SOME CONTRIBUTIONS TO THE SAMPLING THEORY USING APRIORI AND APOSTERIORI INFORMATION

Pulakesh Maiti

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

ACKNOWLEDGEMENTS

As everyone possibly experiences, writing this portion of a thesis seems to be the most difficult. The power of expressing one's gratitude in black and white seems to get choked by emotions! For me, submitting a Ph.D. dissertation was beyond any stretch of my imagination up until about 5 years ago when I joined the Indian Statistical Institute. As that "undreamt dream" has finally shaped into a reality, I am deeply moved in recalling the happy and strange emotional experiences I have had after joining this Institute, which I guess were the motive forces behind this thesis. Friends — some more close, some lesser — at this Institute rendered invaluable moral support, cordiality, warmth and encouragement. "Being grateful" is too weak a phrase to express my gratitude to them. I am very grateful to Professors J. Roy, N. Bhattacharya and B.P. Adhikari who often went out of their way to help me channelise my efforts, — especially to Professor J.K. Ghosh who has never been run out of patience while listening to me.

It was indeed a source of great pleasure to work for this thesis under the guidance and supervision of Dr. T.P. Tripathi. His overall friendliness and the smile of welcome that I received even when I knocked at the door of his residence at wee hours of the morning with a problem — whether academic or non-academic — was always very reassuring. My grasp of the English language is stunningly insufficient to express in words my indebtedness to him. I also thank him and Dr. S.D. Sharma for permitting me to include our joint work in the thesis.

Thanks are also due to Dr. Arijit Chaudhury, who took great pain in going through the manuscript and offering many fruitful suggestions.

Shri K.B. Goswami was a particularly important driving force for my being able to complete this dissertation. He was a source of constant encouragement and his accommodative attitude helped me complete this work faster than it could possibly have been with the burden of my normal office work. He also provided the necessary facilities for doing this work. I greatly appreciate his help.

And finally, I am grateful to :

— my colleagues in the Sankhyā office for all necessary cooperation and help – Sarvashree Arun Das and Harish Chandra for doing a wonderful job of bringing the manuscripts to the present form.

Pulakesh Maiti

CONTENTS

INDEX TO PRINCIPAL NOTATION

The following notation will be used throughout this work.

N	:	Number of units in the population
n	:	Number of units in the sample
K	:	$(N-n)/[n(N-1)]$
SRSWR	:	Simple random sampling with replacement
SRSWOR	:	Simple random sampling without replacement
PPSWR	:	Probability proportional to size with replacement
MSE	:	Mean square error
iff	:	if and only if

For variates y and x, in the population :

θ : $(y_1, y_2, \ldots, y_N)$

$Y = \sum_{i=1}^{N} y_i$: Population total for the character y

$\bar{Y}$: Population mean for the character y

σ_y^2 : Population variance $= \sum_{i=1}^{N} (y_i - \bar{Y})^2/N = (N-1)S_y^2/N$

$C_y = (\sigma_y/\bar{Y})$: Coefficient of variation for the variate y

$\mu_r(y)$: r^{th} central moment of the character y

β_1 : μ_3^2/μ_2^3

β_2 : μ_4/μ_2^2

μ_2 : σ_y^2, Population variance

X : Population total for the character x

$\bar{X}_i$: Population mean for the i^{th} auxiliary character x_i

$\tilde{X}_i$: Population Harmonic mean for the i^{th} auxiliary character x_i

ρ : Coefficient of correlation between the variates

For variates y and x in the sample :

$\bar{y} = \sum_{i=1}^{n} y_i/n$: Sample mean of the variate y

$s_y^2 = \sum_{i=1}^{n} (y_i-\bar{y})^2/(n-1)$: Sample variance of the variate y

$c_y = s_y/\bar{y}$: Coefficient of variation

$\bar{x} = \sum_{i=1}^{n} x_i/n$: Sample mean for the character x

CHAPTER I

INTRODUCTION

1.1 General

The practice of keeping records on socio-economic aspects etc. in terms of the data from real or actual populations by the States etc. is age old. There are evidences of such practice in India even at about 500 B.C. where the King Courts/Tribe Heads/ Gramini (village Heads) used to keep the records on military force, human population, cattle population, land use etc. The records were in the form of data on both the quantitative and qualitative characteristics. It is well known that the Great Moghul Empire, in India, also used to keep the records in form of data on the characteristics related to socio-economic, military affairs and land ownership etc. No doubt such a practice was in force not only in India, but in many other countries, especially the European Countries. As early as in 1662, Graunt published his work on social-statistics, based on the data collected in an arbitrary or haphazard manner. However such practice was not well-organised and it is known very little about the methods used for conducting such enquiries and they were of limited value. With the development of agriculture, trade, commerce and industry, especially after the industrial renaissance in Europe, the necessity for such enquiries in depth and breadth also increased and the collection of data in the form of complete enumeration (census) on various social, economic, demographic and biological characteristics came into practice in the 19th century through organised bodies in a number of countries. Such a practice came into existence in India too, where the British Government started the operation of census around 1881 and which is continuing with a periodicity of every ten years, so far almost uninterrupted. Conducting of census has provided the countries - where such practice has been in force - a treasure of varied type of data and helped the Governments in formulating the strategy of development etc.

The expanding demand of time further made it necessary to consider the collection of data for a part of the natural population in case of those characteristics and in those spheres (areas) where complete enumeration was not feasible or the data were required within short span of time for making certain decisions. This necessitated the use of some kind of a sample survey approach (collection of data for a small portion of the whole natural population about which some conclusions were to be drawn), which started in some countries in the last quarter of 19th Century. It is possibly about hundred years ago from today that the practice of conducting 'sample surveys' came into being in many countries of the world especially in Europe. Thus use of a sample (a part) for making certain decisions or drawing conclusions about the real/actual/natural populations (The whole from which sample was taken) is relatively a modern development. However these enquiries can not be rightly termed as scientific enquiries as they were not conducted in a strictly scientific manner, in the sense, as we understand today, and no one discussed how the samples were taken and what was the accuracy of the results etc. These methods of enquiries (based on sample) could earn little trust until the turn of the last Century.

The rapid growth of interest in 'Sampling methods' and the conclusions made depending on the sample data, possibly started after Kiaer (1895) introduced the concept of 'random sampling' to study the socio-economic problems (characteristics), to replace the usual approach of complete enumeration, and emphasized the value of a representative sample which he defined as a "photograph which reproduces the details of the original in its true relative proportions". The discussion by Kiaer (1895) on what is a representative sample etc. is quite interesting.

The discussion by Bowley, A.L. (1906) about the use of 'random sample' for making inferences about some actual/natural/finite populations characteristics, such as mean etc., is quite interesting. Further interesting discussion concerning some surveys conducted in Germany during 19th Century is found in Schott. S. (1923), [Quoted by Godambe, V.P. (1976)].

The use of artificial devices for drawing a random sample seems to come in practice only in the early 20th Century. These were used in Sweden in 1912 to study the housing conditions [Dalenius, T, (1957)].

It may be noted that initially the units in the sample, the information on which was used for making inferences (mostly of estimation type) about some characteristics of the actual populations (finite), were used to be selected purposively, in a haphazard and arbitrary manner. And later they used to be selected using artificial, but crude methods of randomization. The foundation of modern sampling theory was yet to be laid down.

The works of Bowley, A.L. (1926) and Neyman, J. (1934) may be said to have laid the foundations of modern sampling theory. Both the authors dealt with stratified random sampling and while the proportional allocation appears to have been discussed by Bowley, Neyman discussed about now well-known 'Neyman allocation' and put forward theoretical criticism of purposive selection. These works gave an impetus and it is after this period especially during forties of this century, that different sampling procedures (methods of drawing random samples from finite populations) which are under common use ~~today~~, viz. cluster sampling, multistage sampling, systematic sampling, pps sampling, double sampling, sampling on successive occasions etc., and different estimation procedures, including well-known ratio and regression methods of estimation were developed and their use came into practice. At the same time the Survey Statisticians became much more conscious about the quality of data and the accuracy of the results obtained by using different selection and estimation procedures developed. The works on non-sampling errors also started during the forties. India witnessed the advent of large scale sample surveys under the guidance of Ld Professor P.C. Mahalanobis, the founder of the Indian Statistical Institute. These developments deepened broadened and made firm the foundations of Theory and practice of sample surveys. The pillars of foundations during this period, may be attributed to be laid down by the contributions of J. Neyman, P.C. Mahalanobis, P.V. Sukhatme, W.G. Cochran, R.J. Jessen, M.H. Hansen, W.N. Hurwitz, F. Yates, W.G. Madow, W.E. Deming *et al.* The significant works done during the above period, related to sampling procedures and estimation procedures, are contained in most of the books on sampling theory, e.g., in [Cochran (1963, 1977)], Sukhatme and Sukhatme (1970), Raj (1968), Murthy (1967).

During 1950's, new avenues of research were unfolded by the works of Horvitz and Thompson (1952), [Godambe (1955), (1960)], [Koop (1957), (1963)] and Basu (1958) etc., whose contributions made firm the philosophical and theoretical foundations of sampling from finite populations and estimation of some population parameters. This resulted in to the development of so called "Unified Theory of Sampling" [Review by Godambe, (1965)].

The works after 1950 are further deeper and diversified and they have broadened the sphere of sampling and estimation methods.

The use of auxiliary information in devising suitable sampling procedures and estimation procedures was recognized well during the thirties and especially forties itself and led to further significan contributions by T. Dalenius, H.O. Hartley, H.D. Patterson, H.Midzuno, I. Olkin, Des Raj, J.N.K. Rao, M.N. Murthy, A.R. Sen, B.D. Tikkiwal, and D. Singh etc., during 1950's and 1960's [Reviews by Sukhatme (1959,66), Dalenius (1962), Murthy (1963a)].

Almost all the works mentioned above dealt with the problems of obtaining suitable sampling procedures and estimation procedures, under various situations for estimating the finite population mean, total, proportion, or ratio. The role of the use of apriori or auxiliary information in devising suitable sampling strategies has been of immense importance.

1.2 Preliminaries

A collection

$$U = \{U_1, U_2, \ldots, U_N\} \tag{1.1.1}$$

of N (given) well defined, identifiable and observable objects under consideration about which certain valid inferences are to be made is called a finite population, N is called the population size, objects U_i are called sampling units and the list of all sampling units, with identification numbers $1,2,\ldots,N$ is known as the sampling frame.

Let a variate (character) y be a real valued function defined on U and let $y_i = y(U_i)$, $(i = 1,2,\ldots,N)$ be the value of the character of y associated with the i^{th} unit of the population.

Let θ denote the set of all N y-values, i.e.,

$$\theta = \{y_1, y_2, \ldots, y_N\} \tag{1.1.2}$$

and let $\Omega = \{\theta\}$ denote the space, described by θ, which may be the N-dimensional Euclidean space or subspace of it; $\Omega = \{\theta\}$ may be referred to as parametric space of y.

A parameter (or parametric function) $\phi(y)$ of a character y is a real valued function of θ. Generally, the parameters of special interest are

(i) $Y = \sum_{i=1}^{N} y_i$: population total of the character y

(ii) $\bar{Y} = Y/N$: population mean of y

(iii) P : the proportion of population units falling in some specified subset of the population U.

(iv) $R = Y/X$: the population ratio; the ratio of population means (or totals) of two characters y and x

(v) $\sigma_y^2 = \sum_{i=1}^{N} (y_i - \bar{Y})^2/N$: population variance of y

(vi) $C_y = \sigma_y/\bar{Y}$: the population coefficient of variation of y.

Obviously, if θ is known completely and accurately (which is possible only if complete enumeration (census) of all population units is done without any non-sampling errors), the required inferences about the population parameters can be made with certainty.

We shall confine our discussion to <u>point estimation</u> of a specified parameter $\phi(y)$, in which case an estimator for $\phi(y)$ may be defined to be a real valued function of the random sample (considering the identity of the units selected, the y-values for selected units and the order in which the units are appearing etc.). The aim is to reach $\phi(y)$ as close as possible with the help of the information contained in the sample.

A measure of closeness (or dispersion) of an estimator d to a parameter $\phi(y)$ may be defined by $M(d) = E_{p*}(d-\phi(y))^2$, the average value (expectation being taken with respect to the sampling procedure or the sampling design according to which the sample is generated) of squared error loss and which is called mean square error (MSE) of an estimator d in estimating $\phi(y)$. The $M(d)$ may also be denoted by $M(\tau)$ and called the MSE of the sampling strategy $\tau \equiv (P^*,d)$ where P^* is the sampling design or the sampling procedure on which the estimator d is based.

The general criterion of optimality may be described to be minimizing the mean square error (maximizing the efficiency of estimator) for a given (fixed) budget and time or minimizing the cost and time for achieving a given level of efficiency.

However, we shall take MSE as the sole criterion of preference (assuming the cost and time involved to be of the same order) and define a strategy $\tau_1 \equiv (P_1^*, d_1)$ (or an estimator d_1) to be better than another strategy $\tau_2 \equiv (P_2^*, d_2)$ (or estimator d_2)

$$\text{iff} \quad M(d_1) \leq M(d_2) \quad \forall \ \theta \in \Omega$$

the inequality holding true for at least one $\theta = \{y_1, y_2, \ldots, y_N\}$.

For a specified sampling procedure P^*, our effort would be to choose an estimator $d \in \mathcal{D}$ from a class of estimators $\mathcal{D}$ such that

$$M(d) \leq M(d') \quad \forall \ \theta \in \Omega \ \text{and} \ \forall \ d'(\neq d) \in \mathcal{D}$$

with inequality holding true for at least one θ. The estimator d, then, would be called as the best estimator in $\mathcal{D}$ for the specified P^*.

1.3 Classes of Linear Estimators

During the decade 1940-1950, a large number of sampling techniques (sampling procedures or methods of selecting the population units into a sample) corresponding to different situations were evolved and the problems of estimating population total, mean, proportion or ratio were considered. The estimators considered were mostly the linear estimators (linear combination of y-values for the units in the sample). The concept of linear estimators was broadened later by Horvitz and Thompson (1952), Godambe (1955), and Koop (1957),(1963)].These authors made an effort to provide a general treatment of problem of point estimation in case of finite populations and as a result there emerged the so called unified theory of sampling.

Koop (1963) described that for any sampling scheme, the three features inherent in the sample formation, which may be stated as axioms, are

(i) the units appear in order

(ii) a given unit is either present or absent in the sample

(iii) the sample itself is one of the possible samples which could be drawn from the population.

A linear estimator for the population total Y(or mean $\bar{Y}$) is a linear combination of the values of the character y observed for the sample and the coefficients in the linear combination may be determined using the axioms (i), (ii) and (iii) in various combinations, thereby having different types of linear estimators. Applying the axioms (i), (ii) and (iii), the most frequently used three types of linear estimators for Y are obtained as follows :

$T_1 = \sum_{r=1}^{n} \alpha_r (y_{i_r} = y_i)$, where $\alpha_r (r=1,2,\ldots,n)$ is the weight to be associated with the y-value for the unit appearing in the sample at the r^{th} draw.

$T_2 = \sum_{i \epsilon s} \beta_i y_i$, where β_i $(i=1,2,\ldots,N)$ is the coefficient attached to the y-value for the i^{th} population unit, selected in the sample.

and $T_3 = \sum_{i \epsilon s} \gamma_s y_i$, where γ_s is the coefficient attached to each of the y-values in the sample s whenever the sample s is selected.

$\sum_{i \epsilon s}$ denoting the sum over all distinct units in the sample s.

The above three classes were presented by Horvitz and Thompson (1952) who discussed the class T_2 in detail.

Likewise, the linear estimator

$$T_5 = \sum_{i \epsilon s} \beta(s,i)\ y_i$$

where weights $\beta(s,i)$ depend on identity of the units and the sample at hand, discussed by Godambe (1955) is one of the seven classes of estimators which can be generated using axioms (i) to (iii). He (1955) proved that there does not exist a uniformly minimum variance unbiased estimator (UMVUE) in the class of all unbiased estimators in T_5 with respect to *any* sampling design. Exception to the above result was later shown by Hanurav (1966) in case of unicluster designs.

It is shown by Koop (1957, 1963) and Ajgaonkar (1967) that there does not exist UMVU estimator in the unbiased subclass of T_1 for all sampling designs. However, if the sampling procedure is that of simple random sampling without replacement (SRSWOR), the sample mean $\bar{y}$ is UMVUE for $\bar{Y}$ in the unbiased subclass of T_1. The question naturally arises : Is there any estimator in the whole class T_1 which is better (in the sense of having smaller MSE) than $\bar{y}$? In Chapter II, we have answered the question and find that there are a large number of biased estimators in T_1 better than $\bar{y}$. The study has been further extended in Chapter III, to non-homogeneous linear T_1 class and non-linear estimators.

It was shown by Horvitz and Thompson (1952) that the estimator $\hat{Y}_{H-T} = \sum_{i \epsilon s} y_i/\pi_i$ (the so called Horvitz-Thompson estimator) is UMVU estimator for Y in the unbiased subclass of T_2, whatever be the sampling design adopted. A natural question arises. Is there any estimator in the whole class T_2 which is better than $\hat{Y}_{H-T}$.? In Chapter IV, we have undertaken to resolve the question and shown that for a number of sampling designs, a large number of biased estimators from T_2 can be found such that they are better than $\hat{Y}_{H-T}$.

1.4 Use of Supplementary Information and Brief Review of the Related Works

The use of supplementary (auxiliary) information in estimating the population total or mean was recognized during the decade 1940-1950 itself. It is now well recognized that if used intelligibly, the auxiliary information increases the efficiency of the estimators.

Broadly speaking, the auxiliary information may be used in three basic ways [Tripathi, (1970),(1976)]

(i) At pre-selection stage or designing stage : for example in stratifying the population or constructing the clusters on the basis of a related character.

(ii) At selection stage : for example in selecting the units with varying probabilities, based on an auxiliary character, with or without replacement.

(iii) At estimation stage : for example in constructing ratio, regression, difference or product-type estimators.

The information may also be used in mixed ways, at any two stages and then at all the three stages, thereby generating seven ways of using auxiliary information. For example, information on a character z may be used in defining the set of inclusion probabilities and that on another character x may be used in constructing regression-type estimator for the population total Y of the main character y [Tripathi; (1969),(1973)].

In this work, we have confined ourselves to the use of auxiliary information at estimation stage or/and selection stage for presenting the estimators for $\bar{Y}$, Y and σ_y^2.

Further the auxiliary information may be available in the following main forms :

(i) All the values $x_1, x_2, \ldots, x_N$ of auxiliary character (s) $\underset{\sim}{x}$ are known or the frequency distribution of x may be known, where x may be a scalar or a vector $\underset{\sim}{x} = (x_1, x_2, \ldots, x_p)$.

(ii) Some parameter (s) of x are known. For example, population means, totals, variances, coefficients of variation, coefficients of kurtosis or skewness for some or all characters in $\underset{\sim}{x} = (x_1, x_2, \ldots, x_p)$ may be known.

(iii) Values of x are known only for some units of the population.

(iv) Parameters of x, say $\psi(x)$ are not known exactly, but some apriori value $\psi_o(x)$ is known or it is known that $\psi_1(x) \leq \psi(x) \leq \psi_2(x)$ where ψ_1 and ψ_2 are known quantities.

(v) An apriori value, say $\phi_o(y)$, of the parameter under estimation $\phi(y)$ is known or the bounds $\phi_1(y) \leq \phi(y) \leq \phi_2(y)$ are known.

(vi) Some y-values, say $y_1, y_2, \ldots, y_k$ for k out of N units in the population are known apriori (before the sample is selected) or aposteriori, (in addition to y-values for sample after the sample has been selected).

(vii) An apriori value $\psi_{o(i)}(y)$ of a parameter $\psi_i(y)$, $i = 1,2,\ldots,p$, other than the parameter $\phi(y)$ under estimation, is known or it is known that $\psi_i(y)$ lies between quantities $\psi_i^{(1)}(y)$ and $\psi_i^{(2)}(y)$.

The problem of estimating the population mean $\bar{Y}$ or total Y of a character y using the information on a single auxiliary character x in stratifying the population or in selecting the units with varying probabilities with or without replacement or in constructing ratio, regression, difference and product-type estimators has been considered in detail after 1940 and the work is now contained in most of the books on sampling theory. Hansen and Hurwitz (1943) were the first to consider varying probability selection, Cochran (1940,1942) was first to discuss ratio, regression estimators, Hansen, Hurwitz and Madow (1953) were the first to discuss the difference-type estimator and Murthy (1963b) was the first to discuss product estimator using the knowledge on the mean $\bar{X}$ of an auxiliary character x.

The use of information on several auxiliary characters for estimating $\bar{Y}$ (or Y) has been considered among others by Olkin (1958), Raj (1965), Singh (1967), Khan and Tripathi (1967), Tripathi [(1970),(1976(a)],(1978)], Srivastava (1965), Rao and Mudholkar (1967), and Das and Tripathi [(1979(a)), (1979(b)]. The form of information used was the knowledge on the population means of the auxiliary characteristics available readily or through the first phase sample. The use of univariate or multivariate auxiliary information for estimating the population ratio $R = \bar{Y}/\bar{X}$ of two means (or totals) has been considered, among others, by Rao and Pereira (1968), Singh (1969), and Tripathi [(1970), (1980)]. The use of multivariate auxiliary information for estimating $\bar{Y}$ (or Y) on successive occasions has been considered, among others, by Tikkiwal (1967), Sen, A.R. (1971), Tripathi (1970), Tripathi and Srivastava (1979), Dhiresh Adhvaryu (1978). The use of multivariate information for estimating the population ratio R on two occasions has been considered, among others, by Tripathi and Sinha (1976b) Das (1982a). Further the use of multiple information in estimating the population proportion P has been considered by Das (1982b).

To the author's best knowledge, Das and Tripathi (1980) were the first to advocate the use of C_x (coefficient of variation of x) or σ_x^2 (variance of x) in estimating $\bar{Y}$, under certain situations especially when x and y are uncorrelated or very poorly correlated. Also the above two authors [Das and Tripathi, (1978a), (1978b)] were the first who considered the problem of estimating σ_y^2 and C_y in case (a) $\bar{X}$ is known, (b) σ_x is known and (c) C_x is known.

Though a lot of work has been done on the use of multivariate auxiliary information at the estimation stage, not much work has been done on the use of information on several auxiliary characters in stratifying the population or in formation of clusters. Similarly, the problem of using the information on several characters in selecting the units with varying probabilities has not attracted much attention.

To the author's best knowledge, Maiti and Tripathi (1976) were the first to consider the problem of estimating population total Y with the help of a sampling strategy based on varying probability selection with replacement (VPSWR) where the set of probabilities $\{p_1,p_2,\ldots,p_N\}$, $p_j = z_j/\sum_{j=1}^{N} z_j$ (> 0) are based on $z = \phi(x_1,x_2)$, a real valued function of two auxiliary characters x_1 and x_2. Some of the results obtained by them are mentioned in Section 7.3 of Chapter VII.

Recently, Agrawal and Singh (1980) have considered the problem of using multivariate information in selecting the units according to VPSWR and developed a set of probabilities for selecting the units. However the set of probabilities proposed by them depends on the exact knowledge of some population parameters involving the character y itself. Thus the general problem remains yet unresolved and which may be described as follows.

To obtain a suitable measure $z = \phi(x_1,x_2,\ldots,x_p)$ as a real valued function of $(x_1,x_2,\ldots,x_p)$ alone such that the sampling strategy τ for Y, based on the probabilities proportional to z is better than each of the strategies τ_i based on probabilities proportional to x_i $(i = 1,2,\ldots,p)$ and is also better than the strategies based on simple random sampling and sample mean or ratio-type estimators etc. <u>In Chapter VII, we have considered the above problem for $p = 2$ confining ourselves to VPSWR.</u>

The above works are related to p-variate ($p \geq 2$) populations. However, the use of ancillary information has also been considered in case of univariate populations too with an effort to improve the usual estimators for the parameter under estimation $\phi(y)$ especially in the case where knowledge on another parameter $\psi(y)$ is available or some apriori value of $\phi(y)$ is available.

In case σ_y^2 is known, a class of estimators for $\bar{Y}$ defined by

$$C = [\hat{\bar{Y}} : \hat{\bar{Y}} = \frac{\bar{y}-t_1(s_y^2-\sigma_y^2)}{\{s_y^2-t_2(s_y^2-\sigma_y^2)\}^{\alpha}}(\sigma_y^2)^{\alpha}]$$

in case of SRSWOR has been discussed by Das and Tripathi (1980(a)). where t_1 and t_2 are suitably chosen statistics and α is a suitable chosen constant.(1) It is found for large samples and SRSWR that for $\alpha = 0$, the best choice of t_1 is given by

$$Et_1 = \mu_3(y) \; / \; [\sigma_y^4\{\beta_2(y)-\tfrac{n-3}{n-1}\}]$$

and then the resulting MSE is given by

$$M(\hat{\bar{Y}}) = (1/n)\;\sigma_y^2\;[1 - \frac{\beta_1(y)}{\beta_2(y)-\frac{n-3}{n-1}}]$$

where n is the sample size.

It is obvious that in case of symmetric distributions, e.g., normal distribution, none of the estimators from the above class would be better than sample mean $\bar{y}$. However, in other situations, one may generate estimators better than the sample mean $\bar{y}$

(1) However, their method does not yield optimum values because they did not obtain the values of Et_1, Et_2 and α by solving the set of equations:

$\partial M(\hat{\bar{Y}})/\partial Et_1 = 0$, $\partial M(\hat{\bar{Y}})/\partial Et_2 = 0$ and $\partial M(\hat{\bar{Y}})/\partial \alpha = 0$, simultaneously.

In case of normal distribution, the estimator for $\bar{Y}$ considered by Mehta and Srinivasan (1971) defined by

$$T = \bar{y}[1 - a \exp\{-nb\bar{y}^2/\sigma_y^2\}]$$

is found to be better than $\bar{y}$ in many situations. However, such estimators are not of much practical use as σ_y^2 may not be known exactly in most of the practical situations. Moreover, computation of MSE etc., are quite cumbersome.

In case C_y^2 is known, Searls (1964) defined an estimator for $\bar{Y}$ given by

$$\hat{T}_S = n\bar{y}/(n+C_y^2)$$

which is found to be uniformly better than $\bar{y}$

A class of estimators depending on the knowledge of C_y^2 discussed by Das and Tripathi (1980b) is defined by

$$T_{DT} = \frac{\bar{y} - t_1(\hat{\sigma}_y^2 - C_y^2)}{\{c_y^2 - t_2(c_y^2 - C_y^2)\}^{\alpha}} \quad (C_y^2)^{\alpha}$$

where $c_y^2 = s_y^2/\bar{y}^2$

They have discussed a number of estimators from the above class and have shown many of them to be better than $\bar{y}$ under some conditions. Some of the estimators are also shown to be better than Searls' estimator. It is found that in case of symmetrical distributions, the above class of estimators with $t_1 = 0$ and $t_2 = \lambda$, a constant always provides the estimators better than $\bar{y}$ if α and λ are chosen such that

$$\alpha(1-\lambda) = -\ 2C_y^2/[4C_y^2+\beta_2(y) - \frac{n-3}{n-1}]$$

and the MSE of such optimum estimators is given by

$$M_o(T_{DT}) = \frac{\sigma_y^2}{n}[1 - \frac{4\ C_y^2}{4\ C_y^2+\beta_2(y) - \frac{n-3}{n-1}}].$$

However, the results are restricted to large samples to $o(n^{-1})$. Further the proposed estimators are not always better than Searle's estimator.

In case C_y^2 is not known exactly, but a good-guessed value C_o^2 of C_y^2 is known, Hirano (1972) modified the Searle's estimator

$$T_H = n\bar{y}/(n+C_o^2)$$

which will not be better than $\bar{y}$ unless $C_o^2/[2+\frac{C_o^2}{n}] < C_y^2$.

Thompson (1968) considered the estimators

$$\tilde{T}_S = n\bar{y}/(n+c_y^2)$$

$$\tilde{T}_{S(\alpha)} = n\bar{y}/(n+\alpha c_y^2)$$

and Pandey and Singh (1978) considered the estimators

$$T_{PS} = \bar{y} - (1/n)\ s_y^2/\bar{y}$$

$$T_{PS(\alpha)} = \bar{y} + (\alpha/n)\ s_y^2/\bar{y}$$

in case no apriori information about C_y^2 is available and which are found to be preferable over $\bar{y}$ in some situations, especially when $\sqrt{n}/C_y$ is small.

Das and Tripathi (1980a) and Das (1982) discussed about

$$d = w_1 \bar{y} + w_2 s_y^2$$

and found it to be better than $\bar{y}$ under some conditions on the weights w_1 and w_2 depending on prior information about C_y and coefficients of skewness and kurtosis.

Further, Mehta and Srinivasan (1971) discussed

$$T = \bar{y}[1-a \exp\{-nb/C_y^2\}]$$

as an estimator for the mean of a normal population. However, the estimator is not of much interest from practical point of view, because it involves complicated computations and may not be better than $\bar{y}$ in many situations. Moreover, the estimator is confined only to normal distribution.

It is noted from the above discussion that exact or good-guessed values of one or more parameters are necessary to obtain the estimators better than $\bar{y}$. In most of the practical situations, it is not very difficult to have the knowledge on some of the quantities $\bar{Y}_{(1)}, \bar{Y}_{(2)}, \sigma^2_{(1)}, \sigma^2_{(2)}, C_{(1)}, C_{(2)}, \beta_1^{(1)}, \beta_1^{(2)}, \beta_2^{(2)}$ which are such that

$$0 < \bar{Y}_{(1)} \leq \bar{Y} \leq \bar{Y}_{(2)}; \quad 0 < C_{(1)} \leq C_y \leq C_{(2)}$$

$$0 \leq \beta_1^{(1)} \leq \beta_1 \leq \beta_1^{(2)}; \quad \beta_2 \leq \beta_2^{(2)}$$

and such a knowledge can be easily utilized in obtaining the estimators better than $\bar{y}$ as evidenced by our discussion in Chapters II and III and to some extent Chapter IV.

In Chapter II, we have identified a large number of linear biased estimators better than $\bar{y}$ depending only on the knowledge of $C_{(1)}$ and shown that in the populations where C_y is large, the gain by using those estimators would be quite appreciable.

In Chapter III, we have obtained a large number of estimators for $\bar{Y}$, better than sample mean $\bar{y}$, T_H, T_{PS} etc., depending on the knowledge of one or more of the above mentioned apriori values. Further, a large number of estimators better than Searle's estimator are also obtained. The estimators discussed are linear and also non-linear ones.

In Chapter IV also, we have found, depending on the knowledge of $C_{(1)}$, some biased estimators under the structure of general sampling designs from T_2-class of linear estimators, which are better than Horvitz-Thompson-type estimators for $\bar{Y}$, in case of a family of sampling designs.

The problem of estimating population variance σ_y^2 of a character y has been considered, among others, by Wakimoto[(1970), (1971)], Singh, Pandey and Hirano (1973), Liu (1974), Pandey and Singh (1977), and Das and Tripathi[(1977),(1978a)] and the estimators due to them have been discussed in the Section 5.1.

In case of bi-variate populations, Das and Tripathi (1978a) have considered some estimators based on SRSWR from the class $\tilde{d}$ discussed by Das (1982) in case of SRSWOR, where

$$\tilde{d} = \frac{\{s_y^2 - t_1(\hat{\psi}-\psi)\}}{\{\psi - t_2(\hat{\psi}-\psi)\}^{\alpha}} \cdot \psi^{\alpha}$$

$\hat{\psi}$ being an estimator of a known parameter ψ. The estimators discussed by them depend on the knowledge of either of $\bar{X}$, σ_x^2 and C_x.

Further, Srivastava and Jhajj (1980), following Das and Tripathi (1978a) considered the estimation of σ_y^2 in case both $\bar{X}$ and σ_x^2 are known.

It is found, in case of symmetrical distributions, that none of the estimators from $\tilde{d}$ with $\psi = \bar{Y}$, $\hat{\psi} = \bar{y}$ would be better than the usual unbiased estimator s_y^2, while for $\psi = C_y^2$, $\hat{\psi} = s_y^2/\bar{y}^2$ in $\tilde{d}$, one may always find a large number of estimators better than s_y^2 provided coefficient of kurtosis is known at least approximately.

Using the exact or approximate knowledge of the coefficient of kurtosis, Pandey, Singh and Hirano (1973) discussed the estimators

$$T_{o1} = (n/\Delta)\, s_y^2 \;; \quad \tilde{T}_{o1} = (n/\Delta^*)\, s_y^2$$

where $\Delta = \beta_2 + \{(n^2-2n+3)/(n-1\}$ and Δ^* is a good-guessed value of Δ.

Further using the apriori knowledge σ_o^2 for σ_y^2, Pandey and Singh(1977) discussed the estimator

$$\sigma_{PS}^2 = w\, s_y^2 + (1-w)\sigma_o^2, \quad 0 \leq w \leq 1.$$

While T_{o1} is always better than s_y^2, the $\tilde{T}_{o1}$ and σ_{PS}^2 would be so only under some conditions depending upon the knowledge of β_2 and that of $\beta_{2.}$ and σ_o/σ_y respectively. It may be shown that in case $n/\Delta^* < 2(n/\Delta) - 1$ or > 1, $\tilde{T}_{o1}$ would be worse than s_y^2 and same is true about σ_{PS}^2 in case

$$\sigma_o^2/\sigma_y^2 < 0.5 \quad \text{or} \quad > 1.5.$$

Further, Das and Tripathi (1980a) and Das (1982) have considered the estimators

$$T' = w_1\, s_y^2 + w_2\, \bar{y}, \quad w_1 + w_2 \neq 1$$

for σ_y^2 and found that, in case of symmetric or negatively skewed distributions, it would be better than $w_1 s_y^2$ (hence than s_y^2 also with proper choice of w_1) if

$$0 < w_2 < 2n\, \bar{Y}_{(1)} C^2_{(1)y}(1-w_1)/[n+C^2_{(1)y}].$$

The estimators for σ_y^2 due to the other authors mentioned above are discussed in Section 5.1 and are not discussed here to avoid repition.

As mentioned some where else in this Section itself, knowledge of some of the quantities $\bar{Y}_{(1)}$, $\bar{Y}_{(2)}$, $C_{(1)}$, $C_{(2)}$, $\sigma_{(1)}$, $\sigma_{(2)}$, $\beta_2^{(1)}$, $\beta_2^{(2)}$ etc., may be easily utilized to improve upon the usual estimators for a parameter $\phi(y)$ under estimation.

In Chapter V, we have discussed a class of estimators, for σ_y^2, to which many interesting estimators in addition to T_{o1}, $\tilde{T}_{o1}$,

σ^2_{PS} and T' and some from $\tilde{d}$ defined above belong. A large number of estimators better than s^2_y and also the Searls'-type estimator T_{01} can be generated out of the above class by merely using the knowledge of one or more of the quantities $\bar{Y}_{(1)}$, $\bar{Y}_{(2)}$, $C_{(1)}$, $C_{(2)}$, $\sigma^2_{(1)}$, $\sigma^2_{(2)}$, $\beta_1^{(1)}$, $\beta_1^{(2)}$, $\beta_2^{(2)}$

In many situations, it is not strange to get exact or approximate information about some population units which have not appeared in the sample, during or after the collection of data for sampled units. While canvassing the schedule or through some other sources, the information may be available (or may be made available through some extra effort) for some population units, in addition to those which have already appeared in the sample. Such an information may be called as a <u>posteriori information.</u> Similarly, in many practical situations, one may have <u>apriori information</u> on some population units before the sample is selected or one may decide to enumerate all those units for which y-values are supposed to be quite high and/or low and to select a sample from the rest of the population units. Such an information may be called as <u>apriori information</u>.

It won't be advisable to throw off either of the above mentioned types of information -whether a posteriori or apriori, for some population units in addition to those in the sample. This motivates us into the detailed investigation of how to use such an additional information together with that contained in the sample, for providing improved estimator (s) for population total Y (or mean $\bar{Y}$).

In Chapter VI, we have undertaken the problem of estimating the population total Y or mean $\bar{Y}$ in case y-values for $\bar{n}$-population units, in addition to those in the sample, are known (a) aposteriori or (b) apriori.

1.5 Contents of the Chapters in Brief

In Chapter II, we have revisited the T_1-class of linear estimators, $T_1 = \sum_{r=1}^{n} a_r y_r$, in search of the estimators better than sample mean $\bar{y}$ in case of simple random sampling without replacement. We find that the Searle's estimator $\hat{T}_S = \bar{y}/[1+K C_y^2]$ with $K = (N-n)/n(N-1)$ is uniformly minimum mean square error (UMMSE) estimator in the class T_1. We find necessary and sufficient conditions to be satisfied by the weights $a_1, a_2, \ldots, a_n$ so that the resulting biased estimators are better than the sample mean $\bar{y}$ However, the estimator $\hat{T}_S$ and the necessary and sufficient conditions depend on the knowledge of C_y^2 As an alternative to this, we have found all those biased estimators in T_1 to be better than $\bar{y}$ for which

$$(N-1)\ (\ell-1)^2/Q \leq C_{(1)}^2 \qquad (1.5.1)$$

where, $\ell = \sum_{r=1}^{n} a_r$, $\ell_o = \sum_{r=1}^{n} a_r^2$ and $Q = \ell^2 + (\frac{N}{n} - 1) - N\ \ell_o > 0$ is satisfied. Further, the modified Searle's estimator

$$\hat{T}_S' = \lambda\bar{y},\ [1-KC_{(1)}^2]/[1+KC_{(1)}^2] \leq \lambda < 1 \qquad (1.5.2)$$

is always found to be better than $\bar{y}$.

A procedure for choosing the weights a_r's satisfying (1.5.1) has also been indicated. We have also found sufficient conditions for some estimators from T_1, with different weights satisfying (1.5.1) to be better than $\hat{T}'_S$. Some numerical illustrations have been provided to justify our points.

We have studied the relative performance of $\hat{T}_S$ over $\bar{y}$ for different values of K and C_y and find that the relative efficiency of $\hat{T}_S$ over $\bar{y}$ increases with C_y for a given K and with K for a given C_y. We find that the estimators from $\hat{T}'_S$ or from T_1 satisfying (1.5.1) should be used only in case $K \geq 0.01$ and $C_y \geq 1$.

<u>In Chapter III</u>, we extend the study, made in Chapter II, for searching estimators better than sample mean, further to a more general class

$$d = \lambda_1 \bar{y} + \lambda_2 t \qquad (1.5.3)$$

based on SRSWR, where t is a suitably chosen statistic, which in particular, may be a constant, say $\bar{Y}_o$-an apriori value of the population mean $\bar{Y}$ and λ's are suitably chosen weights which need not add to unity.

We find that the estimators from the non-homogeneous T_1-class, $d^* = T_1^* + \lambda_2 \bar{Y}_o$ would always be better than the sample mean $\bar{y}$ as well as the modified Searle's estimator $T_1^* = n\bar{y}/(n+C_*^2)$ with $C_*^2 \leq C_y^2$, provided λ_2 lies between

$$0 \text{ and } 2[C_*^2/(n+C_*^2)]\ \bar{Y}_{(1)}/\bar{Y}_o \qquad (1.5.4)$$

The general properties of the proposed class of estimators in (1.5.3) are studied and the necessary and sufficient condition for d to be better than $\lambda_1 \bar{y}$, for a given λ_1, is found to be that λ_2 lies between

$$0 \text{ and } 2[(1-\lambda_1)\ \bar{Y}\ E(t) - \lambda_1\ Cov(\bar{y},t)]/\ Et^2 \qquad (1.5.5)$$

The general technique of shrinking 'preference-interval' has been discussed and the necessary and sufficient conditions and the sufficient conditions under which the non-linear estimators

$$d_3^* = T_1^* + \lambda_2 s_y^2,\ d_4^* = T_1^* + \lambda_2 (s_y^2/\bar{y}^2),\ d_5^* = T_1^* + \lambda_2 s_y^2 \bar{y}$$

$$d_6^* = T_1^* + \lambda_2 \bar{y}^2 \text{ and } d_7^* = T_1^* + \lambda_2 (s_y^2/\bar{y})$$

are better than T_1^* as well as $\bar{y}$ and also the Searle's estimator $\hat{T}_S$, in case C_y^2 is known, have been derived. These sufficient conditions depend merely on the knowledge of quantities $\bar{Y}_{(1)}$, $\bar{Y}_{(2)}$, $C_{(1)}$, $C_{(2)}$, $\beta_1^{(1)}$ and $\beta_2^{(2)}$

Further, a subclass

$$d' = \bar{y} + \lambda_2 t \qquad (1.5.6)$$

of d has got special attention in our discussion and a set of necessary and sufficient conditions, under which the estimators from d' are better than $\bar{y}$ are obtained, for some choices of the statistic t in d'.

The relative performance of the estimators proposed by us and those given by Searle's (1964), Hirano (1972), Pandey and Singh (1979) has been discussed. Further an empirical study has been made considering the natural populations and percent relative efficiency of

$d' = \bar{y} + \lambda_2 t$ over $\bar{y}$ and that of $d^* = \hat{T}_S + \lambda_2 t$ over $\hat{T}_S$ has been obtained for $t = s_y^2, \bar{y}^2, s_y^2 \bar{y}$ etc.

We find that for population II, in which case, the coefficient of variation C_y is large, the percent relative efficiency of some estimators in d^* over $\hat{T}_S$ is as high as 290%.

In Chapter IV, the T_2-class of linear estimators, $T_2 = \sum_{i \varepsilon s} \beta_i y_i$ for population total Y is revisited in search of the estimators (biased) better than the well-known Horvitz-Thompson estimator $\hat{Y}_{H-T} = \sum_{i \varepsilon s} y_i/\pi_i$, whatever be the sampling design and find that there does not exist UMMSE estimator for Y in T_2.

We find that for some populations and for a family of sampling designs, the estimator $T_2' = \lambda \hat{Y}_{H-T}$ is better than $\hat{Y}_{H-T}$. We give a numerical illustration showing that for some choices of λ, the percent relative efficiency of T_2' over $\hat{Y}_{H-T}$, under a specified design, may be higher than even 400% for some populations.

Further, we carry the study of relative performance of the estimators T_2' and $T_2^* = \lambda^* \sum_{i \varepsilon s} y_i/p_i$, $p_i = x_i / \sum_{i=1}^{N} x_i (> o)$ over $\hat{Y}_{H-T}$ under an apriori distribution characterised by the usual super-population model (Godambe, 1955),

$$E(y_i|x_i) = a\, x_i$$

$$V(y_i|x_i) = \sigma^2 x_i^2 \qquad (1.5.7)$$

$$\text{and} \quad Cov(y_i, y_j|x_i, x_j) = 0$$

and the following results have been noted :

(i) If $\sigma^2/c^2 = C^2(y|x)$, be known exactly, then $T^*_{o2} = \lambda^*_o \sum_{i=1}^{n} y_i/r_i$

th
$$\lambda^*_o = \frac{1}{n}[(n+C^2(y|x)\sum_{i=1}^{N}\pi_i r_i)/(n+C^2(y|x)]$$

∂_1-better than $\hat{Y}_{H-T}$ in the T_2-class if

$$\pi_i \leq 1/2 \quad and \quad t \geq (1-2\sum_{i=1}^{N}\pi_i P_i) \qquad (1.5.8)$$

ere, $t = V(\hat{X}_{H-T})/[E(\hat{X}_{H-T})]^2$.

