A New Approximate Formula for Variance of Horvitz–Thompson Estimator using first order Inclusion Probabilities

Muhammad Qaiser Shahbaz
Department of Statistics
GC University, Lahore, Pakistan
drshahbaz@gcu.edu.pk

Muhammad Hanif
Department of Mathematics
LUMS, Lahore, Pakistan
hanif@lums.edu.pk

Abstract

A new approximate formula for sampling variance of Horvitz–Thompson (1952) estimator has been obtained. Empirical study of the approximate formula has been given to see its performance.

Keywords: Unequal probability sampling. Horvitz–Thompson estimator, First order inclusion probabilities, Joint inclusion probabilities.

1. Introduction

Horvitz and Thompson (1952) propose an unbiased estimator of population total for unequal probability sampling without replacement:

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

with variance formula

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

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

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

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 expressions (1.2) and (1.3) require the calculations of joint inclusion probabilities π_{ij} which become tedious as the sample size increase. Many attempts have been made to approximate the sampling variance of Horvitz and Thompson estimator by using the first order inclusion probabilities. These attempts were based on approximation of joint inclusion probabilities π_{ij} with a suitable function of first order inclusion probabilities.

The first attempt was made by Hartley and Rao (1962) using random systematic sampling procedure:

$$Var\left(y_{HT}^{/}\right) \approx \sum_{i=1}^{N}\pi_i\left(1-\frac{n-1}{n}\pi_i\right)\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)^2-\frac{n-1}{n^2}\sum_{i=1}^{N}\left(2\pi_i^3-\frac{\pi_i^2}{2}\sum_{j=1}^{N}\pi_j^2\right)\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)^2+\frac{2(n-1)}{n^3}\left(\sum_{i=1}^{N}\pi_iY_i-\frac{Y}{n}\sum_{j=1}^{N}\pi_j^2\right)^2 \quad (1.4)$$

The expression (1.4) is correct to order N^0. Rao (1963) further showed that the asymptotic variance formula to order N^0 for a sample of size 2 is given as:

$$Var\left(y_{HT}^{/}\right) = \sum_{i=1}^{N}\pi_i\left(1-\frac{\pi_i}{2}\right)\left(\frac{Y_i}{\pi_i}-\frac{Y}{2}\right)^2-\frac{1}{2}\sum_{i=1}^{N}\left(\pi_i^3-\frac{\pi_i^2}{4}\sum_{j=1}^{N}\pi_j^2\right)\left(\frac{Y_i}{\pi_i}-\frac{Y}{2}\right)^2+\lambda\left(\sum_{i=1}^{N}\pi_iY_i-\frac{Y}{2}\sum_{j=1}^{N}\pi_j^2\right)^2 \quad (1.5)$$

Rao (1963) further showed that the approximate formula to the order N^1 for a sample of size n is:

$$Var\left(y_{HT}^{/}\right) \approx \sum_{i=1}^{N}\pi_i\left(1-\frac{n-1}{n}\pi_i\right)\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)^2 \quad (1.6)$$

Shahbaz and Hanif (2003) has given some more approximate formulae of variance of Horvitz–Thompson (1952) estimator that depends only on first order inclusion probabilities.

2. The Approximate Formula

In this section the approximate formula for variance of Horvitz–Thompson (1952) estimator has been obtained. To obtained this approximate formula consider an alternative expression for variance of Horvitz–Thompson (1952) estimator as:

$$Var\left(y_{HT}^{/}\right) = \sum_{i=1}^{N}\pi_i\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)^2-\sum_{i=1}^{N}\pi_i^2\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)^2+\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}\left(\pi_{ij}-\pi_i\pi_j\right)\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)\left(\frac{Y_j}{\pi_j}-\frac{Y}{n}\right) \quad (2.1)$$

