ON 'HORVITZ AND THOMPSON ESTIMATOR'

By T. V. HANURAV
Indian Statistical Institute

SUMMARY. In this paper we consider the problem of finding optimal sampling designs for the use of 'Horvitz and Thompson' estimator (1952) to estimate (unbiasedly, unless otherwise stated) the population total of a character Y, when auxiliary information on a correlated character X is available for all the units. Since there does not exist a design in which the variance is uniformly minimum, the optimal designs are obtained by minimising the expected variance under a realistic superpopulation set-up. These turn out to be designs in which the effective sample size is constant for all samples of the design. It is further proved that with the same criterion for comparison, the Horvitz-Thompson estimator for these optimal class of designs is uniformly superior to Des Raj's (1956) estimator in the symmetrised form for the sampling designs considered by him when the average effective sample size is 2.

INTRODUCTION

Consider a finite population, consisting of N units

$$u_1, u_2, \ldots, u_N. \qquad \ldots \quad (1.1)$$

Let Y be a quantitative character, taking the value y_i (which is unknown, *a priori*) on u_i, $(1 \leqslant i \leqslant N)$. Let $D = D(S, P)$ be a *sampling design* consisting of a set S of samples s from (2.1), with a probability measure P defined on it. We define

$$\pi_i = \sum_{s \supset u_i} P_s, \qquad (1 \leqslant i \leqslant N) \qquad \ldots \quad (1.2)$$

(where the summation extends over all samples s of S that contain u_i), to be the *inclusion probability* of u_i in D. Similarly, we define

$$\pi_{ij} = \sum_{s \supset u_i, u_j} P_s, \qquad (1 \leqslant i \neq j \leqslant N) \qquad \ldots \quad (1.3)$$

to be the inclusion probability of the pair (u_i, u_j). An unbiased estimator of the population total

$$T = \sum_{i=1}^{N} y_i \qquad \ldots \quad (1.4)$$

as proposed by Horvitz and Thompson, is then given by

$$\hat{T}_1 = \sum_{i \epsilon s} \frac{y_i}{\pi_i} \qquad \ldots \quad (1.5)$$

where the summation extends over all *distinct* units u_i belonging to s (i.e., we ignore repetitions). The variance of T_1 is given by

$$V(\hat{T}_1) = \sum_{i=1}^{N} y_i^2 \left(\frac{1-\pi_i}{\pi_i} \right) + \sum_{i \neq j}^{N}\sum y_i y_j \left\{ \frac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j} \right\}. \qquad \ldots \quad (1.6)$$

In many situations of practical interest, we have *a priori*, the values x_i, which another quantitative character X, highly correlated with Y, takes on u_i (for $1 \leqslant i \leqslant N$). In such cases, this information is used to construct sampling designs and estimators of, T say, which result in a greater precision than those which do not make use of this information. Examples of such procedures are pps estimator, Des Raj's estimator, ratio estimator, regression estimator and so on. The amount of gain in precision due to these estimators depends on the degree of correlation between X and Y, and to assess this more fully, one needs to assume some broad statistical relationship between X and Y, so far as the population (1.1) is concerned.

When the value x_i is known for the unit u_i, it is reasonable to assume, in many practical cases, that the corresponding y_i (which, however, is unknown), is the outcome of a random variable Z whose expectation is proportional to x_i and whose variance is either partly or fully unknown. The realised value $y = (y_1, ..., y_N)$ can thus be considered as the realisation of N-length random vector from a super-population. This concept has been introduced by Cochran (1946) and since then has been successfully used by many others. We notice here that we tacitly make this assumption when dealing with pps estimator, ratio estimator, and almost all estimators used in designs of varying probability sampling, while when using regression estimator we make a similar but slightly weaker assumption. We explicitly formulate our model, writing E_1 and V_1 to denote the conditional expectation and variance, given x_i's, thus :

$$E_1(y_i/x_i) = a\, x_i \qquad \dots \quad (1.7)$$

and

$$V_1(y_i/x_i) = \sigma_i^2 \quad \text{(say)} \qquad \dots \quad (1.8)$$

where, a and σ_i^2's are unknown constants. We also assume that y_i and y_j are independent for all $i \neq j$, for given x_i and x_j. In particular this implies that