(ii) If $C^2(y|x)$ is not known exactly, but it is known that

$$C^2_{(1)}(y|x) \leq C^2(y|x) \leq C^2_{(2)}(y|x),$$

a sufficient condition for T^*_2 to be ∂_1-better than $\hat{Y}_{H-T}$ has en found.

(iii) In case $C^2(y|x)$ is known exactly, then the estimator from T'_2 with optimum choice of λ, i.e.,

$$T'_{o2} = [(1+C^2(y|x)\sum_{i=1}^{N}P_i^2)/(1+C^2(y|x)\sum_{i=1}^{N}P_i^2/\pi_i)]\ \hat{Y}_{H-T}$$

has been found to be ∂_1-better than $\hat{Y}_{H-T}$. However, in case $C^2(y|x)$ is not known exactly, similar types of results as in (ii) have been obtained.

Further an empirical study has been made and the percent relative efficiency of T'_2 and T^*_2 over $\hat{Y}_{H-T}$ are obtained. It is interesting to note that in some cases percent relative efficiency is even more than 5000%.

The results obtained in Chapter IV obviously apply (with obvious modifications) to the problem of estimation of population mean $\bar{Y}$, too, in case of general sampling designs, the estimators being T_2-type of linear class of estimators for $\bar{Y}$.

In Chapter V, we discuss the problem of estimation of the population variance σ_y^2 in case of SRSWR. We find that an estimator

$$T_1 = \lambda_1 \, s_y^2$$

would be better than s_y^2

$$\text{if} \quad 2(n/\Delta_{(1)}) - 1 < \lambda_1 < 1, \quad \Delta_{(1)} > n$$

$$\text{or if} \quad n/\Delta_{(1)} < \lambda_1 < 1$$

where $n < \Delta_{(1)} = \beta_2^{(1)} + \frac{(n^2-2n+3)}{(n-1)} \leq \Delta$.

Further, we study the general properties of the class of estimators

$$d = \lambda_1 \, s_y^2 + \lambda_2 t$$

and find that for a specified λ_1, estimators from d would be better than $T_1 = \lambda_1 s_y^2$

iff λ_2 lies between 0 and $2[(1-\lambda_1)\sigma_y^2 Et - \lambda_1 Cov(s_y^2, t)]/Et^2$.

We obtain the necessary and sufficient conditions under which the estimators $d^* = T_1^* + \lambda_2 t$, for $t = \sigma_o^2$, an apriori value of σ_y^2 $t = \bar{y}$, $t = s_y^2/\bar{y}$ etc., are better than s_y^2 and the modified estimator $T_1^* = (n/\Delta^*)s_y^2$ with $\Delta^* \leq \Delta$ and also than $T_{o1} = (n/\Delta)s_y^2$ in case β_2 is known. We also find the sufficient conditions which are dependent merely on the knowledge of the lower or upper bounds of the parameters involved.

We discuss the relative performance of the estimators proposed by us and those by Pandey, Singh and Hirano (1973), Pandey and Singh (1977) and Das and Tripathi (1978a).

Further, we undertake an empirical study based on a natural population data and obtain the percent relative efficiency of the estimators $d' = s_y^2 + \lambda_2 t$ over s_y^2 and that of $d^* = T_{o1} + \lambda_2 t$ over $T_{o1} = (n/\Delta)\, s_y^2$ for some choices of t. In some cases, the percent relative efficiency is found to be even more than 200%.

Finally, we find that in case of normal populations, the estimators

$$d_1^{**} = T_{o1} + [2/(n+1)]\ \sigma_{(1)}^2$$

$$d_3^{**} = T_{o1} + 2\ \bar{Y}_{(1)}[n/(n+1)][C_{(1)}^2/(n+C_{(1)}^2)]\bar{y}$$

$$d_4^{**} = T_{o1} + [(n+C_{(1)}^2)/(n+6C_{(1)}^2)][C_{(1)}^2/(n+1)]\bar{y}^2$$

and $$d_6^{**} = T_{o1} + \frac{6[(n-1)/(n+1)]C_{(1)}^2\bar{Y}_{(1)}^2 c_y^2}{10(n-1)C_{(1)}^2+n(n+1)}$$

would be better than each of s_y^2, $s_y^{*2} = \sum_{i=1}^{n}(y_i-\bar{y})^2/n$, the maximum likelihood estimator and the Searls' type estimator T_{o1}.

In Chapter VI, we consider the utilization of the set of information $\{y_1, y_2, \ldots, y_k\}$, the y-values for k units in the population, in addition to those in the sample, for improving the usual estimators for population total Y (or $\bar{Y}$). We consider the two situations, the first being that this set of information is available, after the sample s has been already drawn from the entire population according to some sampling procedure (*aposterior* use of information) and the second being the situation that the above set of information is available in addition to the information for units in the sample drawn from remaining $N-k$ units of the population (*apriori* use of information).

We consider the aposterior use of information under the structure of general sampling designs and provide an unbiased class of estimators

$$d = w\,\hat{Y}_{H-T} + (1-w)\,d^{*}$$

with $d^{*} = \sum_{i \in s} y_i + \sum_{i \in s_k} y_i/\pi_{is}$ (Ref. 6.2)

for the population total Y, based on the information contained in the sample s as well as the additional set of information $s_k = (i, y_i)$, $i = 1, 2, \ldots, k$ from $U-s$. We find a necessary and sufficient condition under which the estimators from the proposed class are better than Horvitz-Thompson estimator $\hat{Y}_{H-T}$, based on the sample s alone.

We find that in case $\alpha = C^2(d^{*})/C^2(\hat{Y}_{H-T}) < 10$, the estimator d based on the optimum choice of $w = \alpha/[1+\alpha]$ are always better than $\hat{Y}_{H-T}$, the relative efficiency of d being 11 times more than that of $\hat{Y}_{H-T}$ for $\alpha = 0.1$. Further, we consider some special cases dealing with different sampling designs through which the set s_k can appear, given the sample s.

Then we discuss about the sampling strategies for Y, based on the apriori information (i,y_i), $i = 1,2,...,k$ in addition to the information contained in the sample s, from the remaining $N-k$ units in the population and the schemes of SRSWOR, VPSWR, and VPSWOR. In every case, we identify some estimators which are better, under very moderate conditions, than the usual estimators based on the sample s alone from the entire population.

At the end, we give numerical illustration regarding the apriori use of information and find that the additional information increases the efficiency of the estimators considerably.

In Chapter VII, we consider the problem of using information on two auxiliary characters x_1 and x_2 in defining the probability distribution $\{p_1^*,p_2^*,...,p_N^*\}$, $p_j^* = \phi(x_{1j},x_{2j})/\sum_{j=1}^{N}\phi(x_{1j},x_{2j}) > 0$, on the population $U = \{1,2,...,N\}$, where $\phi(x_1,x_2) = z$ is a real valued function of x_1 and x_2. We find the necessary and sufficient conditions under which the strategy

$$\tau \equiv \{PPZWR, \hat{Y}\},\ \hat{Y} = (1/n)\sum_{j=1}^{n}(y_j/p_j^*)$$

is better than the strategies

$$\tau_i \equiv \{PPX_iWR, \hat{Y}_i\},\ \hat{Y}_i = (1/n)X_i\sum_{j=1}^{n}(y_j/x_{ij}),\ i=1,2$$

and the strategy

$$\tau_o \equiv \{SRSWR, \hat{Y}_o\},\ \hat{Y}_o = N\bar{y}.$$

Further, we find that if

$$z_j = x_{2j} - \lambda_j x_{1j}$$

with λ_j such that

$$(x_{2j}-K\, D_{(1)}/D_{(2)})/x_{1j} \leq \lambda_j \leq (x_{2j}-1)/x_{1j}$$

where

$$D_{(1)} \leq \min_j y_j^2 \leq \max_j y_j^2 \leq D_{(2)}; \; K = \min_i \bar{X}_i/\tilde{X}_i$$

and $K\, D_{(1)}/D_{(2)} > 1$

then the strategy τ will always be better than τ_1 and τ_2.

We also give a set of necessary and sufficient conditions for the strategy τ to be η_z-better than τ_1, τ_2 and τ_0 under a super-population model η_z

$$E(y_j|z_j) = \alpha + \beta\, z_j$$

$$V(y_j|z_j) = \psi(z_j)$$

and $$Cov(y_j, y_{j'}|z_j, z_{j'}) = 0$$

where, $z_j = \phi(x_{1j}, x_{2j})$, $j = 1, 2, \ldots, N$.

In case, $\psi(z_j) = \gamma \cdot z_j^g$ and $\alpha = 0$, we find that the strategy τ with $z_j = x_{2j} - \lambda_j x_{1j} \geq 1$ and $x_{ij} > 0$ $(i=1,2)$ is η_z-better than both of τ_1 and τ_2,

if $(x_{2j}-K)/x_{1j} \leq \lambda_j \leq (x_{2j}-1)/x_{1j}$ for $0 \leq g \leq 1$

and if $(x_{2j}-K^{1/g})/x_{1j} \leq \lambda_j \leq (x_{2j}-1)/x_{1j}$ for $g \geq 1$

where, $K = \min_{1 \leq i \leq 2} \{\bar{X}_i/\tilde{X}_i\} > 1$.

Further, in case $C_j^2 = V(y_j|x_{1j},x_{2j})/[E(y_j|x_{2j},x_{1j})]^2 = C^2$, constant or $g = 2$ in the n_z-model with $\alpha = 0$, we find τ to always better than τ_1, τ_2 and τ_o. In other cases, a set of ficient conditions are found for τ to be better than τ_1, τ_2 τ_o. The strategy τ is also compared with the strategies ed on SRSWR and two-variate ratio-estimators due to Olkin (1958) Tripathi (1978).

We give empirical evidences in support of some of the results.

CHAPTER II

SOME RESULTS ON T_1 CLASS OF LINEAR ESTIMATORS

2.1 Introduction and Summary

Let $U = \{1,2,\ldots,N\}$ be a finite population of N (given) units labelled through 1 to N and y be a variate (real) which takes value y_i for the ith unit of the population $(i = 1,2,\ldots,N)$. Let $\bar{Y} = \sum_{i=1}^{N} y_i/N$, $\sigma_y^2 = \sum_{i=1}^{N}(y_i-\bar{Y})^2/N$ and $C_y = \sigma_y/\bar{Y}$ be the population mean, variance and coefficient of variation of y respectively. It is desired to estimate the mean $\bar{Y}$ on the basis of a sample of n units drawn according to simple random sampling without replacement (SRSWOR).

The T_1-class of linear estimators for $\bar{Y}$ based on a sample of size n may be defined by

$$\hat{T}_1 = \sum_{r=1}^{n} a_r\, y_r \qquad (2.1.1)$$

where $a_r (r = 1,2,\ldots,n)$ is the weight (constant) associated with the y-value of the unit appearing at the rth draw,[Horvitz and Thompson, (1952); Koop (1957), (1963)].

When $a_r = \lambda/n$, for all $r = 1,2,\ldots,n$, then the estimator in (2.1.1) reduces to

$$\hat{T}_1^* = \lambda\bar{y} \qquad (2.1.2)$$

and the optimum value of λ which minimises mean square error, $M(\hat{T}_1^*)$ of $\hat{T}_1^*$ in (2.1.2) is found to be

$$\lambda_o = 1/[1+K\, C_y^2] \qquad (2.1.3)$$

case of simple random sampling without replacement, where $= \frac{(N-n)}{n(N-1)}$. The resulting estimator $\lambda_o \bar{y}$ was discussed by Searle (1964) and is defined by

$$\hat{T}_S = \bar{y}/[1+K\, C_y^2]$$

with its bias and MSE as

$$B(\hat{T}_S) = -K\, C_y^2\, \bar{Y}/[1+K\, C_y^2] \qquad (2.1.4)$$

$$\text{and} \quad M(\hat{T}_S) = K\, \bar{Y}^2 C_y^2/[1+K\, C_y^2] .$$

(2) Obviously, $\hat{T}_S$, a member of T_1-class, is better than the sample mean $\bar{y}$ (better in the sense of having smaller mean square error) and the relative efficiency of Searls estimator $\hat{T}_S$ over $\bar{y}$, R.E., is found to be $R.E. = [1 + KCy^2]$.

It is well known that in case of general sampling designs, there does not exist best linear unbiased estimator in any unbiased sub-class of the class of linear estimators $\hat{T}_1$ [Koop, (1957), (1963); Ajgaonkar (1969)]. However in case of SRSWOR, sample mean $\bar{y}$ is found to be the best in the unbiased sub-class of the class $\hat{T}_1$ in (2.1.1). The question arises: Does there exist the best linear (uniformly minimum mean square error, UMMSE) estimator in the entire linear class $\hat{T}_1$? Further, are there some biased estimators in $\hat{T}_1$ better than $\bar{y}$?

(2) It may be mentioned here that in the regression analysis Hoerl and Kennard (1970(a), 1970(b)) suggested the "ridge" estimator for the regression coefficients which for the one regressor case reduces to the expression similar to Searle's estimator for $\bar{Y}$, the population mean.

In this chapter, we answer the questions confining ourselves to SRSWOR.

We find that the strategy $\tau_1 \equiv (SRSWOR, \hat{T}_S)$ is the UMMSE-strategy in the class of strategies $\tau_2 \equiv (SRSWOR, \hat{T}_1)$ and identify a subclass of τ_2 which contains the strategies better than the strategy $\tau_o \equiv (SRSWOR, \bar{y})$. We find that a large number of biased estimators in $\hat{T}_1$ better than $\bar{y}$ can be generated depending merely upon the knowledge of a quantity $C_{(1)}$ which is such that $C^2_{(1)} \leq C^2_y$. The relative efficiency of $\hat{T}_S$ over $\bar{y}$ is studied for different values of C_y and some numerical illustrations have also been given.

2.2 Existence of the UMMSE-estimator in $\hat{T}_1$

Theorem 2.2.1 If C_y is known exactly, then the sampling strategy $(SRSWOR, \hat{T}_S)$ is the best in the class of strategies $(SRSWOR, \hat{T}_1)$ for $\bar{Y}$.

Proof. The mean square error (MSE) of the estimator $\hat{T}_1$ defined in (2.1.1) is found to be

$$M(\hat{T}_1) = N\sigma_y^2 \sum_{r=1}^{n} a_r^2/(N-1) - \sigma_y^2(\sum_{r=1}^{n} a_r)^2/(N-1) + \bar{Y}^2(\sum_{r=1}^{n} a_r - 1)^2 \qquad (2.2.1)$$

It may be shown that $M(\hat{T}_1)$ would be minimum for

$$a_r = 1/[n(1+KC_y^2)] \qquad (2.2.2)$$

For these choices of a_r's, $\hat{T}_1$ clearly reduces to $\hat{T}_S$ and hence the result.

Remarks

(i) It may be shown that $\hat{T}_1^*$ in (2.1.2) would be better than $\bar{y}$ under simple random sampling without replacement,

$$\text{iff} \quad [1 - KC_y^2]/[1 + KC_y^2] < \lambda < 1 \tag{2.2.3}$$

and hence a sufficient condition for $\hat{T}_1^*$ in (2.1.2) to be better than $\bar{y}$ would be

$$[1 - KC_{(1)}^2]/[1+KC_{(1)}^2] \leq \lambda < 1 \tag{2.2.4}$$

which may be modified to $1/[1 + KC_{(1)}^2] \leq \lambda < 1$.

Let us call $\hat{T}_1^*$ in (2.1.2) with λ satisfying (2.2.4), a modified Searls' estimator $\hat{T}_S'$, i.e.,

$$\hat{T}_S' = \lambda\bar{y}, \quad \lambda\epsilon[(1-KC_{(1)}^2)/(1+KC_{(1)}^2),1) \text{ or } \lambda\epsilon[1/(1+KC_{(1)}^2),1) \tag{2.2.5}$$

(ii) To find a λ in modified Searls's estimator, one needs only apriori information $C_{(1)}$ about C_y. Though it is not impossible, but may be very very difficult to find weights differing from each other at least for some r such that $\hat{T}_1$ is better than both of $\bar{y}$ and $\hat{T}_S'$ and the gain in efficiency of $\hat{T}_1$ over $\hat{T}_S'$ is appreciable (we are going to study this in the Section 2.4), in fact one may safely use $a_r = \lambda/n$ for all r where λ is such that $\hat{T}_S'$ is better than $\bar{y}$ and bias $(\lambda-1)\bar{Y}$ in $\hat{T}_S'$ is not appreciable. It may be noted that on th average $\hat{T}_S'$ is always an underestimate of $\bar{Y} > 0$ and $B(\hat{T}_S') < 0$. It is always preferable to use λ near 1, but not very close to 1 as gain by using $\hat{T}_S'$ would not be appreciable over the use of $\bar{y}$.

(iii) The percent relative efficiency R.E. = $[V(\bar{y})/M(\hat{T}_S)]\times 100$ of $\hat{T}_S$ over $\bar{y}$ and the absolute relative bias, R.B. = $|B(\hat{T}_S)/\bar{Y}|$ of $\hat{T}_S$ is given in the Table 2.2.1 for various values of C_y and K. It is found that the efficiency of $\hat{T}_S$ over $\bar{y}$ is almost negligible in case both $K < 0.01$ and $C_y < 1$. Thus, the Searls' estimator should be used only in the other situations.

Table 2.2.1: Percent Relative Efficiency of the Searle's estimator over sample mean for $K = 0.005, 0.01, 0.05\ (0.05), 0.4$ and $C_y = 0.1, 0.5(0.5), 3.0$.

K		C_y						
		0.1	0.5	1.0	1.5	2.0	2.5	3.0
(1)		(2)	(3)	(4)	(5)	(6)	(7)	(8)
.005	R.E.	100.005	100.125	100.500	101.125	102.000	103.125	104.500
	R.B.	0.00005	0.00120	0.00500	0.01110	0.01960	0.03030	0.04310
0.01	R.E.	100.010	100.250	101.000	102.250	104.000	106.250	109.000
	R.B.	0.00009	0.00250	0.00990	0.02200	0.03850	0.05880	0.08260
0.05	R.E.	100.050	101.250	105.000	111.250	120.000	131.250	145.000
	R.B.	0.00050	0.01230	0.04760	0.10110	0.15670	0.23810	0.31030
0.10	R.E.	100.100	102.500	110.000	122.500	140.000	162.500	190.000
	R.B.	0.00100	0.02440	0.09090	0.18370	0.28570	0.38460	0.47370
0.15	R.E.	100.150	103.750	115.000	133.750	160.000	193.750	235.000
	R.B.	0.00150	0.03610	0.13040	0.25230	0.37500	0.48390	0.57450
0.20	R.E.	100.200	105.000	120.000	145.000	180.000	225.000	280.000
	R.B.	0.00200	0.04760	0.16670	0.31030	0.44444	0.55560	0.64290
0.25	R.E.	100.250	106.250	125.000	156.250	200.000	256.250	325.000
	R.B.	0.00250	0.05880	0.20000	0.36000	0.50000	0.60980	0.69230
0.30	R.E.	100.300	107.500	130.000	167.500	220.000	287.500	370.000
	R.B.	0.00301	0.06980	0.23080	0.40300	0.54550	0.65220	0.72970
0.35	R.E.	103.500	108.750	135.000	178.750	240.000	318.750	415.000
	R.B.	0.00350	0.08050	0.25930	0.44060	0.58330	0.68630	0.75901
.40	R.E.	104.000	110.000	140.000	190.000	260.000	350.000	460.000
	R.B.	0.00401	0.09090	0.28570	0.47370	0.61540	0.71430	0.78260

2.3 Estimators in $\hat{T}_1$ better than the sample mean

In this section we search for biased estimators in $\hat{T}_1$, based on SRSWOR, better than $\bar{y}$.

2.3.1 Let $\ell = \sum_{r=1}^{n} a_r$, $\ell_o = \sum_{r=1}^{n} a_r^2$ and $Q = \ell^2 + (\frac{N}{n} - 1) - N\,\ell_o$.

Next, we prove the following

Theorem 2.3.1. Let $(a_1, a_2, \ldots, a_n)$ be chosen such that $Q > 0$. Then a necessary and sufficient condition for the sampling strategy (SRSWOR, $\hat{T}_1$) to be better than the strategy (SRSWOR, $\bar{y}$) is given by

$$(N-1)\;(\ell-1)^2/Q \leq C_y^2.$$

Proof. From (2.2.1), the mean square error (MSE) of the estimator $\hat{T}_1$ defined in (2.1.1) is found to be

$$M(\hat{T}_1) = \bar{Y}^2[(\ell-1)^2 + \frac{(N\ell_o - \ell^2)\;C_y^2}{(N-1)}] \tag{2.3.1}$$

and

$$V(\bar{y}) = \bar{Y}^2\;K\;C_y^2 \tag{2.3.2}$$

Obviously,

$$M(\hat{T}_1) \leq V(\bar{y})$$

iff
$$\frac{(N\,\ell_o - \ell^2)C_y^2}{(N-1)} + (\ell-1)^2 - K\;C_y^2 \leq 0$$

i.e., iff
$$C_y^2[N\,\ell_o - \ell^2 - \frac{N}{n} + 1] + (N-1)(\ell-1)^2 \leq 0$$

i.e., iff
$$(N-1)\;(\ell-1)^2 - Q\;C_y^2 \leq 0. \tag{2.3.3}$$

Now taking $Q > 0$ in (2.3.3), the result of the Theorem follows.

Remarks

(i) Obviously, the inequality (2.3.3) can never be satisfied if $Q \leq 0$ whatever be the choice of a_r's.

(ii) It may be shown that in case $a_r = 1/n(1+K\,C_y^2)$, $Q > 0$ and $(N-1)\ (\ell-1)^2 < Q\,C_y^2$ always holds true.

(iii) For the inequality in (2.3.3) to be satisfied, exact knowledge of C_y^2 is not always required, if $C_{(1)}^2$ be a quantity such that $C_{(1)}^2 \lesssim C_y^2$, then a sufficient condition for $\hat{T}_1$ defined in (2.1.1) better than $\bar{y}$ would be given by (2.3.3) with C_y^2 replaced by $C_{(1)}^2$. Thus merely the approximate knowledge of C_y^2 will enable us to find at least one set of values $(a_1, a_2, \ldots, a_n)$ such that the sampling strategy (SRSWOR, $\hat{T}_1$) would be better than the strategy (SRSWOR, $\bar{y}$).

(iv) In the Table 2.3.1, the percent relative efficiencies, R.E. of $\hat{T}_S$ and $\hat{T}_1$ over the sample mean $\bar{y}$ and absolute relative biases, R.B., are given for the different populations, where coefficient of variations are greater than or equal to 1. We fix the population size $N = 25$ and sample size $n = 5$. The weights a_r's in $\hat{T}_1$ are taken arbitrarily with $\ell = \sum_{r=1}^{5} a_r = 0.8$ and such that $Q > 0$ and (2.3.4) is satisfied with $C_{(1)} = 1.0$.

This Table (2.3.1) shows that one may generate estimators from $\hat{T}_1$, with arbitrary weights, better than $\bar{y}$ even when C_y is not known exactly, the case in which Searls' estimator $\hat{T}_S$ can not be used.

Table 2.3.1 Percent relative efficiency of $\hat{T}_1$ over y for arbitrary weights. N=25, n=5, $C_y \geq 1$; $\ell = 0.8$
$a_1 = 0.1$, $a_2 = 0.2$, $a_3 = 0.2$, $a_4 = 0.1$, $a_5 = 0.2$

R.E & R.B.	C_y						
	1.0	1.5	2.0	2.5	3.0	3.5	4.0
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$[V(\bar{y})/M(\hat{T}_S)] \times 100$	116.66	137.48	166.64	204.12	249.94	304.08	366.56
$\lvert B(\hat{T}_S)/\bar{Y}\rvert$	0.1431	0.2731	0.4004	0.5107	0.6005	0.6717	0.7277
$[V(\bar{y})/M(\hat{T}_1)] \times 100$	104.29	121.22	128,51	132.20	134.29	135.58	136.44
$\lvert B(\hat{T}_1)/\bar{Y}\rvert$	0.20	0.20	0.20	0.20	0.20	0.20	0.20

2.3.2 A Procedure to Choose a_r's.

Now in what follows, we give a procedure for making choices of a_r's in $\hat{T}_1$ such that the results stated in Theorem 2.3.1 may be implemented in practice. We also provide some numerical illustrations.

From Theorem 2.3.1, $\hat{T}_1$ defined in (2.1.1) would be better than $\bar{y}$ if

$$(N-1)\ (\ell-1)^2/Q \leq C^2_{(1)} \qquad (2.3.4)$$

Let $a_r = r/\lambda$, (2.3.5)

where $r(1\leq r\leq n)$ is a positive integer and λ is any real number satisfying $Q > 0$. Then from (2.3.4), we have the following inequality

$$(N-1)[\ell^2-2\ell+1] \leq C^2_{(1)}[\ell^2+\frac{N}{n} - 1 - N\,\ell_o]$$

which, after observing that $\ell = \Sigma a/\lambda = n(n+1)/2$ and $\ell_o = n(n+1)\,(2n+1)/6\lambda^2$, can be written as

$$(N-1)[\frac{n^2(n+1)^2}{4\lambda^2} - \frac{n(n+1)}{\lambda} + 1] \leq C^2_{(1)}[\frac{n^2(n+1)^2}{4\lambda^2} + \frac{N}{n} - 1 - \frac{Nn(n+1)(2n+1)}{6\lambda^2}] \quad (2.3.6)$$

The inequality (2.3.6) can again be expressed equivalently by

$$q(\lambda) \leq 0 \quad (2.3.7)$$

where

$$q(\lambda) = \alpha\lambda^2 + \beta\lambda + \gamma$$

$$\alpha = (N-1)\ (1-K\ C^2_{(1)})$$

$$\beta = -(N-1)\ n(n+1)$$

and

$$\gamma = \frac{n(n+1)}{2}[\frac{n(n+1)}{2}(N-1-C^2_{(1)}) + \frac{N(2n+1)C^2_{(1)}}{3}] \quad (2.3.8)$$

Let D be the discriminant of $q(\lambda) = 0$ and let $f = \frac{n}{N}$, be the sampling fraction. Then D is found to be

$$= (N-1)^2 n^2 (n+1)^2 - 4(N-1)^2 (1-NC^2_{(1)}) \left[\frac{n^2(n+1)^2}{4}\left(1-\frac{C^2_{(1)}}{N-1}\right) + \frac{n(n+1)(2n+1)C^2_{(1)}}{6} \times \frac{N}{N-1}\right]$$

$$= (N-1)^2 \Big[\{Nf(Nf+1)\}^2 - \left\{4 - \frac{4(1-f)C^2_{(1)}}{(N-1)f}\right\}\left\{\frac{Nf(Nf+1)}{2}\right\}^2 \left\{1 - \frac{C^2_{(1)}}{N-1}\right\} + \left\{\frac{N^2 f(Nf+1)(2Nf+1)C^2_{(1)}}{6(N-1)}\right\}\Big]$$

$$= (N-1)^2 \Big[\{Nf(Nf+1)\}^2 \frac{C^2_{(1)}}{N-1} + \left\{(1-f)C^2_{(1)}\left(1 - \frac{C^2_{(1)}}{N-1}\right)\frac{(Nf(Nf+1))^2}{(N-1)f}\right\} + \left\{\frac{4}{6} \cdot \frac{N}{N-1} Nf(Nf+1)(2Nf+1)\, C^2_{(1)} \left(\frac{(1-f)C^2_{(1)}}{(N-1)f} - 1\right)\right\}\Big]$$

$$= (N-1)^2 \Big[\left\{(Nf(Nf+1))^2 \frac{C^2_{(1)}}{N-1} \left(\frac{1}{f} - \frac{(1-f)C^2_{(1)}}{(N-1)f}\right)\right\} - \left\{4N^2 f(Nf+1)(2Nf+1)\, C^2_{(1)} \left(1 - \frac{(1-f)C^2_{(1)}}{(N-1)f}\right)/6(N-1)\right\}\Big]$$

$$= (N-1)^2 \Big[\left\{\frac{N^2C^2_{(1)} f(Nf+1)}{(N-1)}\right\} \left\{(Nf+1) - \frac{2(2Nf+1)}{3}\right\} - \left\{\frac{N^2C^4_{(1)}(1-f)}{f(N-1)^2}\right\} f(Nf+1) \left\{f(Nf+1) - \frac{2(2Nf+1)}{3}\right\}\Big]$$

$$= (N-1)^2 \frac{N^2C^2_{(1)} f(Nf+1)}{(N-1)} \Big[\left\{(Nf+1) - \frac{2(2Nf+1)}{3}\right\} - \frac{C^2_{(1)}(1-f)}{f(N-1)} \cdot \left\{f(Nf+1) - \frac{2(2Nf+1)}{3}\right\}\Big]$$

$$= (N-1)\, N^2 C^2_{(1)} f(Nf+1) \Big[(Nf+1) \left\{1 - \frac{C^2_{(1)}(1-f)}{f(N-1)}\right\} - \frac{2(2Nf+1)}{3}\left\{1 - \frac{C^2_{(1)}(1-f)}{f(N-1)}\right\}\Big] \tag{2.3.9}$$

and hence, it may be shown that a sufficient condition for $q(\lambda) = 0$ to admit two real roots is given by

$$(Nf+1)[(N-1)f - (1-f)(3f-2)C^2_{(1)}] - 2Nf[f(N-1+C^2_{(1)}) - C^2_{(1)}] > 0 \tag{2.3.10}$$

obviously, for

$$f < \min\{\frac{2}{3}, \frac{C^2_{(1)}}{N-1+C^2_{(1)}}\} \tag{2.3.11}$$

the condition (2.3.10) is satisfied.

Let λ_1 and λ_2 be two roots of $q(\lambda) = 0$; we note that

$$\begin{aligned} &\alpha < 0 \quad \text{iff} \quad C^2_{(1)}/(N-1+C^2_{(1)}) > f \\ \text{and} \quad &\alpha < 0 \quad \text{iff} \quad C^2_{(1)}/(N-1+C^2_{(1)}) < f \end{aligned} \tag{2.3.12}$$

obviously, if (2.3.11) is satisfied, then for $\lambda < \lambda_1$ or $\lambda > \lambda_2$, (2.3.7) will always be satisfied as α would be negative. Further in case $\alpha > 0$, one may generate the values of f with $C^2_{(1)}/(N-1+C^2_{(1)}) < f < 1$, such that (2.3.10) is satisfied.

Let $R_{0\lambda}$, $R_{1\lambda}$, $R_{2\lambda}$ and $R_{3\lambda}$ denotes the ranges for λ for which $Q > 0$, $\lambda_1 < \lambda$, $\lambda > \lambda_2$, and $\lambda_1 < \lambda < \lambda_2$ respectively. Then obviously, from Theorem 2.3.1, the estimators

$$\hat{T}'_1 = \frac{1}{\lambda} \sum_{r=1}^{n} r\, y_r \tag{2.3.13}$$

will be better than $\bar{y}$

if $\lambda \in R_{o\lambda} \cap R_{1\lambda}$ or $\lambda \in R_{o\lambda} \cap R_{2\lambda}$ (2.3.14)

provided $n < n_o$

and $\lambda \in R_{o\lambda} \cap R_{3\lambda}$ in case $n > n_o$ (2.3.15)

where $n_o = N\, C^2_{(1)}/(N-1+C^2_{(1)})$.

Remarks

(i) As a general procedure to generate the weights a_r's so that $\hat{T}_1$ is better than $\bar{y}$ we proceed as follows : For given N, n and $C_{(1)}$ we find a λ such that $q(\lambda) < 0$ in (2.3.7) is satisfied. Then for $a_r = r/\lambda$, $(r=1,2,\dots,n)$ in $\hat{T}_1$, the resulting estimator will be better than $\bar{y}$.

(ii) Though the expression for $q(\lambda)$ in (2.3.7) looks some what complicated, but once N, $C^2_{(1)}$ and n are known, the coefficients α, β, and γ can easily be computed and hence the roots λ_1, λ_2 of λ such that $q(\lambda) = 0$ may be obtained without any difficulty.

(iii) In estimating the mean square errors of the estimators, we require that $n \geq 2$. Then for $n_o \geq 2$, one must have $2(N-1)/(N-2) \leq C^2_{(1)} \leq C^2_y$ otherwise, we can not use (2.3.11) and (2.3.14). However, in case of using (2.3.14), no restriction on $C^2_{(1)}$ is needed.

(iv) The choices λ^* of λ for which $\hat{T}_1'$ is better than $\bar{y}$ would be infinite; Noting that bias due to $\hat{T}_1'$ i.e. $B(T_1') = [n(n+1)/2\lambda - 1]\bar{Y}$, we should take that λ^* which is closer to $n(n+1)/2$, so that bias is also controlled to some extent.

Numerical Illustrations

For the application of our procedures discussed in 2.3.2, we give the following two numerical examples.

(a) Let us consider a population with $N=51$, $C_y > 4$. Let us take $C_{(1)}^2 = 10$. To satisfy (2.3.11), we must have $n < \min(\frac{102}{3}, \frac{51}{6})$. Let us take $n = 5$. This gives

$$Q = 9.2 - (2580/\lambda^{2} .$$

Obviously, for all $\lambda \geq 17$ or $\lambda \leq -17$, we shall have $Q > 0$. Now the roots of $q(\lambda) = 0$ are given by

$$\lambda_1 = -54.45 \quad \text{and} \quad \lambda_2 = 16.95.$$

Therefore, from (2.3.14), for any

$$\lambda > \max(17, 16.95) \text{ or } \lambda < \min(-17, -54.45)$$

the estimator in $\hat{T}_1'$ in (2.3.13) will be better than sample mean $\bar{y}$. Further taking $\lambda = 20$, we have

$$|B(\hat{T}_1')/\bar{Y}| = 0.25$$

and the relative efficiency of $\hat{T}_1'$ over $\bar{y}$ as

$$[V(\bar{y})/M(\hat{T}_1')] = 1 + [102C_y^2 - 125]/[258C_y^2+125].$$

Obviously, $[V(\bar{y})/M(\hat{T}_1')]$ is an increasing function in C_y^2. Thus larger the coefficient of variation, higher would be the relative efficiency of $\hat{T}_1'$ over $\bar{y}$.

(b) Let there be another population with $N = 20$ and $C_y > 1$. Let $C_{(1)}^2 = 1$. Then to have $n > n_o$, we must have, from (2.3.15), $f > 0.5$. We find that for $f = 0.2$, the relation (2.3.10) is satisfied. So, let $f = 0.2$, i.e., $n = 4$. This gives

$$Q = 4 - (500/\lambda^2).$$

Obviously for $\lambda \geq 11.2$ and $\lambda \leq -11.2$, we shall have $Q > 0$. It is found that the roots of $\varsigma(\lambda) = 0$ are

$$\lambda_1 = 12.04 \text{ and } \lambda_2 = 13.29.$$

Hence from (2.3.15), for any $\lambda \varepsilon (12.1, 13.2)$, the estimator T_1' would be better than $\bar{y}$ Further taking $\lambda = 12.5$, we have

$$|B(\hat{T}_1')/\bar{Y}| = 0.20$$

and $$[V(\bar{y})/M(\hat{T}_1')] = 1 + [20C_y^2 - 19]/[80C_y^2 + 19].$$

2.4 Comparison of Choosing Different Weights in $\hat{T}_1$ Againts the Use of Equal Weights

Theorem 2.3.1 assures the superiority of an estimator $\hat{T}_1 = \sum_{r=1}^{n} a_r y_r$ over $\bar{y}$, the sample mean, but it does not guarantee whether $\hat{T}_1$ will be better than modified Searle's estimator $\hat{T}_S'$. Moreover, we note that optimum choice of a_r's does not depend on r. This indicates that one may take $a_r = \lambda_1$, a constant for all $r = 1,2,\ldots,n$ and study the performance of choosing different weights a_r's in $\hat{T}_1$ compared to the same weight $a_r = \lambda_1$ for all $r = 1,2,\ldots,n$ which will reduce to $\hat{T}_1$ to

$$T_1^* = n\,\lambda_1\,\bar{y} = \lambda\,\bar{y}. \tag{2.4.1}$$

Obviously, we should chose λ such that $[1-KC_y^2]/[1+KC_y^2] < \lambda < 1$, as for other choices of λ, $\bar{y}$ will always be better than $\hat{T}_1^*$ The optimum choice of λ would be $\lambda_0 = 1/[1+KC_y^2]$ and for which $\hat{T}_1^*$ would reduce to $\hat{T}_S$ defined in (2.1.4).

In this section, we shall observe that there always exists at least one set of choices $(a_1,a_2,\ldots,a_n)$ with all $a_r \neq \lambda (\neq \lambda_0)$ such that the strategy (SRSWOR, $\hat{T}_1$) is better than the strategy (SRSWOR, $\hat{T}_S'$) and hence the strategy (SRSWOR, $\bar{y}$).

Let ℓ and ℓ_o be the same as in Theorem 2.3.1 and let

$$[2/(1+KC^2_{(1)})] - \lambda < \ell < \lambda \qquad (2.4.2)$$

then we have the following

Theorem 2.4.1 A sufficient condition that the strategy (SRSWOR, $\hat{T}_1$) is better than the strategy (SRSWOR, $\hat{T}'_S$) and hence the strategy (SRSWOR, $\bar{y}$) would be

$$\ell^2/n < \ell_o < \frac{1}{n}[\lambda^2 - \{2(\lambda-\ell)/(1+KC^2_{(1)})\}] \qquad (2.4.3)$$

Proof. We have, from (2.2.1),

$$M(\hat{T}_1) = \bar{Y}^2[\frac{N}{N-1}\ell_o C^2_y - \frac{\ell^2 C^2_y}{N-1} + (\ell-1)^2]$$

and $\quad M(\hat{T}'_S) = \bar{Y}^2[\lambda^2 KC^2_y + (\lambda-1)^2]$.

It is found that $M(\hat{T}_1) \leq M(\hat{T}'_S)$,

$$\text{iff} \quad \frac{N}{N-1} C^2_y \ell_o + \ell^2(1 - \frac{C^2_y}{N-1}) < \lambda^2(1-KC^2_y) - 2(\lambda-\ell) \qquad (2.4.4)$$

Since, $\ell_o \geq \ell^2/n$ a sufficient condition for (2.4.4) to be true is obtained through replacing ℓ^2 by $n\ell_o$, where it is assumed that $C^2_y < (N-1)$. Thus $M(\hat{T}_1) < M(\hat{T}'_S)$

$$\text{if} \quad \ell_o < (1/n)[\lambda^2 - 2(\lambda-\ell)/(1+KC^2_y)] \qquad (2.4.5)$$

and hence

$$\text{if} \quad \ell_o < (1/n)[\lambda^2 - 2(\lambda-\ell)/(1+KC^2_{(1)})] \qquad (2.4.6)$$

provided $\lambda > \ell$.

Remarks

(i) We can not take $\ell = \lambda$, because in that case (2.4.5) would yield $\ell_o < \ell^2/n$ which is never true.

(ii) Choice of λ should be guided by (2.2.4). Further since $\ell_o \geq \ell^2/n$ we must choose ℓ such that R.H.S. of (2.4.5) is greater than ℓ^2/n. Thus ℓ should be chosen to satisfy (2.4.2).

Numerical Illustration

Let $C_y \geq 0.5$, $N = 25$ and $n = 5$. Let $C_{(1)} = 0.5$ and λ and $\ell = \sum_{r=1}^{n} c_r$ are such that they satisfy (2.2.4) and (2.4.2) respectively. Let λ and ℓ be fixed at $\lambda = 0.98$ and $\ell = 0.97$. Then from (2.4.3), we must have

$$0.18810 \leq \ell_o < 0.18826$$

let us take $\ell_o = 0.18825$ and let

$$a_1 = 0.1919,\ a_2 = 0.1919,\ a_3 = 0.1918,\ a_4 = 0.1918,\ a_5 = 0.2026.$$

Percent Relative Efficiency	C_y^2			
	0.25	1.00	2.25	4.00
(1)	(2)	(3)	(4)	(5)
$[V(\bar{y})/M(\hat{T}_S)] \times 100$	104.16	116.66	137.49	166.66
$[V(\bar{y})/M(\hat{T}_1)] \times 100$	103.86	109.97	110.34	110.47
$[V(\bar{y})/M(\hat{T}_S')] \times 100$	103.09	103.86	104.01	104.05

CHAPTER III

USE OF PRIOR INFORMATION ON SOME PARAMETERS IN ESTIMATING POPULATION MEAN

3.1 Introduction and Summary

It is well-known that the use of supplementary information in a suitable manner at the designing or pre-selection stage and at the stage of sample selection and/or estimation stage generally results into improved estimators of the population parameters. The usual techniques in this respect assume that the values of one or more supplementary (auxiliary) variables related to the characteristic of interest are known or may be made to be known without much difficulty for each unit of the population. In many cases, however, such detailed apriori information may not be available or may be quite costly to collect. On the other hand, some summary information, for instance, an apriori value of the parameter θ, under consideration, quite close to its true value, may be known to the experimenter. Such an information may be available from census, surveys, or even from expert guesses by the specialists in the concerned field. It may also happen that the upper and lower limits of θ may be known (Dalenius, 1965) in which case a simple or modified average (depending on the expected skewness of the distribution of the variate under consideration or the apriori distribution of θ) may provide a good approximation to θ.

Many times some apriori information may be available about some other parameters as well in addition to that about θ, the parameter under estimation. We have already discussed in Section 1.4 the various forms in which apriori information may be available. It seems worthwhile to develop estimators utilizing this information about θ or some other parameters, so that their mean square error is considerably smaller than the variance of the usual unbiased estimator of θ. In this chapter we propose such a class of estimators at first for any θ in case of general sampling design and then in particular for population mean $\bar{Y}$ in case of simple random sampling and discuss about the practical applicability of such estimators.

Searls (1964) discussed about an estimator $\hat{T}_S = n\bar{y}/(n+C_y^2)$ for $\bar{Y}$ defined in (2.1.4) and gave the bias and mean square error, in case, coefficient of variation C_y of y is known exactly.

Hirano (1972) considered an estimator for $\bar{Y}$, utilising the knowledge of a good-guessed value of C_y^2, say C_o^2 His proposed estimator is

$$T_H = n\bar{y}/(n+C_o^2) \qquad (3.1.1)$$

Thompson (1968) considered some 'shrinkage techniques' in an effort to improve the uniformly minimum variance unbiased estimator (UMVUE) for θ through permitting some amount of bias in estimation.

In case an apriori value θ_o of θ is available, it may be reasonable to take the usual estimator $\hat{\theta}$ (which may be UMVUE or the maximum likelihood estimator or some other estimator for θ) and move it closer to θ_o. Thus Thompson proposed to use the shrunken estimator

$$\hat{\theta}_S = [(\hat{\theta}-\theta_o)^2/\{(\hat{\theta}-\theta_o)^2 + \hat{Var}(\hat{\theta})\}](\hat{\theta}-\theta_o) + \theta_o \quad (3.1.2)$$

as an estimator for θ.

In particular, he has examined the problem of estimating the means of univariate normal, binomial, Poisson and Gamma distributions. The various procedures employed shrink the usual (MVULE's) towards μ_o, an apriori value of the mean μ. It is observed by him that the shrunken estimator

$$\hat{\mu} = [(\bar{y}-\mu_o)^2/\{(\bar{y}-\mu_o)^2 + s_y^2/n\}](\bar{y}-\mu_o) + \mu_o \quad (3.1.3)$$

where s_y^2 is the sample variance, yields increased efficiency near $(\mu-\mu_o)/\sigma_{\bar{y}} = 0$ at the expense of poorer performance for moderate values of $|(\mu-\mu_o)/\sigma_{\bar{y}}|$. For $|(\mu-\mu_o)/\sigma_{\bar{y}}|$ large, the above shrunken estimator gives MSE values only infinitesimally greater than those of $\bar{y}$.

