Journal of Reliability and Statistical Studies (ISSN: 0974-8024)
Vol. 2, Issue 2(2009): 59-66

IMPROVED ESTIMATION UNDER MIDZUNO – LAHIRI – SEN TYPE SAMPLING SCHEME

Shashi Bhushan[1], R. Karan Singh[2] and Shubhra Katara[3]

1. Department of Statistics, PUC Campus, Mizoram University, Aizawl – 796001, Mizoram, India.
E Mail: shashi.py@gmail.com
2. Department of Statistics, University of Lucknow, Lucknow, U P, India.
3. Department of Statistics, Bareilly College, Bareilly, U P, India.
E Mail: shubhraonthenet@gmail.com.

Abstract

The generalized ratio type estimator for estimating the mean or total of finite population proposed by Walsh (1970) is reconsidered following Midzuno-Lahiri-Sen type sampling scheme. The unbiasedness of the generalized ratio estimator under the proposed Midzuno-Lahiri-Sen type sampling scheme is established and the expression of mean square error of the generalized ratio estimator under the proposed Midzuno-Lahiri-Sen type sampling scheme is derived. Further the optimum value of the parameter involved and the minimum mean square error under this optimum value of the parameter are also given. A comparative study of the proposed sampling strategy as compared to mean per unit estimator, ratio estimator, product estimator, linear regression estimator and generalized ratio estimator (Walsh) is made. The concluding remarks show that using suitable range information we can get estimators which under the proposed Midzuno-Lahiri-Sen type sampling scheme are better than the usual mean per unit estimator, ratio estimator, product estimator, linear regression estimator and generalized ratio estimator (Walsh) in the sense of unbiasedness and smaller mean square error. Finally an empirical study is included for illustration.

Key Words: Ratio type estimator, simple random sampling, Midzuno-Lahiri-Sen type sampling scheme, unbiasedness, mean square error, range prior information.

1. Introduction

In sampling from a finite population, the use of information on an auxiliary variable for increasing the efficiency of sampling strategy is quite well established. But, in most cases, the use of such auxiliary information results in biased estimation of population parameters. One such generalized class of estimators was proposed by Walsh (1970) for estimating population mean, which, despite all its advantages, has one serious drawback that it was biased whenever the optimum value of the characterizing scalar was not attained. In his paper, Walsh has proposed to use the estimated value of the characterizing scalar, when the optimizing value is unknown, which will result in a biased estimator. A sampling design was proposed separately by three authors, namely, Midzuno (1952), Lahiri (1951) and Sen (1952), under which the traditional ratio estimator is unbiased. In this paper, an attempt has been made to improvise Walsh estimator by using a Midzuno – Lahiri – Sen type sampling design. Recently, some attempts have been made by various authors, including Senapati et al (2006), Singh et al (2005) among others, to improve the existing sampling strategies by using auxiliary information.

Let y be the main (study) variable with population mean $\bar{Y}$ and population variance σ_y^2 and x be the auxiliary variable with population mean $\bar{X}$ and population variance σ_x^2. Thus, for a finite population of size N with population values Y_i and X_i for the ith unit $(i=1,2,...,N)$ of the population on y and x respectively, we have

$$\bar{Y}=\frac{1}{N}\sum_{i=1}^{N}Y_i \qquad \bar{X}=\frac{1}{N}\sum_{i=1}^{N}X_i$$

$$\sigma_y^2=\frac{1}{N}\sum_{i=1}^{N}(Y_i-\bar{Y})^2=\frac{N-1}{N}S_y^2 \qquad \sigma_x^2=\frac{1}{N}\sum_{i=1}^{N}(X_i-\bar{X})^2=\frac{N-1}{N}S_x^2$$

Further, let $\sigma_{xy}=\frac{1}{N}\sum_{i=1}^{N}(X_i-\bar{X})(Y_i-\bar{Y})=\frac{N-1}{N}S_{xy}$ be the population covariance between x and y, $C_x=S_x/\bar{X}$ and $C_y=S_y/\bar{Y}$ be the population coefficient of variation of x and y respectively, and $\rho=\sigma_{xy}/\sigma_x\sigma_y=S_{xy}/S_xS_y$ be the population correlation coefficient between y and x. Further let $\bar{y}_s$ and $\bar{x}_s$ are the sample means of x and y respectively for a sample s.

