RESEARCH ARTICLE OPEN ACCESS

A Generalized Class of Ratio Type Estimator for Population Variance under Midzuno-Sen Type Sampling Scheme

Shashi Bhushan[a], Chandni Kumari[b] and Ratan Kumar Thakur[b]
[a]*Department of Mathematics & Statistics, Dr. Shakuntala Mishra National Rehabilitaion University, Lucknow, 226017, India.*
[b]*Department of Statistics, Bahasaheb Bhimrao Ambedkar University,(A Central University), Lucknow, 226025, India.*

ABSTRACT
In this paper, we present the family of estimators for estimating the population variance under Midzuno-Lahiri-Sen (MLS) type sampling design P(s). The properties of the proposed sampling strategy are derived up to the first order of approximation. Further, the comparison of proposed sampling strategy with respect to the some important estimators made. Finally, numerical illustration is also given in support of the present study.

Keywords: Simple random sampling, Ratio type estimator, Midzuno-Lahiri-Sen type sampling design.

Date of Submission: 22-04-2020 Date of Acceptance: 06-05-2020

I. INTRODUCTION

In many surveys, information on an auxiliary variable that is highly correlated with the variable under study is readily available which can be used for improving the sampling strategy. Isaki (1983), Das and Tripathi (1978) proposed a generalized class of biased estimators for estimating population variance. Also, a sampling design was proposed by Midzuno (1952), Lahiri (1951) and independently by Sen (1952), under which the traditional ratio estimator is unbiased. An attempt has been made by the authors in the direction of Bhushan (2012, 2013, 2016), Bhushan et al. (2009), Bhushan and Pandey (2010), Bhushan and Katara (2010), Kumari et al. (2018), to improve the existing sampling strategies by using auxiliary information which modifies both the estimator as well as the sampling scheme. These works were restricted to unbiased estimation of population mean. It is also noteworthy that very few works have been done in this direction.

Let the population consists of N units. Yi and Xi denote the i[th] characteristics of the population. The population mean of the study variable is denoted by $\bar{Y}$ and population mean of the auxiliary variable which is known is denoted by $\bar{X}$. The population variance of the study variable and the auxiliary variable is denoted by S_y^2 and S_x^2. Let $\mu_{pq} = N^{-1}\sum_{i=1}^{N}\left(X_i - \bar{X}\right)^p \left(Y_i - \bar{Y}\right)^q$ be the population product moment between x and y . C_x and C_y be the coefficient of variation of auxiliary and study variable respectively. Thus, ρ be the correlation coefficient between the variable under study and auxiliary variable which measures the degree of linear relationship between two variables and it is given as $\rho = Cov\left(Y, X\right)/\sigma_Y\sigma_X$. Now, let us take a random sample of size n drawn without replacement. y_i be the i[th] characteristics of the study variable of the sample and x_i be the i[th] characteristics of the auxiliary variable of the sample. The sample mean of study variable for estimating the population mean is denoted by $\bar{y}\left(= n^{-1}\sum_{i=1}^{n} y_i\right)$. The sample mean of auxiliary variable is denoted by $\bar{x}\left(= n^{-1}\sum_{i=1}^{n} x_i\right)$. When the random sample s is selected by simple random sampling without replacement.

II. PROPOSED SAMPLING STRATEGY

Taking clue from Kumari et al. (2018) the generalized ratio estimator for estimating population variance S_y^2 is modified to a more general class of estimators using prior information is proposed to be

$$\hat{S}^2_{A\Upsilon} = \frac{\Upsilon\, s_y^2\, (\alpha\,\bar{X} + \beta)}{\alpha\,\bar{X} + \beta + A\alpha(\,\bar{x} - \bar{X})}$$

where A is the characterizing scalar to be chosen suitably; α and β represents the prior information in the form of the parameters based on auxiliary characters discussed later. We now consider this generalized ratio estimator under the proposed modified Midzuno-Lahiri-Sen type sampling design. The proposed MLS type sampling design for selecting a sample s of size n deals with selecting the first unit of the sample by PPS type sampling scheme described below and selecting the remaining $(n-1)$ units in a sample from $(N-1)$ units in the population by simple random sampling without replacement.

