Jour. Ind. Soc. Ag. Statistics
Vol. XLII, No. 1 (1990), pp. 108-117

A GENERALISED CHAIN RATIO ESTIMATOR FOR MEAN OF FINITE POPULATION

S. RANI SRIVASTAVA, B. B. KHARE
and
S. R. SRIVASTAVA
Banaras Hindu University, Varanasi 221005,
(Received : February, 1987)

SUMMARY

A generalised chain ratio estimator for the mean of a finite population has been proposed. Expressions for bias and mean square error have been derived and a comparison has been made with suitable estimators. The efficiency of the proposed estimator has also been studied with the help of numerical illustration.

Key words : Chain ratio estimator; Double sampling; Mean square error; Relative bias; Relative efficiency.

Introduction

Using information on the auxiliary variable x, we often use classical ratio and product estimators depending upon the condition $\rho_{yx} > \frac{1}{2}\frac{C_x}{C_y}$ and $\rho_{yx} < -\frac{1}{2}\frac{C_x}{C_y}$ respectively, where C_y, C_x denote the coefficients of variation of the variable y and x and ρ_{yx} is the correlation coefficient between y and x.

Let $\bar{Y}_N$, $\bar{X}_N$ and $\bar{y}_n$, $\bar{x}_n$ be the population and sample means of the main and auxiliary variables y and x based on a sample of size n with SRSWOR, where N is the size of the population. In practice, when $\bar{X}_N$ is not known a priori, the technique of double sampling is used. In such case $\bar{X}_N$ is replaced by its estimate $\bar{x}_{n_1}$, based on a preliminary larger sample of size $n_1 > n$. The double sampling ratio and product estimators in this case will be given by

$$T_1 = \frac{\bar{y}_n}{\bar{x}_n}\bar{x}_{n1} \quad \text{and} \quad T_2 = \frac{\bar{y}_n\,\bar{x}_n}{\bar{x}_{n1}}. \tag{1.1}$$

If $\bar{X}_N$ is not known, but $\bar{Z}_N$, the population mean of another cheaper auxiliary variable z closely related to x but compared to x remotely related to y (i.e. $\rho_{yx} > \rho_{yz}$) is also available. In this case, Chand [1] defined the chain ratio estimator

$$T_3 = \frac{\bar{y}_n}{\bar{x}_n}\,\frac{\bar{x}_{n1}}{\bar{z}_{n1}}\,\bar{Z}_N \tag{1.2}$$

where $\bar{z}_{n1}$ denote the sample mean based on $n_1 > n$ units of the auxiliary variable z.

If y and x are negatively correlated then similar type of estimator T_4 may be defined as

$$T_4 = \frac{\bar{y}_n\,\bar{x}_n\,\bar{z}_{n1}}{\bar{x}_{n1}\,\bar{Z}_N} \tag{1.3}$$

Kiregyera [2], [3] have also studied a chain ratio type estimator and regression type estimator using two auxiliary variables under a super population model.

In the present context a generalised chain ratio estimator is proposed using another auxiliary variable z when the population mean of the main auxiliary variable x is not known. Expressions for bias and mean square error have been derived to the first degree of approximation and a comparison is made with suitable estimators. Recommendations regarding the choice of constant α_1 and α_2 is made for minimum bias and mean square error. The proposed estimators are also studied with the help of numerical illustration in both the cases, when cost of the survey is fixed or unlimited.

2. The Estimator, Bias and Mean Square Error

We propose a generalised chain ratio estimator T_5 for the population mean of the variable y using additional auxiliary variable z related to main auxiliary variable x, which is given as

$$T_5 = \bar{y}_n \left(\frac{\bar{x}_{n1}}{\bar{x}_n}\right)^{\alpha_1} \left(\frac{\bar{Z}_N}{\bar{z}_{n1}}\right)^{\alpha_2} \tag{2.1}$$

where α_1 and α_2 are constants. The optimum values of α_1 and α_2 are obtained by minimising the mean square error of T_5.

