ON THE OPTIMUM CHARACTER OF SOME ESTIMATORS USED IN MULTISTAGE SAMPLING PROBLEMS

By D. BASU

(Research Fellow of the National Institute of Sciences)
Indian Statistical Institute, Calcutta

1. Introduction

In the National Sample Survey which is being conducted by the Indian Statistical Institute, the country is first divided into a number of strata the object being to get estimates for a very large number of characteristics of each stratum. Each stratum is divided into a large number of primary units out of which a small number of units are randomly selected and the whole sampling enquiry is then localised in those selected primary units. From the estimated characteristics of the selected primary units we have to build up an over-all estimate for the whole stratum and also to give an estimate of the error of our estimate. The problems considered here are quite general and the solutions are independent of the sampling technique used in the primary units. As a matter of fact, any sampling technique that is capable of yielding unbiassed estimates for the primary units may be used, the only condition being that the mode and size of the sample for each primary unit must be given in advance. We consider the following three situations where

(A) the primary units are chosen with replacement and with different probabilities. If a particular unit appears r times then we get r independent estimates for that unit;

(B) the primary units are chosen without replacement and with equal probabilities;

(C) the primary units are chosen as in (A) but if a particular unit appears r times we put the same estimate r times.

The method of approach used here is very simple. The unbiassed estimators given in Deming's book entitled *Some Theory of Sampling* for situations (A) and (B) are very easily deduced and it is incidentally shown that these are indeed the best estimators within a class of estimators. Unbiassed estimators under situation (C) were obtained by S. Raja Rao of the Indian Statistical Institute simultaneously with the author. The optimum character of these estimators also is demonstrated here. We also consider the problem of choosing the probabilities of selection in (A) and (C) and get the optimum theoretical solutions which can be approximated to in actual practice.

2. The Mathematical Set-up

Let $u_1, u_2, \ldots, u_M$ be the M primary units and let θ_j† be some characteristic of u_j in which we are interested. We can estimate θ_j by taking a sample of some type and size from u_j. We assume that the mode and size of the sample that we are allowed

† Throughout this discussion j runs through the values $1, 2, \ldots, M$ and i runs through the values $1, 2, \ldots, m$.

to take from each u_j is determined at the outset. Thus with each u_j is associated a chance variable t_j which is the best (in some sense) unbiassed estimator of θ_j and another chance variable s_j^2 which is the best unbiassed estimator of $\sigma_j^2 = V(t_j)$.

Our problem is to estimate the linear function

$$\theta = a_1\theta_1 + a_2\theta_2 + \ldots + a_M\theta_M$$

and also the variance of our estimator by selecting at random m of the M primary units and then getting observations on the corresponding t_j's and s_j^2s. Without any loss of generality we can assume that $a_1 = a_2 = \ldots = a_M = 1$ for we can always write θ_j for $a_j\theta_j$.

3. SAMPLING WITH REPLACEMENT

The m primary sample units $u_1^*, u_2^*, \ldots, u_m^*$ are chosen with replacement from the M primary units $u_1, \ldots, u_M$ with probabilities $\pi_1, \pi_2, \ldots, \pi_M$ ($\Sigma\pi_j = 1$). The ith sample unit u_i^* can be any one of the u_j's. The corresponding observations on t_j and s_j^2 after being multiplied by π_j^{-1} are denoted by t_i^* and s_i^{*2}.

$$\left.\begin{aligned} &\text{Thus} && t_i^* = \pi_1^{-1}t_1,\ \pi_2^{-1}t_2,\ \ldots \text{ or } \pi_M^{-1}t_M \\ &\text{and} && s_i^{*2} = \pi_1^{-1}s_1^2,\ \pi_2^{-1}s_2^2, \ldots \text{ or } \pi_M^{-1}s_M^2 \end{aligned}\right\} \quad \ldots (3.1)$$

with probabilities $\pi_1, \pi_2, \ldots, \pi_M$ respectively.

