Pak J. Statist.
2003 Vol. 19(1) pp 151 – 160

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

Muhammad Qaiser Shahbaz
Department of Statistics, Govt. College University Lahore.

and

Muhammad Hanif
National College of Business Administration and Economics, Lahore

ABSTRACT

A new selection procedure has been developed for use with the Hurvitz – 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, Hurvitz – Thompson estimator, Brewer and Durbin Selection procedures, Variance estimators.

1. INTRODUCTION

The concept of sampling with unequal probabilities with replacement was given by Hansen and Horwitz (1943) and unequal probability sampling without replacement was first introduced by Madow (1949) and subsequently by Narian (1951) but no theoretical framework was given. Hurvitz and Thompson (1952) were the first to give theoretical framework of unequal probability sampling without replacement. The estimator proposed by Hurvitz and Thompson (1952) was:

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

where p_i is probability of inclusion of i-th unit to be in the sample.

The variance of (1.1) was:

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

with an unbiased variance estimator

$$var_{HT}(y'_{HT}) = \sum_{i=1}^{n} \frac{1-\pi_i}{\pi_i^2} y_i^2 + \sum_{\substack{i=1 \\ j \neq i}}^{n} \sum_{j=1} \frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j \pi_{ij}} y_i y_j \tag{1.3}$$

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

$$V(y'_{HT}) = \sum_{j>i}^{N} \sum (\pi_i \pi_j - \pi_{ij}) \left(\frac{Y_i}{\pi_i} - \frac{Y_j}{\pi_j} \right)^2 \tag{1.4}$$

with an unbiased variance estimator:

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

Since the time of introduction of Hurvitz 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) and Brewer and Hanif (1983). In the following section new selection procedure has been developed.

2. NEW SELECTION PROCEDURE FOR n = 2

In this section we have given a new selection procedure for use with the Hurvitz – Thompson (1952) estimator with sample size 2.

- Select first unit with probability proportional to $\frac{p_i}{1-2p_i}$ and without replacement
- Select second unit with probability proportional to size, p_i of the remaining units

The probability of inclusion π_i for the i-th unit to be in sample:

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

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

$$= \frac{p_i}{d}\left[\frac{1}{1-p_i} + \sum_{j=1}^{N}\frac{p_j}{(1-p_j)(1-2p_j)}\right], \qquad (2.1)$$

where $d = \sum_{i=1}^{N}\frac{p_i}{1-2p_i}$ (2.2)

The joint probability of inclusion for i-th and j-th units in the sample for this selection procedure is given as:

$$\pi_{ij} = p_i\, p_{j|i} + p_j\, p_{i|j}$$

$$= \frac{\frac{p_i}{1-2p_i}}{\sum_{i=1}^{N}\frac{p_i}{1-2p_i}}\cdot\frac{p_j}{1-p_i} + \frac{\frac{p_j}{1-2p_j}}{\sum_{j=1}^{N}\frac{p_j}{1-2p_j}}\cdot\frac{p_i}{1-p_j}$$

$$= \frac{p_i\ p_j}{d}\left[\frac{1}{(1-p_i)(1-2p_i)}+\frac{1}{(1-p_j)(1-2p_j)}\right] \qquad (2.3)$$

a. SOME RESULTS FOR NEW SELECTION PROCEDURE

Result–1: The values of p_i and p_{ij} reduces to the standard results of simple random sampling for $p_i = p_j = \frac{1}{N}$.

Putting $p_i = p_j = \frac{1}{N}$ in (2.1), we get:

$$\pi_i = \frac{1/N}{d}\left[\frac{1}{1-1/N}+\sum_{j=1}^{N}\frac{1/N}{(1-1/N)(1-2/N)}\right] = \frac{2}{d(N-2)} \qquad (3.1)$$

$$d = \sum_{i=1}^{N}\frac{p_i}{1-2p_i} = \sum_{i=1}^{N}\frac{1/N}{1-2/N} = \frac{N}{N-2} \qquad (3.2)$$

From (3.2) and (3.1) we get: $\pi_i = \frac{2}{N}$. This is for simple random sampling without replacement.

For π_{ij}, substituting $p_i = p_j = \frac{1}{N}$ and the value of d, we get:

$$\pi_{ij} = \frac{1/N^2}{N/(N-2)}\left[\frac{1}{(1-1/N)(1-2/N)}+\frac{1}{(1-1/N)(1-2/N)}\right] \qquad (3.3)$$

which is the joint probability of inclusion for ith and jth units to be in the sample for a sample of size two in case of simple random sampling without replacement.

