Metrika (1994) 41:21-27

Metrika

Stratification, Midzuno-Sen Strategy and Nonnegative Unbiased Variance Estimation

VIDYASAGAR RAMCHANDRA R. PADMAWAR

Stat-Math Division, Indian Statistical Institute, 8th Mile, Mysore Road, Bangalore 560059, India

Abstract: When the ratio method is appropriate for estimating the population total one is faced with the problem of nonavailability of uniformly nonnegative unbiased variance estimators (*nnuve*). Here we highlight the twofold role of stratification in that it not only improves the efficiency of the ratio method of estimation but it also enhances the chances of getting uniformly nonnegative unbiased variance estimators.

Keywords: Midzuno-Sen strategy, variance estimators, unbiasedness, uniorm nonnegativity, stratification, measure of uncertainty, measurisation.

Subject classification: 62D05

1 Introduction

Consider a finite population of size N. Let y be the study variable taking value y_i on unit i, $1 \leq i \leq N$. Let x be an auxiliary variable, positively correlated with y, taking value $x_i > 0$ on unit i, $1 \leq i \leq N$. The sampling strategy, for estimating the population total, $Y = \sum_{i=i}^{N} y_i$, that consists of Midzuno-Sen sampling design and the ratio estimator uses advantageously the auxiliary information both at the time of selection of units and at the estimation stage. This strategy is quite efficient especially when the ratio method of estimation is appropriate. The main drawback of this strategy, however, is the possible nonavailability of uniformly nonnegative unbiased variance estimator (*nnuve*), which is a measure of the efficiency. J. N. K. Rao and K. Vijayan (1977) studied the problem of obtaining *nnuves* for this strategy. Padmawar (1990) studied the problem of *nnuves* for a general class of strategies.

In this note we investigate the role of stratification, *vis-a-vis* a superpopulation model, to improve the efficiency of the ratio method of estimation and at the same time enhance the chances of getting uniformly *nnuves*.

Appropriateness of the ratio estimator can best be translated in the language of superpopulation models by hypothesising the following superpopulation model ξ.

Let $y_1, y_2, \ldots, y_N$ be a realisation of random variables $Y_1, Y_2, \ldots, Y_N$ with a joint distribution that is specified, though not completely, by the first two moments as

$$\begin{aligned} E_\xi(Y_i) &= \beta x_i & i &= 1, 2, \ldots, N \\ V_\xi(Y_i) &= \sigma^2 x_i & i &= 1, 2, \ldots, N \\ \mathrm{Cov}_\xi(Y_i, Y_j) &= 0 & i &\neq j = 1, 2, \ldots, N \end{aligned} \tag{1.1}$$

where $\sigma^2 > 0$ and β are the unknown parameters of the superpopulation model ξ. Subscript ξ in (1.1) is to indicate that the operations are carried out *w.r.t.* the model ξ.

For a sample s of size n let $y_s = \sum_{i \in s} y_i$, $x_s = \sum_{i \in s} x_i$. Let $X = \sum_{i=1}^{N} x_i$ and $M_r = \binom{N-r}{n-r}$, $r = 0, 1, 2$. The Midzuno-Sen sampling design p_M assigns probability $\frac{x_s}{M_1 X}$ to sample s, $s \in S$, S being the collection of all samples of size n and the ratio estimator is of course given by

$$t_R = \frac{y_s}{x_s} X \tag{1.2}$$

(p_M, t_R) denotes the Midzuno-Sen strategy which is unbiased for estimating the population total.

Rao and Vijayan (1977) obtained a useful expression for the variance of (p_M, t_R). They deduced the necessary form of *nnuve* and proposed a set of sufficient conditions for the uniform nonnegativity of such an estimator. We list down their findings

i) The expression for the variance,

$$V = V(p_M, t_R) = -\frac{1}{2} \sum_{i=1}^{N} \sum_{j=i}^{N} a_{ij} x_i x_j \left(\frac{y_i}{x_i} - \frac{y_j}{x_j} \right)^2 \tag{1.3}$$

where

$$\begin{aligned} a_{ii} &= \frac{X}{M_1} \sum_{s \ni i} \frac{1}{x_s} - 1 & i &= 1, 2, \ldots, N \, . \\ a_{ij} &= \frac{X}{M_1} \sum_{s \ni i,j} \frac{1}{x_s} - 1 & i &\neq j = 1, 2, \ldots, N \, . \end{aligned} \tag{1.4}$$

