OPTIMUM ESTIMATORS AND STRATEGIES IN SURVEY SAMPLING

M. K. RAMAKRISHNAN

RESTRICTED COLLECTION

A thesis submitted to the Indian Statistical Institute in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Calcutta
1970

To my parents

Parvathy and M. S. K.

ACKNOWLEDGEMENTS

I am grateful to Prof. C. R. Rao, F.R.S., for his valuable guidance and encouragement during the course of my investigations. I would also like to express my indebtedness to Prof. J. N. K. Rao for his constant encouragement and supervision that enabled me to complete my dissertation soon and for his permission to include the results of our joint work in Section (2.5). I am grateful to Drs. M. N. Murthy and T. V. Hanurav for their interest in this work.

Thanks are also due to all the members of the faculty of the Research and Training School whose help I received at various stages of the work: and to Mr. G. M. Das for the quick and elegant typing of the manuscript.

M. K. R.

C O N T E N T S

Page

===========

I N T R O D U C T I O N

This thesis consists of nine chapters. In the first chapter we give the basic concepts and definitions and also a broad review of the literature related to the problems considered in this thesis. The second and third chapters are devoted to a detailed modified comparison of sampling with and without replacement, for the case of equal and unequal probability sampling respectively. In the fourth and fifth chapters we discuss the criteria of hyper-admissibility and linear sufficiency for the choice of an optimum estimator for a given sampling design. The succeeding three chapters have as their main objective the central problem of choice of an optimum sampling strategy subject to some cost restrictions. In the last chapter we give some results concerning Horvitz-Thompson estimator and its variance estimation. Since each chapter has its own summary at the beginning, we shall not give a detailed chapter-wise summary here. Instead we shall briefly describe the course of development of the theory of survey sampling to the present date.

Though the concept of sampling has been in vogue from time immemorial, it is only during the thirties and forties that a more systematic development of the theory of sample surveys has taken place with the advent of sampling without replacement, probability sampling and the theory of stratification. But the large number of techniques - sometimes ingenious - developed and practised during that period were mostly based on empirical and intuitive considerations. It is only of late that attention has been paid to the purely theoretical aspects of the subject.

A clear formulation of the central problem of the theory of sampling from finite populations is due to Godambe [14]. In that paper he proposed a unified theoretical set-up in which most of the problems of sampling from finite populations could be discussed. Since then tremendous progress has been made towards developing a unified theory of sampling. Whether or not the present status of the theory is satisfactory is a controversial issue and we propose not to enter into a discussion about it here.

CHAPTER I

CONCEPTS AND DEFINITIONS

In this chapter we give the necessary concepts and definitions and also a broad review of the literature related to the problems considered in this thesis.

1.1 <u>The Unified Theory</u>: We consider a finite population consisting of N units

$$U = (1, 2, \ldots, N). \tag{1.1.1}$$

It is understood that the units of the population are distinguishable. A list of such units as (1.1.1) is termed as a sampling frame and N is called the size of the population.

A finite sequence of units from U is called a sample and is denoted by

$$s = (i_1, i_2, \ldots, i_{r(s)}) \tag{1.1.2}$$

where $i_j \in U$, $j = 1, 2, \ldots, r(s)$. $r(s)$ is called the 'sample size' and $n(s)$, the number of distinct units in s, is called the 'effective sample size'. Let

$$S = \{ s \} \tag{1.1.3}$$

denote the totality of samples from U and p a real set function on S such that

$$p(s) \geq 0, \quad \text{for all } s \in S \text{ and } \sum_s p(s) = 1 . \tag{1.1.4}$$

The pair (S,p) is called a (sampling) design and is denoted by

$$d = d(S, p) = (S, p). \tag{1.1.5}$$

Sometimes we may denote a design simply by p.

In practice, a design is not implemented by listing all possible samples with the corresponding probabilities and then making a draw with the appropriate probabilities p(s). In this connection Hanurav [21] defined a 'sampling mechanism' of drawing units i from the population U in (1.1.1) one after another with varying probabilities. Specifically

Definition 1.1.1: A sampling mechanism is a function $q(i, s_{k-1}, k)$ with arguments i, $i \in U$, s_{k-1} the sequences of the type (1.1.2) with k-1 terms and k = 1,2,... such that

$q(i, s_{k-1}, k) \geq 0$ for all i, k, s_{k-1} and $\sum_{i \in U} q(i, s_{k-1}, k) = 1$ for all k and s_{k-1}. Hanurav [21] then proved the following fundamental

Theorem 1.1.1: There exists a one to one correspondence between sampling designs and sampling mechanisms.

The importance of the above theorem lies in the fact that, when one is concerned with a search for some sort of optimum sampling method, one can make the search in the unified framework of sampling designs rather than the seemingly diverse types of sampling methods, all of which cannot be handled with a single tool.

1.1.1 Construction of designs

Given a design $d(S, p)$, let

$$\pi_i = \pi_i(p) = \sum_{s \supset i} p(s), \quad 1 \leq i \leq N \tag{1.1.6}$$

and

$$\pi_{ij} = \pi_{ij}(p) = \sum_{s \supset i,j} p(s), \quad 1 \leq i \neq j \leq N \tag{1.1.7}$$

where in (1.1.6) the sum on the r.h.s. is over all samples that contain the unit i and in (1.1.7) the sum is over all samples that contain the units i and j. The π_i's and

π_{ij}'s are called the first and second order inclusion probabilities respectively and they play an important role in the choice of optimum sampling methods, as can be seen from Chapter VIII. Immediately from the definitions we have

$$0 \leq \pi_i \leq 1, \qquad 1 \leq i \leq N,$$

and

$$0 \leq \pi_{ij} \leq \min(\pi_i, \pi_j), \qquad 1 \leq i \neq j \leq N,$$

where min (a, b) denotes the smaller of a and b. Let μ be the expected effective sample size of a design p, defined by

$$\mu = \mu(p) = \sum_{s \in S} n(s)p(s) \quad . \tag{1.1.8}$$

The following relations are well known:

i) $\pi_{ij}(p) \geq \pi_i(p) + \pi_j(p) - 1$

ii) $\sum_{i=1}^{N} \pi_i(p) = \mu(p)$ (Godambe [14])

iii) $\sum\sum_{i \neq j}^{N} \pi_{ij}(p) = \mu(p)(\mu(p) - 1) + V_p(n(s))$ (Hanurav [21])

iv) $V_p(n(s)) \geq f(1 - f)$ (Hanurav [21])

where f denotes the fractional part of $\mu(p)$.

v) For a design p for which

$$s \in S,\ p(s) > 0 \Longrightarrow n(s) = \mu$$

$$\sum_{j \neq i} \pi_{ij}(p) = (\mu - 1)\ \pi_i$$

and $$\sum_{i \neq j} \sum \pi_{ij}(p) = \mu(\mu - 1).$$ (Yates and Grundy [66]).

Some interesting problems of internal consistency of given set of inclusion probabilities of various orders now arise. We discuss them in detail in Chapter VIII.

1.1.2 Estimation

We consider a real valued variable $\mathcal{Y}$ defined over U and which takes value y_i on i, $1 \leq i \leq N$. Let y denote the vector

$$y = (y_1, y_2, \ldots, y_N). \qquad (1.1.9)$$

The y_i's are unknown apriori and in fact our parameter of interest is y which is assumed to be a point in R_N, the N-dimensional Euclidian space. In this thesis we restrict ourselves to the problem of estimating the particular parametric function, conventionally called the population total, defined by

$$T(y) = Y = \sum_{i=1}^{N} y_i \tag{1.1.10}$$

by observing the values of y_i for all $i \in s$, where s is a sample drawn according to a design (S,p).

Definition 1.1.2: An estimator $e(s,y)$ is a real valued function defined on $S \times R_N$ which depends on y only through those y_i's for which $i \in s$. That is for any two y, y' such that $y_i = y_i'$ for all $i \in s$, $e(s,y) = e(s,y')$.

From practical considerations it is evident that the estimate $e(s, y)$ need not be defined for those samples for which $p(s) = 0$. However, it may be mentioned that the above definition of an estimator is not most general, in the sense that there are estimators in common use and in the licrature which are not special cases of $e(s, y)$ in Definition (1.1.2). But from the sufficiency of the effective sample (i.e. the set of distinct units together with their y-values) it follows that we may restrict to Definition (1.1.2) without any loss of generality in our search for an optimum estimator. In this connection we may refer to the papers of Basu ([4],[6]), Des-Raj and Khamis [11], Roy and Chakraborthy [54], Pathak ([37], [39]), Godambe and Joshi [17], Murthy [34] and Basu and Ghosh [5]. An estimator is said to be linear if it is of the form

$$e(s,y) = \sum_{i \in s} b(s, i) y_i \tag{1.1.11}$$

where b is a function on $S \times U$ such that $b(s,i) = 0$ if $i \notin s$.

Definition 1.1.3: For a design $d(S, p)$, an estimator $e(s, y)$ is said to be unbiased for a parametric function $\tau(y)$ if

$$\sum_{s} e(s, y)\, p(s) = \tau(y) \tag{1.1.12}$$

for all $y \in R_N$. We denote by $L(d)$ (or $L(p)$) and $L_o^*(d)$ (or $L_o^*(p)$), the class of all linear estimators and the class of all linear unbiased estimators of the population total Y in (1.1.10) respectively. An estimator $e(s, y)$ of $\tau(y)$ which is not unbiased for $\tau(y)$ is called a biased estimator of $\tau(y)$ and the bias is defined by

$$B(e) = E(e(s, y)) - \tau(y). \tag{1.1.13}$$

In this thesis we consider the mean square error as our loss function though most of the definitions and conclusions given here apply broadly to any convex loss function. The mean square error (mse, for brevity) of an estimator e of τ is

$$M(e,y) = E(e-\tau)^2 = \sum_s e^2(s,y)p(s) - 2\tau(y) \sum_s e(s,y)p(s) + \tau^2(y). \qquad (1.1.14)$$

When $e(s,y)$ is unbiased for $\tau(y)$, then $M(e,y)$ in (1.1.14) denotes the variance of $e(s,y)$, namely

$$V(e,y) = \sum_s e^2(s,y)p(s) - \tau^2(y). \qquad (1.1.15)$$

Godambe [14], then proved the following celebrated -

<u>Theorem 1.1.2</u>: For any design $d(S, p)$ there does not exist a uniformly minimum variance (UMV, for short) estimator in $L_o^*(d)$.

However, later Godambe [16], Hége [24] and Hanurav [21] pointed out some exceptions to the theorem and gave some nontrivial designs where best estimators exist. Such designs were called uni-cluster designs by Hanurav [21].

<u>Definition 1.1.4</u>: A design (S, p) is said to be unicluster if $s_1, s_2 \in \bar{S}$ implies either s_1 is equivalent to s_2

(i.e. they contain the same units) or $s_1 \cap s_2 = \emptyset$ (i.e. they do not contain any unit in common) where $\bar{S}$ is the set of all samples s for which $p(s) > 0$.

Godambe's [14] negative result mentioned in Theorem (1.1.2) necessitated the weeding out of 'bad' estimators and the criterion of admissibility was introduced in this connection.

Definition 1.1.5: With respect to a given design d, an estimator $e(s,y)$ belonging to a class $\mathcal{C}$ of estimators of T is said to be admissible in $\mathcal{C}$ if there does not exist any other estimator in $\mathcal{C}$ which is uniformly better than $e(s,y)$, i.e. given any $e'(s,y) (\neq e(s,y)) \in \mathcal{C}$, there exists at least one point $y_o \in R_N$ such that

$$M(e, y_o) < M(e', y_o) \qquad (1.1.16)$$

where both the mse's in (1.1.16) are evaluated at $y = y_o$.

The usual definition of admissibility used in decision theory (see Wald [65]) which is slightly different from the above definition is

Definition 1.1.6: With respect to a given design d, an estimator $e(s,y)$ belonging to a class $\mathcal{C}$ of estimators is said to be admissible in $\mathcal{C}$, if

$$e'(s,y) \in \mathcal{C},\ M(e',y) \leq M(e,y), \text{ for all } y \in R_N$$

$$\Rightarrow M(e',y) = M(e,y) \quad \text{for all } y \in R_N. \tag{1.1.17}$$

Murthy and Singh [36] raised as an open problem some conditions under which the above two definitions are equivalent. The well-known result that when $\mathcal{C}$ is convex (i.e. $e_1, e_2 \in \mathcal{C} \Rightarrow \frac{e_1+e_2}{2} \in \mathcal{C}$) the above two definitions are equivalent seems to have escaped their notice. For any design d we denote by $A^*(d)$ and $A(d)$ the class of all unbiased estimators and the class of all estimators respectively for the population total. A particular unbiased estimator suggested by Horvitz and Thompson [26] seems to have received much attention from different authors recently. It is defined as

$$\bar{e}(s, y) = \sum_{i \in s} \frac{y_i}{\pi_i(d)} \tag{1.1.18}$$

where $\pi_i(d)$'s are the first order inclusion probabilities defined in (1.1.6). $\bar{e}(s,y)$ will be called the Horvitz-Thompson estimator (HT estimator, for short) of the population total.

Godambe [15] proved that for any design d, $\bar{e}(s,y)$ is admissible in $L_0^*(d)$. Godambe and Joshi [17] generalised this result to the class $A^*(d)$ and deduced as a corollary the non-existence of a UMV estimator in $A^*(d)$. An interesting observation can be made in this connection. A careful examination of the proof of admissibility of $\bar{e}(s,y)$ in $A^*(d)$ [See Godambe and Joshi [17]) will show that, nowhere in the proof have they made use of the fact that $d(S,p)$ is a design

$$\text{i.e.} \quad \sum_{s} p(s) = 1.$$

We can in fact generalise the definition of a design to include the case where there is a positive probability of not choosing any unit in the sample. To give an example of such a design, consider the following sampling method: Conduct N independent binomial trials with probability of success for i-th trial being equal to π_i. The i^{th} unit is included in the sample if and only if the i-th trial results in a success. The sample consists of the units selected arranged in the increasing order of i. Clearly in the above design there is a positive probability, namely $\prod_{i=1}^{N} (1 - \pi_i)$, of not selecting any unit in the sample. From the observation made above it follows that even for such designs $\bar{e}(s,y)$ remains admissible in the class of all unbiased estimators. An alternative elegant proof of the above statement is given in Section (2.1) using induction on N, the population size. One can note that the Godambe-Joshi [17] proof of admissibility of $\bar{e}(s,y)$ depends heavily on the form of the loss function considered, namely the variance. It will be a good problem to see whether $\bar{e}(s,y)$ remains admissible in $A^*(d)$ for any general convex loss function.

1.1.3 Estimation of $V(\bar{e},y)$ Straight-forward calculation shows that

$$V(\bar{e},y) = \sum_{i=1}^{N} \frac{1-\pi_i}{\pi_i} y_i^2 + \sum_{i\neq j}^{N}\sum \frac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j} y_i y_j . \quad (1.1.19)$$

Two unbiased estimators have been proposed in the literature for $V(\bar{e}, y)$ in (1.1.19), namely

$$v_{HT} = \sum_{i\in s} \frac{1-\pi_i}{\pi_i^2} y_i^2 + \sum_{i\neq j\in s}\sum \frac{\pi_{ij}-\pi_i\pi_j}{\pi_{ij}} \frac{y_i}{\pi_i} \frac{y_j}{\pi_j} , \quad (1.1.20)$$

$$v_{YG} = \sum_{i \neq j \in s}\sum \frac{\pi_i \pi_j - \pi_{ij}}{\pi_{ij}} \left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j}\right)^2, \qquad (1.1.21)$$

where π_{ij}'s are as defined in (1.1.6) and (1.1.7) respectively. The estimator v_{HT} has been proposed by Horvitz and Thompson [26] and v_{YG} by Yates and Grundy [66] and they are called HT-variance estimator and the YG-variance estimator respectively. While v_{HT} is applicable in any design in which $V(\bar{e}, y)$ is estimable, v_{YG} is applicable only in fixed sample size designs in which $V(\bar{e}, y)$ is estimable. In Section (9.3) we modify the YG-variance estimator to suit **non-fixed sample size** designs **also** and prove the existence of designs where the modified YG-variance estimator is non-negative.

1.1.4 Choice of estimators

The non-existence of a UMV estimator in $L_o^*(d)$ for non-unicluster designs, and $A^*(d)$ for any design leads to the problem of choice of optimum estimators from the class of all admissible estimators. Many optimality criteria have been put forward in the literature for the choice of an estimator. We review them briefly below. The criteria of hyper-admissibility due to Hanurav [21] and, linear sufficiency and distribution - free sufficiency due to Godambe [18] are not included here since they are dealt with in detail in Chapters IV and V respectively.

Linear invariance: This concept introduced by Roy and Chakraborthy [54], demands that the estimator should be invariant under linear transformations of y. Linear invariance seems reasonable for the case of equal probability sampling. Most of the known estimators in varying probability sampling are not linear invariant. Also, the class of linear invariant estimators is very big and hence the criterion fails to give an optimum estimator.

Regularity: This criterion is also due to Roy and Chakraborthy [54]. An estimator t is regular if and only if its variance is proportional to the population variance σ_y^2.

$$\text{i.e.} \qquad V(t) = k\,\sigma_y^2$$

where $\sigma_y^2 = \frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{Y})^2$ and k is a constant. Designs which are balanced in a certain sense are shown to possess a best estimator in the class of all regular estimators.

Necessary bestness: This curious criterion introduced by Ajgaonkar [2] claims to choose 'a serviceable estimator from the practical point'. We first give the following

Definition 1.1.7: (Ajgaonkar [2]): Between two unbiased estimators T and T' of the population total Y with variances

$$V(T) = \sum_{i}^{N} a_i y_i^2 + 2 \sum_{i<j}^{N}\sum a_{ij}\, y_i\, y_j$$

and $$V(T') = \sum_{i}^{N} b_i\, y_i^2 + 2 \sum_{i<j}^{N}\sum b_{ij}\, y_i\, y_j$$

the estimator T is 'necessary better' than T' if $a_i < b_i$ for all i. The estimator T (in a class $\mathcal{C}$) is 'necessary best in $\mathcal{C}$ if it is necessary better than every other estimator in $\mathcal{C}$.

Ajgaonkar [2] considered a subclass of $L_o^*(d)$ termed T_5-class defined by

$$T_5 = b(s_m) \sum_{i \in s} b_i\, y_i \qquad (1.1.22)$$

where $b(s_m)$ is a function of s_m, the set of distinct units observed in the sample and $b_1, b_2, \ldots, b_N$ are pre-assigned constants and proved that the HT estimator $\bar{e}(s,y)$ in (1.1.18) is the unique necessary best estimator in the T_5-class of estimators. Hege [25] extended this result to $L_o^*(d)$ and Singh [60] claims to have extended it further to the class H_u of all unbiased estimators $e(s, y)$ defined by

$$e(s,y) \in H_u \text{ if and only if } e(s,y) = 0 \text{ if } y_i = 0 \text{ for all } i \in s \text{ and } s \in S.$$

It is easily seen that all the above results are incorrect unless in the Definition (1.1.7) of necessary bestness we replace $a_i < b_i$, $1 \leq i \leq N$ by $a_i \leq b_i$, $1 \leq i \leq N$. Once this change is made in the definition all the above results follow easily by noting that necessary bestness is equivalent to bestness in hyper-planes of dimension one and that the HT-estimator is best in hyper-planes of dimension one in H_u. However Singh's [60] claim of its uniqueness remains incorrect even with the modified definition since any estimator that reduces to the HT-estimator in hyper-planes of dimension one is a necessary best estimator.

Bayesness : Here we assume that all our prior knowledge about the population or the variate under study, can be formulated in some sort of a prior distribution ∂ of the vector parameter y. The role of the prior distribution ∂ is solely to choose between different estimators (and also designs) and has nothing to do with the final inference about y, which will exclusively depend on the observed sample s and the variate values y_i, $i \in s$. This approach introduced by Cochran [9] has been fruitfully exploited by Godambe [14], Hajek [20], Aggarwal [1] and many others. Our sole criterion of judgement in terms of a prior distribution would be the expected variance of an estimate as defined in (1.1.24), to follow.

With respect to a given prior distribution ∂ on R_N we define

$$\begin{aligned} \mu_i &= \int y_i \, d\partial \\ \sigma_i^2 &= \int (y_i - \mu_i)^2 \, d\partial \qquad (1.1.23) \\ \sigma_{ij} &= \int (y_i - \mu_i)(y_j - \mu_j) \, d\partial \, . \end{aligned}$$

If as in (1.1.15), V(e, y) denotes the variance of an unbiased estimator e(s, y), the expected variance of e w.r.t ∂ is

$$\mathcal{E}\, V(e) = \int V(e, y)\, d\, \partial. \qquad (1.1.24)$$

Godambe [14] then proved

Theorem 1.1.3: For any fixed sample size (= n) design d(S, p) for which $\pi_i > 0$, $1 \leq i \leq N$ and for any prior distribution ∂ on R_N such that

i) the variates y_i, $1 \leq i \leq N$ are uncorrelated w.r.t. ∂

ii) $$\mu_i = \frac{\left(\sum_{i=1}^{N} \mu_i\right) \pi_i}{n} \qquad (1 \leq i \leq N).$$

(i.e. μ_i is proportional to π_i),

the HT estimator $\bar{e}(s, y)$ in (1.1.18) is a Bayes' solution in $L_o^*(d)$.

Godambe and Joshi [17] extended the above result to the class $A^*(d)$. Before proceeding to state the results regarding the choice of designs through Bayes' approach, we introduce some more basic concepts which are needed in this thesis.

1.1.5 Sampling strategies and their optimum choice

In the theory of estimation from infinite populations the distribution is assumed to be known but for a certain number of parameters and accordingly for random sampling from such populations the probability of getting any particular sample is completely determined by the sample size, but for those unknown parameters and the problem is to estimate some functions of the parameters on the basis of a sample. However, in finite population theory, the probability with which a sample is to be selected is completely at the choice of the statistician As a result, the problem in survey sampling is not merely to choose an 'optimum' estimator, as in the case of classical theory for infinite populations, but to choose an optimum combination of sampling and estimation procedures and this makes the problem more complicated. One point seems worth mentioning here. While there is inherent randomness in the process in the classical set-up, there is only artificial randomness in the survey set-up, injected in by the statistician. Why this artificial randomness, is a logical question to be answered and in this connection we refer to the papers of Basu and Ghosh [5], Basu ([6], [7]) and Zacks [67]. Much need be explored in this area and there is every chance that a new survey sampling theory which does not involve probability theory may emerge in the near future.

Definition 1.1.8: A design $d(S,p)$ together with an estimator e of τ, defined over d is called a 'sampling strategy' or simply a 'strategy' for the estimation of τ. Thus we denote a strategy by

$$H = H(d, e) = H(S, p, e) \tag{1.1.25}$$

This definition is due to Hajek [19] and the importance of this terminology is stressed by Hanurav ([21], [22]). A strategy $H(S, p, e)$ is said to be unbiased for τ if e is an unbiased estimator of τ. Otherwise it is called a 'biased strategy' and the bias of the strategy is defined as

$$B(H) = E(e) - \tau. \tag{1.1.26}$$

The expectation, variance or mean square error of a strategy are defined as the expectation, variance or mean square error of the corresponding estimator. Analogous to the Definition (1.1.6) of admissibility of an estimator we have

Definition 1.1.9 : A strategy H_o belonging to a class $\mathcal{C}(H)$ of strategies is said to be admissible in $\mathcal{C}(H)$ if there does not exist another H_1 belonging to $\mathcal{C}(H)$ which is uniformly superior to H_o.

i.e. $H_1 \in \mathcal{C}(H)$, $M(H_1, y) \leq M(H_o, y)$, for all $y \in R_N$

$$\Longrightarrow M(H_1, y) = M(H_o, y), \text{ for all } y \in R_N \qquad (1.1.27)$$

where M stands for mean square error. When only unbiased strategies are considered we replace M by V in (1.1.27) to denote variance.

For a comparison between two designs d_1 and d_2 (in both of which τ is estimable) we should first compare their costs in terms of a reasonable cost function. We assume a linear cost function which seems appropriate in most of the cases of unistage sampling. Under this, the cost of a sample is assumed to be given by

$$c(s) = An(s) + B \qquad (1.1.28)$$

where A and B are constants independent of s and $n(s)$ is the effective sample size of s. The expected cost of a strategy $H(S,p,e)$ or equivalently the design $d(S,p)$ is then defined as

$$C(H) = C(d) = \sum_s c(s)p(s) = A\mu(d) + B \qquad (1.1.29)$$

where $\mu(d)$ is the expected effective sample size defined in (1.1.8). Hence under our cost function, two designs (or strategies) are equally costly if and only if they have the same expected sample size.

1.1.6 Choice of designs

If 'best' estimators e_1 and e_2 exist for two equicost designs then we can say d_1 is better than d_2 if and only if

$$M(H_1,y) \leq M(H_2,y), \quad \text{for all} \quad y \in R_N \qquad (1.1.30)$$

where $H_1 = (d_1,e_1)$ and $H_2 = (d_2,e_2)$ and the strict inequality in (1.1.30) for at least one $y \in R_N$. Unfortunately, as seen before, there are not many designs in which 'best' estimators exist even in the class of homogeneous linear estimators. A natural definition for d_1 to be better than d_2 is then

Definition 1.1.10: d_1 is better than d_2 if and only if given any e_2 defined over d_2, there exists and e_1 defined over

d_1 (which may depend upon e_2) such that

$$M(H_1, y) \leq M(H_2, y) \quad \text{for all} \quad y \in R_N$$

with the strict inequality occurring for at least one y.

1.2 <u>Sampling with and without replacement</u>: For estimating a population parametric function (such as the population total or mean) two procedures of selecting a sample may be distinguished:

(a) Sampling with replacement. A fixed number n of units is selected with replacement.

(b) A fixed number n of units is selected without replacement or selection with replacement is continued till the desired number n of distinct units is obtained. Though 'sampling without replacement' is usually referred to selection of units without replacement, so that all the units in the sample are distinct, the two schemes of selection of n distinct units mentioned in (b) above are equivalent in the sense that the probability of drawing a specified set of n distinct units is same for both.

It was a common belief among survey practitioners that sampling without replacement was better than sampling with replacement. Basu [4] and, Des Raj and Khamis [11] proved that the usual estimator (overall sample mean) in simple random sampling with replacement is inadmissible and hence it is unfair to compare the sample means in simple random sampling with and without replacement. Basu [4] moreover mentioned the importance of cost consideration in making efficiency comparisons between sampling strategies.

It has not been possible so far to prove that sampling without replacement is better (Definition 1.1.10) than sampling with replacement even for simple classes of designs like those generated by simple random sampling. As a result various authors have attempted to study the relative efficiencies of sampling with and without replacement for the same expected cost on the basis of admissible estimators among both. Such an attempt is both logical and natural and should not be criticised though there has been a tendency on the part of certain authors to look down upon such results. But in all the comparisons, so far made, it is assumed that the number of units to be selected in sampling without replacement in order to equalise the expected cost is an integer. It can be easily seen that this assumption will not be true in general. In the next chapter we suggest a modification to cover the situations when the assumption is not satisfied and make modified efficiency comparisons for many equal probability sampling schemes. In Chapter II we extend this modified efficiency comparison for some interesting unequal probability sampling schemes.

1.3 Choice of Strategies: We give the following

Definition 1.3.1: In a class, $\mathcal{C}(H)$, of equicost unbiased strategies, an $H_o \in \mathcal{C}(H)$, is said to be '∂-optimum' in $\mathcal{C}(H)$, if it minimises

$$\int V(H, y) d\, \partial,$$

for H varying over $\mathcal{C}(H)$, uniformly with respect to all the parameters of the distribution ∂. An H_o which is ∂-optimum in $\mathcal{C}(H)$ for every $\partial \in \Delta$, a class of apriori distributions, is said to be 'Δ-optimum' in $\mathcal{C}(H)$.

Let Δ_g be the class of all prior distributions ∂ of y satisfying

i) $\mu_i = a\, x_i \qquad 1 \leq i \leq N,$

ii) $\sigma_i^2 = \sigma^2 x_i^g \qquad 1 \leq i \leq N,$

iii) $\sigma_{ij} = 0 \qquad 1 \leq i \neq j \leq N,$

where μ_i, σ_i^2 and σ_{ij} are as defined by (1.1.23) and $x_1, x_2, \ldots, x_N$ are the known-values of an auxiliary variate $\mathcal{X}$ highly related to the study-variate $\mathcal{Y}$ and a, σ^2 and g are unknown constants. This Δ_g-class of prior distributions has been widely used in survey sampling. Empirical investigations conducted by Mahalanobis [32], Smith [62] and Jessen [27] have shown that g lies between 1 and 2 in many practical situations. Godambe [14] then proved the following

<u>Theorem 1.3.1:</u> If μ is an integer $(\mu < \frac{X}{\max x_i})$ any strategy $H_o(d_o, \bar{e})$ where d_o is a design of fixed sample size μ with $\pi_i(d_o) = \mu x_i / X$, $1 \leq i \leq N$, is Δ_2-optimum in the class $L_1(H, \mu)$ of all fixed sample size $(= \mu)$, linear, unbiased strategies of the population total.

Hanurav [21] showed that the above strategy is Δ_2-optimum in the class $L(H, \mu)$ of all linear unbiased strategies with expected sample size μ. Afterwards Godambe and Joshi [17] generalised the above theorem by dropping the adjective 'linear' from it. Vijayan [64] showed that the result remains true even if we extend the Godambe-Joshi class to embrace all sampling designs with given expected sample size.

One should not be too much dogmatic about the above results as they are true only for Δ_2-class i.e. Δ_g with $g = 2$. Even if g is slightly different from 2 the validity of the above results ceases to hold. The above results of Godambe [14] and, Godambe and Joshi [17] for μ an integer, prompted the author to see whether they remain true with corresponding changes even when μ is not an integer. This question is answered in the negative in Section (9.2).

CHAPTER II

COMPARISON OF SAMPLING WITH AND WITHOUT REPLACEMENT: EQUAL PROBABILITY SAMPLING

2.0 Summary

In this chapter, after a brief introduction, we make modified efficiency comparisons - comparisons that take into consideration the fact that the expected number of units in a with-replacement sample need not be an integer - among some admissible unbiased estimators in simple random sampling with replacement (srswr, for short) and simple random sampling without replacement (srswor, for short), for the same expected cost. These results are extended to two-stage sampling and interpenetrating subsampling. Some numerical results are also provided.

Assuming a guessed value for the population mean $\bar{Y}$, satisfying $S^2 > N(\bar{Y} - C)^2$, where S^2 is the population mean square error for the characteristic, the usual estimator in srswor is improved. This improved estimator is shown to be better than that proposed by Pathak [41] for srswr which also uses the same prior information and the percent gain in efficiency has been tabulated.

Recently, Chaudhuri [8] has made some comparisons between srswr and srswor. He has proposed an estimator for a with-replacement sampling scheme and has claimed that it is better than the usual sample mean for srswor. We show that his claim is false and that most of his comparisons are either special cases of a theorem due to Seth and Rao [59] (and Rao [48]) or incorrect.

2.1 Introduction

In srswr of fixed size n from a population of N units, the usual estimator of the population mean $\bar{Y}$, based on all the units in the sample is

$$\bar{y} = n^{-1} \sum_{i=1}^{n} y_i . \tag{2.1.1}$$

Basu [4] and, Des Raj and Khamis [11] have shown that $\bar{y}$ is inadmissible by proving that the estimator

$$\bar{y}_m = m^{-1} \sum_{i=1}^{m} y_i \tag{2.1.2}$$

based on the m distinct units in the sample is uniformly more efficient than $\bar{y}$.

Consider the following class of estimators based only on the distinct units in the sample:

$$\bar{y}_{[f(m)]} = \frac{f(m)}{Ef(m)} \bar{y}_m \qquad (2.1.3)$$

where $f(m)$ is a function of m. Assuming that the cost of a with-replacement sample is proportional to the number of distinct units in it, the expected cost of a sample of n units drawn with srswr is equivalent to that of a sample of size $E(m)$ drawn with srswor. Seth and Rao [59] have shown that the sample mean $\bar{y}_{E(m)}$ in srswor is more efficient than $\bar{y}_{[f(m)]}$ if $S^2/\bar{Y}^2 < N$.

However, this comparison is not very satisfactory because $E(m)$ is not necessarily an integer. In Section 2, we develop a method which takes the above fact into account by using a randomised estimator. We prove that the randomised estimator is more efficient that $\bar{y}_m$. In Section 3, we extend the results to two-stage sampling and interpenetrating sub-sampling.

Pathak [41] considered a situation in which an apriori value C is available, such that

$$S^2 > N(\bar{Y} - C)^2.$$

Under this assumption he has shown that the estimator

$$\bar{y}_s = \frac{m}{E(m)} (\bar{y}_m - C) + C \tag{2.1.4}$$

in srswr is more efficient than $\bar{y}_{E(m)}$ in srswor. He cites this as an instance in which sampling with replacement is better than sampling without replacement.

This comparison is not justifiable because the apriori information C is not utilised in $\bar{y}_{E(m)}$. In Section 4 we get an estimator better than $\bar{y}_{E(m)}$ by using the prior information C. This estimator is better than $\bar{y}_s$ in almost all the situations encountered in practice.

2.2 Modified comparison for simple random sampling

In srswor we take the sample size as $[E(m)]$ or $[E(m)]+1$ with probabilities p_1 and p_2 respectively where p_1 and p_2 are chosen so that the expected sample size is $E(m)$ and $[x]$ denotes the integral part of x.

The above modification is well-defined as it is easily seen that there exists only one such choice of p_1 and p_2, namely

$$p_1 = 1 - E(m) + [E(m)] \quad \text{and} \quad p_2 = E(m) - [E(m)] \qquad (2.2.1)$$

The unbiased randomised estimator in srswor, we propose, is

$$\bar{y}^*_{E(m)} = \begin{cases} \bar{y}_{[E(m)]} & \text{with probability } p_1 \\ \bar{y}_{[E(m)]+1} & \text{with probability } p_2 . \end{cases} \qquad (2.2.2)$$

It is clear that $\bar{y}^*_{E(m)}$ reduces to $\bar{y}_{E(m)}$ when $E(m)$ is an integer. We now prove

<u>Theorem 2.2.1:</u> The variance of $\bar{y}^*_{E(m)}$ is greater than or equal to the variance of $\bar{y}_{E(m)}$.

<u>Proof:</u> We have

$$V(\bar{y}^*_{E(m)}) = \left(\frac{1-E(m)+[E(m)]}{[E(m)]} + \frac{E(m)-[E(m)]}{[E(m)]+1} - \frac{1}{N}\right)S^2 \qquad (2.2.3)$$

and

$$V(\bar{y}_{E(m)}) = (\frac{1}{E(m)} - \frac{1}{N})S^2 \quad (2.2.4)$$

where V denotes the variance.