Result – 2: $\sum_{i=1}^{N} p_i = n$

Summing both sides of (2.1)

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

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

$$= \frac{2}{d}\sum_{i=1}^{N}\frac{p_i}{1-2p_i}=2. \qquad (3.4)$$

Result – 3: The quantity π_{ij}, satisfies the relation $\sum_{\substack{j=1\\ j\neq i}}^{N}\pi_{ij} = (n-1)\pi_i$.

For this summing π_{ij} given in (2.3)

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

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

$$= \frac{p_i}{d}\left[\frac{1}{(1-p_i)}+\sum_{j=1}^{N}\frac{p_j}{(1-p_j)(1-2p_j)}\right] \qquad (3.5)$$

Comparing (3.5) with (2.1), it can be seen that $\sum_{\substack{j=1\\ j\neq i}}^{N}\pi_{ij} = \pi_i$. Since n is 2,it can be easily verified that $\sum_{\substack{j=1\\ j\neq i}}^{N}\pi_{ij} = (n-1)\pi_i$.

Result – 4: The quantity π_{ij}, also satisfies the relation $\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}} \boldsymbol{p}_{ij} = n(n-1)$,

where n is the sample size.

Applying double summation on both sides of (2.3):

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

$$= \sum_{i=1}^{N}\left[\sum_{\substack{j=1\\ j\neq i}}^{N}\left\{\frac{p_i\, p_j}{d}\left(\frac{1}{(1-p_i)(1-2\,p_i)}+\frac{1}{(1-p_j)(1-2\,p_j)}\right)\right\}\right] \tag{3.6}$$

Also

$$\sum_{\substack{j=1\\ j\neq i}}^{N} \boldsymbol{p}_{ij} = \sum_{\substack{j=1\\ j\neq i}}^{N}\left[\frac{p_i\, p_j}{d}\left\{\frac{1}{(1-p_i)(1-2\,p_i)}+\frac{1}{(1-p_j)(1-2\,p_j)}\right\}\right] = \boldsymbol{p}_i \tag{3.7}$$

Substituting (3.6) in (3.7)

$$\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}} \boldsymbol{p}_{ij} = \sum_{i=1}^{N} \boldsymbol{p}_i = 2 \tag{3.8}$$

Since n = 2, therefore equation (3.8) can be written as $\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}} \boldsymbol{p}_{ij} = n(n-1)$.

Result 5: The Sen – Yates – Grundy variance estimator is always positive under this selection procedure: i.e. $\pi_i\, \pi_j - \pi_{ij} > 0$

or

$$\frac{p_i}{d}\left[\frac{1}{1-p_i}+\sum_{h=1}^{N}\frac{p_h}{(1-p_h)(1-2p_h)}\right]\cdot\frac{p_j}{d}\left[\frac{1}{1-p_j}+\sum_{h=1}^{N}\frac{p_h}{(1-p_h)(1-2p_h)}\right]$$
$$-\frac{p_i\,p_j}{d}\left[\frac{1}{(1-p_i)(1-2p_i)}+\frac{1}{(1-p_j)(1-2p_j)}\right]>0$$

Writing $A=\sum_{h=1}^{N}\frac{p_h}{(1-p_h)(1-2p_h)}$ above equation becomes

$$\frac{1}{d}\left[\frac{1}{1-p_i}+A\right]\left[\frac{1}{1-p_j}+A\right]-\left[\frac{1}{(1-p_i)(1-2p_i)}+\frac{1}{(1-p_j)(1-2p_j)}\right]>0$$

or $$\frac{1}{d}\left[\frac{1}{(1-p_i)(1-p_j)}+\frac{A(2-p_i-p_j)}{(1-p_i)(1-p_j)}+A^2+\frac{d}{(1-p_i)}+\frac{d}{(1-p_j)}\right]$$
$$-\left[\frac{2}{(1-2p_i)}+\frac{2}{(1-2p_j)}\right]>0$$

or

$$\frac{1}{d}\left[A^2+\frac{(A+d)(2-p_i-p_j)+1}{(1-p_i)(1-p_j)}\right]-\frac{4(1-p_i-p_j)}{(1-2p_i)(1-2p_j)}>0 \quad (3.9)$$

Equation (3.9) is true for all values of p_i and p_j. Hence Sen – Yates – Grundy variance estimator is positive for all samples.

