Asian Journal of Statistical Sciences
Vol. 1, No-2, (2021), pp. 151-158

ARF INDIA
Academic Open Access Publishing
Journal homepage link:
http://arfjournals.com/ajss

A Note on bias reduction using almost unbiased estimators of population mean in finite population sampling

A. K. P. C. Swain[a] and Prabin Kumar Panigrahi[b]

[a] *Utkal University, Bhubaneswar. India* [b]*Indian Institute of Management Indore, M.P, India*

ARTICLE HISTORY
Compiled December 12, 2021

Received 24 March 2021; Accepted 14 October 2021

ABSTRACT
In this paper we present a short review and discussion of some characterizations of almost unbiased ratio type estimators constructed at estimation stage . In such estimators the bias of $O(1/n)$ is removed and the ultimate bias becomes of $O(1/n^2)$,where n is the sample size.

KEYWORDS
Almost Unbiased estimators; Finite population; Ratio type estimators; Simple random sampling

1. Introduction

In large scale sample surveys to estimate the population mean/total of the study variable a researcher invariably collects information on some correlated auxiliary variables along with observations on the study variable. The reason is to construct an improved estimator compared to the mean per unit estimator of the study variable without use of auxiliary information. Ratio method of estimation is the most popular and simplest of all methods used in practice . However, the classical ratio estimator (Cochran,1940) is a biased estimator ,although the bias becomes negligible for large samples. For small sample sizes the bias may be substantial, more so in stratification with ratio method of estimation in each stratum when the accumulated bias of the estimator makes the estimator unacceptable to be of any practical use (Cochran,1977) . This discerning picture of classical ratio estimator has aroused continuing interest among many research workers during last seven decades to construct unbiased or almost unbiased ratio type estimators In this paper we present some characterizations of classes of almost unbiased ratio type estimators constructed at estimation stage to be used in practice. In such estimators the bias of $O(1/n)$ is removed and the ultimate bias becomes of $O(1/n^2)$, where n is the sample size.

CONTACT A. K. P. C. Swain[a]. Email:akpcs@rediffmail.com

2. Preliminaries

Consider a finite population $U = \{U_1, U_2, ..., U_N\}$ of size N distinct and identifiable units indexed by the values of characteristic y(study variable) and x(auxiliary variable) $\{Y_{i,} X_i\}, i = 1, 2, ..., N$.

Select a simple random sample without replacement $u = \{u_1, u_2, ..., u_n\}$ of size n and the observed values of characteristics on the sample units are $\{y_{i,} x_i\}, i = 1, 2, ..., n$.

Let $\bar{Y}$ and$\bar{X}$ be the population means of y and x respectievely. Let S_y^2, S_x^2 and S_{xy} be finite population variance of y , finite population variance x and finite population co-variance between $x$and y respectively.

$C_y = \frac{S_y}{\bar{Y}}$ coefficient of variation of y , $C_x = \frac{S_x}{\bar{X}}$ coefficient of variation of x

$C_{xy} = \frac{S_{xy}}{\bar{Y}\bar{X}}$=$\rho C_x C_y$, ρ being the correlatin coefficient between x and y

$R = \frac{\bar{Y}}{\bar{X}}$

Define $\bar{y}$, $\bar{x}$, s_y^2, s_x^2 and s_{xy} as sample mean of y ,sample mean of x, sample variance of y, sample variance of x and sample co-variance between x and y respectively. Define $c_y = \frac{s_y}{\bar{y}}$, $c_x = \frac{s_x}{\bar{x}}$, $c_{xy} = \frac{s_{xy}}{\bar{x}\bar{y}}$, $r = \frac{\bar{y}}{\bar{x}}$. Put $\theta = (\frac{1}{n} - \frac{1}{N})$

3. Formulation of a class of Almost unbiased Ratio estimators

The ratio estimator of the population ratio $R = \frac{\bar{Y}}{\bar{X}}$, given by $r = \frac{\bar{y}}{\bar{x}}$ is a biased estimator having first order bias, Bias (r)= $R\theta(C_x^2 - C_{xy})$, (Sukhatme,et al., 1984), being of $O(\frac{1}{n})$, which may be negligible for large samples and tends to zero as $n \to \infty$. . For small samples the bias may be substantial so as to make the estimator unacceptable to be used in practice. Cochran (1977) has shown that if the coefficient of variation of x is less than 0.1,then the bias of the simple ratio estimator is small relative to standard error.