When the random sample s is selected by simple random sampling without replacement, the generalized ratio estimator by Walsh (1970) for estimating population mean $\bar{Y}$ is

$$\bar{y}_A=\frac{\bar{y}_s\bar{X}}{\bar{X}+A(\bar{x}_s-\bar{X})} \tag{1.1}$$

where A is the characterizing scalar to be chosen suitably.

We now consider the generalized ratio estimator $\bar{y}_A$ under Midzuno (1952)-Lahiri (1951)-Sen (1952) type sampling scheme and denote it by $\bar{y}_{AM}$ which has the following two objectives of

(i) making $\bar{y}_A$ to be unbiased for all values of characterizing scalar A, and

(ii) finding better estimators in some sense in the class $\bar{y}_{AM}$ utilizing some stable range prior information in practice.

The proposed Midzuno-Lahiri-Sen type sampling scheme for selecting a sample s of size n deals with

(i) selecting first unit with probability proportional to $\bar{X}+A(x_i-\bar{X})$, where x_i is the size of the first selected unit,

$$P(i)=\frac{\bar{X}+A(x_i-\bar{X})}{X} \tag{1.2}$$

and

(ii) selecting the remaining $n-1$ units in the sample from $N-1$ units in the population by simple random sampling without replacement.

Thus,

$P(s)=\sum_{i=1}^{n}$ {Probability of selecting ith sample unit at first draw} X {probability of selecting $n-1$ units out of $N-1$ units by simple random sampling without replacement}

$$= \sum_{i=1}^{n} \frac{\bar{X} + A(x_i - \bar{X})}{X} \frac{1}{^{N-1}C_{n-1}} \qquad \text{(where } X = N\bar{X} \text{)}$$

$$= \frac{\bar{X} + A(\bar{x}_s - \bar{X})}{\bar{X}.\,^{N}C_n} \qquad (1.3)$$

2. Unbiasedness of the proposed sampling strategy

Consider the expectation of the proposed estimator under the proposed sampling design

$$E(\bar{y}_{AM}) = E\{\frac{\bar{y}_s \bar{X}}{\bar{X} + A(\bar{x}_s - \bar{X})}\}$$

$$= \sum_{s=1}^{^NC_n} \frac{\bar{y}_s \bar{X}}{\bar{X} + A(\bar{x}_s - \bar{X})} P(s)$$

$$= \sum_{s=1}^{^NC_n} \frac{\bar{y}_s}{^NC_n}$$

$$= E(\bar{y}_s) \qquad \text{(under SRSWOR)}$$

$$= \bar{Y} \qquad (2.1)$$

showing that $\bar{y}_{AM}$ is an unbiased estimator of population mean $\bar{Y}$ for all values of A under the proposed Midzuno-Sen type sampling scheme.

3. Mean Square Error of the proposed sampling strategy

Since $\bar{y}_{AM}$ is unbiased estimator of population mean hence we have

$$MSE(\bar{y}_{AM}) = V(\bar{y}_{AM}) = E(\bar{y}_{AM} - \bar{Y})^2$$

$$= E\{\frac{\bar{y}_s \bar{X}}{\bar{X} + A(\bar{x}_s - \bar{X})} - \bar{Y}\}^2$$

$$= \sum_{s=1}^{^NC_n} [\frac{\bar{y}_s \bar{X} - \bar{Y}\left\{\bar{X} + A(\bar{x}_s - \bar{X})\right\}}{\bar{X} + A(\bar{x}_s - \bar{X})}]^2 P(s)$$

