CHAPTER I :

INTRODUCTION

1.1 LITERATURE RELATED TO THE PROBLEMS CONSIDERED IN THIS THESIS

In the traditional theory of estimation it has been long customary to put forward some desirable properties like unbiasedness, minimum variance etc., that a good estimator should possess. In the sampling theory for finite populations - often called sample surveys, eventhough the problem involved was that of estimation, the practice was to consider adhoc estimators for a given sampling method and to compute their variances. No attempt was made to select estimators on the basis of well defined criteria. The first attempts in this direction can be found in the work of Horvitz-Thompson [7] and the first formulation of the problem is due to Godambe [2] who generalised the concepts of sampling design and linear estimators and proved the important result that for no sampling design does there exist a uniformly minimum variance unbiased linear estimator of the population total (equivalently, the mean). This result has some exceptions. Hanurav [5] and

Hege [6] proved independently that unicluster designs admit uniformly minimum variance unbiased estimators. Barring this exception, this result pointed to the inadequacy of the then existing method of applying the famous Markov's theorem of least squares to derive best linear unbiased estimators of the population total, and made clear the main difference between the classical theory of estimation for theoretical populations and the theory for finite populations. Because of the negative result of Godambe [2] , it was natural to apply the weaker criterion of admissibility. Godambe [3] and Roy-Chakravarti [17] independently proved that the Horvitz Thompson (HT) estimator is admissible within the homogeneous linear unbiased class for any sampling design. Roy-Chakravarti [17] also constructed a complete class $\mathcal{C}$ of linear estimators, which do not take into account either the order of occurence or the number of repetitions of any unit in the sample. However, Godambe and Joshi [8] and Dharmadhikari [1] have constructed examples of inadmissible estimator in $\mathcal{C}$. Thus, in general, $\mathcal{C}$ is not minimal complete.

The above provides the basic framework for the problems of this thesis. We also have to use the concept of linear invariant estimators introduced by Roy-Chakravarti [17] and the concept of necessary best estimators (NBE) introduced by Prabhu-Ajgaonkar [12] .

Finally, we refer to two papers of Murthy [10] , [11] and a paper of Midzuno [9] . These propose estimators for the population total for special varying probability sampling schemes.

1.2 DEFINITIONS

Consider a finite population $\mathcal{U}$ of N identifiable units

$$\mathcal{U} = \{1, 2, \ldots, N\} . \qquad \ldots(1.2.1)$$

A list of units as (1.2.1) is called a sampling frame and N is called the size of the population.

By a sampling design for $\mathcal{U}$ we mean a pair (S, P), where

(a) S is a class of subsets of $\mathcal{U}$ whose union is $\mathcal{U}$,

and (b) $P = \{P(s),\ s \in S\}$ $\qquad \ldots(1.2.2)$

is such that $p(s) > 0$ for all s and $\sum_{s \in S} p(s) = 1$.

We may write

$$D = (S, P). \qquad \ldots(1.2.3)$$

Note that samples of zero probability have been eliminated. Also repetitions and orders of occurrence of the units have been ignored.

In practice, we draw a sample by drawing units one after other and not by listing all possible samples and then selecting a sample s with specific probability p(s). Hanurav [5] defined a sampling mechanism of drawing units from the population $\mathcal{U}$ in (1.2.1) one after another with varying probabilities. He also proved that there exists a one to one correspondence between sampling designs and sampling mechanisms. This enables us to search for optimum estimators in the unified frame work of sampling design rather than in the seemingly diverse types of sampling methods.

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

$$\underset{\sim}{Y} = (Y_1, Y_2, \ldots, Y_N). \qquad \ldots(1.2.4)$$

The Y_i's are unknown a priori and our parameter of interest is $\underset{\sim}{Y}$ which is assumed to be a point in R_N, the N-dimensional

Euclidian space. We consider the problem of estimating the particular parametric function, called the population total, defined by

$$Y = \sum_{i=1}^{N} Y_i \qquad \ldots(1.2.5)$$

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

<u>Definition</u> 1.2.1 : An <u>estimator</u> $t(s,\underset{\sim}{Y})$ is a real valued function defined on $S \times R_N$ which depends on $\underset{\sim}{Y}$ only through those Y_i for which $i \in s$.

<u>Definition</u> 1.2.2 : An estimator $t(s,\underset{\sim}{Y})$ is called <u>linear</u> if it has the form

$$t(s,\underset{\sim}{Y}) = \sum_{i \in s} b(s,i)Y_i. \qquad \ldots(1.2.6)$$

Remark : Some authors also introduce non-homogeneous linear estimators by adding a constant $\alpha(s)$ to the right side of (1.2.6). We will have no occasion to use such estimators in this thesis. For us, a linear estimator will always have the form (1.2.6).

<u>Definition</u> 1.2.3 : For a design (S,P) an estimator $t(s,\underset{\sim}{Y})$ is said to be <u>unbiased</u> for a parametric function $\mathcal{T}(\underset{\sim}{Y})$ if

$$\sum_{s \in S} t(s,\underset{\sim}{Y})\ p(s) = \Upsilon(\underset{\sim}{Y}) \qquad \ldots(1.2.7)$$

for all $\underset{\sim}{Y}$ in R_N.

