Jour. Ind. Soc. Ag. Statistics
43 (2), 1991 : 127—135

Class of Almost Unbiased Dual to Product Estimators in Sample Surveys

K.S. Kushwaha
J.N.K.V.V., Jabalpur
(Received : June, 1987)

Summary

Ratio and product estimators are generally biased estimators. In the present paper, the author has defined a general class of ratio estimators alternative to product estimators and tried to reduce/eliminate its bias using jackknife technique introduced by Quenouille [5] to make the class of estimators to be almost unbaised. The expressions for the bias and mean squared error (MSE) of the proposed class of estimators have been derived to the first order of approximation under simple random sampling without replacement (SRSWOR) strategy and the optimum estimator in the class is also identified. The results are illustrated by a simple numerical example.

Key words : Bias, mean squared error, dual to product estimator, optimum estimator.

Introduction

The ratio estimator and its jackknife version have been considered by Rao, and webster [9], Rao [8], Rao and Rao [10] and Hutchison [3] among others using a super population model. However, ratio estimator is known to be inferior to the product estimator when correlation coefficient between two positive variates y and x is highly negative (see Murthy [4]. Some attention to the product estimators has been given by Goodman [2], Srivastava [11], Wu and Chang [15] and Srivastava *et al.* [12] among others. The product estimator has not acquired the popularity as the ratio estimator due to the misconcept about non occurrence of negative correlation in practice and the apparent superiority of ratio estimator (under positive correlation) under design based comparisons. It is true that the positive correlations are often encountered in practice but negative correlations are not uncommon at all. The negative correlations are often induced through inverse transformation of regressor variate x. It has been found that the correlation between (yield and alkali) and (yield and days to maturity) is negative.

Consider a finite population (x_i , y_i), i = 1, 2, . . . , N with population means ($\overline{X}$, $\overline{Y}$), population variances (S_x^2 , S_y^2) and covariance S_{xy}. In forming a product estimator $\hat{Y}_p$ of $\bar{Y}$ from a simple random sample (x_i, y_i), $i = 1, 2, \ldots, n$ of size n, we obtain

$$\hat{Y}_p = \bar{y}\,\bar{x}/\bar{X} \tag{1.1}$$

which is due to Robson [6] and rediscovered by Murthy [4] where $\bar{x}$ and $\bar{y}$ are the sample means.

Consider the estimator

$$\bar{x}^* = (N\bar{X} - n\bar{x})/(N - n) \tag{1.2}$$

which is also an unbiased estimator for $\bar{X}$. Using $\bar{x}^*$ Srivenkataramana [13] proposed a dual to product estimator to estimate $\bar{Y}$ as

$$\hat{Y}_r^* = \frac{\bar{y}}{\bar{x}^*}\,\bar{X} \tag{1.3}$$

Obviously, $\hat{Y}_r^*$ is a biased estimator for $\bar{Y}$. To reduce/eliminate the bias of $\hat{Y}_r^*$, we use Jackknife technique due to Quenouille [5] and obtain a general class of estimators for $\bar{Y}$ and study the properties to the first order approximation.

2. *Class of Estimators*

In this section, we define a class of almost unbiased dual to product estimators using jackknife technique and also an approach adopted by Rao [7]. Here, we take $n = pm$ and split the sample at random into p subsamples of m units each. Let $\bar{y}_j$ and $\bar{x}_j$ $(j = 1, 2, \ldots, p)$ be unbiased estimators of $\bar{Y}$ and $\bar{X}$ based on subsamples and $\bar{y}, \bar{x}$ those based on entire sample of size n. Define the complementary subsample means $\bar{y}_j'$ and $\bar{x}_j'$ $(j = 1, 2, \ldots, p)$ as follows :

$$\begin{aligned} \bar{y}_j' &= (n\bar{y} - m\bar{y}_j)/(n - m) = (n\bar{y} - m\bar{y}_j)/m' \\ \bar{x}_j' &= (n\bar{x} - m\bar{x}_j)/(n - m) = (n\bar{x} - m\bar{x}_j)/m' \end{aligned} \tag{2.1}$$

which are subsample means based on $m' = (n - m)$ units obtained by omitting jth $(j = 1, 2, \ldots, p)$ group from the entire sample of size n.

