CONTRIBUTIONS TO THEORIES OF REPETITIVE SURVEY SAMPLING STRATEGIES

To Prof P. K. Bose with best regards
Raghunath Arnab
3/5/79

RAGHUNATH ARNAB

A thesis submitted to the
Indian Statistical Institute in fulfilment of the
primary requirements for the award of the degree of

DOCTOR OF PHILOSOPHY

CALCUTTA.
1979

P R E F A C E

This thesis is being submitted to the Indian Statistical Institute in fulfilment of the primary requirements for the award of the degree of Doctor of Philosophy.

No part of this thesis was submitted to any other Institute or University for any degree, diploma, prizes etc.

In this thesis we have presented some results relating to problems of sampling finite populations on repeated occasions with the principal objective of estimating the population total for the current occasion. We have also incidentally considered auxiliary topics relating to strategies for sampling on a single occasion alone in getting insights into problems relating to our main concern about repetitive surveys.

Certain parts of the thesis are based on the results which have been and are being published in journals under the following titles:

(i) "On the relative efficiencies of a few strategies of sampling with varying probabilities on two occasions" : jointly with Arijit Chaudhuri, Cal. Statist. Assoc. Bull. 26, 25-38 (1977).

(ii) "On the role of sample-size in determining efficiency of Horvitz-Thompson estimators" jointly with Arijit Chaudhuri. Sankhyā Ser C, 40, 104-109 (1978).

(iii) "On estimating the mean of a finite population sampled on two occasions with varying probabilities" jointly with Arijit Chaudhuri, accepted for publication in Aust. Jour. Statist (to appear in 1979).

(iv) "On the relative efficiencies of sampling strategies under a super population model" jointly with Arijit Chaudhuri, accepted for publication in Sankhyā Ser C, 41 (to appear in 1979).

(v) "On the relative efficiencies of some sampling strategies under a super population model" accepted for publication in Jour. Ind. Soc. Agri. Statist (to appear in 1979).

(vi) "On non-negative variance-estimation" jointly with Arijit Chaudhuri re-submitted to Metrika after implementing referee's suggestion for revision.

(vii) "A few strategies for two-stage sampling over two occasions and their relative efficiencies" re-submitted to Aust. Jour. Statist. after implementing referee's suggestions for revision.

(viii) "Surveying a finite population on more than two occasions" Tech. report ASC/16. Indian Statistical Institute.

This dissertation was prepared under the active guidance of Dr. Arijit Chaudhuri to whom I owe a deep debt of gratitude for his supervision, constant encouragement throughout the work and permission to include some of our joint works in the thesis.

I must also gratefully acknowledge the encouragement and support received from Prof. J. Roy in getting an opportunity to work for a Ph.D programme at the Indian Statistical Institute, Calcutta.

I am also grateful to the Principal and the Head of the Department of Statistics, R.K. Mission Residential College, Narendrapur, who generously granted me a study leave to work at the ISI.

I also wish to thank Messrs. M. Paul, P. Bose and C.R. Nayak for their help in certain stages of the preparation of this thesis and finally my heartfelt thanks go to Mr. M.K. Chandra who typed the thesis with care and competence.

Raghunath Arnab
March, 1979

Applied Statistics, Surveys
and Computing Division,
Indian Statistical Institute,
Calcutta.

C O N T E N T S

CHAPTER 1

FORMULATION OF THE PROBLEM AND SUMMARY OF RESULTS

1.1 INTRODUCTION

A common practical problem to which a survey - sampler has frequently to address himself is one of sampling a given finite population on successive occasions. One of the relevant issues requiring one's attention then is to adopt a suitable sampling strategy to estimate the population total of a variate of interest on the current occasion in an optimal manner. Here one has of necessity to take care to utilize the accumulated data on that variate procured in course of the survey along with other auxiliary information on one or more additional variables that may also incidentally be available. Several such strategies are well known in the literature. In this context significant contributions have been from Jessen (1942), Patterson (1950), Tikkiwal (1950, 1951, 1964, 1965), Keyfitz (1951), Hansen-Hurwitz and Madow (1953), Kulldorff (1963), Rao, J.N.K. and Graham (1964), Raj (1965), Ghangurde and Rao J.N.K. (1969), Fellegi (1966), Singh (1968), Avadhani and Sukhatme (1970), Ravindra Singh (1972), Sen (1973), Chotai (1974) and Lanke (1975). Most of the strategies involve selection of units with equal probabilities on each occasion in one or more stages. Only those due mainly to Raj (1965), Ghangurde and Rao J.N.K. (1969) and Chotai (1974) call for selection with varying probabilities. Since the problem is formulated in a finite population set-up we cannot have a best strategy in general. So, what we may aim at is to have one that may be claimed to be optimal in a certain sense and to get equipped with tools with which we may judge the circumstances in which a proposed one may be anticipated to fare better or worse than its possible competitors and thus to be able to choose a course of action to follow in particular situations. So, a major portion of the thesis is devoted to suggesting additional sampling strategies for estimating finite population totals on a current occasion, by choosing samples in two or more occasions in one or more stages with varying probabilities of selection using auxiliary information in the form of what are commonly known as size-measures along with knowledge of values of the variate of interest on previous occasions as well. For the suggested strategies we mainly seek to recommend proportions of units to be matched on successive occasions and compare the efficiencies of the resulting strategies relative to their rivals often under various models popularly treated in survey sampling literature. Incidentally, as we observe that an appropriate sampling strategy to employ on two occasions depends on the appropriateness of the strategies for sampling separately on the current and the previous occasions, we also pay heed to how one should choose an optimal sampling strategy on a single occasion alone. Literature is replete with results on this issue. Yet we find it worthwhile for thesake of completeness to extend the study of relative efficiencies of several well-known strategies for sampling on a single occasion. Further, we note a few associated problems in the context of sampling on a single occasion, namely, determination of optimal-sample size, removing the bias of product estimators without losing efficiency, non-negative estimation of variance, estimation of variance in two-stage sampling. Results of such queries are presented in the chapters 2 and 3. These are followed by our study of strategies applicable on two occasions for selecting the samples only in one stage, presented in the chapter 4. The chapter 5 takes care of those strategies applied on successive occasions in two stages. In chapter 6 we have proposed some strategies for sampling on more than two occasions. In the final chapter we take up another specific problem of uni-stage sampling on two occasions. Here some aspects of the sampling schemes to be adopted for the two occasions are supposed to be fixed and subject to the consequential constraints on them, the problem we treat is about maximization of the expected number of common units to be surveyed on both the occasions - a step obviously to relieve the pressure on the budget. We offer a few alternatives to the earlier results in this area due to Keyfitz (1951), Fellegi (1966) and Lanke (1977) and make a few comparisons.

In what follows we lay down a brief summary of our treatment of various issues elaborated in the succeeding chapters, preceded by a short review in outlines of the salient aspects of the literature on the subject.

1:2 A BRIEF REVIEW OF SOME ASPECTS OF THE LITERATURE ON SURVEY SAMPLING ON SUCCESSIVE OCCASIONS

From the records we find that the pioneering work on repeated sampling was done by Jessen (1942). To estimate a finite population mean on a current occasion he considered a sample drawn on a previous occasion, retained a sub-sample from it and selected an independent sample from the entire population [supposed, as throughout the thesis, to remain in tact over all occasions considered], the total sample-size being maintained fixed on both occasions [this also will be the general practice for this thesis]. Each sample is a simple random sample chosen without replacement. Then he considered two separate estimators, namely (1) the sample mean of the new units only (2) and a double-sampling regression estimator based on the current as well as the previous values of the common units, and of the values of the units surveyed earlier only. Finally, he considered the two estimators with weights chosen to minimize the variance of the combined estimator in terms of the variances of the separate estimators. The resulting optimal estimator, however, is usable only when some information on the population characteristics is available and some conditions are imposed on the nature of variate-values for the two occasions. This is a disturbing feature to be observed throughout the literature on this subject of repetitive survey sampling problems. Nevertheless, the study seems fascinating. Because, one should be curious about how the theory develops although it may not often be put into practice owing to limitations of resources.

Yates (1949) extended Jessen's (1942) study to the case when sampling is repeated h (≥ 2) times. But he worked under the restriction of a constant 'total sample size' and 'matching sampling fraction' over the occasions, and postulating the correlation coefficients of the values for the units over k-occasions-apart to be of the form ρ^k (for $0 \leq k \leq h$). Some of his restrictions were later relaxed by Patterson (1950) and Tikkiwal (1951). For example, ρ^k was replaced by a more general ρ_k (for $k = 1,\ldots,h$) and the later by $\rho_{r,k} = \prod_{t=r}^{r+k} \rho_{t,t+1}$ with some starting point $r = 1,\ldots h-1$; $k = 1,2,\ldots h-1$. Eckler (1955) stretched the ideas further to develop the theory of rotation - sampling and construct suitable linear combinations of the sample observations for various occasions to estimate the current mean. In these investigations although the problem is to estimate a finite population mean, the optimalities of suggested sampling strategies are established under the postulation of an underlying infinite population set-up. But Rao and Graham (1964) developed a truly finite population theory for composite estimators of the current mean and changes in the mean values over various occasions with rotation sampling schemes. Kulldorff (1963) modified Jessen's (1942) method to choose the un-matched sample on the second occasion from the remainder of the population leaving aside the units chosen earlier. Pathak and Rao, T.J. (1967) effected an improvement on Jessen's (1942) strategy by Rao-Blackwellization on the latter's estimator noting that Jessen's 'un-matched' sample may have some units coincident with the 'matched' one. Ravindra Singh (1972) and Rao-Blackwellized also suggested another modification by taking the unmatched sample from the population leaving aside only the 'matched' portion. Sen (1973) allowed difference in the total sample-size on two occasions and considered the situations when on the matched units values of p (≥ 2) auxiliary variables on the first occasion are available. Following Olkin (1958) and Goswami and Sukhatme (1965) for the matched sample he used a multivariate double sampling ratio estimator and combined it with the mean of the unmatched sample for the second occasion using weights minimizing the variance of the combined estimator. In all the papers mentioned thus far selections are made with equal probabilities. Raj (1965) for the first time recommended selection with varying probabilities on two occasions. He assumed [as we shall, throughout the thesis] that normed size-measure p_i's are available for the units on both the occasions. His initial sample is chosen with probabilities proportional to size with replacement (PPSWR) out of which a simple random 'matched' sub-sample is taken without replacement (SRSWOR) on the second occasion and that is supplemented by a fresh PPSWR sample from the entire population, the total sample size on the two occasions being taken as the same. For the matched portion he used a different estimator of the Hansen-Hurwitz (1943) type using data for the initial sample on the first occasion and for both occasions on the matched units. This is then linearly combined with the Hansen-Hurwitz type estimator based on the 'unmatched' sample for the second occasion using weights chosen to minimize the variance of the combined estimator. As usual the problem is optimal choise of matching proportion and weights for the final estimator and the solution is feasible under stringent restrictions on the structure of variate-values and availability of information about values of certain parameters. An additional disadvantage stems from the possibility of the same units occurring repeatedly in the initial sample rendering artificiality on the 'matched sub-sampling' thereform. Pathak and Rao (1967) applied Rao-Blackwellization to improve upon the efficiency of Raj's strategy just as they did on Jensen's (1942), seizing the opportunity of isolating the distinct units. Ravindra Sing (1972) also approached likewise to effect a further improvement by seggregating the repeated units from the 'matched'sample. Ghangurde and Rao, J.N.K. (1969) for the first time selected units with varying probabilities without replacement from a finite population sampled on two occasions with a time-lag. As a matter of fact we draw inspiration more from the joint article than from any other in setting out to prepare this dissertation. They choose the initial sample employing the well-known Rao-Hartley-ochran (RHC, in brief) scheme. The matched sample is an RSWOR sub-sample therefrom and the un-matched sample is nother RHC sample drawn from the entire population on the second occasion keeping in tact the total sample-size for both the occasions What would be more desirable, however, as one may note, was to choose the un-matched sample from the remainder of the population leaving aside either the initial sample or the 'matched' sub-sample at least. But either scheme is difficult to implement as the authors themselves admitted! We also concede the point and humbly submit that this will not be attempted in what follows. Their estimator is of the same form as Raj's except that they use the RHC-type estimator. The resulting strategy, as expected, is more efficient than Raj's in most situations. But the resulting optimal strategy ought to have a limited applicability as of course any other in this area. Chotai (1974) modified their strategy by choosing the 'matched' sub-sample again employing RHC scheme and investigated relative efficiencies of the three strategies involving selection with varying probabilities.

Avadhani and Sukhatme (1970) recommended a procedure when the initial SRSWCR is chosen and surveyed to yield variate-values providing normed size-measures to be utilized in drawing therefrom a matched sample according to RHC scheme. The un-matched sample is then taken as an SRSWOR from the residual population minus the initial sample, the total sample sizes for the two occasions not being identical. The estimator employed is of the same form as in Ghangurde-Rao (1969). In an alternative strategy they chose all three samples as SRSWOR but in the final estimator RHC estimator is replaced by the usual double sampling ratio estimator. They have studied the relative efficiencies of their two strategies and to facilitate their study introduced a new model which we shall throughout call model II (to be written as M_2) as distinct from the model I (to be denoted as M_1), introduced in survey sampling literature earlier by Cochran (1946) and effectively studied by Hanurav (1967), Rao (1967) and many others. These are the principal references which stimulated our research interest in this area. Besides, there are a few subsidiary ones due to Singh (1968), Abraham, Khosla and Kathuria (1969), Singh and Kathuria (1969), Kathuria (1975) among others who employed multi-stage sampling techniques while repeatedly sampling a population, the selection being usually with equal probabilities without replacement with or without stratification. In subsequent chapters we shall have more elaborate discussions on these strategies which will have direct bearings on our works. So, now let us briefly narrate the directions of our own research efforts.

1.3 A BRIEF (CHAPTER-WISE) RESUME' OF THE RESULTS FOUND

1.3.1 Chapter 2

A reasonable strategy to estimate a mean on the current occasion from the current and past data is first to consider two appropriate strategies for the two occasions -, current and previous - and to combine them in an effective manner. So, it is needful to ascertain an optimal strategy for a single survey alone. Various results are already there concerning the relative performances of the well-known strategies due to Horvitz-Thompson (1952), Rao-Hartley-Cochran (1962) and Midzuno-Sen (1952, 1953). We add a few more relating to them under certain familiar models considered in survey sampling literature due mainly to Cochran (1946), Avadhani and Sukhatme (1970). Chaudhuri's (1977) observations about irregularities in the behaviour of variances of HTE vis-a-vis sample-size led us here to enquire about the problem in greater depths. Turning to the role of product estimators in case the ancillary character has a negative correlation with the one under study methods have been given to obtain unbiased product estimators with no loss in asymptotic efficiencies for general sampling schemes.

1.3.2 Chapter 3

A few 'negative' results are derived concerning the problem of 'non-negative' estimation of variances of a few estimators. Limitations of some earlier results in this context due to Rao, T.J. (1972, 1977) and Chaudhuri (1976) are emphasized. For multi-stage sampling the problem of unbiased estimation of variances has got some interesting features as observed by Durbin (1953, 1967) and Raj (1968). For specific sampling schemes it is known that if the first stage units are chosen with replacement then one can estimate the variance of the estimator for any scheme of sampling in later stages. But the situation is restrictive when the first stage units are chosen without replacement. This point has been carried further and results applicable under more general circumstances have been derived.

1.3.3 Chapter 4

Several strategies of sampling a finite population on two occasions with varying probabilities have been suggested as alternatives mainly to those due to Raj (1965), Ghangurde and Rao (1969), Avadhani and Sukhatme (1970) and

Chotai (1974). Situations under which they fare promisingly have been studied in details under assumptions of various popular models. Techniques of double sampling for stratification have also been employed in dealing with the problem of sampling on two occasions.

1.3.4. Chapter 5

Various strategies for sampling a population on two occasions in two stages are well-known in the literature. These are further generalized and a few particular techniques are proposed as possible alternatives to the existing ones. Situations in which they are more efficient have been investigated and a number of positive results have been derived.

1.3.5 Chapter 6

Strategies proposed by Raj (1965), Ghangurde and Rao (1969) and Chotai (1974) for sampling a finite population with varying probabilities on two occasions have been here extended to suggest methods for sampling on any number of occasions with varying probabilities to yield current estimates. The relative efficiencies of the resulting extended strategies have been investigated in details

1.3.6 Chapter 7

Following Keyfitz (1951), Fellegi (1966) and Lanke (1975) alternative devices have been suggested for ensuring a sizeable overlap between samples on two occasions under prior constraints on inclusion-probabilities of units on two occasions in accordance with their size-measures varying over the occasions. Optimal results are not obtained but alternatives have been suggested along with clues for choosing among the competitors

CHAPTER - 2

SURVEYING A POPULATION ONLY ONCE - ESTIMATING POPULATION MEANS (TOTALS)

2.1 MOTIVATION AND SUMMARY

In estimating the current mean (or total) value of a finite population surveyed on the current and a previous occasion, one can profitably utilize the outcomes of both the surveys. To choose the required strategy for the purpose in an optimal manner one way is to devise optimal sampling strategies separately for estimating population mean (or total) on the current as well as on the previous occasion. Various such strategies for sampling a finite population only once are well-known. Their relative efficiencies have also been extensively investigated in the literature. But we consider it needful to supply a few supplementary results concerning the strategies due to Horvitz-Thompson (1952)(HT), Rao-Hartley-Cochran (1962) (RHC) and Midzuno-Sen (1952,1953) (MS). It is well known that the HT estimator (HTE) is studded with many optimal properties. But one of its disquieting features is that when based on some particular schemes it may not behave nicely [vide Chaudhuri (1977)] in the sense of having its variance decreasing monotonically with increasing sample size. So here we study this problem a little further. The above-mentioned strategies are useful in case the variate of interest and the auxiliary variate are correlated positively. In case this correlation is negative, usually one employs 'product estimators'. But these are generally biased. So we present a modification to eliminate this bias without sacrificing asymptotic efficiency.

2.2 INTRODUCTION, FORMULATION AND STATEMENT OF THE PROBLEM

If t_1 and t_2 be two uncorrelated unbiased estimators for a parameter θ having variances σ_1^2 and σ_2^2, then, for a weight ϕ with $0 \leq \phi \leq 1$, the combined estimator

$$t = \phi t_1 + (1-\phi) t_2$$

is unbiased for θ. The choice of ϕ that minimizes var(t) is

$$\phi_{opt} = \frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2}$$

and the corresponding t has the minimum variance

$$V_{min}(t) = \frac{\sigma_1^2 \ \sigma_2^2}{\sigma_1^2 + \sigma_2^2} = V \text{ (say)}.$$

If t_1' and t_2 be two alternative uncorrelated unbiased estimators for θ with variances $\sigma_1'^2$ and $\sigma_2'^2$ the corresponding linear combination

$$t' = \psi t_1' + (1-\psi) t_2'$$

will naturally have a minimum variance

$$V_{min}(t') = \frac{\sigma_1'^2 \sigma_2'^2}{\sigma_1'^2 + \sigma_2'^2} = V'.$$

It follows that if $\sigma_1'^2 \geq \sigma_1^2$ and $\sigma_2'^2 \geq \sigma_2^2$, then $V' \geq V$. For the current mean (or total) of a finite population surveyed twice, an estimator of the form t is often employed (as we shall see in details in chapters 4 and 5), t_1 and t_2 being uncorrelated unbiased estimators for the current and previous totals based only on data of the respective occasions. For the repetitive surveys, the strategies we shall consider in this thesis (as well as those generally studied by our predecessors) will involve not only uncorrelated stimators t_1 and t_2 but also correlated ones. But their correlation (as we shall see in chapters 4 and 5) will be of such a form, that essentially the problem will be to control a quantity of the form $\sigma_1^2\sigma_2^2/(\sigma_1^2+\sigma_2^2)$ by choosing t_1 and t_2 with controlled magnitudes of their respective variances. It follows, therefore, that the best course to follow in estimating the current total from data on both the surveys is to adopt the best available strategies for estimating separately the totals on the two occasions. But, as it is well known that in sampling a finite population 'best strategies are generally not available at all, we shall be in quest for the appropriately optimal rather than the ellusive 'best' strategies in various circumstances.

So, while treating the problem of devising strategies for sampling on successive occasions one has to pause to pay heed to the estimation problem when surveying a finite population once only.

To save space we shall eschew detailed discussions on the unified theory of inference relating to sampling a finite population. Rather, we shall only recall that the strategies due to Horvitz-Thompson (HT) (1952), Rao, Hartley, Cochran (RHC) (1962) and Midzuno, Sen (MS) (1952, 1953) occupy prominent places in it. We shall denote them respectively by H_1, H_2, H_3 the corresponding estimators for the population total $T = \sum_{i=1}^{N} y_i$ by t_1, t_2, t_3 and their respective variances by v_1, v_2 and v_3. Let us also note the following model M_1 the socalled superpopulation model of Cochran (1946) which we shall assume in studying their relative efficiencies as were earlier done by Rao, J.N.K. (1966 a,b), Hanurav (1967), Rao, T.J. (1967) among many others

By y_i, x_i ($i = 1, \ldots, N$) we denote the values assumed by the variate y under study and an auxiliary variate x for the ith unit of a finite population of size N. We shall assume throughout that x_i (> 0) is known for every $i = 1, \ldots, N$. The model M_1, in brief, is as follows.

$$\text{Model } M_1 : \quad y_i = \beta x_i + e_i, \quad i = 1, \ldots, N$$
$$\varepsilon(e_i | x_i) = 0, \quad i = 1, \ldots, N$$
$$\varepsilon(e_i^2 | x_i) = \sigma^2 x_i^g, \quad i = 1, \ldots, N$$
$$\varepsilon(e_i e_j | x_i, x_j) = 0, \quad i \neq j = 1, \ldots, N$$

where β, σ^2 and g are unknown constants, e_i's are values of a random variate e having a probability distribution with $\varepsilon(.|.)$ as the operator for its conditional expectation for given x-values. From extensive empirical evidences [Smith, H.F. (1938), Mahalanobis (1944), Brewer et al (1977)] it has been found reasonable to assume that $0 \leq g \leq 2$ and quite frequently, $1 \leq g \leq 2$. We too shall assume as such but in a particular case also examine what happens if g exceeds 2. The (model-) expected value of v_i will be denoted by E_i ($i = 1,2,3$). From Hanurav (1967) and Rao, T.J. (1967) we know that

$$\left.\begin{array}{l} E_1 \lesseqgtr E_2 \\ E_1 \lesseqgtr E_3 \end{array}\right\} \text{ according as } g \gtreqless 1.$$

and

both inequalities reducing to equalities in case $g = 1$.

So, if a sampler has reasons to believe on the evidence of the data he visualizes that $g > 1$, he should prefer H_1 to H_2 and H_3. If he realises that $g < 1$, then surely he should not use H_1 but has to choose between H_2 and H_3. But the question is 'how can be do so'? We provide an answer in what follows by establishing that

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

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

all inequalities reducing to equalities when $g = 1$. This facilitates a unique choice out of H_1, H_2, H_3 under model M_1 incase one can guess about g being inside or outside $[0, 1]$.

Our results are exact. But Rao, J.N.K. (1966,a) observed the following asymptotic (for large population size N and and large sample size n) result in this context, viz, $E_3 \gtrless E_2$ according as $g \gtrless 1$. In case in the model M_1, $\sigma^2 x_i^g$ is replaced by $v(x_i)$ ($i = 1, \ldots, N$) for some function v of x, we denote the model as M_1' and the E_j's by E_j''s

($j = 1,2,3$). Under M_1' we have further results concerning the relative efficiencies of H_1, H_2 and H_3 as shown below.

Several authors like Quenouille (1956), Rao and ebster (1966), Durbin (1967), Rao, P.S.R.S. (1971) and Chakrabarty (1973), Singh (1975) among others considered an extended model M_1'' (say), where in M_1, one has $y_i' = \alpha + \beta x_i + e_i$ with α as an additional unknown additive constant and allows x to be a random variable having specific distributions. Frequently, x is supposed to have a gamma distribution with a single unknown (mean-) parameter or else, x,y are supposed to have a joint bivariate normal distribution with unknown parameters. Avadhani and Sukhatme (1970) ntroduced another model to be called Model M_2 - a truly inite population model under which one stipulates the following :

$$\text{Model } M_2: \quad y_i = \beta x_i + e_i, \quad i = 1, \ldots N$$

$$\text{such that} \quad \sum_1^N e_i = 0 = \sum_1^N e_i x_i$$

$$\text{and} \quad \bar{e}_j^2 = \sigma^2 x_j^g,$$

here σ^2 and g are unknown constants and $\bar{e}_j^2$ is the mean of e_{ji}^2's, viz. the squared residuals for the jth array which s the collection of sampling units for which the value of x is x_j. We shall consider the case where each x_j is distinct so as to write $e_j^2 = \sigma^2 x_j^g$ for every $j = 1, \ldots N$. Here also g is supposed to lie between 0 and 2 although no empirical evidence is known to support this postulation. Several results concerning relative efficiencies are known in the literature under all these models - to save space we shall not report them in detail. We shall only offer a few of our own and consider, besides the three strategies already mentioned, one more viz H_4, the one involving the ratio-estimator based on SRSWOR method.

Let us present the results with accompanying clarifications about motivations and notations.

2.3 COMPARISON OF STRATEGIES UNDER CERTAIN MODELS

2.3.A. RELATIVE EFFICIENCIES OF STRATEGIES UNDER M_1 and M_1'

The problem is to estimate $Y = \sum_1^N y_i$ using the known values x_i, $i = 1, \ldots N$ for which $X = \sum_1^N x_i$ and $p_i = \frac{x_i}{X}$, $i = 1, \ldots N$. We consider the following strategies viz

H_1: The Horvitz-Thompson (1952) strategy (HT-strategy):

The units are selected in any manner such that inclusion-probabilities are

$\pi_i = \nu P_i$ for the ith unit $(i = 1, \ldots N)$

(ν being the expected sample-size, frequently, we shall restrict to the case when the effective sample-size is a fixed interger n).

The estimator used is

$$t_1 = \sum_{i \in s} \frac{y_i}{\pi_i}$$

where a sample s is selected following any scheme satisfying the above condition and being called a 'πps scheme'. By π_{ij} we shall denote the corresponding inclusion probabilities of the pair of ith and jth units $(i \neq j = 1, 2, \ldots N)$.

H_2: The Rao-Hartley-Cochran (1962) strategy (RHC strategy)

The population is randomly divided into n (sample-size) groups For the ith group G_i $(i = 1, \ldots, n)$ we write $P_i = \sum_{j \in G_i} p_j$. One unit is selected independently from each group such that the jth unit in the ith group has a selection probability $\frac{p_{ji}}{P_i}$ $(i = 1, \ldots, N)$ attached to it using the notation p_{ji} for normed size-measure of jth unit falling in ith group. The estimator is of the form

$$t_2 = \sum_{i=1}^{n} \frac{y_i}{p_{ji}} P_i .$$

For the sake of simplicity we shall assume that each group consists of $\frac{N}{n} = K$ (an integer) units.

H_3: The Midzuno-Sen (1952, 1953) strategy (MS-strategy)

– On the first draw ith unit is chosen with probability p_i and is set aside, in subsequent (n-1) draws selections are made with equal probabilities without replacement from among units not already drawn. The estimator is the ratio estimator namely

$$t_3 = \frac{\sum_s y_i}{\sum_s x_i} X = \frac{\bar{y}_s}{\bar{x}_s} X$$

for a sample s chosen in the above manner, $\bar{y}_s, \bar{x}_s$ denoting sample means, y_s, x_s denoting sample totals, with or without subscript, calculated from the sample s.

The fourth strategy H_4 we shall consider is one employing the ratio estimator of the form t_3 when the sample is chosen by SRSWOR method. We shall denote this estimator by t_4 when we denote the strategy as H_4.

The variances for the above four strategies are

$$V_1 = V(t_1) = \sum_i \frac{y_i^2}{\pi_i} + \sum_{i \neq j}\sum \frac{y_i \; y_j}{\pi_i \; \pi_j} \pi_{ij} - Y^2$$

$$V_2 = V(t_2) = \frac{N-n}{n(N-1)} \left[\sum_i \frac{y_i^2}{p_i} - Y^2\right]$$

$$V_3 = V(t_3) = \sum_i T_i \; y_i^2 + \sum_{i \neq j}\sum T_{ij} \; y_i \; y_j - Y^2$$

where $T_i = \frac{X}{\binom{N-1}{n-1}} \Sigma_i (x_i + x_{i_2} + \dots + x_{i_n})^{-1}$

for $i = 1, \dots, N$

$$T_{ij} = \frac{X}{\binom{N-1}{n-1}} \Sigma_{ij} (x_i + x_j + x_{i_3} + \dots + x_{i_n})^{-1}$$

for $i \neq j = 1, \dots, N$.

Here $i_2, i_3, \dots, i_n$ are $(n-2)$ distinct integers out of $1, \dots, N$. Σ_i denotes sum over distinct integers out of $1, \dots, N$ $(\neq i)$ and Σ_{ij} denotes sum over distinct integers out of $1, \dots, N$ $(\neq i \neq j)$.

Writing E_i generically for expected values of V_i $(i = 1,2,3)$ for the various models, we get the following theorem

Theorem 2.1

Under the model M_1', it follows that

$E_1 < E_2 < E_3$ when $\frac{v(x)}{x}$ is an increasing function of x,

$E_1 > E_2 > E_3$ when $\frac{v(x)}{x}$ is a decreasing function of x,

and $E_1 = E_2 = E_3$ when $\frac{v(x)}{x}$ is a constant.

Proof.

$$E_1 = \varepsilon (V_1)$$
$$= \varepsilon \left[\sum_i \frac{Y_i^2}{\pi_i} + \sum_{i \neq j} \sum \frac{Y_i}{\pi_i} \frac{Y_j}{\pi_j} \pi_{ij} - Y^2 \right]$$

$$= \sum_i \frac{\beta^2 x_i^2 + v(x_i)}{\pi_i} + \sum_{i \neq j}\sum \frac{\beta^2 x_i x_j}{\pi_i \pi_j} \pi_{ij} - \beta^2 X^2 - \sum_i v(x_i)$$

$$= \sum_i V(x_i) \left(\frac{1}{\pi_i} - 1\right)$$

$$= \sum_i \frac{V(x_i)}{x_i} \left(\frac{X}{n} - x_i\right) \qquad (2.3.1)$$

$$E_2 = \varepsilon (V_2)$$

$$= \frac{N-n}{n(N-1)} \; \varepsilon \left(\sum_i \frac{y_i^2}{p_i} - Y^2\right)$$

$$= \frac{N-n}{n(N-1)} \left(\sum \frac{\beta^2 x_i^2 + v(x_i)}{p_i} - \beta^2 X^2 - \sum_i v(x_i)\right)$$

$$= \frac{N-n}{n(N-1)} \sum_i v(x_i) \left(\frac{1}{p_i} - 1\right)$$

$$= \frac{N-n}{n(N-1)} \sum_i \frac{V(x_i)}{x_i} (X - x_i) \qquad (2.3.2)$$

$$E_3 = \varepsilon (V_3)$$

$$= \varepsilon \left[\sum_i T_i y_i^2 + \sum_{i \neq j}\sum T_{ij} y_i y_j - Y^2\right]$$

$$= \sum_i T_i \left[\beta^2 x_i^2 + v(x_i)\right] + \sum_{i \neq j}\sum T_{ij} \left[\beta^2 x_i x_j - \beta^2 X^2 - \sum_i v(x_i)\right]$$

$$= \sum_i (T_i - 1) \; v(x_i)$$

$$= \sum_i \frac{v(x_i)}{x_i} (T_i x_i - x_i) \qquad (2.3.3)$$

[Since $\sum_i T_i x_i^2 + \sum_{i \neq j}\sum T_{ij} x_i x_j = X^2$, vide Rao, T.J.(1967)]

Again

$$E_1-E_2 = \sum_i \frac{v(x_i)}{x_i} \left[\frac{X}{n} - x_i - \frac{N-n}{n(N-1)} X + \frac{N-n}{n(N-1)} x_i\right]$$

$$= -\frac{(n-1)}{n(N-1)} N \sum_i \frac{v(x_i)}{x_i} (x_i - \bar{X})$$

$$= -\frac{(n-1)}{n(N-1)} N^2 \operatorname{cov}\left(\frac{v(x)}{x}, x\right) \qquad (2.3.4)$$

Thus $E_1-E_2 \lesseqgtr 0$ according as $\frac{v(x)}{x}$ is an increasing, constant or decreasing function of x respectively.

Also

$$E_1-E_3 = \sum_i \frac{v(x_i)}{x_i} \left(\frac{X}{n} - x_i - T_i x_i + x_i\right)$$

$$= -\sum_i \frac{v(x_i)}{x_i} \left(T_i x_i - \frac{X}{n}\right)$$

$$= -N \operatorname{cov}\left(\frac{v(x_i)}{x_i}, T_i x_i\right) \qquad (2.3.5)$$

[Since $\sum_i T_i x_i = \frac{N}{n} X$, vide Rao, T.J. (1967), an alternative proof of this occurs in chapter 3, Section (3.3), p.66]

Thus $E_1-E_3 \lesseqgtr 0$ according as $\frac{v(x)}{x}$ is an increasing, constant or a decreasing function of x, because $T_i x_i$ is an increasing function of x_i as Rao, T.J. (1967) observed in his unpublished thesis, and is not difficult to check.

Finally,

$$E_2 - E_3 = \sum_i \frac{v(x_i)}{x_i} \left(\frac{N-n}{n(N-1)} X - \frac{N-n}{n(N-1)} x_i - T_i x_i + x_i \right)$$

$$= \sum_i \frac{v(x_i)}{x_i} \left(\frac{N-n}{n(N-1)} X - T_i x_i + \frac{N(n-1)}{n(N-1)} x_i \right)$$

$$= \sum_i \frac{v(x_i)}{x_i} z_i \qquad (2.3.6)$$

where we write

$$z_i = \frac{N-n}{n(N-1)} X - T_i x_i + \frac{N(n-1)}{n(N-1)} x_i \quad \forall\ i = 1, \ldots, N$$

so that

$$\sum_i z_i = \frac{(N-n)N}{n(N-1)} X - \frac{N}{n} X + \frac{N(n-1)}{n(N-1)} X$$

$$= 0$$

and hence $E_2 - E_3 = N \operatorname{cov}\left(\frac{V(x_i)}{x_i}, z_i \right)$ (2.3.7)

It follows then that

$$\frac{\partial z_i}{\partial x_i} = 1 - \left(T_i + x_i \frac{\partial T_i}{\partial x_i} \right)$$

Now $T_i + x_i \frac{\partial T_i}{\partial x_i}$

$$= \frac{1}{\binom{N-1}{n-1}} \left[X \sum_i (x_i + x_{i_2} + \ldots + x_{i_n})^{-1} + x_i \sum_i \{ (x_i + x_{i_2} + \ldots + x_{i_n})^{-1} - X (x_i + x_{i_2} + \ldots + x_{i_n})^{-2} \} \right]$$

$$= \frac{1}{\binom{N-1}{n-1}} \sum_i \left[x_i (x_i + x_{i_2} + \ldots + x_{i_n}) + (x_{i_2} + \ldots + x_{i_n}) X \right] \cdot (x_i + x_{i_2} + \ldots + x_{i_n})^{-2}$$

$$> \frac{1}{\binom{N-1}{n-1}} \Sigma_i \frac{x_i(x_i+x_{i_2}+\dots+x_{i_n}) + (x_{i_2}+\dots+x_{i_n})(x_i+x_{i_2}+\dots+x_{i_n})}{(x_i+x_{i_2}+\dots+x_{i_n})^2}$$

$$= 1, \qquad (2.3.8)$$

This implies $\frac{\partial z_i}{\partial x_i} < 0$ which in its turn implies z_i is a decreasing function of x_i and it follows that $E_2 \lesseqgtr E_3$ according as $\frac{v(x)}{x}$ is an increasing, constant or a decreasing function of x.

So combining (2.3.4), (2.3.5), (2.3.7) and (2.3.8) one can check the theorem.

When $v(x_i) = \sigma^2 x_i^g$, the model M_1' reduces to the model M_1 and hence we have the following corollary :

<u>Corollary 2.1</u>

Under the model M_1 we have

$$\left.\begin{array}{ll} E_1 < E_2 < E_3 & \text{if } g > 1 \\ E_1 > E_2 > E_3 & \text{if } g < 1 \\ \text{and } E_1 = E_2 = E_3 & \text{if } g = 1 \end{array}\right\} \qquad (2.3.9)$$

2.3B RELATIVE EFFICIENCIES OF STRATEGIES UNDER MODEL M_1''

For our results we shall assume the following :

(i) $\alpha = 0$

(ii) x_i's are independently identically distributed gamma variates with a common density function of the form

$$f(x) = \frac{1}{\Gamma h} e^{-x} x^{h-1}$$

with an unknown (mean-) parameter h,

(iii) g is negligible compared to N_h.