$$E_1(y_i y_j/x_i \text{ and } x_j) = a^2 x_i x_j. \qquad \dots \quad (1.9)$$

2. Optimal designs

Under the assumptions (1.7), (1.8) and (1.9), we shall proceed to find the sampling designs best suited for the use of $\hat{T}_1$ as given by (1.5), to estimate T. The criteria that we choose for the best are (1) unbiasedness and (2) minimum variance. Clearly, the increase of sample size, while increasing the precision of estimates, increases the cost also. Assuming that the cost of drawing and inspecting a sample s is proportional to the number of distinct units in the sample (which we shall call the 'effective size' of s and denote by $\nu(s)$, henceforth), we search for the best designs to use (1.5) in the class of all designs having a fixed given value for $\nu(s)$ for all s in the design. However, we may relax the later condition to include designs in which $\nu(s)$ can vary from sample to sample by demanding that the expected value of $\nu(s)$ be equal to a given value. This means that the expected cost of sampling is to be fixed, which is a reasonable condition unless the variation of $\nu(s)$ in our optimal design is too large. We note that

$$E\nu(s) = \sum_{s \in S} \nu(s) P_s = \sum_{i=1}^{N} \pi_i = \nu_0 \quad \text{say.} \qquad \dots \quad (2.1)$$

Given the auxiliary information on X, we shall consider only those designs for which the inclusion probability π_i is proportional to $x_i (1 \leqslant i \leqslant N)$. (This is not only reasonable but is probably better as hinted in Section 3). Given the expected cost, ν_0 is fixed and our domain of search then becomes the class of all designs in which

$$\pi_i = c x_i \qquad (1 \leqslant i \leqslant N) \qquad \dots \quad (2.2)$$

where c is a constant determined by (2.1) thus

$$c = \frac{\nu_0}{\sum\limits_{i=1}^{N} x_i} = \frac{\nu_0}{M}, \quad \text{say.} \qquad \dots \quad (2.3)$$

It is seen that in these designs, the variance of (1.5), as given by (1.6) depends on y_i's and π_{ij}'s. Hence it is the set of π_{ij}'s that are at our disposal in the choice of optimal designs. If there exists a subclass of the above designs, for which (1.6) is minimum, uniformly with respect to all possible values of $y = (y_1, ..., y_N)$, clearly, we have obtained the best designs. But we prove that (1.6) cannot be uniformly minimised with respect to y_i's. Setting

$$Q = \sum_{i=1}^{N} y_i^2 \left(\frac{1-\pi_i}{\pi_i} \right) + \sum\sum_{i \neq j}^{N} \frac{y_i}{\pi_i} \frac{y_j}{\pi_j} (\pi_{ij} - \pi_i \pi_j).$$

which is continuous in y 's the conditions for Q to be minimum require

$$\frac{\delta Q}{\delta y_i} = 2y_i \frac{(1-\pi_i)}{\pi_i} + \sum_{\substack{j=1 \\ j \neq i}}^{N} y_j \frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j} = 0$$

for all i and for all sets of y_i's. Clearly, this is violated by setting $y_i = 1$ for any fixed i and $y_j = 0$ for $j \neq i$. Hence we proceed to do the next best, which is to see whether the expected value of (1.6), when conditioned over given x_i's can be minimised, uniformly with respect to all values of x_i's, a and σ_i^2's. When even this is not possible, we then have to restrict our population of x_i's to some specific models. However, we shall prove that there exists an optimum class of designs with given π_i's for which the conditional expectation of (1.6) is uniformly minimised. From (1.7), (1.8), (1.9), (2.2) and (2.3), we have

$$E_1 V(\hat{T}_1) = \sum_{i=1}^{N} \left(\frac{1-\pi_i}{\pi_i}\right)(a^2 x_i^2 + \sigma_i^2) + \sum\sum_{i \neq j}^{N} \left(\frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j}\right) a^2 x_i x_j$$

$$= \sum_i \left(\frac{1-\pi_i}{\pi_i} \right) \left(\frac{a^2 \pi_i^2}{c^2} + \sigma_i^2 \right) + \frac{a^2}{c^2} \sum\sum_{i \neq j} (\pi_{ij} - \pi_i \pi_j). \qquad \dots \quad (2.4)$$

Hence $E_1 V(\hat{T}_1)$ is minimum, for given values of π_i's, a, c and σ_i^2's, when and only when

$$\sum\sum_{i \neq j} \pi_{ij} \qquad \dots \quad (2.5)$$

is minimum. Defining auxiliary random variables R_{si} by

$$R_{si} = \begin{cases} 1 & \text{if } u_i \epsilon s \\ 0 & \text{otherwise} \end{cases}$$

we have

$$\pi_i = \sum_{s \epsilon S} R_{si} P_s$$

$$\pi_{ij} = \sum_{s \epsilon S} R_{si} R_{sj} P_s$$

and $$\nu(s) = \sum_{i=1}^{N} R_{si} = \sum_{i} R_{si}^2.$$

Hence $$\sum\sum \pi_{ij} = \sum\sum_{i \neq j} \{ \sum_{s} R_{si} R_{sj} P_s \}$$
$$= \sum_{s \in S} \{ \sum^{r}\sum_{i \neq j} R_{si} R_{sj} P_s \}$$
$$= \sum_{s \in S} \{\nu^2(s) - \nu(s)\} P_s$$
$$= \nu_0^2 + V(\nu(s)) - \nu_0. \qquad \dots (2.6)$$