The expression for the relative bias and mean square error of the estimator T_5 upto the terms of 0 (n^{-1}) are given by

$$\text{RB } (T_5) = \frac{E(T_5) - \bar{Y}_N}{\bar{Y}_N} = \frac{1}{2}\Big[f_0 C_x^2 \, \alpha_1 \left(\alpha_1 - \overline{2\,k_1 - 1}\right)$$

$$+ \; f_1 \, C_z^2 \, \alpha_2 \left(\alpha_2 - \overline{2k_2 - 1}\right)\Big] \tag{2.2}$$

$$\text{MSE } (T_5) = \bar{Y}_N^2 \Big[f_0 \, V_5 + f_1 \, V_6 \Big], \quad \text{where } f_0 = \left(\frac{1}{n} - \frac{1}{n_1}\right)$$

$$\text{and } f_1 = \left(\frac{1}{n_1} - \frac{1}{N}\right) \tag{2.3}$$

The MSE (T_5) will be minimum for optimum value of α_1 and α_2 which is obtained as

$$\alpha_{1_{opt}} = \rho_{yx} \frac{C_y}{C_x} = k_1 \quad \text{and} \quad \alpha_{2_{opt}} = \rho_{yz} \frac{C_y}{C_z} = k_2$$

The minimum value of MSE (T_5) is

$$\text{MSE } (T_5)_{opt} = \bar{Y}_N^2 \cdot V_0 \left[f_0 \, (1 - \rho_{yx}^2) + f_1 \, (1 - \rho_{yz}^2) \right] \tag{2.4}$$

where

$$V_0 = C_y^2, \; V_5 = C_y^2 + \alpha_1 \; C_x^2 - 2 \, \alpha_1 \, \rho_{yx} \, C_y \, C_x$$

and

$$V_6 = C_y^2 + \alpha_2^2 \, C_z^2 - 2 \, \alpha_2 \, \rho_{yz} \, C_y \, C_z$$

3. Comparison of Bias and Mean Square Error of the Proposed Estimator T_5 with Suitable Estimators

The expressions for relative bias of estimators T_1, T_2, T_3 and T_4 are given as follows :

$$\text{RB}(T_1) = f_0 B_1, \ \text{RB}(T_2) = f_0 D_1, \tag{3.1}$$

$$\text{RB}(T_3) = f_0 B_1 + f_1 B_2, \tag{3.2}$$

$$\text{RB}(T_4) = f_0 D_1 + f_1 D_2, \tag{3.3}$$

where

$B_1 = C_x^2 (1 - k_1)$, $B_2 = C_z (1 - k_2)$, $D_1 = \rho_{yx} C_y C_x$ and $D_2 = \rho_{yz} C_y C_z$.

The estimator T_5 will be unbiased upto the terms of $0 (n^{-1})$ if $\alpha_1 = \alpha_2 = 0$, or $\alpha_1 = (2k_1 - 1)$, $\alpha_2 = (2k_2 - 1)$.

RB $(T_5) \leqslant$ RB (T_3) if

$$\frac{1}{2} f_0 C_x^2 \{\alpha_1^2 - (2k_1 - 1)\alpha_1 - 2(1 - k_1)\} + \frac{1}{2} f_1 C_z^2 \{\alpha_2^2 - (2k_2 - 1)\alpha_2 - 2(1 - k_2)\} \leqslant 0 \tag{3.4}$$

We have two quadratic equations in α_1 and α_2, the sum of which must be negative. One of the possibilities of (3.4) to be negative is that

$$\{\alpha_1^2 - (2k_1 - 1)\alpha_1 - 2(1 - k_1)\} < 0 \tag{3.5}$$

and

$$\{\alpha_2^2 - (2k_2 - 1)\alpha_2 - 2(1 - k_2)\} \leqslant 0 \tag{3.6}$$

Now solving the expressions (3.5) and (3.6) for α_1 and α_2, we have the following sufficient conditions :

$$2(k_1 - 1) \leqslant \alpha_1 < 1, \ 2(k_2 - 1) < \alpha_2 \leqslant 1$$

$$\text{if} \quad 1 < k_1 \leqslant 1.5, \ 1 \leqslant k_2 < 1.5. \tag{3.7}$$

The expression for mean square error of the estimators T_1, T_2, T_3 and T_4 upto the terms of 0 (n^{-1}) are given) as follows :

$$\text{MSE}(T_1) = \bar{Y}_N^2 (f_0 V_1 + f_1 V_0), \tag{3.8}$$

$$\text{MSE}(T_2) = \bar{Y}_N^2 (f_0 V_2 + f_1 V_0), \tag{3.9}$$

$$\text{MSE}(T_3) = \bar{Y}_N^2 (f_0 V_1 + f_1 V_3). \tag{3.10}$$

$$\text{MSE}(T_4) = \bar{Y}_N^2 (f_0 V_2 + f_1 V_4), \tag{3.11}$$

where

$$V_1 = C_y^2 + C_x^2 - 2\rho_{yx} C_y C_x,\ V_2 = C_y^2 + C_x^2 + 2\rho_{yx} C_y C_x,$$

$$V_3 = C_y^2 + C_z^2 - 2\rho_{yz} C_y C_z, \text{ and } V_4 = C_y^2 + C_z^2 + 2\rho_{yz} C_y C_z.$$

The MSE (T_3) < MSE (T_1) and MSE (T_4) < MSE (T_2) if $\rho_{yz} > \frac{1}{2}\frac{C_z}{C_y}$ and $\rho_{yz} \leqslant -\frac{1}{2}\frac{C_z}{C_y}$ respectively.

