ON ALMOST UNBIASED FAMILY OF RATIO AND PRODUCT TYPE ESTIMATORS

Shashi Bhushan[1]

Abstract: A family of biased estimators was proposed by Khoshnevisan et al. (2007), using some known parameters of the auxiliary variable, for estimating the population mean of study variable in a finite population. This family contained various known ratio and product type biased estimators. A family of Jack-Knife estimators, based on Khoshnevisan et al. (2007), is proposed for estimating the population mean. Some new estimators are also included in the proposed study. The proposed sampling strategies are unbiased and there is substantial gain in relative efficiency. A comparative study is made with some important estimators and some concluding remarks are given.

Keywords: Ratio type estimators, product type estimators, Jack-Knife estimator, relative efficiency, sampling strategy.

1. INTRODUCTION

In most of the sample surveys, the main aim of survey statistician is to estimate the population mean (total) of the character under study. The ratio or product type estimators of population mean (total) of the study variable which has high (positive or negative) correlation with the auxiliary variable provide a substantial gain in efficiency. Recently, various authors have proposed various ratio and product type estimators when the prior information in the form of some auxiliary parameter, like correlation coefficient, coefficient of variation, coefficient of skewness, coefficient of kurtosis, standard deviation, quartiles etc., is available. For such type of prior information, a number of alternative estimators have been suggested with their mean square error (like Sidodia and Dwivedi (1981), Pandey and Dubey (1988), Upadhyaya and Singh (1999), Singh (2003), Singh and Taylor (2003), Singh et al (2004), Al-Omari (2009)).

Koshnevesan et al. (2007) proposed a family of such biased estimators which included many biased ratio and product type estimators as special cases. In this paper, we propose an unbiased family of ratio and product type estimators involving Jack-Knife technique by Gray and Schucany (1972). The present study has been done with three objectives in mind:

(i) to improve upon the present classes of biased estimators;
(ii) to provide some new estimators of the population mean;
(iii) to present a better alternative family of unbiased strategy.

These comparisions will help survey practitioners to choose among various alternative estimators, knowing survey population parameters such as correlation coefficient, coefficient of variation, coefficient of skewness, coefficient of kurtosis, standard deviation, quartiles etc.

2. PRELIMINARIES AND SOME ALTERNATIVE BIASED ESTIMATORS

Let x and y be real valued characteristics defined on a finite population $U = (U_1, U_2, \ldots, U_N)$ having N identifiable units, and, let $\bar{Y}$ and $\bar{X}$ be the population means of study variable y and auxiliary variable x respectively. Let $\bar{y}$ and $\bar{x}$ be the sample means of y and x respectively. Also, let S_y^2 and S_x^2 be the population variances, and, s_y^2 and s_x^2 be the sample variances of y and x respectively. Further, let $S_{xy} = \rho S_x S_y$ and ρ be the population covariance and the correlation coefficient between x and y respectively. The ratio estimator for estimating the population mean of the study variable y is given by $\bar{y}_R = (\bar{y}/\bar{x})\bar{X} = \hat{R}\bar{X}$ where $\hat{R} = \bar{y}/\bar{x}$ and $\bar{X}$ is assumed to be known. If some parameters of x denoted by α_x and β_x are known, apart from $\bar{X}$, then using Walsh (1970), Reddy (1973) and Srivastava (1967); Koshnevesan et al (2007) proposed to study a generalized family of ratio type estimators for estimating the population mean $\bar{Y}$ of the study variable given by

$$t = \bar{y}\left[\frac{\alpha_x \bar{X} + \beta_x}{A(\alpha_x \bar{x} + \beta_x) + (1-A)(\alpha_x \bar{X} + \beta_x)}\right]^g \tag{2.1}$$

The following well known ratio and product estimators, arranged chronologically, can be obtained by suitable choice of A, g, α_x and β_x in (2.1):