Therefore,

$\mathrm{P}(s) \; = \sum_{i=1}^{n} P$ {selecting i^th^ sample unit at first draw} X

P {selecting (n-1) units out of (N-1) units by SRSWOR}

$$=\sum_{i=1}^{n}\frac{\left(\alpha\bar{X}+\beta\right)+A\alpha\left(x_i-\bar{X}\right)}{{}^{N-1}C_{n-1}\left(\alpha\bar{X}+\beta\right)}$$

$$=\frac{\left(\alpha\bar{X}+\beta\right)+A\alpha\left(\bar{x}-\bar{X}\right)}{{}^{N}C_{n}\left(\alpha\bar{X}+\beta\right)}$$

III. MEAN SQUARE ERROR OF $\left(S^2_{A\gamma}\right)$

$$MSE\left(s^2_{A\gamma}\right)=E\left[s^2_{A\gamma}-S^2_{A\gamma}\right]^2$$

$$=E\left[\frac{\gamma s^2_{A\gamma}\left(\alpha\bar{X}+\beta\right)}{\left(\alpha\bar{X}+\beta\right)+A\alpha\left(\bar{x}-\bar{X}\right)}-S^2_{A\gamma}\right]^2$$

$$=E\left[\frac{\gamma s^2_{A\gamma}\left(\alpha\bar{X}+\beta\right)-S^2_{A\gamma}\left(\alpha\bar{X}+\beta\right)-S^2_{A\gamma}A\alpha\left(\bar{x}-\bar{X}\right)}{\left(\alpha\bar{X}+\beta\right)+A\alpha\left(\bar{x}-\bar{X}\right)}\right]^2$$

$$=\sum_{S=1}^{{}^{N}C_{n}}\left[\frac{\gamma s^2_{A\gamma}\left(\alpha\bar{X}+\beta\right)-S^2_{A\gamma}\left(\alpha\bar{X}+\beta\right)-S^2_{A\gamma}A\alpha\left(\bar{x}-\bar{X}\right)}{\left(\alpha\bar{X}+\beta\right)+A\alpha\left(\bar{x}-\bar{X}\right)}\right]^2 P(s)$$

$$=\sum_{S=1}^{{}^{N}C_{n}}\left[\frac{\gamma s^2_{A\gamma}\left(\alpha\bar{X}+\beta\right)-S^2_{A\gamma}\left(\alpha\bar{X}+\beta\right)-S^2_{A\gamma}A\alpha\left(\bar{x}-\bar{X}\right)}{\left(\alpha\bar{X}+\beta\right)+A\alpha\left(\bar{x}-\bar{X}\right)}\right]^2\frac{\left\{\left(\alpha\bar{X}+\beta\right)+A\alpha\left(\bar{x}-\bar{X}\right)\right\}}{{}^{N}C_{n}\left(\alpha\bar{X}+\beta\right)}$$

$$=\sum_{S=1}^{{}^{N}C_{n}}\left[\frac{\left\{\gamma s^2_{A\gamma}\left(\alpha\bar{X}+\beta\right)-S^2_{A\gamma}\left(\alpha\bar{X}+\beta\right)-S^2_{A\gamma}A\alpha\left(\bar{x}-\bar{X}\right)\right\}^2}{\left(\alpha\bar{X}+\beta\right)+A\alpha\left(\bar{x}-\bar{X}\right)}\right]\frac{1}{{}^{N}C_{n}\left(\alpha\bar{X}+\beta\right)}$$

$$=\sum_{S=1}^{{}^{N}C_{n}}\left[\frac{\left\{\gamma s^2_{A\gamma}\left(\alpha\bar{X}+\beta\right)-S^2_{A\gamma}\left(\alpha\bar{X}+\beta\right)-S^2_{A\gamma}A\alpha\left(\bar{x}-\bar{X}\right)\right\}^2}{\left\{1+\frac{A\alpha\left(\bar{x}-\bar{X}\right)}{\left(\alpha\bar{X}+\beta\right)}\right\}}\right]\frac{1}{{}^{N}C_{n}\left(\alpha\bar{X}+\beta\right)^2}$$

