

GENERALIZED EXPONENTIAL ESTIMATOR FOR ESTIMATING THE POPULATION MEAN USING AUXILIARY VARIABLE

¹Poonam Singh, ²Carlos Bouza and ^{†1}Rajesh Singh

¹Department of Statistics, Banaras Hindu University (U.P), India

²University of Havana, Havana

[†] Corresponding Author

Email: ppiyu123@gmail.com, bouza@matcom.uh.cu, rsinghstat@gmail.com

Abstract:

In this paper we have proposed a generalized exponential estimator for estimating population mean of a variable using an auxiliary variable under the simple random sampling. Mean square error and bias of the proposed estimator have been derived up to the first order of approximation. We have compared the efficiency of the proposed estimator with existing estimators and also calculated the optimal conditions under which our proposed estimator performs better than the existing estimators. We have also derived condition under which our estimator reduces to regression estimator. Results of this study are analyzed theoretically as well as empirically. Improvement over ratio, exponential and regression estimator has been shown using four real data sets.

Keywords: Auxiliary Variable, Efficiency, Mean Square Error, Exponential estimator, Regression estimator.

1. Introduction

The use of auxiliary information in estimation of population parameters increases the efficiency of the estimators. Cochran (1940) was the first to propose well known ratio estimator for estimating the yield of cereal experiments using auxiliary information. Several authors including Singh et al. (2005), Khosnevisan et al. (2007), Singh et al. (2009), Singh and Kumar (2011), Sharma et al. (2013), Sharma et al. (2013), Singh and Malik (2014), Singh et al. (2018) and Singh et al. (2019) proposed improved estimators using auxiliary information. Here, we have considered a finite population $\psi = \{\psi_1, \psi_2, \dots, \psi_N\}$ of size N. Let us consider Y be the study variable having mean \bar{Y} and X be the auxiliary variable having mean \bar{X} . A sample of size n is drawn from population $\psi = \{\psi_1, \psi_2, \dots, \psi_N\}$ using SRSWOR (simple random sampling without replacement) and obtained \bar{x} and \bar{y} as the sample means of the auxiliary variable and study variable respectively.

Some of the existing estimators (which we have considered for the comparison) with their Bias and MSE (mean square error) up to the first order of approximation are:

- I. MSE of the Usual unbiased estimator \bar{y} for estimating mean

$$\text{MSE}(\bar{y}_{\text{usual}}) = \bar{Y}^2 \vartheta C_y^2 \quad (1.1)$$

- II. Cochran (1940) ratio estimator

$$\bar{y}_{\text{ratio}} = \frac{\bar{y}}{\bar{X}} \bar{X} \quad (1.2)$$

$$\text{Bias}(\bar{y}_{\text{ratio}}) = \bar{Y} \vartheta (C_x^2 - \rho C_y C_x) \quad (1.3)$$

$$\text{MSE}(\bar{y}_{\text{ratio}}) = \bar{Y}^2 \vartheta (C_y^2 + C_x^2 - 2\rho C_y C_x) \quad (1.4)$$

- III. Bahal and Tuteja (1991) exponential ratio type estimator

$$\bar{y}_{\text{exp r}} = \bar{y} \exp\left(\frac{\bar{X} - \bar{x}}{\bar{X} + \bar{x}}\right) \quad (1.5)$$

$$\text{Bias}(\bar{y}_{\text{exp r}}) = \bar{Y} \vartheta \left(\frac{3}{8} C_x^2 - \frac{1}{2} \rho C_y C_x \right) \quad (1.6)$$

$$\text{MSE}(\bar{y}_{\text{exp r}}) = \bar{Y}^2 \vartheta \left(C_y^2 + \frac{1}{4} C_x^2 - \rho C_y C_x \right) \quad (1.7)$$

- IV. Regression estimator

$$\bar{y}_{\text{reg}} = \bar{y} + \beta(\bar{X} - \bar{x}) \quad (1.8)$$

$$\text{MSE}(\bar{y}_{\text{reg}}) = \bar{Y}^2 \vartheta (1 - \rho^2) C_y^2 \quad (1.9)$$

Where $\vartheta = \frac{N-n}{Nn}$, $S_y^2 = \frac{1}{N-1} \sum_{i=1}^N (Y_i - \bar{Y})^2$, $C_y = \frac{S_y}{\bar{Y}}$, $C_x = \frac{S_x}{\bar{X}}$

C_y : is coefficient of variation related to auxiliary variable X.

C_x : is coefficient of variation related to auxiliary variable Y.

