

ESTIMATION OF POPULATION RATIO ON TWO OCCASIONS

By

AJIT KUMAR DAS

Indian Statistical Institute, Calcutta

(Received : January, 1978)

SUMMARY

A sampling strategy based on a scheme of partial replacement has been considered for estimating the population ratio on second occasion. The proposed strategy has been compared with those of Tripathi and Sinha [5], Rao [3], Rao and Pereira [4] and with other sampling strategies. The values of optimum matched proportion has been tabulated and an empirical study is made to study the relative performance of the proposed strategy compared to some other strategies.

1. INTRODUCTION

In practice, we need the estimates for ratio $R_h = \bar{Y}_h / \bar{X}_h = Y_h / X_h$ ($h=1, 2, \dots$) of population means (or totals) of two characters y and x , on successive occasions.

Let D_0 denote the sampling procedure of drawing independent samples on both the occasions, using simple random sampling without replacement (SRSWOR), and D_1 of retaining the entire sample on the second occasion drawn at the first occasion. A simplest type of sampling strategy for R_h may be given by (D_0, \hat{r}_h^*) , where

$$\hat{r}_h^* = r_h = \bar{y}_{hn} / \bar{x}_{hn} \quad \dots(1.1)$$

\bar{y}_{hn} and \bar{x}_{hn} are the sample means (sample of size n) of y and x respectively on h -th occasion. Rao [3] and Rao and Pereira [4] considered the sampling strategies (D_1, \hat{r}_2) and (D_1, r_2) respectively where

$$\hat{r}_2 = (r_2/r_1) R_1 \text{ and } \tilde{r}_2 = r_2 r_1 / R_1 \quad \dots(1.2)$$

Tripathi and Sinha [5] considered a sampling strategy (D_λ, \hat{R}_2^*) for R_2 which does not require the knowledge on R_1 where the sampling scheme D_λ is similar to that given by Kulldorff [2] and

Ghangurde and Rao [1] and \hat{R}_2^* is the estimator for R_2 defined as

$$\hat{R}_2^* = W^* \hat{R}_{2m}^* + (1 - W^*) \hat{R}_{2u} \quad \dots(1.3)$$

where $\hat{R}_{2m}^* = \frac{\bar{y}_{2m} - b(\bar{y}_{1m} - \bar{y}_{1n})}{\bar{x}_{2m} - b^*(\bar{x}_{1m} - \bar{x}_{1n})}$ and $\hat{R}_{2u} = \frac{\bar{y}_{2u}}{\bar{x}_{2u}}$

are the 'regression-type ratio estimator' for R_2 based on the matched sample s_1 of m units from the sample s of n units drawn at the first occasion and the usual ratio estimator for R_2 based on the replaced (unmatched) sample s_2 of u units from the $(N - n)$ units in the population not included in s respectively.

The estimator \hat{R}_{2m}^* , for R_2 , based on matched portion, suffers from a weakness that to estimate \bar{Y}_2 in R_2 , it utilizes information only on y_1 (in addition to y_2), and to estimate \bar{X}_2 only on x_1 (in addition to x_2); while the information on both the characters y and x on previous occasion may in fact be used to estimate both of \bar{Y}_2 and \bar{X}_2 .

In this paper we use information y_1 and x_1 both to estimate \bar{Y}_2 , \bar{X}_2 in setting up estimator for R_2 based on matched portion and study the properties of proposed sampling strategy. We also compare it with other sampling strategies.