Koop (1951), Quenouille (1956), Beale (1962) and Tin (1965) suggested improved ratio type estimators ,whose first order bias is removed and the ultimate bias becomes of $O(\frac{1}{n^2})$, Tin (1965) has showed that to $O(\frac{1}{n^2})$ Beale's estimator is less biased and has equal variance to that of his estimator and in large samples there is marginal loss of efficiency. His Monte Carlo simulations indicate better performances of Beale's estimator compared to Tin's and Quenouille's estimators. Hutchison (1971) compared Quenouille 's , Beale's and Tin's estimators by Monte Carlo simulations and concluded that Beale's and Tin's estimators happened to be more efficient than Quenouille's under certain models. Beale's ratio estimator has been fruitfully applied to hydrological studies with least bias (Richards and Holloway, 1987; Richards, 1987).

Chakrabarty (1979) and Mussa (1999) suggested bias-reducing composite estimators involving simple expansion estimator and almost unbiased ratio estimators such as those of Quenouille's and Tin's estimators. Lui (2020) discussed the use of composite estimator simple mean estimator and ratio estimator with optimal weights to reduce the variance/mean square error of the ratio estimator. The bias and mean square error of composite estimators (with estimated optimal weight and known optimal weight) were examined both by Monte Carlo simulations and with applications to natural populations.

S.Singh and R.Singh (1993) devised a novel technique to arrive at a weighted combination of biased estimators, which is almost unbiased having variance equal to the large sample variance of the linear regression estimator.

3.1. A generalized class of almost unbiased ratio estimators

Swain(2015) suggested a generalized class of almost unbiased ratio estimators for Srivastava's(1971) class of ratio estimators

$$t_g = \bar{y}H(u)$$

under certain regularity conditions, where $u = \bar{x}/\bar{X}$. The suggested class of almost unbiased ratio estimators were

$$t_{gB} = t_g \frac{(1-\theta H_1 \hat{C}_{yx})}{(1+\theta H_2 \hat{C}_x^2)} \quad (Beale\ type)$$

$$t_{gT} = t_g \left[1 - \theta \left(H_2 \hat{C}_x^2 + H_1 \hat{C}_{yx}\right)\right] (Tin\ type)$$
$$t_{gK} = \frac{t_g}{\left(1 + \theta \left(H_2 \hat{C}_x^2 + H_1 \hat{C}_{yx}\right)\right)} (Koop\ type)$$

where $\hat{C}_x^2 = \frac{s_x^2}{\bar{x}^2}$ and $\hat{C}_{yx} = \frac{s_{yx}}{\bar{y}\bar{x}}$; where $H_1 = \left(\frac{\partial H}{\partial u}\right)_{u=1}$ and $H_2 = \frac{1}{2}\left(\frac{\partial^2 H}{\partial u^2}\right)_{u=1}$ are known constants.
The biases of t_{gB}, t_{gT} and t_{gK} are of $O(\frac{1}{n^2})$ first order biases being removed. To $O(\frac{1}{n^2})$, they have the same asymptotic variance.
Thus ,we have the following theorem:

Theorem 1
The class of estimators $\bar{y}H(u)$, where $u = \bar{x}/\bar{X}$ generates a class of almost unbiased estimators of the population mean $\bar{Y}$ of the study variable y, x being the auxiliary variable, if $H(u)$ satisfies the regularity conditions:

(1) H(1)=1
(2) $H(u)$ continuous and bounded. The first and second derivatives of H with respect to u exists and are known constants at a given point $u = 1$ at $|u - 1| < 1$.
(3) $H(u)$ is expandable under Taylor's series expansion subject to the required regularity conditions as:

$$\begin{aligned} H(u) = & \quad H(1 + (u-1)) \\ = & \quad H(1) + (u-1)\left(\frac{\partial H}{\partial u}\right)_{u=1} + (u-1)^2 \frac{1}{2}\left(\frac{\partial^2 H}{\partial u^2}\right)_{u=2} + ... \end{aligned}$$