Consider

$$\begin{aligned} \bar{x}_j^{*\prime} &= (N\bar{X} - n\bar{x}_j')/(N - n) \\ \text{and} \quad \bar{x}^* &= (N\bar{X} - n\bar{x})/(N - n) \end{aligned} \tag{2.2}$$

Now, consider the ratio type estimators as

$$\hat{Y}_r^* = \frac{\bar{y}}{\bar{x}^*} X \quad \text{and} \quad \hat{Y}_{r.}^* = \frac{1}{p} \sum_{j=1}^{p} \frac{\bar{y}_j}{\bar{x}_j^{*\prime}} X \tag{2.3}$$

based on entire sample of size n and subsample of size (n – m) respectively. As motivated by Rao [7], we propose a general class of almost unbaised dual to product estimator for estimating $\bar{Y}$ as

$$T_R = \alpha \hat{Y}_r^* + \{1 - E(f(\alpha))\} \hat{Y}_{r.}^* \tag{2.4}$$

Where α is a random variable independent of $\hat{Y}_r^*$ and f (α) is a function of α. Then

$$E(T_R) = \bar{Y}$$

if

$$E[\alpha \hat{Y}_r^* - E(f(\alpha)) \hat{Y}_r^*] = E(\bar{y} - \hat{Y}_{r.}^*) \tag{2.5}$$

For (2.5), $\alpha = \bar{x}^*/\bar{X}$ and $f(\alpha) = \alpha$ is a solution in the sample mean.

Introducing a constant 'q' in the right hand side of (2.5), we can write (2.5) as

$$E[\alpha \hat{Y}_r^* - E(f(\alpha)) \hat{Y}_{r.}^*] = E[\bar{y} - \hat{Y}_{r.}^* - q \hat{Y}_r^* + q \hat{Y}_r^*] \tag{2.6}$$

Following Sukhatme *et al.* [14] the biases of $\hat{Y}_r^*$ and $\hat{Y}_{r.}^*$ to the terms of order of approximation $O(n^{-1})$ can be given as

$$B(\hat{Y}_r^*) = \bar{Y} \frac{C_x^2}{N} \left(\frac{n}{N-n} + K\right) \tag{2.7}$$

$$B(\hat{Y}_{r.}^*) = \bar{Y} C_x^2 \frac{n(N-n+m)}{N(N-n)(n-m)} \left(\frac{n}{N-n} + K\right) \tag{2.8}$$

Where $K = \rho \frac{C_y}{C_x}$ and C_y, C_x are the coefficients of variation for the variates y and x and ρ is the correlation coefficient between the two variates.

From (2.7) and (2.8), it can be shown that

$$\frac{B(\hat{Y}_r^*)}{B(\hat{Y}_{r.}^*)} = \frac{(N-n)(n-m)}{n(N-n+m)} = \frac{1}{g}\,\frac{m'}{(N-m')} = (1-\delta), \quad \text{(say)}$$

$$\Rightarrow \qquad B(\hat{Y}_r^*) = (1-\delta)\,B(\hat{Y}_{r.}^*)$$

where $(1-\delta) = \dfrac{(N-n)}{n}\,\dfrac{m'}{(N-m')}$ and $g = \dfrac{n}{(N-m)}$

$$\Rightarrow \qquad E(\hat{Y}_r^*) - \bar{Y} = (1-\delta)\left\{E(\hat{Y}_{r.}^*) - \bar{Y}\right\}$$

$$\Rightarrow \qquad E(\hat{Y}_r^*) = \delta\,\bar{Y} + (1-\delta)\,E(\hat{Y}_{r.}^*)$$

$$= \delta\,E(\bar{y}) + (1-\delta)\,E(\hat{Y}_{r.}^*)$$

$$\Rightarrow \qquad \hat{Y}_r^* = \delta\,\bar{y} + (1-\delta)\,\hat{Y}_{r.}^* \tag{2.9}$$

