NECESSARY BEST ESTIMATORS IN CERTAIN CLASSES OF LINEAR ESTIMATORS FOR IKEDA-SEN SAMPLING

By

A. K. Yogi* and P. C. Gupta**

(Received: November, 1975)

1. Introduction

Let a population consist of a finite number N of units. For estimating the population total

$$T=\sum_{i=1}^{N} x_i,$$

Horvitz & Thompson [6] introduced three classes of linear estimators. Godambe [4] formulated his most general class of linear estimator in addition to Horvitz–Thompson's estimators and proved the non-existence of Minimum variance Linear Unbiased estimator (MVLUE) in his most general class. Koop [7] following the logic of Horvitz–Thompson gave an axiomatic system of sample formation and defined seven classes of linear estimators, within which almost all the linear unbiased estimators of the population total lie. He examined the Godambe's formulation and commented on page 91 of his paper as follows:

"Godambe [4] formulated four classes of estimators of T *viz.* what in this system of notations turns out to be T_1, T_2, T_6, and T_5 (his most general estimator) without posting the three features of sample formation as axioms...The advantage of the systematic approach, having its being in the axioms, is that three more classes, including T_7 whch is more general than Godambe's T_5 have been brought to light'.

Further Koop gave a proof that the best linear unbiased estimator does not exist in the most general class (T_7) of linear estimators.

Thus while estimating the population total, if the MVLUE does not exist the use of the 'necessary best estimators' (NBE) was

* Deputy Director, N.B.O. Ministry of Works & Housing, New Delhi.

** Reader in South Gujarat University, Surat.

recommended by Prabbu–Ajgaonkar [1], as serviceable estimators purely from the practical point of view. He [2] defined the higher order NBE's as follows:

'Let C be a class linear unbiased estimators and t' belong to C, then it is easily seen that the variance of t' can be expressed in the following form:

$$\text{Var}\,(t')=\sum_{i=1}^{N} A_{ii}\, Y_i^2+\sum_{i\neq j=1}^{N} A_{ij}\, Y_i Y_j \qquad \ldots(1.1)$$

Where the quantities A_{ij} $(i, j=1, 2\ldots\ldots N)$ involve the known functions of probabilities and coefficients of the class.

Further let t'_1 be another unbiased estimator belonging to the class C and that the variance of t'_1 be given by

$$\text{Var}\,(t'_1)=\sum_{i=1}^{N} b_{ii}\, Y_i^2+\sum_{i\neq j=1}^{N} b_{ij}\, Y_i\, Y_j \qquad \ldots(1.2)$$

consider the quantity

$$Q=\text{Var}\,(t')-\text{Var}\,(t')$$

$$=\sum_{i=1}^{N}(A_{ii}-b_{ii})\, Y_i^2+\sum_{i\neq j=1}^{N}(A_{ij}-b_{ij})\, Y_i\, Y_j$$

obtained from equations (1.1) and (1.2)

Definition: The estimator t_1' is a necessary best estimator of order for the class C if for every t' in C principal minors of expressions of the form Q upto the order r, are positive. Prabhu-Ajgaonkar [2] says "The necessary best estimator of the first order minimises the variance of the population for which $Y_1\neq0$ and $Y_i=0, i=2.3\ldots N$. Similarly a necessary best estimator of the second order is obtained by minimising $V(t)$ for the population vector for which $Y_1\neq0$, $Y_2=0$ but $Y_1=0$ for $i=3,4\ldots\ldots N$. Further he proved the existence of second order NBE in the most general class and NBE is the same as that of Horvitz-Thompson estimator; which is infact not true as remarked by Rao [9]. This has been established by Yogi and Gupta [11] utilising the approach of obtaining first and second order NBE of the author (1969) and further proved the non-existence of second and higher order NBE in T_3, T_4, T_5 and T_6 classes of linear unbiased estimators.