Following Cochran (1963), if we neglect the error in writing $\bar{X}$ for $\bar{x}_s$ for every s [valid, if we assume (as we do) large samples] - the usual [vide Cochran (1963)] approximate expression for the mean square error

$$\frac{1}{\binom{N}{n}} \sum_s \left(\frac{\bar{y}_s}{\bar{x}_s} X - Y\right)^2 \text{ of } t_4 \text{ is}$$

$$V_4 = \frac{N(N-n)}{n(N-1)} \sum_i (y_i - R x_i)^2, \quad (\text{writing } R = \frac{Y}{X} = \frac{\bar{Y}}{\bar{X}})$$

$$= \frac{N(N-n)}{n(N-1)} \left[\sum_i e_i^2 + \frac{(\sum_i e_i)^2 \sum_i x_i^2}{X^2} - \frac{2}{X} (\sum_i e_i x_i)(\sum_i e_i)\right]$$

Let us use the following notations :

$\mathcal{E}(\cdot|x)$ ≡ conditional expectation operator when x is held fixed,

$\varepsilon_x(.)$ ≡ expectation - operator over distribution of

ε ≡ the over-all expectation - operator

$= \varepsilon_x(.)\ \varepsilon\ (.|x)$.

$$\varepsilon\ (V_4|x) = \frac{N(N-n)}{n(N-1)}\ \sigma^2\ [\sum_i x_i^g + \frac{1}{X^2}\sum_i x_i^{g+2} + \sum_{i\neq j}\sum \frac{x_i^g x_j^2}{X^2} - \frac{2}{X}\sum_i x_i^{g+1}]$$

and $E_4 = \varepsilon(V_4) = \varepsilon_x\ \ \varepsilon\ (V_4|x)$

$$= \frac{N(N-n)}{n(N-1)}\ \sigma^2\ \varepsilon_x\ [N\ x_i^g + \frac{1}{X^2}\ (N\ x_i^{g+2} + N(N-1)x_i^g x_j^2) - \frac{2N}{X}\ \ x_i^{g+1}] .$$

We may recall from Rao and Webster (1966) that for any non-negative integers a, b, c and the mutually identically independently distributed gamma variates z_i's with a mean-parameter h one has for every i, j $(i \neq j)$

$$E\ \frac{z_i^a\ z_j^b}{(\sum_{i=1}^{n} z_i)^c} = \frac{\Gamma(a+h)\,\Gamma(b+h)}{(\Gamma h)^2\ \prod_{t=1}^{c}(m+a+b-t)},\ \text{where } m = nh.$$

One may readily check that this is true also when a, b are non-integers. Using this result we get

$$E_4 = \varepsilon(V_4)$$

$$= \frac{N^2(N-n)}{n(N-1)}\ \sigma^2\ [\frac{\Gamma(g+h)}{\Gamma h} + \frac{\Gamma(g+h+2)}{\Gamma h\ (Nh+g+1)(Nh+g)} + \frac{(N-1)\ \Gamma(g+h)\ \Gamma(2+h)}{(\Gamma h)^2(Nh+g+1)(Nh+g)} - 2\ \frac{\Gamma(g+h+1)}{\Gamma h\ (Nh+g)}\ ]$$

which simplifies to

$$N^2\sigma^2 \frac{(N-n)}{n} \frac{\Gamma g+h}{\Gamma h} \frac{h(Nh+2)}{(Nh+g)(Nh+g+1)}$$

Under this model, we further get

$$E_3 = \varepsilon (V_3|x) = \sigma^2 \sum_i x_i^{g-1} (T_i x_i - x_i)$$

Noting that

$$\varepsilon_x(T_i x_i^g) = \varepsilon_x(x_i^g) + (N-n)h\, \varepsilon_x\left(\frac{x_i^g}{x_{i_1}+x_{i_2}+\cdots+x_{i_n}}\right)$$

$$= \varepsilon_x(x_i^g) + (N-n)h \frac{\Gamma g+h}{\Gamma h} \frac{1}{(nh+g-1)},$$

simplifies to $\sigma^2 N (N-n)h \frac{\Gamma g+h}{\Gamma h} \frac{1}{(nh+g-1)}$

Also, $E_1 = \varepsilon(V_1) = \varepsilon_x\, \varepsilon\, (V_1|x)$

$$= \sigma^2 \varepsilon_x[\sum_n x_i^g (\frac{X}{nx_i} - 1)]$$

$$= \sigma^2 \varepsilon_x[\sum_i x_i^g (\frac{1}{n} - 1) + \sum_{i \neq j}\sum \frac{x_i^{g-1} x_j}{n}]$$

hich simplifies to

$$\sigma^2[N(\frac{1}{n} - 1) \frac{\Gamma g+h}{\Gamma h} + \frac{N(N-1)}{n} \frac{\Gamma g+h-1}{\Gamma h}.h]$$

and the latter to

$$\sigma^2 \frac{N}{n} \frac{\Gamma g+h}{\Gamma h} \frac{(N-n)h - (n-1)(g-1)}{g+h-1}.$$

It follows that

$$E_4 - E_3{}' = \sigma^2 N^2 \; \frac{(N-n)}{n} \; \frac{\Gamma(g+h)}{\Gamma(h)} \; \frac{h(Nh+2)}{(Nh+g)(Nh+g+1)}$$

$$- \sigma^2 N \; \frac{(N-n)}{\Gamma(h)} \; h \; \frac{\Gamma(g+h)}{(nh+g-1)}$$

$$= \sigma^2 Nh(N-n) \; \frac{\Gamma(g+h)}{\Gamma(h)} \left[\frac{N(Nh+2)}{n(Nh+g)(Nh+g+1)} - \frac{1}{nh+g-1} \right]$$

$$= \sigma^2 \; \frac{N}{n} \; (N-n) \; \frac{\Gamma(g+h)}{\Gamma(h)} \left[\frac{(1 + \frac{2}{Nh})}{(1 + \frac{g}{Nh})(1 + \frac{g+1}{Nh})} - \frac{1}{1 + \frac{g-1}{nh}} \right]$$

$$= \sigma^2 \; \frac{N}{n} \; (N-n) \; \frac{\Gamma(g+h)}{\Gamma(h)} \; \Delta(g),$$

writing $\Delta(g)$ for the expression in the square bracket. Now neglecting the terms involving $(\frac{1}{Nh})^j$ and $(\frac{g-1}{nh})^j$ for $j \geq 2$, we get, approximately, [for large N, n and small excess (relative to h) in value of g over unity, which conditions we assume in what follows next]

$$\Delta(g) = (1 - \frac{2g+1}{Nh} + \frac{2}{Nh}) - (1 - \frac{g-1}{nh})$$

$$= (\frac{1}{Nh} - \frac{1}{nh}) + g\,(\frac{1}{nh} - \frac{2}{Nh})$$

$$= \delta(g) \quad \text{(say)}.$$

So, we get the theorem.

Theorem 2.2

$$\delta(g) < 0 \quad \text{for } 0 \leq g \leq 1,$$

$$\delta(g) < 0 \quad \text{for } g > 1,\ f = \frac{n}{N} > \frac{1}{2},$$

and $$\delta(g) > 0 \quad \text{if } g > \frac{1-f}{1-2f} \text{ and } f < \frac{1}{2}.$$

Remark 2.1

This theorem gives information about how to compare he relative efficiencies of H_3 and H_4 involving the ratio-estimator based on Midzuno-Sen and SRSWOR sampling scheme r spectively.

Next we have,

$$E_4 - E_1 = \sigma^2 \frac{N}{n} \frac{\overline{g+h}}{\overline{h}} \quad f(g) \qquad \text{(say)}$$

where we write

$$f(g) = (N-n) \frac{Nh\ (Nh+2)}{(Nh+g)(Nh+g+1)} - \frac{(N-n)h - (n-1)(g-1)}{h+g-1}$$

$$= (N-n)\ [(1 + \frac{2}{Nh})(1 + \frac{g}{Nh})^{-1}(1 + \frac{g+1}{Nh})^{-1} - (1 - \frac{(n-1)(g-1)}{(N-n)h})(1 + \frac{g-1}{h})^{-1}]$$

$$\simeq (N-n)\ [(1 + \frac{2}{Nh} - \frac{2g+1}{Nh}) - (1 - \frac{g-1}{h} - \frac{(n-1)(g-1)}{(N-n)h})]$$

$$= \frac{N-n}{h}[g(1 - \frac{2}{N} + \frac{n-1}{N-n}) - (1 + \frac{n-1}{N-n} - \frac{1}{N})] .$$

So, we get the

Theorem 2.3

$$E_4 < E_1 \quad \forall \; g.\ni. \quad 0 \leq g < 1 + \frac{1}{N-2 + \frac{N(n-1)}{N-n}}$$

$$E_4 > E_1 \quad \forall \; g.\ni. \quad g > 1 + \frac{1}{N-2 + \frac{N(n-1)}{N-n}}$$

$$\text{and} \quad E_4 = E_1 \quad \text{if } g = 1 + \frac{1}{N-2 + \frac{N(n-1)}{N-n}}$$

Since in practice in most cases $g > 1$, it follows that with this super-population model M_1'', HTE may often fare better than the ratio estimator based on SRSWOR scheme.

Here we donot compare E_1 with E_3 because their relative magnitudes have been considered in the previous section without requiring to assume any specific distributions for x_i's.

2.3.C RELATIVE EFFICIENCIES OF THE SAMPLING STRATEGIES UNDER MODEL M_2

Under M_2 it is algebraically cumbrous to get expressions for E_1 and E_3 in general. So, for simplicity we assume that H_1 is based on Sampford's (1967) πps-sampling scheme and for the variance of t_1 we use its asymptotic expression due to Asok and Sukhatme (1976), valid on assuming (which we now do) large sample and population sizes. This expression denoted as V_1' (under Model M_2 it is to be written as E_1') turns out as

$$V_1' = \frac{1}{n} \sum_i p_i r_i^2 \, [1 - (n-1) p_i] \quad , \text{ where } r_i = \frac{y_i}{p_i} - Y$$
$$i = 1, \ldots, N.$$

Then $$V_2 - V_1' = \frac{n-1}{n} \sum_i p_i \, r_i^2 \, (p_i - \frac{1}{N-1})$$
$$\approx \frac{n-1}{n} \sum_i p_i \, r_i^2 \, (p_i - \frac{1}{N}),$$

approximating $\frac{1}{N-1}$ by $\frac{1}{N}$, and because of Model M_2, it follows that

$$E_2 - E_1' = \frac{n-1}{n} \sum_i \frac{e_i^2}{p_i} (p_i - \frac{1}{N})$$
$$= \sigma^2 \frac{n-1}{n} \sum_i x_i^{g-1} (x_i - \frac{X}{N})$$
$$= \sigma^2 \frac{n-1}{n} N \text{ Cov } (x_i^{g-1}, x_i).$$

Hence follows the

Theorem 2.4

Under Model M_2, it follows that

$$E_2 \gtreqless E_1' \quad \text{according as} \quad g \gtreqless 1.$$

In case the HTE is based on Midzuno's (modified) πps sampling scheme the variance of HTE will be denoted by V_1 as before but the value of V_1 under the model M_2 will be devoted by E_1''. Then, we get the following

Theorem 2.5

If the model M_2 holds with $g \geq 2$, then

$$E_1'' \leq E_2.$$

<u>Proof.</u> Following Chaudhuri (1974) we have

$$V_1 - V_2 \leq \frac{(n-1)}{n(N-1)(N-2)} [\Sigma \, (\frac{y_i}{p_i} - Y)^2 - N \, \Sigma \, p_i \, (\frac{y_i}{p_i} - Y)^2].$$

Under M_2, the RHS reduces to

$$(\frac{n-1}{n}) \frac{1}{(N-1)(N-2)} \, [X \, \sigma^2 \{ X \, \Sigma \, x_i^{g-2} - N \, \Sigma \, x_i^{g-1} \}] \leq 0, \text{ if } g \geq 2$$

Hence the result.

2.4 THE PROBLEM OF CHOOSING THE SAMPLE-SIZE FOR HORVITZ-THOMPSON ESTIMATION ON . CONSIDERATION OF EFFICIENCIES

Chaudhuri (1977) in a recent paper noted that the variance of HTE does not go on diminishing with increasing sample sizes when it is based on arbitrary sampling schemes and also suggested schemes for which it behaves contrarily. Here we derive some further results pursuing with this problem. For our discussion we shall use generic symbols $\pi_i(\nu)$, $\pi_{ij}(\nu)$'s to denote inclusion probabilities of first two orders for sampling schemes involving ν as expected effective sample size and $\pi_i^*(r)$, $\pi_{ij}^*(r)$ as the corresponding entries, when r draws are involved. The corresponding HTE's based on $\pi_i(\nu)$ and $\pi_i^*(r)$ will be denoted as e_ν and e_r^*, with their variances as $V(\nu) = V(e_\nu)$ and $V^*(r) = V(e_r^*)$. Considering two sampling schemes which are same in form in all respects except that they are based on different sample-sizes we have

$$V_{(\nu)} - V_{(\nu')} = \sum_i y_i^2 \left(\frac{1}{\pi_i(\nu)} - \frac{1}{\pi_i(\nu')}\right)$$

$$+ \sum\sum_{i \neq j} y_i\, y_j \left(\frac{\pi_{ij}(\nu)}{\pi_i(\nu)\pi_j(\nu)} - \frac{\pi_{ij}(\nu')}{\pi_i(\nu')\pi_j(\nu')}\right)$$

$$= \sum_i y_i^2 \Big[\frac{1}{\pi_i(\nu)} \Big(1 + \sum_{j \neq i} \frac{\pi_{ij}(\nu)}{\pi_j(\nu)}\Big)$$

$$- \frac{1}{\pi_i(\nu')} \Big(1 + \sum_{j \neq i} \frac{\pi_{ij}(\nu')}{\pi_j(\nu')}\Big)\Big]$$

$$+ \sum\sum_{i<j} (y_i - y_j)^2 \left(\frac{\pi_{ij}(\nu')}{\pi_i(\nu')\pi_j(\nu')} - \frac{\pi_{ij}(\nu)}{\pi_i(\nu)\pi_j(\nu)}\right)$$

We shall suppose that $\nu' > \nu$.

A sufficient condition that V_ν decreases monotonically with increasing ν is that

$\frac{\pi_{ij}(\nu)}{\pi_i(\nu)\pi_j(\nu)}$ should be non-decreasing with increasing ν (2.4.1)

and $\frac{1}{\pi_i(\nu)} \left(1 + \sum_{j \neq i} \frac{\pi_{ij}(\nu)}{\pi_j(\nu')}\right)$ should be non-increasing with increasing ν (2.4.2)

and at least one of them being strictly so.

Let us see the applicability or otherwise of these criteria to a few well-known sampling strategies :

Illustration (i):- Poisson sampling scheme [vide Hájek (1964)].

Here $\pi_{ij}(\nu) = \pi_i(\nu)\ \pi_j(\nu)$ for every $i \neq j = 1, \ldots, N$

$$\text{But } (2.4.2) = \frac{1}{\pi_i(\nu)} \{1 + (N-1)\, \pi_i(\nu)\}$$
$$= (N-1) + \frac{1}{\pi_i(\nu)} .$$

In general we cannot say whether it is monotone. Also, in this case

$$V(\nu) - V(\nu') = \sum_i y_i^2 \left(\frac{1}{\pi_i(\nu)} - \frac{1}{\pi_i(\nu')} \right).$$

If it so happends that $y_j = 0 \quad j \neq i$ but $y_i \neq 0$ then

$$V(\nu) - V(\nu') = y_i^2 \; \frac{\pi_i(\nu') - \pi_i(\nu)}{\pi_i(\nu)\; \pi_i(\nu')}$$

which may be either positive or negative implying that V(HTE) may not be monotonically decreasing with ν. But if one has, for example,

$$\pi_i(\nu') = \frac{\nu'}{\nu}\, \pi_i(\nu) \text{ (which is a reasonable choice)}$$

so that (2.4.2) is monotonically decreasing it follows clearly that $V(\nu)$ is also monotonically decreasing with increasing ν.

Illustration (ii) - SRSWOR scheme.

It is trivial to check the issue in this case, so omit elaboration.

Illustration (iii):- Linear systematic sampling with sampling interval $K = \frac{N}{n}$ which is assumed to be an integer, $n = \nu$ being a preassigned sample size.

Here (2.4.1) $\left.\begin{array}{l} = K \\ \\ = 0 \end{array}\right\}$ if $j = i + rk$ for $r = 0, 1, \ldots, n-1$, $i = 1, \ldots, K$; otherwise.

which decreases monotonically with n.

But (2.4.2) $= K\,(1+n-1) = N$ = Constant.

In this case our sufficiency condition does not hold but $V(\nu)$ is well-known to monotonically decrease with ν as one can check from standard text books [vide, e.g., Cochran (1963)].

Again, in case the sampling designs be such that each sample contains only distinct units, the number of such distinct units being an integer n, then we may write Var (HTE) as

$$V(n) = \sum_{i<j}\sum \{\pi_i(n)\pi_j(n) - \pi_{ij}(n)\} \left\{\frac{y_i}{\pi_i(n)} - \frac{y_j}{\pi_j(n)}\right\}^2 .$$

Further, if we assume that the sampling scheme has the πps property so that

$$\pi_i(n) = n\, p_i \quad \forall\, i = 1, \ldots, N$$

then we can write

$$V(n) = \sum_{i<j}\sum p_i\, p_j \left(\frac{y_i}{p_i} - \frac{y_j}{p_j}\right)^2 - \sum_{i<j}\sum \frac{\pi_{ij}(n)}{n^2}\left(\frac{y_i}{p_i} - \frac{y_j}{p_j}\right)^2$$

So, in this case a sufficient condition for $V(n)$ to monotonically decrease with n is that

$$\frac{\pi_{ij}}{n^2} \qquad (2.4.3)$$

should increase monotonically with n.

In order to see the applicability or otherwise of this criterion we may consider,

(a) SRSWOR sampling scheme for which

$$\frac{\pi_{ij}(n)}{n^2} = \frac{n(n-1)}{N(N-1).n^2} = \frac{1}{N(N-1)}\left(1 - \frac{1}{n}\right) \text{ which}$$

increases with n enabling us to conclude that V(n) decreases with increasing n,

(b) Linear systematic sampling scheme with $K = \frac{N}{n}$ = an integer. In this case

$$\frac{\pi_{ij}(n)}{n^2} = \begin{cases} \dfrac{1}{Kn^2} = \dfrac{1}{Nn} & \text{for } j = i+rk \\ & \quad r = 0,\ldots,n-1 \\ & \quad k = 1,\ldots,k \\ = 0 & \text{otherwise} \end{cases}$$

and hence no conclusion follows from this criterion as such. Instead of applying these criteria in examining the availability of designs with the indicated monotonicity property we shall now investigate if we can find particular sampling designs (involving selections with varying probabilities) with this required property and those without it.

Midzuno-Sen's (1952, 1953) scheme of sampling :

$$\pi_i(n) = \frac{N-n}{N-1} p_i + \frac{n-1}{N-1}, \quad i = 1, \ldots, N$$

and $$\pi_{ij}(n) = \frac{n-1}{N-1} \cdot \frac{N-n}{N-2} (p_i + p_j) + \frac{n-1}{N-1} \cdot \frac{n-2}{N-2}$$

$$i \neq j = 1, \ldots N.$$

Considering the particular case where

$N = 30$, $n = 3$, $p_1 = p_2 = .4$, p_j's $(j = 3, 4, \ldots N)$ are arbitrary except that $\sum_i p_i = 1$, $0 < p_i < 1 \; \forall i = 1, \ldots, N$, $y_1 = y_2 = Y$, $y_j = 0 \; \forall j > 2$ and employing Midzuno's sampling scheme we have

$$\pi_1(3) = \pi_2(3) = .44138, \quad \pi_{12}(3) = .055665,$$

$$\pi_1(4) = \pi_2(4) = .46207, \quad \pi_{12}(4) = .084236.$$

giving us $V(4) - V(3) = .0147\, Y^2 > 0$ and thus demonstrating that V(HTE) *for Midzuno's scheme may have larger variance* with a larger sample-size.

However observing that for any sampling scheme

$$V(HTE) = \sum_i \frac{y_i^2}{\pi_i(\nu)} + \sum_{i \neq j} \sum y_i y_j \frac{\pi_{ij}(\nu)}{\pi_i(\nu)\, \pi_j(\nu)} - Y^2$$

and the $\pi_i(\nu)$, $\pi_{ij}(\nu)$'s are subject to the consistency conditions

$$\sum_i \pi_i(\nu) = \nu, \quad \sum_{j(\neq i)} \pi_{ij}(\nu) = \sum_{s \ni i} \nu(s)\, p(s) - \pi_i(\nu)$$

$$\forall \; i = 1, \ldots, N,$$

where p(s) is the probability of selecting a sample s and $\nu(s)$ is the effective size of s and

$$\sum_{i \neq j}\sum \pi_{ij}(\nu) = \nu(\nu-1) + \text{var}\{\nu(s)\},$$

it is difficult to visualize the behaviour of $V(\nu)$ with changing ν for any arbitrary sampling scheme without knowing the values of the y_i's. Unfortunately, so far it has not been possible for us to come up with any general result concerning the situation where $V(\nu)$ 'misbehaves' by increasing with increasing ν for certain types of y_i values and certain range of ν-values. But we have come accross a few specific cases with particular values of y_i's, ν's, $\pi_{ij}(\nu)$'s and $\pi_i(\nu)$'s for some specific sampling schemes like those due to Midzuno (1952) (as illustrated above, in particular) Rao (1961), Chaudhuri (1975) and Seth (1966) where $V(\nu')$ may exceed $V(\nu)$ even though $\nu' > \nu$. Naturally, therefore, we are inclined to seek to identify known and simple but 'varying probability' sampling schemes which may be found to behave properly yielding $V(\nu)$ decreasing monotonically with increasing ν and one such is found as follows .

Midzuno's πps sampling scheme :

Suppose in a particular situation x_i's are such that we have

$$\frac{n-1}{n(N-1)} < p_i < 1 \quad \forall\, i = 1,\ldots N \qquad (2.4.4)$$

for every sample-size n that we may contemplate to use. Then as considered by Chaudhuri (1974) we may adopt a modified form of Midzuno sampling scheme having the πps property where on the first draw the ith unit is selected with a probability

$$\theta_i = \frac{N-1}{N-n} \; n\, p_i - \frac{n-1}{N-n} \quad (i = 1, \ldots, N),$$

is set aside and further selections are made on subsequent (n-1) draws with equal probabilities without replacement from among units not drawn earlier so that for this scheme we have

$$\pi_i(n) = n\, p_i \qquad \forall\; i = 1, \ldots, N$$

and $$\pi_{ij}(n) = \frac{n(n-1)}{N-2} \{ (p_i+p_j) - \frac{1}{N-1} \},\ \forall\, i \neq j = 1, \ldots, N.$$

Clearly, it follows, [noting (2.4.4)] that condition (2.4.3) is satisfied for this scheme for every n > 2, so that V(n) 'behaves' desirably falling off regularly with increasing n (provided n > 2).

It may be of interest, however, to apply a direct check on this property for this scheme as given below as we feel that this may be instructive and helpful in examining the behaviour of V(n) function in respect of n for other varying probability sampling schemes as well. For this scheme [vide Chaudhuri (1974)] we have

$$V(n) = \frac{N-2n}{N-2} \; \frac{1}{n} \; \sum_{i} p_i \left(\frac{y_i}{p_i} - Y\right)^2$$

$$\frac{n-1}{n(N-1)(N-2)} \; \sum_{i \neq j}\sum \left(\frac{y_i}{p_i} - Y\right)\left(\frac{y_j}{p_j} - Y\right)$$

hen it follows that

$$\frac{d}{dn} V(n) = \frac{1}{(N-2)} \left(- \frac{1}{n^2}\right) \Big[N \sum_{i} p_i \left(\frac{y_i}{p_i} - Y\right)^2$$

$$+ \frac{1}{N-1} \sum_{i \neq j}\sum \left(\frac{y_i}{p_i} - Y\right)\left(\frac{y_j}{p_j} - Y\right)\Big]$$

$$< 0$$

$$\Big[\text{because,} \quad N \sum_{i} p_i \left(\frac{y_i}{p_i} - Y\right)^2 + \frac{1}{N-1} \sum_{i \neq j}\sum \left(\frac{y_i}{p_i} - Y\right)\left(\frac{y_j}{p_j} - Y\right)$$

$$> \frac{n-1}{n} \; \frac{N}{N-1} \; \sum_{i} \left(\frac{y_i}{p_i} - Y\right)^2 + \frac{1}{N-1} \sum_{i \neq j}\sum \left(\frac{y_i}{p_i} - Y\right)\left(\frac{y_j}{p_j} - Y\right)$$

$$\geq 0 \qquad \Big]$$

bserving that $\frac{n-1}{n} N - 1 > 0$ for every $n > 1$ and

$_i > \frac{n-1}{n} \frac{1}{N-1} \quad \forall \; i = 1, \ldots, N$

us for this sampling scheme V(HTE) decreases monotonically

th increasing sample size.

A further modification of the above sampling scheme

e to Sankaranarayanan (1969), which is applicable, if

stead of (2.4.4) only the condition

$$\min_{s} \sum_{i \in s} p_i > \frac{n-1}{N-1} \qquad (2.4.5)$$

is satisfied, is worth consideration in this context. This sampling scheme has the same $\pi_i(n)$, $\pi_{ij}(n)$ and $V(n)$ values as the ones considered above and hence for this scheme $V(n)$ 'behaves' like wise.

Let us finally note a result of sheer academic interest to see that V(HTE) behaves irregularly with increasing number of draws in samples chosen (simple) randomly with replacement.

Simple random sampling with replacement (SRSWR)

For this sampling scheme we have

$\pi_i^*(n) = 1-t^n$, $\pi_{ij}^*(n) = 1-2t^n+u^n$, where

we write $t = 1-\frac{1}{N}$, $u = 1-\frac{2}{N} = t - \frac{1}{N}$

$$V^*(n) = V(e^*(n)) = \frac{1}{(1-t^n)} \sum_i y_i^2 + \frac{1-2t^n+u^n}{(1-t^n)^2} \sum_{i \neq j}\sum y_i y_j - Y^2.$$

Then $\Delta(n,n+1) = V^*(n) - V^*(n+1)$

$$= \alpha \sum_i y_i^2 + \beta \sum_{i \neq j}\sum y_i y_j$$

which is a quadratic form with the discriminant

$$D = \{\alpha + (N-1)\beta\}(\alpha-\beta)^{N-1}$$

where $$\alpha = \frac{1}{(1-t^n)} - \frac{1}{(1-t^{n+1})}$$

and $$\beta = \frac{1}{1-t^n} - \frac{1}{1-t^{n+1}} - \left\{ \frac{t^n-u^n}{(1-t^n)^2} - \frac{t^{n+1}-u^{n+1}}{(1-t^{n+1})^2} \right\}.$$

A necessary and sufficient condition for positive definiteness of $\Delta(n,n+1)$ for every n is that $(\alpha-\beta)^{k-1}$ and $\alpha + (k-1)\beta$ for every $k = 1,\ldots,N$ should have a common sign.

Now we see that

$$\alpha-\beta = \frac{t^n - u^n}{(1-t^n)^2} - \frac{t^{n+1} - u^{n+1}}{(1-t^{n+1})^2}$$

$$= \frac{t^n-u^n}{(1-t^{n+1})^2}\left[\left(\frac{1-t^{n+1}}{1-t^n}\right)^2 - \frac{t^{n+1}-u^{n+1}}{t^n-u^n}\right]$$

$$= \left[\{t + (1-t)(1-t^n)^{-1}\}^2 - \frac{t^{n+1}-u^{n+1}}{t^n-u^n}\right] A(n)$$

$$\left[\text{writing } A(n) = \frac{t^n-u^n}{(1-t^{n+1})^2} > 0\right]$$

$$> \left[(1-t)^2(1-t^n)^{-2} + 2t(1-t)(1-t^n)^{-1} + \frac{t^n u - t^{n+1}}{t^n-u^n}\right] A(n)$$

$$(\text{since } t^2 > u)$$

$$= \left[(1-t)^2(1-t^n)^{-2} + 2t(1-t)(1-t^n)^{-1} - t^n(t-u)(t^n-u^n)^{-1}\right] A(n)$$

$$= \left[\frac{1}{N^2}(1-t^n)^{-2} + 2\left(1-\frac{1}{N}\right)\frac{1}{N}(1-t^n)^{-1} - \frac{1}{N}\left\{1-\left(\frac{u}{t}\right)^n\right\}^{-1}\right] A(n)$$

$$> \left[\frac{1}{N^2}(1-t^n)^{-2} + \frac{1}{N}\left(1-\frac{2}{N}\right)(1-t^n)^{-1}\right] A(n)$$

$$\left[\text{since } \left\{1-\left(\frac{u}{t}\right)^n\right\}^{-1} < (1-t^n)^{-1}\right]$$

$$> 0$$

So, the sign of $\alpha + (k-1)\beta$ will decide whether $\Delta(n, n+1)$ may be positive definite. But we shall see below that there exist situations where $\alpha + (k-1)\beta$ may be negative implying thereby that $\Delta(n, n+1)$ may not be a positive definite quadratic form.

Let $N = 5$ so that $t = .8$, $u = .6$. In this case considering $\Delta(2,3)$ corresponding to $n = 2$, we find the value of $\alpha + (N-1)\beta$ to equal $.7286 + 4(-.1889) = -.0270 < 0$ and get an example where $\Delta(n, n+1)$ may fail to be positive definite implying failure of $V(n)$ to decrease monotonically with increasing n, the number of draws in this case

For large n and N, however, it may be checked that the situation is reversed. That, of course, is not surprising because asymptotically SRSWR scheme is equivalent to SRSWOR scheme.

The algebra is worked out below.

Let us write

$$V(n) = C(n) \sum_i y_i^2 + D(n) \sum_{i<j}\sum (y_i - y_j)^2 - Y^2$$

where $$C(n) = \frac{1}{(1-t^n)^2}\{N(1-t^n) + (N-1)(u^n - t^n)\}$$

and $$D(n) = \frac{2t^n - u^n - 1}{(1-t^n)^2}$$

Then we have

$$\frac{d}{dn} C(n) = \frac{1}{(1-t^n)^3} [\{t^n + (1-2N)t^{2n} + 2(N-1)u^n t^n\} \log t + (N-1)u^n (1-t^n) \log u]$$

Now if both N and n be so large that we may neglect the errors in making the approximations (i) $\log t \approx \log u$ and (ii) $u^n \approx t^n$, then we may write

$$\frac{d}{dn} C(n) \approx \frac{1}{(1-t^n)^3} [t^n + (1-2N)t^{2n} + 2(N-1) t^{2n} + (N-1) t^n (1-t^n)] \log t$$

$$= \frac{1}{(1-t^n)^3} N t^n (1-t^n) \log t < 0$$

$$\frac{d}{dn} D(n) = \frac{1}{(1-t^n)^3} [2t^n(t^n-u^n)\log t - u^n(1-t^n) \log u]$$

$$\approx \frac{\log t}{(1-t^n)^3} [2t^n (t^n-u^n) - u^n(1-t^n)]$$

[using (i) above]

$$= \frac{\log t}{(1-t^n)^3} [t^n (1-u^n) + (t^2)^n - u^n]$$

< 0, because $0 < u^n < 1$ and $t^2 > u$

the asymptotic conditions assumed above we have

$$= \frac{d}{dn} C(n) (\sum_i y_i^2) + \frac{d}{dn} D(n) \sum_{i<j}\sum (y_i-y_j)^2$$

< 0

So, for simple random sampling schemes with replacement V(HTE) decreases with increasing number of draws provided that the number of draws as well as the population size are so large that the approximations (i) and (ii) above entail negligible error.

2.5 ON PRODUCT METHOD OF ESTIMATION

Suppose x is negatively correlated with y. By s we denote a sample of size n chosen from this population following any probability sampling scheme. Whether it is drawn with or without replacement, the n units in it (distinct or not) will be given n different identifying labels $u_1, .. u_i, .. u_n$ (say) and denoting by $s = (u_1, .. u_i, .. u_n)$, the sample s will be divided into k-subgroups the ith subgroup consisting of n_i labels out of u_i's ($n_i \geq 1$, $i = 1, \ldots, k$, $\sum_i n_i = n$). On the basis of s we shall consider linear unbiased estimators of Y and X of the form

$$\hat{Y} = \sum_i b_{s_i} y_i$$

$$\hat{X} = \sum_i b_{s_i} y_i$$

where $b_{s_i} = 0$ if $i \not\in s$ (and b_{s_i}'s are determined not in terms of y_i's and x_i's for the sampled units) so that the conventional product estimator for Y based on s is

$$\hat{Y}^{(P)} = \frac{\hat{Y}\,\hat{X}}{X} \qquad (2.5.1)$$

which is generally employed when y and x are negatively correlated. By s_j' we shall mean the jth sub sample from s of size n_j selected (from s) at random with equal probability and without replacement ($j = 1, \ldots, k$).

Writing

$$\left.\begin{array}{l} \xi_i = b\,(s,i)\;y_i \\ \eta_i = b\,(s,i)\;x_i \end{array}\right\} \text{ for } i = 1, \ldots, N$$

$$\bar{\xi}^{(j)} = \frac{1}{n_j} \sum_{i\varepsilon s_j'} \xi_i$$

$$\bar{\eta}^{(j)} = \frac{1}{n_j} \sum_{i\varepsilon s_j'} \eta_i \,,$$

$$\hat{Y}_j = n\;\bar{\xi}^{(j)} \quad \text{and} \quad \hat{X}_j = n\;\bar{\eta}^{(j)}$$

We shall consider

$$\hat{Y}_j^{(P)} = \frac{\hat{Y}_j\,\hat{X}_j}{X} \qquad (2.5.2)$$

as the product estimator for Y based on s_j' ($j = 1, \ldots, k$). Denoting by w_i's a set of as yet un-specified real numbers not depending in any way on which units are in the sample let us consider an estimator $\hat{Y}_u^{(P)}$ of the form

$$\hat{Y}_u^{(P)} = \sum_i w_i\,\hat{Y}_i^{(P)} + (1 - \sum_i w_i)\,\hat{Y}^{(P)} \qquad (2.5.3)$$

Let,

$$e = \frac{\hat{Y} - Y}{Y}, \qquad e' = \frac{\hat{X} - X}{X}$$

$$e_j = \frac{\hat{Y}_j - Y}{Y}, \qquad e_j' = \frac{\hat{X}_j - X}{X} \quad \text{for } j = 1, \ldots, k.$$

Then

$$E(\hat{Y}_u^{(P)}) = \sum_i w_i \, E(\hat{Y}_i^{(P)}) + (1 - \sum_i w_i) \, E(\hat{Y}^{(P)})$$

$$= Y \sum_i w_i \, E(1+e_i)(1+e_i') +$$

$$Y (1 - \sum_i w_i) \, E(1+e)(1+e')$$

$$= Y [\Sigma w_i \, E(e_i \, e_i') + (1 - \Sigma w_i) E(ee') + 1]$$

[Noting that $0 = E(e_i') = E(e_i) = E(e) = E(e')$

$i = 1, \ldots, k$]

So, the bias of the estimator $\hat{Y}_u^{(P)}$ is

$$B(\hat{Y}_u^{(P)}) = E(\hat{Y}_u^{(P)} - Y)$$

$$= Y [\Sigma w_i \, E(e_i \, e_i') + (1 - \Sigma w_i) E(ee')] .$$

Writing $C_i = \dfrac{E(e_i, e_i')}{E(e, e')}$, $\forall i = 1, \ldots, k$,

it follows that assuming [$E(ee') \neq 0$, if this is not true, then the results will be still more simplified] a necessary and sufficient condition for $\hat{Y}_u^{(P)}$ to be unbiased for Y is that

$$\sum_i w_i \, C_i + (1 - \sum_i w_i) = 0 \qquad (2.5.4)$$

gain, we have the MSE of $\hat{Y}_u^{(P)}$ as

$$E(\hat{Y}_u^{(P)} - Y)^2 = E[\sum_i w_i \hat{Y}_i^{(P)} + (1 - \sum_i w_i) \hat{Y}^{(P)} - Y]^2$$

$$= Y^2 E[\sum_i w_i (1+e_i)(1+e_i') + (1 - \sum_i w_i)(1+e)(1+e') - 1]^2$$

$$= Y^2 E[\sum_i w_i (e_i+e_i'+e_ie_i') + (1 - \sum_i w_i)(e+e'+ee')]^2.$$

If we assume

$E(e_i^{\ell} e_j'^{r})$ and $E(e^{\ell} e'^{r})$ to be negligible for $0 < \ell, r, \ell+r > 2$ for every $i \neq j = 1, 2, \ldots, k,$ (3.3.5)

then to this order of approximation (to be called 2nd order approximations) we shall have the MSE of $\hat{Y}_u^{(I)}$ as

$$E(\hat{Y}_u^{(P)} - Y)^2 = Y^2 E[\sum_i w_i (e_i+e_i') + (1 - \sum_i w_i)(e+e')]^2.$$

riting $z = \sum_i w_i (e_i+e_i') + (1 - \sum_i w_i)(e+e')$

nd $g(s) = E(z|s) \equiv$ the conditional expectation of z given e initial sample s we have

$$g(s) = \sum_i w_i E\{(e_i+e_i')|s\} + (1 - \sum_i w_i)(e+e')$$

en, it follows that

$$E(z) = E_1 E_2(z|s) = E_1 g(s)$$

$$= E_1[\sum_i w_i (e+e') + (1 - \sum_i w_i)(e+e')]$$

[Here $E_2 \equiv$ operator for taking conditional expectation when the initial sample s is fixed,

$E_1 \equiv$ operator for taking unconditional expectation over the random variation of s] .

Then, $g(s) = \sum_i w_i\ (e+e') + (1 - \sum_i w_i)(e+e')$

$$= (e+e') .$$

Now it follows that

$$E(\hat{Y}_u^{(P)} - Y)^2 = Y^2\ E_1\ \{Var(z|s) + E_2^{\ 2}(z|s)\}$$
$$\geq Y^2\ E_1\ \{g(s)\}^2$$
$$= Y^2\ E_1\ (e+e')^2$$
$$= \text{Mean square error of } \hat{Y}^{(P)}$$

[to the order of approximation mentioned by the assumption (2.5.5)]

So, we get the theorem

Theorem 2.6

Under the assumption (2.5.5) an unbiased estimator for Y of the form (2.5.3) based on random sub-samples cannot have a variance less than the MSE (to the second order of approximation) of a product estimator based on the initial sample.

Now we note that a modified product estimator $\hat{Y}_u^{(P)}$ of the form (2.5.3) will have a variance equal to the MSE of $\hat{Y}^{(P)}$ to the order of approximation (2.5.5) if and only if w_i's are chosen to satisfy

$$\sum_i w_i \, (e_i + e_i') = (e+e') \sum_i w_i \qquad (2.5.6)$$

with probability one.

Noting that $\hat{Y}_i = n \, \bar{\xi}^{(i)}$ and $\hat{X}_i = n \, \bar{\eta}^{(i)}$

so that $\sum_i \frac{n_i}{n} Y_i = \sum_{i \in s} \xi_i = Y$

and $\sum_i \frac{n_i}{n} \hat{X}_i = \sum_{i \in s} \eta_i = \hat{X}$, we get

$$(e+e') = \sum \frac{n_i}{n} (e_i + e_i') .$$

So, the condition (2.5.6) reduces to

$$\sum_i w_i \, (e_i + e_i') = \sum_i \frac{n_i}{n} (e_i + e_i') \sum_i w_i$$

Now this is realised if we choose w_i such that

$$w_i = \frac{n_i}{n} \sum_i w_i \quad \text{for } i = 1, \ldots k \qquad (2.5.7)$$

This observation leads to the following theorem, viz

Theorem 2.7

A modifed product estimator of the form $\hat{Y}_u^{(P)}$ (2.5.3) is unbiased for Y and is as efficient as $\hat{Y}^{(P)}$ in (2.5.1) in terms of mean-square error to the order of approximation as specified under the assumption (2.5.6) if the weights w_i's are chosen subject to the condition (2.5.4) and (2.5.7).

In order that an estimator $\hat{Y}_u^{(P)}$ may be usable, w_i's involved in it must be independent of individual y_i's, and x_i's. We show below that this need is fulfilled when the initial sample is chosen either as an SRSWOR or as a PPSWR sample.

Case - 1

Initial sample s is an SRSWOR sample.