ρ : is the population correlation coefficient between study variable Y and auxiliary variable X.

This article proposes new estimator and derive its bias and MSE up to the first order of approximation. And compared the efficiency of the proposed estimator with the existing ones. Result is theoretically as well as empirically validated.

2. PROPOSED ESTIMATOR

We have proposed the following generalized exponential estimator for estimating population mean:

$$\bar{y}_{ra} = \bar{y} A^{\left[\frac{\bar{X}-\bar{x}}{\bar{X}+\bar{x}}\right]} \quad \text{for } A > 0 \tag{2.1}$$

Where A be any constant.

Exponential estimator $\bar{y}_{exp r}$ is particular case of proposed estimator \bar{y}_{ra} for $A=e=2.71\dots$

For calculating Bias and Mean square error of the estimator \bar{y}_{ra} we need to find the exact expression for $E(\bar{y} A^{\left[\frac{\bar{X}-\bar{x}}{\bar{X}+\bar{x}}\right]})$, which is difficult to find.

So, we approximate them using following values:

$$\begin{aligned} \epsilon_0 &= \frac{\bar{x} - \bar{X}}{\bar{X}} \Rightarrow \bar{x} = \bar{X}(1 + \epsilon_0) \\ \epsilon_1 &= \frac{\bar{y} - \bar{Y}}{\bar{Y}} \Rightarrow \bar{y} = \bar{Y}(1 + \epsilon_1) \end{aligned}$$

Since we are estimating population mean under SRSWOR, so we have

$$\begin{aligned} E(\epsilon_0) &= 0 \\ E(\epsilon_1) &= 0 \\ E(\epsilon_0^2) &= E\left(\frac{\bar{y} - \bar{Y}}{\bar{Y}}\right)^2 = \vartheta C_y^2 \\ E(\epsilon_1^2) &= E\left(\frac{\bar{x} - \bar{X}}{\bar{X}}\right)^2 = \vartheta C_x^2 \\ E(\epsilon_0 \epsilon_1) &= E\left[\left(\frac{\bar{x} - \bar{X}}{\bar{X}}\right)\left(\frac{\bar{y} - \bar{Y}}{\bar{Y}}\right)\right] = \vartheta \rho C_y C_x \end{aligned}$$

Writing proposed estimator \bar{y}_{ra} in terms of ϵ 's, we get

$$\bar{y}_{ra} = \bar{Y}(1 + \epsilon_0) A^{\left[\frac{\bar{X}-\bar{X}(1+\epsilon_1)}{\bar{X}+\bar{X}(1+\epsilon_1)}\right]} \tag{2.2}$$

$$= \bar{Y}(1 + \epsilon_0) A^{\left[\frac{-\epsilon_1}{2} \left(1 + \frac{\epsilon_1}{2}\right)^{-1}\right]} \tag{2.3}$$

The term $\left(1 + \frac{\epsilon_1}{2}\right)^{-1}$ can be expanded and it would be convergent only when $-1 < \epsilon_1 < 1$, which means that $0 < \bar{x} < 2\bar{X}$. This holds true if the sample size is fairly large i.e. variation in \bar{x} is small. With these assumptions, we have

$$\bar{y}_{ra} = \bar{Y} \left(1 + \epsilon_0 - \frac{\epsilon_1}{2} \log_e A + \frac{\epsilon_1^2}{4} \log_e A + \frac{\epsilon_0 \epsilon_1}{2} \log_e A + \frac{\epsilon_1^2}{8} (\log_e A)^2 \right) \quad (2.4)$$

Under the assumption of $-1 < \epsilon_1 < 1$, the terms of ϵ_0 and ϵ_1 involving powers more than two are negligibly small. So, we ignored the ϵ 's higher order terms.

Subtracting \bar{Y} from both sides of the equation (2.4) we get the following equation:

$$(\bar{y}_{ra} - \bar{Y}) = \bar{Y} \left(\epsilon_0 - \frac{\epsilon_1}{2} \log_e A + \frac{\epsilon_1^2}{4} \log_e A + \frac{\epsilon_0 \epsilon_1}{2} \log_e A + \frac{\epsilon_1^2}{8} (\log_e A)^2 \right) \quad (2.5)$$

Now taking expectation on both sides of the equation (2.5), we get Bias of the estimator \bar{y}_{ra} as follows:

$$Bias(\bar{y}_{ra}) = \bar{Y} \mathcal{G} \left[\frac{C_x^2}{4} \log_e A + \frac{C_x^2}{8} (\log_e A)^2 + \frac{\rho C_y C_x}{2} \log_e A \right] \quad (2.6)$$