Prabhu-Ajgaonkar [3], while considering the Ikeda Sen system of sampling proved the existence of NBE of order one in T_2 and T_4 classes of linear estimators. In the present paper the authors have investigated the non-existence of second and higher order NBE's in T_4, T_5 and T_6 classes of linear estimators when Ikeda-Sen (1952) system of sampling is undertaken, following the same approach as in the author's earlier paper (1969).

1. Ikeda-Sen Sampling Scheme

A sampling procedure was proposed by Ikeda, a student of Midzuno and the procedure was then generalised by Midzuno [8] Sen [10] independently gave the identical system of sampling and hence in view of these facts the present sampling system is referred to as Ikeda-Sen sampling Scheme. The scheme consists in drawing the first unit of the sample with probability proportional to size and the rest of the (n-1) units with equal probability without replacement from the remaining (N-1) units of the population.

In the above sampling procedure, the probability of drawing ith unit at rth draw and the probability of selecting ith unit at the rth draw and jth unit at the sth draw are respectively given by

$$P[(X_r=X_i)]=P_{ir}=\begin{cases} P_i & \text{for } r=1 \\ \dfrac{1-p_i}{N-1}=C_i & \text{for } r=2, 3,\ldots N \end{cases} \qquad \ldots(2.1)$$

$$P\ (X_r=X_i \text{ and } X_s=X_j$$

$$=p\ (i_r,\ j_s)=\begin{cases} \dfrac{p_i}{N-1} & \text{when } r=1 \text{ and } s\geqslant 2 \\ \dfrac{1-p_i-p_j}{(N-1)\ (N-2)} & \text{when } r,s\geqslant 2 \end{cases} \qquad \ldots(2.2)$$

We know that

$$S=\binom{N}{n}\ n! \qquad \ldots(2.3)$$

is the total number of ordered samples of size in from the population of size N in without replacement sampling. Then the probability of the 1th sample in which the ith ($i=1, 2,\ldots N$) unit appears at the first draw is given by

$$\frac{P_i}{\binom{N-1}{n-1}\ (n-1)\ !}=k_i \text{ (say)} \qquad \ldots(2.4)$$

Also we define the following

$$k=\binom{N-1}{n-1}(n=i)! \qquad \ldots(2.5)$$

is the number of samples in which the first place in the sample is fixed for ith $(i=1, 2, \ldots N)$ unit.

$$k'=\binom{N-2}{n-2}(n-1)! \qquad \ldots(2.6)$$

is the number of samples in which jth $(\neq i)$ unit appears at the first draw and the ith unit appears anywhere else in the sample.

3. Necessary Best Estimators in T_6 Class

The Ikeda-Sen estimator in this class for the population total T is defined as

$$T_6^M=\sum_{r=1}^{n} \beta_r^t x_r \qquad \ldots(3.1)$$

where β_r^t is the weight associated to the rth draw of the ith sample

$$E\left(T_6^M\right)=\sum_{i=1}^{N}\sum_{r=1}^{n}\sum_{t\supset i_r} \beta_r^t (x_r=x_i)\, k_i \qquad \ldots(3.2)$$

collecting the coefficients of each x_i, the expression in equation (3.2) can be written as,

$$x_1\Bigg[K_1\sum_{t\supset 1_1}^{K}\beta_1^t+k_2\Bigg(\sum_{t\supset(2_1, 1_2)}\beta_2^t+\sum_{t\supset 2_1, 1_3)}\beta_3^t+\ldots+\sum_{t\supset(2_1, 1_n)}\beta_n^t\Bigg)$$

$$+k_3\Bigg(\sum_{t\supset(3_1, 1_2)}\beta_2^t+\sum_{t\supset(3_1, 1_3)}\beta_3^t+\ldots+\sum_{t\subset(3_1, 1_n)}\beta_n^t\Bigg)$$