Here $$C_i = \frac{E\,(e_i\; e_i')}{E\,(ee')} = 1 + n^2\,\frac{E\,(\bar{\xi}^{(i)}-\bar{\xi})(\bar{\eta}^{(i)}-\bar{\eta})}{E\,(\hat{Y}-Y)(\hat{X}-X)}$$

$$\left(\text{where } n\,\bar{\xi} = \sum_{i=1}^{n} \xi_i \text{ and } n\,\bar{\eta} = \sum_{i=1}^{n} \eta_i\right)$$

$$= 1 + \frac{n^2\left(\frac{1}{n_i} - \frac{1}{n}\right)\cdot\frac{1}{n-1}\,E_1 \sum_{j=1}^{n}(\xi_j-\bar{\xi})(\eta_j-\bar{\eta})}{E_1\,(\hat{Y} - Y)(\hat{X} - X)}$$

or $$C_i - 1 = \left(\frac{1}{n_i} - \frac{1}{n}\right)\Big/\left(\frac{1}{n} - \frac{1}{N}\right) \quad \forall\ i = 1, \ldots, k.$$

Here E_1 denotes expectation-operator over selection of the original sample of size n. So, in the case (2.5.7) and (2.5.4) respectively reduce to

$$w_i = \frac{n_i}{n} \sum_i w_i$$

and

$$\frac{1}{(\frac{1}{n} - \frac{1}{N})} \sum_i w_i \left(\frac{1}{n_i} - \frac{1}{n}\right) + 1 = 0$$

Then, w_i's can be determined free from x_i's and y_i's and one may readily get $\hat{Y}_u^{(P)}$

If in particular $n_i = \frac{n}{k}$ for $i = 1, \ldots, k$, then $w_i = \frac{1}{k} \sum_i w_i$ when $\Sigma w_i = -\frac{1}{k-1} \frac{N-n}{N}$ so that

$$w_i = -\frac{1}{k(k-1)} \frac{N-n}{N} \quad \forall \; i = 1, \ldots, k.$$

In particular, for $k = 2$ and $n_i = n/2$, $i = 1, 2$, we have $w_i = -\frac{N-n}{2N}$ for $i = 1,2$ which agrees with Shukla's (1976) result. If $k = 2$ but n_i's are unequal then $\hat{Y}_u^{(P)}$ becomes unbiased with the least variance equalling MSE $(\hat{Y}^{(P)})$ subject to (2.5.7) if we choose

$$w_i = \frac{n_i}{n} (w_1 + w_2) \qquad \text{for } i = 1,2 \quad \text{such that}$$

$$w_1 \frac{(\frac{1}{n_1} - \frac{1}{n})}{(\frac{1}{n} - \frac{1}{N})} + w_2 \frac{(\frac{1}{n_2} - \frac{1}{n})}{(\frac{1}{n} - \frac{1}{N})} + 1 = 0$$

i.e if we choose $w_i = -\frac{n_i}{n} \frac{N-n}{N}$, $i = 1,2$

Case 2

Initial sample s is a PPSWR sample.

Here $\xi_i = \frac{y_i}{np_i}$, $\eta_i = \frac{x_i}{np_i}$, where p_i's are some normed size-measure (assumed all positive). These p_i's are not same as x_i/X's as considered earlier.

Then, we have

$$c_i - 1 = n^2\left(\frac{1}{n_i} - \frac{1}{n}\right) \frac{1}{n-1} \frac{E_1 \sum_i (\xi_i - \bar{\xi})(\eta_i - \bar{\eta})}{E_1(\hat{Y} - Y)(\hat{X} - X)}$$

$$= n\left(\frac{1}{n_i} - \frac{1}{n}\right) = \frac{n - n_i}{n_i} \quad \text{for } \forall\ i = 1, \ldots, k$$

so that c_i's and hence w_i's are free from x_i's and y_i's and $\hat{Y}_u^{(P)}$ is available readily in this case.

If in particular $n_i = \frac{n}{k}\ \forall\ i = 1, \ldots, k$ we get

$$c_i - 1 = (k-1) \quad \forall\ i = 1, \ldots, k$$

so that $w_i = -\frac{1}{k(k-1)} \quad \forall\ i = 1, \ldots, k$

and with this choice $\hat{Y}_u^{(P)}$ will be an unbiased estimator for Y with the least variance equalling MSE $(\hat{Y}^{(P)})$ subject to the assumption (2.5.5).

In case $k = 2$ but n_i's are unequal we have the best choice of w_i's as $w_i = -\frac{n_i}{n}$ for $i = 1,2$ and if in particular $k = 2$ and n_i's are equal then $w_i = -\frac{1}{2}$ for $i = 1,2$.

CHAPTER - 3

SURVEYING A POPULATION ONLY ONCE
UNBIASED ESTIMATION OF VARIANCES

3.1 MOTIVATION AND SUMMARY

Estimating the variance of an estimator of a finite population total is a problem to which now we turn our attention - first in uni-stage and then in multi-stage sampling schemes. An added problem is estimating it non-negatively. This latter problem was considered by many authors. Among them Rao,T.J. (1972, 1977) and Chaudhuri (1976) considered the problem of non-negative unbiased estimation of the variance of the ratio-estimator based on Midzuno-Sen (1952, 1953) sampling scheme. Working along this line we find that some of the non-negativity conditions suggested by them cannot be effectively realised except, if at all, under trivial circumstances. Also on examining the consequences of carrying through these methods we note that positive results are unavailing. Trying to extend the similar approach to find non-negative estimators for the variances of alternative estimators also we end up only with negative results that such are not easy to come by. Only Lanke (1974), Vijayan (1975), Rao and Vijayan (1977) appear to establish positive results but they too are too specific.

For multi-stage sampling designs we have generalized Durbin's (1953, 1967) and Raj's (1968) results to show that a general class of linear homogeneous unbiased estimators for a finite population total based on arbitrary sampling schemes, in multi-stages, selecting the first stage units (f.s.u's) with replacement admits an unbiased estimator for its variance in terms of homogeneous quadratic functions of the estimators (based on sampling in stages following the first) of f.s.u. totals. In case the f.s.u's are chosen without replacement or the estimators are based on distinct f.s.u's alone selected with replacement it is also shown that such variance-estimators are available if the variances of estimators for the f.s.u totals based on sampling in subsequent stages possess unbiased estimators but not otherwise.

3.2 INTRODUCTION, FORMULATION AND STATEMENT OF THE PROBLEM

An important problem in adopting Midzuno-Sen (1952, 1953) strategy is to find a non-negative unbiased estimator of the variance of the estimator. This question was considered by Rao T.J. (1972,1977), Chaudhuri (1976) and Rao and Vijayan (1977). Rao (1972, 1977) and Chaudhuri (1977) proposed some variance estimators and derived some sufficient conditions for their non-negativity. Here we shall demonstrate that some of the conditions for non-negativity proposed by Rao, T.J. (1972, 1977) cannot be realized at all and some of his other conditions and the condition proposed by Chaudhuri (1976) cannot hold except when the Midzuno-Sen scheme degenerates to SRSWOR scheme, the size-measures being all equal in rare and uninteresting situations. Similar consequences are observed in connection with estimators proposed and studied by Sharma (1970) and Bandyopadhyaya, Chattapadhyaya and Kundu (1977). As a matter of added interest we point out that any attempt at extending Rao's (1977) technique to cover the situation concerning Horvitz-Thompson method of estimation is also destined to a similar failure. Positively effective results in this context are hard to come by Rao and Vijayan (1977), however, produced some encouraging results to which thus far we have remained unable to add further contributions.

For multi-stage sampling it is well known [vide Durbin (1953, 1967), Raj (1968), Cochran (1963), Stuart (1963)] that some particular linear unbiased estimator for a finite population total admit unbiased variance-estimators which do or donot involve unbiased estimators (based on sampling at subsequent stages) of the variances of the estimators of the f.s.u totals according as the f.s.u's are selected without or with replacement respectively, provided the f.s.u's are selected according to the method of sampling with probability proportional to size (pps) in the latter case. Here we extend these results to accommodate more general situations where the estimators belong to classes wider than those so far considered in the literature and in case of sampling with replacement the f.s.u's are selected not necessarily following the pps method. In this connection one may note that the results presented by Stuart (1963) are not strictly valid in the generality in which they are apparently claimed because of a few obvious algebraic mistakes committed in the paper.

This generalization has a theoretical necessity in view of the recent growth in the literature concerning general classes of estimators in uni-stage sampling and we show that the earlier results in this area are covered as special cases.

3.3 NON-NEGATIVE VARIANCE-ESTIMATION

3.3.1 THE MAIN RESULTS

For Midzuno-Sen strategies the estimator employed to estimate the population ratio R is

$$t_5 = \frac{\sum_{i \varepsilon s} y_i}{\sum_{i \varepsilon s} x_i} = \frac{y_s}{x_s}$$

Now following Rao (1972, 1977) and Chaudhuri (1976) we have alternative expressions for thevariance of t_5 as

$$V_1\ (t_5) = \sum_i \lambda_i\ y_i^2 + \sum_{i \neq j}\sum \lambda_{ij} Y_i Y_j\ ,$$

$$V_2\ (t_5) = \frac{1}{X^2}\left[\sum_i (T_i-1) y_i^2 + \sum_{i \neq j}\sum (T_{ij}-1) y_i\ y_j\right]$$

$$= \sum_i (nT_i-M) r_i^2 + \sum_{i<j}\sum (1-T_{ij})(r_i-r_j)^2$$

$$V_3\ (t_5) = \frac{1}{2} \sum_{i \neq j}\sum Q_{ij}$$

and $$V_4\ (t_5) = \frac{1}{2} \sum_{i \neq j}\sum R_{ij}$$

where $$\lambda_i = \frac{1}{MX} \sum_{s \ni i} \frac{1}{X_s} - \frac{1}{X^2}\ ,$$

$$\lambda_{ij} = \frac{1}{MX} \sum_{s \ni ij} \frac{1}{X_s} - \frac{1}{X^2}\ ,$$

$$M = \binom{N-1}{n-1},\quad r_i = \frac{y_i}{x_i}\ ,\quad T_i = 1 + \lambda_i\ X^2,\quad T_{ij} = 1 + \lambda_{ij} X^2,$$

$$Q_{ij} = \frac{T_i-1}{N-1} r_i^2 + 2(T_{ij}-1) r_i r_j + \frac{T_j-1}{N-1} r_j^2 ,$$

$$R_{ij} = \left(\frac{T_{ij}}{n-1} - \frac{1}{N-1}\right) r_i^2 + 2(T_{ij}-1) r_i r_j + \left(\frac{T_{ij}}{n-1} - \frac{1}{N-1}\right) r_j^2$$

The estimators for variance of t_5 namely

$$V(t_5) = \sum_s t_5^2 P(s) - R^2$$

where $P_s = \frac{X_s}{M X}$,

as suggested by Rao (1972, 1977) and Chaudhuri (1976) are given by

$$\hat{V}_1(t_5) = \sum_{i \in s} \lambda_i \frac{y_i^2}{\pi'_i} + \sum\sum_{i \neq j \in s} \lambda_{ij} \frac{y_i y_j}{\pi'_{ij}}$$

$$\hat{V}_2(t_5) = \sum_{i \in s} r_1^2 \frac{(nT_i - N)}{\pi'_i} + \sum\sum_{i<j \in s} \frac{(1-T_{ij})}{\pi'_{ij}} (r_i - r_j)^2$$

$$\hat{V}_3(t_5) = \frac{1}{2} \sum\sum_{i \neq j \in s} \frac{Q_{ij}}{\pi'_{ij}}$$

and

$$\hat{V}_4(t_5) = \frac{1}{2} \sum\sum_{i \neq j \in s} \frac{R'_{ij}}{\pi'_{ij}}$$

where

π'_i = inclusion probability of the ith unit

$$= \frac{n-1}{N-1} + \frac{N-n}{N-1} \frac{x_i}{X}, \quad \forall\ i = 1, \ldots, N,$$

and

π'_{ij} = inclusion probability of the ith and jth units

$$= \frac{n-1}{N-1} \cdot \frac{n-2}{N-2} + \frac{N-n}{N-1} \frac{n-1}{N-2} \left(\frac{x_i}{X} + \frac{x_j}{X}\right) \forall\ i \neq j = 1, \ldots, N.$$

Rao T.J.'s (1972) condition for non-negativity of $\hat{V}_1(t_5)$ is [assuming tacitly that $y_i \geq 0$ for $i = 1, \ldots N$, which condition he inadvertantly failed to mention but sought to justify the omission in his (1977) paper by noting the origin-invariance property of his estimator]

$$\lambda_{ij} \geq 0 \quad \forall\ i, j = 1, \ldots, N\ (i \neq j) \qquad (3.3.1)$$

But

$$(3.3.1) \Rightarrow \frac{1}{M} \sum_{s \ni ij} \frac{1}{P_s} \geq 1 \quad \text{for every } i \neq j = 1, \ldots, N.$$

$$\Rightarrow \frac{1}{M} \sum_{j \neq i} \sum_{s \ni ij} \frac{1}{P_s} \geq (N-1)$$

$$\Rightarrow \frac{1}{M} \sum_{i} p_i \sum_{j \neq i} \sum_{s \ni ij} \frac{1}{P_s} \geq (N-1) \qquad (3.3.2)$$

But the L.H.S. in (3.3.2) equals

$$\frac{1}{M} \sum_{s} \frac{1}{P_s} \sum\sum_{i \neq j \in s} p_i$$

$$= \frac{1}{M} \sum_{s} \frac{(n-1)}{P_s} \sum_{i \in s} p_i$$

$$= \frac{(n-1)}{M} \binom{N}{n}$$

$$< (N-1) \qquad (3.3.3)$$

So, (3.3.1) can never hold.

Infeasibility of (3.3.1) was also demonstrated differntly by Rao and Vijayan (1977).

Chaudhuri's (1976) sufficiency condition for non-negativity of $\hat{V}_2\,(t_5)$ is

$$\left.\begin{array}{ll} T_i \geq \frac{N}{n} & \forall\ i = 1, \ldots, N \\ \text{and} \quad T_{ij} \leq 1 & \forall\ i \neq j = 1, \ldots, N \end{array}\right\} \qquad (3.3.4)$$

Now $T_i \geq \frac{N}{n}$ $\forall$ i implies

$$\sum_i T_i\, p_i \geq \frac{N}{n} \qquad (3.3.5)$$

Now the L.H.S. of (3.3.5) is

$$\sum_i T_i\, p_i = \sum_i p_i \left\{ \left(\frac{1}{MX} \sum_{s \ni i} \frac{1}{x_s} - \frac{1}{X^2} \right) X^2 + 1 \right\}$$

$$= \frac{X}{M} \sum_i p_i \sum_{s \ni i} \frac{1}{X_s}$$

$$= \frac{1}{M} \sum_i x_i \sum_{s \ni i} \frac{1}{X_s}$$

$$= \frac{1}{M} \sum_s \frac{1}{x_s} \sum_{i \in s} x_i$$

$$= \frac{1}{M} \binom{N}{n} = \frac{N}{n} \qquad (3.3.6)$$

[The equation (3.3.6) was earlier obtained by Rao, T.J. (1967*) in a different manner.]

Thus (3.3.5) holds only if $T_i = \frac{N}{n}$ for every i = 1, ..., N.

Now when $T_i = T_j$ for $i \neq j$ we have

$$\sum_{s \ni i} \frac{1}{x_s} = \sum_{s \ni j} \frac{1}{x_s}$$

i.e. $$\sum_{\substack{s \ni i \\ s \not\ni j}} \frac{1}{x_s} = \sum_{\substack{s \ni j \\ s \not\ni i}} \frac{1}{x_s}$$

i.e. $$\Sigma_{ij} \frac{1}{x_i + x_{i_2} + \ldots + x_{i_n}} = \Sigma_{ij} \frac{1}{x_j + x_{i_2} + \ldots + x_{i_n}}$$

[Here Σ_{ij} denotes, as earlier, the sum over (n-1) distinct numbers $i_2, \ldots, i_n$ other than i and j out of $1, \ldots, N$]

i.e. $$\Sigma_{ij} \frac{(x_j - x_i)}{(x_i + x_{i_2} + \ldots + x_{i_n})(x_j + x_{i_2} + \ldots + x_{i_n})} = 0$$

for $i \neq j = 1, \ldots, N$

Thus $T_i = \frac{N}{n}$ $\forall i = 1, \ldots, N$ implies x_i = constant $\forall i = 1, \ldots, N$. i.e. $p_i = \frac{1}{N}$ for $i = 1, \ldots, N$. Hence the applicability of Chaudhuri's (1976) result holds only in trivial circumstances.

Rao's (1977) sufficient condition for non-negativity of $\hat{V}_3(t_5)$ is that

$$Q_{ij} \geq 0 \;\; \forall \; i, j = 1, \ldots, N \; (i \neq j) \qquad (3.3.7)$$

equilaventl y that

$$\left.\begin{array}{l} T_i > 1 \quad \forall\ i = 1,\ldots,N \\ \text{ong with} \\ (T_{ij}-1)^2 - \dfrac{(T_i-1)(T_j-1)}{(N-1)^2} \leq 0 \\ \qquad\qquad \forall\ i,j = 1,\ldots N \quad (i\neq j) \end{array}\right\} \qquad (3.3.8)$$

t the relation (3.3.8) requires that

$$\lambda_{ij}^2 \leq \frac{1}{(N-1)^2} \lambda_i \lambda_j \quad \forall\ i,j = 1,\ldots,N \quad (i\neq j)$$

ch, in its turn requires that

$$\sum_{i\neq j}\sum \lambda_{ij}^2 p_i^2 p_j^2 - \sum_{i\neq j}\sum \frac{\lambda_i \lambda_j p_i^2 p_j^2}{(N-1)^2} \leq 0.$$

the LHS in (3.3.9) is

$$\sum_{i\neq j}\sum \lambda_{ij}^2 p_i^2 p_j^2 - \frac{1}{(N-1)^2} \sum_i \lambda_i p_i^2 (\sum_i \lambda_i p_i^2 - \lambda_i p_i^2)$$

$$\geq \sum_{i\neq j}\sum \lambda_{ij}^2 p_i^2 p_j^2 - \frac{1}{(N-1)^2} (\sum_i \lambda_i p_i^2)^2 + \frac{(\sum_i \lambda_i p_i^2)^2}{N(N-1)^2}$$

$$\geq \frac{(\sum_{i\neq j}\sum \lambda_{ij} p_i p_j)^2}{N(N-1)} - \frac{(\sum_i \lambda_i p_i^2)^2}{N(N-1)} = 0 \qquad (3.3.10)$$

use

$$X^2 \sum_{i\neq j}\sum p_i p_j \lambda_{ij} = \sum_{i\neq j}\sum p_i p_j (\frac{1}{M} \sum_{s\neq ij} \frac{1}{p_s} - 1)$$

$$= \frac{1}{M} \sum_{i \neq j}\sum p_i \, p_j \sum_{s \ni ij} \frac{1}{P_s} - \sum_{i \neq j}\sum p_i \, p_j$$

$$= \frac{1}{M} \sum_{i} p_i \{ \sum_{j \neq i} p_j \sum_{s \ni ij} \frac{1}{P_s} \} - 1 + \sum_{i} p_i^2$$

$$= \frac{1}{M} \sum_{i} p_i \{ \sum_{s \ni i} \frac{1}{P_s} \sum_{j \in s, j \neq i} p_j \} - 1 + \sum_{i} p_i^2$$

$$= \frac{1}{M} (\sum_{i} p_i \sum_{s \ni i} \frac{P_s - p_i}{P_s}) - 1 + \sum_{i} p_i^2$$

$$= \frac{1}{M} (M - \sum_{i} p_i^2 \sum_{s \ni i} \frac{1}{P_s}) - 1 + \sum_{i} p_i^2$$

$$= - \sum_{i} p_i^2 (\frac{1}{M} \sum_{s \ni i} \frac{1}{P_s} - 1)$$

$$= - X^2 \sum_{i} \lambda_i \, p_i^2$$

or, $\sum_{i \neq j}\sum \lambda_{ij} \, p_i \, p_j = - \sum \lambda_i \, p_i^2$

Thus (3.3.7) can never hold.

Rao's (1977) sufficient condition for non-negativity of $\hat{V}_4(t_5)$ is that

$$T_{ij} > \frac{n-1}{N-1} \quad \forall \quad i \neq j = 1, \ldots, N \qquad (3.3.11)$$

and

$$(T_{ij}-1)^2 - (\frac{T_{ij}}{n-1} - \frac{1}{N-1})^2 \leq 0 \; \forall \; i \neq j = 1, \ldots N \quad (3.3.12)$$

(3.3.12) is true if, and only if

$$\left.\begin{array}{l} T_{ij} > \frac{N}{2(N-1)} \\ \frac{N}{n}\cdot\frac{n-1}{N-1} < T_{ij} < \frac{N-2}{n-2}\cdot\frac{n-1}{N-1} \end{array}\right\} \begin{array}{l} \text{if } n = 2 \ \forall\, i \neq j = 1, \ldots, N \\ \text{if } (n>2)\ \forall\, i, j = 1, \ldots, N \ (i \neq j) \end{array}$$

(3.3.13)

satisfied.

$$\sum_{j(\neq i)} T_{ij} = \sum_{j(\neq i)} \{ 1 + \frac{X}{M} \sum_{s \ni ij} \frac{1}{X_s} - 1 \}$$

$$= \sum_{j(\neq i)} \frac{X}{M} \sum_{s \ni i, j} \frac{1}{X_s}$$

$$= \frac{X}{M} \sum_{s \ni i} \frac{1}{X_s} \sum_{\substack{j \in s \\ j \neq i}}$$

$$= (n-1) \frac{X}{M} \sum_{s \ni i} \frac{1}{X_s}$$

$$= (n-1)\, T_i \qquad (3.3.14)$$

relation (3.3.14) was however shown by Rao T.J. (1967*) in different way]

s from the relation (3.3.14) it follows that (3.3.11) (3.3.12) cannot hold except under the trivial situation $p_i = \frac{1}{N}$ $\forall\, i = 1, \ldots, N$.

3.3.2 SUBSIDIARY RESULTS

First we note how an extension of Rao's (1977) technique for finding a non-negative unbiased estimator $\hat{V}_3(t_5)$ for $V(t_5)$ proves abortive in yielding a non-negative estimator for the variance of Horvitz-Thompson (1952) estimator for **a finite population** total based on any fixed sample-size sampling design.

With usual notations the variance of HTE viz $t_1 = \sum_{i \in s} \frac{y_i}{\pi_i}$ is

$$V_1 = V(t_1) = \sum_i y_i^2 \left(\frac{1}{\pi_i} - 1\right) + \sum_{i \neq j} \sum y_i y_j \left(\frac{\pi_{ij}}{\pi_i \pi_j} - 1\right)$$

$$= \sum_{i<j} \sum \left\{ \frac{\left(\frac{1}{\pi_i} - 1\right)}{N-1} y_i^2 + 2\left(\frac{\pi_{ij}}{\pi_i \pi_j} - 1\right) y_i y_j + \frac{\left(\frac{1}{\pi_j} - 1\right)}{N-1} y_j^2 \right\}$$

$$= \sum_{i<j} \sum \Delta_{ij}$$

On writing

$$\Delta_{ij} = \frac{\left(\frac{1}{\pi_i} - 1\right)}{N-1} y_i^2 + 2 y_i y_j \left(\frac{\pi_{ij}}{\pi_i \pi_j} - 1\right) + \frac{\left(\frac{1}{\pi_j} - 1\right)}{N-1} y_j^2$$

With the usual assumption that $\pi_{ij} > 0 \ \forall \ i,j \ (i \neq j)$, a readily available unbiased estimator for V_1 is

$$\hat{V}_1 = \hat{V}_1(t_1) = \sum_{i<j \in s}\sum \frac{\Delta_{ij}}{\pi_{ij}}$$

which is uniformly non-negative if Δ_{ij} be a non-negative definite quadratic form (for $i \neq j = 1, \ldots, N$) in y_i's. Following Rao (1977) one may consider the sufficient condition for non-negative definiteness of Δ_{ij} namely

$$\frac{1}{(N-1)^2}\left(\frac{1}{\pi_i} - 1\right)\left(\frac{1}{\pi_j} - 1\right) \geq \left(\frac{\pi_{ij}}{\pi_i \pi_j} - 1\right)^2 \ \forall i \neq j = 1, \ldots, N \tag{3.3.15}$$

and may have $\hat{V}_1$ uniformly non-negative if (3.3.15) holds.

But in order that (3.3.15) may hold it is necessary that

$$\sum_{i \neq j}\sum \left(\frac{\pi_{ij}}{\pi_i \pi_j} - 1\right)^2 \pi_i^2 \pi_j^2 - \frac{1}{(N-1)^2} \sum_{i \neq j}\sum \left(\frac{1}{\pi_i} - 1\right)\left(\frac{1}{\pi_j} - 1\right) \pi_i^2 \pi_j^2$$
$$\leq 0 \tag{3.3.16}$$

But the LHS of (3.3.16) is not less than (by Cauchy's inequality)

$$\frac{\{\sum_{i\neq j}\sum(\frac{\pi_{ij}}{\pi_i\pi_j} - 1)\pi_i\pi_j\}^2}{N(N-1)} - \frac{1}{(N-1)^2}\left[\sum_i\{(\frac{1}{\pi_i} - 1)\pi_i^2\}^2 - \sum_i(\frac{1}{\pi_i} - 1)^2\pi_i^4\right]$$

$$\geq \frac{1}{N(N-1)}\left[\{\sum_{i\neq j}\sum(\frac{\pi_{ij}}{\pi_i\pi_j} - 1)\pi_i\pi_j\}^2 - \{\sum_i(\frac{1}{\pi_i} - 1)\pi_i^2\}^2\right]$$

$$= 0$$

because

$$\sum_{i\neq j}\sum(\pi_{ij} - \pi_i\pi_j) = n(n-1) - \sum_i\pi_i(n-\pi_i)$$

$$= -\sum_i\pi_i(1-\pi_i)$$

as is well known.

Thus with this approach one cannot get a uniformly non-negative unbiased estimator for $V(t_1)$.

Considering a sampling scheme with varying probabilities of selection in n draws without replacement giving a selection probability q(s) to a typical sample, Sharma (1970) considered the unbiased estimator for Y namely

$$\hat{Y} = \frac{y_s}{L\,q(s)} \qquad (3.3.17)$$

having the variance

$$V(\hat{Y}) = \frac{1}{L^2}\sum_s\frac{y_s^2}{q(s)} - Y^2$$

and suggested for it an unbiased estimator

$$\hat{V}_1(\hat{Y}) = \hat{Y}^2 - \frac{1}{(L^2 - L_{12})} \frac{1}{q(s)} \{ y_s^{\,2} - \frac{L_{12} - L_{11}}{L} \; y_s^* \}$$

where $y_s^* = \sum_{i \in s} y_i^2$, $L = \sum_s m_i(s)$, $L_{11} = \sum_{s \neq s'} \sum m_i(s) \; m_i(s')$

and $L_{12} = \sum_{s \neq s'} \sum m_i(s) \; m_j(s')$ for $i \neq j$

$m_i(s)$ = number of times ith unit occurs in s

which are all independent of i, j's ($i, j = 1, \ldots, N$, $i \neq j$) for this particular sampling scheme.

In case we consider the unordered samples, we have

$$L = \binom{N-1}{n-1}, \quad L_{11} = \binom{N-1}{n-1} \{ \binom{N-1}{n-1} - 1 \},$$

$$L_{12} = \binom{N-1}{n-1}^2 - \binom{N-2}{n-2}$$

and

$$\hat{V}_1(\hat{Y}) = \hat{Y}^2 - \frac{1}{\binom{N-2}{n-2} \; q(s)} \; \{ y_s^{\,2} - \frac{N-n}{N-1} \; y_s^* \}$$

$$= a(s) \sum_{i \in s} y_i^2 + b(s) \sum\sum_{i \neq j \in s} y_i \, y_j \qquad (3.3.18)$$

where

$$a(s) = \frac{1}{q(s)} \{ \frac{1}{L^2 \; q(s)} - \frac{1}{\binom{N-2}{n-2}} (1 - \frac{N-n}{N-1}) \} .$$

$$= \frac{1}{q(s)} \; \frac{1}{L} \{ \frac{1}{L \; q(s)} - 1 \}.$$

and

$$b(s) = \frac{1}{q(s)} \left\{ \frac{1}{L^2 \; q(s)} - \frac{1}{\binom{N-2}{n-2}} \right\}$$

Now, a necessary condition that (3.3.18) be uniformly non-negative is that

$$|\Delta(s)| = \begin{vmatrix} a(s) & b(s) & \cdots & b(s) \\ b(s) & a(s) & \cdots & b(s) \\ \vdots & \vdots & & \vdots \\ b(s) & b(s) & \cdots & a(s) \end{vmatrix} \geq 0 \quad s$$

but $|\Delta(s)| = \{a(s) + (n-1)\, b(s)\} \{a(s) - b(s)\}^{(n-1)}$

However, it may be seen that $a(s) - b(s) \geq 0 \;\forall\; s$ but $a(s) + (n-1)\, b(s) \geq 0 \;\forall\; s$ implies $q(s) \leq \frac{1}{\binom{N}{n}} \;\forall\; s$. Hence $\hat{V}(\hat{Y})$ can be uniformly non-negative only in the trivial case of SRSWOR. This result extends readily to the case of ordered samples.

Pursuing the not-too promising line of approach due to Chaudhuri (1976) we may suggest an alternative possibility of getting a uniformly non-negative estimator for $V(\hat{Y})$ as follows.

We may write $V(\hat{Y})$ as

$$V(\hat{Y}) = \sum_i y_i^2 (a_i + \sum_{j(\neq i)} a_{ij}) - \sum_{i<j}\sum a_{ij} (y_i - y_j)^2$$

where $\quad a_i = \frac{1}{L^2} \sum_{s \ni i} \frac{1}{q(s)} - 1, \quad a_{ij} = \frac{1}{L^2} \sum_{s \ni i,j} \frac{1}{q(s)} - 1$

Hence the suggested unbiased estimators are

$$\hat{V}_2(\hat{Y}) = \sum_{i \varepsilon s} \frac{y_i^2}{\pi_i} (a_i + \sum_{j(\neq i)} a_{ij}) - \sum\sum_{i<j \varepsilon s} \frac{a_{ij}}{\pi_{ij}} (y_i - y_j)^2$$

$$\hat{V}_3(\hat{Y}) = \frac{1}{L} \sum_{i \varepsilon s} \frac{y_i^2 (a_i + \sum_{j(\neq i)} a_{ij})}{q(s)}$$

$$- \frac{1}{\binom{N-2}{n-2}} \sum\sum_{i<j \varepsilon s} \frac{a_{ij} (y_i - y_j)^2}{q(s)}$$

$$\hat{V}_4(\hat{Y}) = \sum_{i \varepsilon s} y_i^2 \frac{a_i}{\pi_i} + \sum\sum_{i \neq j \varepsilon s} y_i y_j \frac{a_{ij}}{\pi_{ij}}$$

$$= \sum_{i \varepsilon s} y_i^2 \{ \frac{a_i}{\pi_i} + \sum_{j(\neq i) \varepsilon s} \frac{a_{ij}}{\pi_{ij}} \} - \sum\sum_{i<j \varepsilon s} \frac{a_{ij}}{\pi_{ij}} (y_i - y_j)^2$$

nd

$$\hat{V}_5(\hat{Y}) = \frac{1}{q(s)} [\sum_{i \varepsilon s} y_i^2 (\frac{a_i}{L} + \frac{1}{\binom{N-2}{n-2}} \sum_{j(\neq i) \varepsilon s} a_{ij})$$

$$- \sum\sum_{i<j \varepsilon s} (y_i - y_j)^2 \frac{a_{ij}}{\binom{N-2}{n-2}}]$$

here

$$\pi_i = \sum_{s \ni i} q(s) \qquad \forall\ i = 1, \ldots, N,$$

$$\pi_{ij} = \sum_{s \ni i, j} q(s) \qquad \forall\ i \neq j = 1, \ldots, N.$$

An obvious set of sufficiency conditions for n-negativity of $\hat{V}_2(\hat{Y})$ and $\hat{V}_3(\hat{Y})$ is

$$\left.\begin{array}{ll} a_i + \sum_{j(\neq i)} a_{ij} \geq 0 & \forall\ i = 1, \ldots, N \\ \text{and} & \\ a_{ij} \leq 0 & \forall\ i \neq j = 1, \ldots, N \end{array}\right\} \quad (3.3.19)$$

Now

$$a_{ij} \leq 0$$

$$\Rightarrow \frac{1}{L^2} \sum_{s \ni ij} \frac{1}{q(s)} \leq 1$$

$$\text{i.e. } \frac{1}{L^2} \sum_{i \neq j} \sum \sum_{s \ni ij} \frac{1}{q(s)} \leq N(N-1)$$

$$\text{i.e } \frac{1}{L^2} \sum_{s} \frac{1}{q(s)} \leq \frac{N(N-1)}{n(n-1)}$$

$$\text{i.e. } \frac{\sum_s \frac{1}{q(s)}}{\binom{N}{n}} \leq \frac{N(N-1)}{n(n-1)} \frac{\binom{N-1}{n-1}^2}{\binom{N}{n}}$$

$$\text{i.e. } \frac{\binom{N}{n}}{\sum_s \frac{1}{q(s)}} \geq \frac{(n-1)N}{n(N-1)} \cdot \frac{1}{\binom{N}{n}} \quad (3.3.20)$$

Thus the conditions (3.3.19) lead to the condition that

$0 \leq$ (Arithmetic mean of $q(s)$'s - Hermonic mean of $q(s)$'s)

$$\leq \frac{N-n}{(N-1)} / \{ n \binom{N}{n} \} \quad (3.3.21)$$

which obviously does not imply any inconsistency.

But for large N and n, (3.3.21) would require $q(s)$'s virtually to correspond to the case of SRSWOR.

Sufficient conditions for non-negativity of $\hat{V}_4(\hat{Y})$ are

$$\left.\begin{aligned} &\frac{a_i}{\pi_i} + \sum_{j(\neq i)\in s} \frac{a_{ij}}{\pi_{ij}} \geq 0 \quad \forall\ i \\ \text{and} \quad &a_{ij} \leq 0 \quad \forall\ i,j = 1,\ldots,N\ (i\neq j) \end{aligned}\right\} \qquad (3.3.22)$$

But whenever (3.3.22) holds, so also does (3.3.19). The former is thus a stronger condition and hence is not likely to hold in practice except in the case of SRSWOR.

Sufficiency conditions for non-negativity of $\hat{V}_5(\hat{Y})$ are

$$\left.\begin{aligned} &\frac{a_i}{L} + \frac{1}{\binom{N-2}{n-2}} \sum_{j(\neq i)\in s} a_{ij} \geq 0 \quad \forall\ i \\ \text{and} \quad &a_{ij} \leq 0 \quad \forall\ i,j = 1,\ldots,N\ (i\neq j) \end{aligned}\right\} \qquad (3.3.23)$$

and whenever (3.3.23) holds, so does (3.3.19) and hence our conclusion stands as above.

Bandyopadhyaya, Chattopadhyaya and Kundu (1977) suggested the use of

$$t = t(s) = \sum_i y_i \frac{1}{\alpha_i} \frac{m_i(s)}{q(s)}$$

as a linear unbiased estimator for Y based on a fixed sample size (n) design for which q(s) is the selection probability of the sample s and where $\sum_s m_i(s) = \alpha_i$'s are constants chosen to make t unbiased for Y.

Writing

$$\beta_{ij} = \sum_{s} m_i(s)\ m_j(s)\ \frac{1}{q(s)}\ ,$$

they proposed

$$\hat{V}(t) = \frac{1}{q(s)} \Big[\sum_{i\in s} y_i^2 \{ \binom{N-1}{n-1}^{-2} \beta_{ii} - 1 \} \binom{N-1}{n-1}^{-1}$$

$$+ \sum_{j(\neq i)\in s} y_i^2 \{ \binom{N-1}{n-1}^{-2} \beta_{ij} - 1 \} \binom{N-2}{n-2}^{-1}$$

$$- \sum\sum_{i\neq j\in s} \{ \binom{N-1}{n-1}^{-2} \beta_{ij} - 1 \} \binom{N-2}{n-2}^{-1} (y_i - y_j)^2 \Big]$$

as an unbiased estimator for Var(t).

These authors asserted that a set of sufficient conditions for non-negativity of $\hat{V}(t)$ is that

$$\beta_{ij} \leq \binom{N-1}{n-1}^2 \qquad \forall\ i \neq j = 1, \ldots, N \qquad (3.3.24)$$

However, this assertion by them is dependent on the validity of their psedo-assumption that for any design q(s) the following is true, namely

$$\left.\begin{array}{l} \beta_{ii} \geq \binom{N}{n} \binom{N-1}{n-1} \quad \forall\ i = 1, \ldots, N \\ \text{and} \\ \beta_{ij} \geq \binom{N}{n} \binom{N-2}{n-2} \quad \forall\ i \neq j = 1, \ldots, N \end{array}\right\} \qquad (3.3.25)$$

But, for any design, the following example shows that (3.3.25) may not necessarily be true implying that (3.3.24) is not a sufficient condition for non-negativity of $\hat{V}(t)$.

An example

Let $N = 4$, $n = 2$ and $p_1 = .4$, $p_2 = p_3 = p_4 = .2$ and the samples be $s_1 = (1,2)$, $s_2 = (1,3)$, $s_3 = (1,4)$, $s_4 = (2,3)$, $s_5 = (2,4)$, $s_6 = (3,4)$.

Let us consider the PPSWOR schemes considered among others by Raj (1956) and Murthy (1957), in this situation. Then

$$q(s_1) = p_1 p_2 \left[\frac{1}{1-p_1} + \frac{1}{1-p_2}\right] = \frac{1.4}{6} = q(s_2) = q(s_3)$$

and $q(s_4) = .1 = q(s_5) = q(s_6)$.

In this case

$$\beta_{11} = \sum_{s \ni 1} \frac{1}{q(s)} = \frac{18}{1.4} < 18$$

and $$\beta_{12} = \sum_{s \ni 12} \frac{1}{q(s)} = \frac{6}{1.4} < 6.$$

So, a sufficient condition for non-negativity of $\hat{V}(t)$ would be (3.3.24) simultaneously with the condition

$$\beta_{ii} > \binom{N-1}{n-1}^2 \quad \forall\ i = 1, \ldots, N \qquad (3.3.26)$$

But observing the preceding findings it may well be conjectured that situations where both (3.3.24) and (3.3.26) hold true together shall be rare.

REMARKS

Here we have presented mostly negative results showing the inadequacy and limitations of some of the published results concerning non-negativity of estimators of positive-valued parameters and pointed out the possibility of failures of the same approaches in similar situations. However we should mention that the results due to Vijayan (1975), and Rao and Vijayan (1977) in this area have got positive implications and hence are worthy of attention. But their results are directly applicable only to variances of linear unbiased estimators expressible in the form

$$V = \sum_{i<j}\sum d_{ij}\, w_i\, w_j\, \left(\frac{y_i}{w_i} - \frac{y_j}{w_j}\right)^2$$

under certain conditions.

