Jour. Ind. Soc. Ag. Statistics
Vol. XXXIX, No. 3 (1987), pp. 280-288

CLASS OF ALMOST UNBIASED RATIO AND PRODUCT TYPE ESTIMATORS FOR FINITE POPULATION MEAN APPLYING QUENOUILLE'S METHOD

HOUSILA P. SINGH
College of Agricultural Engineering, J. N. Agricultural University, Jabalpur

(Received : October, 1986)

SUMMARY

This paper proposes a general class of almost unbiased ratio and product estimators for estimating the finite population mean $\overline{Y}$ using jackknife technique envisaged by Quenouille [2]. The expression for mean squared error (MSE) is obtained to the first degree of approximation. The optimum estimator in the class is also identified. It is also seen that the estimator reported by Quenouille [2] is a particular case of the proposed class of estimators. Discussions are made under simple random sampling without replacement (SRSWOR) throughout the investigation.

Keywords : Population mean, almost unbiased ratio and product estimators, Jackknife estimator, optimum estimator.

Introduction

Use of auxiliary information for improving the precision of the estimates in the field of sample surveys has been well recognized. Some of the important applications are in the areas of ratio and product methods of estimation. When the character y under investigation and auxiliary character x are positively correlated the ratio method is quite effective. On the other hand if the correlation between y and x is high but negative, Robson [5] and Murthy [1] propounded a complementary method to the ratio method, called product method of estimation.

Let the variates y, x take values (y_i, x_i) on the ith unit $(i = 1, 2, \ldots, N)$ in a finite population. First assume $y_i, x_i \geqslant 0$, since the survey variates are generally non-negative with only occasional exceptions like savings and profit. Section 4 covers such cases. A common situation in surveys is that the population mean $\bar{X} = N^{-1} \sum_{i=1}^{N} x_i$ of the auxiliary character x is known and we are to estimate $\bar{Y} = N^{-1} \sum_{i=1}^{N} y_i$, the population mean of the study character y. For illustration, consider a simple random sample of n ($\leqslant N$) units without replacement from the population. Let $(\bar{y}, \bar{x})$ be unbiased estimators of $(\bar{Y}, \bar{X})$ based on n observations. Then ratio estimator of $\bar{Y}$ is

$$\hat{\bar{Y}}_r = \bar{y}\,(\bar{X}/\bar{x}) \tag{1.1}$$

The bias and mean squared error (MSE) of $\hat{\bar{Y}}_r$ do not have closed form expressions. The usual approximation (to the first degree of approximation) to these are

$$B(\hat{\bar{Y}}_r) = \left(\frac{1-f}{n}\right) \bar{Y}(1-K)\, C_x^2 \tag{1.2}$$

and

$$M(\hat{\bar{Y}}_r) = \left(\frac{1-f}{n}\right) \bar{Y}^2 \left[C_y^2 + (1-2K)\, C_x^2\right] \tag{1.3}$$

where $K = \rho\,(C_y/C_x)$, $\rho = \text{Cov}\,(y, x)/(\sigma_y \cdot \sigma_x)$, $C_y = \sigma_y/\bar{Y}$,

$C_x = \sigma_x/\bar{X}$, $\sigma_v^2 = N^{-1} \sum_{i=1}^{N} (v_i - \bar{V})^2$, $v = x, y$ and $f = n/N$.

2. The Class of Almost Unbiased Ratio Estimator

It is obvious from (1.2) that the usual ratio estimator is biased. It is therefore, desirable to reduce or completely eliminate it. Quenouille [2] has suggested a technique now well known as jackknife estimation, of making ratio estimator unbiased to the desired degree of approximation, which is further generalized by Schucany, Gray and Owen [6]. We use this method for reducing the bias in ratio-type estimator and also an approach adopted by Rao [3] for defining a class of estimators. Here the simple random sample of size $n = gm$ drawn without replacement, is split at random into g subsamples, each of size m. We, now, define the ratio estimator

$$\bar{y}_r^{(j)} = \frac{\bar{y}_j'}{\bar{x}_j'}\,\bar{X} \tag{2.1}$$

$\bar{y}'_j$ and $\bar{x}'_j$ being the sample mean based on a sample of $(n - m)$ units obtained by omitting the jth group.

Motivated by Rao [3] we consider a class of estimator for $\bar{Y}$ as

$$T = \lambda \hat{\bar{y}}_r + \{1 - E(f(\lambda))\} \hat{\bar{y}}_{rJ}, \tag{2.2}$$

where $\hat{\bar{Y}}_{rJ} = \frac{1}{g} \sum_{j=1}^{g} \bar{y}'_j \bar{X}/\bar{x}'_j$, λ is a random variable and $f(\lambda)$ is a function of λ.