$$+\ldots+K_N\Bigg(\sum_{t\supset(N_1, 1_2)}\beta_2^t+\sum_{t\supset(N_1, 1_3)}\beta_3^t+\ldots+\sum_{t\supset(N_1, 1_n)}\beta_n^t\Bigg)\Bigg]$$

$$+x_2\Bigg[K_2\sum_{t\supset(2_1)}^{K}\beta_2^t+k_1\Bigg(\sum_{t\supset(1_1, 2_2)}\beta_2^t+\sum_{t\supset(1_1, 2_3)}\beta_3^t+\ldots+\sum_{t\supset(1_1, 2_n)}\beta_n^t\Bigg)$$

$$+\ldots+K_N\Bigg(\sum_{t\supset(N_1, 2_2)}\beta_2^t+\ldots+\sum_{t\supset(N_1, 2_n)}\beta_n^t\Bigg)\Bigg]$$

$$+x_n \Big[K_N \sum_{t\supset N_1}^{K} \beta_N^t + K_1 \Big(\sum_{t\supset(1_1, N_2)} \beta_2^t + \sum_{t\supset(1_1, N_3)} \beta_3^t + \ldots + \sum_{t\supset(11_1, N_n} \beta_n^t \Big)$$

$$+ \ldots + K_{N-1} \Big(\sum_{t\supset[(N-1)_1, N_2]} \beta_2^t + \ldots + \sum_{t\supset[(N-1)_1, N_n)} \beta_n^t \Big)\Big] \quad \ldots(5.3)$$

where the summation have the following meaning.

(*i*) $\sum_{r\supset i_1} \beta_1^t$

is the sum of all those weights associated to the first draw and to the ith sample in which ith unit appears at the first draw.

(*ii*) $\sum_{t\supset(j_1, i_r)} \beta_r^t$ for $j=1,2, \ldots N$

$(i\neq j)=1,2, \ldots N$ and $r=1, 2, \ldots n$ is the sum of all those weights associated to the rth draw and to the 1th sample in which jth unit appears at the first draw and i $(\neq j)$th unit appears at the rth draw.

The conditions of unbiasedness of $\left(T_6^M\right)$ for the population total T can be easily had from the equation (3.3) by equating the bracketed terms to unity. Let us denote them by

$[i]$ for $i=1, 2, \ldots N$ ...(3.4)

Further

$$V\left(T_6^M\right) = E\left(T_6^M\right)^2 - T^2 \quad \ldots(3.5)$$

Thus to minimize $V\left(T_6^M\right)$ subject to the conditions of unbiasedness in equation (3.4) ; we consider the function ϕ_1 as follows

$$\phi_1 = x_1^2 \Big[K_1 \sum_{t\supset(1_1)} \left(\beta_1^t\right)^2 + k_2 \Big(\sum_{t\supset(2_1, 1_2)} \left(\beta_2^t\right)^2 + \ldots$$

$$+ \sum_{t\supset(2_1, 1_n)} \left(\beta_n^t\right)^2 + K_3 \Big(\sum_{t\supset(3_1, 1_2)} \left(\beta_2^t\right)^2 + \sum_{t\supset(3_1, 1_3)} \left(\beta_3^t\right)^2 + \ldots$$

$$+ \sum_{t\subset(3_1, 1_n)} \left(\beta_n^t\right)^2 + \ldots + K_N \Big(\sum_{t\supset(N_1, 1_2)} \left(\beta_2^t\right)^2 + \ldots + \sum_{t\supset(N_1, 1_n)} \left(\beta_n^t\right)^2 \Big]$$

$$+x_2^2\Big[K_2 \sum_{t\supset 2_1} \left(\beta_1^t\right)^2 + K_1 \sum_{t\supset(1_1, 2_2)} \left(\beta_2^t\right)^2 + \ldots$$

$$+ \sum_{t\supset(1_1, 2_n)} \left(\beta_n^t\right)^2 + \ldots + K_N\Big(\sum_{t\supset(N_1, 2_2)} \left(\beta_2^t\right)^2 + \ldots$$