But their results are not immediately available to provide solutions for estimators due to Sharma (1970) and Bandyopadhyaya, Chattopadhyaya and Kundu (1977) in the general situations as one may check with a little labour. So further works are necessary to get more comprehensive solutions, but as yet we are unable to establish any positive results in this region. Rao and Vijayan (1977) have shown that if there exists a set of non-zero constants w_i's so that the variance of a linear unbiased estimator e for Y namely

$$V(\ell) = \sum_i a_i y_i^2 + \sum_{i \neq j}\sum a_{ij} y_i y_j$$

becomes zero, when $y_i \alpha w_i$, then $V(\ell)$ can be written as

$$V(\ell) = \sum_{i<j}\sum d_{ij} w_i w_j \left(\frac{y_i}{w_i} - \frac{y_j}{w_j}\right)^2 \qquad (3.3.27)$$

From this they concluded the necessary forms of unbiased estimators for $V(\ell)$ and **deduced** conditions for the non-negativity of the latter.

However, their approach does not work if the proportionality of y_i's with some constants w_i's does not make $V(\ell)$ zero. Obviously, the converse to Rao and Vijayan's r.sult is also true. Going a step further we may assert that the form (3.3.27) of $V(\ell)$ is available if and only if $\exists$ w_i's such that $y_i \alpha w_i$ and when this is true the value of ℓ for the corresponding y_i's is a constant for every s with its selection.probability $q(s) > 0$.

From this it follows that the Horvitz-Thompson estimator has its variance expressible in the form (3.3.27) if and only if it is based on a fixed sample size design.

However for Sharma's (1970) estimator for Y the ariance is expressible in the form (3.3.27) iff $\exists$ w_i's uch that

$$y_i \propto w_i \;\Rightarrow\; \frac{\sum_{i \epsilon s} y_i}{L\, q(s)} = C \frac{\sum_{i \epsilon s} w_i}{L\, q(s)} = \text{constant}$$

(where C is a constant)

i.e. iff $(\sum_{i \epsilon s} w_i) \propto q(s)$

i.e. iff $q(s) = K \sum_{i \epsilon s} w_i$ (3.3.28)

Naturally, then, $V(\hat{Y})$ is not expressible as (3.3.27) and consequently, Rao and Vijayan's (1977) approach is not applicable in estimating $V(\hat{Y})$ unbiasedly and non-negatively, if $\hat{Y}$ is based on pps sampling without replacement scheme for which (3.3.28) does not hold. A similar remark applies to the estimator due to Bandyopadhyaya, Chattopadhyaya and Kundu (1977).

3.4 UNBIASED VARIANCE-ESTIMATION IN MULTI-STATE SAMPLING STRATEGIES

3.4.1 SAMPLING STRATEGIES A, B AND C

For a population of N first-stage units (f.s.u), let us have a scheme of sampling when on the rth draw the ith f.s.u is selected with probability $p_i(r)$, draws being made with replacement ($r = 1, \ldots, n$, $i = 1, \ldots, N$) such that $\Sigma p_i(r) = 1 \; \forall r = 1, \ldots, n$ and every f.s.u so selected is sub-sampled, as it is drawn, in subsequent stages employing arbitrary selection probabilities. Denoting by Y_i, the ith f.s.u total, the customary problem is to unbiasedly estimate $Y = \sum_i Y_i$, the population total, on the basis of a sample so drawn. Here we propose to employ a general linear estimator of the following form namely,

$$T = T(s) = \sum_i \sum_r b_{s_i}(r) \; T_i(r)$$

where

$$b_{s_i}(r) \begin{cases} = 0 & \text{if the rth draw does not yield the ith f.s.u of the sample s} \\ \neq 0 & \text{otherwise} \end{cases}$$

such that $T_i(r)$ is a statistic calculated on the basis of the sample values of the variable y under study defined in respect of the sampling units of the ultimate stage sub-sampled from the ith f.s.u when it is selected on the rth draw such that with respect to the sampling design adopted in sub-sequent stages $T_i(r)$ is an unbiased estimator for Y_i for each r having the variance $V(T_i(r)) = \sigma_i^2(r)$ $(r = 1, \ldots, n, \quad i = 1, \ldots, N)$ (these expectations and variances relate to the sampling designs adopted in sub-sampling the selected f.s.u's). Also $b_{s_i}(r)$'s are independent of y-values and are so chosen that T is an unbiased estimator for Y. The strategy so described will be called the strategy A.

An alternative strategy to be denoted as strategy B is one where a sample is chosen in the same manner as above but the linear unbiased estimator we consider is based only on the distinct f.s.u's included in the selected sample and is of the following form .

$$T^{+} = T^{+}(s) = \sum_{i \epsilon s} C_{s_i} \; T_i^{*}$$

where T_i^* is a function of the sample-values of y for the units of the ultimate stage sub-sampled from the ith f.s.u whenever it is included in a sample such that with respect to the sampling design in subsequent stages its expectation and variance become Y_i and σ_i^2 respectively for every i, C_{s_i}'s are so chosen that T^+ is unbiased for Y, the symbol

$\sum_{i \in s}$ denoting summation over distinct f.s.u's only in ample s. Here it is supposed that from the consideration of 'sufficiency' once the f.s.u's are selected with replacement we retain the distinct f.s.u's alone and discard the nformation about which draws produced which f.s.u's and each selected f.s.u is sub-sampled just once in an independent manner.

Finally we consider a strategy C where the f.s.u's are selected with replacement as in strategy A and B but the homogenous linear unbiased estimator for Y is of the form

$$T_+ = T_+(s) = \sum_{i \in s} C_{s_i} \; T_i^*(\ell)$$

which is based on the distinct f.s.u's (selected with replacement) only in s but here we retain the information about the particular draws ℓ (= 1, ...λ_i (say)) on which the ith f.s.u (i = 1, ...N) occurs inthe sample s of f.s.u's and decide to sub-sample the ith f.s.u with a probability $q_i(\ell)$ [where $0 < q_i(\ell) < 1$, such that $\sum_{\ell} q_i(\ell) = 1$ for $i = 1, \ldots, N$] when it is selected on the ℓ th draw in which case we estimate Y_i by $T_i^*(\ell)$ on the basis of the sub-samples so drawn.

3.4.2 UNBIASED VARIANCE ESTIMATION IN STRATEGY A

Denoting by E_1, E_L and V_1, V_L the operators for mathematical expectations and variances in the initial and subsequent stages of sampling respectively, and by E and V the corresponding operators for the overall sampling scheme we have

$$E(T) = E_1 [\sum_i \sum_r b_{s_i}(r) \; E_L(T_i(r))]$$

$$= E_1 [\sum_i \sum_r b_{s_i}(r) \; Y_i]$$

$$= E_1 [\sum_i Y_i \, d_{si}] \text{, on writing } d_{si} = \sum_r b_{si}(r),$$

$$= \sum_i Y_i \sum_s d_{si} \; P(s),$$

P(s) denoting the probability of selecting a sample s of f.s.u's in n draws. Obviously, because of the unbiasedness condition we require [redacted] the b_{s_i}'s satisfy the condition

$$\sum_s d_i(s) \; P(s) = 1 \quad \forall \, i = 1, \ldots, N \qquad (3.4.1)$$

Assuming throughout that the condition (3.4.1) is satisfied we have

$$V(T) = E_1 [\sum_i \sum_r b^2_{s_i}(r) \; \sigma_i^2(r)] + V_1 [\sum_i Y_i \, d_{si}]$$

[remembering that $T_i(r)$ and $T_j(r')$ are uncorrelated whenever $i \neq j$ or $r \neq r'$ or both at a time]

$$= \sum_s \left(\sum_i \sum_r b_{si}^2(r)\ \sigma_i^2(r)\right) P(s) + \sum_i Y_i^2\left(\sum_s d_{si}^2\ P(s) - 1\right)$$

$$+ \sum_{i \neq j}\sum \left(\sum_s d_{si}\ d_{sj}\ P(s) - 1\right) Y_i\ Y_j$$

$$= \sum_i \sum_r \sigma_i^2(r)\ \alpha_i(r) + \sum_i Y_i^2 \beta_i + \sum_{i \neq j}\sum Y_i\ Y_j\ \gamma_{ij}$$

On writing

$$\alpha_i(r) = \sum_s b_{si}^2(r)\ P(s)$$

$$\beta_i = \sum_s d_{si}^2\ P(s) - 1,\ \forall\ i = 1, \ldots, N$$

$$\gamma_{ij} = \sum_s d_{si}\ d_{sj}\ P(s) - 1\ \forall\ i \neq j = 1, \ldots, N.$$

Our problem is to estimate V(T) unbiasedly from the sampled data such that the resulting estimator does not involve any unbiased estimator of $\sigma_i^2(r)$'s, but a function of $T_i(r)$'s alone. At this stage let us note the special case for which a simple well-known solution exists as described below. Suppose the f.s.u's are selected in n draws with replacement with suitable selection probabilities and the selected f.s.u's are independently sub-sampled such that for every rth draw of an f.s.u, t_r is an unbiased estimator for Y for each r such that t_r's are independently distributed. Then T becomes

$$T = \frac{1}{n} \sum_r t_r$$

and $E(T) = \frac{1}{n} \sum_r E(t_r) = Y$

nd it follows that $\frac{1}{n(n-1)} \sum_r (t_r - T)^2$ is an unbiased estimator for V(T) because

$$E \sum_r (t_r - T)^2 = E \sum_r [(t_r - Y) - (T - Y)]^2$$

$$= \sum_r E (t_r - Y)^2 + n\, E\, (T - Y)^2 - \frac{2}{n} E\, (t_r - Y)$$

$$= \sum_r \sigma_r^2 + \frac{1}{n} \sum_r \sigma_r^2 - \frac{2}{n} \sum_r \sigma_r^2$$

$$[\text{where } \sigma_r^2 = E\, (t_r - Y)^2\,]$$

$$= \frac{n-1}{n} \sum_r \sigma_r^2$$

and

$$V(T) = E(T - Y)^2 = E\,[\frac{1}{n} \sum_r (t_r - Y)]^2$$

$$= \frac{1}{n^2} \sum_r \sigma_r^2 .$$

so that

$$E \frac{1}{n(n-1)} \sum_r (t_r - T)^2 = \frac{1}{n^2} \sum_r \sigma_r^2 = V(T) .$$

However, we shall consider our linear unbiased estimator T so general that $b_{si}(r)$ is subject just to (3.4.1). So, in getting an unbiased estimator for V(T) in terms of $T_i(r)$'s we shall seek one within the following class of homogeneous quadratic functions of $T_i(r)$'s as

$$t_1 = \sum_r C_r\, T_r^2 + \sum_{r \neq r'}\sum C_{rr'}\, T_r\, T_{r'} \qquad (3.4.2)$$

where

$C_r = q_i(r)$ if rth draw produces the ith f.s.u with probability $P_i(r)$ in which case $T_r = T_i(r)$

$_{rr'} = e_{ii}$ if rth as well as r'th draw produces the ith f.s.u with probability $p_i(r)\ p_i(r')$

$_{rr'} = f_{ij}$ if rth draw produces ith f.s.u and r'th draw produces jth f.s.u $i \neq j = 1, \ldots N$ with probability for this being $p_i(r)\ p_j(r')$.

ow,

$$E(t_1) = \sum_r \sum_i g_i(r)\ (Y_i^2 + \sigma_i^2(r))\ p_i(r)$$

$$+ \sum\sum_{r \neq r'} \sum_i e_{ii}\ Y_i^2\ p_i(r)\ p_i(r')$$

$$+ \sum\sum_{r \neq r'} \sum\sum_{i \neq j} f_{ij}\ Y_i\ Y_j\ p_i(r)\ p_j(r').$$

$$= \sum_i \sum_r \sigma_i^2(r)\ g_i(r)\ p_i(r)$$

$$+ \sum_i Y_i^2 \{\sum_r g_i(r)\ p_i(r) + e_{ii} \sum\sum_{r \neq r'} p_i(r) p_i(r')\}$$

$$+ \sum\sum_{i \neq j} Y_i\ Y_j\ f_{ij} \sum\sum_{r \neq r'} p_i(r)\ p_j(r').$$

Now, if we choose

$$g_i(r) = \frac{\sigma_i(r)}{p_i(r)} \quad \forall\ i = 1, \ldots, N\ ;\ r = 1, \ldots, n\ ,$$

$$e_{ii} = \frac{\beta_i - \sum_r \alpha_i(r)}{\sum\sum_{r \neq r'} p_i(r)\ p_i(r')} \quad \forall\ i = 1, \ldots, N.$$

and

$$f_{ij} = \frac{Y_{ij}}{\sum\sum_{r \neq r'} p_i(r)\ p_j(r')} \quad \forall\ i \neq j = 1, \ldots, N,$$

then t_1 becomes unbiased for V(T) and such choices are clearly possible. We may cite as examples the following

<u>Example - 1</u>

As a generalization of the estimator due to Bandyopadhyaya, Chattopadhyaya and Kundu (1977) we may consider the estimator

$$e = \sum_i \frac{n_i(s)}{K_i(s)} \frac{Y_i}{P(s)}$$

for Y in case of uni-stage sampling scheme to choose a sample s with a probability P(s) with or without replacement such that

$$\sum_s \frac{n_i(s)}{K_i(s)} = 1 \quad \forall\ i = 1, \ldots, N \qquad (3.4.3)$$

where $n_i(s)$ = frequency of ith f.s.u in the sample s of f.s.u. In multi-stage sampling an estimator of this general type should be chosen as

$$T(s) = \sum_i \sum_r a_i(s) \frac{T_i(r)}{K_i(s)\ P(s)}$$

where $a_i(s)$ = 1, if ith f.s.u is included in s and zero otherwise, that

$$E(T(s)) = E_1 \left[\sum_i Y_i \sum_r \frac{a_i(s)}{K_i(s)\; P(s)} \right]$$

$$= E_1 \sum_i Y_i \frac{n_i(s)}{K_i(s)\; P(s)}$$

$$= Y \quad [\text{ provided (3.4.3) is assumed }]$$

Here clearly, the condition for unbiasedness of $T(s)$ is condition of the type (3.4.1) and an unbiased estimator for $V(T(s))$ of the form (3.4.2) is clearly available as one may check following the line discussed above.

Example - 2

If $p_i(r) = p_i$ for every r, then t_1 will be determined on choosing

$$g_i(r) = \frac{\alpha_i(r)}{p_i} \quad \forall\ i = 1, \ldots, N$$
$$\forall\ r = 1, \ldots, n,$$

$$e_{ii} = \frac{\beta_i - \sum_r \alpha_i(r)}{n(n-1)\ p_i^2} \quad \forall\ i = 1, \ldots, N,$$

$$f_{ij} = \frac{\gamma_{ij}}{n(n-1)\ p_i\ p_j} \quad \forall\ i \neq j = 1, \ldots, N.$$

f, in addition, $\sigma_i^2(r) = \sigma_i^2\ \forall\ r = 1, \ldots, n$ then we may choose

$$g_1(r) = g_i = \frac{\alpha_i}{n\ p_i},$$

$$e_{ii} = \frac{\beta_i - \alpha_i}{n(n-1)\ p_i^{\,2}} \quad \forall\ i = 1, \ldots, N$$

with $\alpha_i = \sum_r \alpha_i(r)$

and $$f_{ij} = \frac{\gamma_{ij}}{n(n-1)\ p_i\ p_j} \quad \forall\ i \neq j = 1, \ldots, N.$$

Example - 3

The situation is well-known [vide Durbin (1953), Raj (1968)] where $p_i(r) = p_i\ \forall\ r = 1, \ldots, n$, $K_i(s) = \frac{np_i}{P(s)}$ $\forall\ i = 1, \ldots, N$, $T_i(r) = T_i$ (say) for each $r = 1, \ldots, n$, $\sigma_i^{\,2}(r) = \sigma_i^{\,2}\ \forall\ r = 1, \ldots, n$.

3.4.3 UNBIASED VARIANCE ESTIMATION IN STRATEGY B

In strategy B we have

$$E(T^{+}) = E_1 \sum_{i \in s} C_{si}\ Y_i = \sum_i Y_i \sum_{s \ni i} C_{si}\ P(s).$$

So, for the sake of unbiasedness of T_1 for Y the sufficiency condition is that

$$\sum_{s \ni i} C_{si}\ P(s) = 1 \quad \forall\ i = 1, \ldots, N \qquad (3.4.4)$$

Assuming this is to be satisfied we have

$$V(T^{+}) = \sum_i \sigma_i^{\,2} \Big(\sum_{s \ni i} C_{si}^2\ P(s) \Big) + \sum_i Y_i^{\,2} \Big(\sum_{s \ni i} C_{si}^2 P(s) - 1 \Big)$$

$$+ \sum_{i \neq j} \sum Y_i\ Y_j \Big(\sum_{s \ni i j} C_{si}\ C_{sj}\ P(s) - 1 \Big)$$

ow, if possible, let an unbiased estimator for $V(T^+)$ be vailable as a quadratic function of T_i^* as

$$v_1 = \sum_{i \in s} h_{si} T_i^{2*} + \sum_{i \neq j \in s}\sum h_{sij} T_i^* T_j^* \qquad (3.4.5)$$

en,

$$E(v_1) = \sum_i (Y_i^2 + \sigma_i^2) \left(\sum_{s \ni i} h_{si} P(s) \right)$$

$$+ \sum_{i \neq j}\sum Y_i Y_j \sum_{s \ni i,j} h_{sij} P(s) .$$

ow in order than v_1 be an unbiased estimator for $V(T_1)$ e must have

$$\sum_{s \ni i} h_{si} P(s) = \sum_{s \ni} C_{si}^2 P(s) \quad \forall \ i = 1, \ldots, N,$$

nd at the same time $\sum_{s \ni i} h_{si} P(s)$ should equal $\sum_{s \ni i} C_{si}^2 P(s) - 1$ $i = 1, \ldots, N$. But this is absurd. So an unbiased estimator f $V(T^+)$ of the form (3.4.5) does not exist. If, however, n unbiased estimator for σ_i^2 is available as $\hat{\sigma}_i^2$ on the asis of units sub-sampled from the ith f.s.u in later tages when it is selected, an unbiased estimator for $V(T^+)$ available as

$$v_1' = v_1 + \sum_{i \in s} \frac{\hat{\sigma}_i^2}{\pi_i}, \quad \text{where} \quad \pi_i = \sum_{s \ni i} P(s)$$

rovided $\pi_i > 0$ and we have

$$\sum_{s \ni i} h_{si} P(s) = \sum_{s \ni i} C_{si}^2 P(s) - 1 \ \forall \ i = 1, \ldots, N.$$

and $\sum_{s \ni i,j} h_{sij} P(s) = \sum_{s \ni ij} C_{si} C_{sj} P(s) - 1 \;\forall\; i \neq j = 1, \ldots, N.$

It readily follows that the result applicable to the strategy B carries over to situations when the f.s.u's are selected without replacement as one may check with reference to well-known literature on the subject.

3.4.4 UNBIASED VARIANCE ESTIMATION WITH STRATEGY C

For the strategy C let us denote by E_2, V_2 the expectations and variances with respect to the probabilities $q_i(\ell)$'s and E_L', V_L' the conditional expectations and variances with respect to the sub-sampling designs given by the estimators $T_i^*(\ell)$ $(\ell = 1, \ldots, \lambda_i, \quad i = 1, \ldots, N)$ actually chosen, we have,

$$E_L(T_i^*(\ell)) = E_2(E_L' \; T_i^*(\ell)) = \sum_{K=1}^{\lambda_i} q_i(K) Y_i = Y_i$$

$$[\text{assuming } E_L' \; T_i^*(K) = Y_i \;\forall\; K = 1, \ldots, \lambda_i]$$

and

$$V_L(T_i^*(\ell)) = E_2 \,[V_L' \; (T_i(\ell))] + V_2 \,[E_L'(T_i(\ell))]$$

$$= \sum_{K=1}^{\lambda_i} q_i(K) \; \sigma_i^2(K) \quad \text{for } i = 1, \ldots, N, \quad \ell = 1, \ldots, \lambda_i$$

$$= \xi_i \quad (\text{say})$$

o,

$$V(T_+) = E_1 \sum_{i\varepsilon s} C_{si}^2 V_L(T_i^*(\ell)) + V_1 \sum_{i\varepsilon s} C_{si} E(T_i^*(\ell))$$

$$= E_1 \sum_{i\varepsilon s} C_{si}^2 \xi_i + V_1 (\sum_{i\varepsilon s} C_{si} Y_i)$$

nd in estimating $V(T_+)$ unbiasedly we encounter the ircumstances similar to those obtaining in the case of trategy B.

CHAPTER - 4

SURVEYING A FINITE POPULATION ON TWO OCCASIONS - UNISTAGE SAMPLING.

4.1 MOTIVATION AND SUMMARY

In this chapter we propose some sampling strategies for estimating the population total for the current occasion based on the information collected on this as well as on a previous occasion. Some of these strategies proposed are compared with the existing sampling strategies due to Raj (1965), Rao and Ghangurde (1969), Avadhani and Sukhatme (1970) and Chotai (1974) for sampling over two occasions and found to fare better under some situations.

4.2 INTRODUCTION

We consider a finite universe U of N units labelled (1,...,N) which is supposed to remain unchanged over two occasions such that x_i and y_i $(i = 1,...,N)$ are the values of the characteristic for the first (x) and second (y) occasions respectively. Our object is to devise a suitable sampling strategy to estimate $Y = \sum_i y_i$, the population total on the second occasion on utilizing results of two consecutive surveys to be undertaken on this population, suffering no change in composition over the time-lag.

Several strategies are available in the literature or the purpose. But inthis chapter we shall consider only a few of those involving schemes of selection with varying probabilities. In particular we shall be concerned with those due to Raj (1965), Pathak and Rao (1967), Rao and Ghangurde (1969), Avadhani and Sukhatme (1970) Chotai (1974) and Singh (1975). We have proposed a few sampling strategies and some of them are found superior to those mentioned earlier under certain situations, as described n subsequent selections of this chapter.

4.3 A BRIEF REVIEW OF EARLIER STRATEGIES RELEVANT TO OUR WORK

STRATEGY A due to RAJ (1965)

Raj (1965) considered a strategy where on the first occasion a sample S_1 of size n is selected with replacement with selection probabilities proportional to uitably chosen normed size-measures p_i $(i = 1, \ldots, N)$ of the units. On the 2nd occasion a simple random sub-sample S_{2m} (say) without replacement (SRSWOR) of $m = n\lambda$ (integer, with $0 \leq \lambda \leq 1$) units is retained out of S_1 and an independent sample S_{2u} of size $u = n-m = n\mu$ (with $\mu = 1-\lambda$) is selected from the entire population U again following the scheme of selection with probability proportional to size with replacement (PPSWR) using the earlier p_i's.

The estimator employed is of the form

$$\hat{Y} = \phi \hat{Y}_{2m} + (1-\phi) \hat{Y}_{2u}$$

with

$$\hat{Y}_{2m} = \frac{1}{m} \sum_{S_{2m}} \frac{y_i - x_i}{p_i} + \frac{1}{n} \sum_{S_1} \frac{x_i}{p_i},$$

$$\hat{Y}_{2u} = \frac{1}{u} \sum_{S_{2u}} \frac{y_i}{p_i},$$

and ϕ as a constant such that $0 \leq \phi \leq 1$ to be so determined as to minimize the variance of $\hat{Y}$.

[$\sum_{Sm}$ denoting sum over the units in S_{2m}, $\sum_{S_1}$, $\sum_{S_{2u}}$ denoting sums over n and u terms corresponding to n and u draws in S_1 and S_{2u} respectively]

Clearly $E(\hat{Y}_{2m}) = E(\hat{Y}_{2u}) = E(\hat{Y}) = Y$.

Writing $z_i = y_i - x_i$ and $Z = \sum_i z_i$

one gets

$$V(\hat{Y}_{2m}) = \frac{1}{n} V_{pps}(y) + (\frac{1}{m} - \frac{1}{n}) V_{pps}(z),$$

$$V(\hat{Y}_{2u}) = \frac{1}{u} V_{pps}(y),$$

and $$\text{Cov}(\hat{Y}_{2m}, \hat{Y}_{2u}) = 0$$

Writing

$$V_{pps}(t) = \sum_i p_i (\frac{t_i}{p_i} - T)^2 \quad \text{and} \quad T = \sum_i t_i$$

where t stands for x, y, z and T for X, Y, Z respectively.

Assuming (as we shall throughout) that

$$V_{pps}(x) = V_{pps}(y) = V_o \quad \text{(say)} \tag{4.3.1}$$

and writing

$$\rho V_o = \sum_i p_i (\frac{y_i}{p_i} - Y)(\frac{x_i}{p_i} - X), \tag{4.3.2}$$

Raj (1965) obtained

$$V(\hat{Y}_{2m}) = \frac{V_o}{n} + 2(\frac{1}{m} - \frac{1}{n})(1-\rho)V_o,$$

and $$V(\hat{Y}_{2u}) = \frac{V_o}{u}.$$

The optimum value of Var $(\hat{Y})$, being the minimum with respect to ϕ for a fixed λ, is

$$V_{opt}(\hat{Y}|\lambda) = \frac{V(\hat{Y}_{2m})\ V(\hat{Y}_{2u})}{V(\hat{Y}_{2m}) + V(\hat{Y}_{2u})}$$

Also, Raj obtained the optimum values of μ as

$$\mu_o \text{ (say)} = [1 + \sqrt{2(1-\rho)}]^{-1}$$

$\lambda = \lambda_o$ (say) $= 1-\mu_o$ and the corresponding minimum (with respect to ϕ and λ) variance of $\hat{Y}$ as

$$V_{min}(\hat{Y}) = \underset{\phi,\lambda}{\text{Min}}\ \text{Var}(\hat{Y}) = \underset{\lambda}{\text{Min}}\ V_{opt}(\hat{Y}|\phi)$$

$$= \frac{V_o}{2n}\left[1 + \sqrt{2(1-\rho)}\right]$$

STRATEGY B: due to GHANGURDE AND RAO (1969)

On the 1st occasion a sample S_1 of size n is chosen following Rao-Hartley-Cochran (1962) (RHC, in brief) method of sampling with normed size measures p_i's (as in strategy A) and on the second occasion a sub-sample S_{2m} of size $m = n\lambda$ (integer, with $0 \le \lambda \le 1$) is chosen out of S_1 following SRSWOR method as in Raj's scheme and an independent sample S_{2u} of size $u = n-m = n\mu$ (with $\mu = 1-\lambda$) is chosen from U following RHC method again using the previous size measures p_i's. Their proposed estimator for Y is of the form

$$\hat{Y}' = \phi\ \hat{Y}'_{2m} + (1-\phi)\ \hat{Y}'_{2u}$$

with

$$\hat{Y}'_{2m} = \frac{n}{m} \sum_{S_{2m}} \frac{y_i - x_i}{p_i} P_i + \sum_{S_1} \frac{x_i}{p_i} P_i$$

$$\hat{Y}'_{2u} = \sum_{S_{2u}} \frac{y_i}{p_i} P_i^{*}.$$

here P_i and P_i^* are the total of p-values of the group containing the ith unit ($i = 1, \ldots, N$) in selection of S_1 and S'_{2u} respectively, the groups being randomly (as usual) formed while applying RHC scheme.

Assuming

$$V_{pps}(y) = V_{pps}(x) = V_o \quad \text{(say), as in Raj (1965)}$$

and $$V(y) = V(x) = \bar{V} \quad \text{(say)} \qquad (4.3.3)$$

where

$$V(t) = N \sum_i (t_i - \bar{T})^2 , \quad \text{with } \bar{T} = \frac{1}{N} \sum_i t_i$$

$$\text{for } t = x, y \text{ respectively}$$

and writing

$$\delta \bar{V} = \frac{1}{N} \sum_i (y_i - \bar{Y})(x_i - \bar{X})$$

one has

$$V(\hat{Y}'_{2m}) = 2\left(\frac{1}{m} - \frac{1}{n}\right) \frac{1}{N-1} \left[(N-n)(1-\rho) V_o + n \bar{V} (1-\delta) \right] + \frac{N-n}{n} \frac{V_o}{N-1} ,$$

$$V(\hat{Y}'_{2u}) = \frac{N-u}{u(N-1)} V_o$$

$$\text{Cov}(\hat{Y}'_{2m}, \hat{Y}'_{2u}) = 0$$

and the optimum value of variance of $\hat{Y}'$ (for variation of ϕ) for a fixed λ is denoted by $V_{opt}(\hat{Y}' | \lambda)$ and is obtained by the authors as

$$V_{opt}(\hat{Y}'|\lambda) = \underset{\phi}{\text{Min}}\ V(\hat{Y}'|\lambda)$$

$$= \frac{N}{n(N-1)} V_o [\mu(1 - \frac{n}{N}\mu)^{-1}$$

$$+ \lambda \{(1 - \frac{n}{N})\lambda + 2\mu (1-\rho) + (1 + \frac{n}{N}\gamma)\}^{-1}]^{-1}$$

where

$$\gamma = \frac{(1-\delta)\bar{V}}{(1-\rho)V_o} - 1.$$

Chotai (1974) obtained the optimum value of μ as

$$\mu_o \text{ (say)} = [1 + \sqrt{2(1-\rho)(1 + \gamma \frac{n}{N})}]^{-1}$$

and the minimum value $V(\hat{Y}')$ for variation of λ as

$$V_{min}(\hat{Y}') = \underset{\lambda}{\text{Min}}\ V_{opt}(\hat{Y}'|\lambda)$$

$$= \frac{N}{(N-1)\ 2n} [(1 - \frac{n}{N}) + \sqrt{2(1-\rho)(1 + \gamma \frac{n}{N})}]\, V_o.$$

STRATEGY C: due to CHOTAI (1974)

Chotai (1974) modified the strategy of Ghangurde and Rao (1969) where S_1 and S_{2u} are selected as in Ghangurde and Rao (1969)-method but S_{2m} is selected from S_1 again following RHC scheme replacing p_i's by P_i. The estimator employed is

$$\hat{Y}'' = \phi\ \hat{Y}''_{2m} + (1-\phi)\ \hat{Y}'_{2u} \qquad (4.3.4)$$

Let

$$\hat{Y}''_{2m} = \sum_{S_{2m}} \frac{y_i - x_i}{p_i} \Delta_i + \sum_{S_1} \frac{x_i}{p_i} P_i\ ,$$

$$\hat{Y}'_{2u} = \sum_{S_{2u}} \frac{y_i}{p_i} P_i^*$$

re Δ_i is the sum of P-values for the group containing ith unit ($i = 1, \ldots, N$) in selecting the sample S_{2m} from following RHC scheme of sampling.

uming $V_{pps}(y) = V_{pps}(x) = V_o$, one has

$$V(\hat{Y}''_{2m}) = \frac{N}{n(N-1)} V_o [(1 - \frac{n}{N}) + 2(\frac{1}{\lambda} - 1)(1-\rho)],$$

$$V(\hat{Y}'_{2u}) = \frac{N-u}{u(N-1)} V_o ,$$

$$Cov(\hat{Y}''_{2m}, \hat{Y}'_{2u}) = 0.$$

optimum value of $V(\hat{Y}'')$ with respect to ϕ for given λ is

$$V_{opt}(\hat{Y}'' | \lambda) = \underset{\phi}{Min}\ V(\hat{Y}''|\lambda)$$

$$= \frac{N}{n(N-1)} V_o [\frac{\mu}{1 - \frac{n\mu}{N}} + \frac{\lambda}{2\mu(1-\rho) + \lambda(1 - \frac{n}{N})}]^{-1} \quad (4.3.5)$$

the corresponding value of $\mu = \mu_o$ (say) is

$$\mu_o = \{1 + \sqrt{2(1-\rho)}\}^{-1} \quad (4.3.6)$$

that of λ is $\lambda_o = 1 - \mu_o$

the minimum value of the variance of $\hat{Y}''$ is

$$V_{min}(\hat{Y}'') = \underset{\phi, \lambda}{Min}\ V(\hat{Y}'') = \underset{\lambda}{Min}\ V_{opt}(\hat{Y}''|\lambda)$$

$$= \frac{N}{2n(N-1)} [1 + \sqrt{2(1-\rho)} - \frac{n}{N}] V_o \quad (4.3.7)$$

Finally, Chotai compared the relative efficiencies of the above three strategies A, B, C by replacing $N(N-1)^{-1}$ by 1 as follows :

(i) $V_{min}(\hat{Y}') < V_{min}(\hat{Y}'') < V_{min}(\hat{Y})$ if $\frac{\tilde{V}}{V_o} < d$

(ii) $V_{min}(\hat{Y}'') < V_{min}(\hat{Y}') < V_{min}(\hat{Y})$ if $d < \frac{\tilde{V}}{V_o} < b$

(iii) $V_{min}(\hat{Y}'') < V_{min}(\hat{Y}) < V_{min}(\hat{Y}')$ if $b < \frac{\tilde{V}}{V_o}$

where

$$b = \frac{1-\rho}{1-\delta} + \frac{n/(2N) + \sqrt{2(1-\rho)}}{1-\delta} \quad \text{and} \quad d = \frac{1-\rho}{1-\delta}$$

STRATEGY D due to PATHAK AND RAO (1967) and STRATEGY E due to SINGH (1972).

Strategies D and E are based on the same sampling scheme due to Raj (1965). Pathak and Rao T.J. (1967) considered the estimator for population total as

$$Z^* = \frac{\phi}{n-m}\left[\sum_{S'_{2u}} \frac{y_i}{p_i} + m_2 \frac{\left(\sum_{S_{2m}} \frac{y_i}{p_i}\right)}{\left(\sum_{S_{2m}} p_i\right)}\right] + (1-\phi)\,\hat{Y}_{2m}$$

where $S_{2u} = S_{2mu} + S'_{2u}$,

where S_{2mu} consists of m_2 units coming from the matched sample S_{2m} and S'_{2u} consists of $u-m_2$ units from U/S_{2m}.

Pathak and Rao, T.J. (1967) applied Rao-Blackwellization to show that variance of the proposed estimator Z* is smaller than that of $\hat{Y}$.

Ravindra Singh (1972) further observed S_{2mu} to be sub-sample from S'_{2m} which is subset of S_{2m} consisting of the distinct units alone in the latter and sought a further improvement by Rao-Blackwellization. He proposed the estimator

$$Z^{**} = \frac{\phi}{n-m}\left[\sum_{S'_{2u}} \frac{y_i}{p_i} + m_2 \sum_{S'_{2m}} y_i \Big/ \sum_{S'_{2m}} p_i\right] + (1-\phi)\hat{Y}_m$$

and showed that Z** possesses a smaller variance than that of Z*.

STRATEGIES F and G: due to AVADHANI AND SUKHATME (1970)

STRATEGY F:

In this strategy, on the first occasion an SRSWOR sample S_1 of size n is taken from U and on the second occasion a subsample S_{2m} of size m is chosen from S_1 by following RHC scheme using normed size measure $p_i' = \frac{x_i}{\sum_{S_1} x_i}$ for the unit $i \in S_1$ and an SRSWOR sample S_{2u} of size u = n-m is chosen from the units of U not included in S_1.

The proposed estimator for the population mean $\bar{Y}$ is given by

$$\hat{\bar{Y}} = \phi\ \bar{y}_{RHC} + (1-\phi)\,\bar{y}_u$$

with

$$\bar{y}_{RHC} = \frac{1}{n} \sum_{S_{2m}} \frac{y_i}{p_i'}\ P_i'$$

$$\bar{y}_u = \frac{1}{u} \sum_{S_{2u}} y_i$$

where P_i' = sum of p_i'-values in the group containing the ith unit ($i = 1, \ldots, N$) in selecting the sample S_{2m} by RHC scheme from S_1.

Then

$$E(\bar{y}_{RHC}) = E\ \bar{y}_u = \bar{Y},$$

$$V(\bar{y}_{RHC}) = (\frac{1}{n} - \frac{1}{N})\ S_y^{\ 2} + (\frac{1}{m} - \frac{1}{n})\ \frac{1}{N(N-1)} \sum_i \frac{1}{p_i}\ (y_i - Rx_i)^2,$$

(where $R = Y/X$),

and

$$\mathrm{Cov}\ (\bar{y}_u,\ \bar{y}_{RHC}) = -\frac{S_y^{\ 2}}{N},\ (\text{where } S_y^{\ 2} = \frac{1}{N-1} \sum_i (y_i - \bar{Y})^2)$$

The value of ϕ which minimizes $V(\hat{\bar{Y}})$ for given λ is $\phi_o = V_1' / (V_1' + V_2')$ and the corresponding optimum value of $V(\hat{\bar{Y}})$ is

$$V_{opt}(\hat{\bar{Y}}|\lambda) = \underset{\phi}{\mathrm{Min}}\ V(\hat{\bar{Y}}|\lambda)$$

$$= \frac{V_1'\ V_2'}{V_1' + V_2'} + V_{12}$$

where $V_1' = V(\bar{y}_{RHC}) - V_{12}$

$$V_2' = V(\bar{y}_u) - V_{12}$$

$$V_{12} = Cov(\bar{y}_{RHC}, \bar{y}_u).$$

Avadhani and Sukhatme (1970) obtained the optimum value of μ for which $V_{opt}(\hat{\bar{Y}})$ becomes minimum as

$$\mu = \mu_o \text{ (say) } = (1 + \sqrt{\delta})^{-1} \text{ with } \delta = \frac{1}{N} \frac{\Sigma \frac{1}{p_i}(y_i - Rx_i)^2}{\Sigma (y_i - \bar{Y})^2}$$

and the minimum variance of $\hat{\bar{Y}}$ as

$$V_{min}(\hat{\bar{Y}}) = \underset{\phi, \lambda}{Min}\, V(\hat{\bar{Y}}) = \left(\frac{1 + \sqrt{\delta}}{2n} - \frac{1}{N}\right) S_y^2 \qquad (4.3.8)$$

Now putting optimum value of μ in ϕ_o we get the optimum value of ϕ as $\phi_{opt} = \frac{1}{2}$ and putting $\phi = \frac{1}{2}$ in $\hat{\bar{Y}}$ we get

$$\hat{\bar{Y}} = \frac{1}{2}[\bar{y}_{RHC} + \bar{y}_u]$$

STRATEGY G:

In this strategy S_1 and S_{2u} are selected as in strategy F but S_{2m} is selected from S_1 by SRSWOR.

The proposed estimator for population mean $\bar{Y}$ is given by

$$\hat{\bar{Y}}' = \phi\, \bar{y}_R + (1-\phi)\, \bar{y}_u$$

with

$$\bar{y}_R = \frac{\sum\limits_{s_{2m}} y_i}{\sum\limits_{s_{2m}} x_i} \sum\limits_{s_1} x_i/n$$

and $\bar{y}_u$ is the same as in strategy F.

