OPTIMAL STRATEGIES UNDER SUPERPOPULATION MODELS

V.R. PADMAWAR

Thesis submitted to the
Indian Statistical Institute in partial fulfilment
of the requirements for the award of the degree of

DOCTOR OF PHILOSOPHY

CALCUTTA
1982

आईस

ACKNOWLEDGEMENTS

The work reported in this thesis was carried out under the supervision of Dr. T.J. Rao. I am thankful to him for his valuable guidance.

I have benefited immensely from innumerable discussions with Dr. Bikas Sinha. I am very grateful to him for his constructive help and encouragement.

I wish to thank Dr. Bikas Sinha and Dr. Parimal Mukhopadhyay for their kind permission to include in this thesis some of our joint work.

I express my thanks to Mr. J.K. Jain and Mr. K. Ganesan for their help in carrying out the programming part for some of the illustrations.

I am thankful to the authorities of the Indian Statistical Institute for the facilities extended to carry out my research.

I would be failing in my gratitude if I did not remember various friends at the I.S.I. for encouragement and help of various kinds, tangible and intangible (nevertheless concrete), academic and extra-academic, which made my stay in the I.S.I. a memorable experience.

Finally, I thank Mr. Dilip Kr. Bardhan for his excellent typing and Mr. Mukta Bag for efficient cyclostyling.

V.R. Padmawar

C O N T E N T S

*****

CHAPTER 0

INTRODUCTION

Though the concept of survey sampling is very old and has always been in vogue it is only during the thirties and forties that a more systematic development of the theory of sample surveys took place with the introduction of ideas like sampling without replacement, probability sampling and stratification. However, a large number of techniques developed and practised during this period had mostly either empirical or intuitive basis. It was much later that attention was paid to the purely theoretical aspect of the development of the survey sampling.

The main problem of sampling from finite population consists of devising an appropriate 'strategy' for estimating the population 'parameter' so as to maximize the 'precision' subject to certain 'cost' constraints or alternatively minimize the cost of survey for achieving a given level of precision.

Until recently, survey sampling and statistical inference theory were viewed as distinct fields. It was only the fifties that brought a change in this outlook. A clear formulation of the central problem of the theory of survey sampling is due to Godambe (1955). This brings forth similarity between the statistical inference theory and sampling theory. He also demonstrates the nonexistence of 'uniformly minimum variance linear unbiased estimator' for estimating the population mean.

Many traditional sampling tools had proven their value from a practical point of view, but had remained ad hoc procedures from a statistical inference point of view. The new trend in survey sampling has devised methods for evaluating the traditional techniques and proposed conditions for optimality of strategies.

Whenever an auxiliary information on a characteristic closely related to the study variate is available, it was first shown by Cochran (1946) that this information can be used to set up a criterion of optimality. This promising approach popularly known as 'superpopulation concept' is an avenue through which important new methods are currently being added to the survey samplers' traditional set of tools and it also contributes significantly towards the better understanding of various survey sampling problems. In recent times tremendous progress has been made in developing sampling theory and bridging the gap between statistical inference and survey sampling.

We shall now present a brief summary of the contents of various chapters of this thesis.

After this introduction we present, in Chapter 1, various definitions and explain basic concepts which are used in the sequel.

In Chapter 2 we work under a well-known superpopulation model. We first obtain an optimal design as well as model unbiased linear estimator for a given design for estimating the population mean in the sense of minimum expected variance. Imposing a condition on model parameters we then obtain an optimal design unbiased estimator for a given design for estimating the population mean again in the sense of minimum expected variance. We then note the limitations of celebrated Gπps strategy, an optimal design unbiased strategy for estimating population mean in the class of strategies of given average size in the sense of minimum average variance, from the realization point of view. We suggest an alternative criterion to obtain a reasonable strategy and show that the strategy so obtained is as good as any πps strategy. We take up comparison of various commonly used strategies in the last section of this chapter.

We work under a random permutation model in Chapter 3. We examine the estimation of symmetric parametric functions when the labels of distinguishable units are noninformative and the sampler has no knowledge, what so ever, of any relationship between the labels and the values associated with the units. The primitive strategies of simple random sampling, sample mean, sample variance and in general symmetric estimators are found to play important roles. We, in fact, obtain various optimality results giving subjective justification for using strategies which hitherto had only intuitive appeal.

In Chapter 4, we obtain various results for two stage population under two stage random permutation model which, essentially, are the analogues of their unistage counterparts.

We take up a study under a continuous survey sampling model in Chapter 5. The idea of continuous survey sampling is due to

Cassel and Särndal (1972). Such an interpretation makes it easier to grasp some of the complexities of modern survey sampling and exact efficiencies of various strategies can often be computed. We finally suggest a criterion to obtain a reasonable strategy and compare the strategy so obtained with some known strategies. We also consider stratification in the continuous set up. Evaluation of efficiency and comparison of different strategies are also taken up. We obtain some further results in Chapter 8.

Chapter 6 is devoted to the problem of estimating the population proportion. Lanke (1975) suggested a superpopulation model approach to utilize the auxiliary information more effectively for estimating the proportion. He compared a few strategies under the proposed model. Here we note that the model suggested by Lanke (1975) has many interesting features. In this chapter we consider model based inference as well as study design based inference. Under model based set up we first derive a few non-existence results and then obtain various estimators for proportion. In design based inference we obtain some optimality results in a reasonable sense.

In Chapter 7 we supplement Rao and Vijayan's (1977) attempt to solve the problem of estimating the variance of sampling strategies, for estimating population mean, nonnegatively. We mainly deal with the strategy that consists of Midzuno-Sen sampling design and ratio estimator. We extend the techniques of Rao and Vijayan of obtaining nonnegative unbiased variance estimators to more general strategies.

In Chapter 8 we have the same set up as that in Chapter 5. We pursue the study taken up in Chapter 5 and obtain optimal estimators, for given sampling design, in certain classes of estimators.

A list of references used in this thesis is given at the end. Each chapter has its own summary, somewhat elaborate, at the beginning. The contents of Section 2.4 are published (J. Roy. Statist. Soc., Ser. B, 43, 1981). The contents of Chapter 3 are to be published (Metrika, 30, 1983). First three sections of Chapter 2 were presented at the conference organized to honour Dr. C.R. Rao on his sixtieth birthday at the Indian Statistical Institute, Delhi in December 1980. The contents of Chapter 6 were presented at the conference 'Applications and New Directions', held in Calcutta during the golden jubilee celebrations of the Indian Statistical Institute in December 1981.

CHAPTER 1

CONCEPTS AND DEFINITIONS

In this chapter we give various definitions and explain some basic concepts that would be used in the sequel.

A finite population of size N $[<\infty]$ is a collection of N units. The size N is always assumed to be known. The units of a finite population are said to be identifiable or distinguishable if they can be labelled by integers $1, 2, \ldots, N$ and the label of each unit is known. We deal only with those populations whose units are identifiable. A label i is often used for a physically existing unit 'u_i', $1 \leq i \leq N$. After identifying unit u_i with label i, $1 \leq i \leq N$, a finite population is denoted by

$$U = \{1, 2, \ldots, N\} . \qquad \ldots(1.1)$$

Let y be a real valued variate defined on U, taking value y_i on unit i, $1 \leq i \leq N$. In the abstract sense symbol y_i denotes 'the value of y, the characteristic of interest, for the given unit i'. It is assumed that as soon as unit i is accessible we can measure y_i without any error. Concretely speaking y_i is just a real number which is a result of measuring unit i. When unit i has been measured the complete observation i.e. data, is recorded as (i, y_i). This enables us to introduce what is called parameter of the finite population. Let,

$$\underset{\sim}{y} = (y_1, y_2, \ldots, y_N) \qquad \ldots(1.2)$$

where y_i, $1 \leq i \leq N$, are unknown apriori, $\underset{\sim}{y}$ is called parameter of interest.

We assume that the ith component of $\underset{\sim}{y}$ is associated with unit i so that the observation (i, y_i) may be replaced by simply y_i. Thus if $\underset{\sim}{y}$ is known the components of $\underset{\sim}{y}$ are in effect labelled by their positions in the vector $\underset{\sim}{y}$. $\underset{\sim}{y}$ is assumed to be a point in R_N, the N-dimensional Euclidean space. Frequently the parameter space is R_N, however, other parametric spaces of practical importance (e.g. in Chapter 6, Y_i's are one-zero variates) are also considered.

Any real valued function of $\underset{\sim}{y}$ is called a parametric function. Inference in finite populations is usually about a parametric function and seldom about the parameter $\underset{\sim}{y}$ itself. Population mean is an important parametric function and is given by

$$\bar{Y} = \frac{1}{N} \sum_{i=1}^{N} y_i . \qquad \ldots(1.3)$$

The inference about the population mean is made on the basis of information obtained from a part of U. Let us first introduce some more concepts. Let

$$S = \{s\} \qquad \ldots(1.4)$$

be a collection of all possible subsets of U. S is called sample space and a typical member $s \in S$ is called a sample which is nothing but a subset of U. The number of units in s i.e. the cardinality of s, denoted by n(s), is called its sample size.

A real valued set function p on S such that

$$p : S \longrightarrow [0,1]$$
$$p(s) \geq 0, \quad \sum_{s \in S} p(s) = 1 \qquad \ldots(1.5)$$

is called a sampling design.

However, in practice, it is extremely cumbersome to list down all possible samples and then choose one at random with the probabilities prescribed by a sampling design. In this connection Hanurav (1962) demonstrated that every sampling design p can be implemented by some practically feasible 'unit drawing mechanisms'. For a given design p the first and second order inclusion probabilities are given by,

$$\pi_i(p) = \pi_i = \sum_{s \ni i} p(s) \qquad 1 \leq i \leq N$$
$$\pi_{ij}(p) = \pi_{ij} = \sum_{s \ni ij} p(s) \qquad 1 \leq i \neq j \leq N \qquad \ldots(1.6)$$

where $\sum_{s \ni i}$ denotes summation over all samples containing unit i and $\sum_{s \ni ij}$ denotes summation over all samples containing units i and j .

A sampling design p is said to be noninformative if p(s) does not depend on the y-values associated with the units having labels in s, $s \in S$. In this thesis we deal only with noninformative sampling designs.

A design p is said to be a fixed size design (FS(n)) if

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

Given a design p its average or expected size (not necessarily an integer) is given by

$$\sum_{s \in S} n(s)p(s) .$$

In traditional statistical inference there is typically a sample of n independent observations $x_1, x_2, \ldots, x_n$ on a random variable X with the hypothetical density function $f(x,\theta)$ which depends on the unknown parameter θ and the problem is often to estimate the parameter θ. Though the statistical inference and survey sampling inference are not opposing theories, because of the identifiability of units the latter has some interesting features of its own.

Let $\theta(\underset{\sim}{y}) = \theta$ be a generic parametric function. An estimator $t(s,\underset{\sim}{y})$ for estimating $\theta(\underset{\sim}{y})$ is a real valued function defined on $S \times R_N$ that depends on $\underset{\sim}{y}$ only through those coordinates y_i's for which $i \in s$. When there is no ambiguity we may denote $t(s,\underset{\sim}{y})$ by t_s or simply t.

A linear estimator is of the form

$$t(s, \underset{\sim}{y}) = b_s + \sum_{i \in s} b(s,i)\, y_i \qquad \ldots(1.7)$$

where b_s and $b(s,i)$, $i \in s$ are constants independent of y-values. When $b_s = 0$ (1.7) is called a linear homogeneous estimator.

A pair (p,t) where p is a design and t is an estimator is called a sampling strategy. The sampling design is used to select a sample $s \varepsilon S$ and based on the data $\{(i, y_i) : i \varepsilon s\}$ the estimator $t(s, \underset{\sim}{y})$ is used to estimate the parameter $\theta(\underset{\sim}{y})$. A strategy may be denoted by $H(p,t)$ or H itself.

An estimator t is said to be design unbiased (or p-unbiased for θ under a sampling design p if

$$E(p,t) = E_p(t) = \sum_{s \varepsilon S} p(s)t(s,\underset{\sim}{y}) = \theta(\underset{\sim}{y}) \quad \forall \underset{\sim}{y} \varepsilon R_N. \qquad ..(1.8)$$

If (1.8) holds then the strategy (p,t) is said to be p-unbiased, ($E_p(t)$ may be replaced by $E(t)$). If a strategy (p,t) does not satisfy (1.8) then it is said to be biased and its bias is given by

$$B(t) = E(t) - \theta(\underset{\sim}{y}).$$

As in the traditional decision theory we introduce loss function to evaluate the performances of various sampling strategies for estimating a given parametric function $\theta(\underset{\sim}{y})$. We shall consider loss functions $\ell(a,\theta)$ which are convex in a for every value of θ. A loss function which is used most often, squared error loss function, is given by

$$\ell(a,\theta) = (a-\theta)^2 \qquad ...(1.9)$$

$\ell(a,\theta)$ may be interpreted as loss incurred in estimating θ by a. Thus if we use a strategy (p,t) to estimate $\theta(\underset{\sim}{y})$, then for a sample s, the squared error loss is given by

$$[t(s, \underset{\sim}{y}) - \theta(\underset{\sim}{y})]^2.$$

Though we use general convex loss functions (Chapters 3 and 4) here we confine ourselves to squared error loss function given by (1.9). For a strategy (p,t) the <u>expected loss</u> or <u>risk</u>, called mean square error (MSE) is defined as

$$MSE(p,t,\theta(\underset{\sim}{y})) = MSE(p,t) = \sum_{s\varepsilon S} p(s)(t_s - \theta(\underset{\sim}{y}))^2 \qquad \ldots(1.10)$$

When (p,t) is unbiased for $\theta(\underset{\sim}{y})$ MSE is called variance and is denoted by $V(p,t) = V(t)$

$$\text{where} \quad V(t) = \sum_{s\varepsilon S} t_s^2\, p(s) - \theta^2(\underset{\sim}{y}) . \qquad \ldots(1.11)$$

Given a design p an estimator t_1 is said to be better than another estimator t_2 for estimating $\theta(\underset{\sim}{y})$ if

$$MSE(p,t_1,\theta(\underset{\sim}{y})) \leq MSE(p,t_2,\theta(\underset{\sim}{y})) \quad \forall\ \underset{\sim}{y} \varepsilon R_N \qquad \ldots(1.12)$$

and the strict inequality holds in (1.12) for at least one $\underset{\sim}{y} \varepsilon R_N$.

An estimator t^* belonging to a class C of estimators for estimating $\theta(\underset{\sim}{y})$ is said to be best in C if there exists no other estimator in C which is better than t^*.

A sampling strategy (p_1, t_1) is said to be better than another strategy (p_2, t_2) for estimating $\theta(\underset{\sim}{y})$ if

$$MSE(p_1, t_1, \theta(\underset{\sim}{y})) \leq MSE(p_2,t_2,\theta(\underset{\sim}{y})) \quad \forall\ \underset{\sim}{y} \varepsilon R_N \qquad \ldots(1.13)$$

and the strict inequality holds in (1.13) for at least one $\underset{\sim}{y}_o \varepsilon R_N$.

A strategy (p^*,t^*) is said to be best in a class $\mathcal{T}$ of strategies for estimating $\theta(\underset{\sim}{y})$ if there exists no other strategy in $\mathcal{T}$ which is better than (p^*,t^*).

Let us first get acquainted with some commonly used designs and estimators.

With units $1, 2, \ldots, N$ associate numbers $x_1, x_2, \ldots, x_N > 0$.

(a) A simple random sampling (srs) of size n is a design that gives equal probability of selection to all possible $\binom{N}{n}$ samples of size n i.e.

$$p(s) = 1/\binom{N}{n} \quad \text{if} \quad n(s) = n \qquad \ldots(1.14)$$
$$= 0 \quad \text{otherwise}.$$

(b) Midzuno-Sen sampling scheme (p_M) of size n is a design that selects samples of size n with probability proportional to the aggregate size $(\sum_{i \varepsilon s} x_i)$ i.e.

$$p(s) = \frac{x_s}{\binom{N-1}{n-1} X} \quad \text{if} \quad n(s) = n \qquad \ldots(1.15)$$
$$= 0 \quad \text{otherwise}$$

where $x_s = \sum_{i \varepsilon s} x_i$ and $X = \sum_{i=1}^{N} x_i$

(c) An inclusion probability proportional to size (πps) sampling design is one for which

$$\pi_i = \sum_{s \ni i} p(s) = \frac{n x_i}{X}, \; 1 \leq i \leq N. \qquad \ldots(1.16)$$

(d) A generalized πps $(G\pi ps)$ design of average size n for $g \geq 0$ is one with

$$\pi_i = \frac{n x_i^{g/2}}{\sum_{i=1}^{N} x_i^{g/2}} , \quad 1 \leq i \leq N$$

...(1.17)

and $p(s) > 0 \Longrightarrow \sum_{i\varepsilon s} x_i^{1-g/2} = nX \Big/ \sum_{i=1}^{N} x_i^{g/2}$.

Note that for $g = 2$ Gπps design reduces to a πps design of fixed size n. Further for a Gπps design to be feasible we must have $\max_{1 \leq i \leq N} x_i^{g/2} < \frac{1}{N} \sum_{i=1}^{N} x_i^{g/2}$.

Let us now consider some commonly used estimators

(a) Sample mean is given by

$$\bar{y} = \frac{1}{n(s)} \sum_{i\varepsilon s} y_i . \qquad ...(1.18)$$

(b) The conventional ratio estimator is given by

$$t_R = \frac{y_s}{x_s} \bar{X} \qquad ...(1.19)$$

where $y_s = \sum_{i\varepsilon s} y_i$ and $\bar{X} = \frac{X}{N} = \frac{1}{N} \sum_{i=1}^{N} x_i$.

(c) For a given design p with $\pi_i = \pi_i(p) > 0 \;\forall\; i = 1,2,...,N$ the well known Horvitz-Thompson estimator is given by

$$t_{HT} = \frac{1}{N} \sum_{i\varepsilon s} \frac{y_i}{\pi_i} . \qquad ...(1.20)$$

As a matter of course, attempts were made to obtain best estimators in different classes of estimators for estimating population mean $\bar{Y}$. However, Godambe (1955) proved a very powerful nonexistence result which was further strengthened by Godambe and Joshi (1965).

For a given design p, an estimator t_0 belonging to some class C of unbiased estimators of population mean is said to be uniformly minimum variance (UMV) estimator if and only if

$$V(p,t_0) \leq V(p,t) \quad \forall\, t \in C \quad \text{and} \quad \forall\, \underset{\sim}{y} \in R_N. \qquad \ldots(1.21)$$

We now have a result due to Godambe and Joshi (1965).

<u>Theorem 1.1</u> (Godambe and Joshi). For a design p there does not exist a UMV estimator in the class of all p-unbiased estimators for population mean $\bar{Y}$.

Godambe (1955) had proved the nonexistence of UMV estimator in the class of linear p-unbiased estimators. As a consequence of the nonexistence results, the interest was then directed towards optimality criteria other than UMV. The criterion of admissibility was introduced to weed out 'bad' estimators.

A strategy (p,t) belonging to a class $\mathcal{T}$ of strategies is said to be <u>admissible</u> in $\mathcal{T}$ if and only if there exists no other strategy in $\mathcal{T}$ which is better than (p,t).

Evidently enough, the criterion of admissibility is rather weak and is useful only in eliminating really bad estimators but is of little help towards obtaining optimal strategies. There are a few more criteria in the literature introduced in the recent past like hyperadmissibility (Hanurav, 1965), necessary bestness (Prabhu-Ajgaonkar, 1965). Both these criteria yield Horvitz-Thompson estimator as 'optimal' one. However, apparently, these criteria were designed with optimality of Horvitz-Thompson estimator in mind and consequently met with a lot of criticism.

A constructive step towards obtaining a reasonable criterion of optimality is due to Cochran (1946). Whenever some auxiliary information on a real valued variate x taking positive value x_i on unit i, $1 \leq i \leq N$, closely related to the study variate y is available, it is possible to use this information for setting up a criterion of optimality. Let us first introduce the concept of stochastic study variate. The vector $\underset{\sim}{y} = (y_1, y_2, \ldots, y_N)$ of population values is assumed to be a realization of a vector $\underset{\sim}{Y} = (Y_1, Y_2, \ldots, Y_N)$ of random variables. ξ, the joint distribution of $\underset{\sim}{Y}$, is assumed to depend on auxiliary values $x_1, x_2, \ldots, x_N$ through some unknown parameters.

This concept is called superpopulation concept. A 'superpopulation model' or simply a model means a specified set of conditions that defines a class of distributions to which ξ is assumed to belong. In its most pure form the idea of superpopulation is that the finite population is actually drawn from a larger universe of populations. In many situations, model summarizes and formalizes the model maker's prior knowledge about the population whether it be based on experience or personal subjective beliefs. A criterion of optimality now may be given as

$$E_\xi \, MSE(p, t) = E_\xi E_p (t - \bar{Y})^2 \qquad \ldots(1.22)$$

where E_ξ denotes expectation w.r.t. the joint distribution ξ.

After the introduction of this new stochastic element one may think of a distribution $p\xi$ which is a joint effect of a design and a model. And consequently it may be possible to set up yet another criterion of optimality based on $p\xi$ distribution as

$$E_p E_\xi (t - \bar{\mu})^2 \qquad \ldots(1.23)$$

where $\bar{\mu} = E_\xi(\bar{Y})$

Though (1.22) is used very widely, (1.23) is more natural than (1.22) whenever the interest lies in inferring about the superpopulation itself. In some practical situations $\bar{\mu}$ is indeed a natural target of inference. The choice for <u>measure of uncertainty</u> (optimality criterion) seems simple, once the objective of the estimation procedure is clearly understood. Though (1.22) is the most commonly used criterion of optimality the criterion (1.23) is also seen in various studies. A detailed discussion on these criteria can be found in Särndal (1980a). Cochran (1977) does not discuss the basis for his procedure to obtain 'reasonable' estimators but apparently he justifies the estimators which minimize $E_\xi(t - \bar{\mu})^2$. It should, however, be noted that when N is large there is little to discriminate between these two measures of uncertainty. In this thesis we make use of both the criteria.

If $M(p,t) = M$ denotes a measure of uncertainty (e.g. (1.22) or (1.23)) then we have :

A strategy (p,t) is said to be better than another strategy (p_1, t_1) for estimating the population mean w.r.t. the measure of uncertainty M if

$$M(p,t) < M(p_1, t_1)$$

and they are equally good if

$$M(p,t) = M(p_1, t_1) .$$

A strategy (p^*, t^*) is said to be a best strategy in a class $\mathcal{T}$ of strategies for estimating the population mean if there exists no other strategy which is better than (p^*, t^*).

While comparing different sampling strategies we must also consider the cost of implementing strategies (or else census is always an optimal thing to undertake). Let us consider the following cost function.

$$c_s = c_o + c_1 n(s) \qquad \ldots(1.24)$$

where c_s denotes cost of sample s of size $n(s)$ with c_o as overhead cost and c_1 is the cost per unit for collecting the data. The cost of a strategy $H(p,t)$ is the expected cost of design and is given by

$$c(H) = c_o + c_1 \sum_{s \in S} n(s) p(s) . \qquad \ldots(1.25)$$

Thus two strategies are equicost if and only if they have the same expected or average size. Therefore when we deal with efficiency evaluation and comparison of sampling strategies we, justifiably, restrict to equicost strategies. In Chapters 3 and 4, however, we also consider a different cost function.

We now present a few important superpopulation models that will be used in this thesis. As mentioned earlier a model is nothing but a set of specifications which determines a class of distributions to which ξ, the joint distribution of $\underset{\sim}{Y}$, is assumed to belong. The specifications may range from a crude formulation, prescribing, for instance a few moments of the distribution ξ to a very detailed description (and sometimes complete) of ξ. Consider a model specified by:

$$E_\xi(Y_i|x_i) = \beta x_i \qquad i=1,2,\ldots,N$$

$$V_\xi(Y_i|x_i) = E_\xi\left[(Y_i-\beta x_i)^2 \mid x_i\right] = \sigma^2 x_i^g \qquad \ldots(1.26)$$

$$E_\xi\left[(Y_i-\beta x_i)(Y_j-\beta x_j)|x_i, x_j\right] = \rho\sigma^2 x_i x_j \quad i \neq j = 1,2,\ldots,N$$

where $\sigma^2 > 0$, β and $-\frac{1}{N-1} \leq \rho \leq 1$ are unknown model parameters and $g \in [0,2]$ may be known or unknown.

Random variables $Y_1, Y_2, \ldots, Y_N$ are said to be exchangeably distributed if for every permutation $\pi_1, \pi_2, \ldots, \pi_N$ (π) of integers $1,2,\ldots,N$, $Y_{\pi_1}, Y_{\pi_2}, \ldots, Y_{\pi_N}$ have the same joint distribution.

We now define <u>random permutation model</u>. A class of distributions ξ such that for any fixed and unknown numbers $y_1, y_2, \ldots, y_N$ the random variables $Y_1, Y_2, \ldots, Y_N$ have an exchangeable distribution such that

$$\text{Prob}\left[Y_i = y_{\pi(i)}\ ;\ 1 \leq i \leq N\right] = \frac{1}{N!} \qquad \ldots(1.27)$$

for every permutation π of integers $1,2,\ldots,N$.

When random variables $Y_1, Y_2, \ldots, Y_N$ are one-zero variates then we can think of the following model (Lanke, 1975)

$$\text{Prob}\left[Y_i = 1 \mid x_i\right] = \beta x_i \qquad \ldots(1.28)$$

$$0 \leq \beta \leq \frac{1}{x_n}, \quad x_n = \max_{1 \leq i \leq N} x_i$$

where β is unknown parameter of the model (1.28). Note that (1.28) is a completely specified model.

We shall also consider a continuous survey sampling model. The idea of continuous survey sampling is due to Cassel and Särndal (1972) in that they try to adapt Godambe's survey sampling set up to continuous framework (See Section 5.1).

An estimator t is said to be model unbiased or ξ-unbiased for the population mean if and only if

$$E_\xi(t(s,\underset{\sim}{Y})) = E_\xi(\overline{Y}) \quad \forall\, s \text{ with } p(s) > 0. \qquad \ldots(1.29)$$

A strategy (p,t) is said to be model-design unbiased or $p\xi$-unbiased for the population mean if and only if

$$E_p E_\xi(t) = E_\xi(\overline{Y}). \qquad \ldots(1.30)$$

We finally give an optimality result due to Ramachandran (1973).

<u>Theorem 1.2</u> (Ramachandran) For estimating the population mean $\overline{Y}$, in the class of all p-unbiased strategies with given expected size, there exists an optimal strategy under the model (1.26) with $\rho = 0$ and $g \in [0,2]$ known in the sense of minimum expected variance (1.22)

The optimal strategy popularly known as $G\pi ps$ strategy consists of a $G\pi ps$ design and corresponding Horvitz-Thompson estimator. However, the above existence theorem is of little practical importance since construction of $G\pi ps$ designs, in mos practical situations, can safely be ruled out. In Chapter 2 we discuss an alternative criterion to obtain a 'reasonable' strateg under model (1.26).

CHAPTER 2

ESTIMATION UNDER REGRESSION MODEL

2.0 Summary

In this chapter we work under a commonly used superpopulation model. We first establish that for a given connected design there exists a best linear design as well as model unbiased estimator in the sense of minimum expected variance. We then propose a sufficient condition for the existence of a best linear design unbiased estimator, for a given design, again in the sense of minimum expected variance. Ramachandran (1978) obtained an optimal strategy, popularly known as $G\pi ps$ strategy, in the class of all design unbiased strategies of a given average size. However in most situations $G\pi ps$ designs cannot be realized. Here we suggest an alternative criterion to obtain a 'reasonable' strategy and show that such a strategy is as good as any πps strategy. We then demonstrate that using our alternative criterion we get the same optimal strategy as that due to Cassel et al (1976). Finally, in the last section, we take up the comparison of certain commonly used strategies under the proposed model.

2.1 Introduction

Godambe (1955) showed that for any sampling design p there does not exist a best estimator for the population mean $\bar{Y}$ in the class of linear p-unbiased estimators for the squared error loss function. Hanurav (1965) and Hege (1965) pointed out some non-trivial exceptions to this where a best estimator exists. Such designs were termed as unicluster designs by Hanurav. Godambe's nonexistence theorem was later on extended to the class of all p-unbiased estimators by Godambe and Joshi (1965). As mentioned before whenever auxiliary information on a (positive valued) characteristic x closely related to the characteristic y under study is available, the information can be used to set up a criterion of optimality for estimating $\bar{Y}$ as shown by Cochran (1946). Here $y_1, y_2, \ldots, y_N$ is a realization of random variables $Y_1, Y_2, \ldots, Y_N$ the joint distribution of which depends on the auxiliary values $x_1, x_2, \ldots, x_N$ and some unknown quantities called parameters. This is termed as superpopulation model ξ.

Superpopulation models have a long history in the sampling literature. Cochran (1946) was one of the foremost users of such ideas. In fact, he was the first to notice that such an idea was used by Laplace around 1800 A.D. in a sampling problem. The idea of superpopulation, in its most pure form, may be explained as the finite population is actually drawn from a bigger universe. In many situations it is natural to let a model summerize and formalise our prior knowledge about the population. Superpopulation models need not be Bayesian in the sense of expressing personal subjective belief. They can be as objective as some of the models used in classical statistical theory (Cassel et al 1977). A model essentially defines a class of distributions ξ.

The criterion of optimality suggested by Cochran (1946) is to minimize the expected variance, the expectation being taken under the distribution ξ. Most often in the literature unbiased strategies with minimum expected variance are investigated. The intuitive appeal of this optimality criterion was further strengthened by recourse to usual Chebychev's inequality by Godambe and Thompson (1973).

Let us consider one such superpopulation model. The joint distribution of $Y_1, Y_2, \ldots, Y_N$ is specified, though not completely, by first two moments as follows. Let E_ξ and V_ξ denote the expectation and variance w.r.t. the model.

$$
\begin{aligned}
E_\xi(Y_i | x_i) &= \beta x_i \\
V_\xi(Y_i | x_i) &= \sigma^2 x_i^g
\end{aligned}
\qquad i = 1, 2, \ldots, N \qquad \ldots(2.1.1)
$$

$$E_\xi \left[(Y_i - \beta x_i)(Y_j - \beta x_j) | x_i, x_j \right] = 0 \quad i \neq j = 1, 2, \ldots, N.$$

In the model (2.1.1) $\sigma^2 > 0$ and β are unknown parameters of the prior distribution ξ whereas $g \in [0, 2]$ may be known or unknown. In many practical situations g is found to lie between 1 and 2. Note that (2.1.1) is a particular case of (1.26) with $\rho = 0$.

If $n(s)$ denotes the number of units in a sample s and if the cost of drawing and inspecting sample s is assumed to be proportional to $n(s)$, the two strategies (p_1, t_1) and (p_2, t_2) will be equi-cost if they have the same expected or average sample size. Under this convention it is justified to compare strategies having a given average size.

Godanbe (1955) proved that the strategy $H_1 = (\pi ps, t_{HT})$ is the best in the class of all linear p-unbiased strategies of a given fixed sample size $(=n)$ for estimating the population mean $\bar{Y}$ in the sense of minimum expected variance under the model (2.1.1) with $g = 2$. πps is a design that gives inclusion probability $\pi_i = \frac{nx_i}{X}$ to unit i, $1 \leq i \leq N$ and $t_{HT} = \frac{1}{N} \sum_{i \varepsilon s} \frac{y_i}{\pi_i}$ is the corresponding Horvitz-Thompson estimator of the population mean. Here n is such that $\max_{1 \leq i \leq N} x_i \leq \frac{X}{n}$. Later Godanbe and Joshi (1965) proved the optimality of the above strategy in the class of all fixed size (n) p-unbiased strategies under the model (2.1.1) with $g = 2$ and an additional assumption of independence of $Y_1, Y_2, \ldots, Y_N$. In the literature there are many sampling procedures that result in the required inclusion probabilities π_i proportional to x_i, $1 \leq i \leq N$. (see Hanif and Brewer, 1980).

Hansen and Hurvitz (1943) demonstrated the profitability of πps designs and indicated methods to determine the probability of selection which minimize the variance of an estimator for a fixed cost. Hansen and Hurvitz (1949) also showed that sampling with probability proportional to the square root of size (x_i) is more efficient than πps sampling under certain conditions. Under the model (2.1.1) T.J. Rao (1971) studied the Horvitz-Thompson estimator of the population mean $\bar{Y}$ with designs wherein the inclusion probability π_i is proportional to the modified size x_i^{α}, $\alpha > 0$, $1 \leq i \leq N$.

<u>Definition 2.1.1</u> (T.J. Rao, 1972). A design p is called a generalized πps or $G\pi ps$ design if

$$\pi_i = \sum_{s \ni i} p(s) \; \alpha \; x_i^{g/2}$$

and $$p(s) > 0 \Longrightarrow \sum_{i \varepsilon s} x_i^{1-g/2} = \frac{nX}{\sum_{i=1}^{N} x_i^{g/2}}$$

where $g \geq 0$ and n, the expected size, is such that

$$\max_{1 \leq i \leq N} n x_i^{g/2} \leq \sum_{i=1}^{N} x_i^{g/2} .$$

For estimating the population mean $\bar{Y}$, T.J. Rao (1971) observes that the $G\pi ps$ strategy (consisting of a $G\pi ps$ design and the corresponding Horvitz-Thompson estimator) is better than the strategy $H_1 = (\pi ps, t_{HT})$ in the sense of smaller expected variance under the model (2.1.1) for all $g \varepsilon [1,2]$. Later Ramachandran (1978) proved the optimality of $G\pi ps$ strategy in the class of all p-unbiased strategies of given average size (=n say) in the sense of minimum expected variance under the model (2.1.1) with $g \varepsilon [0,2]$ known and the additional assumption of independence of $Y_1, Y_2, \ldots, Y_N$. As a consequence for $g = 2$ we get

Godambe and Joshi's (1965) result and T.J. Rao's (1971, 1972) results also follow immediately.

However, it should be noted that the whole idea of $G\pi ps$ design is quite artificial and one must recognize that the birth of $G\pi ps$ design is more on algebraic considerations rather than anything else. T.J. Rao (1971) and Ramachandran (1978) have given some examples of $G\pi ps$ designs, for $g = 1$, to show the nonvaccuousness of the concept. The present author too knows some more example of $G\pi ps$ designs in slightly relaxed set up (fixed size and $\forall\ g \geq 0$) but all these examples are specific to the (artificial) x-values under consideration and cannot convey anything positive regarding the practicability of $G\pi ps$ designs. The existence of $G\pi ps$ designs, in most practical situations, can safely be ruled out.

Let us first give the preliminaries and develop necessary concepts before outlining the results of this chapter.

Apart from model (2.1.1) we will also deal with the following 'transformation' model considered by Cassel, Särndal and Wretman (1976).

The joint distribution of $Y_1, Y_2, \ldots, Y_N$ is specified by

$$\begin{aligned} E_\xi(Y_i) &= \mu_i = \mu a_i + b_i \\ V_\xi(Y_i) &= \sigma_i^2 = \sigma^2 a_i^2 \end{aligned} \qquad i = 1, 2, \ldots, N \qquad \ldots(2.1.2)$$

$$E_\xi\left[(Y_i - \mu_i)(Y_j - \mu_j)\right] = \sigma^2 \rho a_i a_j \qquad i \neq j = 1, 2, \ldots, N$$

where a_i, b_i, $1 \leq i \leq N$, are known real numbers such that $a_i > 0$,

$1 \leq i \leq N$ and $\sum_{i=1}^{N} a_i = N$. μ, σ^2, ρ are unknown parameters such that $\sigma^2 > 0$ and $-\frac{1}{N-1} \leq \rho \leq 1$.

In this model the choice of a_i, b_i ; $1 \leq i \leq N$ is such that the model maker is ready to hypothesize that the transformed variables $\frac{Y_i - \mu_i}{a_i}$, $1 \leq i \leq N$, have common mean and variance.

We now define different types of unbiasedness.

<u>Definition 2.1.2.</u> A strategy (p,t) is said to be

(a) p-unbiased (design unbiased) if

$$E(p,t) = \sum_{s \in S} p(s) t(s, \underset{\sim}{y}) = \overline{Y} \quad \forall \; \underset{\sim}{y} \in R_N$$

(b) ξ-unbiased (model unbiased) if

$$E_\xi (t(s, \underset{\sim}{Y})) = E_\xi (\overline{Y}) \quad \forall \; s \text{ with } p(s) > 0$$

(c) $p\xi$-unbiased (model-design unbiased) if

$$E_p E_\xi (t - \overline{Y}) = 0 .$$

Next we define what is called a connected design.

<u>Definition 2.1.3.</u> (Patel and Dharmadhikari, 1977). A design p is said to be connected if for any two units $i \neq j \in U$ there exist units $i_1, i_2, \ldots, i_{n-1}$ and samples $s_1, s_2, \ldots, s_n \in S$ such that $p(s_r) > 0 \; \forall \; r = 1, 2, \ldots, n$ and $s_1 \ni i, i_1$; $s_2 \ni i_1, i_2, \ldots$ and $s_n \ni i_{n-1}, j$.

Note that most of the designs used in practice are connected designs.

Our main problem is to infer about the population mean $\bar{Y}$, the other aspect of inference being the estimation of model parameters μ and β themselves. To compare the performances of various sampling strategies we introduce the following two measures of uncertainty. For a strategy (p,t) let

$$M_1(p,t) = E_\xi E(t - \bar{Y})^2 \quad \text{...(2.1.3)}$$

and $$M_2(p,t) = E_\xi E(t - E_\xi \bar{Y})^2. \quad \text{...(2.1.4)}$$

Note that if (p,t) is p-unbiased then

$$M_1(p,t) = EV_\xi(t) + E(E_\xi(t-\bar{Y}))^2 - V_\xi(\bar{Y}) \quad \text{...(2.1.5)}$$

and if (p,t) is p as well as ξ-unbiased then

$$M_1(p,t) = EV_\xi(t) - V_\xi(\bar{Y}). \quad \text{...(2.1.6)}$$

Once the goal of the estimation procedure is clearly understood the choice of the measure of uncertainty seems rather simple. M_2 may be used in the practical situations which call for an inference about the superpopulation process itself whereas an inference about a specific outcome of the process motivates the use of M_1. Further for a p-unbiased strategy M_1 and M_2 are effectivel same. (They differ only by a quantity $V_\xi(\bar{Y})$), Särndal (1980a).

Comparisons of various sampling strategies under a superpopulation model w.r.t. a certain measure of uncertainty has been one of the main problems of interest to survey statisticians whenever a globally optimal strategy does not exist. Such investigations have been carried out by various authors in the literature. We will consider the following strategies in course of the discussions.

$H_1 = (\pi ps, t_{HT})$, $H_2 = (p_M, t_{HT})$, $H_3 = (p_{RHC}, t_{RHC})$, $H_4 = (G\pi ps, t_{HT})$

where H_1 is a πps strategy, H_2 consists of Midzuno-Sen sampling scheme and Horvitz-Thompson estimator H_3 is the well-known Rao-Hartley-Cochran (1962) strategy and H_4 is a $G\pi ps$ strategy.

In this chapter we first prove that for a given connected design there exists a best p- as well as ξ-unbiased linear estimator in the sense of minimum M_1 or M_2 under the model (2.1.1) with g known. Next we deal with the existence of an optimal p-unbiased linear estimator for a given design p under the model (2.1.1) when g known and the ratio σ^2/β^2 also known w.r.t. the measure of uncertainty M_1. We then suggest a criterion for obtaining a strategy as an alternative to $G\pi ps$ strategy. In the class of all linear p-unbiased strategies Cassel et al (1976) obtained an optimal strategy under the transformation model (2.1.2) in the sense of minimum M_1 or M_2. We obtain the same strategy by different arguments. Finally we compare the strategies H_1, H_2, H_3 and H_4 under the model (2.1.1) w.r.t. the measure of uncertainty M_1. We first show that the sufficient condition, for H_1 to be superior to H_2 due to Chaudhuri (1976), can never hold. We then suggest a sufficient condition for H_1 to be better than H_2 which is an improvement over the condition due to T.J. Rao (1967). Further we disprove the claim that H_3 is superior to H_2 for $g > 1$ by Ajgaonkar and Pedgaonkar and then give a sufficient condition for

H_3 to be better than H_2. We finally make a remark on a result due to Pedgaonkar and Ajgaonkar (1978).

2.2 Optimal Estimators

Most of the designs that are used in practice are connected. In this section we first investigate the following: Given a connected design p does there exist a best linear p as well as ξ-unbiased estimator of the population mean $\bar{Y}$ in the sense of minimum expected variance i.e. w.r.t. the measure of uncertainty M_1 of (2.1.3) under the model (2.1.1) with g known?