$$+ \sum_{t\supset(N_1, 2_n)} \left(\beta_n^t\right)^2\Big]$$

$$+2x_1 x_2\Big[K_1\Big(\sum_{t\supset(1_1, 2_2)} \beta_1^t \;\beta_2^t + \sum_{t\supset(1_1, 2_3)} \beta_1^t \;\beta_3^t + \ldots + \sum_{t\supset(1_1, 2_n} \beta_1^t \;\beta_n^t\Big)$$

$$+K_2\Big(\sum_{t\supset(2_1\, 1_2)} \beta_1^t \;\beta_2^t + \sum_{t\supset(2_1, 1_3)} \beta_1^t \beta_3^t + \ldots + \sum_{t\supset(2_1, 1_n)} \beta_1^t \beta_n^t\Big)\Big]$$

$$-T^2 - 2\,\lambda_1\,[1] - 2\,\lambda_2\,[2] \qquad \ldots(3.6)$$

where

(1) $\sum_{t\supset(1_1, 2_r)}$

denotes the sum over those samples in which first unit appears at the first draw and second unit appears at the rth draw for $r=2, 3, \ldots\ldots n$.

(2) $\sum_{t\supset(2_1, 1_r)}$

denotes the sum over those samples in which second unit appears at the first draw and first unit appears at the rth draw for $r=2, 3, \ldots n$.

Diff. ϕ_1 w.r.t. β_1^t .

when $t\supset(1_1)$; β_2^t

when $t\supset(2_1, 1_2)$ ..., β_n^t

when $t\supset(2_1, 1_n)$, β_1^t

when $t \supset (2_1) \ldots, \beta_2^t$

when $t \supset (3_1, 2_2) \ldots, \beta_n^t$

when $t \supset (3_1, 2_n) \ldots, \beta_1^t$

when $t \supset N_t, \beta_2^t$

when $t \supset (N_t 2_2); \beta_n^t$

when $t \supset (N_1, 1_n)$

and then

equating the differential coefficients to zero we get a set of equations. The n considering the vector $X=(x_1 \neq 0, x_1=0; i=2, 3, \ldots N)$ and the above equations so obtained and making use of proper sums, along with the conditions of unbiasedness we get,

$$\lambda_1 = \frac{x_1^2}{\pi_1} \qquad \ldots(3.6)$$

Substituting the value of λ_1 in the above equation we get

$$\beta_r^t = \frac{1}{\pi_1} = 0 \qquad \ldots(3.8)$$

for all samples which include first unit at any draw for $r=1, 2 \ldots, n$

Now if we consider the population vector $X=(x_1 \neq 0, x_2 \neq 0$ and $x_i=0; i=3, 4, \ldots, N)$ and making use of equations (3.7) and (3.8) and the first equation of the diferentials we get

$$\sum_{r=1}^{n} \beta_r^t = 0 \qquad \ldots(3.9)$$

for those samples containing 1st unit at the first draw. But from equation (3.8) each β_r^t is greater than zero. Thus (3.8) and (3.9) suggest that

$$\beta_r^t = 0 \text{ for all } r \text{ and for } t \supset 1_1 \qquad \ldots(3.10)$$

which is a contradiction to the existence of NBE of order one and hence it proves that NBE of order two under T_6 class of linear estimators does not exist when Ikeda-Sen sampling scheme is adopted.

4. Necessary Best Estimators in T_4 and T_5 Classes

The Ikeda-Sen estimator for population total T in cases of T_4 and T_5 classes of Linear Estimators is defined respectively as:

$$T_4^M = \sum_{r=1}^{n} \beta_{ir} \, x_r \qquad \ldots(4.1)$$

and

$$T_5^M = \sum_{i=1}^{n} \beta_i^t \, x_i \qquad \ldots(4.2)$$

where β_{ir} is the weight associated to the ith unit, whenever it appears at the rth draw and β_i^t is the weight associated to the ith unit whenever it is included in the t^{th} sample.

