Ann. Inst. Statist. Math.
33 (1981), Part A, 225-231

ON A CLASS OF ALMOST UNBIASED RATIO ESTIMATORS

T. J. RAO

(Received Aug. 26, 1980; revised Mar. 10, 1981)

Summary

Murthy and Nanjamma [4] studied the problem of construction of almost unbiased ratio estimators for any sampling design using the technique of interpenetrating subsamples. Subsequently, Rao [7], [8] has given a general method of constructing unbiased ratio estimators by considering linear combinations of the two simple estimators based on the ratio of means and the mean of ratios. However, it is difficult to choose an optimum weight (Rao [9]) which minimizes the variance of the combined estimator since the weights are random in certain cases. In this note, we consider a different method of combining these estimators and obtain a general class of almost unbiased ratio estimators of which Murthy and Nanjamma's is a particular case and derive an optimum in this class. The case of simple random sampling where a similar class of almost unbiased ratio estimators can be developed is briefly discussed. The results are illustrated by means of simple numerical examples.

1. Introduction

Consider a finite population of size N consisting of units $(U_1, U_2, \cdots, U_N)$. Let $\boldsymbol{y}$ be the characteristic of interest taking values Y_i on the units U_i, $i=1, 2, \cdots, N$. Information on an auxiliary characteristic $\boldsymbol{x}$, related to $\boldsymbol{y}$, taking values X_i on the units U_i, $i=1, 2, \cdots, N$ is available in most of the sample survey situations. We are interested in estimating the parameters such as population total $Y=\sum_{i=1}^{N} Y_i$ or the population mean $\bar{Y}=Y/N$. In order to increase the precision of the estimates, it is often found advantageous to make use of auxiliary information at the estimation stage or for the selection of the sample or both. One such method where auxiliary information is made use of at the estimation stage is the method of ratio estimation which is very simple and is found to be more efficient under well known circumstances than the conventional method of estimation using the sample mean which is not based on such information. The method of ratio estimation for estimating the population mean $\bar{Y}$ (or equivalently the total Y) of the study variate $\boldsymbol{y}$, consists in getting an estimator $\hat{R}$ of the population ratio $R=Y/X=\sum_{i=1}^{N} Y_i \Big/ \sum_{i=1}^{N} X_i$ and multiplying this estimator by the known population mean $\bar{X}=X/N$ of the $\boldsymbol{x}$-variate (or by X when estimating the total Y). Ratio estimators are in general biased and a number of authors have studied the bias and mean square error in detail both theoretically and empirically (for example, see Murthy [3] and Sukhatme [10]). During the last couple of decades, the problem of construction of unbiased ratio estimators attracted the attention of many samplers. Murthy and Nanjamma [4] have extensively studied the problem of construction of unbiased ratio estimators for any sample design using the technique of interpenetrating subsamples. Rao [7] has presented a general method of constructing unbiased ratio estimators by considering linear combinations of the two simple estimators $r\bar{X}=(\bar{y}/\bar{x})\bar{X}$ and $\bar{r}\bar{X}=\bar{X}\sum(y_i/x_i)/n$. Certain theoretical investigations regarding these estimators are presented in Rao [8] while the efficiencies of the linear combination estimators for different values of the weights are discussed in Rao [9]. As illustrated here, it is difficult to choose an optimum weight which minimizes the variance of the combined estimator since the weights happen to be stochastic in certain cases. Motivated by this, we consider a different type of combination of the two estimators $r\bar{X}$ and $\bar{r}\bar{X}$ and obtain a class of 'almost unbiased ratio estimators' which includes Murthy and Nanjamma's [4] estimator. We then find an optimum estimator in this class. Finally we illustrate the results by means of simple examples of sampling from two populations.

2. A general class of almost unbiased ratio estimators

First we discuss the interpenetrating subsamples model considered by Murthy and Nanjamma [4]. Let (y_i, x_i) be unbiased estimates of the population totals Y and X respectively from the ith independent interpenetrating subsample, $i=1, 2, \cdots, n$. Let $r=\bar{y}/\bar{x}$ and $\bar{r}=\sum(y_i/x_i)/n$.