It is easily verified that

$$\frac{1}{E(m)} \leq \frac{1 - E(m) + [E(m)]}{[E(m)]} + \frac{E(m) - [E(m)]}{[E(m)]+1} . \quad (2.2.5)$$

q.e.d.

Remark 2.2.1: From the above theorem it is clear that, when $E(m)$ is not an integer we are overestimating the gain in efficiency by taking the estimator $\bar{y}_{E(m)}$ in srswor.

2.2.1 Comparison of $\bar{y}_m$ with $\bar{y}^*_{E(m)}$

We now prove

Theorem 2.2.2: The estimator $\bar{y}^*_{E(m)}$ in srswor is uniformly more efficient than $\bar{y}_m$ in srswr.

Proof: Since

$$V(\bar{y}_m) = (E(\frac{1}{m}) - \frac{1}{N})S^2 \quad (2.2.6)$$

it is enough to show

$$E\left(\frac{1}{n}\right) \geq \frac{1+[E(n)]-E(n)}{[E(n)]} + \frac{E(n)-[E(n)]}{[E(n)]+1}. \qquad (2.2.7)$$

Now the function $L(x)$ defined by

$$L(x) = \frac{1+[x]-x}{[x]} + \frac{x-[x]}{[x]+1}, \quad x \geq 1$$

is convex because it is piecewise linear with increasing slope. Hence, applying Jensen's inequality for the convex function L we have

$$EL(n) \geq LE(n). \qquad (2.2.8)$$

Since $n(\geq 1)$ is an integer-valued random variable inequality (2.2.8) reduces to (2.2.7).

q.e.d.

Table 1 gives the percent gain in efficiency of $\bar{y}^*_{E(n)}$ over $\bar{y}_n$ for selected values of n and N.

Table 1. Percent gain in efficiency of $\bar{y}^*_{E(m)}$ over $\bar{y}_m$.

n \ N	4	7	10	13	15	20	23	25	27
3	10.5	2.6	1.2	0.7	0.5	0.3	0.2	0.2	0.1
6		8.1	4.4	4.4	3.6	1.9	1.4	1.2	1.0
8			5.2	4.3	3.2	2.2	2.1	2.2	2.0
10				4.6	3.4	3.0	2.2	1.8	1.6
12				5.3	3.6	2.8	2.2	2.0	1.9
15						2.9	2.5	2.1	2.0

It can be seen from the table that the gains are slightly smaller than those given in Rao([48] -Table 1). Also we see that $\bar{y}^*_{E(m)}$ leads to moderate gains in efficiency for small size populations. Moreover it is clear from the table that as N increases the gains become small even when the sampling fraction n/N is not small. An analytical proof of this fact follows from Rao [48] and Theorem(2.2.1).

2.2.2 Comparison of $\bar{y}_{(1)}$ with $\bar{y}^*_{E(m)}$

Pathak [37] has considered the estimator

$$\bar{y}_{(1)} = \frac{m}{E(m)} \bar{y}_m \tag{2.2.9}$$

which belongs to the class (2.1.3).

Since

$$V(\bar{y}_{(1)}) = \left(\frac{1}{E(m)} - \frac{1}{N}\right) S^2 + \frac{V(m)}{E^2(m)} \left(\bar{Y}^2 - \frac{S^2}{N}\right) \tag{2.2.10}$$

it is clear from (2.2.3) and (2.2.10) that

$$V(\bar{y}^*_{E(m)}) \leq V(\bar{y}_{(1)})$$

if and only if (iff, for short)

$$S^2 / \bar{Y}^2 \leq N \left[1 + \frac{N\, E(m)\,(E(m) - [E(m)])\,(1 - E(m) + [E(m)])}{V(m)\,[E(m)]\,([E(m)] + 1)}\right]^{-1} . \tag{2.2.11}$$

Table 2 gives the values of the right hand side of the above inequality (2.2.11) for selected values of n and N.

Table 2: Values of the right hand side of inequality (2.2.11).

N / n	$\frac{n}{.30}$ (rounded)	$\frac{n}{.25}$ (rounded)	$\frac{n}{.20}$	$\frac{n}{.15}$ (rounded)	$\frac{n}{.10}$	$\frac{n}{.05}$
3	1.95	1.96	1.96	1.97	1.98	1.99
4	3.32	3.23	3.17	3.12	3.07	3.03
5	5.30	4.91	4.66	4.46	4.28	4.13
6	8.12	7.23	6.56	6.04	5.62	5.28
7	13.42	10.65	9.05	7.98	7.12	6.46
8	25.75	16.23	12.50	10.32	8.82	7.77
9	19.55	26.95	17.57	13.24	10.75	9.11

The table shows that the inequality may not be satisfied for small sampling fractions. For example if $n = 3$, the inequality is not satisfied if $S^2 / \bar{Y}^2$ is greater than 2. But in such cases (i.e. where the inequality is not satisfied), there is no point in using simple random sampling, with or without replacement as the sampling error expressed as a percentage of the mean will be very high making the estimates altogether unusable.

2.3 Two extensions

2.3.1 Two-stage sampling with equal probabilities at both stages

Let M denote the number of primaries and N the number of secondaries in each primary. Let a sample of m primaries be drawn with srswr and from each of the η distinct primaries in the sample, n secondaries be drawn with srswr. Let w_i denote the number of distinct secondaries in the i^{th} distinct primary. Under a linear cost function the expected cost of a with-replacement sample is

$$C(m, n) = C_1 E(\eta) + C_2 E(\eta) E(w).$$

Ghosh [13] has proposed the estimator

$$\bar{\bar{y}}_{\eta, w_i} = \frac{1}{\eta} \sum_{1}^{\eta} \frac{1}{w_i} \sum_{1}^{w_i} y_{ij} . \qquad (2.3.1)$$

Minimising $V(\bar{\bar{y}}_{\eta, w_i})$ with respect to m and n subject to a fixed cost C_o, we can find m_{opt}, n_{opt} and $V_{min}(\bar{\bar{y}}_{\eta, w_i})$.

Now the expected cost of a with-replacement sample is equivalent to the cost of sampling $E(\eta) = M \left\{1 - \left(\frac{M-1}{M}\right)^{m_{opt}}\right\}$ primaries with srswor and from each selected primary

$E(w) = N\{1 - (\frac{N-1}{N})\}^{n_{opt}}$ secondaries with srswor, so that the estimator in sampling without replacement is

$$\bar{\bar{y}}_{E(\eta),E(w)} = \frac{1}{E(\eta)} \sum_{1}^{E(\eta)} \frac{1}{E(w)} \sum_{1}^{E(w)} y_{ij}. \qquad (2.3.2)$$

Rao [48] has shown that

$$V(\bar{\bar{y}}_{E(\eta),E(w)}) \leq V_{min}(\bar{\bar{y}}_{\eta,w_i}). \qquad (2.3.3)$$

This comparison is, however, not very satisfactory since $E(\eta)$ and $E(w)$ are not necessarily integers. As in Section 2, we therefore make the following modification:

Select $[E(\eta)]$ or $[E(\eta)]+1$ primaries with probabilities $1-E(\eta)+[E(\eta)]$ and $E(\eta)-[E(\eta)]$ respectively with srswor and from each selected primary select $[E(w)]$ or $[E(w)]+1$ secondaries with srswor with probabilities $1-E(w)+[E(w)]$ and $E(w)-[E(w)]$ respectively. We have four possible cases as follows:

Case:	$([E(\eta)],[E(w)])$	$([E(\eta)],[E(w)]+1)$	$([E(\eta)]+1,[E(w)])$	$([E(\eta)]+1,[E(w)]+1)$
Prob:	$(1-f_1)(1-f_2)$	$f_2(1-f_1)$	$f_1(1-f_2)$	$f_1 f_2$

(2.3.4)

where $f_1 = E(\eta) - [E(\eta)]$, $f_2 = E(w) - [E(w)]$.

Let us denote the modified estimator by $\bar{\bar{y}}^*_{E(\eta),E(w)}$.

<u>Theorem 2.3.1</u>: The variance of $\bar{\bar{y}}^*_{E(\eta),E(w)}$ is smaller than V_{min} of $\bar{\bar{y}}_{\eta,w_i}$.

<u>Proof:</u> Considering the four cases listed in (2.3.4) we get

$$V(\bar{\bar{y}}^*_{E(\eta),E(w)}) = \left(\frac{1-f_1}{n_1} + \frac{f_1}{n_1+1} - \frac{1}{M}\right)S_p^2$$

$$+ \left(\frac{1-f_1}{n_1} + \frac{f_1}{n_1+1}\right)\left(\frac{1-f_2}{n_2} + \frac{f_2}{n_2+1} - \frac{1}{N}\right)\frac{1}{M}\sum_1^M S_i^2 \quad (2.3.5)$$

where f_1 and f_2 are as defined above , $n_1 = [E(\eta)]$, $n_2 = [E(w)]$, S_p^2 is the population mean square of the primary means $\bar{Y}_i$ and S_i^2 is the i-th primary population mean square. Also the variance of $\bar{\bar{y}}_{\eta,w_i}$ is

$$V(\bar{\bar{y}}_{\eta,w_i}) = \left(E\left(\frac{1}{\eta}\right) - \frac{1}{M}\right)S_p^2 + E\left(\frac{1}{\eta}\right)\left(E\left(\frac{1}{w}\right) - \frac{1}{N}\right)\frac{1}{M}\sum_1^M S_i^2 . \quad (2.3.6)$$

Since η and w are non-negative integer-valued random variables, the theorem follows from (2.3.5) and (2.3.6) by making use of (2.2.7).

q.e.d.

2.3.2 Interpenetrating subsampling

Let k subsamples each of size n be drawn with srswor from a population of size N, independently of each other. Pathak [39] has shown that $\bar{y}_m$, the mean of the m distinct units in the combined sample of nk units, is uniformly more efficient than any linear function $\sum_{i=1}^{k} c_i \bar{y}_i$, $\sum c_i = 1$, of the sub-sample means.

Since

$$V(\bar{y}_m) = (E(\tfrac{1}{m}) - \tfrac{1}{N})S^2$$

it follows immediately from inequality (2.2.7) that $\bar{y}^*_{E(m)}$ in srswor is more efficient than $\bar{y}_m$. Though the idea of using interpenetrating subsamples is not to reduce sampling error, the author makes this comparison in view of Pathak's [39] result.

2.4 Utilisation of a guessed value of $\bar{Y}$

When a guessed value C of $\bar{Y}$ is available such that

$$S^2 > N(\bar{Y} - C)^2 \qquad (2.4.1)$$

we propose the estimator

$$\bar{y}_C = C + \frac{E(m)}{N}(\bar{y}_{E(m)} - C) \qquad (2.4.2)$$

for srswor. It is clear that $\bar{y}_C$ is a biased estimator of $\bar{Y}$ with bias

$$B(\bar{y}_C) = (\frac{E(m)}{N} - 1)(\bar{Y} - C). \qquad (2.4.3)$$

The bias decreases as the sample size increases or the guessed value C gets closer to $\bar{Y}$ and it is zero when the guess is perfect, i.e., $\bar{Y} = C$. We now prove

<u>Theorem 2.4.1</u>: Under the condition (2.4.1) $\bar{y}_C$ is more efficient than $\bar{y}_{E(m)}$ in the sense of having smaller mean square error (mse, for short).

<u>Proof:</u> Now

$$\text{mse}\ (\bar{y}_C) = \frac{E^2(m)}{N^2}(\frac{1}{E(m)} - \frac{1}{N})S^2 + (1 - \frac{E(m)}{N})^2(\bar{Y} - C)^2 \qquad (2.4.4)$$

and using (2.2.4) we get

$$V(\bar{y}_{E(m)}) - \text{mse}\ (\bar{y}_C) \geq \frac{(1 - f_o)^2}{f_o}(\bar{Y} - C)^2 \geq 0$$

where $f_o = \frac{E(m)}{N}$.

q.e.d.

Now we compare $\bar{y}_C$ with the estimator $\bar{y}_s$ in srswr which also uses the prior information (2.4.1). We prove

<u>Theorem 2.4.2:</u> The estimator $\bar{y}_C$ in srswor is more efficient than the estimator $\bar{y}_s$ in srswr if $E^2(m) \leq N(N-1)$.

<u>Proof:</u> Since

$$V(\bar{y}_s) = \left(\frac{1}{E(m)} - \frac{1}{N}\right)S^2 - \frac{V(m)}{E^2(m)}\left\{\frac{S^2}{N} - (\bar{Y} - C)^2\right\} \qquad (2.4.5)$$

from (2.4.4) and (2.4.5) we get

$$V(\bar{y}_s) - \text{mse}\ (\bar{y}_C) = aS^2 + b(\bar{Y} - C)^2 \qquad (2.4.6)$$

where

$$a = \frac{1}{E(m)} - \frac{1}{N} - \frac{V(m)}{NE^2(m)} - \frac{E(m)}{N^2} + \frac{E^2(m)}{N^3} \qquad (2.4.7)$$

and

$$b = \frac{V(m)}{E^2(m)} + \frac{2E(m)}{N} - \frac{E^2(m)}{N^2} - 1. \qquad (2.4.8)$$

If $a \geq 0$, using (2.4.1) we get -

$$\begin{aligned} V(\bar{y}_s) - \text{mse}\ (\bar{y}_C) &\geq (aN + b)(\bar{Y} - C)^2 \\ &= ((N - E(m))^2/NE(m))(\bar{Y} - C)^2 \\ &\geq 0. \end{aligned}$$

Hence to prove the theorem, it suffices to show that $a \geq 0$ if $E^2(m) \leq N(N-1)$. Using

$$V(m) \leq E(m)\left(1 - \frac{E(m)}{N}\right) \text{ in } (2.4.7) \text{ we have}$$

$$a \geq \frac{1}{E(m)} - \frac{1}{N} - \frac{1}{NE(m)} + \frac{1}{N^2} - \frac{E(m)}{N^2} + \frac{E^2(m)}{N^3}$$

$$= \frac{(N-E(m))}{N^2}\left\{\frac{N-1}{E(m)} - \frac{E(m)}{N}\right\}.$$

Therefore, $a \geq 0$ if

$$E^2(m) \leq N(N-1).$$

q.e.d.

Table 3 gives the percent gain in efficiency of $\bar{y}_C$ over $\bar{y}_s$ for selected values of N, n and $S^2/(\bar{Y} - C)^2 (>N)$. The table shows that, ordinarily, gains increase as N, n and $S^2/(\bar{Y} - C)^2$ increase. Moreover, the gains are very high.

Table 3: Percent gain in efficiency of $\bar{y}_C$ over $\bar{y}_s$.

N	$\frac{n}{.25}$ (rounded)				$\frac{n}{.20}$				$\frac{n}{.10}$			
$\frac{S^2}{(\bar{Y}-C)^2}$ / n	1.5N	2N	2.5N	3N	1.5N	2N	2.5N	3N	1.5N	2N	2.5N	3N
3	484	605	706	790	633	799	941	1064	1379	1785	2156	2497
4	491	614	716	802	639	808	953	1077	1385	1793	2167	2510
5	495	619	723	810	643	813	958	1084	1389	1798	2173	2518
6	497	623	727	815	646	816	963	1089	1392	1802	2177	2523
7	499	625	730	818	647	819	966	1093	1394	1804	2180	2526
8	501	627	732	821	649	821	968	1095	1395	1806	2183	2529
9	502	629	734	823	650	822	969	1098	1396	1808	2185	2531

When $a < 0$, using (2.4.1) one can easily show $b > 0$. Therefore from (2.4.6)

$$V(\bar{y}_s) - \text{mse}\ \bar{y}_C \geq 0 \quad \text{iff}$$

$$S^2/(\bar{Y}-C)^2 \leq N + (N - E(m))^2/(-a).N.E(m). \qquad (2.4.9)$$

Noting that the second term in (2.4.9) is greater than or equal to N, it follows that $\bar{y}_C$ will be more efficient than $\bar{y}_s$ if

$$S^2/(\bar{Y} - C)^2 \leq 2N .$$

For $a < 0$ and $S^2/(\bar{Y} - C)^2 \geq N + (N - E(m))^2/(-a)NE(m)$ Pathak's estimator $\bar{y}_s$ is better than $\bar{y}_C$.

In the above comparison we have ignored the fact that $E(m)$ is not necessarily an integer. However, as in Section 2, the comparison can be modified.

Remark 2.4.1: Since $E(m) < n$ it is clear that Theorem (2.4.2) is true if the sampling fraction $n/N < 1$, which will be usually satisfied in practice.

Remark 2.4.2: It is clear from the theorem of Seth and Rao [59] that their conclusions were based only on a sub-class of estimators in srswr defined by (2.1.3). We can consider a wider class of unbiased estimators in srswr than that considered in (2.1.3), namely,

$$\bar{y}_{f_i}(m) = \sum_{i=1}^{m} \frac{f_i(m)}{Ef_i(m)} \frac{y_i}{m}$$

where $f_1, f_2, \ldots, f_N$ are pre-assigned functions of m. It will be worthwhile to see whether the theorem of Seth and Rao [59] is true for this class of estimators.

2.5 On a paper [8] of Chaudhuri

In this section we discuss Chaudhuri's [8] paper in which he has made some comparisons between srswr and srswor. We show that most of his comparisons are either special cases of a theorem due to Seth and Rao [59] (and Rao [48]) or incorrect.

To avoid confusion, we follow Chaudhuri's notations closely and consider the following schemes of sampling from a finite population of N units $u(1), u(2), \ldots, u(N)$:

R: A fixed number of units is selected with srswr.

R* : A fixed number of units is selected with srswor.

R**: Selection with srswr is continued till the desired (fixed) number of distinct units is obtained.

Chaudhuri used the notation R for both R and R**, but we adopt different notations to avoid confusion. It is well-known that R* and R** are equivalent in the sense of having the same probability of selecting a sample of distinct units.

Suppose R** is based on r draws (random variable) such that exactly n(fixed) distinct units are realised $(r \geq n)$. Chaudhuri proposed the following estimator of the population mean $\mu = N^{-1} \sum_{1}^{N} x(i)$ for R**:

$$t^{**} = [N\{1-(1-\frac{1}{N})^r\}]^{-1} \sum_{\lambda \in s} x(\lambda) \qquad (2.5.1)$$

where the summation is over the distinct units in the sample s and $x(i)$ is the value of a character for the population unit $u(i)$. He claimed that, for a given $r(\geq n)$, R^{**} is identical to R based on r draws and, hence that t^{**} is a Horvitz- Thompson (H-T) estimator of μ since the probability of inclusion of $u(\lambda)$ for a given r, say $\pi(\lambda|r)$, would then be equal to $1-(1-1/N)^r$. However, this is obviously wrong because the number of distinct units for R is a random variable ν $(\leq r)$ whereas the number of distinct units for R^{**} (for a given r) is a fixed number n. Consequently, the estimator t^{**} is not the H-T estimator of μ (it is in fact biased for μ) and his conclusion that the variance of t^{**} is smaller than that of the sample mean $\bar{x}(n, R^*)$ based on n units for R^* is incorrect.

It may be of interest now to derive the H-T estimator of μ for R^{**} for a given $r(\geq n)$ - note it would be conditionally unbiased. The probability of drawing a set s_n of n distinct units in r draws for R^{**} is given by

$$P(s_n, r) = \Sigma' \frac{n(r-1)!}{t_1! \cdots t_{n-1}!} \frac{1}{N^r} = \frac{n}{N^r} \Delta^{n-1} 0^{r-1} \qquad (2.5.2)$$

where Σ' denotes summation over all integers $t_1, \ldots, t_{n-1}$ such that $t_i > 0$, $\Sigma t_i = r-1$ and $\Delta^p 0^k = [\Delta^p x^k]_{x=0}$ for positive integers p and k. Further, the probability distribution of r is

$$P(r) = \binom{N-1}{n-1} \frac{1}{N^{r-1}} \Delta^{n-1} 0^{r-1}, \quad r \geq n \qquad (2.5.3)$$

so that the conditional probability of s_n given r is

$$P(s_n | r) = \frac{P(s_n, r)}{P(r)} = \frac{1}{\binom{N}{n}} = P(s_n) \qquad (2.5.4)$$

which is independent of r. Similarly

$$P(r | s_n) = P(r) \qquad (2.5.5)$$

which is independent of s_n. Now

$$\pi(\lambda | r) = \sum_{s \supset \lambda} P(s|r) = \sum_{s_n \supset \lambda} \Sigma' P(s|r)$$

$$= \sum_{s_n \supset \lambda} P(s_n | r) = \frac{n}{N}, \quad \lambda = 1, \ldots, N \qquad (2.5.6)$$

where $\sum_{s \supset \lambda}$ denotes summation over all **samples** s of size r containing $u(\lambda)$. Consequently, the H-T estimator of μ for R** (for a given r) is in fact equal to the sample mean

$$\bar{x}\ (n,\ r,\ R^{**}) = \frac{1}{n} \sum_{\lambda \in s} x(\lambda) \qquad (2.5.7)$$

which is independent of r. Also the joint probability of inclusion of units $u(\lambda)$ and $u(\lambda')$ in the sample, for a given r, is

$$\pi(\lambda,\lambda'|r) = \sum_{s \supset \lambda,\lambda'} P(s|r) = \sum_{s_n \supset \lambda,\lambda'} P(s_n|r) = \frac{n(n-1)}{N(N-1)},$$

$$\lambda \neq \lambda' = 1, \ldots, N \qquad (2.5.8)$$

so that the variance of $\bar{x}(n,\ r,\ R^{**})$ is identical to that of the sample mean for R* based on n draws.

One could construct a class of unbiased (but not conditionally unbiased) estimators of R** which depend on r and $\sum_{\lambda \in s} x(\lambda)$ as follows:

$$f(n,\ r,\ R^{**}) = \frac{f(r)}{Ef(r)}\ \bar{x}\ (n,\ r,\ R^{**}) \qquad (2.5.9)$$

where $f(r)$ is a function of r. However, by using (2.5.5), we get

$$E[f(n, r, R^{**})|s_n] = \bar{x}(n, r, R^{**}) \tag{2.5.10}$$

so that $\bar{x}(n, r, R^{**})$ is uniformly better than $f(n, r, R^{**})$, by the Rao-Blackwell theorem.

We now turn to Chaudhuri's comparison of the H-T estimators for R and R*, both based on n draws. First, it must be pointed out that his statement 'with no loss of generality we shall take all $x(\lambda)$'s to be positive because if some $x(\lambda)$'s were positive and some negative then before undertaking the survey we might add a high positive value M to each value of the variate so that all the values become positive and then our problem will be to estimate $\mu + M$ of which the component M is known to us' is incorrect because the H-T estimator for R is not origin-invariant. Therefore, his results are valid only when all $x(\lambda) > 0$. He proved that the variance of the H-T estimator for R* (i.e., variance of the sample mean $\bar{x}(n, R^*)$) is smaller than that of the H-T estimator for R(namely, $\sum_{\lambda \in s} x(\lambda)/E(\nu)$ where $E(\nu) = N[1-(1-1/N)^n]$ and ν = number of distinct units in the R-sample) if $n \geq 3$ and N so large that $o(N^{-2})$ terms can be neglected. However, a more general result (in the sense of being valid for all N) follows as a special case of the following general theorem of Seth and Rao [59] (and Rao [48]) **noting** that all $x(\lambda) > 0 \Rightarrow (C.\ V.\ x)^2 < N$ and $V[\ \bar{x}(E(\nu),\ R^*)\ ] < V[\ \bar{x}(n,\ R^*)\ ]$, where $C.V.x$ = coefficient of variation of x:

Theorem [Seth and Rao [59]]: Variance of $\bar{x}\ [E(\nu), R^*]$ for R^* based on $E(\nu)$ draws is smaller than the variance of any estimator belonging to the class

$$f(\nu,\ n,\ R) = \frac{f(\nu)}{Ef(\nu)}\ \bar{x}(\nu,\ n,\ R) \qquad (2.5.11)$$

for R based on n draws (which includes the H-T estimator for R) if $(C.\ V.\ x)^2 < N$.

Next, Chaudhuri compared $\bar{x}[E(\nu),\ R^*]$ for R^* and $\sum_{\lambda \in s} x(\lambda)/E(\nu)$ for R so that the expected costs for R and R^* are equal. He proved that the variance of $\bar{x}[E(\nu),\ R^*]$ is smaller than that of $\sum_{\lambda \in s} x(\lambda)/E(\nu)$ for R to terms $o(N^{-3})$ provided all $x(\lambda) > 0$ and n is very small compared to N. However, a more general result (valid for all n and N) immediately follows as a special case of the above theorem of Seth and Rao. It is indeed strange that Chaudhuri states that the results of Rao [48] are 'asymptotic' and 'isolated' when in fact his own results are asymptotic (unlike Rao's) and special cases of Rao's.

Finally, we turn to Chaudhuri's comparison of R and R*, <u>both</u> based on the same number n (fixed) of distinct units. Clearly this comparison is impossible since the number of distinct units for R(a random variable) cannot be fixed. It appears that this mixing up of R and R** has led to this impossible comparison - R was based on m = E(r) draws where E(r) is the expected number of draws to get n distinct units in the sample for R**. It may be noted, however that one could make the expected number of distinct units in the sample equal to n for R by selecting d units with srswr where d is given by

$$N[1 - (1 - \frac{1}{N})^{d}] = n$$

$$\text{or} \quad d = [\ln (1 - n/N)] / [\ln (1 - 1/N)].$$

For this case also the comparison of H-T estimators for R and R* follows as a special case of the above theorem of Seth and Rao with R* and R based on n and d draws respectively.

CHAPTER III

COMPARISON OF SAMPLING WITH AND WITHOUT REPLACEMENT : UNEQUAL PROBABILITY SAMPLING

3.0 Summary

In this chapter, we have extended the modified efficiency comparison considered in the previous chapter to the case of unequal probability sampling. For unequal probability sampling, in the special case where the units are, or, can be grouped with respect to the selection probabilities p_i such that units in a group have the same p-value, Stevens' [63] estimator is shown to be better than that proposed by Pathak [38]. For the case of two-stage sampling with unequal probabilities we have derived without-replacement strategies that are uniformly more efficient than the with-replacement strategies considered by Pathak [39], for the same expected cost.

The usual modification of srswr which makes the customary ratio estimator unbiased for the population ratio is to choose the first unit in the sample with probability

proportional to size x (ppx, for brevity) and the remaining units with srswr from the whole population. Pathak [40] has shown that the ratio of the means for the distinct units in the sample has uniformly smaller variance than the ratio of the overall means. In the last section of this chapter, for the same expected cost, we get a more efficient estimator by using Midzuno-Sen scheme, under a super-population set-up.

3.1 Introduction

It is well-known that, under certain circumstances, selection of units with unequal probabilities provides more efficient estimators than equal probability sampling and this type of sampling is known as unequal or varying probability sampling. In the most commonly used varying probability sampling scheme, the units are selected with probability proportional to a given measure of size x (ppx) where the size measure x is the value of an auxiliary variable $\mathcal{X}$ related to the study characteristic $\mathcal{Y}$ and this type of sampling is termed 'probability proportional to size' sampling.

It is generally observed that sampling without replacement provides more efficient estimators than sampling with replacement, since the effective sample size is more in the former than in the latter. There has been tremendous development in the field of sampling with varying probabilities without replacement since 1950. But most of the suggested procedures, estimators and variance estimators are rather complicated and so they are not commonly used in practice, especially in large scale sample surveys with a small sampling fraction, since in such cases the efficiencies of sampling with and without replacement are not likely to differ much. Unequal probability sampling with replacement together with estimators based on all the units in the sample has been widely used in practice, mainly due to the simplicity of the estimators and their unbiased variance estimators. But it is worthwhile to compare varying probability sampling with and without replacement, for the same expected cost. As in the previous chapter we assume a linear cost function. It would be interesting, however, to make efficiency comparisons under reasonable non-linear cost functions.

For the special case of unequal probability sampling where the units are, or, can be grouped with respect to the selection probabilities p_i such that units in a group have the same p-value, Rao [48] has shown that, Stevens' [63] estimator is more efficient than that suggested by Pathak [38]. However this comparison ignores the fact that the number of units selected from a group in the case of Stevens' estimator need not be an integer. As in the earlier chapter we introduce a randomised estimator, and make modified efficiency comparison in the next section.

For the general case of unequal probability sampling, it has not been possible to make efficiency comparisons between estimators in sampling with and without replacement (based on the distinct units) since the comparisons depend on the y-values (and p-values too). Rao [48] has considered criteria other than sampling efficiency for the choice of estimators in sampling with and without replacement such as (a) the ease with which a sample can be drawn, (b) the simplicity of the estimator and (c) the availability of a non-negative variance estimator.

3.2 Unequal probabilities: Special case

Suppose the population is divided into k non-overlapping groups such that in the t-th group there are N_t units having the same p-value, ($\Sigma N_t = N$). Let n_t denote the number of units falling in the t-th group in a sample drawn with probabilities p_i and with replacement. For the groups with $n_t \geq 1$, let m_t denote the number of distinct units among the n_t units. Pathak [38] proposed the following unbiased estimator of Y:

$$\hat{Y}_1 = \frac{1}{n} \sum_{1}^{k'} \frac{n_t}{p_t} \bar{y}_{n_t} \qquad (3.2.1)$$

where k' is the number of groups with $n_t \geq 1$ and $\bar{y}_{n_t}$ is the mean of the n_t distinct units belonging to the t^{th} group. It can be shown that the estimator $\hat{Y}_1$ in (3.2.1) is more efficient than the conventional unbiased estimator in unequal probability sampling with replacement

$$\hat{Y}_2 = \frac{1}{n} \sum_{i=1}^{n} y_i / p_i$$

if and only if $n \geq 3$ and at least three population units have the same p-value (i.e., $N_t \geq 3$ for at least one group).

In the above set up, Stevens [63] has given a simple method of unequal probability sampling without replacement for which the estimator of Y is more efficient than $\hat{Y}_2$. Stevens' technique consists in selecting n groups with probabilities $N_i\, p_i$ with replacement and if the t^{th} group is selected n_t^* times, selecting n_t^* units from it with simple random sampling without replacement. To avoid the possibility of the non-applicability of Stevens' technique we assume that $N_t \geq n$, for all t. The unbiased estimator of Y considered is

$$\hat{Y}_3 = \frac{1}{n} \sum_{1}^{n} \frac{y_t}{p_t} .$$

It is easily seen that the joint distribution of $(n_1, n_2, \ldots, n_k)$ and $(n_1^*, n_2^*, \ldots, n_k^*)$ are identical, namely, multinomial with probabilities $N_t\, p_t$. Under the assumed linear cost function, it is evident that the expected cost in Stevens' procedure is greater than that in Pathak's procedure. Hence, in order to make efficiency comparisons Rao [48] modified Stevens' procedure as follows: If the t-th group is selected n_t^* times select

$$E(m_t^* \mid n_t^*) = N_t \left\{1 - \left(\frac{N_t - 1}{N_t}\right)^{n_t^*}\right\}.$$

units from it with srswor. The estimator considered is

$$\hat{Y}_4 = \frac{1}{n} \sum_{1}^{k^*} \frac{n_t^*}{p_t} \bar{y}_{E(m_t^* \mid n_t^*)}$$

where k^* is the number of groups with $n_t^* \geq 1$. Rao [48] has shown that the variance of $\hat{Y}_4$ is always smaller than the variance of $\hat{Y}_1$ assuming that $E(m_t^* \mid n_t^*)$ is an integer.

However, it can be easily seen that $E(m_t^* \mid n_t^*)$ need not be an integer and in such cases $\bar{y}_{E(m_t^* \mid n_t^*)}$ ceases to have any meaning. As in the previous chapter we modify the scheme by taking the sample sizes to be μ_t and μ_t+1 with probabilities $1 - f_t$ and f_t respectively

where

$$\mu_t = [E(m_t^* | n_t^*)], \quad f_t = E(m_t^* | n_t^*) - \mu_t$$

and $[x]$ denotes the integral part of x. Denoting the unbiased randomised estimator corresponding to the above modified scheme by $\hat{Y}_4^*$, we prove

<u>Theorem 3.2.1:</u> The variance of $\hat{Y}_4^*$ is uniformly smaller than the variance of $\hat{Y}_1$ in (3.2.1).

<u>Proof:</u> The variance of $\hat{Y}_4^*$ is

$$V(\hat{Y}_4^*) = \underset{n_t^*}{E}\, V(\hat{Y}_4^* | n_1^*, \ldots, n_k^*) + \underset{n_t^*}{V}\, E(\hat{Y}_4^* | n_1^*, \ldots, n_k^*)$$

$$= \underset{n_t^*}{E} \frac{1}{n^2} \sum_{1}^{k^*} \frac{n_t^{*2}}{p_t} \left(\frac{1 - f_t}{\mu_t} + \frac{f_t}{\mu_t + 1} - \frac{1}{N_t}\right) S_t^2 + \underset{n_t^*}{V} \left(\frac{1}{n} \sum_{1}^{k^*} \frac{n_t^*}{p_t} \bar{Y}_t\right) \tag{3.2.2}$$

where $\bar{Y}_t$ and S_t^2 are respectively the population mean and mean square for the t-th group. Now using inequality (2.2.7) we get

$$E(\frac{1}{m_t^*} \mid n_t^*) \geq \frac{1-f_t}{\mu_t} + \frac{f_t}{\mu_t+1} . \tag{3.2.3}$$

Using (3.2.3) in (3.2.2) we have

$$V(\hat{Y}_4^*) \leq \underset{n_t^*}{E} \frac{1}{n^2} \sum_1^{k^*} \frac{n_t^{*2}}{p_t^2} (E(\frac{1}{m_t^*} | n_t^*) - \frac{1}{N_t}) S_t^2$$

$$+ \underset{n_t^*}{V} (\frac{1}{n} \sum_1^{k^*} \frac{n_t^*}{p_t} \bar{Y}_t) . \tag{3.2.4}$$

Also

$$V(\hat{Y}_1) = \underset{n_t}{E} \frac{1}{n^2} \sum_1^{k'} \frac{n_t^2}{p_t^2} (E(\frac{1}{m_t} | n_t) - \frac{1}{N_t}) S_t^2$$

$$+ \underset{n_t}{V} (\frac{1}{n} \sum_1^{k'} \frac{n_t}{p_t} \bar{Y}_t). \tag{3.2.5}$$

As $(n_1, n_2, \ldots, n_k)$ and $(n_1^*, n_2^*, \ldots, n_k^*)$ have the same distribution, m_t and m_t^* have the same distribution as also k' and k^*. Hence comparing (3.2.4) and (3.2.5) we see that *the right hand side* of (3.2.4) is equal to $V(\hat{Y}_1)$.

q.e.d.