$$=\sum_{S=1}^{{}^{N}C_{n}}\left[\left\{\gamma s^2_{A\gamma}\left(\alpha\bar{X}+\beta\right)-S^2_{A\gamma}\left(\alpha\bar{X}+\beta\right)-S^2_{A\gamma}A\alpha\left(\bar{x}-\bar{X}\right)\right\}^2\left\{1+\frac{A\alpha\left(\bar{x}-\bar{X}\right)}{\left(\alpha\bar{X}+\beta\right)}\right\}^{-1}\right]\frac{1}{{}^{N}C_{n}\left(\alpha\bar{X}+\beta\right)^2}$$

$$=E\left[\left\{\gamma s_{A\gamma}^2\left(\alpha\bar{X}+\beta\right)-S_{A\gamma}^2\left(\alpha\bar{X}+\beta\right)-S_{A\gamma}^2A\alpha\left(\bar{x}-\bar{X}\right)\right\}^2\left\{1+\frac{A\alpha\left(\bar{x}-\bar{X}\right)}{\left(\alpha\bar{X}+\beta\right)}\right\}^{-1}\right]\frac{1}{\left(\alpha\bar{X}+\beta\right)^2}$$

Let $\varepsilon_o=\left(s_{A\gamma}^2-S_{A\gamma}^2\right)$ and $\varepsilon_1=\left(\bar{x}-\bar{X}\right)$

and V(ε_o)=E(ε_0^2)= $\frac{\mu_{04}-\mu_{02}^2}{n}$

V(ε_1)= E(ε_1^2)= $\frac{\mu_{20}}{n}$ and E($\varepsilon_0\varepsilon_1$)= $\frac{\mu_{12}}{n}$

Thus (1) becomes,

$$=E\left[\left\{\gamma\left(S_{A\gamma}^2+\varepsilon_o\right)\left(\alpha\bar{X}+\beta\right)-S_{A\gamma}^2\left(\alpha\bar{X}+\beta\right)-S_{A\gamma}^2A\alpha\varepsilon_1\right\}^2\left\{1+\frac{A\alpha\varepsilon_1}{\left(\alpha\bar{X}+\beta\right)}\right\}^{-1}\right]\frac{1}{\left(\alpha\bar{X}+\beta\right)^2}$$

$$=\frac{1}{\left(\alpha\bar{X}+\beta\right)^2}E\left[\left\{(\gamma-1)S_{A\gamma}^2\left(\alpha\bar{X}+\beta\right)+\gamma\varepsilon_0\left(\alpha\bar{X}+\beta\right)-S_{A\gamma}^2A\alpha\varepsilon_1\right\}^2\left\{1-\frac{A\alpha\varepsilon_1}{\left(\alpha\bar{X}+\beta\right)}+\frac{A^2\alpha^2\varepsilon_1^2}{\left(\alpha\bar{X}+\beta\right)^2}-...\right\}\right]$$

$$=\frac{1}{\left(\alpha\bar{X}+\beta\right)^2}E\left[\left\{(\gamma-1)^2\left(S_{A\gamma}^2\right)^2\left(\alpha\bar{X}+\beta\right)^2+\gamma^2\varepsilon_0{}^2\left(\alpha\bar{X}+\beta\right)^2+\left(S_{A\gamma}^2\right)^2A^2\alpha^2\varepsilon_1{}^2\right.\right.$$

$$+2(\gamma-1)\gamma\left(S_{A\gamma}^2\right)^2\left(\alpha\bar{X}+\beta\right)^2\varepsilon_0-2(\gamma-1)\left(S_{A\gamma}^2\right)^2\left(\alpha\bar{X}+\beta\right)A\alpha\varepsilon_1-2\gamma S_{A\gamma}^2\left(\alpha\bar{X}+\beta\right)A\alpha\varepsilon_1\varepsilon_0\Big\}$$

$$\left\{1-\frac{A\alpha\varepsilon_1}{\left(\alpha\bar{X}+\beta\right)}+\left\{\frac{A\alpha\varepsilon_1}{\left(\alpha\bar{X}+\beta\right)}\right\}^2-...\right\}$$ (ignoring higher power terms)