Now T is unbiased for $\bar{Y}$ if

$$E(T) = \bar{Y} \qquad \text{or}$$

$$\text{if} \quad E[\lambda \hat{\bar{Y}}_r - E\{f(\lambda)\} \hat{\bar{Y}}_{rJ}] = E(\bar{y} - \hat{\bar{Y}}_{rJ}) \tag{2.3}$$

for which $\lambda = \bar{x}/\bar{X}$ and $f(\lambda) = \lambda$ is a solution.

Introducing a constant 'p' in right hand side of (2.3) we write (2.3) as

$$E[\lambda \hat{\bar{Y}}_r - E\{f(\lambda)\} \hat{\bar{Y}}_{rJ}] = E[\bar{y} - \hat{\bar{Y}}_{rJ} + p \hat{\bar{Y}}_r - p \hat{\bar{Y}}_r] \tag{2.4}$$

The bias of $\hat{\bar{Y}}_{rJ}$ to the first degree of approximation is given by

$$B(\hat{\bar{Y}}_{rJ}) = \frac{(N - n + m)}{(n - m) N} \bar{Y} (1 - K) C_x^2 \tag{2.5}$$

From (1.2) and (2.5) we have

$$B(\hat{\bar{Y}}_r) = \frac{(N - n)(n - m)}{n(N - n + m)} B(\hat{\bar{Y}}_{rJ}) \tag{2.6}$$

$$\Rightarrow E(\hat{\bar{Y}}_r) = \bar{Y} + \delta B(\hat{\bar{Y}}_{rJ})$$

$$\Rightarrow E(\hat{\bar{Y}}_r) = (1 - \delta) \bar{Y} + \delta E(\hat{\bar{Y}}_{rJ}); \; \delta = \frac{(N - n)(n - m)}{n(N - n + m)} \tag{2.7}$$

Using (2.7) in (2.4) we find that

$$E[\lambda \hat{\bar{Y}}_r - E(f(\lambda)) \hat{\bar{Y}}_{rJ}] = E\left[\left\{P - (1 - P(1 - \delta)) \frac{\bar{x}}{\bar{X}}\right\} \hat{\bar{Y}}_r - (1 + P\delta) \hat{\bar{Y}}_{rJ}\right] \tag{2.8}$$

from which it follows that

$$\lambda = \left[P + \{1 - P(1 - \delta)\} \frac{\bar{x}}{\bar{X}}\right] \quad \text{and } f(\lambda) = \lambda, \tag{2.9}$$

where $E(f(\lambda)) = (1 + p\delta)$.

Substituting (2.9) in (2.2) we obtain a general class of almost unbiased ratio estimators as

$$T = [\{1 - P(1 - \delta)\}\,\bar{y} + P\bar{y}\left(\frac{\bar{X}}{\bar{x}}\right) - P\delta\,\frac{1}{g}\sum_{j=1}^{g} \bar{y}'_j\,(\bar{X}/\bar{x}'_j)] \qquad (2.10)$$

Thus we have the following theorem.

THEOREM 2.1 : *The class of estimators*

$$T = \lambda\,\hat{\bar{Y}}_r + E\{1 - f(\lambda)\}\,\hat{\bar{Y}}_{rJ}$$

would be almost unbiased if $\lambda = \left[p + \{1 - p(1 - \delta)\}\,\frac{\bar{x}}{\bar{X}}\right]$ and $f(\lambda) = \lambda$ for which $E(f(\lambda)) = (1 + p\delta)$.

When N is very large or the population is infinite the class of almost unbiased ratio estimator T boils down to :

$$T^* = [\{1 - p(1 - \delta^*)\}\,\bar{y} + p\,\bar{y}\left(\frac{\bar{X}}{\bar{x}}\right) - p\,\delta^*\,\frac{1}{g}\sum_{j=1}^{g} \bar{y}'_j\,(\bar{X}/\bar{x}'_j)] \qquad (2.11)$$

where $\delta^* = (g - 1)/g$.