Motivated by the study of Rao [9], we now consider the linear combination

$$\hat{\bar{Y}}=\lambda r\bar{X}+(1-\mathrm{E}(\lambda))\bar{r}\bar{X}\ , \tag{2.1}$$

where λ is a random variable. $\hat{\bar{Y}}$ is unbiased for $\bar{Y}$ if

$$\mathrm{E}(\hat{\bar{Y}})=\bar{Y}$$

i.e.

$$\mathrm{E}\,(\lambda r\bar{X}-\mathrm{E}\,(\lambda)\,\bar{r}\bar{X})=\mathrm{E}\,(\bar{y}-\bar{r}\bar{X})\,. \tag{2.2}$$

An obvious solution for λ, from (2.2) is given by $\lambda=\bar{x}/\bar{X}$, which gives the unbiased estimator $\bar{y}$. Rewriting (2.2) by introducing $r\bar{X}$ in the r.h.s. we have

$$\mathrm{E}\,(\lambda r\bar{X}-\mathrm{E}\,(\lambda)\,\bar{r}\bar{X})=\mathrm{E}\,(kr\bar{X}+\bar{y}-\bar{r}\bar{X}-kr\bar{X}) \tag{2.3}$$

where k is a constant.

Noticing that for the interpenetrating subsamples model we have

$$\mathrm{E}\,(r\bar{X})=\bar{Y}+B(r\bar{X})\sim\bar{Y}+(1-c)B(\bar{r}\bar{X})\,,$$

where $c=(n-1)/n$, we observe that

$$\mathrm{E}\,(r\bar{X})\sim c\bar{Y}+(1-c)\,\mathrm{E}\,(\bar{r}\bar{X})\,.$$

We then have from (2.3)

$$\begin{aligned}\mathrm{E}\,(\lambda r\bar{X}-\mathrm{E}\,(\lambda)\,\bar{r}\bar{X})&=\mathrm{E}\,(kr\bar{X}+\bar{y}-\bar{r}\bar{X}-k\{\bar{y}c+(1-c)\bar{r}\bar{X}\})\\ &=\mathrm{E}\,(\{k+(1-ck)\bar{x}/\bar{X}\}r\bar{X}-\{1+(1-c)k\}\bar{r}\bar{X})\,. \qquad (2.4)\end{aligned}$$

From (2.4) it is immediate that

$$\lambda=k+(1-ck)\bar{x}/\bar{X}$$

is a solution which generates a class of 'almost unbiased ratio estimators.' Thus we have the following

THEOREM 2.1. *$\hat{\bar{Y}}^{U}=\lambda r\bar{X}+(1-\mathrm{E}\,(\lambda))\bar{r}\bar{X}$ is unbiased (almost) for $\bar{Y}$ where $\lambda=k+(1-ck)\bar{x}/\bar{X}$ for any constant k and $c=(n-1)/n$.*

Remark 2.1. Notice that $k=0$ gives the conventional unbiased estimator $\hat{\bar{Y}}=\bar{y}$. With $k=1$, we have the estimator given by

$$\hat{\bar{Y}}=[1+(1-c)\bar{x}/\bar{X}]r\bar{X}-[1-c]\bar{r}\bar{X}=(1+(\bar{x}/n\bar{X}))r\bar{X}-(1/n)\bar{r}\bar{X}\,, \tag{2.5}$$

while $k=c^{-1}$ gives the estimator

$$\hat{\bar{Y}}^{MN}=c^{-1}r\bar{X}+(1-c^{-1})\bar{r}\bar{X}=(n/n-1)r\bar{X}-(1/n-1)\bar{r}\bar{X}\,, \tag{2.6}$$

which is the estimator due to Murthy and Nanjamma [4]. The form of the general class of estimators is given by

$$\hat{\bar{Y}}^{U}=(1-ck)\bar{y}+kr\bar{X}-((1-c)k)\bar{r}\bar{X}\,. \tag{2.7}$$

3. Choice of an optimum estimator in the class

From (2.7) we have

$$\begin{aligned} \mathrm{V}\,(\hat{\bar{Y}}^{U}) &= (1-ck)^2\,\mathrm{V}\,(\bar{y}) + k^2\,\mathrm{V}\,(r\bar{X}) + (1-c)^2k^2\,\mathrm{V}\,(\bar{r}\bar{X}) \\ &\quad + 2k(1-ck)\,\mathrm{Cov}\,(\bar{y}, r\bar{X}) - 2k^2(1-c)\,\mathrm{Cov}\,(r\bar{X}, \bar{r}\bar{X}) \\ &\quad - 2k(1-c)(1-ck)\,\mathrm{Cov}\,(\bar{y}, \bar{r}\bar{X})\ . \end{aligned}$$

