Stability of Variance Estimator of Modified Murthy Estimator

Muhammad Qaiser Shahbaz
Department of Statistics, GC University
Lahore, Pakistan

Mariyam Hafeez
Department of Statistics, GC University
Lahore, Pakistan

Abstract

Stability of the estimator of sampling variance is very important in using the estimator for estimation of sampling variance. In this paper the stability of the estimator of sampling variance of Modified Murthy (1957) estimator given by Shahbaz (2004) has been carried out by using the super population model.

Keywords: Unequal Probability Sampling, Stability of Variance Estimator.

1. Introduction

Horvitz and Thompson (1952) were the first who developed the theoretical ground of unequal probability sampling without replacement and provide the following estimator of population total:

$$y_{HT}^{/} = \sum_{i \in S} \frac{y_i}{\pi_i} \tag{1.1}$$

along with the following designed based variance:

$$V\left(y_{HT}^{/}\right) = \sum_{i=1}^{N} \frac{\left(1-\pi_i\right)}{\pi_i} Y_i^2 + \sum_{\substack{i=1 \\ j \neq i}}^{N} \sum_{j=1} \frac{\left(\pi_{ij} - \pi_i \pi_j\right)}{\pi_i \pi_j} Y_i Y_j \tag{1.2}$$

Sen (1953) and, independently, Yates and Grundy (1953) obtained the following model based variance of the Horvitz–Thompson (1952) estimator:

$$V\left(y_{HT}^{/}\right) = \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j > i}} \left(\pi_i \pi_j - \pi_{ij}\right) \left(\frac{Y_i}{\pi_i} - \frac{Y_j}{\pi_j}\right)^2 \tag{1.3}$$

Under the linear stochastic model:

$$\left. \begin{array}{l} y_i = \beta z_i + \varepsilon_i \\ with \quad E\left(\varepsilon_i\right) = 0 \\ \qquad E\left(\varepsilon_i, \varepsilon_j\right) = \begin{cases} \sigma_i^2 & i = j \\ 0 & otherwise \end{cases} \\ where \quad \sigma_i^2 = \sigma^2 z_i^{2\gamma} \quad \frac{1}{2} \leq \gamma \leq 1 \end{array} \right\} \tag{1.4}$$

Bayless and Rao (1970) have shown that the expected variance of $y_{HT}^{/}$ is:

$$E_M\left[Var\left(y'_{HT}\right)\right]=\sigma^2\left(\frac{Z}{n}\right)^{2\gamma}\sum_{I=1}^{N}\left(1-\pi_I\right)\pi_I^{2\gamma-1} \tag{1.5}$$

Rao and Bayless (1969) and Bayless and Rao (1970) have studied the stability of the variance estimator given by Sen (1953) and Yates and Grundy (1953).

Murthy (1957) symmetrize the Raj (1956) estimator to obtain the following unbiased estimator of population total:

$$t_{symm}=\frac{1}{2-p_i-p_j}\left[\frac{y_i}{p_i}\left(1-p_j\right)+\frac{y_j}{p_j}\left(1-p_i\right)\right] \tag{1.6}$$

The variance of (1.6), obtained by Murthy (1957), is:

$$Var\left(t_{symm}\right)=\frac{1}{2}\sum_{\substack{i=1\\j\neq i}}^{N}\sum_{j=1}^{N}\frac{P_iP_j\left(1-P_i-P_j\right)}{2-P_i-P_j}\cdot\left(\frac{Y_i}{P_i}-\frac{Y_j}{P_j}\right)^2 \tag{1.7}$$

The expected variance of (1.6) under model (1.4), obtained by Rao and Bayless (1969), is:

$$E_M\left[Var\left(t_{symm}\right)\right]=\sigma^2\left(\frac{Z}{2}\right)^{2\gamma}\sum_{\substack{I=1\\J\neq I}}^{N}\sum_{J=1}^{N}\pi_I^{2\gamma-1}\pi_J\left(\frac{2-\pi_I-\pi_J}{4-\pi_I-\pi_J}\right) \tag{1.8}$$

Rao and Bayless (1969) and Bayless and Rao (1970) have conducted extensive empirical study for the expected variances given in (1.5) and (1.8).