Let us first prove some lemmas. For a given sampling design p define the following $N \times N$ matrix A as

$$a_{ii} = \pi_i - \eta_i p_i \sum_{s \ni i} \frac{p(s)}{d(s)} \qquad i = 1,2,\ldots,N$$

$$\text{and} \quad a_{ij} = - \eta_i p_j \sum_{s \ni ij} \frac{p(s)}{d(s)} \qquad i \neq j = 1,2,\ldots,N \qquad \ldots(2.2.1)$$

where $\eta_i > 0$; $p_i > 0$; $1 \leq i \leq N$ and $d(s) = \sum_{j \in s} \eta_j p_j$.

<u>Lemma 2.2.1</u> (Patel and Dharmadhikari, 1977 ; 1978) If the sampling design p is connected then Rank (A) = N-1 where A is given by (2.2.1).

<u>Proof</u> Let $C = D_1 A D_2$

where $D_1 = \text{diag}(p_1, p_2, \ldots, p_N)$ and $D_2 = \text{diag}(\eta_1, \eta_2, \ldots, \eta_N)$ are the diagonal matrices.

The entries of the matrix C are given by

$$c_{ii} = \eta_i p_i \pi_i - \eta_i^2 p_i^2 \sum_{s \ni i} \frac{p(s)}{d(s)} \quad i = 1,2,\ldots,N$$

and $$c_{ij} = - \eta_i p_i \eta_j p_j \sum_{s \ni i,j} \frac{p(s)}{d(s)} \quad i \neq j = 1,2,\ldots,N.$$

Note that since $\eta_i p_i > 0 \;\forall\, i = 1,2,\ldots,N$ and $d(s) > 0 \;\forall\, s$ it follows that $c_{ij} \leq 0$ for $i \neq j = 1,2,\ldots,N$. Further it is easy to check that $\sum_{j=1}^{N} c_{ij} = 0 \;\forall\, i = 1,2,\ldots,N$ and finally that C is symmetric. We now show that Rank $(C) = N-1$.

Let $\underset{\sim}{e}$ be a column N-vector of 1's then $\sum_{j=1}^{N} c_{ij} = 0$ $\forall\, i = 1,2,\ldots N$ can be written as $C\underset{\sim}{e} = \underset{\sim}{0}$. Suppose now that $\underset{\sim}{z}$, some other N-vector, is such that $C\underset{\sim}{z} = \underset{\sim}{0}$. We show that $\underset{\sim}{z}$ is a multiple of $\underset{\sim}{e}$. This would establish that Rank $(C) = N-1$.

Define $M = \max_{1 \leq i \leq N} z_i$ and $m = \min_{1 \leq i \leq N} z_i$.

If possible, let $m < M$. Let $U_1 = \{j : z_j = M\}$. Then U_1 is a nonempty proper subset of $U = \{1,2,\ldots,N\}$. Since the design p is connected there exist an $i_0 \in U_1$ and a $j_0 \in U - U_1$ such that they belong to some sample s with $p(s) > 0$. For this pair i_0, j_0 we have $c_{i_0 j_0} < 0$.

Now $\quad j \in U_1 \implies z_j = M \implies c_{i_0 j} z_j = M c_{i_0 j}$

and $\quad j \in U - U_1 \implies z_j < M \implies c_{i_0 j} z_j \geq M c_{i_0 j}$

with strict inequality for $j = j_0$.

Consequently for $i = i_o$, $\sum_{j=1}^{N} c_{ij} z_j > M \sum_{j=1}^{N} c_{ij} = 0$.

Thus $C\underset{\sim}{z} \neq \underset{\sim}{0}$, which is a contradiction. This proves that $M = m$ i.e. $\underset{\sim}{z}$ is a scalar multiple of $\underset{\sim}{e}$. Thus Rank (C) = N-1. Finally, since D_1 and D_2 are full rank diagonal matrices we get

$$\text{Rank } (A) = \text{Rank } (C) = N-1 .$$

This completes the proof of the lemma.

We now prove the following.

Lemma 2.2.2. (Patel and Dharmadhikari, 1978). For a connected sampling design p the following system of equations is consistent

$$A\underset{\sim}{z} = \underset{\sim}{d} \qquad \ldots(2.2.2)$$

where A is given by (2.2.1) and $\underset{\sim}{d}$ is a column N-vector with

$$d_i = 1 - \eta_i \sum_{s \ni i} \frac{p(s)}{d(s)} . \qquad \ldots(2.2.3)$$

Proof. Let $\underset{\sim}{p} = (p_1, p_2, \ldots, p_N)$ be a row N-vector. Then $\underset{\sim}{p} A = \underset{\sim}{0}$. Since by our previous lemma Rank (A) = N - 1 the system (2.2.2) is consistent as soon as $\underset{\sim}{p}\,\underset{\sim}{d} = 0$ which is indeed true. Hence the lemma.

Using these two results we proceed to prove our main result.

Theorem 2.2.1 : For a given connected design p there exists a best linear p as well as ξ-unbiased estimator of the population mean w.r.t. the measure of uncertainty $M_1(p,t)$ of (2.1.3) under the model (2.1.1) with g known.

Proof : We would, in fact, establish the existence of a unique linear estimator, for a given connected design p, that minimizes $M_1(p,t)$ of (2.1.3) subject to the conditions of

(i) p-unbiasedness

and (ii) ξ-unbiasedness.

For a linear estimator $t(s,\underset{\sim}{y}) = \sum_{i \in s} b(s,i)y_i$, the condition of p-unbiasedness is equivalent to

$$\sum_{s \ni i} b(s,i)p(s) = \frac{1}{N} \quad \forall\, i = 1,2,\ldots,N \qquad \ldots(2.2.4)$$

and the condition of ξ-unbiasedness is equivalent to

$$\sum_{i \in s} b(s,i)x_i = \overline{X} \quad \forall\, s \text{ with } p(s) > 0\,. \qquad \ldots(2.2.5)$$

The attempt now is to minimize $M_1(p,t) = E_\xi E(t - \overline{Y})^2$ subject to the conditions (2.2.4) and (2.2.5) but because of (2.1.6) it is enough to minimize $EV_\xi(t)$ subject to (2.2.4) and (2.2.5) since $V_\xi(\overline{Y})$ does not depend on any particular estimator t.

Now $\quad EV_\xi(t) = \sigma^2 \sum_{i=1}^{N} x_i^g \sum_{s \ni i} b^2(s,i)p(s)\,.$

Let $\quad a(s,i) = Nb(s,i),\; p_i = \frac{x_i}{X},\; N^2 q_i = x_i^g\,.$

The problem now reduces to the following minimization problem.

$$\left.\begin{array}{l} \text{Minimize } \sum_{i=1}^{N} q_i \sum_{s \ni i} a^2(s,i)p(s) \\ \text{subject to (i) } \sum_{s \ni i} a(s,i)p(s) = 1 \quad i = 1,2,\ldots,N \\ \text{and (ii) } \sum_{i \in s} a(s,i)p_i = 1 \quad \forall\, s \text{ with } p(s) > 0. \end{array}\right\} \ldots(2.2.6)$$

We use Lagrangian multipliers technique to solve the minimization problem (2.2.6). Let

$$Q = \sum_{i=1}^{N} q_i \sum_{s \ni i} a^2(s,i)p(s) - 2\sum_{i=1}^{N} \lambda_i \left[\sum_{s \ni i} a(s,i)p(s) - 1\right]$$

$$- 2 \sum_{s\varepsilon S} \alpha_s \left[\sum_{i\varepsilon s} a(s,i)p_i - 1\right] \quad \ldots(2.2.7)$$

where λ_i, $1 \leq i \leq N$ and α_s, $s \varepsilon S$ with $p(s) > 0$ are Lagrangian multipliers.

Solving (2.2.6) is equivalent to minimizing Q unconditionally. Differentiating and simplifying one has to solve the following system of equations.

$$\left.\begin{aligned} q_i a(s,i)p(s) &= \lambda_i p(s) + \alpha_s p_i \\ \sum_{s \ni i} a(s,i)p(s) &= 1 \\ \sum_{i\varepsilon s} a(s,i)p_i &= 1 \end{aligned}\right\} \quad \begin{aligned} &1 \leq i \leq N \\ &i\varepsilon s,\ s\varepsilon S \text{ with } p(s) > 0. \end{aligned} \quad \ldots(2.2.8)$$

After a little algebra one gets

$$a(s,i) = z_i + \eta_i \left(1 - \sum_{j\varepsilon s} z_j p_j\right)/d(s) \quad \ldots(2.2.9)$$

where $\eta_i = p_i / q_i$ and $\underset{\sim}{z}$ is a solution to the system of equations (2.2.2) with $p_i = x_i / X$ and $\eta_i = p_i/q_i$.

But by Lemma 2.2.2, the system (2.2.2) is consistent as soon as p is a connected design. Further it is easy to check that the solution (2.2.9) does not depend on any particular solution $\underset{\sim}{z}$ of (2.2.2).

This proves the existence and uniqueness of the best linear p as well as g-unbiased estimator for a given connected design p.

Example 2.2.1.

Let p be any connected πps design of fixed size n. Fortunately, for a πps design, (2.2.2) can be solved explicitly, where again $p_i = x_i/X$ and $\eta_i = p_i / q_i = N^2 x_i^{1-g}/X$. A solution is given by $z_i = X / nx_i$, for ith entry of $A\underset{\sim}{z}$ is given by

$$1 - \frac{N^2 x_i^{1-g}}{nX} \sum_{s \ni i} \frac{p(s)}{d(s)} - \frac{N^2 x_i^{1-g}}{nX} \sum_{j \neq i}^{N} \sum_{s \ni i,j} \frac{p(s)}{d(s)}$$

$$= 1 - \frac{N^2 x_i^{1-g}}{nX} \sum_{s \ni i} \frac{p(s)}{d(s)} - \frac{N^2 x_i^{1-g}}{nX} \sum_{s \ni i} \frac{p(s)}{d(s)} (n-1)$$

$$= 1 - \frac{N^2 x_i^{1-g}}{X} \sum_{s \ni i} \frac{p(s)}{d(s)}$$

$$= d_i .$$

And thus the optimal estimator, as expected, is given by $\frac{X}{N} \sum_{i \in s} \frac{y_i}{nx_i}$.

The next problem that we investigate is the following : Given a design p does there exist a best inear p-unbiased estimator of the population mean $\overline{Y}$ under the model (2.1.1) with g known and the ratio $\sigma^2/\beta^2 = k$ (say) also known w.r.t. The measure of uncertainty M_1 of (2.1.3) ?

Let us first state a lemma without proof.

Lemma 2.2.3 An $N \times N$ matrix $C = (c_{ij})$ is nonsingular if

$$c_{ii} - \sum_{j \neq i}^{N} |c_{ij}| > 0 \quad \forall\, i = 1,2,\ldots,N .$$

We now have

Theorem 2.2.2 For a design p with $\pi_i(p) > 0$, $1 \leq i \leq N$, there exists a best linear p-unbiased estimator for the population mean $\overline{Y}$ under the model (2.1.1) with g and k known w.r.t. the measure of uncertainty M_1 provided the parameter k satisfies

$$k > k(p) \qquad \ldots(2.2.10)$$

where $k(p) = \inf Z$ and

$$Z = \left\{ \alpha : \pi_i = \sum_{s \ni i} p(s) > \left[\sum_{s \ni i} p(s) \frac{\sum_{j \varepsilon s} x_j^{1-g}}{\alpha + \sum_{j \varepsilon s} x_j^{2-g}} \right],\ 1 \leq i \leq N \right\}. \qquad \ldots(2.2.11)$$

Remark 2.2.1 Note that π_i, $1 \leq i \leq N$, are independent of α whereas the bracketed expression in the definition of Z in (2.2.11) is decreasing in α $\forall\ i = 1,2,\ldots,N$.

Proof of Theorem 2.2.2 A linear estimator is of the type $\sum_{i \varepsilon s} b(s,i) y_i$. It is easy to see that obtaining the required estimator is equivalent to solving the following minimization problem.

$$\text{Minimize} \quad k \sum_{i=1}^{N} x_i^g \sum_{s \ni i} b^2(s,i)p(s) + \sum_{s \varepsilon S} p(s) \Big(\sum_{i \varepsilon s} b(s,i) x_i \Big)^2 \qquad (2.2.12)$$

$$\text{Subject to} \quad \sum_{s \ni i} b(s,i)p(s) = \frac{1}{N}, \quad 1 \leq i \leq N .$$

Let
$$Q_1 = k \sum_{i=1}^{N} x_i^g \sum_{s \ni i} b^2(s,i)p(s) + k\sum_{s \in S} p(s)\Big(\sum_{i \in s} b(s,i)x_i\Big)^2 - 2\sum_{i=1}^{N} \lambda_i \sum_{s \ni i} b(s,i)p(s)$$

where λ_i , $1 \leq i \leq N$, are Lagrangian multipliers.

Now solving (2.2.12) is equivalent to minimizing Q_1 unconditionally. After differentiating and simplifying we get

$$b(s,i) = \frac{1}{kx_i^g}\left[\lambda_i - x_i \frac{\sum_{j \in s} \lambda_j x_j^{1-g}}{k + \sum_{j \in s} x_j^{2-g}}\right] \qquad \ldots(2.2.13)$$

where $\underset{\sim}{\lambda}$ is a solution to the following system of equations.

$$C\underset{\sim}{\lambda} = \underset{\sim}{\delta} \qquad \ldots(2.2.14)$$

with $c_{ii} = \pi_i - x_i^{2-g} \sum_{s \ni i} p(s)/e(s) \quad i = 1,2,\ldots,N$

$$c_{ij} = -x_i x_j^{1-g} \sum_{s \ni ij} p(s)/e(s) \quad i \neq j = 1,2,\ldots,N \qquad \ldots(2.2.15)$$

and $\delta_i = kx_i^g / N \qquad i = 1,2,\ldots,N$

further $e(s) = k + \sum_{i \in s} x_i^{2-g}$.

Let (2.2.10) be true. We then demonstrate the following,

$$c_{ii} > \sum_{j \neq i}^{N} |c_{ij}| \quad \forall \; i = 1,2,\ldots,N$$

where c_{ii} and c_{ij} are given by (2.2.15).

$$c_{ii} - \sum_{j \neq i}^{N} |c_{ij}| = \pi_i - x_i^{2-g} \sum_{s \ni i} \frac{p(s)}{e(s)} - x_i \sum_{j \neq i}^{N} x_j^{1-g} \sum_{s \ni i,j} \frac{p(s)}{e(s)}$$

$$= \pi_i - x_i^{2-g} \sum_{s \ni i} \frac{p(s)}{e(s)} - x_i \sum_{s \ni i} \frac{p(s)}{e(s)} \sum_{j \neq i \in s} x_j^{1-g}$$

$$= \pi_i - x_i \sum_{s \ni i} \frac{p(s)}{e(s)} \sum_{j \in s} x_j^{1-g}$$

$$> 0 \qquad \text{by (2.2.10).}$$

Thus, in view of Lemma 2.2.3 $\det C \neq 0$; i.e. the system of equations (2.2.14) is consistent and has a unique solution.

Therefore (2.2.13) gives the required optimal estimator. This completes the proof.

Remark 2.4.2 Observe that the left hand side of the equation $\det C = 0$ can be thought of as a polynomial in k of degree m (say). Therefore for a given design p there can be at most m values $k_1, k_2, \ldots, k_m$ of k for which the best linear p-unbiased estimator may not exist.

2.3. p-unbiasedness of an Optimal ξ-unbiased Estimator.

In Section 2.2 we minimized $EV_\xi(t)$ under the model (2.1.1) subject to p as well as ξ-unbiasedness for a given connected design p. Ramachandran (1978) obtained an optimal strategy that minimizes $M_1(p,t)$ in the class of all p-unbiased strategies of given expected size under the model (2.1.1) with g known and the additional assumption of independence of $Y_1, Y_2, \ldots, Y_N$. The optimal strategy, known as Gmps strategy, is, as mentioned before, usually difficult to realize. The existence of $G\pi ps$ designs, in most of the practical situations, can safely be ruled out. Ramachandran minimized $EV_{\xi}(t)$ subject to the conditions of p-unbiasedness and fixed average size and tried to impose the condition of ξ-unbiasedness on the estimator so obtained only to end up with $G\pi ps$ strategy. Here, as a compromise, an attempt is made to obtain a strategy first by minimizing $EV_{\xi}(t)$ subject to ξ-unbiasedness and then making the estimator so obtained p-unbiased. Further it is shown that the strategy so obtained is as good as the strategy $(\pi ps, t_{HT})$ w.r.t. the measure of uncertainty M_1. The approach adopted here can also be interpreted as follows: When the measure of uncertainty M_2 is appropriate we first minimize M_2 in the class of linear ξ-unbiased estimators and then try to obtain a design that makes the estimator so obtained p-unbiased. For large N the two measures of uncertainty are practically the same and in the continuous set up (Chapter 5) for the approach taken here, using M_1 is equivalent to using M_2.

We first state a result due to Farkas (vide Mangasarian, 1969, p.34).

<u>Theorem 2.3.1</u> (Farkas) Let B be any $n \times n$ matrix then

either $B\underset{\sim}{z} = \underset{\sim}{b}$, $\underset{\sim}{z} \geq \underset{\sim}{0}$ has a solution

or $B'\underset{\sim}{u} \leq \underset{\sim}{0}$, $\underset{\sim}{b}'\underset{\sim}{u} > 0$ has a solution.

Here 'or' is used in the exclusive sense.

We now prove a lemma that would be used in proving the main result of this section.

<u>Lemma 2.3.1</u>. Let $x_1, x_2, \ldots, x_N > 0$ be such that

$$x_n^{g-1} = \max_{1 \le i \le N} x_i^{g-1} \le \frac{1}{n} \sum_{i=1}^{N} x_i^{g-1}$$

Then if $u_1, u_2, \ldots, u_N, u_{N+1}$ are such that for every n distinct labels $1 \le i_1, i_2, \ldots, i_n \le N$

$$\sum_{j=1}^{n} u_{i_j} + u_{N+1} \sum_{j=1}^{n} x_{i_j}^{2-g} \le 0 \qquad \ldots(2.3.1)$$

$$\text{and} \quad \sum_{i=1}^{N} x_i^{g-1} u_i + u_{N+1} \sum_{i=1}^{N} x_i > 0 \qquad \ldots(2.3.2)$$

then for every $n-1$ distinct labels $1 \le i_1, i_2, \ldots, i_{n-1} \le N$

$$\sum_{j=1}^{n-1} u_{i_j} + u_{N+1} \sum_{j=1}^{n-1} x_{i_j}^{2-g} \le 0 .$$

<u>Proof</u>. Our x_i, $1 \le i \le N$, values are such that

$$x_n^{g-1} \le \frac{1}{n} \sum_{i=1}^{N} x_i^{g-1}$$

$$\Longrightarrow x_n^{g-1} \le \sum_{i=1}^{N} x_i^{g-1} - (n-1) x_n^{g-1}$$

$$\le \sum_{i=1}^{N} x_i^{g-1} - \sum_{j=1}^{n-1} x_{i_j}^{g-1} \quad \text{for any set of } (n-1) \text{ labels } 1 \le i_1, i_2, \ldots, i_{n-1} \le N,$$

$$= \sum_{j=n}^{N} x_{i_j}^{g-1} ,$$

Thus we have for any set of $N-n+1$ labels $1 \le i_n, i_{n+1}, \ldots, i_N \le N$

$$x_n^{g-1} \le \sum_{j=n}^{N} x_{i_j}^{g-1} . \qquad \ldots(2.3.3)$$

First observe that, because of (2.3.1) there can be at most (n-1) labels for which

$$u_i + u_{N+1} x_i^{2-g} > 0 . \qquad \ldots(2.3.4)$$

Without loss of generality we call them 1,2,...,(n-1). Again because of (2.3.1) and (2.3.4) for at most these (n-1) labels we can have

$$\lambda = \sum_{i=1}^{n-1} u_i \quad u_{N+1} \sum_{i=1}^{n-1} x_i^{2-g} > 0 .$$

Now from (2.3.1), we have

$$\lambda \leq - u_i - u_{N+1} x_i^{2-g} \quad \forall \; i \geq n .$$

$$\text{or} \quad \lambda \sum_{i=n}^{N} x_i^{g-1} \leq - \sum_{i=n}^{N} u_i x_i^{g-1} - u_{N+1} (X - \sum_{i=1}^{n-1} x_i) . \qquad \ldots(2.3.5)$$

Further from (2.3.2), we get

$$\sum_{i=1}^{n-1} u_i x_i^{g-1} + u_{N+1} (\sum_{i=1}^{n-1} x_i) > - \sum_{i=n}^{N} u_i x_i^{g-1} - u_{N+1} (X - \sum_{i=1}^{n-1} x_i) . \qquad \ldots(2.3.6)$$

Combining (2.3.5) and (2.3.6), we get

$$\lambda \sum_{i=n}^{N} x_i^{g-1} < \sum_{i=1}^{n-1} u_i x_i^{g-1} + u_{N+1} \sum_{i=1}^{n-1} x_i$$

$$\leq \sum_{i=1}^{n-1} u_i (\sum_{i=n}^{N} x_i^{g-1}) + U_{N+1} \sum_{i=1}^{n-1} x_i^{2-g} (\sum_{i=n}^{N} x_i^{g-1})$$

$$= \lambda \sum_{i=n}^{N} x_i^{g-1}$$

or $\lambda < \lambda$ which is a contradiction.

Hence $\lambda \leq 0$.

Thus for every set of $(n-1)$ distinct labels $1 \leq i_1, i_2, \ldots, i_{n-1} \leq N$

$$\sum_{j=1}^{n-1} u_{i_j} + u_{N+1} \sum_{j=1}^{n-1} x_{i_j}^{2-g} \leq 0 .$$

This proves the lemma.

Let us now try to obtain our compromise strategy. We first minimize $EV_\xi(t)$ subject to the condition of ξ-unbiasedness. Equivalently we solve the following minimization problem

$$\text{Minimize} \quad \sum_{i=1}^{N} x_i^g \sum_{s \ni i} b^2(s,i)p(s)$$

$$\text{subject to} \quad \sum_{i \varepsilon s} b(s,i)x_i = \bar{X} .$$

This readily yields the optimal solution as

$$b(s,i) = \bar{X} x_i^{1-g} \Big/ \sum_{i \varepsilon s} x_i^{2-g} . \qquad \ldots(2.3.7)$$

Thus $EV_\xi(t)$ can easily be minimized subject to the condition of ξ-unbiasedness. However, the main problem of interest is to obtain a design that makes the estimator with $b(s,i)$ in (2.3.7) p-unbiase

For $s \varepsilon S$, define $I_s(\cdot)$, the indicator function, as

$$I_s(i) = 1 \text{ if } i \varepsilon s$$
$$= 0 \text{ if } i \notin s .$$

For convenience let us denote the samples by $1, 2, \ldots, s, \ldots, M$. Also let $d(s) = \sum_{i \varepsilon s} x_i^{2-g}$.

Now the problem of finding a design that makes the estimator with $b(s,i)$ in (2.3.7) p-unbiased is equivalent to solving the following system of equations

$$B\underset{\sim}{z} = \underset{\sim}{b}$$
$$\underset{\sim}{z} \geq \underset{\sim}{0} \qquad \ldots(2.3.8)$$

where

$$B = \begin{bmatrix} \frac{I_1(1)}{d(1)} & \frac{I_2(1)}{d(2)} & \cdots & \cdots & \frac{I_M(1)}{d(M)} \\ \frac{I_1(2)}{d(1)} & \frac{I_2(2)}{d(2)} & \cdots & \cdots & \frac{I_M(2)}{d(M)} \\ \vdots & & & & \\ \frac{I_1(N)}{d(1)} & \frac{I_2(N)}{d(2)} & \cdots & \cdots & \frac{I_M(N)}{d(M)} \\ 1 & 1 & \cdots & \cdots & 1 \end{bmatrix}, \quad \underset{\sim}{z} = \begin{bmatrix} p(1) \\ p(2) \\ \vdots \\ \vdots \\ \vdots \\ p(M) \end{bmatrix} \text{ and } \underset{\sim}{b} = \frac{1}{X}\begin{bmatrix} x_1^{g-1} \\ x_2^{g-1} \\ \vdots \\ \vdots \\ \vdots \\ x_N^{g-1} \\ X \end{bmatrix} \qquad \ldots(2.3.9)$$

<u>Theorem 2.3.2</u>. Let $x_1, x_2, \ldots, x_N > 0$ be such that

$$x_n^{g-1} \leq \frac{1}{n} \sum_{i=1}^{N} x_i^{g-1} .$$

Then the system of equations (2.3.8) is consistent.

<u>Proof</u>. In view of Theorem 2.3.1 it is enough to show that $B'\underset{\sim}{u} \leq \underset{\sim}{0}$ and $\underset{\sim}{b}'\underset{\sim}{u} > 0$ is not consistent where B and $\underset{\sim}{b}$ are given by (2.3.9). If possible let there exist $u_1, u_2, \ldots, u_N, u_{N+1}$ such that $B\underset{\sim}{u} \leq \underset{\sim}{0}$ and $\underset{\sim}{b}'\underset{\sim}{u} > 0$.

$$\underset{\sim}{b}'\underset{\sim}{u} > 0 \iff \sum_{i=1}^{N} u_i x_i^{g-1} + u_{N+1} \sum_{i=1}^{N} x_i > 0 . \qquad \ldots(2.3.10)$$

Further we have, for any n distinct labels $i_1, i_2, \ldots, i_n$, because of $B'\underset{\sim}{u} \leq \underset{\sim}{0}$,

$$\sum_{j=1}^{n} u_{i_j} + u_{N+1} \sum_{j=1}^{n} x_{i_j}^{2-g} \leq 0 \qquad \ldots(2.3.11)$$

Now with the repeated application of Lemma 2.3.1, from (2.3.10) and (2.3.11) we have

$$u_i + u_{N+1}\, x_i^{2-g} \leq 0 \quad \forall\, i = 1, 2, \ldots, N$$

$$\Longrightarrow \sum_{i=1}^{N} u_i x_i^{g-1} + u_{N+1} \sum_{i=1}^{N} x_i \leq 0 .$$

This is a contradiction to (2.3.10). This proves that the system (2.3.8) is consistent.

This enables us to obtain a design p that makes the estimator (2.3.7) p-unbiased.

The proof of existence of a design that makes the estimator (2.3.7) p-unbiased, in fact, establishes the existence of a fixed size (n) design that satisfies (2.3.8). Let p_1 be any fixed size (n) design satisfying (2.3.8) and t_1 be the estimator give by (2.3.7). We now have the following theorem.

<u>Theorem 2.3.3</u>. The strategy (p_1, t_1) is unique upto design in the sense that $M_1(p_1, t_1)$, under the model (2.1.1), is same for all fixed size (n) designs p_1 satisfying (2.3.8). Further the strategy (p_1, t_1) is as good as the strategy $(\pi ps,\ t_{HT})$ under the model (2.1.1) w.r.t. the measure of uncertainty M_1.

Proof. It is enough to check that $EV_\xi(t)$ is same for all (p_1, t_1).

Now $EV_\xi(t) = \sigma^2 \sum_{i=1}^{N} x_i^g \sum_{s \ni i} b^2(s,i) p(s)$.

Hence for a strategy (p_1, t_1)

$$EV_\xi(t_1) = \sigma^2 \sum_{i=1}^{N} x_i^g \sum_{s \ni i} \bar{X}^2 x_i^{2-2g} p_1(s)/(d(s))^2$$

$$= \sigma^2 \bar{X}^2 \sum_{s \in S} p_1(s)/d(s)$$

where $d(s) = \sum_{i \in s} x_i^{2-g}$.

But note that (p_1, t_1) is p-unbiased, hence

$$\sum_{s \ni i} \bar{X} x_i^{1-g} p_1(s)/d(s) = \frac{1}{N} \quad \forall i = 1, 2, \ldots, N$$

$$\Rightarrow \sum_{s \ni i} p_1(s)/d(s) = x_i^{g-1}/X$$

$$\Rightarrow \sum_{i=1}^{N} \sum_{s \ni i} p_1(s)/d(s) = \sum_{i=1}^{N} x_i^{g-1}/X$$

or $\sum_{s \in S} p_1(s)/d(s) = \frac{1}{nX} \sum_{i=1}^{N} x_i^{g-1}$, p_1 being a fixed size (n) design.

which is independent of choice of p_1.

Thus $M_1(p_1, t_1)$ is same for all (p_1, t_1), p_1 being fixed size (n) design satisfying (2.3.8). Further the value of $EV_\xi(t)$ for the strategy $(\pi ps, t_{HT})$ is known to be

$$\frac{\sigma^2 \bar{X}}{nN} \sum_{i=1}^{N} x_i^{g-1} .$$

Hence. $M_1(p_1, t_1) = M_1(\pi ps, t_{HT})$.

This completes the proof.

Remark 2.3.1. As a consequence of our approach when $g = 1$ in the model (2.1.1) we get the strategy (p_M, t_R) where p_M is the Midzuno-Sen sampling design and $t_R = \bar{X} \sum_{i \epsilon s} y_i / \sum_{i \epsilon s} x_i$ is the conventional ratio estimator of the population mean.

For $g = 2$, we, of course, get the strategy $(\pi ps, t_{HT})$.

Remark 2.3.2. The approach we adopted is a kind of insurance against possible model break-downs. Since the strategy (p_1, t_1) is p-unbiased even if the model breaks down (p_1, t_1) remains at least pξ-unbiased. Thus the step of obtaining p_1 even after getting the best ξ-unbiased linear estimator is justified. However, it must be noted that the strategy (p_1, t_1) depends on the model parameter g which we assumed to be known. Thus when g is not known we cannot think of getting strategy (p_1, t_1). Theorem 2.3.3 gives a kind of robustness property of the strategy $(\pi ps. t_{HT})$ since it is independent of the model parameter g and is as good as the strategy (p_1, t_1) $\forall g \in [0, 2]$.

We now comment on an optimality result due to Cassel, Särndal and Wretman (1976). Cassel et al (1976) obtained the following optimality result.

Theorem 2.3.4. Under the model (2.1.2)

$$M_1(p, t) \geq M_1(p_o, t_{GDO})$$

$\forall (p, t)$ such that p is a fixed size (n) design and t is a linear p-unbiased estimator and the strategy (p_o, t_{GDO}) consists of a fixed size (n) design p_o that gives inclusion probability $\frac{na_i}{N}$ to unit i, $1 \leq i \leq N$ and the estimator

$$t_{GDO} = \sum_{i \varepsilon s} \frac{y_i - b_i}{na_i} + \frac{1}{N} \sum_{i=1}^{N} b_i$$

In a criticism of this result Smith (1976) questioned the logic behind using p-unbiasedness as a constraint. He further queried why an optimality criterion based on both design and model should be subject to the constraint based on the design alone?

Whether the inference should be based on just design or just model or both model and design has always been a controversial issue. The condition of p-unbiasedness is a safe-guard against model break downs. Here we try to give an alternative justification to the optimality result due to Cassel et al (Theorem 2.3.4). We make an attempt first to minimize $M_1(p,t)$ subject to ξ-unbiasedness in the class of linear estimators and then to obtain a design that makes the estimator so obtained p-unbiased.

<u>Theorem 2.3.5</u>. The strategy so obtained is same as the optimal strategy (p_o, t_{GDO}), due to Cassel et al, of the Theorem 2.3.4.

<u>Proof</u>. We omit the proof. The technique involved is same as in the earlier results of this section. The covariance terms in (2.1.2) do not cause any trouble.

2.4 Comparison of Strategies

As seen before the optimal $G\pi ps$ strategy is not usually available. Even the alternative criterion that we adopted in the previous section does not yield any globally optimal strategy unless $g = 2$ in the model (2.1.1). In view of this it is in order to compare few well-known strategies used for estimating the population mean. Comparison of sampling strategies under the super population set up w.r.t. certain measure of uncertainty, as mentioned earlier, has been one of the main problems of interest to survey statisticians whenever a globally optimal strategy is not available. Such investigations have been carried out by various authors in the literature. In this section we compare the strategies H_1 ; H_2 ; H_3 ; H_4, explained in Section 2.1, under the model (2.1.1) w.r.t. the measure of uncertainty M_1.

Let p be any design with $\sum_{s \ni i} p(s) = \alpha_i > 0$, $1 \le i \le N$, and $\alpha_{ij} = \sum_{s \ni ij} p(s)$ $i \ne j = 1, 2, \ldots, N$. Then the sampling variance of the corresponding Horvitz-Thompson estimator is given by

$$V(t_{HT}) = \frac{1}{N^2}\left[\Sigma\, y_i^2 \left(\frac{1}{\alpha_i} - 1\right) + \sum_{i \ne j = 1}^{N} y_i y_j \left(\frac{\alpha_{ij}}{\alpha_i \alpha_j} - 1\right)\right]. \quad \ldots(2.4.2)$$

<u>In this section Σ will denote</u> $\sum_{i=1}^{N}$, <u>unless otherwise specified</u>.

The sampling variance of the Rao-Hartley-Cochran strategy, H_3, is given by

$$V(t_{RHC}) = \frac{1}{nN^2}\,\frac{N-n}{N-1}\left[\Sigma\, x_i\, \Sigma\, \frac{y_i^2}{x_i} - (\Sigma\, y_i)^2\right] \quad \ldots(2.4.2)$$

where $\frac{N}{n}$ is an integer.

It is well-known that under the model (2.1.1)

$$E_1 = M_1(H_1) = \frac{\sigma^2}{N^2} \Sigma x_i^g \left(\frac{1}{\pi_i} - 1\right) \quad \text{(T.J. Rao, 1967)} \quad ..(2.4.3)$$

where $\pi_i = \frac{nx_i}{X}$ $1 \leq i \leq N$.

$$E_2 = M_1(H_2) = \frac{\sigma^2}{N^2} \Sigma x_i^g \left(\frac{1}{\pi_i'} - 1\right) + \frac{\beta^2}{N^2} V\left(\sum_{i \epsilon s} \frac{x_i}{\pi_i'}\right) \quad \text{(T.J. Rao, 1967)} \quad ...(2.4.4)$$

where $\pi_i' = \frac{n-1}{N-1} + \frac{N-n}{N-1} \frac{x_i}{X}$ and $\frac{1}{N^2} V\left(\sum_{i \epsilon s} \frac{x_i}{\pi_i'}\right)$ is the sampling variance of the strategy H_2 at $\underline{x} = (x_1, x_2, \ldots, x_N)$. Finally,

$$E_3 = M_1(H_3) = \frac{N-n}{N-1} \frac{\sigma^2}{nN^2} [X \Sigma x_i^{g-1} - \Sigma x_i^g] \quad \text{(Hanurav, 1965)} \quad ...(2.4.5)$$

where n is such that $\frac{N}{n}$ is an integer.

(a) <u>Comparison between H_1 and H_2</u>

Here we first show that the sufficient condition for H_1 to be superior to H_2, obtained by Chaudhuri (1976), can never be satisfied. We then suggest a sufficient condition which is an improvement over the condition due to T.J. Rao (1967).

<u>Theorem 2.4.1</u>. The condition $D = \Sigma\left(\frac{1}{\pi_i'} - \frac{1}{\pi_i}\right) > 0$ due to Chaudhuri (1976) can never hold. π_i and π_i' are given by (2.4.3) and (2.4.4) respectively, $1 \leq i \leq N$.

<u>Proof</u>. $D = \Sigma \left[1 / \left(\frac{n-1}{N-1} + \frac{N-n}{N-1} \frac{x_i}{X}\right) - 1 / \left(\frac{nx_i}{X}\right)\right] = \frac{n-1}{n(N-n)} \Sigma \frac{Np_i - 1}{p_i \lambda_i}$

where $p_i = x_i / X$ and $\lambda_i = p_i + \frac{n-1}{N-n}$.

Hence $D = \frac{(n-1)N^2}{n(N-n)} \operatorname{Cov}(p_i, 1/p_i\lambda_i)$. Now $p_i\lambda_i$ is increasing in p_i

Hence $\operatorname{Cov}(p_i, 1/p_i\lambda_i) \leq 0$ or $D \leq 0$.

Thus the condition $D > 0$ can never be satisfied.

<u>Remark 2.4.1</u>. In the numerical example considered by Choudhuri to show that D can be positive, we observe that D is, in fact, negative.

For $N = 4$, $n = 2$, $p_1 = p_2 = 0.2$ and $p_3 = p_4 = 0.3$, we have $D = -25/84$.

We now prove the following theorem.

<u>Theorem 2.4.2</u>. A sufficient condition for H_1 to be superior to H_2 under the model (2.1.1), for a given value of $g \geq 1$ is given by

$$n \geq n_o = \frac{N(2-g)p + (g-1)}{(2-g)p + (g-1)} \qquad \ldots(2.4.6)$$

where $p = \max\limits_{1 \leq i \leq N} p_i$.

<u>Proof</u>. Following T.J. Rao (1967) we have

$$\frac{E_2 - E_1}{\sigma^2} \geq \frac{n-1}{n(N-n)} X^g \operatorname{Cov}(p_i, p_i^{g-1}/\lambda_i).$$

Note that p_i^{g-1}/λ_i is nondecreasing in p_i for $g = 2$ and for $1 \leq g < 2$ it is nondecreasing if and only if

$$p_i \leq \frac{g-1}{2-g} \frac{n-1}{N-n}.$$

Thus $\operatorname{Cov}(p_i, p_i^{g-1}/\lambda_i)$ is non-negative for $g = 2$ and for $1 \leq g < 2$

it is non-negative if

$$p \leq \frac{g-1}{2-g}\,\frac{n-1}{N-n} \qquad \ldots(2.4.7)$$

It is a matter of verification that (2.4.6) is equivalent to (2.4.7) Hence under (2.4.7) $E_2 - E_1 \geq 0$, i.e. H_1 is superior to H_2. This proves the theorem.

<u>Remark 2.4.2.</u> The condition (2.4.6) is an improvement over the sufficient condition due to T.J. Rao (1967) in that n_o in (2.4.6) is smaller than that of T.J. Rao whenever $n_o p \leq 1$. This is because Rao's sufficient condition is based on the fact that np_i must be ≤ 1 for πps sampling to be possible. He requires right hand side of (2.4.7) to be greater than $\frac{1}{n}$ whereas if we know that $p \leq \frac{1}{n}$ an improved sufficient condition in (2.4.6) is obtained.

<u>Remark 2.4.3.</u> For a given value of n such that $np \leq 1$, the strategy H_1 is superior to H_2 if the parameter g of the model (2.1.1) satisfies the following condition

$$g \geq g_o = 1 + p\Big/\left(p + \frac{n-1}{N-n}\right). \qquad \ldots(2.4.8)$$

Again (2.4.8) is an improvement over the corresponding sufficient condition due to T.J. Rao (1967).

Table 2.4.1 presents the minimum sample size, $[n_o] + 1$, given by the condition (2.4.6). It also gives the corresponding values of $[n_o] + 1$ due to T.J. Rao, where $[n_o]$ is the largest integer strictly smaller than n_o.

Table 2.4.1

Giving minimum sample size for H_1 to be superior to H_2

N	10				50				
g \ p	0.125	0.2	0.25	( )	0.025	0.05	0.1	0.125	( )
1.1	6	*	*	(7)	10	17	*	*	(18)
1.2	4	5	*	(5)	6	10	*	*	(13)
1.3	4	4	*	(5)	4	7	*	*	(11)
1.4	3	4	4	(4)	3	5	8	*	(9)
1.5	2	3	3	(4)	3	4	6	7	(8)
1.6	2	3	3	(3)	2	3	5	5	(6)
1.7	2	2	2	(3)	2	3	4	4	(5)
1.8	2	2	2	(2)	2	2	3	3	(4)
1.9	2	2	2	(2)	2	2	2	2	(3)

N	100					200				
g \ p	.0125	.025	.05	.1	( )	.00625	0125	.025	.05	( )
1.1	12	20	*	*	(27)	12	22	38	*	(39)
1.2	6	10	18	*	(19)	6	11	19	*	(27)
1.3	4	7	12	*	(15)	4	7	12	*	(21)
1.4	3	5	8	*	(12)	3	5	9	15	(18)
1.5	3	4	6	10	(10)	3	4	6	11	(15)
1.6	2	3	5	8	(9)	2	3	5	8	(12)
1.7	2	3	4	6	(7)	2	3	4	6	(10)
1.8	2	2	3	4	(6)	2	2	3	4	(8)
1.9	2	2	2	3	(4)	2	2	2	3	(6)

(i) Figures in parantheses are the corresponding values of $[n_o]+1$ due to Rao. They are useful only when $np \leq 1$.