The conditions of unbiasedness for T_4^M for population total in T_4 class can be seen as

$$=(B_{i1} \; p_1 + \sum_{r=2}^{n} c_i \, B_{ir}) = 1 \qquad \ldots(4.3)$$

for all $i = 1, 2, \ldots, N$

where as the conditions of unbiasedness for T_5^M for the population total in T_5 class are

$$[g_i] = K_i \sum_{t \supset i_1}^{N} \beta_i^t + \sum_{j(\neq i)=1}^{n} K_j \sum_{t \supset j_1} \beta_i^t \qquad \ldots(4.4)$$

For $i = 1, 2, \ldots N$.

The results similar to that of T_6 class can be easily proved for T_4 and T_5 classes, following the same routine algebra. The outlines of the proof are however, given below:

1. Obtain the variance expressions

$$V\left(T_4^M\right) \text{ and } V\left(T_5^M\right)$$

2. For minimization of $V\left(T_4^M\right)$ w.r.t. the conditions in equation (4.3), define the function

$$\phi_2' = V\left(T_4^M\right) - 2\sum_{i=1}^{N} \lambda_i\,(q_i - 1),$$

and for minimization of $V\left(T_5^M\right)$ w.r.t. the conditions in equation (4.4), define the function

$$\phi_3' = V\left(T_5^M\right) - 2\sum_{i=1}^{N} \lambda_i'\,(g_i - 1).$$

where λ_s are the Langragian undetermined multipliers.

3. Define functions ϕ_2 from ϕ_2' and ϕ_3 from ϕ_3' containing the terms involving $x_1\,x_2$ only. The so obtained functions will be of the following form,

$$\phi_2 = x_1^2 \sum_{r=1}^{n} \beta_{1r}^2\, p_{1r} + x_1 x_2 \sum_{s(\neq r)=1}^{n} \beta_{1r}\,\beta_{2s}\, p(1r,\ 2s)$$

$$+ x_2^2 \sum_{r=1}^{n} \beta_{2r}^2\, p_{2r} + x_1\, x_2 \sum_{s\,(\neq r)=1}^{n} \beta_{2r}\,\beta_{1s}\, p(2r,\ 1s) - 2\sum_{i=1}^{2} (q_i - 1)$$

and

$$\phi_3 = x_1^2 \left[K_1 \sum_{t \supset 1_1}^{K} \left(\beta_1^t\right)^2 + K_2 \sum_{t \supset 2_1}^{K^1} \left(\beta_1^t\right)^2 + \ldots\ K_N \sum_{t \supset N_1}^{K^1} \left(\beta_1^t\right)^2 \right]$$

$$+x_2^2\left[K_2\sum_{t\supset 2_1}^{K}\left(\beta_2^t\right)^2+K_1\sum_{t\supset 1_1}^{K'}\left(\beta_2^t\right)^2+\ldots+K_N\sum_{t\supset N_1}^{K'}\left(\beta_2^t\right)^2\right]$$

$$2x_1x_2\left[K_1\sum_{t\supset(1_1,2)}\beta_1^t\,\beta_2^t+K_2\sum_{t\supset(2_1,1)}\beta_2^t\beta_1^t+K_3\sum_{t\supset(3_1,1,2)}\beta_2^t\beta_1^t+\ldots\right.$$

$$\left.+K_N\sum_{t\supset(N_1,1,2)}\beta_2^t\beta_1^t\right]-2\sum_{i=1}^{-2}\lambda_i^1(g_i-1)$$

where

$$\sum_{t\supset(i_1,j)}$$

is the sum over those weights of all the samples in which ith unit appears at the first draw and the jth unit appears anywhere in the samples.

$$\sum_{t=j_1,1,2)}$$

is the sum over those weights of all the samples in which jth ($j\neq 1, 2$) unit appears at the first draw and first and second units appear anywhere in the samples.