Avadhani and Sukhatme (1970) obtained the minimum variance of $\hat{\bar{Y}}'$ w.r.t. ϕ as

$$V_{opt}(\hat{\bar{Y}}'|\lambda) = \underset{\phi}{\text{Min}}\ V(\hat{\bar{Y}}'|\lambda) = \frac{V_1''\ V_2'}{V_1'' + V_2'} + V_{12}$$

$$V_1'' = \frac{1}{n} S_y^2 + (\frac{1}{m} - \frac{1}{n})\frac{1}{N-1} \sum_i (y_i - Rx_i)^2$$

V_2' and V_{12} are same as in strategy F.

Finally, they obtained the minimum variance of $\hat{\bar{Y}}'$ as

$$V_{min}(\hat{\bar{Y}}') = \underset{\lambda\ \phi}{\text{Min}}\ V(\hat{\bar{Y}}') = \underset{\lambda}{\text{Min}}\ V_{opt}(\hat{\bar{Y}}'|\lambda)$$

$$= [\frac{1 + \sqrt{1-\rho^2}}{2n} - \frac{1}{N}] S_y^2 \qquad (4.3.6)$$

For comparison of the strategies F and G Avadhani and Sukhatme (1970) considered the finite population model M_2 we discussed inchapter 2. Assuming the model M_2 they found that

$$V_{min}(\hat{\bar{Y}}) \underset{>}{\leqq} V_{min}(\hat{\bar{Y}}') \quad \text{according as} \quad g \underset{<}{\geqq} 1.$$

.4 A PROPOSED STRATEGY

TRATEGY I :

On the first occasion a sample S_1 of size n is osen following Midzuno's modified πps scheme with p_i's subject to

$$\frac{1}{n} > p_i > \frac{n-1}{n} \frac{1}{N-1} \quad \forall\ i = 1, \ldots, N \qquad (4.4.1)$$

as normed size measures and on the second occasion a sub-sample S_{2m} of m units is chosen from S_1 by SRSWOR method and an independent sample S_{2u} of size u is taken from U following Midzuno's modified πps scheme with earlier p_i's [assuming in addition to (4.4.1) that we have

$$\frac{1}{u} > p_i > \frac{u-1}{u} \cdot \frac{1}{N-1} \qquad \forall\ i = 1, \ldots, N$$

as well and observing, of course, that this **is implied by** (4.4.1) whenever $u < n$.]

Our proposed unbiased estimator for Y is

$$\hat{Y}^o = \phi\ \hat{Y}^o_{2m} + (1-\phi)\ \hat{Y}^o_{2u} \qquad (4.4.2)$$

with

$$\hat{Y}^o_{2m} = \frac{n}{m} \sum_{S_{2m}} \frac{y_i - x_i}{\pi_i} + \sum_{S_1} \frac{x_i}{\pi_i},$$

$$\hat{Y}^o_{2u} = \sum_{S_{2u}} \frac{y_i}{\pi_i'}$$

where $\pi_i = np_i$ and $\pi_i' = up_i$ for $i=1, \ldots, N$.

Denoting by E_1 the operator for taking expectation with espect to the sampling design for sampling on first occasion and E_2 the operator for taking conditional expectaion with respect to the sampling design adopted on the 2nd occasion when the results of sampling on the first occasion are held fixed and by v_1 and v_2 the corresponding variance nd conditional variance-operators, we have the expectations nd variances as

$$E\,\hat{Y}^{o}_{2m} = E\,\hat{Y}^{o}_{2u} = Y,$$

$$\text{Var}\,(\hat{Y}^{o}_{2m}) = E_1\,v_2\,(\hat{Y}^{o}_{2m}) + v_1\,E_2\,(\hat{Y}^{o}_{2m}).$$

Now,

$$v_1\,E_2\,(\hat{Y}^{o}_{2m}) = v_1\,\Big(\sum_{S_1} \frac{y_i}{\pi_i}\Big) = V_{mid}(n, y) \quad \text{(say)}.$$

om Chaudhuri (1974) we note that

$$V_{mid}(n, y) = \frac{N-2n}{n(N-2)} \sum_i p_i \Big(\frac{y_i}{p_i} - Y\Big)^2$$

$$- \frac{n-1}{n(N-1)(N-2)} \sum_{i \neq j}\sum \Big(\frac{y_i}{p_i} - Y\Big)\Big(\frac{y_j}{p_j} - Y\Big)$$

$$\leq \frac{N-n}{n(N-2)} \sum_i p_i \Big(\frac{y_i}{p_i} - Y\Big)^2$$

$$= \frac{N-n}{n(N-2)}\, V_{pps}(y) \qquad (4.4.3)$$

$$E_1 V_2(\hat{Y}^o_{2m}) = n^2 E_1 \left[\left(\frac{1}{m} - \frac{1}{n}\right) \frac{1}{n-1} \left\{ \sum_{S_1} \left(\frac{y_i - x_i}{\pi_i}\right)^2 - \frac{\left(\sum_{S_1} \frac{y_i - x_i}{\pi_i}\right)^2}{n} \right\} \right]$$

$$E_1 V_2(\hat{Y}^o_{2m}) = n^2 \left(\frac{1}{m} - \frac{1}{n}\right) \frac{1}{n-1} \left[\sum_i \frac{z_i^2}{\pi_i} - \frac{V_{mid}(Z,n) + Z^2}{n} \right]$$

$$= n^2 \left(\frac{1}{m} - \frac{1}{n}\right) \frac{1}{n-1} \left[\frac{1}{n} \left(\sum_i \frac{z_i^2}{p_i} - Z^2 \right) - \frac{V_{mid}(Z,n)}{n} \right]$$

$$= n \left(\frac{1}{m} - \frac{1}{n}\right) \frac{1}{n-1} \left[V_{pps}(Z) - V_{pps}(Z,n) \right]$$

$$= \left(\frac{1}{m} - \frac{1}{n}\right) V_{pps}(Z) + \left(\frac{1}{m} - \frac{1}{n}\right) \frac{n}{n-1} \left[\frac{V_{pps}(Z)}{n} - V_{mid}(Z,n) \right] \quad (4.4.4)$$

Thus

$$V(\hat{Y}^o_{2m}) = V_{mid}(n,y) + \left(\frac{1}{m} - \frac{1}{n}\right) V_{pps}(z) + \left(\frac{1}{m} - \frac{1}{n}\right) \frac{n}{n-1} \left[\frac{V_{pps}(z)}{n} - V_{mid}(z,n) \right]$$

$$= f_n(y)\, V_{pps}(y) + \left(\frac{1}{m} - \frac{1}{n}\right) V_{pps}(z) + \left(\frac{1}{m} - \frac{1}{n}\right) \frac{n}{n-1} \left[\frac{V_{pps}(z)}{n} - f_n(z)\, V_{pps}(z) \right]$$

where we write

$$f_n(t)\, V_{pps}(t) = V_{mid}(n,t)$$

for $t = y, z$ and note that

$$f_v(t) \leq \frac{N-v}{v(n-2)} \quad \text{where } v = n \text{ or } u \quad (4.4.5)$$

Now, assuming (4.3.1) we have

$$V(\hat{Y}^o_{2m}) = f_n(y)\ V_o + 2\left(\frac{1}{m} - \frac{1}{n}\right) V_o (1-\rho)$$

$$+ 2\left(\frac{1}{m} - \frac{1}{n}\right) \frac{n}{n-1} \left\{\frac{1}{n} - f_n(z)\right\} V_o (1-\rho) .$$

Also, we have

$$V(\hat{Y}^o_{2u}) = V\left(\sum_{S_{2u}} \frac{y_i}{\pi_i}\right) = V_{mid}(u, y) \qquad \text{(say)}$$

Then the optimum value of $V(\hat{Y}^o)$ for a given λ will be

$$V_{opt}(\hat{Y}^o|\lambda) = \frac{V(\hat{Y}^o_{2m})\ V(\hat{Y}^o_{2u})}{V(\hat{Y}^o_{2m}) + V(\hat{Y}^o_{2u})} .$$

4.5 COMPARISON OF RELATIVE EFFICIENCIES OF THE PROPOSED AND OTHER STRATEGIES

A. COMPARISON WITHOUT ASSUMING A SUPER-POPULATION MODEL

If we assume that

$$f_n(y) = f_n(z) = f_n \quad \text{(say)} \qquad (4.5.1)$$

then we can derive a number of results concerning the efficiency of strategy I (relative to strategies A, B, C) which are given below. If it so happens that for the chosen normed size-measures we have

$$p_i = \frac{2y_i - x_i}{2Y - X} \quad \forall\, i = 1, \ldots, N \qquad (4.5.2)$$

then obviously (4.5.1) hold exactly, instead, if $y_i \propto x_i$, then also (4.5.1) holds exactly. Even in other cases as well (4.5.1) may hold exactly or approximately.

In the Appendix (4.1) given at the end of this chapter we have given an example from live data where, in fact (4.5.1) holds approximately with remarkable closeness.

I. COMPARISON WITH STRATEGY A DUE TO RAJ (1965)

Under the assumptions (4.3.1) and (4.5.1) and writing

$\Lambda_{n,m}(\rho) = \frac{2(n-m)}{m(n-1)}(1-\rho)$, we get

$$\Delta = V(\hat{Y}_{2m}) - V(\hat{Y}^{o}_{2m})$$

$$= (\frac{1}{n} - f_n)\,[1 - \Lambda_{n,m}(\rho)]\,V_o \quad .$$

Now, remembering (4.4.5) we have

$$f_n \leq \frac{N-n}{n(N-2)} \leq \frac{1}{n} \qquad \text{for } n \geq 2 \qquad (4.5.3)$$

and as such $\Delta \geq 0$ for $n \geq 2$ if $\Lambda_{n,m}(\rho) \leq 1$ (4.5.4)

But $\Lambda_{n,m}(\rho)$ is a monotonically decreasing function of ρ and moreover

$$\Lambda_{n,m}(\rho) < 1 \left\{ \begin{array}{l} \text{for } \;= \frac{1}{2} \quad \forall\; n,m \\ \text{and also for } \rho = -1, \\ \text{provided } \; n > m \geq 4 \end{array} \right. \qquad (4.5.5)$$

Again, by Chaudhuri's (1974) results we have

$V(\hat{Y}^{o}_{2u}) < V(\hat{Y}_{2u})$ and hence the

Theorem 4,1

$$V_{opt}(\hat{Y}_2^o|\lambda) < V_{opt}(\hat{Y}_2|\lambda)$$

if either (4.3.1), (4.4.1), (4.5.1) hold jointly along with the condition $n > m \geq 4$ or if (4.3.1), (4.4.1), (4.5.1) hold jointly and in addition $\rho \geq \frac{1}{2}$.

II. COMPARISON WITH STRATEGY B DUE TO GHANGURDE AND RAO (1969)

Assuming (4.3.1), (4.3.3) and (4.5.1) to hold simultaneously, one may check that

$$B = V(\hat{Y}'_{2m}) - V(\hat{Y}^o_{2m})$$

$$= [\{\frac{N-n}{n(N-1)} - f_n\} + 2(\frac{1}{m} - \frac{1}{n})n(1-\rho)\{\frac{N-n}{n(N-1)} - \frac{1-f_n}{n-1}\}]V_o$$

$$+ 2(\frac{1}{m} - \frac{1}{n})\frac{n}{N-1}(1-\delta)\bar{V}. \qquad (4.5.6)$$

Now B is a monotonically decreasing function of f_n for $m \geq 4$ whatever ρ may be. Remembering that $f_n \leq \frac{N-n}{n(N-2)}$ and seeking to check if $B \geq 0$, let us consider the worst situation, namely, when

$$f_n = \frac{N-n}{n(N-2)} .$$

Let us assume N so large that we may neglect the error in writing $\frac{1}{N-1}$ for $\frac{1}{N-2}$ (the assumption Δ) and write

$$f_n = \frac{N-n}{n(N-1)} \qquad (4.5.7)$$

Using (4.5.7), the RHS expression in (4.5.6) is

$$(\frac{1}{m} - \frac{1}{n}) \frac{n}{N-1} [N \sum_i (z_i - \bar{Z})^2 - \sum_i p_i (\frac{z_i}{p_i} - Z)^2] \qquad (4.5.8)$$

If $z_i \alpha p_i$, then (4.5.8) is positive. From this one may expect that if z_i's be approximately proportional to p_i's even than (4.5.8) should be positive. So if y_i's and x_i's be approximately proportional to p_i's which we assume to be the case in practice, then (4.5.8) is expected to be positive. With the assumption A_4 it follows also that

$$V(\hat{Y}'_{2u}) > V(\hat{Y}^o_{2u})$$

So we have

Theorem 4.2

If (4.3.1), (4.3.3) and (4.5.1) hold together and y_i's and x_i's be (at least approximately) proportional to p_i's and if one may neglect the error in writing $\frac{1}{N-1}$ for $\frac{1}{N-2}$ for large N, then

$$V_{opt}(\hat{Y}_2' | \lambda) > V_{opt}(\hat{Y}^o_2 | \lambda) .$$

III. COMPARISON WITH STRATEGY C DUE TO CHOTAI (1974)

Assuming (4.5.1), we get, on simplification,

$$V(\hat{Y}''_{2m}) - V(\hat{Y}^{o}_{2m}) = \{ 1 - 2 \cdot \frac{n-m}{m(n-1)} (1-\rho) \}\{ \frac{N-n}{n(N-1)} - f_n \} V_o$$

> 0, for $m \geq 4$ provided the assumption Δ holds.

By Δ it follows that $V(\hat{Y}^{*}_{2u}) < V(\hat{Y}^{o}_{2u})$.

Hence follows the

<u>Theorem 4.3</u>

If (4.5.1) holds and Δ is assumed, then

$V_{opt}(\hat{Y}^{*}_{2}|\lambda) \geq V_{opt}(\hat{Y}^{o}_{2}|\lambda)$, provided $m \geq 4$.

B. COMPARISON UNDER A SUPER-POPULATION MODEL

Following Cochran (1963) and Hanurav (1967) we shall now assume the model M_1 for y and x (vide model M_1 in chapter 2, p. 19) under which U is a random sample from an infinite super-population so that y and x are stochastic variables having certain probability distributions. We shall denote by ε and v the operators for conditional expectations and variances in respect of these distributions for given values of the size-measures w_i's (where $p_i = \frac{w_i}{W}$, $W = \sum_i w_i$).

For the model M_1 we assume that

$$\varepsilon(x_i|w_i) = a\, w_i, \qquad \varepsilon(y_i|w_i) = b\, w_i$$

$$v(x_i|w_i) = v(y_i|w_i) = \sigma^2\, w_i^{\,g}$$

where

$$a,\ b,\ \sigma^2 > 0 \quad \text{and} \quad 0 \leq g \leq 2,$$

then

$$\varepsilon(z_i|w_i) = C\, w_i \qquad \text{where } C = a-b,$$

$$v(z_i|w_i) = 2\sigma^2(1-\tau)\, w_i^{\,g} = k\, \sigma^2\, w_i^{\,g} \quad \text{(say)}$$

on writing

$$K = 2(1-\tau)$$

where

$$\tau = \frac{\varepsilon\{y_i - \varepsilon(y_i|w_i)\}\{x_i - \varepsilon(x_i|w_i)\}}{V(y_i|w_i)}$$

i) STRATEGY I Vs STRATEGY 'C' UNDER M_1

On simplification,

$$V(\hat{Y}''_{2m}) - V(\hat{Y}^{o}_{2m}) = \{V_{RHC}(y,n) - V_{mid}(y,n)\}$$

$$- \left(\frac{1}{m} - \frac{1}{n}\right) \frac{n}{n-1} \{V_{RHC}(z,n) - V_{mid}(z,n)\}$$

where $V_{RHC}(t,r)$ = variance of RHC estimator for T based on the sample of size r, with $T = \sum_i t_i$.

Then, using Hanurav's (1967) results we get

$$\varepsilon\, [V(\hat{Y}''_{2m}) - V(\hat{Y}^{o}_{2m})]$$

$$= \sigma^2 \left[1 - \left(\frac{1}{m} - \frac{1}{n}\right) \frac{n}{n-1} K\right] \frac{n-1}{n(N-1)} \left[N \sum_i x_i^{\,g} - X \sum_i x_i^{\,g-1}\right]$$

Now, $\{1 - (\frac{1}{m} - \frac{1}{n}) \frac{n}{n-1} K\} > 0$ for every τ whenever $m \geq 4$.

So,

$$\varepsilon [V(\hat{Y}''_{2u}) - V(\hat{Y}^o_{2u})] \begin{cases} \geq 0 & \text{if } g \geq 1 \\ < 0 & \text{if } g < 1 \end{cases}$$

and likewise

$$\varepsilon [V(\hat{Y}'''_{2u}) - V(\hat{Y}^o_{2u})] \begin{cases} \geq 0 & \text{if } g \geq 1 \\ < 0 & \text{if } g < 1 \end{cases}$$

So, we have

Theorem 4.4

Under model M_1

$$\varepsilon [V_{opt}(\hat{Y}^*_2|\lambda) - V_{opt}(\hat{Y}^o_2|\lambda)] \begin{cases} \geq 0 & \text{if } g \geq 1 \\ < 0 & \text{if } g < 1 \end{cases}$$

provided $m \geq 4$.

Remark I

In practice, the situations with $g \geq 1$ are more frequent than those with $g < 1$ and so under M_1 our proposed strategy I should often fare better than the strategy C.

Remark II

The results concerning the relative efficiencies, under Model M_1, apply to situations where S_1 and S_{2u} are lected according to any πps-scheme and not by Midzuno's odified) πps scheme alone, because as shown by Hanurav 1967) the expected value of V(HTE) with respect to the odel M_1 is same for every πps-scheme.

ii) STRATEGY I Vs STRATEGY 'B'

On simplifications, we get

$$V(\hat{Y}'_{2m}) = \left(\frac{1}{m} - \frac{1}{n}\right) \frac{1}{N-1} \left[(N-n)\ V_{pps}(z) + n\ \bar{V}\ (z)\right] + \frac{N-n}{n(N-1)}\ V_{pps}(y)$$

nd

$$V(\hat{Y}^{o}_{2m}) = V_{mid}(n, y) + \left(\frac{1}{m} - \frac{1}{n}\right) \frac{n}{n-1} \left[V_{pps}(z) - V_{mid}(z, n)\right].$$

Then

$$V(\hat{Y}'_{2m}) - V(\hat{Y}^{o}_{2m})$$

$$= \frac{N-n}{n(N-1)}\ V_{pps}(y) - V_{mid}(y, n) + \left(\frac{1}{m} - \frac{1}{n}\right) \left[\frac{n}{N-1} \left(\bar{V}(z) - V_{pps}(z)\right) + \frac{n}{n-1} \left\{V_{mid}(z, n) - \frac{N-n}{n(N-1)}\ V_{pps}(z)\right\}\right].$$

Now,

$$\varepsilon \frac{\bar{V}(z)}{N} = \frac{N-1}{N} \quad \sigma^2\ K \sum_i w_i^{\,g} + C^2 \sum_i (w_i - \bar{w})^2$$

$$\left(\text{with } \bar{w} = \frac{1}{N} \sum_i w_i\right).$$

$$\varepsilon\, V_{mid}(z,n) = \sigma^2 \; K \sum_i w_i^{g}\left(\frac{1}{np_i} - 1\right).$$

So,

$$\varepsilon\, [V(\hat{Y}'_{2m}) - V(\hat{Y}^{o}_{2m})]$$

$$= \sigma^2 \Big[\frac{n-1}{n(N-1)} \sum_i \left(N - \frac{1}{p_i}\right) w_i^{g}$$

$$+ \left(\frac{1}{m} - \frac{1}{n}\right) \frac{n-1}{N-1} \; K \sum_i w_i^{g} \left(N - \frac{1}{p_i}\right)\Big]$$

$$+ \left(\frac{1}{m} - \frac{1}{n}\right) \frac{n\; C^2}{N-1} \; N \sum_i (w_i - \bar{w})^2 .$$

So, we get

<u>Theorem 4.5</u>

If M_1 holds, then

$$\varepsilon[V_{opt}(\hat{Y}'|\lambda) - V_{opt}(\hat{Y}^{o}|\lambda)] \geq 0 \quad \text{if } g \geq 1$$

(iii) STRATEGY I Vs STRATEGY A.

$$\varepsilon[\, V(\hat{Y}_{2m}) - V(\hat{Y}^{o}_{2m})\,] = \sigma^2 \; \frac{n-1}{n} \sum_i w_i^{g} \left(1 - \frac{n-m}{m(n-1)} K\right)$$

$$\geq 0 \text{ whenever } m \geq 4$$

(Since $0 \leq K \leq 4$ as we have seen earlier)

So, we get

eorem 4.6

If M_1 holds, then

$$\varepsilon\ [V_{opt}(\hat{Y}|\lambda) - V_{opt}(\hat{Y}^o|\lambda)] \geq 0 \quad \text{for } m \geq 4.$$

ark III

If we consider an alternative strategy for which is chosen as in strategy C, S_{2u} as in strategy I and is chosen from S_1 by using Midzuno's (modified) πps-eme with inclusion probabilities

$$\pi_i^* = m\, p_i \ \forall\ i = 1, \ldots, N$$

d the estimator is

$$\hat{Y}^* = \phi\ \hat{Y}^*_{2m} + (1-\phi)\hat{Y}^o_{2u}$$

ere

$$\hat{Y}_{2m}^* = \sum_{S_{2m}} \frac{y_i - x_i}{p_i} \frac{p_i}{\pi_i^*} + \sum_{S_1} \frac{x_i}{p_i} p_i,$$

en, recalling Chaudhuri's (1974) results it follows, we neglect the error in writing

$$\frac{1}{n-1} \text{ for } \frac{1}{n-2}$$

he assumption B), that

$$V(\hat{Y}_{2m}^*) \leq V(\hat{Y}_{2m}'') .$$

Also, by λ $\quad V(\hat{Y}^{o}_{2u}) \leq V^{*}(\hat{Y}''_{2u})$.

Then, it follows, on assuming $\underset{\sim}{A}$ and $\underset{\sim}{B}$ that

$$V_{opt}(\hat{Y}^{*}|\lambda) \leq V_{opt}(\hat{Y}''|\lambda)$$

Remark IV

If we consider a strategy where S_{2u} and S_1 are chosen as in strategy B or C but S_{2m} is chosen from S_1 by using Midzuno's (modified) πps-sampling scheme then also an appropriate unbiased estimator is available with variance smaller than $V_{opt}(\hat{Y}''|\lambda)$ provided $\underset{\sim}{A}$ and $\underset{\sim}{B}$ hold.

4.6 MODIFICATION OF AVADHANI-SUKHATME'S STRATEGY (G)
STRATEGY II

On the first occasion an SRSWOR sample S_1 of size n is taken from U and on the second occasion a sub-sample S_{2m} of size m is chosen therefrom by following a suitable πps-design with $p_i' = x_i / \sum_{S_1} x_i$ (for $i \in S_1$) as normed size-measure for the ith unit and an unmatched sample S_{2u} of size u is chosen from U/S_1 by SRSWOR.

The proposed estimator for the population mean $\bar{Y}$ is given by

$$\hat{\hat{\bar{Y}}} = \phi \hat{\hat{\bar{Y}}}_{2m} + (1-\phi) \bar{y}_u$$

where
$$\hat{\hat{\bar{Y}}}_{2m} = \frac{1}{n} \sum_{s_{2m}} \frac{y_i}{\pi_i'}$$

$$\bar{y}_u = \frac{1}{u} \sum_{s_{2u}} y_i$$

and
$$\pi_i' = m\, p_i', \quad i = 1, \ldots, n.$$

The resulting strategy we call strategy II.

Now
$$V(\hat{\hat{\bar{Y}}}_{2m}) = E_1 V_2(\hat{\hat{\bar{Y}}}_{2m}) + V_1 E_2(\hat{\hat{\bar{Y}}}_{2m})$$

$$= \frac{1}{n^2} E_1 \left[\sum_{s_1} y_i^2 \left(\frac{1}{\pi_i'} - 1\right) + \sum_{s_1 (i \neq j)} \sum \left(\frac{\pi'_{ij}}{\pi_i' \pi_j'} - 1\right) y_i y_j \right]$$

$$+ \left(\frac{1}{n} - \frac{1}{N}\right) S_y^2 \qquad (4.6.1)$$

where π'_{ij} s are the inclusion-probabilities of pairs of units i and j ($i \neq j$) in sub-sampling from s_1 according to a πps design. The term inside the square bracket of (4.6.1) for different πps designs will be different and for many of the usual πps designs it is difficult to have an elegant expression for it but we shall note in what follows that it is still possible to study the performances of the strategy II relative to F and G due to Avadhani and ukhatme (1970). Later still we shall supply an elegant xpression for (i) the expectation of $V(\hat{\hat{\bar{Y}}}_{2m})$ with respect o a super population model M_1 (vide Chapter 2, p 19) nd (ii) for an unbiased estimator for $V(\hat{\hat{\bar{Y}}}_{2m})$.

Now variance of $\bar{y}_u$ is

$$V(\bar{y}_u) = (\frac{1}{u} - \frac{1}{N}) S_y^2$$

nd

$$Cov(\bar{y}_u, \hat{\hat{\bar{Y}}}_{2m}) = - \frac{S_y^2}{N} .$$

o, the minimum value of $Var(\hat{\hat{\bar{Y}}})$ over variation of ϕ for iven λ ($= \frac{m}{n}$) is given by

$$V_{opt}(\hat{\hat{\bar{Y}}}|\lambda) = \frac{V_1' V_2'}{V_1' + V_2'} + V_{12}$$

here

$$V_1^{**} = V(\hat{\hat{\bar{Y}}}_{2m}) - V_{12}$$

$$V_2' = V(\bar{y}_u) - V_{12}$$

$$V_{12} = Cov(\bar{y}_u, \hat{\hat{\bar{Y}}}_{2m}) = - \frac{S_y^2}{N} .$$

RATEGY III

In the strategy III we are going to describe low S_1 and S_{2u} are selected in the same manner as in rategy II but the matched sample S_{2m} of size $m = n\lambda$

(an integer with $0 \leq \lambda \leq 1$) is selected from S_1 by Midzuno-Sen (1952, 53) scheme of sampling using the normed size measures p_i' 's $(i = 1, \ldots, n)$.

Here the proposed estimator for the population mean $\bar{Y}$ is given by

$$\hat{\hat{\bar{Y}}}' = \phi \, \hat{\hat{\bar{Y}}}'_{2m} + (1-\phi) \, \bar{y}_u$$

where

$$\hat{\bar{Y}}'_{2m} = \frac{\sum_{S_{2m}} y_i}{\sum_{S_{2m}} x_i} \sum_{S_1} x_i/n$$

Finally, we get

$$V(\hat{\hat{\bar{Y}}}'_{2m}) = E_1 V_2 \, (\hat{\hat{\bar{Y}}}'_{2m}) + (\frac{1}{n} - \frac{1}{N}) \, S_y^2 \qquad (4.6.2)$$

An explicit expression for $v_2(\hat{\hat{\bar{Y}}}'_{2m})$ is readily available e.g. from Rao T.J. (1967) and it is of the form V_3 in Chapter 2, p. 24).

Now, Cov $(\hat{\hat{\bar{Y}}}'_{2m}, \bar{y}_u) = - \frac{S_y^2}{N}$.

So the optimum value for the variance of $\hat{\hat{\bar{Y}}}'$ for given λ is given by

$$V_{opt}(\hat{\hat{\bar{Y}}}' | \lambda) = \frac{V_1^{*}{}'' \, V_2'}{V_1'' + V_2'} + V_{12}$$

where

$$V_1^{*}{}'' = V(\hat{\hat{\bar{Y}}}'_{2m}) - V_{12},$$

V_2' and V_{12} are the same as in strategy II.

COMPARISON BETWEEN THE RELATIVE EFFICIENCIES OF THE STRATEGIES II, III, F AND G

The suggested estimators for the population mean $\bar{Y}$ for the above four strategies are of the form

$$t = \phi e + (1-\phi) e'$$

where e is an estimator for population mean $\bar{Y}$ based on the matched sample and e' is an estimator for $\bar{Y}$ based on the un-matched sample. Here for all the four strategies e' is the same and equal to $\bar{y}_u$. The covariance between e and e' is also the same for the above four strategies. So, the relative efficiencies of the above four strategies are determined by the relative magnitudes of $V(\hat{\hat{Y}}_{2m})$, $V(\hat{Y}'_{2m})$, $V(\bar{Y}_{RHC})$ and $V(\bar{Y}_R)$ only. Again, the initial sample s_1 being chosen following the same design for each of the four strategies their relative efficiencies are determined by the relative magnitudes of the conditional variances $v_2(\hat{Y}_{2m})$, $v_2(\hat{\hat{Y}}'_{2m})$, $v_2(\bar{y}_{RHC})$ and $v_2(\bar{y}_R)$, for any given sample s_1, chosen on the first occasion. Bearing these points in mind we state below the results concerning the relative efficiencies of the strategies stated above.

Now if S_{2m} is selected according to a modified πps sampling scheme, with normed size-measures p_i' attached to the ith unit [vide Chaudhuri (1974), Mukhopadhyaya (1972), Sankaranarayanan (1969)] we have from (4.4.3)

$$v_2(\hat{\bar{Y}}_{2m}) \leq \frac{n-m}{n^2 m(n-2)} \sum_{i \varepsilon S_1} p_i' \left(\frac{y_i}{p_i'} - \sum_{i \varepsilon S_1} y_i\right)^2 \qquad (4.6.3)$$

If we assume that n is so large that we may neglect the error in writing $\frac{1}{n-2}$ for $\frac{1}{n-1}$, then using the relation (4.6.3) we may assert that

$$v_2(\hat{\bar{Y}}_{2m}) < v_2(\bar{y}_{RHC})$$

and we get the following

<u>Theorem 4.7</u>

If S_{2m} of strategy II is selected by Midzuno's modified πps sampling scheme with normed size measures p_i's, then the strategy II is superior to F provided we ignore the error in writing $\frac{1}{n-2}$ by $\frac{1}{n-1}$.

If we assume (which we shall call the assumption <u>c</u>) the model M_2 (vide chapter 2, p. 24) not only for the entire population U but for every sample S_1 of size n taken from U on the first occasion by SRSWOR method also, then we get the following results

Theorem 4.8

If the assumption c holds and neglect the error in writing $\bar{x}_{S_{2m}} = \frac{1}{m} \sum_{S_{2m}} x_i$ for $\dot{x}_1 = \frac{1}{n} \sum_{S_1} x_i$ for every sub-sample S_{2m} of S_1 for every S_1, then we have recalling the results due to Avadhani and Srivastava (1972) and Chaudhuri (1977) that

(i) $V_{opt}(\hat{\hat{\hat{Y}}}'|\lambda) = V_{opt}(\hat{\hat{Y}}'|\lambda)$

(ii) $V_{opt}(\hat{\hat{\hat{Y}}}'|\lambda) = V_{opt}(\hat{\hat{Y}}'|\lambda) \geq V_{opt}(\hat{\hat{Y}})$ for $g \geq 1$

and $V_{opt}(\hat{\hat{Y}}'|\lambda) < V_{opt}(\hat{\hat{Y}})$ for $0 \leq g < 1$

Theorem 4.9

If theassumption c holds and if we ignore the error in writing 1/n for $\frac{1}{n-1}$, then assuming the asymptotic relationship considered by Asok and Sukhatme (1977) and using the results (vide theorem 2.4) we have under Sampford's πps sampling scheme

$$V_{opt}(\hat{\hat{\hat{Y}}}) \lesseqqgtr V_{opt}(\hat{\hat{Y}}) \quad \text{according as} \quad g \gtreqqless 1.$$

Theorem 4.10

If the assumption $\underline{c}$ holds, then we have [from theorem 2.5] $V_{opt}(\hat{\hat{Y}}) < V_{opt}(\hat{Y})$ if $g \geq 2$, provided S_{2m} is selected following (modified) Midzuno's πps sampling scheme.

As opposed to the model M_2 we may use model M_1 (vide chapter 2, p.19) to compare the relative efficiencies of the strategies F, G, II and III by comparing the minimum (with respect to ϕ, the value of λ remaining fixed) expected variances which we write as $\varepsilon_{opt} V(\cdot|\lambda)$. Thus we get the following theorems (for which the easily available proofs we omit)

Theorem 4.11

If the model M_1 holds, then it follows that

$$\varepsilon_{opt} V(\hat{\hat{\hat{Y}}}|\lambda) \lesseqgtr \varepsilon_{opt} V(\hat{\bar{Y}}|\lambda) \lesseqgtr \varepsilon_{opt} V(\hat{\hat{\bar{Y}}}'|\lambda)$$

according as $g \gtreqless 1$.

Theorem 4.12

If model M_1 holds, then,

$$\varepsilon_{opt} V(\hat{\hat{Y}}') > \varepsilon_{opt} V(\hat{\hat{\hat{Y}}}) \quad \text{if } g \geq 1$$

[vide theorem 4 in Chaudhuri (1977)]

Theorem 4.13

If M_1 holds, and if we are justified to neglect the error in writing $\bar{x}_{S_{2m}}$ for $\bar{x}_{s1}$ for every $S_{2m} \subset S_1$ for each S_1 then

$$\varepsilon_{opt} V(\hat{\hat{Y}}' | \lambda) \gtreqless \varepsilon_{opt} V(\hat{\hat{Y}} | \lambda) \quad \text{according as } g \gtreqless 1$$

[recalling theorem 5, Chaudhuri (1977)]

A FEW REMARKS

Remark I

Writing $V_{opt}(t|\lambda)$, $\varepsilon_{opt} V(t|\lambda)$ for the minimum [with respect to ϕ values of $V(t)$ and $\varepsilon V(t)$] for a given λ and $V_{min}(t)$, $\varepsilon_{min} V(t)$ their minimum values with respect to variations in λ it readily follows that for any alternative estimator t' of the form t we have

$$V_{opt}(t|\lambda) \lesseqgtr V_{opt}(t'|\lambda) \Rightarrow V_{min}(t) \lesseqgtr V_{min}(t')$$

and

$$\varepsilon_{opt} V(t|\lambda) = \varepsilon_{opt} V(t'|\lambda) \Rightarrow \varepsilon_{min} V(t) = \varepsilon_{min} V(t').$$

So, the results relating to the comparative efficiencies of the strategies II, III, F and G given earlier have the obvious extended interpretations.

Remark II

If M_1 holds, then for any πps design we have

$$\varepsilon V(\hat{\hat{\hat{Y}}}_{2m}) = \varepsilon[E_1 v_2 (\hat{\hat{\hat{Y}}}_{2m}) + v_1 E_2 (\hat{\hat{\hat{Y}}}_{2m})]$$

$$= \varepsilon [E_1 \{\sum_{S_1} y_i^2 (\frac{1}{\pi_i'} - 1) + \sum_{S_1(i \neq j)}\sum y_i y_j (\frac{\pi_{ij}'}{\pi_i' \pi_j'} - 1)\}$$

$$+ (\frac{1}{n} - \frac{1}{N}) s_y^2]$$

$$= E_1 [\varepsilon \sum_{S_1} y_i^2 (\frac{1}{\pi_i'} - 1) + \varepsilon \sum_{S_1(i \neq j)}\sum y_i y_j (\frac{\pi_{ij}'}{\pi_i' \pi_j'} - 1)]$$

$$+ (\frac{1}{n} - \frac{1}{N}) \varepsilon (s_y^2)$$

$$= \frac{\sigma^2}{n^2} E_1 \sum_{S_1} x_i^g (\frac{1}{\pi_i'} - 1) + (\frac{1}{n} - \frac{1}{N}) (\frac{\sigma^2}{N} \sum_i x_i^g + \beta^2 S_x^2$$

[where $(N-1) S_x^2 = \sum_i (x_i - \bar{X})^2$ and $\pi_i' = m \frac{x_i}{\sum_{S_1} x_i}$]

$$= \frac{\sigma^2}{Nn} [\frac{1}{m} \{\sum_i x_i^g + \frac{n-1}{N-1} \sum_{i \neq j}\sum x_i^{g-1} x_j\} - \sum_i x_i^g]$$

$$+ (\frac{1}{n} - \frac{1}{N}) \{ \frac{\sigma^2}{N} \sum_i x_i^g + \beta^2 S_x^2 \}$$

Remark III

An unbiased estimator for $V(\hat{\hat{\hat{Y}}}_{2m})$ is given by

$$\hat{V}(\hat{\hat{\hat{Y}}}_{2m}) = [1 - (\frac{1}{n} - \frac{1}{N}) \frac{N}{N-1}]^{-1} [\frac{1}{n^2} \sum_{S_1(i<j)}\sum \frac{\pi_i'\pi_j' - \pi_{ij}'}{\pi_{ij}'}$$

$$(\frac{y_i}{\pi_i'} - \frac{y_j}{\pi_j'})^2 + (\frac{1}{n} - \frac{1}{N}) \frac{N}{N-1} \{ \frac{1}{n} \sum_{S_{2m}} \frac{y_i^2}{\pi_i'}$$

$$- (\sum_{S_{2m}} \frac{y_i}{n \pi_i'})^2 \}]$$

Proof.

$$E\hat{V}(\hat{\hat{Y}}_{2m}) = [1 - \frac{N}{N-1}(\frac{1}{n} - \frac{1}{N})]^{-1} [\frac{1}{n^2} E_1 \sum_{S_1} \sum_{(i<j)} (\pi_i'\pi_j' - \pi_{ij}') (\frac{y_i}{\pi_i'} - \frac{y_j}{\pi_j'})^2 + (\frac{1}{n} - \frac{1}{N}) \frac{N}{N-1} \{\frac{1}{N} \Sigma y_i^2 - V(\sum_{S_{2m}} \frac{y_i}{n\pi_i'}) - \bar{Y}^2\}]$$

$$= [1 - (\frac{1}{n} - \frac{1}{N}) \frac{N}{N-1}]^{-1} [\frac{1}{n^2} E_1 \sum_{S_1} \sum_{(i<j)} (\pi_i'\pi_j' - \pi_{ij}') (\frac{y_i}{\pi_i'} - \frac{y_j}{\pi_j'})^2 + (\frac{1}{n} - \frac{1}{N}) \frac{N}{N-1} \frac{1}{N} \Sigma (y_i - \bar{Y})^2 - (\frac{1}{n} - \frac{1}{N}) \frac{N}{N-1} V(\sum_{S_{2m}} \frac{y_i}{n\pi_i'})]$$

$$= [1 - (\frac{1}{n} - \frac{1}{N}) \frac{N}{N-1}]^{-1} [\frac{1}{n^2} E_1 \sum_{S_1} \sum_{(i<j)} (\pi_i'\pi_j' - \pi_{ij}') (\frac{y_i}{\pi_i'} - \frac{y_j}{\pi_j'})^2 + (\frac{1}{n} - \frac{1}{N}) S_y^2 - (\frac{1}{n} - \frac{1}{N}) \frac{N}{N-1} V(\sum_{S_{2m}} \frac{y_i}{n\pi_i'})]$$

$$= V(\hat{\hat{Y}}_{2m}) .$$

MODIFICATION OF A FEW WELL-KNOWN STRATEGIES FOR SAMPLING ON TWO OCCASIONS TO IMPROVE EFFICIENCIES

As we have seen [vide section (4.3)] the sampling ategies due to Raj (1965) and Chotai (1974) are useful whenever the correlation coefficient-like quantity ρ defined in (4.3.2) for the variables measured on the same units on two occasions exceeds 1/2 and provided, incase of the latter one, in addition, the population-size exceeds twice the size of the sample chosen on the first occasion. But, for Ghangurde and Rao's (1969) strategy we have no knowledge about simple parallel results. In this section we propose alternatives to each of the three above-mentioned strategies by replacing the respective estimators by modified ones and show them to possess uniformly smaller variance than those for the respective estimators considered in the original strategies.