(ii) * : Sufficient condition cannot be satisfied since $np > 1$.

(b) Comparison between H_2 and H_3

Chaudhuri (1976) claimed the following. If $p < \frac{2}{N}$, $\frac{N}{n}$ is an integer and $\Sigma(\frac{1}{\pi_i'} - \frac{1}{\pi_i}) > 0$ then

$$E_1 < E_3 < E_2 \quad \text{if} \quad g > 1 .$$

It should be noted that since $\Sigma(\frac{1}{\pi_i'} - \frac{1}{\pi_i}) \leq 0$ the claim is irrelevant. However the claim $E_3 < E_2$ whenever $g > 1$ (Ajgaonkar and Pedgaonkar (see footnote, Caudhuri, 1976, p.124)) can be shown to be incorrect by simple counter examples.

We shall now give a sufficient condition for H_3 to be superior to H_2. We first have the following lemma.

Lemma 2.4.1. The function $f(p) = p^{g-1}(1-p)/(p+\theta)$ where $0 < p < 1$, $\theta = (n-1)/(N-n)$ and $g > 1$, is non-decreasing in p if and only if

$$n \geq n_1 = N - (N-1)(g-1-gp)/(g-1)(1-p)^2. \qquad \ldots(2.4.9)$$

Proof. Differentiating $f(p)$ we get

$$\partial f/\partial p = - p^{g-2}(g-1)(p^2+bp-\theta)/(p+\theta)^2$$

where $b = (2+g(\theta-1))/(g-1)$.

Hence $\partial f/\partial p \geq 0$ if and only if $p^2+bp-\theta \leq 0$.

Now $p^2+bp-\theta \leq 0$

$$\iff [p^2(g-1) + (2-g)p] \leq [(g-1) - gp] \frac{n-1}{N-n}$$

$$\iff n \geq n_1 = N - (N-1)(g-1-gp)/(g-1)(1-p)^2 .$$

Thus f is nondecreasing in p if and only if (2.4.9) holds.

We now prove the following theorem

<u>Theorem 2.4.3</u>. Let $\frac{N}{n}$ be an integer. Then a sufficient condition for H_3 to be superior to H_2 under the model (2.1.1) when $g > 1$ is given by (2.4.9) where $p = \max\limits_{1 \leq i \leq N} p_i$.

<u>Proof</u>. Let $\lambda_i = p_i + \frac{n-1}{N-n}$ and $p_i' = \frac{N-n}{n(N-1)} \lambda_i$ then $\pi_i' = np_i'$. Now

$$D_1 = \frac{E_2 - E_3}{\sigma^2} N^2$$

$$\geq \Sigma \frac{x_i^g}{np_i'} - \Sigma x_i^g - \frac{N-n}{n(N-1)} \left[\Sigma x_i \Sigma x_i^{g-1} - \Sigma x_i^g \right]$$

$$= \frac{1}{n(N-1)} \left[(N-n)(N \Sigma x_i^g - \Sigma x_i \Sigma x_i^{g-1}) - (N-1)(N \Sigma x_i^g - \Sigma x_i^g / p_i') \right]$$

$$= \frac{N^2 X^g}{n(N-1)} \left[(N-n) \mathrm{Cov}(p_i, p_i^{g-1}) - (N-1) \mathrm{Cov}(p_i, p_i^g / \lambda_i) \right]$$

$$= \frac{N^2 X^g}{n(N-1)} \mathrm{Cov}(p_i, (1-p_i) p_i^{g-1} / \lambda_i).$$

Now using Lemma 2.4.1, under (2.4.9) $D_1 \geq 0$, i.e. H_3 is superior to H_2.

This proves the theorem.

Table 2.4.2 presents the values of $[n_1] + 1$, where $[n_1]$ is the largest integer strictly smaller than n_1.

Table 2.4.2

Giving the values of $[n_1] + 1$

N	10			50			
ε \ p	0.125	0.2	0.25	0.025	0.05	0.1	0.125
1.1	*	*	*	13	26	*	*
1.2	8	*	*	7	12	26	34
1.3	5	9	*	5	8	16	21
1.4	4	6	8	3	6	11	14
1.5	3	5	6	3	4	8	10
1.6	3	4	5	2	3	6	8
1.7	2	3	4	2	3	5	6
1.8	2	3	3	2	2	4	4
1.9	2	2	3	2	2	3	3

N	100				200			
ε \ p	.0125	.025	.05	.1	.00625	.0125	.025	.05
1.1	13	24	51	*	13	24	49	*
1.2	7	12	24	52	7	12	23	46
1.3	4	8	15	31	4	7	14	28
1.4	3	5	10	21	3	5	9	19
1.5	3	4	7	15	3	4	7	13
1.6	2	3	5	11	2	3	5	9
1.7	2	3	4	8	2	3	4	7
1.8	2	2	3	6	2	2	3	5
1.9	2	2	2	4	2	2	2	3

(i) * : Sufficient condition cannot be satisfied.

(ii) For any $n \geq [n_1] + 1$ such that $\frac{N}{n}$ is an integer H_3 is superior to H_2.

(c) A remark on Pedgaonkar and Ajgaonkars' result.

Pedgaonkar and Ajgaonkar (1978) proved that the $G\pi ps$ strategy (H_4) is superior to H_3 for $g > 1$. In fact, much more is known about the $G\pi ps$ strategy. Whenever a $G\pi ps$ strategy is realizable, it is optimal in the class of all p-unbiased strategies, for a given expected sample size, under the model (2.1.1) (Ramachandran, 1978).

Let $E_4 = M_1(H_4)$

$$= \frac{\sigma^2}{N^2}\left[\frac{1}{n}(\Sigma x_i^{g/2})^2 - \Sigma x_i^g\right] \quad \text{(T.J. Rao, 1971)} \qquad \ldots(2.4.10)$$

The same result can also be proved using the following two results

$E_3 \geq E_1$ for $g > 1$ (J.N.K. Rao, 1966)

$E_1 \geq E_4$ for $0 \leq g \leq 2$ (T.J. Rao, 1971)

Hence $E_3 \geq E_4$ for $g > 1$.

However for the sake of completeness we outline the proof of the resu

$E_3 \geq E_4$ for $0 \leq g \leq 2$.

$$\frac{nN^2(E_3-E_4)}{\sigma^2} = \frac{N-n}{N-1}(\Sigma x_i \Sigma x_i^{g-1} - \Sigma x_i^g) - (\Sigma x_i^{g/2})^2 + n\Sigma x_i^g$$

$$= \frac{N-n}{N-1}\left[\Sigma x_i \Sigma x_i^{g-1} - (\Sigma x_i^{g/2})^2\right] - (1 - \frac{N-n}{N-1})(\Sigma x_i^{g/2})^2 + (n - \frac{N-n}{N-1})\Sigma x_i^g$$

$$= \frac{N-n}{N-1}\left[\Sigma x_i \Sigma x_i^{g-1} - (\Sigma x_i^{g/2})^2\right] + \frac{(n-1)N^2}{N-1}\,\mathrm{Cov}(x_i^{g/2}, x_i^{g/2})$$

$$\geq 0$$

Since in the right hand side, the first term is non-negative by Cauchy-Schwarz inequality and the second term is always nonnegative.

Hence $E_3 \geq E_4$ for $0 \leq g \leq 2$.

Thus the $G\pi ps$ strategy is always superior to H_3.

CHAPTER 3

ESTIMATION OF SYMMETRIC PARAMETRIC FUNCTIONS UNDER RANDOM PERMUTATION MODEL

3.0 Summary

In this chapter we work under a random permutation model for the uni-stage set up. When the labels are noninformative and the population is 'homogeneous' w.r.t. the study variate, symmetric estimators (strategies) are found to fare better than their nonsymmetric counterparts in the sense of smaller average risk for convex loss functions. In Section 3.2 we derive some results based on p-distributions. We first prove the 'completeness' of the class $\mathcal{T}^* = \{(p,\bar{y}) : p$ is a simple random sampling and $\bar{y} = \frac{1}{n(s)} \sum_{i \in s} y_i\}$ in the class $\mathcal{T} = \{(p,t) : t$ is linearly invariant$\}$. It is known that the order statistic is complete for a fixed size design (Royall, 1968). Here we show that this result cannot be extended to varying size designs. In Section 3.3 we consider $p\xi$-distributions. We first prove the optimality of a strategy that consists of a fixed size (m) design and a symmetric estimator for its $p\xi$-expectation w.r.t. any convex loss function in the sense of minimum average risk in the class of all $p\xi$-unbiased strategies such that $p(s) > 0 \Longrightarrow n(s) \leq m$. Thus, when the population is 'homogeneous', the labels are noninformative and the sampler's budget allow him to sample at most m units, we give a subjective justification for the use of symmetric estimators accompanied by any fixed size (m) design.

3.1 Introduction

The concept of exchangeability was introduced by de Finetti (1937) in probability theory. The notion of exchangeability in inference from finite populations was first used and popularized by Ericson (1965, 1969a,b). He used 'exchangeable priors' in the Bayesian framework. However the concept of exchangeability is useful in the non-Bayesian framework as well. Exchangeability express a kind of prior knowledge that the labels of units, though observab do not carry any information about the characteristic values associated with the units. The idea of exchangeable priors approximate situations where statisticians believe that simple random sampling would be most appropriate.

In this chapter we deal with discrete exchangeable superpopulation models also known as random permutation models. The idea which found an early application in Madow and Madow (1944) was brought to notice and shown to be useful in inference from finite populations by Kempthorne (1969). We first develop concepts for uni-stage sampling and later on extend the ideas to two stage sampl ing set up.

In uni-stage set up a finite population of size N consists of N distinguishable units. With each unit is associated a distinct integer called label. Without loss of generality we assume that $1,2,\ldots,N$ constitute the labels. The variate value of a typical unit is denoted by y_i, $1 \leq i \leq N$, and $\underset{\sim}{y} = (y_1, y_2, \ldots, y_N)$ is a parametric vector. Let $\bar{R}_N$ be the parametric space.

Let S be the collection of all possible samples and $S(m) \subseteq S$ be a subcollection of samples of size m, $1 \leq m \leq N$. Any probability function p defined on S is a sampling design with $p(s)$ denoting the probability of selecting sample s. An estimator $t(s,\underset{\sim}{y})$ is a real valued function which depends on $\underset{\sim}{y}$ only through the y-values of the units with labels in s. A strategy (p,t) consists of a design and an estimator. The problem of interest is to estimate a symmetric parametric function $\theta(\underset{\sim}{y})$. In particular we shall consider estimation of the population mean $\bar{Y} = \frac{1}{N} \sum_{i=1}^{N} y_i$.

Let $\pi 1, \pi 2, \ldots, \pi N$ be a permutation of integers $1,2,\ldots,N$. As in Stenger (1979), we relabel the units, associating label πi to the unit originally labelled i. This change of labels does not, however, affect the variate values of the units, i.e. with πi is now associated the same variate value which was associated with the unit originally labelled i. If $\pi \underset{\sim}{y}$ denotes the parameter vector after relabelling, then

$$(\pi \underset{\sim}{y})_{\pi i} = y_i \quad \text{or equivalently} \quad (\pi \underset{\sim}{y})_i = y_{\pi^{-1}(i)}$$

i.e. the πith component of $\pi \underset{\sim}{y}$ is identical to the ith component of $\underset{\sim}{y}$. By relabelling, subset s of labels is transformed to πs where $\pi s = \{\pi i : i \in s\}$. Let Π denote the set of all $N!$ permutations of integers $1,2,\ldots,N$.

We shall assume that the parametric space $\tilde{R}_N$ is a subset of N dimensional Euclidean space R_N and is an N-fold product of some set Z of real numbers i.e.

$$y \in \bar{R}_N \iff y_i \in Z \quad \forall\, i = 1,2,\ldots,N$$

for some set Z of real numbers.

Let for $(s,\underset{\sim}{y})$ the coordinates y_i, $i \in s$ arranged in non-decreasing order be $y_{(1)}, y_{(2)}, \ldots, y_{(n(s))}$. The statistic $\eta_s = (\eta_s, \underset{\sim}{y}) = (y_{(1)}, y_{(2)}, \ldots, y_{(n(s))})$ is called an order statistic.

A parametric function $\theta(\underset{\sim}{y})$ is said to be symmetric if

$$\theta(\pi \underset{\sim}{y}) = \theta(\underset{\sim}{y}) \quad \forall\, \pi \in \prod \text{ and } \forall\, \underset{\sim}{y} \in \bar{R}_N .$$

An estimator $t(s,\underset{\sim}{y})$ is said to be symmetric if it depends on $\underset{\sim}{y}$ only through the order statistic η_s i.e. if

$$t(\pi s, \pi \underset{\sim}{y}) = t(s,\underset{\sim}{y}) \quad \forall\, \pi \in \prod \text{ and } \forall\, \underset{\sim}{y} \in \bar{R}_N .$$

Here, and subsequently, s denotes a subset of labels sampled from the 'basic situation' in which a unit labelled i has value y_i. From an estimator t associated with strategy (p,t) a symmetric estimator $\bar{t}_p$ is constructed as follows, (Royall, 1970),

$$\bar{t}_p(s,\underset{\sim}{y}) = \sum_{\pi} t(\pi s, \pi \underset{\sim}{y})\, p(\pi s) \,/\, m!\,(N-m)!\,p(m), \quad p(m) \neq 0 \qquad \ldots(3.1.1)$$

where $n(s) = m$ and $p(m) = \sum_{s \in S(m)} p(s)$

if $p(m) = 0$ then $\bar{t}_p(s,\underset{\sim}{y})$ may be assigned arbitrary values.

A design p is said to be symmetric if

$$p(\pi s) = p(s) \quad \forall\, \pi \in \prod \text{ and } \forall\, s \in S .$$

If p_m, $1 \leq m \leq N$, is a simple random sampling design of fixed size m, i.e.

$$p_m(s) = \frac{1}{\binom{N}{m}} \quad \text{if} \quad s \varepsilon S(m)$$
$$= 0 \quad \text{otherwise} . \qquad \ldots(3.1.2)$$

Then a design p is symmetric if and only if it is a probability mixture of $p_1, p_2, \ldots, p_N$.

Symmetry of a design p is therefore equivalent to the existence of a probability vector $\underset{\sim}{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_N)$ such that

$$p(s) = \sum_{m=1}^{N} \alpha_m p_m(s) \ , \quad s \varepsilon S$$

where p_m is given by (3.1.2).

Clearly the design $p = \sum_{m=1}^{N} \alpha_m p_m$ has an average size $\sum_{m=1}^{N} m\alpha_m$.

With $\underset{\sim}{\alpha}$ we associate another probability vector $\underset{\sim}{\alpha}^* = (\alpha_1^*, \alpha_2^*, \ldots, \alpha_N^*)$ by defining

$$\alpha_i^* = 1 - (\Sigma j\alpha_j - [\Sigma j\alpha_j]) \quad \text{for} \quad i = [\Sigma j\alpha_j]$$
$$= \Sigma j\alpha_j - [\Sigma j\alpha_j] \quad \text{for} \quad i = [\Sigma j\alpha_j] + 1 \qquad \ldots(3.1.3)$$
$$= 0 \quad \text{otherwise}$$

where $[x]$ denotes the largest integer not exceeding x.

It is easy to see that the symmetric design $p^* = \sum_{m=1}^{N} \alpha_m^* p_m$ has the same average size $\sum_{m=1}^{N} m\alpha_m$ as that of design $p = \Sigma \alpha_m p_m$. Further $p^*(s_1) > 0, \ p^*(s_2) > 0 \implies |n(s_1) - n(s_2)| \leq 1$. If $\Sigma m\alpha_m$ is an integer, say n, then $p^* = p_n$ and if $\Sigma m\alpha_m$ is not an integer then p^* is a probability mixture of $p_{[\Sigma m\alpha_m]}$ and $p_{[\Sigma m\alpha_m]+1}$. Note that for the sampling design p^* the sample size does not differ by more than unity which corresponds to the least possible variation to realize the average sample size $\Sigma m\alpha_m$ of the design. Design p^* is called a simple random sampling of almost fixed si

Given a design p of average size $n+\theta$, $1 \leq n < N$ and $0 \leq \theta \leq 1$, we associate, with p, designs $\bar{p}$ and p^* as follows

$$\bar{p}(s) = \frac{p(n)}{\binom{N}{n}} \quad \forall \quad s \in S(n),\ 1 \leq n \leq N \qquad \ldots(3.1.4)$$

and $$p^*(s) = (1-\theta)p_n(s) + \theta p_{n+1}(s)\ ,\ s \in S \qquad \ldots(3.1.5)$$

Let D be the class of all symmetric designs. Further let

$\delta_m = \{p : p(s) > 0 \implies n(s) \leq m\}$ and $\rho_m = \{p : p(s) > 0 \implies n(s) = m\}$

From an estimator t associated with strategy $(p,t), p \in \delta_m$ let us construct an estimator t_p^* as follows, (Joshi, 1979).

$$t_p^*(s_m, \underset{\sim}{y}) = \sum_{k=1}^{m} p(k) / \binom{m}{k} \sum_{s_k \subset s_m} \bar{t}_p(s_k, \underset{\sim}{y})\ ;\ s_m \in S(m) \quad \ldots(3.1.6$$

and for $s \notin S(m)$, t_p^* may be assigned any arbitrary values, $\bar{t}_p(s, \underset{\sim}{y})$ being given by (3.1.1).

We now proceed to an important concept of this chapter. As mentioned earlier we will be dealing with the special case of exchangeable models namely random permutation models. In its simplest form, the model is equivalent to an assumption that the units which bear fixed but unknown values $y_1, y_2, \ldots, y_N$ have been labelled at random i.e. all the $N!$ ways of labelling the given set of N units are equally likely. This reflects the situation when labels are used only to identify the units and the sampler has no knowledge, what so ever, of any relationship between the labels and values of the units. Under this set up all possible distinct vectors obtained by permuting the vector $\underset{\sim}{y}$ are equiprobable i.e. one could say that the fixed but unknown numbers y_1, $y_2, \ldots, y_N$ have been randomly assigned to the N units.

Formally, random permutation model is a class of distributions ξ such that, for any fixed, unknown numbers $y_1, y_2, \ldots, y_N$ the random variables $Y_1, Y_2, \ldots, Y_N$ have an exchangeable distribution such that

$$\text{Prob}\left[Y_i = y_{\pi^{-1}(i)},\ 1 \leq i \leq N\right] = \frac{1}{N!}\quad \forall\ \pi \in \prod .$$

Thus random permutation model expresses the prior belief that the labels are uninformative.

Though the random permutation model can be viewed in two ways namely the units which bear fixed but unknown values are labelled at random and the fixed set of N values are assigned to the units at random, we would, for the analysis in this chapter, adopt the former characterization of the random permulation model.

For a strategy (p,t) we now define various expectations.

(a) p-expectation (design expectation)

$$E_p(t(s,\underset{\sim}{y})) = E_p(t) = \sum_{s \in S} p(s)t(s,\underset{\sim}{y})$$

(b) ξ-expectation (model expectation)

$$E_\xi(t(s,\underset{\sim}{y})) = E_\xi(t) = \frac{1}{N!}\sum_{\pi} t(s,\pi \underset{\sim}{y}) \; ; \; p(s) > 0.$$

(c) $p\xi$-expectation (model-design expectation)

$$E_p E_\xi(t(s,\underset{\sim}{y})) = E_p E_\xi(t) = \frac{1}{N!}\sum_{\pi}\sum_{s\in S} p(s)t(s,\pi \underset{\sim}{y}). \qquad \ldots(3.1.7)$$

We assume a loss function $\ell(a, \theta)$ which is convex in a for every value of θ. Let us now define average risk for a strategy (p,t) for estimating a symmetric parametric function $\theta(\underset{\sim}{y})$ corresponding to a convex loss function $\ell(a,\theta)$ as

$$\bar{R}(t,p,\underset{\sim}{y}) = \frac{1}{N!}\sum_{\pi} R(t,p,\pi \underset{\sim}{y})$$
$$= \frac{1}{N!}\sum_{\pi}\sum_{s\in S} p(s)\ell(t(s,\pi \underset{\sim}{y}), \theta(\underset{\sim}{y})). \qquad \ldots(3.1.8)$$

<u>Definition 3.1.1</u> An estimator t associated with strategy (p,t) is said to be linearly invariant if t is linear i.e. of the type

$$t(s,\underset{\sim}{y}) = \sum_{i\in s} b(s,i)y_i$$

and $\forall\, s$ with $p(s) > 0$, $\sum_{i\in s} b(s,i) = 1$.

Let $\mathcal{T}_0$ and $\mathcal{T}$ be two classes of strategies such that $\mathcal{T}_0 \subseteq \mathcal{T}$.

<u>Definition 3.1.2.</u> $\mathcal{T}_0$ is said to be complete in $\mathcal{T}$ if for every strategy $(p,t) \in \mathcal{T} - \mathcal{T}_0$ there exists a strategy $(p',t') \in \mathcal{T}_0$ such that

$$\bar{R}(p',t',\underset{\sim}{y}) \le \bar{R}(p,t,\underset{\sim}{y}) \;\forall\; \underset{\sim}{y} \in \bar{R}_N.$$

In this chapter we demonstrate optimality of symmetric stimators, under random permutation model, in a reasonable sense. We first derive some results based on p-distribution and then proceed to obtain results based on pξ-distribution.

3.2 Estimation of the Population Mean

Royall (1968) proved the following result regarding the completeness of order statistic.

Theorem 3.2.1. For any fixed size design p, the order statistic is complete. i.e. if for a function $\phi(\eta_s, \underset{\sim}{y})$ of order statistic $\sum_{s \in S} p(s)\phi(\eta_s, \underset{\sim}{y}) = 0 \;\; \forall \; \underset{\sim}{y} \in \bar{R}_N$ then $\phi(\eta_s, \underset{\sim}{y}) = 0 \;\; \forall \; \underset{\sim}{y} \in \bar{R}_N$ and $\forall$ s with $p(s) > 0$.

We state yet another result,

Theorem 3.2.2. For a strategy (p,t) for estimating $\theta(\underset{\sim}{y})$

$$\bar{R}(\bar{t}_p, p, \underset{\sim}{y}) \leq \bar{R}(t, p, \underset{\sim}{y}) \quad \forall \; \underset{\sim}{y} \in R_N \qquad \ldots(3.2.1)$$

and $$\bar{R}(\bar{t}_p, \bar{p}, \underset{\sim}{y}) = \bar{R}(\bar{t}_p, p, \underset{\sim}{y}) \quad \forall \; \underset{\sim}{y} \in R_N \qquad \ldots(3.2.2)$$

(3.2.1) is due to Royall (1970) and (3.2.2) is due to Joshi (1979). The interpretation of Theorem 3.2.2 is that for a given design there is gain in using symmetric estimators over non-symmetric estimators. However, once a symmetric estimator is used there is no additional gain by symmetrizing the accompanying sampling design.

In this section we consider estimation of the population mean $\bar{Y} = \frac{1}{N} \sum_{i=1}^{N} y_i$.

Let $\mathcal{T} = \{(p,t) : t$ is linearly invariant and p is arbitrary$\}$

and $\mathcal{T}^* = \{(p^*, \bar{y}) : p^*$ is a simple random sampling of almost fixed size and $\bar{y} = \frac{1}{n(s)} \sum_{i \in s} y_i\}$.

Let $\ell(a,\theta)$ be strictly convex in $a \in I$, an open interval, and let the second order partial derivative $\partial^2 \ell(a,\theta)/\partial a^2$ exist for all $a \in I$ and be convex. ...(3.2.3)

For the following Theorem 3.2.3, Corollary 3.2.1 and Corollary 3.2.2 we assume that $\ell(a,\theta)$ satisfies (3.2.3) and the parametric space $\bar{R}_N$ is N-fold product of the open interval I.

We now have the following,

<u>Theorem 3.2.3</u> $\mathcal{T}^*$ is complete in $\mathcal{T}$.

<u>Proof</u>. First we observe that for a strategy (p,t) if t is linearly invariant then $\bar{t}_p = \bar{y}$. Let $t(s,\underset{\sim}{y}) = \sum_{i \in s} b(s,i) y_i$. For a sample s such that $n(s) = n$ and $p(s) > 0$, we have,

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

$$= \sum_{i \in s} y_i \left[\sum_{s_1 \in S(n)} p(s_1) \sum_{i \in s_1} b(s_1, i) \right] (n-1)!(N-n)!$$

$$= (n-1)!(N-n)!\, p(n) \sum_{i \in s} y_i$$

since $\sum_{i \in s} b(s,i) = 1$ for linearly invariant estimator t.

Now using (3.1.1), we get, $\bar{t}_p = \bar{y}$.

By Theorem 3.2.2, combining (3.2.1) and (3.2.2), we have, for any strategy (p,t), for any convex loss function and for any N-fold product space $\bar{R}_N$,

$$\bar{R}(\bar{t}_p, \bar{p}, \underset{\sim}{y}) \leq \bar{R}(t,p,\underset{\sim}{y}) \;\forall\; \underset{\sim}{y} \in \bar{R}_N. \qquad \ldots(3.2.4)$$

Following Stenger and Gabler (1981) we can establish that for a symmetric sampling design p

$$\bar{R}(\bar{y},p^*,\underset{\sim}{y}) \leq \bar{R}(\bar{y},p,\underset{\sim}{y}) \;\forall\; \underset{\sim}{y} \in \bar{R}_N. \qquad \ldots(3.2.5)$$

Combining (3.2.4) and (3.2.5), in view of the fact that for any linearly invariant (p,t), $\bar{t}_p = \bar{y}$, we get that for a linearly invariant strategy (p,t)

$$\bar{R}(\bar{y},p^*,\underset{\sim}{y}) \leq \bar{R}(t,p,\underset{\sim}{y}) \;\forall\; \underset{\sim}{y} \in \bar{R}_N.$$

This establishes the completeness of the class of strategies $\mathcal{T}^*$ in the superclass of strategies $\mathcal{T}$ for any convex loss function satisfying (3.2.3) and the parametric space $\bar{R}_N$, the N-fold product of open interval I .

<u>Corollary 3.2.1</u> For a given $n+\theta$, $1 \leq n < N$ and $0 \leq \theta < 1$, for estimating the population mean $\bar{Y}$, $(p^*_{n+\theta}, \bar{y})$ is the best strategy in the class of strategies $\mathcal{T}_{n+\theta} = \{(p,t) : t$ is linearly invariant and $\sum_{s \in S} n(s)p(s) = n+\theta\}$ for any loss function satisfying (3.2.3) in the sense that

$$\bar{R}(\bar{y}, p^*_{n+\theta}, \underset{\sim}{y}) = \min_{(p,t)\in\mathcal{T}_{n+\theta}} \bar{R}(t,p,\underset{\sim}{y}) \;\forall\; \underset{\sim}{y} \in \bar{R}_N$$

where $p^*_{n+\theta}$ is simple random sampling of almost fixed size $n+\theta$.

<u>Corollary 3.2.2</u> For a given integer m, $1 \leq m \leq N$, for estimating the population mean $\bar{Y}$, $(p_m, \bar{y})$ is the best strategy in the class $\mathcal{T}_m = \{(p,t) : t$ is linearly invariant and $\sum_{s \in S} n(s)p(s) = m\}$ for any loss function satisfying (3.2.3) in the sense that

$$\bar{R}(\bar{y}, p_m, \underset{\sim}{y}) = \min_{(p,t) \in \mathcal{T}_m} \bar{R}(t, p, \underset{\sim}{y}) \quad \forall\ \underset{\sim}{y} \in \bar{R}_N.$$

<u>Remark 3.2.1</u> Stenger (1979) considers the following symmetrizatio Given an estimator t, he defines

$$\bar{t}(s, \underset{\sim}{y}) = \frac{1}{N!} \sum_{\pi} t(\pi s, \pi \underset{\sim}{y}). \qquad \ldots(3.2.6)$$

For a linearly invariant strategy (p,t) for any convex loss function and for any N-fold product space $\bar{R}_N$, Stenger (1979) proves that

$$\max_{\pi \in \Pi} R(\bar{y}, \bar{p}, \pi \underset{\sim}{y}) \leq \max_{\pi \in \Pi} R(t, p, \pi \underset{\sim}{y}) \ \forall\ \underset{\sim}{y} \in \bar{R}_N. \qquad \ldots(3.2.7)$$

However, note that, for a linearly invariant estimator t both types of symmetrization $\bar{t}_p$ of (3.1.1) and $\bar{t}$ of (3.2.6) yield sample mean. Hence using (3.2.4) for the case of linearly invariant strategy (p,t), we have,

$$\bar{R}(\bar{y}, \bar{p}, \underset{\sim}{y}) \leq \bar{R}(t, p, \underset{\sim}{y}) \ \forall\ \underset{\sim}{y} \in \bar{R}_N.$$

Further $R(\bar{y}, \bar{p}, \pi \underset{\sim}{y}) = \bar{R}(\bar{y}, \bar{p}, \underset{\sim}{y}) \ \forall\ \pi \in \Pi$ and $\forall\ \underset{\sim}{y} \in \bar{R}_N$.

Therefore we have,

$$\max_{\pi \in \Pi} R(\bar{y}, \bar{p}, \pi \underset{\sim}{y}) \leq \max_{\pi \in \Pi} R(t, p, \pi \underset{\sim}{y}) \quad \forall\ \underset{\sim}{y} \in \bar{R}_N.$$

Thus Stenger's (1979) result (3.2.7) is a particular case of (3.2.4)

Remark 3.2.2. The above results are proved for linearly invariant estimators. As a matter of fact, for any design p, for estimating the population mean $\bar{Y}$, the only estimator that is symmetric as well as linearly invariant is the sample mean $\bar{y}$.

Proposition 3.2.1. The family of sampling distributions of order statistic generated by the class of all symmetric designs is complete i.e. if

$$E_p \emptyset(\eta_s, \underset{\sim}{y}) = 0 \quad \forall\, p \in D \quad \text{and} \quad \forall\, \underset{\sim}{y} \in \bar{R}_N$$

then
$$\emptyset(\eta_s, \underset{\sim}{y}) = 0 \quad \forall\, s \in S \quad \text{and} \quad \forall\, \underset{\sim}{y} \in \bar{R}_N .$$

Proof. The proof follows from the fact that for any fixed size design, in particular, for p_m, $1 \leq m \leq N$, the order statistic is complete. (Theorem 3.2.1, Royall, 1968).

Remark 3.2.3 As a consequence of the above proposition sample mean is the only estimator that is design unbiased for every symmetric sampling design.

After proving the completeness of order statistic for the class of fixed size designs a natural question that arises is that 'does the same result hold even if we relax the condition of fixed size ?' The answer to this question is 'no' and can be seen as follows.

Proposition 3.2.2. Let p be any varying size design i.e. there are integers m_1 and m_2, $1 \leq m_1 \neq m_2 \leq N$, with $p(m_1) > 0$ and $p(m_2) > 0$. The order statistic η_s is not complete for p.

Proof. Let $\phi(\eta_s, \underset{\sim}{y})$ be a function of the order statistic defin[ed] for every $\underset{\sim}{y} \in R_N$, as follows

$$\begin{aligned}\phi(\eta_s, \underset{\sim}{y}) &= a n_1 \quad \forall\, s \in S(n_1) \text{ and } \forall\, \underset{\sim}{y} \in R_N,\ a \neq 0 \\ &= -a n_1 \frac{p(n_1)}{p(n_2)} \quad \forall\, s \in S(n_2) \text{ and } \forall\, \underset{\sim}{y} \in R_N \\ &= 0 \qquad \text{otherwise}.\end{aligned}$$

Clearly $\sum_{s \in S} p(s)\phi(\eta_s, \underset{\sim}{y}) = 0 \quad \forall\, \underset{\sim}{y} \in R_N$

but ϕ is not zero identically.

This proves that the order statistic is not complete for varying size sampling designs.

3.3. Estimation of a Parametric Function $\theta(\underset{\sim}{y})$

In Section 3.2 we derived a few results based on p-distribution. In this section we obtain some results based on pξ-distribution.

Suppose for a finite population the mean value of certain characteristic is to be estimated and according to prior knowledg[e] the population is 'homogeneous' w.r.t. the study variate. Furthe[r] suppose that the sampler's budget allows him to sample at most n units. Then what is a 'reasonable' strategy for him? In practic[e] one invariably goes for a simple random sampling of size n and sample mean. Though this strategy has an intuitive appeal, first non-intuitive justification was given by Joshi (1979). He proved that the above strategy is the best in the class of all p-unbiased strategies $(p(s) > 0 \implies n(s) \leq m)$ in the sense of minimum average risk for any convex loss function.

Here we generalize the above result. We prove optimality of a strategy that consists of any fixed size (m) design and a symmetric estimator for its $p\xi$-expectation in the class of all $p\xi$-unbiased strategies $(p(s) > 0 \implies n(s) \leq m)$ in the sense of minimum average risk for any convex loss function.

Let us first prove a lemma regarding the completeness of order statistic in the $p\xi$-distribution sense.

Lemma 3.3.1. For any fixed size (m) design p, the order statistic is complete in the $p\xi$-sense i.e. if

$$E_p E_\xi \phi(\eta_s, \underset{\sim}{y}) = 0 \quad \forall \underset{\sim}{y} \in \bar{R}_N \text{ then}$$

$$\phi(\eta_s, \underset{\sim}{y}) = 0 \quad \forall \underset{\sim}{y} \in \bar{R}_N \text{ and } \forall s \text{ with } p(s) > 0.$$

Proof. Note that,

$$E_p E_\xi \phi(\eta_s, \underset{\sim}{y}) = \frac{1}{N!} \sum_{\pi} \sum_{s \in S} p(s) \phi(\eta_s, \pi \underset{\sim}{y})$$

$$= \frac{1}{\binom{N}{m}} \sum_{s \in S(m)} \phi(\eta_s, \underset{\sim}{y}) .$$

Since the order statistic is complete for p_m in the p-sense its completeness for a fixed size (m) design p in the $p\xi$-sense now follows.

Here we consider a convex loss function $\ell(a, \theta)$ and an N-fold product space $\bar{R}_N$.

Lemma 3.3.2

(i) $E_p E_\xi(t) = E_{\bar{p}} E_\xi(\bar{t}_p)$ for any strategy (p,t)

(ii) $E_q E_\xi(t_p^*) = E_{\bar{p}} E_\xi(\bar{t}_p)$ for any strategy (p,t) with $p \in \delta_m$ and for any $q \in \rho_m$.

Proof.

(i)
$$E_{\bar{p}} E_\xi(\bar{t}_p) = \frac{1}{N!} \sum_\pi \sum_{s \in S} \bar{p}(s) \bar{t}_p(s, \pi \underset{\sim}{y})$$

$$= \sum_{n=1}^{N} \frac{p(n)}{\binom{N}{n}} \sum_{s \in S(n)} \bar{t}_p(s, \underset{\sim}{y})$$

$$= \sum_{n=1}^{N} \frac{p(n)}{\binom{N}{n}} \sum_{s \in S(n)} \sum_\pi \frac{t(\pi s, \pi y) p(\pi s)}{n!\,(N-n)!\,p(n)}$$

$$= \frac{1}{N!} \sum_\pi \sum_{n=1}^{N} \sum_{s \in S(n)} p(s) t(s, \pi \underset{\sim}{y})$$

$$= E_p E_\xi(t).$$

(ii)
$$E_q E_\xi(t_p^*) = \frac{1}{N!} \sum_\pi \sum_{s \in S} q(s) t_p^*(s, \pi \underset{\sim}{y})$$

$$= \frac{1}{\binom{N}{m}} \sum_{s \in S(m)} t_p^*(s, \underset{\sim}{y})$$

$$= \frac{1}{\binom{N}{m}} \sum_{s \in S(m)} \sum_{k=1}^{m} \frac{p(k)}{\binom{m}{k}} \sum_{s_k \subset s} \bar{t}_p(s_k, \underset{\sim}{y})$$

$$= \sum_{k=1}^{m} \frac{p(k)}{\binom{N}{k}} \sum_{s \in S(k)} \bar{t}_p(s, \underset{\sim}{y})$$

$$= E_{\bar{p}} E_\xi(\bar{t}_p).$$

This proves the lemma.

We prove yet another lemma before going to our main theorem.

<u>Lemma 3.3.3</u>. For any strategy (p,t) with $p \in \delta_n$ and $q \in \rho_n$, for estimating any symmetric function $\theta(\underset{\sim}{y})$, we have,

$$\bar{R}(t_p^*, q, \underset{\sim}{y}) \leq \bar{R}(\bar{t}_p, \bar{p}, \underset{\sim}{y}) \leq \bar{R}(t, p, \underset{\sim}{y}) \quad \forall \underset{\sim}{y} \in \bar{R}_N$$

<u>Proof</u>.

$$\bar{R}(t_p^*, q, \underset{\sim}{y}) = \frac{1}{N!} \sum_{\pi} \sum_{s \in S} q(s)\, \ell[t_p^*(s, \pi\underset{\sim}{y}), \theta(\underset{\sim}{y})]$$

$$= \frac{1}{\binom{N}{n}} \sum_{s \in S(n)} \ell[t_p^*(s, \underset{\sim}{y}), \theta(\underset{\sim}{y})]$$

$$\leq \sum_{k=1}^{n} \frac{p(k)}{\binom{N}{k}} \sum_{s \in S(k)} \ell[t_p^*(s, \underset{\sim}{y}), \theta(\underset{\sim}{y})] \quad \text{(since } \ell \text{ is convex)}$$

$$= \bar{R}(\bar{t}_p, \bar{p}, \underset{\sim}{y}).$$

The proof is now complete using (3.2.4).

<u>Remark 3.3.1</u>. Note that $\bar{R}(t_p^*, q, \underset{\sim}{y})$ is same for all $q \in \rho_n, \forall \underset{\sim}{y} \in \bar{R}_N$.

We are now in a position to state and prove our main theorem. Let t^* be any function of order statistic η_s. Let $\theta(\underset{\sim}{y})$ be the common value of $E_q E_\xi(t^*)$, $q \in \rho_n$, and suppose that $\theta(\underset{\sim}{y})$ is the parametric function of interest. Let $H_n = \{(p,t) : p \in \delta_n$ and $E_p E_\xi(t) = \theta(\underset{\sim}{y})\}$. We now have

<u>Theorem 3.3.1</u> For estimating $\theta(\underset{\sim}{y})$,

$$\bar{R}(t^*, q, \underset{\sim}{y}) = \min_{(p,t) \in H_n} \bar{R}(t, p, \underset{\sim}{y}) \quad \forall \underset{\sim}{y} \in \bar{R}_N$$

where q is any design in ρ_n.

Proof. The proof follows from Lemmas 3.3.1, 3.3.2 and 3.3.3.

Corollary 3.3.1. Strategy (p_m, t^*) is the best p-unbiased strategy in the subclass of p-unbiased strategies of H_m, in the sense that for any convex loss function (p_m, t^*) minimizes the average risk $\bar{R}(t,p,\underset{\sim}{y})$ in that subclass.

Proof. The proof follows from Theorem 3.3.1 and the fact that (p_m, t^*) is p-unbiased for $\theta(\underset{\sim}{y})$.

Corollary 3.3.2. As special cases of Corollary 3.3.1 we get

(a) (Joshi, 1979). For $t^* = \bar{y}$, $\theta(\underset{\sim}{y}) = \bar{Y}$. Therefore for estimati the population mean, $(p_m, \bar{y})$ is the best strategy in the su class of p-unbiased strategies of H_m.

(b) For $t^* = s^2 = \frac{1}{n-1} \sum_{i \varepsilon s} (y_i - \bar{y})^2$, $\theta(\underset{\sim}{y}) = S^2 = \frac{1}{N-1} \sum_{i=1}^{N} (y_i - \bar{Y})^2$.

Theorem 3.3.1 provides a subjective justification for the use of symmetric estimators.

CHAPTER 4

ESTIMATION UNDER TWO STAGE RANDOM PERMUTATION MODEL

.0 Summary

In this chapter we deal with two stage random permutation model defined for an $L \times M$ complete array of which the present population is assumed to be a truncation. When the population is 'homogeneous' and the labels are noninformative, symmetric estimators are found to fare better than nonsymmetric estimators. In this chapter we first define order statistic for a two stage sample and prove its completeness, in the $p\xi$ sense, for fixed size designs. We then prove the optimality of a strategy that consists of a fixed size (k,m) design and the over all sample mean, for estimating the population mean, in the sense of minimum average risk for convex loss functions in a class of $p\xi$-unbiased strategies (p selects at most k primary sampling units (psu's) and at most m secondary sampling units (ssu's) per sampled psu). Finally, we establish the optimality of a strategy consisting of an almost fixed size $(k+\theta_1, m+\theta_2)$ design, $0 \leq \theta_1, \theta_2 \leq 1$, and $\bar{\bar{y}}$, the mean of means per ssu in a class $\{(p, \bar{\bar{y}}) : p$ selects, on an average, $k+\theta_1$ psu's and $m+\theta_2$ ssu's per sampled psu$\}$ for estimating the population mean again in the sense of minimum average risk but now for the squared error loss function.

