COMBINATION OF RATIO AND PPS ESTIMATORS

By
S.K. Agarwal* and Pranesh Kumar,
I. A. S. R. I., NewDelhi-110 012.
(Received : January, 1978)

1. Introduction

It is well known that suitable use of auxiliary information in probability sampling results in considerable reduction in variance of the estimator of population mean/total. The auxiliary information may be used either at the stage of designing or at the stage of estimation, depending upon the form in which such information is available for increasing the precision of the estimate of the population characteristics under study. If there are more than one auxiliary characters, the problem remains as to how the entire information can be utilized in a better way. Multi-variate ratio and regression methods of estimation, two way stratification etc. provide some alternative solution to the problem. Singh [3] [4] also suggested a method of using the two auxiliary variates by considering a ratio cum product estimator for estimating the population total of the study variable. However, the estimators suggested by him would be more efficient than the usual ratio estimator only under certain conditions.

In present paper two auxiliary characters have been used in different ways, viz. one for the purpose of selection of the sample and the other for the purpose of estimation and then suitably combining the pps and ratio estimators so obtained, to estimate the population mean, so that the mean squared error of the estimator is minimum. Following this approach an estimator, combining ratio and pps estimators of the population mean has been proposed and it has been proved that the proposed ertimator would always be more precise then that of either pps estimator or ratio estimator under pps sampling scheme.

2. Notations

Let y_i be the value of the character under study for the i-th unit of the population under consideration ; x_{1i} and x_{2i} be the values

*Lecturer, Department of Mathematics and Statistics, University of Jodhpur, Jodhpur.

of the two auxiliary characters on the same unit ; and Y, X_1 and X_2 be the population totals for y, x_1 and x_2 respectively. Further let,

$$u_i=\frac{y_i}{Np_{1i}}\ ;\quad v_i=\frac{x_{2i}}{Np_{1i}}\ ;\quad p_{1i}=\frac{x_{1i}}{X_1}$$

$$\bar{u}_n=n^{-1}\sum_{i=1}^{n} u_i=\bar{y}_{1pps}$$

$$\bar{v}_n=n^{-1}\sum_{i=1}^{n} v_i=\bar{x}_{2pps}$$

$$\sigma_u^2=\sum_{i=1}^{N} p_{1i}\,(u_i-\bar{Y})^2$$

$$\sigma_v^2=\sum_{i=1}^{n} p_{1i}\,(v_i-\bar{X}_2)^2$$

$$\bar{y}_R=\frac{\bar{u}_n}{\bar{v}_n}\cdot\bar{X}_2\ ;\ C_u=\frac{\sigma_u}{\bar{Y}}\ ;\ C_v=\frac{\sigma_v}{\bar{X}_2}$$

$$v\,(\bar{y}_R)=n^{-1}\,\bar{Y}^2\left(C_u^2+C_v^2-2\,\rho_{uv}\,C_uC_v\right)$$

where
$$\rho_{uv}=\sum_{i=1}^{N}\frac{p_{1i}\,(u_i-\bar{Y})\,(v_i-\bar{X}_2)}{\sigma_u\,\sigma_v}$$

3. Proposed Estimator

Consider the pps sampling scheme with replacement based on x_1. It is assumed that the information on x_1 for every unit of the population and for x_2, on the population mean $\bar{X}_2$ is available. Then, the proposed estimator of $\bar{Y}$ is given by

$$\text{To}=k\,\bar{y}_R+(1-k)\,\bar{y}_{1pps}, \qquad ...(3.1)$$

where k is a constant to be determined so that M (To), the mean squared error of To is minimum. The expression for bias and MSE of To are

$$B\,(\text{To})=kB\,(\bar{y}_R) \qquad ...(3.2)$$

and

$$M\,(\text{To})=k^2\,M\,(\bar{y}_R)+(1-k)^2\,V\,(\bar{y}_{1pps})+2k\,(1-k)\,\text{Cov}\,(\bar{y}_R,\bar{y}_{1pps}) \qquad ...(3.3)$$

where $\text{Cov}\,(\bar{y}_R,\bar{y}_{1pps})=n^{-1}\,\bar{Y}^2\,(C_u-\rho_{uv}\,C_u\,C_v)$ ...(3.4)

