Thailand Statistician
July 2013; 11(2): 133-141
http://statassoc.or.th
Contributed paper

A Simplified Selection Scheme for Unequal Probability Sampling Without Replacement

Neeraj Tiwari and Akhil Chilwal*

Department of Statistics, Kumaun University, Almora-263601, Uttarakhand, India.

* Corresponding author; e-mail: akhil.stat@gmail.com

Received: 17 December 2012

Accepted: 20 March 2013

Abstract

In this paper we present a new inclusion probability proportional size (IPPS) sampling scheme which have some desirable properties of IPPS sampling and providing an unbiased and non-negative variance estimator under the Horvitz-Thompson model. I also consider an empirical study to examine the performance of the proposed sampling scheme compared to some standard IPPS sampling schemes.

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

1. Introduction

A sample is a selection of units from a well defined population. In survey research a sample is a selection of persons or households from the resident population in private households. Survey researchers expect that the selected persons or households are representative for the population. The goal is to get such a sample which is an unbiased and representative for the population. For this purpose, Hansen and Hurvitz [1] published their theory of multi-stage samples. In the first stage, primary sampling units are selected with probabilities proportional to their size. After that, within selected primary units, fixed numbers of secondary units are selected.

The concept of unequal probability sampling without replacement was first introduced by Madow [2], without any framework. Horvitz & Thompson [3] were the first to give theoretical framework of unequal probability sampling without replacement. They also proposed their selection scheme and an estimator of population total. In 1950, Midzuno [4] proposed a system of sampling. It was very simple for computational point of view but it has some restrictions. Narian [5] introduced a sampling scheme which has no restriction; although at the same time leads to a more efficient estimate of population value. Brewer & Undy [6] have considered this method of sampling in detail for the case n=2 and have given a faster iterative method of solution. Midzuno [7], Durbin [8], Stevens [9], Rao et al. [10], Cochran [11], Brewer [12], Rao [13], Connor [14], Fellegi [15], Hartley [16], Hanurav [17], Durbin [18], Sampford [19], Singh [20], Deshpande & Aigaonkar [21] and Chao [22] had proposed some new methods of selecting units with varying probabilities and without replacement.

Brewer & Hanif [23], Chaudhuri & Vos [24], Sahoo et al. [25], Sahoo et al. [26], Senapati et al. [27], Sahoo et al. [28], Sahoo et al. [29], Sahoo et al. [30] had made elaborate discussion on a number of IPPS schemes. A majority of these methods are restricted to sample size 2 only, because the calculation of second order inclusion probability (π_{ij}) is very cumbersome when the sample size is greater than 2. An IPPS sampling scheme must satisfy the following properties viz.,

(i) $$\sum_{i=1}^{N} \pi_i = n$$

(ii) $$\sum_{i \neq j}^{N} \pi_{ij} = (n-1)\pi_i$$

(iii) $$\sum_{i}\sum_{j<i} \pi_{ij} = \frac{1}{2} n(n-1)$$

(iv) $$\pi_i \pi_j \leq \pi_{ij}\,, \qquad \textit{for all } i \neq j,$$

in order to make Sen [31] and Yates & Grundy [32] unbiased estimator of $\mathrm{Var}\left(\hat{y}_{HT}\right)$ given as $\mathrm{var}(\hat{y}_{HT}) = \sum_{i=1}^{N} \sum_{j>i=1}^{N} \left(\frac{\pi_i \pi_j - \pi_{ij}}{\pi_{ij}}\right)\left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j}\right)^2$ positive.

The main objective of this paper is to propose an IPPS sampling scheme for n = 2, which have desirable properties of IPPS scheme, easy in calculation and also perform better when compared to popular probability sampling schemes for a number of natural populations with different characteristics.

2. The Proposed Sampling Schemes:

Let us consider $p_i = x_i / \sum_{i=1}^{N} x_i$ as the initial probability of selection of i^{th} unit and x_i's are known and positive for all *i*. Then, corresponding to the set of initial probabilities $p_1, p_2, p_3, \ldots, p_N$ for the *N* population units, suppose the set of revised probabilities are $P_1, P_2, P_3, \ldots, P_N$, where P_i is defined as:

$$P_i = \frac{p_i(1-p_i)^2}{(1-2p_i)} \qquad (i = 1, 2, \ldots, N) \qquad (2.1)$$

The suggested sampling scheme consist the following steps:

Step (i): Draw the first unit with the revised probability P_i and without replacement.