4.1 Introduction

In the previous chapter we obtained certain 'optimal' strategies, to estimate symmetric parametric functions, under the random permutation model all in unistage set up. In this chapter an attempt is made to obtain some similar results in two stage sampling set up.

The problem of finding 'optimal' strategies in two stage sampling set up has been discussed in literature with great deal of interest, notably by Scott and Smith (1969, 1975), J.N.K. Rao (1975), Bellhouse, Thompson and Godambe (1977), J.N.K. Rao and Bellhouse (1978). In some of those studies various two stage exchangeable superpopulation models have been invoked to obtain 'optimal' strategies. J.N.K. Rao was the first to present a discrete exchangeable superpopulation model, also called random permutation model, for two stage set up. Here we adopt the two stage random permutation model suggested by Bellhouse et al (1977). They established 'optimality' of a strategy that consists of scaled sample mean and a design that makes it p-unbiased in the class of all fixed size strategies. Essentially it is a generalization of a result due to Godambe and Thompson (1973) that establishes optimality of a strategy that consists of a πps design and Horvitz-Thompson estimator, in the class of all fixed size strategies, under 'exchangeability' for uni-stage sampling.

A finite two stage population U consists of L primary sampling units (psu's), consisting of $M_1, M_2, \ldots, M_L$ secondary sampling units (ssu's) respectively. Without loss of generality we assume that $\{(i,j) : j = 1, 2, \ldots, M_i$ and $i = 1, 2, \ldots, L\}$ is the set of labels. Variate value of a typical unit is denoted by y_{ij}, $1 \leq j \leq M_i$; $1 \leq i \leq L$. $\underset{\sim}{y} = (y_{ij})$ denotes the parametric array.

It is convenient to represent population characteristic in the generally incomplete $L \times M$ array $\underset{\sim}{y} = (y_{ij})$, where $M = \max_{1 \leq i \leq L} M_i$. $\underset{\sim}{y}$ is unknown before sampling and y_{ij} may be any real numbers.

Two stage samples and two stage sampling designs for U may be defined through corresponding $L \times M$ complete or rectangular population say U^o. Effectively we are assuming the existence of some additional units at hypothetical level so that each of the psu's consists of M ssu's. A two stage sampling design is one that selects a subset of primary units at first stage and conditional on this selection, it selects sets of secondary units at the second stage independently from each sampled psu. Let S be the collection of all possible two stage samples for the complete $L \times M$ population. Let $S(n_1, n_2, \ldots, n_k) \subseteq S$ be the collection of samples that contain $n_1, n_2, \ldots, n_k$ ssu's from k different psu's. If $\{Q = (i,j) : j > M_i, 1 \leq i \leq L\}$, then a sampling design for the original population U is defined as

$$p : S \longrightarrow [0, 1]$$

$$\sum_{s \in S} p(s) = 1$$

and $p(s) = 0$ if s contains a label in Q.

Let $p(n_1,n_2,\ldots,n_k) = \sum_{s \in S(n_1,n_2,\ldots,n_k)} p(s)$ ...(4.1.1)

Given a two stage sampling design p for U (implementable design) we associate a design $\bar{p}$ for U^o as

$$\bar{p}(s) = \frac{p(n_1,n_2,\ldots,n_k)}{c(n_1,n_2,\ldots,n_k)}, \quad s \in S(n_1,n_2,\ldots,n_k) \qquad \ldots(4.1.2)$$

where $c(n_1,n_2,\ldots,n_k)$ is the cardinality of $S(n_1,n_2,\ldots,n_k)$. If $(n_1,n_2,\ldots,n_k)$ is such that there are r distinct values wit frequencies $d_1,d_2,\ldots,d_r$; $\sum_{i=1}^{r} d_i = k$, then define

$$I(n_1,n_2,\ldots,n_k) = \frac{k!}{\prod_{i=1}^{r} d_i!} \qquad \ldots(4.1.3)$$

The cardinality $c(n_1,n_2,\ldots,n_k)$ of $S(n_1,n_2,\ldots,n_k)$ is given by

$$c(n_1,n_2,\ldots,n_k) = \binom{L}{k} I(n_1,n_2,\ldots,n_k) \prod_{i=1}^{k} \binom{M}{n_i} . \qquad \ldots(4.1.4)$$

Clearly $\bar{p}$ is a design only at the hypothetical level. $\bar{p}$ is not implementable for U unless it is complete. We would viev $\bar{p}$ as a design for U^o associated with an implementable design p for U.

An estimator $t(s, \underset{\sim}{y})$ is a real valued function that depends on $\underset{\sim}{y}$ only through the y-values of units with labels in s A strategy (p,t) consists of an estimator t and a two stage design p. The problem of interest is to estimate the mean of a character for a two stage population.

Let $(\pi(i,j))$ denote a permutation of complete $L \times M$ array of labels $((i,j))$, $1 \leq j \leq M$, $1 \leq i \leq L$, obtained by first permuting the rows of $((i,j))$ and then permuting the labels within each row independently. We shall relabel the units associating label $\pi(i,j)$ to the unit originally labelled (i,j). This change of labels does not, however, affect the variate values of the units, i.e., with $\pi(i,j)$ is now associated the same variate value as that associated with the unit originally labelled (i,j). If $\pi \underset{\sim}{y}$ denotes the parametric array obtained by relabelling, then

$$(\pi \underset{\sim}{y})_{\pi(i,j)} = y_{ij}$$

i.e. the $\pi(i,j)$th component of $\pi \underset{\sim}{y}$ is identical to the (i,j)th component of $\underset{\sim}{y}$. Relabelling transforms set s of labels to $\pi s = \{\pi(i,j) : (i,j) \in s\}$. Let $\prod$ be the collection of all possible $L!\,(M!)^L = \lambda$ (say) permutations.

We now proceed to define two stage random permutation model for the case of $L \times M$ complete population. In its simplest form the two stage random permutation model means that the units which bear fixed but unknown values y_{ij} have been labelled randomly in two stages, first the psu's are labelled at random and then the ssu's within each of the psu's are labelled at random independently i.e. all possible λ ways of labelling the given $L \times M$ population, in two stages, are equally likely. This reflects the situation when labels are used only to identify the units and the sampler has no knowledge, what so ever, of any relationship between labels and values of the units. In this set up all possible distinct $L \times M$ arrays, obtained by permuting first the rows of $\underset{\sim}{y} = (y_{ij})$ and then the values in rows independently, are equiprobable. Equivalently one could say that there are L sets of fixed but unknown numbers which have been associated to L psu's at random and the values in these sets are associated to ssu's within psu's at random independently.

Thus two stage random permutation model is a class of distributions ξ such that for any fixed but unknown array $\underset{\sim}{y} = (y_{ij})$, $1 \leq j \leq M$, $1 \leq i \leq L$, the random variables Y_{ij}, $1 \leq j \leq M$, $1 \leq i \leq L$, have an exchangeable distribution such that

$$\text{Prob}\left[Y_{ij} = y_{\pi^{-1}(i,j)} \; ; \; \begin{matrix} 1 \leq j \leq M \\ 1 \leq i \leq L \end{matrix}\right] = \frac{1}{\lambda} \quad \forall \; \pi \in \Pi .$$

We adopt the following formulation of the two stage random permutation model, for the analysis in this chapter; all possible λ ways of labelling a given $L \times M$ population in two stages, are equally likely. However in practice two stage populations are mostly incomplete. Let $C = \{\pi_t \underset{\sim}{y} : \pi \in \Pi\}$ where $\pi_t \underset{\sim}{y}$ is a truncation of $\pi \underset{\sim}{y}$ obtained by dropping the portion corresponding to labels in Q. The assumption embodied in the superpopulation model is that the actual incomplete population is a point from C. Note that to define two stage random permutation model we introduc necessary hypothetical ssu's and obtain a complete array. Two stage random permutation model is defined for a complete array of which the incomplete array is assumed to be a truncation.

This definition of two stage random permutation model has been adapted from Bellhouse et al (1977).

A two stage order statistic for $s \in S(n_1, n_2, \ldots, n_k)$ is defined as follows.

We shall let $\eta_s = (\eta_{uv})$ be a $k \times n$ incomplete array of sampled values of $\underset{\sim}{y}$, where $n = \max_{1 \leq i \leq k} n_i$, satisfying the following three conditions

(i) for each u, the values $\eta_{u1}, \eta_{u2}, \ldots, \eta_{un_u}$ correspond to the ssu's from the same psu

(ii) for each u, $\eta_{u1} \geq \eta_{u2} \geq \cdots \geq \eta_{un_u}$

and finally,

(iii) starting at the bottom right, rows are in increasing lexicographic order, i.e. if $u < w$ and there are n_1 and n_2 entries in the uth and wth rows respectively and $m = \min(n_1, n_2)$ then look at the smallest m entries from each of the rows and let them be

$$\eta_{u1} \geq \eta_{u2} \geq \cdots \geq \eta_{um}$$

$$\eta_{w1} \geq \eta_{w2} \geq \cdots \geq \eta_{wm}$$

then either $\eta_{uv} = \eta_{wv}$ for $v = 1, 2, \ldots, m$ and $n_1 \geq n_2$ or there exists $v_o \leq m$ such that $\eta_{uv_o} > \eta_{wv_o}$ and $\eta_{uv} = \eta_{wv}$ for all v, if any, greater than v_o.

This/a is generalization of the two stage order statistic defined for samples $s \in S(m, m, \ldots, m)$ by Bellhouse et al (1977).

Example 4.1.1

Let $k = 5$, $n_1 = 3$, $n_2 = 4$, $n_3 = n_4 = 5$, $n_5 = 6$.

Let the observations be

$$\begin{bmatrix} & & & 2 & 3 & 2 \\ & & 5 & 2 & 4 & 1 \\ & 5 & 3 & 2 & 6 & 5 \\ & 1 & 6 & 2 & 1 & 7 \\ 6 & 5 & 4 & 6 & 8 & 5 \end{bmatrix}$$

using (ii) we get

$$\begin{bmatrix} & & & 3 & 2 & 2 \\ & & 5 & 4 & 2 & 1 \\ & 6 & 5 & 5 & 3 & 2 \\ & 7 & 6 & 2 & 1 & 1 \\ 8 & 6 & 6 & 5 & 5 & 4 \end{bmatrix}$$

and using (iii) we get the order statistic as

$$\begin{bmatrix} 8 & 6 & 6 & 5 & 5 & 4 \\ & 6 & 5 & 5 & 3 & 2 \\ & & & 3 & 2 & 2 \\ & & 5 & 4 & 2 & 1 \\ & 7 & 6 & 2 & 1 & 1 \end{bmatrix}$$

An estimator $t(s,\underset{\sim}{y})$ is said to be symmetric if it depends on $\underset{\sim}{y}$ only through the order statistic η_s i.e. if for all $\underset{\sim}{y}$

$$t(\pi s, \pi \underset{\sim}{y}) = t(\eta_s, \underset{\sim}{y}) \quad \forall\, s \in S \text{ and } \forall\, \pi \in \prod .$$

Here, and subsequently, s denotes a set of labels sampled from the 'basic situation' in which a unit labelled (i,j) has value y_{ij}.

A design p defined on S, not necessarily implementable, is said to be symmetric if

$$p(\pi s) = p(s) \quad \forall\, s \in S \text{ and } \forall\, \pi \in \Pi .$$

As mentioned earlier with an implementable design p we associate a design $\bar{p}$ on S given by

$$\bar{p}(s) = \frac{1}{\lambda} \sum_{\pi} p(\pi s) , \quad s \in S .$$

Clearly, by construction itself, $\bar{p}$ is a symmetric design.

Consider the following classes of implementable designs

$\delta^1_{(k,m)}$ = {p : p(s) > 0 ⟹ s contains at most k psu's and at most m ssu's from each of the ($\leq k$) psu's}.

$\delta_{(k,m)}$ = {p : p(s) > 0 ⟹ s contains exactly k psu's and at most m ssu's from each of the k psu's}.

$\rho_{(k,m)}$ = {p : p(s) > 0 ⟹ s contains exactly k psu's and exactly m ssu's from each of the k psu's}.

Given a strategy (p, t) define an estimator $\bar{t}_p$ as follows

$$\bar{t}_p(s,\underset{\sim}{y}) = \frac{1}{\sum_{\pi} p(\pi s)} \sum_{\pi} p(\pi s) t(\pi s, \pi \underset{\sim}{y}) \quad \text{if } \sum_{\pi} p(\pi s) \neq 0 \qquad \ldots(4.1.5)$$

and $\bar{t}_p(s,\underset{\sim}{y})$ may be assigned any arbitrary value if $\sum_{\pi} p(\pi s) = 0$.

Further for a strategy (p, t), $p \in \delta_{(k,m)}$, define

$$t^*_p(s,\underset{\sim}{y}) = \sum_{(n_1,\ldots,n_k)} \frac{p(n_1,\ldots,n_k)}{I(n_1,\ldots,n_k) \prod_{i=1}^{k} \binom{m}{n_i}} \sum_{\substack{s_1 \in S(n_1,\ldots,n_k) \\ s_1 \subset s}} \bar{t}_p(s_1, \underset{\sim}{y}) \qquad \ldots(4.1.6)$$

if $s \in S(m,m,\ldots,m)$,

and $t_p^*(s, \underset{\sim}{y})$ may be assigned arbitrary values for samples not belonging to $S(m,m,\ldots,m)$, where $p(n_1,\ldots,n_k)$, $I(n_1,\ldots,n_k)$ and $\bar{t}_p$ are given by (4.1.1), (4.1.3) and (4.1.5) respectively and $\sum_{(n_1,\ldots,n_k)}$ is a summation over all possible <u>distinct</u> <u>sets</u> of integers $\{n_1,\ldots,n_k\}$, $1 \leq n_i \leq M$, $1 \leq i \leq k$.

A strategy (p,t), p implementable, is said to be p-unbiased or design unbiased for the population mean if

$$\sum_{s \in S} p(s)t(s, \underset{\sim}{y}) = \frac{1}{M_o} \sum_{i=1}^{L} \sum_{j=1}^{M_i} y_{ij} \quad \forall \text{ incomplete } \underset{\sim}{y}$$

where $M_o = \sum_{i=1}^{L} M_i$.

Recall that for a complete $L \times M$ array $\underset{\sim}{y}$ we defined $C = \{\pi_t \underset{\sim}{y} : \pi \in \prod\}$. Now for each of these truncated arrays, (λ in number) $\pi_t \underset{\sim}{y}$;. $\pi \in \prod$, we can think of population mean. The model or ξ-expectation of population mean is just an average of these λ population means. As a matter of fact, what the mode says is that there are λ incomplete populations (not all distinc and the population which we can sample from is one of these incomplete populations occuring with equal chance. Clearly the ξ-expectation of the population mean is given by

$$\theta(\underset{\sim}{y}) = \frac{1}{LM} \sum_{i=1}^{L} \sum_{j=1}^{M} y_{ij} \, . \qquad \ldots(4.1.7)$$

A strategy (p,t), p implementable, is said to be $p\xi$-unbiased or model design unbiased for the population mean if

$$E_\xi E_p(t(s, \underset{\sim}{y})) = \frac{1}{\lambda} \sum_\pi \sum_{s\varepsilon S} p(s)t(s, \pi \underset{\sim}{y}) = \theta(\underset{\sim}{y}) \quad \forall \underset{\sim}{y}$$

where $\theta(y)$ is given by (4.1.7).

We assume a loss function $\ell(a,\theta)$, of real variables a and θ, which is convex in a for all values of θ.

For a strategy (p,t), p implementable, to estimate the population mean, the average risk corresponding to convex loss function $\ell(a,\theta)$ is defined to be

$$\bar{R}(t, p, \underset{\sim}{y}) = \frac{1}{\lambda} \sum_\pi \sum_{s\varepsilon S} p(s)\ell(t(s, \pi \underset{\sim}{y}), \theta(\underset{\sim}{y})) \qquad \ldots(4.1.8)$$

where $\theta(\underset{\sim}{y})$ is given by (4.1.7).

Note that the average risk (4.1.8) is based on $p\xi$ distribution. [Compare with $M_2(p,t) = E_p E_\xi (t - E_\xi \bar{Y})^2$ of (2.1.4) and $M_2(p,t) = E_p E_\xi (t - \mu_Y)^2$ of (5.1.5)].

For squared error loss function it would be

$$\frac{1}{\lambda} \sum_\pi \sum_{s\varepsilon S} p(s) [t(s, \pi \underset{\sim}{y}) - \theta(\underset{\sim}{y})]^2.$$

The risk based on $p\xi$-distribution has been used in the literature by various authors e.g. Särndal (1978, 1980,a) ; Cassel et al (1977), J.N.K. Rao and Bellhouse (1978), just to mention a few.

4.2 Estimation of Population Mean

In the preceding section we introduced a transition from incomplete set up to complete set up. Our problem is to estimate the population mean efficiently, efficiency being judged by the average risk (4.1.8) based on $p\xi$-distribution. If the sampler can afford to select at most $k(\leq L)$ psu's and at most m ssu's from each of the psu's selected at the first stage then what is a 'reasonable' strategy for him? In this section we attempt to answer this question.

We first state a result regarding completeness of order statistic η_s for the two stage simple random sampling.

<u>Theorem 4.2.1</u> (Bellhouse et al, 1977). For the design that select k psu's by simple random sampling without replacement and then selects m ssu's again by simple random sampling without replacement from each of the sampled primaries independently order statist is complete. i.e. if $\phi(\eta_s, \underset{\sim}{y})$ is a function with zero sampling expectation for all possible $\underset{\sim}{y}$ then it must be identically zero.

We use the above result to prove the completeness of the order statistic η_s in the $p\xi$-sense, but now for any sampling design $q \in \rho_{(k,m)}$.

<u>Theorem 4.2.2</u>. The order statistic η_s is complete in $p\xi$-sense fo any sampling design $q \in \rho_{(k,m)}$ i.e. if for a design $q \in \rho_{(k,m)}$, $\phi(\eta_s, \underset{\sim}{y})$ is a function of order statistic with zero $p\xi$-expectatio for all possible $\underset{\sim}{y}$ then it must be identically zero.

<u>Proof</u>. Note that for a $q \in \rho_{(k,m)}$,

$$E_q E_\xi(\phi(\eta_s , \underset{\sim}{y})) = \frac{1}{\lambda} \sum_\pi \sum_{s\varepsilon S} q(s)\phi(\eta_s , \pi \underset{\sim}{y})$$

$$= \frac{1}{\lambda} \sum_{s\varepsilon S(m,\ldots,m)} q(s) \sum_\pi \phi(\eta_s , \pi \underset{\sim}{y})$$

$$= \frac{1}{\lambda} \sum_{s\varepsilon S(m,m,\ldots,m)} q(s) \sum_{s\varepsilon S(m,m,\ldots,m)} \phi(\eta_s , \underset{\sim}{y}) \frac{\lambda}{\binom{L}{k}\binom{M}{m}^k}$$

$$= \frac{1}{\binom{L}{k}\binom{M}{m}^k} \sum_{s\varepsilon S(m,m,\ldots,m)} \phi(\eta_s , \underset{\sim}{y})$$

The completeness of order statistic, in the $p\xi$-sense, for any $q \varepsilon \rho_{(k,m)}$, now follows from the fact that the order statistic is complete for the two stage simple random sampling in the p-sense (Theorem 4.2.1).

Note that the definitions of $p\xi$-expectation and average risk can very well be applied to designs which are not implementable. This is easy to see because, afterall, any implementable design is defined on S only. With this in mind we prove the following results.

<u>Lemma 4.2.1</u> For all $\underset{\sim}{y}$ and an implementable strategy (p,t),

$$\bar{R}(\bar{t}_p , p , \underset{\sim}{y}) \le \bar{R}(t , p , \underset{\sim}{y}) \quad \ldots(4.2.1)$$

and $$\bar{R}(\bar{t}_p , \bar{p} , \underset{\sim}{y}) = \bar{R}(\bar{t}_p , p , \underset{\sim}{y}) \quad \ldots(4.2.2)$$

<u>Proof</u>. $\lambda\bar{R}(\bar{t}_p , p , \underset{\sim}{y}) = \sum_\pi \sum_{s\varepsilon S} p(s)\ell[\bar{t}_p(s,\pi \underset{\sim}{y}), \theta(\underset{\sim}{y})]$

$$\sum_{k=1}^{L} \sum_{(n_1,\ldots,n_k)} \sum_{s\varepsilon S(n_1,\ldots,n_k)} p(s) \sum_\pi \ell[\bar{t}_p(s, \pi \underset{\sim}{y}), \theta(\underset{\sim}{y})]$$

$$= \sum_{k=1}^{L} \sum_{(n_1,\ldots,n_k)} \sum_{s\varepsilon S(n_1,\ldots,n_k)} p(s) \sum_{s\varepsilon S(n_1,\ldots,n_k)} \ell[\bar{t}_p(s,\underset{\sim}{y}),\theta(\underset{\sim}{y})] \frac{\lambda}{c(n_1,n_2,\ldots n_k)}$$

$$= \lambda \sum_{k=1}^{L} \sum_{(n_1,\ldots,n_k)} \sum_{s\varepsilon S(n_1,\ldots,n_k)} \bar{p}(s)\, \ell[\bar{t}_p(s,\underset{\sim}{y}),\theta(\underset{\sim}{y})]$$

$[= \lambda \bar{R}(\bar{t}_p, \bar{p}, \underset{\sim}{y})$, which proves (4.2.2)]

$$\leq \sum_{k=1}^{L} \sum_{(n_1,\ldots,n_k)} \sum_{s\varepsilon S(n_1,\ldots,n_k)} \lambda\bar{p}(s) \sum_{\pi} \frac{p(\pi s)\ell[t(\pi s,\pi y),\theta(y)]}{\sum_{\pi} p(\pi s)}$$

(since ℓ is con

$$= \sum_{\pi} \sum_{k=1}^{L} \sum_{(n_1,\ldots,n_k)} \sum_{s\varepsilon S(n_1,\ldots,n_k)} p(s)\, \ell[t(s,\pi \underset{\sim}{y}),\theta(\underset{\sim}{y})]$$

$$= \lambda \bar{R}(t, p, \underset{\sim}{y}) .$$

This proves (4.2.1).

Though we started with an implementable design p, we can conceive of design $\bar{p}$. Lemma 4.2.1 says that symmetric estimators fare better than corresponding non-symmetric estimators and for a symmetric estimator there is no additional gain by symmetrizing the accompanying design.

We now prove,

<u>Lemma 4.2.2</u> For a strategy (p,t), $p \varepsilon \delta_{(k,m)}$, and a design $q \varepsilon \rho_{(k,m)}$, we have,

$$\bar{R}(t_p^*, q, \underset{\sim}{y}) \leq \bar{R}(t, p, \underset{\sim}{y}) \quad \forall\, \underset{\sim}{y} \qquad \ldots(4.2.3$$

<u>Proof</u>: Let $p_{(k,m)}$ be two stage simple random sampling, on the complete $L \times M$ set up, that selects exactly k psu's at first

stage by simple random sampling and exactly m ssu's, at second stage, from each of the psu's selected, by simple random sampling, independently. Let $s(k,m)$ be a typical sample that contains m ssu's each from k different psu's. Using convexity of the loss function $\ell(a,\theta)$, we have,

$$\ell[t_p^*(s(k,m),\underset{\sim}{y}),\theta(y)] \leq \sum_{(n_1,\ldots,n_k)} \frac{p(n_1,\ldots,n_k)}{I(n_1,\ldots,n_k)\prod_{i=1}^{k}\binom{M}{n_i}} \sum_{\substack{s\in S(n_1,\ldots,n_k)\\ s\subset s(k,m)}} \ell[\bar{t}_p(s,\underset{\sim}{y}),\theta(\underset{\sim}{y})]$$

where $\bar{t}_p$ and t_p^* are given by (4.1.5) and (4.1.6) respectively. Multiplying both sides by $\frac{1}{c(m,\ldots,m)}$ and taking summation over all possible samples $s(k,m)\in S(m,\ldots,m)$ and noting that in the summation a sample $s\in S(n_1,\ldots,n_k)$ occurs $\prod_{i=1}^{k}\binom{M-n_i}{m-n_i}$ times, we obtain,

$$\frac{1}{c(m,\ldots,m)} \sum_{s\in S(m,\ldots,m)} \ell[t_p^*(s,\underset{\sim}{y}),\theta(\underset{\sim}{y})] \leq \sum_{(n_1,\ldots,n_k)} \sum_{s\in S(n_1,\ldots,n_k)} \bar{p}(s)\ell[\bar{t}_p(s,\underset{\sim}{y}),\theta(\underset{\sim}{y})]$$

This proves that

$$\bar{R}(t_p^*, p_{(k,m)}, \underset{\sim}{y}) \leq \bar{R}(\bar{t}_p, \bar{p}, \underset{\sim}{y}) \quad \forall\, \underset{\sim}{y}$$

Now using (4.2.2), for a design $q \in \rho_{(k,m)}$

$$\bar{R}(t_p^*, q, \underset{\sim}{y}) = \bar{R}(t_p^*, p_{(k,m)}, \underset{\sim}{y}) \quad \forall\, \underset{\sim}{y}$$

since $\bar{q} = p_{(k,m)}$.

Thus $\bar{R}(t_p^*, q, \underset{\sim}{y}) \leq \bar{R}(\bar{t}_p, \bar{p}, \underset{\sim}{y}) \quad \forall \underset{\sim}{y}$.

This completes the proof of the lemma.

It is easy to see that if a strategy (p,t) ; $p \in \delta_{(k,m)}$ is $p\xi$-unbiased for the population mean then a strategy (q,t_p^*) also $p\xi$-unbiased for the population mean $\forall\, q \in \rho_{(k,m)}$.

Let $\bar{y}$ be overall sample mean and $H_{(k,m)}$ be the class of all $p\xi$-unbiased sampling strategies (p,t), $p \in \delta_{(k,m)}$ i.e.

$$H_{(k,m)} = \{(p,t) : p \in \delta_{(k,m)} \text{ and } E_p E_\xi(t) = \theta(\underset{\sim}{y})\}.$$

We now have the following theorem,

<u>Theorem 4.2.3</u>. For estimating the population mean, $(q,\bar{y})$ is best strategy in $H_{(k,m)}$ in the sense that

$$\bar{R}(\bar{y}, q, \underset{\sim}{y}) = \min_{(p,t) \in H_{(k,m)}} \bar{R}(t, p, \underset{\sim}{y}) \quad \forall \underset{\sim}{y}$$

where q is any design in $\rho_{(k,m)}$.

<u>Proof</u>. The proof follows from Theorem 4.2.2 and Lemmas 4.2.1 and 4.2.2.

We prove yet another lemma before going to our main theorem.

<u>Lemma 4.2.3</u>.

$$\bar{R}(\bar{y}, p_{(k,n)}, \underset{\sim}{y}) \leq \bar{R}(\bar{y}, p_{(k-1,m)}, \underset{\sim}{y}) \quad \forall \underset{\sim}{y} \quad \ldots(4.2.$$

<u>Proof</u>. $\bar{y}(s(k,m),\underset{\sim}{y})$ can be expressed as an average of $\bar{y}(s(k-1,m),\underset{\sim}{y})$ as follows,

$$\bar{y}(s(k,m)\,,\,\underset{\sim}{y}) = \frac{1}{k} \sum_{s(k-1,m) \subset s(k,m)} \bar{y}(s(k-1,m)\,,\,\underset{\sim}{y})$$

Now by convexity of the loss function $\ell(a,\theta)$, we have,

$$\ell\left[\bar{y}(s(k,m),\underset{\sim}{y}),\theta(\underset{\sim}{y})\right] \leq \frac{1}{k} \sum_{s(k-1,m) \subset s(k,m)} \ell\left[\bar{y}(s(k-1,m),\underset{\sim}{y}),\theta(\underset{\sim}{y})\right]$$

Hence multiplying both sides by $\frac{1}{\binom{L}{k}\binom{M}{m}^k}$, taking summation over all samples $s(k,m)$ and noting that a sample $s(k-1,m)$ occurs $(L-k+1)\binom{M}{m}$ times, we obtain,

$$\bar{R}(\bar{y}\,,\,p_{(k,m)}\,,\,\underset{\sim}{y}) \leq \bar{R}(\bar{y}\,,\,p_{(k-1,m)}\,,\,\underset{\sim}{y}) \quad \forall\, \underset{\sim}{y}\,.$$

Hence the lemma.

We now have our main result.

Let $\quad H^1_{(k,m)} = \{(p,t) : p \in \delta^1_{(k,m)}$ and $E_p E_t(t) = \theta(\underset{\sim}{y})\}\,.$

<u>Theorem 4.2.4</u>. For estimating the population mean, $(q,\bar{y})$ is a best strategy in $H^1_{(k,m)}$ in the sense that for all $\underset{\sim}{y}$ it minimizes the average risk i.e.

$$\bar{R}(\bar{y}\,,\,q\,,\,\underset{\approx}{y}) = \min_{(p,t) \in H^1_{(k,m)}} \bar{R}(t\,,\,p\,,\,\underset{\sim}{y}) \quad \forall\, \underset{\sim}{y}$$

where q is any design in $\rho_{(k,m)}$.

Theorem 4.2.4 can be interpreted as follows :
If on certain considerations like resources, precision etc., the sampler can take at most k psu's and at most m ssu's from each

of the $(\leq k)$ psu's then the overall sample mean coupled with any fixed size design in $\rho_{(k,m)}$ is a best strategy in the class of pξ-unbiased strategies, for estimating the population mean, in the sense of minimum average risk.

Let $p_o \varepsilon \rho_{(k,m)}$ be a design that selects k psu's with inclusion probabilities proportional to their sizes, $(M_i, 1 \leq i \leq L)$, and m ssu's, within each of them, by simple random sampling independently. For this design p_o sample mean is p-unbiased for the population mean. Hence we have the following corollary.

<u>Corollary 4.2.1</u>. For estimating the population mean, the strategy $(p_o, \bar{y})$ is the best p-unbiased strategy, in the subclass of $H^1_{(k,m)}$ of strategies that are p-unbiased for the population mean, in the sense of minimum average risk.

<u>Proof</u>. Follows from Theorem 4.2.4 and the fact that $(p_o, \bar{y})$ is p-unbiased for the population mean.

For a given sample s that has ssu's from k psus let

$$\bar{\bar{y}} = \frac{1}{k} \sum_{i=1}^{k} \bar{y}_i \qquad \ldots(4.2.5)$$

where $\bar{y}_i$ is the mean per ssu for the ith psu in the sample. $\bar{\bar{y}}$ is same as $\bar{y}$, the over all sample mean, if the sample is from $S(m,m,\ldots,m)$ i.e. it contains equal number of ssu's from different psu's sampled.

Let, for $1 \leq k < L$, $1 \leq m < \min_{1 \leq i \leq L} M_i$, $0 \leq \theta_1, \theta_2 \leq 1$,

$\delta^{a}_{(k+\theta_1, m+\theta_2)}$ be the class of designs, implementable, that select on an average $k+\theta_1$ psu's and $m+\theta_2$ ssu's, on an average, from each of the psu's sampled, independently. Further let

$$H^{a}_{(k+\theta_1, m+\theta_2)} = \{(p,\bar{\bar{y}}) : p \in \delta^{a}_{(k+\theta_1, m+\theta_2)}\}.$$

Theorem 4.2.5. $(q,\bar{\bar{y}})$ is a best strategy in $H^{a}_{(k+\theta_1, m+\theta_2)}$ in the sense of minimum average risk for the squared error loss function, where q is any implementable design in $\delta^{a}_{(k+\theta_1, m+\theta_2)}$ for which the associated hypothetical design (4.1.2) is

$$p^{*}_{(k+\theta_1, m+\theta_2)} = (1-\theta_1)\left[(1-\theta_2)p_{(k,m)} + \theta_2 p_{(k,m+1)}\right] + \theta_1\left[(1-\theta_2)p_{(k+1,m)} + \theta_2 p_{(k+1,m+1)}\right]. \quad \ldots(4.2.6)$$

First we state a lemma due to Stenger and Gabler (1981).

A sequence $h_1, h_2, \ldots, h_N$ of real numbers is said to be strictly convex if for $i = 2,3,\ldots,N-1$

$$2h_i < h_{i-1} + h_{i+1}.$$

Lemma 4.2.4. (Stenger and Gabler, 1981). For a strictly convex sequence $h_1, h_2, \ldots, h_N$ and a probability vector $\underset{\sim}{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_N)$

$$\sum_{i=1}^{N} \alpha_i h_i \geq \sum_{i=1}^{N} \alpha_i^{*} h_i. \quad \ldots(4.2.7)$$

Equality holds if and only if $\underset{\sim}{\alpha} = \underset{\sim}{\alpha}^{*}$.

[$\underset{\sim}{\alpha}^{*}$ is given by (3.1.3)].

Proof of Theorem 4.2.5. In proving the theorem it is enough to consider complete $L \times M$ array and two stage symmetric designs on that because of (4.2.2).

A typical symmetric two stage design p, for $L \times M$ complet array, may be thought of as - for the first stage a probability mixture $(\alpha_1, \alpha_2, \ldots, \alpha_L)$ of symmetric designs that select exactly $1, 2, \ldots, L$ psu's respectively and for the second stage, within each selected primary, a probability mixture $(\beta_1, \beta_2, \ldots, \beta_M)$ of symmetric designs that select exactly $1, 2, \ldots, M$ ssu's respectivel

Let p be a two stage symmetric design for $L \times M$ complete array such that

$$\sum_{i=1}^{L} i\,\alpha_i = k+\theta_1 \quad \text{and} \quad \sum_{j=1}^{M} j\beta_j = m+\theta_2 .$$

It is easy to evaluate $\bar{R}(\bar{\bar{y}}, p, \underset{\sim}{y})$ for the squared error loss function and is given by,

$$\bar{R}(\bar{\bar{y}}, p, \underset{\sim}{y}) = \sum_{i=1}^{L} \alpha_i \left[\left(\frac{1}{i} - \frac{1}{L}\right) S_b^2 + \frac{S_w^2}{i} \sum_{j=1}^{M} \beta_j \left(\frac{1}{j} - \frac{1}{M}\right) \right]$$

where $S_b^2 = \frac{1}{L-1} \sum_{i=1}^{L} (\bar{Y}_i - \bar{Y})^2$, $\bar{Y}_i = \frac{1}{M} \sum_{j=1}^{M} y_{ij}$, $\bar{Y} = \frac{1}{ML} \sum_{i=1}^{L} \sum_{j=1}^{M} y_{ij}$,

$$S_w^2 = \frac{1}{L} \sum_{i=1}^{L} S_i^2 , \qquad S_i^2 = \frac{1}{M-1} \sum_{j=1}^{M} (y_{ij} - \bar{Y}_i)^2 .$$

Observe that the sequence $1, \frac{1}{2}, \frac{1}{3}, \ldots, \frac{1}{N}$, for any integer N is strictly convex, for

$$\frac{1}{r-1} + \frac{1}{r+1} = \frac{2r}{r^2-1} > \frac{2}{r} \quad \forall\, r > 1 .$$

Hence using (4.2.7) of Lemma 4.2.4, we get

$$\bar{R}(\bar{\bar{y}}, p, \underset{\sim}{y}) \leq \sum_{i=1}^{L} \alpha_i^* \left[\left(\frac{1}{i} - \frac{1}{L}\right) S_b^2 + \frac{S_w^2}{i} \sum_{j=1}^{M} \beta_j^* \left(\frac{1}{j} - \frac{1}{M}\right) \right]$$

and the equality is attained if and only if $p = p^*_{(k+\theta_1, n+\theta_2)}$ of (4.2.6).

This completes the proof of the theorem.

Thus in this chapter we have given a subjective justification to use, for symmetric estimators, the classical formulae based on symmetric designs, but now for any design, when the only prior knowledge available is the 'exchangeability' of the unknown variate values of the units.

CHAPTER 5

ESTIMATION UNDER A CONTINUOUS SURVEY SAMPLING MODEL

5.0 Summary

In this chapter we take up a study under a continuous survey sampling model. We first show that there does not exist, in general, a globally optimal p-unbiased strategy under the proposed model in the sense of minimum expected variance. We then suggest an alternative criterion to obtain a 'reasonable' strategy and investigate the properties of this 'reasonable' strategy. Next we compare certain sampling strategies under the proposed model. Finally, we deal with the stratified set up and compare a few more strategies. A few related optimality results are obtained in Chapter 8.

5.1 Introduction

The main objective of this chapter is to present a different analytical treatment of sampling and estimation for appropriate models. Assuming that N, the size of the finite population, is very large, a stage is set for moving from the finite set up to a continuous framework. A continuous variable formulation is substituted for the usually cumbersome finite population algebra. Such formulations (Cassel and Särndal 1972, 1974 and Särndal 1980), are an attempt to steer Godambe's survey sampling set up in continuous terms. Such an interpretation makes it easier to grasp some of the complexities of modern survey sampling theory; exact efficiencies of various sampling strategies can often be computed.

Consider a population of infinitely many pairs $(y(x),x)$, $x \geq 0$, such that the joint distribution of $y(x)$, $x \geq 0$, is unknown. For convenience we assume that $y(x)$, $x \geq 0$ are defined on some probability space (Ω, A, ξ). We further assume that the distribution of X, whose observed values are x, is continuous. Let $F(x)$ be the probability, assumed known for every $x \geq 0$, that X does not exceed x and let

$$F(x) = \int_0^x f(u)du .$$

In the continuous set up, the label of a population unit is a continuous index λ, where for convenience $\lambda \in [0,1]$. A more specific ordering is imposed on λ by identifying the label λ with the λth quantile of the X distribution, so that $\lambda = F(x)$

or $x = F^{-1}(\lambda)$. Having drawn and observed one unit the data is recorded as $(y_1(\lambda),\lambda)$ or, equivalently, as $(y(x),x)$. In the latter case, the unit drawn is identified by its x-value. The two functions $y_1(\cdot)$ and $y(\cdot)$ are related by $y_1(\lambda) = y(h(\lambda))$, where $h(\lambda) = F^{-1}(\lambda)$ is the inverse function of $F(x)$. The problem under consideration is to estimate the population mean for the variate Y, namely,

$$m_Y = E_f(Y) = \int_0^\infty y(x)f(x)dx = \int_0^1 y(h(\lambda))d\lambda. \qquad \ldots(5.1.1)$$

This, incidentally, defines the operator E_f.

In the continuous set up in the presence of auxiliary variabl x, members of a sample of size n are identified by the vector of labels drawn, $\underset{\sim}{\lambda} = (\lambda_1,\lambda_2,\ldots,\lambda_n)$, where $0 \leq \lambda_i \leq 1$, $1 \leq i \leq n$, or by the corresponding vector $\underset{\sim}{x} = (x_1,x_2,\ldots,x_n)$, where $\lambda_i = F(x_i)$, $1 \leq i \leq n$. The y-value is observed for each of the labels drawn and the data is recorded as $(y(x_i),x_i)$, $1 \leq i \leq n$, or $(y(\underset{\sim}{x}),\underset{\sim}{x})$.

Let $\mathcal{B}$ be the Borel σ-algebra of $R_n^+ = \{\underset{\sim}{x} : x_i \geq 0,\ 1 \leq i \leq n\}$. Any continuous probability measure Q on $\mathcal{B}$ is called a <u>sampling design</u>. $Q(\underset{\sim}{x})$ is the probability of drawing a sample such that the auxiliary variate value in the ith draw does not exceed x_i, $1 \leq i \leq n$. Let $q(\underset{\sim}{x}) = dQ(\underset{\sim}{x}) / d\underset{\sim}{x}$. Then $q(\underset{\sim}{x})$ can be expressed as $q(\underset{\sim}{x}) = p(\underset{\sim}{x})f(\underset{\sim}{x})$, where $f(\underset{\sim}{x}) = \prod_{i=1}^{n} f(x_i)$. We shall refer to the function $p(\underset{\sim}{x})$ introduced here as the <u>design function</u> giving rise to the sampling design $Q(\underset{\sim}{x})$. Since a unit can be identified by its x-value, the sampler can choose the function $p(\underset{\sim}{x})$ suitably.