From (2.9) and (2.6), we have

$$E[\alpha\hat{Y}_r^* - E(f(\alpha))\hat{Y}_{r.}^*] = E\left[\left\{q + (1-\delta q)\frac{\bar{x}^*}{\bar{X}}\right\}\hat{Y}_r^* - \{1 + (1-\delta)q\}\hat{Y}_{r.}^*\right] \tag{2.10}$$

which shows that a solution is now given as

$$\alpha = q + (1-\delta q)\frac{\bar{x}^*}{\bar{X}} \quad \text{and} \quad f(\alpha) = \alpha \tag{2.11}$$

for which $E(f(\alpha)) = 1 + (1-\delta)q$

Thus, using (2.11) in (2.4), a general class of almost unbiased dual to product estimators can be obtained as

$$T_R = \left\{q + (1-\delta q)\frac{\bar{x}^*}{\bar{X}}\right\}\hat{Y}_r^* - (1-\delta)\,q\,\hat{Y}_{r.}^*$$

$$= q\,\hat{Y}_r^* + (1-\delta q)\,\bar{y} - (1-\delta)\,q\,\hat{Y}_{r.}^* \tag{2.12}$$

Thus, the following theorem can be stated here

Theorem 2.1 The class of estimators

$$T_R = \alpha \hat{Y}_r^* + E\{1 - E(f(\alpha))\} \hat{Y}_{r.}^* \quad \text{would be almost unbiased if}$$

$$\alpha = q + (1 - \delta q)\frac{\bar{x}^*}{\bar{X}} \quad \text{and} \quad f(\alpha) = \alpha \quad \text{for which}$$

$$E(f(\alpha)) = 1 + (1 - \delta)q$$

When N is very large or population is infinite, the class of almost unbiased dual to product estimators in (2.12) reduces to

$$T_R^* = q\hat{Y}_r^* + (1 - \delta^* q)\bar{y} - (1 - \delta^*)q\hat{Y}_{r.}^*$$

where $\delta^* = \frac{1}{p}$ and $(1 - \delta^*) = \frac{p-1}{p}$

Remark The class of estimators T_R in (2.12) reduces to the following set of unbiased estimators for $\bar{Y}$ for suitable choices of q.

(i) For $q = 0$, $T_{R_1} = \bar{y}$, the usual sample mean.

(ii) For $q = \frac{1}{\delta}$, $T_{R_2} = \frac{\hat{Y}_r^*}{\delta} - \left(\frac{1-\delta}{\delta}\right)\hat{Y}_{r.}^*$ (2.14)

or $T_{R_2} = \frac{(N-n+m)p}{N}\hat{Y}_r^* - \frac{(N-n)(p-1)}{n}\hat{Y}_{r.}^*$

(See Sukhmatme *et al.* [14])

(iii) When N is very large or population is infinite, the estimate $T_{R_2}^*$ turns out to be

$$T_{R_2} = p\hat{Y}_r^* + (1-p)\hat{Y}_{r.}^* \tag{2.15}$$

which is Quenouille [5] type estimator

(iv) For $q = \frac{1}{1-\delta}$

$$T_{R_3} = (1 - 2\delta)(1-\delta)^{-1}\bar{y} + (1-\delta)^{-1}\hat{Y}_r^* - \hat{Y}_{r.}^* \tag{2.16}$$

Several other estimators can be cited for other suitable choices of q in (2.12).