$t_0 = \bar{y}$	(mean per unit)
$t_{1R} = (\bar{y}/\bar{x})\bar{X}$	(customary ratio)
$t_{1P} = (\bar{y}.\bar{x})/\bar{X}$	(customary product)
$t_{2R} = \bar{y}\dfrac{\bar{X}+C_x}{\bar{x}+C_x}$	(Sisodia and Dwivedi (1981) ratio)
$t_{2P} = \bar{y}\dfrac{\bar{x}+C_x}{\bar{X}+C_x}$	(Pandey and Dubey (1988) product)
$t_{3P} = \bar{y}\dfrac{\beta_{2x}\bar{x}+C_x}{\beta_{2x}\bar{X}+C_x}$	(Upadhyaya and Singh (1999) product)
$t_{4P} = \bar{y}\dfrac{C_x\bar{x}+\beta_{2x}}{C_x\bar{X}+\beta_{2x}}$	(Upadhyaya and Singh (1999) product)
$t_{5P} = \bar{y}\dfrac{\bar{x}+\sigma_x}{\bar{X}+\sigma_x}$	(Singh (2003) product)
$t_{6P} = \bar{y}\dfrac{\beta_{1x}\bar{x}+\sigma_x}{\beta_{1x}\bar{X}+\sigma_x}$	(Singh (2003) product)

ISBN - 978-93-81583-56-2

$t_{7P} = \bar{y}\dfrac{\beta_{2x}\bar{x}+\sigma_x}{\beta_{2x}\bar{X}+\sigma_x}$ (Singh (2003) product)

$t_{8R} = \bar{y}\dfrac{\bar{X}+\rho}{\bar{x}+\rho}$ (Singh and Taylor (2003) ratio)

$t_{8P} = \bar{y}\dfrac{\bar{x}+\rho}{\bar{X}+\rho}$ (Singh and Taylor (2003) product)

$t_{9R} = \bar{y}\dfrac{\bar{X}+\beta_{2x}}{\bar{x}+\beta_{2x}}$ (Singh et al (2004) ratio)

$t_{9P} = \bar{y}\dfrac{\bar{x}+\beta_{2x}}{\bar{X}+\beta_{2x}}$ (Singh et al (2004) product)

$t_{10P} = \bar{y}\dfrac{\bar{x}+q_{1x}}{\bar{X}+q_{1x}}$ (Al-Omari (2009) product)

$t_{11P} = \bar{y}\dfrac{\bar{x}+q_{3x}}{\bar{X}+q_{3x}}$ (Al-Omari (2009) product)

Further, in order to make our study more comprehensive we also include the following ratio estimators:

$$t_{3R} = \bar{y}\frac{\beta_{2x}\bar{X}+C_x}{\beta_{2x}\bar{x}+C_x}$$

$$t_{4R} = \bar{y}\frac{C_x\bar{X}+\beta_{2x}}{C_x\bar{x}+\beta_{2x}}$$

$$t_{5R} = \bar{y}\frac{\bar{X}+\sigma_x}{\bar{x}+\sigma_x}$$

$$t_{6R} = \bar{y}\frac{\beta_{1x}\bar{X}+\sigma_x}{\beta_{1x}\bar{x}+\sigma_x}$$

$$t_{7R} = \bar{y}\frac{\beta_{2x}\bar{X}+\sigma_x}{\beta_{2x}\bar{x}+\sigma_x}$$

$$t_{10R} = \bar{y}\frac{\bar{X}+q_{1x}}{\bar{x}+q_{1x}}$$

$$t_{11R} = \bar{y}\frac{\bar{X}+q_{3x}}{\bar{x}+q_{3x}}$$

Khoshnevesan et al (2007) considered estimator t under Simple random sampling without replacement. Let $\bar{y} = \bar{Y} + e_0$ and $\bar{x} = \bar{X} + e_1$ such that $E(e_0) = E(e_1) = 0$. Putting these values in the estimator and simplifying, we get

$$t-\bar{Y}=e_0-gA\eta\bar{Y}\frac{e_1}{\bar{X}}+\frac{g(g+1)}{2}A^2\eta^2\bar{Y}\frac{e_1^2}{\bar{X}^2}-gA\eta\frac{e_0e_1}{\bar{X}}+\dots \quad (2.2)$$

