Survey Methodology, December 1987
Vol. 13, No. 2, pp. 221–224
Statistics Canada

Comparison of the Horvitz-Thompson Strategy with the Hansen-Hurwitz Strategy

S.G. PRABHU-AJGAONKAR[1]

ABSTRACT

The Hansen-Hurwitz (1943) strategy is known to be inferior to the Horvitz-Thompson (1952) strategy associated with a number of IPPS (inclusion probability proportional to size) sampling procedures. The present paper presents a simpler proof of these results and therefore has some pedagogic interest.

KEY WORDS: Sampling strategies; Inclusion probability proportional to size; Positive definite quadratic form.

1. INTRODUCTION

Let U be a finite population consisting of N identifiable units $[U_1, U_2, \ldots, U_N]$. With the i-th unit of the population U_i are associated two numbers X_i and Y_i, where X_i's are known and Y_i's are fixed but unknown. Generally, X_i represents a measure of size of U_i which is highly correlated with Y_i.

For estimating the population total $T_y = Y_1 + Y_2 + \ldots + Y_N$, the Hansen and Hurwitz (1943) strategy consists of selecting with replacement n population units with probability proportional to X_i, and using the unbiased estimator

$$t_{HH} = \frac{1}{n} \sum_{r=1}^{n} \frac{y_r}{p_r}$$

where $p_r = X_r / T_x$, $T_x = X_1 + X_2 + \ldots + X_N$, and y_r $(r=1, 2, \ldots, n)$ represents the outcome at the r-th draw. It is easy to show, noting that $\Sigma Z_i = 0$,

$$\mathrm{Var}(t_{HH}) = \sum_{i=1}^{N} \frac{Z_i^2}{np_i} \tag{1}$$

where $Z_i = Y_i - p_i T_y$, $i=1, 2, \ldots, N$.

When population units are selected without replacement, Horvitz and Thompson (1952) proposed the unbiased estimator

$$t_{HT} = \sum_{i=1}^{n} \frac{y_i}{\pi_i}$$

[1] S.G. Prabhu-Ajgaonkar, Department of Mathematics and Statistics, Marathwada University, Aurangabad 431004, India.

where π_i ($i=1, 2, \ldots, N$) denotes the probability of including the i-th population unit U_i in the sample. Further, when π_i is proportional to X_i, the sampling procedure is termed an IPPS scheme. For such a sampling procedure,

$$\mathrm{Var}(t_{HT}) = \sum_{i=1}^{N} \frac{Z_i^2}{np_i} + \sum_{i \neq j=1}^{N} Z_i Z_j \frac{\pi_{ij}}{n^2 p_i p_j} \tag{2}$$

where Z_i is given in (1), and π_{ij} ($i \neq j=1, 2, \ldots, N$) represents the joint probability of including the i-th and j-th population units in the sample. When an IPPS procedure is specified, π_{ij} can be further simplified.

From (1) and (2),

$$\phi = \mathrm{Var}(t_{HT}) - \mathrm{Var}(t_{HH}) = \sum_{i \neq j=1}^{N} Z_i Z_j \frac{\pi_{ij}}{n^2 p_i p_j}. \tag{3}$$

2. COMPARISON OF STRATEGIES

Midzuno (1952), Sen (1952) and Sankaranarayanan (1969) proposed IPPS sampling schemes for estimating T_y, using the Horvitz-Thompson estimator t_{HT}. The Midzuno-Sen scheme is feasible if

$$p_i = \frac{X_i}{T_x} > \frac{n-1}{n(N-1)}, \; i=1, \ldots, N, \tag{4}$$

Sankaranarayanan's scheme requires the weaker condition

$$\sum_{j \in s} p_j > (n-1)/(N-1) \text{ for all } s \in S.$$

For both the schemes, the joint inclusion probabilities are given by

$$\pi_{ij} = \frac{n(n-1)}{N-2}\left(p_i + p_j - \frac{1}{N-1}\right).$$

Hence, from (3),

$$\phi = \frac{n(n-1)}{n^2(N-2)}\left[\sum_{i=1}^{N} \frac{Z_i^2}{p_i}\left(2 - \frac{1}{(N-1)p_i}\right) + \frac{1}{(N-1)}\left(\sum_{i=1}^{N} \frac{Z_i}{p_i}\right)^2\right]. \tag{5}$$

The above expression is nonnegative if

$$P_i > \frac{1}{2(N-1)}, \; i=1, 2, \ldots, N,$$

in which case the Horvitz-Thompson strategy is superior to the Hansen-Hurwitz strategy. The above restriction on X_i^2 was first derived by Rao (1963) when $n=2$ and Midzuno-Sen scheme is employed, but it is interesting to note from (5) that the restriction remains the same even when n is greater than 2.

Chaudhuri (1975) and Mukhopadhyay (1975) independently derived the above for the Midzuno-Sen scheme.

Brewer (1963), Rao (1965) and Durbin (1967) proposed different IPPS schemes, for the case $n=2$, with the same inclusion probabilities,

$$\pi_{ij} = \frac{2p_i p_j}{1+k}\left(\frac{1}{1-2p_i} + \frac{1}{1-2p_j}\right) \text{ where } k = \sum_{i=1}^{N} \frac{p_i}{1-2p_i}.$$

These schemes are free from the restrictions on the p_i's of the previous schemes. From (3),

$$\phi = \frac{1}{1+k}\sum_{i=1}^{N} \frac{Z_i^2}{1-2p_i} \geq 0,$$

so that the Hansen-Hurwitz strategy is again inferior to the Horvitz-Thompson strategy.

ACKNOWLEDGEMENTS

The author is indebted to the Editor, M.P. Singh, and a referee for their many helpful comments.

REFERENCES

BREWER, K.R.W. (1963). A model of systematic sampling with unequal probabilities. *Australian Journal of Statistics*, 5, 5-13.

CHAUDHURI, A. (1975). On some properties of the sampling scheme due to Midzuno. *Bulletin of Calcutta Statistical Association*, 23, 1-19.

DURBIN, J. (1967). Design of multi-stage surveys for the estimation of sampling errors. *Applied Statistics*, 16, 152-164.

HANSEN, M.H., and HURWITZ, W.N. (1943). On the theory of sampling from finite population. *Annals of Mathematical Statistics*, 14, 333-362.

HORVITZ, D.G., and THOMPSON, D.J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47, 663-685.

MIDZUNO, H. (1952). On the sampling system with probability proportionate to sum of sizes. *Annals of Institute of Statistical Mathematics,* 3, 99-107.

MUKHOPADHYAY, P. (1975). PPS sampling schemes to base HTE. *Bulletin of Calcutta Statistical Association*, 23, 21-44.

RAO, J.N.K. (1963). On two systems of unequal probability sampling without replacement. *Annals of Institute of Statistical Mathematics*, 15, 67-72.

RAO, J.N.K. (1965). On two simple schemes of unequal probability sampling without replacement. *Journal of the Indian Statistical Association*, 3, 173-180.

SANKARANARAYANAN, K. (1969). An IPPS sampling scheme using Lahiri's method of selection. *Journal of the Indian Society of Agricultural Statistics*, 21, 58-66.

SEN, A.R. (1952). Further developments of the theory and application of primary sampling units with special reference to the North Carolina agricultural population. Ph.D. Thesis, North Carolina State College, Raleigh.