AN ALTERNATIVE TO RAJ'S (1965) ESTIMATOR BASED ON RAJ'S (1965) SAMPLING SCHEME

Let us consider

$$\hat{Y}^* = a_1 \frac{1}{m} \sum_{S_{2m}} \frac{y_i}{p_i} + b_1 \frac{1}{m} \sum_{S_{2m}} \frac{x_i}{p_i} + \frac{c_1}{n} \sum_{r=1}^{n} \frac{x_r}{p_r} + \frac{d_1}{u} \sum_{r=1}^{u} \frac{y_r}{p_r} \qquad (4.7.1)$$

[$\sum_{S_{2m}}$ denotes summation over the units in S_{2m} and $\sum_{r=1}^{n}$ denotes summation over n terms corresponding to n draws] here a_1, b_1, c_1 and d_1 are weights to be determined so s to make $\hat{Y}^*$ unbiased with the minimum variance.

Now in order that $\hat{Y}^*$ may be unbiased for Y we ust have

$$E(\hat{Y}^*) = (a_1 + d_1) Y + (b_1 + c_1) X = Y$$

which implies $a_1 + d_1 = 1$ and $b_1 + c_1 = 0$.

, we write

$$\hat{Y}^* = a_1 \left(\frac{1}{m} \sum_{S_{2m}} \frac{y_i}{p_i}\right) + (1-a_1)\left(\frac{1}{u} \sum_{r=1}^{u} \frac{y_r}{p_r}\right) + b_1\left(\frac{1}{m} \sum_{S_{2m}} \frac{x_i}{p_i} - \frac{1}{n} \sum_{r=1}^{n} \frac{x_r}{p_r}\right) \qquad (4.7.2)$$

en,

$$V(\hat{Y}^*) = a_1^2 V\left(\frac{1}{m} \sum_{S_{2m}} \frac{y_i}{p_i}\right) + (1-a_1)^2 V\left(\frac{1}{u} \sum_{r=1}^{n} \frac{y_r}{p_r}\right) + b_1^2 V\left(\frac{1}{m} \sum_{S_{2m}} \frac{x_i}{p_i} - \frac{1}{n} \sum_{r=1}^{n} \frac{x_r}{p_r}\right) + 2a_1 b_1 \operatorname{Cov}\left\{\frac{1}{m} \sum_{S_{2m}} \frac{y_i}{p_i}, \frac{1}{m} \sum_{S_{2m}} \frac{x_i}{p_i} - \frac{1}{n} \sum_{r=1}^{n} \frac{x_r}{p_r}\right\}$$

Since

$$\operatorname{Cov}\left\{\frac{1}{u} \sum_{r=1}^{n} \frac{y_r}{p_r}, \frac{1}{m} \sum_{S_{2m}} \frac{y_i}{p_i}\right\} = 0 =$$

$$\operatorname{Cov}\left\{\frac{1}{u} \sum_{r=1}^{n} \frac{y_r}{p_r}, \frac{1}{m} \sum_{S_{2m}} \frac{x_i}{p_i} - \frac{1}{n} \sum_{r=1}^{n} \frac{x_r}{p_r}\right\}]$$

$$= a_1^2 V_1 + (1-a_1)^2 V_2 + b_1^2 V_3 + 2a_1 b_1 V_{13} \qquad (4.7.3)$$

$$V_1 = V\left(\frac{1}{m} \sum_{s_{2m}} \frac{y_i}{p_i}\right)$$

$$= E_1\, V_2\left(\frac{1}{m} \sum_{s_{2m}} \frac{y_i}{p_i}\right) + V_1\, E_2\left(\frac{1}{m} \sum_{s_{2m}} \frac{y_i}{p_i}\right)$$

$$= E_1\left[\left(\frac{1}{m} - \frac{1}{n}\right) \frac{1}{n-1} \left\{ \sum_{r=1}^{n} \frac{y_r^2}{p_r^2} - \frac{\left(\sum_{r=1}^{n} \frac{y_r}{p_r}\right)^2}{n} \right\}\right]$$

$$+ V_1\left(\frac{1}{n} \sum_{r=1}^{n} \frac{y_r}{p_r}\right)$$

$$= \left(\frac{1}{m} - \frac{1}{n}\right) \frac{1}{n-1} \left\{ n \sum_{i} \frac{y_i^2}{p_i} - nY^2 - n\, V_1\left(\frac{1}{n} \sum_{r=1}^{n} \frac{y_r}{p_r}\right)\right\}$$

$$+ V_1\left(\frac{1}{n} \sum_{r=1}^{n} \frac{y_r}{p_r}\right)$$

$$= \frac{1}{m} \sum_{i} p_i \left(\frac{y_i}{p_i} - Y\right)^2$$

$$= \frac{1}{n\lambda} V_{pps}(y) \qquad (\text{since } m = n\lambda)$$

$$V_2 = V\left(\frac{1}{u} \sum_{r=1}^{n} \frac{y_r}{p_r}\right)$$

$$= \frac{V_{pps}(y)}{n\mu} \qquad (\text{since } u = n\mu)$$

$$V_3 = \left(\frac{1}{m} \sum_{s_{2m}} \frac{x_i}{p_i} - \frac{1}{n} \sum_{r=1}^{n} \frac{x_r}{p_r}\right)$$

$$= E_1 \, V_2 \left(\frac{1}{m} \sum_{S_{2m}} \frac{x_i}{p_i} - \frac{1}{n} \sum_{r=1}^{n} \frac{x_r}{p_r} \right)$$

$$\left[\text{since } E_2\left(\frac{1}{m} \sum_{S_{2m}} \frac{x_i}{p_i} - \frac{1}{n} \sum_{r=1}^{n} \frac{x_r}{p_r} \right) = 0 \right]$$

$$= \left(\frac{1}{m} - \frac{1}{n} \right) \frac{1}{n-1} \, E_1 \left\{ \sum_{r=1}^{n} \frac{x_r^{\,2}}{p_r^{\,2}} - \frac{\left(\sum_{r=1}^{n} \frac{x_r}{p_r} \right)^2}{n} \right\}$$

$$= \left(\frac{1}{m} - \frac{1}{n} \right) V_{pps}(x) .$$

$$= \frac{\mu}{n\lambda} \, V_{pps}(x) .$$

$$V_{13} = \mathrm{Cov} \left(\frac{1}{m} \sum_{S_{2m}} \frac{y_i}{p_i} , \frac{1}{m} \sum_{S_{2m}} \frac{x_i}{p_i} - \frac{1}{n} \sum_{r=1}^{n} \frac{x_r}{p_r} \right)$$

$$= E_1 \, \mathrm{Cov}_2 \left(\frac{1}{m} \sum_{S_{2m}} \frac{y_i}{p_i} , \frac{1}{m} \sum_{S_{2m}} \frac{x_i}{p_i} \right)$$

$$= \frac{1}{2} E_1 \left[V_2 \left(\frac{1}{m} \sum_{S_{2m}} \frac{y_i}{p_i} \right) + V_2 \left(\frac{1}{m} \sum_{S_{2m}} \frac{x_i}{p_i} \right) - V_2 \left(\frac{1}{m} \sum_{S_{2m}} \frac{z_i}{p_i} \right) \right]$$

(where $z_i = y_i - x_i$)

$$= \frac{1}{2} \left(\frac{1}{m} - \frac{1}{n} \right) \left[V_{pps}(y) + V_{pps}(x) - V_{pps}(z) \right]$$

where

$$V_{pps}(z) = \sum_i p_i \left(\frac{y_i - x_i}{p_i} - \overline{Y-X}\right)^2$$

$$= V_{pps}(x) + V_{pps}(y) - 2\rho\sqrt{(V_{pps}(x)\ V_{pps}(y))}$$

with

$$\rho\sqrt{(V_{pps}(x)\ V_{pps}(y))} = \sum_i p_i \left(\frac{y_i}{p_i} - Y\right)\left(\frac{x_i}{p_i} - X\right).$$

Thus

$$V_{13} = \left(\frac{1}{m} - \frac{1}{n}\right)\sqrt{(V_{pps}(x)\ V_{pps}(y))}\ .\rho$$

$$= \frac{\mu}{n\lambda}\rho\sqrt{V_{pps}(x)\ V_{pps}(y)}$$

Now $\frac{\partial}{\partial a_1} V(\hat{Y}^*) = 0$ and $\frac{\partial}{\partial b_1} V(\hat{Y}^*) = 0$ yields the optimum values of a_1 and b_1 as

$$a_{1\ opt}\ (\text{say}) = \left[V_1 + V_2 - \frac{V_{13}^2}{V_3}\right]^{-1} V_2$$

$$= \lambda\left[1 - \rho^2\mu^2\right]^{-1} \qquad (4.7.4)$$

and

$$b_{1\ opt}\ (\text{say}) = -a_{1\ opt}\frac{V_{13}}{V_3}$$

$$= -\lambda\left[1 - \rho^2\mu^2\right]^{-1}\rho\sqrt{\frac{V_{pps}(y)}{V_{pps}(x)}} \qquad (4.7.5)$$

Now putting the optimum values of a_1 and b_1 in (4.7.3) optimum value of $V(\hat{Y}^*)$ for fixed λ comes out as

$$V_{opt}(\hat{Y}^*|\lambda) = (1 - a_{1\ opt})\ V_2$$

$$= \frac{1 - \mu\rho^2}{1 - \mu^2\rho^2}\ \frac{V_{pps}(y)}{n} \qquad (4.7.6)$$

Now minimizing $V_{opt}(\hat{Y}^*|\lambda)$ with respect to λ we get the optimum values of μ and λ as

$$\frac{1}{1 + \sqrt{1-\rho^2}} \text{ and } \frac{\sqrt{1-\rho^2}}{1 + \sqrt{1-\rho^2}} \text{ respectively.}$$

Putting the optimum values of λ and μ in $V_{opt}(\hat{Y}^*|\lambda)$ and in $a_{1\ opt}$, $b_{1\ opt}$, we get the minimum value of $V(\hat{Y}^*)$ and the corresponding estimator for $\hat{Y}^*$ as

$$V_{min}(\hat{Y}^*) = \frac{1 + \sqrt{1-\rho^2}}{2n}\ V_{pps}(y)$$

and

$$\left.\begin{aligned} \hat{Y}^* = \frac{1}{2}\Big\{\frac{1}{m}\sum_{S_{2m}}\frac{y_i}{p_i} + \frac{1}{u}\sum_{r=1}^{u}\frac{y_r}{p_r}\Big\} \\ - \frac{1}{2}\rho\Big\{\frac{1}{m}\sum_{S_{2m}}\frac{x_i}{p_i} - \frac{1}{n}\sum_{r=1}^{n}\frac{x_r}{p_r}\Big\}\sqrt{\frac{V_{pps}(y)}{V_{pps}(x)}} \end{aligned}\right\} \quad (4.7.7)$$

Remark I

If we assume $V_{pps}(x) = V_{pps}(y) = V_o$, then $V_{min}(\hat{Y}^*)$ becomes $\frac{1 + \sqrt{1-\rho^2}}{2n}\ V_o$ which is less than $V_{min}(\hat{Y}) = \frac{1}{2n}\{1 + \sqrt{2(1-\rho)}\}\ V_o$ unless $\rho = 1$, in which use the Raj's (1965) original estimator coincides with the modified one.

Remark 2

If a completely un-matched sample of size n is selected and the Hansen-Hurwitz (1943) estimator is used for estimating the population total on the second occasion then Raj's (1965) strategy will be more efficient than Hansen-Hurwitz (1943) strategy provided $\rho > \frac{1}{2}$. But the proposed estimator with Raj's scheme of sampling results in gain in efficiency over Hansen-Hurwitz strategy for all values of ρ.

AN ALTERNATIVE TO GHANGURDE AND RAO'S (1969) ESTIMATOR BASED ON GHANGURDE AND RAO'S SAMPLING SCHEME

Here also we consider the estimator of the form

$$\hat{Y}'^{*} = a_2 \left(\frac{n}{m} \sum_{S_{2m}} \frac{y_i}{p_i} P_i\right) + (1-a_2) \sum_{S_{2u}} \frac{y_i}{p_i} P_i^{*}$$
$$+ b_2 \frac{n}{m} \left\{ \sum_{S_{2m}} \frac{x_i}{p_i} P_i - \sum_{S_1} \frac{x_i}{p_i} P_i \right\} \quad (4.7.8)$$

Clearly, $\hat{Y}'^{*}$ is unbiased for Y, so long as a_2, b_2 are constants independent of samples. They are to be chosen to minimize $V(\hat{Y}'^{*})$.

Now

$$V(\hat{Y}'*) = a_2^{\ 2}\ V\ (\frac{n}{m} \sum_{S_{2m}} \frac{y_i}{p_i} P_i) + (1-a_2)^2\ V\ (\sum_{S_{2u}} \frac{y_i}{p_i} P_i*)$$

$$+ b_2^{\ 2}\ V\ \{\frac{n}{m} \sum_{S_{2m}} \frac{x_i}{p_i} P_i - \sum_{S_1} \frac{y_i}{p_i} P_i\}$$

$$+ 2a_2b_2\ Cov\ \{\frac{n}{m} \sum_{S_{2m}} \frac{x_i}{p_i} P_i - \sum_{S_1} \frac{y_i}{p_i} P_i,$$

$$\frac{n}{m} \sum_{S_{2m}} \frac{y_i}{p_i} P_i\}$$

[Since $Cov\ \{\sum_{S_{2u}} \frac{y_i}{p_i} P_i*,\ \frac{n}{m} \sum_{S_{2m}} \frac{y_i}{p_i} P_i\}$

$$= 0 = Cov\ \{\sum_{S_{2u}} \frac{y_i}{p_i} P_i,\ \frac{n}{m} \sum_{S_{2m}} \frac{x_i}{p_i} P_i - \sum_{S_1} \frac{x_i}{p_i} P_i\}]$$

$$= a_2^{\ 2}V_1{}' + (1-a_2)^2\ V_2{}' + b_2^{\ 2}V_3{}' + 2a_2b_2V_{23}' \qquad (4.7.9)$$

where

$$V_1{}' = V\ (\frac{n}{m} \sum_{S_{2m}} \frac{y_i}{p_i} P_i)$$

$$= E_1V_2(\frac{n}{m} \sum_{S_{2m}} \frac{y_i}{p_i} P_i) + V_1E_2(\frac{n}{m} \sum_{S_{2m}} \frac{y_i}{p_i} P_i)$$

w

$$E_1V_2(\frac{n}{m} \sum_{S_{2m}} \frac{y_i}{p_i} P_i)$$

$$= n^2\ (\frac{1}{m} - \frac{1}{n})\ \frac{1}{n-1}\ E_1(\sum_{S_1} \frac{y_i^{\ 2}}{p_i^{\ 2}} P_i^{\ 2} - \frac{(\sum_{S_1} \frac{y_i}{p_i} P_i)^2}{n})$$

$$= n^2 \left(\frac{1}{m} - \frac{1}{n}\right) \frac{1}{n-1} \left\{ \frac{n-1}{n(N-1)} \bar{V}(y) + \frac{N-n}{n(N-1)} V_{pps}(y) + \frac{Y^2}{n} - \frac{V_1\left(\sum_{S_1} \frac{y_i}{p_i} P_i\right) + Y^2}{n} \right\}$$

ince, $E_1 \sum_{S_1} \frac{y_i^2}{p_i^2} P_i^2 = \underset{G}{E} \; E\left\{ \sum_{S_1} \frac{y_i^2}{p_i^2} P_i^2 \middle| G \right\}$

where

$\underset{G}{E}$ = unconditional expectation over groups that were formed in selecting the sample S_1 by RHC method,

$E(.|G)$ = conditonal expectation for given G.

$$E_1 \sum_{S_1} \frac{y_i^2}{p_i^2} P_i^2 = \underset{G}{E} \left\{ \sum_{S_1} \sum_{j \in G_i} \frac{y_j^2}{p_j} \sum_{j \in G_i} p_j \right\}$$

[where G_i denotes the ith group u; $(i = 1, \ldots, n)$]

$$= \underset{G}{E} \left\{ \sum_{S_1} \sum_{j \in G_i} y_i^2 + \sum_{S_1} \sum_{j \neq k \in G_i} \sum \frac{y_i^2}{p_i} p_j \right\}$$

$$= \sum_i y_i^2 + \frac{N-n}{n(N-1)} \sum_i \frac{y_i^2}{p_i} (1-p_i)$$

$$= \frac{n-1}{n(N-1)} \bar{V}(y) + \frac{N-n}{n(N-1)} V_{pps}(y) + \frac{Y^2}{n} \qquad (4.7.10)$$

Thus,

$$E_1 v_2 \left(\frac{n}{m} \sum_{S_{2m}} \frac{y_i}{p_i} P_i\right)$$

$$= \frac{1}{n(N-1)} \frac{\mu}{\lambda} [\, n\, \bar{V}(y) + (N-n)\, V_{pps}(y) \,] \qquad (4.7.11)$$

o,

$$V_1' = \frac{1}{n(N-1)} \frac{\mu}{\lambda} \{n\, \bar{V}(y) + (N-n)\, V_{pps}(y)\} + \frac{N-n}{n(N-1)} V_{pps}(y)$$

(4.7.12)

$$V_2' = V\left(\sum_{S_{2u}} \frac{y_i}{p_i} P_i^*\right) = \frac{N-n\mu}{n\mu(N-1)} V_{pps}(y) \qquad (4.7.13)$$

$$V_3' = V\left(\frac{n}{m} \sum_{S_{2m}} \frac{x_i}{p_i} P_i - \sum_{S_1} \frac{x_i}{p_i} P_i\right)$$

$$= E_1 v_2 \left(\frac{n}{m} \sum_{S_{2m}} \frac{x_i}{p_i} P_i - \sum_{S_1} \frac{x_i}{p_i} P_i\right)$$

$$= E_1 v_2 \left(\frac{n}{m} \sum_{S_{2m}} \frac{x_i}{p_i} P_i\right)$$

$$= \frac{1}{n(N-1)} \frac{\mu}{\lambda} \{n\, \bar{V}(x) + (N-n)\, V_{pps}(x)\} \qquad (4.7.14)$$

(using 4.7.11)

$$V_{13}' = \mathrm{Cov}\left\{\frac{n}{m} \sum_{S_{2m}} \frac{x_i}{p_i} P_i - \sum_{S_1} \frac{x_i}{p_i} P_i,\ \frac{n}{m} \sum_{S_{2m}} \frac{y_i}{p_i} P_i\right\}$$

$$= E_1\, \mathrm{Cov}_2 \left\{\frac{n}{m} \sum_{S_{2m}} \frac{x_i}{p_i} P_i,\ \frac{n}{m} \sum_{S_{2m}} \frac{y_i}{p_i} P_i\right\}$$

$$= \frac{1}{2} E_1 \{ V_2 (\frac{n}{m} \sum_{s_{2m}} \frac{x_i}{p_i} P_i) + V_2 (\frac{n}{m} \sum_{s_{2m}} \frac{x_i}{p_i} P_i) - V_2 (\frac{n}{m} \sum_{s_{2m}} \frac{z_i}{p_i} P_i) \}$$

(where $z_i = y_i - x_i$) (4.7.15)

using (4.7.11) we get

$$_{13} = \frac{1}{2} \frac{1}{n(N-1)} \frac{\mu}{\lambda} \{ n \bar{V}(y) + (N-n) V_{pps}(y) + n \bar{V}(x) + (N-n) V_{pps}(y) - n \bar{V}(z) - (N-n) V_{pps}(z) \}$$

$$(z) = N \sum_i \{ (y_i - \bar{Y}) - (x_i - \bar{X}) \}^2$$

$$= \bar{V}(y) + \bar{V}(x) - 2\delta \{ \bar{V}(y) \bar{V}(x) \}^{1/2}$$

$$V_{pps}(z) = \sum_i p_i \left(\frac{y_i - x_i}{p_i} - \overline{Y-X} \right)^2$$

$$= V_{pps}(y) + V_{pps}(x) - 2\rho \{ V_{pps}(x) V_{pps}(y) \}^{1/2}$$

s finally we get

$$V'_{13} = \frac{1}{n(N-1)} \frac{\mu}{\lambda} [n\delta \{ \bar{V}(y) \bar{V}(x) \}^{1/2} + (N-n) \{ V_{pps}(x) V_{pps}(y) \}^{1/2} \rho]$$

the expression (4.7.9) is similar to that of (4.7.3) have the optimum values of a_2, b_2 and $V_{opt}(\hat{Y}^* | \lambda)$ as

$$a_{2\ opt} = [V_1' + V_2' - \frac{V_{13}'^2}{V_3'}]^{-1} V_2' ,$$

$$b_{2\ opt} = - a_{2\ opt} \frac{V_{13}'}{V_3'} ,$$

$$V_{opt}(\hat{Y}'*|\lambda) = (1 - a_{2\ opt}) V_2'$$

$$= [\frac{1}{V_1' - \frac{V_{13}'^2}{V_3'}} + \frac{1}{V_2'}]^{-1} .$$

Now

$$V_1' - \frac{V_{13}'^2}{V_3'}$$

$$= \frac{1}{n(N-1)} \frac{\mu}{\lambda} \{n \bar{V}(y) + (N-n) V_{pps}(y)\} + \frac{N-n}{n(N-1)} V_{pps}(y)$$

$$- \frac{1}{n(N-1)} \frac{\mu}{\lambda} \frac{[n\delta\{\bar{V}(x) \bar{V}(y)\}^{1/2} + (N-n)\{V_{pps}(y) V_{pps}(x)\}^{1/2}\rho]^2}{n \bar{V}(x) + (N-n) V_{pps}(x)}$$

$$= \frac{1}{n(N-1)} \frac{\mu}{\lambda} [n \bar{V}(y) + (N-n) V_{pps}(y)] (1-\eta^2) + \frac{N-n}{n(N-1)} V_{pps}(y)$$

Where

$$\eta = \frac{n\delta \{\bar{V}(x) \bar{V}(y)\}^{1/2} + (N-n)\rho\{V_{pps}(x) V_{pps}(y)\}^{1/2}}{\{n \bar{V}(x) + (N-n) V_{pps}(x)\}^{1/2}\{n\bar{V}(y) + (N-n) V_{pps}(y)\}^{1/2}}$$

Thus,

$V_{opt}(\hat{Y}'*|\lambda)$ comes out as

$$V_{opt}(\hat{Y}'^{*}|\lambda)$$

$$[\frac{1}{\frac{1}{n(N-1)}\ \frac{\mu}{\lambda}\{n\bar{V}(y) + (N-n)V_{pps}(y)\}\ (1-\eta^2) + \frac{N-n}{n(N-1)}\ V_{pps}(y)}$$

$$+ \frac{1}{\frac{N-n\mu}{n\mu\ (N-1)}\ V_{pps}(y)}]^{-1}$$

$$= \frac{N\ V_{pps}(y)}{n(N-1)}\ [\frac{\lambda}{\mu\{\frac{n}{N}\ w + 1 - \frac{n}{N}\}\ (1-\eta^2) + \lambda(1-\frac{n}{N})} + \frac{\mu}{1-\frac{n\mu}{N}}]^{-1}$$

$$(\text{where}\quad w = \frac{\bar{V}(y)}{V_{pps}(y)})$$

us write

$$(\frac{n}{N}\ w + 1 - \frac{n}{N})(1-\eta^2) = 1-\xi^2 > 0.$$

, we have

$$V_{opt}(\hat{Y}'^{*}|\lambda)$$

$$\frac{N}{n(N-1)}\ [\frac{\lambda}{\mu(1-\xi^2) + \lambda(1-\frac{n}{N})} + \frac{\mu}{1-\frac{n\mu}{N}}]^{-1}\ V_{pps}(y)$$

(4.7.16)

$$a_{2\ opt} = [V_1' + V_2' - \frac{V_{13}^{2\,\prime}}{V_3'}]^{-1}\ V_2'$$

$$= [\frac{\mu}{\lambda}\ (1-\xi^2) + (1-\frac{n}{N}) + \frac{(1-\frac{n\mu}{N})}{\mu}]^{-1}\ \frac{(1-\frac{n\mu}{N})}{\mu}$$

$$= \lambda\ (1-\frac{n\mu}{N})\ [1 - \mu^2\xi^2 - \frac{2n}{N}\ \mu\lambda]^{-1}$$

So,

$$b_{2\ opt} = -\lambda [\, 1 - \mu^2 \xi^2 - \tfrac{2n}{N}\, \mu\lambda]^{-1}\ (1 - \tfrac{n\mu}{N})$$

$$\eta \left\{ \frac{n\ \bar{V}(y) + (N-n)\ V_{pps}(y)}{n\ \bar{V}(x) + (N-n)\ V_{pps}(x)} \right\}$$

It can be easily checked that $V_{opt}(\hat{Y}'^{*}|\lambda)$ is of the same form as $V_{opt}(\hat{Y}'|\lambda)$ given in (4.3.5). From (4.3.6) and (4.3.7) we get the optimum values of μ, λ and minimum value of variance of $\hat{Y}'^{*}$ as

$$\mu_{opt} = [1 + \sqrt{1-\xi^2}\,]^{-1}, \quad \lambda_{opt} = \frac{\sqrt{1-\xi^2}}{1 + \sqrt{1-\xi^2}} \qquad (4.7.17)$$

and

$$V_{min}(\hat{Y}'^{*}) = \frac{N}{2n(N-1)}\ \{1 + \sqrt{1-\xi^2} - \tfrac{n}{N}\}\ V_{pps}(y) \qquad (4.7.18)$$

Putting these optimum values of $\mu = \mu_{opt}$ and $\lambda = \lambda_{opt}$ in the expression for $a_{2\ opt}$ and $b_{2\ opt}$ we have

$$a_{2\ opt} = a_{20}\ (\text{say}) = \frac{1}{2}$$

$$b_{2\ opt} = b_{20}\ (\text{say}) = -\frac{1}{2}\ \eta\ \left\{ \frac{n\ \bar{V}(y) + (N-n)\ V_{pps}(y)}{n\ \bar{V}(x) + (N-n)\ V_{pps}(x)} \right\}$$

Remark I

The improved estimator $\hat{Y}'^{*}$ with Ghangurde and Rao's (1969) scheme of sampling will be superior to customary estimator for population total Y based on completely un-matched sample of size n, selected by RHC scheme of sampling provided $\xi^2 > \frac{n}{N}(2 - \frac{n}{N})$.

Remark II

If we assume, as we have done so far in following Bhangurde and Rao (1969), that $\bar{V}(y) = \bar{V}(x) = \bar{V}$ and $V_{pps}(y) = V_{pps}(x) = V_o$, then we get

$$V_{min}(\hat{Y}') - V_{min}(\hat{Y}'^*)$$

$$= \frac{N}{2n(N-1)} \left[\sqrt{2(1-\rho)(1 + \frac{\gamma n}{N})} - \sqrt{1-\xi^2} \right] V_o .$$

Now

$$2(1-\rho)(1 + \frac{\gamma n}{N}) - (1-\xi^2)$$

$$= 2(1-\rho)(1 - \frac{n}{N} + \frac{n}{N} \frac{1-\delta}{1-\rho} w) - \frac{(N-n+nw)}{N} (1-\eta^2)$$

$$= \frac{2}{N} \{ (N-n)(1-\rho) + n(1-\delta) w \}$$

$$- \frac{(nw+N-n)^2 - \{ n\delta w + (N-n)\rho \}^2}{N(nw+N-n)}$$

$$= \frac{1}{N(nw+N-n)} [2(nw+N-n) \{ (N-n)(1-\rho) + n(1-\delta) w \}$$

$$- (nw+N-n)^2 + \{ n\delta w + (N-n)\rho \}^2]$$

$$= \frac{1}{N(nw+N-n)} [(nW+N-n)^2 + (n\delta w + \overline{N-n}\,\rho)^2$$

$$- 2(nw+N-n)(\overline{N-n}\,\rho + n\delta w)]$$

$$= \frac{1}{N(nw+N-n)} [nw(1-\delta) + N-n\ (1-\rho)]^2.$$

Thus the proposed estimator is more efficient than Ghangurde and Rao's original estimator both based on the same sampling scheme, unless

$$nw(1-\delta) + (N-n)(1-\rho) = 0 \qquad (4.7.19)$$

AN ALTERNATIVE TO CHOTAI'S ESTIMATOR BASED ON CHOTAI'S SAMPLING SCHEME

Here also we propose the estimator of the following type

$$\hat{Y}^{**} = a_3\Big(\sum_{S_{2m}} \frac{y_i}{p_i}\ \Delta_i\Big) + (1-a_3)\Big(\sum_{S_{2u}} \frac{y_i}{p_i}\ p_i^*\Big) + b_3\Big(\sum_{S_{2m}} x_i \frac{\Delta_i}{p_i} - \sum_{S_1} \frac{x_i}{p_i}\ p_i\Big) \qquad (4.7.20)$$

Here the weights a_3 and b_3 are constants independent of the samples, and are to be chosen so as to minimize the variance of $\hat{Y}^{**}$.

ow

$$V(\hat{Y}^{**}) = a_3^2\ V\Big(\sum_{S_{2m}} \frac{y_i}{p_i}\ \Delta_i\Big) + (1-a_3)^2\ V\Big(\sum_{S_{2u}} \frac{y_i}{p_i}\ p_i^*\Big) + b_3^2\ V\Big(\sum_{S_{2m}} \frac{x_i\ \Delta_i}{p_i} - \sum_{S_1} \frac{x_i}{p_i}\ p_i\Big)$$

$$+ 2a_3b_3 \text{ Cov } \left(\sum_{s_{2m}} \frac{y_i}{p_i} \Delta_i , \sum_{s_{2m}} \frac{x_i \Delta_i}{p_i} - \sum_{s_1} \frac{x_i}{p_i} P_i \right)$$

Since $\text{Cov} \left(\sum_{s_{2u}} \frac{y_i}{p_i} P_i^*, \sum_{s_{2m}} \frac{y_i}{p_i} \Delta_i \right)$

$$= 0 = \text{Cov} \left(\sum_{s_{2u}} \frac{y_i}{p_i} P_i^*, \sum_{s_{2m}} \frac{x_i \Delta_i}{p_i} - \sum_{s_1} \frac{x_i}{p_i} P_i \right)]$$

$$= a_3^2 V_1'' + (1-a_3)^2 V_2'' + b_3^2 V_3'' + 2a_3b_3V_{13}'' \qquad (4.7.21)$$

where

$$V_1'' = V \left(\sum_{s_{2m}} \frac{y_i}{p_i} \Delta_i \right)$$

$$= E_1 V_2 \left(\sum_{s_{2m}} \frac{y_i}{p_i} \Delta_i \right) + V_1 E_2 \left(\sum_{s_{2m}} \frac{y_i}{p_i} \Delta_i \right)$$

$$= \frac{n-m}{m(n-1)} \left\{ E_1 \sum_{s_1} \frac{y_i^2}{p_i^2} P_i - V_1 \left(\sum_{s_1} \frac{y_i}{p_i} P_i \right) - Y^2 \right\}$$

$$+ V_1 \left(\sum_{s_1} \frac{y_i}{p_i} P_i \right)$$

$$= \frac{n-m}{m(n-1)} \left\{ \sum_i \frac{y_i^2}{p_i} - Y^2 \right\} + \left(1 - \frac{n-m}{m(n-1)}\right) V_1 \left(\sum_{s_1} \frac{y_i}{p_i} P_i \right)$$

$$= \frac{1}{m(n-1)} \left[(n-m) + \frac{(m-1)(N-n)}{N-1} \right] V_{pps}(y)$$

$$= \frac{1}{n(N-1)} \left[\frac{N}{\lambda} - n \right] V_{pps}(y) .$$

$$V_2'' = V\left(\sum_{S_{2u}} \frac{y_i}{p_i} p_i^*\right) = \frac{N-u}{u(N-1)} V_{pps}(y)$$

$$= \frac{1}{n(N-1)} \left(\frac{N}{\mu} - n\right) V_{pps}(y)$$

$$V_3'' = V\left(\sum_{S_{2m}} \frac{x_i}{p_i} \Delta_i - \sum_{S_1} \frac{x_i}{p_i} P_i\right)$$

$$= E_1 v_2 \left(\sum_{S_{2m}} \frac{x_i}{p_i} \Delta_i\right)$$

$$= \frac{N}{N-1} \frac{\mu}{n\lambda} V_{pps}(x) \qquad (4.7.22)$$

$$V_{13}'' = \mathrm{Cov}\left(\sum_{S_{2m}} \frac{y_i}{p_i} \Delta_i, \sum_{S_{2m}} \frac{x_i}{p_i} \Delta_i - \sum_{S_1} \frac{x_i}{p_i} P_i\right)$$

$$= E_1 \mathrm{Cov}_2\left(\sum_{S_{2m}} \frac{y_i}{p_i} \Delta_i, \sum_{S_{2m}} \frac{x_i}{p_i} \Delta_i\right)$$

$$= \frac{1}{2} E_1\left\{ v_2\left(\sum_{S_{2m}} \frac{y_i}{p_i} \Delta_i\right) + v_2\left(\sum_{S_{2m}} \frac{x_i}{p_i} \Delta_i\right) - v_2\left(\sum_{S_{2m}} \frac{z_i}{p_i} \Delta_i\right)\right\}$$

(where $z_i = y_i - x_i$)

N using (4.7.22) we get

$$V_{13}'' = \frac{N}{2(N-1)} \frac{\mu}{n\lambda} \left[V_{pps}(y) + V_{pps}(x) - V_{pps}(z) \right]$$

$$= \frac{N}{N-1} \frac{\mu}{n\lambda} \rho \left\{V_{pps}(x) V_{pps}(y)\right\}^{1/2}$$

$$V_{pps}(z) = \sum_i p_i \left(\frac{y_i - x_i}{p_i} - \overline{Y-X}\right)^2$$

$$= V_{pps}(y) + V_{pps}(x) - 2\rho \{V_{pps}(x)\ V_{pps}(y)\}^{1/2}]$$

expression for $V(\hat{Y}''^*)$ in (4.7.21) is of the same

in (4.7.3) we have the optimum values of a_3, b_3

$_{pt}(\hat{Y}''^*|\lambda)$ by using (4.7.4), (4.7.5) and (4.7.6) as

$$a_3 = a_{3\ opt}\ (\text{say}) = [V_1'' + V_2'' - \frac{V_{13}^{2}{}''}{V_3''}]^{-1} V_2''$$

$$= \lambda(1 - \frac{n\mu}{N})\ (1 - \mu^2\rho^2 - 2n\frac{\mu\lambda}{N})^{-1},$$

$$b_3 = b_{3\ opt}\ (\text{say}) = - a_{3\ opt} \frac{V_{13}''}{V_3''}$$

$$= - \lambda(1 - \frac{n\mu}{N})\ (1 - \mu^2\rho^2 - 2n\frac{\mu\lambda}{N})^{-1} \rho\left(\frac{V_{pps}(y)}{V_{pps}(x)}\right)^{1/2}$$

$$V_{opt}(\hat{Y}''^*|\lambda) = (1 - a_{3\ opt})\ V_2''$$

$$= \left[\frac{\lambda}{\mu(1-\rho^2) + \lambda(1 - \frac{n}{N})} + \frac{\mu}{(1 - \frac{n\mu}{N})}\right]^{-1} \frac{N}{n(N-1)} V_{pps}(y) \qquad (4.7.23)$$

we note that the form $V_{opt}(\hat{Y}'^*|\lambda)$ is the same as

$\hat{Y}'|\lambda)$ given in (4.3.5). So, following (4.3.7) and

(6) we have the minimum variance of $\hat{Y}''^*$, the optimum

s of μ and λ as

$$V_{min}(\hat{Y}''*) = \frac{N}{2n(N-1)} \{1 - \frac{n}{N} + \sqrt{1-\rho^2}\} \ V_{pps}(y) \qquad (4.7.24)$$

$$\mu_{opt} = \{1 + \sqrt{1-\rho^2}\}^{-1} \quad \text{and} \quad \lambda_{opt} = \frac{\sqrt{1-\rho^2}}{1 + \sqrt{1-\rho^2}} \qquad (4.7.25)$$

Now putting the optimum values of μ and λ in the expressions for $a_{3\ opt}$ and $b_{3\ opt}$ we have

$$a_{3\ opt} = a_{30}\ (\text{say}) = \frac{1}{2}$$

$$b_{3\ opt} = b_{30}\ (\text{say}) = \frac{1}{2} \left\{ \frac{V_{pps}(y)}{V_{pps}(x)} \right\}^{1/2} \rho$$

[(4.7.24) was also derived by Chotai (1974) in a different way]

Remark I

Here we note that $V_{min}(\hat{Y}''*) \leq V_{min}(\hat{Y}'')$ unless $\rho = 1$, i.e. the modified estimator with Chotai's sampling scheme is superior to Chotai's original estimator for his scheme of sampling unless $\rho = 1$.

Remark II

The strategy employing modified estimator with Chotai's scheme of sampling is superior to RHC strategy based on the completely un-matched sample of the same size.

4.8 COMPARISON AMONG THE ESTIMATORS IMPROVED UPON THOSE DUE TO RAJ (1965), GHANGURDE AND RAO (1969) AND CHOTAI (1974)

We have

$$V_{min}(\hat{Y}^*) - V_{min}(\hat{Y}^{*\prime\prime})$$

$$= \frac{1}{2n}\left[1 + \sqrt{1-\rho^2} - \frac{N-n}{N-1} - \frac{N}{N-1}\sqrt{1-\rho^2}\right] V_{pps}(y)$$

$$= \frac{1}{2n(N-1)}\left[n - 1 - \sqrt{1-\rho^2}\right] V_{pps}(y)$$

$$\geq 0 \qquad \text{(since } n \geq 2\text{)}$$

Thus modified estimator with Chotai's scheme of sampling is superior to that based on Raj's scheme of sampling.

