Thailand Statistician
January 2011; 9(1) : 93-102
http://statassoc.or.th
Contributed paper

A New IPPS Sampling Scheme of Two Units

L.N. Sahoo* [a], S.R. Nayak [b] and R.K. Sahoo [a]

[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: 16 March 2010

Accepted: 31 May 2010

Abstract

This paper makes an attempt to construct a new IPPS sampling scheme of sample size two for estimating the total of a finite population.

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

1. Introduction and Features of the IPPS Sampling Schemes

Let $(y_i, x_i),\ i = 1,2,...,N$ denote observations on the i th unit of a finite population U in (y, x) with totals $(Y, X),$ such that y is the survey variable and x is an auxiliary variable used as a size measure. In order to estimate $Y,$ assume that a sample s of n distinct units is selected from U according to some unequal probability sampling without replacement scheme with known inclusion probabilities of first and second orders π_i and π_{ij} , respectively. The most commonly used estimator in this context is the Horvitz and Thompson (HT) estimator [1] defined by $t_{HT} = \sum_{i \in s} \frac{y_i}{\pi_i}$. It is well known that if π_i 's are proportional to y_i 's, then there is a considerable reduction in

$Var(t_{HT})$. But, in the absence of knowledge on $y-$values, there is no scope to investigate such a relationship at the design stage. Hence the sampling schemes which ensure $\pi_i \propto x_i$ are usually used in practice. Such schemes are termed as inclusion probability proportional to size (IPPS) or πps sampling schemes. As the IPPS schemes are used under the HT model, they satisfy some conditions *viz*., (i) $\sum_{i=1}^{N} \pi_i = n$, (ii) $\sum_{i \neq j}^{N} \pi_{ij} = (n-1)\pi_i$, (iii) $\sum_{i} \sum_{j<i} \pi_{ij} = \frac{1}{2}n(n-1)$, and (iv) $\pi_i \pi_j \leq \pi_{ij}$, for all $i \neq j$, in order to make Sen [2], and Yates and Grundy [3] unbiased estimator of $Var(t_{HT})$ given by

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

positive.

Many successful attempts have been made in the literature to develop IPPS schemes (cf, Durbin [4, 5], Brewer [6], Sampford [7], Singh [8], Sahoo et al. [9]). Brewer and Hanif [10] and Chaudhuri and Vos [11] have also made elaborate discussions on a number of IPPS schemes. But a majority of such schemes are restricted to $n = 2$ only as the calculation of π_{ij} becomes cumbersome when $n > 2$. However, IPPS schemes with $n = 2$ are useful in stratified sampling, where stratification is sufficiently 'deep' (cf, Chaudhuri and Vos [11, p.148]).

In this paper, we suggest an IPPS sampling scheme for $n = 2$ satisfying the above mentioned desirable properties in terms of π_i and π_{ij}, and also performing well as compared to some other IPPS schemes for a number of natural populations.

2. Suggested IPPS Sampling Scheme

Assuming that x_i's are known and positive for all i, let us define $p_i = \frac{x_i}{X}$ as the initial probability of selection of ith unit. Then, corresponding to the set of initial probabilities $\{p_1, p_2, ..., p_N\}$ for the N population units, consider the set of revised probabilities $\{P_1, P_2, ..., P_N\}$, where P_i is defined by

$$P_i = \frac{(2Zp_i - \gamma z_i)(Z - z_i)}{Z(Z - 2z_i)} \ , \ i = 1,2,...,N, \tag{2.1}$$

such that $z_i = p_i^2$, $Z = \sum_{i=1}^{N} z_i$ and $\gamma = \sum_{i=1}^{N} \frac{Zp_i}{Z - 2z_i} \Big/ \sum_{i=1}^{N} \frac{z_i(Z - z_i)}{Z(Z - 2z_i)}$, determined so as to make $\sum_{i=1}^{N} P_i = 1$, *i.e.*, by solving the equation

$$\sum_{i=1}^{N} \frac{p_i\{Z + (Z - 2z_i)\}}{Z - 2z_i} - \gamma \sum_{i=1}^{N} \frac{z_i(Z - z_i)}{Z(Z - 2z_i)} = 1, \tag{2.2}$$

for γ.