3.3 Two-stage sampling with unequal probabilities

Let $U_1, U_2, \ldots, U_N$ be the N first-stage units (fsu) of a population. Suppose that U_j consists of M_j second-stage units (ssu) and U_{jh} stands for the h^{th} ssu of U_j. Consider a two-stage design in which the fsu's are selected with probabilities P_j and with replacement and if U_j is selected λ_j times, λ_j subsamples of m_j units each are drawn there-from independently of each other by srswor. Denote $Z_{jh} = (Y_{jh}/P_j)(M_j/\Sigma M_j)$ as the Z-value of U_{jh}. Let $u_{(1)}, u_{(2)}, \ldots, u_{(d)}$ be the $d (\leq n)$ distinct fsu's in the sample arranged in the increasing order of their unit indices. Let $\lambda_{(i)}$ be the number of times $u_{(i)}$ is selected in the sample $(\Sigma \lambda_{(i)} = n)$. Also, let $u_{(i1)}, u_{(i2)}, \ldots, u_{(id_{(i)})}$ the $d_{(i)}$ $(\leq \lambda_{(i)} m_{u_{(i)}})$ distinct ssu's of $u_{(i)}$ arranged in increasing order of their unit indices. Pathak [39] has proposed the following estimator of $\bar{Y}$:

$$\bar{Z}_d^* = n^{-1} \Sigma \lambda_{(i)} \bar{Z}_{d_{(i)}}$$

where $\bar{Z}_{d_{(i)}} = [d_{(i)}]^{-1} \Sigma Z_{(ir)}$, $Z_{(ir)}$ being the Z-value of $u_{(ir)}$.

We now give an estimator more efficient than $\bar{Z}_d^*$ for the same expected cost, by using the following sampling scheme. Choose n fsu's as before. If $u_{(i)}$ is selected $\lambda_{(i)}$ times select n_3 or $n_3 + 1$ ssu's from it by srswor, with probabilities $1 - f_3$ and f_3 respectively where

$$n_3 = [E(d_{(i)} | \lambda_{(i)})], \quad f_3 = E(d_{(i)} | \lambda_{(i)}) - n_3.$$

Then the estimator

$$\bar{Z}_1 = n^{-1} \Sigma \lambda_{(i)} \bar{Z}^*_{E(d_{(i)})}$$

is an unbiased estimator of $\bar{Y}$, where $\bar{Z}^*_{E(d_{(i)})}$ denotes the mean of the Z-values of the units selected from $u_{(i)}$. We now prove

<u>Theorem 3.3.1:</u> The estimator $\bar{Z}_1$ is more efficient than the estimator $\bar{Z}_d^*$.

Proof: Now

$$V(\bar{Z}_1) = n^{-2} E \{\Sigma \lambda^2_{(i)} V(\bar{Z}^*_{E(d_{(i)})} | \lambda_{(i)}) \}$$

$$+ V \{n^{-1} \Sigma \lambda_{(i)} \bar{Z}_{(i)} \} \qquad (3.3.1)$$

where $\bar{Z}_{(i)} = M^{-1}_{(i)} \sum_{h=1}^{M_{(i)}} Z_{(ih)}$ and $M_{(i)}$ is the number of ssu's in $u_{(i)}$. Also

$$V(\bar{Z}^*_d) = n^{-2} E \{\Sigma \lambda^2_{(i)} V(\bar{Z}_{d_{(i)}} | \lambda_{(i)}) \}$$

$$+ V \{n^{-1} \Sigma \lambda_{(i)} \bar{Z}_{(i)} \}. \qquad (3.3.2)$$

Comparing (3.3.1) and (3.3.2) it suffices to show that

$$V(\bar{Z}^*_{E(d_{(i)})} | \lambda_{(i)}) \leq V(\bar{Z}_{d_{(i)}} | \lambda_{(i)})$$

which follows by an application of inequality (2.2.7).

q.e.d.

Pathak [39] has suggested another estimator for $\bar{Y}$ which is more efficient than $\bar{Z}^*_d$ but computationally more cumbersome:

$$\bar{Z}_d = \sum_{i=1}^{d} C_{(i)} \bar{Z}_{d_{(i)}}$$

where

$C_{(i)} = E[(\lambda_{(i)}/n) | T]$ and T is the

sufficient statistic defined by

$$T = [\{u_{(i)}; u_{(i1)}, \ldots, u_{(id_{(i)})}\} \quad i = 1,2,\ldots, d].$$

Theorem (3.3.2) follows immediately from Theorem (3.3.1).

Theorem 3.3.2: An estimator more efficient than $\bar{Z}_d$ is given by

$$Z_2 = \sum_{i=1}^{d} C_{(i)} \cdot \bar{Z}^*_{E(d_{(i)})} .$$

3.4 Comparison for Midzuno-Sen scheme

Let $y_1, \ldots, y_N$ denote the values of a characteristic y for the units in a finite population U and $x_1, \ldots, x_N$ be the corresponding values of an auxiliary characteristic x which are assumed to be known. Pathak [40] has considered the following sampling scheme:

Scheme A: Draw one unit from the population with probability proportional to x. Draw (n-1) units from the whole population using a srswr scheme.

Let m denote the number of distinct units in the sample. Let $\bar{y}$ and $\bar{x}$ denote the sample means of y and x and $\bar{y}_m$ and $\bar{x}_m$ denote the corresponding sample means of the m distinct units. Pathak [40] has shown that the estimator $\bar{y}_m / \bar{x}_m$ is uniformly more efficient than the estimator $\bar{y} / \bar{x}$ for estimating the population ratio $(\Sigma y_i)/(\Sigma x_i)$. Let E(m) stand for the expected number of distinct units when the sample is selected according to scheme A. Now consider the Midzuno-Sen Scheme which we shall denote by scheme B.

Scheme B: Draw one unit with ppx. Draw $n_1 - 1$ or n_1 units with probabilities $1 - f_1$ and f_1 respectively with simple random sampling without replacement from the remaining (N-1) units where $n_1 = [E(m)]$ and $f_1 = E(m) - n_1$ and [x] denotes the integral part of x.

Assuming, as before, that the cost of a sample is proportional to the number of distinct units in it, it can be easily checked that the expected costs for schemes A and B are equal. It is known that $(\bar{y} / \bar{x})$ is unbiased for the population ratio for scheme B. In the following we show that under a superpopulation model $\bar{y} / \bar{x}$ for scheme B is on the average more efficient than $\bar{y}_m / \bar{x}_m$ for scheme A, under certain conditions on the distribution of x.

To show that

$$V_B (\bar{y} / \bar{x}) \leq V_A(\bar{y}_m / \bar{x}_m) \qquad (3.4.1)$$

it is enough we show

$$E_B(\bar{y} / \bar{x})^2 \leq E_A (\bar{y}_m / \bar{x}_m)^2$$

$$= \underset{m}{E}\, E_A [(\bar{y}_m / \bar{x}_m)^2 | m].$$

i.e.

$$\frac{(1-f_1)}{\binom{N}{n_1}} \sum_{s=1}^{\binom{N}{n_1}} \frac{\bar{y}_s^2}{\bar{x}_s} + \frac{f_1}{\binom{N}{n_1+1}} \sum_{s=1}^{\binom{N}{n_1+1}} \frac{\bar{y}_s^2}{\bar{x}_s} \leq \underset{m}{E} \left\{ \frac{1}{\binom{N}{m}} \sum_{s=1}^{\binom{N}{m}} \frac{\bar{y}_m^2}{\bar{x}_m} \right\} \qquad (3.4.2)$$

where $\sum_{s=1}^{\binom{N}{r}}$ denotes summation over all samples of size r drawn with srswor from the N units.

Now consider the function

$$f(z) = \frac{1-z+[z]}{\binom{N}{[z]}} \sum_{1}^{\binom{N}{[z]}} \frac{\bar{y}^2_{[z]}}{\bar{x}_{[z]}} + \frac{z-[z]}{\binom{N}{[z]+1}} \sum_{1}^{\binom{N}{[z]+1}} \frac{\bar{y}^2_{[z]+1}}{\bar{x}_{[z]+1}},$$

$$1 \leq z \leq N.$$

If $f(z)$ is a convex function (3.4.2) will follow from an application of Jensen's inequality.

Now between two integers r and $r+1$ the graph of $f(z)$ is a straight line with slope

$$m_r = -\frac{1}{\binom{N}{r}} \sum_{s=1}^{\binom{N}{r}} \frac{\bar{y}^2_r}{\bar{x}_r} + \frac{1}{\binom{N}{r+1}} \sum_{s=1}^{\binom{N}{r+1}} \frac{\bar{y}^2_{r+1}}{\bar{x}_{r+1}}$$

$$r = 1,2,\ldots, N-1.$$

Hence the function f will be convex if

$$m_{r+1} \geq m_r, \quad r = 1,2,\ldots, N-1. \tag{3.4.3}$$

Since m_r **involves** $x_1,\ldots, x_N$ and also the unknown y_i's (3.4.3) may not be true in general. We shall now make the assumptions as in Rao and Webster [50] and P. S. Rao [52] and we further need the following

<u>Lemma (Rao and Webster [50]):</u> Let $Z_1, Z_2, \ldots, Z_n$ be independent gamma variates with parameter h. Then, for $i \neq j$,

$$E\left\{ \frac{Z_i^a \, Z_j^b}{(\Sigma Z_t)^c} \right\} = \frac{\Gamma(a+h) \cdot \Gamma(b+h)}{\Gamma^2(h)} \cdot \frac{1}{\prod_{t=1}^{c} (m+a+b-t)}$$

where $m = nh$, c is an integer greater than zero and a and b are integers greater than or equal to zero.

It can be easily seen that the lemma is true even when a and b are not integers.

We now have

<u>Theorem 3.4.1:</u> Under the assumptions that

i) The population U is itself a random sample from an infinite super-population with the model

$$y_i = \alpha + \beta x_i + e_i$$

where $E(e_i|x_i) = 0$, $E(e_i^2|x_i) = \sigma^2 x_i^g$ and $E(e_i e_j|x_i, x_j) = 0$

$i \neq j$ and

ii) $x_1, x_2, \ldots, x_N$ *are independently* distributed as gamma with same parameter h, the estimator $\bar{y} / \bar{x}$ corresponding to scheme B is, on the average, more efficient than the estimator $\bar{y}_n / \bar{x}_n$ for scheme A.

<u>Proof:</u> To prove the theorem it suffices to show that $E(m_r)$ is increasing i.e.

$$E(m_{r+1}) \geq E(m_r), \quad r = 1, 2, \ldots, N-1.$$

Now

$$\frac{\bar{y}_r^2}{\bar{x}_r} = \frac{(\alpha + \beta . \bar{x}_r + \bar{e}_r)^2}{\bar{x}_r} .$$

Taking expectation over the super-population for given $x_1, x_2, \ldots, x_N$, we get

$$E(m_r|x_1, \ldots, x_N) = \alpha^2 \left\{ - \frac{1}{\binom{N}{r}} \sum_{s=1}^{\binom{N}{r}} \frac{1}{\bar{x}_r} + \frac{1}{\binom{N}{r+1}} \sum_{s=1}^{\binom{N}{r+1}} \frac{1}{\bar{x}_{r+1}} \right\} + m_r'$$

where

$$m_r' = \sigma^2 \Big[-\frac{1}{r\binom{N}{r}} \sum_{s=1}^{\binom{N}{r}} \frac{\sum_{i \in s} x_i^g}{\sum_{i \in s} x_i} + \frac{1}{(r+1)\binom{N}{r+1}} \sum_{s=1}^{\binom{N}{r+1}} \frac{\sum_{i \in s} x_i^g}{\sum_{i \in s} x_i} \Big] \text{ and}$$

$\sum_{s=1}^{\binom{N}{r}}$ denotes summation over all samples of size r drawn with srswor from the N units. Again taking expectation over the distribution of $x_1, x_2, \ldots, x_N$ and using the lemma we get,

$$E(m_r) = \alpha^2 \left\{ -\frac{r}{rh-1} + \frac{r+1}{(r+1)h-1} \right\} + m_r''$$

$$\text{where } m_r'' = E(m_r') = \sigma^2 \frac{\Gamma(g+h)}{\Gamma(h)} \cdot \frac{(-h)}{(rh+g-1)(\overline{r+1}\; h+g-1)} .$$

One can easily check that $E(m_r)$ is increasing with r.

q.e.d.

CHAPTER IV

HYPER-ADMISSIBILITY

4.0 Summary

For any non-unicluster design, Hanurav [21] established the unique hyper-admissibility of the Horvitz-Thompson estimator (HT-estimator, for short) in the class of all general polynomial unbiased estimators of the population total Y. In this chapter we extend the above result to the class of all unbiased estimators of Y. Similar result is established for the variance estimator of it suggested by Horvitz and Thompson [26]. Further we show that the HT-estimator remains as the unique choice even if principal hyper-surfaces of lower dimensions are ignored. The new criterion has been termed as 'k-hyperadmissibility'.

4.1 Introduction

We have already mentioned earlier that a UMV estimator does not exist in the class $L_o^*(p)$ for any non-unicluster design. p. Godambe and Joshi [17] generalised this result to the class $A^*(p)$, of all unbiased estimators of Y. However, it may be mentioned that their result regarding the non-existence of a UMV estimator in $A^*(p)$ is true even for unicluster designs. An elegant and short alternative proof of the above result can be seen in Basu [7].

Let us recall the definition of admissibility. We have

Definition 4.1.1: In a class, $\mathcal{C}$, of unbiased estimators of Y, an estimator t_1 belonging to $\mathcal{C}$ is said to be admissible in $\mathcal{C}$ if for every other estimator t in $\mathcal{C}$, $V(t_1) < V(t)$ for at least one $y \in R_N$. The usual definition of admissibility used in decision theory (see Wald [65]) which is slightly different from the above definition is

Definition 4.1.2: In a class, $\mathcal{C}$, of unbiased estimators of Y, an estimator t_1 belonging to $\mathcal{C}$ is said to be admissible in $\mathcal{C}$, if for any $t \in \mathcal{C}$,

$$v(t) \leq v(t_1), \text{ for all } y \in R_N \Rightarrow v(t) = v(t_1),$$
$$\text{for all } y \in R_N.$$

We have also stated in Chapter I that when $\mathcal{C}$ is convex the above two definitions of admissibility are equivalent. So while stating any result on admissibility pertaining to convex class of estimators we will not distinguish between the above two definitions.

Godambe [15] and, Roy and Chakravarthy [54] proved that the HT-estimator $\bar{e}(s,y)$ defined in (1.1.18) is admissible in $L_0^*(p)$ for any design p, with $\pi_i > 0$ for all $1 \leq i \leq N$. Godambe and Joshi [17] generalised this result to the class, $A^*(p)$, of all unbiased estimators of Y and further they remarked that $\bar{e}(s,y)$ remains admissible even when the parameter space is restricted to any sphere containing the origin of R_N. They have also shown that a variance estimator, proposed by Horvitz and Thompson [26] is admissible in the class of all unbiased estimators of the variance of $\bar{e}(s,y)$.

The criterion of admissibility has, however, not been conclusive. In view of this and the non-existence of a UMVU estimator, the recent literature has been flooded with new criteria which give rise to a unique choice of estimator. We have broadly reviewed the related literature in Chapter I. It is worth noting here that, however, all of them have not led to the same choice. In this chapter we consider one such criterion, namely 'hyper-admissibility' due to Hanurav [21] which is stronger than admissibility and weaker than uniform minimum variance.

A principal hyper-surface (phs, for short) of R_N is defined as a linear subspace of all points $y = (y_1, y_2, \ldots, y_N)$ with

$$y_{i_1} = y_{i_2} = \ldots = y_{i_k} = 0$$

where $0 \leq k < N$ and $(i_1, i_2, \ldots, i_k)$ is a subset of $(1, 2, \ldots, N)$. Clearly the whole space R_N corresponds to the case $k = 0$, and there are, in all, 2^N-1 phs's of R_N. Let $\mathcal{C}$ be a class of unbiased estimators of Y.

Definition 4.1.3: (Hanurav [21]): $t_1 \in \mathcal{C}$, is hyper-admissible in $\mathcal{C}$, if it is admissible (Definition 4.1.2) in $\mathcal{C}$, when we restrict y to any of the 2^N-1 phs's of R_N.

For any non-unicluster design, Hanurav [21] established the unique hyper-admissibility of the HT-estimator $\bar{e}(s,y)$ in the class, $M^*(p)$, of all unbiased polynomial estimators of Y. Rao and Singh [51], following an approach which demonstrates the vital role played by the N phs's of dimension one, have proved that $\bar{e}(s,y)$ is uniquely hyper-admissible in the wider class of unbiased estimators

$$G = \bigcup_{n=1}^{\infty} G(n) \tag{4.1.1}$$

where $G(n)$ is defined by

$$G(n): t(s,y) = T_1 + T_2 + \dots + T_n \tag{4.1.2}$$

where $T_1 = \sum_{i \in s} e_i(s, y_i)$, $T_2 = \sum\sum_{i \neq j \in s} e_{ij}(s, y_i, y_j)$,

$T_3 = \sum\sum\sum_{i \neq j \neq k \in s} e_{ijk}(s, y_i, y_j, y_k)$ etc

such that $e_{ij\dots q}(s, y_i, y_j, \dots, y_q) = 0$ if y-value is zero for at least one unit in it. Further they have shown that this criterion of optimum choice leads to the Horvitz-Thompson estimator v_1 in (4.2.6) of variance of $\bar{e}(s, y)$

for a wide class of unbiased estimators.

Before proceeding further let us point out an error committed initially by Hanurav [21 ,23] and followed subsequently by Rao and Singh [51]. In the Definition (4.1.3) of hyper-admissibility, admissibility is understood in the sense of Definition (4.1.1). One can easily construct linear unbiased estimators ($\neq \bar{e}(s,y)$) which reduce to $\bar{e}(s, y)$ in some phs's of dimension one. So in such phs's $\bar{e}(s, y)$ is not admissible according to Definition (4.1.1) and as a result $\bar{e}(s, y)$ is not hyper-admissible according to Definition (4.1.3). However, the results of Hanurav [21] and, Rao and Singh [51] remain true if in Definition (4.1.3) of hyper-admissibility, admissibility is understood in the sense of Definition (4.1.2) and hereafter hyper-admissibility will be understood in this sense. The above-mentioned error pointed out to Hanurav in an oral discussion, has later been noted by Basu [7] also.

In the next section we prove the unique hyper-admissibility of $\bar{e}(s, y)$ in the class, $A^*(p)$, for any non-unicluster design p. We prove a similar result for the HT-estimator v_1 in (4.2.6) of the variance of $\bar{e}(s, y)$.

4.2 Unique hyper-admissibility of $\bar{e}(s, y)$

In order to prove the unique hyper-admissibility of $\bar{e}(s, y)$ in $A^*(p)$, for any non-unicluster design p, we first extend the result of Rao and Singh [51] to the class $G'(\supset G)$ defined by

$$G' : t'(s, y) = t(s, y) + k(s) \qquad (4.2.1)$$

where $t(s, y) \in G$ and $k(s)$'s are constants independent of y such that $\sum k(s)p(s) = 0$, for any non-unicluster design p. This result was obtained by the author in collaboration with Singh [60]. We prove

Theorem 4.2.1: For any non-unicluster design for which $\pi_i > 0$ for all i, $1 \leq i \leq N$, $\bar{e}(s, y)$ is the unique hyper-admissible estimator in the class G' in (4.2.1) of unbiased estimators of Y.

Proof: Consider $t'(s, y) \in G'$ and the i-th phs of dimension one i.e. $(0, 0, \ldots, 0, y_i, 0 \ldots 0)$. Inside this phs

$$t'(s,y) = \begin{cases} c_i(s, y_i) + k(s), & \text{if } s \in S_i \\ k(s), & \text{if } s \in \bar{S}_i \end{cases} \qquad (4.2.2)$$

where S_i is the set of samples containing the i-th unit and $\bar{S}_i$ is the complement of S_i. Now if $t'(s, y)$ is to be hyper-admissible, it should be admissible inside the i-th phs of dimension one. Thus $t'(s, y)$ should satisfy the necessary condition for being admissible in the i-th phs of dimension one (see Hanurav [23]). Noting that inside this phs S_i and $\bar{S}_i$ consist of effectively equivalent samples, we get from (4.2.2)

$$k(s) = \begin{cases} k_1(i), & \text{if} \quad s \in S_i \\ k_2(i), & \text{if} \quad s \in \bar{S}_i \ . \end{cases} \tag{4.2.3}$$

Since i is arbitrary, (4.2.2) and (4.2.3) hold for all i, $1 \leq i \leq N$, and hence

$$k(s) = \begin{cases} k_1(j), & \text{if} \quad s \in S_j \\ k_2(j), & \text{if} \quad s \in \bar{S}_j \ . \end{cases} \tag{4.2.4}$$

Now since the design is not unicluster, it can be easily seen that there exist samples s_1, s_2 and units i_o, j_o such that

$$s_1 \in S_{i_o} \cap S_{j_o} \quad \text{and} \quad s_2 \in S_{i_o} \cap \bar{S}_{j_o} \ .$$

Making use of this and (4.2.3) with $i = i_o$ and (4.2.4) with $j = j_o$ it follows that $k_1(j_o) = k_2(j_o)$. That is $k(s) = k$, a constant for all $s \in S$. Now from the condition of unbiasedness of $t'(s, y)$, we have $\Sigma\, k(s)p(s) = 0$ which implies $k(s) = 0$ for all $s \in S$.

This proves that any possible hyper-admissible estimator in G' must belong to G. The proof of our theorem is complete by invoking the result of Rao and Singh [51] stated in the previous section.

q.e.d.

Rao and Singh [51] have considered the following class of unbiased estimators of $V(\bar{e}(s,y))$ given by $G_v = \bigcup_{m=1}^{\infty} G_v(m)$, where an estimator in $G_v(m)$ is expressible as

$$v(s,y) = \sum_{i \in s} v_i(s,y_i) + \sum\sum_{i \neq j \in s} v_{ij}(s,y_i,y_j) + \sum\sum\sum_{i \neq j \neq k \in s} v_{ijk}(s,y_i,y_j,y_k) + \ldots \qquad (4.2.5)$$

where $v_{ij\ldots q}(s,y_i,y_j,\ldots,y_q)$ vanish when at least one y-value of the units in it is zero. They have shown that the HT-estimator of $V(\bar{e}(s,y))$ given by

$$v_1 = \sum_{i \in s} \frac{1-\pi_i}{\pi_i^2} y_i^2 + \sum\sum_{i \neq j \in s} \frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j \pi_{ij}} y_i y_j \qquad (4.2.6)$$

where $\pi_{ij} = \sum_{s \supset i,j} p(s)$, is the unique hyper-admissible estimator in G_v. Now we consider a wider class of unbiased estimator $G_v^!$ given by

$$G_v^! : v'(s, y) = v(s, y) + k(s) \qquad (4.2.7)$$

where $v(s,y) \in G_v$ and $k(s)$'s are constants independent of y_i's such that $\sum k(s)p(s) = 0$. Along the lines of Theorem (4.2.1), the following theorem can be established.

Theorem 4.2.2: v_1 is uniquely hyper-admissible in the class $G_v^!$ of unbiased estimators of $V(\bar{e}(s,y))$.

We are now in a position to prove our main result contained in the following

Theorem 4.2.3: For any non-unicluster design p, $\bar{e}(s,y)$ is the unique hyper-admissible estimator in the class $A^*(p)$ of all unbiased estimators of Y.

Proof: First we shall prove that any possible hyper-admissible estimator in $A^*(p)$, belongs to G'. Consider $e(s,y) \in A^*(p)$ and the i-th hyper-surface of dimension one i.e. $y^{(i)} = (0,\ldots, 0, y_i, 0 \ldots 0)$. Since S_i and $\bar{S}_i$ consist of effectively equivalent samples, for $e(s,y)$ to be hyper-admissible it has to be admissible in the $y^{(i)}$ plane and using the necessary condition for this, we have,

$$e(s,y) = \begin{cases} k(y_i), \text{ a constant for } & s \in S_i \\ k_1 \quad\quad \text{ for } & s \in \bar{S}_i, \end{cases}$$

where k_1 is some constant independent of y_i.

Now, using the condition of unbiasedness

$$k(y_i)\pi_i + k_1(1-\pi_i) = y_i$$

which gives $k(y_i) = \dfrac{y_i}{\pi_i} - \dfrac{k_1(1-\pi_i)}{\pi_i}$.

$\therefore$ In the $y^{(i)}$ plane

$$e(s,y) = \bar{e}(s,y) + k(s) \text{ where } k(s) = \begin{cases} \dfrac{-k_1(1-\pi_i)}{\pi_i}, & \text{if } s \in S_i \\ k_1, & \text{if } s \in \bar{S}_i. \end{cases}$$

(4.2.8)

Since i is arbitrary (4.2.8) is true for $1 \leq i \leq N$.

Now consider a typical hyper-plane of dimension two, say

$$y^{(i,j)} = (0,\ldots,0,y_i,0\ldots 0, y_j, 0\ldots 0).$$

Noting that, in this phs, $S_i \cap S_j$, $S_i \cap \bar{S}_j$, $\bar{S}_i \cap S_j$ and $\bar{S}_i \cap \bar{S}_j$ consist of effectively equivalent samples, for $e(s,y)$ to be admissible in the $y^{(i,j)}$ plane,

$$e(s,y) = \begin{cases} k(y_i,y_j), & \text{a constant if } s \in S_i \cap S_j \\ \bar{e}(s,y) + k_2(i), & \text{if } s \in S_i \cap \bar{S}_j \\ \bar{e}(s,y) + k_2(j), & \text{if } s \in \bar{S}_i \cap S_j \\ k_3 & \text{if } s \in \bar{S}_i \cap \bar{S}_j \end{cases}$$

where $k_2(i)$, $k_2(j)$ and k_3 are constants independent of the values y_i and y_j. (For $s \in S_i \cap \bar{S}_j$, $e(s,y)$ is independent of y_j and we already know that in the $y^{(i)}$ plane

$e(s,y) = \bar{e}(s,y) + k_2(i)$ where $k_2(i)$ is independent of y_i etc.). It can be easily seen that $k_2(i) = k_2(j) = k_3$, for by putting $y_i = 0$, $(S_i \cap \bar{S}_j) \cup (\bar{S}_i \cap \bar{S}_j)$ consists of effectively equivalent samples and hence $k_2(i) = k_3$

and similarly $k_2(j) = k_3$.

Using the condition of unbiasedness, we get

$$k(y_i,y_j)\pi_{ij} + (\pi_i - \pi_{ij})(\frac{y_i}{\pi_i} + k_3) + (\pi_j - \pi_{ij})(\frac{y_j}{\pi_j} + k_3)$$
$$+ (1 - \pi_i - \pi_j + \pi_{ij})k_3 = y_i + y_j$$

which gives

$$k(y_i,y_j) = \frac{y_i}{\pi_i} + \frac{y_j}{\pi_j} - \frac{1-\pi_{ij}}{\pi_{ij}} k_3.$$

∴ In the $y^{(i,j)}$ plane $e(s,y)$ reduces to

$e(s,y) = \bar{e}(s,y) + k(s)$

$$\text{where} \qquad k(s) = \begin{cases} -\frac{1-\pi_{ij}}{\pi_{ij}} k_3, & \text{if } s \in S_i \cap S_j \\ k_3, & \text{if } s \in \bar{S}_i \cup \bar{S}_j . \end{cases} \qquad (4.2.9)$$

Since i and j are arbitrary (4.2.9) is true for $1 \leq i \neq j \leq N$. Similarly for hyper-plane of higher order also it can be shown that

$$e(s,y) = \bar{e}(s,y) + k(s)$$

where $k(s)$'s are constants independent of y_i's such that $\Sigma\, k(s)p(s) = 0$.

Hence any possible hyper-admissible estimator $e(s,y)$ is of the form

$$e(s,y) = \bar{e}(s,y) + k(s) \quad \text{where} \quad \Sigma\, k(s)p(s) = 0$$

$$\text{i.e.} \quad e(s,y) \subset G'.$$

That in fact, $\bar{e}(s,y)$ is the unique hyper-admissible estimator in $A^*(p)$ follows as a consequence of Theorem (4.2.1) and the fact that $\bar{e}(s,y)$ is hyper-admissible in $A^*(p)$ which follows from Godambe and Joshi [17] (see Rao and Singh [51]).

q.e.d.

Along the lines of the proof of the above theorem, we can establish the following theorem regarding the HT-variance estimator v_1 in (4.2.6).

Theorem 4.2.4: For any non-unicluster design p, v_1 is the unique hyper-admissible estimator in the class of all unbiased estimators of $V(\bar{e}(s,y))$.

4.3 k-hyper-admissibility

We have proved in the proceding section that, 'hyper-admissibility' as a criterion of optimum choice leads to a unique choice of an estimator, namely $\bar{e}(s,y)$, in the class of all unbiased estimators of Y, for any non-unicluster design p. For a unicluster design Hanurav [22] has shown that an unbiased estimator $e(s,y)$ for Y is hyper-admissible if and only if $e(s,y) = \bar{e}(s,y) + k(s)$ where $\Sigma\, k(s)p(s) = 0$ and $k(s_1) = k(s_2)$ if s_1 and s_2 contain the same units. As a justification for the criterion of 'hyper-admissibility' Hanurav [23] has stated that, in practice, often one is interested in estimating not only Y but also the totals and means of sub-populations or 'domains' and that these totals or means would all be admissibly estimable by means of a single estimator, if we start with a hyper-admissible estimator. But this justification has been criticised on the ground that, it is unrealistic to consider all the 2^N-1 possible 'domains' (or phs's) because the number of domains of practical interest will be much smaller than 2^N-1 and the sizes of such domains will be fairly large, although the actual number of units in a domain may not be known. Further, we can note the vital role played by the domains of size one (i.e. the phs's of dimension one) in the above proofs in arriving at a unique hyper-admissible estimator (see Rao and Singh [51] and Basu [7]). On Professor J. N. K. Rao's suggestion, the author made some investigations regarding the consequences of dropping phs's of lower dimensions. The following results are the outcomes of these investigations. The proof of Theorem (4.3.1) is due to Hanurav. We make the following definition.

Definition 4.3.1: For any positive integer $k \leq N$, an estimator t belonging to a class of unbiased estimators of Y is said to be k-hyper admissible (k-HA, for short) in that class if it is admissible (Definition 4.1.2) in every phs of dimension greater than or equal to k.

It follows from the definition that if an estimator is k_o-HA then it is k-HA for any $k \geq k_o$. Note also that 1-hyper admissibility is the same as hyper-admissibility.

For any design $d(S,p)$ let

$$k_1 = \min_{s_1, s_2 \in S} \{ N - n(s_1) - n(s_2) + n(s_1 \cap s_2) + 1 \} \qquad (4.3.1)$$

and

$$k_2 = \min_{s_1, s_2 \in \bar{S}} \{ N - n(s_1) - n(s_2) + 2n(s_1 \cap s_2) \} \qquad (4.3.2)$$

where $n(s)$ denotes the number of distinct units in s and $\bar{S}$ is the set of all samples s for which $p(s) > 0$. ~~Clearly $k_2 \geq k_1$.~~ In practice, k_1 and k_2 will be moderately large since $n(s)$ will be small compared to N. We now prove

Theorem 4.3.1: For any design p, $\bar{e}(s,y)$ is the unique k_1-HA estimator in $L_o^*(p)$, where k_1 is as defined in (4.3.1).

Proof: Consider $t = \sum_{i \in s} b(s,i) y_i \in L_o^*(p)$.

If t is k_1-HA then we shall show that $b(s,i)$ is independent of s, i.e. $i \in s_1 \cap s_2 \Rightarrow b(s_1,i) = b(s_2,i)$. Consider the phs for which $y_j = 0$, for all $j \neq i \in s_1 \cup s_2$. Clearly the dimension of this phs is greater than or equal to k_1. If t is k_1-HA, then it should be admissible in the above phs. Since s_1 and s_2 are effectively equivalent in this phs we have $t_{s_1} = t_{s_2}$ which gives $b(s_1,i) = b(s_2,i)$. Therefore $b(s,i)$ is independent of s, say $b(i)$. Using the condition of unbiasedness of t we see $b(i) = \frac{1}{\pi_i}$. Hence the only possible k_1-HA estimator in $L_o^*(p)$ is given by $\bar{e}(s,y)$. That, in fact, it is k_1-HA follows from its hyper-admissibility.

q.e.d.

Considering the phs for which $y_j = 0$, for all $j \in s_1 \cup s_2 - s_1 \cap s_2$, one can prove ~~a stronger result contained in~~

<u>Theorem 4.3.2</u>: For any design p, $\bar{e}(s,y)$ is the unique k_2-HA estimator in $L_o^*(p)$, where k_2 is as defined in (4.3.2).

<u>Remark 4.3.1:</u> One can consider the following class of estimators

$$L_o^{*\prime}(p) \; : \; t'(s,y) = t(s,y) + k(s)$$

where $t(s,y) \in L_o^*(p)$ and $k(s)$'s are constants independent of the y_i''s such that $\Sigma\, k(s)p(s) = 0$. Theorem (4.3.1) and (4.3.2) can be shown to be true for $L_o^{*\prime}(p)$, if p is a non-unicluster design.

<u>Remark 4.3.2:</u> For a unicluster design it is easy to show that an estimator $t'(s,y) \in L_o^{*\prime}(p)$ is k_1-HA or k_2-HA if and only if it is of the form

$t'(s,y) = \bar{e}(s,y) + k(s)$ where $k(s)$'s are constants not depending on y_i's such that $\Sigma\, k(s)p(s) = 0$ and $k(s_1) = k(s_2)$ if s_1 and s_2 contain the same units.

Remark 4.3.3: If $k_2 > k_1$ - which is the case for e.g. if the design p is such that no two samples with positive probabilities are disjoint-then one can easily see that Theorem (4.3.2) is a stronger version of Theorem (4.3.1).

Now we generalise the results contained in Theorems (4.3.1) and (4.3.2) to the class G in (4.1.1) considered by Rao and Singh [51]. We prove

Theorem 4.3.3: For any design p, $\bar{e}(s,y)$ is the unique k_1-HA estimator in G.

Proof: Let $t(s,y)$ as given by (4.1.2) be a typical element of G and s_1, s_2 be two samples containing the unit i. Consider the phs for which $y_j = 0$ for all $j \neq i \in s_1 \cap s_2$. Clearly the dimension of this phs is greater than or equal to k_1.

If t is k_1-HA, then it should be admissible in the above phs and hence for any y belonging to this phs

$$e_i(s_1, y_i) = t(s_1, y) = t(s_2, y) = e_i(s_2, y_i)$$

since s_1 and s_2 are equivalent in this phs. So it follows that $e_i(s,y_i)$ is independent of s, say $e_i(y_i)$. Now using the condition of unbiasedness of t we get

$$e_i(y_i) = \frac{y_i}{\pi_i}.$$

Similarly considering the phs for which $y_k = 0$ for all $k \neq i, j \in s_1 \cup s_2$ and using the necessary condition for admissibility of t in this phs we can see that $e_{ij}(s,y_i,y_j)$ is independent of s, say $e_{ij}(y_i,y_j)$ and the condition of unbiasedness of t gives $e_{ij}(y_i,y_j) = 0$. One can similarly show that

$$e_{ij\ldots q}(s,y_i,y_j,\ldots,y_q) = 0, \text{ for all } s \in S.$$

Thus we have shown that any k_1-HA estimator in G is necessarily of the form

$$t(s,y) = \sum_{i \in s} \frac{y_i}{\pi_i}$$

which is nothing but $\bar{e}(s,y)$. That, in fact, it is k_1-HA in G, follows from its hyper-admissibility in G.

q.e.d.