In case of normal distributions, he considered the estimators

$$\tilde{T}_S = n\bar{y}/(n+c_y^2)$$
$$\tilde{T}_S(\alpha) = n\bar{y}/(n+\alpha\, c_y^2) \quad (3.1.4)$$

for $\bar{Y}$, where $c_y^2 = s_y^2/\bar{y}^2$, is an estimate of C_y^2 and α is a constant, and studied their performance compared to that of $\bar{y}$. Similar types of estimators were proposed and discussed by him in case of binomial and Poisson populations, as special cases of general shrunken estimator in (3.1.3).

The problem of shrinking the maximum likelihood estimator (MLE) $\hat{\mu}$ of the mean μ of various populations towards a natural origin μ_o was also studied by Mehta and Srinivasan (1971). Their proposed estimators are of the general form :

$$T = \hat{\mu} - a(\hat{\mu}-\mu_o)\ \exp\{-b(\hat{\mu}-\mu_o)^2/\sigma^2_{\hat{\mu}}\} \qquad (3.1.5)$$

where a and b are positive constants to be chosen suitably and $\sigma^2_{\hat{\mu}}$ denotes the variance of $\hat{\mu}$. They found their estimators to be better in an interval around μ_o.

Further Mehta and Srinivasan (1971) studied the performance of an estimator

$$T = \bar{y}[1 - a\ \exp\{-nb/C^2_y\}] \qquad (3.1.6)$$

for $\bar{Y}$ in case of normal populations and found that $MSE(T)/MSE(\bar{y})$ is a symmetric bounded function of $\delta = \sqrt{n}\bar{Y}/\sigma_y$ and that it converges to 1 as $\delta \to \pm\infty$. They also compared the performance of T with that of $\bar{y}$ and the estimators due to Thompson in (3.1.4). The discussion about some estimators of the form T in (3.1.6) was also made by them in case of binomial, Poisson and Gamma populations.

Finally, Pandey and Singh (1978)* considered the estimators

$$T_{PS} = \bar{y} - (1/n)\ s^2_y/\bar{y}$$
$$\text{and} \quad T_{PS}(\alpha) = \bar{y} + (\alpha/n)\ s^2_y/\bar{y} \qquad (3.1.7)$$

for $\bar{Y}$ and also discussed about $\tilde{T}_{S(\alpha)}$ in (3.1.4).

Das and Tripathi (1980a) and Das (1982) have considered the general problem of estimation of a population parameter using information on several statistics based on the same (or different) sample and also discussed the estimation of $\bar{Y}$, in particular. They found that the estimators $w\bar{y}_R$, $w\bar{y}_p$* $w\bar{y}_d$ and $w\bar{y}_{rg}$ are better than the usual ratio, product, difference and regression estimators $\bar{y}_R$, $\bar{y}_p$, $\bar{y}_d$ and $\bar{y}_{rg}$ respectively depending on the knowledge of $\bar{X}$, the population mean of an auxiliary character x, provided apriori information about the location of ρ, C_y and C_x is available. They hlso studied the properties of the estimators

$$d = w_1\bar{y} + w_2 s_y^2$$
$$d = w_1\bar{y} + w_2(\bar{x}-\bar{X}) \qquad (3.1.8)$$

which in fact are particular members of the very wide class of estimators discussed by us in this chapter.

In Chapter II, we have already identified a large number of biased estimators for $\bar{Y}$ better than $\bar{y}$ in case of SRSWOR, provided a quantity $C_{(1)}^2 \leq C_y^2$ is known. These estimators belong to T_1-class of homogeneous linear estimators. The study made in this chapter may in fact be regarded as an extension of our study made in Chapter II. In this chapter, we have identified a large number of estimators(biased) for $\bar{Y}$ better than $\bar{y}$, including non-homogeneous linear estimators as well as non-linear estimators.

The estimators due to Searls (1964), Hirano (1972), Pandey and Singh (1978) mentioned above, in addition to many other estimators, are identified as particular members of the class discussed by us. We have also discussed the relative performance of various estimators including those mentioned above.

We require the knowledge in an effort to generate estimators better than $\bar{y}$, $\hat{T}_S$ etc., of one or more of the quantitius $\bar{Y}_{(1)}, \bar{Y}_{(2)}, C_{(1)}, C_{(2)}, \beta_1^{(1)}, \beta_2^{(2)}$ which are such that

$$0 < \bar{Y}_{(1)} \leq \bar{Y} \leq \bar{Y}_{(2)} \; ; \quad 0 \leq C_{(1)} \leq C_y \leq C_{(2)}$$
$$0 \leq \beta_1^{(1)} \leq \beta_1 \leq \beta_1^{(2)} \; ; \; \beta_2 \leq \beta_2^{(2)}.$$

2 Searls Estimator for General Sampling Design

Let $\hat{\theta}$ be an unbiased estimator for the parameter θ. The general Searls-type estimator for θ may be defined by

$$T_1 = \lambda_1 \hat{\theta} \tag{3.2.1}$$

where λ_1 is a suitably chosen constant. The natural question arises : for what choice of λ_1, the estimator T_1 is better than $\hat{\theta}$ and what is the best choice of λ_1?

We have,

$$M(T_1) = \theta^2[\lambda_1^2(1+C^2(\hat{\theta})) - 2\lambda_1 + 1]. \tag{3.2.2}$$

Clearly, the optimum choice λ_o, which minimises (3.2.2) is found to be

$$\lambda_o = 1/[1+C^2(\hat{\theta})] \tag{3.2.3}$$

and then the resulting estimator and its optimum mean square error would be given by

$$\hat{T}_S = \hat{\theta}/[(1+C^2(\hat{\theta})]$$

$$M_o(\hat{T}_S) = \theta^2.C^2(\hat{\theta})/[1+C^2(\hat{\theta})] \tag{3.2.4}$$

and Relative Efficiency = $[1 + C^2(\hat{\theta})]$.

Obviously, the reduction in MSE is then

$$\Delta' = V(\hat{\theta})/[1 + \frac{1}{C^2(\hat{\theta})}] \tag{3.2.5}$$

where, $C^2(\hat{\theta})$ is the coefficient of variation of $\hat{\theta}$ and Δ' is a monotone function of $C(\hat{\theta})$, increasing from 0 to $V(\hat{\theta})$ as $C(\hat{\theta})$

increases from 0 to ∞ and λ_1 decreases from 1 to zero. Large reductions occur when $C(\hat{\theta})$ is large. $C(\hat{\theta})$ is usually an inverse function of sample size, and $\to 0$ as $n \to \infty$ while $\lambda_1 \to 1$.

In general, $C(\hat{\theta})$ (and therefore λ_1) is a function of θ, so $\lambda_1 \hat{\theta}$ is not a statistic usable for estimation. Thus, e.g., if θ is the population mean $\bar{Y}$ and $\hat{\theta}$ is the sample mean $\bar{y}$, $C^2(\hat{\theta}) = C_y^2/n$.

Thus our next endeavour is to find a λ_1, independent of the exact value of $C^2(\hat{\theta})$, but depending on some ranges of $C^2(\hat{\theta})$, so that the estimator T_1 in (3.2.1) would be better than $\hat{\theta}$.

Comparing $M(T_1)$ in (3.2.2) with $V(\hat{\theta})$, it may be shown that T_1 would be better than $\hat{\theta}$ for all λ_1 satisfying

$$[1 - C^2(\hat{\theta})]/[1+ C^2(\hat{\theta})] < \lambda_1 < 1 \qquad (3.2.6)$$

and hence a sufficient condition for T_1 to be better than $\hat{\theta}$ would be

$$[1 - C^2_{(1)}(\hat{\theta})]/[1 + C^2_{(1)}(\hat{\theta})] \leq \lambda_1 < 1 \qquad (3.2.7)$$

which may be reduced to

$$1/[1 + C^2_{(1)}(\hat{\theta})] \leq \lambda_1 \leq 1. \qquad (3.2.8)$$

We shall call the estimator T_1 in (3.2.1) with λ_1 satisfying (3.2.8) modified Searls' estimator $\hat{T}_S'$. The Searls' estimator in case of SRSWOR has already been discussed in detail in Chapter II.

In case of SRSWR, the estimator (3.2.4) reduces to

$$\hat{T}_S = n\bar{y}/(n+C_y^2) \qquad (3.2.9)$$

We note that we can always generate estimators, whatever be the sampling procedure, better than $\hat{\theta}$ by merely having the knowledge of some value $C_{(1)}(\hat{\theta})$ less than $C(\hat{\theta})$; closer this value to $C(\hat{\theta})$, higher would be the efficiency. Further $\hat{T}_S$ should be used in case of populations for which $C(\hat{\theta})$ is large.

3.3 A Class of Estimators for a Parameter θ, Better than an Unbiased estimator $\hat{\theta}$ and Searls-type Estimator.

In Section 3.2, we have observed that any unbiased estimator $\hat{\theta}$ for θ can be improved by modified Searls-type estimator $\hat{T}_S'$, provided the required information is available. In this section, we consider the problem of generating estimators bettef than $\hat{\theta}$, $\hat{T}_S$ and $\hat{T}_S'$, by using the information contained in an additional statistic.

We consider a class of weighted estimators, for θ, defined by

$$C_{\lambda,v}(\theta) = \{d: d = \lambda' v\} \tag{3.3.1}$$

where $v' = (\hat{\theta}, t)$ and $\lambda' = (\lambda_1, \lambda_2)$, $\hat{\theta}$ being an unbiased estimator for θ and t, being a suitably chosen statistic such that σ_t^2 exists and λ_1, λ_2 being suitably chosen constants.

We study the properties of this class and identify an optimum subclass d_o which contains the estimators having the smaller mean square error than those which are in the class d, but not in d_o.

The estimators in the proposed class $C_{\lambda,v}$ are in general biased and their bias and mean square errors are given by

$$B(d) = \lambda'\psi - \theta$$

$$\text{and} \qquad M(d) = \lambda' G\lambda - 2\theta\lambda'\psi + \theta^2 \tag{3.3.2}$$

respectively, where

$$G = \begin{pmatrix} E(\hat{\theta}^2) & E(\hat{\theta}t) \\ E(\hat{\theta}t) & E(t^2) \end{pmatrix}, \quad \psi = \begin{pmatrix} \theta \\ E(t) \end{pmatrix}$$

It may be shown that the optimum choice λ_o of λ which minimises $M(d)$, is a solution of

$$G\lambda = \theta\psi \tag{3.3.3}$$

Following Das and Tripathi (1980a), it may be shown that the equation (3.3.3) is consistent i.e., it always yields a solution for λ and that

$$M(d) = M(d)_{\lambda=\lambda_o} + (\lambda - \lambda_o)' G (\lambda - \lambda_o) \geq M(d)_{\lambda=\lambda_o}$$

where λ_o is a solution of (3.3.3). We then have

<u>Theorem 3.3.1</u>. The weight vector $\lambda(\lambda_1+\lambda_2 \neq 1)$ which minimises $M(d)$ in (3.3.2) is given by

$$\lambda_o = \theta\, \bar{G}\, \psi \tag{3.3.4}$$

and then the resulting MSE and bias would be given by

$$M_o(d) = \theta^2 [1 - \psi'(\bar{G})'\psi] \tag{3.3.5}$$

$$B_o(d) = -M_o(d)/\theta \tag{3.3.6}$$

where $\bar{G}$ is a g-inverse of the matrix G.

It may be noted that the matrix G, in general, is a non-negative definite matrix and would be non-singular, excluding the trivial case $\bar{y} = 0$ a.s. and $t = b\bar{y}$ where b is a constant. In case G is a positive definite matrix, $\lambda_o = (\lambda_{o1}, \lambda_{o2})'$ and $M_o(d)$ would be given by

$$\begin{aligned} \lambda_{o1} &= \theta[E(\hat{\theta}).\ E(t^2) - E(t).E(t\hat{\theta})]/D(\hat{\theta},t) \\ \lambda_{o2} &= \theta[E(t).\ E(\hat{\theta}^2) - E(\hat{\theta}).E(\hat{\theta}t)]/D(\hat{\theta},t) \end{aligned} \qquad (3.3.7)$$

$$\text{and} \quad M_o(d) = \theta^2[1 - \{N(\hat{\theta},t)/D(\hat{\theta},t)\}] \qquad (3.3.8)$$

where,

$$\begin{aligned} D(\hat{\theta},t) &= E(\hat{\theta}^2)\ E(t^2) - (E(\hat{\theta}t))^2 \\ &= \theta^2(E(t))^2[(1-\rho^2_{\hat{\theta},t})C^2(\hat{\theta})\ C^2(t) + C^2(\hat{\theta}) + C^2(t) \\ &\qquad - 2\rho_{\hat{\theta},t}\ C(\hat{\theta})C(t)] \end{aligned}$$

$$N(\hat{\theta},t) = \theta^2(E(t))^2[C^2(\hat{\theta}) - 2\rho_{\hat{\theta},t}\ C(\hat{\theta})\ C(t) + C^2(t)] \qquad (3.3.9)$$

$C(t)$ = Coefficient of variation of t

and $\rho_{\hat{\theta},t}$ = Correlation coefficient between $\hat{\theta}$ and t.

It may be noted from (3.3.8) that for any fixed ρ, the $M_o(d)$ is an increasing function of $C(t)$. Thus if t' and t'' are two choices of t, both having the same correlation with $\bar{y}$, then the use of t' in d would be preferable over that of t'' iff $C(t') < C(t'')$.

In practice, λ_o would not be known, as it depends upon a number of parameters. The following technique would help, in that case, to generate estimators from d better than $\hat{\theta}$, T_1 in (3.2.1) and $\hat{T}_S$ in (3.2.4).

From (3.3.2), we may write

$$M(d) = M(T_1) + \lambda_2^2\, Et^2 - 2\lambda_2\{\theta\, ET - \lambda_1\, E(\hat{\theta}t)\} \tag{3.3.10}$$

where, $M(T_1)$ has been defined in (3.2.2).

Thus for a specified λ_1, the estimator $d = T_1 + \lambda_2 t$ would be better than T_1,

$$\text{iff} \quad \lambda_2 \text{ lies between } 0 \text{ and } 2\,\lambda_{o2}^* \tag{3.3.11}$$

where

$$\lambda_{o2}^* = [(1-\lambda_1)\,\theta\, E(t) - \lambda_1.\text{Cov}(\hat{\theta},t)]/E(t^2)$$

is the optimum choice of λ_2, for a fixed λ_1 in d; obviously the resulting MSE of d, in this case, would be

$$M_o(d) = M(T_1) - \lambda_{o2}^{*2}\, E(t^2). \tag{3.3.12}$$

A subclass of d which may be quite interesting in some situations for generating estimators better than $\hat{\theta}$, is

$$d' = \hat{\theta} + \lambda_2' t \tag{3.3.13}$$

$$\text{with} \quad M(d') = V(\hat{\theta}) + 2\lambda_2'\, \text{Cov}(\hat{\theta},t) + \lambda_2'^2\, Et^2. \tag{3.3.14}$$

Obviously, the estimators from d' would be unbiased for θ, in case, t is such that $Et = 0$ or $\lambda_2' = 0$. The optimum choice λ_{o2}' of λ_2' and the resulting MSE would be given by

$$\lambda_{o2}' = -\,\text{Cov}(\hat{\theta},t)/E(t^2) \tag{3.3.15}$$

$$\text{and} \quad M_o(d') = V(\hat{\theta}) - [\text{Cov}(\hat{\theta},t)]^2/E(t^2) \tag{3.3.16}$$

It follows from (3.3.11) that estimators in d' would be better than $\hat{\theta}$ iff

$$\begin{array}{lll} \text{either} & 0 < \lambda_2' < 2\lambda_{o2}' & \text{in case } \rho_{\hat{\theta},t} < 0 \\ \text{or} & 2\lambda_{o2}' < \lambda_2' < 0 & \text{in case } \rho_{\hat{\theta},t} > 0. \end{array} \qquad (3.3.17)$$

In absence of the exact knowledge of λ_{o2}', the condition (3.3.17) may be used to generate estimators better than $\hat{\theta}$.

From (3.3.11), we note that even if we have only approximate good-guessed values for λ_{o2}^*, we may generate estimators from d better than $\hat{\theta}$, $\hat{T}_S'$ and even $\hat{T}_S$. Similarly estimators better than $\hat{\theta}$ may be generated from the subclass d' through the approximate knowledge of λ_{o2}' only. In the next sections, we shall use a technique of shrinking the intervals $(0,\lambda_{o2}^*)$ and $(0,\lambda_{o2}')$, at a loss of some efficiency of course, requiring lesser restrictive conditions for generating estimators of the type d for population mean and for specific choices of the statistic t. The above general treatment would be utilized in the particular situations of next sections.

<u>Remarks</u>

(i) if $\lambda_o = (\lambda_{o1}, \lambda_{o2})'$ be known exactly, then the optimum estimator $d_o = \lambda_{o1}\hat{\theta} + \lambda_{o2}t$ would always be better than $\hat{T}_S$ in (3.2.4) and hence $\hat{T}_S'$ and $\hat{\theta}$ also.

(ii) It is found from (3.3.14) or otherwise, that in the subclass d', we should never choose t to be uncorrelated with $\hat{\theta}$ as for such choices of t, $\hat{\theta}$ would be uniformly better than the corresponding estimators from d'.

(iii) It may be noted from (3.3.10) or otherwise, that none of the estimators in d would be better than $T_1 = \lambda_1 \hat{\theta}$ with $\lambda_1 = \theta\, E(t)/E(\hat{\theta} t)$ whatever be t.

(iv) It is found with the help of (3.3.2), that in [redacted] (a) $\rho \geq 0,\ \lambda_1 > 1,\ \lambda_2 \geq 0$, (b) $\rho \leq 0,\ \lambda_1 > 1,\ \lambda_2 \leq 0$, (c) $\rho \geq 0,\ \lambda_1 < -1,\ \lambda_2 \leq 0$, and (d) $\rho \leq 0,\ \lambda_1 < -1,\ \lambda_2 \geq 0$, none of the members in the class d would be better than $\hat{\theta}$ and hence $\hat{T}_S$ also, whatever be the choice of t.

(v) In case $t = \alpha$, is a constant in $d = \lambda_1 \hat{\theta} + \lambda_2 t$, then it may be shown that $\lambda_{o1} = 0,\ \lambda_{o2} = \theta/\alpha$ and $M_o(d) = 0$. For example t may be some apriori value, based on past experience or through other sources of knowledge, of θ. However, in practice, it would not be possible to obtain λ_{o2}.

3.4 A Class of Estimators Better than Sample Mean and Searls' Estimator

In the previous section, we have discussed general properties of the estimator d for any parameter θ and found that whatever be the choices of the auxiliary statistic t, we can always (except possibly in few situations) generate estimators better than $\hat{\theta}$, an unbiased one for θ and Searls' estimator $\hat{T}_S$, provided approximate good-guessed values of the parameters involved are available.

In this section and onwards, we shall confine ourselves to the estimation of population mean $\bar{Y}$ of a character y and find some estimators better than sample mean $\bar{y}$, modified Searls' and Searls' estimator in case of simple random sampling without replacement where finite population correction factor (f.p.c.) is to be ignored.

The estimator defined in (3.3.1) in this case reduces to

$$C_{\lambda,v}(\bar{Y}) = \{d : d = \lambda' v\} \tag{3.4.1}$$

where, $v' = (\bar{y}, t)$, $\lambda' = (\lambda_1, \lambda_2)$. Obviously the estimators defined in (2.1.2), (2.1.4), (3.1.1) and (3.1.7) may be identified as members of the class, $d' = \bar{y} + \lambda_2' t$ which itself is a subclass of the class defined in (3.4.1). Our main effort is to present estimators better (in the sense of having smaller MSE) than those in $T_1 = \lambda_1 \bar{y}$ (and hence better than $\bar{y}$, $\hat{T}_S = n\bar{y}/(n+C_y^2)$ and T_H in (3.1.1), in case of simple random sampling (where f.p.c. is ignored).

Let μ_r denote the r^{th} central moment of the character y and

$$\beta_1 = \mu_3^2/\mu_2^3, \quad \beta_2 = \mu_4/\mu_2^2, \quad \mu_2 = \sigma_y^2. \tag{3.4.2}$$

Depending upon the situation, we shall sometimes assume the knowledge of some of the quantities $\bar{Y}_{(1)}, \bar{Y}_{(2)}, C_{(1)}, C_{(2)}, \beta_1^{(1)}, \beta_2^{(2)}$ which are such that

$$\begin{aligned} &0 < \bar{Y}_{(1)} \le \bar{Y} \le \bar{Y}_{(2)} \ ; \quad 0 < C_{(1)} \le C_y \le C_{(2)} \\ &0 \le \beta_1^{(1)} \le \beta_1 \le \beta_1^{(2)} \ ; \quad \beta_2 \le \beta_2^{(2)} \end{aligned} \tag{3.4.3}$$

The estimators in the proposed class $C_{\lambda,v}(\bar{Y})$ in (3.4.1) are in general biased. Using the general results of the previous section, we have

$$B(d) = \lambda'\psi - \bar{Y}$$
$$\text{and} \qquad M(d) = \lambda' G\lambda - 2\bar{Y}\lambda'\psi + \bar{Y}^2 \tag{3.4.4}$$

respectively, where

$$G = \begin{pmatrix} E(\bar{y}^2) & E(\bar{y}t) \\ E(\bar{y}t) & E(t^2) \end{pmatrix}; \quad \psi = \begin{pmatrix} \bar{Y} \\ E(t) \end{pmatrix}.$$

The optimum choice λ_o of λ which minimises $M(d)$ is a solution of

$$G\lambda = \bar{Y}\psi \tag{3.4.5}$$

Further, we have

<u>Theorem 3.4.1</u>. The weight vector $\lambda(\lambda_1+\lambda_2\neq 1)$ which minimises $M(d)$ in (3.4.4) is given by

$$\lambda_o = \bar{Y}\,\bar{G}\,\psi \tag{3.4.6}$$

and the resulting bias and MSE of d are given by

$$B_o(d) = -M_o(d)/\bar{Y} \tag{3.4.7}$$

$$M_o(d) = \bar{Y}^2[1 - \psi'\,\bar{G}\,\psi] \tag{3.4.8}$$

where $\bar{G}$ is a g-inverse of the matrix G. It may also be noted alternatively that

$$M_o(d) = \bar{Y}^2[1 - \lambda_o'\,\psi/\bar{Y}]$$

In case G is a positive definite matrix, it is found that

$$\lambda_{o1} = [\bar{Y}\,E(\bar{y})\,E(t^2) - E(t).E(t\bar{y})]/D(\bar{y},t)$$
$$\lambda_{o2} = \bar{Y}[E(t)\,E(\bar{y}^2) - E(\bar{y}).E(t\bar{y})]/D(\bar{y},t) \tag{3.4.9}$$

$$\text{and} \qquad M_o(d) = \bar{Y}^2[1 - \frac{N(\bar{y},t)}{D(\bar{y},t)}] \tag{3.4.10}$$

here

$$D(\bar{y},t) = \bar{Y}^2(E(t))^2[(1-\rho^2_{\bar{y},t})C^2(\bar{y})C^2(t)+C^2(\bar{y})+C^2(t) - 2\rho_{\bar{y},t}C(\bar{y})C(t)]$$

$$N(\bar{y},t) = \bar{Y}^2(E(t))^2[C^2(\bar{y})-2\rho_{\bar{y},t}C(\bar{y})C(t)+C^2(t)]$$

$C(\bar{y})$ = Coefficient of variation of $\bar{y}$

$\rho_{\bar{y},t}$ = Correlation coefficient between $\bar{y}$ and t.

From (3.3.10), it may be noted that for $\theta = \bar{Y}$, and $\hat{\theta} = \bar{y}$,

$$M(d) = M(T_1)+\lambda_2^2Et^2-2\lambda_2\bar{Y}\ E(t)\{(1-\lambda_1)-\lambda_1\ \rho_{\bar{y},t}C(\bar{y})C(t)\} \quad (3.4.11)$$

It is found that $M(d)$ is an increasing or decreasing function of $\rho_{\bar{y},t}$ according as λ_1 and λ_2 are of different or same sign. It may be noted that for any fixed ρ, the $M_o(d)$ is an increasing function of $C(t)$. Thus, if t' and t'' are two choices of t, both having the same correlation with $\bar{y}$, then the use of t' in d would be preferable over than that of t'', iff $C(t') < C(t'')$.

In practice λ_o would not be known, as it depends upon a number of parameters. The following technique would help, in that case, to generate estimators from d better than $\bar{y}$, T_1 and $\hat{T}_S$.

From (3.3.11) or (3.4.11), for a specified λ_1, the estimator $d = T_1 + \lambda_2 t$ would be better than $T_1 = \lambda_1\ \bar{y}$

$$\text{iff} \quad \lambda_2 \text{ lies between } 0 \text{ and } 2\lambda^*_{o2} \quad (3.4.12)$$

where

$$\lambda^*_{o2} = [(1-\lambda_1)\bar{Y}\ E(t) - \lambda_1 Cov(\bar{y},t)]/E(t^2)$$

is the optimum choice of λ_2, for fixed λ_1, in d. Obviously, the resulting MSE of d in this case would be

$$M_o(d) = M(T_1) - \lambda^{*2}_{o2} E(t^2) \qquad (3.4.13)$$

Let $q = \bar{Y}/E(t) > 0$. From (3.4.12) we find that in both the cases (a) $\rho \leq 0$, $0 < \lambda_1 < 1$ and (b) $\rho \geq 0$, $\lambda_1 < 0$, a set of sufficient conditions for d to be better than T_1 would be

$$0 < \lambda_2 < 2q_{(1)} (1-\lambda_1)/[1+C^2_{(2)}(t)] \qquad (3.4.14)$$

$$0 < \lambda_2 < -2\lambda_1 q_{(1)} C_{(1)}(\bar{y}) C_{(1)}(t)\rho^*/[1+C^2_{(2)}(t)] \qquad (3.4.15)$$

where $q_{(1)}$, $C_{(1)}(t)$ etc., are known quantities such that

$$0 < q_{(1)} < q,\ 0 < C_{(1)}(t) < C(t) < C_{(2)}(t),\ 0 < C_{(1)}(\bar{y}) < C(\bar{y})$$

and $\rho^* = \rho_{(1)}$ or $\rho_{(-2)}$ such that $0 < \rho_{(1)} < \rho,\ \rho < \rho_{(-2)} < 0$.

Thus by using $\lambda_2 t$ in addition to $T_1 = \lambda_1 \bar{y}$ in the form of $d = T_1 + \lambda_2 t$, we can always generate (except when $\lambda_1 = \bar{Y}E(t)/E(\bar{y}t)$) the estimators from d better than T_1 even when exact values of the parameters involved in λ_o are not known and only appropriate bounds are known.

A subclass of d, which we get special attention in our discussion, is

$$d' = \bar{y} + \lambda_2' t \qquad (3.4.16)$$

with

$$M(d') = V(\bar{y}) + 2\lambda_2' \, Cov(\bar{y},t) + \lambda_2'^2 E(t^2). \qquad (3.4.17)$$

Obviously the estimators from d' would be unbiased for $\bar{Y}$ in case t is such that $E(t) = 0$ or $\lambda_2' = 0$. The optimum choice λ_{o2}' of λ_2' and the resulting MSE would be given by

$$\lambda_{o2}' = -\,Cov(\bar{y},t)/E(t^2) \qquad (3.4.18)$$

$$\text{and} \quad M_o(d') = V(\bar{y}) - [Cov(\bar{y},t)]^2/E(t^2) \qquad (3.4.19)$$

It follows from (3.4.12) or otherwise, that the estimators in d' would be better than $\bar{y}$ iff

$$\begin{aligned} &\text{either} \quad 0 < \lambda_2' < 2\lambda_{o2}' \quad \text{in case} \quad \rho_{\bar{y},t} < 0 \\ &\text{or} \quad 2\lambda_{o2}' < \lambda_2' < 0 \quad \text{in case} \quad \rho_{\bar{y},t} > 0. \end{aligned} \qquad (3.4.20)$$

In absence of the exact knowledge λ_{o2}', the conditions (3.4.20) or (3.4.12) with $\lambda_1 = 1$, may be used to generate estimators from d' better than $\bar{y}$.

Remarks

(i) It may be shown that a set of sufficient conditions for T_1 to be better than $\bar{y}$ would be

$$[n - C_{(1)}^2]/[n+C_{(1)}^2] < \lambda_1 < 1 \qquad (3.4.21)$$

which may be written as

$$n/[n+C_{(1)}^2] \leq \lambda_1 < 1.$$

The choice of λ_1 in d such that (3.4.12), (3.4.14) and (3.4.15) should be guided by this remark, and for some choices of λ_1, the class d would generate estimators better than $\bar{y}$ even if $\rho_{\bar{y},t} = 0$. Further if $\lambda_1 = n/(n+C_y^2)$ the class d under (3.4.12), (3.4.14) and (3.4.15) would generate estimators better than the Searls' estimator $\hat{T}_S$.

(ii) From remark (ii) of Section 3.3 or otherwise, it is found that the subclass $d' = \lambda_2' t$ of d, we should never choose t to be uncorrelated with $\bar{y}$ as for such choices of t, the sample mean $\bar{y}$ would be uniformly better than the corresponding estimators from d'.

(iii) Remark (iv) of Section 3.3 is also true in this case.

(iv) Using the expression $M_o(\hat{T}_S)$ in (3.2.4) for the Searls' estimator and from (3.4.10),

$$M_o(\hat{T}_S) - M_o(d) = \bar{Y}^2 C^2(\bar{y})(C(\bar{y}) - \rho_{\bar{y},t} C(t))^2/(1+C^2(\bar{y}))D(\bar{y},t) \tag{3.4.22}$$

and since $V(\bar{y}) = \bar{Y}^2 C^2_{\bar{y}}$, we have, then

$$V(\bar{y}) - M_o(d) = \bar{Y}^2 C^2(\bar{y})[(C(\bar{y}) - \rho_{\bar{y},t} C(t))^2 + C^2(t) C^2(\bar{y})(1-\rho^2_{\bar{y},t})]/D(\bar{y},t) \tag{3.4.23}$$

Thus from (3.4.22), we observe that if t is such that $\rho_{\bar{y},t} = C(\bar{y})/C(t)$, then $\hat{T}_S$ and optimum d will be equally efficient and hence in such a case there would not be any gain by using the component $\lambda_2 t$ in d.

3.5 Some Specific Estimators Better than Sample Mean and Searls' Estimator

In the previous section, we have discussed general properties of the estimator d and found that whatever be the choice of the auxiliary statistic t, we can always (except possibly in few situations) generate estimators better than the sample mean $\bar{y}$ and Searls' estimator $\hat{T}_S$ provided some approximate good-guessed values f the parameters involved are available. In this section, we shall dentify the estimators from d, better than $\bar{y}$ and $\hat{T}_S$, for some pecific choices of the statistic t.

For a specific choice of the statistic t and specified λ_1, we hall find, as in (3.4.12), that, the estimator d would be better than $\bar{y}$ or $\hat{T}_S$ iff λ_2 lies between 0 and $2\lambda_o^*$, say. Obviously, λ_o^* would be a function of some unknown population parameters, say ϕ and the vector ϕ can be decomposed into component vectors say, ϕ_1, ϕ_2, ϕ_3 such that $|\lambda_o^*|$ is non-decreasing in each component of ϕ_1 and non-increasing in each component of ϕ_2. If $\phi_1^*, \phi_2^*, \phi_3^*$ are known quantities such that $\phi_1 \geq \phi_1^*$, $\phi_2 \leq \phi_2^*$, $\phi_3 = \phi_3^*$ hold and moreover $Sgn[\lambda_o^*(\phi)]$ is known, then $\mu^* = Sgn[\lambda_o^*(\phi)]|\lambda_o^*(\phi^*)|$ is a known quantity and then it is obvious that, for a given λ_1, we shall have $M(d) \leq V(\bar{y})$ or $M(d) \leq M(\hat{T}_S)$ for all μ such that either $0 < \mu < 2\mu^*$ or $2\mu^* < \mu < 0$ holds.

In this section, we have used the above technique of shrinking the effective interval $(0, 2\lambda_o^*)$ or $(2\lambda_o^*, 0)$, to obtain some estimators from d, for some specific choices of the statistic t, which are better than $\bar{y}$ and $\hat{T}_S$.

The Table 3.5.1 gives the expressions for $E(t)$, $E(t^2)$ and $Cov(\bar{y}, t)$, upto order n^{-1}, in case of simple random sampling with replacement, for some choices of t.

Let $\lambda_1^* = n/[n+C_*^2]$, where $C_*^2 < C_y^2$ so that the estimator $T_1^* = \lambda_1^* \bar{y}$ is better than the sample mean $\bar{y}$ (we refer to Remark (i) of Section 3.4 in this context).

Table 3.5.1. Expressions for $E(t)$, $E(t^2)$ and $Cov(\bar{y},t)$ for different choices of t.

Choices of t	$E(t)$	$E(t^2)$	$Cov(\bar{y},t)$
(1)	(2)	(3)	(4)
α	α	α^2	0
$\bar{y}$	$\bar{Y}$	$\bar{Y}^2[n+C_y^2]/n$	$\bar{Y}^2 C_y^2/n$
s_y^2	$\bar{Y}^2 C_y^2$	$\bar{Y}^4 C_y^4 \Delta/n$	$\bar{Y}^3 C_y^3 a\sqrt{\beta_1}/n$
$s_y^2/\bar{y}^2$	$C_y^2[n+3C_y^2-2a\sqrt{\beta_1}C_y]/n$	$C_y^4[10C_y^2-8a\sqrt{\beta_1}C_y+\Delta]/n$	$\bar{Y} C_y^3(a\sqrt{\beta_1}-2C_y)/n$
$s_y^2\bar{y}$	$C_y^2 \bar{Y}^3[n+a\sqrt{\beta_1}C_y]/n$	$\bar{Y}^6 C_y^4[C_y^2+4a\sqrt{\beta_1}C_y+\Delta]/n$	$\bar{Y}^4 C_y^3(a\sqrt{\beta_1}+C_y)/n$
$\bar{y}^2$	$\bar{Y}^2[n+C_y^2]/n$	$\bar{Y}^4(n+6C_y^2)/n$	$2\bar{Y}^3C_y^2/n$
$s_y^2/\bar{y}$	$\bar{Y} C_y^2[n+C_y^2-a\sqrt{\beta_1}C_y]/n$	$\bar{Y}^2C_y^4[3C_y^2-4a\sqrt{\beta_1}C_y+\Delta]/n$	$\bar{Y}^2C_y^2(a\sqrt{\beta_1}-C_y)/n$
$1/\bar{y}$	$[n+C_y^2]/n\bar{Y}$	$[n+3C_y^2]/n\bar{Y}^2$	$-C_y^2/n$
$1/s_y^2$	$\Delta/n\bar{Y}^2C_y^2$	$[3\Delta-2n]/n\bar{Y}^4C_y^4$	$-a\sqrt{\beta_1}/n\bar{Y}C_y$

The notations "Δ" and "a" have been defined in (3.5.2)

$$d^* = T_1^* + \lambda_2 t \tag{3.5.1}$$

Throughout we shall consider that $\bar{Y} > 0$. The results in case < 0 may be obtained likewise.

We shall make frequent use of the notation (3.4.3) and

$$\begin{aligned} \Delta &= \beta_2 + (n^2-2n+3)/(n-1) \\ \Delta_{(2)} &= \beta_2^{(2)} + (n^2-2n+3)/(n-1) \end{aligned} \tag{3.5.2}$$

$$\alpha = \begin{cases} +1 & \text{for positively skewed distributions} \\ 0 & \text{for symmetrical distributions} \\ -1 & \text{for negatively skewed distributions} \end{cases}$$

using (3.4.11), (3.4.12) and the expressions in Table 3.5.1, it may be shown that

$$\begin{aligned} &(i) \quad d_1^* = T_1^* + \lambda_2 \alpha \\ &(ii) \quad d_2^* = T_1^* + \lambda_2 s_y^2 \\ &(iii) \quad d_3^* = T_1^* + \lambda_2 (s_y^2/\bar{y}^2) \\ &(iv) \quad d_4^* = T_1^* + \lambda_2 s_y^2 \bar{y} \\ &(v) \quad d_5^* = T_1^* + \lambda_2 \bar{y}^2 \\ &(vi) \quad d_6^* = T_1^* + \lambda_2 (s_y^2/\bar{y}) \end{aligned} \tag{3.5.3}$$

would be better than T_1^* (and hence $\bar{y}$ too) iff λ_2 lies between

$$0 \quad \text{and} \quad 2\bar{Y}C_*^2/(n + C_*^2)\alpha$$

$$0 \quad \text{and} \quad 2n(C_*^2 - \alpha\sqrt{\beta_1}C_y)/(n+C_*^2)\,\Delta\,\bar{Y}C_y^2$$

$$0 \quad \text{and} \quad \frac{2\bar{Y}[n(C_*^2+2C_y^2) + 3C_y^2C_*^2 - \alpha\sqrt{\beta_1}C_y(n+2C_*^2)]}{C_y^2\ \Delta^*(n+C_*^2)}$$

$$0 \quad \text{and} \quad \frac{2[n(C_*^2-C_y^2) + \alpha\sqrt{\beta_1}C_y(C_*^2-n)]}{(n+C_*^2)\ \bar{Y}^2C_y^2\ \ \Delta'} \qquad (3.5.4)$$

$$0 \quad \text{and} \quad 2[C_*^2(n+C_y^2) - 2nC_y^2][\bar{Y}(n+C_*^2)(n+6C_y^2)]^{-1}$$

$$\text{dnd } 0 \quad \text{and} \quad \frac{2[n(C_*^2+C_y^2) + C_y^2C_*^2 - \alpha\sqrt{\beta_1}C_y(C_*^2+n)]}{(n+C_*^2)\ C_y^2\ \Delta''}$$

respectively, where α is a constant and say, in particular, be an approximate value of $\bar{Y}$ and $\Delta' = \Delta + C_y^2 + 4\alpha\sqrt{\beta_1}C_y$, $\Delta'' = \Delta + 3C_y^2 - 4\alpha\sqrt{\beta_1}C_y$, $\Delta^* = \Delta + 10C_y^2 - 8\alpha\sqrt{\beta_1}C_y$.

We now generate a set of sufficient conditions, corresponding to the above ones, for d_1^* to d_6^* to be better than T_1^* and $\bar{y}$. The estimator d_1^* would be better than T_1^* (and $\bar{y}$)

$$\text{if} \quad 0 < \lambda_2 < 2\bar{Y}_{(1)}C_*^2/(n+C_*^2)\alpha, \text{ in case } \alpha \geq 0 \qquad (3.5.5)$$

$$\text{of} \quad \text{if} \quad 2\bar{Y}_{(1)}C_*^2/(n+C_*^2)\alpha < \lambda_2 < 0 \text{ in case } \alpha < 0.$$

In case of symmetrical distributions, the estimators d_2^*, d_3^*, d_4^* and d_6^* would be better than T_1^* (and $\bar{y}$) if

(a) $0 < \lambda_2 < 2nC_*^2/[(n+C_*^2)\ \bar{Y}_{(2)}C_{(2)}^2\ \ \Delta_{(2)}]$

(b) $$0 < \lambda_2 < \frac{2\bar{Y}_{(1)}[n(C_*^2+2C_{(1)}^2) + 3C_{(1)}^2C_*^2]}{C_{(2)}^2(n+C_*^2)(\Delta_{(2)} + 10C_{(2)}^2)}$$

(c) $0 < \lambda_2 < 2[n(C_*^2-C_{(2)}^2)][(n+C_*^2)\bar{Y}_{(2)}^2C_{(2)}^2(\Delta_{(2)}+C_{(2)}^2)]^{-1}$

in case $C_*^2 > C_{(2)}^2$ (3.5.6)

or $-2n(C_{(1)}^2-C_*^2)[(n+C_*^2)\bar{Y}_{(2)}^2C_{(2)}^2(\Delta_{(2)}+C_{(2)}^2)]^{-1} < \lambda_2 < 0$

in case $C_*^2 < C_{(1)}^2$

and (d) $$0 < \lambda_2 < \frac{2[n - (C_*^2 + C_{(2)}^2) + C_{(2)}^2C_*^2]}{(n + C_*^2) - C_{(2)}^2(\Delta_{(2)}+3C_{(2)}^2)}$$

respectively. The sufficient conditions in case of positively and negatively skewed distributions may be obtained likewise by using the appropriate bounds of involved parameters.

Further, the estimator d_5^* would be better than T_1^* (and $\bar{y}$)

$$\text{if} \quad 0 < \lambda_2 < \frac{2[C_*^2(n+C_{(2)}^2)-2nC_{(2)}^2]}{\bar{Y}_{(2)}(n+C_*^2)(n+6C_{(2)}^2)}$$

$$\text{or} \quad -\frac{2[2nC_{(1)}^2-(n+C_{(1)}^2)C_*^2]}{\bar{Y}_{(2)}(n+C_*^2)(n+6C_{(2)}^2)} < \lambda_2 < 0 \qquad (3.5.7)$$

according as $C_*^2 > 2nC_{(2)}^2/(n+C_{(2)}^2)$ or $C_*^2 < 2nC_{(1)}^2/(n+C_{(1)}^2)$ respectively.

It may be noted that for $C_* = 0$, the estimators T_1^* and d^* with λ_2 replaced by λ_2' would be reduced to $\bar{y}$ and $d' = \bar{y} + \lambda_2' t$ respectively and hence the necessary and sufficient conditions for the estimators

$$
\begin{aligned}
d_2' &= \bar{y} + \lambda_2' \, s_y^2 \\
d_3' &= \bar{y} + \lambda_2'(s_y^2/\bar{y}^2) \\
d_4' &= \bar{y} + \lambda_2' \, s_y^2\bar{y} \qquad (3.5.8) \\
d_5' &= \bar{y} + \lambda_2 \, \bar{y}^2 \\
\text{and} \quad d_6' &= \bar{y} + \lambda_2(s_y^2/\bar{y})
\end{aligned}
$$

to be better than $\bar{y}$ may be derived from those in (3.5.4) for corresponding choice of t in d^*, by substituting $C_* = 0$.

Thus the necessary and sufficient condition for d_5' to be better than $\bar{y}$ iff λ_2' lies between

$$0 \quad \text{and} \quad -4C_y^2/\bar{Y}(n+6C_y^2) \qquad (3.5.9)$$

and hence a sufficient condition for d_5' to be better than $\bar{y}$ would be

$$-4C_{(1)}^2/\bar{Y}_{(2)}(n+6C_{(1)}^2) < \lambda_2' < 0. \qquad (3.5.10)$$

For symmetrical distributions, the estimators d_3', d_4' and d_6' would be better than $\bar{y}$ iff λ_2 lies between

$$
\begin{aligned}
&0 \quad \text{and} \quad 4\bar{Y}/\Delta^* \\
&0 \quad \text{and} \quad -2/\bar{Y}^2\Delta' \qquad (3.5.11) \\
&0 \quad \text{and} \quad 2/\Delta''
\end{aligned}
$$

d hence the sufficient conditions for the estimators d_3', d_4' and $_6$ to be better than $\bar{y}$ may be found as

$$0 < \lambda_2' < 4\bar{Y}_{(1)}/[\Delta_{(2)} + 10C^2_{(2)}]$$

$$-2/\bar{Y}^2_{(2)}[\Delta_{(2)} + C^2_{(2)}] < \lambda_2' < 0 \qquad (3.5.12)$$

and

$$0 < \lambda_2' < 2/[\Delta_{(2)} + 3C^2_{(2)}]$$

espectively.

t is noted here and may be observed otherwise also that we can get o estimator from d_2' better than $\bar{y}$ in case of symmetrical distributions.