Hence, for a given ν_0, the expected sample size, (2.6) is minimised when $V(\nu(s))$ is minimised. In other words, the design should contain, as far as possible, samples of the same effective size. When ν_0 is an integer

$$\min \; V(\nu(s)) = 0$$

while if $$\nu_0 = [\nu_0] + f, \qquad 0 < f < 1,$$

$[\nu_0]$ being the greatest integer not greater than ν_0,

$$\min \; V(\nu(s)) = f(1-f) \qquad \dots (2.7)$$

since we should have

$$\nu(s) = [\nu_0] \quad \text{with probability} \quad f$$
$$[\nu_0]+1 \text{ with probability } 1-f.$$

In practice, the considerable degree of freedom we have in the choice of ν_0 allows us to have ν_0 as an integer. Even otherwise, (2.7) is negligible in comparison with $(\nu_0^2 - \nu_0)$ for practical values of ν_0.

The practical problem of the actual construction of these designs is not solved fully, but the author (Hanurav, 1962)* gave a method of selecting units from (1.1) which results in prescribed general values of π_i's. It is pointed therein that the resulting design has a very stable value of $\nu(s)$, serving as a good approximation for our purpose. We therefore assume that the minimum of (2.6), viz., $(\nu_0^2 - \nu_0)$ can be closely attained in practice, so that for purposes of comparison we can take

$$\min \; E_1 V(\hat{T}_1) = \frac{a^2}{c^2} \sum_{i=1}^{N} \pi_i(1-\pi_i) + \sum_{i=1}^{N} \left(\frac{1-\pi_i}{\pi_i}\right) \sigma_i^2 + \frac{a^2}{c^2}(\nu_0^2 - \nu^2 - \sum\sum_{i \neq j} \pi_i \pi_j)$$
$$= \frac{a^2}{c^2}\left(\nu_2 - \sum \pi_i^2\right) + \sum \left(\frac{1-\pi_i}{\pi_i}\right)\sigma_i^2 + \frac{a^2}{c^2}\{\nu_0^2 - \nu_0 - (\nu_0^2 - \sum \pi_i^2)\}$$
$$= \sum\left(\frac{1-\pi_i}{\pi_i}\right)\sigma_i^2 \qquad \dots (2.8)$$

*Published in the same issue, pp. 421-428.

3. Comparison with symmetrised Des Raj's estimator

For the purpose of this section, we shall make one further assumption beside (1.7) and (1.9), and this is regarding the conditional variances σ_i^2's. A commonly advocated assumption is that σ_i^2's are all equal but unknown. However, in many cases of practical interest (especially when the variates Y and X are positive as is the case in most of the sample surveys) it is more realistic to assume that the conditional coefficients of variation are more or less same for all units so that the conditional variance increases with the conditional mean. We explicitly write this as

$$V_1(y_i/x_i) = K.\, a^2x_i^2 = \sigma^2x_i^2 \qquad \ldots \quad (3.0)$$

where σ^2 is an unknown constant.

We now compare the estimator $\hat{T}_1$ used in the optimal class of designs derived above, with the symmetrised Desraj's (1956) estimator, when $\nu_0 = 2$, under the assumptions (1.7), (1.9) and (3.0). We briefly describe this later estimator.

We draw a sample size n say, without replacement. At each draw, the probability p_i of selecting u_i is proportional to x_i, if it is not already selected. Here p_i will be taken proportional to x_i, where x_i has the same meaning as in Sections 1 and 2. If u_i is selected in the first draw, the probability of selecting u_j in the second draw is

$$p^{(2)}(i) = \begin{cases} \dfrac{p_j}{1-p_i} & \text{if } j \neq i \\ 0 & \text{if } j = i. \end{cases}$$