The value of k which minimises (3.3) is

$$k_{opt}=\rho_{uv}\ \frac{C_u}{C_v} \qquad \text{...(3.5)}$$

k_{opt} involves certain unknown population parameters. Therefore in practical applications approximate value of k_{opt} will have to be determined. One way would be to substitute sample values of ρ_{uv}, C_u and C_v (ignoring the sampling error). This would be plausible for large samples only. In case of small samples, a modified estimator T_o', using $k=1/2$, is suggested which is found to be more efficient than $\bar{y}_R$ or $\bar{y}_{1pps}$ under certain conditions.

Theorem 1 : For k_{opt}, the bias and MSE of T_o, to the first degree of approximation, are

$$B\,(\text{To})=n^{-1}\ \rho_{uv}\ \sigma_u\ (C_v-\rho_{uv}\ C_u) \qquad \text{...(3.6)}$$

and

$$M\,(\text{To})=n^{-1}\ \sigma_u^2\ \ (1-\rho_{uv}^2) \qquad \text{.. (3.7)}$$

respectively.

Proof: Substitution of k_{opt} from (3.5) and Cov $(\bar{y}_R, \bar{y}_{1pps})$ from (3.4) in (3.3) leads to (3.7).

From (3.7) it can be seen that $M\,(T_o)$ is of the same form as the variance of the regression estimator under ppswr sampling. Thus it can be estimated approximately on the lines of the variance of regression estimator under ppswr sampling.

Theorem 2 : For $k=1/2$, the bias and MSE of T_o', to the first degree of approximaton, are give by

$$B\,(T_o')=\tfrac{1}{2}\ B\,(\bar{y}_R) \qquad \text{...(3.8)}$$

and

$$M\,(T_o')=n^{-1}\ \sigma_u^2 \left(1+\frac{C_v^2}{4\ C_u^2}-\rho_{uv}\ \frac{C_v}{C_u}\right) \qquad \text{...(3.9)}$$

respectively.

4. Efficiency Comparison

In this section the variances of T_o and T_o' are compared with those of $\bar{y}_{1pps}$ and $\bar{y}_R$.

4.1 Comparison of To with $\bar{y}_{1pps}$ and $\bar{y}_R$:

The percentage gain in efficiency of To over $\bar{y}_{1pps}$ and $\bar{y}_R$ are

$$\frac{\rho_{uv}^2}{1-\rho_{uv}^2}\times 100 \qquad \text{...(4.1.1)}$$

and

$$\frac{\rho_{uv}^2}{1-\rho_{uv}^2}\left(\frac{1}{K_{opt}}-1\right)^2 \times 100 \qquad \text{...(4.1.2)}$$

respectively.

4.2. Comparison of T_o' with $\bar{y}_{1pps}$ and $\bar{y}_R$:

It is seen that T_o' is more efficient than

(*i*) $\bar{y}_{1pps}$ if *kopt* $>.25$ ...(4.2.1)

and

(*ii*) $\bar{y}_R$ if *kopt* $<.75$ ...(4.2.2)

Thus *kopt* if lies in the range of .25 to .75, T_o' will be more efficient thon $\bar{y}_{1pps}$ or $\bar{y}_R$. In practice an approximate value of *kopt* can be obtained by sample value.

4.3 Comparison of To with $\bar{y}_{1pps}$ and $\bar{y}_R$ when k departs from k_{opt}.

We will examine the case when we do not use *kopt* as it requires a prior knowledge of certain population parameters which is generally not known in advance.