Clearly $$E(t_i^*) = \Sigma\pi_j(E\overline{\pi_j^{-1}t_j}) = \Sigma\theta_j = \theta \quad \ldots (3.2)$$

and $$E(s_i^{*2}) = \Sigma\sigma_j^2. \quad \ldots (3.3)$$

Also $$V(t_i^*) = E(t_i^{*2}) - \theta^2 = \Sigma\pi_j^{-1}E(t_j^2) - \theta^2 = \Sigma\pi_j^{-1}(\theta_j^2 + \sigma_j^2) - \theta^2. \quad \ldots (3.4)$$

Since $t_1^*, t_2^*, \ldots, t_m^*$ are independently and identically distributed chance variables it follows (Basu, 1952) that the best unbiassed linear estimator of the common mean θ is

$$\bar{t}^* = \frac{1}{m}(t_1^* + \ldots + t_m^*) \quad \ldots (3.5)$$

and the best unbiassed quadratic estimator of the common variance (3.4) is

$$s^2(t^*) = \frac{1}{m-1}\Sigma(t_i^* - \bar{t}^*)^2 \quad \ldots (3.6)$$

and therefore the best unbiassed quadratic estimator of $V(\bar{t}^*) = \frac{1}{m}V(t_i^*)$ is

$$\frac{1}{m}s^2(t^*). \quad \ldots (3.7)$$

The estimators (3.5) and (3.7) are best not only when we take square of the error as our loss function but with any arbitrary convex loss function.

Now the variance of l^*, the best estimator of θ, will be minimum if (3.4) is minimum.

But
$$\Sigma n_j^{-1}(\theta_j^2+\sigma_j^2) = \Sigma(n_j^{-1}\sqrt{\theta_j^2+\sigma_j^2})^2.\Sigma(n_j)^2 \geq \{\Sigma(n_j^{-1}\sqrt{\theta_j^2+\sigma_j^2}.n_j^{\frac{1}{2}})\}^2 = (\Sigma\sqrt{\theta_j^2+\sigma_j^2})^2$$

for all $n_1, n_2, \ldots, n_M$, where the sign of equality can hold if and only if

$$n_j^{\frac{1}{2}} = \lambda\, n_j^{-\frac{1}{2}}\sqrt{\theta_j^2+\sigma_j^2}$$

i.e. if
$$n_j = \lambda\sqrt{\theta_j^2+\sigma_j^2}. \qquad \ldots \quad (3.8)$$

Now if σ_j^2 be small compared with θ_j^2 which we can assume to be so if the sampling proportions for each of the primary units u_j be fairly large, then in order to make the variance of l^* small we should choose the probabilities n_j's to be nearly proportional to the unknown constants θ_j's. For instance, if θ_j be the total agricultural income in the j-th district u_j then perhaps we can choose n_j proportional to the total area of the land under cultivation in u_j (if information on this is available) or to the total agricultural population of u_j or some other like characteristic on which we have past information.

4. Sampling without replacement

The deduction in the earlier section was very much simplified because the m primary units were chosen with replacement which made the t_i^*'s identically distributed and independent of one another. Here we consider the case of without replacement but in the particular situation where

$$n_1 = n_2 = \ldots = n_M = 1/M. \qquad \ldots \quad (4.1)$$

We get the set of chance variables $t_1^*, \ldots, t_m^*$ and $s_1^{*2}, \ldots, s_m^{*2}$ as in (3.1); (here the n_j's are all equal to $1/M$) but clearly now they are not independent of one another.