Similarly, in the third draw we have

$$p_k^{(3)}(i, j) = \begin{cases} \dfrac{p_k}{1-p_i-p_j} & \text{if } \; k \neq i, k \neq j \\ 0 & \text{otherwise} \end{cases}$$

where the notations are clear. An unbiased estimator of T in such cases, as given by Desraj (1956) is

$$\hat{T}_{2,\,asym} = \frac{1}{n} \sum_{s=1}^{n} t_s \qquad \ldots \quad (3.1)$$

where $$t_s = y_{i_1}+y_{i_2}+\ldots+y_{i_{s-1}}+\frac{y_{i_s}}{p_{i_s}}(1-p_{i_1}-p_{i_2}\ldots-p_{i_{s-1}}),$$

$u_{i_1}, u_{i_2}, \ldots, u_{i_{n-1}}, u_{i_n}$ being the units successively obtained in the sample, the suffix "asym" denoting that the estimator is asymmetric in the observed values. Restricting ourselves to an important practical case of $n = 2$, we write (3.1) thus

$$\hat{T}_{2,\,asym} = y_i+\frac{y_j}{p_j}\,(1-p_i) \qquad \ldots \quad (3.2)$$

where u_i is selected in the first draw and u_j in the second draw. It is well known that this estimator can always be improved by taking the weighted mean of different asymmetric estimators, for given unordered sample, the weights being the respective conditional probabilities of obtaining the ordered samples given the unordered sample (Halmos, 1946). Denoting this improved symmetric estimator by $\hat{T}_2$, we have

$$\hat{T}_2 = \frac{1}{2-p_i-p_j}\left\{(1-p_j)+\frac{y_i}{p_i}\;(1-p_i)\frac{y_j}{p_j}\right\} \qquad \ldots \quad (3.3)$$

which is symmetric in i and j as it should be. We have

$$V(\hat{T}_2) = \sum_{(i\neq j)}\sum_{1}^{N} p_ip_j\left(\frac{1-p_i-p_j}{2-p_i-p_j}\right)\left[\frac{y_i}{p_i}-\frac{y_j}{p_j}\right]^2$$

$$= \sum_{i=1}^{N}\frac{y_i^2}{p_i}\left\{\sum_{j\neq i} p_j\frac{(1-p_i-p_j)}{(2-p_i-p_j)}\right\} - \sum_{(i\neq j)}\sum_{1}^{N} y_iy_j\left\{\frac{1-p_i-p_j}{2-p_i-p_j}\right\}. \qquad \ldots \quad (3.4)$$

In order to compare $\hat{T}_1$ and $\hat{T}_2$ we should take π_i's for $\hat{T}_1$ such that $\nu_0 = 2$, so that the expected effective sample size remains the same in both cases. Since $\sum_{1}^{N} p_i = 1$, and both p_i's and π_i's are proportional to x_i's,

$$p_i = \frac{\pi_i}{2}, \qquad \ldots \quad (3.5)$$

so that from (1.7), (1.9), (2.8), (3.0), (3.4) and (3.5),

$$E_1V(\hat{T}_2) = \sum_{i=1}^{N}\frac{2}{\pi_i}\cdot\frac{\pi_i^2(a^2+\sigma^2)}{c^2}\left\{\sum_{j\neq i}\frac{\pi_j(2-\pi_i-\pi_j)}{2(4-\pi_i-\pi_j)}\right\} - \sum_{(i\neq j)}\sum\frac{a^2}{c^2}\,\pi_i\pi_j\left[\frac{2-\pi_i-\pi_j}{4-\pi_i-\pi_j}\right]$$

$$= \frac{\sigma^2}{c^2}\sum_{(i\neq j)}\sum \pi_i\pi_j\left[\frac{2-\pi_i-\pi_j}{4-\pi_i-\pi_j}\right]. \qquad \ldots \quad (3.6)$$

We assume that the minimum variance of $\hat{T}_1$ as given by (2.8) can be closely attained and that we can neglect the component of variance arising due to the slight variations in the effective sample size. Hence we shall proceed to compare (3.6) and (2.8). We have

$$E_1V(\hat{T}_2) - \min E_1V(\hat{T}_1)$$

$$= \frac{\sigma^2}{c^2}\left[\sum_{(i\neq j)}\sum\frac{\pi_i\pi_j(2-\pi_i-\pi_j)}{(4-\pi_i-\pi_j)} - 2 + \sum \pi_i^2\right]$$

$$= \frac{2\sigma^2}{c^2}\left[1-\sum_{(i\neq j)}\sum\frac{\pi_i\pi_j}{(4-\pi_i-\pi_j)}\right]$$

$$= \frac{2\sigma^2}{c^2}\left[1-2\sum_{(i\neq j)}\sum\frac{p_ip_j}{(2-p_i-p_j)}\right]. \qquad \ldots \quad (3.7)$$