The variance of an unbiased estimator $t(s,\underset{\sim}{Y})$ of $\Upsilon(\underset{\sim}{Y})$ is

$$\text{Var}\ [t(s,\underset{\sim}{Y})] = \sum_{s \in S} t^2(s,\underset{\sim}{Y})\ p(s) - \Upsilon^2(\underset{\sim}{Y}) \qquad \ldots(1.2.8)$$

<u>Definition</u> 1.2.4 : An estimator $t(s,\underset{\sim}{Y})$ belonging to a class $\mathcal{C}$ of unbiased estimators of $\Upsilon(\underset{\sim}{Y})$ is said to be <u>admissible</u> in $\mathcal{C}$ with respect to a given design (S,P) if there does not exist any other estimator in $\mathcal{C}$ which is uniformly better than $t(s,\underset{\sim}{Y})$, i.e. if given $t'(s,\underset{\sim}{Y}) \neq t(s,\underset{\sim}{Y})$ in $\mathcal{C}$, there exist atleast one $\underset{\sim}{Y}_o$ in R_N such that

$$\text{Var}\ [t(s,\underset{\sim}{Y}_o)] < \text{Var}\ [t'(s,\underset{\sim}{Y}_o)] \qquad \ldots(1.2.9)$$

where both the variances in (1.2.9) are evaluated at $\underset{\sim}{Y} = \underset{\sim}{Y}_o$.

Prabhu-Ajgaonkar [12] introduced the criteria of necessary best estimators of various orders.

If an unbiased estimator $t(s,\underset{\sim}{Y})$ of Y is of the form given in (1.2.6) then

Var $[t(s,\underset{\sim}{Y})]$ in (1.2.8) is given by

$$\text{Var}\,[t(s,\underset{\sim}{Y})] = \sum_{i=1}^{N} \sum_{s \ni i} b^2(s,i)\, Y_i^2\, p(s) + \sum_{i \neq j}^{N} \sum_{s \supset \{i,j\}} b(s,i)b(s,j)Y_iY_j p(s) - Y^2. \qquad \ldots(1.2.10)$$

<u>Definition</u> 1.2.5 : An unbiased estimator $t(s,\underset{\sim}{Y})$ of Y of the form given in (1.2.6) is called a <u>necessary best estimator</u> (NBE) of order r if, for every unbiased estimator $t'(s,\underset{\sim}{Y})$ of Y of the form (1.2.6), the quadratic form

$Q = \text{Var}\,[t'(s,\underset{\sim}{Y})] - \text{Var}\,[t(s,\underset{\sim}{Y})]$ has all leading principal minors of order 1,2, ..., r non-negative.

It is clear from the definition that a NBE of order N, if it exists, is a UMV unbiased estimator.

Roy-Chakravarti [17] introduced the criteria of linear invariant estimators and regular estimators.

<u>Definition</u> 1.2.6 : An estimator $t(s,\underset{\sim}{Y})$ in (1.2.6) is called a <u>linear invariant</u> estimator for the population total Y if

$$\sum_{i \in s} b(s,i) = N \text{ for all } s \in S. \qquad \ldots(1.2.11)$$

Definition 1.2.7 : An unbiased estimator $t(s,\underset{\sim}{Y})$ in (1.2.6) is called a regular estimator if

$$\mathrm{Var}\left[t(s,\underset{\sim}{Y})\right] = k\,\sigma_Y^2 \qquad \ldots(1.2.12)$$

where k is a constant and $\sigma_Y^2 = \frac{1}{N} \sum_{i=1}^{N} (Y_i-\overline{Y})^2$.

Definition 1.2.7 : A design (S,P) is said to be unicluster if

$$s_1 \in S, \quad s_2 \in S \implies s_1 = s_2 \quad \text{or} \quad s_1 \cap s_2 = \emptyset .$$

In other words a unicluster design divides the population into clusters and then selects just one of these clusters as the sample.

1.3 SUMMARY OF THE THESIS

In Chapter II we discuss the concept of NBE estimator defined by definition 1.2.5. We prove that NBE of order 1 exists and coincides with Horvitz-Thompson estimator. It is also established that NBE of order 2 and higher do not exist for any non-unicluster design.

In Chapter III, necessary and sufficient conditions on the design are given for the existence of a linear invariant unbiased estimator of the population total. It is also proved that the same condition implies the existence of an admissible linear invariant unbiased estimator. We also prove that a connected design satisfies these conditions.

Chapter IV deals with admissibility of Murthy's [10],[11] estimators for the population total. We actually construct a large class of admissible linear estimators for the scheme of sampling two units without replacement and with varying probabilities. Also a linear invariant admissible estimator for this design is obtained. We also prove that the usual unbiased ratio estimator for the Midzuno scheme is admissible.

In Chapter V we construct the minimal complete class of homogeneous linear estimators of the population mean for the simple case of sampling two units from a population of size 3. We also give a sufficient condition on the design under which the class $\mathcal{C}$ of Roy-Chakravarti [17] is not minimal complete.

The thesis is based on the four published papers of the author with Dharmadhikari [13],[14], [15], and [16].