The MSE (T_5) = MSE (T_3) if $\alpha_1 = \alpha_2 = 1$ or $\alpha_1 = (2k_1 - 1)$ and $\alpha_2 = (2k_2 - 1)$.

The MSE (T_5) < MSE (T_3), if

$$[f_0 C_x^2 (1 - \alpha_1)(\alpha_1 - \overline{2k_1 - 1}) + f_1 C_z^2 (1 - \alpha_2)(\alpha_2 - \overline{2k_2 - 1})] > 0 \tag{3.12}$$

The ranges for α_1 and α_2 may be given under the sufficient conditions which statisfies the expression (3.12) are as follows :

$$(2k_1 - 1) \leqslant \alpha_1 \leqslant 1,\ (2k_2 - 1) < \alpha_2 < 1 \text{ if } \tfrac{1}{2} < k_1 < 1,\ \tfrac{1}{2} < k_2 < 1$$

$$1 \leqslant \alpha_1 < (2k_1 - 1),\ 1 < \alpha_2 \leqslant (2k_2 - 1) \text{ if } k_1 > 1,\ k_2 > 1$$

$$(2k_1 - 1) < \alpha_1 < 1,\ 1 \leqslant \alpha_2 < (2k_2 - 1) \text{ if } \tfrac{1}{2} < k_1 < 1,\ k_2 > 1$$

$$1 < \alpha_1 < (2k_1 - 1),\ (2k_2 - 1) \leqslant \alpha_2 < 1 \text{ if } k_1 > 1,\ \tfrac{1}{2} \leqslant k_2 < 1. \tag{3.13}$$

Similarly in the case of negative correlation,

$$\text{MSE}(T_5) = \text{MSE}(T_4) \text{ if } \alpha_1 = \alpha_2 = -1 \text{ or } \alpha_1 = (2k_1 + 1) \text{ and } \alpha_2 = (2k_2 + 1).$$

$$\text{MSE}(T_5) \lhd \text{MSE}(T_4) \text{ if } [f_0 C_x^2 (1 + \alpha_1)(\overline{2k_1 + 1} - \alpha_1) + f_1 C_z^2 (1 + \alpha_2)(\overline{2k_2 + 1} - \alpha_2)] > 0 \tag{3.14}$$

Similarly the range of α_1 aad α_2 can be obtained which satisfies the condition (3.14).

Hence we see that for the above values of α_1 and α_2 in the specified ranges of k_1 and k_2, the proposed estimator T_5 will always be more efficient than T_1, T_2, T_3 and T_4.

For the purpose of numerical illustration, consider data used by Sukhatme and Chand [4].

$$\bar{Y}_N = 0.293458 \times 10^4 \quad \rho_{yx} = 0.93 \quad C_y^2 = 0.402004 \times 10^1$$

$$\bar{X}_N = 0.103182 \times 10^4 \quad \rho_{yz} = 0.84 \quad C_x^2 = 0.255280 \times 10^1$$

$$\bar{Z}_N = 0.365149 \times 10^4 \quad \rho_{xz} = 0.77 \quad C_z^2 = 0.209379 \times 10^1$$

Here, take $n = 30$, $n_1 = 60$, and $N = 120$.

TABLE 1 : RELATIVE EFFICIENCY OF T_3 AND T_5 w.r.t. T_1

Estimator	*Bias*	*MSE*	*R. E. w.r.t* T_1(%)
T_1	0.710745×10^{-2}	0.37667×10^6	100.00
T_3	0.996779×10^{-2}	0.17715×10^6	212.63
T_5	0.476196×10^{-2}	0.16348×10^6	230.41
$(\alpha_{1opt} = \alpha_{2opt} = 1.2)$			

From the above table it is clear that T_5 is a better estimator than T_1 and T_3 in the sense of bias and mean square error for the optimum value. Further it is also observed that T_5 has minimum mean square error and bias for a big range in the neighbourhood of the optimum value of α_1 and α_2.

