Empirical Comparison of Some Approximate Variance Formulae of Horvitz-Thompson Estimator

[1]*Yasir Hassan,* [2]*Muhammad Qaiser Shahbaz and* [3]*Muhammad Hanif*

[1]Department of Mathematics, The University of Lahore, Lahore, Pakistan
[2]Department of Mathematics, COMSATS Institute of Information Technology, Lahore, Pakistan
[3]Department of Mathematics, LUMS, Lahore, Pakistan

Abstract: In this paper we have compared various approximate formulae for variance of Horvitz-Thompson estimator using first order inclusion probabilities. The empirical study has been conducted with a view to pick the most appropriate approximation for exact variance of the Horvitz-Thompson estimator. The empirical study is based upon two most popular selection procedures for unequal probability sampling.

Key words: Unequal probability sampling . Horvitz-Thompson estimator . Approximate Variance formulae

INTRODUCTION

Unequal probability sampling is a widely used method of drawing a sample from a population to draw certain inferences. In this method various units of the population are given different probabilities of selection as well as of inclusion in the sample. The method leads to more efficient results in terms of smaller mean square error of the estimate. Hansen and Hurwitz [3] are thought to be pioneer in this branch of sampling when they proposed their tor of population total in unequal probability sampling with replacement. Horvitz and Thompson [4] developed the historical theoretical ideas of unequal probability sampling without replacement when they developed the following estimator of population total in unequal probability sampling without replacement:

$$y_{HT}^{/} = \sum_{i \in s} \frac{y_i}{\pi_i} \qquad (1.1)$$

Following expression for variance of (1.1) has been developed by in [7] and in [9]:

$$Var\left(y_{HT}^{/}\right) = \sum_{j>i}^{N}\sum \left(\pi_i \pi_j - \pi_{ij}\right)\left(\frac{Y_i}{\pi_i} - \frac{Y_j}{\pi_j}\right)^2 \qquad (1.2)$$

with following variance estimator:

$$var\left(y_{HT}^{/}\right) = \sum_{j>i}^{N}\sum \frac{\left(\pi_i \pi_j - \pi_{ij}\right)}{\pi_{ij}}\left(\frac{Y_i}{\pi_i} - \frac{Y_j}{\pi_j}\right)^2 \qquad (1.3)$$

where π_i is the probability of the ith population unit to be included in the sample and π_{ij} is the joint probability of ith and jth population units to be included in the sample.

The variance expression (1.2) and variance estimator (1.3) depend upon the joint inclusion probabilities π_{ij} that are very hard to compute with increase in the sample size. The complexities in computing (1.2) and (1.3) can be avoided by expressing π_{ij} in terms of π_i and π_j. Numerous survey statisticians have proposed approximate expression for variance of Horvitz-Thompson [4] estimator using only first order inclusion probabilities. The following approximate formula has been proposed in [6] that is correct to order N^1:

$$Var\left(y_{HT}^{/}\right) \approx \sum_{i=1}^{N} \pi_i\left(1 - \frac{n-1}{n}\pi_i\right)\left(\frac{Y_i}{\pi_i} - \frac{Y}{n}\right)^2 \qquad (1.4)$$

Hanif and Ahmad [2] and Shahbaz and Hanif [8] have proposed various approximate formulae for variance of Horvitz-Thompson [4] estimator using only first order inclusion probabilities.

In the following section we have given the approximate variance formulae that are compared in the empirical study.

APPROXIMATIONS COMPARED

Various approximations for variance of Horvitz and Thompson [4] estimator have been proposed from time to time that are based upon only first order inclusion probabilities. In this paper we have compared

Corresponding Author: Dr. Yasir Hassan, Department of Mathematics, The University of Lahore, Lahore, Pakistan

the following approximate variance formulae for variance of Horvitz-Thompson [4] estimator:

- Hanif and Ahmad [2]:

$$\text{Var}\left(y'_{HT}\right) \approx \sum_{i=1}^{N} \pi_i \left(1 - \frac{\pi_i}{2-\pi_i}\right) \left\{\frac{Y_i}{\pi_i} - \frac{Y}{n}\right\}^2 \qquad (2.1)$$

- Shahbaz and Hanif [8]:

$$\text{Var}\left(y'_{HT}\right) \approx \sum_{i=1}^{N} \pi_i \left(1 - \frac{2\pi_i}{2-\pi_i}\right) \left\{\frac{Y_i}{\pi_i} - \frac{Y}{n}\right\}^2 \qquad (2.2)$$

- Hanif and Ahmad [2]:

$$\text{Var}\left(y'_{HT}\right) \approx \sum_{i=1}^{N} \pi_i \left\{1 - (2\pi_i)\Big/\left(4 - \sum_{k=1}^{N} \pi_k^2\right)\right\} \left\{\frac{Y_i}{\pi_i} - \frac{Y}{n}\right\}^2 \quad (2.3)$$

- Shahbaz and Hanif [8]:

$$\text{Var}\left(y'_{HT}\right) \approx \sum_{i=1}^{N} \pi_i \left\{1 - (4\pi_i)\Big/\left(4 - \sum_{k=1}^{N} \pi_k^2\right)\right\} \left\{\frac{Y_i}{\pi_i} - \frac{Y}{n}\right\}^2 \quad (2.4)$$

- Khawaja [5]:

$$\begin{aligned}
\text{V}\left(y'_{HT}\right) = &\sum_{i=1}^{N} \pi_i \left(1 - \frac{\pi_i}{(2-\pi_i)^2}\right)\left(\frac{Y_i}{\pi_i} - \frac{Y}{n}\right)^2 \\
&+ \left\{\sum_{i=1}^{N}\left(\frac{\pi_i}{2-\pi_i}\right)\left(\frac{Y_i}{\pi_i} - \frac{Y}{n}\right)\right\}^2
\end{aligned} \qquad (2.5)$$

The following section contains the empirical study of these approximate variance formulae.