As before we can prove a stronger result contained in

<u>Theorem 4.3.4</u>: $\bar{e}(s,y)$ is the unique k_2-HA estimator in G for any design p.

For any non-unicluster design we can extend the results contained in Theorems (4.3.3) and (4.3.4) to the class G' defined by (4.2.1). For a unicluster design it is possible to show that an estimator $t'(s,y) \in G'$ is k_1-HA or k_2-HA if and only if it is of the form $t'(s,y) = \bar{e}(s,y) + k(s)$ where $k(s)$'s are constants not depending on y_i's such that $\Sigma k(s)p(s) = 0$ and $k(s_1) = k(s_2)$ is s_1 and s_2 contain the same units.

Proofs of the above statements are omitted as they are straight-forward and follow a routine pattern.

CHAPTER V

LINEAR SUFFICIENCY

5.0 Summary

Godambe [18] defined 'linear sufficiency' for survey sampling and has shown that for any design p with fixed sample size n for which $p(s) > 0$ for all the $\binom{N}{n}$ samples, the unique unbiased, linear sufficient estimator for the population total which also satisfies the principle of censoring is given by

$$e_b = [\binom{N-1}{n-1} \; p(s)]^{-1} \sum_{i \in s} y_i \quad .$$

In this chapter, we show that Godambe's definition of linear sufficiency is different from the original definition of linear sufficiency due to Barnard [3] which is also applicable to survey sampling. It is shown that the HT-estimator is an unbiased linear sufficient estimator in the class, L(p), of all linear estimators of the population total.

That it need not be unique is illustrated with an example. Under an alternative definition, non-existence of a linear sufficient estimator in L(p) is established. Consequently it is noted that there does not exist a uniformly minimum mean square error estimator in L(p). After comparing the two definitions of linear sufficiency in the Gauss-Markov set-up we consider extensions to wider classes of estimators.

5.1 Introduction

In the previous chapter we considered the criterion of hyper-admissibility due to Hanurav [21] for the choice of an optimum estimator. In this chapter we consider two other criteria, namely, linear sufficiency and distribution-free sufficiency introduced by Godambe [18] in survey sampling. We have defined a sample, s, as a finite sequence of units i from the population U and a design as a probability measure on the set of all samples from U. Following the arguments of Basu [4] and Hajek [20], Godambe and Joshi [17] have pointed out that, so far as the problem of estimation is concerned, without any loss of generality, we can restrict our attention to estimators which depend on s only through the set of distinct units in s. Hence without any loss of generality we can define a sample as a nonempty subset of units drawn from the population and a design as a probability measure on the set of all nonempty subsets of U and in this chapter a sample and a design will be understood in this sense. We now give

Definition 5.1.1: If n(s) denotes the total number of units i such that $i \in s$, then p is said to be a fixed sample size (=n) design if for every s

$$n(s) \neq n \implies p(s) = 0.$$

As shown by Godambe [18], the concept of a linear estimator is more general in survey sampling than in the general statistical theory and his celebrated negative result mentioned in Theorem (1.1. 2) clearly points out the inapplicability of the Gauss-Markov set-up for the problem of estimation in sampling from finite populations. This made Godambe [18] redefine the concept of linear sufficiency - originally introduced by Barnard [3] in connection with the Gauss-Markov set-up of estimation - to suit the problem of estimation in survey sampling.

For the sake of ready reference let us recall some definitions. Any real-valued function $e\,(s,\,y)$ on $S \times R_N$ which depends on y through only those y_i for which $i \in s$, is called an estimator. It is said to be linear if it is of the form

$$e_b\,(s,\,y) = \sum_{i=1}^{N} b(s,\,i)\,y_i \qquad (5.1.1)$$

where $b(s,\,i) = 0$ if $i \notin s$. As before for any given design p, let $L(p)$ and $A(p)$ denote the class of all linear estimators and the class of all estimators respectively and let $L_0^*(p)$ and $A^*(p)$ denote the corresponding classes of unbiased estimators for the population total Y. A design p is said to be unicluster if and only if

$s_1, s_2 \in S$, $p(s_1) > 0$, $p(s_2) > 0 \implies s_1 \cap s_2 = \emptyset$, i.e. s_1 and s_2 do not contain any unit in common.

5.2 Definitions

Before proceeding to give the various definitions due to Godambe [18], we wish to make a remark about the following fundamental assumption made by him.

Assumption (Godambe [18]): The vector $y = (y_1, y_2, \ldots, y_N)$ of the variate values associated with different units of the population is such that whatever may be our knowledge about some of the co-ordinates of y, it cannot impart any knowledge about any of the remaining co-ordinates of y. It is not clear what exactly is meant by this assumption. If it means that there is no mathematical relation among the various co-ordinates of y, it is clear from our basic assumption - namely, R_N is our parameter space - that the above assumption is redundant, for, if there is a mathematical relation among the various co-ordinates of y, then the parameter space will not be the whole of R_N but only a subset of it. Hence we will not bother about the above assumption of Godambe [18]. We now give the various definitions due to him.

Definition 5.2.1: Any two linear functions of the co-ordinates of $y = (y_1, y_2, \ldots, y_N)$, say,

$$gy = \sum_{i=1}^{N} g_i y_i \quad \text{and} \quad g'y = \sum_{i=1}^{N} g_i' y_i,$$

are said to be independent of each other whenever the vectors g and g' are orthogonal, i.e. whenever $gg' = 0$ which means $\sum_{i=1}^{N} g_i g_i' = 0$.

We can write the linear estimator e_b in (5.1.1) in a form of scalar product of two vectors as

$$e_b\,(s,\ y) = b(s)y \quad (s \in S\ ,\ y \in R_N),$$

where $b(s)$ denotes the vector with its i-th co-ordinate

$b(s, i)$ as in (5.1.1). Godambe [18] then gave

Definition 5.2.2: For a given sampling design p, two estimators e_b and $e_{b'}$ are said to be independent of each other if, for all samples with $p(s) > 0$, the corresponding vectors $b(s)$ and $b'(s)$ are orthogonal.

Definition 5.2.3: For a given sampling design p, a linear estimator e_b is said to be linearly sufficient for the linear function gy if (i) it is unbiased for gy and (ii) for every other estimator $e_{b'}$ which is independent of e_b (Definition 5.2.2), the expected value of $e_{b'}$ is a linear function $g'y$ independent (Definition 5.2.1) of the linear function gy.

Godambe [18] has stated that the above definition of linear sufficiency includes Barnards [3] definition of linear sufficiency. Before proceeding further we give Barnard's [3]

Definition 5.2.4: The (vector) available observable T is linearly sufficient for the unknown (vector) θ if the true value of any available observable U which is orthogonal to T is orthogonal to θ.

[In the above definition an available observable means a linear estimator of the form $\sum_{i=1}^{n} \lambda_i y_i$, true value means expected value, orthogonality between two available observables means that they are uncorrelated and two vectors of unknowns are orthogonal means that no linear compound of one is expressible as a linear compound of the other].

It can be easily checked that Definition (5.2.2) of independence is equivalent to uncorrelation, in the Gauss-Markov set up. But that they are not equivalent in general can be easily seen by considering the model $y = A\theta + \epsilon$ where ϵ is distributed with mean zero and dispersion matrix Σ which is not diagonal. It follows that two correlated estimators can be independent according to Definition (5.2.2) due to Godambe [18]. To give a simple example in survey sampling, consider a population of size $N = 3$ and the following design p:

sample s		probability p(s)
1	2	.5
2	3	.5
		$\Sigma\, p(s) = 1.$

It can be easily checked that the estimators

$$T_1(1, 2) = y_1 + y_2 \qquad T_2(1, 2) = y_2 - y_1$$
$$\text{and}$$
$$T_1(2, 3) = y_2 + y_3 \qquad T_2(2, 3) = y_2 - y_3$$

are independent (Definition 5.2.2). But they are not uncorrelated since

$$\text{cov}(T_1, T_2) = -\frac{1}{4}(y_1 - y_3)^2.$$

Clearly Godambe's definition of independence of estimators looks artificial. The same artificiality is present in his Definition (5.2.3) of linear sufficiency also. Before stating Godambe's [18] main result we give

Definition 5.2.5: The principle of censoring: For a given fixed sample size design p any inference about the population parameter y should be exclusively in terms of the observation $(s; y_i, i \in s)$ and its probability $p(s)$. That is, the inference should not depend on the probabilities of the undrawn samples s. Godambe [18] has proved

Theorem 5.2.1: For any fixed sample size (=n) design p having $p(s) > 0$ for all the $\binom{N}{n}$ samples, the only unbiased estimator of the population total which is *linearly sufficient* (Definition 5.2.3) and which satisfies the principle of censoring is given by

$$e_{\bar{b}} = \frac{1}{\binom{N-1}{n-1} p(s)} \sum_{i \in s} y_i \, . \qquad (5.2.1)$$

He recommended the use of $e_{\bar{b}}$ for every fixed sample size ($= n$) design. In particular, when $p(s)$ is proportional to $\sum_{i \in s} x_i$, where $x_1, \ldots, x_N$ are the known values of an auxiliarly variable $\mathcal{X}$ highly correlated with $\mathcal{Y}$, $e_{\bar{b}}$ is nothing but the Lahiri's estimator

$$\hat{Y}_L = (\sum_{i \in s} y_i / \sum_{i \in s} x_i)(\sum_{i=1}^{N} x_i) .$$

Rao [49] has shown that under the usual model

$$y_i = \beta x_i + e_i , \quad 1 \leq i \leq N$$

where $E(e_i | x_i) = 0$, $E(e_i^2 | x_i) = a x_i^g$, $E(e_i e_j | x_i, x_j) = 0$

$$a > 0, \; g \geq 0$$

$$E \, V (\hat{Y}_L, y) \geq E V (\bar{e}, y)$$

where $\bar{e}$ is the HT-estimator. This result of Rao [49] shows that the suggestion of $e_{\bar{b}}$ for every fixed sample size design is not sound.

Below, we give yet another definition of linear sufficiency and compare it with Barnard's Definition (5.2.4) in Section (5.5). The motivation for this definition stems from an exercise in Rao [46 , page 250] which gives a logical justification of least square estimation in Gauss-Markov set-up without appealing to unbiasedness, linearity of the estimator or minimum variance.

Definition 5.2.6: A linear estimator T is said to be linearly sufficient for a (vector) parameter θ, if for every linear estimator U,

$$E(U) \equiv 0 \implies Cov\ (T, U) \equiv 0.$$

5.3 Barnard's definition and allied results

In survey sampling, the parametric function of interest is a scalar, namely, the population total Y and the Definition (5.2.4) of linear sufficiency becomes

<u>Definition 5.3.1</u>: For a given sampling design p, a linear estimator $e_b \in L(p)$ is said to be linearly sufficient for the population total if

$$e_{b'} \in L(p),\ \mathrm{Cov}\,(e_b, e_{b'}) \equiv 0 \implies E(e_{b'}) \equiv 0.$$

First we establish the following

<u>Lemma 5.3.1</u>: For any design p, given any estimator $e_b \in L(p)$ there exists another $e_{b'}\,(\neq 0) \in L(p)$ such that $\mathrm{Cov}\,(e_b, e_{b'}) \equiv 0$ if

$$\sum_{s \in \bar{S}} n(s) > \frac{N(N+1)}{2} \qquad (5.3.1)$$

where $n(s)$ is the number of units i such that $i \in s$ and $\bar{S}$ is the set of all samples s for which $p(s) > 0$.

<u>Proof</u>:

Now

$$\mathrm{Cov}\,(e_b, e_{b'}) \equiv 0 \iff$$

$$\sum_{s \in S_i} b(s,i)b'(s,i)p(s) - a_i \sum_{s \in S_i} b'(s,i)p(s) = 0,$$

$$1 \leq i \leq N$$

$$\sum_{s \in S_i \cap S_j} b(s,i)b'(s,j)p(s) + \sum_{s \in S_i \cap S_j} b(s,j)b'(s,i)p(s)$$

$$- a_i \sum_{s \in S_j} b'(s,j)p(s) - a_j \sum_{s \in S_i} b'(s,i)p(s) = 0,$$

$$1 \leq i \neq j \leq N,$$

where $a_i = \sum_{s \in S_i} b(s,i)p(s)$, $1 \leq i \leq N$ and S_i is the set of all samples containing the i^{th} unit. The above forms a set of linear equations in $\{ b'(s,i): i \in s, s \in \bar{S} \}$. Note that the number of unknowns is equal to $\sum_{s \in \bar{S}} n(s)$ and the number of independent equations, say m, is less than or equal to $N(N + 1)/2$.

q.e.d.

<u>Remark 5.3.1:</u> It may be noted that for fixed sample size $(= n)$ designs, for which $p(s) > 0$ for all the $\binom{N}{n}$ samples, (5.3.1) is satisfied if $N \geq 4$. For the design corresponding to srswor of size 2 from a population of size 3, one can easily check that there are linear estimators for which there do not exist any other non-zero linear estimator which is identically uncorrelated with it.

Note: It is clear from the lemma that $\sum_{s \in \bar{S}} n(s)$ should be greater than m for the applicability of Definition (5.3.1). For the sake of convenience we assume that $\sum_{s \in \bar{S}} n(s) > N(N+1)/2$ for the rest of this section.

We now prove

Theorem 5.3.1: For any design p for which $0 < \pi_i < 1$, $1 \leq i \leq N$, any estimator of the form

$$e_b(s,y) = \sum_{i \in s} b(i)\, y_i$$

where $b(i) \neq 0$, $1 \leq i \leq N$, is linearly sufficient according to Definition (5.3.1).

Proof: Let $e_{b'} \in L(p)$ be such that $\text{Cov}\,(e_b, e_{b'}) \equiv 0$. Now the coefficient of y_i^2 in $\text{Cov}\,(e_b, e_{b'})$ is equal to $b(i)(1-\pi_i) \sum_{s \in S_i} b'(s,i)p(s)$. Since $b(i) \neq 0$ and $0 < \pi_i < 1$, $1 \leq i \leq N$, we get

$$\text{Cov}\,(e_b, e_{b'}) \equiv 0 \implies \sum_{s \in S_i} b'(s,i)p(s) = 0, \quad 1 \leq i \leq N.$$

q.e.d.

Imposing unbiasedness criterion we get

<u>Theorem 5.3.2:</u> For any design p, the HT-estimator $\bar{e}(s,y)$ is an unbiased linear sufficient (Definition 5.3.1) estimator for the population total in the class, $L(p)$, of all linear estimators.

That it need not be the unique unbiased linearly sufficient estimator in $L(p)$, can be illustrated with an example. Consider the following design and estimator for a population of two units.

Sample s	Probability p(s)	Estimator T(s)
1	1/3	$\frac{3}{2} x_1$
2	1/3	$3x_2$
1, 2	1/3	$\frac{3}{2} x_1$

One can easily check that the above estimator T (which is not identically equal to $\bar{e}(s,y)$) is unbiased and linearly sufficient (Definition 5.3.1) for the population total.

5.4 Alternative definition and allied results

In the case of survey sampling the alternative Definition (5.2.6) of linear sufficiency becomes

Definition 5.4.1: For a given sampling design p, a linear estimator $e_b \in L(p)$ is said to be linearly sufficient for the population total if

$$e_{b'} \in L(p),\ E(e_{b'}) \equiv 0 \implies \text{Cov}\ (e_b, e_{b'}) \equiv 0.$$

Throughout this section 'linearly sufficient' will be understood in the sense of Definition (5.4.1). We now prove

Theorem 5.4.1: For any design p, a necessary condition for the estimator $e_b \in L(p)$, to be linearly sufficient for the population total is that

$$b(s,i) = b(i), \quad \text{for all} \quad s \in S_i, \quad 1 \leq i \leq N.$$

In other words, any linearly sufficient estimator is of the form

$$e_b(s, y) = \sum_{i \in s} b(i) y_i.$$

<u>Proof:</u> Suppose there exist samples $s_1, s_2 \in \bar{S}$ and a unit $i_o \in s_1 \cap s_2$ such that $b(s_1, i_o) \neq b(s_2, i_o)$. We shall show that e_b is not linearly sufficient by exhibiting an $e_{b'} \in L(p)$, satisfying

$$E(e_{b'}) \equiv 0 \qquad \text{and} \qquad Cov\,(e_b, e_{b'}) \neq 0.$$

Define a new set of coefficients $\{ b'(s,i);\ i \in s,\ s \in S \}$ as follows:

$$b'(s,i) = 0 \text{ for all } (s,i) \text{ except } (s_1, i_o) \text{ and } (s_2, i_o).$$

$b'(s_1,i_o)$ and $b'(s_2,i_o)$ are respectively chosen as the solutions y_o and z_o of the linear equations

$$yp(s_1) + zp(s_2) = 0$$

$$b(s_1,i_o)\, yp(s_1) + b(s_2,i_o)\, zp(s_2) = K$$

where K is an arbitrary non-zero real number. (Note that the above equations are consistent due to the assumption $b(s_1, i_o) \neq b(s_2, i_o)$). It is easy to check that the estimator defined by

$$e_{b'}(s,y) = \sum_{i \in s} b'(s, i)\, y_i$$

is such that

$$E(e_{b'}) \equiv 0 \quad \text{and} \quad Cov\,(e_b, e_{b'}) \not\equiv 0.$$

q.e.d.

For a unicluster design one can easily see that any estimator of the form $\sum_{i \in s} b(i)y_i$ is linearly sufficient. Imposing the condition of unbiasedness we get

Theorem 5.4.2: For a unicluster design, p, the HT-estimator $\bar{e}(s,y)$ is the unique unbiased linearly sufficient estimator for the population total in the class, L(p), of all linear estimators.

For a non-unicluster design we prove the following interesting negative result.

Theorem 5.4.3: If p is a non-unicluster design, there does not exist a linearly sufficient estimator $e_b(s, y)$, in the class L(p), which depends on y through every y_i for which $i \in s$.

[Though it is clear what is meant by the underlined part of the above theorem, for the sake of exactness, let us explain it. An estimator $e(s, y)$ is said to be independent of a unit i_o if there exists a sample $s \in \bar{S}$ such that $i_o \in s$ and $e(s, y)$ is unaltered by changing y_{i_o} alone (i.e. keeping all the other co-ordinates fixed). If there does not exist any unit i for which $e(s, y)$ is independent of i, we say that $e(s, y)$ depends on y through every y_i for which $i \in s$. Clearly a linear estimator

$$\sum_{i \in s} b(i)\, y_i$$

is dependent on y through every y_i for which $i \in s$, if and only if $b(i) \neq 0, \; 1 \leq i \leq N$.]

Proof: First we shall show that for a non-unicluster design p, any estimator of the form

$$e_b(s, y) = \sum_{i \in s} b(i) y_i, \quad (b(i) \neq 0, \; 1 \leq i \leq N)$$

is not linearly sufficient. As in Theorem (5.4.1) the proof here is again accomplished by exhibiting an $e_{b'} \in L(p)$ such that

$$E(e_{b'}) \equiv 0 \quad \text{and} \quad \mathrm{Cov}\,(e_b, e_{b'}) \neq 0.$$

Since the design is non-unicluster there exist samples $s_1, s_2 \in \bar{S}$ and units i_o, j_o such that

$$s_1 \in S_{i_o} \cap S_{j_o}, \quad s_2 \in \bar{S}_{i_o} \cap S_{j_o}.$$

Define

$$b'(s,i) = 0, \text{ for all } (s,i) \text{ other than } (s_1, j_o) \text{ and } (s_2, j_o).$$

$b'(s_1, j_o)$ and $b'(s_2, j_o)$ are respectively chosen as the solutions y_o and z_o of the equations

$$y\,p(s_1) + z p(s_2) = 0$$

$$b(i_o)\; y\; p(s_1) = K$$

where K is an arbitrary non-zero real number. As before one can easily check that the estimator

$$e_{b'}(s,y) = \sum_{i \in s} b'(s,i) y_i$$

is such that $E(e_{b'}) \equiv 0$ and $\mathrm{Cov}\,(e_b, e_{b'}) \not\equiv 0$. The **proof of** Theorem (5.4.3) is complete by invoking Theorem (5.4.1).

q.e.d.

It is worthwhile to note that in the case of unbiased estimators the concepts of 'linear sufficiency' and 'minimum variance' are equivalent. So from Theorem (5.4.3) it follows that for any non-unicluster design p, there does not exist a UMV estimator in the class $L_0^*(p)$ of all linear unbiased estimators of the population total. Godambe [14] has already proved this result using a different approach and we have stated it in Theorem (1.1.2). But our approach, in fact, establishes the more general result contained in

Theorem (5.4.4): For a non-unicluster design p, there does not exist a uniformly minimum mean square error estimator in $L'(p)$.

In order to prove the theorem we require the following lemma.

Lemma 5.4.1: A necessary condition for T_0 to be a minimum mean square error estimator of a parameter θ, is that it is linearly sufficient.

Proof: Let T_0 be a minimum mean square error estimator for θ and Z be a zero-function

$$\text{i.e.,} \quad E(Z) \equiv 0.$$

We shall show that T_o and Z are uncorrelated. Since T_o is a minimum msc estimator we have

$$M(T_o) \leq M(T_o + \epsilon Z), \text{ for all } \epsilon.$$

i.e. $$V(T_o) + B^2(T_o) \leq V(T_o + \epsilon Z) + B^2(T_o + \epsilon Z), \text{ for all } \epsilon \quad (5.4.1)$$

where V and B stand for variance and bias respectively. As $B(T_o) = B(T_o + \epsilon Z)$ since Z is a zero-function, (5.4.1) reduces to

$$2\epsilon \text{ Cov } (T_o, Z) + \epsilon^2 V(Z) \geq 0, \text{ for all } \epsilon$$

which can be true only if Cov $(T_o, Z) \equiv 0$.

q.e.d.

Proof of Theorem (5.4.4): Immediately follows from Lemma (5.4.1) and Theorem (5.4.3).

q.e.d.

For a unicluster design p, Hege [24] and Hanurav [21] have shown that the HT-estimator $\bar{e}(s,y)$ is the UMV estimator in $L_o^*(p)$. It may be noted that this result also follows from Theorem (5.4.2).

5.5 Comparison of the two definitions

In this section we propose to compare Barnard's Definition (5.2.4) and the alternative Definition (5.2.6) of linear sufficiency in the Gauss-Markov set up. Let

$$y = A\,\theta + \epsilon$$

where A is the design matrix of order $n \times k$ with full rank $(k \leq n)$, θ is the $(k \times 1)$ vector parameter and ϵ, the error vector of order $k \times 1$, distributed with mean 0 and dispersion matrix $\sigma^2 I$. Barnard [3] has shown that the conventional least-square estimate of θ, namely

$$T_2 = (A'A)^{-1} A'y \qquad (5.5.1)$$

is linearly sufficient (Definition 5.2.4) for θ. Moreover, since, when variables are jointly normal orthogonality is equivalent to statistical independence, it is evident that, with the normality assumption, the notion of linear sufficiency becomes equivalent to the ordinary notion of sufficiency. (See Barnard [3]). That this is not the case with the alternate Definition (5.2.6) will follow as a consequence of

Theorem 5.5.1: In the Gauss-Markov set-up $y = A\theta + \epsilon$.

where $\epsilon \frown N(0, \sigma^2 I)$, the only sufficient statistic which is linearly sufficient (Definition 5.2.6) for θ is given by the minimal sufficient statistic T_2 in (5.5.1).

Proof: Let T be any linearly sufficient statistic according to the alternate Definition (5.2.6). We shall show that $T = T_2$ if T is sufficient for θ in the conventional sense, which will prove the above theorem.

Now any statistic U which is independent of T_2 has expection zero, since T_2 is linearly sufficient (Definition 5.2.4). Hence such a U is independent of T since T is linearly sufficient (Definition 5.2.6) by our assumption. Let T_2 and T be written in the form

$$T_2 = \begin{bmatrix} \lambda_1' \\ \lambda_2' \\ \vdots \\ \lambda_k' \end{bmatrix} y \qquad T = \begin{bmatrix} m_1' \\ m_2' \\ \vdots \\ m_k' \end{bmatrix} y$$

where λ_i and m_j, $1 \leq i \leq k$, $1 \leq j \leq r$, are $n \times 1$ vectors, and η' denotes the transpose of η. Also let $\mathcal{V}$ be the vector space generated by the column vectors $\lambda_1, \lambda_2 \ldots \lambda_k$. If for some j $(= 1,2,\ldots, r)$, $m_j \notin \mathcal{V}$ write

$$m_j = m_{j1} + m_{j2}$$

where $m_{j1} \in \mathcal{V}$ and $m_{j2} \in \mathcal{V}^{\perp}$, where $\mathcal{V}^{\perp}$ stands for the orthogonal complement of $\mathcal{V}$. Clearly $m_{j2}' y$ and T_2 are independent but $m_{j2}' y$ and T are not independent-a contradiction. Hence $m_j \in \mathcal{V}$, $1 \leq j \leq r$. That is components of T are linear functions of components of T_2. Therefore if T is sufficient then $T = T_2$ since T_2 is known to be minimal sufficient.

q.e.d.

Remark 5.5.1: A linearly sufficient (Definition 5.2.6) estimator need not be sufficient in the ordinary sense. For example, take $T = \lambda_i' y$, any component of T_2. Clearly

$$E(U) = 0 \implies \text{Cov}(T, U) = 0.$$

Therefore T is linearly sufficient (Definition 5.2.6) but is clearly not sufficient for θ.

Remark 5.5.2: A sufficient statistic for θ need not be linearly sufficient according to the alternate Definition (5.2.6). For example let T be any statistic bigger than T_2, say,

$$T = \begin{pmatrix} T_2 \\ T_3 \end{pmatrix}$$

where components of T_3 belong to $\mathcal{U}^{\perp}$. It is easy to check that T is sufficient but not linearly sufficient (Definition 5.2.6) for θ.

5.6 Extensions

In part II of his paper, Godambe [18] has extended his definition of linear sufficiency - which he termed distribution-free sufficiency - to non-linear estimators. This extension is based on the assumption that our prior knowledge (denoted by K) of y could be formulated as a class of prior distributions. One can easily observe that the above assumption is necessary for the extension because his Definition (5.2.2) of independence of two linear estimators is not applicable to the case of non-linear estimators.

Godambe [18] has shown that the linear estimator e_b in (5.2.1) is distribution-free sufficient (DF-sufficient, for short) for the population total. He has claimed that every DF-sufficient estimator must necessarily be of the form

$$a(s) \sum_{i \in s} y_i + b(s)$$

where $a(s)$ and $b(s)$ are constants independent of y. Joshi [30] has shown that the above claim is false by showing that an estimator is DF-sufficient if and only if it depends on y through every y_i for which $i \in s$. Clearly, the complete class of DF-sufficient estimates is too large which shows the limitations of this concept. Joshi [30] has shown that Godambe's claim remains valid with an alternative definition of DF-sufficiency. However, as has been pointed out in Section (5.2) for the case of linear sufficiency, the author feels that, both Godambe's and Joshi's definitions of DF-sufficiency contain some artificiality due to the fact that two correlated estimators can be K-independent according to their definitions (See Godambe [18] and Joshi [30] for the various definitions).

The Definitions (5.3.1) and (5.4.1) of linear sufficiency can readily be applied to non-linear estimators and in what follows we investigate the consequences of it. For any design p, first we consider the non-homogeneous class, $L'(p)$, of linear estimators defined by

$$L'(p) : e_b'(s, y) = e_b(s, y) + k(s)$$

where $e_b(s, y) \in L(p)$ and $k(s)$'s are constants independent of y. We now prove

Theorem 5.6.1: For any design p, a necessary condition for the estimator $e_b^{\prime} \in L^{\prime}(p)$, to be linearly sufficient (Definition 5.4.1) for the population total is that

$$b(s, i) = b(i), \text{ for all } s \in S_i, \quad 1 \leq i \leq N,$$

and

$$k(s) = 0, \quad \text{for all } s \in \bar{S}.$$

In other words, any linearly sufficient (Definition: 5.4.1) estimator is of the form

$$e_b^{\prime}(s, y) = \sum_{i \in s} b(i) y_i .$$

Proof: Suppose there exists samples $s_1, s_2 \in \bar{S}$ and a unit $i_o \in s_1 \cap s_2$ such that

$$b(s_1, i_o) \neq b(s_2, i_o). \tag{5.6.1}$$

We shall show that $e_b^{\prime}$ is not linearly sufficient by exhibiting an $e_{\alpha}^{\prime} \in L^{\prime}(p)$, satisfying

$$E(e_{\alpha}^{\prime}) \equiv 0 \quad \text{and} \quad \text{Cov}\,(e_b^{\prime}, e_{\alpha}^{\prime}) \not\equiv 0.$$

Define a new set of coefficients $\{\alpha(s, i);\ i\in s,\ s\in\overline{S}\}$ as follows:

$\alpha(s, i) = 0$ for all (s, i) except (s_1, i_o) and (s_2, i_o).

$\alpha(s_1, i_o)$ and $\alpha(s_2, i_o)$ are respectively chosen as the solutions y_o and z_o of the linear equations

$$yp(s_1) + zp(s_2) = 0$$

and $$yb(s_1, i_o)p(s_1) + zb(s_2, i_o)p(s_2) = C$$

where C is an arbitrary non-zero real number. [Note that the above equations are consistent due to (5.6.1)]. It is easy to check that the estimator

$$e_\alpha(s, y) = \sum_{i\in s} \alpha(s, i)y_i$$

where $\alpha(s, i)$'s are defined as above belongs to $L'(p)$ and that $E(e_\alpha) \equiv 0$. Also it is easy to see that

$$\mathrm{Cov}\ (e_b', e_\alpha)\Big|_{(0,\ldots 0, y_{i_o}, 0\ldots 0)} = C\, y_{i_o}^2$$

which is not equal to zero if $y_{i_o} \neq 0$. Hence it follows that any linearly sufficient (Definition 5.4.1) **estimator is of** the form

$$e_b'(s, y) = \sum_{i \in s} b(i) y_i + k(s).$$

Now we shall show that $k(s) = 0$, for all $s \in \bar{S}$. Suppose $k(s_1) \neq 0$. Define constants $\{ L(s);\ s \in \bar{S} \}$ as follows:

$$L(s) = 0, \quad \text{for all} \quad s \neq s_1, s_2$$

$L(s_1)$ and $L(s_2)$ **are chosen** so as to satisfy

$$L(s_1)p(s_1) + L(s_2)p(s_2) = 0 \qquad (5.6.2)$$

and

$$L(s_1)k(s_1)p(s_1) = C \qquad (5.6.3)$$

where C is an arbitrary non-zero real number.

[Note that the above choice is possible since $k(s_1) \neq 0$].

Clearly the estimator

$$t(s, y) = L(s)$$

belongs to $L'(p)$ and moreover from (5.6.2) we have $E(t) \equiv 0$.

Also

$$\operatorname{Cov}(e_b', t) \Big|_0 = k(s_1)L(s_1)p(s_1)$$

$$= C \qquad \text{from } (5.6.3)$$

which is not equal to zero. Hence any linearly sufficient (Definition 5.4.1) estimator in $L'(p)$ is of the form

$$e_b'(s,y) = \sum_{i \in s} b(i) y_i .$$

q.e.d.

From Theorems (5.6.1) and (5.4.3) we have, immediately the following

Theorem 5.6.2: For any non-unicluster design, p, there does not exist a linearly sufficient (Definition 5.4.1) estimator, e_b', in the class $L'(p)$, which depends on y through every y_i for which $i \in s$.

Using Theorem (5.6.1) one can easily check that Theorem (5.6.2) is true even for unicluster designs. This result together with Lemma (5.4.1) yields immediately

Remark 5.6.1: For any design p, there does not exist a uniformly minimum mean square error estimator in $L'(p)$.

For any design p, let $G_o(p)$ denote the class of all estimators $e(s, y)$ such that

$$e(s, y) = 0 \quad \text{if} \quad y_i = 0, \quad \text{for all} \quad i \in s.$$

For a given k $(0 \leq k \leq N)$ let $R_N(k)$ be the subset of R_N defined by

$$R_N(k) = \{ y \in R_N : y_i \neq 0 \text{ for exactly } i \text{ co-ordinates} \} \tag{5.6.4}$$

It is clear that $\bigcup_{k=0}^{N} R_N(k) = R_N$. For a given $y \in R_N(k)$ let $S_i(y)$ be the set of samples defined by

$$S_i(y) = \{ s \in \overline{S} : y_i \neq 0 \text{ for exactly } i \text{ units in } s \} \tag{5.6.5}$$

Clearly $\bigcup_{i=0}^{k} S_i(y) = \overline{S}$. We now prove

<u>Theorem 5.6.3</u>: A necessary condition for an estimator $e(s, y) \in G_o(p)$ to be linearly sufficient (Definition 5.4.1) is that for any k $(0 \leq k \leq N)$ and any $y \in R_N(k)$

$$e(s_1, y) = e(s_2, y), \quad \text{for every } s_1, s_2 \in S_k(y),$$

where $R_N(k)$ and $S_k(y)$ are as defined in (5.6.4) and (5.6.5) respectively.

<u>Proof</u>: Suppose there exist a k $(0 < k \leq N)$, a

$y' = (y_1', y_2', \ldots, y_i', \ldots, y_N') \in R_N(k)$ and samples $s_1, s_2 \in S_k(y')$ such that

$$e(s_1, y') \neq e(s_2, y').$$

We shall show that e is not linearly sufficient (Definition 5.4.1) by exhibiting a $t \in G_o(p)$ such that

$$E(t) \equiv 0 \quad \text{and} \quad \text{Cov}\,(e, t) \not\equiv 0.$$

Define $t(s,y) = 0$, for all $s \neq s_1, s_2$, and for all $y \in R_N$.

For any $y \in R_N$ such that $y_i = y_i'$ for all the k non-zero co-ordinates of y', define

$$t(s_1, y) = -\,c/(e(s_2, y') - e(s_1, y'))\; p(s_1)$$

and
$$t(s_2, y) = c/(e(s_2, y') - e(s_1, y'))\,p(s_2)$$

where c is an arbitrary non-zero fixed real number. For remaining $y \in R_N$, define

$$t(s_1, y) = t(s_2, y) = 0.$$

Since $t(s,y)$ depends on y only through those y_i's for which $i \in s$, it is clear that t is an estimator. Moreover directly from the definition of t, we have

$$E(t) = 0, \text{ for all } y \in R_N$$

and at any point y for which $y_i = y_i'$ for the k non-zero co-ordinates of y' we have

$$\mathrm{Cov}\ (e,\ t)\Big|_y = C \neq 0.$$

q.e.d.