2. SAMPLING STRATEGY UNDER CONSIDERATION

Let $C_h(y)$ and $C_h(x)$ denote the coefficient of variation of y_h and x_h respectively and $\rho_{hh'}(y, x)$ denote the correlation coefficient between y_h and $x_{h'}$ ($h, h' = 1, 2$). We consider the sampling strategy for R_2 as $(D\lambda, \hat{R}_2)$ where

$$\hat{R}_2 = W_1 \hat{R}_{2m} + W_2 \hat{R}_{2u} \quad \dots(2.1)$$

$$\hat{R}_{2m} = \frac{\bar{y}_{2m} - b_{y_2 y_1, x_1} (\bar{y}_{1m} - \bar{y}_{1n}) - b_{y_2 x_1, y_1} (\bar{x}_{1m} - \bar{x}_{1n})}{\bar{x}_{2m} - b_{x_2 y_1, x_1} (\bar{y}_{1m} - \bar{y}_{1n}) - b_{x_2 x_1, y_1} (\bar{x}_{1m} - \bar{x}_{1n})}$$

and \hat{R}_{2u} , defined in (1.3) being the estimators for R_2 based on the matched portion and unmatched portion respectively, W_1 and W_2 the constants such that $W_1 + W_2 = 1$ and b 's are the partial regression coefficients based on the matched sample.

3. MEAN SQUARE ERROR (MSE) OF THE SAMPLING STRATEGY

Assuming $|(\bar{x}_{2u} - \bar{X}_2)/\bar{X}_2| < 1$, the MSE of \hat{R}_{2u} , to the terms $O(u^{-1})$, is given by

$$M(\hat{R}_{2u}) = R_2^2 \left(\frac{1}{u} - \frac{1}{N} \right) J_2 \quad \dots(3.1)$$

where $J_2 = C_2^2(y) + C_2^2(x) - 2\rho_{22}(y, x) C_2(y) C_2(x)$.

Let $|(Q^* - \bar{X}_2)/\bar{X}_2| < 1$ where Q^* is the denominator of \hat{R}_m . The MSE of \hat{R}_{2m} , to the terms $O(m^{-1})$, is given by

$$M(\hat{R}_{2m}) = R_2^2 \left[\left(\frac{1}{m} - \frac{1}{n} \right) J_{(1,2)} + \left(\frac{1}{n} - \frac{1}{N} \right) J_2 \right] \quad \dots(3.2)$$

where

$$J_{(1,2)} = q_{11} C_2^2(y) + 2q_{12} C_2(y) C_2(x) + q_{22} C_2^2(x)$$

with

$$q_{11} = 1 - q [\rho_{21}^2(y, x) + \rho_{12}^2(y, y) - 2\rho_{11}(y, x) \rho_{21}(y, x) \rho_{12}(y, y)]$$

$$q_{22} = 1 - q [\rho_{12}^2(y, x) + \rho_{12}^2(x, x) - 2\rho_{11}(y, x) \rho_{12}(y, x) \rho_{12}(x, x)]$$

$$q_{12} = -[\rho_{22}(y, x) + q \{ \rho_{11}(y, x) \rho_{21}(y, x) \rho_{12}(y, x) + \rho_{11}(y, x)$$

$$\rho_{12}(y, y) \rho_{12}(x, x) - \rho_{21}(y, x) \rho_{12}(x, x) - \rho_{12}(y, x) \rho_{12}(y, y) \}]$$

and $q = [1 - \rho_{11}^2(y, x)]^{-1}$.

Using the optimum weights in (2.1) the resulting MSE of \hat{R}_2 , with large sample approximation, is given by

$$M_0(\hat{R}_2) = \frac{1}{n} R_2^2 J_2 \frac{\lambda J_2 + (1-\lambda)J_{(1,2)} - \mu f[\lambda J_2 + (1-\lambda)J_{(1,2)}] - f\lambda J_2 + f^2\lambda\mu J}{\lambda J_2(1+\mu) + \mu(1-\lambda)J_{(1,2)} - 2\lambda\mu f J_2} \quad \dots(3.3)$$

where $\lambda = m/n$, $\mu = u/n$ and $f = n/N$.

If $u = n - m$, and if the sampling fraction f is ignored, then

$$M_0(\hat{R}_2) = \frac{1}{n} R_2^2 J_2 \frac{\lambda J_2 + (1-\lambda)J_{(1,2)}}{\lambda(2-\lambda)J_2 + (1-\lambda)^2 J_{(1,2)}} \quad \dots(3.4)$$

where λ will now be called as matched proportion of the sample at second occasion.