Now, the case where $p(\underset{\sim}{x}) \equiv 1 \;\; \forall \underset{\sim}{x} \in R_n^+$, which corresponds to the mathematical statistician's standard assumption of 'independent and identically distributed' observations x_i, $1 \leq i \leq n$, represents unrestricted simple random sampling from the population and then $q(\underset{\sim}{x}) \equiv f(\underset{\sim}{x})$. Otherwise, $q(\underset{\sim}{x}) \neq f(\underset{\sim}{x})$. In general, $q(\underset{\sim}{x})$ is the density generated by the randomization device $p(\underset{\sim}{x})$. Godambe and Thompson (1971) used the term 'frequency distribution generated by randomization' with a similar connotation.

The marginal density of the ith component of $\underset{\sim}{x}$ is given by

$$q_i(x_i) = \int_{R^+_{n-1}} q(\underset{\sim}{x}) \prod_{j \neq i}^{n} dx_j \, , \quad 1 \leq i \leq n . \qquad \ldots(5.1.2)$$

The expected x-value in draw i is $\int_0^\infty x q_i(x) dx$. Särndal (1980) gave an example of a design function $p(\underset{\sim}{x})$ that produces draws with correlated labels.

In this chapter, we consider a specific superpopulation model induced by the probability space $(\underline{\Omega}, A, \xi)$. This model is essentially a class of distributions satisfying the following

$$Y(x) = \beta x^b + Z(x) \;, \; x \geq 0$$

where for every fixed $x \geq 0$

$$E_\xi(Z(x)) = 0 \;\; \text{and} \;\; E_\xi(Z^2(x)) = \sigma^2 x^g \qquad \ldots(5.1.3)$$

and for every fixed $x \neq x' \; ; \; x, x' \geq 0$

$$E_\xi(Z(x) \cdot Z(x')) = 0 \, ,$$

where $\sigma^2 > 0$ and β are unknown whereas $b \geq 0$ and $g \varepsilon [0,2]$ may be known or unknown.

Any particular infinite population value, $y(x) = \beta x^b + z(x)$, is to be interpreted as βx^b plus a realization of the random variable $Z(x)$.

Any function t of the observed data, $(y(\underset{\sim}{x}),\underset{\sim}{x})$, is called an estimator of m_Y, the population mean (5.1.1), whereas (p,t), a design function p together with an estimator t, is called a strategy. Note that specifying p is same as specifying the corresponding sampling design; therefore in a strategy we may specify the design function or the corresponding sampling design.

We now give the definitions of various types of unbiasedness.

(a) A strategy (p,t) is said to be p-unbiased (design-unbiased) if

$$E_p(t) = \int_{R_n^+} t(y(\underset{\sim}{x}),\underset{\sim}{x})p(\underset{\sim}{x})f(\underset{\sim}{x})d\underset{\sim}{x} = \int_{R^+} y(x)f(x)dx = m_Y$$

for every real valued F-integrable function $y(x)$.

This, incidentally, defines the operator $E_p(\cdot)$.

(b) A strategy (p,t) is said to be ξ-unbiased (model unbiased) if

$$E_\xi(t(Y(\underset{\sim}{x}),\underset{\sim}{x}) - m_Y) = 0 \quad \text{a.e. } [Q]$$

where Q is the sampling design corresponding to the design function $p(\underset{\sim}{x})$.

(c) A strategy (p,t) is said to be $p\xi$-unbiased (model-design-unbiased) if

$$E_p E_\xi (t(Y(\underset{\sim}{x}), \underset{\sim}{x}) - E_\xi (m_Y)) = 0 .$$

For the comparison of various sampling strategies we introduce the following measures of uncertainty. For a strategy (p,t) define

$$M_1(p,t) = E_\xi E_p (t - m_Y)^2 \quad \ldots(5.1.4)$$

and $$M_2(p,t) = E_p E_\xi (t - \mu_Y)^2 \quad \ldots(5.1.5)$$

where $$\mu_Y = E_\xi (m_Y) = E_\xi \int_0^\infty y(x) f(x) dx . \quad \ldots(5.1.6)$$

We assume that $Y(X)$ is square integrable w.r.t. the product probability $(F \times \xi)$. In actually computing $M_1(p,t)$ or $M_2(p,t)$, we, in fact assume that the population conforms to the model (5.1.3) with $b = 1$ and $g \in [0, 2]$ may be known or unknown. We also assume that F, the distribution function of X, is a gamma distribution with parameter α.

We also consider stratified sampling involving L strata. A unit is said to belong to hth stratum if its x-value belongs to the interval $[x_{h-1}, x_h)$, where $0 = x_0 < x_1 < x_2 < \ldots < x_L = \infty$ are the given stratification points. For the stratified sampling we have to modify our basic set up accordingly. For the hth stratum define $f_h(x)$, the restriction of $f(x)$ to the interval $[x_{h-1}, x_h)$, analogous to $f(x)$ on R^+, as

$$f_h(x) = f(x)/W_h \quad \text{if } x \in [x_{h-1}, x_h)$$

$$= 0 \qquad \text{otherwise}$$

where $W_h = F(x_h) - F(x_{h-1}) = \int_{x_{h-1}}^{x_h} f(u)du$.

One can now think of a design function $p_h(\underset{\sim}{x})$ for the hth stratum, where $\underset{\sim}{x} = (x_1, x_2, \ldots, x_{n_h})$; n_h being the number of units to be selected from the hth stratum. The sampling design can now be defined as $q_h(\underset{\sim}{x}) = p_h(\underset{\sim}{x}) f_h(\underset{\sim}{x})$ where $f_h(\underset{\sim}{x}) = \prod_{i=1}^{n_h} f_h(x_i)$. The overall stratified sampling design is then given by $\prod_{h=1}^{L} q_h(\underset{\sim}{x})$.

We will be using some of the designs and estimators quite often in our discussions, which are given below.

Simple random sampling (srs) corresponds to the design function $p(\underset{\sim}{x}) \equiv 1$. $ppx^{g/2}$ is the sampling design that corresponds to the design function $\prod_{i=1}^{n} p_i(x_i)$ with $p_i(x) \propto x^{g/2}$. The continuous analogue of Midzuno-Sen sampling scheme corresponds to the design function $p(\underset{\sim}{x}) = \sum_{i=1}^{n} x_i / n\mu$, where $\mu = E_f(X)$, let us denote this by p_M.

Given a sampling design $q(\underset{\sim}{x})$ the Horvitz-Thompson estimator takes the form

$$t_{HT} = \frac{1}{n} \sum_{i=1}^{n} y(x_i) f(x_i) / q_i(x_i) \qquad \ldots(5.1.7)$$

Finally, the usual ratio estimator takes the form

$$t_R = \mu \sum_{i=1}^{n} y(x_i) / \sum_{i=1}^{n} x_i . \qquad \ldots(5.1.8)$$

As mentioned earlier, the main purpose of the continuous variable formulation is to understand some of the complexities of the finite population sampling theory. In this chapter we first obtain continuous versions of the results derived in Section 3 of Chapter 2. We then compare certain sampling strategies under the model (5.1.3). We finally deal with the stratified set up. Some of the related results will be obtained in Chapter 8.

5.2 p-unbiasedness of Optimal ξ-unbiased Estimator

In this section we first note that an attempt to minimize $M_1(p,t)$ in the class of all linear p-unbiased strategies fails unless $g=2$ in the model (5.1.3). (As mentioned earlier we assume that the population conforms to the model (5.1.3) with $b=1$ whereas $g \in [0,2]$ may be known or unknown). To strike a compromise we try to minimize $M_1(p,t)$ in the class of all linear ξ-unbiased estimators and then obtain a design that makes an estimator so obtained p-unbiased. We further show that such a strategy is as good as the strategy (ppx, t_{HT}).

We had adopted the same approach in Chapter 2. However there we minimized, as the first step, a part of $M_1(p,t)$ (or equivalently $M_2(p,t)$) subject to ξ-unbiasedness and then obtained a design that made the estimator so obtained p-unbiased. Here we are minimizing the entire quantity $M_1(p,t)$ in the class of linear ξ-unbiased estimators. The next step is, of course, to obtain an appropriate design. Further our compromise strategy happens to be an analogue of its counter-part in the finite set up. An additional advantage over the finite set up is that we are able to find the functional form of the design that makes the best ξ-unbiased estimator p-unbiased. This, when the finite population size, N, is very large, justifies the approach adopted in Chapter 2.

A linear estimator t is of the type

$$t(y(\underset{\sim}{x}),\underset{\sim}{x}) = \sum_{i=1}^{n} a_i(\underset{\sim}{x})y(x_i) \qquad \ldots(5.2.1)$$

where $a_1, a_2, \ldots, a_n$ are $\mathcal{B}$-measurable functions.

The condition of p-unbiasedness of a linear strategy (p,t) can be given by

$$\sum_{i=1}^{n} \phi_i(x) = 1 \quad \text{a.e. } (F) \qquad \ldots(5.2.2)$$

$$\text{where } \phi_i(x_i) = \int_{R^+_{n-1}} a_i(\underset{\sim}{x})p(\underset{\sim}{x}) \prod_{j \neq i}^{n} f(x_j)dx_j \, , \; 1 \leq i \leq n . \qquad \ldots(5.2.3)$$

If $\phi_i(x)$ is same for all $i = 1,2,\ldots,n$ say $\phi(x)$, and also continuous then the condition of p-unbiasedness reduces to

$$\phi(x) = \frac{1}{n} \; ; \; x \geq 0 . \qquad \ldots(5.2.4)$$

Now for a p-unbiased strategy (p,t)

$$M_1(p,t) = E_p E_\xi t^2 - E_\xi m_Y^2 .$$

Under the model (5.1.3),

$$E_p E_\xi t^2 = \sigma^2 \int_{R^+_n} \left[\sum_{i=1}^{n} a_i^2(\underset{\sim}{x}) x_i^g \right] p(\underset{\sim}{x}) f(\underset{\sim}{x}) d\underset{\sim}{x}$$

$$+ \beta^2 \int_{R_n} \left[\sum_{i=1}^{n} a_i(\underset{\sim}{x}) x_i \right]^2 p(\underset{\sim}{x}) f(\underset{\sim}{x}) d\underset{\sim}{x} .$$

Let us first try to minimize

$$\int_{R_n^+} \left[\sum_{i=1}^{n} a_i^2(\underset{\sim}{x}) x_i^g \right] p(\underset{\sim}{x}) f(\underset{\sim}{x}) d\underset{\sim}{x} \qquad \ldots(5.2.5)$$

subject to the condition (5.2.4).

Observe that by Cauchy-Schwartz inequality,

$$\int_{R_{n-1}^+} a_i^2(\underset{\sim}{x}) p(\underset{\sim}{x}) \prod_{j \neq i}^{n} f(x_j) dx_j \int_{R_{n-1}^+} p(\underset{\sim}{x}) \prod_{j \neq i}^{n} f(x_j) dx_j$$

$$\geq \left[\int_{R_{n-1}^+} a_i(\underset{\sim}{x}) p(\underset{\sim}{x}) \prod_{j \neq i}^{n} f(x_j) dx_j \right]^2$$

$$= \frac{1}{n^2} \text{, using (5.2.4)}$$

or $\displaystyle \int_{R_{n-1}^+} a_i^2(\underset{\sim}{x}) p(\underset{\sim}{x}) \prod_{j \neq i}^{n} f(x_j) dx_j \geq \frac{f(x_i)}{n^2 q_i(x_i)}$

where $q_i(x_i)$ is given by (5.1.2),

and the equality is attained when $\displaystyle a_i(\underset{\sim}{x}) = \frac{f(x_i)}{n q_i(x_i)} \qquad \ldots(5.2.6)$

Now

$$\int_{R_n^+} \left[\sum_{i=1}^{n} a_i^2(\underset{\sim}{x}) x_i^g \right] p(\underset{\sim}{x}) f(\underset{\sim}{x}) d\underset{\sim}{x} = \sum_{i=1}^{n} \int_{R^+} x_i^g \int_{R_{n-1}^+} a_i^2(\underset{\sim}{x}) p(\underset{\sim}{x}) \prod_{j \neq i}^{n} f(x_j) dx_j f(x_i) dx_i$$

$$\geq \sum_{i=1}^{n} \int_{R^+} x_i^g \frac{f^2(x_i)}{n^2 q_i(x_i)} dx_i \qquad \ldots(5.2.7)$$

and the equality is attained for $a_i(\underset{\sim}{x})$, $1 \leq i \leq n$, given by (5.2.6).

Now note that (5.2.7) depends on $q_i(x_i)$, $1 \leq i \leq n$, hence there is a scope to minimize it further.

Again using Cauchy-Schwartz inequality

$$\int_{R^+} x^g \frac{f^2(x)}{q_i(x)} dx \int_{R^+} q_i(x)dx \geq \left[\int_{R^+} x^{g/2} f(x)dx \right]^2$$

$$\text{or} \quad \int_{R^+} x^g \frac{f^2(x)}{q_i(x)} dx \geq \left[E_f(X^{g/2}) \right]^2$$

and the equality is attained for $q_i(x) = x^{g/2} f(x) / E_f(X^{g/2})$.

Thus (5.2.5) is minimized, subject to the condition (5.2.4), for the strategy $(ppx^{g/2}, t_{HT})$.

Unfortunately, the strategy $(ppx^{g/2}, t_{HT})$ does not minimize the second term in the expression for $E_p E_\xi t^2$, unless $g=2$. On the other hand there are certain strategies for which the second term is minimized but they do not minimize the first term. Thus for $g < 2$ we opt for a strategy obtained using new approach and for $g=2$ we have the following theorem.

<u>Theorem 5.2.1</u>. Under the model (5.1.3), with $b=1$ and $g=2$, (ppx, t_{HT}) is the best strategy in the class of all linear p-unbiased strategies w.r.t. either measure of uncertainty $M_1(p,t)$ or $M_2(p,t)$.

<u>Proof</u>. The proof is immediate. Note that for a p-unbiased strategy $M_1(p,t)$ and $M_2(p,t)$ differ only by a constant term independent of the strategy.

As mentioned before for $g < 2$ we try to minimize $M_1(p,t)$ subject to the condition of ξ-unbiasedness and then try to obtain a design which makes the estimator so obtained p-unbiased. Since we have assumed the square integrability of $Y(X)$ w.r.t. the product probability $F \times \xi$ the order of integration, in the following steps, can be interchanged.

The condition of ξ-unbiasedness, under model (5.1.3) with $b=1$, for the linear estimator (5.2.1) can be given by

$$\sum_{i=1}^{n} a_i(\underset{\sim}{x})x_i = \mu = E_f(X) \quad \forall\ \underset{\sim}{x} \in R_n^+ \qquad \ldots(5.2.8)$$

Note that $E_\xi E_p(t-m_Y)^2 = E_\xi E_f p(\underset{\sim}{x})(t-m_Y)^2$

$$= E_f p(\underset{\sim}{x}) E_\xi (t-m_Y)^2 .$$

Hence it suffices to minimize $E_\xi(t-m_Y)^2$ subject to (5.2.8).

Now $E_\xi(t-m_Y)^2 = E_\xi(t^2+m_Y^2-2tm_Y)$

Note that $E_\xi(t\,m_Y) = E_\xi \sum_{i=1}^{n} a_i(\underset{\sim}{x})Y(x_i)E_f Y(X)$

$$= \sum_{i=1}^{n} a_i(\underset{\sim}{x})E_\xi E_f Y(x_i)Y(X)$$

$$= \sum_{i=1}^{n} a_i(\underset{\sim}{x})E_f E_\xi Y(x_i)Y(X)$$

but $E_\xi(Y(x_i)Y(x)) = \beta^2 x_i x$ a.e. (F)

Hence $E_\xi(t\,m_Y) = \sum_{i=1}^{n} a_i(\underset{\sim}{x})E_f(\beta^2 x_i X)$

$$= \beta^2 \sum_{i=1}^{n} a_i(\underset{\sim}{x})x_i . \mu$$

$$= \beta^2 \mu^2 \qquad \text{(using (5.2.8))}.$$

Thus it is enough to minimize $E_\xi t^2$ subject to the condition (5.2.8). Equivalently, we have to solve the following minimization problem:

$$\text{Minimize} \quad \sum_{i=1}^{n} a_i^2(\underset{\sim}{x}) x_i^g$$
$$\text{Subject to} \quad \sum_{i=1}^{n} a_i(\underset{\sim}{x}) x_i = \mu . \qquad \ldots(5.2.9)$$

The optimization problem (5.2.9) readily yields the following solution

$$a_i(\underset{\sim}{x}) = \frac{\mu x_i^{1-g}}{\sum_{i=1}^{n} x_i^{2-g}}, \; 1 \leq i \leq n . \qquad \ldots(5.2.10)$$

Hence the estimator

$$t_g = \frac{\mu}{\sum_{i=1}^{n} x_i^{2-g}} \sum_{i=1}^{n} x_i^{1-g} y(x_i) \qquad \ldots(5.2.11)$$

minimizes $M_1(p,t)$ in the class of all linear ξ-unbiased estimators. However, our main problem of interest is to obtain a design for which the estimator in (5.2.11) becomes p-unbiased. Following theorem gives us a required sampling design.

<u>Theorem 5.2.2</u>. There exists a sampling design that makes the estimator t_g in (5.2.11) p-unbiased and corresponds to the design function

$$p_g(\underset{\sim}{x}) = A \left[\prod_{i=1}^{n} x_i^{g-1} \right] \sum_{i=1}^{n} x_i^{2-g} , \qquad \ldots(5.2.12)$$

where $A = \frac{1}{n^{\alpha}} \left[\frac{\Gamma\alpha}{\Gamma(\alpha+g-1)} \right]^{n-1}$

Proof. It is enough to check the following two conditions:

$$\text{(i)} \quad \int_{R_n^+} p_g(\underset{\sim}{x}) f(\underset{\sim}{x}) d\underset{\sim}{x} = 1$$

$$\text{and (ii)} \quad \sum_{i=1}^{n} \int_{R_{n-1}^+} a_i(\underset{\sim}{x}) p_g(\underset{\sim}{x}) \prod_{j \neq i}^{n} f(x_j) dx_j = 1$$

where $a_i(\underset{\sim}{x})$, $1 \leq i \leq n$, are given by (5.2.10).

Note that

$$\int_{R_n^+} p_g(\underset{\sim}{x}) f(\underset{\sim}{x}) d\underset{\sim}{x} = A \int_{R_n^+} \left(\sum_{i=1}^{n} x_i^{2-g} \right) \prod_{i=1}^{n} x_i^{g-1} f(x_i) dx_i$$

$$= \frac{A}{(\Gamma\alpha)^n} \int_{R_n^+} \left(\sum_{i=1}^{n} x_i^{2-g} \right) \prod_{i=1}^{n} x_i^{\alpha+g-2} e^{-x_i} dx_i$$

$$= \frac{A}{(\Gamma\alpha)^n} n \left(\Gamma(\alpha+g-1) \right)^{n-1} \Gamma(\alpha+1)$$

$$= An\alpha \left[\frac{\Gamma(\alpha+g-1)}{\Gamma\alpha} \right]^{n-1}$$

$$= 1 .$$

$$\text{And} \int_{R_{n-1}^+} \frac{\mu x_i^{1-g}}{\sum_{i=1}^{n} x_i^{2-g}} p_g(\underset{\sim}{x}) \prod_{j \neq i}^{n} f(x_j) dx_j = \mu x_i^{1-g} \int_{R_{n-1}^+} A x_i^{g-1} \prod_{j \neq i}^{n} x_j^{g-1} f(x_j) dx_j$$

$$= A\mu \left[\frac{\Gamma(\alpha+g-1)}{\Gamma\alpha} \right]^{n-1}$$

$$= \frac{1}{n} \; ; \; (\mu = \alpha) .$$

Hence the theorem.

Thus we have obtained a strategy (p_g, t_g) where t_g is the best linear ξ-unbiased estimator and p_g is the design for which t_g is p-unbiased. Here we had assumed that $b=1$ and $g \in [0,2]$ known in the model (5.1.3).

We now prove our next theorem.

Theorem 5.2.3. Under the model (5.1.3) with $b=1$ and $g \in [0,2]$ the strategy (p_g, t_g) is as good as the strategy (ppx, t_{HT}) w.r.t either measure of uncertainty $M_1(p,t)$ or $M_2(p,t)$.

Proof. Since both the strategies are p as well as ξ-unbiased M_1 and M_2 differ only by a constant.

$$\text{Now } M_2(p_g, t_g) = E_{p_g} E_\xi t_g^2 - \beta^2\mu^2$$

$$= E_{p_g} \frac{\sigma^2\mu^2}{\left(\sum_{i=1}^{n} x_i^{2-g}\right)}$$

$$= A\sigma^2\mu^2 \int_{R_n^+} \prod_{i=1}^{n} x_i^{g-1} f(x_i)\,dx_i$$

$$= A\sigma^2\mu^2 \left[\frac{\Gamma(\alpha+g-1)}{\Gamma\alpha}\right]^n$$

$$= \frac{\sigma^2\mu^2}{n} \frac{\Gamma(\alpha+g-1)}{\Gamma(\alpha+1)} \qquad \ldots(5.2.13)$$

On the other hand,

$$M_2(ppx, t_{HT}) = \frac{\sigma^2\mu^2}{n} \frac{\Gamma(\alpha+g-1)}{\Gamma(\alpha+1)} \quad \text{(Särndal, 1980)}$$

This completes the proof.

Corollary 5.2.1. In particular when $g=1$, we get the strategy (p_M, t_R) and when $g=2$, we, of course, get the strategy (ppx, t_{HT}).

Remark 5.2.1. Observe that since the strategy (p_g, t_g) is p-unbiased even if the model breaks down it remains pξ-unbiased. Thus the step of obtaining the design p_g even after getting the best linear ξ-unbiased estimator, a kind of insurance against the possible model break downs, is justified. However, the strategy (p_g, t_g) depends on the model parameter g that, contrary to our assumption, may be unknown. In which case the strategy (p_g, t_g) is not feasible. Theorem 5.2.3 states that the strategy (ppx, t_{HT}) is as good as the strategy (p_g, t_g) $\forall\, g \in [0,2]$. Further the strategy (ppx, t_{HT}) is independent of the model parameter g. This gives a kind of robustness property of the strategy (ppx, t_{HT}). Thus when the parameter g is not known the strategy (p_g, t_g) can be substituted by the equally good strategy (ppx, t_{HT}).

5.3 Comparison of Strategies

In the previous section we saw how the attempt of obtaining a best sampling strategy w.r.t. the measure of uncertainty $M_1(p,t)$ under the model (5.1.3) fails. Our compromise strategy, too, is not globally optimal unless $g=2$ in the model (5.1.3). In view of these observations we compare certain well-known strategies used for estimating the population mean. Comparison of sampling strategies, in the absence of globally optimal strategy, under the superpopulation set up, w.r.t. certain measure of uncertainty, has been one of the main problems of interest to the survey statisticians. Such investigations are carried out in the finite set up by various authors. In this section we compare some of the strategies, introduced in Section 5.1, under the model (5.1.3) with $b=1$.

Särndal (1980) studied the strategy (srs, t_R). He notes that the strategy (srs, t_R) is not p-unbiased and when $b \neq 1$ in the model (5.1.3) it is not even pξ-unbiased. Here we note that the strategy (p_M, t_R) consisting of Midzuno-Sen sampling design and the conventional ratio estimator is always p-unbiased and even if the model parameter b is different from 1 it remains at least pξ-unbiased. Apart from this advantage over (srs, t_R) we shall, in fact, prove that the strategy (p_M, t_R) is always superior to (srs, t_R). Let us first prove the following lemma.

Lemma 5.3.1

$$J = \int_{R_n^+} \frac{x_1^g}{\left(\sum_{i=1}^{n} x_i\right)^m} e^{-\Sigma x_i} \prod_{i=1}^{n} x_i^{\alpha-1} = \frac{\Gamma(n\alpha+g-m)\,\Gamma(g+\alpha)}{\Gamma(n\alpha+g)} (\Gamma\alpha)^{n-1} \quad ..(5.3.1)$$

where m is a real number.

Proof. Consider the following transformation

$x_1 = u_1(1-u_2)$, $x_2 = u_1 u_2 (1-u_3) \ldots \ldots$,

$x_{n-1} = u_1 u_2 \ldots u_{n-1}(1-u_n)$, $x_n = u_1 u_2 \ldots u_{n-1} u_n$.

For this transformation,

$$0 \leq u_1 < \infty \quad \text{and} \quad 0 \leq u_i \leq 1 \quad i = 2,3,\ldots,n.$$

The Jacobian of the transformation is $u_1^{n-1} u_2^{n-2} \ldots\ldots u_{n-1}$. Hence

$$J = \int_0^\infty \int_0^1 \cdots \int_0^1 \frac{[u_1(1-u_2)]^g}{u_1^m} e^{-u_1} \left[u_1(1-u_2)u_1u_2(1-u_3)\ldots(u_1\ldots u_n)\right]^{\alpha-1} \cdot u_1^{n-1} u_2^{n-2} \ldots u_{n-1} \prod_{i=1}^{n} du_i$$

$$= \int_0^\infty e^{-u_1} u_1^{n\alpha+g-m-1} du_1 \int_0^1 u_2^{(n-1)\alpha-1} (1-u_2)^{g+\alpha-1} du_2$$

$$\int_0^1 u_3^{(n-2)\alpha-1} (1-u_3)^{\alpha-1} du_3 \ldots \int_0^1 u_{n-1}^{2\alpha-1} (1-u_{n-1})^{\alpha-1} du_{n-1}$$

$$\int_0^1 u_n^{\alpha-1} (1-u_n)^{\alpha-1} du_n$$

$$= \Gamma(n\alpha+g-m) \frac{\Gamma((n-1)\alpha)\,\Gamma(g+\alpha)}{\Gamma(n\alpha+g)} \frac{\Gamma((n-2)\alpha)\,\Gamma(\alpha)}{\Gamma((n-1)\alpha)} \ldots\ldots \frac{\Gamma(2\alpha)\,\Gamma(\alpha)}{\Gamma(3\alpha)} \frac{\Gamma(\alpha)\,\Gamma(\alpha)}{\Gamma(2\alpha)}$$

$$= \frac{\Gamma(n\alpha+g-m)\,\Gamma(g+\alpha)}{\Gamma(n\alpha+g)} (\Gamma(\alpha))^{n-1}.$$

Hence the lemma.

We now have the following theorem.

<u>Theorem 5.3.1.</u> Under the model (5.1.3) with $b = 1$ and $g \in [0,2]$ the strategy (p_M, t_R) is always superior to the strategy (srs, t_R) w.r.t. the measure of uncertainty $M_2(p,t)$ of (5.1.5).

<u>Proof.</u> Let p be any design function then

$$M_2(p\,,\,t_R) = E_pE_\xi(t_R - \beta\mu)^2$$

$$= E_pV_\xi t_R \qquad (V_\xi \text{ denotes variance under the model } \xi)$$

$$= \sigma^2\mu^2\, E_p \frac{\sum_{i=1}^{n} x_i^g}{(\sum_{i=1}^{n} x_i)^2}$$

For $p(\underset{\sim}{x}) \equiv 1$

$$E_p\left[\frac{\sum_{i=1}^{n} x_i^g}{(\sum_{i=1}^{n} x_i)^2}\right] = \frac{n}{(\Gamma\alpha)^n} \int_{R_n^+} \frac{x_1^g}{(\sum_{i=1}^{n} x_i)^2}\, e^{-\Sigma x_i} \prod_{i=1}^{n} x_i^{\alpha-1}\, dx_i$$

$$= \frac{n}{(\Gamma\alpha)^n} \frac{\Gamma(n\alpha+g-2)\,\Gamma(g+\alpha)}{\Gamma(n\alpha+g)} (\Gamma\alpha)^{n-1} ,$$

using Lemma 5.3.1 with $m = 2$.

Thus $M_2(srs, t_R) = \sigma^2\mu^2 \dfrac{n\,\Gamma(g+\alpha) / \Gamma\alpha}{(g+n\alpha-1)(g+n\alpha-2)}$ (see also Särndal, 1980).

Similarly for $p(\underset{\sim}{x}) = \sum_{i=1}^{n} x_i / n\mu$, using (5.3.1) with $m = 1$, we get

$$M_2(p_M\,,\,t_R) = \sigma^2\mu^2 \frac{\Gamma(g+\alpha) / \Gamma(\alpha+1)}{(g+n\alpha-1)} \qquad \ldots(5.3.2)$$

Now let $\eta = M_2(p_M\,,\,t_R) / M_2(srs\,,\,t_R)$

then $\eta = n\alpha / [n\alpha + (g-2)]$. $\qquad \ldots(5.3.3)$

This shows that $(p_M\,,\,t_R)$ is always superior to $(srs\,,\,t_R)$.

Remark 5.3.1. Thus apart from a safeguard against model break-downs the strategy $(p_M\,,\,t_R)$ is always superior to the strategy

(srs, t_R). Note that for $g = 2$ the two strategies are, in fact, equally good and for $g < 2$ $\lim_{n \to \infty} \eta = 1$, where η is given by (5.3.3). This is true otherwise also since $\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} x_i = \mu$ or $p_M(\underset{\sim}{x}) \to 1$ as $n \to \infty$ i.e. for large values of n the Midzuno-Sen sampling design is approximately equal to the simple random sampling.

In our next theorem we compare the strategies (p_M, t_R) and (ppx, t_{HT}).

Theorem 5.3.2. Under the model (5.1.3) with $b = 1$, we have for $n \geq 2$ and $g + \alpha - 1 > 0$

$$M_r(p_M, t_R) \lesseqgtr M_r(ppx, t_{HT}) \text{ according as } g \lesseqgtr 1, \; r = 1, 2.$$

Proof. Since both the strategies are p as well as ξ-unbiased it is enough to prove the result for $r = 2$.

Now observe that using (5.2.13) and (5.3.2) we get

$$M_2(p_M, t_R) / M_2(ppx, t_{HT}) = n(\alpha + g - 1) / (g + n\alpha - 1)$$

$$= 1 + (n-1)(g-1) / (g + n\alpha - 1).$$

Now for $n \geq 2$ $g + n\alpha - 1 > 0$

and $(n-1)(g-1) \gtreqless 0$ according as $g \gtreqless 1$.

Thus $M_2(p_M, t_R) \lesseqgtr M_2(ppx, t_{HT})$ according as $g \lesseqgtr 1$.

Hence the theorem.

Remark 5.3.2. The above theorem says that when the sampler has to choose between the above two strategies, there is a clear demarcation of the range of the parameter g of the model (5.1.3). If there are reasons to believe that the parameter g is less than unity then he should go for the strategy (p_M, t_R). On the other hand if he speculates g to be greater than one then surely he should prefer the strategy (ppx, t_{HT}) to (p_M, t_R).

<u>Remark 5.3.3</u>. Theorem 5.3.2 is in complete agreement with a result due to T.J. Rao (1967).

In the next section we compare a few more strategies in the stratified set up.

5.4 Stratified Sampling

In this section we deal with the stratified sampling involving L strata, introduced in Section 5.1. We are basically interested in comparing certain strategies in the stratified set up. Särndal (1980) proposed strategy $(srst, t_{st})$ where srst is the stratified simple random sampling and the estimator t_{st} is given by

$$t_{st} = \sum_{h=1}^{L} W_h \bar{y}_h \qquad \ldots(5.4.1)$$

where $W_h = F(x_h) - F(x_{h-1})$, $1 \leq h \leq L$, and letting n_h to be the number of units sampled from hth stratum, $1 \leq h \leq L$, $\sum_{h=1}^{L} n_h = n$, $\bar{y}_h$ is the sample mean for the hth stratum

$$\bar{y}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} y(x_i) \, , \; 1 \leq h \leq L \, . \qquad \ldots(5.4.2)$$

One may think of strategy $(ppxst, t^*_{HT})$ where ppxst denotes the stratified ppx sampling and the estimator t^*_{HT} is given by

$$t^*_{HT} = \sum_{h=1}^{L} W_h \frac{\mu_h}{n_h} \sum_{i=1}^{n_h} \frac{y(x_i)}{x_i} \qquad \text{...(5.4.3)}$$

where $$\mu_h = \frac{1}{W_h} \int_{x_{h-1}}^{x_h} xf(x)dx \,. \qquad \text{...(5.4.4)}$$

It is easy to check that the strategy $(ppxst, t^*_{HT})$ is p as well as ξ-unbiased. For this strategy let us evaluate $M_2(p,t)$. Let x_{hi} denote ith unit in hth stratum, $1 \le i \le n_h$ and $1 \le h \le L$.

$$M_2(p,t) = E_p V_\xi (t^*_{HT})$$

$$= E_p \sum_{h=1}^{L} W_h^2 \frac{\mu_h^2}{n_h^2} \sum_{i=1}^{n_h} \frac{\sigma^2 x_{h_i}^g}{x_{h_i}^2}$$

$$= \sigma^2 \sum_{h=1}^{L} \frac{W_h \mu_h}{n_h} \int_{x_{h-1}}^{x_h} x^{g-1} f(x)dx \,. \qquad \text{...(5.4.5)}$$

For the allocation $$n_h = \frac{nW_h \mu_h}{\mu}\,,\quad 1 \le h \le L, \qquad \text{...(5.4.6)}$$

$$M_2(ppxst, t^*_{HT}) = \frac{\sigma^2 \mu}{n} \int_0^\infty x^{g-1} f(x)dx$$

$$= \frac{\sigma^2 \mu^2}{n} \frac{\Gamma(\alpha+g-1)}{\Gamma(\alpha+1)} \,. \qquad \text{...(5.4.7)}$$

For the optimal allocation,

$$n_h^2 \propto W_h \mu_h \int_{x_{h-1}}^{x_h} x^{g-1} f(x)dx \,,\quad 1 \le h \le L, \qquad \text{...(5.4.8)}$$

$$M_2(ppxst, t^*_{HT}) = \frac{\sigma^2}{n} \left[\sum_{h=1}^{L} \left(W_h \mu_h \int_{x_{h-1}}^{x_h} x^{g-1} f(x)dx\right)^{1/2} \right]^2 \qquad \text{..(5.4.9)}$$

We now have the following theorem.

<u>Theorem 5.4.1</u> Under the model (5.1.3) with $b=1$, we have

(a) for the allocation (5.4.6)

$$M_2(ppxst, t^*_{HT}) = M_2(ppx, t_{HT})$$

and (b) for the allocation (5.4.8)

$$M_2(ppxst, t^*_{HT}) \leq M_2(ppx, t_{HT})$$

where the equality holds if and only if $g=2$.

<u>Proof</u>. Proof is immediate.

We now proceed to comment on a result due to Särndal (1980). Let us first evaluate $M_2(srst, t_{st})$.

$$\begin{aligned} M_2(p,t) &= E_p\left[V_\xi(t_{st}) + (E_\xi t_{st})^2\right] - \beta^2\mu^2 \\ &= E_p\left[\sigma^2 \sum_{h=1}^{L} \frac{W_h^2}{n_h^2} \sum_{i=1}^{n_h} x_{h_i}^g + \beta^2\left(\sum_{h=1}^{L} W_h \bar{x}_h\right)^2\right] - \beta^2\mu^2, \quad \left(\bar{x}_h = \frac{1}{n_h}\sum_{i=1}^{n_h} x_{h_i}\right) \\ &= \sigma^2 \sum_{h=1}^{L} \frac{W_h}{n_h} \int_{x_{h-1}}^{x_h} x^g f(x)dx + \beta^2 \sum_{h=1}^{L} \frac{1}{n_h}\left[W_h \int_{x_{h-1}}^{x_h} x^2 f(x)dx - \left(\int_{x_{h-1}}^{x_h} x f(x)dx\right)^2\right]. \end{aligned} \quad \ldots(5.4.10)$$

For the proportional allocation, $n_h = nW_h$, $1 \leq h \leq L$, we have

$$M_2(p,t) = \frac{\sigma^2\mu^2}{n}\frac{\Gamma(\alpha+g)}{\alpha\,\Gamma(\alpha+1)} + \beta^2 \sum_{h=1}^{L} \frac{1}{nW_h}\left[W_h \int_{x_{h-1}}^{x_h} x^2 f(x)dx - \left(\int_{x_{h-1}}^{x_h} x f(x)dx\right)^2\right] \quad \ldots(5.4.11)$$

Särndal (1980) mentions that for the proportionally allocated stratified random sample, ($n_h = nW_h$) and under 'maximum benefit from stratification' (many strata with optimally located boundaries),

$$M_2(srst, t_{st}) = \frac{\sigma^2\mu^2}{n} \frac{\sqrt{\alpha+g}}{\alpha\sqrt{\alpha+1}} \qquad \ldots(5.4.12)$$

However, it is easy to check that the coefficient of β^2 in (5.4.10) as well as (5.4.11) is always positive. This follows by Cauchy-Schwartz inequality. Hence even under 'maximum benefit from stratification' (5.4.12) cannot be true. This automatically invalidates the comparison between the strategies (srs, t_R) and $(srst, t_{st})$ due to Särndal (1980). Instead we have the following theorem.

<u>Theorem 5.4.2</u> Under the model (5.1.3) with $b = 1$ and $g \geq 1$ for the proportional allocation the strategy $(srst, t_{st})$ is inferior to both the strategies (p_M, t_R) and (ppx, t_{HT}).

<u>Proof</u>. Using (5.3.2) and (5.4.11) we get

$$\frac{M_2(srst, t_{st})}{M_2(p_M, t_R)} \geq 1 + \frac{g-1}{n\alpha}$$

Therefore for $g \geq 1$ (p_M, t_R) and hence (ppx, t_{HT}) are both superior to the strategy $(srst, t_{st})$.

The aspect of pξ-bias introduced due to possible model failures has been considered by Särndal (1980). He advocates a 'conservative' approach to use p-unbiased or approximately p-unbiased strategies. This is a safeguard against model break-downs. All the strategies considered in this chapter are, indeed, so.

CHAPTER 6

ESTIMATION OF POPULATION PROPORTION

6.0 Summary

For the comparison of certain sampling strategies for estimating the population proportion Lanke (1975) suggested the use of superpopulation model approach. In this chapter we observe that the superpopulation model suggested by Lanke (1975) has many interesting features which remained unnoticed. Here we consider both the aspects of inference namely model-based inference and model-design-based inference for estimating the population proportion. While dealing with the model-based inference we first obtain some results regarding the nonexistence of estimators satisfying certain conditions for the only parameter of the proposed model. We then obtain estimators for the same parameter using different principles of estimation like 'maximum likelihood' generalized least squares', 'least absolute value' etc. An attempt is made to compare all these estimators in a 'reasonable' sense. Finally, while considering the role of designs we obtain some optimality results in the sense of minimum expected variance.

6.1 Introduction

In practice sometimes we are interested in estimating the total number or proportion of units in a population possessing a particular characteristic. Many of the results of surveys or censuses are of this nature e.g. number of unemployed persons or percentage of people happy with the present government or percentage of people aged sixty and above etc. Such surveys also simplify 'measurements' since a questionnaire can be formulated so as to introduce a classification that is answered by simple 'yes' or 'no'. Even if the original measurements are more or less continuous e.g. respondent's income to the nearest integer, the percentage of population having income 5000 and more can be tabulated.

In the general superpopulation model approach when the characteristic y under study is roughly continuous and in addition an auxiliary information i.e. the values of an auxiliary variate x taking positive values x_i, $1 \leq i \leq N$ and closely related to the study variate y, is available a suitable model is assumed and a criterion of optimality is set up for estimating the population mean. However, when Y_i's are one-zero variates it is not possible to relate the auxiliary information with the study variate in the 'traditional' way. Nevertheless it is possible to set up a reasonable sort of dependence of y on x.

Let $y_1, y_2, \ldots, y_N$ be a realization of N independent one-zero variates $Y_1, Y_2, \ldots, Y_N$ having a joint distribution that is specified as follows ' We assume that the values of a positive valued nonstochastic variate x are available as size measures. The model hypothesizes that the probability of success for ith variate Y_i i.e. Prob $[Y_i = 1]$, $1 \leq i \leq N$, in some way depends on the auxiliary values x_i, $1 \leq i \leq N$. A reasonable form of dependence is

$$\text{Prob}\,[\,Y_i = 1 \mid x_i\,] = \beta x_i \quad 1 \leq i \leq N$$

$$\text{or} \quad E_\xi(Y_i \mid x_i) = \beta x_i$$

$$E_\xi[(Y_i - \beta x_i)^2 \mid x_i] = \beta x_i(1 - \beta x_i) \qquad i = 1, 2, \ldots, N \qquad (6.1.1)$$

$$E_\xi[(Y_i - \beta x_i)(Y_j - \beta x_j) \mid x_i, x_j] = 0 \quad i \neq j = 1, 2, \ldots, N$$

where β, $0 \leq \beta \leq 1/x_m$; $x_m = \max\limits_{1 \leq i \leq N} x_i$, is the unknown model parameter. Model (6.1.1) is due to Lanke (1975). He used the model to compare certain sampling strategies for estimating the population proportion $\overline{Y} = \frac{1}{N} \sum\limits_{i=1}^{N} y_i$. We note that the model (6.1.1) is interesting in itself and has many appealing features. Here an attempt is made to explore the model more extensively. Note that the model (6.1.1), unlike usual superpopulation models, is completely specified in the sense that the joint distribution of $Y_1, Y_2, \ldots, Y_N$ is completely known. In this chapter we take up model based inference as well as examine the role of designs in estimating the proportion.