$$= \sum_{s=1}^{^NC_n} \frac{\left[\bar{y}_s \bar{X} - \bar{Y}\left\{\bar{X} + A(\bar{x}_s - \bar{X})\right\}\right]^2}{\left\{\bar{X} + A(\bar{x}_s - \bar{X})\right\}\bar{X}} \frac{1}{^NC_n}$$

$$= \sum_{s=1}^{^NC_n} \frac{\left[\left(\bar{Y} + e_0\right)\bar{X} - \bar{Y}\left(\bar{X} + Ae_1\right)\right]^2}{\left(\bar{X} + Ae_1\right)\bar{X}} \frac{1}{^NC_n} \qquad \text{(where } \bar{y}_s = \bar{Y} + e_0 \text{ and } \bar{x}_s = \bar{X} + e_1 \text{)}$$

$$= \sum_{s=1}^{^NC_n} \frac{\left(\bar{X}e_0 - A\bar{Y}e_1\right)^2}{\bar{X}^2} \left(1 + \frac{A}{\bar{X}} e_1\right)^{-1} \frac{1}{^NC_n}$$

$$= \sum_{s=1}^{^NC_n} \left[\left\{e_0^2 - 2A\left(\frac{\bar{Y}}{\bar{X}}\right)e_0 e_1 + A^2\left(\frac{\bar{Y}}{\bar{X}}\right)^2 e_1^2\right\}\left(1 - \frac{A}{\bar{X}} e_1 + ...\right)\right] \frac{1}{^NC_n}$$

$$= E\left\{ e_0^{\,2} - 2A\left(\frac{\bar{Y}}{\bar{X}}\right) e_0 e_1 + A^2\left(\frac{\bar{Y}}{\bar{X}}\right)^2 e_1^{\,2} - \frac{A}{\bar{X}} e_1 e_0^{\,2} + \ldots \right\} \qquad \text{(under SRSWOR)}$$

$$= E\left(e_0^{\,2}\right) - 2A\left(\frac{\bar{Y}}{\bar{X}}\right) E\left(e_0 e_1\right) + A^2\left(\frac{\bar{Y}}{\bar{X}}\right)^2 E\left(e_1^{\,2}\right) - \frac{A}{\bar{X}} E\left(e_1 e_0^{\,2}\right) + \ldots$$

Let the sample size be so large that $|e_i|, i = 0,1$ becomes so small that terms of e_i's having powers greater than two may be neglected. Also we know that

$$E(e_0^{\,2}) = (\frac{1}{n} - \frac{1}{N}) S_y^{\,2} \qquad E(e_1^{\,2}) = (\frac{1}{n} - \frac{1}{N}) S_x^{\,2} \qquad E(e_0 e_1) = (\frac{1}{n} - \frac{1}{N}) S_{xy}$$

Therefore, we have

$$MSE(\bar{y}_{AM}) = (\frac{N-n}{N.n})\{S_y^{\,2} - 2A\left(\frac{\bar{Y}}{\bar{X}}\right) S_{xy} + A^2\left(\frac{\bar{Y}}{\bar{X}}\right)^2 S_x^{\,2}\}$$

$$= (\frac{N-n}{N.n})\{S_y^{\,2} - 2ARS_{xy} + A^2 R^2 S_x^{\,2}\} \qquad (\text{where } R = \frac{\bar{Y}}{\bar{X}}) \qquad (3.1)$$

$$= \bar{Y}^2 (\frac{N-n}{N.n})\{C_y^{\,2} - 2A\rho C_x C_y + A^2 C_x^{\,2}\} \qquad (3.2)$$

Further, the expression (3.1) or (3.2) is minimum when

$$AR = \rho \frac{S_y}{S_x} \text{ or } A = \rho \frac{C_y}{C_x} = C \text{ (say)} \qquad (3.3)$$

and the minimum mean square error of $\bar{y}_{AM}$ is given by

$$MSE(\bar{y}_{AM})_{\min} = (\frac{N-n}{N.n})(1-\rho^2) S_y^{\,2} = \bar{Y}^2 (\frac{N-n}{N.n})(1-\rho^2) C_y^{\,2} \qquad (3.4)$$