Let
$$F^*(\tau_1, \tau_2, \ldots, \tau_m) = P(t_1^* \leq \tau_1, \ldots, t_m^* \leq \tau_m) \qquad \ldots \quad (4.2)$$

be the distribution function of the chance vector $(t_1^*, \ldots, t_m^*)$ and let $F_j(\tau)$ be the d.f. of t_j. The probability that $u_1^*, u_2^*, \ldots, u_m^*$ are the same as $u_{j_1}, u_{j_2}, \ldots, u_{j_m}$ is equal to $m!/M!$ and is the same for all distinct sets $j_1, j_2, \ldots, j_m$ of integers from the set $1, 2, \ldots, M$. And in this situation the r.h.s. of (4.2) is equal to

$$P(Mt_{j_1} \leq \tau_1, Mt_{j_2} \leq \tau_2, \ldots, Mt_{j_m} \leq \tau_m)$$
$$= F_{j_1}\left(\frac{\tau_1}{M}\right) F_{j_2}\left(\frac{\tau_2}{M}\right) \ldots F_{j_m}\left(\frac{\tau_m}{M}\right).$$
$$\therefore\ F^*\left(\tau_1, \tau_2, \ldots, \tau_m\right) = \frac{m!}{M!}\sum F_{j_1}\left(\frac{\tau_1}{M}\right) \ldots F_{j_m}\left(\frac{\tau_m}{M}\right)$$

where the summation is to be taken over all sets of distinct integers $(j_1, j_2, \ldots, j_m)$ from the set $1, 2, \ldots, M$. Clearly, therefore, F^* is a symmetric function of $\tau_1, \tau_2, \ldots, \tau_m$. Similarly the joint d.f. of $s_1^{*2}, s_2^{*2}, \ldots, s_m^{*2}$ is a symmetric function. Hence it follows

(Basu, 1952) that, whenever we are working with convex loss functions, for any function of the $t_i^{\bullet}$'s and $s_i^{\bullet 2}$'s to be an admissible estimator it is necessary that the estimator be symmetric in the $t_i^{\bullet}$'s as well as in the $s_i^{\bullet 2}$'s. Corresponding to any unbiassed non-symmetric estimator there always exists an unbiassed symmetric estimator with a uniformly smaller risk function.

It at once follows that the best unbiassed linear estimator of θ is

$$t^{\bullet} = \frac{1}{m}(t_1^{\bullet} + \ldots + t_m^{\bullet}). \qquad \ldots (4.3)$$

Now, because of symmetry of $F^{\bullet}$ we have

$$V(t^{\bullet}) = \frac{1}{m} V(t_1^{\bullet}) + \frac{m-1}{m} \mathrm{Cov}(t_1^{\bullet}, t_2^{\bullet}). \qquad \ldots (4.4)$$

The conditional expectation of $t_1^{\bullet} t_2^{\bullet}$ when $u_1^{\bullet}$ is u_j and $u_2^{\bullet}$ is $u_{j'}$ is $M^2 \theta_j \theta_{j'}$ $(j \neq j')$.

$$\therefore \quad E(t_1^{\bullet} t_2^{\bullet}) = \frac{M^2}{M(M-1)} \Sigma \theta_j \theta_{j'}$$

and hence

$$\mathrm{Cov}(t_1^{\bullet}, t_2^{\bullet}) = \frac{M}{M-1} \Sigma \theta_j \theta_{j'} - \theta^2. \qquad \ldots (4.5)$$

From (4.4), (4.5) and (3.4) we have

$$\begin{aligned} V(t^{\bullet}) &= \frac{1}{m}\left[M \Sigma (\theta_j^2 + \sigma_j^2) - \theta^2 \right] + \frac{m-1}{m}\left[\frac{M}{M-1} \Sigma \theta_j \theta_{j'} - \theta^2 \right] \\ &= \frac{M-m}{m} \sum \theta_j^2 - \frac{M-m}{m(M-1)} \sum \theta_j \theta_{j'} + \frac{M}{m} \sum \sigma_j^2. \qquad \ldots (4.6) \end{aligned}$$

We have to set up an unbiassed estimator for (4.6). In the class of all unbiassed estimators of the form

$$T = \Sigma a_i t_i^{\bullet 2} + \sum_{i \neq i'} b_{ii'} t_i^{\bullet} t_{i'}^{\bullet} + \Sigma c_i s_i^{\bullet 2}$$

the best estimator must be (because of symmetry) of the form

$$T_0 = a \Sigma t_i^{\bullet 2} + b \Sigma t_i^{\bullet} t_{i'}^{\bullet} + c \Sigma s_i^{\bullet 2}.$$