An estimator b is said to be model unbiased for β if

$$E_\xi(b(s, \underset{\sim}{Y})) = \beta \quad \forall\, \beta \in [0, 1/x_m] \; ; \; p(s) > 0,$$

whereas an estimator t is said to be model unbiased for the population proportion if

$$E_\xi(t(s, \underset{\sim}{Y}) - \bar{Y}) = 0 \quad \forall \beta \in [0, \frac{1}{x_m}], \quad p(s) > 0 .$$

A strategy (p,t) is said to be design unbiased or p-unbiased for estimating the population proportion if

$$E(t) = \sum_{s \in S} p(s)t(s,\underset{\sim}{y}) = \bar{Y} \quad \forall \quad \underset{\sim}{y} \in R_N .$$

To compare the performances of various sampling strategies we use the following measure of uncertainty

$$M_1(p,t) = E_\xi E_p(t - \bar{Y})^2 . \qquad \ldots(6.1.2)$$

Note that the measure of uncertainty (6.1.2) is same as (2.1.3).

6.2 Model-Based Inference.

The prediction approach allows a model an essential role in inference. When the population total is expressed as the sum of the sample total $\sum_{i \in S} y_i$ and the total of the unsampled residuum $\sum_{i \notin s} y_i$, the problem of estimating $\sum_{i=1}^{N} y_i$ is recognized as one of predicting the sum of the unobserved random variables $\sum_{i \notin s} Y_i$. The assumed model plays a role to link the observed and unobserved values. From the observed values an inference is made about the model which is then used to predict the values of unobserved variates. (Royall and Herson, 1973).

Following Royall's prediction approach we may use $\left[\frac{1}{N}\sum_{i \varepsilon s} y_i + b \sum_{i \not\varepsilon s} x_i\right]$ to estimate $\bar{Y}$ where b is an estimate of the model parameter β based on the sample actually drawn. Usually b is taken to be the generalized least square estimator (glse) of β based on the sample. However, since $V_\xi(Y_i)$, in (6.1.1), also involves the same parameter β all known 'nice' results cannot be applied as they are. We first prove a nonexistence theorem.

<u>Theorem 6.2.1</u>. Based on a sample s such that $x_i < x_m$ $\forall$ $i \varepsilon s$ there does not exist an estimator b of β satisfying

$$0 \leq b \leq 1/x_m \qquad \text{...(6.2.1)}$$

and $E_\xi(b) = \beta \quad \forall \beta \varepsilon [0, 1/x_m]$. ...(6.2.2)

Note that bx_i is the estimated probability of success for the ith variate, $1 \leq i \leq N$ and (6.2.1) assures that the estimated probability lies between zero and one. (6.2.2) is the condition of unbiasedness. Thus Theorem 6.2.1 states that based on a sample s of the type indicated above no unbiased estimate of β can lie within the bounds zero and $1/x_m$.

<u>Proof</u>. Without loss of generality (wlg) let $s = \{1,2,...,n\}$. Hence $0 < x_i < x_m$ for $1 \leq i \leq n$. Let $B = \{\underset{\sim}{y} = (y_1, y_2, ..., y_n) : y_i = 1 \text{ or } 0, \ 1 \leq i \leq n\}$. If possible let there exist an estimator b of β satisfying (6.2.1) and (6.2.2). The condition (6.2.2) demands that

$$E(b) = \sum_{\underset{\sim}{y} \in B} b(y) \prod_{i=1}^{n} (\beta x_i)^{y_i} (1 - \beta x_i)^{1-y_i} \quad \forall \beta \in [0, 1/x_m].$$

In particular for $\beta = 1/x_m$ we have

$$\sum_{\underset{\sim}{y} \in B} b(\underset{\sim}{y}) \prod_{i=1}^{n} \left(\frac{x_i}{x_m}\right)^{y_i} \left(1 - \frac{x_i}{x_m}\right)^{1-y_i} = \frac{1}{x_m}$$

Now note that $0 < \prod_{i=1}^{n} \left(\frac{x_i}{x_m}\right)^{y_i} \left(1 - \frac{x_i}{x_m}\right)^{1-y_i} < 1 \quad \forall \underset{\sim}{y} \in B$ since

$0 < x_i < x_m \quad \forall i = 1, 2, \ldots, n.$

Hence in view of the condition (6.2.1), $E_\xi(b | \beta = 1/x_m) = 1/x_m$ if and only if $b(\underset{\sim}{y}) = 1/x_m \ \forall \underset{\sim}{y} \in B$. And once $b(\underset{\sim}{y})$ = constant the condition (6.2.2) cannot be satisfied. This contradicts the existence of b satisfying (6.2.1) and (6.2.2). Hence the theorem.

Now consider a sample s such that $x_i = x_m$ for some r (≥ 1) units and $x_i < x_m$ for remaining units in the sample. Again wlg assume that $s = \{1, 2, \ldots, n\}$, $x_1 = x_2 = \ldots = x_r = x_m$ and $x_i < x_m$ for $r+1 \leq i \leq n$. If $r = 1$ then we have the following theorem.

<u>Theorem 6.2.2</u>. For $r = 1$, based on the sample s, Y_1/x_m is the only estimator satisfying (6.2.1) and (6.2.2).

<u>Proof</u>. This follows from the completeness of Y_1/x_m. Next for $2 \leq r \leq n-1$ we try to find estimators of β satisfying (6.2.1) and (6.2.2). Consider estimators of the following types :

(i) $b_1(Y_1, Y_2, \ldots, Y_r)$ (that depends only on $Y_1, Y_2, \ldots, Y_r$) satisfying (6.2.1) and (6.2.2) and

(ii) $b_2(Y_1,Y_2,\ldots,Y_n)$ satisfying (6.2.1) and (6.2.2) such that

$$\sum_{y_r} \cdots \sum_{y_1} b_2(y_1,y_2,\ldots,y_n) \prod_{i=1}^{r} (\beta x_i)^{y_i}(1-\beta x_i)^{1-y_i} = \beta \quad \forall\ \beta \in [0, 1/x_m]$$

and $\forall\ (y_{r+1}, y_{r+2}, \ldots, y_n)$.

First we give a nontrivial example of an estimator of type (ii).

Example 6.2.1. For $n=3$ and $r=2$ define,

$$b_2(Y_1, Y_2, Y_3) = Y_1/x_m \quad \text{if} \quad Y_3 = 1$$

$$= Y_2/x_m \quad \text{if} \quad Y_3 = 0$$

clearly $\bar{b}_2(Y_3) = \sum_{y_1}\sum_{y_2} b_2(y_1, y_2, Y_3) \prod_{i=1}^{2} (\beta x_i)^{y_i}(1-\beta x_i)^{1-y_i}$

$= \beta$ whether $Y_3 = 1$ or $Y_3 = 0$.

Further $b_2(Y_1, Y_2) = \sum_{y_3} b_2(Y_1, Y_2, y_3)(\beta x_3)^{y_3}(1-\beta x_3)^{1-y_3}$

$$= \beta x_3 Y_1/x_m + (1-\beta x_3)Y_2'/x_m$$

and $\sum_{y_1}\sum_{y_2} \bar{b}_2(y_1, y_2) \prod_{i=1}^{2} (\beta x_i)^{y_i}(1-\beta x_i)^{1-y_i} = \beta$.

Let C be the collection of estimators of type (i) and type (ii), we then have,

Theorem 6.2.3. $\frac{1}{rx_n}\sum_{i=1}^{r} Y_i$ is the best estimator in C in the sense of minimum variance under the model (6.1.1).

Proof. From the completeness of order statistic $\frac{1}{rx_n}\sum_{i=1}^{r} Y_i$ is the best estimator of type (i). Further any estimator of type (ii) can be improved upon by $\frac{1}{rx_n}\sum_{i=1}^{r} Y_i$ since for any fixed $y_{r+1}, y_{r+2}, \ldots, y_n$ it is essentially an estimator of type (i).

Let C_1 ($\subseteq C$) consist of estimators of type (i). We then have the following

Theorem 6.2.4. There exists no linear estimator outside C_1 satisfying (6.2.1) and (6.2.2).

Proof. If possible let there exist a linear estimator b of β outside C_1 satisfying (6.2.1) and (6.2.2).

A linear estimator b is of the type $b = \sum_{i=1}^{n} a_i Y_i$. In view of $b \geq 0$ we have $a_i \geq 0$ $\forall\, i = 1, 2, \ldots, n$ and since $b \notin C_1$ we must have $a_j > 0$ for some j such that $r + 1 \leq j \leq n$. Further the condition (6.2.2) is equivalent to $\sum_{i=1}^{n} a_i x_i = 1$.

Now $b(1, 1, \ldots, 1) = \sum_{i=1}^{n} a_i$ and $\sum_{i=1}^{n} a_i > \frac{1}{x_n}$ because of (6.2.2) but this contradicts (6.2.1). Hence the theorem.

Remark 6.2.1. There may, however, exist nonlinear estimators outside C satisfying (6.2.1) and (6.2.2). It is also easy to check that the class C of estimators is not complete. Consider the following

Example 6.2.2.

For $n = 3$ and $r = 2$ let

$$b(1,1,1) = b(1,1,0) = 1/x_m, \quad b(0,0,1) = \alpha, \quad b(0,0,0) = 0$$

$$b(1,0,1) = b(0,1,1) = [1 + \alpha(x_m - x_3)]/2x_m$$

$$b(1,0,0) = b(0,1,0) = (1 - \alpha x_3)/2x_m, \quad 0 < \alpha \leq 1/x_m$$

and $\bar{b} = (Y_1 + Y_2)/2x_m$.

It is easy to check that $b \notin C$ and it satisfies (6.2.1) and (6.2.2) For $\alpha = .1$, $x_3 = 1$ and $x_m = 1.1$ the range for β is $0 \leq \beta \leq \frac{10}{11} = .9090$ and for $0 \leq \beta \leq \frac{1680}{1859} = .9037$, $V_\xi(b) < V_\xi(\bar{b})$.

Here we note that the main reason behind the various non-existence results is that the parametric range for the probability of success for the ith variate is $[0, x_i/x_m]$ and not the natural range $[0,1]$, the two being same for $x_i = x_m$. If we interpret $\hat{\beta}x_i$ as the estimated value of y_i rather than the estimated probability of success we may be willing to relax the condition (6.2.1). Further estimation of β is just an intermediate step in the estimation of the proportion. So what we must demand is that the estimate of the proportion should not exceed one. This in turn gives bounds for $\hat{\beta}$ as

$$0 \leq \hat{\beta} \leq (N-n)/(X-x_s) \qquad \ldots(6.2.3)$$

where $X = \sum_{i=1}^{N} x_i$, $x_s = \sum_{i \in s} x_i$.

Even after relaxing the condition (6.2.1) to (6.2.3) we can have similar existence and nonexistence results in the linear set up. Going a step further even if we relax the condition (6.2.1) complete we still have the following nonexistence theorem.

Theorem 6.2.5. For any sample s such that $x_i \neq x_j$ for at least one pair $(i,j), i \neq j \in s$, there does not exist uniformly minimum variance linear unbiased estimator of β under the model (6.1.1).

Proof. The proof follows from the fact that the optimal choice of linear estimator depends on the parameter β.

Remark 6.2.2. If $x_i = x \;\forall\, i \in s$ then uniformly minimum variance linear unbiased estimator exists and is given by $\frac{1}{nx} \sum_{i \in s} Y_i$. Further, using completeness of order statistic, it is the best unbiased estimator of β in the sense of minimum variance.

We now proceed to obtain various estimators for β. As mentioned earlier the model (6.1.1), unlike usual superpopulation models, is completely specified, hence we can think of maximum likelihood estimate (mle) of β.

(a) Maximum likelihood estimator.

Let $s = \{1,2,\ldots,n\}$ and $f(\underset{\sim}{y}) = \text{Prob}\,[Y_i = y_i,\; 1 \leq i \leq n]$. Then

$$f(\underset{\sim}{y}) = \prod_{i=1}^{n} (\beta x_i)^{y_i}(1-\beta x_i)^{1-y_i} \quad \text{and}$$

$$L = \log f(\underset{\sim}{y}) = \sum_{i=1}^{n} y_i \log x_i + \left(\sum_{i=1}^{n} y_i\right)\log\beta + \sum_{i=1}^{n} (1-y_i)\log(1-\beta x_i).$$

If $y_i = 1 \;\forall\, i \in s$ then the mle is given by $b_1 = (N-n)/(X-x_s)$. Otherwise,

$$\frac{\partial L}{\partial \beta} = \frac{1}{\beta}\sum_{i=1}^{n} y_i - \sum_{i=1}^{n} \frac{(1-y_i)x_i}{1-\beta x_i} \quad \text{and} \quad \frac{\partial^2 L}{\partial \beta^2} = -\frac{1}{\beta^2}\sum_{i=1}^{n} y_i - \sum_{i=1}^{n} \frac{x_i^2(1-y_i)}{(1-\beta x_i)^2}.$$

Hence for $0 \leq \beta \leq 1/x_{mo}$, L is a concave function

where $x_{mo} = \max\limits_{\substack{1 \leq i \leq n \\ y_i = 0}} x_i$.

Now $\frac{\partial L}{\partial \beta} = 0 \iff \sum\limits_{i=1}^{n} y_i = \sum\limits_{i=1}^{n} \frac{\beta x_i (1-y_i)}{(1-\beta x_i)}$. ...(6.2.6)

Clearly (6.2.6) has a solution in $[0, 1/x_{mo}]$. Hence in view of concavity of L and the condition (6.2.3), the mle is given by

$$b_1 = \min [b_1', (N-n)/(X-x_s)]$$

where $b_1' \epsilon [0, 1/x_{mo}]$ is a solution to (6.2.6).

(b) Generalized least square estimator (glse).

Under many superpopulation models, the best linear unbiased estimator (in the sense of minimum variance) is obtainable using the principle of weighted least square estimation. We shall see, however, that the same principle does not yield a linear estimator in case of model (6.1.1). This is because of the fact that in (6.1.1), $V_\xi(Y_i)$ too involves the parameter β.

Let $s = \{1, 2, \ldots, n\}$ and wlg let a typical sample observation be $y_i = 1$, $1 \leq i \leq r$ and $y_i = 0$, $r < i \leq n$.

Then the glse is obtained by minimizing $Q_1 = \sum\limits_{i=1}^{n} \frac{(y_i - \beta x_i)^2}{\beta x_i (1-\beta x_i)}$.

For our typical observation, Q_1 simplifies to

$$Q_1 = \sum_{i=1}^{r} \frac{1-\beta x_i}{\beta x_i} + \sum_{i=r+1}^{n} \frac{\beta x_i}{1-\beta x_i}.$$

f r = n then the glse is given by $b_2 = (N-n)/(X-x_s)$. Otherwise,

$$\frac{\partial Q_1}{\partial \beta} = -\sum_{i=1}^{r} \frac{1}{\beta^2 x_i} + \sum_{i=r+1}^{n} \frac{x_i}{(1-\beta x_i)^2} \quad \text{and}$$

$$\frac{\partial^2 Q_1}{\partial \beta^2} = \sum_{i=1}^{r} \frac{2}{\beta^3 x_i} \quad \sum_{i=r+1}^{n} \frac{2x_i^2}{(1-\beta x_i)^3} .$$

learly for $\beta \varepsilon [0, 1/x_{mo}]$, $\partial^2 Q_1 / \partial \beta^2 > 0$. Hence Q_1 is a onvex function of β for $\beta \varepsilon [0, 1/x_{mo}]$. Further

$$\frac{\partial Q_1}{\partial \beta} = 0 \iff \sum_{i=1}^{r} \frac{1}{x_i} = \sum_{i=r+1}^{n} \frac{\beta^2 x_i}{(1-\beta x_i)^2} \qquad \ldots(6.2.7)$$

It is easy to check that (6.2.7) has a unique solution in $[0, 1/x_{mo}]$. Thus in view of the condition (6.2.3) and the convexity of Q_1 the glse is given by

$$b_2 = \min(b_2', (N-n)/(X-x_s))$$

where $b_2' \varepsilon [0, 1/x_{mo}]$ is a solution to (6.2.7).

We now consider yet another interesting estimator.

(c) Least absolute value estimator (lave)

To our knowledge, this particular principle of estimation has not been used in the model based inference so far. Consider a sample s as in (b). This principle is based on minimization of $Q_2 = \sum_{i=1}^{n} |y_i - \beta x_i|$. For our typical sample observation, we have

$$Q_2 = \sum_{i=1}^{r} |1-\beta x_i| + \sum_{i=r+1}^{n} \beta x_i .$$

Let wlg $x_1 \leq x_2 \leq \ldots \leq x_r$. Define $I_1 = [1/x_1, \infty)$, for $2 \leq t \leq r$ $I_t = [1/x_t, 1/x_{t-1}]$ and $I_{r+1} = [0, 1/x_r]$.

Clearly $\bigcup_{t=1}^{r+1} I_t = [0, \infty)$.

Now for $\beta \varepsilon I_t$, we have,

$$Q_2(\beta) = 2(t-1) - r + \beta \left[\sum_{i=t}^{n} x_i - \sum_{i=1}^{t-1} x_i \right].$$

Thus it is easy to check that $\min_{\beta \varepsilon I_1} Q_2(\beta)$ is attained at $1/x_1$ and for $2 \leq t \leq r+1$, $\min_{\beta \varepsilon I_t} Q_2(\beta)$ is attained at one of the end points of I_t. Further we have the condition $\hat{\beta} \leq \frac{N-n}{X-x_s}$. Let $\frac{N-n}{X-x_s} \varepsilon I_{t_o}$. Define $I^*_{t_o} = \{a : a \varepsilon I_{t_o}$ and $a \leq \frac{N-n}{X-x_s}\}$. Even for the interval $I^*_{t_o}$, $\min_{\beta} Q_2(\beta)$ is attained at one of the end points of $I^*_{t_o}$. Thus we conclude that $\min_{\beta \leq \frac{N-n}{X-x_s}} Q_2(\beta)$ is attained at one of the points $0, \frac{1}{x_r}, \frac{1}{x_{r-1}}, \ldots, \frac{1}{x_{t_o}}$ and $\frac{N-n}{X-x_s}$. Thus for a given sample one can easily obtain the lave.

Along with above estimators we consider some more estimators as follows.

$$b_4 = \text{lse} = \min \left[\frac{N-n}{X-x_s}, \frac{\sum_{i \varepsilon s} x_i Y_i}{\sum_{i \varepsilon s} x_i^2} \right]$$

$$b_5 = \min\left[\frac{N-n}{X-x_s}, \frac{\sum_{i \in s} Y_i}{\sum_{i \in s} x_i}\right]$$

$$b_6 = \min\left[\frac{N-n}{X-x_s}, \frac{1}{n}\sum_{i \in s}\frac{Y_i}{x_i}\right]$$

$$b_7 = \min\left[\frac{N-n}{X-x_s}, \frac{1}{X-x_s}\sum_{i \in s}\frac{X-nx_i}{nx_i}Y_i\right].$$

Note that the untruncated b_5 corresponds to the ratio estimator for the population proportion and untruncated b_7 corresponds to the mean of the ratios estimator for population proportion.

We now compare the various estimators listed above. For a given population of N units, values $x_1, x_2, \ldots, x_N$ and a value β_o of the parameter β let $y_1, y_2, \ldots, y_N$ be the most probable y-values i.e.

$$y_i = 1\,(0) \quad \text{if} \quad \beta_o x_i \geq \frac{1}{2} \left(< \frac{1}{2}\right), \; 1 \leq i \leq N.$$

For an estimator b_r, $1 \leq r \leq 7$, let s_r be the sample for which $\min_{s \in S} |b_r(s) \sum_{i \notin s} x_i - \sum_{i \notin s} y_i|$ is attained. We compare the 'strategies' (b_r, s_r), $1 \leq r \leq 7$.

For the given values of $x_1, x_2, \ldots, x_N$ and β_o a 'best' strategy is one for which $A = \min_{1 \leq r \leq 7} |b_r(s_r) \sum_{i \notin s_r} x_i - \sum_{i \notin s_r} y_i|$ is attained. Here we consider four different populations. Population 1 consists of arbitrary x-values whereas populations 2,3 and 4 are samples from Gamma population f(x) with parameter $\alpha = 0,3$ and 8 respectively.

$$f(x) = e^{-x} x^{\alpha-1} / \Gamma\alpha, \quad x > 0, \; \alpha \geq 0.$$

In all the examples we have $N=10$ and $n=3$. In the Table 6.2.1 a row following the most probable y-values gives the 'best' strategy and the corresponding value of A.

Table 6.2.1

Population 1											
β	i	1	2	3	4	5	6	7	8	9	10
	x_i	1	2	3	5	6	7	7	9	9	10
0.051	y_i	0	0	0	0	0	0	0	0	0	1
	$b_3 = 0.00$			$s_3 = (2,8,10)$				$A = 0.00$			
0.066	y_i	0	0	0	0	0	0	0	1	1	1
	$b_2 = 0.056$			$s_2 = (5,6,10)$				$A = 3.06 \times 10^{-3}$			
0.080	y_i	0	0	0	0	0	1	1	1	1	1
	$b_3 = 0.10$			$s_3 = (3,5,10)$				$A = 1.91 \times 10^{-6}$			
0.089	y_i	0	0	0	0	1	1	1	1	1	
	$b_3 = 0.11$			$s_3 = (2,3,8)$				$A = 9.54 \times 10^{-7}$			

Table 6.2.1 (Contd.)

Population 2											
β ↓	i	1	2	3	4	5	6	7	8	9	10
	x_i	7	7	11	48	76	88	100	140	170	190
.0027	y_i	0	0	0	0	0	0	0	0	0	1
	b_3 = 0.00			s_3 = (4,9,10)				A = 0.00			
.0032	y_i	0	0	0	0	0	0	0	0	1	1
	b_5 = .0024			s_5 = (6,8,10)				A = .0024			
.0043	y_i	0	0	0	0	0	0	0	1	1	1
	b_2 = .004			s_2 = (5,6,9)				A = .0142			
Population 3											
	x_i	25	66	75	89	92	101	115	122	122	143
.0036	y_i	0	0	0	0	0	0	0	0	0	1
	b_3 = 0.00			s_3 = (1,8,10)				A = 0.00			
.0042	y_i	0	0	0	0	0	0	0	1	1	1
	b_6 = .0027			s_6 = (1,3,8)				A = .0109			
.0051	y_i	0	0	0	0	0	1	1	1	1	1
	b_6 = .0051			s_6 = (5,8,10)				A = .0025			
.0061	y_i	0	0	0	1	1	1	1	1	1	1
	$b_7 = 6.72 \times 10^{-3}$			s_7 = (1,4,5)				$A = 7.25 \times 10^{-5}$			
.0067	y_i	0	0	1	1	1	1	1	1	1	1
	$b_3 = 9.13 \times 10^{-3}$			s_3 = (1,2,5)				$A = 9.54 \times 10^{-7}$			

Table 6.2.1 (Contd.)

Population 4											
$\beta \downarrow$	i	1	2	3	4	5	6	7	8	9	10
	x_i	48	64	64	66	68	87	94	97	100	112
.0046	y_i	0	0	0	0	0	0	0	0	0	1
	$b_3 = 0.00$			$s_3 = (1,4,10)$				$A = 0.00$			
.0051	y_i	0	0	0	0	0	0	0	0	1	1
	$b_7 = 0.00$			$s_7 = (1,2,10)$				$A = 0.381$			
.0053	y_i	0	0	0	0	0	0	0	1	1	1
	$b_5 = .0038$			$s_5 = (5,6,10)$				$A = .0038$			
.0062	y_i	0	0	0	0	0	1	1	1	1	1
	$b_7 = .0056$			$s_7 = (1,9,10)$				$A = .0476$			
.0081	y_i	0	1	1	1	1	1	1	1	1	1
	$b_1 = .0122$			$s_1 = (1,2,10)$				$A = 9.54 \times 10^{-7}$			

For all the populations considered above the estimator b_3 seems to fare well, especially, when the parameter β is very small However, the study presented here is not quite conclusive.

6.3 Role of Designs

So far we considered inference based only on the model. We now introduce sampling designs so as to get strategies for estimating the proportion. To compare the performances of various strategies we use the measure of uncertainty (6.1.2). Let

$$D = \{p : p(s) > 0 \Rightarrow \sum_{i \in s} x_i / \pi_i = X\}$$

where $\pi_i = \sum_{s \ni i} p(s)$ is the the inclusion probability of the ith unit, $1 \leq i \leq N$. We now have the following theorem.

Theorem 6.3.1 For a design $p \in D$ there exists a best estimator for the population proportion $\bar{Y}$ in the class of all p-unbiased linear estimators in the sense of minimum $M_1(p,t)$.

Proof. A linear estimator is of the type $t = \sum_{i \in s} b(s,i)y_i$. Under the design p the condition of p-unbiasedness is equivalent to

$$\sum_{s \ni i} b(s,i)p(s) = 1/N \quad \forall\, i = 1,2,\ldots,N. \qquad \ldots(6.3.1)$$

Further, for a p-unbiased strategy, (6.1.2) simplifies to

$$M_1(p,t) = EV_\xi(t) + E(E_\xi(t) - \beta\bar{X})^2 - V_\xi(\bar{Y}).$$

$$\text{Now } EV_\xi(t) = \sum_{i=1}^{N} v(x_i) \sum_{s \ni i} b^2(s,i)p(s), \quad [v(x) = \beta x(1-\beta x)]$$

$$\geq \sum_{i=1}^{N} \frac{v(x_i)}{\pi_i N^2}$$

using condition (6.3.1) and the Cauchy-Schwartz inequality. The equality is attained for $b(s,i) = 1/N\pi_i \ \forall\, i$ and $\forall\, s$. Further for this choice of $b(s,i)$, by virtue of the fact that $p \in D$ the condition of model-unbiasedness is automatically satisfied. Hence for any design $p \in D$ the Horvitz-Thompson estimator is the best linear p-unbiased estimator in the sense of minimum $M_1(p,t)$.

Corollary 6.3.1. In particular, for a design giving inclusion probabilities proportional to the size measures x_i's (πps design) the Horvitz-Thompson estimator is the best linear p-unbiased estimator.

We now go on to the stratified set up. Let N units be grouped into L strata such that $x_{hi} = x_h \ \forall\, i \in S_h$, the hth stratum. In this set up, Lanke (1975) compared the simple random sampling with replacement (srswr) and probability proportional to size (pps) sampling scheme for proportional as well as optimal allocation. In the stratified set up we first obtain an optimality result. Fix an allocation $n_1, n_2, \ldots, n_L$ such that $\sum_{h=1}^{L} n_h = n$. Let D_1 be the class of designs such that $p \in D_1$ selects on an average n_h units from hth stratum, $1 \leq h \leq L$ and the selection of units from a stratum is independent of selection of units from remaining $L-1$ strata.

Theorem 6.3.2. For a given allocation $n_1, n_2, \ldots, n_L$; $\sum_{h=1}^{L} n_h = n$; there exists a best strategy in the class of strategies $H = \{(p,t) : p \in D_1$ and t is p-unbiased$\}$ in the sense of minimum $M_1(p,t)$ and is given by (p_o, t_o), where p_o is a design that gives inclusion probability

$$\frac{n_h}{N_h} \quad \forall\, i \in S_h\,, \quad 1 \le h \le L \quad \text{and} \quad t_o = \sum_{h=1}^{L} \frac{N_h}{N}\, \bar{y}_h \qquad \ldots (6.3.2)$$

with $\bar{y}_h = \frac{1}{n_h} \sum_{i \in s_h} y_i$, the sample mean for the hth stratum.

Proof. For a p-unbiased strategy (p,t)

$$M_1(p,t) = EV_\xi(t) + E\,[E_\xi(t) - \beta \bar{X}]^2 - V_\xi(\bar{Y})\,.$$

Now following Godambe and Joshi (1965) for any design p

$$EV_\xi(t) \ge EV_\xi(t_{HT})$$

where t_{HT} is the Horvitz-Thompson estimator and t is any other p-unbiased estimator. For any $p \in D_1$ under the model (6.1.1)

$$EV_\xi(t_{HT}) = \frac{1}{N^2} \sum_{h=1}^{L} v(x_h) \sum_{i \in S_h} \frac{1}{\pi_i}\,.$$

Using the condition $\sum_{i \in S_h} \pi_i = n_h$, $1 \le h \le L$, and the Cauchy-Schwartz inequality one easily gets that

$$EV_\xi(t_{HT}) \ge \frac{1}{N^2} \sum_{h=1}^{L} v(x_h) \frac{N_h^2}{n_h}\,.$$

And the equality is attained for the design p_o of (6.3.2). Now the Horvitz-Thompson estimator corresponding to p_o is nothing other than t_o. Further it is a matter of verification that the estimator t_o is model unbiased. Hence (p_o, t) of (6.3.2) is the best strategy in H.

Corollary 6.3.1. For $L = 1$ and $x_i = x \ \forall\, i = 1,2,\ldots,N$; $(p_0, \bar{y})$ is the best p-unbiased strategy for a given average size n in the sense of minimum $M_1(p,t)$, where p_0 now is a design giviving constant inclusion probability n/N to all the units and $\bar{y}$ is the sample mean $\frac{1}{n} \sum_{i \in s} y_i$.

Remark 6.3.1. After obtaining the optimal strategy (p_0, t_0) of (6.3.2) for a given allocation, one may think of an overall optimal allocation. However, as expected, the optimal allocation will involve the model parameter β.

Remark 6.3.2. In view of the Theorem 6.3.2 both the strategies considered for comparison by Lanke (1975) are 'inadmissible' ones.

6.4 Some Further Results.

In the same spirit, as in model (6.1.1), we can think of random variables Y_i taking finite or countable values. For instance x_i may be the income of the ith individual and y_i may be the number of luxury items he possesses. In such a situation we can think of the following two models.

Let $Y_1, Y_2, \ldots, Y_N$ be independent random variables taking values $0,1,2,\ldots$ with Poisson probability which depends on auxiliary variate value x_i as follows :

$$\text{Prob}\,[Y_i = r] = e^{-\beta x_i}(\beta x_i)^r / r! \quad r = 0,1,2\ldots \text{ and } i = 1,2,\ldots,N \qquad \ldots(6.4.1)$$

$$\beta > 0.$$

Under this model one can show that the mle and the glse of β are given by

$$\frac{\sum_{i\epsilon s} Y_i}{\sum_{i\epsilon s} x_i} \quad \text{and} \quad \left[\frac{\sum_{i\epsilon s} Y_i^2/x_i}{\sum_{i\epsilon s} x_i} \right]^{1/2} \quad \text{respectively.}$$

In the second case we assume that $Y_1, Y_2, \ldots, Y_N$ are independent and

$$\text{Prob}\,[\, Y_i = r\,] = (1-\beta x_i)(\beta x_i)^r \qquad r = 0,1,2,\ldots$$

$$0 < \beta < 1/x_m \qquad i = 1,2,\ldots,N. \qquad \ldots(6.4.2)$$

The model (6.4.2) can also be studied along similar lines.

CHAPTER 7

NONNEGATIVE VAR-ANCE ESTIMATIO-

7.0 Summary

This chapter is devoted to the problem of estimation of variance of estimators of population mean nonnegatively. We mainly deal with the strategy (p_M, t_R) that consists of Midzuno-Sen sampling scheme and the conventional ratio estimator of the population mean. However, analogous treatment can be given to some other strategies. In this chapter we first propose a general class of estimators, that includes all known estimators, having the necessary form of nonnegative unbiased variance estimators (nnuves) (Rao and Vijayan, 1977). We also give sufficient conditions for the uniform nonnegativity, which are weaker than the conditions known hitherto, for the various proposed estimators. Chaudhuri and Arnab (1981) listed down different variance estimators proposed prior to Rao-Vijayan's (1977) result regarding the necessary form of nnuve and demonstrated that for none of the estimators the sufficient conditions for the uniform nonnegativity can be satisfied. However, they tacitly ignored the fact that none of the estimators they considered satisfy the necessary form of nnuve, the notion which was well-known by then. After noting this we see how the use of transformation on auxiliary values obviates all the problems of nonnegative variance estimation. Next we consider biased nonnegative estimators and the aspect of reduction in mean square error (MSE). We then see how stratification technique can be employed to improve the chances of obtaining a uniformly nnuve retaining the advantages of the ratio estimator. We conclude the chapter by extending Rao-Vijayan's (1977) technique of obtaining nnuves to more general strategies.

.1 Introduction

In this chapter we consider the problem of obtaining a nonnegative estimator of the variance of an estimator of finite population mean. This problem has received a great deal of attention in the literature. Among others T.J. Rao (1972a, 1977), Chaudhuri (1976a, 1979), J.N.K. Rao and Vijayan (1977), Chaudhuri and Arnab (1981) considered the problem of estimating nonnegatively the variance of strategy (p_M, t_R) that consists of Midzuno-Sen sampling scheme and the conventional ratio estimator. In this chapter we mainly consider the estimation of variance of the strategy (p_M, t_R). However, an analogous theory can be developed for the strategy $(\pi ps, t_{HT})$ that consists of a πps design [π_i, the inclusion probability of the ith unit proportional to its size] and the Horvitz-Thompson estimator.

Consider a finite population of size N. Let y_i be the value of a real characteristic y on unit i, $1 \leq i \leq N$. A primary problem in survey sampling is to estimate the mean $\overline{Y} = \frac{1}{N} \sum_{i=1}^{N} y_i$ of the variate y. When values x_i, $1 \leq i \leq N$, of a real variate x which is positively correlated with the study variate y are available as size measures it is advantageous to make use of these values for selection of units as well as estimation of $\overline{Y}$. (p_M, t_R) is one such strategy; however, the main drawback of this strategy is the nonavailability of uniformly nonnegative unbiased variance estimator (nnuve). Here we deal with the problem of obtaining nonnegative unbiased variance estimators for the strategy (p_M, t_R).

The probability $p(s)$ of selecting a typical sample s of n distinct units under Midzuno-Sen sampling scheme is given by

$$p(s) = x_s / M_1 X \qquad \ldots(7.1.1)$$

where $x_s = \sum_{i \varepsilon s} x_i$, $X = \sum_{i=1}^{N} x_i$ and $M_r = \binom{N-r}{n-r}$, $r = 0,1,2$.

The conventional ratio estimator is given by

$$t_R = \bar{X} y_s / x_s \qquad \ldots(7.1.2)$$

where $y_s = \sum_{i \varepsilon s} y_i$ and $\bar{X} = \frac{1}{N} \sum_{i=1}^{N} x_i = \frac{X}{N}$.

Clearly the strategy (p_M, t_R) is unbiased for the population mean. Further its sampling variance is given by

$$V(p_M, t_R) = V(t_R) = \frac{\bar{X}}{NM_1} \sum_{s \varepsilon S} \frac{y_s^2}{x_s} - \bar{Y}^2. \qquad \ldots(7.1.3)$$

Rao and Vijayan (1977) gave an alternative expression for $V(t_R)$,

$$V(t_R) = - \sum_{i < j=1}^{N} a_{ij} x_i x_j \left(\frac{y_i}{x_i} - \frac{y_j}{x_j}\right)^2 \qquad \ldots(7.1.4)$$

where $N^2 a_{ii} = \frac{X}{M_1} \sum_{s \ni i} \frac{1}{x_s} - 1 \qquad i = 1,2,\ldots,N.$

and $N^2 a_{ij} = \frac{X}{M_1} \sum_{s \ni ij} \frac{1}{x_s} - 1 \quad i \neq j = 1,2,\ldots,N. \qquad \ldots(7.1.5)$

The problem of obtaining nnuve has been considered by various authors. T.J. Rao (1972a, 1977) and Chaudhuri (1976a) proposed a few estimators for $V(t_R)$ and also gave sufficient conditions for their uniform nonnegativity. However these conditions cannot be satisfied unless $x_1 = x_2 = \ldots = x_N$. J.N.K. Rao and Vijayan (1977) obtained necessary form of nnuve. They proposed two estimators having the necessary form of nnuve and also suggested sufficient conditions for uniform nonnegativity. (By uniform nonnegativity we mean nonnegativity for all $y_1, y_2, \ldots, y_N$ and for all samples). Later Chaudhuri (1979) proposed a few more estimators having the necessary form of nnuve. He also gave conditions for their uniform nonnegativity. In this chapter we propose a general class of nnuves and explore various avenues to improve the chances of obtaining uniformly nnuves.

7.2 Nonnegative Unbiased Variance Estimators.

Vijayan (1975) found the necessary form of nnuve for the Horvitz-Thompson estimator of population mean for any design with fixed sample size n; the method, however, as noted by Rao and Vijayan (1977), is applicable to any linear strategy (p,t) as in the following theorem.

Theorem 7.2.1 (Rao and Vijayan, 1977). For a linear strategy (p,t) if $MSE(p,t) = \sum_{i=1}^{N} \sum_{j=1}^{N} d_{ij} y_i y_j$ becomes zero when the ratios $z_i = \frac{y_i}{w_i}$, $1 \leq i \leq N$, are all equal for some known constants $w_i (\neq 0)$, $1 \leq i \leq N$, and d_{ij}'s are idependent of y-values then

(i) $MSE(p,t)$ reduces to

$$- \sum_{i<j=1}^{N} d_{ij} w_i w_j \left(\frac{y_i}{w_i} - \frac{y_j}{w_j}\right)^2 \qquad \ldots(7.2.1)$$

and (ii) a nonnegative quadratic unbiased estimator of MSE (p,t) is necessarily of the form

$$\sum_{i<j\in s} d_{ij}(s) w_i w_j \left(\frac{y_i}{w_i} - \frac{y_j}{w_j}\right)^2 \qquad \ldots(7.2.2)$$

where $d_{ij}(s)$ are independent of y-values and

$$\sum_{s \ni i,j} d_{ij}(s) p(s) = d_{ij} \quad \forall\ i \neq j = 1, 2, \ldots, N .$$

Using the above theorem it is easy to see that the necessary form v of nnuve for the strategy (p_M, t_R), is given by

$$v(s) = - \sum_{i<j\in s} a_{ij}(s) x_i x_j \left(\frac{y_i}{x_i} - \frac{y_j}{x_j}\right)^2 \ ; \ s \in S \qquad \ldots(7.2.3)$$

where $a_{ij}(s) = 0$ if $s \not\ni i$ or $s \not\ni j$

and $\sum_{s \ni ij} a_{ij}(s) p(s) = a_{ij} \quad i \neq j = 1, 2, \ldots, N$

a_{ij}'s are given by (7.1.3), S is the collection of all M_0 samples. Here and subsequently, in Section 7.2, p(s) denotes the probability of a sample s under Midzuno-Sen scheme given by (7.1.1).

We first list down various estimators, having the form (7.2.3), already known in the literature. Note that an estimator v in (7.2.3) is completely specified by $a_{ij}(s)$, $i \neq j = 1, 2, \ldots, N$; $s \in S$. Therefore it suffices to specify the form of $a_{ij}(s)$.

$$v_1 : a_{ij}(s) = a_{ij} / \pi_{ij} \qquad \ldots(7.2.4)$$

where $\pi_{ij} = \sum_{s \ni ij} p(s) \quad i \neq j = 1, 2, \ldots, N$

$$v_2 : a_{ij}(s) = \frac{X}{N^2 x_s}\left(\frac{X}{x_s} - \frac{N-1}{n-1}\right) \qquad \ldots(7.2.5)$$

$$v_3 : a_{ij}(s) = \frac{e_{ij}}{M_2 p(s)} - \frac{N^{-2}}{\pi_{ij}} \qquad \ldots(7.2.6)$$

where $e_{ij} = a_{ij} + N^{-2}$

$$v_4 : a_{ij}(s) = \frac{e_{ij}}{\pi_{ij}} - \frac{N^{-2}}{M_2 p(s)} \qquad \ldots(7.2.7)$$

$$v_5 : a_{ij}(s) = a_{ij}/M_2 p(s) \qquad \ldots(7.2.8)$$

Estimators (7.2.4) and (7.2.5) are due to Rao and Vijayan (1977) and estimators (7.2.6), (7.2.7) and (7.2.8) are due to Chaudhuri (1979). We propose yet another estimator before introducing a general class of estimators.

$$v_6 : a_{ij}(s) = \frac{X^2}{N^2 x_s^2} - \frac{N^{-2}}{\pi_{ij}} \qquad \ldots(7.2.9)$$

We now define a general class C of estimators having form (7.2.3)

$$C : a_{ij}(s) = e_{ij} \frac{w_{ij}(s)}{\sum\limits_{s \ni ij} w_{ij}(s)p(s)} - \frac{u_{ij}(s)}{N^2 \sum\limits_{s \ni ij} u_{ij}(s)p(s)} \qquad \ldots(7.2.10)$$

where $u_{ij}(s)$ and $w_{ij}(s)$, $i \neq j = 1,2,\ldots,N$, $s \in S$ are real constants.