3. *Optimum Estimator in the Class*

From (2.12), we have

$$T_R = q\,\hat{Y}_r^* + (1-\delta q)\,\bar{y} - (1-\delta)\,q\,\hat{Y}_{r.}^*$$

and

$$V(T_R) = q^2 V(\hat{Y}_r^*) + (1-\delta q)^2 V(\bar{y}) + (1-\delta)^2 q^2 V(\hat{Y}_{r.}^*) + 2q(1-\delta q)\,\mathrm{Cov}(\hat{Y}_r^*, \bar{y}) - 2q(1-\delta q)(1-\delta)\,\mathrm{Cov}(\bar{y}, \hat{Y}_{r.}^*) - 2q^2(1-\delta)\,\mathrm{Cov}(\hat{Y}_r^*, \hat{Y}_{r.}^*) \tag{3.1}$$

Following Sukhatme *et al.* [14], to the terms of the order of approximation $O(n^{-1})$, it can be checked that

$$V(\bar{y}) = \frac{N-n}{nN}\bar{Y}^2 C_y^2$$

$$V(\hat{Y}_r^*) = V(\hat{Y}_{r.}^*) = \mathrm{Cov}(\hat{Y}_r^*, \hat{Y}_{r.}^*) = \frac{N-n}{n}\bar{Y}^2 [C_y^2 + gC_x^2(g+2K)]$$

$$\mathrm{Cov}(\bar{y}, \hat{Y}_r^*) = \mathrm{Cov}(\bar{y}, \hat{Y}_{r.}^*) = \frac{N-n}{n}\bar{Y}^2 [C_y^2 + gKC_x^2] \tag{3.2}$$

Putting the results (3.2) in (3.1) and simplifying it, we get

$$V(T_R) = \delta^2 q^2 V(\hat{Y}_r^*) + (1-\delta q)^2 V(\bar{y}) + 2\delta q(1-\delta q)\,\mathrm{Cov}(\hat{Y}_r^*, \bar{y})$$

$$= V[\delta q\,\hat{Y}_r^* + (1-\delta q)\,\bar{y}]$$

$$= \frac{N-n}{nN}\bar{Y}^2 [C_y^2 + \delta q g C_x^2(\delta q g + 2K)] \tag{3.3}$$

which is minimum for

$$q = -\frac{k}{\delta g} = q_{opt} \quad \text{(say)} \tag{3.4}$$

From (3.3) and (3.4), the optimum (minimum) $V(T_R)$ can be obtained as

$$V_{opt}(T_R) = \frac{N-n}{nN}\bar{Y}^2 C_y^2 (1-\rho^2) \tag{3.5}$$

which is equivalent to the approximate variance of usual biased linear regression estimator $\hat{Y}_{lr} = \bar{y} + b(\bar{X} - \bar{x})$, where b is the sample regression coefficient of y on x. From (3.4) and (2.12), an optimum estimator in the class (2.12) is obtained as

$$T_{RO} = (1 + Kg^{-1})\bar{y} - K(\delta g)^{-1}\hat{Y}_r^* + (1-\delta)K(\delta g)^{-1}\hat{Y}_r^* \quad (3.6)$$

with variance given at (3.5).

Remark

(i) In (3.6), $\bar{X}$ is known but $K = \rho\frac{C_y}{C_x} = \frac{\bar{X}}{\bar{Y}}\beta$ is rarely known where β is the population regression coefficient of y on x and β may be assessed with the help of a scatter diagram of y or x for the data from a pilot survey or a study based on past data or a part of the data from the current study. Thus, the value of K may be assessed to be used to obtain the feasible estimators.

(ii) It is to be pointed out that the estimator T_R in (2.12) would be more efficient than the usual sample mean $\bar{y}$ and T_{R2} according if

$$\text{either} \quad 0 < q < \frac{2K}{\delta} \qquad \text{or} \quad -\frac{2K}{\delta} < q < 0 \quad (3.7)$$

$$\text{and either } \delta^{-1} < q < (2K-1)\delta^{-1} \qquad \text{or} \quad (2K-1)\delta^{-1} < q < \delta^{-1} \quad (3.8)$$

(iii) The variance/mean squared error of any estimator belonging to the class T_R (2.12) can be easily obtained from (3.3) to the terms of order $O(n^{-1})$.

4. *Empirical Study*

We now illustrate the results by empirical example considering the data on the number of peach trees is an orchard denoted by x and estimated production in bushels of peach by y. Summarised data (see Cochran, W.G.[1], pp.172) is given below.