We now extend the result contained in Theorem (5.4.3) to the class $G_o(p)$, in the following

Theorem 5.6.4: For any non-unicluster design p there does not exist a linearly sufficient (Definition 5.4.1) estimator in $G_o(p)$, which depends on y through every y_i for which $i \in s$.

Proof: From Theorem (5.6.3), it suffices to show that any estimator $e(s, y) \in G_o(y)$ such that $e(s, y)$ is a constant for any $y \in R_N(k)$ and $s \in S_k(y)$ is not linearly sufficient in $G_o(p)$. Now since the design is non-unicluster there exist units i_o, j_o and samples s_o, s_1 such that

$$s_0 \in S_{i_0} \cap S_{j_0}\,, \quad s_1 \in \overline{S}_{i_0} \cap S_{j_0}$$

where S_i is the set of samples in $\overline{S}$ containing unit i and $\overline{S}_i$ stands for the complement of S_i with respect to $\overline{S}$. Define an estimator t as follows:

$$t(s,y) = \begin{cases} -\dfrac{\pi_{j_0} - \pi_{i_0 j_0}}{\pi_{i_0 j_0}}\, y_{j_0} & \text{if } s \in S_{i_0} \cap S_{j_0} \\ y_{j_0} & \text{if } s \in \overline{S}_{i_0} \cap S_{j_0} \\ 0 & \text{otherwise}. \end{cases}$$

It is clear that $E(t) = 0$, for all $y \in R_N$. Considering the hyperplane $y^{(2)} = (0,\ldots,0,y_{i_0},0,\ldots,0,y_{j_0},0,\ldots,0)$ and using the assumption that $e(s,y)$ is a constant for any $y \in R_N(k)$ and $s \in S_k(y)$, we can write

$$e(s,y^{(2)}) = \begin{cases} e(y_{i_0},y_{j_0}) & \text{if } s \in S_{i_0} \cap S_{j_0} \\ e(y_{j_0}) & \text{if } s \in \overline{S}_{i_0} \cap S_{j_0} \\ e(y_{i_0}) & \text{if } s \in S_{i_0} \cap \overline{S}_{j_0} \\ 0 & \text{if } s \in \overline{S}_{i_0} \cap \overline{S}_{j_0}. \end{cases}$$

Direct calculation yields

$$\text{Cov }(e,t)\Big|_{y(2)} = y_{j_o}(\pi_{j_o}-\pi_{i_o j_o})(e(y_{j_o})- e(y_{i_o},y_{j_o})). \tag{5.6.6}$$

Noting that $\pi_{j_o} > \pi_{i_o j_o}$ and using the assumption that $e(s,y)$ depends on y through every y_i for which $i \in s$ it is clear that (5.6.6) cannot be equal to zero for every value of y_{i_o} and y_{j_o}.

q.e.d.

<u>Remark 5.6.2:</u> The underlined condition in Theorem (5.6.4) is crucial for its validity. One can easily construct linearly sufficient (Definition 5.4.1) estimators which do not depend on y through every y_i for which $i \in s$.

<u>Remark 5.6.3:</u> From Lemma (5.4.1) and Theorem (5.6.4) immediately follows the non-existence of a uniformly minimum mean square error estimator for Y in the class $G_o(p)$. Noting that if there exists a uniformly minimum mean square error estimator <u>in</u> $A(p)$, then it should belong to $G_o(p)$, we get the non-existence of a uniformly minimum mean square error estimator in $A(p)$.

Analogous to Theorem (5.6.2) we can prove

Theorem 5.6.5: For any non-unicluster design there does not exist a linearly sufficient (Definition 5.4.1) estimator in the class $G_o'(p)$ defined by

$$G_o'(p): e'(s,y) = e(s,y) + k(s) \quad \text{where} \quad e(s,y) \in G_o(p)$$

and $k(s)$'s are constants independent of y.

Theorem (5.6.5) can be shown to be true even for unicluster designs. We omit the proofs. So far we considered extensions for the case of alternative Definition (5.4.1) of linear sufficiency. Now we shall try to extend Barnard's Definition (5.3.1) for the case of non-linear estimators. For the linear case, we have shown that, for any design p for which $0 < \pi_i < 1$, $1 \leq i \leq N$, any estimator of the form $e_b(s,y) = \sum_{i \in s} b(i)y_i$, $(b(i) \neq 0,\ 1 \leq i \leq N)$ is linearly sufficient (Definition 5.3.1). Clearly the corresponding generalisation of this result to the class $G_o(p)$ is as follows: Is an estimator $e(s,y) \in G_o(p)$ satisfying

i) $e(s_1,y) = e(s_2,y)$, for any $y \in R_N(k)$ and $s_1, s_2 \in S_k(y)$

and ii) $e(s,y)$ depends on y through every y_i for which $i \in s$

linearly sufficient (Definition 5.3.1) in $G_o(p)$, for any p for which $0 < \pi_i < 1$, $1 \leq i \leq N$? *We give below a simple* counter-example to illustrate the falsity of this generalisation.

Counter-example: Consider a population of size $N = 2$ with the design p and estimators T and U as follows:

Sample s	Probability p(s)	Estimator e	t
1	$\frac{1}{3}$	$\frac{3y_1}{2}$	$\frac{3y_1^2}{2}$
2	$\frac{1}{3}$	$\frac{3y_2}{2}$	$\frac{3y_2^2}{2}$
1,2	$\frac{1}{3}$	$\frac{3(y_1+y_2)}{2}$	$\frac{3}{2}(3y_1y_2 - y_1^2 - y_2^2)$ if $y_1+y_2 \neq 0$ and 0 if $y_1+y_2 = 0$

One can easily check that

$$E(e) = y_1 + y_2$$

$$E(t) = \begin{cases} \frac{3}{2} y_1 y_2 & \text{if } y_1 + y_2 \neq 0 \\ \frac{1}{2}(y_1^2 + y_2^2) & \text{if } y_1 + y_2 = 0 \end{cases}$$

$$\text{Cov}(e, t) = 0, \text{ for all } (y_1, y_2).$$

Clearly e and t belong to $G_o(p)$ and e satisfies the conditions (i) and (ii) mentioned above. But e is not linearly sufficient (Definition 5.3.1).

CHAPTER VI

CHOICE OF AN OPTIMUM STRATEGY

6.0 Summary

In this chapter we discuss the problem of an optimum sampling strategy from the class $L(H, \mu)$ of all equi-cost linear unbiased strategies. It is shown that any Horvitz-Thompson strategy (HT-strategy, for short) $(d, \bar{e}(s,y))$ belonging to $L(H, \mu)$ is admissible in $L(H, \mu)$ and that the class $L(HT, \mu)$, of all such strategies, is not complete in $L(H, \mu)$ in situations of practical interest. The last result shows that we cannot exclude strategies other than the HT-strategies from the point of view of minimum variance criterion alone. After noting that there does not exist a hyper-admissible strategy in $L(H, \mu)$, we have introduced a new criterion called 'strong admissibility' which is stronger than admissibility and weaker than hyper-admissibility. It is shown that $L(HT, \mu)$ is complete in $L(H, \mu)$ with respect to strong admissibility. Extensions to very wide classes of unbiased strategies are also given.

6.1 Introduction

So far we were mostly concerned with the choice of an 'optimum' estimator for a given design. But the central problem of survey sampling, as has been pointed out in Chapter I, is not merely to choose an optimum estimator for a given design but to choose an optimum combination of a design and an estimator (i.e. a strategy) subject to a fixed budget. As before we assume a linear cost function and this implies that two strategies are equally costly if and only if they have the same expected effective sample size. In this chapter we propose to study the central problem of the choice of an optimum strategy systematically.

Analogous to the definition of admissibility of an estimator, admissibility of a strategy has already been defined in Chapter I. Similarly corresponding to the definition of hyper-admissibility of an estimator we have

Definition 6.1.1: In a class, $\mathcal{C}(H)$, of unbiased strategies for the estimation of Y, an $H_o \in \mathcal{C}(H)$ is said to be hyper-admissible in $\mathcal{C}(H)$, if it is admissible in $\mathcal{C}(H)$, when y is restricted to any of the 2^N-1 phs's in R_N.

<u>Definition 6.1.2:</u> A subclass $\mathcal{C}_1(H)$ of $\mathcal{C}(H)$ is said to be complete in $\mathcal{C}(H)$ if and only if for any given $H_1 \in \bar{\mathcal{C}}_1(H)$, the set-theoretic complement of $\mathcal{C}_1(H)$ in $\mathcal{C}(H)$, there exists an $H_1^* \in \mathcal{C}_1(H)$ such that

$$M(H_1^*, y) \leq M(H_1, y), \quad \text{for all } y \in R_{II}. \qquad (6.1.1)$$

The intersection of all complete subclasses of strategies in $\mathcal{C}(H)$, if it exists, is called the minimal complete class of strategies in $\mathcal{C}(H)$.

Evidently if one wants to search for an optimum sampling strategy from among members of $\mathcal{C}(H)$ one need restrict one's attention only to any complete subclass of $\mathcal{C}(H)$. We will replace M by V if $\mathcal{C}(H)$ consists of only unbiased strategies.

6.2 Linear strategies

A strategy $H(S, p, t)$ is said to be linear unbiased if $t \in L_o^*(p)$. The expected effective sample size of a strategy is defined as

$$\mu(H) = \sum_s n(s)p(s) \qquad (6.2.1)$$

Where $n(s)$ is the number of distinct units in s. Let $L(H, \mu)$ denote the class of all linear unbiased strategies H for which $\mu(H) = \mu$, a given number. Under our cost function $L(H, \mu)$ consists of equi-cost strategies. Defining a strategy to be unicluster if and only if the corresponding design is unicluster (Definition 1.1.4) it is clear that

$$L(H, \mu) = L_{NUC}(H, \mu) \cup L_{UC}(H, \mu) \qquad (6.2.2)$$

Where $L_{NUC}(H, \mu)$ is the class of all non-unicluster strategies contained in $L(H, \mu)$ and $L_{UC}(H, \mu)$ consists of all unicluster strategies contained in $L(H, \mu)$. With this background we prove

Theorem 6.2.1: The class $L(H, \mu)$ is complete in ~~$L(H, \mu)$~~ $L(H, \mu) \cup \underline{H}(B)$ where $\underline{H}(B)$ is the set of all biased strategies, if and only if $\mu = N$.

Proof: Let $L(H, \mu)$ be complete in $L(H, \mu) \cup \underline{H}(B)$. Consider $H_1(d_1, t_1) \in \underline{H}(B)$ where

$$t_1 \equiv \partial, \text{ a constant not equal to zero.}$$

From hypothesis, there exists an $H_o \in L(H, \mu)$ such that

$$V(H_o, y) \leq M(H_1, y) = (\partial - Y)^2 \qquad (6.2.3)$$

for all $y \in R_N$. If $H_o = (d_o, t_o)$ where $d_o = (S, p)$ and

$t_o = \sum_{i \in s} b(s,i) y_i$ then

$$V(H_o,y) = \sum_{i=1}^{N} \Big(\sum_{s \supset i} b^2(s,i)p(s)-1 \Big) y_i^2 +$$

$$\sum_{i \neq j}^{N} \sum \Big(\sum_{s \supset i,j} b(s,i)b(s,j)p(s)-1 \Big) y_i y_j . \qquad (6.2.4)$$

At the point $y^{(i)} = (0,0\ldots,\partial,\ldots,0)$ where $y_i = \partial$ from (6.2.3) we have $V(H_o,y)\Big|_{y(i)} = 0$. Hence from the expression for $V(H_o,y)$ in (6.2.4) it follows that

$$\sum_{s \supset i} b^2(s,i)p(s) = 1 . \qquad (6.2.5)$$

But from the condition of unbiasedness of t_o it can be seen that

$$\sum_{s \supset i} b^2(s,i)p(s) \geq \frac{1}{\pi_i(d_o)} \qquad (6.2.6)$$

which together with (6.2.5) gives $\pi_i(d_o) = 1$. Since i is arbitrary we have $\pi_i(d_o) = 1,\ 1 \leq i \leq N$. Hence $\mu = \sum_{i=1}^{N} \pi_i(d_o) = N$, which proves the 'only if' part. If $\mu = N$, clearly $L(H, \mu)$ has a member H_o with

$$V(H_o, y) = 0, \quad \text{for all} \quad y \in R_N.$$

q.e.d.

<u>Remark 6.2.1</u>: From Theorem (6.2.1) we see that $L(H,\mu)$ will not be complete in $L(H,\mu) \cap \underline{H}(B)$ except in the trivial case $\mu = N$ (i.e. a complete census). In practice $\mu < N$ and so one cannot exclude biased strategies from the point of view of mean square criterion alone.

6.3 Horvitz-Thompson strategies

Any strategy $H(d, \bar{e})$ where $\bar{e}$ is the HT-estimator of the population total Y is called a Horvitz-Thompson strategy (HT-strategy for short) for the estimation of Y. Clearly an HT-strategy is unbiased for Y and its variance is given by

$$V(H,y) = \sum_{i=1}^{N} \frac{1-\pi_i(d)}{\pi_i(d)} y_i^2 + \sum_{i \neq j}^{N} \sum \left(\frac{\pi_{ij}(d)}{\pi_i(d)\pi_j(d)} - 1 \right) y_i y_j. \qquad (6.3.1)$$

We now prove

Theorem 6.3.1: Any strategy $H_o(d_o, \bar{e}) \in L(H, \mu)$ is admissible in $L(H, \mu)$.

Proof: Let $H_1(d_1, t_1) \in L(H, \mu)$ where $d_1 = (S, p)$ and $t_1 = \sum_{i \in s} b(s,i) y_i$. We shall show that either there exist points y such that $V(H_o, y) < V(H_1, y)$ or else $V(H_o, y) = V(H_1, y)$, for all $y \in R_N$. At points $y^{(i)} = (0, \ldots, y_i, \ldots, 0)$ $y_i \neq 0$

$$V(H_1, y^{(i)}) = \Big(\sum_{s \supset i} b^2(s,i) p(s) - 1\Big) y_i^2$$

and

$$V(H_o, y^{(i)}) = \Big(\frac{1}{\pi_i(d_o)} - 1\Big) y_i^2 .$$

Now if there exists an i_o such that $\pi_{i_o}(d_o) \neq \pi_{i_o}(d_1)$ one can easily check that there exists a j such that

$$V(H_o, y^{(j)}) < V(H_1, y^{(j)})$$

since

$$\sum_{s \supset i} b^2(s,i) p(s) \geq \frac{1}{\pi_i(d_1)}, \quad 1 \leq i \leq N \tag{6.3.2}$$

and

$$\sum_{1}^{N} \pi_i(d_o) = \sum_{1}^{N} \pi_i(d_1) \quad (= \mu). \tag{6.3.3}$$

In case $\pi_i(d_o) = \pi_i(d_1)$, $1 \leq i \leq N$ and strict inequality in (6.3.2) for some j, then also

$$V(H_o,\ y^{(j)}) < V(H_1,\ y^{(j)}).$$

If $\pi_i(d_o) = \pi_i(d_1)$, $1 \leq i \leq N$ and equality holds in (6.3.2) for $1 \leq i \leq N$, then in the hyper plane

$$y^{(i,j)} = (0,\ldots,0,y_i 0\ldots 0,y_j,0\ldots,0),$$

We have from (6.2.4) and (6.3.1)

$$V(H_1,y^{(i,j)}) - V(H_o,y^{(i,j)}) =$$

$$2\Big(\sum_{s \supset i,j} b(s,i)b(s,j)p(s) - \frac{\pi_{ij}(d_o)}{\pi_i(d_o)\pi_j(d_o)}\Big)y_i y_j. \tag{6.3.4}$$

If there exist units i and j such that the quantity inside the brackets in (6.3.4) is not zero, then it is clear that there exist points y in $y^{(i,j)}$ plane such that $V(H_o,y) < V(H_1,y)$. If

$$\sum_{s \supset i,j} b(s,i)\,b(s,j)\,p(s) = \frac{\pi_{ij}(d_o)}{\pi_i(d_o)\,\pi_j(d_o)} \;, \quad 1 \leq i \neq j \leq N \tag{6.3.5}$$

then clearly

$$V(H_o, y) = V(H_1, y), \quad \text{for all} \quad y \in R_N.$$

q.e.d.

Remark 6.3.1: From Theorem (6.3.1) it follows that there does not exist a best strategy (in the sense of uniformly minimum variance) in $L(H,\mu)$, the class of all equi-cost linear unbiased strategies, for there are at least two - in fact infinitely many (vide Section 8.2) - HT-strategies belonging to $L(H,\mu)$.

Let $L(HT,\mu)$ denote the class of all HT-strategies contained in $L(H,\mu)$. The following theorem shows that in situations of practical interest we cannot exclude strategies other than the HT-strategies using criterion of minimum variance alone.

Theorem 6.3.2: The class $L(HT,\mu)$ is complete in $L(H,\mu)$ if and only if $\mu = 1$ or N.

Before proving the theorem we digress a little to prove a generalisation of Joshi's [29] result. Removing the condition of unbiasedness and considering mean square error as our loss function, we establish the admissibility of a particular strategy in the class $C(H,\mu)$, of all strategies with expected sample size less than or equal to μ. Joshi [29] has proved the following two theorems.

Theorem 6.3.3: The estimate $e^*(s,y)$ given by

$$e^*(s,y) = \frac{N}{n(s)} \sum_{i \in s} y_i \tag{6.3.6}$$

where $n(s)$ is the number of units $i \in s$ is admissible for Y in $A(p)$, for any sampling design p.

Theorem 6.3.4: If μ is an integer, any strategy $H^*(d^*,e^*)$ where d^* is a fixed sample size $(=\mu)$ design is admissible in $C(H,\mu)$.

We now generalise Joshi's [29] result contained in Theorem (6.3.4) to cover the cases when μ is not an integer. Let $\mu = m+f$ where m denotes the greatest integer not exceeding μ and f denotes the fractional part of μ, namely, $\mu - m$. We now prove

Theorem 6.3.5: Any strategy $H^*(d^*,e^*) \in C(H,\mu)$, where d^* is such that

$$\mu(d^*) = \mu .$$

and

$$V_{d^*}(n(s)) = f(1 - f)$$

is admissible in $C(H,\mu)$.

In order to prove the theorem we require the following lemma due to Joshi [28].

Lemma 6.3.1 (Joshi [28]): If

(a) $y_1, y_2, \ldots, y_N$ are independently and identically distributed real random variables,

(b) for every $n = 1,2,\ldots,N$, $\phi_n(y)$ is a real function of $y_1, y_2, \ldots, y_n$,

(c) for every $n = 1,2,\ldots,N$, $\bar{y}_n = \frac{1}{n} \sum_{i=1}^{n} y_i$,

(d) for every common finite discrete frequency function w of $y_1, y_2, \ldots, y_N$,

$$\sum_{n=1}^{N} A_n^2 \, E_w(\phi_n(y) - \theta(w))^2 \leq \sum_{n=1}^{N} A_n^2 \, E_w(\bar{y}_n - \theta(w))^2 ,$$

E_w denoting the expectation, $\theta(w)$ the common mean of $y_1, y_2, \ldots, y_N$ and A_n, $n = 1,2,\ldots, N$, being arbitrary real *constants, then for every* $y = (y_1, y_2, \ldots, y_N) \in R_N$, $\emptyset_n(y) = \bar{y}_n$ for all n, $n = 1,2,\ldots,N$ for which $A_n \neq 0$.

Proof of Theorem 6.3.5: Suppose the theorem is not true. Then there exists an

$$H_1 = (d_1, e_1(s, y)) \in C(H, \mu)$$

such that

$$\mu(d_1) \leq \mu \tag{6.3.7}$$

and

$$M(H_1, y) \leq M(H^*, y) \tag{6.3.8}$$

where strict inequality holds either in (6.3.7) or for at least one $y \in R_N$ in (6.3.8). For the sampling design d_1 let

$p_1(s)$ = probability of sample s

$n_1(s)$ = effective size of sample s

and $\bar{S}_1$: the subset of all those samples s for which $p_1(s) > 0$ and let $p^*(s)$, $n^*(s)$ and $\bar{S}^*$, denote the corresponding terms for the sampling design d^*. Evidently

$$V_{d^*}(n^*(s)) = f(1-f) \iff n^*(s) = \begin{cases} m & \text{with probability } 1-f \\ m+1 & \text{with probability } f \end{cases} \tag{6.3.9}$$

From (6.3.7) and (6.3.8)

$$\sum_{s \in \bar{S}_1} n_1(s)p_1(s) \leq \sum_{s \in \bar{S}^*} n^*(s)p^*(s) = m+f \qquad (6.3.10)$$

and

$$\sum_{s \in \bar{S}_1} p_1(s)[e_1(s,y) - Y]^2 \leq \sum_{s \in \bar{S}^*} p^*(s)[e^*(e,y) - Y]^2 \qquad (6.3.11)$$

where strict inequality holds either in (6.3.10) or for at least one $y \in R_N$ in (6.3.11). Taking expectations on both sides of (6.3.11) with respect to a prior distribution on R_N, under which $y_i (i = 1,2,\ldots, N)$ are distributed independently and identically with common mean θ and variance σ^2, we get

$$\sum_{s \in \bar{S}_1} p_1(s)E[e_1(s,y) - Y]^2 \leq \sum_{s \in \bar{S}^*} p^*(s)E[e^*(s,y)-Y]^2. \qquad (6.3.12)$$

Defining $g_1(s,y) = [N - n_1(s)]^{-1}[e_1(s,y) - \sum_{i \notin s} y_i]$ (6.3.13)

and making use of the fact that y_i's are independently distributed we get

$$E[e_1(s,y)-Y]^2 = (N - n_1(s))^2 E[g_1(s,y) - \theta)]^2 + (N-n_1(s))\sigma^2 \qquad (6.3.14)$$

Similarly

$$E[e^*(s,y)-Y]^2=(N-n^*(s))^2E(\bar{y}_s-\theta)^2+(N-n^*(s))\sigma^2 \qquad (6.3.15)$$

where $\bar{y}_s = [n^*(s)]^{-1} \sum_{i\in s} y_i$.

Now substituting (6.3.14) and (6.3.15) in (6.3.12) and using the relation

$$\sum_{s\in S_1} p_1(s) = \sum_{s\in \bar{S}^*} p^*(s) = 1$$

and cancelling out the common term, (6.3.12) becomes

$$\sum_{s\in \bar{S}_1} p_1(s)(N-n_1(s))^2 E[g_1(s,y)-\theta]^2-\sigma^2 \sum_{s\in \bar{S}_1} p_1(s)n_1(s)$$

$$\leq \sum_{s\in \bar{S}^*} p^*(s)(N-n^*(s))^2E(\bar{y}_s-\theta)^2-\sigma^2 \sum_{s\in \bar{S}^*} p^*(s)n^*(s). \qquad (6.3.16)$$

Putting $g_1(s,y) = \bar{y}_s+h_1(s,y)$ and noting that $E(\bar{y}_s-\theta)^2=\frac{\sigma^2}{n(s)}$, we have from (6.3.16), after cancelling out the common term

$$\sum_{s\in \bar{S}_1} p_1(s)(N-n_1(s))^2E(h_1^2(s,y))+2\sum_{s\in \bar{S}_1} p_1(s)(N-n_1(s))^2$$

$$\times E[h_1(s,y)(\bar{y}_s-\theta)]+\sigma^2N^2 \sum_{s\in \bar{S}_1} \frac{p_1(s)}{n_1(s)} \leq \sigma^2N^2\left(\frac{1-f}{m}+\frac{f}{m+1}\right).$$

$$(6.3.17)$$

Now we will show that

$$\sum_{s\in S_1} \frac{p_1(s)}{n_1(s)} \geq \frac{1-f}{m} + \frac{f}{m+1} \quad . \tag{6.3.18}$$

Let

$$\sum_{s\in \bar{S}_1} n_1(s)p_1(s) = \mu' = m' + f' \tag{6.3.19}$$

where m' is the greatest integer not exceeding μ' and f' is the fractional part of μ', namely, $\mu'-m'$. Clearly $\sum_s \frac{p_1(s)}{n_1(s)}$ can be written as

$$\sum_s \frac{p_1(s)}{n_1(s)} = \sum_{i=1}^{N} \frac{p_i}{i} \tag{6.3.20}$$

where $p_i = \Sigma^{(i)} p_1(s)$ where the summation is taken over all those samples s which contain exactly i units. Using inequality (2.2.7) we get

$$\sum_{i=1}^{N} \frac{p_i}{i} \geq \frac{1-f'}{m'} + \frac{f'}{m'+1} \quad . \tag{6.3.21}$$

Hence to establish inequality (6.3.18) it suffices to show

$$\frac{1-f'}{m'} + \frac{f'}{m'+1} \geq \frac{1-f}{m} + \frac{f}{m+1} \quad .$$

i.e.

$$\frac{m'+1-f'}{m'(m'+1)} \geq \frac{m+1-f}{m(m+1)} . \tag{6.3.22}$$

Since $m'+f' \leq m+f$ due to (6.3.10) we get $m' \leq m$. If $m'=m$ (6.3.22) is obvious since $f' \leq f$. When $1 \leq m' \leq m-1$, it can be checked that

$$\frac{m'+1-f'}{m'(m'+1)} - \frac{m+1-f}{m(m+1)} = \frac{m(m+1)(1-f')+m'(m+1)(m-m'-1)+m'(m'+1)f}{mm'(m+1)(m'+1)}$$

$$\geq 0$$

The sign of equality holds if and only if $m'=m$ and $f'=f$.

Combining (6.3.17) and (6.3.18), we have

$$\sum_{s\in S_1} p_1(s)(N-n_1(s))^2 E[h_1^2(s,y)]$$

$$+ 2 \sum_{s\in S_1} p_1(s)(N-n_1(s))^2 E[h_1(s,y)(\bar{y}_s - \theta)] \leq 0. \tag{6.3.23}$$

The inequality (6.3.23) is equivalent to the inequality contained in clause (d), in Lemma (6.3.1) and hence using the lemma, it follows that for all $s\in\bar{S}_1$

$$h_1(s, y) = 0 \tag{6.3.24}$$

so that $g_1(s, y) = \bar{y}_s$ and by (6.3.6) and (6.3.13)

$$e_1(s, y) = e^*(s, y). \tag{6.3.25}$$

Because of (6.3.24) the first two terms in the left hand side of (6.3.17) vanish and hence due to (6.3.18) the sign of equality must hold in both (6.3.17) and (6.3.18) so that d_1 is also a sampling design such that

$$\mu(d_1) = \mu \quad \text{and} \quad V_{d_1}(n(s)) = f(1-f).$$

Hence the strict inequality cannot hold in (6.3.7). We shall next show that the strict inequality in (6.3.8) cannot hold either. Let the inclusion probabilities for the units for d_1 and d^* be given by

$$\pi_i = \sum_{s \supset i} p_1(s), \qquad \pi_i^* = \sum_{s \supset i} p^*(s), \quad 1 \leq i \leq N.$$

Clearly π_i and π_i^* can be written as

$$\pi_i = \pi_{1i} + \pi_{2i} \quad \text{and} \quad \pi_i^* = \pi_{1i}^* + \pi_{2i}^*$$

where

$$\pi_{1i} = \sum_{s_m \supset i} p_1(s), \quad \pi_{1i}^* = \sum_{s_m \supset i} p^*(s)$$

$$\pi_{2i} = \sum_{s_{m+1} \supset i} p_1(s), \quad \pi_{2i}^* = \sum_{s_{m+1} \supset i} p^*(s)$$

where s_m denotes a typical sample of size m and s_{m+1} that of size m+1. e* being a linear estimator the mean square errors of H_1 and H^* are quadratic forms in $y_1,\ldots,y_N$ given by

$$M(H_1,y) = \sum_{i,j=1}^{N}\sum a_{ij}y_iy_j$$

$$M(H^*,y) = \sum_{i,j=1}^{N}\sum a^*_{ij}y_iy_j \,.$$

Since $M(H_1,y) - M(H^*,y) \leq 0$, for all $y \in R_N$ (6.3.26)

it follows that

$$a_{ii} - a^*_{ii} \leq 0, \quad 1 \leq i \leq N. \qquad (6.3.27)$$

One can easily check that

$$a_{ii} - a^*_{ii} = \frac{N^2}{m^2}(\pi_{1i}-\pi^*_{1i}) + \frac{N^2}{(m+1)^2}(\pi_{2i}-\pi^*_{2i})$$

$$- \frac{2N}{m}(\pi_{1i}-\pi^*_{1i}) - \frac{2N}{m}(\pi_{2i}-\pi^*_{2i}) \,. \qquad (6.3.28)$$

Since

$$\sum_{i=1}^{N}\pi_{1i} = \sum_{i=1}^{N}\pi^*_{1i} = m(1-f)$$

$$\sum_{i=1}^{N}\pi_{2i} = \sum_{i=1}^{N}\pi^*_{2i} = (m+1)f$$

It follows from (6.3.28) that

$$\sum_{i=1}^{N} (a_{ii} - a_{ii}^*) = 0$$

which together with (6.3.27) gives

$$a_{ii} = a_{ii}^*, \quad 1 \leq i \leq N. \tag{6.3.29}$$

From (6.3.26) and (6.3.29) we have

$$a_{ij} = a_{ij}^*, \quad 1 \leq i \neq j \leq N \text{ also.}$$

Hence $M(H_1, y) \equiv M(H^*, y)$, for all $y \in R_N$ and thus the strict inequality in (6.3.8) cannot hold.

q.e.d.

Remark 6.3.2: We get Joshi's Theorem (6.3.4) by putting $f = 0$ in Theorem (6.3.5).

6.3.1 Necessity of the condition $V_{d*}(n(s)) = f(1-f)$

If the strategy $H^*(d^*, e^*)$ of Theorem (6.3.5) is such that

$$\mu(d^*) = \mu \quad \text{and} \quad V_{d*}(n(s)) > f(1-f)$$

then the strategy $H^*(d^*, e^*)$ may become inadmissible. We illustrate this with an example.

Example: Consider srswr with size $n \geq 3$ and let the set s_m of distinct units in the sample s form a sample for d^* with the same probability, namely, $\frac{1}{N^n}$. From Section (2.5) we know that

$$\mu(d^*) = \mu = N[1 - (\frac{N-1}{N})^n] = m+f, \text{ say.}$$

Since $n(s)$ takes all values from 1 to n with positive probability, it is clear that

$$V_{d^*}(n(s)) > f(1-f) .$$

Now we show that the strategy $H^*(d^*,e^*)$, where d^* is as defined above, and e^* is given by (6.3.6) is inadmissible in $C(H,\mu)$.

Proof: Consider the strategy $H_1(d_1,e^*)$ where d_1 is the design obtained by srswor with size m or $m+1$ with probabilities $1-f$ and f respectively. This is the same strategy described in Section (2.2) and it follows from Theorem (2.2.1) proved therein that $H_1(d_1,e^*)$ is uniformly better than $H^*(d^*,e^*)$. It appears that condition, $V_{d^*}(n(s)) = f(1-f)$ is also necessary for the validity of

Theorem (6.3.5).

For the sake of future use we denote the strategy $H_1(d_1, e^*)$, above, by

$$H_1(d_1,\ N\bar{y},\ \mu,\ N) \qquad (6.3.30)$$

where $d_1 = (S,\ p_1)$, μ denotes the expected sample size and N, the size of the population. We are now in a position to prove Theorem (6.3.2).

Proof of Theorem 6.3.2: If $\mu = 1$ or N, for any strategy $H(d,\ t) \in L(H, \mu)$, the design d will be unicluster and hence the HT-estimator $\bar{e}(s,y)$ is the UMV estimator in $L_o^*(d)$. In either case it is seen that $L(HT, \mu)$ is complete in $L(H, \mu)$. This proves the 'if' part of the theorem.

In case $1 < \mu < N$ and μ is not an integer it follows as a consequence of Theorem (6.3.5) that the strategy $H_1(d_1,\ N\bar{y},\ \mu,\ N)$ in (6.3.30) is admissible in $L(H, \mu)$. Since the above strategy does not belong to $L(HT,\ \mu)$, it follows that $L(HT,\ \mu)$ is not complete in $L(H,\ \mu)$.

The only case left out is: μ an integer, $2 \leq \mu \leq N-1$ and $N \geq 3$. Choose and fix ∂ such that $0 < \partial < 1$ so that

$$1 < \mu_o = \mu - \partial < N - 1. \qquad (6.3.31)$$

From the strategy $H_1(d_1, (N-1)\bar{y}, \mu_o, N-1)$ corresponding to the population consisting of the first $(N-1)$ units, we construct a strategy $H'(d', t')$ where $d' = (S, p')$ corresponding to the population of N units as follows:

Any sample s for d_1, goes into two samples (s, N) and s for d' with probabilities $\partial\, p_1(s)$ and $(1 - \partial)p_1(s)$ respectively. Also define

$$t'((s,N),y) = (N-1)\,\bar{y}_s + \frac{y_N}{\partial}$$

and

$$t'(s, y) = (N-1)\,\bar{y}_s.$$

Since $\pi_N(d') = \partial$ and $(N-1)\,\bar{y}_s$ is unbiased for $\sum_{i=1}^{N-1} y_i$ for d_1, it is clear that t' is an unbiased estimator for $Y\ (= \sum_1^N y_i)$ under d'. Also since $t' \neq \bar{e}$

$$H'(d', t') \in L(H, \mu) - L(HT, \mu).$$

Next we show that given any $H(d, \bar{e}) \in L(HT, \mu)$, there exists a point y_o (which may depend on H and H') $\in R_N$ such that

$$V(H', y_o) < V(H, y_o),$$

which will show that $L(HT, \mu)$ is not complete in $L(H, \mu)$.

If $\pi_N(d) < \partial$, one can easily check that

$$V(H', y^{(N)}) < V(H, y^{(N)})$$

where $y^{(N)} = (0, \ldots, 0, y_N)$ and y_N is any non-zero real number. If $\pi_N(d) \geq \partial$ and d gives positive probability to the sample consisting of unit N alone, then also it is easy to check that

$$0 = V(H', y_o) < V(H, y_o)$$

where $y_o = (k, k, \ldots, k, 0)$ and k is any non-zero real number. If $\pi_N(d) \geq \partial$ and d gives zero probability to the sample consisting of unit N alone, then construct the strategy $H_2(d_2, \bar{e})$ corresponding to the population of the first $(N-1)$ units where d_2 is obtained from d by removing unit N from all those samples for d_1, containing it, the probability structure remaining unchanged. Since $\pi_N(d) \geq \partial$ and $\sum_1^N \pi_i(d) = \mu$,

$$\sum_1^{N-1} \pi_i(d_2) = \sum_1^{N-1} \pi_i(d) \leq \mu - \partial = \mu_o .$$

Hence by Theorem (6.3.5), there exists a point $y_o' = (y_{01}', \ldots, y_{0N-1}') \in R_{N-1}$ such that

$$V(H_1, y_0'), < V(H_2, y_0'). \tag{6.3.32}$$

Since

$$V(H_1, y_0') = V(H', y_0) \tag{6.3.33}$$

and

$$V(H_2, y_0') = V(H, y_0) \tag{6.3.34}$$

where $y_0 = (y'_{01}, \ldots, y'_{0N-1}, 0)$, we have, on comparing (6.3.32), (6.3.33) and (6.3.34)

$$V(H', y_0) < V(H, y_0).$$

q.e.d.