4. Comparision of the proposed sampling strategy with mean per unit estimator under SRSWOR

As we know that

$$MSE(\bar{y}) = \bar{Y}^2 (\frac{1}{n} - \frac{1}{N}) C_y^{\,2}$$

so that $MSE(\bar{y}_{AM}) < MSE(\bar{y})$ i.e. if, $-2A\rho C_x C_y < -A^2 C_x^{\,2}$ (4.1)

or

(i) if $A > 0$ and $R > 0$, the efficiency condition (4.1) reduces to

$$A < 2\rho \frac{C_y}{C_x} = 2C \quad \text{that is} \quad C > \frac{A}{2} \qquad (4.2)$$

(ii) if $A < 0$ and $R > 0$, the efficiency condition (4.1) reduces to

$$A > 2\rho \frac{C_y}{C_x} = 2C \quad \text{that is} \quad C < \frac{A}{2} \qquad (4.3)$$

5. Comparision of the proposed sampling strategy with ratio estimator under SRSWOR

Further we know that mean square error of ratio estimator $\bar{y}_R$ to be

$$MSE(\bar{y}_R) = \bar{Y}^2(\frac{N-n}{N.n})\{C_y^2 + C_x^2 - 2\rho C_x C_y\}$$

Thus, $MSE(\bar{y}_{AM}) < MSE(\bar{y}_{RM})$

if and only if $(A^2-1)C_x^2 - 2(A-1)\rho C_y C_x < 0$ or $(A-1)(A+1-2C) < 0$ (5.1)

(i) When A is chosen such that $A > 1$, the efficiency condition (5.1) reduces to

$$C > \frac{A+1}{2} \quad (5.2)$$

(ii) When A is chosen such that $A < 1$, the efficiency condition (5.1) reduces to

$$C < \frac{A+1}{2} \quad (5.3)$$

6. Comparision of the proposed sampling strategy with product estimator under SRSWOR

We know that the mean square error of the product estimator $\bar{y}_P$ is

$$MSE(\bar{y}_P) = \bar{Y}^2(\frac{N-n}{N.n})\{C_y^2 + C_x^2 + 2\rho C_x C_y\}$$

Therefore, $MSE(\bar{y}_{AM}) < MSE(\bar{y}_P)$

Iff $(A^2-1)C_x^2 - 2(A+1)\rho C_y C_x < 0$ or $(A+1)(A-1-2C) < 0$ (6.1)

(i) When A is chosen such that $A > -1$, the efficiency condition (6.1) is reduced to

$$C > \frac{A-1}{2} \quad (6.2)$$

(ii) When A is chosen such that $A < -1$, the efficiency condition (6.1) is reduced to

$$C < \frac{A-1}{2} \quad (6.3)$$

7. Comparision of the proposed sampling strategy with linear regression estimator under SRSWOR

The mean square error of linear regression estimator $\bar{y}_{lr}$ is

$$MSE(\bar{y}_{lr}) = \bar{Y}^2(\frac{N-n}{N.n})(1-\rho^2)C_y^2$$

Hence $MSE(\bar{y}_{AM})_{\min} = MSE(\bar{y}_{lr})$ to the first degree of approximation. Also $\bar{y}_{AM}$ is unbiased for all values of A under the proposed Midzuno-Sen type sampling scheme while the linear regression estimator $\bar{y}_{lr}$ is biased estimator under simple random sampling without replacement to the first degree of approximation.

8. Comparision of the proposed sampling strategy with generalized ratio estimator (Walsh) under SRSWOR

The mean square error of generalized ratio estimator (Walsh) $\bar{y}_A$ under simple random sampling without replacement is

$$MSE(\bar{y}_A)_{SRS} = \bar{Y}^2(\frac{N-n}{N.n})\{C_y^2 + A^2C_x^2 - 2A\rho C_x C_y\}$$

Although $MSE(\bar{y}_A)_{SRS} = MSE(\bar{y}_{AM})$ for all values of the characterizing scalar A but the generalized ratio estimator (Walsh) $\bar{y}_A$ under simple random sampling without replacement is almost unbiased at only one point $A = C$ whereas the proposed sampling strategy is unbiased for all values of A.