The bias and mean square error of t under simple random sampling are given by

$$Bias(t)=\gamma_n\bar{Y}\left\{\frac{g(g+1)}{2}A^2\eta^2C_x^2-gA\eta\rho C_xC_y\right\} \quad (2.3)$$

$$MSE(t)=\gamma_n\bar{Y}^2\left\{C_y^2+gA\eta C_x^2\left(gA\eta-2K\right)\right\} \quad (2.4)$$

where C_y and C_x are the population coefficient of variation of y and x respectively, $\eta=\frac{\alpha_x\bar{X}}{\alpha_x\bar{X}+\beta_x}$, $K=\rho\frac{C_y}{C_x}$. Further, the optimum value of A minimizing the MSE(t) of the proposed family is given by

$A=\frac{K}{g\eta}=A_{opt}$ (say) and the minimum mean square error under optimizing value of $A=A_{opt}$ is

$$MSE(t)_{\min}=\gamma_n\bar{Y}^2(1-\rho^2)C_y^2 \quad (2.5)$$

which is same as the mean square error of the linear regression estimator.

We now consider this generalized estimator under the Simple random sampling without replacement sampling scheme with the Jack – Knife technique. The proposed generalized family of Jack – Knife sampling strategies aims at getting some classes of better sampling strategies than the existing ones in the sense of unbiasedness and lesser mean square error.

3. UNBIASEDNESS AND VARIANCE OF THE PROPOSED STRATEGY:

Let us now apply Quenouille's (1956) method of Jack-Knife such that the sample of size $n=2m$ from a population of size N is split up at random into two sub samples of size m each. For further details one may refer to Gray and Schucany (1972). Let us define

$$t_i=\bar{y}_i\left[\frac{\alpha_x\bar{X}+\beta_x}{A\left(\alpha_x\bar{x}_i+\beta_x\right)+(1-A)\left(\alpha_x\bar{X}+\beta_x\right)}\right]^g;\ i=1,2$$

$$t=\bar{y}\left[\frac{\alpha_x\bar{X}+\beta_x}{A\left(\alpha_x\bar{x}+\beta_x\right)+(1-A)\left(\alpha_x\bar{X}+\beta_x\right)}\right]^g \quad (3.1)$$

where $\bar{y}_1$, $\bar{y}_2$ and $\bar{y}$ denote the sample means based on two sub samples of size m and the entire sample of size $n=2m$ for characteristic y ; $\bar{x}_1$, $\bar{x}_2$ and $\bar{x}$ denote the

ISBN - 978-93-81583-56-2

sample means based on two sub samples of size m and the entire sample of size $n = 2m$ for characteristic x.

It can be easily seen that

$$Bias\left(t_i\right) = \gamma_m \bar{Y}\left\{\frac{g\left(g+1\right)}{2}A^2\eta^2C_x^2 - gA\eta\rho C_x C_y\right\};\ i = 1,2$$

$$Bias\left(t\right) = \bar{Y}\left\{\frac{g\left(g+1\right)}{2}A^2\eta^2C_x^2 - gA\eta\rho C_x C_y\right\} = B_1(say) \quad (3.2)$$

Let us define $t' = \left(t_1 + t_2\right)/2$ as an alternative estimator of the population mean $\bar{Y}$.

The bias of t' is

$$Bias(t') = \gamma_m \bar{Y}\left\{\frac{g\left(g+1\right)}{2}A^2\eta^2C_x^2 - gA\eta\rho C_x C_y\right\} = B_2(say) \quad (3.3)$$

We propose the jackknife estimator t_J for estimating population mean $\bar{Y}$ given by

$$t_J = \frac{t - Rt'}{1-R} = \frac{t - \left\{\frac{N-2m}{2(N-m)}\right\}t'}{1 - \left\{\frac{N-2m}{2(N-m)}\right\}} \text{ where } R = \frac{B_1}{B_2} \qquad (3.4)$$

Taking expectation of (3.4) and using (3.2) and (3.3) we obtain

$E(t_J) = \bar{Y}$ showing that t_J is an unbiased estimator of population mean $\bar{Y}$ to the first order of approximation.