Step (ii): Draw the second unit from the remaining (*N*-1) units with conditional probability:

$$P_{(j|i)} = \frac{p_j}{1-p_i} \qquad (2.2)$$

By definition, the probability of inclusion for the i^{th} unit in the sample for this selection scheme is given as:

$$\pi_i = \frac{P_i}{\sum_{i=1}^{N} P_i} + \sum_{j \neq i=1}^{N} \left(\frac{P_j}{\sum_{j=1}^{N} P_j} \times \frac{p_i}{1-p_j} \right)$$

$$= \frac{1}{B}\left[\frac{p_i(1-p_i)^2}{(1-2p_i)} \right] + \frac{1}{B} \sum_{j \neq i}^{N} \left(\frac{p_j(1-p_j)^2}{(1-2p_j)} \times \frac{p_i}{1-p_j} \right)$$

$$= \frac{1}{B}\left[p_i(1-p_i) + p_i \sum_{j=1}^{N} \frac{p_j(1-p_j)}{(1-2p_j)} \right] \qquad (2.3)$$

where B = $\sum_{i=1}^{N} P_i$.

The joint probability of inclusion for the i^{th} and j^{th} units in the sample for this selection scheme is given as:

$$\pi_{ij} = P_i P_{(j|i)} + P_j P_{(i|j)} \qquad (2.4)$$

$$\pi_{ij} = \frac{p_i p_j}{B}\left[\frac{(1-p_i)}{(1-2p_i)} + \frac{(1-p_j)}{(1-2p_j)}\right] \tag{2.5}$$

3. Desirable Properties of the Proposed Scheme

In this section we have verified some of the results for the quantities π_i and π_{ij} obtained under the proposed selection scheme. These results are very important for validity and application of a selection scheme.

(i)

$$\sum_{i=1}^{N}\pi_i = \sum_{i=1}^{N}\frac{1}{B}\left[p_i(1-p_i) + p_i\sum_{j=1}^{N}\frac{p_j(1-p_j)}{(1-2p_j)}\right]$$

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

$$= \frac{1}{B}\left[\sum_{i=1}^{N}\frac{p_i(1-p_i)(1-2p_i+1)}{(1-2p_i)}\right]$$

$$= \frac{2}{B}\left[\sum_{i=1}^{N}\frac{p_i(1-p_i)^2}{(1-2p_i)}\right]$$

$$= 2$$

(ii)

$$\sum_{j\neq i}\pi_{ij} = \sum_{j\neq i}\left[\frac{p_i p_j}{B}\left\{\frac{(1-p_i)}{(1-2p_i)} + \frac{(1-p_j)}{(1-2p_j)}\right\}\right]$$

$$= \frac{1}{B}\left[\frac{p_i(1-p_i)}{(1-2p_i)}\sum_{j\neq i}p_j + p_i\sum_{j\neq i}\frac{p_j(1-p_j)}{(1-2p_j)}\right]$$

$$= \frac{1}{B}\left[\frac{p_i(1-p_i)^2}{(1-2p_i)} - \frac{p_i^2(1-p_i)}{(1-2p_i)} + p_i\sum_{j=i}\frac{p_j(1-p_j)}{(1-2p_j)}\right]$$

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

$$= \pi_i$$

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

(iv) After simplification for any arbitrary *i* and *j* with the help of Konijn [33], we have:

$$\pi_i\pi_j - \pi_{ij} = \frac{p_i p_j (1-p_i)(1-p_j)}{B^2(1-2p_i)(1-2p_j)}\left(\sum_{k\neq i\neq j} p_k\right)^2 + \frac{p_i p_j}{B^2}\left(\sum_{k\neq i\neq j}\frac{p_k(1-p_k)}{(1-2p_k)}\right)^2$$

$$+\frac{\pi_{ij}}{B}\sum_{k\neq i\neq j}\frac{p_k^2(1-p_k)}{(1-2p_k)}$$

$$\pi_i\pi_j - \pi_{ij} > 0$$

Thus an unbiased and positive estimator of the variance of the Horvitz-Thompson (HT) estimator can always be obtained under the suggested sampling scheme.