The application of the above theorem gives

$$\bar{y}H(u) = Y\left(H(1) + \theta H_2 C_x^2 + \theta H_1 C_{yx} + O\left(1/n^2\right)\right)$$

where $H_1 = \left(\frac{\partial H}{\partial u}\right)_{u=1}$ and $H_2 = \frac{1}{2}\left(\frac{\partial^2 H}{\partial u^2}\right)_{u=1}$

The first order bias is Bias $= \bar{Y}\left(\theta H_2 C_x^2 + \theta H_1 C_{yx}\right)$

Thus we can construct Beale's estimator as $t_B = \bar{y}\frac{\bar{X}}{\bar{x}}\left[\frac{1-\theta H_1 C_{yx}}{(1+\theta H_2 C_x^2}\right]$

And Tin's estimator as $t_T = \bar{y}\frac{\bar{X}}{\bar{x}}\left[1-\theta\left(H_1C_{yx}+H_2C_x^2\right)\right]$

Koop's estimator as $t_K = \bar{y}\frac{\bar{X}}{\bar{x}}\left[1+\theta\left(H_1C_{yx}+H_2C_x^2\right)\right]^{-1}$

Depending on the functional form of $H(u)$, the values of H_1 and H_2 will vary.

3.2. An alternative class of almost unbiased ratio estimators

In the following we derive an alternative simple method for the characterization of a class of almost unbiased estimators with less restrictive assumptions.
Define the generalized ratio estimator as

$$r_g = \frac{\bar{y}+\theta a}{\bar{x}+\theta b} \tag{1}$$

where a and b may be either real constants or random variables converging in probability to constants.
Let $\bar{y} = \bar{Y}(1+e_0)$ and $\bar{x} = \bar{X}(1+e_1)$,

$E(e_0) = E(e_1) = 0$ and $V(e_0) = \theta C_y^2$, $V(e_1) = \theta C_x^2$

$Cov(e_0, e_1) = \theta C_{yx}$.

Now expanding r_g in a binomial series with assumption $|e_1 + \frac{\theta}{\bar{X}}| < 1$ for all C_n^N possible samples, and retaining terms up to and including degree two in e_0 ande_1, we have

$$\begin{aligned} r_g = R[1 &- \frac{\theta b}{\bar{X}} + e_1^2 + \frac{\theta^2 b^2}{\bar{X}^2} - 2e_1\frac{\theta b}{\bar{X}} - e_0e_1 + e_0(e_1^2 + \frac{\theta^2 b^2}{\bar{X}^2} - 2e_1\frac{\theta b}{\bar{X}}) + \frac{\theta a}{\bar{Y}} - \frac{\theta^2 ab}{\bar{Y}\bar{X}} \\ &+ \frac{\theta a}{\bar{Y}}(e_1^2 + \frac{\theta^2 b^2}{\bar{X}^2} - 2e_1\frac{\theta b}{\bar{X}})] \end{aligned} \tag{2}$$

Taking expectation of r_g and keeping terms up to $O(1/n)$, we have

$$E(r_g) = R\left[1 - \frac{\theta b}{\bar{X}} + \theta C_x^2 - \theta C_{yx} + \frac{\theta a}{\bar{Y}}\right] \tag{3}$$

$$Bias(r_g) = E(r_g) - R \tag{4}$$

In order that the first order bias is zero we put,

$$\frac{\theta a}{\bar{Y}} - \frac{\theta b}{\bar{X}} = \theta\left(C_{yx} - C_x^2\right) \tag{5}$$

or

$$\frac{a}{\bar{Y}} - \frac{b}{\bar{X}} = \left(C_{yx} - C_x^2\right)$$

For a chosen value of b or a we get a value for a or b.
Hence we may state the following theorem.