4. Differentiate ϕ_2 and ϕ_3 w.r.t. the weighting coefficients involved in them and equate the differential coefficients to zero. Making the appropriate sums and using the population vector $X=(x_1\neq o,\ x_i=0$ for $i=2, 3, \ldots N)$

One would get,

$$B_{1r}=\frac{1}{\pi_1}\text{ (in } T_4 \text{ case) for all } r=1, 2, \ldots n$$

$$B_1^t=\frac{1}{\pi_1}\text{ (in } T_5 \text{ case) for all } t=1, 2, \ldots K$$

This proves the existence of NBE of order one in T_4 & T_5 classes of linear unbiased estimators.

Also by utilizing the population vector

$X=(x_1=0\ \ x_2=0\ \ x_i=0$ for $i=3, 4, \ldots N)$ and the appropriate sums, one would get,

$\beta_{1r}=0$ (in T_4 case) for all $r=1, 2, \ldots n$

$\beta_1^t=0$ (in T_5 case) for all $t=1, 2, \ldots K$

This leads to contradiction that all the weighting coefficients are zero and hence proves the non-existence of NBE of order two in T_4 and T_5 classes of linear unbiased estimators under Ikeda-Sen sampling scheme,

Remarks

Since MVLUE is the NBE of order N, the results of sections 3, and 4 suggest that the MVLUE in T_4, T_5 and T_6 classes of linear estimators does not exist when Ikeda-Sen system of sampling is being employed.

Summary

The non-existence of necessary best estimators of order two has been established in T_4, T_5 and T_6 classes of linear estimators when Ikeda-Sen sampling Scheme is undertaken.

Acknowledgements

The authors are extremely thankful to Dr. B.D. Tikkiwal, Senior Professor in the Department of Stat., Rajasthan University, Jaipur and also to Dr. Arijit Choudhry for their valuable suggestions. The authors also offer their sincere thanks to the referee for his suggestions for the improvement of the paper.

References

[1] Prabhu Ajgaonkar, S.G. (1965). On a class of linear estimators in sampling with varying probabilities of selection and without replacement J.A.S.A., 60, pp 637-642.

[2] Prabhu Ajgaonkar, S.G. (1969). On the non-existence of a best estimator for the entire class of linear estimators, Sankhya A, Vol. 31, pp. 453-462.

[3] Prabhu Ajgaonkar, S.G. (1970). A contribution to the theory of unique estimators for Horvitz and Thompson's classes of linear estimation, Metrika.

[4] Godambe. V.P. (1955). A unified theory of sampling from finite population, J.R.S.S., B. 17, pp 269-278.

[5] Hege, V.S. (1967). An optimum property of the Horvitz—Thompson estimate, J.A.S.A., 62, pp 1013-1017.

[6] Horvitz, D.G. and Thompson, D.J. 1952). A generalisation of sampling without replacement from a finite Universe, J.A.S.A., 47, 663-685.

[7] Koop, J.C. (1963). On the axioms of sample formation and their bearing on the construction of linear estimators in sampling theory for finite populations, Metrika 7, (2 and 3) : B—114, 165-204.

[8] Midzuno, H. (1952). 'On the sampling system with probability proportional to sum of sizes,, Ann. Inst. Statist. Math., Japan. 3, 99-107.

[9] Rao, J.N.K. (1971). Some thoughts on foundations of survey sampling, JISAS, Vol. XXIII, No. 2 (December), pp. 69-82.

[10] Sen, A.R. (1952) 'Further developments of the theory of the selection of primary sampling units with special reference to N.C. Agricultural Population', Ph.D. thesis N.C. State College.

[11] Yogi, A.K. and Gupta, P.C. (1975). 'On the non-existence of second and higher order necessary best estimators' Jour. Ind. Soc. Agril. Stat., Vol. XXVII, No. 1, pp. 55-66'.