4. Empirical Evaluation

In this section, we consider some numerical examples to demonstrate the utility of the proposed sampling scheme and compare it with the existing sampling schemes. We considered the efficiency of the proposed sampling scheme, I (say), compared to some various IPPS sampling schemes. For this purpose, we have a numerical study with the help of 10 natural populations (given in Table 1) by considering eight other IPPS sampling schemes viz., schemes due to Durbin [8], Brewer [12], Singh [20], Deshpande & Ajgaonkar [21], Chao [22], Sahoo et al. [25], Sahoo et al. [28] and Sahoo et al. [30] denoted by A, B, C, D, E, F, G and H respectively. Relative efficiency of the HT estimator under the nine competing IPPS sampling schemes, compared to the estimator $\hat{Y}_{pps} = (1/n)\sum_{i\in s}(y_i / p_i)$ under probability proportional to size with replacement (PPSWR) sampling scheme, are presented in Table 2. Our mathematical calculations are based on all possible samples of size 2 drawn from a population. Examination of the results shown in Table 2 indicates that the proposed scheme is the best performer for 7 populations. On the other hand, schemes A, F, and H are the best performer for population 2, 5, and 6. Other IPPS sampling schemes under comparison appear to be inferior to G, H and I. On the whole, our proposed scheme turns out to be most efficient.

5. Conclusion

On the basis of our analytical and empirical study, we found that the proposed sampling scheme has satisfied all the properties of IPPS sampling and the efficiency of the suggested scheme, compared to other leading sampling schemes is high and the performance of the proposed sampling scheme is much satisfactory. The calculation of proposed sampling scheme is very easy. The computation of revised probabilities for some standard schemes is very laborious but with the help of proposed sampling scheme we can get it easily. The comparison of efficiencies gives an indication that the proposed IPPS scheme compares well with other popularized IPPS schemes under consideration.

Table 1. Description of populations.

Pop.	Source	N	*y*	*x*
1	Singh and Chaudhary, [34], (p.155)	17	No. of milch animals in survey	No. of milch animals in census
2	Cochran, [11], (p.34)	17	Food cost	Family size
3	Murthy, [35], (p.399)	17	Area under wheat in 1964	Cultivated area in 1961
4	Ashok and Sukhatme, [36], (1-17)	17	Acreage under oats in 1957	Recorded acreage of crops and grass for 1947
5	Mukhopadhyay, [37], (p.110)	10	Output	No. of workers
6	Cochran, [11], (p.203)	10	Actual weight of peaches	Estimated weight of peaches
7	Cochran, [11], (p.325)	10	No. of persons	No. of rooms
8	Sukhatme and Sukhatme, [38], (p.166, 1-10)	10	No. of banana bunches	No. of banana pits
9	Sukhatme and Sukhatme, [38], (p.166, 11-20)	10	No. of banana bunches	No. of banana pits
10	Singh and Chaudhary, [34], (p.107)	12	Catch of fish in a day	No. of boats landing a day

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

Pop.	Sampling Scheme								
	A	B	C	D	E	F	G	H	I (Proposed)
1	105.85	106.73	106.73	106.34	105.99	106.74	106.87	106.74	**107.47**
2	**113.96**	108.00	107.56	108.33	111.67	112.22	113.74	112.31	108.78
3	108.21	107.87	108.00	107.90	107.99	108.32	108.12	108.22	**113.56**
4	109.00	108.02	108.04	108.01	108.66	108.99	109.00	109.21	**109.92**
5	116.81	109.87	105.11	108.77	117.00	**121.47**	118.36	119.48	111.90
6	109.01	112.11	112.09	112.12	112.38	112.94	112.92	**113.00**	108.82
7	111.47	111.66	111.67	111.56	111.81	113.66	112.65	113.79	**113.93**
8	113.67	113.74	113.76	113.72	113.83	113.85	113.81	113.91	**113.97**
9	112.45	112.30	112.38	112.31	112.22	112.72	112.86	112.78	**113.90**
10	108.36	110.92	110.45	111.00	109.32	110.95	110.88	111.35	**111.90**

References

[1] Hansen, M.H., and Hurvitz, W.N. On a sample procedure of unequal probability sampling without replacement. J. Roy. Stat. Soc. Ser. B, 1943; 24: 482-491.

[2] Madow, W.G. On the theory of systematic sampling II, Ann. Math. Stat., 1949; 20: 333-354.