Remark 2.1. We observe that $p = 0$, gives the usual unbiased estimator $\hat{\bar{Y}} = \bar{y}$ while $p = (1 - \delta)^{-1}$ yields the estimator

$$T_1 = \frac{(N - n + m)}{N}\,g\,\hat{\bar{Y}}_r - \left(\frac{(N - n)}{N}\right)(g - 1)\,\frac{1}{g}\sum_{j=1}^{g} \bar{y}'_j\,(\bar{X}/\bar{x}'_j). \qquad (2.12)$$

[see Sukhatmes and Ashok ([8] equation (67), pp. 207)]
When N is very large or the population is infinite the estimator T_1 turns out to be

$$T_1^* = g\,\hat{\bar{Y}}_r - \frac{(g - 1)}{g}\sum_{j=1}^{g} \bar{y}'_j\,(\bar{X}/\bar{x}'_j) \qquad (2.13)$$

which is due to Quenouille [2].
For $p = \delta^{-1}$, T reduces to another estimator

$$T_2 = \delta^{-1}\,\hat{\bar{Y}}_r - \frac{1}{g}\sum_{j=1}^{g} \bar{y}'_j\,(\bar{X}/\bar{x}'_j) + (2 - \delta^{-1})\,\bar{y}$$

Many other almost unbiased ratio estimators can be obtained by putting various suitable values of p in (2.10).

3. Optimum Estimator in the Class (2.10)

We have from (2.10) that

$$V(T) = V(\bar{y}) + p^2\ V(t_1) - 2p\ \text{Cov}\ (\bar{y}, t_1) \tag{3.1}$$

where $t_1 = [(1-\delta)\bar{y} - t_2]$ and $t_2 = \hat{\bar{Y}}_r - \delta\ \hat{\bar{Y}}_{rJ}$.

To terms of order $0(n^{-1})$, it is easy to verify that

$$\left.\begin{aligned} &V(\hat{\bar{Y}}_r) = V(\hat{\bar{Y}}_{rJ}) = \text{Cov}\ (\hat{\bar{Y}}_r, \hat{\bar{Y}}_{rJ}) = M\ (\hat{\bar{Y}}_r) \\ &\text{Cov}\ (\bar{y}, \hat{\bar{Y}}_r) = \text{Cov}\ (\bar{y}, \hat{\bar{Y}}_{rJ}) = \left(\frac{1-f}{n}\right) \bar{Y}^2\ (C_y^2 - K\ C_x^2) \\ &V(\bar{y}) = \left(\frac{1-f}{n}\right) \bar{Y}^2\ C_y^2 \end{aligned}\right\} \tag{3.2}$$

Substituting (3.2) in (3.1) we get

$$V(T) = \left(\frac{1-f}{n}\right) \bar{Y}^2\ [C_y^2 + p\ (1-\delta)\ \{p\ (1-\delta) - 2\ K\}\ C_x^2\ ] \tag{3.3}$$

which is minimized for

$$p = K\ (1-\delta)^{-1} = P_{\text{opt}}\ \ \text{(say)} \tag{3.4}$$

Substituting (3.4) in (3.3) we obtain the minimum variance of T as

$$\text{Min} \cdot V(T) = \left(\frac{1-f}{n}\right) \bar{Y}^2\ C_y^2\ (1-\rho^2) \tag{3.5}$$

which is equivalent to the approximate variance of usual biased regression estimator $\hat{\bar{Y}}_{lr} = \bar{y} + b\ (\bar{X} - \bar{x})$, where b is the sample regression coefficient of y on x.

Substituting the value $P_{\text{opt}} = K\ (1-\delta)^{-1}$

for p in (2.10) we obtain optimum estimator in the class (2.10) defined by

$$T_0 = \left[(1-K)\ \bar{y} + K\ (1-\delta)^{-1}\ \bar{y}(\bar{X}/\bar{x}) - K\delta\ (1-\delta)^{-1}\ \frac{1}{g} \sum_{j=1}^{g} \bar{y}_j' \left(\frac{\bar{X}}{\bar{x}_j'}\right)\right] \tag{3.6}$$

with the variance as given in (3.5).

4. Negatively Correlated Variates

When auxiliary information on a variate x is available which is negatively correlated with y, it was shown that this information could be profitably used to construct a 'product estimator' for population mean $\overline{Y}$ which is in certain cases far superior to the conventional estimator which does not use the information on x. The usual product estimator

$$\hat{\overline{Y}}_p = \bar{y}\ (\bar{x}/\overline{X}) \tag{4.1}$$

is in general biased. Murthy [1] and Rao [4] have reported some unbiased product estimators, based on interpenetrating subsample design/replicated samples. For the same purpose Shukla [7] has applied the method developed by Quenouille [2]. We use this latter method for getting rid of bias in the product estimator $\hat{\overline{Y}}_p$. The jackknife product estimator is defined by

$$\hat{\overline{Y}}_{PJ} = \frac{1}{g} \sum_{j=1}^{g} \bar{y}'_j\ (\bar{x}'_j\ /\ \overline{X}), \qquad j = 1, 2, \ldots, g \tag{4.2}$$

where $\bar{y}'_j = (n\,\bar{y} - m\,\bar{y}_j)/(n - m)$ and $\bar{x}'_j = (n\,\bar{x} - m\bar{x}_j)/(n - m)$; $(\bar{y}, \bar{x})$ and $(\bar{y}_j, \bar{x}_j)$ are the sample means based on entire sample of size n and subsample of size $m = n/g$.