Consider

$$MSE(t_J) = E\left(\frac{t - Rt'}{1-R} - \bar{Y}\right)^2$$

$$= \frac{1}{(1-R)^2}\left\{E(t-\bar{Y})^2 + R^2E(t'-\bar{Y})^2 - 2RE(t-\bar{Y})(t'-\bar{Y})\right\} \qquad (3.5)$$

Also,

$$E(t-\bar{Y})^2 = MSE(t) = \gamma_n \bar{Y}^2\left\{C_y^2 + gA\eta C_x^2\left(gA\eta - 2K\right)\right\} \qquad (3.6)$$

Further,

$$E(t'-\bar{Y})^2 = E\left(\frac{t_1 + t_2}{2} - \bar{Y}\right)^2$$

$$= \frac{1}{4}\left\{E(t_1-\bar{Y})^2 + E(t_2-\bar{Y})^2 + 2E(t_1-\bar{Y})(t_2-\bar{Y})\right\} \qquad (3.7)$$

Since

$$E(t_i-\overline{Y})^2 = MSE(t_i) = \gamma_m \overline{Y}^2 \left\{C_y^2 + gA\eta C_x^2\left(gA\eta - 2K\right)\right\};\, i=1,2 \tag{3.8}$$

Let $\overline{y}_i = \overline{Y} + e_0^{(i)}$ and $\overline{x}_i = \overline{X} + e_1^{(i)}$ such that $E\left(e_0^{(i)}\right) = E\left(e_1^{(i)}\right) = 0$, $i=1,2$

Consider

$$E(t_1-\overline{Y})(t_2-\overline{Y}) = E\left(e_0^{(1)} - gA\eta\overline{Y}\frac{e_1^{(1)}}{\overline{X}}\right)\left(e_0^{(2)} - gA\eta\overline{Y}\frac{e_1^{(2)}}{\overline{X}}\right)$$

$$= E(e_0^{(1)}e_0^{(2)}) - gA\eta\frac{\overline{Y}}{\overline{X}}\left\{E(e_0^{(2)}e_1^{(1)}) + E(e_0^{(1)}e_1^{(2)})\right\} + g^2A^2\eta^2\frac{\overline{Y}^2}{\overline{X}^2}E(e_1^{(1)}e_1^{(2)})$$

Substituting the results in Sukhatme and Sukhatme

$$E(e_0^{(1)}e_0^{(2)}) = -\frac{1}{N}\overline{Y}^2C_y^{\,2}$$

$$E(e_1^{(1)}e_1^{(2)}) = -\frac{1}{N}\overline{X}^2C_x^{\,2}$$

$E(e_0^{(1)}e_1^{(2)}) = E(e_0^{(2)}e_1^{(1)}) = -\frac{1}{N}\overline{X}\overline{Y}\rho C_xC_y$ we have

$$E(t_1-\overline{Y})(t_2-\overline{Y}) = -\frac{1}{N}\overline{Y}^2\left\{C_y^2 + gA\eta C_x^2\left(gA\eta - 2K\right)\right\} \tag{3.9}$$

Putting the values from (3.8) and (3.9) in (3.7) we have

$$E(t'-\overline{Y})^2 = \overline{Y}^2\frac{1}{4}\left[2\gamma_m - \frac{2}{N}\right]\left\{C_y^2 + gA\eta C_x^2\left(gA\eta - 2K\right)\right\}$$

$$= \gamma_n\overline{Y}^2\left\{C_y^2 + gA\eta C_x^2\left(gA\eta - 2K\right)\right\} \tag{3.10}$$

Now consider

$$E(t-\overline{Y})(t'-\overline{Y}) = E\{(t-\overline{Y})(\frac{t_1+t_2}{2} - \overline{Y})\}$$

$$= \frac{1}{2}\{E(t-\overline{Y})(t_1-\overline{Y}) + E(t-\overline{Y})(t_2-\overline{Y})\}$$