Theorem 2
The necessary and sufficient condition for r_g to be almost unbiased for $\mathbf{R}$ is

$$\frac{a}{\bar{Y}} - \frac{b}{\bar{X}} = \left(C_{yx} - C_x^2\right)$$

Case I Put $b = \bar{X}C_x^2$
Thus, we have $\frac{a}{\bar{Y}} = \left(C_{yx} - C_x^2\right) + \frac{b}{\bar{X}} = \left(C_{yx} - C_x^2\right) + \frac{\bar{X}C_x^2}{\bar{X}} = C_{yx}$,
which simplifies to $a = \bar{Y}C_{yx} = \frac{S_{yx}}{\bar{X}}$
a and b are functions of unknown population parameters which are usually unknown. These may be replaced by the corresponding design consistent estimates, which does not affect order of approximations. Thus, $\hat{a} = \frac{s_{yx}}{\bar{x}}$ and $\hat{b} = \frac{s_x^2}{\bar{x}}$.
After substitution of $\hat{a}$ and $\hat{b}$, in place of a and b, r_g is rewritten as
$r_{gB} = \frac{\bar{y}+\theta s_{yx}/\bar{x}}{\bar{x}+\theta s_x^2/\bar{x}} = r\left[\frac{1+\theta c_{yx}}{1+\theta c_x^2}\right]$, which is r_B, an almost unbiased ratio type estimator proposed by Beale (1962).

Case II Put $b = 0$, then $\frac{a}{\bar{Y}} = \left(C_{yx} - C_x^2\right)$
The design consistent estimator $\hat{a} = y\left(c_{yx} - c_x^2\right)$
Hence r_G is rewritten as
$r_{gT} = r\left[1 + \theta\left(c_{yx} - c_x^2\right)\right] = r_T$, which was proposed by Tin (1965).

Case III Put $a = 0$, then

$$b = \bar{X}C_x^2 - \bar{X}C_{yx}.$$

The design consistent estimator $\hat{b} = \bar{x}c_x^2 - \bar{x}c_{yx}$
Then r_g is rewritten as

$$r_{gk} = \frac{r}{\left[1 + \theta\left(\frac{s_x^2}{\bar{x}^2} - \frac{s_{yx}}{\bar{y}\bar{x}}\right)\right]} = \frac{r}{\left[1 + \theta\left(c_x^2 - c_{yx}\right)\right]} = r_k$$

which was suggested by Koop (1951) by dividing the estimator by the estimate of its bias component.
Thus we see that we can generate large class of almost unbiased ratio estimator by suitable choice of either a or b.

4. Generalized almost unbiased (Linear variety) separation of bias precipitates method(Singh-Singh,1993)

Construct an estimator as

$$t = w_1t_1 + w_2t_2 + w_3t_3$$

$$w_1 + w_2 + w_3 = 1 \tag{6}$$

$$t_1 = \bar{y}, t_2 = \bar{y}H(u), t_3 = \bar{y}G(u)$$

$H(u)$ and $G(u)$ are continuous bounded functions of $u = \bar{x}/\bar{X}$, and $H(u)$ and $G(u)$ are not equivalent and. satisfy following assumptions

(1) $H(1) = 1, G(1) = 1.$
(2) First and second derivatives of $H(u)$ and $G(u)$ exist.

Writing $H(u) = H(1 + u - 1)$, $G(u) = G(1 + u - 1)$ and expanding $H(u)$ and $G(u)$ in Taylor's series at the point $u = 1$ with $|u - 1| < 1$, we have

$$\begin{aligned} H(u) &= H(1) + (u-1) H'(u)_{u=1} + \frac{1}{2}(u-1)^2 H''(u)_{u=1} + ... \\ G(u) &= G(1) + (u-1) G'(u)_{u=1} + \frac{1}{2}(u-1)^2 G''(u)_{u=1} + ... \end{aligned}$$

where $H'(u)$ and $G'(u)$ are first derivatives of $H(u)$ and $G(u)$ respectively ,$H''(u)$ and $G''(u)$ are second derivatives respectively.
Writing $H'(u)_{u=1} = H$ and $\frac{1}{2}H''(u)_{u=1} = H_2$
$G'(u)_{u=1} = G_1$ and $\frac{1}{2}G''(u)_{u=1} = G_2$,
we have

$$\begin{aligned} H(u) &= 1 + (u-1) H_1 + \frac{1}{2}(u-1)^2 H_2 + ... \\ G(u) &= 1 + (u-1) G_1 + \frac{1}{2}(u-1)^2 G_2 + ... \end{aligned}$$