Before stating a more general theorem, let us define

$$L(HT) = \bigcup_{1 \leq \nu \leq \mu} L(HT, \nu) \tag{6.3.35}$$

and

$$L(H) = \bigcup_{1 \leq \nu \leq \mu} L(H, \nu) \tag{6.3.36}$$

where μ is some given number such that $1 \leq \mu \leq N$. Under a linear cost function, it is clear that, L(H) consists of all linear unbiased strategies whose expected costs are less than or equal to a fixed budget C_0 and L(HT) is the set of all HT-strategies contained in L(H). Using Theorem (6.3.2) we get

<u>Theorem 6.3.6:</u> L(HT) is complete in L(H) if and only if $\mu = 1$ or N.

6.4 Strong admissibility

After having proved the nonexistence of a best strategy (in the sense of uniformly minimum variance) in $L(H, \mu)$ and that the complete class of admissible strategies is wider than $L(HT, \mu)$ in most of the situations, our next step is to impose further criteria which will give a narrow enough class of strategies. One can easily check there does not exist a hyper-admissible strategy in $L(H, \mu)$. In the following we weaken this criterion and characterise the class of all strategies in $L(H, \mu)$ that satisfy the new criterion.

Definition 6.4.1: In a class $C(H)$ of unbiased strategies for Y, a strategy $H_1 \in C(H)$ is said to be 'strongly admissible' in $C(H)$ if it is admissible in $E_1, E_2, \ldots, E_N$ separately, where $E_r = \bigcup_{i=1}^{\binom{N}{r}} R_i^r$ and R_i^r is the i-th phs of dimension r.

The definition of a strongly admissible estimator is straightforward. For the case of a single design, it can be noted that the criteria of strong admissibility and hyper-admissibility are effectively equivalent in arriving at an optimum estimator. While there exists no hyper-admissible strategy, there exist strongly admissible strategies in $L(H, \mu)$ and we characterise the set of all strongly admissible strategies in $L(H, \mu)$ in the following

Theorem 6.4.1: $L(HT, \mu)$ is precisely the set of all strongly admissible strategies in $L(H, \mu)$. In other words $L(HT, \mu)$ is complete in $L(H, \mu)$ with respect to strong admissibility.

Proof: For any design d, the unique strongly admissible estimator in $L_o^*(d)$ is given by $\bar{e}(s, y)$, which shows that the set of all strongly admissible strategies in $L(H, \mu)$ is contained in $L(HT, \mu)$. Further, it can be easily noted that, from the proof of Theorem (6.3.1), in fact follows the strong admissibility of any strategy $H \in L(HT, \mu)$ in $L(H, \mu)$.

q.e.d.

The criterion of strong admissibility has some practical implications. For example, in case of estimation of a domain total (or mean) where the domain size is known (say r) but the domain frame is not available (a unit can be classified into that domain only after surveying it) it is easily seen that the parameter ($y = (y_1, \ldots, y_N)$ where exactly $N-r$ co-ordinates have fixed zero values) space is given by E_r. So if we start with a strongly admissible estimator, such domain totals (or means) can be admissibly estimated with the same estimator.

6.5 Extensions

For a given design d let $G_o(d)$ be the class of all unbiased estimators of the population total Y, satisfying

$$t(s,y) \in G_o(d) \Rightarrow t(s,y) = 0 \text{ if } y_i = 0, \text{ for all } i \in s. \tag{6.5.1}$$

Let

$$G(H,\mu) = \{ H(d,t) : \mu(d) = \mu, \; t \in G_o(d) \}. \tag{6.5.2}$$

Clearly $L(H, \mu) \subset G(H, \mu)$. In this section we extend some results of the earlier sections to the class $G(H, \mu)$. We prove

Theorem 6.5.1: Any strategy $H_o(d_o, \bar{e}) \in L(HT, \mu)$ is admissible in $G(H, \mu)$.

Proof: Let $H = (d, t) \in G(H, \mu)$ where $d = (S, p)$ be such that

$$V(H, y) \leq V(H_o, y), \quad \text{for all } y \in R_N. \tag{6.5.3}$$

We shall show that strict inequality cannot hold in (6.5.3) at any point $y \in R_N$. Consider the hyperplane

$$y^{(i)} = (0, \ldots, 0, y_i, 0, \ldots, 0).$$

Condition of unbiasedness of t gives

$$\sum_{s \in S_i} t(s, y^{(i)})\, p(s) = y_i, \quad \text{for all } y^{(i)} \in R_N. \qquad (6.5.4)$$

Using (6.5.4) one can easily check that in the $y^{(i)}$-plane

$$V(H, y^{(i)}) - V(H_o, y^{(i)}) = \sum_{s \in S_i} \left(t(s, y^{(i)}) - \frac{y_i}{\pi_i(d)}\right)^2 p(s) + $$
$$+ y_i^2 \left(\frac{1}{\pi_i(d)} - \frac{1}{\pi_i(d_o)}\right) \qquad (6.5.5)$$

where $S_i \subset S$, is the set of samples containing the i-th unit. From (6.5.3) and (6.5.5) we have

$$V(H, y^{(i)}) - V(H_o, y^{(i)}) \leq 0, \quad \text{for all } y^{(i)} \in R_N. \qquad (6.5.6)$$

Since $\Sigma\, \pi_i(d) = \Sigma\, \pi_i(d_o) (= \mu)$ and (6.5.5) is true for $1 \leq i \leq N$, we have from (6.5.5) and (6.5.6)

$$\pi_i(d) = \pi_i(d_o), \qquad 1 \leq i \leq N \qquad (6.5.7)$$

and

$$t(s, y^{(i)}) = \frac{y_i}{\pi_i(d_1)}, \quad 1 \leq i \leq N. \qquad (6.5.8)$$

From (6.5.6), (6.5.7) and (6.5.8) we see that $t(s,y)$ reduces to the HT-estimator $\bar{e}(s,y)$ in hyper-planes of dimension one and

$$V(H, y^{(i)}) = V(H_o, y^{(i)}).$$

Now consider a typical hyper-plane of dimension two, say,

$$y^{(i,j)} = (0,\ldots,0,y_i,0,\ldots,0,y_j,0,\ldots,0).$$

Since t is unbiased, we have

$$\sum_{s \in S_i \cap S_j} t(s,y^{(i,j)})p(s) = \pi_{ij}(d)\left(\frac{y_i}{\pi_i(d)} + \frac{y_j}{\pi_j(d)}\right). \qquad (6.5.9)$$

Using (6.5.9), after a little algebra, one can see

$$V(H,y^{(i,j)}) - V(H_o,y^{(i,j)}) = \sum_{s \in S_i \cap S_j} \left[t(s,y^{(i,j)}) - \frac{y_i}{\pi_i(d)} - \frac{y_j}{\pi_j(d)}\right]^2$$

$$+ 2(\pi_{ij}(d) - \pi_{ij}(d_o)) \frac{y_i}{\pi_i(d)} \frac{y_j}{\pi_j(d)}.$$

(6.5.10)

From (6.5.3) and (6.5.10), we get

$$V(H,y^{(i,j)}) - V(H_o,y^{(i,j)}) \leq 0, \quad \text{for all } y^{(i,j)} \in R_N.$$

(6.5.11)

Comparing (6.5.10) and (6.5.11) we see that (6.5.11) can be true only if

$$\pi_{ij}(d) = \pi_{ij}(d_o) \tag{6.5.12}$$

and

$$t(s, y^{(i,j)}) = \frac{y_i}{\pi_i(d)} + \frac{y_j}{\pi_j(d)} . \tag{6.5.13}$$

Since i and j are arbitrary, (6.5.12) and (6.5.13) are true for $1 \leq i \neq j \leq N$. From (6.5.11), (6.5.12) and (6.5.13) we see that in hyper-planes of dimension two $t(s,y)$ reduces to the HT-estimator and

$$V(H, y^{(i,j)}) = V(H_o, y^{(i,j)}), \quad 1 \leq i \neq j \leq N.$$

Similarly, it can be shown that $t(s,y)$ reduces to the HT-estimator $\bar{e}(s,y)$ in any hyper-plane of higher order and $V(\dot{H},y) = V(H_o,y)$ for every y belonging to that hyper-plane. The proof our theorem is complete since the N-dimensional hyper-plane is the whole of R_N.

q.e.d.

<u>Remark 6.5.1:</u> For any design d let

$$A(H,\mu) = \{ H(dt) : \mu(d) = \mu, \ t \in A^*(d) \} \tag{6.5.14}$$

where $A^*(d)$ denotes the class of all unbiased estimators of the population total. Clearly $G(H,\mu) \subset A(H,\mu)$. Since for any $H \in A(H,\mu) - G(H,\mu)$, $V(H,0) > 0$, Theorem (6.5.1), in fact, establishes the admissibility of any $H_0 \in L(HT,\mu)$ in $A(H,\mu)$.

Analogous to Theorem (6.4.1), we have

Theorem 6.5.2: $L(HT,\mu)$ is minimal complete in $G(H,\mu)$ with respect to strong admissibility.

Proof: It is not difficult to establish that for any given design, the unique strongly admissible estimator in $G_0(d)$ is given by the HT-estimator $\bar{e}(s,y)$. [We only outline the proof here. As in the case of hyperadmissibility in Chapter IV we can first show that the only possible strongly admissible estimator in $G_0(d)$ is $\bar{e}(s,y)$. That it is strongly admissible follows from its hyperadmissibility]. Hence it follows that the set of all strongly admissible strategies is contained in $L(HT,\mu)$. Moreover, it is easy to see that from the proof of Theorem (6.5.1), in fact, follows the strong admissibility of any strategy $H_0 \in L(HT,\mu)$ in $G(H,\mu)$.

q.e.d.

Remark 6.5.2: From Theorem (6.5.1) and using the fact that for any non-unicluster design d, the unique strongly admissible estimator in $A^*(d)$, is given by the HT-estimator $\bar{e}(s,y)$, it follows that $L_{NUC}(HT,\mu)$ is minimal complete in $A_{NUC}(H,\mu)$ with respect to strong admissibility where $L_{NUC}(HT,\mu)$ and $A_{NUC}(H,\mu)$ denote the classes of non-uniclustor strategies contained in $L(HT,\mu)$ and $A(H,\mu)$ respectively.

Remark 6.5.3: For a unicluster design, d, Hanurav [22] has shown that the complete class of hyperadmissible estimators in $A^*(d)$ is given by

$$L'(\bar{e},d):\ e(s,y) = \bar{e}(s,y) + k(s) \qquad (6.5.15)$$

where $k(s)$'s are constants independent of y and $k(s_1) = k(s_2)$ if s_1 and s_2 are equivalent. Since, for the case of a single design, hyper-admissibility and strong admissibility are equivalent, $L'(\bar{e},d)$ is the set of all strongly admissible estimators in $A^*(d)$, for any unicluster design d. Hence from Theorem (6.4.1) and Remark (6.4.1) it follows that, the minimal complete class of strongly admissible strategies in $A(H,\mu)$ is given by $L_{NUC}(HT,\mu) \cup L'_{UC}(HT,\mu)$ where

$$L'_{UC}(HT,\mu) = \{H(d,t):\ \mu(d) = \mu,\ d \text{ unicluster},\ t \in L'(\bar{e},d)\}.$$

We have proved in Section (6.3) that $L(HT,\mu)$ is complete in $L(H,\mu)$ if and only if $\mu = 1$ or N. We now prove

Theorem 6.5.5: $L(HT,\mu)$ is complete in $A(H,\mu)$ if and only if $\mu = N$.

Proof: If $\mu = N$, it is obvious that $L(HT,\mu)$ is complete in $A(H,\mu)$. If $1 < \mu < N$, from Theorem (6.3.2) it follows that $L(HT,\mu)$ is not complete in $A(H,\mu)$. To complete the proof of our theorem it is enough we show $L(HT,1)$ is not complete in $A(H,1)$. Let d_o be any design such that

$$\pi_i(d_o) > 0, \quad i \leq i \leq N \quad \text{and} \quad \Sigma\, \pi_i(d_o) = 1$$

and let

$$t_o(s,y) = \bar{e}(s,y) - \bar{e}(s,y_o) + Y_o$$

where $y_o = (y_{1o}, y_{2o}, \ldots, y_{No})$ is an arbitrary fixed point in R_N, other than the origin and $Y_o = \Sigma\, y_{io}$. Clearly the strategy

$$H_o(d_o,t_o) \in A(H,1) - L(HT,1)$$

and $$V(H_o, y_o) = 0 \ .$$ (6.5.16)

Now

$$H(d, \bar{c}) \in L(HT, 1) \Rightarrow V(H, y) = \sum_{i=1}^{N} \left(\frac{1}{\pi_i(d)} - 1 \right) y_i^2 \ .$$

(6.5.17)

From (6.5.16), (6.5.17) and noting that

$0 < \pi_i(d) < 1, \quad 1 \leq i \leq N,$

$$0 = V(H_o, y_o) < V(H, y_o) = \sum_{i=1}^{N} \left(\frac{1}{\pi_i(d)} - 1 \right) y_{io}^2 \ .$$

q.e.d.

CHAPTER VII

CHOICE OF AN OPTIMUM STRATEGY-CONTD.

7.0 Summary

We derive some optimum properties of the sample mean in srswor under a super-population model generated by permutations of values attached to the distinct units. Recently Royal [55] has suggested the use of a purposive sampling plan together with a ratio estimator for the estimation of the population total Y. In this chapter we arrive at the same strategy suggested by Royal [55], using an alternative approach based on finite parametrization of the population vector $y = (y_1, \ldots, y_N)$, similar to those used by Kempthorne [31] and Rao [47]. The optimum strategy corresponding to a more general model than that considered by Royal [55] is also derived. It is interesting to note that the new strategy is also purposive and uses a regression-type estimator.

7.1 Introduction

After the non-existence of a UMV estimator in the *class of all linear unbiased estimators* was established by *Godambe [14]*, for most of the designs, *different optimality* criteria have been put forward for the choice of an optimum estimator and we have given a broad review of the related literature in Chapter I. It has been pointed out there that the criterion of admissibility is used to weed out bad estimators. Kempthorne [31], while passing some critical remarks on statistical inference in finite sampling suggested a more reasonable criterion of optimality, for the case of srswor. Under this criterion, we suppose that there exists an underlying unknown set of N numbers, say Z_1, Z_2,...., Z_N and that these are associated with the labels of the units in an unknown way and we try to minimise the average variance of an unbiased estimator, for permutations of the values attached to the units.

Kempthorne [31], then gave a justification for the sample mean in srswor by proving

Theorem 7.1.1: The sample mean in srswor has average minimum variance, for permutations of values attached to the units, in the general class of linear unbiased and translation invariant estimators of the population mean $\bar{Y}$.

Although the above approach can be viewed as a super-population approach, it differs from the conventional method in that the super-population considered here is not infinite. Any way, the above formulation seems quite reasonable and appealing since simple random sampling is to be resorted to, theoretically at least, only when there is complete absence of any information about the population.

Recently Royal [55] suggested the use of a purposive sampling plan together with a ratio estimator for the estimation of the population total Y. The above suggestion was made under the following model: The numbers $y_1, y_2, \ldots, y_N$ are realised values of independent random variables $Y_1, Y_2, \ldots, Y_N$ where Y_i has mean βx_i and variance $\sigma^2 v(x_i)$. The function v is assumed to be known, with $v(x) > 0$ for $x > 0$; the constants β and σ^2 are unknown. Denoting the joint probability law of $Y_1, Y_2, \ldots, Y_N$ by ξ, Royal [55] defined ξ-unbiasedness of an estimator as follows:

<u>Definition 7.1.1:</u> An estimator t is said to be ξ-unbiased for Y if

$$E_{\xi}(t_s - Y) = 0, \quad \text{for every } s \in S$$

where E_{ξ} denotes the expected value with respect to the probability law ξ. He proved that for any sampling design, a best linear ξ-unbiased estimator of Y is given by

$$t^* = \sum_s y_i + \hat{\beta}^* \sum_{\bar{s}} x_i \tag{7.1.1}$$

where

$$\hat{\beta}^* = \sum_s \frac{x_i\, y_i}{v(x_i)} \Big/ \frac{x_i^2}{v(x_i)} \tag{7.1.2}$$

and $\sum_s$ and $\sum_{\bar{s}}$ denote summation of units in the sample and not in the sample respectively. Assuming $v(x)$ is non-decreasing, and $v(x)/x^2$ is non-increasing he has shown that the optimum fixed sample-size $(=n)$ design corresponding to t^* is p^* which entails the selection of s which maximises $\sum_s x_i$, with certainty. **Under** the particular case $v(x) = x^2$, t^* reduces to

$$t_1^* = \sum_s y_i + \left(\sum_{\bar{s}} x_i\right)\frac{1}{n} \sum_s y_i/x_i . \tag{7.1.3}$$

7.2 An Optimal property of sample mean in srswor

Rao [47] has shown that the condition of translation invariance is redundant for the validity of Theorem (7.1.1). Now taking average mean square error, for permutations of values attached to the units, as our criterion we show that the condition of unbiasedness is redundant in Theorem (7.1.1) in the following

Theorem 7.2.1: The usual estimator (namely $N\bar{y}$) in srswor has average minimum mean square error, for permutations of values attached to the units, in the general class of linear translation invariant estimators of Y.

Proof: Let

$$t = \sum_{i \in s} b(s, i) y_i . \qquad (7.2.1)$$

The condition of translation invariance of t is clearly given by

$$\sum_{i \in s} b(s, i) = N \quad \text{for all} \quad s \in S \qquad (7.2.2)$$

[Note that here S consists of all the $\binom{N}{n}$ samples without replacement.] Also

$$mse(t) = E(t-Y)^2 = E[\sum_{i\in s} b(s,i)-1)\, y_i - \sum_{i\in \bar{s}} y_i]^2$$

so that

$$\mathcal{E}\, mse(t) = \mathcal{E}\, E[\sum_{i\in s} (b(s,i)-1)y_i - \sum_{i\in \bar{s}} y_i]^2$$

$$= E.\,\mathcal{E}\,[\sum_{i\in s} (b(s,i)-1)y_i - \sum_{i\in \bar{s}} y_i]^2 \qquad (7.2.3)$$

where $\mathcal{E}$ denotes averaging over all permutations of $(y_1, y_2, \ldots, y_N)$. Expanding the square term in the r.h.s of (7.2.3) and averaging over the permutations

$$\mathcal{E}\, mse(t) = E[\sum_{i\in s} (b(s,i)-1)^2 S_1 + \sum_{i\neq j\in s} (b(s,i)-1)(b(s,j)-1) S_{11}$$

$$- 2 \sum_{\substack{i\in s\\ j\in \bar{s}}} (b(s,i)-1) S_{11}] + K_1 \qquad (7.2.4)$$

where $K_1 = E.\,\mathcal{E}(\sum_{i\in \bar{s}} y_i)^2$ is a term independent of $b(s,i)$'s and

$$S_1 = \frac{1}{N} \sum_{i=1}^{N} y_i^2 \quad \text{and} \quad S_{11} = \frac{1}{N(N-1)} \sum_{i\neq j}^{N}\sum y_i y_j . \qquad (7.2.5)$$

Using the condition for translation invariance given in (7.2.2), we see from (7.2.4) that the only terms in $\mathcal{E}$ mse(t) which depend on $b(s,i)$'s are

$$E[S_1 \sum_{i\in s} b^2(s,i) + S_{11} \sum\sum_{i\neq j\in s} b(s,i)b(s,j)]$$

$$= \frac{(S_1 - S_{11})}{\binom{N}{n}} \sum_s \sum_{i\in s} b^2(s,i) + K_2 \qquad (7.2.6)$$

where K_2 is also independent of $b(s,i)$'s. In order to minimise $\mathcal{E}$ mse(t) in the class of all linear translation invariant estimators, it is clear from (7.2.6) that we have to minimise $\phi = \sum_s \sum_{i\in s} b^2(s,i)$ subject to condition (7.2.2). It is easily seen that ϕ is minimised when $b(s,i) = \frac{N}{n}$.

q.e.d.

In the following we drop the condition of translation invariance also. We prove

<u>Theorem 7.2.2:</u> For srswor, the estimator which minimises the average mean square error, for permutations of values attached to the units, in the class of all linear estimators is given by

$$t_C = \frac{N}{n(1+\partial)} \sum_{i\in s} y_i$$

where $\partial = \frac{N-n}{N-1} \frac{C^2}{n}$ and C is population coefficient of variation, which is assumed to be known.

<u>Proof</u>: Let t be as given in (7.2.1). From (7.2.4) we see that in order to minimise $\mathcal{E}$ mse(t), we have to minimise

$$\phi = (S_1 - S_{11}) \sum_s \sum_{i \in s} (b(s,i)-1)^2 + S_{11} \sum_s [\sum_{i \in s} (b(s,i)-1)]^2$$

$$- 2(N-n) S_{11} \sum_s \sum_{i \in s} (b(s,i)-1). \qquad (7.2.7)$$

Differentiating ϕ w.r.t. $b(s,i)$ and equating to zero we have

$$\frac{\partial \phi}{\partial b(s,i)} = 2(S_1 - S_{11})(b(s,i)-1) + 2S_{11} \sum_{i \in s} (b(s,i)-1)$$

$$- 2(N-n)S_{11} = 0 . \qquad (7.2.8)$$

Summing (7.2.8) over $i \in s$

$$(S_1 - S_{11}) \sum_{i \in s} (b(s,i)-1) + nS_{11} \sum_{i \in s} (b(s,i)-1) - n(N-n)S_{11} = 0$$

which gives

$$\sum_{i \in s} (b(s,i)-1) = \frac{(N-n)nS_{11}}{S_1 + (n-1)S_{11}} . \qquad (7.2.9)$$

Substituting (7.2.9) in (7.2.8) and simplifying we get

$$b(s,i) = \frac{N}{n} \cdot \frac{1}{1+\partial}$$

where $\partial = \frac{N-n}{N-1} \cdot \frac{C^2}{n}$ and C is the population coefficient of variation.

q.e.d.

Remark 7.2.1: When C is known Searls [57] suggested the estimator

$$t_C' = \frac{N}{n(1+\partial')} \sum_{i \in s} y_i$$

for srswr, where $\partial' = \frac{C^2}{n}$ and **showed** that it has uniformly smaller mean-square error than the conventional unbiased estimator $N\bar{y}$. It can be easily verified that Searls' method of improving the usual estimator - when C is known - leads to the estimator t_C in Theorem (7.2.2), for srswor. Theorem (7.2.2) gives another justification for the use of t_C in srswor. In practice, C will not be known exactly and a guessed value $\hat{C}$ of C may be available (from previous censuses or surveys) which can be used instead of C. From Singh and Roy [61] we see that $t_{\hat{C}}$ may lead to substantial gain in efficiency over the conventional unbiased estimator.

Remark 7.2.2: Clearly t_C is biased for Y and the bias in t_C is given by

$$B(t_C) = \frac{Y}{1+\partial} - Y = -\frac{\partial}{1+\partial} Y$$

so that

$$|B(t_C)| = \frac{1}{1+1/\partial} \; |Y|.$$

As n increases $1/\partial$ increases and consequently the absolute bias in t_C decreases. When $n = N$, $\partial = 0$ and hence $t = Y$, so that t_C is a consistent estimator of Y.

7.3 An optimum property of the strategy (p^*, t_1^*)

Rao [47], while proving an optimum property of the Horvitz-Thompson estimator considered the set of parameter values $(\pi_1 Z_{i_1}, \ldots, \pi_N Z_{i_N})$ obtained by permutations of $Z_1, Z_2, \ldots, Z_N$, keeping $\pi_1, \pi_2, \ldots, \pi_N$ fixed, where $Z_i = y_i / \pi_i$ and π_i's are the first order inclusion probabilities for the design. The above parametrization seems to be natural because in the case of unequal probability sampling, presumably large values of y_i are given large probabilities of inclusion. In the case of simple random sampling, all the π_i's are the same, and the above parametrization reduces to that considered in the previous two sections.

The above parametrization is dependent on the design. In order to avoid the difficulties arising at the stage of selection of an optimum design, corresponding to an estimator in our method, we modify the above parametrization to make it independent of the design as follows:

We assume that values $x_1, x_2, \dots, x_N$ of an auxiliary variable correlated with y are available for all the units in the population. We consider the set of parameter values $(x_1 Z_{i_1}, \dots, x_N Z_{i_N})$ obtained by permutations of $Z_1, Z_2, \dots, Z_N$, keeping $x_1, x_2, \dots, x_N$ fixed where $Z_i = y_i/x_i$, $1 \leq i \leq N$.

Let L_{ox} be the class of all linear estimators

$$t = \sum_{i \in s} b(s,i) y_i \qquad (7.3.1)$$

such that t_s reduces to $X = \sum_{i=1}^{N} x_i$ at the point $x = (x_1, x_2, \dots, x_N)$

i.e.

$$t_s(x) = \sum_{i \in s} b(s,i) x_i = X, \text{ for all } s \in S. \qquad (7.3.2)$$

Now the mean square error of t is given by

$$mse(t) = V(t) + B^2(t)$$

where V and B stand for variance and bias respectively. Denoting by $\mathcal{E}'$, averaging over all permutations of $Z_1, Z_2, \ldots, Z_N$, one can check that for any $t \in L_{ox}$

$$\mathcal{E}' \, mse(t) = (\alpha - \beta) \sum_{i=1}^{N} (b_i - 2a_i + 1) x_i^2 \qquad (7.3.3)$$

where

$$\alpha = \frac{1}{N} \sum_{i=1}^{N} \frac{y_i^2}{x_i^2}, \qquad \beta = \frac{1}{N(N-1)} \sum_{i \neq j} \sum \frac{y_i}{x_i} \frac{y_j}{x_j}$$

$$a_i = \sum_{s \supset i} b(s,i)p(s) \quad \text{and} \quad b_i = \sum_{s \supset i} b^2(s,i)p(s).$$

We now prove

<u>Theorem 7.3.1:</u> For any design $d(S,p)$, the best estimator (in the sense of having minimum $\mathcal{E}'$ mse) in L_{ox} is given by

$$t_1^* = \sum_s y_i + \left(\sum_{\bar{s}} x_i \right) \frac{1}{n(s)} \sum_s y_i/x_i \,. \qquad (7.3.4)$$

<u>Proof:</u> From (7.3.3) and noting that $(\alpha - \beta) \geq 0$, to obtain the best estimator in L_{ox}, we have to minimise $\sum_{i=1}^{N} (b_i - 2a_i) x_i^2$

subject to condition (7.3.2) with respect to the coefficients $b(s,i)$'s. Let

$$\phi = \sum_{i=1}^{N} (b_i - 2a_i) x_i^2 - \sum_{s \in S} \lambda(s) \left(\sum_{i \in s} b(s,i) x_i - X \right) \qquad (7.3.5)$$

where $\lambda(s)$, $s \in S$ are Lagrangian multipliers.

$$\frac{\partial \phi}{\partial b(s,i)} = 0 \quad \text{gives}$$

$$b(s,i) x_i p(s) - \text{[illegible]}(s) = \lambda(s). \qquad (7.3.6)$$

Summing (7.3.5) over $i \in s$ and using (7.3.6), we get

$$\lambda(s) = \left(\sum_{i \in \bar{s}} x_i \right) \frac{p(s)}{n(s)} .$$

Substituting this value of $\lambda(s)$ in (7.3.6), it follows that the optimum values of $b(s,i)$'s are given by

$$b'(s,i) = 1 + \frac{1}{x_i} \frac{\left(\sum_{\bar{s}} x_i \right)}{n(s)} .$$

Hence the optimum estimator is given by

$$t_1^* = \sum_{i \in s} b'(s,i) y_i = \sum_{i \in s} y_i + \left(\sum_{i \in \bar{s}} x_i \right) \frac{1}{n(s)} \sum_{i \in s} y_i / x_i .$$

q.e.d.

Having decided to use t_1^* for any design, our problem is to choose that design for which $\mathcal{E}'\, mse(t_1^*)$ is a minimum. We will restrict our attention to fixed sample size(=n) designs. We prove

Theorem 7.3.2: The optimum fixed sample size (=n) design for the use of t_1^*, is to select that s which maximises $\sum_{i \in s} x_i$, with probability one.

Proof: From (7.3.3) it is clear that we have to choose p(s) for all the $\binom{N}{n}$ samples of size n which minimises

$$f = \sum_{i=1}^{N} \left[\sum_{s \supset i} b'^2(s,i)p(s) - 2 \sum_{s \supset i} b'(s,i)p(s) \right] x_i^2$$

subject to $p(s) \geq 0$ and $\sum_{s=1}^{\binom{N}{n}} p(s) = 1$. It is easy to check that

$$f = \sum_{s=1}^{\binom{N}{n}} C(s)p(s) - \frac{n^2 - 1}{n} X^2 \qquad (7.3.7)$$

where

$$C(s) = - \sum_{i \in s} x_i^2 + n(X - \bar{x})^2. \qquad (7.3.8)$$

To minimise f, clearly we have to select that s for which C(s) is a minimum with probability one. From (7.3.8), this is equivalent to choosing that s which maximises $\sum_{i \in s} x_i$ with probability one.

q.e.d.

Thus we have arrived at the strategy (p^*, t_1^*) (vide Section 7.1) obtained by Royal [55] using a different approach.

Remark 7.3.1: We considered only those linear estimators which have zero mean square error at the point $(x_1, x_2, \ldots, x_N)$. This assumption seems reasonable, since we assume y and x are correlated and in the ideal case of perfect correlation we wish to estimate Y with zero mean square error.

7.4 Unbiasedness

For any fixed sample size (=n) design p, the condition of unbiasedness of t_1^* in (7.3.4) for Y is given by

$$\sum_{s \supset i} \left(x_i + \frac{X}{n} - \bar{x}_s\right) p(s) = x_i, \quad 1 \leq i \leq N. \tag{7.4.1}$$

Also we have

$$p(s) \geq 0 \quad \text{and} \quad \sum_{s} p(s) = 1. \tag{7.4.2}$$

Considering only fixed sample size designs for which t_1^*

is unbiased for Y, the optimum design corresponding to t_1^* is obtained by minimising

$$\sum_s C(s)p(s)$$

with respect to $p(s)$'s subject to the linear (in $p(s)$) restrictions (7.4.1) and (7.4.2). This is a typical linear programming problem and can be theoretically solved by simplex method. But it may be remarked that the available algorithms in linear programming are of little practical use in solving the above survey problem even for moderately large N and n.

7.5 An optimum strategy for Regression estimation

Royal [55] considered a super-population in which the regression of y on x was assumed to be a straight line passing through the origin. In this section we consider a more general model which seems appropriate for regression estimation. We assume the following super-population model: The numbers $y_1, \ldots, y_N$ are realised values of independent random variables $Y_1, Y_2, \ldots, Y_N$ where Y_i has mean $ax_i + b$, and variance $\sigma^2 v(x_i)$. The function v is known and $v(x) > 0$ for $x > 0$, and the constants a, b and σ^2 are unknown. We denote the joint probability law of $Y_1, \ldots, Y_N$ by ξ. We now prove

<u>Theorem 7.5.1</u>: For any design $d(S,p)$ the best ξ-unbiased (Definition 7.1.1) linear estimator is given by

$$t_2^* = \sum_s y_i + \beta^*(s) \sum_{\bar{s}} x_i \tag{7.5.1}$$

where $\beta^*(s)$ is obtained from the equation

$$2\beta^*(s)\left(\sum_{\bar{s}} x_i\right)p(s) = \lambda'(s) \sum_s \frac{x_i\, y_i}{v(x_i)} + \mu'(s) \sum_s \frac{y_i}{v(x_i)} \tag{7.5.2}$$

and

$$\lambda'(s) = \frac{2(N-n(s))p(s) \sum_s \frac{x_i}{v(x_i)} - 2\left(\sum_{\bar{s}} x_i\right)p(s) \sum_s \frac{1}{v(x_i)}}{\left[\sum_s \frac{x_i}{v(x_i)}\right]^2 - \left[\sum_s \frac{x_i^2}{v(x_i)}\right]\left[\sum_s \frac{1}{v(x_i)}\right]} \tag{7.5.3}$$

$$\mu'(s) = \frac{2(N-n(s))p(s) - \left[\sum_s \frac{x_i}{v(x_i)}\right]\lambda'(s)}{\sum_s \frac{1}{v(x_i)}} . \tag{7.5.4}$$

<u>Proof:</u> A typical linear estimator

$$t = \sum_{i \in s} b(s,i)\, y_i$$

can be written in the form

$$t = \sum_s y_i + \beta(s) \sum_s x_i$$

where $\beta(s) = \dfrac{\sum_s (b(s,i) - 1)y_i}{\sum_s x_i}$.

Now t is ξ-unbiased for Y if and only if

$$E_\xi [\sum_s Y_i - \beta(s) \sum_s x_i] = 0, \text{ for every } s \in S.$$

i.e. $\sum_s (ax_i + b) - \sum_s (b(s,i)-1)(ax_i+b) = 0$, for every $s \in S$

i.e. $a(X - \sum_s b(s,i)x_i) + b(N - \sum_s b(s,i)) = 0$, for every $s \in S$.

(7.5.5)

Since a and b are unknown constants, from (7.5.5), the condition for ξ-unbiasedness of t becomes

$$\left.\begin{aligned} \sum_s b(s,i)\, x_i &= X \\ \sum_s b(s,i) &= N \end{aligned}\right\} \text{ for every } s \in S. \qquad (7.5.6)$$

Now

$$mse(t) = E(t - Y)^2 = E[\sum_s (b(s,i)-1)y_i - \sum_{\bar{s}} y_i]^2 \qquad (7.5.7)$$

Taking expectation over ξ on both sides of (7.5.7) and using the condition (7.5.6) for ξ-unbiasedness, we have

$$E_{\xi}\ \text{mse}(t) = \sum_{s\in S}\ \sum_{i\in s} (b(s,i)-1)^2 v(x_i)p(s)$$

$$+ \text{ terms not containing } b(s,i)\text{'s.} \qquad (7.5.8)$$

From (7.5.8), to minimise the expected mean square error of t we have to minimise

$$\sum_{s\in S}\ \sum_{i\in s} (b(s,i)-1)^2\ v(x_i)p(s) \qquad (7.5.9)$$

subject to (7.5.6).