The estimators (7.2.4) to (7.2.9), as can be seen easily, are particular cases of (7.2.10). For example setting

$w_{ij}(s) = \frac{1}{M_2 p(s)}$ and $u_{ij}(s) = u(\neq 0)$, $i \neq j = 1,2,\ldots,N$; $s \in S$

we get (7.2.6). Further note that a convex combination of estimators satisfying (7.2.3) also satisfies (7.2.3).

We now turn our attention to sufficient conditions for the uniform nonnegativity of the various proposed estimators. To satisfy uniform nonnegativity, attempts have been made to obtain a set $\{a_{ij}(s)\}$, $i \neq j = 1,2,\ldots,N$; $s \in S$, satisfying the following conditions :

$$\text{(i)} \quad a_{ij}(s) \leq 0 \quad i \neq j = 1,2,\ldots,N \text{ ; } s \in S$$

$$\text{and (ii)} \quad \sum_{s \ni ij} a_{ij}(s)p(s) = a_{ij} \quad i \neq j = 1,2,\ldots,N. \qquad \ldots(7.2.11)$$

However such an attempt fails whenever any of the a_{ij}'s are positive. This is because of the following lemma.

<u>Lemma 7.2.1</u> The following two statements are equivalent.

(a) $\exists$ a set $\{a_{ij}(s) : i \neq j = 1,2,\ldots,N,\ s \in S\}$ satisfying (7.2.11)

(b) $a_{ij} \leq 0 \quad i \neq j = 1,2,\ldots,N$.

<u>Proof</u>. (a) => (b)

$$a_{ij}(s) \leq 0 \quad \forall\ i \neq j = 1,2,\ldots,N \text{ ; } s \in S$$

$$\Rightarrow \quad \sum_{s \ni ij} a_{ij}(s)p(s) \leq 0 \iff a_{ij} \leq 0 \quad \forall\ i \neq j = 1,2,\ldots,N.$$

(b) => (a)

$a_{ij} \leq 0 \quad 1 \leq i \neq j \leq N$. Define $a_{ij}(s) = a_{ij}/M_2 p(s) \quad 1 \leq i \neq j \leq N,\ s \in S$.

Hence the lemma.

In view of the above lemma it is interesting to note that the sufficient conditions (7.2.11) can be further relaxed. In the following theorem we also give a set of necessary conditions for the uniform nonnegativity of an estimator v having the form (7.2.3).

<u>Theorem 7.2.2</u> (a) A set of sufficient conditions for the existence of a uniformly nnuve is given by : For any n distinct labels $1,2,\ldots,n$ (say) from among the labels $1,2,\ldots,N$

$$\sum_{k=1}^{n-2} b_{i_k j_k} + \frac{1}{2} b_{i_{n-1} j_{n-1}} \leq 0 \qquad \ldots(7.2.12)$$

for any $(n-1)$ distinct pairs (i_k, j_k), $1 \leq i_k < j_k \leq n$ with $b_{ij} = a_{ij} x_i x_j$.

(b) A set of necessary conditions for the uniform nonnegativity of an estimator v in (7.2.3) is given by : For any n distinct labels $1,2,\ldots,n$ (say) from among the labels $1,2,\ldots,N$ and a label i, $1 \leq i \leq n$, there can be at most $(n-2)$ labels j, $1 \leq j \neq i \leq n$ such that $a_{ij}(s)x_i x_j > 0$.

<u>Proof</u>. (a) Let (7.2.12) be true. Define $r_i = y_i / x_i$ and $b_{ij}(s) = b_{ij} / M_2 p(s)$ then

$$- M_2 p(s) v(s) = \sum_{i < j \in s} b_{ij} (r_i - r_j)^2$$

Because of (7.2.12) there can be at most $(n-2)$ distinct pairs (i,j) ; $i < j \in s$ with $b_{ij} > 0$. Let $b_{i_1 j_1}, b_{i_2 j_2}, \ldots, b_{i_{n-1} j_{n-1}}$ be all positive, $i_k < j_k \in s$, $k=1,2,\ldots,n-2$; further without loss of generality, let $b_{i_1 j_1} = \max_{1 \leq k \leq n-2} b_{i_k j_k}$. Therefore

$$- M_2 p(s)v(s) \leq \left(\sum_{k=1}^{n-2} b_{i_k j_k}\right)(r_{i_1} - r_{j_1})^2 + \sum_{\substack{i_k < j_k \in s \\ k > n-2}} b_{i_k j_k}(r_{i_k} - r_{j_k})^2 .$$

Now there exists a label $i \in s$, $i \neq i_1, j_1$ such that $b_{i_1 i}$ and $b_{j_1 i}$ are nonpositive. If not then $\forall\, i \in s$, $i \neq i_1, j_1$ either $b_{i_1 i} > 0$ or $b_{j_1 i} > 0$. (Note that $b_{ij} = b_{ji}$; $i,j = 1,2,\ldots,N$.)

$\Rightarrow$ There are at least $(n-2)$ distinct pairs for which $b_{ij} > 0$, $i < j \in s$. Further $b_{i_1 j_1}$ is also positive. Thus, in all, there are at least $(n-1)$ distinct pairs for which $b_{ij} > 0$, $i < j \in s$. This contradicts (7.2.12). Therefore

$$- M_2 p(s)v(s) \leq \left[b_{i_1 i} + 2 \sum_{k=1}^{n-2} b_{i_k j_k} \right] (r_{i_1} - r_i)^2$$

$$+ \left[b_{j_1 i} + 2 \sum_{k=1}^{n-2} b_{i_k j_k} \right] (r_{j_1} - r_i)^2$$

+ remaining terms which are already nonpositive,

since $(a-b)^2 \leq 2(a-c)^2 + 2(b-c)^2$.

Now using (7.2.12) we get $v(s) \geq 0$.

This proves the sufficiency.

(b) Let v be any uniformly nnuve. If possible, let for a set of labels $1,2,\ldots,n$ (say) and a label i_0, $1 \leq i_0 \leq n$

$$b_{i_o j}(s) > 0, \; 1 \leq i_o \neq j \leq n$$

where $b_{ij}(s) = a_{ij}(s)x_i x_j$.

Without loss of generality, let $i_o = 1$ and $\min_{2 \leq j \leq n} (r_1 - r_j)^2$ be $(r_1 - r_2)^2$. Now

$$-v(s) = \sum_{i < j \in s} b_{ij}(s)(r_i - r_j)^2$$

$$= \sum_{j=2}^{n} b_{1j}(s)(r_1 - r_j)^2 + \sum_{2 \leq i < j \in s} b_{ij}(s)(r_i - r_j)^2$$

$$\geq (r_1 - r_2)^2 \sum_{j=2}^{n} b_{1j}(s) + \sum_{2 \leq i < j \in s} b_{ij}(s)(r_i - r_j)^2 .$$

Observe that the first term in the above expression depends on r_1 whereas the second term is independent of r_1. Therefore for sufficiently large r_1 the first term can be made to dominate the second term i.e. $v(s)$ can be made negative. But this leads to a contradiction. Hence the necessity.

Corollary 7.2.1 A set of sufficient conditions for the uniform nonnegativity of an estimator v of the form (7.2.3) is given by: For a sample $s = \{i_1, i_2, \ldots, i_n\}$, $s \in S$ and for any $(n-1)$ distinct pairs (i_k, j_k), $i_k < j_k \in s$, $k = 1, 2, \ldots, (n-1)$

$$2 \sum_{k=1}^{n-2} b_{i_k j_k}(s) + b_{i_{n-1} j_{n-1}}(s) \leq 0 \qquad \ldots(7.2.13)$$

where $b_{ij}(s) = a_{ij}(s)x_i x_j$.

Remark 7.2.1 Note that (i) of (7.2.11) clearly implies (7.2.13). In other words the set of sufficient conditions (7.2.13) is less stringent than the set of sufficient conditions (i) of (7.2.11).

Though the conditions (7.2.13) appear rather involved they can be verified easily. For a given sample s let

$$b_{i_1 j_1}(s) \geq b_{i_2 j_2}(s) \geq \dots \geq b_{i_{n-1} j_{n-1}}(s) \;;\; i_k < j_k \varepsilon\, s \text{, be}$$

the first (n-1) maximum values. Verifying (7.2.13) is equivalent to checking if the following is true :

$$2 \sum_{k=1}^{n-2} b_{i_k j_k}(s) + b_{i_{n-1} j_{n-1}}(s) \leq 0 .$$

Remark 7.2.2. We may think of modifying class C to C' as follows.

$$C' : a_{ij}(s) = \frac{a_{ij} w_{ij}(s)}{\sum_{s \ni ij} w_{ij}(s) p(s)} \quad \text{if} \quad a_{ij} \leq 0 \qquad \dots (7.2.14)$$

$$= \text{same as (7.2.10)} \quad \text{if} \quad a_{ij} > 0$$

where the constants $w_{ij}(s)$ are chosen to be nonnegative whenever $a_{ij} \leq 0$. The advantage of (7.2.14) over (7.2.10) is that whenever $a_{ij} \leq 0$, $a_{ij}(s)$ are also nonpositive in (7.2.14) which may not be true for (7.2.10).

Remark 7.2.3. An analogous theory can be developed for the strategy (πps, t_{HT}).

Remark 7.2.4. Arnab (1979) and Chaudhuri and Arnab (1981) considered various unbiased estimators of $V(t_R)$ along with sufficient conditions for their nonnegativity, proposed in the literature by various authors prior to Rao and Vijayan's (1977) result regarding the necessary form of nnuve. They proved that for none of the estimators they considered the proposed conditions for uniform nonnegativity can be satisfied ($x_i \neq$ constant). However, they tacitly ignored the fact that none of the estimators they considered have the necessary form of nnuve (7.2.3). It is a matter of simple verification that for $n = 2$ none of the estimators agree with the unique nnuve. In view of this it is no wonder that the sufficient conditions for the uniform nonnegativity are not satisfied for any of the estimators they considered. Further, if the main objective is to obtain uniformly nnuves then there is no point in studying the estimators not having the necessary form (7.2.3) of nnuve.

Remark 7.2.5 It is interesting to note how a proper transformation on auxiliary variate helps in getting uniformly nnuves. Let $z_i = x_i + d$ be new auxiliary variate. If $Z = \sum_{i=1}^{N} z_i$ and $z_s = \sum_{i \in s} z_i$, $s \in S$, then

$$\frac{Z}{M_1} \sum_{s \ni i j} \frac{1}{z_s} = \frac{X + Nd}{M_1} \sum_{s \ni ij} \frac{1}{x_s + nd} \leq \frac{X + Nd}{x_{s_o} + nd} \frac{n-1}{N-1}$$

where $x_{s_o} = \min_{s \in S} x_s$

Thus for a Midzuno-Sen strategy based on new variate $z(d)$, $d \geq d_o = [(n-1)X - (N-1)x_{s_o}]/(N-n)$ all the a_{ij}'s, $i \neq j$, become negative and that ensures the existence of uniformly nnuves.

In using the new Midzuno-Sen strategy we do not have any idea about the change in efficiency. However, it is worthwhile to use new Midzuno-Sen strategy if loss in efficiency is insignificant.

7.3 Biased Estimators.

Nonnegative variance estimators can be easily constructed from any unbiased variance estimators by replacing the negative values by some nonnegative quantities. Clearly such modified estimators would be biased. An often used modification is to truncate estimators at zero. However such a truncation is not satisfactory when true value of variance is significantly large. Here we consider nonnegative estimators obtained from unbiased estimators by proper substitutions for negative values. For some of the substitutions we actually establish that there is indeed a reduction in MSE.

To find proper substitutions we make use of superpopulation model. Following superpopulation model is reasonable in situations where the ratio estimator is appropriate. Let $y_1, y_2, \ldots, y_N$ be a realization of N random variables $Y_1, Y_2, \ldots, Y_N$ with a joint distribution ξ which is specified, though not completely, by the first two moments,

$$\left.\begin{aligned} E_\xi(Y_i) &= \beta x_i \\ E_\xi(Y_i - \beta x_i)^2 &= \sigma^2 x_i \end{aligned}\right\} \quad i = 1,2,\ldots,N \qquad \ldots(7.3.1)$$

and $\quad E_\xi(Y_i - \beta x_i)(Y_j - \beta x_j) = 0 \quad i \neq j = 1,2,\ldots,N$

where $\sigma^2 > 0$ and β are unknown parameters of the model and E_ξ denotes the expectation w.r.t. the model ξ.

Note that model (7.3.1) is a particular case $(g=1)$ of the model (2.1.1).

Based on a sample s, the least square estimator of $\beta\bar{X}$ under the model (7.3.1) is given by $\frac{y_s}{x_s}\bar{X}$. i.e. the ratio estimator t_R is the least square estimator for $\beta\bar{X}$. Further,

$$V_\xi(t_R) = E_\xi(t_R - \beta\bar{X})^2 = \sigma^2 \bar{X}^2/x_s \, .$$

The least square estimator of σ^2 under (7.3.1) is given by

$$\frac{1}{n-1} \sum_{i\varepsilon s} \frac{1}{x_i} (y_i - \frac{y_s}{x_s} x_i)^2 .$$

Therefore the least square estimator of $V_\xi(t_R)$ is given by

$$w(s) = \frac{\bar{X}^2}{x_s} \frac{1}{n-1} \sum_{i\varepsilon s} \frac{1}{x_i} (y_i - \frac{y_s}{x_s} x_i)^2 \qquad \ldots(7.3.2)$$

Rao and Vijayan (1977) considered the following substitution. Whenever any nnuve is negative, replace it by the least square estimator (7.3.2) of the model variance, $V_\xi(t_R)$, of the ratio estimator. We shall actually demonstrate how the use of this modification results in a reduction in MSE. Corresponding to the estimator v_2, Rao and Vijayan (1977) considered the following modification :

$$v(s) = v_2(s) \quad \text{if} \quad v_2(s) > 0$$
$$= w(s) \quad \text{if} \quad v_2(s) \leq 0 \qquad \ldots(7.3.3)$$

where $v_2(s)$ and $w(s)$ are given by (7.2.5) and (7.3.2) respectively

Let us consider the following modification which is almost the same as (7.3.3).

$$v^*(s) = v_2(s) \quad \text{if} \quad \frac{X}{x_s} \le \frac{N-1}{n-1}$$

$$= \alpha_s w(s) \quad \text{if} \quad \frac{N-1}{n-1} < \frac{N-1}{n-1-\alpha_s} = \frac{X}{x_s} < \frac{N-1}{n-2} \; ; \; 0 < \alpha_s < 1 \qquad \ldots(7.3.4)$$

$$= w(s) \quad \text{if} \quad \frac{N-1}{n-2} \le \frac{X}{x_s} .$$

We consider yet another modification.

$$v^{**}(s) = v_2(s) \quad \text{if} \quad \frac{X}{x_s} \le \frac{N-1}{n-1}$$

$$= \delta_s \frac{X-x_s}{X} w(s) \quad \text{if} \quad \frac{N-1}{n-1} < \frac{N-1-\delta_s}{n-1-\delta_s} = \frac{X}{x_s} < \frac{N-2}{n-2} \qquad \ldots(7.3.5)$$

$$0 < \delta_s < 1$$

$$= \frac{X-x_s}{X} w(s) \quad \text{if} \quad \frac{N-2}{n-2} \le \frac{X}{x_s} .$$

We need the following lemma to demonstrate the reduction in MSE.

Let X and Y be two discrete random variables on some probability space as follows,

$$X = x_1, x_2, \ldots, x_r, x_{r+1}, \ldots, x_n$$

$$Y = y_1, y_2, \ldots \quad \ldots \quad \ldots \quad , y_n$$

such that $\text{Prob}\,[X = x_i] = \text{Prob}\,[Y = y_i] = p_i$, $1 \le i \le n$.

Let $x_i \le 0$ for $1 \le i \le r$ and $x_i > 0$ for $r < i \le n$ and $y_i > 0$ for $1 \le i \le n$.

Define a new random variable Z on the same probability space as

$$Z = X \quad \text{if} \quad X > 0$$

$$= Y \quad \text{if} \quad X \leq 0 .$$

Clearly Z has the following distribution

$$Z = y_1, y_2, \ldots, y_r, x_{r+1}, \ldots, x_n$$

$$\text{Prob} = p_1, p_2, \ldots, p_r, p_{r+1}, \ldots, p_n .$$

Let $E(X) = \sum_{i=1}^{n} p_i x_i = \mu$ be positive.

Let $V(X) = \sum_{i=1}^{n} p_i (x_i - \mu)^2$ and $MSE(Z) = \sum_{i=1}^{r} p_i (y_i - \mu)^2 + \sum_{i=r+1}^{n} p_i (x_i - \mu)^2$.

Lemma 7.3.1 In the above set up if $y_i \leq -x_i$ for $1 \leq i \leq r$ then

$$MSE(Z) \leq V(X).$$

Proof. $x_i \leq 0 \leq y_i \leq -x_i$ for $1 \leq i \leq r$ and $\mu > 0$.

Hence $|y_i - \mu| \leq \mu - x_i$, $1 \leq i \leq r$

or $(y_i - \mu)^2 \leq (\mu - x_i)^2$, $1 \leq i \leq r$

$$\Rightarrow \sum_{i=1}^{r} p_i (y_i - \mu)^2 \leq \sum_{i=1}^{r} p_i (x_i - \mu)^2$$

$$\Rightarrow MSE(Z) \leq V(X) .$$

This proves the lemma.

We now prove the following theorem regarding the reduction in MSE.

Theorem 7.3.1. For the modifications v^* and v^{**} given by (7.3.4) and (7.3.5) respectively we have,

(a) $MSE(v^*) \leq V(v_2)$

(b) $MSE(v^{**}) \leq V(v_2)$

where $v_2(s)$ is given by (7.2.5) and MSE and V are taken w.r.t. the Midzuno-Sen sampling scheme (7.1.1).

<u>Proof</u>. (a) Observe that

$$\sum_{i<j\varepsilon s} x_i x_j \left(\frac{y_i}{x_i} - \frac{y_j}{x_j}\right)^2 = x_s \sum_{i\varepsilon s} \frac{y_i^2}{x_i} - y_s^2 .$$

Therefore $v_2(s)$ can be expressed as

$$v_2(s) = \frac{X}{N^2 x_s}\left(\frac{N-1}{n-1} - \frac{X}{x_s}\right)\left[x_s \sum_{i\varepsilon s} \frac{y_i^2}{x_i} - y_s^2\right] .$$

Also simplifying $w(s)$ we have,

$$w(s) = \frac{X^2}{N^2(n-1)x_s^2}\left[x_s \sum_{i\varepsilon s} \frac{y_i^2}{x_i} - y_s^2\right].$$

Hence $w(s) \leq -v_2(s)$ is equivalent to

$$\frac{X^2}{(n-1)x_s^2} \leq -\frac{X}{x_s}\left(\frac{N-1}{n-1} - \frac{X}{x_s}\right)$$

or $\frac{N-1}{n-2} \leq \frac{X}{x_s}$.

Further $\alpha_s w(s) \leq -v_2(s)$ is equivalent to

$$\frac{N-1}{n-1} \frac{1}{1 - \frac{\alpha_s}{n-1}} \leq \frac{X}{x_s} .$$

Thus using Lemma 7.3.1, we have,

$$MSE(v^{*}) \leq V(v_2) .$$

On the same lines we can prove

$$MSE(v^{**}) \leq V(v_2) .$$

Hence the theorem.

We empirically investigate the relative performances of the estimators v_r, $1 \leq r \leq 6$, defined in Section 7.2. We choose four populations, numbered 15, 23, 24 and 25, considered by Rao and Vijayan (1977). We consider the cases $n = 2$ and $n = 4$. Table 7.3.1 gives the values of RE, PR, RE^* and B^* where

$$RE(r) = V(v_1)/V(v_r) , \quad PR(r) = \text{Prob}\,[v_r \leq 0] , \quad 1 \leq r \leq 6 ,$$

$$RE^*(r) = MSE(v_1^*)/MSE(v_r^*) \text{ and finally,}$$

$$B^*(r) = |E(v_r^*) - V(t_R)| \Big/ [MSE(v_r^*)]^{1/2}$$

with

$$v_r^*(s) = v_r(s) \text{ if } v_r(s) > 0 \qquad 1 \leq r \leq 6 ,$$
$$= w(s) \text{ if } v_r(s) \leq 0$$

$w(s)$ is given by (7.3.2).

Our investigations indicate that the estimator v_4 is most efficient in the sense of minimum variance and probability of it being negative is also very small. Estimator v_4 is closely followed by the estimator v_5.

Table 7.3.1

Population no.15					N = 10			
n→	3				4			
r↓	RE	PR	RE*	B*	RE	PR	RE*	B*
1	1.00	0.09	1.00	0.13	1.00	0.06	1.00	0.09
2	0.84	0.13	0.98	0.21	0.58	0.18	0.77	0.31
3	0.74	0.16	0.85	0.21	0.51	0.21	0.63	0.31
4	1.37	0.00	1.35	0.00	1.90	0.00	1.83	0.00
5	1.10	0.07	1.10	0.12	1.15	0.05	1.16	0.10
6	0.56	0.17	0.72	0.25	0.29	0.28	0.43	0.42
Population no.23					N = 8			
1	1.00	0.02	1.00	0.03	1.00	0.00	1.00	0.00
2	0.78	0.09	0.88	0.14	0.48	0.16	0.62	0.31
3	0.71	0.11	0.78	0.13	0.43	0.18	0.52	0.30
4	1.32	0.00	1.33	0.00	1.38	0.00	1.38	0.00
5	1.10	0.01	1.09	0.03	1.15	0.00	1.15	0.00
6	0.49	0.18	0.61	0.25	0.21	0.32	0.29	0.48
Population no.24					N = 8			
1	1.00	0.04	1.00	0.05	1.00	0.03	1.00	0.04
2	0.85	0.09	0.93	0.12	0.57	0.16	0.69	0.26
3	0.77	0.11	0.83	0.12	0.50	0.18	0.58	0.24
4	1.31	0.00	1.23	0.00	1.45	0.00	1.39	0.00
5	1.09	0.04	1.10	0.05	1.13	0.03	1.14	0.04
6	0.58	0.14	0.66	0.18	0.27	0.27	0.34	0.38

Table 7.3.1 (contd.)

Population no. 25					N = 8			
$n \rightarrow$	3				4			
$r \downarrow$	RE	PR	RE*	B*	RE	PR	RE*	B*
1	1.00	0.09	1.00	0.14	1.00	0.06	1.00	0.13
2	0.65	0.14	0.81	0.25	0.29	0.22	0.42	0.46
3	0.56	0.19	0.68	0.28	0.27	0.25	0.34	0.45
4	1.24	0.02	1.34	0.01	0.79	0.06	0.87	0.12
5	1.07	0.06	1.07	0.13	1.15	0.06	1.14	0.14
6	0.45	0.23	0.45	0.35	0.11	0.32	.17	0.54

7.4 Use of Stratification

In this section we see how the principle of stratification helps in getting nnuve. As mentioned in Section 7.3 the model (7.3.1) is reasonable in situations where the ratio estimator is appropriate. If we stratify the population into L strata and use Midzuno-Sen sampling scheme and the ratio estimator within each stratum then for 'optimal allocation' there is a gain due to stratification i.e. the above strategy is superior to the overall Midzuno-Sen strategy (p_M, t_R) in the sense of smaller expected variance under the model (7.3.1). Equivalently we are using the measure of uncertainty $M_1(p,t)$ of (1.22) which is given by

$$M_1(p,t) = E_{\xi} E_p (t - \overline{Y})^2 \qquad \ldots(7.4.1)$$

where E_ξ denotes expectation under the model (7.3.1).

We further establish that if the stratification is done so as to make the strata homogeneous w.r.t. the x-values then the chances of getting a uniformly nnuve are improved.

Let the population be divided into L strata of sizes $N_1, N_2, \ldots, N_L$ respectively. Let

$$\bar{X}_h = \frac{X_h}{N_h} = \frac{1}{N_h} \sum_{i \in S_h} x_i$$

where S_h, $1 \leq h \leq L$, denotes the hth stratum.

Consider a strategy $H_2(p_{Mst}, t_{Rst})$ that consists of a design that selects units independently from different strata employing Midzuno-Sen sampling scheme within each stratum and the estimator

$$t_{Rst} = \sum_{h=1}^{L} \frac{N_h}{N} t_{Rh} \qquad \ldots(7.4.2)$$

where $t_{Rh} = \bar{X}_h \sum_{i \in s_h} y_i \Big/ \sum_{i \in s_h} x_i$, $\quad 1 \leq h \leq L$

with s_h denoting the set of labels of the units sampled from the hth stratum, $1 \leq h \leq L$.

Now the variance of the strategy $(p_M, t_R) = H_1$ (say) is given by (7.1.4) and it is easy to see that

$$V(H_2) = -\sum_{h=1}^{L} \frac{N_h^2}{N^2} \sum_{i < j \in S_h} a_{ij}^{(h)} x_i x_j \left(\frac{y_i}{x_i} - \frac{y_j}{x_j}\right)^2 \qquad \ldots(7.4.3)$$

where $a_{ij}^{(h)} = \dfrac{X_h}{N_h^2 \binom{N_h - 1}{n_h - 1}} \sum_{s_h \ni i, j} \dfrac{1}{x_{s_h}} - \dfrac{1}{N_h^2}$ $\quad 1 \leq i \neq j \leq N_h$ and $1 \leq h \leq L$

with n_h , $1 \leq h \leq L$, denoting the number of units sampled from the hth stratum and $x_{s_h} = \sum_{i \in s_h} x_i$.

Under the model (7.3.1)

$$E_\xi V(H_1) = -\sum_{i<j}^{N} a_{ij} x_i x_j E_\xi (Y_i / x_i - Y_j / x_j)^2$$

$$= -\sigma^2 \sum_{i<j}^{N} a_{ij} x_i x_j (1/x_i + 1/x_j)$$

$$= -\sigma^2 \sum_{i<j}^{N} a_{ij} (x_i + x_j) .$$

Simplifying we get,

$$E_\xi V(H_1) = (N - n)\sigma^2 X / nN^2 . \qquad \ldots(7.4.4)$$

Similarly we have,

$$E_\xi V(H_2) = \frac{\sigma^2}{N^2} \left[\sum_{h=1}^{L} \frac{N_h X_h}{n_h} - X \right] \qquad \ldots(7.4.5)$$

It is easy to see that the optimal allocation (that minimizes (7.4.5) subject to the condition $\sum_{h=1}^{L} n_h = n$) is given by

$$n_h^2 \propto (N_h X_h) . \qquad \ldots(7.4.6)$$

Under the optimal allocation (7.4.6) we have,

$$E_\xi V(H_2) = \frac{\sigma^2}{N^2} \left[\frac{1}{n} \left(\sum_{h=1}^{L} \sqrt{N_h X_h} \right)^2 - X \right] . \qquad \ldots(7.4.7)$$

The following theorem establishes that there is indeed a gain due to stratification.

<u>Theorem 7.4.1</u> For the optimal allocation (7.4.6) the strategy H_2 is superior to the overall Midzuno-Sen strategy H_1 w.r.t. the measure of uncertainty (7.4.1) (i.e. in the sense of smaller expected variance) under the model 7.3.1.

<u>Proof</u>. Comparing (7.4.4) and (7.4.7) we have, for the optimal allocation (7.4.6),

$$E_\xi [V(H_1) - V(H_2)] = \frac{\sigma^2}{nN^2} \left[NX - \left(\sum_{h=1}^{L} \sqrt{N_h X_h} \right)^2 \right]$$

which is always nonnegative by Cauchy-Schwartz inequality.

This proves the theorem.

We now give an example to see how stratification helps in getting a uniformly nnuve.

<u>Example 7.4.1</u>. We choose a population, numbered 15, considered by Rao and Vijayan (1977). We rearrange the units in increasing order of x-values.

i	1	2	3	4	5	6	7	8	9	10
x	75	101	125	163	254	254	326	359	442	559

Consider the case $n = 4$ and the estimator v_2 given by (7.2.5). Note that $x_{s_o} = \min_{s \varepsilon S} x_s = 464$ and $X = 2658$.

Therefore the condition $x_{s_o} \geq \frac{n-1}{N-1} X$ is not satisfied.

However if we stratify the population into two strata of sizes $N_1 = 6$ and $N_2 = 4$ then $X_1 = 972$ and $X_2 = 1586$.

The optimal allocation (7.4.6) to the nearest integers is given by

$$n_1 = n_2 = 2$$

Further $x_{s_1} = \min_{s \in S_1} x_s = 176$ and $x_{s_2} = \min_{s \in S_2} x_s = 685$.

Let v_{2st} be the analogue of v_2 in the stratified set up.

Observe that $x_{s_1} > \frac{n_1 - 1}{N_1 - 1} X_1$ and $x_{s_2} > \frac{n_2 - 1}{N_2 - 1} X_2$ are both satisfied.

Therefore v_{2st} is uniformly nonnegative whereas v_2 may not be so. Though the allocation $n_1 = n_2 = 2$ is a little different from the optimal allocation, we have for $n_1 = n_2 = 2$

$$E_\xi [V(H_1) - V(H_2)] > 0 .$$

Thus in this section we have seen how the stratification technique improves the chances of getting a uniformly nnuve and at the same time retains the advantages of the ratio estimator.

7.5 Extension to Positive Definite Matrices.

As observed by Rao and Vijayan (1977), analogous treatment can be given to any linear strategy (p,t) with $MSE(p,t) = \sum\sum_{ij}^{N} d_{ij} y_i y_j$ where $D = (d_{ij})$ is an $N \times N$ nonnegative definite matrix. However, if D happens to be strictly positive definite then their technique cannot be applied to get nnuve. In this section, we deal with positive definite matrices.

Here we shall use the following notations. Let $\underset{\sim}{0}$ be $n \times 1$ zero vector and $\underset{\sim}{x}$ be any $n \times 1$ vector then

$\underset{\sim}{x}$ is positive, $\underset{\sim}{x} > \underset{\sim}{0} \iff x_i > 0$, $1 \leq i \leq n$,

$\underset{\sim}{x}$ is nonnegative, $\underset{\sim}{x} \geqq \underset{\sim}{0} \iff x_i \geq 0$, $1 \leq i \leq n$,

$\underset{\sim}{x}$ is semipositive, $\underset{\sim}{x} \dot{\geq} \underset{\sim}{0} \iff \underset{\sim}{x} \geqq \underset{\sim}{0}$ and $\underset{\sim}{x} \neq \underset{\sim}{0}$.

We first state a lemma due to Gale (vide, Mangasarian, 1969, p.35).

Lemma 7.5.1 (Gale, 1960) Given an $m \times n$ matrix A exactly one of the following is true.

$$\begin{aligned} &\text{(a)} \quad A\underset{\sim}{z} \leq \underset{\sim}{0}, \quad \underset{\sim}{z} \geq \underset{\sim}{0} \text{ has a solution } \underset{\sim}{z} \in R_n \\ &\text{(b)} \quad A'\underset{\sim}{x} \geq \underset{\sim}{0}, \quad \underset{\sim}{x} > \underset{\sim}{0} \text{ has a solution } \underset{\sim}{x} \in R_n . \end{aligned} \qquad \ldots(7.5.1)$$

Using above lemma we prove our main lemma.

Lemma 7.5.2 Given an $N \times N$ positive definite matrix D there exists a vector $\underset{\sim}{x} \in R_N$ such that $\underset{\sim}{x} > \underset{\sim}{0}$ and $D\underset{\sim}{x} \geq \underset{\sim}{0}$.

Proof. In view of the Lemma 7.5.1 it is enough to prove that for a positive definite matrix, (a) of (7.5.1) is not true.

If possible let there exist $\underset{\sim}{z} \geq \underset{\sim}{0}$ such that

$$D\underset{\sim}{z} \leq \underset{\sim}{0}.$$

Premultiplying both sides by $\underset{\sim}{z}'$ we get

$$\underset{\sim}{z}'D\underset{\sim}{z} \leqq 0 \quad \text{since} \quad \underset{\sim}{z}' \geq \underset{\sim}{0}.$$

But this is a contradiction since D is a positive definite matrix. Thus for a positive definite matrix, (b) of (7.5.1) is true i.e. there exists a vector $\underset{\sim}{x} \in R_N$ such that $\underset{\sim}{x} > \underset{\sim}{0}$ and $D\underset{\sim}{x} \geq \underset{\sim}{0}$. Hence the lemma.

Let (p,t) be any linear strategy for estimating the population mean $\bar{Y} = \frac{1}{N} \sum_{i=1}^{N} y_i$ with $MSE(p,t) = MSE(t) = \sum_{i}^{N} \sum_{j}^{N} d_{ij} y_i y_j$ where $D = (d_{ij})$ is an $N \times N$ positive definite matrix.

By Lemma 7.5.2 there exists a vector $\underset{\sim}{x} > \underset{\sim}{0}$ such that $D\underset{\sim}{x} \geq \underset{\sim}{0}$. Let $D\underset{\sim}{x} = \underset{\sim}{c}$ (say).

$$\text{Define} \quad Q = -\sum_{i<j=1}^{N} d_{ij} x_i x_j \left(\frac{y_i}{x_i} - \frac{y_j}{x_j}\right)^2 . \qquad \ldots(7.5.2)$$

Simplifying we get

$$Q = \sum_{i=1}^{N} \sum_{j=1}^{N} d_{ij} y_i y_j - \sum_{i=1}^{N} \frac{y_i^2}{x_i} c_i$$

$$\text{or} \quad MSE(t) = Q + \sum_{i=1}^{N} \frac{y_i^2 c_i}{x_i} . \qquad \ldots(7.5.3)$$

We may now think of estimating $MSE(t)$ unbiasedly by

$$mse(t) = -\sum_{i<j \in s} d_{ij}(s) x_i x_j \left(\frac{y_i}{x_i} - \frac{y_j}{x_j}\right)^2 + \sum_{i \in s} \frac{y_i^2}{x_i} \frac{c_i}{\pi_i} \qquad \ldots(7.5.4)$$

where $d_{ij}(s)$ are such that

$$d_{ij}(s) = 0 \text{ if } s \not\ni i \text{ or } s \not\ni j$$

$$\text{and} \quad \sum_{s \ni ij} d_{ij}^{(s)} p(s) = d_{ij} \quad i \neq j = 1, 2, \ldots, N$$

further $\pi_i = \sum_{s \ni i} p(s), \quad 1 \leq i \leq N$.

Note that the second term in the right hand side of (7.5.4) is already nonnegative. Therefore a set of sufficient conditions for the nonnegativity of the estimator $mse(t)$ in (7.5.4) can be given as follows :

For a sample $s \varepsilon S$, for any $(n-1)$ distinct pairs (i_k, j_k), $i_k < j_k \varepsilon s$

$$\sum_{k=1}^{n-2} d_{i_k j_k}(s) x_{i_k} x_{j_k} + \frac{1}{2} d_{i_{n-1} j_{n-1}} x_{i_{n-1}} x_{j_{n-1}} \leq 0. \quad \ldots(7.5.5)$$

Let us now consider an application of the above technique.

Example 7.5.1. Consider a strategy (q, t_1) that consists of a sampling scheme with varying probabilities of selection in n draws without replacement giving a selection probability $q(s)$ to a typical sample s and an estimator $t_1 = y_s / NM_1 q(s)$ (Sharma, 1970)

We now see how the above technique can be used to obtain nnuve for the strategy (q, t_1).

Clearly (q, t_1) is unbiased for the population mean $\bar{Y}$ and its variance is given by

$$V(t_1) = \sum_{i=1}^{N} \sum_{j=1}^{N} a_{ij} y_i y_j$$

where $$N^2 a_{ii} = \frac{1}{M_1^2} \sum_{s \ni i} \frac{1}{q(s)} - 1 \qquad i = 1, 2, \ldots, N$$

$$\ldots(7.5.6)$$

$$N^2 a_{ij} = \frac{1}{M_1^2} \sum_{s \ni ij} \frac{1}{q(s)} - 1 \quad i \neq j = 1, 2, \ldots, N.$$

Note that the Midzuno-Sen strategy is a particular case of the strategy (q, t_1) with $q(s) = x_s / M_1 X$. Also note that the matrix A given by (7.5.6) is positive definite except for Midzuno-Sen strategy. For if A is nonnegative definite then there exists a vector $\underset{\sim}{x}$ such that $A\underset{\sim}{x} = \underset{\sim}{0}$. Hence for the vector x the estimator t_1 becomes constant i.e. $x_s / NM_1 q(s) = \lambda$ (constant) or $q(s) \propto x_s$ i.e. $q(s)$ is a Midzuno-Sen sampling design. On the other hand for Midzuno-Sen strategy A is known to be nonnegative definite. Thus A is nonnegative definite if and only if (q, t_1) is a Midzuno-Sen strategy.

For positive definite matrix A, we rewrite $V(t_1)$ as

$$V(t_1) = - \sum_{i<j=1}^{N} a_{ij} x_i x_j \left(\frac{y_i}{x_i} - \frac{y_j}{x_j}\right)^2 + \sum_{i=1}^{N} \frac{y_i^2 c_i}{x_i} \quad \ldots(7.5.7)$$

where $\underset{\sim}{x} > \underset{\sim}{0}$ is such that $A\underset{\sim}{x} = \underset{\sim}{c} \geq \underset{\sim}{0}$.

Existence of such an $\underset{\sim}{x}$ is guaranteed by Lemma 7.5.2.

We estimate the variance (7.5.7) by

$$v(t_1) = - \sum_{i<j \in s} a_{ij}(s) x_i x_j \left(\frac{y_i}{x_i} - \frac{y_j}{x_j}\right)^2 + \sum_{i \in s} \frac{y_i^2}{x_i} \frac{c_i}{\pi_i} \quad \ldots(7.5.8)$$

where $a_{ij}(s) = 0$ if $s \not\ni i$ or $s \not\ni j$

and $\sum_{s \ni ij} a_{ij}(s) q(s) = a_{ij} \quad i \neq j = 1, 2, \ldots, N$.

a_{ij}'s are given by (7.5.6) and $\pi_i = \sum_{s \ni i} q(s), \quad 1 \leq i \leq N$.

Observe that $\sum_{i \in s} \frac{y_i^2}{\pi_i} \frac{c_i}{x_i}$ is already nonnegative. Further we treat $- \sum_{i<j=1}^{N} a_{ij} x_i x_j \left(\frac{y_i}{x_i} - \frac{y_j}{x_j}\right)^2$ analogous to $V(t_{\pi}$ in (7.1.4).

Finally, (7.5.5), with $d_{ij}(s)$ replaced by $a_{ij}(s)$, will work as a set of sufficient conditions for the uniform nonnegativity of the estimator (7.5.8).

Thus in this chapter we have explored various possibilities of improving the chances of getting nonnegative variance estimators.

CHAPTER 8

OPTIMAL ESTIMATORS IN THE CONTINUOUS SET UP

8.0 Summary

In Chapter 5 we took up a study under a continuous survey sampling model. Here we work in the same set up. In this Chapter we deal with some of the problems which were left out in Chapter 5. Under the regression model we use 'expected variance' as the criterion for the comparison of strategies. Here we have an opportunity to note how the practical problems in survey sampling can be viewed in the abstract set up. We see here how the problems of obtaining certain optimal estimators can be treated as the optimization problems on Banach spaces.

In this chapter we first investigate whether there exists a best p as well as ξ-unbiased linear estimator under the proposed model in the sense of minimum expected variance for a given design. We then find conditions for the existence of a best linear p-unbiased estimator again in the sense of minimum expected variance under the proposed model for a given design. We also give examples to get an idea about the optimal estimators.