Similarly

$$V_{min}(\hat{Y}^*) - V_{min}(\hat{Y}'^*)$$

$$= \frac{1}{2n}\left[1 + \sqrt{1-\rho^2} - \frac{N-n}{N-1} - \frac{N}{N-1}\sqrt{1-\xi^2}\right] V_{pps}(y)$$

$$\approx \frac{1}{2n}\left[\sqrt{1-\rho^2} - \sqrt{1-\xi^2} + \frac{n-1}{N-1}\right]$$

$$\text{(assuming } \frac{N}{N-1} \approx 1\text{)}.$$

Thus if $|\xi| > |\rho|$, then estimator improved upon Ghangurde and Rao's (1969) is more efficient than that on Raj's (1965) at least for large populations.

Now,

$$V_{min}(\hat{Y}''*) - V_{min}(\hat{Y}'*)$$

$$= \frac{N}{2n(N-1)} [\sqrt{1-\rho^2} - \sqrt{1-\xi^2}].$$

Thus $V_{min}(\hat{Y}'*)$ <, = or > $V_{min}(\hat{Y}''*)$ according as

$$\rho^2 >, = \text{ or } < \xi^2$$

$$\text{i.e., } \frac{1-\rho^2}{1-\eta^2} <, = \text{ or } > \frac{n}{N}(w-1)+1$$

We are giving a table below to compare the relative efficiencies of $\hat{Y}'*$ and $\hat{Y}''*$ under the assumption that

$$\bar{V}(x) = \bar{V}(y) = \bar{V} \quad \text{and} \quad V_{pps}(x) = V_{pps}(y) = V_o$$

Table (4.1) showing the values of $V_{min}(\hat{Y}''*)/V_{min}(\hat{Y}'*)$ FOR GIVEN VALUES OF δ, , w and n/N

	ρ = .6				ρ = .7			ρ = .8		ρ = .9
w \ δ	.6	.7	.8	.9	.7	.8	.9	.8	.9	.9
				$\frac{n}{N} = .1$						
1.00	1.000	1.004	1.009	1.014	1.000	1.006	1.012	1.000	1.009	1.000
2.00	.978	.986	.995	1.004	.979	.990	1.003	.981	.998	[illegible]
3.00	.957	.969	.981	.995	.959	.975	.994	.963	.987	.970
4.00	.938	.953	.968	.987	.941	.962	.985	.947	.977	[illegible]
5.00	.921	.937	.956	.979	.925	.949	.977	.932	.968	[illegible]
6.00	.904	.923	.946	.971	.910	.936	.969	.918	.958	[illegible]
7.00	.889	.910	.935	.964	.895	.942	.961	.904	.950	.920
8.00	.875	.898	.925	.958	.881	.914	.954	.892	.942	[illegible]
				$\frac{n}{N} = .25$						
1.00	1.000	1.013	1.027	1.042	1.000	1.018	1.037	1.000	1.026	1.000
2.00	.943	.963	.987	1.016	.945	.975	1.010	.950	.995	[illegible]
3.00	.890	.922	.954	.994	.901	.938	.986	.909	.968	.923
4.00	.857	.887	.925	.974	.864	.907	.965	.875	.944	.894
5.00	.824	.857	.899	.956	.832	.879	.945	.845	.921	[illegible]
6.00	.795	.830	.876	.939	.864	.855	.927	.818	.901	[illegible]
7.00	.770	.806	.855	.924	.779	.832	.911	.795	.882	.824
8.00	.746	.784	.835	.910	.757	.812	.896	.774	.865	[illegible]

4.9 SAMPLING ON TWO OCCASIONS ON STRATIGYING AN INITIAL SAMPLE

We suppose that a population is stratified according to some criteria but is not known which units falls in which particular stratum. On the first occasion, therefore, an SRSWOR of size n (say) is chosen from the population. But once it is observed we know the distribution of the sample values into the pre-specified strata, then, on thesecond occasion we suggest that it may be profitable to draw sub-samples independently from the respective strata by suitable methods, m_h being the number of matched units drawn from the n_h units falling in the hth stratum out of the n-units chosen initially. Here $m_h = \nu_h n_h$ where ν_h is a pre-assigned positive proper fraction for $h = 1, \ldots, L$ (the number of strata being L). Finally an un-matched sample of size $u = n - \sum_{h=1}^{L} m_h = n-m$ (say) is chosen by SRSWOR method from the part of the population leaving aside the initially chosen sample. For further discussion of our approach we consider the following methods of choosing the matched stratified sample and theestimators based on the matched samples for the two occasions. In this context we may mention Cochran's (1963) method of double sampling for stratification subsequently improved upon by Rao, J.N.K. (1973). They stratified the population on the basis of values observed on a large initial sample of size n' (say) taken by SRSWOR method. From the n_h' units falling in the hth stratum so formed sub-samples of sizes n_h are drawn by SRSWOR method independently for each h = 1, ..., L. About Cochran's result for optimal choice of n' and n_h 's Rao observed that n_h's chosen are bounded by random variables, namely, n_h' 's. So, he modified Cochran's procedure by requiring n_h's to be pre-assigned fractions ν_h's of n_h' 's and got alternative results.

First we note the following two strategies IV and V described below.

In both m_h sub-samples are chosen by SRSWOR method. The estimators for $\bar{Y}$ based on the matched samples are proposed as

$$t_4 = \sum_h w_h \bar{y}_{mh} \qquad (4.9.1)$$

$$t_5 = \sum_h w_h \frac{\bar{y}_{mh}}{\bar{x}_{mh}} \bar{x}_{nh} \qquad (4.9.2)$$

respectively for strategies IV and V.

By $\bar{X}$, $\bar{Y}$, $\bar{X}_h$, $\bar{Y}_h$ we shall mean the population and strata means and by $\bar{x}_{nh}$, $\bar{x}_{mh}$, $\bar{y}_{mh}$ and w_h the sample and sub-sample means and sampling fractions for the strata of unknown sizes N_h's (with $N_h = N\ W_h$ and $n_h = n\ w_h$) for the 1st and second occasion and x_{hj}'s y_{hj}'s denote the respective variate values.

$$\text{By} \quad t = \bar{y}_u$$

we mean the un-matched sample mean.

The proposed composite estimators for the population mean for the second occasion for the strategies are

$$\left.\begin{aligned} T_4 &= \phi_1\ t_4 + (1-\phi_1)\,t \\ \text{and} \qquad & \\ T_5 &= \phi_2\ t_5 + (1-\phi_2)\,t \end{aligned}\right\} \qquad (4.9.3)$$

with $0 \leq \phi_1, \ \phi_2 \leq 1$, to be specified to minimize variances of T_i $(i = 4, 5)$.

Following Rao (1973) we get the following

<u>Theorem 4.14</u>

t_4 is an unbiased estimator for $\bar{Y} = \Sigma\, W_h\ \bar{Y}_h$ with

$$V(t_4) = \left(\frac{1}{n} - \frac{1}{N}\right) S_y^{\ 2} + \sum_h \frac{W_h}{n} \left(\frac{1}{\nu_h} - 1\right) S_{hy}^2$$

with usual notations for finite population and stratum variances.

Proof

Using conditional [given a vector $n = (n_1, \ldots, n_L)$] operators for expectations, variances and covariances viz. E_2, v_2 and Cov_2 and usual operators E_1 and v_1 for expectation and variance over the initial sample we have

$$E(t_4) = \sum_h E_1 (w_h) \; E_2 (\bar{y}_{mh}) = \sum_h W_h \bar{Y}_h$$

and

$$V(t_4) = E_1 \sum_h w_h^{\,2} \left(\frac{1}{\nu_h n_h} - \frac{1}{N_h}\right) S_{hy}^{\,2} + v_1 (\sum_h w_h \bar{Y}_h)$$

$$= \left(\frac{1}{n} - \frac{1}{N}\right) S_y^{\,2} + \frac{\Sigma W_h S_{hy}^{\,2}}{n} \left(\frac{1}{\nu_h} - 1\right)$$

since

$$v_2(w_h) = \frac{N-n}{(N-1)} \; \frac{W_h (1-W_h)}{n} \qquad \text{for } h = 1, \ldots, L$$

$$Cov_2(w_h, w_k) = - \frac{N-n}{N-1} \; \frac{W_h W_k}{n} \qquad \text{for } h \neq k = 1, \ldots, L.$$

For proportional allocation $\nu_h = \frac{m}{n}$ and we have

$$V_{prop}(t_4) = \left(\frac{1}{n} - \frac{1}{N}\right) S_y^{\,2} + \left(\frac{1}{m} - \frac{1}{n}\right) \sum_h W_h S_{hy}^{\,2}$$

$$= V_4 \quad \text{(say)} \qquad (4.9.4)$$

Similarly, for the estimator t_5, following Cochran (1963) if we assume that m_h and n_h are so large that for every h,

$$E_2\left(\frac{\bar{y}_{mh}}{\bar{x}_{mh}}\ \bar{x}_{nh}\right) = \bar{Y}_h$$

and

$$V_2\left(\frac{\bar{y}_{mh}}{\bar{x}_{mh}}\ \bar{x}_{nh}\right) = \left(\frac{1}{m_h} - \frac{1}{n_h}\right)\frac{1}{N_h-1}\ \sum_j (y_{hj} - R_h\ x_{hj})^2,$$

$$(\text{where}\ \ R_h = \bar{Y}_h/\bar{X}_h),$$

then it follows that

$$E\ (t_5) = \bar{Y}$$

and

$$V\ (t_5) = E_1 \sum_h w_h^2\ \{\left(\frac{1}{m_h} - \frac{1}{n_h}\right)\ \frac{1}{N_h-1}\ \sum_j (y_{hj} - R_h\ x_{hj})^2$$

$$+ \left(\frac{1}{n_h} - \frac{1}{N_h}\right)\ S_{hy}^2\} + V_1\left(\sum_h w_h\ \bar{Y}_h\right)$$

$$= \frac{1}{n}\ \sum_h \left(\frac{W_h}{\nu_h} - W_h\right)\ \frac{1}{N_h-1}\ \sum_j (y_{hj} - R_h\ x_{hj})^2$$

$$+ \sum_h \left(\frac{W_h}{n} - \frac{W_h^2 + \frac{N-n}{n(N-1)}\ W_h\ (1-W_h)}{N_h}\right)\ S_{hy}^2$$

$$+ \frac{N-n}{n(N-1)}\ \sum_h W_h\ (\bar{Y}_h - \bar{Y})^2$$

$$= \left(\frac{1}{n} - \frac{1}{N}\right) S_y^2 + \sum_h \frac{W_h}{n}\ \left(\frac{1}{\nu_h} - 1\right)\frac{1}{N_h-1}\ \sum_j (y_{hj}-R_h\ x_{hj})^2 .$$

In particular, for proportional allocation we have $\nu = \frac{m}{n}$ and

$$V_{prop}(t_5) = (\frac{1}{n} - \frac{1}{N}) S_y^2 + (\frac{1}{m} - \frac{1}{n}) \sum_h W_h \frac{1}{N_h-1} \sum_j (y_{hj} - R_h x_{hj})^2$$

$$= V_5 \quad \text{(say)} \qquad (4.9.5)$$

lso, it follows that

$$V(\bar{t}) = (\frac{1}{u} - \frac{1}{N}) S_y^2 = V \quad \text{(say)} \qquad (4.9.6)$$

nd

$$Cov(t, t_i) = - \frac{S_y^2}{N} \quad \text{for } i = 4,5$$

$$= V_{1i} \quad \text{(say)} \quad \text{for } i = 4,5 \qquad (4.9.7)$$

r fixed $\lambda\ (= \frac{m}{n})$, the minimum (w.r.t ϕ_i, i = 1,2)

riances of T_i's are

$$V_{opt}(T_i|\lambda) = [\frac{1}{V_i'} + \frac{1}{V'}]^{-1} - \frac{S_y^2}{N} \quad \text{for } i = 4,5$$

.....(4.9.8)

ere

$$V_i' = V_i - V_{1i} = V + \frac{S_y^2}{N} \quad \text{for } i = 4,5$$

$$V' = V - V_{1i} = V + \frac{S_y^2}{N}$$

so, the choices $\lambda_{io} = \frac{\sqrt{\alpha_i}}{1 + \alpha_i}$ minimizes $V_{opt}(T_i|\lambda)$

ving the minimum variance

$$V_{min}(T_i) = [\frac{1 + \sqrt{\alpha_i}}{2n} - \frac{1}{N}] S_y^2 \qquad (4.9.9)$$

th

$$(\frac{1}{m} - \frac{1}{n}) \alpha_i = \frac{V_i' - S_y^2/n}{S_y^2} = \frac{V_i'}{S_y^2} - \frac{1}{n} \quad \text{for } i = 4,5$$

Strategy VI

Here the sub-samples are taken by RHC method using size measures proportional to x_{hj}'s, normed over the respective strata for the initial sample.

The estimator for $\bar{Y}$ based on the matched sample is proposed as

$$t_6 = \sum_h w_h \bar{y}_h' \qquad (4.9.10)$$

with $\bar{y}_h'$ as the usual RHC estimator for $\bar{Y}_h$.
The final composite unbiased estimator for $\bar{Y}$ is

$$T_6 = \phi_3 t_6 + (1-\phi_3) t, \quad \text{with } 0 \leq \phi_3 \leq 1 \qquad (4.9.11)$$

Here

$$V(t_6) = E_1 \sum_h v_2 (\bar{y}_h') + v_1 (\sum_h w_h \bar{Y}_h)$$

$$= E_1 \sum_h w_h^2 \{ (\frac{1}{n_h} - \frac{1}{N_h}) S_{hy}^2$$

$$+ (\frac{1}{m_h} - \frac{1}{n_h}) \frac{1}{N_h(N_h-1)} \sum_j p_{hj}' (\frac{y_{hj}}{p_{hj}'} - Y_h)^2 \}$$

$$+ v_1 (\sum_h w_h \bar{Y}_h)$$

The term inside the second bracket is obtained by following Avadhani and Sukhatme (1970)]

$$(\text{where } p_{hj}' = \frac{x_{hj}}{X_h}, \quad X_h = N_h \bar{X}_h, \; Y_h = N_h \bar{Y}_h)$$

Thus,

$$V(t_6) = (\frac{1}{n} - \frac{1}{N}) S_y^2 + \sum_h (\frac{1}{\nu_h} - 1) \frac{W_h}{n\, N_h(N_h-1)} \sum_j p'_{hj}(\frac{Y_{hj}}{p'_{hj}} - Y_h)^2 \tag{4.9.12}$$

For proportional allocation $\nu_h = \frac{m}{n}$ $\forall\ h = 1, \ldots, L$

we have

$$V_{prop}(t_6) = (\frac{1}{n} - \frac{1}{N}) S_y^2 + (\frac{1}{m} - \frac{1}{n})\frac{1}{N} \sum_h \frac{1}{N_h - 1} \sum_j p'_{hj}(\frac{Y_{hj}}{p'_{hj}} - Y_h)^2$$

$$= V_6 \quad (\text{say}) \tag{4.9.13}$$

Also,

$$\text{Cov}\,(t_6, t) = - \frac{S_y^2}{N} \tag{4.9.14}$$

The choice $\lambda_{60} = \frac{\sqrt{\alpha_6}}{1 + \sqrt{\alpha_6}}$ minimizes the minimum (w.r.t ϕ) value of the variance of T_6 for fixed λ giving the over-all minimum as

$$V_{min}(T_6) = [\frac{1 + \sqrt{\alpha_6}}{2n} - \frac{1}{N}] S_y^2 \tag{4.9.15}$$

where

$$(\frac{1}{m} - \frac{1}{n})\alpha_6 = \frac{V_6'}{S_y^2} - \frac{1}{n} \tag{4.9.16}$$

with $V_6' = V_6 + S_y^2/N$

COMPARISON AMONG THE STRATEGIES IV, V, VI and F and G due to Avadhani and Sukhatme (1970)

From what precedes it is clear that magnitudes of $V(t_i)$, decide the magnitudes of $V_{min}(T_i)$, $i = 4, 5, 6$

(i) Strategy IV Vs Strategy V

We have

$$D_1 = V_4 - V_5 = (\frac{1}{m} - \frac{1}{n}) \{ \sum_h W_h S_{hy}^2 - \sum_h \frac{W_h}{N_h - 1} \sum_j (y_{hj} - R_h x_{hj})^2 \}$$

$$= (\frac{1}{m} - \frac{1}{n}) \sum_h W_h \{ 2 \rho_h R_h S_{hx} S_{hy} - R_h^{2} S_h$$

[where ρ_h is the finite population correlation coefficient for the hth stratum]

Now if

$$\left.\begin{array}{ll} \text{either} & \rho_h > \frac{R_h}{2} \frac{S_{hx}}{S_{hy}} \text{ along with } R_h > 0 \\ \text{or} & \rho_h < \frac{R_h}{2} \frac{S_{hx}}{S_{hy}} \text{ along with } R_h < 0 \end{array}\right\} \text{ for all } h=1, \ldots, L \quad (4.9.17)$$

then $D_1 > 0$ implying strategy V to be more efficient than strategy IV. Thus in situations where ratio estimator is an improvement upon the mean per unit estimator for each stratum, the strategy V fares better than strategy IV.

(ii) Strategy V Vs Strategy VI

Here

$$D_2 = V_5 - V_6 = (\frac{1}{m} - \frac{1}{n}) \frac{1}{N} \sum_h \frac{1}{N_h - 1} \{ N_h \sum_j (y_{hj} - R_h x_{hj})^2 - \sum_j p'_{hj} (\frac{y_{hj}}{p'_{hj}} - Y_h)^2 \} .$$

Now if we assume the model M_2' for which

$$\left.\begin{aligned} & y_{hi} = \beta_h x_{hi} + e_{hi} \\ & \sum_i e_{hi} = 0 = \sum_i e_{hi} x_{hi} \\ & \text{and } e_{hi}^2 = \sigma_h^2 x_{hi}^g \end{aligned}\right\} \quad \forall \; h = 1, \ldots, L$$

then recalling Avadhani and Sukhatme (1970) it is easy to check that

$$V_5 \geq, = \text{ or } < V_6 \text{ according as } g \geq, = \text{ or } < 1.$$

(iii) Strategy V Vs Strategy F

Here

$$V_5 - V(\bar{Y}_{RHC}) = \left(\frac{1}{m} - \frac{1}{n}\right)\frac{1}{N}\left[\sum_h \frac{N_h}{N_h - 1} \sum_j (y_{hj} - R_{hj} x_{hj})^2 - \frac{1}{N-1}\left(\sum_h \sum_i \frac{y_{hi}^2}{p_{hi}} - Y^2\right)\right]$$

Now if we assume the model M_2'' under which

$$\left.\begin{aligned} & y_{hj} = \beta x_{hj} + e_{hj} \\ & \sum_j e_{hj} = 0 = \sum_j e_{hj} x_{hj} \\ & \text{and } e_{hj}^2 = \sigma^2 x_{hj}^g \end{aligned}\right\} \forall \; h = 1, \ldots L$$

then neglecting the errors in writing $\frac{1}{N_h - 1} = \frac{1}{N_h}$ $\forall$ h and $\frac{1}{N-1} = \frac{1}{N}$, we have approximately,

$$V_5 - V(\bar{Y}_{RHC}) \approx (\frac{1}{m} - \frac{1}{n}) \frac{1}{N} \sigma^2 \sum_h \sum_j x_{hj}^{g-1}(x_{hj} - \bar{X}).$$

So F is superior of inferior to strategy V according as $g >$ or < 1 and they become equivalent when $g = 1$.

Now if we assume the model M_2' instead of M_2'' and assuming $\frac{1}{N_h - 1} \approx \frac{1}{N_h}$ $\forall$ h and $\frac{1}{N-1} \approx \frac{1}{N}$ we have

$$V_5 - V(\bar{Y}_{RHC}) \approx (\frac{1}{m} - \frac{1}{n}) \frac{1}{N} [\sigma^2 \sum_h \sum_i x_{hi}^{g-1}(x_{hi} - \bar{X})$$

$$- \frac{1}{N} \sum_{h<h'} \sum X_h X_{h'} (\beta_h - \beta_{h'})^2]$$

$$= (\frac{1}{m} - \frac{1}{n}) \frac{1}{N} f_1(g)$$

where

$$f_1(g) = \sigma^2 \sum_h \sum_i x_{hi}^{g-1} (x_{hi} - \bar{X})$$

$$- \frac{1}{N} \sum_{h<h'} \sum X_h X_{h'} (\beta_h - \beta_{h'})^2.$$

But $f_1(1) < 0$ and $f_1(g)$ is an increasing function of g. So under M_2', $V_5 <, = > V(\bar{Y}_{RHC})$ according as $g <, =$ or $> g_o$ where $g_o > 1$.

(v) Strategy VI Vs strategy G

Here $V(\bar{Y}_R) - V_6$

$$= (\frac{1}{m} - \frac{1}{n}) [\frac{1}{N-1} \sum_h \sum_j (y_{hj} - R x_{hj})^2$$

$$- \frac{1}{N} \sum_n \frac{1}{N_h - 1} \sum_j p'_{hj} (\frac{y_{hj}}{p'_{hj}} - Y_h)^2]$$

Neglecting the errors in writing $\frac{1}{N-1} = \frac{1}{N}$ and $\frac{1}{N_h-1} = \frac{1}{N_h}$ $\forall$ h (h = 1, ...L), we have

$$V(\bar{Y}_R) - V_6 \approx (\frac{1}{m} - \frac{1}{n}) \frac{1}{N} [\sum_h \sum_j (y_{hj} - R x_{hj})^2$$

$$- \sum_h \frac{1}{N_h} \sum_j p'_{hj} (\frac{y_{hj}}{p'_{hj}} - Y_h)^2].$$

Then, under M_2'',

$V(\bar{Y}_R) - V_6$ >, = or < 0 according as g >, = or < 1.

(v) Strategy V Vs strategy G

We have

$$V(\bar{Y}_R) - V_5 = (\frac{1}{m} - \frac{1}{n}) [\frac{1}{N-1} \sum_h \sum_j (y_{hj} - R x_{hj})^2$$

$$- \sum \frac{W_h}{N_h-1} \sum_j (y_{hj} - R_h x_{hj})^2]$$

$$= (\frac{1}{m} - \frac{1}{n}) [\frac{1}{N-1} \sum_h (N_h-1)$$

$$\{ S_{yh}^2 + R^2 S_{xh}^2 - 2 \rho_h R S_{xh} S_{yh} \}$$

$$- \sum_h W_h \{ S_{yh}^2 + R_h^2 S_{xh}^2 - 2\rho_h R_h S_{xh} S_{yh} \}]$$

Now assuming $\frac{N_h-1}{N-1} = W_h$ we have

$$V(\bar{Y}_R) - V_5 = (\frac{1}{m} - \frac{1}{n}) \sum_h W_h \{ (R^2 - R_h^2) S_{xh}^2 - 2 \rho_h (R - R_h) S_{xh} S_{yh} \}$$

The positivity and negativity of the quantity inside the square bracket respectively are known to imply superiority and inferiority (in terms of efficiency through variances) of separate ratio estimator over combined ratio estimator. Thus, strategy V is superior, inferior or equivalent to the strategy G according as separate ratio estimator is superior, inferior or equivalent to the combined ratio estimator.

(vi) Strategy Vi Vs strategy F.

We have

$$V_6 - V(\bar{Y}_{RHC}) = (\frac{1}{m} - \frac{1}{n}) \frac{1}{N} [\sum_h \frac{1}{N_h - 1} \sum_j p'_{hj} (\frac{y_{hj}}{p'_{hj}} - Y_h)^2$$

$$- \frac{1}{N-1} (\sum_h \sum_j \frac{y_{hj}^2}{p_{hj}} - Y^2)]$$

$$(\text{where } p_{hj} = \frac{x_{hj}}{X})$$

Neglecting the error in writing $\frac{1}{N_h - 1} = \frac{1}{N_h}$ $\forall$ h and $\frac{1}{N-1} = \frac{1}{N}$ we have, under M_2 "

$$V_6 - V(\bar{Y}_{RHC}) \approx (\frac{1}{m} - \frac{1}{n}) \frac{1}{N} [\sum_h \frac{1}{N_h} \sum_j p'_{hj} (\frac{y_{hj}}{p'_{hj}} - Y_h)^2$$

$$- \frac{1}{N} \sum_h \sum_j p_{hj} (\frac{y_{hj}}{p_{hj}} - Y)^2]$$

$$= \left(\frac{1}{m} - \frac{1}{n}\right) \frac{1}{N} \sigma^2 \sum_h \sum_j x_{hj}^{g-1} (\bar{X}_h - \bar{X})$$

$$= \left(\frac{1}{m} - \frac{1}{n}\right) \frac{\sigma^2}{N^2} \frac{1}{2} \sum_{j \neq h'} \sum (\bar{X}_h - \bar{X}_{h'}) \sum_{i \neq j} \sum (x_{hi}^{g-1} - x_{h'j}^{g-1})$$

Now if we assume

$0 < x_{11} \leq x_{12} \leq \ldots\ldots\ldots \leq x_{L N_L}$ (call this the assumption Γ) , then $V_6 - V(\bar{Y}_{RHC})$ >, = or < 0 according as g > = or < 1 provided all x_{hj}'s are not equal. If all x_{hj}'s are equal then $V_6 = V(\bar{Y}_{RHC})$.

If we assume (further) the model M_2' with $\frac{1}{N_h - 1} \approx \frac{1}{N_h}$ $\forall$ h and $\frac{1}{N-1} \approx \frac{1}{N}$ we get

$$V_6 - V(\bar{Y}_{RHC}) \approx \left(\frac{1}{m} - \frac{1}{n}\right) \frac{1}{N} f_2(g)$$

where

$$f_2(g) = \sigma^2 \sum_h \sum_j x_{hj}^{g-1} (\bar{x}_h - \bar{X})$$

$$- \frac{1}{N} \sum_{h<h'} \sum X_h X_{h'} (\beta_h - \beta_{h'})^2$$

Here $f_2(1) < 0$ and $f_2(g)$ is an increasing function of g (for g > 1). Thus under the assumption Γ with the approximations mentioned above there exists a g_0 (> 1), such that $V_6 - V(\bar{Y}_{RHC}) \gtrless$ according as $g \gtrless g_0$. But this final result is of little consequence because one can hardly check this condition on $f_2(g)$ because $f_2(g)$ involves too many unknowns.

APPENDIX 4.1

We consider below the data recorded at Agrarian Research Station, Seva Bharati, P.O. Kapgari, Dist - Midnapur, West Bengal, relating to the yields of a rabi crop called Dolichos Biflorus. Eight plants were considered and the yields of the crops were weighted in kilograms as q, x and y respectively after 70, 80 and 85 days following the data of sowing. Writing $p = q_i / \sum_i q_i$ the data are presented in the table below

TABLE 4.2 SHOWING YIELDS (IN KILOGRAM) OF A CROP PER PLANT ON THREE DIFFERENT DATES OF HARVESTING

Serial No. of plants	p	x	y
1	0.1315	.350	.250
2	0.1081	.150	.500
3	0.1622	.700	1.000
4	0.1081	.500	1.700
5	0.1351	.300	.200
6	0.1081	.200	.500
7	0.1081	1.100	.100
8	0.1351	.400	1.200

Now considering samples of sizes n = 2, 3, 4 we calculate the values of $f_n(y)$ based on p_i's and y_i's and $f_n(z)$ based on p_i's and $(y_i - x_i) = z_i$ and present them in the following table.

TABLE 4.3 SHOWING THE CLOSE AGREEMENT AMONG $f_n(y)$ and $f_n(z)$ VALUES FOR n = 2, 3, 4

sample size n	$f_n(y)$	$f_n(z)$	$f_n(y) - f_n(z)$
2	.4367	.4402	.0035
3	.2490	.2535	.0045
4	.1351	.1603	.0052

CHAPTER - 5

A FEW STRATEGIES FOR TWO-STAGE SAMPLING OVER TWO OCCASIONS

5.1 SUMMARY

In this chapter we propose a general sampling strategy for estimating a finite population total for the current occasion based on a two-stage sampling scheme applied on two occasions. This is in generalization of the earlier sampling schemes considered among others by Singh (1968), Abraham, Khosla and Kathuria (1969), Singh and Kathuria (1969) and Kathuria (1975). We also consider a few particular cases of the general scheme and examine relative efficiencies of several comparable strategies. Some of our proposed strategies are found to fare better than their predecessors. In investigating efficiencies Avadhani and Sukhatme's (1970) Model M_2 is assumed towards the end.

5.2 A BRIEF REVIEW OF EARLIER WORKS

The uni-stage sampling strategies over successive occasions were extended to multi-stages by Tikkiwal (1958, 1964, 1965), Singh (1968), Abraham, Khosla and Kathuria (1969), Singh and Kathuria (1969) and Kathuria (1975)

STRATEGY A . due to Singh (1968)

Singh (1968) considered the following two-stage sampling scheme over two occasions .

On the first occasion n f.s.u's are drawn by SRSWOR from a population consisting of N f.s.u's each of which consists of M s.s.u's and from each of the selected f.s.u's a sub-sample of m s.s.u's is selected by SRSWOR method. On the second occasion a simple random sub-sample of $n\lambda$ (such that $n\lambda$ is an integer and $0 \leq \lambda \leq 1$) f.s.u's out of the n f.s.u's selected on the first occasion is retained along with the sample of s.s.u's drawn from them on the first occasion and this is supplemented by an SRSWOR of $n\mu$ ($\mu = 1-\lambda$) f.s.u's selected from the f.s.u's not chosen on the first occasion from each of which an SRSWOR of m.s.u's is chosen.

Singh (1968) considered the following estimator for the population mean $\bar{Y}$ of the s.s.u's for the second occasion :

$$\hat{\bar{Y}}_1 = a\,\bar{y}' + b\,\bar{x}' + c\,\bar{x}'' + d\,\bar{y}'' \qquad (5.2.1)$$

where

$\bar{y}'(\bar{x}')$ = mean per s.s.u on the 2nd (1st) occasion for the λmn units which are common to the two occasions

$\bar{y}''$ = mean per s.s.u on the 2nd occasion from μnm units which are based on the 2nd sample only

$\bar{x}''$ = mean per s.s.u on the first occasion from the μnm units which are in the first sample only.

Here the constants a, b, c and d are so chosen that $\hat{\bar{Y}}_1$ becomes unbiased for $\bar{Y}$ and $V(\hat{\bar{Y}}_1)$ attains its minimum value. Finally the optimum value of λ and the corresponding minimum value of $V(\hat{\bar{Y}}_1)$ are determined.

STRATEGY B . due to Singh and Kathuria (1969)

Singh and Kathuria (1969) considered the following sampling scheme :

The sample on the first occasion is selected in the same manner as in strategy A. On the second occasion all f.s.u's chosen in the preceding occasion are retained but from each f.s.u only a fraction α of the s.s.u's chosen earlier is retained and a fraction $\beta = 1-\alpha$ of s.s.u's is chosen afresh from among the remaining (M-m) s.s.u's not selected on the first occasion.

They considered the following unbiased estimator for the population mean $\bar{Y}$ of the s.s.u's on the second occasion viz,

$$\hat{\bar{Y}}_2 = a\,\bar{y}' + (1-a)\,\bar{y}''' + b(\bar{x}''' - \bar{x}) \qquad (5.2.2)$$

where

$\bar{y}'$ = mean per s.s.u on the 2nd occasion based on $nm\alpha$ units common on the two occasions

$\bar{y}'''$ = mean per s.s.u on the 2nd occasion based on $nm\beta$ units selected afresh.

$\bar{x}''$ = mean per s.s.u on the first occasion based on $nm\beta$ units not common with the 2nd occasion

$\bar{x}$ = mean per s.s.u on the 1st occasion based on mn units.

They selected the constants a and b such that $\hat{\bar{Y}}_2$ has the least variance.

STRATEGIES C AND D . due to Abraham, Khosla and Kathuria (1969)

The strategies C and D based on the sampling scheme are as follows :

The sample on the first occasion is taken in the same way as in strategies A and B. On the second occasion a matched sample of $n\lambda$ ($0 \leq \lambda \leq 1$) f.s.u's is selected from the n f.s.u's selected on the first occasion and from each of the selected f.s.u's a sub-sample of $m\alpha$ ($0 \leq \alpha \leq 1$) s.s.u's is selected from the m s.s.u's selected on the first occasion and a fresh $m\beta$ (with $\beta = 1-\alpha$) s.s.u's are selected from the remaining M-m s.s.u's not selected on the first occasion. Finally an un-matched sample of $n\mu$ f.s.u's is selected from N-n f.s.u's that were not selected on the first occasion and from each of these selected f.s.u's a simple random sub-sample of m f.s.u's is taken.

Abraham, Khosla and Kathuria (1969) considered the following two estimators I and II for the estimation of the population mean on the second occasion. The estimators I and II based on the above sampling scheme constitute the strategies C and D respectively.

Estimator I is taken as

$$\hat{\bar{Y}}_3 = a\,\bar{x}' + b\,\bar{x}'' - (a+b)\,\bar{x}''' + d\,\bar{y}' + e\,\bar{y}'' + (1-d-e)\,\bar{y}''' \qquad (5.2.3)$$

here

$\bar{y}'(\bar{x}')$ = mean per s.s.u on the 2nd (1st) occasion based on $n\lambda m\alpha$ s.s.u's which are common to the two occasions

$\bar{y}''(\bar{x}'')$ = mean per s.s.u on the 2nd (1st) occasion based on $n\lambda\ m\beta$ s.s.u's which are not common to the two occasions but belong to the common s.s.u's on both occasions.

$\bar{y}'''$ = mean per s.s.u on the 2nd occasion based on $n\mu m$ units selected afresh on the 2nd occasion.

$\bar{x}'''$ = mean per s.s.u on the first occasion based on $n\mu m$ units which are in the sample on the first occasion only.

he constants a, b, d and e are so chosen that $V(\hat{\bar{Y}}_3)$ attains he minimum value.

stimator II

$$\hat{\bar{Y}}_4 = P\ \bar{y}_c + (1-P)\ \bar{y}''' + P'\ (\bar{x}_c - \bar{x}''') \qquad (5.2.4)$$

here

$$\bar{y}_c = k\,[\,\bar{y}' + \rho_2(\bar{x}_c - \bar{x}')\,] + (1-k)\ \bar{y}''$$

with

$\bar{x}_c$ = mean per s.s.u on the first occasion, based on $n\lambda$ f.s.u's common to the two occasions.

$$\rho_2 = \frac{\sum_i \sum_j (y_{ij}-\bar{Y}_i)(x_{ij}-\bar{X}_i)}{S_w^2}$$

y_{ij} (x_{ij}) = value of y(x) for the jth s.s.u of the ith f.s.u (j = 1, ...N, i = 1, ..., N) on the 2nd (1st) occasion.

$$\bar{Y}_i = \frac{1}{M} \sum_j y_{ij} \quad \text{and} \quad \bar{X}_i = \frac{1}{M} \sum_j x_{ij}$$

S_w^2 = the mean square between the s.s.u's assumed to be the same (for x and y) on each of the two occasions.

Here k is chosen so as to minimize the variance of $\bar{y}_c$ and the constants P and P' are so chosen that the variance $\hat{\bar{Y}}_4$ is the minimum.

It may be mentioned in passing that in all the strategies considered so far the optimal estimators are usable only when some knowledge about specific population parameters is available. In the strategies we are going to introduce for further investigations the same situation will also be seen to arise as is common with all sampling strategies for two occasions available so far in the literature.

STRATEGIES E AND F : due to Kathuria (1975)

Strategies E and F due to Kathuria (1975) both based on the same sampling scheme is described below.

On the first occasion a sample S_1 of n f.s.u's is selected following PPSWR method using some normed size-measures p_i's assumed available. From each of the selected f.s.u's an independent sub-sample of size m is taken by SRSWOR (if the ith f.s.u U_i appears λ_i times, then from it λ_i independent sub-samples each consisting of m s.s.u's are taken). On the 2nd occasion a sub-sample S_1' of $n\lambda$ $(0 \leq \lambda \leq 1)$ f.s.u's is selected from S_1 by SRSWOR. From each of the selected $n\lambda$ matched f.s.u's a sub-sample of $m\alpha$ s.s.u's is taken from the m s.s.u's selected earlier and in addition to that a fresh sub-sample of $m\beta$ (with $\beta = 1-\alpha$) is selected from the remaining M-m s.s.u's from the respective matched f.s.u's by SRSWOR. Finally an independent sample S_{2u} of $n\mu$ f.s.u's is selected from the entire population by PPSWR method using the previous normed size measures p_i's. From each of these selected f.s.u's an independent sample of m s.s.u's is taken by SRSWOR repeating the process as many times as the f.s.u is selected in the original sample by PPSWR method.

Kathuria (1975) considered the following two timators III and IV for the estimation of the population an $\bar{Y}$ on the second occasion. The estimators III and IV sed on the above sampling scheme constitute the strategies and F respectively.

estimator III is

$$\hat{\bar{Y}}_5 = k\,(\bar{t}_1 + \bar{t}_{12} - \bar{t}_{\lambda 1}) + (1-k)\,\bar{t}_2'' \qquad (5.2.5)$$

ere

$$\bar{t}_{12} = Q\,(\bar{t}_{\lambda 1} + \bar{t}_2' - \bar{t}_1') + (1-Q)\,\bar{t}_2''$$

th

$$\bar{t}_1 = \frac{1}{nN} \sum_i \sum_r L_i(r) \frac{\bar{x}_i(r)}{p_i}$$

$$\bar{t}_{\lambda 1} = \frac{1}{n\lambda N} \sum_i \sum_r L_i'(r) \frac{\bar{x}_i(r)}{p_i}$$

$$\bar{t}_1' = \frac{1}{n\lambda N} \sum_i \sum_r L_i'(r) \frac{\bar{x}_i'(r)}{p_i}$$

$$\bar{t}_2' = \frac{1}{n\lambda N} \sum_i \sum_r L_i'(r) \frac{\bar{y}_i'(r)}{p_i}$$

$$\bar{t}_2'' = \frac{1}{n\lambda N} \sum_i \sum_r L_i'(r) \frac{\bar{y}_i''(r)}{p_i}$$

$$\bar{t}_2''' = \frac{1}{n\mu N} \sum_i \sum_r L_i^*(r) \frac{\bar{y}_i'''(r)}{p_i}$$

$L_i(r) = 1$ if rth draw produces ith unit in S_1
$\quad = 0$ otherwise.

$L_i'(r) = 1$ if rth draw produces ith unit in S_1'
$\quad = 0$ otherwise.

$L_i''(r) = 1$ if rth draw produces ith unit in S_{2u}
$\quad = 0$ otherwise.

$\bar{x}_i(r)$ = mean of the m s.s.u's sub-sampled from ith f.s.u selected on the rth draw in S_1.

$\bar{y}_i'(r)(x_i'(r))$ = mean per s.s.u on 2nd (1st) occasion based on mα matched s.s.u's which were selected from the ith matched f.s.u appearing at the rth draw in selecting S_1'.

$\bar{y}_i''(r)$ = mean per s.s.u on the 2nd occasion based on mβ un-matched s.s.u's, selected from the ith matched f.s.u that appeared at rth draw in S_1'.