The suggested sampling scheme 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

$$P_{j/i} = \frac{z_j}{Z - z_i}. \tag{2.3}$$

Remark 2.1 **:** It must be noted here that the computation of revised probabilities of our scheme is restricted only to those situations for which $p_i \le \frac{2Z}{\gamma}$ and $z_i \le \frac{Z}{2}$ $\forall i$. But, this is not surprising as most of the IPPS sampling schemes developed in the survey sampling literature suffer from this draw back (cf, Durbin [5], Brewer [6], Sampford [7], Deshpande and Prabhu Ajgaonkar [12]). In some cases, the restrictions on the revised probabilities may be less severe than those of others. However, for our scheme, we verified that the conditions are fulfilled for a number of populations available in the text books and research papers on survey sampling. Some of them are also included in our numerical study, reported in section 4.

Remark 2.2 **:** There may be a confusion that the P_i's of this scheme are equal to those of the IPPS sampling scheme considered by Sahoo et al. [13]. But, for both the cases these probabilities are different as two different functions of p_i^2 are considered for their computation.

3. Inclusion Probabilities and Properties of the Scheme

By definition,

$$\pi_i = P_i + \sum_{j \neq i} P_j \frac{z_i}{Z - z_j}$$
$$= 2p_i - z_i \left[\frac{\gamma}{Z} \left(1 + \sum_{j=1}^{N} \frac{z_j}{Z - z_j} \right) - 2 \sum_{j=1}^{N} \frac{p_j}{Z - 2z_j} \right]. \quad (3.1)$$

Again from (2.2), on simplification, we also have

$$\gamma \left(1 + \sum_{i=1}^{N} \frac{z_i}{Z - z_i} \right) - 2 \sum_{i=1}^{N} \frac{Zp_i}{Z - 2z_i} = 0. \quad (3.2)$$

Hence, from (3.1) and (3.2) we obtain

$$\pi_i = 2p_i. \quad (3.3)$$

The second order inclusion probabilities are given by

$$\pi_{ij} = P_i P_{j/i} + P_j P_{i/j}$$
$$= \frac{(2Zp_i - \gamma z_i) z_j}{Z(Z - 2z_i)} + \frac{(2Zp_j - \gamma z_j) z_i}{Z(Z - 2z_j)}. \quad (3.4)$$

In light of the above discussions, we now establish the following properties:

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

(b) $$\sum_{j \neq i}^{N} \pi_{ij} = \frac{2Zp_i - \gamma z_i}{Z(Z - 2z_i)} \sum_{j \neq i}^{N} z_j + z_i \sum_{j \neq i}^{N} \frac{2Zp_j - \gamma z_j}{Z(Z - 2z_j)}$$
$$= 2p_i - z_i \left[\frac{\gamma}{Z} \left(1 + \sum_{j=1}^{N} \frac{z_j}{Z - z_j} \right) - 2 \sum_{j=1}^{N} \frac{p_j}{Z - 2z_j} \right]$$
$$= 2p_i \quad \text{[from (3.2)]}$$

$$= \pi_i$$

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

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

$$\pi_i \pi_j - \pi_{ij} = \frac{(2Zp_i - \gamma z_i)(2Zp_j - \gamma z_j)}{(Z - 2z_i)(Z - 2z_j)} \left(\sum_{k>2} \frac{z_k}{Z} \right)^2$$

$$+ \frac{z_i z_j}{Z^2} \left[\sum_{k>2} \frac{2Zp_k - \gamma z_k}{Z - 2z_k} \right]^2 + \pi_{ij} \sum_{k>2} \frac{(2Zp_k - \gamma z_k) z_k}{Z(Z - 2z_k)}$$

$$\geq 0 .$$

This implies that for the proposed scheme we must have $v(t_{HT}) \geq 0$.

4. Performance of the Scheme

A desirable further goal is to study efficiency of the proposed sampling scheme, S say, compared to some other IPPS sampling schemes. For this purpose, to avoid mathematical difficulties, we have undertaken a numerical study with the help of 30 natural populations (described in Table 1.) by considering seven other IPPS sampling schemes viz., schemes due to Durbin [4], Brewer [6], Singh [8], Deshpande and Prabhu Ajgaonkar [12], Chao [14], Sahoo et al. [13] and Sahoo et al. [9] denoted by A, B, C, D, E, F and G respectively. We have not considered IPPS schemes of Rao [15], Durbin [5] and Sampford [7], because they give the same π_i and π_{ij} values which are identical to that of Brewer's scheme.

Relative efficiency of the HT estimator under the eight competing IPPS sampling schemes, compared to the conventional estimator $\hat{Y}_{pps} = \frac{1}{n} \sum_{i \in s} \frac{y_i}{p_i}$ under probability proportional to size with replacement (PPSWR) sampling scheme, are presented in Table 2. Our calculations are based on all $C(N, n)$ possible samples of $n = 2$ drawn from a population. The entries for the most efficient cases are boldly printed and those for the second best cases are underlined.

