ISSN 1684-8403
Journal of Statistics
Volume 14, 2007, pp. 13-19

A General Class of Selection Procedures and Modified Murthy Estimator

Abdul Basit[1] and Muhammad Qaisar Shahbaz[2]

Abstract

A new selection procedure for unequal probability sampling without replacement has been obtained for sample size "n". Some results regarding probability of inclusion and joint probability of inclusion have been verified. A new estimator of population total has been derived using the idea of Murthy (1957). Empirical study has also been carried out.

Keywords

Unequal probability sampling, Murthy estimator, Sen Midzuno selection procedures

1. Introduction

Unequal probability sampling was firstly introduced in early forties by Hansen and Hurwitz (1943). The Horvitz and Thompson (1952) were the first to give theoretical framework of unequal probability sampling without replacement. The estimator of population total proposed by Horvitz and Thompson (1952) is:

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

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

[1] Statistical Officer, State Bank of Pakistan, Karachi
Email: basit_ravian917@hotmail.com
[2] Department of Mathematics, COMSAT Institute of Information Technology, Lahore
Email: qshahbaz@gmail.com

The variance of (1.1) obtained by Sen (1953) and independently by Yates and Grundy (1953) is given as:

$$V(y'_{HT}) = \sum_{j>i}^{N}\sum\left(\pi_i\pi_j - \pi_{ij}\right)\left(\frac{Y_i}{\pi_i} - \frac{Y_j}{\pi_j}\right)^2 \tag{1.2}$$

where π_{ij} is joint probability of inclusion of i-th and j-th unit in the sample.

Large number of selection procedures have been developed for use with (1.1). A comprehensive review can be found in Brewer and Hanif (1983).

Murthy (1957) proposed his estimator of population total and is given as:

$$t_{symm} = \frac{1}{P(S)}\sum_{i=1}^{n} P(S/i)\,y_i \tag{1.3}$$

where $P\left(S\,|\,i\right)$ is probability of sample given that i-th unit is selected at the first draw and $P\left(S\right)$ is probability of the sample. Murthy (1957) estimator for a sample size 2, under Yates – Grundy (1953) procedure is:

$$t_{symm} = \frac{1}{2 - p_i - p_j}\left[\frac{y_i}{p_i}\left(1 - p_j\right) + \frac{y_j}{p_j}\left(1 - p_i\right)\right] \tag{1.4}$$

with

$$Var\left(t_{symm}\right) = \frac{1}{2}\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}\frac{P_i\,P_j\left(1 - P_i - P_j\right)}{2 - P_i - P_j}\cdot\left(\frac{Y_i}{P_i} - \frac{Y_j}{P_j}\right)^2 \tag{1.5}$$

2. The New Selection Procedure

In this section we have given a new selection procedure and we have also verified some important results for the probability of inclusion and joint probability of inclusion.

2.1 The Procedure

The new selection procedure is given as under:

- Select first unit with probability proportional to $p_i^{\alpha}\ \left(1 - p_i^{\beta}\right)\big/\left(1 - 2\ \ p_i^{\beta}\right)$ and without replacement
- Select a random sample of size (n-1) from the remaining (N-1) units.

The probability of inclusion for the i-th unit in the sample for this selection procedure is given as:

$$\pi_i = \frac{p_i^{\alpha}\left(1-p_i^{\beta}\right)/\left(1-2\,p_i^{\beta}\right)}{d} + \frac{n-1}{(N-1)\,d}\sum_{j\neq i=1}^{N}\frac{p_j^{\alpha}\left(1-p_j^{\beta}\right)}{\left(1-2\,p_j^{\beta}\right)\left(1-p_j\right)} \tag{2.1}$$

$$\pi_i = \frac{1}{d}\left[\frac{p_i^{\alpha}\left(1-p_i^{\beta}\right)}{\left(1-2\,p_i^{\beta}\right)}\left\{1-\frac{n-1}{N-1}\right\}+\frac{(n-1)}{N-1}d\right]$$

$$\pi_i = \frac{1}{d}\left[\frac{p_i^{\alpha}\left(1-p_i^{\beta}\right)}{\left(1-2\,p_i^{\beta}\right)}\left\{\frac{N-n}{N-1}\right\}+\frac{(n-1)}{N-1}d\right] \tag{2.2}$$

where $d = \sum_{i=1}^{N}\frac{p_i^{\alpha}\left(1-p_i^{\beta}\right)}{\left(1-2\,p_i^{\beta}\right)}$