Minimization of this leads to the optimum value of k given by

$$k^{\mathrm{opt}} = A/B$$

where

$$A = c\,\mathrm{V}(\bar{y}) - \mathrm{Cov}\,(\bar{y}, r\bar{X}) + (1-c)\,\mathrm{Cov}\,(\bar{y}, \bar{r}\bar{X})$$

and

$$\begin{aligned} B &= c^2\,\mathrm{V}\,(\bar{y}) + \mathrm{V}\,(r\bar{X}) + (1-c)^2\,\mathrm{V}\,(\bar{r}\bar{X}) - 2c\,\mathrm{Cov}\,(\bar{y}, r\bar{X}) \\ &\quad - 2(1-c)\,\mathrm{Cov}\,(r\bar{X}, \bar{r}\bar{X}) + 2c(1-c)\,\mathrm{Cov}\,(\bar{y}, \bar{r}\bar{X})\ . \end{aligned}$$

Remark 3.1. For the case of Simple Random Sampling With Out Replacement (SRSWOR), let y_i and x_i denote respectively the y and x values for the ith sampled unit, $i=1, 2, \cdots, n$. Now, let $r=\bar{y}/\bar{x}$ and $\bar{r}=\sum (y_i/x_i)/n$. We have

$$B(r\bar{x}) = -\mathrm{Cov}\,(\bar{y}/\bar{x}, \bar{x})\ .$$

When this is approximated by the quantity

$$B(r\bar{X}) \sim -\frac{N-n}{Nn}\,\frac{N}{N-1}\,\mathrm{Cov}\,(y_i/x_i, x_i) \tag{3.1}$$

we have the relation

$$B(\bar{r}\bar{X}) = -\mathrm{Cov}\,(y_i/x_i, x_i) = (1-c')^{-1}B(r\bar{X})$$

where $c'=N(n-1)/n(N-1)$. Thus we have

$$\mathrm{E}\,(r\bar{X}) \sim c'\,\bar{Y} + (1-c')\,\mathrm{E}\,(\bar{r}\bar{X})$$

using the approximation (3.1). As before this then leads to the (almost) unbiased estimator for $\bar{Y}$ given by

$$\hat{\bar{Y}} = \lambda r\bar{X} + (1-\mathrm{E}\,(\lambda))\bar{r}\bar{X}\ ,$$

where $\lambda=k+(1-c'k)\bar{x}/\bar{X}$ for any constant k and $c'=N(n-1)/n(N-1)$. Notice that $k=0$ gives the usual unbiased estimator $\hat{\bar{Y}}=\bar{y}$, while $k=(c')^{-1}$ gives the estimator $\hat{\bar{Y}}^{MN}=c'^{-1}r\bar{X}+(1-c'^{-1})\bar{r}\bar{X}$ (see Rao [9]). The form of the general class of estimators is given by

$$\hat{\bar{Y}}^{U}=(1-c'k)\bar{y}+kr\bar{X}-((1-c')k)\bar{r}\bar{X}\,. \tag{3.2}$$

Next one can easily obtain the optimum value of k that minimizes $\mathrm{V}(\hat{\bar{Y}}^{U})$ in a similar way as above, details of which we omit here. It may also be noted that for large samples $\hat{\bar{Y}}^{U}$ and its variance reduce to the expressions given earlier.

4. Illustrations

We now illustrate the results by two empirical examples of sampling 2 units from small populations, one of which (Population I) was considered by Goodman and Hartley [1] and Nieto de Pascual [5] for comparison of various ratio estimators and the other (Population II) was considered by Sukhatme ([10], p. 165). First, consider the following Population I:

Unit	U_1	U_2	U_3	U_4
x-values	2	2	4	6
y-values	2	6	6	10

for which $c=(n-1)/n=1/2$. The table below gives the variance of $\hat{\bar{Y}}^{U}=(1-ck)\bar{y}+kr\bar{X}-((1-c)k)\bar{r}\bar{X}$ for different choices of k and the variance of a few other ratio-type estimators of $\bar{Y}$ and their biases considered by Murthy [3] for the interpenetrating subsamples model.

Table 4.1 M.S.E. of $\hat{\bar{Y}}^{U}$ for different k compared with other estimators

k	$\mathrm{V}(\hat{\bar{Y}}^{U})$	$(\mathrm{Bias})^2$	rmse
0	2.6667	0	0.0741
1	0.6812	0.0024	0.0190
1.68641	0.2870	0.0069	0.0082
2	0.3693	0.0097	0.0105
$\hat{\bar{Y}}_{\mathrm{HR}}^{\mathrm{IPS}}$	0.4487	0.0081	0.0127
$\hat{\bar{Y}}_{\mathrm{N}}^{\mathrm{IPS}}$	0.4002	0.0089	0.0114
$r\bar{X}$	0.9093	0.0075	0.0255
$\bar{r}\bar{X}$	2.2331	0.0733	0.0641

In this table the last column gives the relative mean square error (rmse) of the estimators. The rmse corresponding to $\hat{\bar{Y}}_{\mathrm{opt}}^{U}$ for the optimum choice of k equal to 1.68641 is the least among those considered which shows that this is the most efficient estimator in this case. In practice, one can substitute the estimated values of the variances and convariances in order to obtain a 'near-optimum' value of k. In the above table, note that the choice of $k=2$ corresponds to $\hat{\bar{Y}}_{MN}$, $\hat{\bar{Y}}_{HR}^{IPS}=\bar{r}\bar{X}+2(\bar{y}-\bar{r}\bar{x})$ and $\hat{\bar{Y}}_{N}^{IPS}=r\bar{X}+(\bar{y}-\bar{r}\bar{x})$.