Examinations of the results displayed in Table 2 indicate that the suggested scheme is the best performer for 21 populations and second best performer for 7 populations. On the other hand, schemes F and G are the best for 4 and 3 populations, and second best for 12 and 6 populations respectively. However, the efficiency difference between F and S in most of the cases appears to be marginal. This is probably because of the construction of their revised probabilities as functions of p_i^2. Other 5 IPPS sampling schemes under comparison appear to be inferior to F, G and S. On the whole, our suggested scheme turns out to be the most efficient.

Table 1. Description of populations.

Pop.	Source	N	y	x
1	Mukhopadhyay [16, p.107]	18	area under paddy	cultivated area
2	Singh and Chaudhary [17, p.155]	17	no. of milch animals in survey	no. of milch animals in census
3	Konijn [18, p.49]	16	food expenditure	total expenditure
4	Singh and Mangat [19, p.73 (1-12)]	12	no. of agricultural labourers	total population
5	Singh and Mangat [19, p.73 (13-24)]	12	no. of agricultural labourers	total population
6	Singh and Mangat [19, p.173 (1-13)]	13	area harvested under cultivation	area under paddy
7	Singh and Mangat [19, p.173 (14-26)]	13	area harvested under cultivation	area under paddy
8	Singh and Mangat [19, p.192 (1-15)]	15	rented value of irrigated land for the current year	assessed rental value five years back
9	Singh and Mangat [19, p.192 (16-30)]	15	rented value of irrigated land for the current year	assessed rental value five years back
10	Cochran [20, p.34 (1-17)]	17	food cost	family size
11	Cochran [20, p.34 (18-33)]	16	food cost	family size
12	Mukhopadhyay [16, p.110]	10	output	no. of workers
13	Mukhopadhyay [16, p.110]	10	output	fixed capital
14	Mukhopadhyay [16, p.114]	9	census population for 1961	census population for 1951
15	Murthy [21, p.399]	17	area under wheat in 1963	cultivated area in 1961
16	Murthy [21, p.399]	17	area under wheat in 1964	cultivated area in 1961

Table 1. Continued

17	Cochran [20, p.203]	10	actual weight of peaches	estimated weight of peaches
18	Cochran [20, p.325]	10	no. of persons	no. of rooms
19	Yates [22, p.169]	17	area under wheat	total acreage of crops and grass
20	Jessen [23, p.151]	16	no. of total catch of fish	no. of tagged fish
21	Jessen [23, p.153]	13	no. of households in 1960	no. of households in 1950
22	Horvitz and Thompson [1]	20	no. of households	eye estimated no. of households
23	Sukhatme and Sukhatme [24, p.166 (1-10)]	10	no. of banana bunches	no. of banana pits
24	Sukhatme and Sukhatme [24, p.166 (11-20)]	10	no. of banana bunches	no. of banana pits
25	Asok and Sukhatme [25 (1-17)]	17	acreage under oats in 1957	recorded acreage of crops and grass for 1947
26	Asok and Sukhatme [25 (18-35)]	18	acreage under oats in 1957	recorded acreage of crops and grass for 1947
27	Singh and Chaudhary [17, p.107]	12	catch of fish in a day	no. of boats landing a day
28	Cochran [20, p.187]	18	population in 1960	population in 1950
29	Raj and Chandhok [26, p.291]	20	actual no. of households	eye estimated no. of households
30	Mukhopadhyay [16, p.96]	20	quantity of raw materials	no. of laborers

Table 2. Relative efficiency of different IPPS schemes compared to PPSWR scheme (in %)