The exact biases of $\hat{\overline{Y}}_p$ and $\hat{\overline{Y}}_{pJ}$ are given by

$$B(\hat{\overline{Y}}_p = \left(\frac{1-f}{n}\right) \overline{Y} K C_x^2 \tag{4.3}$$

and

$$B(\hat{\overline{Y}}_{pJ}) = \frac{(N - n + m)}{(n - m)\ N} \overline{Y} K C_x^2 \tag{4.4}$$

It follows from (4.3) and (4.4) that

$$B(\hat{\overline{Y}}_p) = \frac{(N-n)\ (n-m)}{n\ (N - n + m)}\ B\ (\hat{\overline{Y}}_{pJ})$$

$$\Rightarrow E\ (\hat{\overline{Y}}_p) = (1 - \delta)\ \overline{Y} + \delta E(\hat{\overline{Y}}_{pJ});\ \delta = \frac{(N-n)\ (n-m)}{n\ (N - n + m)}, \tag{4.5}$$

Motivated by Rao [4] we define a class of product estimators for $\overline{Y}$ as

$$T = \mu\ \hat{\overline{Y}}_p + \{1 - E(f(\mu))\}\ \hat{\overline{Y}}_{pJ}, \tag{4.6}$$

where μ is stochastic and $f(\mu)$ is a function of μ. Proceeding as in section 2, it is easy to check that

$$E[\mu \hat{\bar{Y}}_p - E(f(\mu)) \hat{\bar{Y}}_{pJ}] = E\left[\left\{q + (1 - (1 - \delta) q)\left(\frac{\bar{X}}{\bar{x}}\right)\right\} \hat{\bar{Y}}_p - (1 + \delta q) \hat{\bar{Y}}_{pJ}\right] \tag{4.7}$$

q being a non-stochastic.
It follows from (4.7) that

$$\mu = \left[q + \{1 - (1 - \delta) q\}\left(\frac{\bar{X}}{\bar{x}}\right)\right] \text{ and } f(\mu) = \mu\left(\frac{\bar{x}}{\bar{X}}\right) \tag{4.8}$$

for which $E(f(\mu)) = (1 + \delta q)$.
Substituting (4.8) in (4.6) we find the general form of unbiased product estimator as

$$T_p = [\{1 - (1 - \delta) q\} \bar{y} + q \bar{y}\left(\frac{\bar{x}}{\bar{X}}\right) - \frac{q\delta}{g} \sum_{j=1}^{g} \bar{y}'_j \ (\bar{x}'_j / \bar{X})] \tag{4.9}$$

Thus we have the following theorem :

THEOREM 4.1 : *The class of product estimators* $T = \mu \hat{\bar{Y}}_p + \{1 - E(f(\mu))\} \hat{\bar{Y}}_{pJ}$; ($\mu$ *is stochastic*), *is unbiased for* $\mu = [q + \{1 - (1 - \delta) q\} (\bar{X}/\bar{x})]$ *and* $f(\mu) = (\bar{x}/\bar{X}) \mu$, *where* $E(f(\mu)) = (1 + \delta q)$, *q being a non-stochastic.*

If the population size N is very large or population is infinite then T_p in (4.9) reduces to

$$T_p^* = [\{1 - (1 - \delta^*) q\} \bar{y} + q \bar{y}\left(\frac{\bar{x}}{\bar{X}}\right) - \frac{\delta^* q}{g} \sum_{j=1}^{g} \bar{y}'_j (\bar{x}'_j / \bar{X})] \tag{4.10}$$

where $\delta^* = (g - 1) g^{-1}$.
The class of unbiased product-type estimators T_p in (4.9) reduces to the following set of unbiased estimators of $\bar{Y}$:

(i) $T_{p1} = \bar{y}$ for $q = 0$

(ii) $T_{p2} = \dfrac{(N - n + m)}{N} g \hat{\bar{Y}}_p - \dfrac{(N - n)}{N} \dfrac{(g - 1)}{g} \sum_{j=1}^{g} \bar{y}'_j (\bar{x}'_j / \bar{X})$,

for $q = (1 - \delta)^{-1}$. This is Shukla [7] type estimator.