It is clear that to T_0 we cannot add terms like $d \Sigma t_i^{\bullet}$ or $e \Sigma t_i^{\bullet} s_{i'}^{\bullet}$, etc., for then we cannot hope to make T_0 unbiassed unless the multiplying constants d, e, etc. be zeros. Thus although to T we can add terms like $\Sigma d_i t_i^{\bullet}$, where because of unbiassedness we put $\Sigma d_i = 0$, the corresponding symmetric function must lack the term $\Sigma t_i^{\bullet}$.

Now

$$E(T_0) = amM \Sigma (\theta_j^2 + \sigma_j^2) + bm(m-1) \frac{M}{M-1} \Sigma \theta_j \theta_{j'} + cm \Sigma \sigma_j^2. \qquad \ldots (4.7)$$

From (4.6) and (4.7) we have at once

$$a = \frac{M-m}{m^2M},\ b = -\frac{M-m}{m^2M(m-1)} \quad \text{and} \quad c = \frac{1}{m},$$

$$\therefore\ T_0 = \frac{M-m}{m^2M}\Big[\sum t_i^{*2} - \frac{1}{m-1}\sum t_i^* t_{i'}^*\Big] + \frac{1}{m}\sum s_i^{*2}$$

$$= \Big(\frac{1}{m} - \frac{1}{M}\Big)s^2(t^*) + \frac{1}{m}\sum s_i^{*2}$$

where $s^2(t^*)$ is given by (3.6).

5. Unequal probabilities

In this section we consider the procedure (sometimes employed in practice) where the primary sample units u_i^*'s are chosen with replacement and with varying probabilities $\pi_1, \pi_2, \ldots, \pi_M$ but where only one observation is taken on a particular t_j irrespective of whether u_j comes in the sample $u_1^*, \ldots, u_m^*$ once or more than once. The difference between this procedure and that considered in section 3 is only this that in the earlier procedure if a particular primary unit u_j were appearing in the sample r times then we were taking r independent observations on the corresponding chance variable t_j whereas we now take one observation only and repeat the same r times.

As in section 3 we get the m chance variables $t_1^*, \ldots, t_m^*$. Now the t_i^*'s are clearly not independent of one another although their marginal distributions are the same as in section 3. The joint distribution of the t_i^*'s is easily seen to be a symmetric one. All the above statements also hold for the set of variables $s_1^{*2}, \ldots, s_m^{*2}$. Hence it follows that the best unbiassed linear estimator for θ is still the sample mean $\bar{t}^*$. Let us now compute the variance of $\bar{t}^*$

$$V(\bar{t}^*) = E(\bar{t}^{*2}) - \theta^2 = \frac{1}{m}E(t_1^{*2}) + \frac{m-1}{m}E(t_1^* t_2^*) - \theta^2. \qquad \ldots\ (5.1)$$

As in (3.4)
$$E(t_1^{*2}) = \Sigma \pi_j^{-1}(\theta_j^2 + \sigma_j^2). \qquad \ldots\ (5.2)$$

Now the conditional expectation of $t_1^* t_2^*$ in the situation where u_1^* and u_2^* both happen to be the same unit u_j (the probability for which is clearly π_j^2) is

$$E(t_1^* t_2^* \mid u_1^* = u_2^* = u_j) = E(\pi_j^{-1} t_j)^2 = \pi_j^{-2}(\theta_j^2 + \sigma_j^2).$$

Similarly $E(t_1^* t_2^* \mid u_1^* = u_j,\ u_2^* = u_{j'}) = E(\pi_j^{-1} t_j\ \pi_{j'}^{-1} t_{j'}) = (\pi_j \pi_{j'})^{-1}\theta_j\theta_{j'} \quad (j \neq j')$.

$$\therefore \qquad E(t_1^* t_2^*) = \Sigma\,(\theta_j^2 + \sigma_j^2) + \sum_{j \neq j'} \theta_j\theta_{j'} = \theta^2 + \sigma^2 \qquad \ldots\ (5.3)$$

where $\theta = \Sigma\theta_j$ and $\sigma^2 = \Sigma\sigma_j^2$.

