ISSN 1684-8403
J. Stat.
Vol. 10, No.1 (2003) PP 47-51

A SIMPLE PROCEDURE FOR UNEQUAL PROBABILITY SAMPLING WITHOUT REPLACEMENT AND A SAMPLE OF SIZE 2.

*Muhammad Qaiser Shahbaz and **Muhammad Hanif

ABSTRACT: A new selection procedure has been developed for use with the Horvitz – Thompson estimator and a sample of size 2. Some important results have been verified for the first order and second order inclusion probabilities. Empirical study has also been carried out to see the performance of the new selection procedure in comparison with some of the famous selection procedures available in the literature.

KEY WORDS: Unequal Probability Sampling, Horvitz–Thompson estimator.

1. INTRODUCTION

Horvitz and Thompson (1952) provide the basis of unequal probability sampling without replacement when they developed following estimator for population total:

$$y'_{HT} = \sum_{i \in s} \frac{Y_i}{\pi_i}, \tag{1.1}$$

where π_i is probability of inclusion of i-th unit in the sample.

Horvitz and Thompson gave following variance formula for estimator (1.1)

$$V(y'_{HT}) = \sum_{i=1}^{N} \frac{(1-\pi_i)}{\pi_i} Y_i^2 + \sum_{\substack{i,j=1 \\ j \neq i}}^{N} \sum \frac{(\pi_{ij} - \pi_i \pi_j)}{\pi_i \pi_j} Y_i Y_j \tag{1.2}$$

an alternative expression, for fixed n, given by Sen (1953) and independently by Yates and Grundy (1953), is:

$$V(y'_{HT}) = \sum_{j>i}^{N} \sum (\pi_i \pi_j - \pi_{ij}) \left(\frac{Y_i}{\pi_i} - \frac{Y_j}{\pi_j} \right)^2 \tag{1.3}$$

An unbiased estimator for variance given in (1.3) was proposed by Horvitz and Thompson. The estimator is given as:

$$var_{HT}(y'_{HT}) = \sum_{i=1}^{n} \frac{1-\pi_i}{\pi_i^2} y_i^2 + \sum_{\substack{i=1 \\ j \neq i}}^{n} \sum_{j=1}^{n} \frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j \pi_{ij}} y_i y_j \tag{1.4}$$

The variance estimator given in (1.4) is unbiased for variance given in (1.2) but it may assume negative values for some of the pairs. An unbiased estimator for variance given in (1.3) was proposed by Sen (1953) and independently by Yates and Grundy (1953). The estimator is given as:

*Department of Statistics, Govt. College University Lahore (Pakistan).
**National College of Business Administration and Economics, Lahore (Pakistan).

$$var_{SYG}\left(y'_{HT}\right)=\frac{1}{2}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}\frac{\pi_i\pi_j-\pi_{ij}}{\pi_{ij}}\left(\frac{y_i}{\pi_i}-\frac{y_j}{\pi_j}\right)^2 \tag{1.5}$$

The estimator given in (1.5) is rarely seems to assume negative values. Since the time of introduction of Horvitz and Thompson estimator a number of selection procedures have been developed which can be used with this estimator. A comprehensive bibliography of these can be found in Hanif and Brewer (1980), Brewer and Hanif (1983) and Shahbaz (2001).

2. NEW SELECTION PROCEDURE FOR n = 2

In this section we have given a new selection procedure for use with the Horvitz – Thompson (1952) estimator and a sample of size 2. The new selection procedure is given as under:

Select first unit with probability proportional to $\dfrac{2p_i(1-p_i)}{1-4p_i}$

Select second unit with probability proportional to size of the remaining units.

The expressions for π_i and π_{ij} are derived as:

$$\pi_i = \frac{\dfrac{2p_i(1-p_i)}{1-4p_i}}{\sum_{i=1}^{N}\dfrac{2p_i(1-p_i)}{1-4p_i}}+\sum_{\substack{j=1\\ j\neq i}}^{N}\frac{\dfrac{2p_j(1-p_j)}{1-4p_j}}{\sum_{j=1}^{N}\dfrac{2p_j(1-p_j)}{1-4p_j}}\cdot\frac{p_i}{1-p_j}$$

After simplification we obtain

$$\pi_i = \frac{2p_i}{b}\left[1+\frac{1}{1-4p_i}+2\sum_{j=1}^{N}\frac{p_j}{1-4p_j}\right] \qquad for \quad p_i,p_j<\frac{1}{4} \tag{2.1}$$

where $b=1+3\sum_{i=1}^{N}\dfrac{p_i}{1-4p_i}$ (2.2)

Again

$$\pi_{ij} = p_i p_{j|i}+p_j p_{i|j} = \frac{\dfrac{2p_i(1-p_i)}{1-4p_i}}{\sum_{i=1}^{N}\dfrac{2p_i(1-p_i)}{1-4p_i}}\cdot\frac{p_j}{1-p_i}+\frac{\dfrac{2p_j(1-p_j)}{1-4p_j}}{\sum_{j=1}^{N}\dfrac{2p_j(1-p_j)}{1-4p_j}}\cdot\frac{p_i}{1-p_j}$$

After some algebra we have:

$$\pi_{ij} = \frac{4p_ip_j}{b}\left[\frac{1}{1-4p_i}+\frac{1}{1-4p_j}\right] \quad for \quad p_i, p_j < \frac{1}{4} \tag{2.3}$$

3. SOME RESULTS FOR NEW SELECTION PROCEDURE

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

Result–1: $\sum_{i=1}^{N}\pi_i = n$ for this selection procedure.