9. Concluding remarks

Some authors propose to use the minimizing $MSE(\bar{y}_{AM})$ value $A = C$ for A but the exact value of C may not always be known. However, since C is a very stable quantity over time hence range information about the stable value of C may be easily known in practice, see Murthy (1967). Therefore, using this information about C, we find efficient estimators in the sense of having lesser mean square error as follows:

As we know that mean per unit estimator $\bar{y}$ is preferred to ratio and product estimators when $0 < C < 1/2$ and $-1/2 < C < 0$ respectively. In such situations the valuable auxiliary information remains unutilized. From efficiency condition (4.2) and (4.3) we can get class of estimators, which are better than the mean per unit estimator even in situations when $0 < C < 1/2$ or $-1/2 < C < 0$. Let the range information about C be known as $C < C_0$ where $0 < C_0 < 1/2$, then we choose A satisfying the efficiency condition (4.2) such that $A/2 = C_0$ or $A = 2C_0$ to get a class of estimators

$$\bar{y}_{2C_0} = \frac{\bar{y}_s \bar{X}}{\bar{X} + 2C_0(\bar{x}_s - \bar{X})} \tag{9.1}$$

which are better than the mean per unit estimator $\bar{y}$ in the sense of having smaller mean square error. Further if it is known that $C > C_1$ where $-1/2 < C_1 < 0$ then we choose A satisfying the efficiency condition (4.3) such that $A = 2C_1 (< 0)$ so that the class of estimators

$$\bar{y}_{2C_1} = \frac{\bar{y}_s \bar{X}}{\bar{X} + 2C_1(\bar{x}_s - \bar{X})} \tag{9.2}$$

are better than $\bar{y}$ in the sense of having lesser mean square error. More specifically, for prior range information $C < 1/3$ choosing $A = 2/3$ satisfying the efficiency (dominance) condition (4.2), we get more efficient unbiased estimator

$$\bar{y}_{2/3} = \frac{\bar{y}_s \bar{X}}{\bar{X} + 2/3(\bar{x}_s - \bar{X})}$$

than $\bar{y}$ in the sense of having smaller mean square error.

Let the range information about C be known as $C > C_0 (> 1)$, then from the efficiency condition (5.2) we choose $(A+1)/2 = C_0$ or $A = 2C_0 - 1$ to get the class of estimators

$$\bar{y}_{2C_0-1} = \frac{\bar{y}_s \bar{X}}{\bar{X} + (2C_0 - 1)(\bar{x}_s - \bar{X})} \tag{9.3}$$

which are better than the ratio estimator in the sense of having lesser mean square error. Further if it is known that $C < C_1 (< 1)$ then we choose A such that $A = 2C_1 - 1$ satisfying the efficiency condition (5.3) so that the class of estimators

$$\bar{y}_{2C_1-1} = \frac{\bar{y}_s \bar{X}}{\bar{X} + (2C_1 - 1)(\bar{x}_s - \bar{X})} \tag{9.4}$$

are better than the ratio estimator in the sense of having lesser mean square error. For example, if it is known that $C > 3/2$ we may choose $A = 2$ satisfying the efficiency (dominance) condition (5.2) to obtain a more efficient unbiased estimator

$$\bar{y}_2 = \frac{\bar{y}_s \bar{X}}{\bar{X} + 2(\bar{x}_s - \bar{X})}$$

than ratio estimator $\bar{y}_R$ in the sense of having lesser mean square error.