The sufficient conditions for the estimators d_3', d_4' and d_6' to be better than $\bar{y}$ for positively and negatively skewed distributions may be obtained likewise by using the appropriate bounds for involved parameters.

Further we note that in case C_y is known exactly, we may use it in place of C_* in d^*. The necessary and sufficient conditions for the estimators $d'' = \hat{T}_S + \lambda_2 t$ to be better than Searls' estimator $\hat{T}_S$ (and hence $\bar{y}$ too) may be obtained likewise from those in (3.5.4) to (3.5.7), for corresponding choice of t in d^*, by substituting C_y^2 in place C_*, $C_{(1)}$ and $C_{(2)}$.

In case $t = 1/\bar{y}$ and $t = 1/s_y^2$, we may again generate estimators from $d_7' = \bar{y} + \lambda_2'/\bar{y}$, $d_8' = \bar{y} + \lambda_2'/s_y^2$ better than $\bar{y}$ and from $d_7^* = T_1^* + \lambda_2/\bar{y}$ and $d_8^* = T_1^* + \lambda_2/s_y^2$ better than T_1^* (and hence $\bar{y}$ too). However, it may be shown that the optimum estimators in d_7' and d_7^* would be dominated by the optimum estimators in

$d_1' = \bar{y}(1+\lambda_2')$ i.e., by Searls' estimator. Similarly, the optimum estimators in d_8' and d_8^* would be dominated by the optimum estimators in $d_2' = \bar{y} + \lambda_2' s_y^2$ and $d_2^* = T_1^* + \lambda_2 s_y^2$ respectively and hence the use of $t = 1/\bar{y}$ and $t = 1/s_y^2$ could be ruled out in favour of $t = \bar{y}$ and $t = s_y^2$ respectively.

3.6 Some Concluding Discussion

As mentioned earlier, the Searls' estimator $\hat{T}_S$ is the best in the class T_1 to which the sample mean $\bar{y}$ also belongs. Further $\hat{T}_S$ is better than T_H. From (3.4.12) to (3.4.15) and the discussion in Section 3.5 we note that all the estimators in T_1, including $\bar{y}$, $\hat{T}_S$, and T_H, may further be improved merely by having some approximate apriori knowledge about the parameters involved in λ_{o2}^* and λ_{o2}' for various choices of t in $d = \lambda_1\bar{y} + \lambda_2 t$.

It may be shown that, (for a specified statistic t) the estimators from $d' = \bar{y} + \lambda_2' t$ would be better than $d_o' = \bar{y} + t$

$$\text{iff} \quad \lambda_2' \text{ lies between } 1 \text{ and } 2\lambda_{o2}' - 1 \qquad (3.6.1)$$

where $\lambda_{o2}' = -\operatorname{Cov}(\bar{y},t)/(Et^2)$. In particular, $T_{PS(\alpha)}$ would be better than T_{PS}

$$\text{iff} \quad \alpha \text{ lies between } -1 \text{ and } (2\lambda_{o2}'/n) + 1$$

where $\lambda_{o2}' = (C_y - \alpha\sqrt{\beta_1})/C_y \Delta''$, for large samples.

A set of sufficient conditions for $T_{PS(\alpha)}$ to be better than T_{PS} would be :

$$-1 < \alpha \leq 1 \quad \text{in case} \quad \lambda'_{o2} > 0$$

$$-1 < \alpha \leq 0 \quad \text{in case} \quad -n/2 < \lambda'_{o2} < 0$$

$$\frac{2\lambda_{o2}^{(-2)}}{n} + 1 < \alpha < -1 \quad \text{in case} \quad \lambda'_{o2} < -n$$

where $\lambda_{o2}^{(-2)}$ is a known quantity such that $\lambda'_{o2} \leq \lambda_{o2}^{(-2)} < -n$. Thus for large samples, in case of symmetric or negatively skewed distributions, $T_{PS(\alpha)}$ with $-1 < \alpha \leq 1$ will be better than T_{PS}.

In general, it may be shown, for a fixed λ_2 and for a specified statistic t, that the estimators from $d = \lambda_1\bar{y} + \lambda_2 t$ would be better than those in $d' = \bar{y} + \lambda_2 t$

iff λ_1 lies between $[(n-C_y^2) - 2\lambda_2(n/\bar{Y}^2)E(\bar{y}t)]/(n+C_y^2)$ and 1.

In particular $d_o = \lambda_1\bar{y} + (\alpha/n)\ s_y^2/\bar{y}$ would be better than $T_{PS(\alpha)}$

iff λ_1 lies between $[n-(1+2\alpha)C_y^2]/[n+C_y^2]$ and 1.

A set of sufficient conditions for d_o to be better than $T_{PS(\alpha)}$ would be that λ_1 lies between

$$[n - (1+2\alpha)C_{(1)}^2]/[n+C_{(1)}^2] \quad \text{and} \quad 1$$

which is possible if $\alpha \neq -1$.

The estimators $\tilde{T}_S$, $\tilde{T}_{S(\alpha)}$, and T should not be used in case $\sqrt{n}/C_y$ is large (i.e., if samples are large and/or C_y is small) as $\bar{y}$ would be better than them in such cases. However, as shown by Thompson (1968), use of $\tilde{T}_S$ and $\tilde{T}_{S(\alpha)}$ may be preferred over $\bar{y}$ if $\sqrt{n}/C_y$ is small or moderate. Further, as shown by Mehta and Srinivasan (1971), T would be preferable over $\bar{y}$ as well as $\tilde{T}_S$ and $\tilde{T}_{S(\alpha)}$

in case $\sqrt{n}/C_y$ is small or moderate. Further it has been observed by Thompson (1968) that in case of normal distribution $\hat{T}_S$ fares better than $\tilde{T}_S$ and $\tilde{T}_{S(\alpha)}$ (except for some choices of α) even when $\sqrt{n}/C_y$ is small or moderate.

We have already discussed in Section 3.5 that we may obtain a large number of estimators better than not only the sample mean but also than the modified Searls' estimator $\hat{T}_S'$ and the Searls' estimator $\hat{T}_S$ which are themselves better than the sample mean.

However it may be noted that none of the estimators in $d' = \bar{y} + \lambda_2 s_y^2$ would be better than $\bar{y}$ in case of the symmetrical distributions. Further none of the estimators in $d^* = T_1^* + \lambda_2\bar{y}$ or $d_o'' = \hat{T}_S + \lambda_2\bar{y}$ would be better than $\hat{T}_S$ and same is the case about d_5^* and $d_5'' = \hat{T}_S + \lambda_2 s_y^2\bar{y}$ in case of the symmetrical distributions.

The discussion in Section 3.5, obviously, helps in generating numerous estimators (biased) better than the uniformly minimum variance unbiased estimators and maximum likelihood estimators for the population mean in case of various distributions, including the normal populations.

3.7 Empirical Study

In this section the performance of various estimators under consideration have been studied on the basis of data for two natural population .

Population I :

Data under consideration has been taken from 1961 census, West Bengal, District Census Hand Book, Midnapore, (Census of India, 1961).

The population consists of 353 villages under Panskura Police Station.

The character y denotes the village population. For this population of 353 villages, we obtained

$$\bar{Y} = 670.076, \quad \sigma_y^2 = 412,624.880, \quad C_y^2 = 0.918$$

$$\mu_3(y) > 0, \quad \beta_1 = 7.789, \quad \beta_2 = 14.541.$$

We pick up the apriori values of the above parameters, as follows :

$$\bar{Y}_{(1)} = 650, \quad \bar{Y}_{(2)} = 700, \quad C_{(1)}^2 = 0.85, \quad C_{(2)}^2 = 1.5$$

$$\sqrt{\beta_1}^{(1)} = 2.0, \quad \sqrt{\beta_1}^{(2)} = 3.0, \quad \beta_2^{(1)} = 10, \quad \beta_2^{(2)} = 16$$

and a sample of size $n = 30$ has been considered.

The Table 3.7.1 gives the relative bias and percentage relative efficiencies of the estimators from the class $d' = \bar{y} + \lambda_2' t$ over $\bar{y}$ and relative bias, percentage relative efficiencies of the estimators from the class $d^* = \hat{T}_S + \lambda_2^* t$ over $\hat{T}_S$ are given in Table 3.7.2. λ_{o2}' in the Table 3.7.1 denotes the optimum choice of λ_2' (for a specified t) and λ_{o2}'' denotes the located value of λ_{o2}' using the above mentioned apriori values of the involved parameters. Similarly, λ_{o2}^* and λ_{o2}^{**} in the Table 3.7.2 denote the optimum choice of λ_2^* (for a specified t) and the located value of λ_{o2}^* respectively.

Table 3.7.1 Relative bias and percent relative efficiency of d' over $\bar{y}$ for various choices of t and some values of λ_2'.

t	λ_2'		Relative bias	Percent relative efficiency of d' over $\bar{y}$ $[V(\bar{y})/M(d')] \times 100$	
	$\lambda_2' = \lambda_{o2}'$	$\lambda_2' = \lambda_{o2}''$	$\lvert B(d')/\bar{Y}\rvert$	$\lambda_2' = \lambda_{o2}'$	$\lambda_{o2}' = \lambda_{o2}''$
(1)	(2)	(3)	(4)	(5)	(6)
		-0.0550	0.055		102.8292
		-0.0400	0.040		102.6871
		-0.0300	0.030		103.0629
$\bar{y}$	-0.0297		0.0297	103.0632	
		-0.0100	0.0100		101.6916
		-0.0090	0.0090		101.5512
		-0.0050	0.0050		100.3168
		-0.000130	0.0801		119.3344
		-0.000120	0.0738		120.6590
		-0.000100	0.0616		121.7448
s_y^2	-0.000099		0.0610	121.7455	
		-0.000080	0.0493		120.7268
		-0.000060	0.0369		117.6964
		-0.000050	0.0308		115.5127
		-0.0000001585	0.0712		134.1634
		-0.0000001580	0.0710		134.1635
$s_y^2\bar{y}$	-0.0000001570		0.0705	134.1640	
		-0.0000001565	0.0703		134.1601
		-0.0000001560	0.0701		134.1592
		-0.0000001555	0.0698		134.1584
		-0.0000001549	0.0714		134.1508
		-0.000138	0.0953		104.4951
		-0.000130	0.0898		106.1902
		-0.000110	0.0759		109.5405
$\bar{y}^2$	-0.000078		0.0539	111.6033	
		-0.000070	0.0483		111.5444
		-0.000060	0.0414		110.9892
		-0.000050	0.3453		110.0224

Table 3.7.2. Relative bias and percent relative efficiency of d^* over $\hat{T}_S$ for various choices of t and some values of λ_2^*.

t	λ_2^*		Relative bias	Percent relative efficiency of d^* over $\hat{T}_S$ $[M(\hat{T}_S)/M(d^*)] \times 100$	
	$\lambda_2^* = \lambda_{o2}^*$	$\lambda_2^* = \lambda_{o2}^{**}$	$\lvert B(d^*)/\bar{Y}\rvert$	$\lambda_2^* = \lambda_{o2}^*$	$\lambda_2^* = \lambda_{o2}^{**}$
(1)	(2)	(3)	(4)	(5)	(6)
$\alpha=\bar{Y}_o$	0.034		.0075		103.0624
=600		0.033	.0001		103.0570
		-0.000066	0.0703		108.0717
		-0.000065	0.0697		108.0815
		-0.000064	0.0691		108.0832
s_y^2	-0.000063		0.0685	108.0859	
		-0.000060	0.0666		108.0601
		-0.000050	0.0605		107.6933
		-0.000040	0.0543		106.9032
		-0.000000152	0.0304		112.4419
		-0.000000142	0.0303		113.4253
		-0.000000132	0.0302		114.1425
		-0.000000122	0.0302		114.5993
$s_y^2\bar{y}$	-0.000000110		0.0301	114.7825	
		-0.000000100	0.0301		114.6378
		-0.000000090	0.0301		114.2097
		-0.000070	0.0780		100.4328
		-0.000050	0.0642		102.1227
$\bar{y}^2$		-0.000040	0.0573		102.4287
	-0.000037		0.0552	102.4484	
		-0.000017	0.0414		101.7295
		-0.000003	0.0317		100.3758

Population II.

The sample of 191 males in Tharu group, which along with many other groups, was collected by Mahalanobis, Rao and Majumder for anthropometric survey in U.P., 1941. (Sankhyā, Volume 9, Parts 2 and 3) has been considered as the second population in this section for our study.

Different parameters relating to the characteristic stature are as follows :

$$\bar{Y} = 163.30, \ \sigma_y = 520.77, \ C_y = 3.17$$
$$\sqrt{\beta_1} = -0.0141, \ \beta_2 = 2.49$$

We pick up the apriori values of the above parameters as follows :

$$C_{(1)} = 2, \ C_{(2)} = 4; \ \sqrt{\beta_1^{(1)}} = -0.05, \ \sqrt{\beta_1^{(2)}} = 0$$
$$\beta_2^{(1)} = 2, \ \beta_2^{(2)} = 4; \ \bar{Y}_{(1)} = 150, \ \bar{Y}_{(2)} = 180.$$

Let $n = 30$.

The Table 3.7.3 gives the relative bias and percentage relative efficiencies of the estimators from the class $d' = \bar{y} + \lambda_2' t$ over $\bar{y}$ and relative bias, percentage relative efficiencies of the estimators from the class $d^* = \hat{T}_S + \lambda_2^* t$ over $\hat{T}_S$ are given in the Table 3.7.4. λ_{o2}' in the Table 3.7.3 denotes the optimum choice of λ_2' (for a specified t) and λ_{o2}'' denotes the located value of λ_{o2}' using the above mentioned apriori values of the involved parameters. Similarly, λ_{o2}^* and λ_{o2}^{**} in the Table 3.7.4 denote the optimum choice of λ_2^* (for a specified t) and the located value of λ_{o2}^* respectively.

Table 3.7.3. Relative bias and percent relative efficiency of d' over $\bar{y}$ for various choices of t and some values of λ_2'.

t	λ_2'		Relative bias	Percent relative efficiency of d' over $\bar{y}$ $[V(\bar{y})/M(d')] \times 100$	
	$\lambda_2' = \lambda_{02}'$	$\lambda_2' = \lambda_{02}''$	$\lvert B(d')/\bar{Y} \rvert$	$\lambda_2' = \lambda_{02}'$	$\lambda_2' = \lambda_{02}''$
(1)	(2)	(3)	(4)	(5)	(6)
		-0.24	0.2400		133.680
		-0.25	0.2500		133.750
$\bar{y}$	-0.2533		0.2533	133.920	
		-0.26	0.2600		133.880
		-0.27	0.2700		133.720
		0.00000084	0.0013		100.001
s_y^2	0.00000083		0.0013	100.001	
		0.00000082	0.0013		100.001
		0.00000078	0.0013		100.001
		0.00000076	0.0013		100.001
$s_y^2/\bar{y}^2$	2.4485		0.3073	144.058	
		2.40	0.3013		144.033
$s_y^2\bar{y}$	-0.000000898		0.2438	132.051	
		-0.000000600	0.1625		127.541
$\bar{y}2$		-0.0016	0.3499		177.884
		-0.0015	0.3280		180.591
	-0.0013687		0.2993	181.791	
		-0.0012	0.2624		179.503
		-0.0010	0.2187		171.471
$s_y^2/\bar{y}$		0.0100	0.1361		116.301
	0.0160		0.2178	119.536	

Table 3.7.4. Relative bias and percent relative efficiency of d^* over $\hat{T}_S$ for various choices of t and some values of λ_2^*.

	λ_2^*		Relative bias	Percent relative efficiency of d^* over $\hat{T}_S$ $[M(\hat{T}_S)/M(d^*)] \times 100$	
	$\lambda_2^* = \lambda_{02}^*$	$\lambda_2^* = \lambda_{02}^{**}$	$\lvert B(d^*)/\bar{Y}\rvert$	$\lambda_2^* = \lambda_{02}^*$	$\lambda_2^* = \lambda_{02}^{**}$
(1)	(2)	(3)	(4)	(5)	(6)
		0.4746	0.2117		108.2566
		0.3012	0.0418		132.6884
$t=\bar{Y}_o$ $=160$	0.2374		0.0207	133.6144	
		0.2350	0.0230		133.1619
		0.2210	0.0367		132.9672
		0.00025	0.1621		133.3425
		0.00020	0.0791		126.4012
s_y^2	0.0001455		0.0115	132.0829	
		0.00010	0.0871		128.0618
		0.00005	0.1702		116.0412
		4.66	0.3317		181.2787
		4.00	0.2488		229.9994
$s_y^2/\bar{y}^2$	3.641		0.2038	252.3884	
		2.00	0.0022		225.1607
		1.00	0.1277		152.9872
$\bar{y}^2$		-0.00061	0.3867		101.2898
	-0.00033		0.3255	103.7792	
		0.043	0.3320		192.8655
$s_y^2/\bar{y}$	0.04205		0.3192	290.2652	
		0.032	0.1824		201.1012
		0.011	0.1035		172.8149

CHAPTER IV

SOME T_2-CLASS OF ESTIMATORS BETTER THAN HORVITZ-THOMPSON ESTIMATOR

4.1 Introduction and Summary

Let $U = \{1,2,...,N\}$ be a finite population of N (given) units and y be a variate under study which takes value y_i for the i^{th} unit of the population. The T_2-class of linear estimators [Horvitz and Thompson, (1952); Koop, (1963)] for population total $Y = \sum_{i=1}^{N} y_i = N\bar{Y}$ based on any sampling design is defined by

$$T_2 = \sum_{i \in s} \beta_i \, y_i \qquad (4.1.1)$$

where sum is over the distinct units in the sample, s and β_i $(i = 1,2,...,N)$ is the weight attached with a specified unit i of the population.

It is well-known [Horvitz and Thompson, 1952] that the Horvitz-Thompson estimator

$$\hat{Y}_{H-T} = \sum_{i \in s} y_i/\pi_i \qquad (4.1.2)$$

with

$$V(\hat{Y}_{H-T}) = \sum_{i=1}^{N} y_i^2(1-\pi_i)/\pi_i + \sum_{i \neq j=1}^{N}\sum \frac{y_i y_j \, (\pi_{ij} - \pi_i\pi_j)}{\pi_i\pi_j} \qquad (4.1.3)$$

is the best in the unbiased subclass of T_2-class of linear estimators, whatever be the sampling design. Further $\hat{Y}_{H-T}$ is also admissible in the unbiased subclass of Godambe's class of linear estimators $T_5 = \sum_{i \in s} \beta(i,s)\, y_i$ as proved by Godambe (1960) and Roy and Chakravorty (1960). However the Horvitz-Thompson estimator, $\hat{Y}_{H-T}$ is the only unbiased estimator in T_2, while there are a large number of biased estimators in T_2.

The question arises : Does there exist a best (UMMSE) estimator in the entire T_2-class? Further are their some biased estimators in T_2 which are better than $\hat{Y}_{H-T}$ whatever be the sampling design?

In this chapter, we have attempted to find answers to these questions. We find that there does not exist UMMSE-estimator in T_2, whatever be the sampling design, not even in the case of SRSWOR, contrary to the result about T_1-class of estimators, discussed in the Chapter II, where the Searls' estimator was found to be UMMSE in T_1 in case of SRSWOR. We also find that for a family of sampling designs, a large number of biased estimators of the type

$$T_2' = \lambda \cdot \hat{Y}_{H-T} \tag{4.1.4}$$

from T_2 are better than $\hat{Y}_{H-T}$, provided a quantity $C^2_{(1)} \leq C^2_y$ is known. Further, we identify the situations, under the usual super-population model δ_1 (Godambe, 1955), in which the estimators T_2' and $T_2^* = \lambda^* \sum_{i \in s} y_i/p_i$, $p_i = x_i / \sum_{i=1}^{N} x_i$, x being an auxiliary character, are δ_1-better than $\hat{Y}_{H-T}$. Numerical examples are given to illustrate our points and it is found that the relative efficiency of T_2' over $\hat{Y}_{H-T}$ under SRSWOR and π-P.S. scheme, is an increasing function of the conditional coefficient of variation of y given under the model (1.5.7). We have also discussed about the non-egativity of the unbiased estimators of $M(T_2')$ and $M(T_2^*)$.

.2 Non-existence of UMMSE-estimator in T_2-class

The MSE due to T_2 defined in (4.1.1) is given by

$$M(T_2) = Y^2[\beta' A\beta - 2\beta' d + 1] \tag{4.2.1}$$

here

$$\beta = (\beta_1, \beta_2, \ldots, \beta_N)',\ A = (a_{ij})_{N\times N},\ a_{ij} = (y_i y_j/Y^2)\pi_{ij}$$

$$d = (d_1, d_2, \ldots, d_N)',\ d_i = (y_i/Y)\pi_i,\ i,j = 1,2,\ldots,N$$

and π_{ij} for $j = i$ is interpreted as π_i.

The optimum choice β_o of β which minimises $M(T_2)$ is a solution of

$$A\beta_o = d \tag{4.2.2}$$

and optimum (the resulting) MSE is found to be

$$M_o(T_2) = Y^2[1 - d' A^{-1} d], \tag{4.2.3}$$

the absolute bias being found as

$$|B(T_2)| = Y|\beta' d - 1|. \tag{4.2.4}$$

It is found that

$$|A| = \prod_{k=1}^{N} (Y_k/Y)^2 \Delta$$

where,

$$\Delta = \begin{vmatrix} \pi_1 & \pi_{12} & \cdots & \pi_{1N} \\ \pi_{21} & \pi_2 & \cdots & \pi_{2N} \\ \cdots & \cdots & \cdots & \cdots \\ \pi_{N1} & \pi_{N2} & & \pi_N \end{vmatrix} \tag{4.2.5}$$

It may be shown that A^{-1} would exist provided $y_i > 0$ for all $i = 1,2,\ldots,N$ and $\Delta > 0$.

Let

Δ_{ii} be the reduced determinant obtained by eliminating i^{th} row and i^{th} column from Δ

and Δ_{ij} be the one obtained by eliminating i^{th} row and j^{th} column from Δ.

Let

$$\Delta'_{ij} = (-1)^{i+j}\,\Delta_{ij},\quad \tilde{\Delta}_{ii} = \Delta_{ii}/\Delta$$

and $\hat{\Delta}_{ij} = \Delta'_{ij}/\Delta$.

Next we state and prove the following

Theorem 4.2.1. For any sampling design satisfying (4.2.5), i.e., $\Delta > 0$, there does not exist the UMMSE-estimator for Y in T_2.

Proof. From (4.2.2), it follows that

$$\beta_o = A^{-1}d \tag{4.2.6}$$

where,

$$A^{-1} = (C_{ij});\ C_{ii} = (Y/y_i)^2\,\Delta_{ii}/\Delta \text{ and } C_{ij}=(Y/y_i)(Y/y_j)\Delta'_{ij}/\Delta$$

and thus

$$\begin{aligned}\beta_{oi} &= (Y/y_i)[\hat{\Delta}_{i1}\,\pi_1 + \hat{\Delta}_{i2}\pi_2 + \ldots + \tilde{\Delta}_{ii}\pi_i + \ldots + \hat{\Delta}_{iN}\,\pi_N \\ &= (Y/y_i).\ f(\pi_i,\ \pi_{ij}).\end{aligned} \tag{4.2.7}$$

Obviously, the optimum choice β_{oi} in (4.1.1) reduces T_2 to

$$T_2' = Y \sum_{i \epsilon s} f(\pi_i, \pi_{ij}) \qquad (4.2.8)$$

We observe from (4.2.7) that there do not exist best choices of β_i's independent of y-values in the population.

Remark

Let $\nu(s)$ be the number of distinct units in a sample s (effective sample size) and $\nu = E\nu(s) = \sum_{s \epsilon S} \nu(s) P(s)$ be the average effective sample size. Then it may be shown from (4.2.7) or otherwise that a particular solution of $\Lambda \beta_o = d$ is

$$\beta_{oi} = (Y/\nu y_i), \quad i = 1, 2, \ldots, N$$

and in particular, in case of SRSWOR, optimum choice β_{oi} of β would be

$$\beta_{oi} = (\dot{Y}/n y_i)$$

and the optimum estimator will be reduced to the parameter Y itself.

4.3 A Class of Biased Estimators Better than H-T Estimation

Since, in general, the best estimator (UMMSEE) in T_2 class, does not exist at all, we look for some other estimators which may be better than $\hat{Y}_{H-T}$.

Getting motivation from remark above, let the weights β_i in T_2 be chosen such that $\beta_i = \lambda/\pi_i$, where λ is a constant. Then T_2 reduces to

$$T_2' = \lambda \sum_{i \epsilon s} y_i/\pi_i \qquad (4.3.1)$$

Let

$$D = \sum_{i=1}^{N} \sum_{j=1}^{N} (y_i y_j / Y^2)(\pi_{ij}/\pi_i \pi_j) \qquad (4.3.2)$$

$$a_1' = \min_{1 \le i \le N} \{1/\pi_i\}$$

and

$$a_2' = \min_{1 \le i \ne j \le N} \{\pi_{ij}/\pi_i \pi_j\}.$$

Noting that,

$$\pi_{ij}/\pi_i \pi_j \le \min(\pi_i, \pi_j)/\pi_i \pi_j = \frac{1}{\pi_i} \text{ or } \frac{1}{\pi_j} \quad \text{for all } i,j$$

we have that

$$a_2' < a_1'.$$

Further, it may be noted that,

$$D = [V(\hat{Y}_{H-T})/Y^2] > 1. \qquad (4.3.3)$$

Let $D_{(1)}$ be any quantity such that $D_{(1)} \le D$. Such a $D_{(1)}$ may be determined by observing that

$$D > [a_1' \sum_{i=1}^{N} y_i^2 + a_2' \sum_{i \ne j=1}^{N} \sum y_i \, y_j]/Y^2$$

$$= [(a_1' - a_2')N(\sigma_y^2 + \bar{Y}^2) + a_2' \, Y^2]/Y^2$$

$$= a_2' + (a_1' - a_2')(1 + C_{(1)}^2)/N \qquad (4.3.4)$$

with

$$C_{(1)}^2 > N(a_1' - a_2')^{-1} (1 - a_2') - 1$$

where $C_{(1)}^2$ is any known quantity such that $C_{(1)}^2 \le C_y^2$. It may be shown that $D_{(1)} > 1$.

Next we prove the following

<u>Theorem 4.3.1</u>. The estimators in T_2' with λ satisfying $2/D_{(1)} - 1 \leq \lambda < 1$ will always be better than $\hat{Y}_{H-T}$ for all sampling designs in which $\pi_i > 0$ and for all populations satisfying (4.3.2) or (4.3.3).

<u>Proof</u>. From (4.1.2) and (4.2.1), it may be shown that

$$M(T_2) \leq V(\hat{Y}_{H-T})$$

iff

$$\sum_{i=1}^{N} \sum_{j=1}^{N} (y_i/Y)(y_j/Y)\, \pi_{ij}(\beta_i \beta_j - \frac{1}{\pi_i \pi_j}) \leq 2 \sum_{i=1}^{N} (y_i/Y)(\beta_i - \frac{1}{\pi_i})\pi_i \qquad (4.3.5)$$

Hence from (4.3.1) and (4.2.1),

$$M(T_2') \leq V(\hat{Y}_{H-T})$$

$$\text{iff} \quad \lambda^2 D - 2\lambda - (D-2)] \leq 0 \qquad (4.3.6)$$

$$\text{i.e., iff} \quad (1-\lambda)[2 - (1+\lambda)D] \leq 0 \qquad (4.3.7)$$

Noting that $D > 1$, (4.3.7) can never be satisfied for $\lambda \geq 1$. Further observing that

$$D > D_{(1)} > 1,$$

a sufficient condition that (4.3.7) is satisfied would be

$$[2/D_{(1)} - 1] \leq \lambda < 1. \qquad (4.3.8)$$

Remarks

(i) From (4.3.6), it is noted that T_2' will be better than $\hat{Y}_{H-T}$ iff λ lies between 2/D-1 and 1, the best choice of λ being $\lambda_o = 1/D$. In that case relative efficiency of T_2' over $\hat{Y}_{H-T}$ would be

$$V(\hat{Y}_{H-T})/M_o(T_2') = C^2(\hat{Y}_{H-T})/[\lambda_o^2 C^2(\hat{Y}_{H-T}) + (1-\lambda_o)^2]$$

$$= (D-1)/(\lambda_o^2 D - 2\lambda_o + 1)$$

$$= D.$$

Thus upper bound for the relative efficiency of the modified H-T estimator T_2' over H-T estimator $\hat{Y}_{H-T}$ would be D.

Further none of the estimators in T_2' would be better than $\hat{Y}_{H-T}$ if $D = 1$.

(ii) The optimum choice $\lambda_o = 1/D$ indicates that we should choose λ in T_2' such that $0 < \lambda < 1$. It may be noted that T_2' with $0 < \lambda < 1$ will always be better than $\hat{Y}_{H-T}$ for all populations and designs for which $D \geq 2$.

(iii) It may be shown that in case of SRSWOR, the best choice of λ_o in T_2' would be $\lambda_o = 1/[1+KC_y^2]$. Thus in this case if C_y^2 is known exactly, the estimator due to Searls (1964) would be the best in T_2' — a result similar to that, we found in Chapter II, while discussing about the T_1-class.

Numerical Illustration.

Let us consider the following π-P-S design constructed by G. Ramachandran (unpublished Ph.D. Thesis, 1978; pp 147).

S	P_s
{1,3}	$1/17^2$
{2,4}	$254/17^2$
{1,2,4}	$17/17^2$
{2,3,4}	$17/17^2$

Now considering the following hypothetical population of 4 units

$$y_1 = 4, \quad y_2 = 2, \quad y_3 = 3 \quad \text{and} \quad y_4 = 1$$

we have D in (4.3.2) as

$$D = 4.1142$$

thus

$$V(\hat{Y}_{H-T})/M(T_2') = (D-1)/[\lambda^2 D-2\lambda+1]$$

The following Table 4.3.1 gives the percent relative efficiency R.E. = $[V(\hat{Y}_{H-T})/M(T_2')] \times 100$ of T_2' over $\hat{Y}_{H-T}$ and absolute relative bias R.B. = $|B(T_2')/Y|$ for some values of λ.

Table 4.3.1. Percent relative efficiency of T_2' over $\hat{Y}_{H-T}$ and absolute relative bias.

Relative Efficiency(%) & Relative Bias	λ								
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
R.E.	370.23	407.32	404.29	362.84	302.78	243.13	192.72	153.18	122.96
R.B.	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1

4 Relative Performance of Some Biased Estimators from T_2-class compared to H-T estimator under a super-population model

Let x be an auxiliary character (which may be some suitably defined real valued function of some other variate, say z), closely related to y, the character under study.

Let

$$T_2^* = \lambda^* \sum_{i \in s} y_i/p_i, \quad p_i = x_i/X, \ X = \sum_{i=1}^{N} x_i \qquad (4.4.1)$$

We shall study the performance of T_2^* in (4.4.1) and of T_2' in (4.3.1), for fixed sample size designs, compared to $\hat{Y}_{H-T}$ under the class of prior distributions ∂_1 due to Godambe (1955). Let $y_1, y_2, \ldots, y_N$ be a random sample from an infinite super-population, specified by

$$E_{\partial_1}(y_i|x_i) = \alpha x_i$$

$$V_{\partial_1}(y_i|x_i) = \sigma^2 x_i^2 \qquad (4.4.2)$$

$$\text{and} \quad Cov_{\partial_1}(y_i, y_j|x_i, x_j) = 0$$

where α and σ^2 are the two parameters of the distribution ∂_1. The ratio $\sigma^2/\alpha^2 = C^2$, say, can be thought of as the square of conditional coefficient of variation of y given x. In practice, σ^2/α^2 may not be known exactly, but it may be known to lie in some interval. Let C^2 be such that it lies in the interval $(C^2_{(1)}, C^2_{(2)})$. Hence, we shall sometimes use in this section the quantities $C^2_{(1)}$ and $C^2_{(2)}$ such that

$$C^2_{(1)} \le C^2 \le C^2_{(2)} \qquad (4.4.3)$$

Let $t = V(\hat{X}_{H-T})/X^2$, $V(\hat{X}_{H-T})$ being given by (4.1.3) with y_i's being replaced by x_i's.

Definition 4.4.1 Let $E_{\partial_1} M(d)$ denote the expected MSE of d with respect to ∂_1. An estimator d_1 is said to be ∂_1-better than d_2

if $E_{\partial_1} M(d_1) \leq E_{\partial_1} M(d_2)$ for all $\theta = (y_1, y_2, \ldots, y_N)$

with strict inequality for at least one θ.

Lemma 4.4.1 For designs of fixed sample size, we have

(a) $$E_{\partial_1} V(\hat{Y}_{H-T}) = a^2X^2[t + C^2 \sum_{i=1}^{N} p_i^2(1-\pi_i)/\pi_i] \qquad (4.4.4)$$

(b) $$E_{\partial_1} M(T_2') = a^2X^2[\lambda^2(1+t+C^2 \sum_{i=1}^{N} p_i^2/\pi_i) - 2\lambda(1+C^2 \sum_{i=1}^{N} p_i^2) + (1+C^2 \sum_{i=1}^{N} p_i^2)] \qquad (4.4.5)$$

$$\lambda_o = [1+C^2 \sum_{i=1}^{N} p_i^2]/[1+t+C^2 \sum_{i=1}^{N} p_i^2/\pi_i] \qquad (4.4.6)$$

$$E_{\partial_1} M_o(T_2') = a^2X^2[(1+C^2\Sigma p_i^2) - \{(1+C^2\Sigma p_i^2)^2/(1+t+C^2\Sigma p_i^2/\pi_i)\}]$$

(c) $$E_{\partial_1} M(T_2^*) = a^2X^2[\lambda^{*2}n(n+C^2) - 2\lambda^*(n+C^2 \sum_{i=1}^{N} \pi_i p_i) + (1+C^2 \sum_{i=1}^{N} p_i^2)] \qquad (4.4.7)$$

$$\lambda_o^* = \frac{1}{n}[(n+C^2 \sum_{i=1}^{N} \pi_i p_i)/(n+C^2)] < 1 \qquad (4.4.8)$$

$$E_{\partial_1} M_o(T_2^*) = a^2X^2[(1+C^2 \sum_{i=1}^{N} p_i^2) - \{(n+C^2\Sigma\pi_i p_i)^2/(n^2+nC^2)\}] \qquad (4.4.9)$$

where λ_o and λ_o^* are the optimum values of λ and λ^* which minimise the expected mean square errors of T_2' and T_2^* respectively.

<u>Proof</u>. (a) From (4.1.3), it may be shown that

$$E_{\partial_1} V(\hat{Y}_{H-T}) = (\sigma^2+a^2)\sum_{i=1}^{N} x_i^2/\pi_i + a^2 \sum_{i\neq j=1}^{N}\sum x_i x_j(\pi_{ij}/\pi_i\pi_j) - \sigma^2 \sum_{i=1}^{N} x_i^2 - a^2X^2$$

$$= \sigma^2 \sum_{i=1}^{N} x_i^2(1-\pi_i)/\pi_i + a^2[\sum_{i=1}^{N} x_i^2/\pi_i + \sum_{i\neq j=1}^{N}\sum x_i x_j(\pi_{ij}/\pi_i\pi_j) - X^2]$$

$$= \sigma^2 \sum_{i=1}^{N} x_i^2(1-\pi_i)/\pi_i + a^2V(\hat{X}_{H-T})$$

$$= a^2X^2\ [t + C^2 \sum_{i=1}^{N} p_i^2(1-\pi_i)/\pi_i]$$

(b) Observing that,

$$M(T_2') = \lambda^2V(\hat{Y}_{H-T}) + (\lambda-1)^2\ Y^2,$$

we have,

$$E_{\partial_1} M(T_2') = \lambda^2a^2X^2[t+C^2 \sum_{i=1}^{N} p_i^2(1-\pi_i)/\pi_i] + (\lambda-1)^2\ E_{\beta_1} Y^2$$

$$= a^2X^2[\lambda^2(1+C^2 \sum_{i=1}^{N} p_i^2/\pi_i+t) - 2\lambda(1+C^2 \sum_{i=1}^{N} p_i^2) + (1+C^2 \sum_{i=1}^{N} p_i^2)\ .$$

Obviously, the value of λ which minimises $E_{\partial_1} M(T_2')$, is

$$\lambda_o = (1+C^2 \sum_{i=1}^{N} p_i^2)/(1+t+C^2 \sum_{i=1}^{N} p_i^2/\pi_i)$$

(c) From (4.2.1) with $\beta_i = \lambda^*/p_i$ in (4.1.1), or otherwise, it may be shown that,

$$M(T_2^*) = \lambda^2[\sum_{i=1}^{N}(y_i^2/p_i^2)\pi_i + \sum_{i\neq j=1}^{N}\sum(y_iy_j/p_ip_j)\ \pi_{ij}]$$

$$-2\lambda(\sum_{i=1}^{N} y_i) \sum_{i=1}^{N}(y_i/p_i)\pi_i + \sum_{i=1}^{N} y_i^2 + \sum_{i\neq j=1}^{N}\sum y_iy_j$$

Now,

$$E_{\partial_1} M(T_2^*) = \lambda^2 [\sum_{i=1}^{N} (x_i^2/p_i^2)\pi_i(\sigma^2+a^2) + a^2 \sum_{i \neq j=1}^{N}\sum (x_i x_j/p_i p_j)\pi_{ij}]$$

$$- 2\lambda [\sum_{i=1}^{N} (x_i^2/p_i)\pi_i(\sigma^2+a^2) + a^2 \sum_{i \neq j=1}^{N}\sum (x_i x_j/p_j)\pi_j] \quad (4.4.10)$$

$$+ \sigma^2 \sum_{i=1}^{N} x_i^2 + a^2 X^2$$

Observing that $\sum_{i=1}^{N} \pi_i = n$ and $\sum_{i \neq j=1}^{N}\sum \pi_{ij} = n(n-1)$ in (4.4.10), for fixed size design, we obtain

$$E_{\partial_1} M(T_2^*) = a^2 X^2 [\lambda^{*2} n(n+C^2) - 2\lambda^*(n+C^2 \sum_{i=1}^{N} \pi_i p_i) + (1+C^2 \sum_{i=1}^{N} p_i^2)] \quad (4.4.11)$$

Obviously, the value of λ^* which minimises $E_{\partial_1} M(T_2^*)$ would be

$$\lambda_o^* = (n + C^2 \sum_{i=1}^{N} \pi_i p_i)/n(n+C^2)$$

and the resulting expected MSE would be

$$E_{\partial_1} M_o(T_2^*) = a^2 X^2 [(1+C^2 \sum_{i=1}^{N} p_i^2) - \{(n+C^2 \sum_{i=1}^{N} \pi_i p_i)^2/(n^2+nC^2)\}]$$

Further, using Cauchy Schwartz inequality $\sum_{i=1}^{N} \pi_i p_i \leq (\sum_{i=1}^{N} \pi_i)(\sum_{i=1}^{N} p_i)$, we have

$$0 < \lambda_o^* < \frac{1}{n} \frac{a^2 n + n a^2}{(a^2 + n a^2)} = \frac{(1+C^2)}{(n+C^2)} < 1.$$

Now with the help of the Lemma 4.4.1, we have the following Theorems :

<u>Theorem 4.4.1</u>. In case C^2 is known exactly, a ∂_1-optimal estimator in T_2^* may be defined as

$$T^*_{o2} = \lambda^*_o \sum_{i=1}^{n} y_i/p_i$$

where λ^*_o is defined in (4.4.8). A sufficient condition for T^*_{o2} to be ∂_1-better (in the sense of having smaller expected MSE under ∂_1) than $\hat{Y}_{H-T}$ would be

$$\pi_i \leq 1/2 \quad \text{and} \quad t \geq (1-2\sum_{i=1}^{N} \pi_i p_i) \tag{4.4.12}$$

Proof. From (4.4.4) and (4.4.9), it may be shown that

$$E_{\partial_1} M_o(T^*_2) - E_{\partial_1} V(\hat{Y}_{H-T}) = \sigma^2 X^2 [1 + C^2 \sum_{i=1}^{N} p_i^2 - \{(n + C^2 \sum_{i=1}^{N} \pi_i p_i)^2/(n^2+nC^2)\} - t - C^2 \sum_{i=1}^{N} p_i^2 (1-\pi_i)/\pi_i]$$

$$= [2nC^4 \sum_{i=1}^{N} p_i^2 + 2n^2C^2 \sum_{i=1}^{N} p_i^2 + nC^2 - C^4(\sum_{i=1}^{N} \pi_i p_i)^2 - 2nC^2 \sum_{i=1}^{N} \pi_i p_i - n^2 t - ntC^2 - nC^4 \sum_{i=1}^{N} p_i^2/\pi_i - n^2C^2 \sum_{i=1}^{N} p_i^2/\pi_i]/(n^2+nC^2) \tag{4.4.13}$$

Thus $E_{\partial_1} M_o(T^*_2) \leq E_{\partial_1} V(\hat{Y}_{H-T})$

$$\text{iff, } 2\sum_{i=1}^{N} p_i^2 + \frac{1}{C^2} + \frac{2n \sum_{i=1}^{N} p_i^2}{C^2} \leq \frac{1}{n}(\sum_{i=1}^{N} \pi_i p_i)^2 + \frac{2}{C^2}\sum_{i=1}^{N} \pi_i p_i + \frac{t}{C^2} + \frac{nt}{C^4} + \sum_{i=1}^{N} p_i^2/\pi_i + \frac{n \sum_{i=1}^{N} p_i^2/\pi_i}{C^2}$$

$$\text{i.e., iff, } \frac{1}{n}(\sum_{i=1}^{N} \pi_i p_i)^2 + \frac{2}{C^2}(\sum_{i=1}^{N} \pi_i p_i) + \frac{(t-1)}{C^2} + \frac{nt}{C^4} + (1+\frac{n}{C^2})\sum_{i=1}^{N} p_i^2 (\frac{1}{\pi_i} - 2) \geq 0$$

$$\text{i.e., iff } \frac{1}{n}(\sum_{i=1}^{N} \pi_i p_i)^2 + \frac{1}{C^2}(t - 1 + 2\sum_{i=1}^{N} \pi_i p_i) + \frac{nt}{C^4} + (1+\frac{n}{C^2})\sum_{i=1}^{N} p_i^2 (\frac{1}{\pi_i} - 2) \geq 0 \tag{4.4.14}$$

Hence from (4.4.14), the sufficient condition follows.

Theorem 4.4.2. For any design of fixed sample size, ∂_1-optimal estimator in T_2', viz.,

$$T_{o2}' = [1+C^2 \sum_{i=1}^{N} p_i^2]\hat{Y}_{H-T}/[1+t+C^2 \sum_{i=1}^{N} p_i^2/\pi_i]$$

is always ∂_1-better than $\hat{Y}_{H-T}$.

Proof. From (4.4.5) and (4.4.6), it may be shown that

$$E_{\partial_1} M(T_{o2}') = E_{\partial_1} V(\hat{Y}_{H-T})[\{1+C^2 \sum_{i=1}^{N} p_i^2\}/\{1+t+C^2 \sum_{i=1}^{N} p_i^2/\pi_i\}]$$

$$< E_{\partial_1} V(\hat{Y}_{H-T})$$

and hence is the result.