4. EMPIRICAL STUDY

In this section we have given the results of the empirical studies to compare the new selection procedure with some of the well-known selection procedures. 15 natural populations and five selection procedure has been selected. These selection procedures are Sen – Midzuno procedure, [Midzuno (1952) and Sen (1953)], Yates – Grundy (1953) draw-by-draw procedure, Yates – Grundy (1953) rejective procedure, Brewer (1963) draw-by-draw procedure and Durbin (1953) rejective procedure. The result of this empirical study is given in table – 4.1. It can be concluded that the new selection procedure perform better than the other selection procedures in 7 out of 15 populations. The Yates – Grundy (1953) rejective procedure perform better in 6 out of 15 populations. From the study of coefficient of variation and correlation coefficient of the populations in which the new selection procedure is performing well than the other we can see that these populations have relatively low coefficient of variation for the measure of size, X, and moderate correlation coefficient between the variables X and Y. Since in practice we do not have any idea about the correlation coefficient between variables X and Y, we therefore use only the information regarding the coefficient of variation for variable X. It may be concluded that the new selection procedure will perform better than the other selection procedures in populations that have relatively low coefficient of variation for variable X. The new selection procedure can therefore be used with the Hurvitz – Thompson (1952) estimator that have low coefficient of variation for measure of size variable.

5. ACKNOWLEDGEMENTS

We are extremely thankful to the referee and Dr. Munir Ahmad for his constructive Comments, which improve the version of this paper.

TABLE – 4.1: VARIANCE OF VARIOUS SELECTION PROCEDURES

Pop. No.	CV (X)	ρ_{XY}	Sen – Mid	YG (dbd)	YG (Rej.)	Brewer	Dur. (Rej.)	New	Best
1.	0.1708	0.9736	1740.59	333.75	437.87	276.14	277.66	249.27	New
2.	0.1892	0.8020	195738.14	135870.34	132221.08	132359.27	132129.33	129430.63	New
3.	0.2636	0.8807	18249.91	8830.26	8777.41	8504.88	8481.25	8229.51	New
4.	0.7519	0.9714	1505043.13	181002.19	183594.23	177091.80	180367.11	180120.25	Brewer
5.	0.3829	0.8795	69.39	30.87	30.71	29.48	29.48	28.68	New
6.	0.4300	0.6726	362.19	388.89	347.40	402.02	415.00	414.75	YG (Rej)
7.	0.4085	0.7401	1654.17	1248.41	1224.65	1215.89	1208.62	1190.16	New
8.	0.3227	0.8080	646497.50	376569.28	364570.84	370586.00	372031.84	365347.06	YG (Rej)
9.	0.4892	0.7657	1326.13	1304.92	1201.04	1341.15	1366.61	1376.70	YG (Rej)
10.	0.4213	0.7342	28879.18	23854.06	22787.72	23755.04	24029.42	23710.00	YG (Rej)
11.	0.0595	0.4069	187.31	177.17	166.24	177.41	177.54	177.82	YG (Rej)
12.	1.1525	0.9888	2908.04	880.51	1256.86	621.98	743.99	536.93	New
13.	0.1225	0.9662	47.20	11.19	10.43	11.90	11.89	12.77	YG (Rej)
14.	0.0941	0.9809	76.35	11.20	12.53	11.79	11.86	12.99	YG (dbd)
15.	0.0074	0.9462	580.77	166.55	997.10	161.59	161.61	159.87	New

REFERENCES

1. Brewer, K. R. W. (1963a). A model of systematic sampling with unequal probabilities. *Aust. J. Stat*. 5, 5 – 13.
2. Brewer, K. R. W. and Hanif, M. (1983). *Sampling with Unequal Probabilities*. Lecture notes to Statistics, No. 15. Springer – Verlag.
3. Durbin, J. (1953a). Some results in sampling when the units are selected with unequal probabilities. *J. Roy. Stat. Soc*, B, 15, 262–269.
4. Hanif, M. and Brewer, K. R. W. (1980). Sampling with unequal probabilities without replacement; a review. *Inter. Stat. Rev*. 48(3), 317 – 335.
5. Hansen, M. H. and Hurwitz, W. N. (1943). On the theory of sampling from a finite population. *Ann. Math. Stat*. 14, 333 – 362.
6. Hurvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *J. Amer. Stat. Assoc*. 47, 663 – 685.
7. Jessen, R. J. (1978). *Statistical Survey Techniques.* John Wiley and Sons Inc., New York.
8. Madow, W. G. (1949). On the theory of systematic sampling II. *Ann. Math. Stat.* 20, 333 – 354.
9. Midzuno, H. (1952). On the sampling system with probability proportionate to sum of size. *Ann. Inst. Math. Stat*. 3, 99 – 107.
10. Narian, R. D. (1951). On sampling without replacement with varying probabilities. J. Indian Soc. Agri. Statist. 3, 169-175.
11. Sen, A. R. (1953). On the estimate of the variance in sampling with varying probabilities. *J. Ind. Soc. Agri. Stat*., 5, 119 – 127.
12. 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.