If the range information about C be known as $C > C_0^{'}(>-1)$ then we choose A such that $(A-1)/2 = C_0^{'}$ or $A = 2C_0^{'} + 1$ satisfying the efficiency condition (6.2) to get the class of estimators

$$\bar{y}_{2C_0^{'}+1} = \frac{\bar{y}_s \bar{X}}{\bar{X} + (2C_0^{'} + 1)(\bar{x}_s - \bar{X})} \tag{9.5}$$

which are better than the product estimator in the sense of having lesser mean square error. Further, if it is known that $C < C_1^{'}(< -1)$ then we choose $A = 2C_1^{'} + 1$ satisfying the efficiency condition (6.3) so that the class of estimators

$$\bar{y}_{2C_1^{'}+1} = \frac{\bar{y}_s \bar{X}}{\bar{X} + (2C_1^{'} + 1)(\bar{x}_s - \bar{X})} \tag{9.6}$$

is more efficient than the product estimator $\bar{y}_P$ in the sense of having smaller mean square error.

Moreover, in the light of already proved results in sections (7) and (8), it can be mentioned that $\bar{y}_{AM}$ can be preferred to both $\bar{y}_A$ and $\bar{y}_{lr}$ in sense of unbiasedness. Also, $\bar{y}_{AM}$ being unbiased and more efficient than the mean per unit estimator $\bar{y}$, ratio estimator $\bar{y}_R$ and product estimator $\bar{y}_P$ can be a better alternative in various practical situations.

10. An Empirical Study

Let us consider the following example considered by Srivastava (1969): In order to estimate the mean yield of fibre ($\bar{Y}$) per plant in jute fibre crops, the auxiliary characteristic height (x) was taken. For the population consisting of fifty jute plants (capsulanes) shown at Jute Agricultural Research Farm, Barreckpore in the year 1962-63, the following values for the population are obtained $\bar{Y}$ = 5.59 gms, $\bar{X}$ = 6.65 feets, C_y^2 = 0.05680, C_x^2 = 0.00846, ρ = 0.7418.

In the above example $C = \rho\, C_y / C_x$ = 1.92>1, hence the appropriate values of A are chosen suitably so that the resulting estimators $\bar{y}_{AM}$ are better than the ratio estimator, product estimator and mean per unit estimators. The mean square error of the resulting estimators for A =1.2, A =1.5, A =1.92 and A =2 are respectively,

$$MSE(\bar{y}_{1.2M}) = \frac{(1-f)}{n} 0.9713 \tag{10.1}$$

$$MSE(\bar{y}_{1.5M}) = \frac{(1-f)}{n} 0.8742 \tag{10.2}$$

$$MSE(\bar{y}_{1.92M}) = \frac{(1-f)}{n} 0.8256 \quad (10.3)$$

$$MSE(\bar{y}_{2M}) = \frac{(1-f)}{n} 0.8288 \quad (10.4)$$

The mean square errors of the ratio estimator, product estimator and mean per unit estimator are respectively,

$$MSE(\bar{y}_R) = \frac{(1-f)}{n} 1.0684 \quad (10.5)$$

$$MSE(\bar{y}_P) = \frac{(1-f)}{n} 3.1653 \quad (10.6)$$

$$MSE(\bar{y}) = \frac{(1-f)}{n} 1.8390 \quad (10.7)$$

From (10.1) to (10.7) it is seen that all specified estimators are better than the linear regression estimator, ratio estimator, product estimator and mean per unit estimators in the sense of unbiasedness and smaller mean square error.

References

1. Lahiri D. B. (1951). A method of sample selection providing unbiased ratio estimates, Bull. Int. Stat. Inst., 3, p. 133-140.
2. Midzuno H. (1952). On the sampling system with probability proportional to the sum of the sizes, Ann. Inst. Stat. Math., 3, p. 99-107.
3. Murthy M. N. (1967). Sampling theory and method, Statistical Publishing Society, Calcutta, India.
4. Sen A. R. (1952). Present status of probability sampling and its use in the estimation of a characteristic, Econometrika, p. 20-103.
5. Walsh J. E. (1970). Generalization of ratio estimator for population total, Sankhya, A, 32, p. 99-106.