It is to be noted that either of the estimators T_{o2}^* and T_{o2}' can never be defined, unless the exact knowledge C^2 is available. We get rid of this limitation in the following

Theorem 4.4.3. For the class of populations and sampling designs specified respectively by (4.4.2) and (4.4.12), the estimator $T_2^* = \lambda^* \sum_{i=1}^{n} y_i/p_i$ will be ∂_1-better than $\hat{Y}_{H-T}$ if λ is chosen such that

$$\lambda_{o2}^* - \frac{\sqrt{A}_{(1)}}{n(n+C_{(2)}^2)} < \lambda^* < \lambda_{o1}^* + \frac{\sqrt{A}_{(1)}}{n(n+C_{(2)}^2)} \qquad (4.4.15)$$

where

$$A_{(1)} = [C_{(1)}^4(\Sigma\pi_i p_i)^2 + n^2 t + nC_{(1)}^2(t-1+2\sum_{i=1}^{N}\pi_i p_i)$$

$$+n\sum_{i=1}^{N} p_i^2(\frac{1}{\pi_i} - 2)C_{(1)}^2(n+C_{(1)}^2)],$$

λ^*_{o1}, λ^*_{o2} are two known quantities satisfying

$$\lambda^*_{o1} \leq \lambda^*_o \leq \lambda^*_{o2}$$

and λ^*_o is the same as defined in (4.4.8).

<u>Proof</u>. From (4.4.4) and (4.4.7), it may be shown that

$$E_{\partial_1} M(T^*_2) - E_{\partial_1} V(\hat{Y}_{H-T}) = \sigma^2 X^2 [\lambda^{*2} n(n+C^2) - 2\lambda^*(n+C^2 \sum_{i=1}^{N} \pi_i p_i)$$
$$+(1+C^2 \sum_{i=1}^{N} p_i^2) - t - C^2 \sum_{i=1}^{N} p_i^2 (1-\pi_i)/\pi_i] \qquad (4.4.16)$$

The right hand side of (4.4.16) is a quadratic equation in λ^* and hence will be negative for all values λ^* satisfying $\lambda^*_{(1)} < \lambda^* < \lambda^*_{(2)}$, λ^*_1 and λ^*_2 being the two roots of the quadratic equation in λ^*, provided they exist, i.e., discriminant is non-negative. Now discriminant

$$D^* = \sigma^4 . A,$$

where

$$A = C^4(\Sigma\pi_i p_i)^2 + n^2 t + nC^2(t-1+2\sum_{i=1}^{N} \pi_i p_i) + n\Sigma p_i^2(\frac{1}{\pi_i} - 2)C^2(n+C^2)$$

and discriminant will be positive under the condition (4.4.12). Then the two roots will be

$$\lambda_1 = \lambda^*_o - \sqrt{D^*}/b \quad \text{and} \quad \lambda_2 = \lambda^*_o + \sqrt{D^*}/b$$

where $b = \sigma^2 n(n+C^2)$.

Thus for all values of λ satisfying $\lambda_1 < \lambda < \lambda_2$, the expression (4.4.16) will always be negative. Obviously λ_1 and λ_2 depend on the knowledge of C^2. This problem can be overcome by shrinking the interval of choices of λ in the following manner. The interval of preference $\lambda_o^* - \sqrt{D^*}/b < \lambda < \lambda_o^* + \sqrt{D^*}/b$ may be shrunken through

$$\lambda_{o2}^* - \sqrt{A_{(1)}}/n(n+C_{(2)}^2) < \lambda < \lambda_{o1}^* + \sqrt{A_{(1)}}/n(n+C_{(2)}^2)$$

where $C_{(1)}^2 < C^2 < C_{(2)}^2$

Thus T_2^* with λ^* satisfying (4.4.15), will be better than $\hat{Y}_{H-T}$, provided the interval in (4.4.15) is consistent.

Following the same line as in the Theorem 4.4.3, a number of estimators from $T_2' = \lambda \sum_{i \in s} y_i/\pi_i$ can be generated merely depending on the knowledge on the bounds of C^2.

Remarks

(i) In case of simple random sampling, the ∂_1-optimal estimator T_{o2}^* in T_2^* may be found to be

$$T_{o2}^* = \frac{1}{n}[(n+fC^2)/(n+C^2)] \sum_{i=1}^{n} y_i/p_i, \qquad (4.4.17)$$

where $f = (n/N)$ and expected mean square error, from (4.4.9) becomes

$$E_{\partial_1} M(T_{o2}^*) = \sigma^2 X^2[(1+C^2 \sum_{i=1}^{N} p_i^2) - \{(n+C^2 f)^2/(n^2+nC^2)\}]. \quad (4.4.18)$$

Similarly, under π-P-S scheme $(\pi_i = np_i)$, the estimator T_{o2}^* becomes

$$T^*_{o2} = [(1+C^2 \sum_{i=1}^{N} p_i^2)/(n+C^2)] \sum_{i=1}^{n} y_i/p_i \quad (4.4.19)$$

with expected MSE, i.e.,

$$E_{\partial_1} M(T^*_{o2}) = a^2X^2[(1+C^2 \sum_{i=1}^{N} p_i^2) - \{n(1+C^2 \sum_{i=1}^{N} p_i^2)^2/(n+C^2)\}]. \quad (4.4.20)$$

(ii) The ∂_1-optimal estimator, T'_{o2} in case of simple random sampling turns out to be

$$T'_{o2} = N[1+C^2\Sigma p_i^2]\bar{y}/[1+t+C^2 \sum_{i=1}^{N} p_i^2/\pi_i] \quad (4.4.21)$$

and the expected MSE is found as

$$E_{\partial_1} M(T'_{o2}) = a^2X^2[(1+C^2 \sum_{i=1}^{N} p_i^2)-\{(1+C^2 \sum_{i=1}^{N} p_i^2)^2/(1+t+\frac{N}{n}C^2 \sum_{i=1}^{N} p_i^2)\}]. \quad (4.4.22)$$

Similarly, under π-P-S scheme, the estimator T'_{o2} becomes

$$T'_{o2} = [1+ C^2 \sum_{i=1}^{N} p_i^2] \hat{Y}_{H-T}/[1+C^2/n]$$

and the expected MSE becomes

$$E_{\partial_1} M(T'_{o2}) = a^2X^2[(1+C^2 \sum_{i=1}^{N} p_i^2) - \{n(1+C^2 \sum_{i=1}^{N} p_i^2)^2/(n+C^2)\}]. \quad (4.4.23)$$

Thus from (4.4.20) and (4.4.23), it is observed that under π-P-S scheme, both T'_{o2} and T^*_{o2} are equally efficient.

(iii) Further, it is interesting to note that although under π-P-S scheme, $\hat{Y}_{H-T}$ is ∂_1-optimal in unbiased T_2-subclass of estimators (Hanurav, T.V., 1966, Sankhyā), it remains no longer ∂_1-optimal even in T_2-subclass, if the condition of unbiassedness gets relaxed.

4.5 Empirical Study

The following Table 4.5.1 gives the number of inhabitants $x_1, x_2, \ldots, x_{64}$ of 64 large cities in United States (Cochran, 1977, pp.92), x_i being the population of i^{th} city.

Table 4.5.1. Number of inhabitants

x : population

797	314	172	121
773	298	172	120
748	296	163	119
734	258	162	118
588	256	161	118
577	243	159	116
507	238	153	113
507	237	144	113
457	235	138	110
438	235	138	110
415	216	138	108
401	208	138	106
387	201	136	104
381	192	132	101
324	180	130	116
315	179	126	100

Let the above 64 cities be regarded as a random sample from a large super population which may be the population of all the cities in U.S.A. Let $y_1, y_2, \ldots, y_{64}$ be the number of inhabitants at some other period such that the model (4.4.2) is satisfied. Let $n = 6$.

For this population we have $t = \frac{C_x^2}{n} = .083816$, and

$$\sum_{i=1}^{N} p_i^2 = .023482.$$

The following Table 4.5.2 gives the percent relative efficiency of T'_{o2} and T^*_{o2} over $\hat{Y}^*_{H-T}$ for some values of $C^2 = \sigma^2/a^2$

Table 4.5.2 Percent relative efficiency of T'_{o2} and T^*_{o2} over $\hat{Y}_{H-T}$ for some values of C^2.

C^2	Under π-P-S Scheme		Under SRSWOR	
	$\left[\frac{E_{\partial_1} V(\hat{Y}_{H-T})}{E_{\partial_1} M(T^*_{o2})}\right]\times 100$	$\left[\frac{E_{\partial_1} V(\hat{Y}_{H-T})}{E_{\partial_1} M(T'_{o2})}\right]\times 100$	$\left[\frac{E_{\partial_1} V(\hat{Y}_{H-T})}{E_{\partial_1} M(T^*_{o2})}\right]\times 100$	$\left[\frac{E_{\partial_1} V(\hat{Y}_{H-T})}{E_{\partial_1} M(T'_{o2})}\right]\times 100$
(1)	(2)	(3)	(4)	(5)
0.01	100.1398	100.1398	5824.1681	108.6066
0.05	100.7147	100.7147	1206.2458	109.5122
0.10	101.4281	101.4281	679.8831	110.6401
0.50	107.0760	107.0760	266.0057	119.5432
1.00	113.9884	113.9884	223.0989	130.4225
2.00	127.3523	127.3523	215.732[illegible]	151.4381
5.00	164.0558	164.0558	252.2047	209.1682
10.00	215.9558	215.9558	321.0065	290.7414

The Table 4.5.2 reveals that in some situations, the gain by using T^*_{o2} over $\hat{Y}_{H-T}$ may be very very high.

4.6 Estimation of Mean Square Error

In this section, we consider the problem of estimating mean square errors of T'_2 and T^*_2 and look for the conditions under which the estimators of mean square error will be non-negative.

From (4.2.1), it may be shown that

$$M(T_2) = \sum_{i=1}^{N} \sum_{j=1}^{N} y_i y_j (\beta_i \beta_j \pi_{ij} + 1 - 2\beta_j \pi_j) \qquad (4.6.1)$$

An unbiased estimator of $M(T_2)$ is given by

$$\hat{M}(T_2) = \sum_{i\epsilon s} \sum_{j\epsilon s} y_i y_j(\beta_i \beta_j \pi_{ij} + 1 - 2\beta_j \pi_j)/\pi_{ij} \qquad (4.6.2)$$

and hence unbiased estimates for $M(T_2')$ and $M(T_2^*)$ can be obtained from (4.6.2) by replacing β_i by λ/π_i and λ^*/p_i respectively.

Let $y_i > 0$ for $i = 1,2,\ldots,N$. A set of sufficient conditions for non-negativity of (4.6.2) would be

$$\beta_i \beta_j + (1-2\beta_j \pi_j)/\pi_{ij} \geq 0 \quad \text{for } i,j = 1,2,\ldots,N \qquad (4.6.3)$$

In fact, the condition (4.6.3) should hold only for $i \neq j$, because for $i = j$, the condition (4.6.3) reduces to

$$\beta_i^2 - 2\beta_i + \frac{1}{\pi_i} \geq 0$$

$$\text{i.e.,} \quad (\beta_i - 1)^2 + \{(1-\pi_i)/\pi_i\} \geq 0,$$

which always holds true.

Thus from (4.6.3), it is found that $\hat{M}(T_2')$ will be non-negative if for all $i \neq j$,

$$\lambda^2/\pi_i \pi_j + (1-2\lambda)/\pi_{ij} \geq 0 \qquad (4.6.4)$$

the condition (4.6.3) will always be true, irrespective of the sampling design, in case $\lambda \leq 1/2$. Thus for all those populations with $C^2(\hat{Y}_{H-T}) \geq 1$, $\hat{M}(T_2')$ will be non-negative in case the optimum estimator T_{o2}' (T_2' with $\lambda = \lambda_o$) is used. Further in case $C_{(1)}^2(\hat{Y}_{H-T}) \geq 1$, in T_2' is such that $\{1/(1+C^2_{(1)}(\hat{Y}_{H-T}))\} \leq \lambda \leq 1/2$, $\hat{M}(T_2')$ will be non-negative.

Further, the condition (4.6.4) may be expressed as

$$\frac{1}{\pi_i \pi_j} [\lambda^2 - \{(2\lambda-1)\ \pi_i \pi_j / \pi_{ij}\}] \geq 0$$

or $(\lambda-1)^2 + \{(2\lambda-1)\ (\pi_{ij}-\pi_i \pi_j)/\pi_{ij} \geq 0$ (4.6.5)

i.e., the same condition $\pi_{ij}/\pi_i \pi_j \geq 1$, for $i \neq j$, under which $\hat{V}(\hat{Y}_{H-T})$ is non-negative [Horvitz and Thompson,(1952)], will make $\hat{M}(\hat{T}_2')$ also non-negative.

Similarly, from (4.6.3), $\hat{M}(\hat{T}_2^*)$ will be non-negative

if $\{\lambda^{*2}/p_i p_j\} + \{(1 - \frac{2\lambda^* \pi_j}{p_j})/\pi_{ij}\} \geq 0$ for $i \neq j$. (4.6.6)

CHAPTER V

A CLASS OF ESTIMATORS FOR POPULATION VARIANCE USING SOME APRIORI INFORMATION

5.1 Introduction and Summary

The problem of estimating population mean $\bar{Y}$ (or equivalently total $Y = N\bar{Y}$) of a character y has been considered extensively in the survey literature. There are available a large number of estimators for $\bar{Y}$ including the usual ratio, regression, difference and product estimators using the knowledge on population mean $\bar{X}$ of an available auxiliary character x. Further Das and Tripathi [(1979a), (1979b), (1980a), (1980b)] have discussed extensively the problem of estimating $\bar{Y}$ in case (a) $\bar{X}$ is known, (b) σ_x^2 is known and (c) C_x^2 is known, (d) any two of the above three quantities are known. In the previous chapters, we have also discussed, how the usual unbiased estimator for $\bar{Y}$ can be improved through a number of biased estimators using some apriori information on some parameters of the character y under study.

In many situations, the problem of estimating the variance σ_y^2 of a character y assumes importance. But the problem of estimating σ_y^2 has not attracted much attention, especially in case of finite populations. The problem has however been considered among others by [Wakimoto (1970), (1971)], [Singh, Pandey and Hirano (1973)], Liu (1974), Pandey and Singh (1977) and [Das and Tripathi (1977), (1978a)].

Let $s_y^2 = \sum_{i=1}^{n}(y_i-\bar{y})^2/(n-1)$, n being the sample size and $\bar{y}$ the sample mean. In case of finite populations, an unbiased estimator of σ_y^2 based on SRSWOR is given by

$$\hat{\sigma}_y^2 = (N-1)s_y^2/N. \qquad (5.1.1)$$

The exact expression for its variance is given by

$$V(\hat{\sigma}_y^2) = K[A\ \beta_2 - B]\ \sigma_y^4 \qquad (5.1.2)$$

where,

$$A = \{(N-1)/N\}^2 A^*,\ B = \{(N-1)/N\}^2 B^*,\ K = (N-n)/n(N-1)$$

with

$$A^* = \{n/(n-1)\}^2 - \{2n(N-2n)/(n-1)^2(N-2)\}$$
$$+ \{(N^2+N-6Nn+6n^2)/(n-1)^2(N-2)(N-3)\}$$

$$B^* = \{nN^2/(n-1)(N-1)(N-2)\} - \{3N(N-n-1)/(n-1)(N-2)(N-3)\}$$

[Sukhatme (1944); Das (1982)].

In case of infinite populations (or finite populations with sampling fraction ignored), it is well known that s_y^2, based on a simple random sample, is unbiased for σ_y^2 and exact expression for its variance is given by

$$V(s_y^2) = (\sigma_y^4/n)[\beta_2 - \{(n-3)/(n-1)\}] \qquad (5.1.3)$$

which is obviously obtainable from (5.1.2) by ignoring the sampling fraction n/N, in which case $A = 1$, $B = (n-3)/(n-1)$ and $K = 1/n$.

[Wakimoto (1970),(1971)] considered the problem of estimating σ_y^2 in case population is stratified and obtained an unbiased estimator of σ_y^2. Let the population has the distribution function $F(y)$ and density function $f(y)$ with finite mean $\bar{Y}$ and variance σ_y^2. Let this population be divided into L sub-populations, known as L strata, each having the distribution functions $F_1(y), F_2(y), \ldots, F_L(y)$ and the density functions $f_1(y), f_2(y), \ldots, f_L(y)$ with finite means $\bar{Y}_1, \bar{Y}_2, \ldots, \bar{Y}_L$ and variances $\sigma_1^2, \sigma_2^2, \ldots, \sigma_L^2$ respectively. For this stratified population, Wakimoto (1970) obtained an unbiased estimator u_s^2 of σ_y^2 defined by

$$u_s^2 = \sum_{i=1}^{L} \{w_i^2/n_i(n_i-1)\} \sum_{k<\ell} (y_{ik}-y_{i\ell})^2 + \sum_{i<j} \{w_i w_j/n_i n_j\} \sum_{k=1}^{n_i} \sum_{\ell=1}^{n_j} (y_{ik}-y_{j\ell})^2 \qquad (5.1.4)$$

based on random samples of sizes n_i drawn from stratum i $(i=1,2,\ldots,L)$ where w_i is a weight for the i^{th} stratum, depending on the relative frequency.

In particular, in case of proportional allocation, $n_i = n\, w_i$, an unbiased estimator of σ_y^2 is found to be

$$u_{s,p}^2 = [\sum_{i=1}^{L} \sum_{kk<\ell}^{n_i} (y_{ik}-y_{i\ })^2 + \sum_{i<j} \sum_{k=1}^{n_i} \sum_{\ell=1}^{n_j} (y_{ik}-y_{j\ell})^2]/n^2. \qquad (5.1.5)$$

He found that

$$V(u_r^2) \geq V(u_{s,p}^2) \qquad (5.1.6)$$

where,

$$u_r^2 = \Sigma(y_i-y_j)^2/n(n-1)$$

Liu (1974), in case of general sampling designs, constructed an admissible general unbiased quadratic estimator for the population variance σ_y^2 in sampling from a finite population. His proposed estimator is found to be

$$\nu_b(s,y) = \sum_{i,j \epsilon s} b(s,i,j)\, y_i y_j \qquad (5.1.7)$$

with

$$\sum_{s \ni i} b(s,i,j)\, P(S) = \frac{1}{N}(1 - \frac{1}{N})$$

and

$$\sum_{\substack{s \ni i,j \\ (i \neq j)}} b(s,i,j)\, P(s) = \frac{1}{N^2} \text{ for all } i,j = 1,\ldots,N.$$

He also constructed an admissible general unbiased estimator for the variance of any unbiased quadratic estimator of the type $\upsilon_b(s,y)$.

Das and Tripathi (1977) have considered the general problem of estimating quadratic forms $Q = \theta'T\theta$, the problem of estimating σ_y^2 being a particular one, in case of finite populations, where $\theta' = (y_1, y_2, \ldots, y_N)$ and $T = (t_{ij})$, $(i,j=1,2,\ldots,N)$ is a matrix of specified elements. They have also observed that the class of quadratic estimators proposed by Liu (1974) is nothing but a particular member of their class. They proved that

$$d_u^*(s,y) = \sum_{i \epsilon s} y_i^2 t_{ii}/\pi_i + \sum_{i,j \epsilon s}\sum y_i y_j\, t_{ij}/\pi_{ij},\ \pi_{ij} > 0 \qquad (5.1.8)$$

is the uniformly minimum variance unbiased quadratic estimator for Q in the class

$$\mathcal{D}^* = \{d^* : d^* = \sum_{i,j \epsilon s}\sum \lambda(i,j)\, y_i y_j\} \qquad (5.1.9)$$

and is unbiased and admissible for Q in the unbiased sub-class of

$$\mathcal{D} = \{d(s,y) : d(s,y) = \sum_{(i,j) \epsilon s} \lambda(s,i,j)\, Y_i Y_j\} \qquad (5.1.10)$$

In case of varying probability sampling with replacement (VPSWR), Das and Tripathi (1977) obtained an unbiased estimator for σ_y^2 as

$$\sigma_{DT}^2 = \{\sum_{i=1}^{n}(y_i^2/p_i) - \sum_{i \neq j=1}^{n}\sum(y_i y_j/p_i p_j)/(n-1)N\}/Nn \qquad (5.1.11)$$

Pandey, Singh and Hirano (1973), following Searls (1964) defined an estimator for σ_y^2 as

$$T_1 = \lambda_1 \, s_y^2 \qquad (5.1.12)$$

with its mean square error, MSE and bias as

$$M(T_1) = \sigma_y^4[\lambda_1^2(\Delta/n) - 2\lambda_1 + 1]$$

$$\text{and} \quad B(T_1) = (\lambda_1 - 1)\,\sigma_y^2 \qquad (5.1.13)$$

where $\Delta = \beta_2 + [(n^2-2n+3)/(n-1)]$

The optimum choice λ_o of λ which minimises $M(T_1)$ is found to be

$$\lambda_o = (n/\Delta)$$

and in case β_2 is known exactly, the resulting estimator T_{o1} and its mean square error, bias and relative efficiency (RE) over s_y^2 are found to be

$$T_{o1} = (n/\Delta)s_y^2 \qquad (5.1.14)$$

$$M(T_{o1}) = \sigma_y^4(\Delta - n)/\Delta \qquad (5.1.15)$$

$$B(T_{o1}) = [(n-\Delta)/\Delta]\sigma_y^2 \qquad (5.1.16)$$

$$\text{and} \quad R.E. = (\Delta/n) \qquad (5.1.17)$$

In case, β_2 is not known exactly, they defined the modified estimator, using the approximate value, $\beta_o = \alpha\,\beta_2(\alpha > 0)$, instead of β_2 in T_{o1}, as

$$\tilde{T}_{o1} = (n/\Delta^{**})s_y^2 \qquad (5.1.18)$$

where Δ^{**} is Δ with β_2 replaced by β_o.

Further, in case of a prior value σ_o^2 of σ_y^2 is available, Pandey and Singh (1977), following Singh (1969), proposed a class of estimators for variance σ_y^2 as

$$\sigma_{PS}^2 = w s_y^2 + (1-w)\sigma_o^2, \; 0 \leq w \leq 1. \qquad (5.1.19)$$

The optimum value w_o of w which minimises $M(\sigma_{PS}^2)$ and the resulting MSE are

$$w_o = \{1 - (\sigma_o/\sigma_y)^2\}^2/[\{1-(\sigma_o/\sigma_y)^2\}^2+\{(\Delta-n)/n\}] \qquad (5.1.20)$$

$$M_o(\sigma_{PS}^2) = \sigma_y^4 \, w_o(\Delta-n)/n \qquad (5.1.21)$$

Further they modified the above estimator by

$$\sigma_{PS}^2 = \hat{w} \, s_y^2 + (1-\hat{w}) \, \sigma_o^2 \qquad (5.1.22)$$

where $\hat{w}$ is obtained from w_o by replacing σ_y^2 by s_y^2 and β_2 by a consistent estimator of it.

It was also observed that the proposed estimator in (5.1.19) is better than usual unbiased estimator s_y^2, if n is small and $0.5 \leq \sigma_o^2/\sigma_y^2 < 1.5$.

Das and Tripathi (1980a) discussed an estimator

$$d = w_1 \, s_y^2 + w_2 \, \bar{y} \qquad (5.1.23)$$

for σ_y^2, where the weights w_1 and w_2 need not add to unity such that $d > 0$. They found that d is better than T_{o1} in case

$$0 < w_2 < 2(1-w_1) \qquad (5.1.24)$$

with $w_1 = (n/\Delta)$ in case of the population in which $\bar{Y} \geq 1$ and $C_y^2 \geq n/(n-1)$.

Further, according to the situation in which $\bar{X}$ or variance σ_x^2 or coefficient of variation C_x of an auxiliary character x is known, Das and Tripathi (1980a), discussed a number of estimators for σ_y^2 based on simple random sampling, including the following ones :

$$T_1' = s_y^2 - \alpha_1(\bar{x}-\bar{X}) \qquad (5.1.25)$$

$$T_2' = s_y^2 - \alpha_2(s_x^2-\sigma_x^2) \qquad (5.1.26)$$

$$\text{and} \quad T_3' = s_y^2 - \alpha_3(c_x^2-C_x^2) \qquad (5.1.27)$$

where $s_x^2 = \sum_{i=1}^{n}(x_i-\bar{x})^2/(n-1)$, $c_x^2 = s_x^2/\bar{x}^2$ and α's being the suitably chosen constants.

In this chapter, we have considered a class of estimators for σ_y^2 based on SRSWOR, ignoring sampling fraction, defined by

$$C_{\lambda,v} = \{d : d = \lambda' v\} \qquad (5.1.28)$$

where $v' = (s_y^2, t)$, $\lambda' = (\lambda_1, \lambda_2)$ λ_1 and λ_2 being suitably chosen constant weights and t being a suitably chosen statistic such that its variance σ_t^2 exists.

The estimators due to Pandey and Singh (1977), Pandey, Singh and Hirano (1973), and Das and Tripathi (1980a) may be identified as particular members of the class defined by (5.1.28).

Some particular members from this class, for some specific choices of t, have been picked up for detailed study and it has been shown that under moderate conditions, these estimators are better than the usual unbiased estimator s_y^2 and those proposed by Pandey and Singh (1977), Pandey, Singh and Hirano (1973) and Das and Tripathi (1978a). These conditions depend either on exact values of some parameters or depend merely on the knowledge of some of the quantities $\bar{Y}_{(1)}, \bar{Y}_{(2)}, C_{(1)}, C_{(2)}, \beta_1^{(1)}$ and $\beta_2^{(2)}$ which are such that

$$0 < \bar{Y}_{(1)} \leq \bar{Y} \leq \bar{Y}_{(2)}, \quad 0 < C_{(1)} \leq C_y \leq C_{(2)}$$
$$\beta_1^{(1)} \leq \beta_1 \leq \beta_1^{(2)} \; ; \quad \beta_2 \leq \beta_2^{(2)} \tag{5.1.29}$$

where $\beta_1 = \mu_3^2/\sigma_y^6$.

5.2 Properties of the Proposed Class of Estimators

Using the general results in (3.3.2), we get

$$B(d) = \lambda'\psi - \sigma_y^2 \tag{5.2.1}$$

and
$$M(d) = \lambda'G\lambda - 2\sigma_y^2 \lambda'\psi + \sigma_y^4 \tag{5.2.2}$$

where,

$$G = \begin{pmatrix} V(s_y^2) + \sigma_y^4 & E(s_y^2\, t) \\ E(s_y^2\, t) & E(t^2) \end{pmatrix}, \quad \psi = \begin{pmatrix} \sigma_y^2 \\ E(t) \end{pmatrix}$$

Further, following the results in Theorem 3.3.1, we have

Theorem 5.2.1. The value of λ which minimises the mean square error of d and the resulting MSE and bias would be given by

$$\lambda_o = \sigma_y^2 \bar{G} \; \psi \tag{5.2.3}$$
$$B_o(d) = -M_o(d)/\sigma_y^2 \tag{5.2.4}$$

and
$$M_o(d) = [1 - \psi'(\bar{G})'\psi]\sigma_y^4 \tag{5.2.5}$$

where $\bar{G}$ is a g-inverse of the matrix G.

Further using the results in (3.3.7) and (3.3.8), in case G is a positive definite matrix, we get,

$$\lambda_{o1} = \sigma_y^2[\sigma_y^2 \; E(t^2) - E(t).E(s_y^2 t)]/D(s_y^2, t)$$
$$\lambda_{o2} = \sigma_y^2[\; E(s_y^2)^2 \; E(t) - E(s_y^2).E(s_y^2.t)]/D(s_y^2, t) \tag{5.2.6}$$

and
$$M_o(d) = \sigma_y^4[1 - \{N(s_y^2, t)/D(s_y^2, t)\}] \tag{5.2.7}$$

where $D(s_y^2,t)$ and $N(s_y^2,t)$ are defined by $D(\hat{\theta},t)$ and $N(\hat{\theta},t)$ respectively in (3.3.9) with $\hat{\theta}$ replaced by s_y^2.

In particular $\lambda_o' = (\lambda_{o1}, \lambda_{o2})$ would not be known, as it depends upon a number of parameters. In that case by observing from (5.1.13) and (5.2.2),

$$M(d) = M(T_1) + \lambda_2^2 \, Et^2 - 2\lambda_2[\sigma_y^2 \, E(t) - \lambda_1 \, E(s_y^2 \, t)] \qquad (5.2.8)$$

we shall be able to generate estimators from d better than s_y^2, $T_1 = \lambda_1 \, s_y^2$ and T_{o1}.

Thus for a specified λ_1, the estimator $d = T_1 + \lambda_2 t$ would be better than T_1

$$\text{iff } \lambda_2 \text{ lies between } 0 \text{ and } 2\lambda_{o2}^* \qquad (5.2.9)$$

where,

$$\lambda_{o2}^* = \{(1-\lambda_1) \, \sigma_y^2 \, E(t) - \lambda_1 \, Cov(s_y^2,t)\}/Et^2$$

is the optimum choice of λ_2 minimising $M(d)$ for a given λ_1.

If we use $\lambda_2 = \lambda_{o2}^*$, then

$$M_o(d) = M(T_1) - \lambda_{o2}^{*2} \, E \, t^2.$$

Thus in absence of the exact knowledge of λ_{o2}^*, the conditions (5.2.9) may be used to generate estimators better than s_y^2, T_1 and T_{o1} from d.

A subclass of d which would draw special attention may be defined as

$$d' = s_y^2 + \lambda_2' t$$

and its mean square error would be given from (5.2.2) with $\lambda_1 = 1$ and λ_2 being replaced by λ_2'.

Further, it may be shown that the estimators in d' would be better than s_y^2

$$\text{iff} \quad 0 < \lambda_2' < 2\lambda_{o2}' \quad \text{in case } \rho_{s_y^2,t} < 0$$
$$\text{or} \quad 2\lambda_{o2}' < \lambda_2' < 0 \quad \text{in case } \rho_{s_y^2,t} > 0 \qquad (5.2.10)$$

where $\rho_{s_y^2,t}$ is the correlation coefficient between s_y^2 and t and

$$\lambda_{o2}' = - Cov(s_y^2,t)/E(t^2)$$

is the optimum choice of λ_2' in d'.

It may be noted that one should never use such a t in d' which is uncorrelated with s_y^2. However such a t is permissible in d.

5.3 Identification of Some Specific Estimators Better than Usual Unbiased and Some Other Known Estimators

In this section, we shall find that for suitable choices of a statistic t such that σ_t^2 exists, we can always generate estimators better than s_y^2, T_1 and T_{o1}.

From (5.1.3) and (5.1.13), it may be shown that T_1 would be better than s_y^2

$$\text{iff} \quad 2(n/\Delta) - 1 < \lambda_1 < 1. \qquad (5.3.1)$$

Obviously, a sufficient condition for T_1 to be better than s_y^2 would be

$$2(n/\Delta_{(1)}) - 1 \leq \lambda_1 < 1 \qquad (5.3.2)$$

where $\Delta_{(1)}$ is a known quantity such that $n < \Delta_{(1)} < \Delta$.

From (3.2.8), another sufficient condition for T_1 to be better than s_y^2 would be

$$n/\Delta_{(1)} \leq \lambda_1 < 1. \qquad (5.3.3)$$

Let $d = \lambda_1 s_y^2 + \lambda_2 \sigma_o^2$ be an estimator for σ_y^2, where σ_o^2 may be a good-guessed value of σ_y^2. Then from (5.2.1) and (5.2.2), we get

$$B(d) = (\lambda_1 - 1)\sigma_y^2 + \lambda_2 \sigma_o^2$$

$$\text{and} \quad M(d) = \lambda_1^2 V(s_y^2) + \{(\lambda_1 - 1)\sigma_y^2 + \lambda_2 \sigma_o^2\}^2 \qquad (5.3.4)$$

Thus $M(d) \leq V(s_y^2)$,

$$\text{iff} \quad (\lambda_1^2 - 1)V(s_y^2) + (\lambda_1 - 1)^2 \sigma_y^4 + \lambda_2^2 \sigma_o^2 + 2\lambda_2 \sigma_o^2 \sigma_y^2 (\lambda_1 - 1) \leq 0 \qquad (5.3.5)$$

which indicates that whatever be λ_2, for $\lambda_1 \in (-\infty, -1]$ and $\lambda_1 \in [1, \infty)$, d would be worse than s_y^2.

It may be shown that for a given $\lambda_1 \in (-1, 1)$, d would be better than s_y^2 for all λ_2 satisfying

$$0 < \lambda_2 \leq 2(1 - \lambda_1)\, \sigma_y^2/\sigma_o^2. \qquad (5.3.6)$$

Let σ_o, the guessed value be $\sigma_{(2)} \geq \sigma_y$, then a sufficient condition for d to be better than s_y^2 would be

$$0 < \lambda_2 \leq 2(1 - \lambda_1)\, \sigma_{(1)}^2/\sigma_{(2)}^2, \qquad (5.3.7)$$

where $\sigma_{(1)} \leq \sigma_y \leq \sigma_{(2)}$. Thus if we have $\sigma_{(1)} \leq \sigma_y \leq \sigma_{(2)}$, then $\sigma_{(2)}$ may be used as a guessed value of σ_y as in d and $\sigma_{(1)}$ can be used to shrink the interval of preference.

Let us consider the same population I as in the empirical study of the Chapter III. (Section 3.7) where σ_y^2=412,624.88 and Δ = 43.611051.

Let $n = 30$.

Let σ_o^2, the good-guessed value of σ_y^2 be taken as $\sigma_2^2 = 420000$ and $T_{o1} = (n/\Delta)\ s_y^2$, Δ being the same as for the population I of Chapter III and n also being the same as considered there.

The following Table 5.3.1 gives the percent relative efficiency of $d = T_{o1} + \lambda_2\ \sigma_2^2$ over T_{o1} for the values of $\sigma_{(1)}/\sigma_{(2)} = 0.1(0.1)\ 0.9$.

Table 5.3.1. Percent relative efficiency of d over T_{o1} for some values of $\sigma_{(1)}/\sigma_{(2)}$ =0.1(0.1)0.9.

$\sigma_{(1)}/\sigma_{(2)}$ λ_2	Range for λ_2	The values of λ_2 taken		100×[M(T_{o1})/M(d)]	
(1)	(2)	(3)		(4)	
0.1	$0 < \lambda_2 < 0.0062$	0.006	0.0031	100.91	100.47
0.2	$0 < \lambda_2 < 0.0249$	0.024	0.0124	103.61	101.87
0.3	$0 < \lambda_2 < 0.0562$	0.056	0.0281	108.32	104.22
0.4	$0 < \lambda_2 < 0.0998$	0.099	0.4499	114.34	122.11
0.5	$0 < \lambda_2 < 0.1561$	0.156	0.0781	121.32	111.48
0.6	$0 < \lambda_2 < 0.2247$	0.224	0.1124	127.36	116.10
0.7	$0 < \lambda_2 < 0.3058$	0.305	0.1529	130.14	120.98
0.8	$0 < \lambda_2 < 0.3995$	0.399	0.1997	126.68	125.54
0.9	$0 < \lambda_2 < 0.5056$	0.505	0.2528	115.56	128.95

Let $T_1^* = \lambda_1^* s_y^2$ where λ_1^* is any particular value of λ_1 lying in the ranges defined by (5.3.2) or (5.3.3).

In particular, λ_1^* may be taken as $\lambda_1^* = n/\Delta^*$, $n \leq \Delta^* \leq \Delta_{(1)} \leq \Delta$. We note that the estimator T_{o1} belongs to T_1^* for $\Delta_1^* = \Delta$ and that T_1^* is better than s_y^2.

We now consider the following classes of estimators.

(i) $d_1^* = T_1^* + \lambda_2 \; s_y^2$

(ii) $d_2^* = T_1^* + \lambda_2 \; \bar{y}$

(iii) $d_3^* = T_1^* + \lambda_2 \; \bar{y}^2$

(iv) $d_4^* = T_1^* + \lambda_2 \; s_y^2\bar{y}$

(v) $d_5^* = T_1^* + \lambda_2 \; s_y^2/\bar{y}^2$

(vi) $d_6^* = T_1^* + \lambda_2(\bar{x}-\bar{X})$

and

(i) $d_1' = s_y^2 + \lambda_2' \; s_y^2$

(ii) $d_2' = s_y^2 + \lambda_2' \; \bar{y}$

(iii) $d_3' = s_y^2 + \lambda_2' \; \bar{y}^2$

(iv) $d_4' = s_y^2 + \lambda_2' \; s_y^2\bar{y}$

(v) $d_5' = s_y^2 + \lambda_2' \; s_y^2/\bar{y}^2$

(vi) $d_6' = s_y^2 + \lambda_2'(\bar{x}-\bar{X})$

Obviously, d_i' $(1 \leq i \leq 6)$ is a subclass of d_i^* for $\lambda_1^* = 1$, i.e., $n = \Delta^*$.

From now, we shall use the following notation

$$\Delta_{(2)} = \beta_2^{(2)} + [(n^2-2n+3)/(n-1)] \geq \Delta$$

$$a = \begin{cases} +1 & \text{for positively skewed distributions} \\ 0 & \text{for symmetrical distributions} \\ -1 & \text{for negatively skewed distributions} \end{cases}$$

and $\mu_{21}^{(1)}(y,x) \leq \mu_{21}(y,x) \leq \mu_{21}^{(2)}(y,x)$, where $\mu_{21}(y,x)$ is the product moment of second and first order of y and x respectively.

From the condition (5.2.9), we get the following

Theorem 5.3.1. A set of necessary and sufficient conditions for the estimators d_1^* to d_6^* to be better than T_1^* is that the corresponding λ_2 lies between

$$0 \quad \text{and} \quad 2n(\Delta^*-\Delta)/\Delta^*\Delta \tag{5.3.8}$$

$$0 \quad \text{and} \quad 2(n/\Delta^*)C_y^2 \ \bar{Y}[\Delta^*-n-a\sqrt{\beta_1}C_y]/(n+C_y^2) \tag{5.3.9}$$

$$0 \quad \text{and} \quad 2C_y^2[(\Delta^*-n)(n+C_y^2)-2na\sqrt{\beta_1}C_y]/\Delta^*(n+6C_y^2) \tag{5.3.10}$$

$$0 \quad \text{and} \quad 2[n(\Delta^* -\Delta)+a\sqrt{\beta_1}C_y(\Delta^*-2n)]/\Delta^*\bar{Y}(C_y^2+4a\sqrt{\beta_1}C_y+\Delta) \tag{5.3.11}$$

$$0 \quad \text{and} \quad \frac{2\bar{Y}^2[\Delta^*(n+3C_y^2)-2a\Delta^* \sqrt{\beta_1}C_y-n(3C_y^2+\Delta)+4na\sqrt{\beta_1}C_y]}{\Delta^*(10C_y^2 - 8a\sqrt{\beta_1}C_y + \Delta)} \tag{5.3.12}$$

$$0 \quad \text{and} \quad -2(n/\Delta^*)\ \mu_{21}(y,x)/\sigma_x^2 \tag{5.3.13}$$

respectively. We now generate a set of sufficient conditions, corresponding to the above ones, for d_1^* to d_6^* to be better than T_1^*. The estimator d_1^* would be better than T_1^*

$$\text{if} \quad 2n(\Delta^*-\Delta_{(2)})/\Delta_{(2)}\Delta^* \leq \lambda_2 < 0. \tag{5.3.14}$$

In case, the population is <u>symmetrical or negatively skewed</u>, a sufficient condition for d_2^* to be better than T_1^* would be

$$0 < \lambda_2 \leq 2(n/\Delta^*)\, C_{(1)}^2\, \bar{Y}_{(1)}(\Delta^*-n)/(n+C_{(1)}^2) \tag{5.3.15}$$

and for <u>positively skewed distributions</u>, it would be given by

$$0 < \lambda_2 \leq 2(n/\Delta^*)C_{(1)}^2\, \bar{Y}_{(1)}[(\Delta^*-n)-\sqrt{\beta_1^{(2)}}\,C_{(2)}]/(n+C_{(1)}^2) \tag{5.3.16}$$

$$\text{or} \quad 2(n/\Delta^*)C_{(1)}^2\, \bar{Y}_{(1)}[(\Delta^*-n)-\sqrt{\beta_1^{(2)}}\, C_{(2)}]/(n+C_{(1)}^2) \leq \lambda_2 < 0$$

according as $(\Delta^*-n) > \sqrt{\beta_1^{(2)}}\, C_{(2)}$ or $(\Delta^*-n) < \sqrt{\beta_1^{(2)}}\, C_{(2)}$.

Further in case of <u>symmetrical distributions</u>, a set of sufficient conditions for d_3^*, d_4^* and d_5^* to be better than T_1^* would be given respectively by

$$0 < \lambda_2 \leq 2(\Delta^*-n)C_{(1)}^2(n+C_{(1)}^2)/\Delta^*(n+6C_{(1)}^2) \tag{5.3.17}$$

$$2n(\Delta^*-\Delta)/\Delta^*\, \bar{Y}_{(2)}\, (C_{(2)}^2+\Delta_{(2)}) \leq \lambda_2 < 0 \tag{5.3.18}$$

$$\text{and} \quad 0 < \lambda_2 \leq 2\bar{Y}_{(1)}^2[3C_{(1)}^2(\Delta^*-n)+n(\Delta^*-\Delta)]/\Delta^*(10C_{(2)}^2+\Delta_{(2)}) \tag{5.3.19}$$

provided $3C_{(1)}^2(\Delta^*-n) > n(\Delta-\Delta^*)$. Further d_6^* would be better than T_1^*, if

$$0 < \lambda_2 < -2(n/\Delta^*)\, \sigma_{(2)}^{-2}(x)\, \mu_{21}^{(2)}(y,x) \tag{5.3.20}$$

$$\text{or} \quad -2(n/\Delta^*)\, \sigma_{(2)}^{-2}(x)\, \mu_{21}^{(1)}(y,x) \le \lambda_2 < 0$$

according as $\mu_{21}(y,x) \le \mu_{21}^{(2)}(y,x) < 0$ or $0 < \mu_{21}^{(1)}(y,x) \le \mu_{21}(y,x)$.

Necessary and sufficient conditions as well as the corresponding sufficient conditions for d_1' to d_6' to be better than s_y^2 can be obtained from (5.3.8) to (5.3.20) by substituting $\Delta^* = n$ and replacing λ_2 by λ_2' therein, and present the sufficient conditions for d_1' to d_6' to be better than s_y^2.

The estimator d_1' would be better than s_y^2

$$\text{if} \quad 2(n-\Delta_{(2)})/\Delta_{(2)} \le \lambda_2' < 0. \tag{5.3.21}$$

In case of <u>positively skewed distribution</u>, the estimator d_2' would be better than s_y^2

$$\text{if} \quad -2C_{(1)}^3\, \gamma_{(1)}\, \sqrt{\beta_1^{(1)}}/(n+C_{(1)}^2) \le \lambda_2' < 0. \tag{5.3.22}$$

It is to be noted that no estimator from d_2' can be found to be better than s_y^2 in case of <u>symmetrical distributions</u>. Similarly one can not have an estimator of the type d_3' to be better than s_y^2 for <u>symmetrical distributions</u>. However, in case of <u>symmetrical distributions</u>, a set of sufficient conditions for d_4' and d_5', better than s_y^2 would be given by

$$2(n-\Delta)/\gamma_{(2)}(C_{(2)}^2 + \Delta_{(2)}) \le \lambda_2' < 0 \tag{5.3.23}$$

$$\text{and} \quad 2(n-\Delta)\, \gamma_{(1)}^2/[10C_{(2)}^2 + \Delta_{(2)}] \le \lambda_2' < 0 \tag{5.3.24}$$

Further, for <u>asymmetrical distributions</u>, the estimator d_6' would be better than s_y^2

$$\text{if} \quad 0 < \lambda_2' \leq -2\sigma_{(2)}^{-2}(x)\, \mu_{21}^{(2)}(y,x)$$
$$\text{or} \quad -2\sigma_{(2)}^{-2}(x)\, \mu_{21}^{(1)}(y,x) \leq \lambda_2' < 0 \qquad (5.3.25)$$

according as $\mu_{21}(y,x) \leq \mu_{21}^{(2)}(y,x) < 0$ or $0 < \mu_{21}^{(1)}(y,x) \leq \mu_{21}(y,x)$. Obviously for <u>bivariate symmetrical distribution</u>, e.g., bivariate normal population, no estimator from d_6' would be better than s_y^2

<u>Remarks</u>.