$$=\frac{1}{\left(\alpha\bar{X}+\beta\right)^2}E\left[\left\{(\gamma-1)^2\left(S_{A\gamma}^2\right)^2\left(\alpha\bar{X}+\beta\right)^2+\left(S_{A\gamma}^2\right)^2(\gamma-1)^2A^2\alpha^2\varepsilon_1{}^2+\gamma^2\left(\alpha\bar{X}+\beta\right)^2\varepsilon_0{}^2+\left(S_{A\gamma}^2\right)^2A^2\alpha^2\varepsilon_1{}^2\right.\right.$$

$$\left.\left.-2\gamma(\gamma-1)S_{A\gamma}^2\left(\alpha\bar{X}+\beta\right)A\alpha\varepsilon_0\varepsilon_1+2\left(S_{A\gamma}^2\right)^2(\gamma-1)A^2\alpha^2\varepsilon_1{}^2-2\gamma S_{A\gamma}^2\left(\alpha\bar{X}+\beta\right)A\alpha\varepsilon_0\varepsilon_1\right\}\right]$$

$$=\frac{1}{\left(\alpha\bar{X}+\beta\right)^2}\left[\left\{(\gamma-1)^2\left(S_{A\gamma}^2\right)^2\left(\alpha\bar{X}+\beta\right)^2+\left(S_{A\gamma}^2\right)^2(\gamma-1)^2A^2\alpha^2E\left(\varepsilon_1{}^2\right)+\gamma^2\left(\alpha\bar{X}+\beta\right)^2E\left(\varepsilon_0{}^2\right)\right.\right.$$

$$+\left(S_{A\gamma}^2\right)^2A^2\alpha^2E\left(\varepsilon_1{}^2\right)-2\gamma(\gamma-1)S_{A\gamma}^2\left(\alpha\bar{X}+\beta\right)A\alpha E\left(\varepsilon_0\varepsilon_1\right)+2\left(S_{A\gamma}^2\right)^2(\gamma-1)A^2\alpha^2E\left(\varepsilon_1{}^2\right)$$

$$\left.\left.-2\gamma\left(S_{A\gamma}^2\right)\left(\alpha\bar{X}+\beta\right)A\alpha E\left(\varepsilon_0\varepsilon_1\right)\right\}\right]$$

$$=\frac{1}{\left(\alpha\bar{X}+\beta\right)^2}\left[\left\{(\gamma-1)^2\left(S_{A\gamma}^2\right)^2\left(\alpha\bar{X}+\beta\right)^2+(\gamma-1)^2\left(S_{A\gamma}^2\right)^2A^2\alpha^2\left(\frac{\mu_{20}}{n}\right)+\gamma^2\left(\alpha\bar{X}+\beta\right)^2\left(\frac{\mu_{04}-\mu_{02}^2}{n}\right)\right.\right.$$

$$+\left(S_y^2\right)^2A^2\alpha^2\left(\frac{\mu_{20}}{n}\right)-2\gamma(\gamma-1)S_y^2\left(\alpha\bar{X}+\beta\right)A\alpha\left(\frac{\mu_{12}}{n}\right)+2\left(S_y^2\right)^2(\gamma-1)A^2\alpha^2\left(\frac{\mu_{20}}{n}\right)$$

$$\left.\left.-2\gamma S_y^2\left(\alpha\bar{X}+\beta\right)A\alpha\left(\frac{\mu_{12}}{n}\right)\right\}\right]$$

$$=\frac{1}{\left(\alpha\bar{X}+\beta\right)^2}\left[(\gamma-1)^2\left(S_y^2\right)^2\left(\alpha\bar{X}+\beta\right)^2+\left(S_y^2\right)^2\gamma^2A^2\alpha^2\left(\frac{\mu_{20}}{n}\right)+\gamma^2\left(\alpha\bar{X}+\beta\right)^2\left(\frac{\mu_{04}-\mu_{02}^2}{n}\right)\right.$$

$$\left.-2\gamma^2S_y^2A\alpha\left(\alpha\bar{X}+\beta\right)\left(\frac{\mu_{12}}{n}\right)\right]$$