The proposed estimator T_5 has also been compared for fixed cost. For this purpose, consider the cost function of the sample design of the form

$$C = C_1 n + C_2 n_1 \tag{3.15}$$

for the two phase sampling using auxiliary character x. Similarly when we use an additional auxiliary variable z then the cost function is given by

$$C = C_1 n + (C_2 + C_3) n_1 \tag{3.16}$$

where C : total cost of the survey and C_1, C_2, C_3 are the costs per unit of collecting information on the main variable y, auxiliary variable x and additional auxiliary variable z respectively. The optimum values of n and n_1 for a fixed cost $C = C_0$ which minimises the mean square error of the estimator T_1, T_2, T_3, T_4 and T_5 are as follows :

Estimatoe	n_{opt}	$n_{1\,opt}$
T_1	$\dfrac{C_0 \sqrt{V_1/C_1}}{(\sqrt{V_1 C_1} + \sqrt{(V_0 - V_1 C_2)}}$	$\dfrac{C_0 \sqrt{(V_0 - V_1)/C_2}}{(\sqrt{V_1 C_1} + \sqrt{(V_0 - V_1) C_2})}$
T_2	$\dfrac{C_0 \sqrt{V_2/C_1}}{(\sqrt{V_2 C_1} + \sqrt{(V_0 - V_2) C_2})}$	$\dfrac{C_0 \sqrt{(V_0 - V_2)/C_2}}{(\sqrt{V_2 C_1} + \sqrt{(V_0 - V_2) C_2}}$
T_3	$\dfrac{C_0 \sqrt{V_1/C_1}}{(\sqrt{V_1 C_1} + \sqrt{(V_3 - V_1)(C_2 + C_3)})}$	$\dfrac{C_0 \sqrt{(V_3 - V_1)/(C_2 + C_3)}}{(\sqrt{V_1 C_1} + \sqrt{(V_3 - V_1)(C_2 + C_3)})}$
T_4	$\dfrac{C_0 \sqrt{V_2/V_1}}{(\sqrt{V_2 V_1} + \sqrt{(V_4 - V_2)(C_2 + C_3)})}$	$\dfrac{C_0 \sqrt{(V_4 - V_2)/(C_2 + C_3)}}{(\sqrt{V_2 C_1} + \sqrt{(V_4 - V_2)(C_2 + C_3)}}$
T_5	$\dfrac{C_0 \sqrt{V_5/C_1}}{(\sqrt{V_5 C_1} + \sqrt{(V_6 - V_5)(C_2 + C_3)})}$	$\dfrac{C_0 \sqrt{(V_6 - V_5)/(C_3 + V_3)}}{(\sqrt{V_5 C_1} + \sqrt{(V_6 - V_5)(C_2 + C_3)}}$

Putting these optimum values of n and n_1 in the expressions of mean square error of T_1, T_2, T_3, T_4 and T_5, we have

$$\text{MSE } (T_1)_{opt} = \bar{Y}_N^2 \left[\frac{(\sqrt{V_1 C_1} + \sqrt{(V_0 - V_1) C_2})^2}{C_0} - \frac{1}{N} V_0 \right] \quad (3.17)$$

$$\text{MSE } (T_2)_{opt} = \bar{Y}_N^2 \left[\frac{(\sqrt{V_2 V_1} + \sqrt{(V_0 - V_2) C_2})^2}{C_0} - \frac{1}{N} V_0 \right] \quad (3.18)$$

$$\text{MSE } (T_3)_{opt} = \bar{Y}_N^2 \left[\frac{\sqrt{V_1 C_1} + \sqrt{(V_3 - V_1)(C_2 + C_3)})^2}{C_0} - \frac{1}{N} V_3 \right] \quad (3.19)$$

$$\text{MSE } (T_4)_{opt} = \bar{Y}_N^2 \left[\frac{(\sqrt{V_2 C_1} + \sqrt{(V_4 - V_3)(C_2 + C_3)})^2}{C_0} - \frac{1}{N} V_4 \right] \quad (3.20)$$

$$\text{MSE } (T_5)_{opt} = \bar{Y}_N^2 \left[\frac{(\sqrt{V_5 C_1} + \sqrt{(V_6 - V_5)(C_2 + C_3)})^2}{C_0} - \frac{1}{N} V_6 \right] \quad (3.21)$$

$$\text{MSE } (T_{5\, \alpha_1\, opt.,\, \alpha_2\, opt.})_{opt}$$

$$= \bar{Y}_N^2 V_0 \left[\frac{(\sqrt{(1 - \rho_{yx}^2) C_1} + \sqrt{(\rho_{yx}^2 - \rho_{yz}^2)(C_2 + C_3)})^2}{C_0} - \frac{1}{N} (1 - \rho_{yz}^2) \right] \quad (3.22)$$

Numerical illustration

Suppose for estimating the mean yield ($\bar{Y}_N$) of fibre per plant in jute fibre crop, the auxiliary characteristics height (x) and the additional characteristic diameter (z) were taken for the population consisting of 1000 jute plants (capsulance). The values obtained are as follows :

$\bar{Y}_N = 5.69$ gms, $C_y^2 = 0.0568$, $C_x^2 = 0.0385$, $C_z^2 = 0.01269$,

$\rho_{yz} = 0.7418$, $\rho_{yx} = 0.5677$, $\rho_{xz} = 0.2063$.