We now turn our attention to the Population II studied earlier by Sukhatme ([10], p. 165). The Population consists of

Unit (Agricultural area unit)	x-values (Mean agricultural area)	y-values (Mean no. of live stock)
1	63.7	25.4
2	155.3	50.1
3	245.7	76.0
4	344.4	99.2
5	491.6	150.8
6	767.5	244.4
7	1604.0	425.1

for which $c=(n-1)/n=1/2$. The table below gives the variance of $\hat{\bar{Y}}^{U}$ considered above for different choices of k besides the variances of a few other estimators and their biases.

Table 4.2 M.S.E. of $\hat{\bar{Y}}^{U}$ for different k compared with other estimators

k	$V(\hat{\bar{Y}}^{U})$	$(\text{Bias})^2$	rmse
0	7029.2155	0	0.3003
1	1458.1745	0.6740	0.0623
1.78167	129.9061	2.1396	0.0056
2	233.8235	2.6960	0.0101
$\hat{\bar{Y}}_{HR}^{IPS}$	667.1085	5.6259	0.0287
$\hat{\bar{Y}}_{N}^{IPS}$	406.5689	3.4656	0.0175
$r\bar{X}$	123.7869	29.6579	0.0066
$\bar{r}\bar{X}$	170.3926	157.0956	0.0140

In this table, the last column gives the relative mean square error (rmse) of the estimators. The rmse corresponding to $\hat{\bar{Y}}^{U}$ for the optimum choice of k equal to 1.78167 is the least among the estimators considered. It may also be noted that the rmse of $\hat{\bar{Y}}_{\text{opt}}^{U}$ is less than the rmse of the estimators considered in J. N. K. Rao ([6], p. 223) for this population of Sukhatme's.

Remark 4.1. Considering the SRSWOR case, it is seen that for the Yates Population the rmse corresponding to the optimum estimator turns out to be 0.0052 while the rmse of the competitors Hartley-Ross estimator, Nieto de Pascual's estimator, $r\bar{X}$ and $\bar{r}\bar{X}$ are respectively 0.0122, 0.0136, 0.0095 and 0.0136. Further, under the approximation

(3.1) it is found that $B(\hat{\bar{Y}}^U)=0.002975k$ and for small k, the square of the bias is negligible.

Acknowledgements

The author wishes to thank the referee for the comments and suggestions.

IOWA STATE UNIVERSITY*

REFERENCES

[1] Goodman, L. A. and Hartley, H. O. (1958). The precision of unbiased ratio-type estimators, *J. Amer. Statist. Ass.*, **53**, 491-508.

[2] Hartley, H. O. and Ross, A. (1954). Unbiased ratio estimators, *Nature*, **174**, 270-271.

[3] Murthy, M. N. (1967). *Sampling Theory and Methods*, Statistical Publishing Society, Calcutta.

[4] Murthy, M. N. and Nanjamma, N. S. (1959). Almost unbiased ratio estimators based on interpenetrating sub-sample estimates, *Sankhyā*, **21**, 381-392.

[5] Nieto de Pascual, J. (1961). Unbiased ratio estimators in stratified sampling, *J. Amer. Statist. Ass.*, **56**, 70-87.

[6] Rao, J. N. K. (1969). Ratio and regression estimators, *New Developments in Survey Sampling* (eds. N. L. Johnson and H. Smith), Wiley-Interscience, 213-234.

[7] Rao, T. J. (1966). On certain unbiased ratio estimators, *Ann. Inst. Statist. Math.*, **18**, 117-121.

[8] Rao, T. J. (1979). Unbiasedness and convexity in ratio estimation, *Technical Report*, Math. Stat. Unit, Indian Statistical Institute, Calcutta.

[9] Rao, T. J. (1980). A note on unbiasedness in ratio estimation, to appear.

[10] Sukhatme, P. V. (1954). *Sampling Theory of Surveys with Applications*, Iowa State College Press, Ames, Iowa.

* On leave from Indian Statistical Institute, Calcutta.