Squaring equation (2.5) and taking expectation on both sides, we get mean square of the estimator \bar{y}_{ra} up to the first order of approximation as follows:

$$MSE(\bar{y}_{ra}) = \bar{Y}^2 \mathcal{G} \left[C_y^2 + \frac{C_x^2}{4} (\log_e A)^2 - \rho C_y C_x (\log_e A) \right] \quad (2.7)$$

3. OPTIMALITY CONDITION FOR THE PROPOSED ESTIMATOR \bar{y}_{ra} .

Differentiating equation (2.7) with respect to A , we get the following equation:

$$\frac{\partial(MSE(\bar{y}_{ra}))}{\partial A} = \frac{C_x^2}{2} \frac{\log_e A}{A} - \frac{\rho C_y C_x}{A} \quad (3.1)$$

Equating equation (3.1) to zero, we get

$$\log_e A = \rho \frac{2C_y}{C_x} = 2\zeta \quad (3.2)$$

$$\Rightarrow A = \exp(2\zeta) \quad (3.3)$$

Where $\zeta = \rho \frac{C_y}{C_x}$

Substituting value of $\log_e A$ in equation (2.7), we get Min MSE (\bar{y}_{ra}):

$$\text{MinMSE}(\bar{y}_{ra}) = \bar{Y}^2 \vartheta (1 - \rho^2) C_y^2 \tag{3.4}$$

Note: MinMSE(\bar{y}_{ra}) is same as Variance of regression estimator i.e. estimator \bar{y}_{ra} reduces to regression estimator for the optimal value of $A = \exp(2\zeta)$.

4. EFFICIENCY COMPARISON WITH THE EXISTING ESTIMATORS

(i) Proposed estimator \bar{y}_{ra} is better than the usual mean estimator if

$$\text{MinMSE}(\bar{y}_{ra}) < \text{MSE}(\bar{y}_{usual})$$

$$\Rightarrow (1 - \rho^2) < 1$$

$$\Rightarrow \rho^2 > 0$$

(ii) Proposed estimator \bar{y}_{ra} is better than the ratio estimator \bar{y}_{ratio} if

$$\text{MinMSE}(\bar{y}_{ra}) < \text{MSE}(\bar{y}_{ratio})$$

$$\Rightarrow (C_x - \rho C_y)^2 > 0$$

$$\Rightarrow C > 0 \text{ or } C < 0$$

(iii) Proposed estimator \bar{y}_{ra} is better than the exponential ratio estimator $\bar{y}_{exp r}$ if

$$\text{MinMSE}(\bar{y}_{ra}) < \text{MSE}(\bar{y}_{exp r})$$

$$\Rightarrow \left(\frac{C_x}{2} - \rho C_y \right)^2 > 0$$

$$\Rightarrow C > \frac{1}{2} \text{ or } C < \frac{1}{2}$$

(iv) Proposed estimator \bar{y}_{ra} is same as the regression estimator \bar{y}_{reg} if we put $A = \exp(2\zeta)$ in MSE expression of \bar{y}_{ra} .

5. EMPIRICAL STUDY

For the empirical study we have considered following four data sets:

Table 5.1: Data sets for the empirical study.

	Population I [Kadilar and Cingi (2006)]	Population II [Kadilar and Cingi (2006, 2006)]	Population III [Kadilar and Cingi (2006)]	Population IV [Koyuncu and Kadilar (2009)]
N	200	106	104	923
n	50	20	20	180
\bar{Y}	500	2212.59	625.37	436.4345
\bar{X}	25	27421.70	13.93	11440.5
C_y	15	5.22	1.866	1.7183
C_x	2	2.10	1.653	1.8645
ρ	0.90	0.86	0.865	0.9543

Using above four data sets we have compared the efficiency of the existing and the proposed estimator by calculating MSE (mean square error) and PRE (percentage relative efficiency).

PRE of the estimators is calculated using following formula.

$$PRE(\bar{y}_i) = \frac{MSE(\bar{y}_{usual})}{MSE(\bar{y}_i)} \times 100 \quad (5.1)$$

where $\bar{y}_i = \bar{y}_{ratio}, \bar{y}_{expr}, \bar{y}_{ra}, \bar{y}_{reg}$

Table 5.2 : Table representing MSE and PRE of existing estimators and the proposed estimator.