$$=\frac{1}{\left(\alpha\bar{X}+\beta\right)^2}\left[(\gamma-1)^2\left(S_y^2\right)^2\left(\alpha\bar{X}+\beta\right)^2+\gamma^2\left\{\left(S_y^2\right)^2A^2\alpha^2\left(\frac{\mu_{20}}{n}\right)+\left(\alpha\bar{X}+\beta\right)^2\left(\frac{\mu_{04}-\mu_{02}^2}{n}\right)\right.\right.$$

$$\left.\left.-2S_y^2A\alpha\left(\alpha\bar{X}+\beta\right)\left(\frac{\mu_{12}}{n}\right)\right\}\right]$$

$$=\left[(\gamma-1)^2\left(S_y^2\right)^2+\gamma^2\left(\frac{\mu_{02}^2}{n}\right)\left\{(\beta_{02}-1)+\frac{A^2\alpha^2\mu_{20}}{\left(\alpha\bar{X}+\beta\right)^2}-\frac{2A\alpha}{\left(\alpha\bar{X}+\beta\right)}\left(\frac{\mu_{12}}{\mu_{02}}\right)\right\}\right]$$

$$=\left[(\gamma-1)^2\mu_{02}^2+\gamma^2\left(\frac{\mu_{02}^2}{n}\right)\left\{(\beta_{02}-1)+\frac{A^2\alpha^2\mu_{20}}{\left(\alpha\bar{X}+\beta\right)^2}-\frac{2A\alpha}{\left(\alpha\bar{X}+\beta\right)}\left(\frac{\mu_{12}}{\mu_{02}}\right)\right\}\right]$$

$$=\mu_{02}^2\left[(\gamma-1)^2+\frac{\gamma^2}{n}\left\{(\beta_{02}-1)+\frac{A^2\alpha^2\mu_{20}}{\left(\alpha\bar{X}+\beta\right)^2}-\frac{2A\alpha}{\left(\alpha\bar{X}+\beta\right)}\left(\frac{\mu_{12}}{\mu_{02}}\right)\right\}\right] \tag{2.1}$$

which is the mean square error of the estimator under the proposed sampling strategy.

Differentiating (2.1) with respect to A and γ and equating to zero, we get the optimum values of these constant by solving the two normal equations.

$$\gamma_{opt}=\frac{1}{\left[1+\frac{1}{n}\left\{(\beta_{02}-1)+\frac{A^2\alpha^2\mu_{20}}{\left(\alpha\bar{X}+\beta\right)^2}-\frac{2A\alpha}{\left(\alpha\bar{X}+\beta\right)}\left(\frac{\mu_{12}}{\mu_{02}}\right)\right\}\right]} \tag{2.2}$$

$$A_{opt}=\frac{\mu_{12}\left(\alpha\bar{X}+\beta\right)}{\alpha\mu_{20}\mu_{02}} \tag{2.3}$$

Substituting (2.3) in (2.2), we get

$$\gamma_{opt}=\frac{1}{\left[1+\frac{1}{n}\left\{(\beta_{02}-1)-\left(\frac{\mu_{12}^2}{\mu_{02}^2\mu_{20}}\right)\right\}\right]} \tag{2.6}$$

Now minimum mean square error under the proposed sampling strategy is obtained by putting (2.6) and (2.3) in (2.1) we get

$$MSE\left(s_{A\gamma}^{2}\right)_{opt}=\frac{\frac{\mu_{02}^{2}}{n}\left\{\left(\beta_{02}-1\right)-\left(\frac{\mu_{12}^{2}}{\mu_{02}^{2}\mu_{20}}\right)\right\}}{\left[1+\frac{1}{n}\left\{\left(\beta_{02}-1\right)-\left(\frac{\mu_{12}^{2}}{\mu_{02}^{2}\mu_{20}}\right)\right\}\right]}$$

IV. EFFICIENCY COMPARISON

3.1 Comparison with mean per unit estimator

$$MSE\left(s_{A\gamma}^{2}\right)<MSE\left(s_{y}^{2}\right)_{wor} \quad (3.1.1)$$