Pop.	Sampling Schemes							
	A	*B*	*C*	*D*	*E*	*F*	*G*	*S*
1	106.078	103.575	103.397	103.448	103.965	106.518	106.334	**106.695**
2	105.845	106.730	106.729	106.343	105.991	106.735	**106.871**	106.740
3	91.306	107.285	107.264	107.464	107.503	**107.648**	107.524	107.598
4	25.663	109.233	108.214	110.912	111.336	112.642	111.359	**112.862**
5	59.760	107.962	107.324	107.856	108.777	110.732	110.458	**110.804**
6	96.483	107.726	108.306	107.881	107.789	108.219	107.668	**108.348**
7	92.226	109.154	108.889	109.053	108.322	110.010	110.995	**111.023**
8	105.281	107.764	107.761	107.753	107.587	107.570	107.643	**107.795**
9	104.813	107.908	107.904	107.618	107.768	107.808	107.889	**107.946**
10	**113.957**	107.997	107.555	108.325	111.666	112.224	113.737	112.309
11	**113.132**	108.615	108.156	109.390	110.386	112.774	112.839	112.831
12	116.809	109.867	105.112	108.773	116.996	**121.472**	118.359	119.475
13	116.487	112.222	107.186	117.389	112.888	119.775	118.356	**120.118**
14	106.533	113.945	113.956	113.718	114.846	114.772	113.118	**114.919**
15	95.325	107.278	106.875	108.447	108.391	110.645	109.889	**110.757**
16	108.205	107.870	107.997	107.899	107.988	**108.322**	108.118	108.220
17	109.009	112.109	112.094	112.119	112.375	112.935	112.917	**112.997**
18	111.470	111.655	111.668	111.563	111.805	113.656	112.652	**113.791**
19	106.588	106.603	106.615	106.582	106.548	106.666	106.803	**106.839**
20	108.624	108.883	108.312	108.620	110.084	**113.514**	112.936	113.485
21	97.339	111.274	110.583	110.051	110.913	115.373	114.374	**115.448**
22	108.113	107.844	107.921	107.841	107.914	108.313	107.998	**108.443**
23	113.665	113.740	113.763	113.721	113.831	113.845	113.810	**113.905**
24	112.446	112.304	112.384	112.315	112.223	112.717	**112.863**	112.776
25	108.995	108.015	108.041	108.005	108.663	108.993	108.999	**109.214**
26	107.204	106.975	106.862	106.976	107.113	107.215	106.996	**107.355**
27	108.359	110.915	110.453	110.995	109.323	110.953	110.875	**111.353**
28	107.119	107.079	107.114	107.065	107.345	108.645	108.073	**109.199**
29	118.547	118.444	118.623	118.045	118.112	118.762	118.885	**118.969**
30	101.339	105.863	105.822	105.843	105.988	106.286	**106.832**	106.450

5. Conclusion

On the basis of the analytical and empirical results derived in this work, we may conclude that the suggested sampling procedure is no way inferior to some standard sampling procedures and can be safely applied in many practical situations. But, no general conclusion can be drawn from the empirical study as the conclusion is based on the results for 30 populations only. However, this comparison gives an indication that the suggested IPPS scheme (if it exists) compares well with other popularized IPPS schemes in terms of efficiency.

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] Durbin, J. Some results in sampling theory when the units are selected with unequal probabilities, Journal of the Royal Statistical Society, 1953; B15: 262-269.

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

[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] Singh, P. The selection of samples of two units with inclusion probabilities proportional to size, Biometrika, 1978; 65: 450-454.

[9] Sahoo, L.N., Mishra, G. and Nayak, S.R. On a πps scheme of sampling of two units, Brazilian Journal of Probability and Statistics, 2007; 21: 165-173.

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

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

[12] Deshpande, M.N. and Prabhu Ajgaonkar, S.G. An IPPS (inclusion probability proportional to size) sampling scheme, Statistica Neerlandica, 1982; 36: 209-212.

[13] Sahoo, L.N., Mishra, G. and Senapati, S.C. A new sampling scheme with inclusion probability proportional to size, Journal of Statistical Theory and Applications, 2005; 4: 361-369.

[14] Chao, M. A general purpose unequal probability sampling plan, Biometrika, 1982; 69: 653-656.

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

[16] Mukhopadhyay, P. Theory and Methods of Survey Sampling, Prentice-Hall of India, New Delhi, 1998.

[17] Singh, D. and Chaudhary, F.S. Theory and Analysis of Sample Survey Designs, Wiley Eastern Limited, 1986.

[18] Konijn, H.S. Statistical Theory of Sample Survey Design and Analysis, North Holland, 1973.

[19] Singh, R. and Mangat, N.S. Elements of Survey Sampling, Kluwer Academic Publishers, The Netherlands, 1996.

[20] Cochran, W.G. Sampling Techniques, 3rd Edition, John Wiley & Sons, 1977.

[21] Murthy, M.N. Sampling Theory and Methods, Statistical Publishing Society, Calcutta, 1967.

[22] Yates, F. Sampling Methods for Censuses and Surveys, Charles Griffin & Company, London, 1953.

[23] Jessen, R.J. Statistical Survey Techniques, New York: Wiley, 1978.

[24] Sukhatme, P.V. and Sukhatme, B.V. Sampling Theory of Surveys with Applications, Asia Publishing House, Calcutta, 1970.

[25] Asok, C. and Sukhatme, B.V. On Sampford's procedure of unequal probability sampling without replacement, Journal of the American Statistical Association, 1976; 71: 912-918.

[26] Raj, D. and Chandhok, P. Sample Survey Theory, Narosa Publishing House, 1998.