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

$$\pi_{ij} = \frac{p_i^{\alpha}\left(1-p_i^{\beta}\right)/\left(1-2\,p_i^{\beta}\right)}{\sum_{i=1}^{N}p_i^{\alpha}\left(1-p_i^{\beta}\right)/\left(1-2\,p_i^{\beta}\right)}\frac{n-1}{N-1} + \frac{p_j^{\alpha}\left(1-p_j^{\beta}\right)/\left(1-2\,p_j^{\beta}\right)}{\sum_{j=1}^{N}p_j^{\alpha}\left(1-p_j^{\beta}\right)/\left(1-2\,p_j^{\beta}\right)}\frac{n-1}{N-1}$$

$$+\left(1-\frac{p_i^{\alpha}\left(1-p_i^{\beta}\right)}{d\left(1-2\,p_j^{\beta}\right)}-\frac{p_j^{\alpha}\left(1-p_j^{\beta}\right)}{d\left(1-2\,p_j^{\beta}\right)}\right)\left(\frac{n-1}{N-1}\right)\left(\frac{n-2}{N-2}\right)$$

$$\pi_{ij} = \frac{(n-1)(N-n)}{d\,(N-1)(N-2)}\left[\frac{p_i^{\alpha}\left(1-p_i^{\beta}\right)}{1-2\,p_i^{\beta}}+\frac{p_j^{\alpha}\left(1-p_j^{\beta}\right)}{1-2\,p_j^{\beta}}\right]+\frac{(n-1)(n-2)}{(N-1)(N-2)} \tag{2.3}$$

For the values of constant $\alpha = 1$ and $\beta = \infty$, (2.2) and (2.3) reduces to the expression of probability of inclusion and joint probability of inclusion of Midzuno (1952) selection procedure.

2.2 Some Results For New Selection Procedure

In this section we have verified some of the common results for the quantities π_i and π_{ij} obtained under the new selection procedure.

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

Result – 2: $\sum_{i=1}^{N} \pi_i = n$ for this selection procedure.

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

Result – 4: 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.

3. Modified Murthy Estimator

In this section a new estimator of population total with unequal probability is derived with its variance. The estimator has been derived by using the new selection procedure in (1.3). Now for new selection procedure, for a sample of size 2, we have:

$$P(S/i) = P(S/j) = \frac{1}{N-1} \quad \text{and}$$

$$P(S) = \frac{1}{d(N-1)}\left[\frac{p_i^{\alpha}\left(1-p_i^{\beta}\right)}{1-2p_i^{\beta}} + \frac{p_j^{\alpha}\left(1-p_j^{\beta}\right)}{1-2p_j^{\beta}}\right] \tag{3.1}$$

Using (3.1) in (1.3) the modified Murthy estimator is obtained as:

$$t_{MM} = \frac{y_i\left(\frac{1}{N-1}\right) + y_j\left(\frac{1}{N-1}\right)}{\frac{1}{d(N-1)}\left[\frac{p_i^{\alpha}\left(1-p_i^{\beta}\right)}{1-2p_i^{\beta}} + \frac{p_j^{\alpha}\left(1-p_j^{\beta}\right)}{1-2p_j^{\beta}}\right]}$$

$$t_{MM} = \frac{d\left(y_i + y_j\right)\left(1-2p_i^{\beta}\right)\left(1-2p_j^{\beta}\right)}{\left[p_i^{\alpha}\left(1-p_i^{\beta}\right)\left(1-2p_j^{\beta}\right) + p_j^{\alpha}\left(1-p_j^{\beta}\right)\left(1-2p_i^{\beta}\right)\right]} \tag{3.2}$$

The Estimator (3.2) reduces to the estimator of simple random sampling for equal probabilities. The estimator (3.2) is an unbiased estimator of population total and it is slight modification of actual Murthy (1957) estimator. The estimator (3.2) is a class of modified Murthy estimators and different estimators can be constructed for different values of α and β. Also for $\alpha = 1$ and $\beta = -\infty$, the estimator (3.2) transforms to the classical ratio estimator for a sample of size 2.
The variance of (3.2) has been obtained in the following section.