Let $$\chi(p_1, \ldots, p_N) = \sum_{i \neq j}\sum \frac{p_i p_j}{(2-p_i-p_j)}.$$

When $$p_i = \frac{1}{N} \text{ for all } i,$$

clearly $$\chi\left(\frac{1}{N}, \frac{1}{N}, \ldots, \frac{1}{N}\right) = \frac{1}{2},$$

so that in this case

$$E_1 V(\hat{T}_3) = \min E_1 V(\hat{T}_1).$$

In order to prove that

$$E_1 V(\hat{T}_3) \geqslant \min E_1 V(\hat{T}_1),$$

we shall prove that χ is actually maximum when all its arguments are equal to $1/N$. We have the restriction

$$\Sigma p_i = 1$$

on the p_i's. Introducing the Lagrangian multiplier λ,

let $$\psi = \sum_{i \neq j}\sum \frac{p_i p_i}{(2-p_i-p_j)} - \lambda(\Sigma p_i - 1).$$

We can verify that at the point where $p_i = \frac{1}{N}$ for $1 \leqslant i \leqslant N$, we have

$$\frac{\delta \psi}{\delta p_i} = 0$$

$$\frac{\delta^2 \psi}{\delta p} = \frac{-N(2N-1)}{4(N-1)^2} = -b_1 \quad \text{say}$$

$$\frac{\delta^2 \psi}{\delta p_i \, \delta p_j} = \frac{N N^2 + (N-1)^2}{4(N-1)^3} = b_2 \quad \text{say}$$

so that the Hessian of ψ is given by the $N \times N$ determinant

$$H(\psi) = \begin{vmatrix} -b_1 & b_2 & b_2 & \ldots & b_2 \\ b_2 & -b_1 & b_2 & \ldots & b_2 \\ . & . & . & \ldots & . \\ b_2 & b_2 & b_2 & \ldots & -b_1 \end{vmatrix}$$

The value of the r-th order principal minor of $H(\psi)$ is

$$(-1)^{r-1}(b_1+b_2)^{r-1}[(r-1)b_2-b_1]. \qquad \ldots \quad (3.8)$$

Since $b_2 < b_1$, it follows that (3.8) changes its sign alternately as r increases and that it is negative when $r = 1$. This shows that χ attains its maximum when all p_i's are equal to $1/N$ and hence it follows that

$$\min\ E_1 V(\hat{T}_1) \leqslant E_1 V(\hat{T}_2)$$

which shows that when we average over the conditional variations of y_i's, $\hat{T}_1$ is uniformly superior to the symmetrised Desraj's estimator, when samples are of average effective size 2.

Remark : The above result justifies the opinion that when auxiliary information X of the type discussed above is available, it is preferable to choose the sampling scheme so as to make the inclusion probabilities π_i's proportional to x_i's instead of choosing the design with probability of selection in each draw proportional to x_i's. We note the assumption involved in taking (2.8) to be the minimum attainable variance when we use (1.5). This amounts to assuming that the given set of π_i's can be partitioned into two subsets such that in each subset, the total of the π_i's is exactly equal to unity. This assumption though need not hold good in general is a good approximation in practical cases especially when π_i's are small quantities as is usually the case. It also seems reasonable to conjecture the validity of our result even when $\nu_0 > 2$.

Acknowledgement

The author wishes to express his thanks to Dr. J. Roy for his valuable suggestions.

References

Cochran, W. G. (1946) : Relative accuracies of systematic and stratified random samples for a certain class of population. *Ann. Math. Stat.*, 17, 164-179.

Des Raj (1956) : Some estimators in sampling with varying probability without replacement. *J. Amer. Stat. Ass.*, 51, 269-284.

Halmos, P. R. (1946) : The theory of unbiased estimation. *Ann. Math. Stat.*, 17, 34-43.

Hanurav, T. V. (1962) : Some sampling schemes in probability sampling. *Sankhyā*, Series A, 24, 421-428.

Horvitz, D. G. and Thompson, D. J. (1952) : A generalization of sampling without replacement from a finite universe. *J. Amer. Stat. Ass.*, 47, 663-685.

Paper received : September, 1961.

Revised : October, 1962.

ACKNOWLEDGEMENT

Thanks are due to the following for their kind help in the editing of the papers published in this issue.

Shri S. John — Dr. C. R. Rao

Shri M. N. Murthy — Dr. J. Roy

Dr. J. Sethuraman

Editor,

Sankhyā : The Indian Journal of Statistics.