(i) From (5.3.8), it is to be noted that none of the estimators from d_1^* would be better than T_{o1}, because in that case $\Delta^* = \Delta$ and thereby having no λ_2 satisfying (5.3.8). The necessary and sufficient conditions and the corresponding sufficient conditions for d_2^*, d_3^* to d_6^* to be better than T_{o1} may be obtained from above conditions through replacing Δ^* by Δ.

(ii) The sufficient conditions for d_3^*, d_4^* and d_5^* to be better than T_1^* in case of <u>positively skewed</u> and <u>negatively skewed distributions</u> may be obtained likewise.

(iii) It is to be noted that we can generate a number of estimators using (5.3.8) to (5.3.12) better than MLE (maximum likelihood estimator) of σ_y^2 in case of normal populations. In case of the bivariate distributions, in which $\mu_{21}(y,x) = 0$, e.g., bivariate normal populations, no estimator from d_6^* is better than T_1^* and hence to improve s_y^2, there is no necessity of making use of the knowledge $\bar{X}$; however s_y^2 can be improved through T_1^*.

(iv) From (5.2.7), we find that if the statistic t in d is such that $Et = 0$, then $M(d)$ would be an increasing function of σ_t^2. Thus if t_1 and t_2 are such that $Et_1 = Et_2 = 0$ and $\sigma_{t_1}^2 < \sigma_{t_2}^2$, one should use t_1 and not the t_2. Further, we should never use such a t which has zero mean and is uncorrelated with s_y^2, as the resulting d would be worse than T_1 in that case.

(v) The estimator d_2^* was discussed by Das (1982) and the conditions (5.3.9) and (5.3.15) was obtained by him.

5.4 Some Discussion. on the Relative Performance of the Estimators.

We have already discussed in the above section that the estimator s_y^2 and the Searls-type estimators T_1 and T_{o1} may be improved by having some prior information merely, which is not very difficult to have, about some parameters. Following similar steps, the estimator $\tilde{T}_{o1}$ in (5.1.18) may also be improved through the class $d^{**} = \tilde{T}_{o1} + \lambda_2 t$ or through $d^* = T_1^* + \lambda_2 t$. It may also be noted that $\tilde{T}_{o1}$ with $n/\Delta^{**} < 2(n/\Delta)-1$ would be worse than s_y^2 while T_1^* is always better than s_y^2.

It is known (Das and Tripathi, 1980b) that in the case corresponding optimum weights are used in the estimators

$$d = \lambda_1 t_1 + \lambda_2 t_2 \quad \text{with} \quad \lambda_1 + \lambda_2 \neq 1$$

$$\text{and} \quad d' = \lambda_1' t_1 + \lambda_2' t_2 \quad \text{with} \quad \lambda_1' + \lambda_2' = 1 \tag{5.4.1}$$

then d would be better than d'. Thus

$$d_o = \lambda_{o1} s_y^2 + \lambda_{o2} \sigma_o^2$$

would be better than

$$\sigma_{PS}^2 = w_o s_y^2 + (1-w_o) \sigma_o^2$$

where w_o has been defined in (5.1.20).

However, in practice, it is not possible to use either of d_o and σ_{PS}^2. In general, it may be shown that $d = \lambda_1 s_y^2 + \lambda_2 \sigma_o^2$ would be better than σ_{PS}^2,

$$\text{iff} \quad \lambda_1^2(\Delta/n-1) + (\lambda_1+\lambda_2\sigma_o^2/\sigma_y^2-1)^2 \leq w^2(\Delta/n-1) + (1-w)^2(\sigma_o^2/\sigma_y^2-1)^2.$$

In particular, let $\lambda_1 = w$, then it may be shown that d would be better than σ_{PS}^2

iff λ_2 lies between $(1-w)$ and $(1-w)\ 2\sigma_y^2/\sigma_o^2 - 1$.

The estimator σ_{PS}^2 in (5.1.22) may likewise be improved through $\hat{d}_o = \hat{\lambda}_{o1} s_y^2 + \hat{\lambda}_{o2}\sigma_o^2$.

In general, it may be shown that for a given λ_2 and t, the estimator $d = \lambda_1 s_y^2 + \lambda_2 t$ would be better than $d' = s_y^2 + \lambda_2 t$

iff λ_1 lies between $2(n/\Delta)(1-\lambda_2 E s_y^2 t/\sigma_y^4)$ and 1.

In particular, the estimator $d_8^* = \lambda_1 s_y^2 - \alpha_1(\bar{x}-\bar{X})$ would be better than T_1' in (5.1.25), defined by Das and Tripathi (1978a),

iff λ_1 lies between $2(1/\Delta)\,[n+\alpha_1\mu_{21}(y,x)/\sigma_y^4]$ and 1.

The estimators T_2' and T_3' in (5.1.26) and (5.1.27) may also be, likewise, improved at least in some situations.

It may be noted that in case of normal populations, $T_{o1} = \sum_{i=1}^{n}(y_i-\bar{y})^2/(n+1)$, which would be better than s_y^2 and as well as the MLE $s_y^{*2} = \sum_{i=1}^{n}(y_i-\bar{y})^2/n$.

It may be shown, with the help of conditions (5.3.14) to (5.3.19) that the estimators

$$d_1^{**} = T_{o1} + [2/(n+1)]\sigma_{(1)}^2$$

$$d_3^{**} = T_{o1} + 2\bar{Y}_{(1)}[n/(n+1)][C_{(1)}^2/(n+C_{(1)}^2)]\ \bar{y}$$

$$d_4^{**} = T_{o1} + [(n+C_{(1)}^2)/(n+6C_{(1)}^2)][C_{(1)}^2/(n+1)]\ \bar{y}2$$

$$d_6^{**} = T_{o1} + \frac{6[(n-1)/(n+1)]\ C_{(1)}^2\ \bar{Y}_{(1)}^2\ c_y^2}{[10(n-1)C_{(1)}^2 + n(n+1)]}$$

would be better than s_y^2, s_y^{*2} and as well as T_{o1} in case of normal distributions.

5.5 Empirical Study

In this section the performances of various estimators under considerations have been studied on the basis of data for the natural population I of the Chapter III in Section 3.7. In this case also the same sample size $n = 30$ and same set of apriori values as in Chapter III have been taken.

The Table 5.5.1 gives the percentage relative efficiencies of the estimators from the class $d' = s_y^2 + \lambda_2' t$ over s_y^2 and the percentage relative efficiencies of the estimators from the class $d^* = T_{o1} + \lambda_2^* t$ over $T_{o1} = (n/\Delta)\, s_y^2$ are given in Table 5.5.2. λ_{o2}' in the Table 5.5.1 denotes the optimum choice of λ_2'(for a specified t) and λ_{o2}'' denotes the located values of λ_{o2}' using the apriori values of the involved parameters in Chapter III. Similarly λ_{o2}^* and λ_{o2}^{**} in the Table 5.5.2 denote the optimum choice of λ_2^* (for a specified t) and the located value of λ_{o2}^* respectively.

Table 5.5.1. Percentage relative efficiency of d' over s^2 for various choices of t and some values of λ_2'.

	λ_2'		Percentage relative efficiency of d' over s^2 $[v(s^2)/M(d')]\times 100$		Absolute relative bias	
	$\lambda_2'=\lambda_{o2}'$	$\lambda_2'=\lambda_{o2}''$	$\lambda_2'=\lambda_{o2}'$	$\lambda_2'=\lambda_{o2}''$	$\lambda_2'=\lambda_{o2}'$	$\lambda_2'=\lambda_{o2}''$
(1)	(2)	(3)	(4)	(5)	(6)	(7)
$\bar{y}$	-53.246	-64.00	101.728	101.643	0.0856	0.1029
$\bar{y}^2$	-.158	-.07	110.721	105.663	0.1771	0.0785
s^2	-.312	-.310	145.370	145.365	0.312	0.310
$s_y^2\bar{y}$	-.0005	-.0002	215.848	182.655	0.3418	0.1367

Table 5.3.2. Percentage relative efficiency of d^* over T_{o1} for various choices of t and some values of λ_2^*.

	λ'_2		Percentage relative efficiency of d^* over T_{o1} $[M(T_{o1})/M(d^*)]\times 100$		Absolute relative bias	
	$\lambda_2^* = \lambda_{o2}^*$	$\lambda_2^* = \lambda_{o2}^{**}$	$\lambda_2^* = \lambda_{o2}^*$	$\lambda_2^* = \lambda_{o2}^{**}$	$\lambda_2^* = \lambda_{o2}^*$	$\lambda_2^* = \lambda_{o2}^{**}$
(1)	(2)	(3)	(4)	(5)	(6)	(7)
$\bar{y}$	149.708	128.6	171.018	159.816	0.5552	0.5209
$\bar{y}^2$	.155	.06	140.239	114.991	0.4859	0.3794
$s_y^2\bar{y}$	-.00003	-.000039	104.037	101.037	0.2902	0.2836

CHAPTER VI

ESTIMATION OF POPULATION TOTAL USING APOSTERIORI AND APRIORI INFORMATION

6.1 Introduction and Summary

In many practical situations, the desired information for some units of the population in addition to those falling in the sample may be readily available or may be procured. This information may be available (a) even before the sample has been drawn or (b) after the sample has been selected and surveyed, partially or wholly. We shall call the first type of information as apriori information and the second as aposteriori one. The situation (a) may also be thought to include the case where a part of the population is enumerated completely and a sample is drawn from the remaining part.

The type of situation stated in (b) above may arise in many field work. For example, after the sample has been drawn, the information on the desired items may be available about some non-sampled units, along with the required information on sampled units. This information may be available through the sampled units or otherwise. Such a situation arises in many socio-economic surveys in rural areas, specially in India. In most of the rural areas in India, people of a particular community are usually clustered closely in a neighbourhood within a village and are very much close to each other with respect to many socio-economic characteristics. So, during the collection of information on the sampled households, the respondents often supply, in addition to the information for themselves, the information on required items for some of the neighbours who are not included in the sample.

Thus for example, if someone wants to study the costs and benefits of rural electrification, in a particular region, he may get information regarding the use of electricity etc., about some neighbouring or other households not included in the sample on the basis of information obtained from the sampled households. In fact such a situation may prevail in case of almost all the surveys to study the effect of village/area development projects being carried on in India. Similarly, in a hospital survey, many times it is found that some of the sampled hospitals also keep the recorded information about some of the neighbouring or other hospitals with the view of comparing their own medical facilities etc., with the others. Thus some of the sampled hospitals may furnish valid information, in addition to the required information for themselves, about some other hospitals which are not included in the sample.

Most of the villages in India are well-knitted units where some or many of the households keep various types of information (e.g., no. of cattles owned, land possessed, no. of orchards owned, the nature of social, economic and political activity etc.), about other households. Thus if such informative households fall in the sample, most of the times the required valid information may also be obtained about the non-sampled households through the sampled ones. Thus, in this case too, the situation cited in (b) above is realized.

Further, in practice, it may happen that after a sample has been drawn according to a specified sampling procedure and a part or the whole of the field work is over, the survey-statistician, because of one reason or the other, may decide to select some units from the remaining part of the population. He may select these units either purposively or using a conditional sampling design (conditioned) upon. the sampling design used for the original sample). For example, some of the sampled households may provide only a partial, but interesting information about some non-sampled households and the statistician may decide to choose these non-sampled households purposively, as the complete information for these households may be quite important from the point of view of the survey being conducted. This is another example where the aposteriori information is available.

As far as the availability of apriori information, as cited in (a) above, is concerned, it may be available or procured in many situations of practical interest, namely (i) the statistician may have an idea about some extreme values (very high or very low ones) associated with some of the population units. He may utilize this apriori information in designing the sample or estimating the parameters, (ii) the statistician may initially decide to make the complete enumeration and collect data for some of the population units. But latter, because of the limitation of budget, time, technical manpower, etc., he may change his decision of complete enumeration and decide to take a sample from remaining units to collect relevant information and thereby having data for some non-sampled units, as well as for sampled ones.

Thus, in many situations of practical interest, the information may be available or may be made to be available on some population units in addition to those which have been drawn into the sample. This information may be either aposteriori or apriori. Naturally, the question arises : whether we should throw away this information or use it in a suitable manner so as to improve usual estimators. Our view is that instead of throwing away this more or less easily available additional information, the statistician should try to find ways and means of using it, so as to obtain better results.

In this chapter, the methods of using the additional information set, say, $\{y_1, y_2, \ldots, y_k\}$ on k-units, in addition to those in the sample, have been discussed to improve the usual estimator for population total Y. We consider both the situations, namely, where aposteriori or apriori information is available. Some estimators in both the cases for estimating finite population total Y of a character y are proposed which are better, always or under some moderate conditions, than the usual unbiased estimators for Y, based on the information on the sample alone.

The case of using aposteriori information has been dealt under the structure of general sampling design considering a weighted estimator, based on the information contained in the sample and the information obtained aposteriori as well. Noting that the aposteriori information may be available to the statistician in various ways, we have dealt some particular cases depending upon the nature of conditional (conditioned by the original sample) sampling design according to which the set s_k, out of the non-sampled ones in the population, is realized. We identify situations, depending on the original as well as conditional sampling design under which the estimator, based on aposteriori information as well as the information from the sample, is better than the usual unbiased Horvitz-Thompson estimator for Y, based on the information contained in the sample alone.

The problem of utilising apriori information has been considered separately for the cases of SRSWOR, probabilities proportional to size with replacement (PPSWR) and varying probability selection without replacement (VPSWOR). The three estimators T_1, T_2 and T have been considered where T_1 is based on apriori information as well as a sample of size n out of the (N-k) units, T_2 is based on apriori information and on a sample of size (n-k) out of (N-k) units and T is based only on the sample of size n out of N units in entire population. We find that in case of SRSWOR and PPSWR, T_1 is better than both of T_2 and T and further T_2 is better than T under some moderate conditions and same is true in case of VPSWOR with π-P-S-scheme under a super-population model. On the basis of the findings, in case of SRSWOR, one may advocate, complete enumeration of those units which have extreme-high and extreme low values and the sampling from remaining units. In case of PPXWR, one may advocate the complete enumeration of those units which are expected to have extreme-low y-values, but have extreme-high x-values and the sampling from remaining units. At the end, an empirical study has been made to high light our results and the relative efficiencies of the proposed estimators over the usual unbiased estimator have been obtained.

6.2 A General Result on the Use of Aposteriori Information

Let $U = \{1,2,\ldots,N\}$ be a population of N(given) units labelled 1 to N and $S = \{s\}$, be the collection of all possible unordered samples s. Let $D(U,S,P)$ be a sampling design where P is the probability measure defined on s, i.e., $P(s) \geq 0$, $\sum_{s \in S} P(s) = 1$. Let y be a variate (a real-valued function defined on U) taking value y_i for the i^{th} unit of the population and let $Y = \sum_{i=1}^{N} y_i = N\bar{Y}$ be the total for y in the population. Let $s_k = \{1^*,2^*,\ldots,k(s)\}$, $k(s) = 1,2,\ldots,N-\nu(s)$ be the set of additional k units of the population which are known after the sample $s \in S$ has been drawn, according to a specified sampling design $D(U,S,P)$. Let $N^* = N - \nu(s)$ and $S^* = \{s_k|s\}$, $k = 1,2,3,\ldots,N^*$ be the collection of all possible sets which could be materialized from $U-s$, the population of the remaining N^* units after s has been drawn. Let $D^*(U-s, S^*, P^*)$ be a sampling design (conditional on the sample s realized) defined on S^*, with $P^*(s_k|s) \geq 0$ and $\sum_{s_k^* \in S^*} P(s_k^*|s) = 1$.

Basing on both of s and s_k, let the estimator d for population total Y be defined as

$$d = w \sum_{i \in s} y_i/\pi_i + (1-w)\, d^* \qquad (6.2.1)$$

where

$$d^* = \sum_{i \in s} y_i + \sum_{i \in s_k} y_i/\pi_{is},$$

$\sum_{i \in s}$ and $\sum_{i \in s_k}$ denoting sums over distinct units in s and s_k respectively.

$\pi_{is} = \sum_{s_k \ni i} P^*(s_k|s)$, denotes the conditional inclusion probability for a specified unit $i \varepsilon U-s$ to be included in s_k given that a particular sample $s \varepsilon S$ has been drawn.

$P^*(s_k|s)$ is the probability of the sample s_k (given that sample $s \varepsilon S$ has already been drawn), providing additional information on k units from U-S.

$\sum_{s_k \ni i}$ denotes the sum over all possible s_k (of sizes $k = 1,2,\ldots,N^*$), containing a specified unit i from U-S.

and w is a suitable chosen constant (weight) independent of s.

We assume that $\pi_{is} > 0$ for all $i = 1,2,\ldots,N^*$.

It may be noted that

$$E\left[\sum_{i\varepsilon s_k} (y_i/\pi_{is})|s\right] = \sum_{s_k \varepsilon S^*} \sum_{i \varepsilon s_k} (y_i/\pi_{is}) P^*(s_k|s) \tag{6.2.2}$$

$$= \sum_{i \varepsilon U-s} (y_i/\pi_{is}) \sum_{s_k \ni i} P^*(s_k|s) = Y - \sum_{i \varepsilon s} y_i$$

Thus d^* would be conditionally (conditioned on a given s) unbiased for Y and hence would be unbiased unconditionally too. It may be observed that the result (6.2.2) will not hold if s_k is purposively chosen. Further $\hat{Y}_{H-T} = \sum_{i \varepsilon s} y_i/\pi_i$ being unbiased for Y, the weighted d would be unbiased fo. Y.

Further,

$$V(d) = w^2 V(\hat{Y}_{H-T}) + (1-w)^2 V(d^*) + 2w(1-w)\,\mathrm{Cov}(\hat{Y}_{H-T}, d^*) \tag{6.2.3}$$

Now,

$$\mathrm{Cov}(\hat{Y}_{H-T}, d^*) = E_1 C_2(\hat{Y}_{H-T}, d^*) + C_1(E_2\ \hat{Y}_{H-T},\ E_2\ d^*) = 0$$

where E_2 and C_2 denote conditional (given the sample s) expectation and conditional covariance and E_1 and C_1 the unconditional ones

$$V(d^*) = V_1\ E_2(d^*) + E_1\ V_2(d^*) = E_1\ V_2(d^*)$$

where V_2 and V_1 denote conditional and unconditional variances.

Let $\sum_{\substack{s_k \ni (i,j) \\ i \neq j}} P^*(s_k|s) = \pi_{ijs}$, for $(i,j) \in U-s$.

Then

$$V_2(d^*) = E_2 \sum_{i \in s_k} y_i^2/\pi_{is}^2 + E_2 \sum_{\substack{(i,j) \in s_k \\ i \neq j}}\sum (y_i/\pi_{is})(y_j/\pi_{js})$$

$$+ (\sum_{i \in s} y_i)^2 + 2 \sum_{i \in s} y_i\ E_2 \sum_{i \in s_k} y_i/\pi_{is} - Y^2$$

$$= \sum_{i \in s} y_i^2 + \sum_{\substack{(i,j) \in s \\ i \neq j}}\sum y_i y_j + \sum_{s_k \in S^*} \sum_{i \in s_k} (y_i^2/\pi_{is}^2)\ P^*(s_k|s)$$

$$+ \sum_{s \in S^*} \sum_{\substack{(i,j) \in s_k \\ i \neq j}} (y_i/\pi_{is})\ P^*(s_k|s) + 2 \sum_{i \in s} y_i(Y - \sum_{i \in s} y_i) - Y^2$$

$$= 2Y \sum_{i \in s} y_i - (\sum_{i \in s} y_i^2 + \sum_{\substack{(i,j) \in s \\ i \neq j}} y_i y_j + \sum_{i \in s} y_i^2/\pi_{is}$$

$$+ \sum_{\substack{(i,j) \in s \\ i \neq j}}\sum (y_i/\pi_{is})(y_j/\pi_{js}) - Y^2 \qquad (6.2.4)$$

Let $\sum\limits_{s(\in S)\not\ni i}$, $\sum\limits_{\substack{s\in(S)\not\ni(i,j)\\ i\neq j}}$ denote the sum over all those samples s from S which do not contain a specified unit $i \in U$ and a pair (i, j), $i \neq j$ of units from U respectively.

Then

$$V(d^*) = E_1\ V_2(d^*) = 2Y \sum_{i=1}^{N} y_i \pi_i - \Big(\sum_{i=1}^{N} y_i^2 \pi_i + \sum_{i\neq j=1}^{N}\sum y_i y_j \pi_{ij}\Big)$$

$$+ \sum_{i=1}^{N} y_i^2 \sum_{s(\in S)\not\ni i} P(s)/\pi_{is} + \sum_{i\neq j=1}^{N}\sum y_i y_j \sum_{\substack{s(\in S)\not\ni(i,j)\\ i\neq j}} (\pi_{ijs}/\pi_{is}\pi_{js}) P(s) - Y^2 \tag{6.2.5}$$

From (6.2.3), we may write,

$$V(d) = [w^2\ V(\hat{Y}_{H-T}) + (1-w)^2\ V(d^*)] \tag{6.2.6}$$

where, $V(\hat{Y}_{H-T})$ is defined by (4.1.3).

The weight w which minimises $V(d)$ would be

$$w_o = V(d^*)/[V(d^*) + V(\hat{Y}_{H-T})] \tag{6.2.7}$$

and the resulting mean square error would be

$$V_o(d) = V(\hat{Y}_{H-T})\ V(d^*)/[V(\hat{Y}_{H-T}) + V(d^*)] \tag{6.2.8}$$

$$\leq V(\hat{Y}_{H-T})$$

and also,

$$V_o(d) \leq V(d^*)$$

It is to be noted that $w_o \in [0, 1]$, whatever be the population and sampling design. The values $w_o = 1$ and $w_o = 0$ which would occur due to $V(\hat{Y}_{H-T}) = 0$ and $V(d^*) = 0$ respectively, would be the extreme situations rarely to occur in practice. In these situations,

$V_o(d) = V(\hat{Y}_{H-T}) = 0$ and $V_o(d) = V(d^*) = 0$ respectively. In fact one should choose w in practice such that $0 < w < 1$. Thus in case $w_o \varepsilon (0,1)$ is known exactly, d_o will be always better than $\hat{Y}_{H-T}$; but practically, it may not be easy to know w_o exactly and hence, it may not be possible to use $d_o = w_o \hat{Y}_{H-T} + (1-w_o)d^*$ in practice.

Let $C(t)$ denote the coefficient of variation of a statistic t. In case w_o is not known exactly, but is expected to lie in some interval, even then a number of estimators from d better than $\hat{Y}_{H-T}$ may be found by using the following

Theorem 6.2.1. A necessary and sufficient condition for d in (6.2.1) to be better than $\hat{Y}_{H-T}$ would be

$$[C^2(d^*) - C^2(\hat{Y}_{H-T})]/[C^2(d^*) + C^2(\hat{Y}_{H-T})] < w < 1$$

i.e.,

$$2w_o - 1 < w < 1$$

and hence a sufficient condition for d to be better than $\hat{Y}_{H-T}$ would be

$$w_o^{(2)} \leq w < 1$$

where

$$w_o^{(2)} = C^2_{(2)}(d^*)/[C^2_{(2)}(d^*) + C^2_{(1)}(\hat{Y}_{H-T})]$$

$C^2_{(2)}(d^*)$ and $C^2_{(1)}(\hat{Y}_{H-T})$ being known quantities such that

$$C^2(d^*) \leq C^2_{(2)}(d^*) \text{ and } C^2_{(1)}(\hat{Y}_{H-T}) \leq C^2(\hat{Y}_{H-T}).$$

Proof. From (6.2.6), it may be shown that

$$[V(\hat{Y}_{H-T})-V(d)] = (1-w)Y^2[w\{C^2(d^*)+C^2(\hat{Y}_{H-T})\}-\{C^2(d^*)-C^2(\hat{Y}_{H-T})\}]. \tag{6.2.9}$$

Thus $V(d) < V(\hat{Y}_{H-T})$

iff $[C^2(d^*) - C^2(\hat{Y}_{H-T})]/[C^2(d^*) + C^2(\hat{Y}_{H-T})] < w < 1$

i.e., iff $2w_o - 1 < w < 1$

Further since,

$$C^2_{(1)}(\hat{Y}_{H-T}) \leq C^2(\hat{Y}_{H-T}) \quad \text{and} \quad 1/C^2_{(2)}(d^*) \leq 1/C^2(d^*)$$

$$\Rightarrow 1 + [C^2_{(1)}(\hat{Y}_{H-T})/C^2_{(2)}(d^*)] \leq 1 + [C^2(\hat{Y}_{H-T})/C^2(d^*)]$$

$$\Rightarrow 1/w_o^{(2)} \leq 1/w_o$$

$$\Rightarrow w_o \leq w_o^{(2)}$$

Thus the sufficient condition follows.

Remarks

(i) The above Theorem obviously identifies the situations in which the use of aposteriori information in d would provide us with better results compared to the results when information only on the original sample s is used.

(ii) Let $C^2(d^*) = \alpha . C^2(\hat{Y}_{H-T})$; then d will always be better than $\hat{Y}_{H-T}$ for all w satisfying

$$(\alpha-1)/(\alpha+1) < w < 1.$$

(iii) Let $\alpha_{(2)}$ be a quantity such that $\alpha \leq \alpha_{(2)}$. Then using the above remark, we find that for all w with $(\alpha_{(2)} - 1)/(\alpha_{(2)} + 1) < w < 1$, d would be better than $\hat{Y}_{H-T}$.

(iv) In case, $V(d^*) \leq V(\hat{Y}_{H-T})$ i.e., $\alpha \leq 1$ (i.e., $w_o \leq 1/2$), the proposed estimator d will always be better than $\hat{Y}_{H-T}$ for all $w \in (0,1)$. In case $V(d^*) > V(\hat{Y}_{H-T})$, (i.e., $\alpha > 1$, $w_o > 1/2$) and $\alpha_{(2)}$ (> α) is a good-guessed value of α, then for w such that $\alpha_{(2)}/(1+\alpha_{(2)}) < w < 1$, d would be better than $\hat{Y}_{H-T}$.

(v) It may be noted that d would be better than d^* iff $0 < w < 2w_o$. In case, $w_o \leq 1/2$, one should choose w such that $0 < w < 2w_o$ and in case, $w_o > 1/2$, one should choose w such that $2w_o - 1 < w < 1$, so that d would be simultaneously better than both of $\hat{Y}_{H-T}$ and d^*.

(vi) The relative efficiency $V(\hat{Y}_{H-T})/V_o(d)$ of d with $w = w_o$ over $\hat{Y}_{H-T}$ is found to be $1/w_o = 1 + 1/\alpha$ which is obviously a monotonically decreasing function.

The following Table 6.2.1 gives the percent relative efficiency of d_o (i.e., d with $w = w_o = \alpha/(1+\alpha)$) over $\hat{Y}_{H-T}$.

Table 6.2.1. Percent relative efficiency of d_o over $\hat{Y}_{H-T}$.

α	w_o	$[V(\hat{Y}_{H-T})/V(d_o)]\times 100$	α	w_o	$[V(\hat{Y}_{H-T})/V(d_o)]\times 100$
(1)	(2)	(3)	(4)	(5)	(6)
0.1	0.0909	1100	2.0	0.3333	150
0.2	0.1666	600	3.0	0.7500	133
0.3	0.2307	433	4.0	0.8000	125
0.4	0.2857	350	5.0	0.8333	120
0.5	0.3333	300	6.0	0.8571	117
0.6	0.3750	266	7.0	0.8750	114
0.7	0.4117	243	8.0	0.8888	113
0.8	0.4444	225	9.0	0.9000	111
0.9	0.4736	211	10.0	0.9090	110
1.0	0.5000	200			

6.3 Use of Different Types of Aposteriori Information

After a sample s (consisting of distinct units) of fixed size n has been selected and canvassed, the possibility of getting the additional set of units s_k is that s_k may contain either only one or two or so on upto (N-n) units. However, in practice, it is very difficult, specially in case of large populations, to realize all the remaining (N-n) units into the additional set s_k. Hence, without loss of generality, we can assume that k may be any number from 1 to N^* which, in turn, of course, may be (N-n) itself. However, for obtaining mean square error and estimates etc., k must be greater or equal to two. Thus, we assume that $2 \leq k \leq N^* \leq (N-n)$.

In this section, we consider the following two types of conditional probability, namely,

$$\begin{aligned} &(a) \quad P^*(s_k|s) = \text{Constant for all } s_k \varepsilon S^* \\ \text{and } &(b) \quad P^*(s_k|s) \neq \text{Constant for all } s_k \varepsilon S^* \end{aligned} \qquad (6.3.1)$$

6.3.1 Since in many situations, it may be very difficult to find out which of the additional sets of sizes $2,3,\ldots,N^*$ will be preferable over the others, we assume that after the sample s of size n has been drawn, all of the sets of sizes $k(s) = 2,3,\ldots,N^*$ (from remaining N-n population units) have equal chance of being realized as the additional set s_k and that all sets of sizes 1, (N^*+1), (N^*+2), ..., $(N-n)$ have zero chance of occurrence as s_k. Obviously, total number of sets of different sizes $2,3,\ldots,N^*$ from the set of (N-n) units would be

$$B = \binom{N-n}{2}+\binom{N-n}{3}+\ldots+\binom{N-n}{N^*} = 2^{N-n}-1-(N-n)-\sum_{r=N^*+1}^{N-n}\binom{N-n}{r} \tag{6.3.2}$$

where $\sum_{r=N^*+1}^{N-n}\binom{N-n}{r} = 0$ if $N^* = N-n$.

Now we obtain various probabilities under the above situation:

Hence the probability to be associated with each of the possible sets $s_k\{k=2,\ldots N^*\}$, given that the sample s has been drawn, would be

$$P^*(s_k|s)=\begin{cases}1/B & \text{for} \quad k(s) = 2,3,\ldots,N^* \\ 0 & \text{for} \quad k(s) = 1,\ N^*+1,\ldots,N-n\end{cases} \tag{6.3.3}$$

and hence,

for $i \in U-s$: $\pi_{is} = \sum_{s_k \ni i} P^*(s_k|s) = \sum_{k=2}^{N^*}\binom{N-n-1}{k-1}/B = L/B$ say,

for $(i,j) \in U-s$, $i \neq j$: $\pi_{ijs} = \sum_{s_k \ni (i,j),\, i\neq j} P^*(s_k|s) = \sum_{k=2}^{N^*}\binom{N-n-2}{k-2}/B = M/B$ say

$$\sum_{s(\in S)\not\ni i} P(s)/\pi_{is} = (B/L)\sum_{s(\in S)\not\ni i} P(s) = (B/L)(1-\pi_i) \tag{6.3.4}$$

$$\pi_{ijs}/\pi_{is}\pi_{js} = B\ M/L^2$$

and $$\Sigma(\pi_{ijs}/\pi_{is}\pi_{js})\ P(s) = (BM/L^2)\sum_{s(\in S)\not\ni(i,j),\, i\neq j} P(s)=(1-\pi_{ij})BM/L^2 \tag{6.3.5}$$

where $L = \sum_{k=2}^{N^*}\binom{N-n-1}{k-1}$ and $M = \sum_{k=2}^{N^*}\binom{N-n-2}{k-2}$.

Substituting the above values and π_i, π_{ij} based on the design through which the original sample has been generated, one can obtain the expressions for $V(d^*)$, $V(d)$ etc., from (6.2.5) to (6.2.8).

From (6.2.5), (6.3.4) and (6.3.5), it may be shown that,

$$V(d^*) = (1+B/L)\sum_{i=1}^{N}(1-\pi_i)y_i^2+(1+BM/L^2)\sum_{i\neq j=1}^{N}\sum(1-\pi_{ij})y_iy_j+2Y\sum_{i=i}^{N}y_i\pi_i-2Y^2 \quad (6.3.6)$$

$$= V(\hat{Y}_{H-T}) + \sum_{i=1}^{N}\{(1-\pi_i)\{1 + B/L) - \frac{1}{\pi_i}\}y_i^2$$

$$+ \sum_{i\neq j=1}^{N}\sum\{(1-\pi_{ij})(1+BM/L^2) - \frac{\pi_{ij}}{\pi_i\pi_j}\} y_iy_j + Y\sum_{i=1}^{N}(2\pi_i-1)y_i$$

$$= V(\hat{Y}_{H-T}) + \sum_{i=1}^{N}\{(1-\pi_i)(1+B/L) - \frac{1}{\pi_i} + 2\pi_i-1\} y_i^2$$

$$+ \sum_{i\neq j=1}^{N}\sum\{(1-\pi_{ij})(1+BM/L^2) - \frac{\pi_{ij}}{\pi_i\pi_j} + 2\pi_j - 1\} y_iy_j$$

$$= V(\hat{Y}_{H-T}) + \delta \quad \text{(say)}.$$

Let $y_i > 0$, then a set of sufficient conditions for $V(d^*) \leq V(\hat{Y}_{H-T})$ would be

$$(1-\pi_i)(1+B/L) - \frac{1}{\pi_i} + 2\pi_i - 1 \leq 0 \quad \text{for } i = 1,2,\ldots,N \quad (6.3.7)$$

$$\text{and} \quad (1-\pi_{ij})(1+BM/L^2) - \frac{\pi_{ij}}{\pi_i\pi_j} + 2\pi_j - 1 \leq 0 \quad \text{for } i,j=1,2,\ldots,N \; (i\neq j)$$

which on further simplification reduces to

$$\pi_i \leq L/(B-L) \quad \text{for } i = 1,2,\ldots,N \quad (6.3.8)$$

$$\text{and} \quad \pi_{ij} \geq \frac{(BM/L^2)+2\pi_j}{1+(BM/L^2)+(1/\pi_i\pi_j)} \quad \text{for } i,j = 1,2,\ldots,N \quad (6.3.9)$$

Then from remark (iv) of Section 6.2., d with all $w \in (0,1)$ would be better than $\hat{Y}_{H-T}$ in case of the class of designs $\mathcal{D}$ characterized by (6.3.8) and (6.3.9).

The above class of designs $\mathcal{D}$ characterized by (6.3.8) and (6.3.9) would be non-empty

$$\text{if} \quad L \geq B/2.$$

In general, in case of all those designs which result $V(d^*) \leq V(\hat{Y}_{H-T})$, the use of aposteriori information in d with $w \in (0,1)$ will always result into increased precision. Further, following remark (iv) of Section 6.2, one may again choose w suitably so that use of aposteriori information in d results into increased precision even in case of those designs which result into $V(\hat{Y}_{H-T}) < V(d^*)$.

Let $\pi^* = \min_{1 \leq i \leq N} \pi_i$, $\pi^{**} = \min_{\substack{1 \leq i,j \leq N \\ i \neq j}} \pi_{ij}$. It is to be noted that, in addition to other situations that

$$\begin{aligned}
C^2(d^*) &= V(d^*)/Y^2 \\
&= (1+B/L)\sum_{i=1}^{N}(1-\pi_i)(y_i/Y)^2+(1+BM/L^2)\sum_{i\neq j=1}^{N}\sum(1-\pi_{ij})(y_i/Y)(y_j/Y) \\
&\quad + 2\sum_{i=1}^{N}(y_i/Y)\pi_i-2 \\
&< (1-\pi^*)(1+B/L)\sum_{i=1}^{N}(y_i/Y)^2+(1-\pi^{**})(1+BM/L^2)\sum_{i\neq j=1}^{N}\sum(y_i/Y)(y_j/Y) \\
&= (1-\pi^{**})(1+BM/L^2)+\{(1-\pi^*)(1+B/L)-(1+BM/L^2)(1-\pi^{**})\sum_{i=1}^{N}\frac{y_i^2}{Y^2} \\
&\quad + 2(\pi_M-1) \\
&= g_1 + 2(\pi_M-1) + g_2(1+C_y^2)/N, \qquad (6.3.10)
\end{aligned}$$

where $\pi_M = \max_{1 \le i \le N} \pi_i$,

$$g_1 = (1-\pi^{**})(1+BM/L^2)$$

and $g_2 = (1+\pi^*)(1+B/L) - (1+BM/L^2)(1-\pi^{**})$

Let

$$D = \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{\pi_{ij}}{\pi_i \pi_j} (y_i/Y)(y_j/Y)$$

be the same as defined in (4.3.2), and let $D_{(1)}$ be the same as defined in (4.3.4) with $1 < D_{(1)} \le D = 1 + C^2(\hat{Y}_{H-T})$.

Let $\alpha = C^2(d^*)/C^2(\hat{Y}_{H-T})$. Then using (6.3.10),

$$\alpha < [g_1 + 2(\pi_M - 1) + g_2(1+C_y^2)/N]/[D_{(1)} - 1] = \alpha^* \text{ say,} \qquad (6.3.11)$$
$$< \alpha_{(2)}$$

where $\alpha_{(2)}$ is α^* with $C_y^2 = C_{(1)}^2$ in case $g_2 < 0$ or with $C_y^2 = C_{(2)}^2$ in case $g_2 > 0$. Then from remark (iii) of Section 6.3.2, d in (6.2.1) would be better than $\hat{Y}_{H-T}$ for all w satisfying

$$(\alpha_{(2)} - 1)/(\alpha_{(2)} + 1) \le w \le 1. \qquad (6.3.12)$$

<u>In case of SRSWOR</u>,

$$V(d^*) = N^2 K \sigma_y^2 + Np(\sigma_y^2 + \bar{Y}^2) + Nq\{(N-1)\bar{Y}^2 - \sigma_y^2\}$$
$$= N\bar{Y}^2[C_y^2(NK+p-q) + (N-1)q] \qquad (6.3.13)$$

$$p = (1-f)(1+B/L) - \frac{1}{f} + 2f - 1$$

$$q = \left(1 - \frac{n(n-1)}{N(N-1)}\right)(1+BM/L^2) - \frac{N(n-1)}{n(N-1)} + 2f - 1$$

and $K = (N-n)/n(N-1)$

and $V(\hat{Y}_{H-T}) = N^2K\ \bar{Y}^2\ C_y^2$. Thus

$$\begin{aligned}\alpha &= [C_y^2(NK + p - q) + p + (N-1)q]/NKC_y^2\\ &= g_1' + (g_2'/C_y^2)\\ &= \alpha^*_{(2)} \leq \alpha_{(2)}\end{aligned}$$

where,

$$g_1' = (NK + p - q)/NK,$$

$$g_2' = [p + (N-1)q]/NK$$

$$\text{and}\quad \alpha_2 = \begin{cases}\alpha_2^* \text{ with } C_y^2 \text{ replaced by } C_{(1)}^2 \text{ in case } g_2' > 0\\ \alpha_2^* \text{ with } C_y^2 \text{ replaced by } C_{(2)}^2 \text{ in case } g_2' < 0\end{cases}$$

thus d would be better than $\hat{Y}_{H-T}$ for all w satisfying

$$(\alpha_{(2)}-1)/(\alpha_{(2)}+1) \leq w < 1.$$

<u>6.3.2</u> As in Section 6.3.1, we shall assume that the sets s_k, for $k = 1, N^* + 1, N^*+2, \ldots, N-n$ have zero chance of being realized. After a sample of size n has been drawn from U according to a specified design, in many situations, all the sets of different sizes $k(s) = 2,3,\ldots,N^*$ consisting of the units $i \in U-s$ out of the remaining $N-n$ units may not be equally likely to be realized as an additional set s_k. In such a situation, let all the sets of a given size $k(s) = k_o$ be equally likely, but the sets of different sizes $k_o = 2,3,\ldots,N^*$ are allowed to have different probabilities. Thus, if s_{k_o} denotes such a set from all possible $\binom{N-n}{k_o}$ sets of size k_o, we assume that

$$P^*(s_{k_o}|s, k=k_o) = 1/\binom{N-n}{k_o},\ k_o = 2,3,\ldots,N^*.$$

Let $\phi(k_o) = P^*(k=k_o|s) = P^*(S_{k_o}|s)$ be the probability of getting the class $S_{k_o} = \{s_{k_o}\}$ of $\binom{N-n}{k_o}$ sets of size k_o each ($k_o = 2,3,\ldots,N^*$). Obviously, $0 < \phi(k_o) < 1$ and $\sum_{k_o=2}^{N^*} \phi(k_o)=1$ and $\phi(k_o)$ is likely to be a monotonically decreasing function of k_o.

Now, it is on the part of the statistician to associate a particular probability say $\phi(k_o)$ based on the experience from the field and survey-situations etc. to the class S_{k_o} of all sets of size k_o.

Since, all the sets of a given size k_o are equally likely, we have

$$P^*(s_{k_o}|s) = \begin{cases} P^*(k=k_o,\ s_k|s) = \phi(k_o)/\binom{N-n}{k_o} \text{ for a given } k_o, \\ \quad \text{for } s_{k_o} \in S_{k_o},\ k_o = 2,3,\ldots,N^* \\ 0 \quad \text{for } k^* = 1,\ N^*+1,\ldots,N-n \end{cases} \tag{6.3.14}$$

$$\pi_{is} = \sum_{s_{k_o} \ni i} P^*(s_{k_o}|s) = \sum_{k_o=2}^{N^*} \phi(k_o).k_o/(N-n)$$

and
$$\pi_{ijs} = \sum_{\substack{s_{k_o} \ni (i,j) \\ i \neq j}} P^*(s_{k_o}|s) = \sum_{k_o=2}^{N^*} \phi(k_o).k_o(k_o-1)/(N-n)(N-n-1)$$

Further, we have, following the steps as in Section 6.3.1,

$$\sum_{s(\epsilon S) \not\ni i} P(s)/\pi_{is} = a(1-\pi_i)$$

and
$$\sum_{\substack{s(\epsilon S)\ (i,j) \\ i \neq j}} P(s) \frac{\pi_{ijs}}{\pi_{is}\ \pi_{js}} = b(1-\pi_{ij})$$

where, $a = (N-n) / \sum_{k_o=2}^{N^*} k_o \phi(k_o)$ and $b = \frac{(N-n)}{(N-n-1)} \frac{\sum_{k_o=2}^{N^*} k_o(k_o-1)\phi(k_o)}{\left(\sum_{k_o=2}^{N^*} k_o \phi(k_o)\right)^2}$

It may also be shown that

$$V(d^*) = V(\hat{Y}_{H-T}) + \sum_{i=1}^{N} y_i^2 \{(a-1)(1-\pi_i) - \frac{1}{\pi_i} + 1\}$$

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

Now proceeding as in 6.3.1, we can find the range for w in d so that d would be better than $\hat{Y}_{H-T}$.

6.4 Various Ways of Using Apriori Information

Let y-values for $k(\geq 2)$ units of the population be known apriori before the sample has been selected and the observations are made. Without loss of generality, let these values be denoted by $y_1, y_2, \ldots, y_k$. In this section, we consider the use of these known k-values in case of SRSWOR, probability proportional to size with replacement (PPSWR), varying probability selection without replacement (VPSWOR).