Shahbaz (2004) have obtained the following modified Murthy estimator by using the Durbin (1953) selection procedure in the general Murthy (1957) estimator:

$$t_{MM(D)}=\frac{1}{2}\left[\frac{y_i}{p_i}+\frac{y_j}{p_j}\right] \tag{1.9}$$

The design based variance of (1.9) obtained by Shahbaz (2004) is:

$$Var\left(t_{MM(D)}\right)=\frac{1}{2}\sum_{\substack{i=1\\j\neq i}}^{N}\sum_{j=1}^{N}P_iP_j\left[1-\frac{\left(1-P_i-P_j\right)}{k\left(1-2P_i\right)\left(1-2P_j\right)}\right]\cdot\left\{\frac{Y_i}{P_i}-\frac{Y_j}{P_j}\right\}^2 \tag{1.10}$$

The estimator of variance given in (1.10) is given as:

$$var\left(t_{MM(D)}\right)=\frac{1}{4}\left[\frac{k\left(1-2p_i\right)\left(1-2p_j\right)}{\left(1-p_i-p_j\right)}-1\right]\left(\frac{y_i}{p_i}-\frac{y_j}{p_j}\right)^2 \tag{1.11}$$

In the following section we have studied the stability of (1.11) under the linear stochastic model (1.4) in order to see whether the estimator is admissible for estimation of variance of (1.9). The estimator (1.11) will be stable if under model (1.4) its variance remain positive for all values of the constant γ.

2. The Stability of Variance Estimator

In this section we have studied the stability of variance estimator given in (1.11). The stability of the variance estimator is judged by using the expression:

$$E_M E_D \left[var(y') - Var(y') \right]^2 = E_M E_D \left[var(y') \right]^2 - \left[E_M \left\{ Var(y') \right\} \right]^2 \qquad (2.1)$$

where $E_M E_D \left[var(y') \right]^2 = E_M \left[\sum_S \left\{ var(y') \right\}^2 P(S) \right]$ $\qquad (2.2)$

Now considering again (1.11) and squaring we have:

$$\left[var(t_{MM(D)}) \right]^2 = \left[\frac{2k(1-\pi_i)(1-\pi_j)}{(2-\pi_i-\pi_j)} - 1 \right]^2 \left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j} \right)^4$$

Applying the design expectation we have:

$$E_D \left[var(t_{MM(D)}) \right]^2 = \frac{1}{2} \sum_{i=1}^N \sum_{\substack{j=1 \\ j \neq i}} \left[\frac{2k(1-\pi_i)(1-\pi_j)}{(2-\pi_i-\pi_j)} - 1 \right]^2 \left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j} \right)^4 \cdot \frac{\pi_i \pi_j (2-\pi_i-\pi_j)}{2k(1-\pi_i)(1-\pi_j)}$$

Applying the model expectation we have:

$$E_M \left[E_D \left\{ var(t_{MM(D)}) \right\}^2 \right] = \frac{1}{2} \sum_{i=1}^N \sum_{\substack{j=1 \\ j \neq i}} \left[\frac{2k(1-\pi_i)(1-\pi_j)}{(2-\pi_i-\pi_j)} - 1 \right]^2 \left(\frac{3\sigma_i^4}{\pi_i^4} + \frac{6\sigma_i^2\sigma_j^2}{\pi_i^2\pi_j^2} + \frac{3\sigma_j^4}{\pi_j^4} \right)$$

$$\cdot \frac{\pi_i \pi_j (2-\pi_i-\pi_j)}{2k(1-\pi_i)(1-\pi_j)}$$

$$E_M \left[E_D \left\{ var(t_{MM(D)}) \right\}^2 \right] = \frac{3}{4} \sum_{i=1}^N \sum_{\substack{j=1 \\ j \neq i}} \left[\frac{2k(1-\pi_i)(1-\pi_j)}{(2-\pi_i-\pi_j)} - 1 \right]^2 \left(\frac{\sigma_i^2}{\pi_i^2} + \frac{\sigma_j^2}{\pi_j^2} \right)^2$$