Estim-ators	Population I		Population II		Population III		Population IV	
	MSE	PRE	MSE	PRE	MSE	PRE	MSE	PRE
\bar{y}_{usual}	843750	100	5411348	100	54993.75	100	2515.07	100
\bar{y}_{ratio}	656250	128.57	2542740	212.82	13869.96	396.49	267.64	939.71
\bar{y}_{expr}	746250	113.06	3758095	143.99	23642.99	232.60	651.04	386.31
\bar{y}_{reg}	160312.5	526.31	1409115	384.02	13846.05	397.18	224.62	1119.68
\bar{y}_{ra}	160312.5	526.31	1409115	384.02	13846.05	397.18	224.62	1119.68

From the Table 5.2, it is clear that on the basis of PRE and MSE of the estimators our proposed estimator \bar{y}_{ra} performs much better than \bar{y}_{usual} , \bar{y}_{ratio} and \bar{y}_{exp} . So exponential estimator should not be restricted to $e=2.71$, as it performs better as we increase the value of A. Min MSE (\bar{y}_{ra}) is achieved at the point $A = \exp(2\zeta)$, and after this point as we increase the value of A MSE (\bar{y}_{ra}) starts increasing. Value of constant A should be $2.71 < A < \exp(2\zeta)$ to achieve more efficient estimator than the exponential estimator. That's why we have shown the point where Min MSE (\bar{y}_{ra}) is achieved.

6. CONCLUSION

In this article we have proposed a generalized exponential estimator \bar{y}_{ra} using auxiliary variable under SRSWOR. We have theoretically as well as empirically proven that our proposed estimator shows the highest efficiency among the other considered estimators. We have also calculated the value of the constant A where MSE of the proposed estimator is minimum i.e. $A = \exp(2\zeta)$. Therefore it is advisable to use our proposed estimator over exponential estimator (in the range $2.71 < A < \exp(2\zeta)$).

7. REFERENCES

- Bahl, S., & Tuteja, R. K. (1991). Ratio and product type exponential estimators. *Journal of information and optimization sciences*, 12(1), 159-164.
- Cochran, W. G. (1940). The estimation of the yields of cereal experiments by sampling for the ratio of grain to total produce. *The journal of agricultural science*, 30(2), 262-275.
- Kadilar, C., & Cingi, H. (2004). Ratio estimators in simple random sampling. *Applied Mathematics and Computation*, 151(3), 893-902.
- Kadilar, C., & Cingi, H. (2006). An improvement in estimating the population mean by using the correlation coefficient. *Hacettepe Journal of Mathematics and Statistics*, 35(1).
- Kadilar, C., & Cingi, H. (2006). New ratio estimators using correlation coefficient. *Interstat*, 4(March), 1-11.
- Khoshnevisan, M., Singh, R., Chauhan, P., Sawan, N. and Smarandache, F. (2007): A general family of estimators for estimating population means using known value of some population parameter(s). *Far east journal of statistics*, 22(2), 181-191.
- Koyuncu, N., & Kadilar, C. (2009). Efficient estimators for the population mean. *Hacettepe Journal of Mathematics and Statistics*, 38 (2).

- Sharma, P., Singh, R. and Kim, J. M. (2013) : Study of some improved ratio type estimators using information on auxiliary attributes under second order approximation. *Jour. Sci. Res.*, 57, 138-146.
- Singh, H.P., Singh, R., Espejo, M. R., Pineda, M .D. and Nadrajah, S.(2005): On the efficiency of a dual to ratio-cum-product estimator in sample surveys. *Mathematical proceedings of the Royal Irish Academy*, 105 A(2), 51-56.
- Singh, R., Chauhan, P., Sawan, N. and Smarandache, F.(2009): Improvement in estimating the population mean using exponential estimator in simple random sampling. *Intern. Jour. Of Stat. and Econ. (BSE)*, 3, 13-18.
- Singh, R. and Kumar, M. (2011): A note on transformations on auxiliary variable in survey sampling. *Mod. Assis. Stat. Appl.*, 6:1, 17-19.
- Singh, R. and Malik, S. (2014): Improved estimator of population variance using information on auxiliary attribute in simple random sampling. *Appl. Math. Comp.*, 235, 43-49.
- Singh, R., Mishra, P. and Bouza, C.N. (2018) : Estimation of population mean using information on auxiliary attribute: A review. *Ranked Set Sampling*. (Eds- C.N. Bouza and Al Omari), Academic Press
- Singh, R., Vishwakarma, G. K. and Mishra, P. (2019): A note on the estimators for coefficient of dispersion using auxiliary information. *Comm. in Stat.-Simul. and Compt.* <https://doi.org/10.1080/03610918.2018.1517889>