We may write

$$Y = \sum_{i=1}^{k} y_i + \sum_{i=k+1}^{N} y_i \qquad (6.4.1)$$

where the first sum $Y_k = k\bar{Y}_k$ is known and the second sum $Y_{n-k} = (N-k)\bar{Y}_{n-k}$ is to be estimated on the basis of a sample from the unknown part $U(N-k)$ of $(N-k)$ units in the population.

Let d_1 and d_2 denote any unbiased estimators for Y_{n-k} based on any sampling design and the sample of sizes n and $(n-k)$ respectively from $U(N-k)$. Obviously, the estimators

$$T_1 = Y_k + d_1 \quad \text{and} \quad T_2 = Y_k + d_2$$

would be unbiased for Y. But it can not be claimed that $V(T_1) \leq V(T_2)$ or the reverse for all sampling designs. However, if definitions of d_1 and d_2 are similar and $V(d_1)$, $V(d_2)$ are decreasing functions of sample size, then one would get $V(T_1) \leq V(T_2)$. Further if T is an unbiased estimator of Y based on a sample of size n from entire population and defined similar to d_1 and d_2, then it can not be claimed in general that $V(T_1) \leq V(T)$ and/or $V(T_2) \leq V(T)$. However, the former may hold true in many situations while the latter need not be. To get a clear picture of related behaviours of the estimators T_1, T_2 and T, we shall consider the cases of SRSWOR, PPSWR, VPSWOR.

6.4.1 Use of Apriori Information in case of SRSWOR

Let $d_1 = (N-k)\bar{y}_n$ and $d_2 = (N-k)\bar{y}_{n-k}$ be unbiased estimators of Y_{n-k} based on simple random samples (without replacement) of sizes n and $(n-k)$ respectively drawn from the $(N-k)$ units of the population for which y-values are not known. Ovbiously,

$$T_1 = Y_k + d_1 \quad \text{and} \quad T_2 = Y_k + d_2 \qquad (6.4.2)$$

are unbiased estimateds of Y. Let $T = N\bar{y}_n$ be the usual unbiased estimator based on a simple random sample without replacement of size

n from the entire population. The estimators T_1 and T_2 would be same as usual stratified unbiased estimators in case the whole population is thought of divided into two strata, the first consisting of these k-units (those for which y-values are known) and the second consisting of remaining $(N-k)$ units and the first strata being enumerated completely while SRSWOR of sizes n and $(n-k)$ respectively are being drawn from the second. The estimator T being that where population is not stratified.

Next we have the following

Theorem 6.4.1

(i) $V(T_1) < V(T_2)$

(ii) $V(T_1) < V(T)$

and

(iii) $V(T_2) < V(T)$ iff $G^* + [\sum_{i=1}^{k} (y_i - \bar{Y}_k)^2 / \sum_{i=k+1}^{N} (y_i - \bar{Y}_{n-k})^2] \geq \frac{k}{(n-k)} \times [1 + \frac{n}{k(N-k-1)}]$

where $G^* = \frac{[k(\bar{Y}_k - \bar{Y})^2 + (N-k)(\bar{Y}_{N-k} - \bar{Y})^2]}{(N-k-1)\, S_{N-k}^2}$,

$$S_{N-k}^2 = \sum_{i=k+1}^{N} (y_i - \bar{Y}_{N-k})^2/(N-k-1).$$

Proof. We have,

$$V(T_1) = (N-k)\ (N-k-n)\ S_{N-k}^2/n \qquad (6.4.3)$$

$$V(T_2) = (N-n)(N-k)\ S_{N-k}^2/(n-k) \qquad (6.4.4)$$

and

$$V(T) = N(N-n)\ S_N^2/n. \qquad (6.4.5)$$

Noting that,

$$(N-1)s_N^2 = (k-1)s_k^2+(N-k-1)s_{N-k}^2+k(\bar{Y}_k-\bar{Y})^2+(N-k)(\bar{Y}_{N-k}-\bar{Y})^2 \qquad (6.4.6)$$

It may be shown that,

$$V(T) = [N(N-n)/n(N-1)][(k-1)s_k^2+(1+G^*)(N-k-1)s_{N-k}^2] \qquad (6.4.7)$$

where $s_k^2 = \sum_{i=1}^{k}(y_i-\bar{Y}_k)^2/(k-1)$.

From (6.4.3) and (6.4.4), it is obvious that $V(T_1) < V(T_2)$. Next from (6.4.3) and (6.4.7), we have

$$\begin{aligned} V(T)-V(T_1) &= \frac{N(N-n)}{n(N-1)}[(k-1)s_k^2+(1+G^*)(N-k-1)s_{N-k}^2] - \frac{(N-k)(N-k-n)}{n} s_{N-k}^2 \\ &= \frac{N(N-n)(k-1)}{n(N-1)}s_k^2+\frac{s_{N-k}^2}{n(N-1)}[N(N-n)(N-k-1)-(N-1)(N-k)(N-k-n) \\ &\qquad + G^*N(N-n)(N-k-1)] \\ &= \frac{N(N-n)(k-1)}{n(N-1)}s_k^2+\frac{s_{N-k}^2}{n(N-1)}[Nk(N-k-2)+k^2+nk+G^*(N-n)N(N-k-1)] \\ &> 0 \end{aligned}$$

since the definition of T_1 and s_{N-k}^2 requires that $N > k+1$, i.e., $N \geq k+2$.

From (6.4.4) and (6.4.7), it is found that

$$V(T) \geq V(T_2)$$

$$\text{iff } \frac{N}{n(N-1)}[(k-1)s_k^2/s_{N-k}^2 + (1+G^*)(N-k-1)] \geq \frac{N-k}{n-k}$$

$$\text{i.e., iff } \frac{N}{(N-1)}\frac{(N-k-1)}{n}[\{\sum_{i=1}^{k}(y_i-\bar{Y}_k)^2/\sum_{i=k+1}^{N}(y_i-\bar{Y}_{N-k})^2\}+G^*+1] \geq \frac{N-k}{n-k}$$

$$\text{i.e., iff } G^*+\{\sum_{i=1}^{k}(y_i-\bar{Y}_k)^2/\sum_{i=k+1}^{N}(y_i-\bar{Y}_{N-k})^2\} \geq \frac{(N-k)}{(N-k-1)}\cdot\frac{(N-1)}{N}\cdot\frac{n}{(n-k)}-1 \qquad (6.4.8)$$

Now, under the assumption, $(N-1)/N \simeq 1$, the R.H.S. of (6.4.8) becomes

$$[k/(n-k)][1+\{n/k(N-k-1)\}] \text{ and hence}$$

$$V(T_2) \leq V(T)$$

$$\text{iff } G^{*}+\{\sum_{i=1}^{k}(y_i-\bar{Y}_k)^2/\sum_{i=k+1}^{N}(y_i-\bar{Y}_{N-k})^2\} \geq [k/(n-k)][1+\{n/k(N-k-1)\}].$$

Remarks

(i) We note that the use of apriori information in the form of using T_2 is preferable over T not always, but only in some situations. From (6.4.4) and (6.4.5), one notes that in case k is not very large and variation within $\{y_{k+1},\ldots,y_N\}$ is small compared to over all variation in the population, then T_2 would be better than T.

(ii) Noting that $\bar{Y}_k = \bar{Y}_{N-k} \Longleftrightarrow \bar{Y}_k = \bar{Y}_{N-k} - \bar{Y} \Longleftrightarrow G^{*} = 0$, in case $y_1, y_2, \ldots, y_k$ do not differ much from $y_{k+1}, y_{k+2}, \ldots, y_N$ (i.e., in case the population is more or less homogeneous in y-variables), T is expected to be better than T_2.

(iii) If mean square between strata, viz., $k(\bar{Y}_k-\bar{Y})^2 + (N-k)(\bar{Y}_{N-k}-\bar{Y})^2$ is appreciably large compared to the mean square S^2_{N-k} with the second strata, similar to the usual requirement need for appreciable gains due to stratification over simple random sampling, then T_2 would fare well compared to T. It is interesting to note that in such a situation, large value for

s_k^2, perhaps contrary to the usual requirement that strata should be more or less homogeneous within, would further increase the precision of T_2 over T.

(iv) From Theorem 6.4.1, one finds that if k is not very large and variation within $\{y_1, y_2, \ldots, y_k\}$ is larger than that within $\{y_{k+1}, y_{k+2}, \ldots, y_N\}$, then T_2 would be better than T. Such a situation would be realized when either $y_1, y_2, \ldots, y_k$ are quite large or $y_1, y_2, \ldots, y_k$ are relatively small compared to $y_{k+1}, \ldots, y_N$, but differing appreciably from each other (so that variation within them is large) while $y_{k+1}, y_{k+2}, \ldots, y_N$ do not differ much from each other.

(v) If the known k-values are extreme values, preferably consisting of both the extreme-high as well as extreme-low values out of $\{y_1, y_2, \ldots, y_N\}$, then T_2 may be very well expected to be better than T, even if the mean square between the two strata is quite small. Thus one may advocate to enumerate completely all those population units which have extreme y-values (the largest and smallest ones) and to take a sample from the remaining population units which are more or less homogeneous.

6.4.2 Use of Apriori Information in Case of PPSWR

Let $d_1 = \frac{1}{n} \sum_{i=1}^{n} y_i/p_i'$ and $d_2 = \frac{1}{n-k} \sum_{i=1}^{n-k} y_i/p_i'$, where

$p_i' = x_i/X_{N-k}$, $i = k+1, k+2, \ldots, N$, $X_{N-k} = \sum_{i=k+1}^{N} x_i$ be unbiased estimates of Y_{N-k} based on probability proportional to size (x) and with replacement (PPSWR) of sizes n and n-k respectively drawn from those (N-k) units of the population for which y-values are not known. Obviously,

$$T_1 = Y_k + d_1 \quad \text{and} \quad T_2 = Y_k + d_2 \tag{6.4.9}$$

are unbiased estimators of Y.

Let $T = \frac{1}{n} \sum_{i=1}^{n} y_i/p_i$, $p_i = x_i/X$, $i = 1,2,\ldots,N$, $X = \sum_{i=1}^{N} x_i$ be usual unbiased estimator for Y based on a PPSWR-sample of size n from the entire population. Now, we have the following

<u>Theorem 6.4.2</u>

(i) $V(T_1) < V(T_2)$

(ii) $V(T_1) < V(T)$

(iii) $V(T_2) < V(T)$, if $X_k/X_{N-k} > k/(n-k)$.

<u>Proof</u>. The result (i) follows trivially by observing that

$$V(T_1) = [\sum_{i=k+1}^{N} y_i^2/p_i' - Y_{N-k}^2]/n \tag{6.4.10}$$

and $$V(T_2) = [\sum_{i=k+1}^{N} y_i^2/p_i' - Y_{N-k}^2]/(n-k) \tag{6.4.11}$$

(ii) we have,

$$V(T) = [\sum_{i=1}^{N} y_i^2/p_i - Y^2]/n \tag{6.4.12}$$

and thus

$$n[V(T)-V(T_1)] = \sum_{i=1}^{N} y_i^2/p_i - Y^2 - \sum_{i=k+1}^{N} y_i^2/p_i' + Y_{N-k}^2 .$$

Using the relation,

$$\sum_{i=1}^{N} y_i^2/p_i = \sum_{i=1}^{k} y_i^2/p_i + \sum_{i=k+1}^{N} (y_i^2/p_i')(X/X_{N-k}),$$

we obtain

$$n[V(T)-V(T_1)] = \sum_{i=1}^{k} y_i^2(X_k/x_i) + \sum_{i=1}^{k} y_i^2(X_{N-k}/x_i) + \sum_{i=k+1}^{N} y_i^2(X_k/x_i)$$

$$-(Y_k+Y_{N-k})^2 + Y_{N-k}^2$$

$$= v_{p'}(y) + X_{N-k} \sum_{i=1}^{k} y_i^2/x_i + X_k \sum_{i=k+1}^{N} y_i/x_i - 2Y_kY_{N-k}$$

$$= v_{p'}(y) + \frac{X_{N-k}}{X_k} \sum_{i=1}^{k} y_i^2 \frac{X_k}{x_i} + \frac{X_k}{X_{N-k}} \sum_{i=k+1}^{N} y_i^2 \frac{X_{N-k}}{x_i}$$

$$- 2Y_k Y_{N-k}$$

$$= Q + \frac{X_{N-k}}{X_k} Y_k^2 + \frac{X_k}{X_{N-k}} Y_{N-k}^2 - 2Y_k Y_{N-k}$$

$$= Q + [(Y_k X_{N-k}-X_k Y_{N-k})^2/X_N X_{N-k}] \qquad (6.4.13)$$

$$\geq 0$$

where $v_{p'}(y) = [\sum_{i=1}^{k} y_i^2 X_k/x_i - Y_k^2] > 0$, and

$$Q = (X/X_k)(\sum_{i=1}^{k} y_i^2 \frac{X_k}{x_i} - Y_k^2) + (X_k/X_{N-k})(\sum_{i=k+1}^{N} y_i^2 \frac{X_{N-k}}{x_i} - Y_{N-k}^2) > 0.$$

(iii) From (6.4.10) to (6.4.12), we have

$$n[V(T)-V(T_2)] = Q + \{(Y_kX_{N-k}-X_kY_{N-k})^2/X_kX_{N-k}\} + X_{N-k} \sum_{i=k+1}^{N} y_i^2/x_i - Y_{N-k}^2$$

$$- \frac{n}{n-k} [\sum_{i=k+1}^{N} y_i^2 \frac{X_{N-k}}{x_i} - Y_{N-k}^2]$$

$$= (X/X_k)\{\sum_{i=1}^{k} y_i^2 X_k/x_i - Y_k^2\} + (\frac{X_k}{X_{N-k}} - \frac{k}{N-k})(\sum_{i=k+1}^{N} \frac{y_i^2 X_{N-k}}{x_i} - Y_{N-k}^2)$$

$$+ (Y_k X_{N-k}-X_k Y_{N-k})^2/X_k X_{N-k}$$

and hence a sufficient condition for $V(T_2) < V(T)$ is $X_k/X_{N-k} > k/(n-k)$.

Remarks

(i) From (6.4.13), we note that in case $y_i = x_i$ for $i = 1,2,\ldots,N$, the estimators T and T_1 would be equally precise. In other situations, T_1 is always preferable over T with appreciable gains in case y_i's depart considerably from the relation $y_i = x_i$. This indicates that if one enumerates those k-units completely for which y-values are quite further away from the relation $y_i = x_i$, $i = 1,2,\ldots,k$ and takes a sample from the remaining units, the gain by using T_1 would be quite considerable over T.

(ii) From (6.4.14), we note that in case $y_i = x_i$, T_2 and T would be equally precise. In most of the other situations, T_2 would be preferable over the others.

(iii) In many situations of practical importance $k \leq n/2$. In such situations, $X_k \geq X_{N-k}$ would guarantee the superiority of T_2 over T. In other situations too, T_2 would be preferable over T provided the k-units for which y-values are known are those which have high x-values compared to x-values for remaining population units. It is also noted that further the values $\{y_1, y_2, \ldots, y_k\}$ depart away from the relation $y_i = x_i$, higher would be gain by using T_2 over T. In general, it may be prescribed that use of apriori information $\{y_1, y_2, \ldots, y_k\}$ at estimation stage in the form of T_2 should be used in case these known y-values are small while the corresponding x-values are high.

(iv) One may safely advocate to enumerate completely those units which are expected to have extreme-low y-values, but extreme-high x-values and to draw a PPSWR sample from the remaining units.

6.4.3 Use of Apriori Information in Case of VPSWOR

Let

$$d_1 = \sum_{i=1}^{n} y_i/\pi_i' \quad \text{and} \quad d_2 = \sum_{i=1}^{n-k} y_i/\pi_i^*$$

be H-T type unbiased estimators of Y_{N-k} based on samples of sizes n and $(n-k)$ respectively drawn without replacement and with varying probabilities from the unknown part of the population consisting of (N-k) units, where π_i' and π_i^* $(i=k+1,\ldots,N)$ are corresponding inclusion probabilities.

Obviously,

$$T_1 = Y_k + d_1 \quad \text{and} \quad T_2 = Y_k + d_2 \qquad (6.4.15)$$

would be unbiased for y. Let $T = \sum_{i=1}^{n} y_i/\pi_i$ be the usual unbiased H-T estimator for Y based on a sample of n units drawn without replacement from the entire population, π_i being the corresponding inclusion probabilities for the pairs (i,j), $i \neq j$, corresponding to T, d_1 and d_2 will be denoted by π_{ij}, π_{ij}' and π_{ij}^* respectively.

We have,

$$V(T) = (D-1)Y^2 \tag{6.4.16}$$

$$V(T_1) = (D^{(1)}-1)Y^2 \tag{6.4.17}$$

and

$$V(T_2) = (D^{(2)}-1)Y^2 \tag{6.4.18}$$

where

$$D = \sum_{i=1}^{N} (y_i/Y)^2 \pi_i + \sum_{i \neq j=1}^{N}\sum (y_i/Y)(y_j/Y)(\pi_{ij}/\pi_i\pi_j)$$

and $D^{(1)}$ is D with π_i replaced by π_i' and $D^{(2)}$ is D with π_i replaced by π_i^* and the sum extending from $(k+1)$ to N in place of 1 to N in D and π_{ij} being modified by the corresponding second order inclusion probabilities. Obviously, it is very difficult to say whether $V(T_1) < V(T_2)$ and $V(T_1) < V(T)$ in general, contrary to the results obtained in Theorems 6.4.1 and 6.4.2.

To fin out the relative p rformance of t e estimators T_1, T_2 and T, we adopt the super-population model approach. Let $\{y_1, y_2, \ldots, y_N\}$ be a realization of random vector $\{Y_1, Y_2, \ldots, Y_N\}$, say, the joint distribution of which is specified (need not be completely) by the first two moments through the model η:

$$E_\eta(y_i|x_i) = \beta x_i, \quad i = 1,2,\ldots,N$$

$$V_\eta(y_i|x_i) = \psi(x_i) \tag{6.4.19}$$

and

$$E_\eta(Y_i-\beta x_i)(Y_j-\beta x_j) = 0, \quad i \neq j = 1,2,\ldots,N$$

where E and V denote the expectation and variance with respect to the super-population model η.

It may be shown that,

$$E_\eta\, V(T) = \beta^2[\sum_{i=1}^{N} \{(1-\pi_i)/\pi_i\}C^2(y_i|x_i) + V(\hat{x})] \qquad (6.4.20)$$

$$E_\eta\, V(T_1) = \beta^2[\sum_{i=k+1}^{N} \{(1-\pi_i')/\pi_i'\}C^2(y_i|x_i) + V_1(\hat{x})] \qquad (6.4.21)$$

and

$$E_\eta\, V(T_2) = \beta^2[\sum_{i=k+1}^{N} \{(1-\pi_i^*)/\pi_i^*\}C^2(y_i|x_i) + V_2(\hat{x})] \qquad (6.4.22)$$

where,

$$C^2(y_i|x_i) = \psi(x_i)/\beta^2 x_i^2$$

$$V(\hat{x}) = \sum_{i=1}^{N} x_i^2(1-\pi_i)/\pi_i + \sum_{i\neq j=1}^{N}\sum x_i x_j(\pi_{ij}-\pi_i\pi_j)/\pi_i\pi_j$$

and $V_1(\hat{x})$ is $V(\hat{x})$ with π_i and π_{ij} being replaced by π_i', π_{ij}' and $V_2(\hat{x})$ is $V(\hat{x})$ with π_i and π_{ij} being replaced by π_i^* and π_{ij}^*, the sum extending to $i = k+1,\ldots,N$ in each case.

Now, we have the following

Theorem 6.4.3

(i) $E_\eta\, V(T_1) \leq E_\eta\, V(T_2)$, if $\pi_i' > \pi_i^*(i=k+1,\ldots,N)$ and $V_2(\hat{x}) \geq V_1(\hat{x})$

(ii) $E_\eta\, V(T_1) \leq E_\eta\, V(T)$, if $\pi_i' > \pi_i(i=k+1,\ldots,N)$ and $V(\hat{x}) \geq V_1(\hat{x})$

and

(iii) $E_\eta\, V(T_2) \leq E_\eta\, V(T)$, if $\pi_i^* > \pi_i(i=k+1,\ldots,N)$ and $V(\hat{x}) \geq V_2(\hat{x})$

Proof. From (6.4.20) and (6.4.21), we observe that

$$E_\eta[V(T)-V(T_1)] = \beta^2[\sum_{i=k+1}^{N}(\frac{1}{\pi_i} - \frac{1}{\pi_i'})C^2(y_i|x_i) + \sum_{i=k+1}^{N}\{(1-\pi_i)/\pi_i\}C^2(y_i|x_i) + \{V(\hat{x}) - V_1(\hat{x})\}]$$

and hence, the sufficient condition (ii) follows.

Similarly, from (6.4.20) and (6.4.22), observing that

$$E_\eta\,[V(T)-V(T_2)]=\beta^2[\sum_{i=k+1}^{N}(\frac{1}{\pi_i}-\frac{1}{\pi_i^*})C^2(y_i|x_i)+\sum_{i=k+1}^{N}\{(1-\pi_i)/\pi_i\}C^2(y_i|x_i)$$

$$+\{V(\hat{x})-V_2(\hat{x})\}],$$

the sufficient condition (iii) follows.

Finally, from (6.4.21) and (6.4.22), the result (i) follows.

Remarks

(i) In case of π-P-S schemes, with $\pi_i = np_i$, $\pi_i' = np_i'$ and $\pi_i^* = (n-k)p_i'$ where $p_i = x_i/x > 0$ $(i = 1,2,\ldots,N)$ and $p_i' = x_i/X_{N-k} > 0$ $(i = k+1,k+2,\ldots,N)$, it is found that each of $V(\hat{x})$, $V_1(\hat{x})$ and $V_2(\hat{x})$ is zero and that the estimator T is η-better than both of T_2 and T without any condition. Further T_2 would be η-better than T if $X_k/X_{N-k} > k/(n-k)$, a condition similar to that required in case of PPSWR and which would be satisfied in case the apriori information set $\{y_1,y_2,\ldots,y_k\}$ is for those k-units which have extreme-high x-values $\{x_1,x_2,\ldots,x_k\}$.

(ii) In case of all the estimators based on SRSWOR, where $\pi_i = n/N$, $i = 1,2,\ldots,N$, $\pi_i' = n/(N-k)$, $\pi_i^* = (n-k)/(N-k)$, $i = k+1,K+2,\ldots,N$; $\pi_{ij}' = n(n-1)/(N-k)(N-k-1)$, $\pi_{ij}^* = (n-k)(n-k-1)/(N-k)(N-k-1)$ $(i \neq j = k+1,k+2,\ldots,N)$, we have

$$V(\hat{x}) = \frac{N^2(N-n)}{Nn}\;S_x^2,\quad V_1(\hat{x}) = (N-k)^2\,\frac{(N-k-n)}{n(N-k)}\,S_{N-k}^2.$$

6.5 Empirical Study

(a) Let us consider the following data of eye estimated number of households (x) and the actual number of households (y) in an area containing 20 blocks [Horvitz and Thompson (1952)]. Units, after arranging in ascending order according to x are given below :

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
x	9	12	12	13	14	15	17	18	19	19	20	21	22	25	25	27	27	35	37	47
y	9	12	12	12	12	14	14	18	18	19	17	24	25	21	26	23	27	24	40	30

Let the y-values for $k = 4$, the larger units, namely 17th, 18, 19th and 20th be known apriori. Let $n = 6$ and the estimators T, T_1 and T_2 be the same as defined earlier in Section 6.4.1.

The following Table 6.4.1 gives the present relative efficiences of T_1 and T_2 over T and that of T_2 over T_1, in case a sample of size n is selected by SRSWOR.

Table 6.4.1. Present relative efficiencies

$100 \times [V(T)/V(T_1)]$	$100 \times [V(T)/V(T_2)]$	$100 \times [V(T_2)/V(T_1)]$
(1)	(2)	(3)
484.36	115.32	419.99

(b) Let us consider the following population of 12 units with the corresponding income (x) and expenditure on food item (y) respectively.

Units	Income (x) (in Rs.)	Household size	Expenditure on food item (y) (in Rs.)
(1)	(2)	(3)	(4)
1	2500	3	350
2	2000	4	400
3	1500	2	250
4	400	5	300
5	550	6	350
6	600	7	400
7	330	4	220
8	400	3	150
9	500	3	200
10	350	2	150
11	200	2	100
12	300	3	150

Let the y-values for 1st and 2nd units be known apriori, let a sample of size 5 be selected by PPSWR and let the estimators T, T_1 and T_2 be the same as defined in Section 6.4.2.

For this population, the condition $X_k/X_{N-K} > k/(n-k)$ is satisfied and the following Table 6.4.2 gives percent relative efficiences of T_1 and T_2 over T and that of T_1 over T_2.

Table 6.4.2. Perecent relative efficiency

$100 \times [V(T)/V(T_1)]$	$100 \times [V(T)/V(T_2)]$	$100 \times [V(T_2)/V(T_1)]$
(1)	(2)	(3)
344.14	206.48	166.66

CHAPTER VII

USE OF INFORMATION ON TWO AUXILIARY VARIATES IN SELECTING THE SAMPLING UNITS

7.1 Introduction and Summary

In the previous chapters, we have already discussed how the information on certain parameters $\psi(y)$, other than the parameter under estimation $\phi(y)$, or on some values of the character y under study, in addition to those in the sample, may be used in obtaining estimators which are better, at least in some situations, than those where no such information is utilised. The information required was some times in the form of knowledge of exact values or the approximate good-guessed values of the parameters $\psi(y)$ or in the form of some known quantities greater or smaller than the actual value of the parameters $\psi(y)$ involved. Though the most of the estimators under discussion, based on apriori information, were biased, but better than the usual unbiased estimators for the parameter under consideration, viz. population mean $\bar{Y}$, total Y, variance σ_y^2 under some very moderate conditions. It is to be noted that the information was used at the estimation stage only, i.e., in defining the suitable estimators and that the parameters $\psi(y)$ for which the apriori information was required were based on the character under study y itself.

Here in this chapter, we have considered the use of information on some auxiliary characters $\underset{\sim}{x}$, related to the study variate y, at the selection stage, i.e., in selecting the population units into the sample with varying probabilities.

The use of auxiliary information in defining a probability distribution on the population $U = \{1,2,\ldots,N\}$ and selecting the units with varying probabilities, was first considered by Hansen and Hurwitz (1943). Using the information on an auxiliary character x_i, an unbiased sampling strategy for Y, the population total of y, based on probability proportional to size and with replacement (ppswr) sample of size n is given by

$$\tau_i \equiv \{ppx_i wr;\ \hat{Y}_i\},\quad \hat{Y}_i = (1/n) \sum_{j=1}^{n} (y_j/x_{ij})X_i \qquad (7.1.1)$$

with variance

$$V(\tau_i) = (1/n)\ V_{x_i}(y),\ V_{x_i}(y) = \sum_{j=1}^{N} (y_j^2/x_{ij})X_i - Y^2 \qquad (7.1.2)$$

where $X_i = \sum_{j=1}^{N} x_{ij}$, the population total of the character x_i.

In case, the information on more than one auxiliary character is available, it may be utilized in selecting the units with varying probabilities. The general problem is to obtain a suitable measure $z = \phi(x_1,x_2,\ldots,x_p)$, as a function of $x_1,x_2,\ldots,x_p$, alone such that the sampling strategy,

$$\tau \equiv \{ppzwr,\ \hat{Y}\},\qquad \hat{Y} = \frac{1}{n} \sum_{j=1}^{n} (y_j/p_j^*) \qquad (7.1.3)$$

based on probabilities $\{p_j^*\}$,

$$p_j^* = z_j/Z,\ z_j = \phi(x_{1j}, x_{2j}, \ldots, x_{pj}), j=1,2,\ldots,N,\ Z = \sum_{j=1}^{N} z_j \tag{7.1.4}$$

and with replacement selection of units, is better than each of τ_i $(i = 1,2,\ldots,p)$ and the strategy

$$\tau_o \equiv \{SRSWR, \hat{Y}_o\},\ \hat{Y}_o = N\bar{y} \tag{7.1.5}$$

based on a simple random sample (with replacement) of size n and usual unbiased estimator $\hat{Y}_o$ of Y. The problem may be formulated on the same lines in case of varying probability selection without replacement.

To the best of our knowledge, Maiti and Tripathi (1976) were first to consider the problem of estimating Y with the help of a sampling strategy τ in (7.1.3) for $p = 2$.

Recently, Agrawal and Singh (1980) have considered the above problem by taking a linear function $z^* = w'\underset{\sim}{x}$ of the auxiliary variates $\underset{\sim}{x} = (x_1, x_2, \ldots, x_p)$ to determine the selection probabilities $p_j^* = z_j^*/Z^*$, $j = 1,2,\ldots,N$, where the weight vector w is to be selected under same criterion. In case w is such that, mean square error of prediction $E(y-w'\underset{\sim}{x})^2$ is minimum, it would be given by

$$w_o = \Sigma^{-1}\sigma \tag{7.1.6}$$

where Σ^{-1} is the inverse of Σ, the variance covariance matrix of the vector $\underset{\sim}{x}$ and σ is a vector of covariances of y with $x_i(i = 1,2,\ldots,p)$. Obviously, the choice of selection variable $z^* = w_o' \underset{\sim}{x}$ requires the knowledge of not only Σ, but also of σ which depends on character under study y itself and possibly may never be known in practice. This observation led them to propose an alternative choice of selection variable z^* defined by

$$z^* = \alpha' \underset{\sim}{x} \tag{7.1.7}$$

where the vector α is to be chosen such that z^* has maximum correlation with y and is given by

$$\alpha = (\Sigma^{-1}Q)/[Q\Sigma^{-1}Q]^{1/2} \tag{7.1.8}$$

where

$$Q = (q_i),\ q_i = (\rho_{oi}/\rho_{op})(\sigma_i/\sigma_p),\ i = 1,2,\ldots,p$$

ρ_{oi} being the correlation coefficient between y and x_i and σ_i^2 is the variance of x_i.

It may be noted that the index z^* in (7.1.7) depends on the quantities (ρ_{oi}/ρ_{op}), the parameters involving the character under study y again. However, the requirement of the knowledge on (ρ_{oi}/ρ_{op}) is less restrictive than the requirement of knowledge on σ in the index defined by (7.1.6). They considered the strategy τ in (7.1.3) using the index given by (7.1.7) and (7.1.8) for $p = 2$ and compared it with τ_1 and τ_2 in (7.1.1) under the usual super-population model,

$$y_j = \beta z_j^* + e_j$$

with

$$E(e_j|z_j^*) = 0, \quad V(e_j|z_j^*) = \gamma z_j^{*g} \qquad (7.1.9)$$

$$E(e_j, e_j'|z_j^*, z_j^{*'}) = 0$$

and showed that τ is better (under the model) than both of τ_1 and τ_2 for $g = 2$. However, for $1 < g < 2$, they failed to arrive at any conclusion and moreover, they did not compare the performance of τ over τ_o.

It may be noted that the strategy considered by Agrawal and Singh (1980) requires the exact knowledge of $\theta = \rho_{o1}/\rho_{o2}$ involving y and which may not be available in many situations of practical importance. Further, the correlation ρ_{y,z^*}, z^* being given in (7.1.7) and (7.1.8) may be quite poor compared to $\rho_{y,\phi(x)}$, the $\phi(x)$ being some suitable functions of $x_1, x_2, \ldots, x_p$, in case of some populations, indicating that the use of the set of the probabilities based on above z^* won't reduce the variance of the strategy τ considerably. In fact, the natural requirement is that the selection variable $z^* = \phi(x_1, x_2, \ldots, x_p)$ do not depend on the parameters involving y-values and that the resulting strategy τ based on such a z is better than τ_1, τ_2 and τ_o simultaneously at the same time. In this chapter, we make an effort in this direction confining ourselves to $p = 2$.

In this chapter, we have given a set of necessary and sufficient conditions under which the strategy τ in (7.1.3) with general function $z = \phi(x_1, x_2)$ is better than both of τ_1 and τ_2 without any model consideration. Further, we have compared the proposed strategy τ with τ_i $(i = 0,1,2)$ under the structure of a super-population model η similar to that of Raj (1958) and found a set of necessary and sufficient condition for τ to be η-better than τ_i $(i=0,1,2)$ and find that for $g = 2$, τ is always η-better than each of τ_0, τ_1 and τ_2. We have also given a general procedure depending on the knowledge of g in the model η, for constructing $p_j^* = z_j/Z(> 0)$, $j = 1,2,\ldots,N$, $z_j = \phi(x_{1j}, x_{2j})$, so that τ is η-better than τ_1 and τ_2. We have undertaken the empirical studies to illustrate our points and at the end compared the strategy τ with two-variate ratio-estimators due to Olkin (1958) and Tripathi (1978) under the model.

7.2 Sampling Strategy and its General Properties

Let $z = \phi(x_1, x_2)$ denote a positive real valued function of the auxiliary characters x_1, x_2 and $z_j = \phi(x_{1j}, x_{2j})$, $j = 1,2,\ldots,N$ be the value of z associated with the j^{th} unit of the population under consideration. We propose a sampling strategy

$$\tau \equiv \{ppzwr, \hat{Y}\}, \quad \hat{Y} = \frac{1}{n} \sum_{j=1}^{n} y_j/p_j^* \qquad (7.2.1)$$

based on a sample of size n, selected with replacement and with probabilities $p_j^* = z_j|Z > 0$, $Z = \sum_{j=1}^{N} z_j = N\bar{Z}$ associated with j^{th} unit of the population.

Obviously, τ is unbiased for Y with its variance given by

$$V(\tau) = (1/n)V_z(y),\ V_z(y) = Z \sum_{j=1}^{N} y_j^2/z_j - Y^2 \tag{7.2.2}$$

and an unbiased estimate of variance is given by

$$\hat{V}(\tau) = (1/n)\ \hat{V}_z(y),\ \hat{V}_z(y) = \sum_{j=1}^{n} (y_j/p_j^* - \hat{Y})^2/(n-1) \tag{7.2.3}$$

Next,we have the following

<u>Theorem 7.2.1</u>. A necessary and sufficient condition for τ to be better than τ_i is

$$\sum_{j=1}^{N} (z_j - \bar{Z})\ y_j^2/z_j > \sum_{j=1}^{N} (x_{ij} - \bar{X}_i)^2\ y_j^2/x_j \tag{7.2.4}$$

<u>Proof</u>. From (7.1.2) and (7.2.2), it is obvious that $V(\tau) < V(\tau_i)$

$$\text{iff}\quad Z \sum_{j=1}^{N} y_j^2/z_j < X_i \sum_{j=i}^{N} y_j^2/x_{ij},\ i = 1,2, \tag{7.2.5}$$

which reduces to (7.2.4) and hence the theorem follows.

Noting that $V(\tau_o) = V(\tau_i)$ if $x_{ij} = c_i$, a constant, τ would be better than τ_o iff the left hand side of (7.2.4) is positive (Raj, 1954). Thus if covariance between z and y^2/z is positive and is greater than those between x_i and y^2/x_i, $i = 1,2$, the strategy τ will be simultaneously better than τ_1, τ_2 and τ_o. However, above criterion is not valuable from application point of view. This compells us to study the relative performance of the strategies under a model relating y with characters x_1, x_2 and z. Before doing so, we at first identify the situations in which the use of the strategy τ is not desirable

<u>Theorem 7.2.2</u>. Let $D_{(1)}$ and $D_{(2)}$ be any two known quantities such that

$$D_{(1)} \leq \min_{1 \leq j \leq N} y_j^2 \leq \max_{1 \leq j \leq N} y_j^2 \leq D_{(2)},$$

then a sufficient condition for τ_i to be better than τ is given by

$$(\bar{Z}/\tilde{z})\, D_{(1)}/D_{(2)} \geq \bar{X}_i/\tilde{X}_i, \quad i = 1,2. \tag{7.2.6}$$

<u>Proof</u>. Observing that $V(\tau_i) < V(\tau)$ iff $\bar{X}_i \sum_{j=1}^{N} y_j^2/x_{ij} < \bar{\bar{z}} \sum_{j=1}^{N} y_j^2/z_j$ and that

$$\bar{Z} \sum_{j=1}^{N} y_j^2/z_j > \bar{Z}\, D_{(1)} \sum_{j=1}^{N} (1/z_j) \tag{7.2.7}$$

$$\text{and} \qquad \bar{X}_i \sum_{j=1}^{N} y_j^2/x_{ij} < \bar{X}_i\, D_{(2)} \sum_{j=1}^{N} (1/x_{ij}),$$

the sufficient condition of the theorem follows.

Thus one should not use that set of z_j values which satisfy (7.2.6). Obviously, using (7.2.6), τ_o would be better than τ if the left hand side of (7.2.6) is greater than unity. Further if $D_{(1)} = D_{(2)}$ or $D_{(1)}/D_{(2)}$ is very close to one (which would be the case if y-values are more or less stable, i.e., variability in y is very small), τ_o would be better than τ whatever be $(z_1, z_2, \ldots, z_n)$. Also τ_o would be better than both of τ_1 and τ_2. In this case, one should resort to simple random sampling.

Theorem 7.2.3. A sufficient condition for τ to be better than both of τ_1 and τ_2 is given by

$$\bar{Z}/\tilde{Z} \leq K\, D_{(1)}/D_{(2)} \qquad (7.2.8)$$

where, $K = \min_{1\leq i\leq 2} \{\bar{X}_i/\tilde{X}_i\}$.

Proof. Noting that $\bar{Z}\, \Sigma\, y_j^2/z_j < \bar{Z}\, D_{(2)} \sum_{j=1}^{N} 1/z_j$ and $\bar{X}_i \sum_{j=1}^{N} y_j^2/x_{ij} > \bar{X}_i\, D_{(1)} \sum_{j=1}^{N} 1/x_{ij}$, the result follows from (7.2.5).

The above result will never hold true in case $D_{(1)}K/D_{(2)} < 1$. In case $D_{(1)}K/D_{(2)} > 1$, we may always generate z_j's such that (7.2.8) is satisfied. For example, without loss of generality, let us assume that $z_j \geq 1$, for each $j = 1,2,\ldots,N$ (in which case $1/\tilde{Z} \leq 1$) and let

$$z_j = x_{2j} - \lambda_j\, x_{1j},\ j = 1,2,\ldots,N \qquad (7.2.9)$$

where λ_j is to be chosen such that $1 \leq z_j \leq K\, D_{(1)}/D_{(2)}$, i.e., λ_j's are such that

$$[x_{2j}-K(D_{(1)}/D_{(2)})]/x_{1j} \leq \lambda_j \leq (x_{2j}-1)/x_{1j},\ j=1,2,\ldots,N \qquad (7.2.10)$$

where, x_{ij} are assumed (without loss of generality) to be positive for each j. From Theorem 7.2.3, it follows that use of $\{z_1,z_2,\ldots,z_N\}$ generated from (7.2.9) in τ makes it better than both of τ_1 and τ_2. ♦

The condition (7.2.8) merely guarantees the superiority of τ over τ_1 and τ_2. The author has failed to find a sufficient condition for τ to be better than τ_0. In case it is known that either of τ_1 and τ_2 is better than τ_o and $K\,D_{(1)}/D_{(2)} > 1$ one may always make τ better than all of τ_o, τ_1 and τ_2 by generating z_j's with the help of (7.2.9) and (7.2.10). Even in case neither of τ_1 and τ_2 is better than τ_o one may sometimes find τ to be better than all of τ_o, τ_1 and τ_2, provided $K\,D_{(1)}/D_{(2)} > 1$, by using (7.2.9) and (7.2.10) as we find in the empirical study below. We note that closer are the z_j values to a constant, higher would be the chance of satisfying (7.2.8). Such a choice of z_j's may however be defective in the sense that it may lead to a very small gain in efficiency (may be even to a loss in efficiency) of τ over τ_o. The above results do not thus help much and it becomes necessary to study the relative performance of the strategies τ_o, τ_1 and τ_2 under a model.

Empirical Study

The following data for analysis have been taken from the District Census Hand Book (Malda District, Part X - A and B), published by District Census of Operation, West Bengal, 1961.

Table 7.2. Data relating to area of village (x_1), total population of village (x_2) and total number of cultivators(y).

Name of the village	Area of village (x_1)	Total population of the village (x_2)	Total number of cultivators(y)
(1)	(2)	(3)	(4)
Khod Malancha	49.64	363	49
Kagasura	185.63	210	49
Kutubpur	44.59	154	46
Mahinagar	115.24	215	49
Dilalbati	187.04	248	49
Methrani	47.44	237	48
Jalangapara	62.19	104	46
Sanbandha	157.67	186	47
Radhanagar	81.13	164	47
Sonapur	89.34	186	46
K-Danga	68.76	289	47
Karanj	50.80	83	49

For this population, we have :

$$\bar{X}_1 = 94.9583, \quad \tilde{X}_1 = 73.5767, \quad \bar{X}_2 = 199.0833,$$

$$\tilde{X}_2 = 172.3878, \quad \bar{Y} = 47.66, \quad \sigma_y^2 = 2.191$$

$$K = \min_{1 \le i \le 2} \{\bar{X}_i/\tilde{X}_i\} = 1.1548573.$$

Let $D_{(1)} = (46)^2 = 2116$ and $D_{(2)} = (49)^2 = 2401$. Hence

$$K\, D_{(1)}/D_{(2)} = 1.017775 > 1.$$

Now, we take the following set of values of z_j as defined in (7.2.9) with λ_j satisfying (7.2.10).

z_j :	1.01023,	1.03631,	1.02063,	1.02237,	1.01148,	1.01921
z_j :	1.01958,	1.01683,	1.01795,	1.01617,	1.01339,	1.01794

We get the following :

$$n\,V(\tau_1) = V_{x_1}(y) = 88310.6, \quad n\,V(\tau_2) = V_{x_2}(y) = 45430.6$$

$$n\,V(\tau_o) = 315.504 \quad \text{and} \quad V_x(y) = 170.9 = n\,V(\tau).$$

Thus it is observed that

$$V(\tau) < V(\tau_o) < V(\tau_2) < V(\tau_1).$$

We note that gains by using τ over both τ_1 and τ_2 are appreciably high and that over τ_o is also high, though by chance in the later case.

7.3 Comparison of Various Sampling Strategies under a Super-population model

In this section, we shall compare the strategies τ, τ_1, τ_2 and τ_o under the super-population model similar to that considered by Raj (1958). Let the finite population $U = \{1,2,\ldots,N\}$ under consideration, be a random sample from a very large population specified by the model

$$y_j = \alpha + \beta w_j + \epsilon_j, \; j = 1,2,\ldots,N$$

with $E(\epsilon_j | w_j) = 0, \quad j = 1,2,\ldots,N$ (7.3.1)

$$V(\epsilon_j | w_j) = \psi(w_j)$$

and $E(\epsilon_j \; \epsilon_{j'} | w_j \; w_{j'}) = 0, \quad j \neq j' = 1,2,\ldots,N$

where w is any suitably chosen variate and in particular may be $z = \phi(x_1, x_2)$ we shall refer to such a model as η_w-model.

Let

$$V_z(w) = Z \sum_{j=1}^{N} w_j^2/z_j - W^2 \tag{7.3.2}$$

and $$V_o(w) = N \sum_{j=1}^{N} w_j^2 - W^2 \tag{7.3.3}$$

where $W = \sum_{j=1}^{N} w_j = N\overline{W}$.