From (5.1), (5.2) and (5.3) it follows that

$$V(\bar{t}^*) = \frac{1}{m}\sum \pi_j^{-1}(\theta_j^2 + \sigma_j^2) - \frac{1}{m}\theta^2 + \frac{m-1}{m}\sigma^2 \qquad \ldots\ (5.4)$$

$$= \frac{1}{m}V(t_1^*) + \frac{m-1}{m}\sigma^2.$$

As in (3.8) we have that $V(l)$ is minimum if π_j be chosen proportional to $\sqrt{\theta_j^2+\sigma_j^2}$ $(j = 1, 2, \ldots, M)$ and that the minimum value is

$$\begin{aligned}\text{min. } V(l^*) &= \frac{1}{m}\left(\Sigma\sqrt{\theta_j^2+\sigma_j^2}\right)^2-\frac{1}{m}\theta^2+\frac{m-1}{m}\sigma^2 \\ &= \frac{1}{m}\left[(\Sigma\sqrt{\theta_j^2+\sigma_j^2})^2-\theta^2-\sigma^2\right]+\sigma^2 \\ &> \sigma^2\end{aligned}$$

because

$$\begin{aligned}(\Sigma\sqrt{\theta_j^2+\sigma_j^2})^2 &= \Sigma(\theta_j^2+\sigma_j^2)+\Sigma\sqrt{\theta_j^2+\sigma_j^2}\sqrt{\theta_{j'}^2+\sigma_{j'}^2} \\ &> \Sigma\theta_j^2+\Sigma\sigma_j^2+\Sigma\theta_j\theta_{j'} \\ &= \theta^2+\sigma^2.\end{aligned}$$

Thus we find that $V(l^*)$ is always greater than σ^2 and that we can bring it arbitrarily near to σ^2 by taking m sufficiently large. Here l^* is not a consistent estimator. The sampling proportions for each of the M primary units $u_1, u_2, \ldots, u_M$ should be set at such high values that $\sigma^2 = \Sigma\sigma_j^2$ is relatively small and then the question of deciding upon a suitable value for m should be considered. All the above remarks also hold for the sampling procedure considered in section 4 where by the very nature of the scheme m cannot be greater than M.

Consider now the problem of estimating $V(l^*)$. In the class of quadratic estimators we must (for the sake of symmetry) confine ourselves to the class

$$T_0 = a\Sigma l_i^{*2}+b\sum_{(i\neq i')} l_i^* l_{i'}^* + c\Sigma s_i^{*2}.$$

From (5.2), (5.3) and (3.3) we have

$$E(T_0) = am\,\Sigma\pi_j^{-1}(\theta_j^2+\sigma_j^2)+bm(m-1)\theta^2+\{bm(m-1)+cm\}\sigma^2. \qquad \ldots \quad (5.5)$$

From (5.4) and (5.5) we have

$$a = \frac{1}{m^2}, \quad b = -\frac{1}{m^2(m-1)} \quad \text{and} \quad c = \frac{1}{m}$$

so that

$$\begin{aligned}T_0 &= \frac{1}{m^2}\left(\sum l_i^{*2} - \frac{1}{m-1}\sum l_i^* l_{i'}^*\right)+\frac{1}{m}\sum s_i^{*2} \\ &= \frac{1}{m}\,s^2(l^*) + \frac{1}{m}\sum s_i^{*2}\end{aligned}$$

where $s^2(l^*)$ is defined as in (3.6).

REFERENCES

BASU, D. (1952): On symmetric estimators in point estimation with convex weight functions. *Sankhyā*, 12, 45.

DEMING, W. E. (1950): *Some Theory of Sampling*, John Wiley & Sons, New York.

Paper received: August, 1952.