Since

$$E(t-\overline{Y})(t_i-\overline{Y}) = E\left(e_0 - gA\eta\overline{Y}\frac{e_1}{\overline{X}}\right)\left(e_0^{(i)} - gA\eta\overline{Y}\frac{e_1^{(i)}}{\overline{X}}\right);\, i=1,2$$

$$= E(e_0e_0^{(i)}) - gA\eta\frac{\overline{Y}}{\overline{X}}\left\{E(e_0^{(i)}e_1) + E(e_0e_1^{(i)})\right\} + g^2A^2\eta^2\frac{\overline{Y}^2}{\overline{X}^2}E(e_1e_1^{(i)})$$

using the following results given in Sukhatme and Sukhatme

$$E(e_0e_0^{(i)}) = \gamma_n\overline{Y}^2C_y^{\,2}$$

$$E(e_1e_1^{(i)}) = \gamma_n\overline{X}^2C_x^{\,2}$$

ISBN - 978-93-81583-56-2

$E(e_0 e_1^{(i)}) = E(e_0^{(i)} e_1) = \gamma_n \bar{X}\bar{Y}\rho C_x C_y$ for $i = 1,2$ we have

$$E(t-\bar{Y})(t_i-\bar{Y}) = \gamma_n \bar{Y}^2 \left\{ C_y{}^2 + \eta C_x{}^2 \left(A^2\eta - 2AK \right) \right\} \tag{3.11}$$

Putting these values from (3.6), (3.10), (3.11) in (3.5) we have

$$MSE(t_J) = \frac{1}{(1-R)^2}\gamma_n \bar{Y}^2 (1+R^2-2R)\left\{ C_y^2 + gA\eta C_x^2 \left(gA\eta - 2K \right) \right\} =$$

$$\gamma_n \bar{Y}^2 \left\{ C_y^2 + gA\eta C_x^2 \left(gA\eta - 2K \right) \right\} \tag{3.12}$$

which is equal to the mean square error of t given by (2.4). The optimizing value of the characterizing scalar A is given by (2.5) and the minimum mean square error under optimizing value of $A = A_{opt}$ is given by (2.6) which is same as the mean square error of the linear regression estimator.

4. COMPARITIVE STUDY

If the minimizing value $A = K/g\eta = A_{opt}$ is known then, we have

$$MSE(t_0) - MSE(t)_{\min} = \gamma_n \bar{Y}^2 \rho^2 C_y{}^2 \geq 0 \tag{4.1}$$

$$MSE(t_{1R}) - MSE(t)_{\min} = \gamma_n \bar{Y}^2 \left(C_x - \rho C_y \right)^2 \geq 0 \tag{4.2}$$

$$MSE(t_{1P}) - MSE(t)_{\min} = \gamma_n \bar{Y}^2 \left(C_x + \rho C_y \right)^2 \geq 0 \tag{4.3}$$

$$MSE(t_{kR}) - MSE(t)_{\min} = \gamma_n \bar{Y}^2 \left(\eta_k C_x - \rho C_y \right)^2 \geq 0 \;\; (\forall k = 2,...,11) \tag{4.4}$$

$$MSE(t_{kP}) - MSE(t)_{\min} = \gamma_n \bar{Y}^2 \left(\eta_k C_x + \rho C_y \right)^2 \geq 0 \;\; (\forall k = 2,...,11) \tag{4.5}$$

where $\eta_k = \dfrac{\alpha_{kx}\bar{X}}{\alpha_{kx}\bar{X} + \beta_{kx}}$, α_{kx}, β_{kx}, are the parameters auxiliary variable x used in the k th estimator t_{kl}, $k = 2,...,11$ and $l = R, P$. Also, if the minimizing value $A = K/g\eta = A_{opt}$ is not known then it can be estimated by the $\hat{A}_{opt} = \dfrac{s_{xy}}{\bar{y}} \dfrac{\bar{X}}{S_x^2 g\eta}$. Using this estimated value, again, leads to same minimum mean square error $MSE(t)_{\min}$. Hence, the proposed sampling strategies are always better than t_0, t_{1R}, t_{1P}, t_{kR}, t_{kP} and $\bar{y}_{lr}$ in sense of unbiasedness and gain in efficiency.