[3] Horvitz, D.G., and Thompson, D.J. A generalization of sampling without replacement from finite universe. J. Amer. Statist. Assoc., 1952; 47: 663-685.

[4] Midzuno, H. An outline of theory of sampling system. Ann. Inst. Stat. Math., Japan, 1950; 1: 149-156.

[5] Narian, R.D. On sampling without replacement with varying probabilities. J. Ind. Soc. Agri. Stat., 1951; 3: 169-174.

[6] Brewer, K.R.W., and Undy, G.C. Samples of two units drawn with unequal probabilities without replacement, Aust. J. Stat., 1962; 4: 89-100.

[7] Midzuno, H. On the sampling system with probability proportional to sums of sizes. Ann. Inst. Stat. Math., Japan, 1952; 3: 99-107.

[8] Durbin, J. Some results in sampling theory when the units are selected with unequal probabilities. Journal of the Royal Statistical Society, Ser. B, 1953; 15: 262- 269.

[9] Stevens, W.L. Sampling without replacement with probability proportional to size. J. Roy. Stat. Soc. Ser. B, 1958; 20: 393-397.

[10] Rao, J.N.K., Hartley, H.O., and Cochran, W.G. On a sample procedure of unequal probability sampling without replacement. J. Roy. Stat. Soc. Ser. B, 1962; 24: 482-491.

[11] Cochran, W.G. Sampling Techniques, 2nd & 3rd Edition, John Wiley & Sons, 1963, 1977.

[12] 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.

[13] Rao, J.N.K. On two simple schemes of unequal probability sampling without replacement. J. Ind. Stat. Asso., 1965; 3: 173-180.

[14] Connor, W.S. An exact formula for the probability that two specified sampling units will occur in a sample drawn with unequal probabilities and without replacement. J. Amer. Stat. Assoc., 1966; 61: 384-390.

[15] Fellegi, I.P. Sampling with varying probabilities without replacement: Rotating and non-rotating samples. J, Amer. Stat. Assoc., 1963; 58: 183-201.

[16] Hartley, H.O. Systematic sampling with unequal probabilities and without replacement. J, Amer. Stat. Assoc., 1966; 61: 739-748.

[17] Hanurav, T.V. Optimum utilization of auxiliary information: πps sampling of two units from a stratum. Journal of the Royal Statistical Society, Ser. B, 1967; 29: 374-391.

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

[19] Sampford, M.R. On sampling with replacement with unequal probabilities of selection. Biometrika, 1967; 54: 499-513.

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

[21] Deshpande, M.N., and Ajgaonkar, S.G.P. An IPPS (Inclusion Probability Proportional to Size) Sampling Scheme. Statistica Neerlandica, 1982; 36: 209-212.

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

[23] Brewer, K.R.W., and Hanif, M. Sampling with Unequal Probabilities. Springer-Verlag, 1983.

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

[25] 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.

[26] Sahoo, L.N., Das, B.C., and Singh, G.N. A note on an IPPS sampling scheme. Allgemeines Statistisches Archiv, 2006; 90: 385-393.

[27] 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.

[28] Sahoo, L.N., Mishra, G., and Nayak, S. R. On a πps scheme of sampling of two units. Bra. J. Prob. and Stat., 2007; 21: 165-173.

[29] Sahoo, L.N., Senapati, S.C., and Mangaraj, A.K. A class of IPPS sampling schemes, 2010; 217-224.

[30] Sahoo, L.N., Nayak, S.R., and Sahoo, R.K. A new IPPS sampling scheme of two units, 2011; 93-102.

[31] Sen, A.R. On the estimation of variance in sampling with varying probabilities. J. Ind. Soc. Agril. Statist., 1953; 5: 119-127.

[32] Yates, F., and Grundy, P.M. Selection without replacement from within strata with probability proportional to size. J. Roy. Statist. Soc., Ser. B, 1953; 15: 253-261.

[33] Konijn, H.S. Statistical Theory of Sample Survey Design and Analysis. North Holland. 1973; p.253.

[34] Singh, D., and Chaudhary, F.S. Theory and Analysis of Sample Survey Designs, 1986.

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

[36] Asok, C., and Sukhatme, B.V. On Sampford's procedure of unequal probability sampling without replacement, J. Amer. Stat. Assoc., 1976; 71: 912-918.

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

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