Consider mean squared error of T_o From (3.3). It is obtained that

$$M\,(\text{To})-M\,(\bar{y}_R)=(k^2-1)\,M\,(\bar{y}_R)+(1-k)^2V(\bar{y}_{1pps}) + 2k\,(1-k)\ \text{Cov}\,(\bar{y}_R, \bar{y}_{1pps}) \qquad \text{...(4.3.1.)}$$

On putting the values of $M\,(\bar{y}_R)$, $V\,(\bar{Y}_{1pps})$ and $\text{Cov}\,(\bar{y}_R, \bar{y}_{1pps})$, (4.3.1) yields

$$M\,(T_o)-M(\bar{y}_R)=C_v^2\ \ k^2-2\,\rho_{uv}\,C_u\,C_v\,k+\left(2\,\rho_{uv}\,C_u\,C_v-C_v^2\right) \qquad \text{...(4.3.2)}$$

It is seen from (4.3.2) that $[M\,(T_o)-M\,(\bar{y}_R)]$ will always be negative when

$$1<k<\left(2\rho_{uv}\,\frac{C_u}{C_v}-1\right) \qquad \text{...(4.3.3)}$$

which in turn means that the proposed estimator T_o for $\bar{Y}$ will always be more efficient than the conventional ratio éstimator $\bar{y}_R$, if

$$\left|K-K_{opt}\right| < \left|1-\rho_{uv}\frac{C_u}{C_v}\right| \qquad \text{...(4.3.4)}$$

Similarly, from (4.3.1), it could be seen that

$$M\,(T_o)-V\,(\bar{y}_{1pps})=k^2\,M\,(\bar{y}_R)+[(1-k)^2-1]\,V\,(\bar{y}_{1pps}) + 2k\,(1-k)\,C_{ov}\,(\bar{y}_R, \bar{y}_{1pps}) \qquad \text{...(4.3.5)}$$

which, on substituting the values, yields

$$M\,(T_o)-V\,(\bar{y}_{1pps})=C_v\,k-2\,\rho_{uv}\,C_u \qquad \text{...(4.3.6)}$$

From (4.3.6), it is obvious that T_o will be more efficient than $\bar{y}_{1pps}$ if

$$K<2\,\rho_{uv}\frac{C_u}{C_v} \qquad \text{...(4.3.7)}$$

or, in other words,

$$\left|K-K_{opt}\right|<\left|\rho_{uv}\frac{C_u}{C_v}\right| \qquad \text{...(4.3.8)}$$

5. Empirical Illustration

Table 1 presents the gain in efficiency of T_o and T_o' over $\bar{y}_{1pps}$

TABLE 1

The percentage gain in efficiency of T_o and T_o' over $\bar{y}_{1pps}$ and $\bar{y}_R$

N	Source	y	x_1	x_2	Gain in efficiency			
					Over $\bar{y}_{1pps}$		Over $\bar{y}_R$	
					Using		Using	
					To	To'	To	To'
18	Goon & et.el.	yield of dry bark	Ht. in inches	Girth at a ht. of 6″	31.50	24.00	110.40	98.51
54	Census handbook Brauch (Gujarat)	No. of cultivators.	No. of female cultivators	No. of Labour	411.70	154.81	2566.61	706.58

and $\bar{y}_R$ for two natural populations. For population 1, the departure of kopt (=.348) from k (=½) is 0.152 while for population 2 (kopt=.286) it is .214. For both the populations T_o and T_o' are considerably more efficient than $\bar{y}_{1pps}$ or $\bar{y}_R$. As expected when k departs from the optimum value, the gain in efficiency reduces but even for k=0.5 there is substantial gain in efficiency over both $\bar{y}_{1pps}$ and $\bar{y}_R$.

Summary

An estimator, combining ratio and PPS estimators has been proposed, which is found to be always more efficient than usual pps estimator or ratio estimator under pps sampling Further, it is shown that the proposed estimator will be superior to the ratio estimator under pps sampling and to the usual pps estimator, if the difference between weights taken and optimum weights is less than

$$\left| 1 - \frac{\rho_{uv} C_u}{C_v} \right| \text{ and } \left| \rho_{uv} \frac{C_u}{C_v} \right| \text{ respectively.}$$

Acknowledgement

The authors are grateful to the referee for his suggestions in improving the presentation of the paper.

References

[1] District Census hand book (1961) : Brauch (Gujarat).

[2] Goon, Gupta and Das Gupta (1968) : Fundamentals of Statistics, Asia Publications, p.p. 315.

[3] Singh, M.P. (1965) : On the estimation of ratio and product of population parameters, Sankhyā, 27, B, 321-28.

[4] Singh, M.P. (1967) : Ratio cum Product method of estimation, Metrika 12.