The expression (2.1) contains the joint inclusion probabilities in the last expression only. To obtain the approximate formula consider the approximation:

$$\pi_{ij} = a_ia_j\pi_i\pi_j \quad (2.2)$$

Substituting the value given in (2.2) in the last term of (2.1) we have:

$$Last\ Term=\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}\left(\pi_{ij}-\pi_i\pi_j\right)\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)\left(\frac{Y_j}{\pi_j}-\frac{Y}{n}\right)$$

$$=\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}\left(a_i a_j\pi_i\pi_j-\pi_i\pi_j\right)\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)\left(\frac{Y_j}{\pi_j}-\frac{Y}{n}\right)$$

$$=\left[\sum_{i=1}^{N}a_i\pi_i\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)\right]^2-\sum_{i=1}^{N}a_i^2\pi_i^2\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)^2+\sum_{i=1}^{N}\pi_i^2\left(\frac{Y_j}{\pi_i}-\frac{Y}{n}\right)^2$$

Substituting the value of last expression in (2.1), the approximate variance formula for Horvitz–Thompson (1952) estimator using first order inclusion probabilities is obtained as:

$$Var\left(y_{HT}^{/}\right)\approx\sum_{i=1}^{N}\pi_i\left(1-a_i^2\pi_i\right)\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)^2+\left[\sum_{i=1}^{N}a_i\pi_i\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)\right]^2 \tag{2.3}$$

The approximation given in (2.3) is a slight modification given by Shahbaz and Hanif (2003).

1. Empirical Study

In this section the empirical study has been carried out to check the performance of some approximations given in section–2. The empirical study has been carried out by using the approximations:

$$Var\left(y'_{HT}\right)=\sum_{i=1}^{N}\pi_i\left\{1-\frac{2\pi_i}{2-\pi_i}\right\}\left(\frac{Y_i}{\pi_i}-\frac{Y}{2}\right)^2 \tag{3.1}$$

and $$Var\left(y'_{HT}\right)=\sum_{i=1}^{N}\pi_i\left\{1-\frac{\pi_i^3}{\left(2-\pi_i\right)^2}\right\}\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)^2+\left[\sum_{i=1}^{N}\frac{\pi_i}{2-\pi_i}\left(\frac{Y_i}{\pi_i}-\frac{Y}{2}\right)\right]^2 \tag{3.2}$$

In this empirical study the exact and approximate variance of Horvitz–Thompson (1952) estimator has been computed. After computing the variances, the relative percentage error of each approximation has been computed by using the expression:

$$RPE=100\left(\frac{Actual-Aproximated}{Actual}\right) \tag{3.3}$$

These relative percentage errors have been given in the table below:

Table 1: Relative Percentage Error of Various Approximations

Pop. No.	Approx 1	Approx 2	Pop. No.	Approx 1	Approx 2
1	6.55	-3.20	**11**	10.12	-5.25
2	0.35	-0.13	**12**	6.37	-3.06
3	-2.14	-1.28	**13**	14.86	-9.06
4	12.20	-5.35	**14**	6.99	-2.94
5	11.82	-5.16	**15**	7.27	-3.46
6	3.91	-1.34	**16**	11.90	-5.41
7	22.21	-10.68	**17**	7.97	-3.28
8	7.60	-3.58	**18**	8.51	-4.06
9	7.70	-3.63	**19**	5.49	-2.58
10	5.17	-2.56	**20**	-8.68	-1.99

From above table we can see that Approximation–2, that is equation (3.2) produce reasonably well results for variance of Horvitz–Thompson estimator.

References

1. Hartley, H. O. and Rao, J. N. K. (1962). Sampling with unequal probabilities and without replacement, Ann. Math. Stat. 33, 350 – 374.
2. Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe, J. Amer. Statist. Assoc. 47, 663 – 685.
3. Rao, J. N. K. (1963). On two systems of unequal probability sampling without replacement, Ann. Inst. Stat. Math., 15, 67 – 72.
4. Sen, A. R. (1953). On the estimate of the variance in sampling with varying probabilities, J. Indian Soc. Agri. Statist. 5, 119 – 127.
5. Shahbaz, M. Q. and Hanif, M. (2003). Variance formulas for Horvitz–Thompson estimator using first order inclusion probabilities. J. Applied Statistical Sciences. Vol. 12 (3), 201–208.
6. Yates, F. and Grundy, P. M. (1953). Selection without replacement from within strata with probability proportional to size, J. Roy. Statist. Soc. B. 15, 153 – 261.