measure for different values of δ in $[1,2]$. However, we observed for all cases that the RE is either maximum or minimum for $\delta = 1.1$ (approx.). But, this cannot be accepted as a unique criterion for all practical purposes, because our numerical study has a limited scope.

5. Conclusions

The foregoing discussions clearly indicate that the proposed generalized scheme is very much attractive in the sense that it can retain its πps properties and provides a non-negative value of $v(t_{HT})$ without imposing any restriction on the choice of the parameter δ although the revised probability P_i itself is a function of δ. Hence, for various choices of δ, the scheme $S(\delta)$ is capable of producing a family of πps sampling schemes for selecting two units from a finite population.

Acknowledgement

The authors are grateful to the referees for providing some valuable comments on an earlier draft of the paper.

References

[1] Horvitz, D.G. and Thompson, D.J., A generalization of sampling without replacement from a finite universe, *Journal of the American Statistical Association*, 1952; 47: 663-685.

[2] Sen, A.R., On the estimator of the variance in sampling with varying probabilities, *Journal of the Indian Society of Agricultural Statistics*, 1953; 5: 119-127.

[3] Yates, F. and Grundy, P.M., Selection without replacement from within strata with probability proportional to size, *Journal of the Royal Statistical Society*, 1953; B15: 235-261.

[4] Brewer, K.R.W. and Hanif, M., Sampling With Unequal Probabilities, Lecture Notes in Statistics, Springer-Verlag, 1983.

[5] Chaudhuri, A. and Vos, J.W.E., Unified Theory and Strategies of Survey Sampling, North Holland, 1988.

[6] Brewer, K.R.W., Ratio estimation in finite populations: Some results deducible from the assumption of an underlying stochastic process, *Australian Journal of Statistics*, 1963; 5: 93-105.

[7] Sampford, M.R., On sampling without replacement with unequal probabilities of selection, *Biometrika*, 1967; 54: 499-513.

[8] Durbin, J., Design of multi-stage surveys for the estimation of sampling errors, *Applied Statistics*, 1967; 16: 152-164.

[9] Midzuno, H., On the sampling system with probability proportionate to sum of sizes, *Annals of the Institute of Statistical Mathematics*, 1952; 3: 99-107.

[10] Rao, J.N.K., On two simple schemes of unequal probability sampling without replacement, *Journal of the Indian Statistical Association*, 1965; 3: 173-180.

[11] Senapati, S.C., Sahoo, L.N. and Mishra, G., On a scheme of sampling of two units with inclusion probability proportional to size, *Austrian Journal of Statistics*, 2006; 35: 445-454.

[12] Sahoo, L.N., Nayak, S.R. and Sahoo, R.K. A new IPPS sampling scheme of two units, *Thailand Statistician*, 2011; 9: 93-102.