EMPIRICAL STUDY

The preference of the approximate variance formulae of variance of Horvitz-Thompson [4] estimator depend upon its performance in approximating the variance. An approximation is more accurate if it closely approximates the actual variance of Horvitz-Thompson [4] estimator given in (1.2). The direct comparison of (1.2) with (2.1) to (2.5) is again based upon the joint inclusion probabilities. We, therefore, have conducted the empirical study to decide about the best approximation. In this empirical study we have used fifty natural populations from literature

Table 1: Absolute relative error of various approximations

Approximation	Average absolute relative error	
	Yates-Grundy	Brewer
1	0.00284	0.11110
2	0.01109	0.10115
3	0.03737	0.00469
4	0.07422	0.01355
5	0.07097	0.03801

on survey sampling. We have first computed the actual variance of Horvitz-Thompson [4] estimator given in (1.2) by using two most popular selection procedure of unequal probability sampling given by Yates and Grundy [9] and by Brewer [1]. We have used these two procedures as in Yates-Grundy [9] procedure the inclusion probabilities as not directly proportional to measure of size whereas in the Brewer [1] procedure the inclusion probabilities are directly proportional to size. We have then computed the approximate variance of the estimator given in (2.1) through (2.5) under the same selection procedures. After computing the exact and approximate variance we have computed the Absolute Relative Error of each approximation by using:

$$\text{ARE} = \frac{\left|\text{ActualVariance} - \text{ApproximateVariance}\right|}{\text{ActualVariance}} \qquad (3.1)$$

We have then computed the Average *ARE*. The Average *ARE* of various approximations under Yates and Grundy [9] and Brewer [1] procedure is given in the following table.

From Table 1 we can readily see that the for Yates-Grundy [9] procedure; which do not provide inclusion probabilities directly proportional to size; the Approximation-1 have smallest Average *ARE*. We can therefore conclude that for selection procedures for which inclusion probabilities are not directly proportional to size, the Approximation-1 should be used. Table 1 also indicates that the Average *ARE* of Approximation-3 is smallest for Brewer procedure which provide inclusion probabilities which are directly proportional to size. We can therefore conclude that the Approximation-3 is useful for selection procedure which provides inclusion probabilities directly proportional to size.

We have also compared the various approximations on the basis of ranks of the *ARE*. For this we have ranked five approximations with respect to there *ARE*. The approximation with smallest *ARE* in a population is assigned a rank of 1 and so on. We

Table 2: Average rank of ARE under Yates-Grundy procedure

CV	App.-1	App.-2	App.-3	App.-4	App.-5
1-10	1.0	2.0	3.4	4.6	4.0
11-20	1.0	2.1	3.6	4.6	3.7
21-30	1.1	2.0	3.7	4.6	3.6
31-40	1.1	2.0	3.6	4.1	4.2
41-50	1.2	2.1	3.5	4.0	4.2

Table 3: Average rank of ARE under Brewer procedure

CV	App.-1	App.-2	App.-3	App.-4	App.-5
1-10	3.1	2.0	1.2	4.6	4.0
11-20	3.0	2.1	1.1	4.6	3.7
21-30	2.8	2.0	1.2	4.6	3.6
31-40	3.1	2.0	1.0	4.1	4.2
41-50	2.0	2.1	1.2	4.0	4.2

have then computed the average rank for various ranges of coefficient of variation of the measure of size for each population. The coefficient of variation of the population is again grouped as smallest 10 to largest 10. The average ranks of *ARE* for various ranges of coefficient of variation are given in the table below:

From Table 2 we again see that Approximation-1 outperform other approximations in the comparison for Yates-Grundy [9] procedure as this has smallest average rank for all ranges of coefficient of variation. For Brewer [1] procedure the Approximation-3 outperforms other approximations.

CONCLUSION AND RECOMMENDATION

The empirical study has been given in the previous section. From this empirical study we can readily see that Approximation-1 and Approximation-3 outperform other approximations in the study. We therefore conclude that for selection procedures that do not provide inclusion probabilities directly proportional to size, the Approximation-1 can be used to compute the variance of Horvitz-Thompson estimator. Also for selection procedures that provide inclusion probabilities directly proportional to size, the Approximation-3 can be used for computing variance of Horvitz-Thompson [4] estimator.

REFERENCES

1. Brewer, K.R.W., 1963. A model of systematic sampling with unequal probabilities. Aust. J. Stat., 5: 5-13.

2. Hanif, M. and M. Ahmad, 2001. Approximate variance formula for Horvitz-Thompson estimator using first order probabilities. Presented at ICCSS-VII held in Lahore.

3. Hansen, M.H. and W.N. Hurwitz, 1943. On the theory of sampling from a finite population. Ann. Math. Stat. 14: 333-362.

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

5. Khawaja, I., 2008. Sampling with unequal probabilities. Unpublished PhD thesis submitted at NCBA&E, Lahore, Pakistan.

6. Rao, J.N.K., 1963. On two systems of unequal probability sampling without replacement. Ann. Inst. Stat. Math., 15: 67-72.

7. Sen, A.R., 1953. On the estimate of the variance in sampling with varying probabilities. J. Indian Soc. Agri. Statist., 5: 119-127.

8. Shahbaz, M.Q. and M. Hanif, 2003. Variance formulas for Horvitz-Thompson estimator using first order inclusion probabilities. J. Applied Statistical Sciences, 12 (3): 201-208.

9. Yates, F. and P.M. Grundy, 1953. Selection without replacement from within strata with probability proportional to size. J. Roy. Statist. Soc. B., 15: 153-261.