3.1 Designed Based Variance of the Modified Estimator

The variance of the modified Murthy estimator is given as:

$$Var\left(t_{MM}\right) = \frac{1}{2}\sum_{i=1}^{N}\sum_{\substack{j=1\\ i\neq j}}^{N} t_{MM}^{2} P\left(S\right) - Y^{2}$$

Substituting the values of t_{MM} from (3.2) and P(S) from (3.1) and simplifying we get:

$$Var\left(t_{MM}\right) = \frac{1}{2\left(N-1\right)^{2}}\sum_{i=1}^{N}\sum_{\substack{j=1\\ i\neq j}}^{N} \left\{\frac{\left(y_i - y_j\right)^{2}}{P\left(S\right)} - \left(N-1\right)\left\{\left(y_i + y_j\right)^{2} - 2Ny_iy_j\right\}\right\} \quad (3.3)$$

or

$$Var\left(t_{MM}\right) = \frac{1}{2\left(N-1\right)^{2}}\sum_{i=1}^{N}\sum_{\substack{j=1\\ i\neq j}}^{N} \left\{\frac{\left(y_i - y_j\right)^{2}}{P\left(S\right)}\left\{1 - \left(N-1\right)P\left(S\right)\right\} + \left(N-1\right)\left(N-2\right)\ 2y_iy_j\right\} \quad (3.4)$$

The variance expression given in (3.4) transforms to the expression of simple random sampling for equal probabilities.

4. Empirical Study

The empirical study has been carried out for the selected 10 natural numbers being the various values of the constant α and β in the range of -5 to 5 with an increment of 1.0. For this empirical study the variance given in (3.4) has been calculated. After calculating the variance we have assigned ranks to each variance on the basis of its magnitude. The pair of constants α and β which produces smallest variance has been assigned a rank of 1; the second smallest variance has been assigned a rank 2 and so on. This procedure is repeated for all the population and finally we have computed the average rank for each pair of the constants α and β. These average ranks for various pairs have been given in the following Table:

Table: Average Ranks of the Variance of the modified Murthy Estimator for the different values of α and β

β \ α	-5	-4	-3	-2	-1	1	2	3	4	5
-5	96.00	97.00	98.00	99.00	100.00	91.00	92.00	93.00	94.00	95.00
-4	85.90	86.90	87.90	88.90	90.00	80.90	81.90	82.90	83.90	84.90
-3	73.60	74.70	75.70	76.70	77.80	67.60	69.60	70.70	71.90	72.60
-2	56.90	57.90	58.90	59.50	61.00	50.10	52.80	53.80	54.80	55.80
-1	36.90	37.75	38.25	38.75	39.65	28.65	32.35	33.85	34.95	36.10
1	***6.50***	***6.60***	***6.80***	***7.10***	***7.70***	***6.60***	***5.40***	***5.80***	***6.00***	***6.10***
2	14.10	13.10	12.00	10.90	9.70	21.70	18.30	17.20	15.70	15.20
3	28.10	27.10	26.10	25.10	24.00	34.20	31.30	31.10	30.10	29.10
4	42.70	44.10	43.10	42.10	40.00	53.80	49.20	48.10	46.90	49.40
5	62.20	61.20	60.20	59.20	58.10	71.80	67.20	65.50	64.40	63.40

Looking at the Table we can readily see that for all the values of β, the value of $\alpha = 1$ produces the least average rank. We therefore pick $\alpha = 1$ as the suitable choice for this constant. Further, from Table we can straight away decide that constant α can be replaced with 1 and we can develop a sub-class of estimators for various values of β. This sub–class has the form:

$$t_{MM} = \frac{c\left(y_i + y_j\right)\left(1-2p_i^{\beta}\right)\left(1-2p_j^{\beta}\right)}{\left[p_i\left(1-p_i^{\beta}\right)\left(1-2p_j^{\beta}\right) + p_j\left(1-p_j^{\beta}\right)\left(1-2p_i^{\beta}\right)\right]} \tag{4.1}$$

where $c = \sum_{i=1}^{N} p_i \left(1-p_i^{\beta}\right) \Big/ \left(1-2\,p_i^{\beta}\right)$

This sub-class can be used by using various values of the constant β in (4.1) for estimation of population total.

References

1. Brewer, K. R. W. and Hanif, M. (1983). Sampling with unequal probabilities. *Lecture notes in Statistics Series, No.* ***15****, Springer Verlag, New York.*
2. Hansen, M. H. and Hurwitz, W. N. (1943). On the theory of sampling from a finite population. *Annals of Mathematical Statistics*, **14**, 333 – 362.
3. Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of American Statistical Association*, **47**, 663 – 685.
4. Midzuno, H. (1952). On the sampling system with probability proportionate to sum of size. *Annals of Institutional Mathematical Statistics*, **3**, 99 – 108.
5. Murthy, M. N. (1957). Ordered and unordered estimators in sampling without replacement. *Sankhya*, **18**, 379 – 390.
6. Sen, A. R. (1953). On the estimate of the variance in sampling with varying probabilities. *Journal of the Indian Society of Agricultural* Statistics, **5**, 119 – 127.
7. Yates, F. and Grundy, P. M. (1953). Selection without replacement from within strata with probability proportional to size. *Journal of Royal Statistical Society, Series B*, **15**, 153 – 161.