Defining $e_0 = \frac{\bar{y}}{\bar{Y}} - 1$ and $e_1 = u - 1$. we write

$$\begin{aligned} E(t) &= w_1 E(t_1) + w_2 E(t_2) + w_3 E(t_3) \\ &= E[w_1 \bar{Y}(1 + e_0) + w_2 \bar{Y}(1 + e_0)(1 + e_1 H_1 + e_1^2 H_2 + ...) \\ &\quad + w_3 \bar{Y}(1 + e_0)(1 + e_1 G_1 + e_1^2 G_2 + ...)] \end{aligned}$$

$$Bias(t) = w_2 \bar{Y}\theta \left[H_2 C_x^2 + H_1 C_{xy}\right] + w_3 \bar{Y}\theta \left[G_2 C_x^2 + G_1 C_{xy}\right]$$

Substituting the bias of t to $O(1/n)$ equal to zero, we have

$$w_2 \left[H_2 C_x^2 + H_1 C_{xy}\right] + w_3 \left[G_2 C_x^2 + G_1 C_{xy}\right] = 0 \tag{7}$$

For optimum variance of t to $O(1/n)$ we require

$$w_2 H_1 + w_3 G_1 = -\rho \frac{C_y}{C_x} = -k, say \tag{8}$$

where

$$k = \rho \frac{C_y}{C_x}$$

We now put the equations (6), (7) and (8) in the following matrix form to solve for the weights w_1, w_2 and w_3 and hence the almost unbiased ratio type estimator t , whose bias is of $O(1/n)$ and the asymptotic variance is equal to that of the linear regression estimator.

$$\begin{bmatrix} 0 & H_1 & G_1 \\ 1 & 1 & 1 \\ 0 & B_2 & B_3 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = \begin{bmatrix} -k \\ 1 \\ 0 \end{bmatrix} \tag{9}$$

where $B_2 = B(t_2) = \theta \left(H_2 C_x^2 + H_1 C_{xy} \right)$ and $B_3 = B(t_3) = \theta \left(G_2 C_x^2 + G_1 C_{xy} \right)$

The solutions of w_1, w_2 and w_3 are
$w_1 = \Delta_1/\Delta$, $w_2 = \Delta_2/\Delta$ and $w_3 = \Delta_3/\Delta$, where
$\Delta_1 = -kB_3 + kB_2 + G_1 B_2 - H_1 B_3$, $\Delta_2 = kB_3$, $\Delta_3 = -kB_2$ and $\Delta = G_1 B_2 - H_1 B_3$.
It may be verified,

$$\begin{aligned} Bias\,(t) &= (\Delta_2/\Delta) \left(H_2 C_x^2 + H_1 C_{xy} \right) + \Delta_3/\Delta \left(G_2 C_x^2 + G_1 C_{xy} \right) \\ &= \frac{k \left(G_2 C_x^2 + G_1 C_{xy} \right)}{G_1 \left(H_2 C_x^2 + H_1 C_{xy} \right) - H_1 \left(G_2 C_x^2 + G_1 C_{xy} \right)} \left(H_2 C_x^2 + H_1 C_{xy} \right) \\ &- \frac{k \left(H_2 C_x^2 + H_1 C_{xy} \right)}{G_1 \left(H_2 C_x^2 + H_1 C_{xy} \right) - H_1 \left(G_2 C_x^2 + G_1 C_{xy} \right)} \left(G_2 C_x^2 + G_1 C_{xy} \right) = 0 \end{aligned} \tag{10}$$

As t happens to be functions of unknown population parameters C_x, C_y and ρ, these may be substituted by their design consistent estimates. The resulting asymptotic variance of the estimator after substituting the estimates of the unknown parameters is equal to the asymptotic variance of the linear regression estimator given by $\theta \bar{Y} C_y^2 \left(1 - \rho^2\right)$.

Examples
$H(u) = \left(\frac{\bar{X}}{\bar{x}} \right)$, $G(u) = e^{-\frac{1}{2}(\bar{x} - \bar{X})\bar{X}}$, $H_1 = -1, G_1 = -\frac{1}{2}, H_2 = 1, G_2 = \frac{3}{8}$
$H(u) = \left(\frac{\bar{X}}{\bar{x}} \right) \alpha$, $G(u) = \frac{\bar{X}}{\alpha \bar{x} + (1-\alpha)\bar{X}}$, $H_1 = -\alpha, G_1 = -\alpha, H_2 = \alpha(\alpha+1)/2, G_2 = \alpha^2$.