(iii) $T_{p3} = \delta^{-1}\ \hat{\bar{Y}}_p - \dfrac{1}{g}\sum_{j=1}^{g} \bar{y}'_j\,(\bar{x}'_j\,/\,\bar{X}) + (2-\delta^{-1})\,\bar{y},$ for $q = \delta^{-1}$.

Several other unbiased product-type estimator can be had by substituting various suitable values of q in (4.9).

5. Optimum Estimator in the Class (4.9)

From (4.9) we have

$$V(T_p) = V(\bar{y}) + q^2\, V(T^{**}) - 2q\,\text{Cov}\,(\bar{y},\, T^{**}) \tag{5.1}$$

where $T^{**} = (1-\delta)\,\bar{y} - T_p^{**}$ and $T_p^{**} = \hat{\bar{Y}}_p - \delta\,\hat{\bar{Y}}_{pJ}$.

To terms of order $0(n^{-1})$, it is easy to verify that

$$V(\hat{\bar{Y}}_p) = V(\hat{\bar{Y}}_{pJ}) = \text{Cov}\,(\hat{\bar{Y}}_p,\, \hat{\bar{Y}}_{pJ}) = M(\hat{\bar{Y}}_p) = \left(\frac{1-f}{n}\right)\bar{Y}^2\,[C_y^2 + (1+2K)\,C_x^2]$$

$$\text{Cov}\,(\bar{y},\, \hat{\bar{Y}}_p) = \text{Cov}\,(\bar{y},\, \hat{\bar{Y}}_{pJ}) = \left(\frac{1-f}{n}\right)\bar{Y}^2\,(C_y^2 + KC_x^2), \tag{5.2}$$

Using (5.2) in (5.1) we obtain the variance of T_p to the desired degree of approximation (i.e. to terms of order $0(n^{-1})$) as

$$V(T_p) = \left(\frac{1-f}{n}\right)\bar{Y}^2\,[C_y^2 + q\,(1-\delta)\,\{q\,(1-\delta) + 2K\}\,C_x^2] \tag{5.3}$$

which is minimized for

$$q = -K(1-\delta)^{-1} = q_{\text{opt}} \quad \text{(say)} \tag{5.4}$$

Hence the minimum MSE/variance of T_p is given by

$$\text{Min}\cdot V(T_p) = \left(\frac{1-f}{n}\right)\bar{Y}^2\,C_y^2\,(1-\rho^2) \tag{5.5}$$

which is equivalent to the approximate variance of usual biased regression estimator $\hat{\bar{Y}}_{1r}$.

Substituting the value of q_{opt} in (4.9) we obtain the optimum estimator

$$T_{PU} = [(1+K)\,\bar{y} - K(1-\delta)^{-1}\,\bar{y}\,(\bar{x}/\bar{X}) + K(1-\delta)\,\delta^{-1}\cdot\frac{1}{g}\sum_{j=1}^{g}\bar{y}'_j\,(\bar{x}'_j\,/\,\bar{X})] \tag{5.6}$$

with minimum variance in the class.

ACKNOWLEDGEMENTS

With pleasure the author acknowledges the helpful comments furnished by Professor T. J. Rao, ISI, Calcutta on an earlier draft of this manuscript. Author is also grateful to the learned referee for his valuable comments/suggestions regarding the improvement of the paper. Grateful thanks are also due to Professor Lal Chand, Head Department of Mathematics and Statistics, JNKVV, Jabalpur, for his encouragements and fruitful discussions during the course of this investigation.

REFERENCES

[1] Murthy, M. N. (1964) : Product method of estimation, *Sankhya*, A, **26** : 69-74.

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

[3] Rao, T. J. (1981) : On a class of almost unbiased ratio estimators, *Ann. Inst. Stat. Math.*, **33** : 225-231.

[4] Rao, T. J. (1983) : A new class of unbiased product estimators. Tech. Rep. No. 15/83, Stat. Math. Div., Indian Statistical Institute, Calcutta, India.

[5] Robson, D. S. (1957) : Applications of Multivariate polykays to the theory of unbiased ratio type estimation, *Jour. Amer. Stat. Assoc.*, **52** : 511-522.

[6] Schucany, W. R., Gray, H. L. and Owen, D. W. (1971) : On bias reduction in estimation, *Jour. Amer. Stat. Assoc.*, **66** : 524-33.

[7] Shukla, N. D. (1976) : Almost unbiased product-type estimators, *Metrika*, **23**: 127-133.

[8] Sukhatme, P. V., Sukhatme, B. V., Sukhatme, S. and Asock, C. (1984) : *Sampling Theory of Surveys with Applications*. Third edition, Iowa State University Press, Ames, Iowa (USA) and Indian Society of Agricultural Statistics, New Delhi, India.