$N = 256, \bar{X} = 44.55, \bar{Y} = 56.47, S_x^2 = 3898$

$S_y^2 = 6409, \; Sxy = 4434, \; \rho = 0.887$

Let the sample size be n = 30, and split the sample into p = 5 subsamples each of size m = 6.

Table (4.1) % RELATIVE EFFICIENCY OF T_R OVER $\bar{y}$

Value of q	Estimator	$\frac{V(T_R)}{F}$, $F = \frac{N-n}{nN} y^2$	% Relative efficiency over $\bar{y}$
0	$\bar{y}$ or T_{R1}	2.0098	100.00
$\frac{1}{\delta}$	T_{R2}	68.9957	2.9129
$\frac{1}{1-\delta}$	T_{R3}	10.9677	18.3247
$q_{opt} = -0.8113$	T_{RO}	0.4286	468.9221
$q < -8.1133$	T_R	> 2.0098	< 100
$q > 0$	T_R	> 2.0098	< 100
$(-8.1133 < q < 0)$	T_R	< 2.0098	> 100
—	$\bar{Y}_R^{*}$ or $\bar{Y}_{R.}^{*}$	68.9957	2.9129

In this table, the last column gives the percentage relative efficiency of the estimators belonging to class T_R (2.12) over sample mean $\bar{y}$ for different choices of q. The efficiency of the estimator T_{RO} for the choice of $q = -0.8113$ is maximum among those considered here, which shows that T_{RO} is the most efficient estimator in the class T_R (2.12). In practice, one can substitute the estimated values of the variances and covariances in order to obtain a "near optimum" value of q. In the above table for the choice of q in an optimum interval ($-8.1133 < q < 0$), the corresponding estimator in the class T_R (2.12) will be always more efficient than those considered in this case. The table also shows that the estimators T_{R2}, $\hat{Y}_r^{*}$ and $\hat{Y}_{r.}^{*}$ are equally efficient in this case.

REFERENCES

[1] Cochran, W.G., 1977. *Sampling Technique*. Wiley eastern Ltd.

[2] Goodman, L.A., 1960. On the exact variance of product. *J.A.S.A.* **55**, 313-321.

[3] Hutchison, M.C., 1971. A Monte Carlo comparison of some ratio estimators. *Biometrika*, **58**, 313-321.

[4] Murthy, M.N., 1964. Product method of estimation. *Sankhya, Sr. A*, **26**, 69-74.

[5] Quenouille, M.N., 1956. Note on bias in estimation. *Biometrika*, **43**, 353-360.

[6] Robson, D.S., 1957. Application of multivariate polykays to the theory of unbiased ratio type estimator. *J.A.S.A*, **52**, 511-522.

[7] Rao, T.J., 1981. On a class of almost unbiased ratio estimator. *Ann. Inst. Stat. Math*, **33**, 225-231.

[8] Rao, J.N.K., 1967. The precision of Mickeyes unbiased ratio estimator. *Biometrika*, **54**, 321-324.

[9] Rao, J.N.K. and Webster, J.T. 1966. On two methods of bias reduction in the estimation of ratios. *Biometrika* **53**, 517-527.

[10] Rao, P.S.R.S. and Rao, J.N.K. 1971. Small sample results for ratio estimators. *Biometrika*, **58**, 625-630.

[11] Srivastava, S.K., 1966. Product estimator. *Jr. Ind. Stat. Asso.* **4**, 29-37.

[12] Srivastava, V.K., Sukla, N.D. and Bhatnagar, S., 1981. Unbiased product estimators. *Metrika*, **28**, 191-196.

[13] Srivastava, T., 1980. A dual to ratio estimator in sample surveys. *Biometrika*, **67**,1,199-204.

[14] Sukhatme, P.V. Sukhatme, B.V., Sukhatme, S. and Asok, C., 1984. *Sampling theory of surveys with applications*, Iowa State Univ. Press, Ames, Iowa, U.S.A.

[15] Wu. C.F. and Chang, D.S., 1981. The asymptotic distribution of the product estimator. *M.R.C. Tech., Report*, Univ. of Winconsin, Madison.