$$\cdot \frac{\pi_i \pi_j (2-\pi_i-\pi_j)}{2k(1-\pi_i)(1-\pi_j)}$$

$$E_M \left[E_D \left\{ var(t_{MM(D)}) \right\}^2 \right] = \frac{3\sigma^4}{4} \left(\frac{Z}{2} \right)^{2\gamma} \sum_{i=1}^N \sum_{\substack{j=1 \\ j \neq i}} \left[\frac{2k(1-\pi_i)(1-\pi_j)}{(2-\pi_i-\pi_j)} - 1 \right]^2$$

$$\cdot \frac{\pi_i \pi_j (2-\pi_i-\pi_j)}{k(1-\pi_i)(1-\pi_j)} \cdot \left(\pi_i^{2\gamma-2} + \pi_j^{2\gamma-2} \right)^2 \qquad (2.3)$$

also $\quad E\left[Var(t_{MM(D)}) \right] = \sigma^2 \left(\frac{Z}{2} \right)^{2\gamma} \sum_{I=1}^N \sum_{\substack{J=1 \\ J \neq I}} \pi_I^{2\gamma-1} \pi_J \left\{ 1 - \frac{2-\pi_I-\pi_J}{2k(1-\pi_I)(1-\pi_J)} \right\}$ $\qquad (2.4)$

In the following section we have conducted the empirical study to see the stability of the variance estimator.

3. Empirical Study

In this section we have conducted the empirical study to see the stability of the variance estimator. The stability has been studied by using various values of parameter γ ranging from 0.5 to 1.0. The relative expected variance in the variance estimator, given as:

$$E_M E_D \left[var\left(t_{MM(D)}\right) - Var\left(t_{MM(D)}\right) \right]^2 \Big/ \left[E_M \left\{ Var\left(t_{MM(D)}\right) \right\} \right]^2 \tag{3.1}$$

The relative expected variances have been given in Table–1 below:

Table 1: Relative Expected Variances

Pop. No	Values of γ					
	0.5	0.6	0.7	0.8	0.9	1.0
1	0.9948	0.4900	0.1247	0.1426	0.3378	0.4850
2	0.4774	0.2031	0.2426	0.5045	0.9406	0.5084
3	0.1920	0.1132	0.3323	0.4928	0.6105	0.6976
4	0.2111	0.5183	0.6978	0.8050	0.8714	0.9137
5	0.5932	0.7464	0.8406	0.8990	0.9355	0.9585
6	0.4471	0.6522	0.7764	0.8533	0.9021	0.9338
7	0.6577	0.8017	0.8820	0.9283	0.9558	0.9724
8	0.7958	0.8951	0.9440	0.9692	0.9827	0.9901
9	0.8106	0.9057	0.9510	0.9736	0.9854	0.9918
10	0.9132	0.9639	0.9841	0.9927	0.9965	0.9983

Looking at the expected variances given in Table–1 we can readily see that for all choices of the constant γ, the variance estimator has stable performance as all the entries are positive and less then 1 which is a desirable property for a variance estimator to be stable.

References

1. Bayless, D. L. and Rao, J. N. K. (1970) "An empirical study of stabilities of estimators and variance estimators in unequal probability sampling for n = 3 and 4", *J. Amer. Stat. Assoc.* 65, 1645 – 1667.
2. Durbin, J. (1953) "Some results in sampling when the units are selected with unequal probabilities", *J. Roy. Stat. Soc*, B, 15, 262–269.
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. Murthy, M. N. (1957) "Ordered and unordered estimators in sampling without replacement", *Sankhya*, 18, 379–390.
5. Raj, D. (1956) "Some estimators in sampling with varying probabilities without replacement", *J. Amer. Stat. Assoc.* 51, 269 – 284.
6. Rao, J. N. K. and Bayless, D. L. (1969) "An empirical study of the stabilities of estimators and variance estimators in unequal probability sampling of two units per stratum", *J. Amer. Stat. Assoc.*, 64, 540 – 549.
7. Sen, A. R. (1953) "On the estimate of the variance in sampling with varying probabilities", *J. Ind. Soc. Agri. Stat.*, 5, 119 – 127.
8. Shahbaz, M. Q. (2004). A new estimator of population total using Durbin's selection procedure for unequal probability sampling. *Pak. J. Statist.* Vol. 20(2), 287–294.
9. 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.