4. OPTIMUM MATCHED PROPORTION AND OPTIMUM MSE

It may be shown that in general $J_2 \geq J_{(1,2)} \geq 0$. Further, from (3, 4),

$$\text{if } J_2 = J_{(1,2)} \text{ then } M_0(\hat{R}_2) = (1/n) R_2^2 J_2 \quad \dots(4.1)$$

We see that in such a case λ does not play any role in reducing the MSE.

In case $J_2 > J_{(1,2)} > 0$, the optimum value of λ ($0 < \lambda < 1$), which minimizes $M_0(\hat{R}_2)$, defined in (3, 4), is given by

$$\lambda_0 = \sqrt{J_{(1,2)}} / [\sqrt{J_{(1,2)}} + \sqrt{J_2}] \quad \dots(4.2)$$

and then the resulting minimum MSE is given by

$$M_{\min}(\hat{R}_2) = (1/2n) R_2^2 (J_2 + \sqrt{J_2 J_{(1,2)}}) \quad \dots(4.3)$$

If $C_2(y) = C_2(x)$, the optimum λ would depend only upon various correlation coefficients. In such a case one may tabulate the values of λ_0 for various given values of correlation coefficients alone. To have a picture of λ_0 and to make certain other specific studies we shall now assume that

$$C_2(y) = C_2(x) = C_2; \quad \rho_{12}(y, y) = \rho_{12}(x, x) = \rho_0 \quad \dots(4.4)$$

$$\rho_{12}(y, x) = \rho_{21}(y, x) = \rho_{11}(y, x) = \rho_{22}(y, x) = \rho_1$$

Under (4.4), we have

$$J_2 = 2C_2^2(1 - \rho_1); \quad J_{(1,2)} = 2C_2^2(1 - \rho_0) [(1 + \rho_0) - 2\rho_1] / (1 - \rho_1) \quad \dots(4.5)$$

$$\lambda_0 = \sqrt{(1 - \rho_0^2) - 2\rho_1(1 - \rho_0)} / [\sqrt{(1 - \rho_0^2) - 2\rho_1(1 - \rho_0)} + (1 - \rho_1)], \quad (\rho_0 = \rho_1) \quad \dots(4.6)$$

$$M_{min}(\hat{R}_2) = (1/n) R_2^2 C_2^2 [(1 - \rho_1) + \sqrt{(1 - \rho_0^2) - 2\rho_1(1 - \rho_0)}]. \quad \dots(4.7)$$

It may be noted that if $\rho_1 = \rho_0 = \rho$ (say), then, $J_2 = J_{(1,2)} = 2C_2^2(1 - \rho)$ in which case, from (4.1), we have that

$$M_0(\hat{R}_2) = 2(1/n) R_2^2 C_2^2 (1 - \rho). \quad \dots(4.8)$$

The values of λ_0 in (4.6) for various values of ρ_0 and ρ_1 have been given in Table 4.1 (given at the end of the paper)

5. REMARKS ABOUT TABLE 4.1

(i) From (4.8), we note that $M_0(\hat{R}_2)$ does not depend on λ at all, in case $\rho_0 = \rho_1$. The mark (x) in the Table, at all the entries on the diagonal of (ρ_0, ρ_1) , shows this situation.

(ii) From (4.5), we note that if $\rho_1 = 1$, then $J_{(1,2)}$ is not defined at all and $J_2 = 0$. This situation is, obviously, to be excluded from the Table,

(iii) From (4.5), we note that $J_{(1,2)} = 0$ iff $\rho_0 = 1$ or if $\rho_0 = 2\rho_1 - 1$, the situation which we have excluded in obtaining optimum λ . The mark (xx) in the Table depicts such situations.

(iv) From (4.5), we note that if $\rho_0 < 2\rho_1 - 1$ then $J_{(1,2)} < 0$ which is not possible. This in fact indicates that such choices of ρ_0 and ρ_1 would be inconsistent. The asterik (*) in the body of the Table reflects such situations.