Proof: To prove this result considers (2.1) as:

$$\pi_i = \frac{2p_i}{b}\left[1+\frac{1}{1-4p_i}+2\sum_{j=1}^{N}\frac{p_j}{1-4p_j}\right] \tag{2.1}$$

Summing both sides:

$$\sum_{i=1}^{N}\pi_i = \sum_{i=1}^{N}\left[\frac{2p_i}{b}\left\{1+\frac{1}{1-4p_i}+2\sum_{j=1}^{N}\frac{p_j}{1-4p_j}\right\}\right]$$

$$= \frac{2}{b}\sum_{i=1}^{N}p_i\left[1+\frac{1}{1-4p_i}+2\sum_{j=1}^{N}\frac{p_j}{1-4p_j}\right] = 2.$$

Result–2: The quantity π_{ij}, obtained under this selection procedure, satisfies the relation $\sum_{\substack{j=1 \\ j\neq i}}^{N}\pi_{ij} = (n-1)\pi_i$.

Proof: Consider (2.3) as:

$$\pi_{ij} = \frac{4p_ip_j}{b}\left[\frac{1}{1-4p_i}+\frac{1}{1-4p_j}\right] \tag{2.3}$$

Summing π_{ij} when $(j \neq i)$:

$$\sum_{\substack{j=1\\j\neq i}}^{N}\pi_{ij}=\sum_{\substack{j=1\\j\neq i}}^{N}\left[\frac{4p_ip_j}{b}\left\{\frac{1}{1-4p_i}+\frac{1}{1-4p_j}\right\}\right]$$

$$=\frac{4p_i}{b}\left[\frac{1}{1-4p_i}\sum_{\substack{j=1\\j\neq i}}^{N}p_j+\sum_{\substack{j=1\\j\neq i}}^{N}\frac{p_j}{1-4p_j}\right]$$

$$=\frac{2p_i}{b}\left[1+\frac{1}{1-4p_i}+2\sum_{j=1}^{N}\frac{p_j}{1-4p_j}\right]=\pi_i$$

Result–3: The quantity π_{ij}, obtained under this selection procedure, satisfies the relation $\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}\pi_{ij}=n(n-1)$ where n is the sample size.

Proof: The proof immediately followed from results (1) and (2).

Result–4: The Sen – Yates – Grundy variance estimator is always positive under this selection procedure:

Proof: The Sen–Yates–Grundy variance estimator given in (1.5). The quantities π_i and π_{ij} under the new selection procedure are given in (2.1) and (2.3). Now, the Sen–Yates–Grundy variance estimator is positive if:

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

Now

$$\pi_i\pi_j-\pi_{ij} =$$

$$\frac{2p_i}{b}\left[1+\frac{1}{1-4p_i}+2\sum_{h=1}^{N}\frac{p_h}{1-4p_h}\right]\cdot\frac{2p_j}{b}\left[1+\frac{1}{1-4p_j}+2\sum_{h=1}^{N}\frac{p_h}{1-4p_h}\right]$$
$$-\frac{4p_ip_j}{b}\left[\frac{1}{1-4p_i}+\frac{1}{1-4p_j}\right]$$

Writing $B=2\sum_{h=1}^{N}\frac{p_h}{1-4p_h}$, above equation can be written as:

$$\pi_i \pi_j - \pi_{ij} =$$

$$\frac{4 p_i p_j}{b}\left[\frac{1}{b}\left\{1+\frac{1}{1-4 p_i}+B\right\}\left\{1+\frac{1}{1-4 p_j}+B\right\}-\left\{\frac{1}{1-4 p_i}+\frac{1}{1-4 p_j}\right\}\right]$$

$$=$$

$$\frac{4 p_i p_j}{b}\left[\frac{1}{b}\left\{(B+1)^2+\frac{2(B+1)(1-2 p_i-2 p_i)+1}{(1-4 p_i)(1-4 p_j)}\right\}-\frac{2(1-2 p_i-2 p_j)}{(1-4 p_i)(1-4 p_j)}\right] \quad (3.1)$$

Now since the first term within the main brackets of equation (3.1) is always greater than or equal to the second term therefore (3.1) is always non-negative for all values of p_i and p_j, making Sen–Yates–Grundy variance estimator non-negative for all samples. Further, it has been numerically checked that the relation $\pi_i \pi_j - \pi_{ij} \geq 0$ is true for all values of π_i, π_j and π_{ij} for this selection procedure.

REFERENCES

1. Brewer, K. R. W. and Hanif, M. (1983) "Sampling with Unequal Probabilities", Lecture notes to Statistics, No. 15, Springer – Verlag.
2. Hanif, M. and Brewer, K. R. W. (1980) "Sampling with unequal probabilities without replacement; a review", Inter. Stat. Rev. 48(3), 317 – 335.
3. Horvitz, D. G. and Thompson, D. J. (1952) "A generalization of sampling without replacement from a finite universe", J. Amer. Stat. Assoc. 47, 663 – 685.
4. Sen, A. R. (1953) "On the estimate of the variance in sampling with varying probabilities", J. Ind. Soc. Agri. Stat., 5, 119 – 127.
5. Shahbaz, M. Q. (2001) "Probability proportional to size sampling without replacement" Unpublished M.Phil. Thesis.
6. Yates, F. and Grundy, P. M. (1953) "Selection without replacement from within strata with probability proportional to size", J. Roy. Stat. Soc., B, 15, 153 – 161.