8.1 Introduction

In Chapter 5 we presented a different analytical treatment of sampling and estimation. In this chapter we essentially pursue the study taken up in Chapter 5. We use the same basic set up as that in Section 5.1. Here we give a 'continuous treatment' to the problems considered in Section 2.2. We assume basic knowledge of Functional Analysis (Luenberger, 1968). In this chapter we see how a problem of obtaining an 'optimal' estimator gives rise to an interesting problem of optimization on Banach spaces. The results presented here show how the practical problems in survey sampling can be viewed in the abstract set up. Such results have their own academic appeal. To avoid frequent back reference we reproduce the essential part of the basic set up given in Section 5.1.

Consider a population of infinitely many pairs $(y(x), x)$, $x \geq 0$, defined on a probability space (Ω, A, ξ) such that the joint distribution of $y(x)$, $x \geq 0$, is unknown. The distribution of X (taking values $x \geq 0$), which is assumed to be continuous and known, is given by :

$$F(x) = \int_0^x f(u)du \, , \; x \geq 0 .$$

The problem under consideration is to estimate the population mean for the variate Y, namely,

$$m_Y = E_f(Y) = \int_0^\infty y(x)f(x)dx .$$

Any continuous probability measure Q on $\mathcal{B}$, the Borel σ-algebra of $R_n^+ = \{\underset{\sim}{x} : x_i > 0, 1 \leq i \leq n\}$, is called a sampling design. If $q(\underset{\sim}{x}) = dQ(\underset{\sim}{x})/d\underset{\sim}{x}$ then $q(\underset{\sim}{x})$ may be expressed as $q(\underset{\sim}{x}) = p(\underset{\sim}{x})f(\underset{\sim}{x})$, where $f(\underset{\sim}{x}) = \prod_{i=1}^{n} f(x_i)$. $p(\underset{\sim}{x})$ is called design function giving rise to the sampling design $Q(\underset{\sim}{x})$.

Here we consider a specific superpopulation model, namely regression model, induced by the probability space (Ω, A, ξ).

$$Y(x) = \beta x^b + Z(x), \; x \geq 0$$

where for every fixed $x \geq 0$

$$E_\xi(Z(x)) = 0, \; E_\xi(Z^2(x)) = \sigma^2 x^g \qquad \ldots(8.1.1)$$

and for every fixed $x \neq x'$, $x, x' \geq 0$

$$E_\xi(Z(x)Z(x')) = 0$$

where $\sigma^2 > 0$ and β are unknown model parameters whereas $b \geq 0$ and $g \in [0,2]$ may be known or unknown. Note that the model (8.1.1) is same as the model (5.1.1).

We assume that $Y(X)$ is square integrable w.r.t. the product probability $(F \times \xi)$. To judge the performance of a strategy (p,t) we use the following measure of uncertainty

$$M_1(p,t) = E_\xi E_p(t - m_Y)^2. \qquad \ldots(8.1.2)$$

In actually computing (8.1.2) we assume that the population conforms to the model (8.1.1) with $b = 1$ and $g \in [0,2]$ known. Note that the measure of uncertainty (8.1.2) is same as (5.1.4).

In this chapter we try to investigate the following :

(a) For a given design does there exist a best p as well as ξ-unbiased linear estimator, under the model (8.1.1) with $b=1$ and g known, w.r.t. the measure of uncertainty (8.1.2)?

(b) For a given design does there exist a best p-unbiased linear estimator, under the model (8.1.1) with $b=1$, g known and the ratio σ^2/β^2 also known, w.r.t. the measure of uncertainty (8.1.2)?

In this chapter Σ will denote $\sum_{i=1}^{n}$ unless otherwise specified.

8.2 Two Existence Theorems.

A linear estimator t is of the form

$$t(\underset{\sim}{y}(x),\underset{\sim}{x}) = \Sigma\, a_i(\underset{\sim}{x})y(x_i) \qquad \ldots(8.2.1)$$

where $a_i(\underset{\sim}{x})$, $1 \leq i \leq n$, are $\mathcal{B}$-measurable functions. The condition of ξ-unbiasedness for the linear estimator (8.2.1) is given by

$$\Sigma\, a_i(\underset{\sim}{x})x_i = \mu = E_f(X) \quad \forall\, \underset{\sim}{x} \in R_n^+ . \qquad \ldots(8.2.2)$$

And the condition of p-unbiasedness for a strategy (p,t) where p is a design function and t is given by (8.2.1), is given by

$$\phi(\underset{\sim}{a}, z) = 1 \quad \forall\, z \geq 0 \qquad \ldots(8.2.3)$$

where $\underset{\sim}{a} = (a_1,a_2,\ldots,a_n)$, $\phi(\underset{\sim}{a}, z) = \Sigma\, \phi_i(\underset{\sim}{a}, z)$

and $$\phi_i(\underset{\sim}{a}, z) = \int_{R_{n-1}^+} a_i(\underset{\sim}{x})p(\underset{\sim}{x}) \prod_{j \neq i}^{n} f(x_j)dx_j . \qquad \ldots(8.2.4)$$

We first attempt to solve the problem (a) posed earlier. For a p as well as ξ-unbiased linear strategy (p,t) the measure of uncertainty (8.1.2) takes a simpler form, namely

$$M_1(p,t) = \sigma^2 \int_{R_n^+} (\Sigma\, a_i^2(\underset{\sim}{x}) x_i^g) p(\underset{\sim}{x}) f(\underset{\sim}{x}) d\underset{\sim}{x} + \beta^2 \mu^2 - E_\xi m_Y^2 \,. \qquad \ldots(8.2.5)$$

Thus for a given design function our problem is to minimize (8.2.5) subject to the conditions (8.2.2) and (8.2.3).

For a design function $p(\underset{\sim}{x})$ define

$$q_i(x_i) = \int_{R_{n-1}^+} p(\underset{\sim}{x}) f(\underset{\sim}{x}) \prod_{j \neq i}^{n} dx_j \,. \qquad \ldots(8.2.6)$$

Let us assume that the given design function $p(\underset{\sim}{x})$ satisfies the following conditions.

For some fixed $\upsilon > 1$ and for every $i = 1,2,\ldots,n$

$$|x|^{2\upsilon / \upsilon-1} \text{ is } q_i(x) \text{ integrable}$$

$$\ldots(8.2.7)$$

and $$|q_i(x) / f(x)|^{2\upsilon-1 / \upsilon-1} \text{ is } f(x) \text{ integrable.}$$

The number υ is chosen as close to 1 as possible so that the conditions (8.2.7) are still satisfied.

Let $a_i(\underset{\sim}{x})$ be $\mathcal{B}$-measurable, $1 \leq i \leq n$ and $q(\underset{\sim}{x}) = p(\underset{\sim}{x}) f(\underset{\sim}{x})$.

Define $U = \{\underset{\sim}{a} : |a_i(\underset{\sim}{x})|^{2\upsilon}$ is $q(\underset{\sim}{x})$ integrable, $1 \leq i \leq n\}$

$V_1 = \{a : |a(\underset{\sim}{x})|^2$ is $q(\underset{\sim}{x})$ integrable$\}$

$V_2 = \{b(x) : |b(x)|^2$ is $f(x)$ integrable$\}$.

Note that with usual 'L_p' norms U, V_1, V_2 are all Banach spaces. Let $V = V_1 \times V_2$ and V^* be the dual space (the space of all bounded linear functionals on V w.r.t. usual L_2-norm) of V. Clearly $V^* = V$.

Let

$$G(\underset{\sim}{a}) = \int_{R_n^+} (\Sigma\, a_i^2(\underset{\sim}{x}) x_i^g) p(\underset{\sim}{x}) f(\underset{\sim}{x}) d\underset{\sim}{x}$$

$$H_1(\underset{\sim}{a}) = I(\underset{\sim}{x})(\Sigma\, a_i(\underset{\sim}{x}) x_i - \mu)$$

$$H_2(\underset{\sim}{a}) = I_1(x)(\emptyset(\underset{\sim}{a}, x) - 1)$$

$$H(\underset{\sim}{a}) = (H_1(\underset{\sim}{a}), H_2(\underset{\sim}{a}))$$

where $I(\underset{\sim}{x}) = 1$ if $\underset{\sim}{x} \in R_n^+$, $= 0$ otherwise, and $I_1(x) = 1$ if $x \geq 0$, $= 0$ if $x < 0$.

We are now in a position to formulate our problem as a familiar minimization problem on vector spaces.

$$\text{Minimise } G(\underset{\sim}{a}) \quad \text{subject to } H(\underset{\sim}{a}) = \theta \qquad \ldots(8.2.8)$$

where θ is the zero vector of V.

It can be checked that G and H are infinitely Fréchet differentiable. To solve (8.2.8) we make use of the Lagrangian multipliers technique. The Lagrangian corresponding to (8.2.8) is given by

$$L(\underset{\sim}{a}, v^*) = G(\underset{\sim}{a}) + v^* H(\underset{\sim}{a}) \qquad \ldots(8.2.9)$$

where $v^* \in V^* = V$.

It is known that if $(\underset{\sim}{a}_o, v_o^*)$ is the unconstrained minimum of (8.2.9) then $\underset{\sim}{a}_o$ solves (8.2.8). Now (8.2.9) can be written as

$$L(\underset{\sim}{a}, v^*) = \int_{R_n^+} (\Sigma a_i^2(\underset{\sim}{x}) x_i^g) p(\underset{\sim}{x}) f(\underset{\sim}{x}) d\underset{\sim}{x} - 2 \int_{R_n^+} \lambda(\underset{\sim}{x}) (\Sigma a_i(\underset{\sim}{x}) x_i - \mu) p(\underset{\sim}{x}) f(\underset{\sim}{x}) d\underset{\sim}{x}$$

$$- 2 \int_{R^+} b(x)(\emptyset(\underset{\sim}{a}, x) - 1) f(x) dx$$

where $v^* = -2(\lambda(\underset{\sim}{x}), b(x))$, $\lambda(\underset{\sim}{x}) \varepsilon V_1$ and $b(x) \varepsilon V_2$.

Let $\delta L(\underset{\sim}{a}, v^* ; \underset{\sim}{h}, w^*)$ be the derivative of $L(\underset{\sim}{a}, v^*)$ with the increment $(\underset{\sim}{h}, w^*)$.

Setting $\delta L(\underset{\sim}{a}, v^* ; \underset{\sim}{h}, w^*) = 0 \ \forall (\underset{\sim}{h}, w^*) \varepsilon U \times V^*$ we get

$$H(\underset{\sim}{a}) = \theta .$$

And $\forall \underset{\sim}{h} \varepsilon U$

$$\int_{R_n^+} (\Sigma a_i(\underset{\sim}{x}) h_i(x) x_i^g) p(\underset{\sim}{x}) f(\underset{\sim}{x}) d\underset{\sim}{x} - \int_{R_n^+} \lambda(\underset{\sim}{x}) (\Sigma h_i(\underset{\sim}{x}) x_i) p(\underset{\sim}{x}) f(\underset{\sim}{x}) d\underset{\sim}{x}$$

$$- \int_{R^+} b(x) \emptyset(\underset{\sim}{h}, x) f(x) dx = 0 . \quad \ldots(8.2.10)$$

Note that $\int_{R^+} b(x) \emptyset_i(\underset{\sim}{h}, x) f(x) dx = \int_{R_n^+} b(x_i) h_i(\underset{\sim}{x}) p(\underset{\sim}{x}) f(\underset{\sim}{x}) d\underset{\sim}{x}$.

Hence from (8.2.10), we get, $\forall \ h_i(\underset{\sim}{x}) \varepsilon U_1 = \{a : |a(x)|^{2\upsilon}$ is $q(\underset{\sim}{x})$ integrable$\}$,

$$\int_{R_n^+} h_i(\underset{\sim}{x}) [a_i(\underset{\sim}{x}) x_i^g - \lambda(\underset{\sim}{x}) x_i - b(x_i)] p(\underset{\sim}{x}) f(\underset{\sim}{x}) d\underset{\sim}{x} = 0 ; \quad 1 \leq i \leq n.$$

Hence $a_i(\underset{\sim}{x}) x_i^g = \lambda(\underset{\sim}{x}) x_i + b(x_i)$. ...(8.2.11)

Now using the constraint (8.2.2) we get

$$\lambda(\underset{\sim}{x}) = [\mu - \Sigma\, b(x_i)x_i^{1-g}]\, d(\underset{\sim}{x}) \qquad \ldots(8.2.12)$$

where $d(\underset{\sim}{x}) = 1/\Sigma\, x_i^{2-g}$.

Substituting the value of $\lambda(\underset{\sim}{x})$ from (8.2.12) in (8.2.11) we get

$$a_i(\underset{\sim}{x}) = b(x_i)x_i^{-g} + [\mu - \Sigma\, b(x_j)x_j^{1-g}]\, x_i^{1-g}\, d(\underset{\sim}{x}) . \qquad \ldots(8.2.13)$$

From (8.2.13) we get

$$\int_{R^+_{n-1}} a_i(\underset{\sim}{x})p(\underset{\sim}{x}) \prod_{j\neq i}^{n} f(x_j)dx_j = b(x_i)\,[x_i^{-g}\, r_i(x_i) - x_i^{2-2g}\, c_i(x_i)]$$

$$+ \mu x_i^{1-g}\, c_i(x_i) - x_i^{1-g} \sum_{j\neq i}^{n} \int_{R^+} x_j^{1-g}\, b(x_j)c_{ij}(x_i,x_j)f(x_j)dx_j \qquad \ldots(8.2.14)$$

where $r_i(x_i) = q_i(x_i)/f(x_i), c_i(x_i) = \int_{R^+_{n-1}} d(\underset{\sim}{x})p(\underset{\sim}{x}) \prod_{j\neq i}^{n} f(x_j)dx_j$

and $c_{ij}(x_i\, , x_j) = \int_{R^+_{n-2}} d(\underset{\sim}{x})p(\underset{\sim}{x}) \prod_{k\neq i,j}^{n} f(x_k)dx_k$.

Now observe that $\sum_{j\neq i}^{n} \int_{R^+} x_j^{1-g}\, b(x_j)c_{ij}(x_i\, , x_j)f(x_j)dx_j$

$$= \sum_{j\neq i}^{n} \int_{R^+} x^{1-g}\, b(x)c_{ij}(x_i,x)f(x)dx$$

$$= \int_{R^+} x^{1-g}\, b(x)D_i(x_i\, , x)f(x)dx$$

where $D_i(x_i\, , x) = \sum_{j\neq i}^{n} c_{ij}(x_i\, , x)$.

Substituting this in (8.2.14) we get,

$$\phi_i(\underset{\sim}{a}, x) = b(x)(x^{-g} r_i(x) - x^{2-2g} c_i(x)) + \mu x^{1-g} c_i(x)$$
$$- x^{1-g} \int_{R^+} b(w) w^{1-g} D_i(x, w) f(w) dw .$$

Now using the constraint (8.2.3) we get

$$1 = b(x) [x^{-g} \Sigma r_i(x) - x^{2-2g} \Sigma c_i(x)] + \mu x^{1-g} \Sigma c_i(x)$$
$$- x^{1-g} \int_{R^+} b(w) w^{1-g} D(x,w) f(w) dw . \qquad \ldots(8.2.15)$$

where $D(x, w) = \Sigma D_i(x, w)$.

Observe that $r_i(x) - x^{2-g} c_i(x) > 0$ for $1 \leq i \leq n$.

Hence (8.2.15) can be compressed to the familiar Fredholm equation as

$$m(x) = b(x) - \int_{R^+} K(x,w) b(w) f(w) dw \qquad \ldots(8.2.16)$$

where $m(x) = [1 - \mu x^{1-g} \Sigma c_i(x)] / [x^{-g} \Sigma r_i(x) - x^{2-2g} \Sigma c_i(x)]$

and $K(x,w) = [x w^{1-g} D(x,w)] / [\Sigma r_i(x) - x^{2-2g} \Sigma c_i(x)]$.

Thus determining the Lagrangian multiplier $b(x)$ is equivalent to solving the equation (8.2.16). If $b(x)$ is a solution to (8.2.16) then by substituting it in (8.2.13) we get n functions $\bar{a}_1, \bar{a}_2, \ldots, \bar{a}_n$. We now show that this $\underset{\sim}{\bar{a}}$ is indeed a solution to (8.2.8) Treating $L(\underset{\sim}{a}, v^*)$ as a functional in a we note that the second derivative with the increment $\underset{\sim}{h}, \underset{\sim}{h}$; $\delta^2 h(\underset{\sim}{a}, v^* ; \underset{\sim}{h}, \underset{\sim}{h}) > 0 \ \forall \underset{\sim}{h} \in U$

and the higher order derivatives are uniformly zero. Hence using Taylor's expansion, namely

$$L(\underset{\sim}{a}+\underset{\sim}{h}, v^*) = L(\underset{\sim}{a}, v^*) + \delta L(\underset{\sim}{a}, v^* ; \underset{\sim}{h}) + \delta^2 L(\underset{\sim}{a}, v^* ; \underset{\sim}{h}, \underset{\sim}{h}) / 2! \quad \sum_{m \geq 3} \delta^m L(\underset{\sim}{a}, v^* ; \underset{\sim}{h}, \underset{\sim}{h}, \ldots, \underset{\sim}{h}) / m! \qquad \ldots(8.2.17)$$

we get for $\underset{\sim}{\bar{a}}$, $L(\underset{\sim}{\bar{a}}) \leq L(\underset{\sim}{\bar{a}} + \underset{\sim}{h}) \quad \forall \underset{\sim}{h} \in U$.

As a matter of fact, depending on solutions to (8.2.16), even if $\underset{\sim}{\bar{a}}$ is not unique, again using (8.2.17) it is clear that the value $L(\underset{\sim}{a}, v^*)$ is same for all of them, i.e. we may get different vectors $\underset{\sim}{\bar{a}}$ leading to the same value of the functional $L(\underset{\sim}{a}, v^*)$.

We now state a theorem (Hochstadt, 1973) which can be used to solve (8.2.16).

<u>Theorem 8.2.1</u>. If $m(x) \in V_2$ and $\int_{R_2^+} K^2(x,w)f(x)f(w)\,dx\,dw < \infty$

then $$b(x) - \lambda \int_{R^+} K(x,w)b(w)f(w)\,dw = m(x) \qquad \ldots(8.2.18)$$

has a unique solution if and only if

$$b(x) - \lambda \int_{R^+} K(x,w)b(w)f(w)\,dw = 0 \qquad \ldots(8.2.19)$$

has only the trivial solution $b(x) = 0$.

If (8.2.19) has at least one nontrivial solution then (8.2.18) will have a solution if $\int_{R^+} m(x)\ell(x)f(x)\,dx = 0 \quad \forall \ell(x)$ satisfying

$$\ell(x) - \lambda \int_{R^+} K(w,x)\ell(w)f(w)\,dw = 0 .$$

We are now in a position to state our first existence theorem.

Theorem 8.2.2. For any design function $p(\underset{\sim}{x})$ satisfying (8.2.7) and for which (8.2.16) has a solution there exists a best p as well as ξ-unbiased linear estimator under the model (8.1.1) with $b = 1$ and g known w.r.t. the measure of uncertainty (8.1.2).

Example 8.2.1. Let us consider an example so as to get an idea about the above result.

$$\text{Let} \quad p(\underset{\sim}{x}) = A \sum_{i=1}^{n} x_i^{2-g} \prod_{i=1}^{n} p(x_i) \qquad \ldots(8.2.20)$$

where A is the normalizing constant and $p(\underset{\sim}{x})$ is such that (8.2.7) is satisfied and $x^{1-g}\, p(x)$ and $x^{g}/p(x)$ belong to V_2.

For $p(\underset{\sim}{x})$ given by (8.2.20), we have

$$\Sigma\, r_i(x) = Ap(x)\left[(n-1)\lambda_1\lambda_2 + x^{2-g}\lambda_3\right], \quad \Sigma\, c_i(x) = Ap(x)\lambda_3,$$

$$D(x,w) = n(n-1)\lambda_2\, Ap(x)p(w), \quad K(x,w) = xw^{1-g}\, p(w)/\lambda_1$$

and $m(x) = \lambda_4 x^{g}/p(x) - \lambda_5 x$

where $\lambda_1 = \int_{R^+} x^{2-g}\, p(x)f(x)dx$, $\lambda_2 = (\int_{R^+} p(x)f(x)dx)^{n-2}$,

$$\lambda_3 = \lambda_2 \int_{R^+} p(x)f(x)dx, \quad \lambda_4 = (n(n-1)\lambda_1\lambda_2 A)^{-1}, \quad \lambda_5 = \mu\lambda_3/(n-1)\lambda_1\lambda_2 .$$

It is easy to check that $A^{-1} = n\lambda_1\lambda_3$.

The equation (8.2.16) reduces to

$$b(x) - \frac{x}{\lambda_1} \int_{R^+} w^{1-g}\, p(w)b(w)f(w)dw = \frac{\lambda_4 x^{g}}{p(x)} - \lambda_5 x . \qquad \ldots(8.2.21)$$

Let $b_1 = \int_{R^+} w^{1-g}\, p(w)b(w)f(w)dw$. Then

$$b(x) - xb_1/\lambda_1 = \lambda_4 x^g / p(x) - \lambda_5 x . \qquad \ldots(8.2.22)$$

Multiplying both sides of (8.2.22) by $x^{1-g}\, p(x)$ and integrating we get

$$b_1 . 0 = \lambda_4 \mu - \lambda_5 \lambda_1 . \qquad \ldots(8.2.23)$$

This shows that (8.2.23) has a solution if and only if the right hand side of (8.2.23) is zero.

Note that $\lambda_4 \mu - \lambda_5 \lambda_1 = \dfrac{\mu}{n(n-1)\lambda_1 \lambda_2 A} - \dfrac{\mu \lambda_3}{(n-1)\lambda_2}$

$$= \frac{\mu}{n(n-1)\lambda_2 A \lambda_1}(1 - nA\lambda_1\lambda_3)$$

$$= 0 .$$

Thus $b_1 = \eta$ is a solution to (8.2.23) and hence

$b(x) = \eta x/\lambda_1 + \lambda_4 x^g / p(x) - \lambda_5 x$ is a solution to (8.2.21)

where η is any real number.

Substituting $b(x)$ in (8.2.13) we get a unique set of n functions $\bar{a}_1, \bar{a}_2, \ldots, \bar{a}_n$. i.e. they do not depend on any particular choice of η. Thus for $p(x)$ in (8.2.20) the best p as well as ξ-unbiased estimator is given by $\Sigma\, \bar{a}_i(\underset{\sim}{x}) y(x_i)$ where for $1 \leq i \leq n$

$$\bar{a}_i(\underset{\sim}{x}) = \lambda_4 / p(x_i) + [\mu - \Sigma\, x_j / p(x_j)]\, x_i^{1-g}\, d(\underset{\sim}{x}) . \qquad \ldots(8.2.24)$$

<u>Remark 8.2.1</u>. In particular, for $p(x_i) = x_i^{g-1}$, we get $\bar{a}_i(\underset{\sim}{x}) = \mu x_i^{1-g}\, d(\underset{\sim}{x})$ and further if $g = 2$ then $\bar{a}_i(\underset{\sim}{x}) = \mu / n x_i$.

We now proceed to solve the problem (b).

For a p-unbiased linear strategy the measure of uncertainty (8.1.2) takes the form

$$M_1(p,t) = \sigma^2 \int_{R_n^+} (\Sigma\, a_i^2(\underset{\sim}{x}) x_i^g) p(\underset{\sim}{x}) f(\underset{\sim}{x}) d\underset{\sim}{x} + \beta^2 \int_{R_n^+} (\Sigma\, a_i(\underset{\sim}{x}) x_i)^2 p(\underset{\sim}{x}) f(\underset{\sim}{x}) d\underset{\sim}{x}$$

$$- E_\xi n_Y^2 .$$

Let $\sigma^2/\beta^2 = k$. Our attempt is to find a best linear p-unbiased estimator for a given design function when k is known. We assume that $p(\underset{\sim}{x})$ satisfies (8.2.7).

Let $G_1(\underset{\sim}{a}) = k \int_{R_n^+} (\Sigma\, a_i^2(\underset{\sim}{x}) x_i^g) p(\underset{\sim}{x}) f(\underset{\sim}{x}) d\underset{\sim}{x} + \int_{R_n^+} (\Sigma\, a_i(\underset{\sim}{x}) x_i)^2 p(\underset{\sim}{x}) f(\underset{\sim}{x}) d\underset{\sim}{x}$.

Thus our problem is to

$$\text{Minimize} \quad G_1(\underset{\sim}{a})$$
$$\text{subject to} \quad H_2(\underset{\sim}{a}) = \theta \qquad \ldots(8.2.25)$$

where $H_2(\underset{\sim}{a})$ is same as in the previous problem and θ is the zero vector of V_2.

The Lagrangian corresponding to (8.2.25) is given by

$$L_1(\underset{\sim}{a}, v^*) = G_1(\underset{\sim}{a}) + v^* H_2(\underset{\sim}{a}) \qquad \ldots(8.2.26)$$

where $v^* \varepsilon V_2^* = V_2$.

It is easy to check that G_1 is infinitely Fréchet differentiable. Proceeding on the lines similar to that used in solving the previous problem we get

$$a_i(\underset{\sim}{x}) = k^{-1} [b(x_i)x_i^{-\varepsilon} - x_i^{1-\varepsilon} e(\underset{\sim}{x}) \Sigma\, b(x_j)x_j^{1-\varepsilon}] \qquad \ldots(8.2.27)$$

where $b(x) \varepsilon V_2$ is the Lagrangian multiplier and $e(\underset{\sim}{x}) = 1/(k + \Sigma\, x_i^{2-\varepsilon})$. To determine $b(x)$ we make use of the constraint (8.2.3). Let

$$r_i(x_i) = q_i(x_i)/f(x_i), c'_{ij}(x_i,x_j) = \int_{R^+_{n-2}} p(\underset{\sim}{x})e(\underset{\sim}{x}) \prod_{\ell \neq i,j}^{n} f(x_\ell)dx_\ell$$

$$c'_i(x_i) = \int_{R^+} c'_{ij}(x_i,x_j)f(x_j)dx_j \;;\; D'_i(x_i,x) = \sum_{j\neq i}^{n} c'_{ij}(x_i,x),$$

$$D\,(x,w) = \Sigma\, D'_i(x,w),\; r(x) = \Sigma\, r_i(x),\; B(x) = \Sigma\, c'_i(x)\,.$$

From (8.2.25) we get

$$\emptyset_i(\underset{\sim}{a},x) = k^{-1}[\, b(x)x^{-\varepsilon}(r_i(x) - x^{2-\varepsilon}\, c'_i(x)) - x^{1-\varepsilon} \int_{R^+} b(w)w^{1-\varepsilon}\, D'_i(x,w)\, f(w)dw].$$

Now using $\Sigma\, \emptyset_i(\underset{\sim}{a}, x) = \emptyset(\underset{\sim}{a}, x) = 1$ we get

$$b(x) - \int_{R^+} K\,(x,w)b(w)f(w)dw = m'(x) \qquad \ldots(8.2.28)$$

where

$$K'(x,w) = \frac{xw^{1-\varepsilon}\, D'(x,w)}{r(x) - x^{2-\varepsilon}\, B(x)} \quad \text{and} \quad m'(x) = \frac{kx^{\varepsilon}}{r(x) - x^{2-\varepsilon}\, B(x)}\,.$$

Now to solve the equation (8.2.28) we make use of Theorem 8.2.1. Thus if $b(x)$ is a solution to (8.2.28) then by substituting it in (8.2.27) we get a vector $\underset{\sim}{\bar{a}}'$ As shown in first problem, we can indeed prove that $\underset{\sim}{\bar{a}}''$ is a solution to (8.2.25) and if there is more than one choice for $\underset{\sim}{\bar{a}}'$ the corresponding value of the functional $L_1(\underset{\sim}{a}, v^*)$ is same for all of them. Thus we have our second existence theorem as follows.

Theorem 8.2.3. For any design function $p(\underset{\sim}{x})$ satisfying (8.2.7) and for which (8.2.28) has a solution there exists a best p-unbiased linear estimator under the model (8.1.1) with $b = 1$, g known and $\sigma^2/\beta^2 = k$ known ; w.r.t. the measure of uncertainty (8.1.2).

Let us consider an example so as to get an idea about the above result.

Example 8.2.2

$$\text{Let} \quad p(\underset{\sim}{x}) = A(k + \sum_{i=1}^{n} x_i^{2-g}) \prod_{i=1}^{n} p(x_i)$$

where A is the normalizing constant. Let $p(\underset{\sim}{x})$ be such that (8.2.7) is satisfied and $x^{1-g}\, p(x)$, $x^g/p(x)$ belong to V_2. It can be checked that

$$m'(x) = \frac{kx^g}{nAp(x)(k\lambda_1 + (n-1)\lambda_2)\lambda_1^{n-2}} = \lambda_3 \frac{x^g}{p(x)} \quad \text{(say)}$$

$$\text{and} \quad K'(x,w) = \frac{(n-1)xw^{1-g}\, p(w)}{(k\lambda_1 + (n-1)\lambda_2)} = \lambda_4 x w^{1-g}\, p(w) \quad \text{(say)}$$

where $\lambda_1 = \int_{R^+} p(x)f(x)dx$ and $\lambda_2 = \int_{R^+} x^{2-g}\, p(x)f(x)dx$.

Thus the equation (8.2.28) reduces to

$$b(x) - \lambda_4 x \int_{R^+} w^{1-g}\, p(w)b(w)f(w)dw = \lambda_3 x^g / p(x). \quad \ldots(8.2.29)$$

Letting $b_1 = \int_{R^+} w^{1-g}\, p(w)b(w)f(w)dw$, multiplying both sides of (8.2.29) by $x^{1-g}\, p(x)$ and integrating we get $b_1(1-\lambda_4\lambda_2) = \lambda_3\mu$

Or $\quad b_1 = \lambda_3\mu\,/(1-\lambda_4\lambda_2)$, (note $\lambda_4\lambda_2 < 1$)

Thus $\quad b(x) = -\lambda_3 x^g / p(x) + \lambda_3\lambda_4\mu x\,/(1-\lambda_4\lambda_2)$

$$= \lambda_3 x^g / p(x) + \lambda_5 x \quad \text{(say)} .$$

Substituting in (8.2.27) we get for $1 \leq i \leq n$,

$$a_i(\underset{\sim}{x}) = \lambda_3 / kp(x_i) + x_i^{1-g}\, e(\underset{\sim}{x})\, [\lambda_5 - \lambda_3 \Sigma x_j / kp(x_j)] . \qquad \ldots(8.2.30)$$

Thus for the above $p(\underset{\sim}{x})$ the best linear p-unbiased estimator is given by $\Sigma\, \bar{a}_i(\underset{\sim}{x}) y(x_i)$

where $\bar{a}_i'(\underset{\sim}{x})$, $i = 1, 2, \ldots, n$; are given by (8.2.30).

<u>Remark 8.2.2</u>. In particular, for $g = 2$ and $p(x_i) = x_i / \mu$, we get $\bar{a}_i'(\underset{\sim}{x})$ independent of k, namely $\bar{a}_i'(\underset{\sim}{x}) = \mu / nx_i$.

REFERENCES

Arnab, R. (1979). Contributions to the theories of repetitive survey sampling. Ph.D. thesis. Indian Statistical Institute, Calcutta.

Bellhouse, D.R., V.P. Godambe and M.E. Thompson (1977). Two-stage sampling with exchangeable prior distributions. Biometrika 64, 97-103.

Cassel, C.M. and C.E. Särndal (1972). A model for studying robustness of estimators and informativeness of labels in sampling with varying probabilities. J.R. Statist. Soc. B, 34, 279-289.

Cassel, C.M. and C.E. Särndal (1974). Evaluation of some sampling strategies for finite populations using a continuous variable framework. Commun. Statist. 3, 373-390.

Cassel, C.M., C.E. Särndal and J.H. Wretman (1976). Some results based on generalized regression estimation for finite populations. Biometrika 63, 615-620.

Cassel, C.M., C.E. Sarndal and J.H. Wretman (1977). Foundations of inference in survey sampling. Wiley-Interscience.

Chaudhuri, A. (1976). On the choice of sampling strategies. Calcutta Statist. Ass. Bull. 25, 119-128.

Chaudhuri, A. (1976a). A nonnegativity criterion for a certain variance estimator. Metrika 23, 201-205.

Chaudhuri, A. (1979). On nonnegative variance estimation in Midzuno strategy. Tech. Rep. ASC/23 C.S. Unit. Indian Statistical Institute, Calcutta.

Chaudhuri, A. and R. Arnab (1981). On nonnegative variance estimation. Metrika 28, 1-12.

Cochran, W.G. (1946). Relative accuracy of systematic and stratified random samples for a certain class of populations. Ann. Math. Statist. 17, 164-177.

Cochran, W.G. (1977). Sampling techniques. 3rd ed. Wiley.

Ericson, W.A. (1965). Optimum stratified sampling using prior information. J. Amer. Statist. Ass. 36, 45-49.

Ericson, W.A. (1969a). Subjective Bayesian models in sampling finite populations. J.R. Statist. Soc. B 31, 195-224.

Ericson, W.A. (1969b). Subjective Bayesian models in sampling finite populations : Stratification. In N.L. Johnson and H. Smith, Eds., New developments in survey sampling. Wiley-Interscience, 326-357.

Finetti, B. de (1937). La prevision: Ses lois logiques, ses sources subjectives. Annales de l' Institut Henri Poincaré 7, 1-68.

Godambe, V.P. (1955). A unified theory of sampling from finite populations. J.R. Statist. Soc. B, 17, 269-278.

Godambe, V.P. and V.M. Joshi (1965). Admissibility and Bayes estimation in sampling finite populations I. Ann. Math. Statist. 36, 1707-1722.

Godambe, V.P. and M.E. Thompson (1973). Estimation in sampling theory with exchangeable prior distributions. Ann. Statist. 1, 1212-1221.

Hanif, M. and Brewer, K.R.W. (1980). Sampling with unequal probability without replacement : A review. Int. Statist. Rev. 48, 317-335.

Hansen, M.H. and W.N. Hurvitz (1943). On the theory of sampling from finite populations. Ann. Math. Statist. 14, 333-362.

Hansen, M.H. and W.N. Hurvitz (1949). On the determination of optimum probabilities in sampling. Ann. Math. Statist. 20, 426-432.

Hanurav, T.V. (1962). On Horvitz and Thompson estimator. Sankhyā A, 24, 429-436.

Hanurav, T.V. (1965). Optimum sampling strategies and some related problems. Ph.D. thesis, Indian Statistical Institute, Calcutta.

Hege, V. (1965). Sampling designs which admit uniformly minimum variance unbiased estimate. Calcutta Statist. Ass. Bull. 14, 160-162.

Hochstadt, H. (1973). Integral equations. Wiley-Interscience, 108-109.

Joshi, V.M. (1979). The best strategy for estimating the mean of a finite population. Ann. Statist. 7, 531-536.

Kempthorne, O. (1969). Some remarks on statistical inference in finite sampling. In N.L. Johnson and H. Smith, Eds., New developments in survey sampling. Wiley-Interscience, 671-695.

Lanke, J. (1975). Some contributions to the theory of survey sampling. Lund : AV-Centralen.

Luenberger, D.G. (1968). Optimization by vector space methods. Wiley.

Madow, W.G. and L.H. Madow (1944). On the theory of symmetric sampling. Ann. Math. Statist. 15, 1-24.

Mangasarian, O.L. (1969). Nonlinear programming. McGraw Hill, Inc.

Padmawar, V.R. (1979). Optimal strategies under superpopulation models. Tech. Rep. 28/79 Stat-Math. Div. Indian Statistical Institute, Calcutta.

Padmawar, V.R. (1980). On nonnegative variance estimation. Tech. Rep. 17/80 Stat-Math. Div. Indian Statistical Institute, Calcutta.

Padmawar, V.R. (1981). A note on the comparison of certain sampling strategies. J.R. Statist. Soc. B, 43, 321-326.

Padmawar, V.R. (1981a). A study under a continuous survey sampling model. Tech. Rep. 18/81, Stat-Math. Div. Indian Statistical Institute, Calcutta.

Padmawar, V.R. (1981b). Two existence theorems. Tech. Rep. 19/81, Stat-Math. Div. Indian Statistical Institute, Calcutta.

Padmawar, V.R. and P. Mukhopadhyay (1981). Estimation under a two stage random permutation model. Tech. Rep. Stat-Math. Div. Indian Statistical Institute, Calcutta.

Padmawar, V.R. and P. Mukhopadhyay (1983). Estimation of symmetric functions of a finite population. Metrika. (to appear).

Padmawar, V.R. and B.K. Sinha (1981). On estimation of population proportion. Tech. Rep. 20/81, Stat-Math. Div. Indian Statistical Institute, Calcutta.

Patel, H.C. and S.W. Dharmadhikari (1977). On linear invariant unbiased estimators in survey sampling, Sankhyā, C, 39, 21-27.

Patel, H.C. and S.W. Dharmadhikari (1978). Admissibility of Murthy's and Midzuno's estimators within the class of linear unbiased estimators of finite population totals. Sankhyā, C, 40, 21-28.

Pedgaonkar, A.M. and S.G. Prabhu-Ajgaonkar (1978). Comparison of sampling strategies. Metrika 25, 149-153.

Prabhu-Ajgaonkar, S.G. (1965). On a class of linear estimators in sampling with varying probabilities without replacement. J. Amer. Statist. Ass. 60, 637-642.

Ramachandran, G. (1978). Choice of strategies in survey sampling. Ph.D. thesis, Indian Statistical Institute, Calcutta.

Rao, J.N.K. (1966). On the relative efficiency of some estimators in pps sampling for multiple characteristics. Sankhyā, A, 23, 61-70.

Rao, J.N.K. (1975). On the foundations of survey sampling. In J.N. Shrivastava, Ed., A survey of statistical design and linear models. The Hague North Holland, 481-505.

Rao, J.N.K. and D.R. Bellhouse (1978). Optimal estimation of a finite population mean under generalized random permutation models. J. Statist. Plan. and Inf. 2, 125-141.

Rao, J.N.K., H.O. Hartley and W.G. Cochran (1962). On a simple procedure of unequal probability sampling without replacement. J.R. Statist. Soc. B, 24, 482-491.

Rao, J.N.K. and K. Vijayan (1977). On estimating the variance in sampling with probability proportional to aggregate size. J. Amer. Statist. Ass. 72, 579-584.

Rao, T.J. (1967). On the choice of a strategy for the ratio method of estimation. J.R. Statist. Soc. B, 29, 392-397.

Rao, T.J. (1971). πps sampling designs and Horvitz-Thompson estimator. J. Amer. Statist. Ass. 66, 872-875.

Rao, T.J. (1972). Horvitz-Thompson and Des Raj estimators revisited. Aust. J. Statist. 14, 227-230.

Rao, T.J. (1972a). On the variance of the ratio estimator. Metrika 18, 209-215.

Rao, T.J. (1977). Estimating the variance of the ratio estimator for the Midzuno-Sen Sampling scheme. Metrika 24, 203-208.

Royall, R.M. (1968). An old approach to finite population sampling theory. J. Amer. Statist. Ass. 63, 1269-1279.

Royall, R.M. (1970). Finite population sampling - on labels in estimation. Ann. Math. Statist. 41, 1774-1779.

Royall, R.M. and J. Herson (1973). Robust estimation in finite populations I. J. Amer. Statist. Ass. 68, 880-889.

Särndal, C.E. (1978). Design based and model based inference in survey sampling. Scand. J. Statist. 5, 27-52.

Särndal, C.E. (1980). A method for assessing efficiency and bias of estimation strategies in survey sampling. S. Afri. J. Statist. 24, 17-30.

Särndal, C.E. (1980a). Two model based arguments in survey sampling. Aust. J. Statist. 22, 341-348.

Scott, A. and T.M.F. Smith (1969). Estimation in multistage surveys. J. Amer. Statist. Ass. 64, 830-840.

Scott, A. and T.M.F. Smith (1975). Minimax designs for sample surveys. Biometrika 62, 353-357.

Sharma, S.C. (1970). On an estimation in T_3-class of linear estimators in sampling with varying probabilities from a finite population. Ann. Inst. Statist. Math. 22, 495-500.

Smith, T.M.F. (1976). Discussion of paper by C.M. Cassel, C.E. Särndal and J.H. Wretman. Biometrika 63, 620.

Stenger, H. (1979). A minimax approach to randomization and estimation in survey sampling. Ann. Math. Statist. 7, 395-399.

Stenger, H. and S. Gabler (1981). On the completeness of the class of fixed size sampling strategies. Ann. Math. Statist. 9, 229-232.

Sugden, R.A. (1979). Inference on symmetric functions of exchangeable populations. J.R. Statist. Soc. B, 41, 269-273.

Vijayan, K. (1975). On estimating the variance in unequal probability sampling. J. Amer. Statist. Ass. 70, 713-716.

*****