Introducing Lagrangian multipliers $\lambda(s)$, $\mu(s)$, $s\in S$, let

$$\emptyset = \sum_{s\in S}\ \sum_{i\in s} (b(s,i)-1)^2 v(x_i)p(s) - \sum_{s\in S} \lambda(s)\Big(\sum_{i\in s} b(s,i)x_i - X\Big)$$

$$- \sum_{s\in S} \mu(s)\Big(\sum_{i\in s} b(s,i) - N\Big). \qquad (7.5.10)$$

Differentiating $\emptyset$ w.r.t. $b(s,i)$ and equating to zero

$$2(b(s,i)-1)v(x_i)p(s) - \lambda(s)x_i - \mu(s) = 0$$

i.e.

$$2(b(s,i)-1)p(s) = \lambda(s)\frac{x_i}{v(x_i)} + \mu(s)\frac{1}{v(x_i)}\ . \qquad (7.5.11)$$

Summing (7.5.11) over $i\in s$ and using (7.5.6)

$$2(N-n(s))p(s) = \lambda(s)\sum_{i\in s}\frac{x_i}{v(x_i)} + \mu(s)\sum_{i\in s}\frac{1}{v(x_i)}. \qquad (7.5.12)$$

Multiplying (7.5.11) by x_i and summing over $i \in s$ and using (7.5.6), we get

$$2. \sum_{i \in \bar{s}} x_i = \lambda(s) \sum_{i \in s} \frac{x_i^2}{v(x_i)} + \mu(s) \sum_{i \in s} \frac{x_i}{v(x_i)} . \tag{7.5.13}$$

Solving (7.5.12) and (7.5.13) for $\lambda(s)$ and $\mu(s)$, one can easily check that the solution is given by $\lambda'(s)$ and $\mu'(s)$ as given in (7.5.3) and (7.5.4) respectively. Substituting these values of $\lambda'(s)$ and $\mu'(s)$ in (7.5.11) we get the optimum values of the $b(s,i)$'s, which gives the optimum value of $\beta(s)$ as $\beta^*(s)$ in (7.5.2).

q.e.d.

Remark 7.5.1: In the particular case $v(x_i) = 1$, $1 \leq i \leq N$ i.e. the variables Y_i, $1 \leq i \leq N$ are distributed independently with same variance σ^2, it can be seen that t_2^* reduces to the ordinary regression estimator, namely

$$\hat{Y}_{reg} = N[\, \bar{y} + \frac{s_{xy}}{s_x^2} (\bar{X} - \bar{x})]. \quad \text{(Here } \bar{y} \text{ and } \bar{x}$$

are the means of the distinct units in the sample).

Having decided to use t_2^*, the best ξ-unbiased linear estimator, our next logical step is to choose the corresponding optimum design. As in the earlier section we will restrict our attention to fixed sample size designs. We prove

Theorem 7.5.2: A ξ-optimum fixed sample size (=n) design corresponding to t_2^* is p_2^*, where p_2^* gives probability one to the sample which minimises

$$K(s) = \frac{\left((N-n)\lambda_2/\lambda_1 + \sum_s x_i - X\right)^2}{\lambda_3 - \frac{\lambda_2^2}{\lambda_1}} - \frac{(N-n)^2}{\lambda_1} + \sum_{i \in \bar{s}} v(x_i) \tag{7.5.14}$$

where $\lambda_1 = \sum_{i \in s} \frac{1}{v(x_i)}$, $\lambda_2 = \sum_{i \in s} \frac{x_i}{v(x_i)}$ and $\lambda_3 = \sum_{i \in s} \frac{x_i^2}{v(x_i)}$.

(7.5.15)

Proof: Now the expected mean square error of t_2^* is given by

$$E_\xi \, mse(t_2^*) = E_\xi \sum_{s \in S} p(s)(t_2^* - Y)^2$$

$$= \sum_{s \in S} p(s) \, E_\xi \, (t_2^* - Y)^2. \tag{7.5.16}$$

After tedious calculations one can show that

$$E_{\xi} \; (t_2^* - Y)^2 = K(s) \qquad (7.5.17)$$

where $K(s)$ is as given in (7.5.14). It follows immediately that (7.5.16) will be a minimum if we choose that sample for which (7.5.17) is minimum, with probability one.

CHAPTER VIII

CHOICE OF AN OPTIMUM STRATEGY - CONTD.

8.0 Summary

Construction of πPS (or I.P.P.S.) designs has received much attention recently. We characterise the class of all πPS designs and show that the cardinality of it is c, where c is the cardinality of the continuum. Also we characterise the class of all πPS designs which have specified second order inclusion probabilities π_{ij}'s. It is noted that $L(HT,\mu)$ contains uncountably infinitely many strategies by noting that there exist c many strategies in any equivalence class of $L(HT,\mu)$, the equivalence relation being that $H_1, H_2 \in L(HT,\mu)$ are equivalent if $\pi_i(H_1) = \pi_i(H_2)$, $1 \leq i \leq N$.

We disprove a conjecture due to Hanurav [22] which is claimed to have been proved by T. J. Rao [53] after pointing out the error in the latter's proof. We study its consequences in the existence of optimal strategies in the equivalence classes of $L(HT, \mu)$.

8.1 Introduction

We have already proved the non-existence of a best (in the sense of uniformly minimum variance) strategy in the class, $L(H,\mu)$, of all equi-cost linear unbiased strategies (vide Remark 6.3.1). Consequently various optimality criteria have been put forward for the choice of a sampling strategy. We have briefly mentioned the super-population approach to the selection of an optimum strategy in the class $L(H,\mu)$, in Chapter I. The criterion of strong admissibility was introduced in Chapter VI and it was proved there that the class, $L(HT, \mu)$, of all HT-strategies contained in $L(H,\mu)$, is complete in $L(H,\mu)$ with respect to strong admissibility. But the class $L(HT, \mu)$, is itself two wide and in fact we have stated in Remark (6.3.1) that there exists uncountably infinite number of strategies in $L(HT, \mu)$. We prove this claim in Section (8.2) by showing that there exists infinitely many strategies in any equivalence class of $L(HT, \mu)$, the equivalence relation being that H_1, $H_2 \in L(HT,\mu)$ are said to be equivalent if $\pi_i(d_1) = \pi_i(d_2)$, $1 \leq i \leq N$, where $H_1 = (d_1,\bar{e})$ and $H_2 = (d_2,\bar{e})$. Hence our final problem of pinpointing a strategy remains unsolved.

One can easily verify that any two strategies belonging to two different equivalence classes of $L(HT,\mu)$, are not comparable (with respect to variance). As a consequence an interesting problem is to see whether there exists a best strategy in an equivalence class at least.

Now H_1, H_2 $\in$ an equivalence class =>

$$V(H_1,y) - V(H_2,y) = \sum_{i \neq j}^{N}\sum \frac{\pi_{ij}(d_1) - \pi_{ij}(d_2)}{\pi_i(d_1)\quad \pi_j(d_1)} y_i y_j . \tag{8.1.1}$$

If we restrict $y \geq 0$ - often realistically - it is immediate from (8.1.1) that a set of necessary and sufficient conditions for H_2 to be superior to H_1 is that

$$\pi_{ij}(d_2) \leq \pi_{ij}(d_1), \quad 1 \leq i \neq j \leq N, \tag{8.1.2}$$

with strict inequality holding for, at least, one $i \neq j$. Given a design d_1, the problem reduces to that of constructing a design d_2 with the same π_i's and with uniformly smaller (or equal) π_{ij}'s. Hereafter, owing to the repeated use of this, we denote this property by $\mathcal{S}$.

Lemma 8.1.1 (Hanurav [22]): For any design $d(S,p)$

$$\sum_{i\neq j}\sum \pi_{ij}(d) = \mu(\mu-1) + V(n(s)) \geq \mu(\mu-1)+f(1-f) \quad (8.1.3)$$

where μ is the expected effective sample size, namely, $\sum_{i=1}^{N} \pi_i(d)$ and f is the fractional part of μ. It immediately follows from the above lemma that it is not possible to construct a design d_2 with the required property $\mathcal{P}$, when $\sum_{i\neq j}\sum \pi_{ij}(d_1)$ attains its lower bound given in (8.1.3), viz., $\mu(\mu-1) + f(1-f)$. *Hanurav [22] conjectured that it would* be possible to construct a design d_2, satisfying $\mathcal{P}$ if

$$\sum_{i\neq j}\sum \pi_{ij}(d_1) > \mu(\mu-1) + f(1-f). \quad (8.1.4)$$

T. J. Rao [53] claimed to have proved Hanurav's conjecture by giving a constructive proof. In Section (8.3) we show that T. J. Rao's claim is wrong and in Section (8.4) we disprove Hanurav's conjecture by supplying a simple counter-example.

8.2 Construction of πPS designs

Given a set of numbers π_i, $1 \leq i \leq N$, such that

$$0 \leq \pi_i \leq 1 \qquad \text{and} \qquad \sum \pi_i \geq 1 \quad (8.2.1)$$

is it possible to construct a design with the given π_i's as its inclusion probabilities for the units?

The answer is known to be 'yes' and there is a huge literature available on constructions of such designs and they are known as πPS designs (in analogy to pps designs) or I.P.P.S. designs (inclusion probability proportional to size designs) in the literature.

Let us recall the definition of a design. A probability function P on S, the set of all finite ordered sequences of units from a finite population U of N units, is called a design. Equivalently we can define a design as follows: Let $\mathcal{A}$ be the σ-algebra of all subsets of S. A probability measure P on $(S, \mathcal{A})$ is called a design. Let

$$S_i = \{ s: i \in s \}, \quad 1 \leq i \leq N. \tag{8.2.2}$$

Clearly S_i is the set of samples containing the i-th unit and $\bigcup_{i=1}^{N} S_i = S$. Given a set of π_i's satisfying (8.2.1) the problem reduces to the construction of a probability P on $(S, \mathcal{A})$ such that

$$P(S_i) = \pi_i, \quad 1 \leq i \leq N. \tag{8.2.3}$$

Let $\mathcal{B}$ be the smallest σ-algebra containing the sets $S_1, S_2, \ldots, S_N$. It can be easily shown that sets of the form

$$B = \bigcap_{i=1}^{N} D_i \quad \text{where} \quad D_i = S_i \quad \text{or} \quad S_i^c \qquad (8.2.4)$$

are precisely the atoms of $\mathcal{B}$ [note that $B_o = \bigcap_{i=1}^{N} S_i^c = (\bigcup_{i=1}^{N} S_i)^c = S^c = \emptyset$ and hence there are, in all $2^N - 1$ nonempty atoms of $\mathcal{B}$]. and that $\cup B = S$. Clearly a probability **Q** on $(S, \mathcal{B})$ is uniquely specified by its values on the atoms B in (8.2.4) and hence the following

<u>Theorem 8.2.1:</u> A necessary and sufficient condition for the existence of a probability Q on $(S, \mathcal{B})$ such that $Q(S_i) = \pi_i$, $1 \leq i \leq N$ is that there exists a non-negative solution to the system of $N+1$ linear equations in 2^N-1 unknown $x(B)$'s, given by

$$\left.\begin{aligned} &\sum_{B} x(B) = 1 \\ &\sum_{B \subseteq S_i} x(B) = \pi_i, \quad 1 \leq i \leq N. \end{aligned}\right\} \qquad (8.2.5)$$

Now we prove

Theorem 8.2.2: If π_i, $1 \leq i \leq N$ satisfy (8.2.1) there exists a non-negative solution to the system of equations (8.2.5).

Proof: Let P be a probability on $(S, \mathcal{A})$ such that $P(S_i) = \pi_i$, $1 \leq i \leq N$, [existence of such a P is guaranteed because we have already noted before that various authors have constructed P's satisfying (8.2.3)] and let Q be the restriction of P to $(S, \mathcal{B})$. From Theorem (8.2.1), we see that, $\{Q(B)\}$ provides a non-negative solution to (8.2.5).

Remark 8.2.1: One can construct examples to show that the solution to (8.2.5) need not be unique.

We are now in a position to characterise the class of all P's on $(S, \mathcal{A})$ satisfying (8.2.3). Let $\{Q(B)\}$ be a non-negative solution to (8.2.5). Allot Q(B) to the countably infinite number of samples s contained in B in any arbitrary manner such that

$$Q(s) \geq 0 \quad \text{and} \quad \sum_{s \in B} Q(s) = Q(B). \tag{8.2.6}$$

Do this for every non-empty atom $B \in \mathcal{B}$. Clearly any particular mode of allotment of Q(B) to the samples $s \in B$, for all $B \in \mathcal{B}$, specifies a probability P on $(S, \mathcal{A})$ such that

$$P(S_i) = Q(S_i) = \pi_i, \quad 1 \leq i \leq N.$$

We say that such a P is generated by Q. Corresponding to any solution Q for (8.2.5) let $\mathcal{P}_Q$ denote the set of probabilities P on $(S, \mathcal{A})$ generated by Q. The following theorem now becomes obvious.

Theorem 8.2.3: $\mathcal{P} = \{ \mathcal{P}_Q : Q \in NS\ (8.2.5) \}$ gives the complete class of designs having the specified inclusion probabilities π_i's where NS (8.2.5) is the set of all non-negative solutions to the system of linear equations (8.2.5).

Remark 8.2.2: It follows that the cardinality of $\mathcal{P}$ is c - where c denotes the cardinality of the continuum - since each $\mathcal{P}_Q$ has cardinality c.

Remark 8.2.3: An equivalence class of $L(HT, \mu)$ as defined earlier contains c many members, follows as a consequence of Remark (8.2.2).

Another important problem concerning the construction of designs is: Given a set of π_i's and π_{ij}'s such that

$$0 \leq \pi_i \leq 1, \; 1 \leq i \leq N \quad \text{and} \quad 0 \leq \pi_{ij} \leq 1, \; 1 \leq i \neq j \leq N \tag{8.2.7}$$

is it possible to construct a design p which has the above π_i's and π_{ij}'s as its first and second order inclusion probabilities respectively?

It is clear that some consistency conditions should be satisfied by the given π_i's and π_{ij}'s. Some necessary conditions to be satisfied by the given π_i's and π_{ij}'s are well-known in literature. For example

i) $\pi_{ij} \leq \min(\pi_i, \pi_j)$

ii) $\pi_{ij} \geq \pi_i + \pi_j - 1$.

No compact set of sufficient conditions to be satisfied by the given π_i's and π_{ij}'s are known in order that there exists at least one design with those as the inclusion probabilities (see Hanurav [22]).

Now there exists a probability Q on $(S, \mathcal{B})$ such that

$$Q(S_i) = \pi_i, \ 1 \leq i \leq N \quad \text{and} \quad Q(S_i \cap S_j) = \pi_{ij}, \quad 1 \leq i \neq j \leq N \qquad (8.2.8)$$

if and only if the following system of $1 + N + \binom{N}{2}$ linear equations in $2^N - 1$ unknown $x(B)$'s has got a non-negative solution:

$$\sum_{B} x(B) = 1,$$

$$\sum_{B \subset S_i} x(B) = \pi_i, \quad 1 \leq i \leq N, \tag{8.2.9}$$

$$\sum_{B \subset S_i \cap S_j} x(B) = \pi_{ij}, \quad 1 \leq i \neq j \leq N.$$

Analogous to Theorem (8.2.3) we have now the following

Theorem 8.2.4: $\mathcal{P}' = \{ \mathcal{P}_Q : Q \in NS(8.2.9) \}$ *gives the complete class of designs having the given* π_i*'s and* π_{ij}*'s as their* inclusion probabilities, where NS(8.2.9) is the set of all non-negative solutions to the system of equations (8.2.9).

8.3 On a proof of Hanurav's conjecture due to T. J. Rao [53]

We describe briefly T. J. Rao's [53] constructive proof of Hanurav's conjecture stated in Section 8.1, before pointing out the fallacy in it. Suppose d_1 is a design satisfying (8.1.4) so that

$$\sum_{i \neq j}^{N} \sum \pi_{ij}(d_1) = \mu(\mu - 1) + f(1 - f) + \Delta , \tag{8.3.1}$$

where $\Delta > 0$.

Let $\pi_{ij}(d_1) = \pi_i(d_1) + \pi_j(d_1) - 1 + \Delta_{ij}(d_1)$, (8.3.2) where $\Delta_{ij}(d_1) \geq 0$, with strict inequality for at least one $i \neq j$, since $\Delta > 0$. Let ∂ be the number of non-zero elements in the matrix

$$A = \left(\left(\begin{matrix}\Delta_{ij}(d_1) \\ i < j\end{matrix}\right)\right) . \tag{8.3.3}$$

Next, he shows that

$$\sum_{\partial} \Delta_{ij}(d_1) > \frac{\Delta}{2} \tag{8.3.4}$$

(Note that this is true only if $\mu \neq N$ or $N-1$) where the summation is taken over the ∂ non-zero elements of the matrix A in (8.3.3) and that there exists a partition of $\frac{\Delta}{2}$ such that $\sum_{i=1}^{\partial} \Delta_i = \Delta/2$, with the property that $\Delta_{ij}(d_2)$, $i < j$, got by subtracting $\Delta_1, \Delta_2, \ldots, \Delta_\partial$ from the non-zero elements of $\Delta_{ij}(d_1)$, $i < j$, are non-negative with at least one positive. One can note that the zero elements of $\Delta_{ij}(d_1)$, $i < j$, remain unchanged while constructing $\Delta_{ij}(d_2)$, $i < j$. Next he constructs the matrix $((\pi_{ij}(d_2)))$, where

$$\pi_{ij}(d_2) = \pi_i + \pi_j - 1 + \Delta_{ij}(d_2)$$

and claims that it satisfies the property Φ, for clearly the new set of numbers $\pi_{ij}(d_2)$ satisfy

i) $\pi_{ij}(d_2) \leq \pi_{ij}(d_1)$ with strict inequality holding at least once,

ii) $\pi_{ij}(d_2) = \pi_{ji}(d_2)$, and

iii) $\pi_{ij}(d_2) \geq \pi_i(d_2) + \pi_j(d_2) - 1$.

But these are only some necessary conditions for the existence of a design d_2 satisfying $\mathcal{P}$ and are not sufficient. The set of numbers $\pi_{ij}(d_2)$ constructed as above, may turn out inconsistent in the sense that there may not exist a design with the given π_i's, namely, $\pi_i(d_1)$'s and the constructed $\pi_{ij}(d_2)$'s as its first and second order inclusion probabilities. This we illustrate with an example in which some $\pi_{ij}(d_2)$'s as constructed by T. J. Rao's [53] method turn out to be negative - obviously inconsistent - which incidentally contradicts an unsubstantiated claim due to T. J. Rao [53] that the $\pi_{ij}(d_2)$'s constructed according to his method are non-negative.

<u>Example</u>: Consider a population of size $N = 4$ and let the sampling design $d_1(s, p)$ be as follows:

sample s	probability $p(s)$
1	.25
1,3	.50
1,2,4	.25
	$\sum_{s \in S} p(s) = 1.$

We have corresponding to this design,

$$\pi_1(d_1) = 1, \quad \pi_2(d_1) = .25, \quad \pi_3(d_1) = .50 \quad \text{and} \quad \pi_4(d_1) = .25$$

so that $\sum_{i=1}^{4} \pi_i(d_2) = \mu = 2$, and

$$\pi_{12}(d_1) = .25 = 1 + .25 - 1 + (0)$$
$$\pi_{13}(d_1) = .50 = 1 + .50 - 1 + (0)$$
$$\pi_{14}(d_1) = .25 = 1 + .25 - 1 + (0)$$
$$\pi_{23}(d_1) = 0 = .25 + .50 - 1 + (.25)$$
$$\pi_{24}(d_1) = .25 = .25 + .25 - 1 + (.75)$$
$$\pi_{34}(d_1) = 0 = .50 + .25 - 1 + (.25)$$

and $\sum_{i \neq j}^{4}\sum \pi_{ij}(d_1) = 2.5.$

$\text{Min} \sum_{i \neq j}^{4}\sum \pi_{ij} = \mu(\mu - 1) + f(1-f) = 2$, so that

$$\Delta = \text{Excess} = 0.5.$$

We have

$$((\Delta_{ij}(d_1)))_{i<j} = \begin{bmatrix} * & 0 & 0 & 0 \\ & * & .25 & .75 \\ & & * & .25 \\ & & & * \end{bmatrix}$$

and $\partial = 3$ and $\frac{\Delta}{2} = 0.25$.

Let $\Delta_1 = .25$, $\Delta_2 = 0$ and $\Delta_3 = 0$.

Then

$$((\Delta_{ij}(d_2)))_{i<j} = \begin{bmatrix} * & 0 & 0 & 0 \\ & * & 0 & .75 \\ & & * & .25 \\ & & & * \end{bmatrix}$$

so that

$$((\Delta_{ij}(d_2))) = \begin{bmatrix} * & 0 & 0 & 0 \\ 0 & * & 0 & .75 \\ 0 & 0 & * & .25 \\ 0 & .75 & .25 & * \end{bmatrix}$$

and consequently

$$((\pi_{ij}(d_2))) = \begin{bmatrix} * & .25 & .50 & .25 \\ .25 & * & -.25 & .25 \\ .50 & -.25 & * & 0 \\ .25 & .25 & 0 & * \end{bmatrix}$$

The above set of $\pi_{ij}(d_2)$ is obviously inconsistent since $\pi_{23}(d_2)$ is negative.

8.4 A counter-example

Before proceeding to give a counter-example to disprove Hanurav's conjecture stated in Section (8.1), we give an alternative equivalent statement of it. We prove

Theorem 8.4.1: Given a design satisfying (8.1.4) Hanurav's conjecture is equivalent to saying that there exists a design d_3 satisfying $\oint$ and moreover

$$\sum_{i \neq j}^{N} \sum \pi_{ij}(d_3) = \mu(\mu - 1) + f(1 - f). \qquad (8.4.1)$$

Proof: One way implication is clear, namely, existence of a design d_3 satisfying $\oint$ and moreover (8.4.1) implies Hanurav's conjecture. Hence to prove the theorem, it suffices to show that Hanurav's conjecture implies that there exists a design d_3 satisfying $\oint$ and (8.4.1).

Using the formulation in Section (8.1) it is clear that assuming Hanurav's conjecture, the system of linear constraints

$$\begin{aligned} & x(B) \geq 0 \\ & \sum_{B} x(B) = 1 \\ & \sum_{B \subseteq S_i} x(B) = \pi_i(d_1) \\ & \sum_{B \subseteq S_i \cap S_j} x(B) \leq \pi_{ij}(d_1) \end{aligned} \qquad (8.4.2)$$

is consistent. Now minimise the linear function

$$\phi(x) = \sum_{i \neq j}^{N} \sum_{B \subset S_i \cap S_j} x(B) \qquad (8.4.3)$$

subject to the linear constraints (8.4.2). This is a typical bounded linear programming problem and hence it has an optimal solution. Let $\{ x_o(B) \}$ denote an optimal solution and ϕ_o the corresponding minimum value of the objective function $\phi(x)$ in (8.4.3). Next we show that

$$\phi(x_o) = \phi_o = \mu(\mu - 1) + f(1 - f).$$

If not, using Hanurav's conjecture there exists an x_1 satisfying (8.4.2) such that

$$\phi(x_1) < \phi(x_o) = \phi_o ,$$

a contradiction to the fact that x_o is optimal.

q.e.d.

Now we give a counter-example to disprove Hanurav's conjecture. Consider a population of size $N = 4$ and let the sampling design $d_1(S, p)$ be as follows:

sample s	probability $p(s)$
4	.1
1,2	.1
1,2,3	.8
	$\Sigma\, p(s) = 1$

Clearly $\mu = 2.7$ so that $f = .7$ and

$$\pi_4(d_1) = .1, \quad \pi_{14}(d_1) = \pi_{24}(d_1) = \pi_{34}(d_1) = 0.$$

It is clear that d_1 satisfies (8.1.4) and that there does not exist a design d_3 satisfying (?) and moreover (8.4.1) [i.e. $V_{d_3}(n(s)) = f(1-f)$], since $\pi_{14}(d_3) = \pi_{24}(d_3) = \pi_{34}(d_3) = 0$ so that d_3 has to give probability 0.1 to the sample containing unit 4 alone in order to make $\pi_4(d_3) = 0.1$.

Remark 8.4: Now the equivalence relation, $H_1 \sim H_2$ if $\pi_i(H_1) = \pi_i(H_2)$, $1 \leq i \leq N$, partitions $L(HT, \mu)$ into equivalence classes. From Theorem (8.4.1) we see that Hanurav's conjecture implies the existence of optimal strategies within each equivalence class, if the parameter y is restricted to the positive quadrant of R_N. But unfortunately, the falsity of Hanurav's conjecture entails the possibility of non-existence of optimal strategies in certain equivalence classes. But whenever it exists, from the preceding discussion we see that, it can be obtained by method of linear programming. We illustrate this with an example below. However, it is pertinent to note here that when N is large, the L.P. method for solving the above problem becomes unmanageably ~~different~~ difficult.

Example: Consider a population of size $N=4$ and let the design $d_1(S, p)$ be as follows:

sample s	probability $p(s)$
14	.231
23	.212
24	.063
124	.345
234	.147
1234	.002
	$\Sigma\, p(s) = 1$

Corresponding to this design we have

$\pi_1(d_1) = .578$	$\pi_{12}(d_1) = .347$
$\pi_2(d_1) = .769$	$\pi_{13}(d_1) = .002$
$\pi_3(d_1) = .361$	$\pi_{14}(d_1) = .578$
$\pi_4(d_1) = .788$	$\pi_{23}(d_1) = .361$
	$\pi_{24}(d_1) = .557$
	$\pi_{34}(d_1) = .149$

so that $\mu = \sum_{i=1}^{4} \pi_i(d_1) = 2.496$ and $\sum\sum_{i \neq j}^{4} \pi_{ij}(d_1) = 3.988.$

Also

$$\sum\sum_{i \neq j}^{4} \pi_{ij}(d_1) = 3.988 > \mu(\mu -1)+f(1 - f) = 3.984$$

so that condition (8.1.4) is satisfied. Since $N = 4$, we will have $2^4-1 = 15$ non-empty atoms B, as given in (8.2.4) and the system of linear constraints (8.4.2) becomes

$$x_i \geq 0, \quad 1 \leq i \leq 15$$

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}		
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	=	1
1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	=	.578
1	1	1	0	1	0	0	0	1	0	1	1	0	0	1	=	.769
1	1	0	1	0	1	0	0	1	1	0	1	0	1	0	=	.361
1	0	1	1	0	0	1	0	1	1	1	0	1	0	0	=	.788
1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	$\leq$	.347
1	1	0	1	0	1	0	0	0	0	0	0	0	0	0	$\leq$	.002
1	0	1	1	0	0	1	0	0	0	0	0	0	0	0	$\leq$	.578
1	1	0	0	0	0	0	0	1	0	0	1	0	0	0	$\leq$	.361
1	0	1	0	0	0	0	0	1	0	1	0	0	0	0	$\leq$	.557
1	0	0	1	0	0	0	0	1	1	0	0	0	0	0	$\leq$	.149

(8.4.4)

The objective function $\emptyset$ in (8.4.3) becomes

$$\emptyset = 2(6x_1+3x_2+3x_3+3x_4+x_5+x_6+x_7+3x_9+x_{10}+x_{11}+x_{12}).$$

Now we minimise $\emptyset$ subject to the linear constraints (8.4.4). Simplex algorithm is used and computations were done using IBM 1401 machine. The optimal solution obtained is:

$$x_{3opt} = .347,\ x_{4opt} = .002,\ x_{7opt} = .229,\ x_{9opt} = .147$$

$$x_{11\ opt} = .063,\ x_{12\ opt} = .212 \text{ and the rest of } x_{iopt} = 0.$$

It is easily verified that

$$\emptyset_{opt} = 3.984 = \mu(\mu - 1) + f(1 - f).$$

A design, dopt, corresponding to the optimal solution $\{ x_{iopt} \}$ is given below:

sample s	probability p(s)
1,4	.229
2,3	.212
2,4	.063
1,2,4	.347
1,3,4	.002
2,3,4	.147
	$\Sigma\, p(s) = 1$

It can be easily checked that

$$\pi_i(\text{dopt}) = \pi_i(d_1), \quad 1 \le i \le 4$$

$$\pi_{ij}(\text{dopt}) \le \pi_{ij}(d_1), \quad 1 \le i \ne j \le 4$$

and $$\sum_{i \ne j}^{4} \sum \pi_{ij}(\text{dopt}) = \mu(\mu - 1) + f(1 - f) = 3.984.$$

CHAPTER IX

SOME FURTHER RESULTS ON $\bar{e}$ (s,y) AND ITS VARIANCE ESTIMATION

9.0 Summary

Godambe and Joshi [17] have proved that the HT-estimator $\bar{e}$ (s, y) is admissible *in the class* of all unbiased estimators of the population total for any design. We give an alternate proof of this result with a more general definition of a design using induction on N in Section (9.1). In Section (9.2) we demonstrate the non-existence of a Δ_2-best strategy when μ is not an integer. In the last section we modify the well-known Yates and Grundy variance estimator to suit varying sample size designs and after deriving a set of sufficient conditions for its non-negativity we prove the existence of such designs in a special but important case.

9.1 Admissibility of the HT-estimator $\bar{e}(s,y)$

We give some definitions and notations for the exclusive use in this section.

Population of size N: $\bar{U}_N = \{1, 2, 3, \ldots, N\}$.

Sample: A subset of integers from U_N. There are, in all, 2^N samples which we denote by S.

Design: A set function p on S such that

$$p(s) \geq 0, \quad \text{for all } s \in S \quad \text{and} \quad \sum_s p(s) = 1.$$

$p(\emptyset)$ may also be greater than zero and selection of $\emptyset$ signifies that no unit is selected in the sample. Clearly this definition of a design is more general the usual one in which $p(\emptyset)$ is defined to be zero.

Estimator: A real valued function $e(s,y)$ on $S \times R_N$ such that (i) $e(\emptyset,y) = 0$ for all $y \in R_N$ and

(ii) $e(s,y)$ depends on y through only those y_i's for which $i \in s$.

$A_N(p)$: Class of all unbiased estimators of $Y = \sum_1^N y_i$, for the design p.

We now prove

Theorem 9.1.1: For any design p, for which $0 < \pi_i(p) \leq 1$, $(1 \leq i \leq N)$ the HT-estimator $\bar{e}(s,y)$ is admissible (Definition 1.1.5) in $A_N(p)$.

Proof: We propose to prove the theorem using induction on N. For $N = 1$, the theorem is evident since $A_1(p)$ consists of a single element, namely, $\bar{e}(s,y)$. We now assume that the theorem is true for $N = N_0 (\geq 1)$. That is, for any population of size N_0 and any design p for which $0 < \pi_i(p) \leq 1$, $(1 \leq i \leq N_0)$ $\bar{e}(s,y)$ is admissible in $A_{N_0}(p)$. We shall show that the theorem is true for $N = N_0+1$. That is, we have to show that given any design p^* corresponding to U_{N_0+1},

$$t^*(s^*,y^*) \in A_{N_0+1}(p^*),\ V(t^*,y^*) \leq V(\bar{e},y^*),\ \text{for all}\ \ y^* \in R_{N_0+1}$$

$$\Rightarrow t^*(s^*,y^*) = \bar{e}(s^*,y^*),\ \text{for all}\ \ s^* \in S^*,\ y^* \in R_{N_0+1} \tag{9.1.1}$$

[* will always correspond to something concerning the population U_{N_0+1}]. Let $U_{N_0}(j)$ denote the population of size N_0 obtained from U_{N_0+1} by removing unit j from it. From p^*, we construct a design p corresponding to $U_{N_0}(j)$ as follows: Remove unit j from every sample s^* containing it and put $p(\emptyset) = p^*(\emptyset) + p^*(j)$ where (j) denotes the sample containing unit j alone. Also define

$$t(s, y) = t^*(s^*, y^*) \qquad (9.1.2)$$

where

$$y = (y_1, \ldots, y_{j-1}, y_{j+1}, \ldots, y_{N_0+1}) \in R_{N_0} \qquad (9.1.3)$$

$$y^* = (y_1, \ldots, y_{j-1}, 0, y_{j+1}, \ldots, y_{N_0+1}) \in R_{N_0+1} \qquad (9.1.4)$$

and s is obtained from s^* as explained above. Since t^* is unbiased for $\sum_1^{N_0+1} y_i$, it is clear that t is unbiased for $\sum_{i(\neq j)}^{N_0+1} y_i$ and moreover

$$V(t, y) = V(t^*, y^*) \qquad (9.1.5)$$

and

$$V(\bar{e}, y) = V(\bar{e}, y^*) \qquad (9.1.6)$$

where y and y^* are as in (9.1.3) and (9.1.4) respectively. From (9.1.1), (9.1.5), (9.1.6) and using the induction assumption, we have

$$t(s, y) = \bar{e}(s, y), \text{ for all } s \in S,\ y \in R_{N_0}. \qquad (9.1.7)$$

Since we can remove any unit j from U_{N_0+1} and noting that an estimator depends only on those y_i's for which $i \in s$, it follows from (9.1.2) and (9.1.7) that

$$t^*(s^*, y^*) = \bar{e}(s^*, y^*), \text{ for all } s^*$$

except perhaps for the sample $s^*_o = (1,2,\ldots, N_o+1)$. Now since t^* is unbiased for $\sum_{1}^{N_o+1} y_i$ it follows that $t^*(s^*_o, y^*)$ should also be linear and hence equal to $\bar{e}(s^*_o, y^*)$.

q.e.d.

Note: It is clear from the construction of p **that** it may happen that two samples s^*_1 and s^*_2 get mapped into the same s. But this should not cause any ambiguity in the definition of t, for if $t^*(s^*_1, y^*) \neq t^*(s^*_2, y^*)$ where y^* is as in (9.1.4), the estimator $t(s,y)$ obtained by averaging $t^*(s^*, y^*)$ over all those samples s^* which get mapped into s will be uniformly better than the HT-estimator $\bar{e}$, for the population $U_{N_o}(j)$ - a contradiction to the induction assumption.

9.2 Nonexistence of a Δ_2-best strategy when μ is not an integer

For any given number μ, $1 \leq \mu \leq N$ let $L_1(H_1, \mu)$ denote the class of all strategies contained in $L(H, \mu)$ satisfying

$$H_o \in L_1(H,\mu) \Rightarrow V_{H_o}(n(s)) = f(1-f) \qquad (9.2.1)$$

where $\mu = n+f$ and, n and f are the integral and fractional parts of μ respectively. When μ is an integer it is clear that $L_1(H,\mu)$ consists of all fixed sample size $(=\mu)$ strategies contained in $L(H,\mu)$. Let $A_1(H,\mu)$ be the class of all strategies belonging to $A(H,\mu)$ such that

$$H_o \in A_1(H, \mu) \Rightarrow V_{H_o}(n(s)) = f(1-f).$$

We consider the same super-population model as was considered earlier (vide Section 1.3). Let Δ_g denote the class of all prior distributions ∂ for which

$$\begin{aligned} \mathcal{E}_\partial(y_i \mid x_i) &= a\, x_i \\ \mathcal{V}_\partial(y_i \mid x_i) &= \sigma^2\, x_i^g \qquad (9.2.2) \\ \mathcal{C}_\partial(y_i, y_j \mid x_i, x_j) &= 0 \end{aligned}$$

where $\mathcal{C}$ denotes covariance and $x_i(> 0)$, $1 \leq i \leq N$ are the known values of an auxiliary variable $\mathcal{X}$ correlated with $\mathcal{Y}$. Godambe has proved

Theorem 9.2.1: When μ is an integer any strategy $H_o(d_o, \bar{e}) \in L_1(H,\mu)$ such that $\pi_i(d_o) = \mu\, x_i/X$, $1 \leq i \leq N$ is Δ_2 - best in $L_1(H, \mu)$.