ii) The necessary form of an *nnuve*, for $s \in S$,

$$v \equiv v_s = -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} a_{ij}(s)x_i x_j\left(\frac{y_i}{x_i}-\frac{y_j}{x_j}\right)^2 \tag{1.5}$$

where $a_{ij}(s) = 0$ if $i \notin s$ or $j \notin s$

$$\sum_{s \ni i,j} a_{ij}(s)p(s) = a_{ij} \qquad i \neq j = 1, 2, \ldots, N \ . \tag{1.6}$$

and finally,
iii) A set of sufficient conditions for the uniform nonnegativity of the estimator in (1.5),

$$a_{ij}(s) \leq 0 \qquad \text{for all } i \neq j \in s \ , \quad s \in S \ . \tag{1.7}$$

Rao and Vijayan (1977) suggested the following estimator having the form (1.5).

$$v_1(s) = -\frac{1}{2}\frac{X}{x_s}\left(\frac{X}{x_s}-\frac{N-1}{n-1}\right)\sum_{i \in s}\sum_{j \in s} x_i x_j\left(\frac{y_i}{x_i}-\frac{y_j}{x_j}\right)^2 , \qquad s \in S \ . \tag{1.8}$$

Sufficient conditions (1.7) for this estimator to be uniformly nonnegative would be

$$\frac{X}{x_s} \leq \frac{N-1}{n-1} \qquad \forall\, s \in S \tag{1.9}$$

Rao and Vijayan (1977) also observed that

$$\sum_{j=1}^{N} a_{ij}x_j = 0 \qquad \forall\, i = 1, 2, \ldots, N \tag{1.10}$$

It is easy to see that {T. J. Rao (1977)}

$$\sum_{i=1}^{N} a_{ii}x_i = \frac{N-n}{n}X \ . \tag{1.11}$$

Let the population be divided into L strata of sizes $N_1, N_2, \ldots, N_L$ respectively, $\sum_{h=1}^{L} N_h = N$. Let (p_{MSt}, t_{RSt}) be the stratified version of the Midzuno-Sen sampling strategy. This strategy employs Midzuno-Sen sampling design within each stratum and

$$t_{RSt} = \sum_{h=1}^{L} t_{Rh}$$

where $t_{Rh} = \dfrac{\sum_{i \in s_h} y_i}{\sum_{i \in s_h} x_i} x_h$ is the ratio estimator for the total $Y_h = \sum_{i \in S_h} y_i$ of the hth stratum, where $X_h = \sum_{i \in S_h} x_i$, S_h denotes the hth stratum and s_h the sample from the hth stratum, $1 \leq h \leq L$.

We compare the performance of the strategies (p_M, t_R) and (p_{MSt}, t_{RSt}) using the following measure of uncertainty.

$$M(p, t) = E_\xi V(p, t) \tag{1.12}$$

2 Stratified Strategy is Better

In view of the above considerations we proceed to compute expected variance (1.12) for the two strategies. From (1.3) we have,

$$\begin{aligned}
E_\xi V(p_M, t_R) &= -\frac{1}{2}\sum_{i=1}^{N}\sum_{j\neq i}^{N} a_{ij}x_i x_j E_\xi\left(\frac{Y_i}{x_i} - \frac{Y_j}{x_j}\right)^2 \\
&= -\frac{1}{2}\sum_{i=1}^{N}\sum_{j\neq i}^{N} a_{ij}x_i x_j V_\xi\left(\frac{Y_i}{x_i} - \frac{Y_j}{x_j}\right) \\
&= -\frac{1}{2}\sum_{i=1}^{N}\sum_{j\neq i}^{N} a_{ij}(\sigma^2 x_j + \sigma^2 x_i) \\
&= -\sigma^2 \sum_{i=1}^{N}\sum_{j\neq i}^{N} a_{ij}x_j \\
&= \sigma^2 \sum_{i=1}^{N} a_{ii}x_i \qquad \{\text{using } (1.10)\} \\
&= \frac{N-n}{n} X\sigma^2 \qquad \{\text{using } (1.11)\}
\end{aligned}$$

Thus

$$E_\xi V(p_M, t_R) = \frac{N-n}{n} X\sigma^2 \tag{2.1}$$

Analogous to (2.1) we obtain, for the allocation $(n_1, n_2, \ldots, n_L)$, $\sum_{h=1}^{L} n_h = n$,

$$E_\xi V(p_{MSt}, t_{RSt}) = \sigma^2 \sum_{h=1}^{L} \frac{N_h - n_h}{n_h} X_h \ . \tag{2.2}$$