(v) Noting that $\sqrt{J_2} \geq \sqrt{J_{(1,2)}}$, from (4.2) we find that $\lambda_0 < 1/2$. This fact is reflected from the body of the Table also for all possible choices of ρ_0 and ρ_1 . We also note, from the Table, that $\lambda_0 > 0.23$ for all possible (ρ_0, ρ_1) .

6. COMPARISON WITH OTHER SAMPLING STRATEGIES

In this section we shall compare $(D\lambda, \hat{R}_2)$, proposed by us, with $(D\lambda, \hat{R}_2^*)$, (D_0, r_2^*) , (D_1, r_2) , and (D_1, \bar{r}_2) which are defined in section I, and with some other sampling strategies for R_2 .

(i) It may be shown, to our order of approximation, that

$$M(\hat{R}_{2m}^*) \geq M(\hat{R}_{2m}) \tag{5.1}$$

which will further imply that

$$M_0(\hat{R}_2) \leq M_0(\hat{R}_2^*) \tag{5.2}$$

(ii) We note that if $u=n-m$ and if f is ignored and if $\lambda=m/n=0$ (no matching) or if $\lambda=1$ (complete matching), then

$$M_0(\hat{R}_2)_{\lambda=0} = M_0(\hat{R}_2)_{\lambda=1} = (1/n) R_2^2 J_2 = M(r_2) = M(\hat{r}_2^*) \tag{5.3}$$

where r_2 and \hat{r}_2^* are defined in (1.2) and (1.1) respectively. We note that $J_2 \geq J(1,2) \geq 0$. From (3.4) and (5.3) we find that in case $J_2 > J(1,2) > 0$, the sampling strategy $(D\lambda, \hat{R}_2)$ of partial replacement ($0 < \lambda < 1$) will be better than those of complete matching ($\lambda=1$) and complete replacement ($\lambda=0$).

(iii) For large sample sizes we shall have

$$M(\hat{r}_2) = (1/n) R_2^2 (1-f) [J_2 + J_1 - 2L] \tag{5.4}$$

and

$$M(\bar{r}_2) = (1/n) R_2^2 (1-f) [J_2 + J_1 + 2L] \tag{5.5}$$

where

$$J_h = C_h^2(y) + C_h^2(x) - 2\rho_{hh}(y, x) C_h(y) C_h(x) \quad (h = 1, 2)$$

$$L = C_{12}(y, y) + C_{12}(x, x) - C_{12}(y, x) - C_{21}(y, x)$$

$$C_{hh'}(y, x) = \rho_{hh'}(y, x) C_h(y) C_{h'}(x), \quad (h, h' = 1, 2) \text{ etc.}$$

and \hat{r}_2 and \bar{r}_2 are defined in (1.2).

$$\text{Let } C_2(y) = C_1(y), C_2(x) = C_1(x). \tag{5.6}$$

In case f is ignored, then under the assumptions (4.4) and (5.6), it can be shown from (5.4), (5.5) and 3.4) that

$$M_0(\hat{R}_2) \leq M(\hat{r}_2)$$

$$\text{if } \lambda (1 - \rho_1)^2 [2 (1 - \rho_0) (2 - \lambda) - (1 - \rho_1)]$$

$$+ (1 - \rho_0) (1 + \rho_0 - 2\rho_1) (1 - \lambda) [2 (1 - \rho_0) (1 - \lambda) - (1 - \rho_1)] \geq 0 \tag{5.7}$$

which is always satisfied if $\lambda < (1 + \rho_1 - 2\rho_0) / 2 (1 - \rho_0)$ and

$$M(\bar{r}_2) > M_0(\hat{R}_2)_{\lambda=0} > M_0(\hat{R}_2)_\lambda \text{ iff } \rho_0 > (3\rho_1 - 1) / 2.$$

It may be noted that if $p_0 < p_1$ and $\lambda \leq 1/2$ then (5.7) is always satisfied.