Let the total fixed cost C_0 = Rs. 5000/-, C_1 = Rs. 25/- per unit, C_2 = Rs. 2/- per unit, and C_3 = Rs. 1/- per unit.

The optimum values of α_1 and α_2 is $k_1 = 1.9176$ and $k_2 = 1.2011$

respectively. In this example we can only use the estimator T_1, T_3 and T_5. The value of bias and mean square error of estimators T_1, T_3 and T_5 are computed for the fixed cost. It is observed that the bias of $(T_5)_{opt.}$ is less than $B(T_1)$, but slightly more than the $B(T_3)$.

The MSE (T_3) < MSE (T_1), which were found to be 5.7636 x 10^{-3} and 6.3373 x 10^{-3} respectively.

The relative efficiency of the estimator T_5 w.r.t. estimator T_3 for different choice of α_1 and α_2 is given in Table 2.

TABLE 2—RELATIVE EFFICIENCY OF T_5 w.r.t. T_3 (in %)

α_2 \ α_1	1.1	1.2	1.3	1.4	1.9	2.6	2.7
1.1	101.96	103.45	105.02	106.59	111.07	103.98	102.46
1.2	102.19	103.66	105.23	106.77	111.27	104.20	102.68
1.3	101.96	103.45	105.02	106.59	111.07	103.99	102.46
1.4	101.28	102.82	104.44	106.02	110.56	103.39	101.81
1.5	100.20	101.82	103.49	105.12	109.71	102.40	100.76
1 6	98.80	100.51	102.25	103.92	108.59	101.12	99.39
1.7	97.15	98.96	100.76	102.49	107.23	99.60	97.78

The relative efficiency of the estimator T_5 was found to be maximum for the optimum values of α_1 and α_2 which is approximately 111.27. As the value of α_1 and α_2 increases, the relative efficiency of T_5 w.r.t. T_3 increases, till it does not attain the optimum value of α_1 and α_2 respectively and later on it starts decreasing as α_1 and α_2 increases beyond $(2k_1 - 1)$ and $(2k_2 - 1)$ respectively. When the values of (α_1 and α_2) is more than ($2k_1 - 1$ and $2k_2 - 1$), the estimator T_5 is inefficient. Hence it is seen that estimator T_5 is better than T_1 and T_3 in a wide range of α_1 and α_2 for fixed cost.

The optimum value of n and n_1 in the estimator T_1, T_3 and T_5 is as follows :

	T_1	T_3	T_5
$n_{opt.}$	161	174	160
$n_{1opt.}$	488	220	330

The total number of observations drawn on the main character is approxmately same in T_1 and T_5 but less than that of T_3. In T_3, the number of observations drawn on the main character is increased and remaining cost is used in collecting information on auxiliary charcters (i.e. main as well as additional auxiliary characters). In T_5, the total number of observations drawn on the main character is less, in comparison to T_3. Therefore in T_5 more information is collected on the auxiliary characters than T_3, which makes T_5 more efficient than T_3.

ACKNOWLEDGEMENT

The authors are grateful to Prof. Lal Chand, J.N.K.V.V., Jabalpur for his suggestions in the present work. The authors are also highly obliged to the referee for his valuable suggestions in the improvement of the paper.

REFERENCES

[1] Chand, L. (1975) : Some ratio type estimator based on two or more auxiliary variables. *Uupublished Ph.D. dissertation.* Iowa State University, Ames, Iowa.

[2] Kiregyera, B. (1980) : A chain ratio-type estimator in finite population double sampling using two auxiliary variables. *Metrika*, 27 : 217-223.

[3] Kiregyera, B. (1984) : Regression type estimators using two auxiliary variables and the model of double sampling from finite population. *Metrika*, 31 : 215-226.

[4] Sukhatme, B. V. and Chand, L. (1977) : Multivariate ratio-type estimators. *Proceedings, Social Statistics Section*, American Statistical Association, 927-931.