This result has been generalised to $L(H, \mu)$, $A_1(H,\mu)$ and $A(H, \mu)$ by Hanurav [21], Godambe and Joshi [17] and

Vijayan [64] respectively. It is known that the theorem is not true for the class Δ_g with $g \neq 2$. We propose to show that the theorem is not true even for Δ_2, if μ is not an integer. Let C_x denote the coefficient of variation for $\mathcal{X}$ and C that for the super-population, given by

$$C_x = \frac{\sigma_x}{\bar{X}} \quad \text{and} \quad C = \frac{\sigma}{a} \tag{9.2.3}$$

where $\sigma_x^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{X})^2$. Consider the strategy $H_1(d_1, t_1)$ where d_1 is the design obtained by srswor of size n and $n+1$ with probabilities $1-f$ and f respectively and t_1 is the usual estimator, namely $N\bar{y}$, for the population total. One can easily check that $H_1(d_1, t_1) \in L_1(H, \mu)$. We now prove

<u>Theorem 9.2.2</u>: If μ is not an integer and

$$\left(\frac{n+1-f}{n(n+1)} - \frac{1}{N}\right) \frac{N}{N-1} \; C_x^2 < \frac{f(1-f)}{\mu^2} \tag{9.2.4}$$

then there exists a value C_o for the super-population parameter C, such that $H_1(d_1, t_1)$ is more precise or less precise (in the Δ_2-expected variance sense) than the strategy $H_o(d_o, \bar{e})$ in Theorem (9.2.1) according as $C < C_o$ or $C > C_o$.

For $\theta = C_o$, the two strategies are equally efficient.

Proof: One can easily check that, for any $\partial \in \triangle_2$

$$\mathcal{E}_{\partial} V(H_1, y) = X^2 a^2 (k_1 + k_2 C^2) \tag{9.2.5}$$

and

$$\mathcal{E}_{\partial} V(H_o, y) = X^2 a^2 (K_3 + k_4 C^2) \tag{9.2.6}$$

where

$$k_1 = \left(\frac{n+1-f}{n(n+1)} - \frac{1}{N} \right) \frac{N}{N-1} C_x^2 \tag{9.2.7}$$

$$k_2 = \left(\frac{n+1-f}{n(n+1)} - \frac{1}{N} \right) (1 + C_x^2) \tag{9.2.8}$$

$$k_3 = \frac{f(1-f)}{\mu^2} \tag{9.2.9}$$

and
$$k_4 = \left(\frac{1}{\mu} - \frac{1 + C_x^2}{N} \right). \tag{9.2.10}$$

From (9.2.5) and (9.2.6)

$$\mathcal{E}_{\partial} V(H_1, y) - \mathcal{E}_{\partial} V(H_o y) = X^2 a^2 (\overline{k_1 - k_3} + \overline{k_2 - k_4}\; C^2) \tag{9.2.11}$$

One can easily check that $k_2 - k_4 > 0$ and (9.2.4) implies that $k_1 - k_3 < 0$. From (9.2.11) it follows that the value of C_o in the theorem is given by

$$C_o = \sqrt{\frac{k_3 - k_1}{k_2 - k_4}} \ .$$

q.e.d.

Remark 9.2.1: It is not difficult to establish that if at all there exists a Δ_2-best strategy in $L_1(H, \mu)$, then it should be of the form $H_o(d_o, \bar{e})$ considered in Theorem (9.2.1). Hence it follows from Theorem (9.2.2) that under condition (9.2.4) there does not exist a Δ_2-best strategy in $L_1(H, \mu)$ when μ is not an integer.

9.3 Estimation of variance of the HT-estimator

For any sampling design p for which $\pi_i(p) > 0$, $1 \leq i \leq N$, the variance of the HT-estimator $\bar{e}(s, y)$ is given by

$$V(\bar{e}(s,y)) = \sum_{i=1}^{N} \left(\frac{1}{\pi_i(p)} - 1\right) y_i^2 + \sum_{i \neq j}^{N}\sum \left(\frac{\pi_{ij}(p)}{\pi_i(p)\pi_j(p)} - 1\right) y_i y_j \ . \tag{9.3.1}$$

Two unbiased estimators have been proposed in the literature for $V(\bar{e}(s,y))$ in (9.3.1), namely

$$v_{HT} = \sum_{i \in s} \frac{1-\pi_i(p)}{\pi_i^2(p)} y_i^2 + \sum_{i \neq j \in s}\sum \frac{\pi_{ij}(p) - \pi_i(p)\pi_j(p)}{\pi_{ij}(p)} \cdot \frac{y_i}{\pi_i(p)} \frac{y_j}{\pi_j(p)} \quad (9.3.2)$$

$$v_{YG} = \frac{1}{2} \sum_{i \neq j \in s}\sum \frac{\pi_i(p)\pi_j(p) - \pi_{ij}(p)}{\pi_{ij}(p)} \left(\frac{y_i}{\pi_i(p)} - \frac{y_j}{\pi_j(p)}\right)^2. \quad (9.3.3)$$

The estimate v_{HT} has been proposed by Horvitz and Thompson [26] and v_{YG} by Yates and Grundy [66]. While v_{HT} can be used for any sampling design in which $V(\bar{e}(s,y))$ is estimable v_{YG} is applicable only for fixed sample-size designs in which $V(\bar{e}(s,y))$ is estimable, since it becomes biased for varying sample size designs. Yates and Grundy [66] have vigorously rejected v_{HT} in (9.3.2), preferring v_{YG} in (9.3.3) on the considerations of sampling fluctuations, mostly based on illustrative examples. From (9.3.3) it is obvious that a set of sufficient conditions for the estimator v_{YG} to be non-negative is that

$$\pi_{ij}(p) \leq \pi_i(p)\pi_j(p), \quad 1 \leq i \neq j \leq N. \quad (9.3.4)$$

Thus we have the following interesting problem:
Given a population

$$U = \{1, 2, \ldots, N\}$$

. .
.

of N units and numbers $\pi_1, \pi_2, \ldots, \pi_N$ such that

$$0 < \pi_i < 1, \quad 1 \leq i \leq N, \quad \sum_{i=1}^{N} \pi_i = \mu > 1 \qquad (9.3.5)$$

μ an integer, does there exist a sampling design p which satisfies

$$\left.\begin{array}{llll} \text{i)} & \pi_i(p) = \pi_i & & (1 \leq i \leq N) \\ \text{ii)} & n(s) = \mu & & \text{for all } s \in S \text{ such that } p(s) > 0 \\ \text{iii)} & \pi_{ij}(p) > 0 & & (1 \leq i \neq j \leq N) \\ \text{iv)} & \pi_{ij}(p) \leq \pi_i \pi_j & & (1 \leq i \neq j \leq N) . \end{array}\right\} \qquad (9.3.6)$$

The answer is known to be 'yes' and various authors have given methods of construction of designs satisfying all the above conditions. Hanurav [21] put forward some more conditions to be satisfied, namely,

v) $\dfrac{\pi_{ij}(p)}{\pi_i(p)\pi_j(p)}$ is not too small $\quad (1 \leq i \neq j \leq N)$

vi) $\pi_{ij}(p)$ is easily computable from a simple formula.

vii) The computations involved in the method are not too heavy.

We will not bother about conditions (v), (vi) and (vii) since they are highly subjective and are not relevant in a rigorous mathematical treatment. Below we modify Yates and Grundy variance estimator v_{YG} in (9.3.3) to suit varying sample size designs also, and after deriving sufficient conditions for its non-negativity, we prove the existence of designs satisfying those conditions in a special case.

9.3.1 Generalised Yates and Grundy variance estimator

For any design $d(S,p)$, direct calculation shows that $V(\bar{e}(s,y))$ in (9.3.1) can be written as

$$V(\bar{e}(s,y)) = \frac{1}{2} \sum_{i \neq j}^{N}\sum (\pi_i(p)\pi_j(p) - \pi_{ij}(p)) \left(\frac{y_i}{\pi_i(p)} - \frac{y_j}{\pi_j(p)}\right)^2$$

$$+ \sum_{i=1}^{N} \left[(1-f)\pi_i(p) + \sum_{j \neq i}^{N} \pi_{ij}(p) - n\pi_i(p)\right] \frac{y_i^2}{\pi_i^2(p)} \qquad (9.3.7)$$

where n is the largest integer not exceeding $\sum_{i=1}^{N} \pi_i(p)$ and f is the fractional part of $\sum_{i=1}^{N} \pi_i(p)$, namely $\sum_{1}^{N} \pi_i(p) - n$. Clearly the first component of $V(\bar{e}(s,y))$ in (9.3.7) is unbiasedly estimated by v_{YG} in (9.3.3). An unbiased estimator of the second component is given by

$$v_s = \sum_{i \in s} a_i(p) \frac{y_i^2}{\pi_i^3(p)} \qquad (9.3.8)$$

where

$$a_i(p) = (1 - f)\pi_i(p) + \sum_{j \neq i}^{N} \pi_{ij}(p) - n\pi_i(p). \quad (9.3.9)$$

The modified YG-variance estimator is given by

$$v_{YGR} = v_{YG} + v_s \quad (9.3.10)$$

is clearly unbiased for $V(\bar{e}(s,y))$ in (9.3.7). Also it is easily seen that a set of sufficient conditions for v_{YGR} to be non-negative is that

i) $\pi_{ij}(p) \leq \pi_i(p)\pi_j(p) \quad (1 \leq i \neq j \leq N)$ (9.3.11)

ii) $a_i(p) \geq 0 \quad (1 \leq i \leq N)$ (9.3.12)

where $a_i(p)$ is as defined in (9.3.9).

Remark 9.3.1: For a fixed sample size design, $f = 0$ and $a_i(p) = 0$ and consequently $v_s = 0$ so that

$$v_{YGR} = v_{YG}.$$

9.3.2 Special case

Now we consider only those designs, p, for which

$$V_p(n(s)) = f(1 - f) \quad (9.3.13)$$

i.e. $$n(s) = \begin{cases} n & \text{with probability } 1-f \\ n+1 & \text{with probability } f . \end{cases} \qquad (9.3.14)$$

Such designs have got the advantage that, under a linear cost function, the variation of cost among samples is a minimum and survey-practitioners consider random costs undesirable. Moreover, we have mentioned in the previous chapter that given any p satisfying (9.3.5) and (9.3.13), it is not possible to construct another design p' with the same expected sample size $\mu(p)$ for which the corresponding HT-estimator has uniformly smaller variance than that for p. For any design satisfying (9.3.13) it can be easily checked that

$$\sum_{j\neq i}^{N} \pi_{ij}(p) = n\pi_i(p) - \pi_{ni}(p) \qquad (9.3.15)$$

where $\pi_{ni}(p) = \sum_{s_n \supset i} p(s)$ and $\sum_{s_n \supset i}$ stands for summation over all samples of size n containing the i-th unit. From (9.3.9) and (9.3.15) we have

$$a_i(p) = (1-f)\pi_i(p) - \pi_{ni}(p). \qquad (9.3.16)$$

Hence the generalised YG-variance estimator becomes

$$v_{YGR} = v_{YG} + \sum_{i \in s} [(1-f)\pi_i(p) - \pi_{ni}(p)] \frac{y_i^2}{\pi_i^3(p)} . \qquad (9.3.17)$$

It follows from (9.3.11), (9.3.12) and (9.3.16) that a set of sufficient conditions for v_{YGR} in (9.3.17) to be non-negative is that

i) $\pi_{ij}(p) \leq \pi_i(p)\pi_j(p) \qquad (1 \leq i \neq j \leq N) \qquad (9.3.18)$

ii) $\pi_{ni}(p) \leq (1-f)\pi_i(p) \qquad (1 \leq i \leq N). \qquad (9.3.19)$

Now we have the generalised problem:

Given numbers π_i such that

$$0 < \pi_i < 1, \ 1 \leq i \leq N \ \text{and} \ \sum_{i=1}^{N} \pi_i = n+f > 1 \qquad (9.3.20)$$

where n and f are respectively the integral and fractional part of $\sum_{i=1}^{N} \pi_i$ does there exist a design $d(S,p)$ which satisfies

$$\left.\begin{array}{lll} \text{i)} & \pi_i(p) = \pi_i & (1 \leq i \leq N) \\ \text{ii)} & V_p(n(s)) = f(1-f) & \\ \text{iii)} & \pi_{ij}(p) > 0 & (1 \leq i \neq j \leq N) \\ \text{iv)} & \pi_{ij}(p) \leq \pi_i(p)\pi_j(p) & (1 \leq i \neq j \leq N) \\ \text{v)} & \pi_{ni}(p) \leq (1-f)\,\pi_i(p) & (1 \leq i \leq N). \end{array}\right\} \qquad (9.3.21)$$

We now answer the question in the affirmative in the following

Theorem 9.3.1: Given π_i's satisfying (9.3.20) there exists a design p which satisfies (9.3.21) and moreover

$$\pi_{ni}(p) > 0, \quad 1 \leq i \leq N. \tag{9.3.22}$$

Proof: We propose to prove the theorem using induction on N. For $N = 2$ the theorem is evident for consider the following design p:

sample s	probability p(s)
1	$1 - \pi_2$
2	$1 - \pi_1$
1,2	$\pi_1 + \pi_2 - 1$
	$\Sigma\, p(s) = 1$

It is easily verified that p as defined above satisfies all the conditions listed in (9.3.21) and also (9.3.22). Now assuming that the theorem is true for some N, we shall show that it is true for N+1. Without loss of generality we assume that

$$\pi_1 \leq \pi_2 \leq \cdots \leq \pi_{N+1}. \tag{9.3.23}$$

Let

$$\sum_{i=1}^{N+1} \pi_i = n^* + f^* \quad . \tag{9.3.24}$$

where n^* and f^* are respectively the integral and fractional parts of $\sum_1^{N+1} \pi_i$.

Case 1: $n^* = 1$ and $\pi_1 \geq f^*$

Since $\pi_1 \geq f^*$, from (9.3.23) it follows that

$$\pi_i \geq f^*, \quad 1 \leq i \leq N+1. \tag{9.3.25}$$

In this case we directly give a design p^* which satisfies all the required conditions. Consider the design p^* defined by

$$p^*(i) = \pi_i - \frac{2f^* \pi_i (1 + f^* - \pi_i)}{\sum_1^{N+1} \pi_i (1 + f^* - \pi_i)} \qquad (1 \leq i \leq N+1) \tag{9.3.26}$$

$$p^*(i,j) = \frac{2f^* \pi_i \pi_j}{\sum_1^{N+1} \pi_i (1+f^* - \pi_i)} \; . \qquad (1 \leq i > j \leq N+1) \tag{9.3.27}$$

Clearly

$$p^*(i, j) \geq 0$$

and $$\sum_{i>j}^{N+1}\sum p^*(i,j) = \frac{f^* \sum_{i=1}^{N+1} \pi_i(1+f^*-\pi_i)}{\sum_{i=1}^{N+1} \pi_i(1+f^*-\pi_i)} = f^* \quad . \qquad (9.3.28)$$

Now using (9.3.25) we have

$$\sum_{i=1}^{N+1} \pi_i(1+f^*-\pi_i) \geq f^* \sum_{i=1}^{N+1} (1+f^*-\pi_i)$$

$$= N f^*(1+f^*)$$

$$\geq 2f^*(1+f^*-\pi_i) \quad \text{since} \quad N \geq 2.$$

Hence $p^*(i) \geq 0$ for all i. Also

$$\sum_{i=1}^{N+1} p^*(i) = 1 + f^* - 2f^* = 1 - f^* \qquad (9.3.29)$$

so that from (9.3.28) and (9.3.29) it follows that

$\sum_s p^*(s) = \sum_1^{N+1} p^*(i) + \sum_{i>j}^{N+1}\sum p^*(i,j) = 1.$ Hence p^*, as defined above is a design. We now check that p^* satisfies all the required conditions listed in (9.3.21) and (9.3.22). Now

$$\pi_i(p^*) = p^*(i) + \sum_{j\neq i}^{N+1} p^*(i,j) = \pi_i,$$ using (9.3.26) and (9.3.27) so that p^* satisfies (i).

From (9.3.29) it is clear that

$V_{p*}(n(s)) = f*(1-f*)$. Hence condition (ii).

Condition (iii) is satisfied obviously since $\pi_{ij}(p*) = p*(i,j)$. To establish (iv) it is enough to show

$$p*(i,j) \leq \pi_i \pi_j$$

i.e.
$$\frac{2f* \pi_i \pi_j}{\Sigma \pi_i (1+f*-\pi_i)} \leq \pi_i \pi_j$$

which is true since $f* \leq \pi_i$, $1 \leq i \leq N+1$, and $N \geq 2$. It remains to check that $p*$ satisfies (v) and (9.3.22). Now

$$\pi_{1i}(p*) = p*(i) = \pi_i[1 - \frac{2f*(1+f*-\pi_i)}{\Sigma_j \pi_j (1+f*-\pi_j)}].$$

$$\therefore \quad \pi_{1i}(p*) \leq (1-f*)\pi_i$$

if and only if

$$\frac{2f*(1+f*-\pi_i)}{\Sigma \pi_j (1+f*-\pi_j)} \geq f*$$

i.e.
$$2(1+f*-\pi_i) \geq \Sigma \pi_j (1+f*-\pi_j)$$
$$= (1+f*)^2 - \Sigma \pi_j^2$$

i.e.
$$(1-\pi_i)^2 + \sum_{j \neq i} \pi_j^2 - f*^2 \geq 0$$

which is true since $\pi_j \geq f^*$, for all j. Condition (9.3.22) is obviously satisfied. Thus proof of case 1 is complete.

Case 2: $n^*=1$ and $\pi_1 < f^*$ or $n^* \geq 2$ and $\pi_2 \leq f^*$

In this case,

$$\sum_{i=2}^{N+1} \pi_i = n^* + (f^* - \pi_1) > 1 . \qquad (9.3.30)$$

Since the theorem is assumed to be true for N, there exists a design p corresponding to the population consisting of the N units $2, 3, \ldots, N+1$ which satisfies

$$\left.\begin{array}{lll}
\text{(a)} & \pi_i(p) = \pi_i & (2 \leq i \leq N+1) \\
\text{(b)} & V_p(n(s)) = f_1(1-f_1) \quad \text{where } f_1 = f^* - \pi_1 & \\
\text{(c)} & \pi_{ij}(p) > 0 & (2 \leq i \neq j \leq N+1) \\
\text{(d)} & \pi_{ij}(p) \leq \pi_i \pi_j & (2 \leq i \neq j \leq N+1) \\
\text{(e)} & \pi_{n^* i}(p) \leq (1-f_1)\pi_i & (2 \leq i \leq N+1) \\
\text{(f)} & \pi_{n^* i}(p) > 0 & (2 \leq i \leq N+1)
\end{array}\right\} \qquad (9.3.31)$$

It is clear from (9.3.31b) that p gives positive probability to samples of size n^* and n^*+1 only and

$$\left.\begin{array}{l}
p(n(s) = n^*) = 1 - f_1 = 1 - f^* + \pi_1 \\
p(n(s) = n^*+1) = f_1 = f^* - \pi_1 \; .
\end{array}\right\} \qquad (9.3.32)$$

Now construct the design p^* corresponding to the population of $(N+1)$ units as follows: To any sample s_{n^*} of size n^* for p, include unit 1 and put

$$p^*(s_{n^*}, 1) = \frac{\pi_1}{1 - f_1} \; p(s_{n^*}). \qquad (9.3.33)$$

Consider the sample s_{n^*} as a sample for p^*, with

$$p^*(s_{n^*}) = (1 - \frac{\pi_1}{1 - f_1}) \; p(s_{n^*}). \qquad (9.3.34)$$

Any sample s_{n^*+1} of size n^*+1 for p is retained as a sample for p^* with the same probability.

i.e. $$p^*(s_{n^*+1}) = p(s_{n^*+1}). \qquad (9.3.35)$$

Since $\pi_1 \leq 1 - f_1$ it is clear that p^* as defined above is a design for the population of $N+1$ units

i.e. $p^*(s) \geq 0$, for all s and $\sum_s p^*(s) = 1$.

Moreover, from the construction of p^*, we have

$$\pi_i(p^*) = \pi_i(p) \quad (2 \leq i \leq N+1) \qquad (9.3.36)$$

and $$\pi_{ij}(p^*) = \pi_{ij}(p) \quad (2 \leq i \neq j \leq N+1) \qquad (9.3.37)$$

Also $$\pi_1(p^*) = \pi_1 \tag{9.3.38}$$

$$\pi_{1i}(p^*) = \frac{\pi_1}{1-f_1}\ \pi_{n^*i}(p) \qquad (2 \le i \le N+1) \tag{9.3.39}$$

and
$$p^*(n(s) = n^*) = (1 - \frac{\pi_1}{1-f_1})(1-f_1) = 1-f^*. \tag{9.3.40}$$

We now check whether p^* satisfies all the required conditions. From (9.3.36) and (9.3.38) we see that p^* satisfies (i). Clearly p^* gives positive probability to samples of size n^* and n^*+1 only and hence from (9.3.40) it follows that condi--tion (ii) is satisfied. Using (9.3.31c), (9.3.31f), (9.3.37) and (9.3.39) it is easily seen that p^* satisfies (iii). Again using (9.3.31d), (9.3.36) and (9.3.37), to see whether p^* satisfies (iv) it is enough if we check

$$\pi_{1i}(p^*) \le \pi_1\pi_i \qquad (2 \le i \le N+1). \tag{9.3.41}$$

Using (9.3.39) in (9.3.41), inequality (9.3.41) becomes

$$\pi_{n^*i}(p) \le (1 - f_1)\pi_i \tag{9.3.42}$$

which is nothing but (9.3.31e) of the induction assumption.

Clearly

$$\pi_{n*1}(p^*) = 0 \qquad (9.3.43)$$

and $$\pi_{n*i}(p^*) = \left(1 - \frac{\pi_1}{1-f_1}\right)\pi_{n*i}(p) = \frac{(1-f^*)\pi_{n*i}(p)}{(1-f_1)} > 0,$$

$$2 \leq i \leq N+1. \qquad (9.3.44)$$

From (9.3.44) it immediately follows that p^* satisfies

(v) iff

$$\pi_{n*i}(p) \leq (1-f_1)\pi_i, \quad 2 \leq i \leq N+1$$

which is true by the induction assumption.

From (9.3.43) and (9.3.44) we see that p^* does not satisfy (9.3.22) since $\pi_{n*1}(p^*) = 0$. We get over this difficulty by the following modification. Since $\pi_2 \leq f^*$, in the same way as we constructed p^* corresponding to π_1, we can construct a design p_2^* for the population of $(N+1)$ units which satisfies all the required conditions except (9.3.22) as $\pi_{n*2}(p_2^*) = 0$. Now define a new design p_1^* as follows:

$$p_1^*(s) = \frac{p^*(s) + p_1^*(s)}{2}$$

It is easily checked that p_1^* satisfies all our required conditions. Thus proof of the theorem for case 2 is complete.

Case 3: $n^* \geq 2$ and $\pi_2 > f^*$

Now $\pi_1 + \sum_{i=3}^{N+1} \pi_i = n^* + f^* - \pi_2$

$$= (n^* - 1) + f_2$$

where $f_2 = 1 + f^* - \pi_2$.

Since the theorem is assumed to be true for N, there exists a design p_3 for the population of N units $1, 3, 4, \ldots, N+1$, satisfying

(a') $\pi_i(p_3) = \pi_i$ $(1 \leq i \neq 2 \leq N+1)$

(b') $V_{p_3}(n(s)) = f_2(1 - f_2)$

(c') $\pi_{ij}(p_3) > 0$ $(1 \leq i \neq j \neq 2 \leq N+1)$

(d') $\pi_{ij}(p_3) \leq \pi_i \pi_j$ $(1 \leq i \neq j \neq 2 \leq N+1)$

(e') $\pi_{(n^*-1)}(p_3) \leq (1 - f_2)\pi_i$ $(1 \leq i \neq 2 \leq N+1)$

(f') $\pi_{(n^*-1)i}(p_3) > 0$. $(1 \leq i \neq 2 \leq N+1)$

Construct the design p_3^* corresponding to the $(N+1)$ units as follows: To any sample s_{n^*-1} of size n^*-1 for p_3, include unit 2 and consider it as a sample for p_3^* with,

$$p_3^* \; (s_{n^*-1}, 2) = p_3(s_{n^*-1}).$$

To any sample s_{n*} of size $n*$ include unit 2 and put

$$p_3^*(s_{n*}, 2) = \frac{f^*}{f_2} p_3(s_{n*}).$$

Also retain the sample s_{n*} as a sample for p_3^* with

$$p_3^*(s_{n*}) = (1 - \frac{f^*}{f_2}) p_3(s_{n*}).$$

As in case 2, one can check that p_3^*, as defined above, satisfies all the required conditions listed in (9.3.21) and (9.3.22). We, however, omit the details here.

q.e.d.

We have proved only the existence of designs - except for case 1 - for which the generalised YG-variance estimator is non-negative. The actual construction of such designs, left as an open problem by the author, has satisfactorily been solved by Sankaranarayanan [56]. However, extensive empirical investigations are needed to judge the relative performance of the generalised YG-variance estimator over the HT-variance estimator.

=================

REFERENCES

[1] Aggarwal, O. P. (1959): Bayes and minimax procedures in sampling from finite and infinite populations-I, Ann. Math. Stat., 30, 206-218.

[2] Ajgaonkar, S. G. Prabhu (1965): On a class of linear estimators in sampling with varying probabilities without replacement, Jour. Amer. Stat. Assn., 60, 637-643.

[3] Barnard, G. A. (1963): The logic of least squares, Jour. Roy. Stat. Soc., B, 25, 124-127

[4] Basu, D. (1958): On sampling with and without replacement, Sankhyā, 20, 287-294.

[5] Basu, D. and Ghosh, J. K. (1967): Sufficient statistics in sampling from a finite universe, Proc. 36th Session of Int. Stat. Inst., 850-859.

[6] Basu, D. (1969): Role of the sufficiency and likelihood principles in sample survey theory, Sankhyā, A, 31, 441-453.

[7] Basu, D. (1970): An essay on the logical foundations of survey sampling, Presented at the Symposium on Foundations of Statistical Inference, University of Waterloo, Canada.

[8] Chaudhuri, A. (1968): A comparison between sampling with and without replacement from finite populations, Bull. Cal. Stat. Assn., 17, 137-160.

[9] Cochran, W. G. (1946): Relative accuracy of systematic and stratified random samples for a certain class of populations, Ann. Math. Stat., 17, 164-177.

[10] Cochran, W. G. (1953, 1963): Sampling Techniques, Wiley, New York, Ed. 1 and Ed. 2.

[11] Des Raj and Khamis, S. H. (1958): Some remarks on sampling with replacement, Ann. Math. Stat., 39, 550-557.

[12] Des Raj (1968): Sampling Theory, McGraw-Hill, London.

[13] Ghosh, S. P. (1963): Estimating the mean by two stage sampling with replacement, Bull., Cal. Stat. Assn., 12, 97-103.

[14] Godambe, V. P. (1955): A unified theory of sampling from finite populations, Jour. Roy. Stat. Soc., B, 17, 269-278.

[15] Godambe, V. P. (1960): An admissible estimator for any sampling design, Sankhyā, 22, 285-288.

[16] Godambe, V. P. (1965): A review of the contributions towards a unified theory of sampling from finite populations, Rev. Int. Stat. Inst., 33, 242-258.

[17] Godambe, V. P. and Joshi, V. M. (1965): Admissibility and Bayes estimation in sampling finite populations-I, Ann. Math. Stat., 36, 1707-1722.

[18] Godambe, V. P. (1966): A new approach to sampling from finite populations-I and II, Jour. Roy. Stat. Soc., B, 28, 310-328.

[19] Hajek, J. (1958): Some contributions to the theory of probability sampling, Bull. Int. Stat. Inst., 36(3), 127-134.

[20] Hajek, J. (1959): Optimum strategy and other problems in probability sampling, Casipis Pest. Math., 84, 387-423.

[21] Hanurav, T. V. (1965): Optimum sampling strategies, Ph.D. Thesis submitted to the Indian Statistical Institute.

[22] Hanurav, T. V. (1966): Some aspects of unified sampling theory, Sankhyā, A, 28, 175-204.

[23] Hanurav, T. V. (1968): Hyper-admissibility and optimum estimators for sampling finite populations, Ann. Math. Stat., 39, 621-642.

[24] Hege, V. (1965): Sampling design which admit uniformly minimum variance unbiased estimates, Bull. Cal. Stat. Assn., 14, 160-162.

[25] Hege, V. (1967): An optimum property of the Horvitz and Thompson estimate, Jour. Amer. Stat. Assn., 62, 1013-1017.

[26] Horvitz, D. G. and Thompson, D. J. (1952): A generalisation of sampling without replacement from a finite universe, Jour. Amer. Stat. Assn., 47 663-685.

[27] Jessen, R. J. (1942): Statistical investigation of a sample survey for obtaining farm facts, Iowa Agr. Exp. Sta. Res. Bull. 304.

[28] Joshi, V. M. (1965): Admissibility and Bayes' estimation in sampling finite populations-II, Ann. Math. Stat., 36, 1723-1729.

[29] Joshi, V. M. (1966): Admissibility and Bayes estimation in sampling finite populations-IV, Ann. Math. Stat., 37,1658-1670.

[30] Joshi, V. M. (1968): Distribution-free sufficiency in sampling finite populations, Jour. Roy. Stat. Soc., B, 30, 551-555.

[31] Kempthorne, O. (1969): Some remarks on statistical inference, New developments in survey sampling Univ. of North Carolina Symposium, 671-692.

[32] Mahalanobis, P. C. (1944): On large scale sample surveys, Roy. Soc. Phil. Trans., B, 231, 329-451.

[33] Midzuno, H. (1952): On the sampling system with probability proportional to the sum of sizes, Ann. Inst. Stat. Math., 3, 99-108.

[34] Murthy, M. N. (1957): Ordered and unordered estimators in sampling without replacement, Sankhyā, 18, 378-390.

[35] Murthy, M. N. (1967): Sampling Theory and Methods, Statistical Publishing Society, India.

[36] Murthy, M. N. and Singh, M. P. (1969): On the concepts of best and admissible estimators in sampling theory, Sankhyā, A, 31, 343-354.

[37] Pathak, P. K. (1962): On simple random sampling with replacement, Sankhyā, A, 24, 287-302.

[38] Pathak, P. K. (1962): On sampling with unequal probabilities, Sankhyā, 24, 315-326.

[39] Pathak, P. K. (1964): Sufficiency in sampling theory, Ann. Math. Stat., 35, 795-808.

[40] Pathak, P. K. (1964): On sampling schemes providing unbiased ratio estimators, Ann. Math. Stat., 35, 222-231.

[41] Pathak, P. K. (1965): Recent advances in sampling theory, Mimeograph, Pub. No.SM65-4, RTS, ISI.

[42] Ramakrishnan, M. K. (1970): Some results on the comparison of sampling with and without replacement, Sankhyā, A, 31, 333-342.

[43] Ramakrishnan, M. K. (1970): A note on the comparison of sampling with and without replacement, Sankhyā, A, 32, 233-235.

[44] Ramakrishnan, M. K. (1970): Some contributions to the theory of sampling with and without replacement. To appear in the Proceedings of the Second Mastech Conference, Madras.

[45] Ramakrishnan, M. K. (1970): Some results on hyper-admissibility, To be presented at the 58th Session of the Indian Science Congress, Bangalore.

[46] Rao, C. R. (1965): Linear statistical inference and its applications, John Wiley and Sons, New York.

[47] Rao, C. R. (1970): Some aspects of statistical inference in problems of sampling from finite populations, Presented at the Symposium on Foundations of Statistical Inference. University of Waterloo, Canada.

[48] Rao, J. N. K. (1966): On the comparison of sampling with and without replacement, Rev. Int. Stat. Inst., 34:2, 125-138.

[49] Rao, J. N. K. (1966): PPS sampling for multiple characteristics, Sankhyā, A, 28, 61-70.

[50] Rao, J.N.K. and Webster, J.T. (1966): On two methods of bias reduction in the estimation of ratios, Biometrika, 53, 571-577.

[51] Rao, J.N.K. and Singh, M.P. (1969): On the criteria of hyper-admissibility and necessary bestness in survey sampling, Tech. Report No. Math. Stat.40/69,ISI.

[52] Rao, S. R. S. (1968): On three procedures of sampling from finite populations, Biometrika, 55, 438-440.

[53] Rao, T. J. (1967): Contributions to the theory of sampling strategies, Ph.D. Thesis submitted to the Indian Statistical Institute.

[54] Roy, J. and Chakravorthy, I. M. (1960): Estimating the mean of a finite population, Ann. Math. Stat., 31, 392-398.

[55] Royal, R. M. (1970): Linear regression models in finite population sampling theory, Presented at the Symposium on Foundations of Statistical Inference, University of Waterloo, Canada.

[56] Sankaranarayanan, K.(1970): On a Generalised Sen-Yates-Grundy variance estimator, To be presented at the 58th Session of the Indian Science Congress, Bangalore.

[57] Searls, D. T. (1964): On the utilisation of a known coefficient of variation in the estimation procedure, Jour. Amer. Stat. Assn., 58, 1225-1226.

[58] Sen, A. R. (1952): Present status of probability sampling and its use in estimation of farm characteristics (abstract), Econometrica, 20, 103.

[59] Seth, G. R. and Rao, J.N.K. (1964): On the comparison of simple random sampling with and without replacement, Sankhyā, A, 26, 85-86.

[60] Singh, M. P. (1970): Some aspects of estimation in sampling from finite populations, Ph.D. Thesis submitted to the Indian Statistical Institute.

[61] Singh, M. P. and Roy, A. S. (1968): On the utilisation of an apriori value of the parameter in the estimation procedure, Tech. Paper No.30 NSS - ISI.

[62] Smith, F. H. (1938): An empirical law governing soil heterogeneity, Jour. Agr. Sci. 28. 1-23.

[63] Stevens, W. L. (1958): Sampling without replacement with probability proportional to size, Jour. Roy. Stat. Soc., B, 20, 393-397.

[64] Vijayan, K. (1966): Some contributions to sampling theory, Ph.D. Thesis submitted to the Indian Statistical Institute.

[65] Wald, A. (1950): Statistical decision functions, John Wiley and Sons, New York.

[66] Yates, F. and Grundy, P. M. (1953): Selection without replacement from within strata with probability proportional to size, Jour. Roy. Stat. Soc., B, 15, 253-261.

[67] Zacks, S. (1969): Bayes sequential designs of fixed-size samples from finite populations, Jour. Amer. Stat. Assn. 64, 1342 - 1349.