<u>Definition</u>. A sampling strategy τ is said to be η_w-better than a strategy τ^* iff

$$E\,V(\tau) \leq E\,V(\tau^*) \quad \text{for all} \quad \theta = (y_1, y_2, \ldots, y_N)$$

with inequality holding for at least one θ, where expectation is taken under the model η_w.

It may be shown that under the model η_w,

$$n\,E[V(\tau)-V(\tau_i)] = N^2[(\overline{Z}/\tilde{Z})-(\overline{X}_i/\tilde{X}_i)]\alpha^2 + 2\alpha\,\beta \sum_{j=1}^{N}\{(Z/z_j)-(X_i/x_{ij})\}w_j$$

$$+ \beta^2[V_z(w)-V_{x_i}(w)] + \sum_{j=1}^{N}\{(Z/z_j)-(X_i/x_{ij})\}\psi(w_j) \tag{7.3.4}$$

where $\tilde{Z}$ and $\tilde{X}_i$ are harmonic means of z_j's and x_{ij}'s($i=1,2$; $j = 1,2,\ldots,N$) respectively.

Also $$nE[V(\tau)-V(\tau_o)] = N^2\alpha^2[(\bar{Z}/\tilde{Z})-1]+2\alpha\beta[Z\sum_{j=1}^{N} w_j/z_j-NW]$$
$$+ \beta^2[V_z(w)-V_o(w)] + \sum_{j=1}^{N}\{(Z/z_j)-N\}\psi(W_j) \qquad (7.3.5)$$

Next, we prove the following

<u>Theorem 7.3.1</u>. A sufficient condition, in case $\alpha = 0$, for the strategy τ to be η_w-better than τ_1, τ_2 and τ_o, is

$$\beta^2 V_z(w) + Z\sum_{j=1}^{N}\psi(w_j)/z_j \leq \beta^2 V(w) + K \qquad (7.3.6)$$

where,

$$V(w) = \min\{V_{x_1}(w),\ V_{x_2}(w),\ V_o(w)\}$$
$$K = \min\{X_1\sum_{j=1}^{N}\psi(w_j)/x_{1j},\ X_2\sum_{j=1}^{N}\psi(w_j)/x_{2j},\ \sum_{j=1}^{N}\psi(w_j)\}.$$

Another sufficient condition that the strategy τ is η_w-better than τ_1,τ_2 and τ_o would be given by

$$\beta^2 V_z(w) + M\sum_{j=1}^{N}\psi(w_j) \leq \beta^2 V(w) + m^*\sum_{j=1}^{N}\psi(w_j) \qquad (7.3.7)$$

where, $m^* = \min(m,N)$, $m = \min_j(X_1/x_{1j},X_2/x_{2j})$ and $M = \max_j(Z/z_j)$.

<u>Proof</u>. For $\alpha = 0$, τ would be η_w-better than τ_i and τ_o

$$\text{iff} \quad \beta^2 V_z(w)+Z\sum_{j=1}^{N}\psi(w_j)/z_j \leq \beta^2 V_{x_i}(w) + X_i\sum_{j=1}^{N}\psi(w_j)/x_{ij} \qquad (7.3.8)$$

$$\text{and} \quad \beta^2 V_z(w)+Z\sum_{j=1}^{N}\psi(w_j)/z_j \leq \beta^2 V_o(w) + N\sum_{j=1}^{N}\psi(w_j)$$

respectively. The sufficient condition (7.3.8) follows immediately. Further for $\alpha = 0$, we note that

$$nE[V(\tau)-V(\tau_i)] \leq \beta^2[V_z(w)-V_{x_i}(w)] + (M-m)\sum_{j=1}^{N}\psi(w_j), \quad i = 1,2 \tag{7.3.9}$$

and $nE[V(\tau)-V(\tau_o)] \leq \beta^2[V_z(w)-V_o(w)] + (M-N)\sum_{j=1}^{N}\psi(w_j)$

and hence, the second sufficidnt condition follows.

It may be noted that the above approach is quite general in nature. From (7.3.4), (7.3.5), (7.3.8) and (7.3.9), one may generate criteria for preference of τ over τ_1, τ_2 and τ_o for specific choices of w and $\psi(w_j)$ in the super-population model η_w. It may also be noted that under η_z-model (i.e., when $w_j=z_j$), one would have $V_z(w) = 0$ and under η_{x_i}-model $V_{x_i}(w) = 0 = V(w)$. Maiti and Tripathi (1976) obtained the results (7.3.4) and (7.3.5) under η_z-model.

From (7.3.4), it is realized that if $z_j = \phi(x_{1j},x_{2j})$ in the model η_z be known, it may be advisable to use that z_j in constructing the set of selection probabilities,

$$p_j^* = z_j/Z = \phi(x_{1j},x_{2j})/\sum_{j=1}^{N}\phi(x_{1j},x_{2j}) > 0 \quad (j=1,\ldots,N).$$

In the η_z-model, let $\alpha = 0$ and $\psi(z_j) = \beta^2 z_j^2 C_j^2$, where $C_j^2 = \mathcal{E}^2(y_j|x_{1j},x_{2j})/\{\mathcal{E}(y_j|x_{1j},x_{2j})\}^2$ is the square of co-efficient of variation of y_j for given x_{1j} and x_{2j} in the super-population model. Then from (7.3.4) and (7.3.5), in case $C_j^2 = C$, a constant, i.e., for stable populations, we have

$$n\,E\,[V(\tau)-V(\tau_i)] = -\,\beta^2(1+C^2)\,V_{x_i}(z)$$

and $$n\,E\,[V(\tau)-V(\tau_o)] = -\,\beta^2[V_o(z)+N'^2C^2\sigma_z^2].$$

Thus, for all stable populations, the proposed strategy τ based on z_j will always be better than all of τ_1, τ_2 and τ_o.

In case, $C_j'^2$ is not constant, then the strategy τ will be better than τ_1 and τ_2

$$\text{iff} \quad -\beta^2V_{x_i}(z) + \beta^2[Z\sum_{j=1}^{N}C_j^2/z_j - X_i\sum_{j=1}^{N}z_j^2C_j^2/x_{ij}] \le 0$$

Let $C^2_{(1)} \le C^2_j \le C^2_{(2)}$, i.e., for all those populations where square of conditional coefficient of variation of y given x_{1j} and x_{2j} is expected to lie between $C^2_{(1)}$ and $C^2_{(2)}$, then τ will be better than τ_1 and τ_2

$$\text{if} \quad -\beta^2V_{x_i}(z)+C^2_{(2)}\,\beta^2\sum_{j=1}^{N}Z/z_j-C^2_{(1)}\beta^2[X_i\sum_{j=1}^{N}z_j^2/x_{ij}-Z^2]$$

$$-\beta^2C^2_{(1)}Z'^2 \le 0$$

$$\text{i.e., if} \quad -(1+C^2_{(1)})\,V_{x_i}(z) + \beta^2[C^2_{(2)}\sum_{j=1}^{N}Z/z_j-C^2_{(1)}Z^2] \le 0$$

and hence a sufficient condition for the strategy τ to be better than τ_i $(i = 1,2)$ would be

$$C^2_{(2)}/C'^2_{(1)} \le Z.\tilde{Z}. \qquad (7.3.10)$$

Thus for non-stable populations, the strategy τ would be better than τ_i if the condition (7.3.10) is satisfied.

7.4 Comparison of Strategies under a Specific Form of $\psi(z_j)$ in η_z-model

In this section, we take a specific form of $\psi(z_j)$, in η_w-model in (7.3.1) with $w_j = z_j = \phi(x_{1j}, x_{2j})$ as

$$\psi(z_j) = \gamma z_j^g, \ \gamma \geq 0, \ g \geq 0. \tag{7.4.1}$$

We obtain, from (7.3.4) and (7.3.5),

$$n\,E\,[V(\tau)-V(\tau_i)] = N^2[(\bar{Z}/\tilde{Z})-(\bar{X}_i/\tilde{X}_i)]\alpha^2+2\alpha\beta\sum_{j=1}^{N}\{(Z/z_j)-(X_i/x_{ij})\}z_j$$
$$-\beta^2 V_{x_i}(z) + \gamma\sum_{j=1}^{N}\{(Z/z_j)-(X_i/x_{ij})\}z_j^g \tag{7.4.2}$$

and $$n\,E\,[V(\tau)-V(\tau_o)] = N^2\alpha^2\{(\bar{Z}/\tilde{Z})-1\}-\beta^2 V_o(z)$$
$$+\gamma\sum_{j=1}^{n}\{(Z/z_j)-N\}z_j^g \tag{7.4.3}$$

From (7.4.2) with $\alpha = 0$, obviously τ would be η_z-better than τ_i

$$\text{iff} \quad \gamma Z\sum_{j=1}^{N} z_j^{g-1} \leq \beta^2 V_{x_i}(z) + \gamma X_i\sum_{j=1}^{N} z_j^g/x_{ij}, \ i = 1,2 \tag{7.4.4}$$

Now, we have the following

Theorem 7.4.1. Let $z_j = \phi(x_{1j}, x_{2j}) = x_{2j}-\lambda_j x_{1j} \geq 1$, and $x_{ij} \geq 0$ $(i=1,2)$ for all $j = 1,2,\ldots,N$ and $\alpha = 0$ in the model (7.1.3) with $w_j = z_j$ and (7.4.1). Let $K = \min_{1 \le i \le 2} \{\bar{X}_i/\tilde{X}_i\} > 1$. Then the strategy τ would be η_z-better than both of τ_1 and τ_2

$$\text{if} \quad (x_{2j}-K)/x_{1j} \le \lambda_j \le (x_{2j}-1)/x_{1j} \quad \text{for} \quad 0 \le g \le 1 \tag{7.4.5}$$

$$\text{and if} \quad (x_{2j}-K^{1/g})/x_{1j} \le \lambda_j \le (x_{2j}-1)/x_{1j} \quad \text{for} \quad g \ge 1 \tag{7.4.6}$$

for $j = 1,2,\ldots,N$.

<u>Proof</u>. Let $\psi(z_j)$ be of the same form as defined in (7.4.1). A sufficient condition from (7.4.4), for τ to be η_z-better than τ_i is given by

$$\bar{Z} \sum_{j=1}^{N} z_j^{g-1} \le \bar{X}_i \sum_{j=1}^{N} z_j^{g}/x_{ij}, \; i = 1,2.$$

Further, noting that,

$$z_j = \phi(x_{1j},x_{2j}) \ge 1 \Rightarrow \sum_{j=1}^{N} 1/x_{ij} \le \sum_{j=1}^{N} z_j^{g}/x_{ij},$$

the condition

$$\bar{Z}(1/N) \sum_{j=1}^{N} z_j^{g-1} \le K \tag{7.4.7}$$

will be sufficient for τ to be η_z-better than both of τ_1 and τ_2. The condition (7.4.5) may be re-written $1 \le x_{2j}-\lambda_j x_{1j}=z_j \le K$ which implies that

$$1 \le \bar{Z} \le K. \tag{7.4.8}$$

Now observing that

$$z_j \ge 1 \Rightarrow z_j^{g} \ge 1 \quad \text{for} \quad 0 \le g \le 1$$

$$\Rightarrow z_j^{g-1} \le 1$$

$$\Rightarrow \frac{1}{N} \sum_{j=1}^{N} z_j^{g-1} \le 1. \tag{7.4.9}$$

Now (7.4.8) and (7.4.9) imply (7.4.5), and hence the first part of the theorem (for $0 \leq g \leq 1$) follows.

The condition (7.4.6), for $g \geq 1$, may be rewritten as

$$1 \leq x_{2j} - \lambda_j x_{1j} = z_j \leq K^{1/g}$$

which implies

$$1 \leq \bar{Z} \leq K^{1/g} \quad \text{and} \quad z_j^{g-1} \leq K^{\frac{g-1}{g}} \quad \text{for} \quad g \geq 1 \qquad (7.4.10)$$

and which in turn, (for $z_j \geq 1$ and $g \geq 1$) implies

$$\bar{Z}(1/N) \Sigma z_j^{g-1} \leq K^{1/g} K^{(g-1)/g} = K$$

the condition (7.4.7) and hence the results in the above theorem.

<u>Remarks</u>

(i) For $\alpha = 0$, a sufficient condition from (7.4.7) or otherwise for τ to be η_z-better than both of τ_1 and τ_2 is given by

$$\bar{Z}/\tilde{Z} \leq \min_{1 \leq i \leq 2} \{\bar{X}_i/\tilde{X}_i\} \quad \text{if} \quad g = 0 \qquad (7.4.11)$$

and

$$\bar{Z} \leq \min_{1 \leq i \leq 2} \{\bar{X}_i/\tilde{X}_i\} \quad \text{if} \quad g = 1 \quad \text{and} \quad z_j \geq 1 \quad \text{for all}$$

$$j = 1,2,\ldots,N. \qquad (7.4.12)$$

(ii) For $\alpha = 0$ and $g = 2$, it may be shown from (7.4.2) that the strategy τ will always be η_z-better than η_i, because in that case

$$n E [V(\tau) - V(\tau_i)] = -(\beta^{*2}+\gamma) V_{x_i}(z). \qquad (7.4.13)$$

(iii) Following Raj (1958), it may be shown (also from 7.4.3), in case $\alpha = 0$, that τ would be η_z-better than τ_o for $g \geq 1$ and for $0 \leq g < 1$, if

$$\rho(z,v) > -(\beta^2/\lambda)\ (\sigma_z/\sigma_v) \qquad (7.4.14)$$

where $v = z^{g-1}$ and σ_v denotes the standard deviation of v, etc.

Thus τ is η_z-better than τ_o, τ_1 and τ_2 for $g = 2$.

(iv) It may be shown that under η_{x_i}-model (i=1,2), the condition (7.3.4) with $\alpha = 0$ reduces to

$$\Sigma\ p_j^*(x_{ij}/p_j^*-X_i)(v_{ij}/p_j^*-V_i) < -(\beta^2/\lambda)\ V_z(x_i) \qquad (7.4.15)$$

where $p_j^* = z_j/Z$, $v_{ij} = x_{ij}^{g-1}$ and $V_i = \sum_{j=1}^{N} v_{ij}$. It may be noted that (7.4.15) is never satisfied for $g = 1$ and $g = 2$.

(v) It may be shown that under η_z-model, the condition(7.3.4) with $\alpha = 0$ reduces to

$$\beta^2 V_{x_i}(z) + \lambda[X_i \sum_{j=1}^{N} v_j z_j/x_{ij} - ZV] > 0,\ i = 1,2,$$

where, $v_j = z_j^{g-1}$, $V = \sum_{j=1}^{N} v_j$.

(vi) Before using τ one may at first find out whether τ_i is better than τ_o. If τ_i is better, we may always improve it through τ using Theorem 7.4.1. otherwise directly we may choose τ using Remark (iii) above in case it is better than τ_o.

(vii) Once we know g in η_z-model, we may always generate, using Theorem 7.4.1, the z_j's (j=1,...,N) resulting into a strategy τ which would be η_z-better than both τ_1 and τ_2. We note that such sets $\{z_1,\ldots,z_N\}$ may be infinite in number. The author has failed to identify the best set.

(viii) From (7.4.2) and (7.4.3) for $\alpha = 0$ we note that a set of sufficient conditions for τ to be η_z-better than τ_i $(i=1,2)$ and τ_o are given by

$$\bar{Z} \sum_{j=1}^{N} z_j^{g-1} \leq \bar{Z}_i \sum_{j=1}^{N} z_j^{g}/x_{ij}$$

and

$$\bar{Z} \sum_{j=1}^{N} z_j^{g-1} \leq \sum_{j=1}^{N} z_j^{g}$$

respectively.

Empirical Study

For illustration, the following population of 10 plots is considered. The variables x_1, x_2 and y are respectively the average height of shoot (in meters), number of canes and yield in kg. per plot of the sugarcane.

Table 7.4. Data relating to the height of shoot (x_1), number of canes (x_2) and yield in kg. per plot of the sugarcane.

Plot No.	x_1	x_2	y
(1)	(2)	(3)	(4)
1	1.10	60	39.00
2	0.91	76	39.75
3	0.61	70	38.80
4	1.07	65	47.00
5	1.15	60	45.50
6	1.30	68	56.00
7	1.30	70	38.25
8	1.66	82	51.00
9	1.64	86	57.50
10	1.36	70	43.00

Let this population be a random realization from a super-population specified by (7.3.1) with $\alpha = 0$ and $w_j = \phi(x_{1j}, x_{2j}) = z_j$ and (7.4.1) therein and $0 < g \leq 1$.

For this population, we have

$$\bar{X}_1 = 1.2, \quad \tilde{X}_1 = 1.1109254$$

and

$$\bar{X}_2 = 70.7, \quad \tilde{X}_2 = 69.803106$$

thus $K = \min\left\{\frac{\bar{X}_1}{\tilde{X}_1}, \frac{\bar{X}_2}{\tilde{X}_2}\right\} = 1.0128489$

Now, from (7.4.5) with $z_j = x_{2j} - \lambda_j x_{1j}$, we get z_j's as

Plot No.	1	2	3	4	5	6	7	8	9	10
z_j	1.0136	1.0132	1.01327	1.01293	1.01305	1.011	1.0132	1.01026	1.007	1.014

This yields $\bar{Z} = 1.012158$, $\tilde{Z} = 1.0121505$, $\bar{Z}/\tilde{Z} = 1.0000074$.

Now from (7.4.2) and (7.4.3) with $\alpha = 0$ and using above z' values, we obtain $n\,E\,[V(\tau) - V(\tau_i)]$ and $n\,E\,[V(\tau) - V(\tau_o)]$ for $g = 0, 0.5$ and 1.

$$g = 0 \begin{cases} n\,E\,[V(\tau) - V(\tau_1)] = -[\beta^2 V_{x_1}(z) + 8.02\gamma] \\ n\,E\,[V(\tau) - V(\tau_2)] = -[\beta^2 V_{x_2}(z) + 1.28\gamma] \\ n\,E\,[V(\tau) - V(\tau_o)] = -\beta^2 V_o(z) + 0.0007\gamma \end{cases}$$

$$g = 0.5 \begin{cases} n\,E\,[V(\tau) - V(\tau_1)] = -[\beta^2 V_{x_1}(z) + 8.08\gamma] \\ n\,E\,[V(\tau) - V(\tau_2)] = -[\beta^2 V_{x_2}(z) + 1.30\gamma] \\ n\,E\,[V(\tau) - V(\tau_o)] = -[\beta^2 V_o(z) + 0.0038\gamma] \end{cases}$$

$$g = 1 \begin{cases} n\,E\,[V(\tau) - V(\tau_1)] = -[\beta^2 V_{x_1}(z) + 99.25\gamma] \\ n\,E\,[V(\tau) - V(\tau_2)] = -[\beta^2 V_{x_2}(z) + 92.41\gamma] \\ n\,E\,[V(\tau) - V(\tau_o)] = -[\beta^2 V_o(z)] \end{cases}$$

We find that τ is better than each of τ_o, τ_1 and τ_2 for $0 < g \leq 1$.

7.5 Comparison of the Proposed Strategy τ with those Based on Multivariate Ratio Estimators.

In the earlier sections, we have discussed how to choose a suitable function $z = \phi(x_1, x_2)$ of two auxiliary variables and identified the situations, where the use of such a choice in constructing the selection probabilities and thereby forming the usual unbiased estimator based on PPZWR would lead to smaller variance than the estimators based on PPx_1WR, PPx_2WR and SRSWR. However, so far we do not know whether our proposed strategy will be better than those based on the information on both of x_1 and x_2 at estimation stage alone. In other words, the problem is to find out whether the use of multi-auxiliary information at selection stage will always be advantageous over when it is used at estimation stage. We make an effort to get an answer to this, though only partially. For this we compare our strategy τ with those based on multi-variate ratio estimators due to Olkin (1958) and Tripathi (1978) under the η_z-model specified by (7.3.1) and (7.4.1) with $g = 2$.

Using information on p-auxiliary variates $x_1, x_2, \ldots, x_p$, Olkin's (1958) estimator for Y, the population total, may be defined as

$$\hat{Y}_r = w'\alpha \qquad (7.5.1)$$

where, $w' = (w_1, w_2, \ldots, w_p)$ are weights such that $\sum_{i=1}^{p} w_i = 1$ and

$$\alpha' = (\alpha_1, \alpha_2, \ldots, \alpha_p); \ \alpha_i = X_i \ \bar{y}/\bar{x}_i .$$

The mean square error (MSE) of $\hat{Y}_r$ to the terms of $0(n^{-1})$ would be

$$M(\hat{Y}_r) = \frac{N^2}{n} w' B w \qquad (7.5.2)$$

in case of simple random sampling with replacement, where $B = (b_{ik}), i, k = 1, \ldots, p,$

$$b_{ik} = \sigma_o^2 - R_i \sigma_{oi} - R_k \sigma_{ok} + R_i R_k \sigma_{ik} \ (i,k = 1,2,\ldots,p), \ R_i = \bar{Y}/\bar{X}_i,$$

$\sigma_{ik} = Cov(x_i, x_k)$ and 0 standing for y and i for x_i.

Obviously, for $p = 2$, $M(\hat{Y}_r)$ is found to be

$$M(\hat{Y}_r) = (N^2/n) \ [w_1^2 B_{(1)} + 2w_1 B_{(1,2)} + B_{(2)}] \qquad (5.5.3)$$

where

$$B_{(1)} = R_1^2 \sigma_{11} - 2 R_1 R_2 \sigma_{12} + R_2^2 \sigma_{22}$$

$$B_{(2)} = \sigma_o^2 - 2 R_2 \sigma_{o2} + R_2^2 \sigma_{22}$$

and $$B_{(1,2)} = R_2\sigma_{o2} - R_1\sigma_{o1} + R_1 R_2 \sigma_{12} - R_2^2 \sigma_{22}$$

It is well known that the optimum choice of weights in (7.5.1) depends on parameters involving y-also and which are rarely expected to be known in practice. In the face of such a situation Tripathi (1978) proposed a modified multivariate ratio estimator using modified optimum weights. The estimator $\hat{Y}_r^*$ proposed by him may again be defined by (7.5.1) with

$$w' = e'D^{-1}/e'D^{-1}e,\ D(d_{ik}),\ d_{ik} = \rho_{ik} C_i C_k,\quad i,k = 1,\dots,p$$

where $e = (1,1,\dots,1)'$ is p-dimensional unit column vector, C_i is the coefficient of variation of x_i $(i = 1,2,\dots,p)$ and ρ_{ik} is the coefficient of correlation between x_i and x_k. It has been shown by Tripathi that there will not be much loss in efficiency by using the above weights compared to the use of optimum weights in (7.5.1).

For $p = 2$, the MSE of modified ratio estimator $\hat{Y}_r^*$ due to Tripathi is again given by (7.5.3) with

$$w_1 = [(C_2^2-\rho_{12} C_1C_2)/(C_1^2+C_2^2-2\rho_{12} C_1C_2)] = 1 - w_2 \qquad (7.5.4)$$

Let ρ_{iz} and ρ_{iu} denote coefficients of correlation between x_i and z, and x_i and z^2 respectively, $C_z = \sigma_z/\bar{Z}$, C.V. of z and let $C^{*2} = \gamma/\beta^2$.

Lemma 7.5.1. Let $w_1 = w_2 = 1/2$ in (7.5.3). Then under the η_z-model in (7.3.1) with (7.4.1) and $g = 2$, we have,

$$\text{(i)}\quad E\, M(\hat{Y}_r) = L + (\beta^2 Z^2/n)\,[C^{*2}\{1 - \frac{(1+C_z^2)}{N}\}] \tag{7.5.5}$$

where,
$$L = (\beta^2 Z^2/n)[(C_1^2+C_2^2+2\rho_{12}C_1C_2)\{1 + \frac{C^{*2}(1+C_z^2)}{N}\}/4$$
$$- (\rho_{1,z}\, C_1 + \rho_{2z}\, C_2)C_z + C_z^2(1+C^{*2})$$
$$- C^{*2}(1+C_z^2)(\rho_{1u}C_1 + \rho_{2u}\ )\, C_u/N]$$

and
$$\text{(ii)}\quad E\, V(\tau) = (\beta^2 Z^2/n)\,[C^{*2}\{1 - \frac{(1+C_z^2)}{N}\}] \tag{7.5.6}$$

Proof. (i) Observing that

$$E\, Y^2 = \beta^2 Z^2[1 + \{C^{*2}(1+C_z^2)/N\}]$$
$$E\, \sigma_o^2 = \beta^2 Z^2[C_z^2 + \frac{(N-1)}{N} C^{*2}(1+C_z^2)] \tag{7.5.7}$$
and
$$E\, R_i\sigma_{oi} = \beta^2 Z^2[\sigma_{iz}C_iC_z + N\, C^{*2} \sum_{j=1}^{N} p_j^{*2}(p_{ij} - \frac{1}{N})]$$

where $p_{ij} = x_{ij}/X_i$ $(i=1,2)$ and $p_j^* = z_j/Z$, $Z = \sum_{j=1}^{N} z_j$, $j = 1,2,\ldots,N$.

We have,

$$E\, B_{(1)} = \beta^2 Z^2\, [\{1 + \frac{C^{*2}}{N}(1+C_z^2)\}(C_1^2 - 2\rho_{12}C_1C_2 + C_2^2)$$
$$E\, B_{(2)} = \beta^2 Z^2[C^{*2}\{\frac{C_2^2(1+C_z^2)}{N} + \frac{(N-1)}{N}(1+C_z^2) - 2N \sum_{j=1}^{N} p_j^{*2}(p_{2j} - \frac{1}{N})\}$$
$$+ C_z^2 - 2\rho_{2z}\, C_2\, C_z + C_2^2]$$
$$\tag{7.5.8}$$
and
$$E\, B_{(1,2)} = \beta^2 Z^2[(\rho_{2z}C_2C_z - \rho_{1z}C_1C_z) + N\, C^{*2} \sum_{j=1}^{N} p_j^{*2}(p_{2j} - p_{1j})$$
$$+ (\rho_{12}\, C_1C_2 - C_2^2)\ \{1 + \frac{C^{*2}}{N}(1 + C_z^2)\}]$$

Finally, we observe that

$$N\, C^{*2} \sum_{j=1}^{N} \{(\tfrac{1}{N}-p_{1j}) + (\tfrac{1}{N}-p_{2j})\} p_j^{*2} = - \frac{C^{*2}(1+C_z^2)(\rho_{1u}C_1+\rho_{2u}C_2)C_u}{N} \tag{7.5.9}$$

where $u = z^2$. Now the result (i) of Lemma follows using (7.5.8) and (7.5.9) from (7.5.3) with $w_1 = w_2 = 1/2$.

Further observing that

$$V(\tau) = (1/n)[Z \sum_{j=1}^{N} y_j^2/z_j - (\sum_{j=1}^{N} y_j^2 + \sum_{j\neq k=1}^{N}\sum y_i y_k)]$$

the result (ii) follows immediately after taking expectation under specified η_z-model.

Theorem 7.5.1. Under the conditions of Lemma 7.5.1, the strategy τ is η_z- better than the strategy (SRSWR, $\hat{Y}_r$) for $w_1 = w_2 = 1/2$

$$\text{iff} \qquad L > 0 \tag{7.5.10}$$

Proof. The result follows immediately, from Lemma 7.5.1, noting that

$$E\, M(\hat{Y}_r) = L + E\, V(\tau).$$

Remarks

(i) The comparison in the above Theorem is valid for large samples, to the terms $O(n^{-1})$, only.

(ii) (7.5.10) indicates that, we may have $E\, V(\tau) \leq E\, M(\hat{Y}_r)$ in many situations and the opposite in many other. It appears to be quite difficult to identify the specific situations in which either of τ and (SRSWR, $\hat{Y}_r$) is preferable over the other.

(iii) It is interesting to note that if ρ_{iz} *and* ρ_{iz^2} $(i=1,2)$ i.e., correlation coefficient between x_i and z and that between x_i and z^2 are zero or negative, then *L* *would be positive and* hence the strategy τ would be better than the strategy (SRSWR, $\hat{Y}_r$) with $w_1 = w_2 = 1/2$.

(iv) In case of the modified ratio estimator $\hat{Y}_r^*$ due to Tripathi (1978), we note from (7.5.4) that $w_1 = w_2 = \frac{1}{2} \Leftrightarrow C_1 = C_2$ (= C say) > 0 and then using the Lemma 7.5.1 and Theorem 7.5.1, a necessary and sufficient condition for τ to be η_z-better than (SRSWR, $\hat{Y}_r^*$) for $w_1 = w_2 \neq \frac{1}{2}$ is given by

$$(1+C^{*2})\, C_z^2 + 4C^2 \{\frac{1+\rho_{12}}{2} - \frac{(\rho_{1z}+\rho_{2z})C_z}{C}\}$$

$$+ \frac{(1+C_z^2)C^2C^{*2}}{N} \{\frac{1+\rho_{12}}{2} - \frac{(\rho_{1u}+\rho_{2u})C_u}{C}\} > 0 \qquad (7.5.11)$$

and a sufficient condition for the same is given by

$$\rho_{12} \geq 2\rho^* - 1 \qquad (7.5.12)$$

where

$$\rho^* = \max \{\frac{(\rho_{1z}+\rho_{2z})C_z}{C},\ \frac{(\rho_{1u}+\rho_{2u})C_u}{C}\}$$

and which always holds true, if

$$\rho_{iz} \leq 0, \quad \rho_{iu} \leq 0 \quad (i = 1,2). \qquad (7.5.13)$$

REFERENCES

Adbharyu, Dhiresh (1978) : Successive sampling using multiauxiliary information. *Sankhyā*, Ser. C, 40, pp. 29-37.

Agrawal, S.K. and Singh, Murari (1980) : Use of multivariate auxiliary information in selection of units in probability proportional to size with replacement. *Jour. Ind. Soc. Agri. Statist.* XXXII, No. 3, pp. 71-81.

Ajgaonkar, S.G.P. (1967) : On unordering best estimator in Horvitz-Thompson's T_1-class of linear estimators, *Sankhyā*, Ser. B, 29, pp. 209-212.

__________ (1969) : On the non-existence of a best estimator for the entire class of linear estimators. *Sankhyā*, Ser. A, 31, pp. 455-462.

Basu, D. (1958) : On sampling with and without replacement. *Sankhyā*, 20, pp. 287-294.

Bowley, A.L. (1906) : Address to Economic Science and Statistics Section of British Association for the Advancement of Science. *JRSS*, 69, pp. 540-557.

__________ (1926) : Measurement of precision attained in sampling. *Bull. Internat. Stat. Inst.*, 22, pp. 1-62.

Cochran, W.G. (1940) : The estimation of the yields of creal exports by sampling for the ratio of grains to total produce. *Jour. Agri. Sci.*, 30, pp. 262-275.

__________ (1942) : Sampling theory when the sampling units are of unequal sizes. *J. Amer. Statist. Assoc.*, 37, pp. 199-212.

__________ (1963) : *Sampling Techniques* (First edition,1953). Jphn Wiley and Sons, New York.

__________ (1977) : *Sampling Techniques* (Second edition,1963). John Wiley and Sons, New York.

Dalenius, T. (1957) : *Sampling in Sweden.* John Wiley and Sons.

__________ (1962) : Recent advances in sample survey theory and methods. *AMS*, 33, pp. 325-349.

Dalenius, T. (1965) : *Current Trends in the Development of Sample Survey Theory and Methods.*

Das, A.K. and Tripathi, T.P. (1977) : Admissible estimators for quadratic forms in finite populations. *Bull. Inter. Stat. Inst.*, 47, *Book* 4, pp. 132-135.

____________________ (1978a) : Use of auxiliary information in estimating the finite population variance. *Sankhyā, Ser. C*, 40, pp. 139-148.

____________________ (1978b) : Use of ancillary information in estimating the coefficient of variation of a finite population. *Stat-Math. Tech. Report No. 35/78, I.S.I., Calcutta.*

____________________ (1979a) : A class of estimators for the population mean when the mean of an auxiliary character is known. *Stat-Math. Tech. Report No. 22/79, I.S.I., Calcutta.*

____________________ (1979b) : A class of sampling strategies for population mean using knowledge on the variance of an auxiliary character. *Stat-Math. Tech. Report No. 30/79, I.S.I., Calcutta.*

____________________ (1980a) : Estimation of a population parameter using information on several statistics. *Stat-Math. Tech. Report No. 13/80, I.S.I., Calcutta.*

____________________ (1980b) : Sampling strategies for population mean, when the coefficient of variation of an auxiliary character is known. *Sankhyā, Ser. C*, 42, *Pts.* 1 & 2, pp. 76-86.

Das, A.K. (1982a) : Estimation of population ratio on two occasions. To appear in *Jour. Ind. Soc. Agri. Stat.*, 34, *No.*2, 1982.

________ (1982b) : On the use of auxiliary information in estimating proportions. To appear in *Jour. Ind. Stat. Assn.*, 20, 1982.

________ (1982c) : Contributions to the theory of sampling strategies based on auxiliary information. *Unpublished Ph.D. Thesis submitted to the Indian Statistical Institute, Calcutta.*

Godambe, V.P. (1955) : A unified theory of sampling from finite populations. *JRSS, Ser. B*, 17, pp. 269-277.

__________ (1960) : An admissible estimator for any sampling design. *Sankhyā*, 22, pp. 285-288.

__________ (1965) : A review of the contributions towards a unified theory of sampling from finite populations. *Bull. Int. Stat. Inst.*, 33, pp. 242-258.

__________ (1976) : "A historical perspective of the recent developments in the theory of sampling from actual populations". — *Dr. Panse memorial lecture organised by Ind. Soc. Agri. Stat., New Delhi, 29th March, 1976.*

Hansen, M.H. and Hurwitz, W.N. (1943) : On the theory of sampling from finite populations, *AMS*, 14, pp. 333-362.

Hansen, M.H., Hurwitz, W.N. and Madow, W.G. (1953) : *Sample Survey Methods and Theory*, Vol.II, John Wiley and Sons.

Hanurav, T.V. (1966) : Some aspects of unified sampling theory. *Sankhyā, Ser.A*, 28, pp. 175-204.

Hirano, K. (1972) : "Using some approximately known coefficient of variation in estimating the mean. *Research memorandum No. 49, Inst. Stat. Math., Tokyo, Japan.*

㊉ Horvitz, D.G. and Thompson, D.J. (1952) : A generalization of sampling without replacement from a finite universe. *J. Amer. Statist. Assoc.*, 47, pp. 663-685.

Khan, S. and Tripathi, T.P. (1967) : The use of multivariate auxiliary information in double sampling. *Jour. Ind. Stat. Assn.* 5, pp. 42-48.

Kiaer, A.N. (1895) : "Observations et expériences concernant les denombrements réprésentatifs". *Bull. Inst. Int. Stat.*, 9, *div. I*, pp. 176.

㊉ Hoerl, A.E, and Kennerd, R.W. (1970a): Ridge regression: Biased estimation of non-orthogonal problems. *Technometrics*, 12, pp. 55-67.

__________ (1970b): Rigde regression: Applications to non-orthogonal problems. *Technometrics*, 12, pp. 69-82.

Koop, J.C. (1957): Contributions to the general theory ~~of~~ sampling finite populations with replacement and with unequal probabilities. *Ph.D. Thesis, N. Carolina State College Library, Raleigh : North Carolina Inst. Stat. Mimeo. Ser.* 296.

__________ (1963): On axioms of sample formation and their bearing on construction of linear estimators in sampling theory of finite universes. *Metrika*, 7, pp.81-114 *and* 165-204.

Liu, T.P. (1974) : A general unbiased estimator for the variance of a finite population. *Sankhyā, Ser. C*, 36, *pp.* 23-32.

Maiti, Pulakesh and Tripathi, T.P. (1976) : The use of multivariate auxiliary information in selecting the sampling units. *Proceedings of the symposium on recent developments in survey methodology, I.S.I., Calcutta, March,* 1976.

__________ (1978) : Use of prior information on some parameters in estimating population mean. *Stat-Math. Tech. Report No. 23/73, I.S.I., Calcutta - Presented before the 33rd annual conference of the Indian Society of Agricultural Statistics, held at Trichur, December,* 1979.

__________ (1980a) : Some results on T_1-class of linear estimators. *Stat-Math. Tech. Report No. 25/80, I.S.I., Calcutta - Presented at the Conference on mathematical statistics and probability theory in honour of Prof. C.R. Rao to mark his 60th birthday at New Delhi, December, 8-13, 1980.*

__________ (1980b) : Estimation of population total using information on some units of the population. *Stat-Math. Tech. Report No. 8/80, I.S.I., Calcutta.*

__________ (1981a) : Some T_2-class of estimators better than H-T estimator. *Presented before 34th Annual Conference of the Indian Society of Agricultural Statistics, held at Lucknow, December 23-25, 1980. Aligarh Journal of Statistics,* 1, *pp.* 52-58.

Maiti, Pulakesh and Tripathi, T.P. (1981): A class of estimators for population variance using some apriori information. *Stat-Math. Tech. Report No. 22/81, I.S.I., Calcutta - Presented at the Golden Jubilee Conference on Statistics : applications and new directions, held at I.S.I., Calcutta*, December, 1981.

Maiti, Pulakesh, Tripathi, T.P. and Sharma, S.D. (1982): Use of prior information on some parameters in estimating population mean. To appear in *Sankhyā*, Sreis A, 1983.

Malda District Hand Census Hand Book (1961) : *Census of India.*

Mehta, J.S. and Srinivasan, R. (1971): Estimation of the mean by shrinkage to a point. *JASA*, 66, pp. 86-90.

Midnapur District Hand Book (1961) : *Census of India.*

Murthy, M.N. (1963a) : Some recent advances in sampling theory, JASA, 58, pp. 737-755.

__________ (1963b) : Generalised unbiased estimation in sampling from finite populations. *Sankhyā, Ser.* B, 25, *pp.* 245-261.

__________ (1967) : *Sampling Theory and Methods.* Statistical Publishing Society, Calcutta.

Neyman, J. (1934) : On the two different aspects of the representative method : the method of stratified sampling and the method of purposive selection. *JRSS*, 97, *pp.* 558-606.

Olkin, I. (1958) : Multivariate ratio estimation for finite populations. *Biometrika*, 45, *pp.* 145-165.

Pandey, B.N., Singh, J. and Hirano, K. (1973): On the utilisation of a known coefficient of kurtosis in the estimation procedure of variance. *Ann. Inst. Stat. Math.*, 25, *pp.* 51-55.

Pandey, B.N. and Singh, J. (1977) : Estimation of variance of normal population using prior information. *Jour. Ind. Stat. Assn.*, 15, pp. 141-150.

Pandey, B.N. and Singh, J. (1978) : On the use of coefficient of variation in estimating the mean. The *Vikram*, XXII, *No. 1 and 3, pp.* 117-122.

Raj, D. (1954) : On sampling with probabilities proportional to size. *Ganita*, 5, *pp.* 175-182.

______ (1958) : On the relative accuracy of some sampling techniques. *JASA*, 53, pp. 98-101.

______ (1965) : On a method of using multiauxiliary information in sample surveys. *JASA*, 60, *pp.* 270-277.

______ (1968) : *Sampling Theory*. Mc. Graw Hill Book Co., New York.

Ramachandran, G. (1978) : Choice of strategies in survey sampling, pp. 147. *Ph.D. thesis submitted to the Indian Statistical Institute, Calcutta.*

Rao, C.R., Mahalanobis, P.C. and Majumder, D.N. (1948-49): Anthropometric survey of the United Provinces : a statistical study. *Sankhyā*, 9, *pp.* 90-324.

Rao, J.N.K. and Pereira, M.P. (1968) : On double ratio estimator. *Sankhyā, Ser. A*, 30, *pp.* 83-90.

Rao, Podur. S.R.S. and Mudholkar, .S. (1967) : Gene. lised multivariate estimators for the mean of a finite population. *JASA*, 62, *pp.* 1009-1012.

Roy, J. and Chakraborty, I.M. (1960) : Estimating the mean of a finite population. *Ann. Math. Stat.*, 31, *pp.* 392-398.

Schott, S. (1923) : Statistik. Teubner, *Leipzig-Berlin*, pp. 43-44.

Sen, A.R. (1971) : Successive sampling with two auxiliary variables. *Sankhyā, Ser. B*, 33, *pp.* 371-378.

Searls, D.T. (1964) : The utilisation of a known coefficient of variation in the estimation procedure. *JASA*, 59, *pp.* 1225-1226.

Singh, M.P. (1967) : Multivariate product method of estimation for finite populations. *Jour. Ind. Soc. Agri. Stat.* *XIX, pp.* 1-10.

________ (1969) : Some aspects of estimation in sampling from finite populations. *Ph.D. thesis submitted to the Indian Statistical Institute, Calcutta.*

Srivastava, S.K. (1965) : An estimate of the mean of a finite population, using several auxiliary variables. J. *Ind. Stat. Assoc.*, 3, pp. 189-194.

Srivastava, S.K. and Jhajj, Harbans Singh (1980) : A class of estimators using auxiliary information for estimating finite population variance, *Sankhyā*, *Ser. C*, 42, *Pts. 1 & 2*, pp. 87-96.

Sukhatme, P.V. (1944) : Moments and product moments of moment statistics for samples of the finite and infinite populations. *Sankhyā*, 6, pp. 363-382.

__________ (1959) : Major developments in the theory and applications of sampling during the last 25 years. *Estadica*, 17, pp. 652-679.

__________ (1966) : *Major Developments in Sampling Theory and Practice, i.e., research papers in Statistics* (Edt. J. Neyman), John Wiley, London, pp. 367-409.

Sukhatme, P.V. and Sukhatme, B.V. (1970) : *Sampling Theory of Surveys with Applications*. Asia Publishing House, India.

Tripathi, T.P. (1969) : A regression type estimator in sampling with pps and with replacement. *Aust. Jour. Stat.*, 11, pp. 140-148.

__________ (1970) : Contributions to the theory using multivariate information. *Ph. D. thesis submitted to the Punjabi University, Patiala.*

__________ (1973) : Double sampling for inclusion probabilities and regression method of estimation. *Jour. Ind. Stat. Assn.*, 10, pp. 33-46.

__________ (1976a) : On double sampling for inclusion probabilities and regression method of estimation. *Jour. Ind. Stat. Assn.*, 28, pp. 33-54.

Tripathi, T.P. and Sinha, S.K.P. (1976b) : Estimation of ratio on successive occasions. Proceedings "*Symposium on recent developments in survey methodology*", *held at I.S.I., Calcutta*, March, 1976.

Tripathi, T.P. (1978) : A note on optimum weights in multivariate ratio, product, and regression estimators. *Jour. Ind. Soc. Agri. Statist.*, XXX, No.1, pp. 101-109.

Tripathi, T.P. and Srivastava, O.P. (1979) : Estimation on successive occasions using ppswr sampling. *Sankhyā, Ser. C*, **41**, pp. 84-91.

Tripathi, T.P. (1980) : A general class of estimators for population ratio. *Sankhyā, Ser. C*, **42**, *Pts. 1 & 2*, pp. 63-75.

Tikkiwal, B.D. (1967) : Theory of multiphase sampling from a finite or an infinite population on successive occasions 1,2. Rev. *Inst. Internat. Stat.*, **35**, pp. 247-263.

Thompson, J.R. (1968) : Some shirnkage techniques for estimating the mean. *JASA*, **63**, pp. 113-123.

Wakimoto, K. (1970) : On unbiased estimation of population variance based on the stratified random sample. *Ann. Inst. Stat. Math.*, **22**, pp. 15-26.

__________ (1971) : Stratified random sampling (I) : Estimation of the population variance. *Ann. Inst. Stat. Math.*, **23**, pp. 233-252.