3.2 Comparison with ratio estimator

$$MSE\left(s_{A\gamma}^{2}\right)<MSE\left(s_{y}^{2}\right)_{ratio} \quad (3.2.1)$$

3.3 Comparison with product estimator

$$MSE\left(s_{A\gamma}^{2}\right)<MSE\left(s_{y}^{2}\right)_{pro} \quad (3.3.1)$$

3.4 Comparison with linear regression estimator

$$MSE\left(s_{A\gamma}^{2}\right)_{opt}<MSE\left(s_{y}^{2}\right) \quad (3.4.1)$$

3.5 Comparison with generalized ratio estimator

$$MSE\left(s_{A\gamma}^{2}\right)_{opt}<MSE\left(s_{y}^{2}\right)_{opt} \quad (3.5.1)$$

3.6 Comparison with generalized ratio estimator

$$MSE\left(s_{A\gamma}^{2}\right)_{opt}<MSE\left(s_{y}^{2}\right)_{opt} \quad (3.6.1)$$

Numerical study

Table 1: Parameters of the three different population

POPULATION 1		**POPULATION 2**		**POPULATION 3**	
N	20	N	20	N	8
n	8	n	8	n	3
U20	1.79	U20	1.70	U20	93.6
U02	173.89	U02	144.69	U02	131.86
U04	102435.1	U04	-96.95	U04	38610.93
U12	50.1066	U12	61094.87	U12	426.84

On the basis of above information, we calculate the mean squared error of various estimator which is given in Tbale 2.

Table 2: MSE of Various Estimator

Estimator	**Pop 1**	**Pop 2**	**Pop 3**
$\hat{S}_y^2$	9024.67	5020.05	7074.68
$\hat{S}_r^2$	8855.97	5926.81	6439.62
$\hat{S}_p^2$	9692.96	4488.07	8663.12
$\hat{S}_A^2$	8849.39	4329.66	6426.46
$\hat{S}_{AY}^2$	**6845.87**	**3587.66**	**4692.17**

V. CONCLUSION

For all the population, the minimum mean square error of the proposed sampling strategy is less than the mean square error of the mean per unit estimator, ratio estimator, product estimator and the linear regression estimator. Thus the proposed sampling strategy is the best among mean per unit, ratio, product and regression estimator.

REFERENCES

[1]. Bhushan, S. and Kumari, C. (2018). An Unbiased Family of Sampling Strategies for Estimation of Population Variance, International Journal of Statistics and Systems, 13 (1), 41-52
[2]. Bhushan, S. (2013). Improved sampling strategies in finite population. Scholars Press, Germany.
[3]. Bhushan S. (2012). Some efficient sampling strategies based on ratio type estimator, Electronic Journal of Applied Statistical Analysis, 5(1), 74-88.
[4]. Bhushan S., Masaldan R. N. and Gupta P. K. (2011). Improved sampling strategies based on modified ratio estimator , International Journal of Agricultural and Statistical Sciences, 7(1), 63-75.
[5]. Bhushan S. and Katara, S. (2010). On the class of unbiased sampling strategies , Journal of Reliability and Statistical Studies, 3(2), 93-101.
[6]. Bhushan S., Singh, R. K. and Katara, S. (2009). Improved estimation under Midzuno-Lahiri-Sen-type sampling scheme, Journal of Reliability and Statistical Studies, 2(2), 59-66.
[7]. Lahiri D. B. (1951). A method of sample selection providing unbiased ratio estimates, Bull. Int. Stat. Inst., 3, p. 133-140.
[8]. Midzuno H. (1952). On the sampling system with probability proportional to the sum of the sizes, Ann. Ints. Stat. Math., 3, p. 99-107.
[9]. Murthy M. N. (1967). Sampling theory and method , Statistical Publishing Society, Calcutta, India.
[10]. Sen A. R. (1952). Present status of probability sampling and its use in the estimation of characteristics, Econometrika, p. 20-103
[11]. Walsh J. E. (1970). Generalization of ratio estimator for population total, Sankhya, A, 32, p. 99-106.
[12]. Das, A. K., and Tripathi, T. P (1978). Use of auxiliary information in estimating the finite population variance. Sankhya, 40(C), 139-148.