Minimising (2.2) subject to $\sum_{h=1}^{L} n_h = n$ we get the optimal allocation as

$$n_h^0 = \frac{n\sqrt{N_h X_h}}{\sum_{h=1}^{L} \sqrt{N_h X_h}}, \qquad 1 \le h \le L \tag{2.3}$$

and the expected variance (2.2) under the optimal allocation as

$$E_\xi V(p_{MSt}, t_{RSt}) = \sigma^2 \left\{ \frac{1}{n} \left(\sum_{h=1}^{L} \sqrt{N_h X_h} \right)^2 - X \right\} \tag{2.4}$$

We are now in a position to prove the following theorem.

Theorem 2.1: The strategy (p_{MSt}, t_{RSt}) under the optimal allocation (2.3) is superior to the unstratified strategy (p_M, t_R) *w.r.t.* the criterion (1.12).

Proof: Using (2.1) and (2.4) we have,

$$E_\xi V(p_M, t_R) - E_\xi V(p_{MSt}, t_{RSt}) = \frac{\sigma^2}{n} \left\{ NX - \left(\sum_{h=1}^{L} \sqrt{N_h X_h} \right)^2 \right\}$$

which is always nonnegative by Cauchy-Schwarz inequality.

Let us now consider an example to see how the stratification helps in getting a uniformly *nnuve*.

Example 2.1: Consider the following population of size 10 for which we rearrange the units in the increasing order of the x-values.

i:	1	2	3	4	5	6	7	8	9	10
x:	89	112	125	163	229	254	326	359	442	559

Here $N = 10$, $X = 2658$. For the case $n = 4$ consider the estimator v_1 given by (1.8). The sufficient condition (1.9) due to Rao and Vijayan for the uniform nonnegativity of v_1, $\min_{s \in S} x_s \geq \frac{n-1}{N-1} X$, is not satisfied as $\frac{n-1}{N-1} X = 886$, $\min_{s \in S} x_s = 489$. However if we stratify the population into 2 strata of sizes $N_1 = 6$ and $N_2 = 4$ then $X_1 = 972$ and $X_2 = 1686$. The optimal allocation to the nearest integers is given by $n_1 = n_2 = 2$. Let v_{1St} be the stratified analogue of v_1. Now observe that $\min_{s_1} x_{s_1}$ and $\min_{s_2} x_{s_2}$ for the two strata are 201 and 685 respectively. Further $\frac{n_1 - 1}{N_1 - 1} X_1 = 194.4$ and $\frac{n_2 - 1}{N_2 - 1} X_2 = 562$. Hence $\min_{s_1} x_{s_1} \geq \frac{n_1 - 1}{N_1 - 1} X_1$; $\min_{s_2} x_{s_2} \geq \frac{n_2 - 1}{N_2 - 1} X_2$, conditions analogous to (1.9), are both satisfied. Hence v_{1St} is uniformly nonnegative. Finally, even for the allocation $n_1 = n_2 = 2$, the strategy (p_{MSt}, t_{RSt}) is superior to the strategy (p_M, t_R) as $E_\xi V(p_M, t_R) = 3987\sigma^2$ and $E_\xi V(p_{MSt}, t_{RSt}) = 3630\sigma^2$.

Remark 2.1: It is easy to give any number of such examples. In view of this the problem of characterising the auxiliary vectors $\mathbf{x} = (x_1, x_2, \ldots, x_N) \in \mathbb{R}^+$, for given values of N, L, N_h, n_h, $1 \leq h \leq L$, using the techniques of majorisation, enhancing the chances of getting uniformly *nnuves*, is presently being studied. The problem of obtaining efficient biased variance estimators is presented in a separate note {*vide* Padmawar (1991)}.

References

Padmawar VR (1990) *Strategies admitting nonnegative unbiased variance estimators*, (Revised for J.S.P.I)

Padmawar VR (1991) *Midzuno-Sen sampling strategy and biased variance estimators*, (Submitted for publication).

Rao JNK, Vijayan K (1977) *On estimating the variance in sampling with probability proportional to aggregate size*, Journal of the American Statistical Association, 72: 579–584

Rao JNK (1979) *On deriving the mean square errors and their nonnegative unbiased estimators in finite population sampling*, Journal of the Indian Statistical Association 17:125–136
Rao TJ (1977) *Estimating the variance of the ratio estimator for the Midzuno-Sen sampling scheme*. Metrika 24:203–208

Received 15.12.1992
Revised version 05.02.1993