(iv) Now we shall compare (D_1, R_1) with (D_2, R_2) and $(D_1, R_1^{(1)})$ and $(D_2, R_2^{(2)})$, for R_1, R_2 , where the estimators $R_1^{(1)}$ and $R_2^{(2)}$ are defined by

$$R_1^{(1)} = wR_1^{(1)} + (1-w)R_{2u}, \quad R_2^{(2)} = w^*R_2^{(2)} + (1-w^*)R_{2u} \quad \dots(5.8)$$

respectively where

$$R_1^{(1)} = \frac{y_{2m}(y_{1m}/y_{1m})}{y_{2m}(y_{1m}/y_{1m})}, \quad R_2^{(2)} = \frac{x_{2m}(x_{1m}/x_{1m})}{y_{2m}(y_{1m}/y_{1m})}$$

The large sample approximation to the MSEs would be given by

$$M(R_1^{(1)}) = R_2^2 \left[\left(\frac{1}{1} \frac{m}{1} - \frac{1}{1} \frac{N}{1} \right) J_2 + \left(\frac{1}{1} \frac{m}{1} - \frac{1}{1} \frac{N}{1} \right) J_1 \right] - 2R_2^2 \left(\frac{1}{1} \frac{m}{1} - \frac{1}{1} \frac{N}{1} \right) T \quad \dots(5.9)$$

$$M(R_2^{(2)}) = R_2^2 \left[\left(\frac{1}{1} \frac{m}{1} - \frac{1}{1} \frac{N}{1} \right) J_2 + \left(\frac{1}{1} \frac{m}{1} - \frac{1}{1} \frac{N}{1} \right) J_1 \right] + 2R_2^2 \left(\frac{1}{1} \frac{m}{1} - \frac{1}{1} \frac{N}{1} \right) T \quad \dots(5.10)$$

Under the assumptions (4.4) and (5.6), it can be shown that $M(R_2^{(2)}) \leq M(R_1^{(1)})$ which will, from (5.1) and (5.2), further imply that

$$M_0(R_2) \leq M_0(R_1) \leq M_0(R_2^{(2)}) \quad \dots(5.11)$$

Again, under the assumptions (4.4) and (5.6), it is found that

$$M(R_2^{(2)}) \leq M(R_2^{(1)}) \quad \dots(5.12)$$

$$M(R_2^{(1)}) \leq M(R_2^{(2)}) \text{ iff } p_1 \leq (1+p_0)/(3-p_0) \quad \dots(5.13)$$

$$M(R_2^{(1)}) \leq M(R_2^{(2)}) \text{ in case } p_1 \leq p_0 \quad \dots(5.14)$$

which will further imply that

$$M_0(R_2) \leq M_0(R_2^{(1)})$$

$$M_0(R_2) \leq M_0(R_1) \leq M_0(R_2^{(1)}) \text{ in case } p_1 \leq p_0$$

$$M_0(R_2) \leq M_0(R_2^{(2)}) \leq M_0(R_1) \leq M_0(R_2^{(1)})$$

$$\text{in case } p_0 \leq p_1 \leq \frac{(1+p_0)}{(3-p_0)}$$

7. EMPIRICAL STUDY

The data under consideration was taken from census 1951 and census 1961, West Bengal, *District Census Hand Book*, Midnapore. The population consists of 368 villages or town/ward under Sabani Police Station (in fact only those villages or towns/wards have been

considered which are common to both, census 1951 and census 1961, lists and which are shown as inhabited). The characters x_1 and x_2 are numbers of houses for 1951 and 1961 respectively and the characters y_1 and y_2 are number of literate persons for 1951 and 1961 respectively. For this population we obtained

$$\begin{aligned}\bar{X}_1 &= 38.3696 & C_1(x) &= 1.3916 & \rho_{12}(x, x) &= 0.7990 \\ \bar{X}_2 &= 50.4321 & C_2(x) &= 1.0585 & \rho_{12}(y, y) &= 0.5392 \\ \bar{Y}_1 &= 31.4321 & C_1(y) &= 2.2129 & \rho_{11}(y, x) &= 0.9187 \\ \bar{Y}_2 &= 42.5761 & C_2(y) &= 1.5048 & \rho_{12}(y, x) &= 0.7028 \\ \rho_{21}(y, x) &= 0.5471, & \rho_{22}(y, x) &= 0.7952.\end{aligned}$$