5. CONCLUDING REMARKS

The proposed family of ratio and product type estimators provides a better utilization of prior information of the auxiliary parameter(s) as it is unbiased and leads to substantial gain in efficiency.

6. REFERENCES

1. Al-Omari A. I. F. (2009): “New Product-Type Estimators for the Population Mean Using Quartiles of the Auxiliary Variable”, International Journal of Business, Economics, Finance and Management Sciences 1:4, pg 261-265.
2. Gray H. L. and Schucany W. R. (1972): The Generalized Jack-knife Statistic, Marcel Dekker, New York.
3. Lahiri D. B. (1951): “A method of sample selection providing unbiased ratio estimates”, Bull. Int. Stat. Inst., 3, pg 133-140.
4. Khoshnevisan M., Singh R., Chauhan P., Sawan N. & Smarandache F. (2007): “A general family of estimators for estimating population mean using known value of some parameter(s)”, Far East Journal of Theoritical Statistics, 22, 181-191.
5. Midzuno H. (1952): “On the sampling system with probability proportional to the sum of the sizes”, Ann. Inst. Stat. Math., 3, pg 99-107.
6. Murthy M. N. (1967): Sampling theory and methods, Statistical Publishing Society, Calcutta, INDIA.
7. Pandey B.N. and Dubey V. (1988): “Modified product estimator using coefficient of variation of auxiliary variate”, Assam Statistical Rev., 2(2), 64-66
8. Quenouille M. H. (1956): “Notes on bias in estimation”, Biometrika, 43, pg 353-360.
9. Reddy V.N. (1973): “On ratio and product methods of estimation”, Sankhya, B, 35(3), 307-316.
10. Searls D.T. (1964): “The utilization of known coefficient of variation in the estimation procedure”, Journal of American Statistical Association, 59, 1125-1126.
11. Searls D.T. and Intarapanich P. (1990): “A note on an estimator for the variance that utilizes the kurtosis”, The American Statistician, 44, 4, 295-296.
12. Sen A. R. (1952): “Present status of probability sampling and its use in the estimation of a characteristic”, Econometrika, 20, 103.
13. Sen A.R. (1978): “Estimation of the population mean when the coefficient of variation is known”, Commun. Statist., Theory – Meth. A (7), 657-672.
14. Singh G.N. (2003): “On the improvement of product method of estimation in sample surveys”, Jour. Ind. Soc. Agri. Statistics, 56(3), 267-275.
15. Singh D. and Chaudhary F. S. (1986): Theory and analysis of sampling survey design, Wiley Eastern Limited, New Delhi.
16. Singh H.P. And Tailor R. (2003): “Use of known correlation coefficient in estimating the finite population mean”, Statistics in Transition, 6,4,555-560.
17. Singh H.P., Tailor R. and Kakaran M.S. (2004): “An estimator of Population mean using power transformation”, J.I.S.A.S., 58(2), 223-230.
18. Singh J. Pandey B.N. and Hirano K. (1973): “On the utilization of a known coefficient of kurtosis in the estimation procedure of variance”, Ann. Inst. Stat. Math., 25, 51-55.
19. Sisodia B.V.S. And Dwivedi V.K. (1981): “A modified ratio estimator using coefficient of variation of auxiliary variable”, Journ. Ind. Soc. Agril. Statist., 33, 2, 13-18.
20. Srivastava S.K. (1967): “An estimator using auxiliary information”, Calcutta Statist. Assoc. Bull., 16,121-132.
21. Upadhyaya L. N. and Singh H. P. (1999): “Use of transformed auxiliary variable in estimating the finite population mean”, Biometrical Journal, 41(5), pg 627-636.
22. Sukhatme P. V. and Sukhatme B. V. (1997): Sampling Theory of Surveys with Applications, Piyush Publications, Delhi.
23. Walsh J. E. (1970): “Generalization of ratio estimator for population total”, Sankhya, A, 32, pg 99-106.

****

----

[1]Shashi Bhushan, Department of Applied Statistics, Babasaheb Bhimrao Ambedkar University, Lucknow 226025, INDIA
shashi.py@gmail.com