We now state the following Theorem as follows

Theorem 3

$$t_1 = \bar{y},\ t_2 = \bar{y} H(u),\ t_3 = \bar{y} G(u)$$

where $H(u)$ and $G(u)$ are continuous and bounded functions, satisfying certain regularity functions stated above ., then

$$t = w_1 t_1 + w_2 t_2 + w_3 t_3$$

with $w_1 + w_2 + w_3 = 1$, $w_2 = \frac{k(G_2 C_x^2 + G_1 C_{xy})}{G_1(H_2 C_x^2 + H_1 C_{xy}) - H_1(G_2 C_x^2 + G_1 C_{xy})}$

and $w_3 = -\frac{k(H_2C_x^2+H_1C_{xy})}{G_1(H_2C_x^2+H_1C_{xy})-H_1(G_2C_x^2+G_1C_{xy})}$

where $k = \rho\frac{C_y}{C_x}$, is an almost unbiased estimator with first order bias $O(1/n)$ removed and the variance of t being asymptotically equal to that of a linear regression estimator.

5. Conclusion

Two classes of generalized almost unbiased ratio type estimators are suggested, whose biases to $O(1/n)$ are removed. The well known Beale and Tin type modified ratio estimators are derived as special cases. In latter case another generalized weighted class of almost unbiased ratio estimators are derived where the bias of $O(1/n)$ is also removed and the asymptotic variance equals to that of the linear regression estimator. The suggested methods can also be applied to formulate almost unbiased ratio estimators for other finite population parametric functions.

References

[1] Beale, E .M. L. (1962). Some uses of Computers in operational research, Industrialle organization, 31, 27-28.

[2] Chakrabarty, R.P. (1979). Some ratio type estimators, Journal of the Indian Society of Agricultural Statistics XXXI, 1, 49-62.

[3] Cochran, W. G.(1940). The estimation of the yields of cereal experiments by sampling for the ratio of grains to total produce ,The Journal of Agricultural Science, 30(2), 262-275.

[4] Cochran , W. G. (1977). Sampling Techniques, New York, John Wiley and Sons (3rd ed).

[5] Hutchison, M. C. (1971). A Monte Carlo comparison of some ratio estimators, Biometrika, 58, 313-322.

[6] Koop , J. C. (1951). A note on the bias of the ratio estimate, Bulletin of the, , 33, 141-146.

[7] Lui, K. J (2020). Note on use of composite estimator :an improvement of the ratio estimator, Journal of Official Statistics, 36(1), 137-149.

[8] Mussa, A. S. (1999). A new bias-reducing modification of the finite population ratio estimator and comparison among proposed alternatives, Journal of Official Statistics, 15(1), 25-38.

[9] Quenouille , M. H. (1956). Notes on bias in estimation, Biometrika, 43, 353-360.

[10] Richards, R. P. and Holloway Jim (1987). Monte Carlo studies of for sampling strategies for estimating Tributary loads,Water Resources Research, 23, 1339-1348.

[11] Richards, R. P. (1998). Estimation of pollutant loads in rivers and streams – A guidance document for N P S Programs, Water quality laboratory, Heidelberg college, OH, USA.

[12] Singh, S. and Singh, R.(1993). Almost filtration of bias precipitates: a new approach, Journal of Indian Society of Agricultural Statistics, 45, 214-218.

[13] Srivastava, S. K. (1971). A Generalized Estimator for the Mean of finite population using auxiliary variables. Journal of the American Statistical Association, 66(334), 404-407.

[14] Swain, A. K. P. C. Swain (2019).On a Generalized class of Almost Unbiased Ratio Estimators, Model Assisted Statistics and Applications, 14(3), 269-278.

[15] Sukhatme, P. V., Sukhatme, B. V. and Asok, C. (1984). Sampling Theory of Surveys with Applications, 3rd, edition, Iowa State University Press, Ames , Iowa, USA .

[16] Tin, M.(1965). Comparison of some ratio estimators, Journal of the American Statistical Association, 60(309), 294-302.