Using the above values of the parameters we obtained

$$J_2 = 0.8517, \quad J_1 = 1.1752, \quad J_{(1, 2)} = 0.8013$$

$$M(\hat{R}_{2m}^*) = R_2^2 \left[\frac{0.8167}{m} + \frac{0.0350}{n} - \frac{0.8517}{N} \right]$$

Using the relations (3.2), (4.2) and (4.3) we get

$$M(\hat{R}_{2m}) = R_2^2 \left[\frac{0.8013}{m} + \frac{0.0504}{n} - \frac{0.8517}{N} \right]$$

$$\lambda_0 = 0.4924, \quad M_{\min}(\hat{R}_2) = (1/n) R_2^2 (0.8389).$$

Using (5.9), (5.10), (5.4) and (5.5) we obtain

$$M(\hat{R}_{2m}^{(1)}) = R_2^2 \left[\frac{1.6659}{m} - \frac{0.8142}{n} - \frac{0.8517}{N} \right]$$

$$M(\hat{R}_{2m}^{(2)}) = R_2^2 \left[\frac{2.3879}{m} - \frac{1.5362}{n} - \frac{0.8517}{N} \right]$$

$$M(\hat{r}_2) = R_2^2 \left(\frac{1}{n} - \frac{1}{N} \right) (1.6659),$$

$$M(\bar{r}_2) = R_2^2 \left(\frac{1}{n} - \frac{1}{N} \right) (2.3879).$$

From the above expressions we find that

$$M(\hat{R}_{2m}) < M(\hat{R}_{2m}^*) < M(\hat{R}_{2m}^{(1)}) < M(\hat{R}_{2m}^{(2)})$$

which will further imply that

$$M_0(\hat{R}_2) < M_0(\hat{R}_2^*) < M_0(\hat{R}_2^{(1)}) < M_0(\hat{R}_2^{(2)}).$$

Further ignoring the terms involving $1/N$ and using (5.3) and (4.3), we find that

$$M_{\min}(\hat{R}_2) \leq M_0(\hat{R}_2)_{0 < \lambda < 1} < M_0(\hat{R}_2)_{\lambda=0} = M_0(\hat{R}_2)_{\lambda=1}$$

$$= M(\bar{r}_2) = M(\hat{r}_2^*) = \frac{1}{n} R_2^2 (0.8517) < M(\hat{r}_2) < M(\bar{r}_2).$$

ACKNOWLEDGEMENT

The author wants to express his sincere thanks to Dr. T.P. Tripathi for his guidance in preparing this paper and is also thankful to the referees for their valuable comments in improving the paper to its present form.

REFERENCES

- [1] Ghangurde. P.D. : Some results on sampling over two occasions.
Rao, J.N.K. (1969) *Sankhyā. A*, **31**, 463-472.
- [2] Kulldorff, G. (1963) : Some problems of optimum allocation for
sampling on two occasions. *Rev. Inter. Stat. Insti.* **31**, 24-57.
- [3] Rao, J.N.K. (1957) : Double ratio estimate in forest surveys. *Jour. Indian Soc. Agri. Stat.*, **9**, 191-204,
- [4] Rao, J.N.K. and : On double ratio estimators. *Sankhyā A*, **30**,
Pereira, N.P. (1968) 83-90.
- [5] Tripathi, T.P. and : Estimation of ratio on successive occasions.*
Sinha, S.K.P. (1976) : Paper presented at the 'Symposium on Recent
Development in Survey Methodology' held at
Indian Statistical Institute, Calcutta, in
March, 1976.