$\bar{y}_i'''(r)$ = mean per s.s.u based on sub-sampling from ith f.s.u if it is selected at rth draw in S_{2u}.

re Q is a constant chosen to minimize Var($\bar{t}_{12}$), the sulting $\bar{t}_{12}$ with the minimum variance is used in $\hat{\bar{Y}}_5$ and is a constant chosen to minimize Var ($\hat{\bar{Y}}_5$).

Estimator IV is

$$\hat{\bar{Y}}_6 = P\{\bar{t}_2 + Q(\bar{t}_1 - \bar{t}_{\lambda 1})\} + (1-P)\ \bar{t}_2''' \qquad (5.2.6)$$

with

$$\bar{t}_2 = g\{\bar{t}_2' + h(\bar{t}_{\lambda 1} - \bar{t}_1')\} + (1-g)\ \bar{t}_2''$$

Here the constants g and h are so chosen that the variance of $\bar{t}_2$ becomes the minimum and putting the resulting optimal value of $\bar{t}_2$ in $\hat{\bar{Y}}_6$ the value of P and Q are chosen such that the variance of $\hat{\bar{Y}}_6$ attains the minimum.

We present now a slightly more general sampling scheme to follow on two occasions in two stages and consider a few special cases.

5.3 SAMPLING SCHEMES OVER TWO OCCASIONS AND NOTATIONS

From a finite population of N first-stage units (f.s.u's) the ith of which consists of M_i second-stage units (s.s.u's) on the first occasion a sample S_1 of n f.s.u's is chosen following a suitable sampling scheme. If the ith f.s.u is selected a sample S_{1i} of m_i s.s.u's is selected again by a suitable method from the M_i s.s.u's in it giving altogether a sample $S^{(1)}$ of s.s.u's. On the second occasion a matched sample S_1' of $n\lambda$ f.s.u's $(0 \leq n\lambda \leq n)$ is suitably selected from among the n f.s.u's selected earlier and a sample S_{2u} (called un-matched sample) of $n\mu$ (such that $\lambda+\mu = 1$) f.s.u's is chosen suitably either from all the N f.s.u's or from the (N-n) left-over f.s.u's. Again on the 2nd occasion from the ith f.s.u selected earlier a matched sub-sample S_{2mi} of $m_i \alpha_i$ s.s.u's is suitably selected $(1 \leq m_i\alpha_i \leq m_i \;\; \forall\, i)$ this being done for each of the m_i s.s.u's selected earlier and a fresh sample S_{2u_i} of $m_i\beta_i$ (with $\alpha_i+\beta_i = 1 \;\; \forall\, i$) s.s.u's is chosen either from all the M_i s.s.u's in the ith f.s.u earlier chosen or from the $(M_i - m_i)$ left-out s.s.u's. Also each of the un-matched $n\mu$ f.s.u's selected on the second occasion is sub-sampled taking m_i s.s.u's suitably from the i-th such f.s.u in it containing M_i s.s.u's, giving a sample S'_{2u_i}.

The following notations will be used in what follows:

$y_{ij}(x_{ij})$ is the value for the 2nd (1st) occasion on the jth s.s.u of the ith f.s.u $(j = 1,\ldots,M_i, \;\; i = 1,\ldots,N)$.

$$Y_i = \sum_j y_{ij}, \quad X_i = \sum_j x_{ij}, \quad \bar{Y}_i = \frac{Y_i}{M_i}, \quad \bar{X}_i = \frac{X_i}{M_i},$$

$$R_i = \frac{Y_i}{X_i}, \quad Y = \sum_i Y_i, \quad \bar{Y} = \frac{Y}{N}, \quad \bar{\bar{Y}} = \frac{Y}{M_o}, \quad M_o = \sum_i M_i,$$

$\bar{y}_i'$, $\bar{y}_i''$, $\bar{y}_i'''$ are the sample means per s.s.u's on the second occasion based on the samples S_{2m_i}, S_{2u_i} and S'_{2u_i} respectively,

$\bar{x}_i'$ and $\bar{x}_i$ are the sample means per s.s.u on the first occasion based on the samples S_{2m_i} and S_{1i}'s respectively.

$$S^2_{y_i} = \frac{1}{(M_i-1)} \sum_j (Y_{ij} - \bar{Y}_i)^2 ,$$

$$S^2_{x_i} = \frac{1}{(M_i-1)} \sum_j (X_{ij} - \bar{X}_i)^2$$

$$S^2_{by} = \frac{1}{(N-1)} \sum_i (Y_i - \bar{Y})^2$$

$$S^2_{bx} = \frac{1}{(N-1)} \sum_i (X_i - \bar{X})^2$$

$$S^2_{wy} = \frac{1}{N} \sum_i M_i^2 \left(\frac{1}{m_i} - \frac{1}{M_i}\right) S^2_{y_i}$$

$$S^2_{wx} = \frac{1}{N} \sum_i M_i^2 \left(\frac{1}{m_i} - \frac{1}{M_i}\right) S^2_{x_i}$$

$$S_{bxy} = \frac{1}{(N-1)} \sum_i (X_i - \bar{X})(Y_i - \bar{Y})$$

$$S_{wxy} = \frac{1}{N} \sum_i \left(\frac{1}{m_i} - \frac{1}{M_i}\right) \rho_i \, S_{x_i} \, S_{y_i}$$

$$\rho_i = \frac{1}{M_i - 1} \sum_j (y_{ij} - \bar{Y}_i)(x_{ij} - \bar{X}_i)/(Sx_i \; Sy_i)$$

$$\delta_i = \frac{1}{M_i} \sum_j \frac{1}{p_{ij}} (y_{ij} - R_i \; x_{ij})^2 / \sum_j (y_{ij} - \bar{Y}_i)^2$$

$$\sigma_i^2(y) = \sum_j p_{ij} \left(\frac{y_{ij}}{p_{ij}} - Y_i\right)^2$$

$$\sigma_i^2(x) = \sum_j p_{ij} \left(\frac{x_{ij}}{p_{ij}} - X_i\right)^2$$

$$V_{pps}(y) = \sum_i p_i \left(\frac{Y_i}{p_i} - Y\right)^2$$

$$V_{pps}(x) = \sum_i p_i \left(\frac{X_i}{p_i} - X\right)^2$$

$$\rho = \sum_i p_i \left(\frac{Y_i}{p_i} - Y\right)\left(\frac{X_i}{p_i} - X\right) / \left[V_{pps}(y) . V_{pps}(x)\right]^{1/2}$$

p_{ij} = normed size-measure associated with jth s.s.u in the ith f.s.u. ($j = 1, \ldots, M_i$, $i = 1, \ldots, N$),

p_i = normed size-measure associated with ith f.s.u ($i = 1, \ldots, N$),

E_1 = unconditional expectation for the selection of sample S_1 of n f.s.u's on the first occasion,

E_2 = conditional expectation over the selection of s.s.u's selected from the respective f.s.

E_3 = conditional expectation over selection of sub-samples of f.s.u's chosen from samples of f.s.u's selected earlier,

E_4 = conditional expectation over the selection of S_{2m_i}'s from fixed S_{1i} and S_{2u_i} for fixed i (i = 1,...,N).

Also, v_1, v_2, v_3 and v_4 are similarly defined operators for variances. By $E_{ijk\ell}$, $V_{ijk\ell}$ we shall denote expectations and variances calculated over different stages of imposing conditions (i, j, k, ℓ = 1, 2, 3, 4 of which anyone may be suppressed with obvious implications).

5.4 A GENERAL LINEAR UNBIASED ESTIMATOR AND ITS VARIANCE

Let T_1 be an unbiased estimator for Y based on the s.s.u's selected from the matched f.s.u's only of the form

$$T_1 = \sum_i bs_i \; t_i$$

where

$$bs_i \neq 0 \quad \text{if the ith f.s.u is in the matched sample } S_1' \text{ of } n\lambda \text{ f.s.u's (writing S for } S_1' \text{ for simplicity)}$$
$$= 0 \quad \text{otherwise,}$$

such that bs_i's are independent of y_{ij}'s and x_{ij}'s and t_i is a function of y_{ij}'s for the units in the matched sample and x_{ij}'s for the units in the initial sample but independent of bs_i's.

First we consider a lemma we shall use later.

Lemma 5.1

If $E_2E_4(t_i) = Y_i \quad \forall\, i \quad (= 1, \ldots, N)$, and $E_1E_3(bs_i) = 1$ $\forall$ s, i (= 1, ..., N) then the variance of T_1 is

$$V(T_1) = V_{13}(\sum_i bs_i \; Y_i) + E_{13}(\sum_i bs_i^{\,2} \; \sigma_i^{\,2})$$

where

$$\sigma_i^{\,2} = V_{24}(t_i)$$

Proof

$$V(T_1) = V_{1234}\,(T_1)$$

$$= V_1E_2E_3E_4(T_1) + E_1V_2E_3E_4(T_1)$$

$$+ E_1E_2V_3E_4(T_1) + E_1E_2E_3V_4(T_1)$$

$$= V_1E_3(\sum_i bs_i \; Y_i) + E_1 \sum_i \{E_3(bs_i)\}^2 \; V_2E_4(t_i)$$

$$+ E_1 \sum_i V_3(bs_i) \; E_2\, \{E_4(t_i)\}^2$$

$$+ E_1 \sum_{i \neq j}\sum \mathrm{Cov}_3\,(bs_i, bs_j) \; E_2\{E_4(t_i) \; E_4(t_j)\}$$

$$+ E_1E_3 \sum_i (bs_i)^2 \; E_2V_4(t_i)$$

$$= V_1 E_3 (\sum_i bs_i Y_i) + E_1 \sum_i E_3 (bs_i)^2 V_2 E_4 (t_i)$$

$$+ E_1 \sum_i V_3 (bs_i) Y_i^2 + E_1 \sum_{i \neq j} \sum \mathrm{Cov}_3 (bs_i, bs_j) Y_i Y_j$$

$$+ E_1 E_3 \sum_i bs_i^2 E_2 V_4 (t_i)$$

$$= V_1 E_3 (\sum_i bs_i Y_i) + E_1 V_3 (\sum_i bs_i Y_i)$$

$$+ E_1 E_3 \sum_i bs_i^2 \{ V_2 E_4 (t_i) + E_2 V_4 (t_i) \}$$

$$= V_{13} (\sum_i bs_i Y_i) + E_{13} \sum_i (bs_i)^2 V_{24} (t_i)$$

$$= V_{13} (\sum_i bs_i Y_i) + E_{13} \sum_i (bs_i)^2 \sigma_i^2$$

Our general strategy for estimating Y from a le chosen following a sampling scheme outlined earlier s follows. First we consider an unbiased estimator T_1 bove for Y, based on $n\lambda$ matched f.s.u's. Then we use ppropriate estimator T_2 for Y based only on the unhed sample selected on the 2nd occasion only. Next ind an estimator T_3 (a function of x_{ij}'s only) based ll the f.s.u's sampled on the first occasion along with .s.u's retained in the second occasion so chosen that conditional expectation given the initial sample is zero.

Our final composite estimator for Y is then

$$T = \phi\, T_1 + (1-\phi)\, T_2 + \psi\, T_3$$

Unlike Raj (1968, Para 1, 7.92, p 156) we are considering here a 'truly' finite population set-up. So we cannot choose T_1, T_2, T_3 as best linear unbiased estimators for their expectations. They are to be chosen in manners as are found suitable from efficiency considerations. We propose the form T for the sake of efficiency by trying to take account of correlations among T_1, T_2, T_3. So, first we choose ϕ, ψ to minimize Var(T) for fixed T_1, T_2, T_3 supposed to be given. Later, we derive particular forms of the formulae for optimal values of ϕ, ψ and Var(T) for different values of T_1, T_2, T_3 as are usually made in the literature for different sampling schemes.

Now the variance of T for a given λ is

$$V(T|\lambda) \text{ (say) } = \phi^2 V_1 + (1-\phi)^2 V_2 + \psi^2 V_3 + 2\phi(1-\phi) V_{12} + 2\phi\,\psi\, V_{13} \qquad (5.4.1)$$

[where we write $V(T_i) = V_i$ for $i = 1,2,3$

$\text{Cov}(T_i, T_j) = V_{ij}$ for $i \neq j = 1,2,3$

and note $V_{23} = 0$ since $E(T_3) = 0 = E(T_2 T_3)$].

Now solving $\frac{\partial V}{\partial \phi} V(T|\lambda) = 0$ and $\frac{\partial}{\partial \psi} V(T|\lambda) = 0$ we get the optimum values of ϕ and ψ as

$$\phi_{opt} = \phi_o \text{ (say)} = (V_2 - V_{12})[V_1 + V_2 - 2V_{12} - \frac{V_{13}^2}{V_3}]^{-1}$$
$$\psi_{opt} = \psi_o \text{ (say)} = -\phi_o \frac{V_{13}}{V_3} \qquad (5.4.2)$$

utting the values of ϕ_o and ψ_o in (5.4.1) we get

$$V_{opt}(T|\lambda) = V_2 - \phi_o (V_2 - V_{12})$$
$$= V_2 - (V_2 - V_{12})^2 [V_1 + V_2 - 2V_{12} - \frac{V_{13}^2}{V_3}]^{-1} \qquad (5.4.3)$$

An optimal value of λ, of course, may be found o as to minimize the value of $V_{opt}(T|\lambda)$. Here we note that $_{opt}(T|\lambda)$ is an increasing function of V_1. So, other things remaining nchanged, one should choose a sampling strategy for which σ_1^2 is optimally ontrolled. To do so one must apply an appropriate uni-stage sampling design ver two occasions. Many attempts at controlling the agnitude of σ_1^2 have been made in the literature and the emarkable among them are due to Kulldorff (1963), Raj (1965), athak and Rao (1967), Ghangurde and Rao (1969), Singh (1972), otai (1974), Chaudhuri and Arnab (1977).

In the next section we discuss methods of utilizing different ways of controlling the magnitude of σ_i^2 so as to obtain useful two-stage sampling designs to be employed on two occasions and present the expression for minimum variance of $V_{opt}(T|\lambda) = V_{min}(T)$ (say) under different sampling strategies.

5.5 A FEW SPECIFIC SAMPLING STRATEGIES AND THE CORRESPONDING VARIANCES OF ESTIMATORS

STRATEGY I

Here S_1 and S_{1i}'s $(i = 1, \ldots, n)$, S_1', S_{2mi}'s are all selected by SRSWOR method. The S_{2ui}'s are selected from among $(M_j - m_j)$ s.s.u's not selected on the first occasion and S_{2u} is selected by SRSWOR from $(N-n)$ f.s.u's not selected on the first occasion and each f.s.u is sub-sampled independently by SRSWOR.

For the estimator

$$T = \phi \, T_1 + (1-\phi) \, T_2 + \psi \, T_3$$

for Y in this case we have

$$\left.\begin{aligned} T_1 &= \frac{N}{n\lambda} \sum_{S_1'} t_i \quad , \quad T_2 = \frac{N}{n\mu} \sum_{S_{2u}} M_i \, \bar{y}_i''' \\ \text{and} \quad T_3 &= N \left(\frac{1}{n\lambda} \sum_{S_1'} M_i \, \bar{x}_i - \frac{1}{n} \sum_{S_1} M_i \, \bar{x}_i \right) \end{aligned}\right\} \quad (5.5.1)$$

with

$$t_i = M_i \{ a_i \bar{y}_i' + (1-a_i)\bar{y}_i'' + c_i(\bar{x}_i' - \bar{x}_i) \}.$$

Here a_i's and c_i's are so chosen that $V_{24}(t_i) = \sigma_i^2$ becomes the least. By a little algebra it can be shown (vide Appendix 5.1 on p.217) that the optimum value of $\sigma_i^2 = \sigma_{io}^2$ (say) is

$$\sigma_{io}^2 = M_i^2 \left[\frac{1 + (1-\rho_i^2)^{1/2}}{2m_i} - \frac{1}{M_i} \right] S_{yi}^2$$

and corresponding optimum values of a_i, c_i and α_i's are

$$a_i = a_{io} \text{ (say) } = \frac{1}{2}$$

$$c_i = c_{io} \text{ (say) } = - \frac{1}{2} \rho_i \frac{S_{yi}}{S_{xi}} = - \frac{1}{2} b_i \text{ (say)}$$

$$\text{(where } b_i = \rho_i \frac{S_{yi}}{S_{xi}} \text{)}$$

and $\alpha_i = \alpha_{io} \text{ (say) } = (1-\rho_i^2)^{1/2} \{1 + (1-\rho_i^2)^{1/2}\}^{-1}$

Substituting these optimum values (assuming parameters to be all known) in t_i one gets

$$t_i = t_{io} \text{ (say) } = M_i \left[\frac{1}{2}\{ \bar{y}_i' + b_i(\bar{x}_i - \bar{x}_i') \} + \frac{1}{2} \bar{y}_i'' \right] \qquad (5.5.2)$$

If, in the above one replace b_i, the population regression coefficient, by its sample analogue then one finds t_{io} to coincide with Kulldorff's (1963) estimation for uni-stage sampling on two occasions - a point one may note incidentally.

Now, applying the lemma (5.1) we get

$$V_1 = V(T_1) = N^2 \, v_{13} \left(\frac{1}{n\lambda} \sum_{S_1'} Y_i\right) + \frac{N^2}{(n\lambda)^2} E_{13} \sum_{S_1'} \sigma_{io}^2$$

$$= N^2 \left(\frac{1}{n\lambda} - \frac{1}{N}\right) S_{by}^2 + \frac{N}{n\lambda} \sum_i \sigma_{io}^2 \qquad (5.5.3)$$

Also,

$$V_2 = V(T_2) = v_1 E_2' E_3'(T_2) + E_1 v_2' E_3'(T_2) + E_1 E_2' v_3'(T_2)$$

where

E_2' = conditional expectation over the selection of S_{2u} from the N-n f.s.u's not selected on the first occasion.

E_3' = conditional expectation over the selection of s.s.u's from the respective f.s.u's belonging to S_{2u}.

v_2' and v_3' are similarly defined operators for variances.

Hence

$$V_2 = v_1 E_2'\left(\frac{N}{n\mu} \sum_{S_{2u}} Y_i\right) + E_1 v_2'\left(\frac{N}{n\mu} \sum_{S_{2u}} Y_i\right)$$

$$+ E_1 E_2'\left(\frac{N^2}{n^2 \mu^2} \sum_{S_{2u}} M_i^2\left(\frac{1}{m_i} - \frac{1}{M_i}\right) s_{yi}^2\right)$$

$$= N^2 \left[\left(\frac{1}{N-n} - \frac{1}{N}\right) S_{by}^2 + \left(\frac{1}{n\mu} - \frac{1}{N-n}\right) S_{by}^2 + \frac{S_{wy}^2}{n\mu}\right]$$

$$= N^2 \left[\left(\frac{1}{n\mu} - \frac{1}{N}\right) S_{by}^2 + \frac{S_{wy}^2}{n\mu}\right] \qquad (5.5.4)$$

Now

$$V_3 = N^2 \; V\left(\frac{1}{n\lambda} \sum_{S_1'} M_i \bar{x}_i - \frac{1}{n} \sum_{S_1} M_i \bar{x}_i\right)$$

$$= N^2 \; E_1 E_2 v_3 \left(\frac{1}{n\lambda} \sum_{S_1'} M_i \bar{x}_i - \frac{1}{n} \sum_{S_1} M_i \bar{x}_i\right)$$

$$\left[\text{since } \; E_3\left(\frac{1}{n\lambda} \sum_{S_1'} M_i \bar{x}_i - \frac{1}{n} \sum_{S_1} M_i \bar{x}_i\right) = 0\right]$$

$$= N^2 \; E_1 E_2 \left(\frac{1}{n\lambda} - \frac{1}{n}\right) \frac{1}{n-1} \sum_{S_1} \left(M_i \bar{x}_i - \frac{1}{n} \sum_{S_1} M_i \bar{x}_i\right)^2$$

$$= N^2 \left(\frac{1}{n\lambda} - \frac{1}{n}\right) E_1 \left(\frac{1}{n} \sum_{S_1} M_i^2 \left(\frac{1}{m_i} - \frac{1}{M_i}\right) s_{xi}^2\right.$$

$$\left. - \frac{1}{n-1} \sum_{S_1} (X_i - \bar{X}_{S1})^2\right)$$

$$\left(\text{where } \quad \bar{X}_{S1} = \frac{1}{n} \sum_{S_1} X_i\right)$$

$$= N^2 \left(\frac{1}{n\lambda} - \frac{1}{n}\right)(S_{wx}^2 + S_{bx}^2) \qquad (5.5.5)$$

$$V_{12} = \text{Cov}\ (T_1, T_2) = E\ T_1 T_2 - Y^2$$

$$= N^2\ E\ (\frac{1}{n} \sum_{S_1} Y_i \ \frac{1}{n\mu} \sum_{S_{2u}} Y_i) - Y^2$$

$$= N^2\ E_1\ \{\frac{1}{n} \sum_{S_1} Y_i \ \frac{Y - \sum_{i \in S_1} Y_i}{N-n}\} - Y^2$$

$$= - N\ S_{by}^{\ 2} \qquad (5.5.6)$$

$$V_{13} = \text{Cov}\ (T_1, T_3)$$

$$= \text{Cov}_1\ [E_2 E_3 E_4 (T_1),\ E_2 E_3 E_4 (T_3)]$$

$$+ E_1\ \text{Cov}_2\ [E_3 E_4 (T_1),\ E_3 E_4 (T_3)]$$

$$+ E_1 E_2\ \text{Cov}_3\ [E_4 (T_1),\ E_4 (T_3)]$$

$$+ E_1 E_2 E_3\ \text{Cov}_4\ (T_1, T_3)$$

$$= E_1 E_2\ \text{Cov}_3\ [E_4 (T_1),\ T_3$$

[since $E_4(T_3) = T_3$ and $E_3(T_3) = 0$]

$$= E_1 E_2\ (\frac{1}{n\lambda} - \frac{1}{n}) N^2\ \frac{1}{n-1}\ [\sum_{S_1} M_i^{\ 2}\ x_i\ \bar{Y}_i$$

$$- \frac{\sum_{S_1} M_i \bar{Y}_i \ \sum_{S_1} M_i \bar{x}_i}{n}]$$

$$= \left(\frac{1}{n\lambda} - \frac{1}{n}\right) N^2 \frac{1}{n-1} \quad E_1 \left[\frac{(n-1)}{n} \sum_{S_1} M_i^2 \left(\frac{1}{m_i} - \frac{1}{M_i}\right) \right.$$

$$\left. \rho_i \, S_{xi} \, S_{yi} + \sum_{S_1} (X_i - \bar{X}_{S_1}) (Y_i - \bar{Y}_{S_1}) \right]$$

(where $\bar{Y}_{s1} = \frac{1}{n} \sum_{S_1} Y_i$, $\bar{X}_{S1} = \frac{1}{n} \sum_{S_1} X_i$)

$$= N^2 \left(\frac{1}{n\lambda} - \frac{1}{n}\right) (S_{wxy} + S_{bxy}) \tag{5.5.7}$$

From equation (5.4.2) we get the optimum value of ϕ for given λ which minimizes $V(T|\lambda)$ is given by

$$\phi_o = \frac{P\lambda}{R\mu + P\lambda - Q\mu^2} \tag{5.5.8}$$

with

$$P = S_{by}^{\;2} + S_{wy}^{\;2}$$

$$Q = (S_{wxy} + S_{bxy})^2 (S_{bx}^{\;2} + S_{wx}^{\;2})^{-1}$$

$$R = S_{by}^{\;2} + \bar{\sigma}^2 \quad \text{where} \quad \bar{\sigma}^2 = \frac{1}{N} \sum_i \sigma_{io}^{\;2}.$$

The optimum value of μ (= $1-\lambda$) minimizing $V_{opt}(T|\lambda)$ is

$$\mu = \mu_o \text{ (say)} = \frac{R - \{P(R-Q)\}^{1/2}}{Q} \tag{5.5.9}$$

We shall assume $R \geq Q$ as is necessary to get a *real solution. The resulting minimum for* $V(T|\lambda)$ *over* variation in λ turns out to be

$$V_{min}(T) = N^2 \frac{PQ}{n} [P + R - 2 \{P(R-Q)\}^{1/2}]^{-1} - N S_{by}^2 \tag{5.5.10}$$

REMARKS

(i) If a completely un-matched sample of n f.s.u's is taken on the second occasion and if the ith f.s.u is selected and we take a sub-sample of m_i s.s.u's from it then the variance of the estimator for Y, say,

$$T' = \frac{N}{n} \sum_i \frac{M_i}{m_i} \Sigma y_{ij}$$

is $$V(T') = N^2 [(\frac{1}{n} - \frac{1}{N}) S_{by}^2 + \frac{1}{n} S_{wy}^2] = N^2 \frac{P}{n} - N S_{by}^2 \tag{5.5.11}$$

With a little algebra one may check that the minimum variance of the proposed estimator T based on the above sampling scheme has smaller variance than that of T' based on completely un-matched sample of the same size.

(ii) The sampling scheme considered above reduces to Abraham, Khosla and Kathuria's (1969) scheme when $M_i = M$, $m_i = m$ and $\sigma_i = \sigma \quad \forall\ i = 1, \ldots, N$. Abraham et al (1969) considered two estimators defined in (5.2.3) and (5.2.4) as

$$\hat{\bar{Y}}_3 = a\ \bar{x}' + b\ \bar{x}'' - (a+b)\ \bar{x}''' + d\ \bar{y}' + e\ \bar{y}'' + (1-d-e)\bar{y}'''$$

and

$$\hat{\bar{Y}}_4 = P\ \bar{y}_c + (1-P)\bar{y}''' + P'(\bar{x}_c - \bar{x}'')$$

Then $\hat{\bar{Y}}_4$ can be written also as

$$\hat{\bar{Y}}_4 = A\ \bar{x}' + B\ \bar{x}'' - (A+B)\ \bar{x}''' + D\ \bar{y}' + E\ \bar{y}' + (1-D-E)\ \bar{y}'''$$

with

$$A = P'\alpha - PK\rho_2\beta, \qquad B = PK\rho_2\beta + P'\beta$$

$$D = PK \quad \text{and} \quad E = P(1-K)$$

where $\alpha = 1-\beta$ as defined in strategy C and D.

Since $\hat{\bar{Y}}_4$ is of the same form as $\hat{\bar{Y}}_3$, the minimum variance of $\hat{\bar{Y}}_3$ can never be greater than that of $\hat{\bar{Y}}_4$. The empirical results due to Abraham et al (1969) however, contradict this. An explanation for the anomaly is that they indiscriminately ignored the f.p.c. Two estimators for Y under Abraham et al's (1969) strategy are $M_o\ \hat{\bar{Y}}_3$ and $M_o\ \hat{\bar{Y}}_4$.

It is easy to see that T reduces to $M_o \hat{\bar{Y}}_4$ when $M_i = M$, $m_i = m$ and $\alpha_i = \alpha \;\forall\; i$. Thus the strategy I is superior to that of Abraham et al (1969) using $M_o \hat{\bar{Y}}_4$. But it is difficult to check whether the strategy I is superior to that of Abraham et al's (1969) using $\hat{\bar{Y}}_3$. Arguing similarly it can be checked that the proposed strategy is superior to that of strategy A due to Singh (1968) and B due to Singh and Kathuria (1969) which are special cases of strategy I with assumptions like $M_i = M$, $m_i = m \;\forall\; i$ and respectively with restrictions $\alpha_i = 1 \;\forall\; i$ and $\alpha_i = \alpha \;\forall\; i$ together with $\lambda = 1$.

STRATEGY II

S_1 and S_{1i}'s are selected first by SRSWOR as in strategy I. Then S_1' is also selected by SRSWOR. Let x_{jr} denote the values of rth s.s.u of S_{1j} corresponding to jth f.s.u that is selected on the first occasion. From S_{1j} is selected a sub-sample S_{2mj} of size $m_j\alpha_j$ by following RHC method using normed size-measure

$$p_{jr}^* = \frac{x_{jr}}{\sum_{S_{1j}} x_{jr}}$$

and from the remaining $M_j - m_j$ s.s.u's of jth f.s.u we select

$m_j\beta_j$ s.s.u's by SRSWOR. S_{2u} and S'_{2ui}'s are selected as in strategy I. The proposed estimator for Y is

$$T^* = \phi T_1^* + (1-\phi) T_2 + \psi T_3 \qquad (5.5.12)$$

Here T_2 and T_3 are the same as in strategy I and

$$T_1^* = \frac{N}{n\lambda} \sum_{S_1'} t_i'$$

where

$$t_i' = M_i \left\{ \frac{a_i}{m_i} \sum_{r=1}^{m_i \alpha_i} \frac{y_{ir}}{P^*_{ir}} P^*_{ir} + (1-a_i) \frac{1}{m_i \beta_i} \sum_{\ell=1}^{m_i \beta_i} y_{i\ell} \right\}$$

Here P^*_{ir} is the sum of p^*_{jr} values over the s.s.u's falling in the ith group formed while selecting S'_{2mj} by RHC scheme. The constants a_i's are so chosen that $v_{24}(t_i')$ becomes minimum. Using Avadhani and Sukhatme's (1970) results discussed in (Section 4.3, p.108) we get the optimum values of a_i and α_i's as

$$a_{i\ opt} = a_{io}\ (\text{say}) = \frac{1}{2}$$

$$\alpha_{i\ opt} = \alpha_{io}\ (\text{say}) = (1 + \sqrt{\delta i}\ )^{-1}$$

respectively and putting these optimum values in the expression for t_i' we get

$$t_i' = t_{io}' \text{ (say) } = \frac{M_i}{2} \left[\frac{1}{m_i} \sum_{r=1}^{m_i \alpha_i} \frac{y_{ir}}{p^*_{ir}} P^*_{ir} + \frac{1}{m_i \beta_i} \sum_{\ell=1}^{m_i \beta_i} y_{i\ell} \right]$$

(5.5.13)

th variance

$$v_{24}(t_{io}') = \sigma^{*2}_{io} = M_i^2 \; S_{yi}^2 \left(\frac{1 + \sqrt{\delta_i}}{2m_i} - \frac{1}{M_i} \right)$$

e variance of T_1^* with $t_i' = t_{io}'$ turns out to be

$$V(T_1^*) = V_1' \text{ (say) } = N^2 \left[\left(\frac{1}{n\lambda} - \frac{1}{N} \right) S_{by}^2 + \frac{1}{n\lambda} \bar{\sigma}^{*2} \right]$$

where

$$\bar{\sigma}^{*2} = \frac{1}{N} \sum_i \sigma^{*2}_{io}$$

and V_2, V_3, V_{12} and V_{13} are the same as in strategy I.

The optimum value of μ minimizing $V(T^*|\lambda)$ is given by

$$\mu = \mu_o \text{ (say) } = \frac{R' - \{P(R' - Q)\}^{1/2}}{Q} \qquad (5.5.14)$$

(assuming $R' > Q$) and consequently the minimum value for $V(T^*)$ as

$$V_{min}(T^*) = N^2 \frac{PQ}{n} \left[P + R' - 2\sqrt{P(R'-Q)} \right]^{-1} - N\, S_{by}^2$$

(5.5.15)

where $R' = S_{by}^2 + \bar{\sigma}^{*2}$, P and Q are the same as in strategy I.

STRATEGY III

Here S_1 is selected by PPSWR method using normed size-measures P_i's. Each f.s.u when selected on any draw sub-sampled independently as in strategy I no matter whether it appears on any other draw. Then, S_1' of $n\lambda$ f.s.u's is selected by SRSWOR from S_1. Here S_{2mj} and S_{2uj}'s are selected as in strategy I. Finally an unmatched sample S_{2u} of $n\mu$ f.s.u's is selected from the entire population by PPSWR method using the previous normed size-measures. If ith f.s.u is selected at any draw then sub-sample of m_i s.s.u's is selected from it by SRSWOR (of course, if ith f.s.u appears λ_i times, λ_i independent sub-samples of size m_i are taken by SRSWOR from it) method.

The estimator for Y is taken as

$$T^{**} = \phi\, T_1^{**} + (1-\phi)\, T_2^{*} + \psi\, T_3^{*} \qquad (5.5.16)$$

where

$$T_1^{**} = \frac{1}{n\lambda} \sum_r \sum_i \frac{L_i'(r)}{P_i}\, t_{1o}(r)$$

where $L_i'(r) = 1$ if rth draw produces ith f.s.u. in S_1'.

$= 0$ otherwise.

$t_{1o}(r) =$ an estimator of the form t_{1o} defined in (5.5.2) when rth draw produces ith unit.

$$T_2^* = \frac{1}{n\mu} \sum_r \sum_i \frac{L_i''(r)}{P_i} M_i \bar{y}_i''(r)$$

re $L_i''(r) = 1$ if rth draw produces ith unit in S_{2u}
$= 0$ otherwise.

$$T_3^* = \frac{1}{n\lambda} \sum_r \sum_i \frac{L_i'(r)}{P_i} M_i \bar{x}_i(r)$$
$$- \frac{1}{n} \sum_i \sum_i L_i(r) \frac{M_i \bar{x}_i(r)}{P_i}$$

$L_i(r) = 1$ if rth draw produces ith unit in S_1
$= 0$ otherwise.

$\bar{y}_i''(r)$ = mean per s.s.u based on the sub-sample taken from ith f.s.u selected at rth draw in selecting the sample S_{2u}.

$\bar{x}_i(r)$ = mean per s.s.u based on the sub-sample taken from ith f.s.u selected at rth draw in selecting the sample S_1.

Now by virtue of our lemma (5.1) we have

$$V(T_1^{**}) = V_{13}\left(\frac{1}{n\lambda} \sum_r \sum_i L_i'(r) \frac{Y_i}{P_i}\right)$$
$$+ E_1 E_3 \frac{1}{(n\lambda)^2} \sum_r \sum_i \frac{L_i'^2(r)}{P_i^2} \sigma_{io}^2$$

$$= \left(\frac{1}{n\lambda} - \frac{1}{n}\right) \frac{1}{n-1} E_1 \sum_i \sum_r \left\{\frac{L_i(r)\; Y_i}{P_i} - \frac{1}{n} \sum_i \sum_r \frac{L_i(r)\; Y_i}{P_i}\right\}^2$$

$$+ V_1 \left(\frac{1}{n} \sum_r \sum_i L_i(r) \frac{Y_i}{P_i}\right) + \frac{1}{n\lambda} \sum_i \frac{\sigma_{io}^2}{P_i}$$

$$= \left(\left(\frac{1}{n\lambda} - \frac{1}{n}\right) + \frac{1}{n}\right) \sum_i P_i \left(\frac{Y_i}{P_i} - Y\right)^2 + \frac{1}{n\lambda} \sum_i \frac{\sigma_{io}^2}{P_i}$$

$$= \frac{1}{n\lambda} \left[\sum_i P_i \left(\frac{Y_i}{P_i} - Y\right)^2 + \sum_i \frac{\sigma_{io}^2}{P_i}\right]$$

$$= V_1' \text{ (say)}, \tag{5.5.17}$$

$$V(T_2^*) = V\left(\frac{1}{n\mu} \sum_r \sum_i \frac{L_i''(r)}{P_i} M_i \bar{y}''(r)\right)$$

$$= \frac{1}{n\mu} \sum_i \frac{M_i^2}{P_i} \left(\frac{1}{m_i} - \frac{1}{M_i}\right) S_{yi}^2 + \frac{1}{n\mu} \sum_i P_i \left(\frac{Y_i}{P_i} - Y\right)^2$$

$$= V_2' \text{ (say)} \tag{5.5.18}$$

$$V(T_3^*) = V\left[\frac{1}{n\lambda} \sum_r \sum_i \frac{L_i'(r)}{P_i} M_i \bar{x}_i(r)\right.$$

$$\left. - \frac{1}{n} \sum_r \sum_i \frac{M_i}{P_i} L_i(r) \bar{x}_i(r)\right]$$

$$= E_1E_2V_3 \left[\frac{1}{n\lambda} \sum_r \sum_i \frac{L_i'(r)}{P_i} M_i \bar{x}_i(r) - \frac{1}{n} \sum_r \sum_i \frac{M_i}{P_i} L_i(r) \bar{x}_i(r) \right]$$

$$= \left(\frac{1}{n\lambda} - \frac{1}{n}\right) \left[\sum_i \frac{M_i^2}{P_i} \left(\frac{1}{m_i} - \frac{1}{M_i}\right) S_{xi}^2 + \sum_i P_i \left(\frac{X_i}{P_i} - X\right)^2 \right] \qquad (5.5.19)$$

$$\text{Cov}(T_1^{**}, T_3) = E_1E_2\text{Cov}_3 [E_4(T_1^*),\ E_4(T_3^*)]$$

$$= E_1E_2 \text{ Cov}_3 \left[\frac{1}{n\lambda} \sum_r \sum_i L_i(r) \frac{\bar{y}_i(r)}{P_i} M_i ,\ \left\{ \frac{1}{n\lambda} \sum_r \sum_i \frac{M_i}{P_i} L_i'(r) \bar{x}_i(r) - \frac{1}{n} \sum_r \sum_i L_i(r) \frac{M_i \bar{x}_i(r)}{P_i} \right\} \right]$$

$$= \left(\frac{1}{n\lambda} - \frac{1}{n}\right) \left[\sum_i P_i \left(\frac{Y_i}{P_i} - Y\right)\left(\frac{X_i}{P_i} - X\right) + \sum_i \frac{M_i^2}{P_i} \left(\frac{1}{m_i} - \frac{1}{M_i}\right) \rho_i S_{xi} S_{yi} \right]$$

$$= V_{13}' \quad \text{(say)} \qquad (5.5.20)$$

Clearly

$$\text{Cov}\ (T_1^{**},\ T_2) = 0 = \text{Cov}\ (T_2^{*},\ T_3^{*}) \tag{5.5.21}$$

The optimum value of $V(T^{**}|\lambda)$ turns out to be

$$V_{opt}(T^{**}|\lambda) = \frac{P''}{n} \{\mu + P''\ \lambda(R'' - \mu Q')^{-1}\}^{-1} \tag{5.5.22}$$

with

$$P'' = n\mu\ V_2',\quad R'' = n\ \lambda\ V_1'\ ,\ Q' = (\frac{1}{n\lambda} - \frac{1}{n})^{-1}\ \frac{V_{13}'^2}{V_3}$$

The optimum value of μ minimizing $V_{opt}(T^{**}|\lambda)$ is

$$\mu = \mu_o\ (\text{say}) = \frac{R'' - \{P''(R'' - Q')\}^{1/2}}{Q'} \tag{5.5.23}$$

(we assume here $R'' > Q'$).

The resulting minimum value of $V_{opt}(T^{**}|\lambda)$ with μ replaced by μ_o, is

$$V_{min}(T^{**}) = \frac{P''Q'}{n}\ [P'' + R'' - 2\sqrt{P''(R'' - Q')}]^{-1} \tag{5.5.24}$$

REMARK

The sampling scheme considered in strategy III reduces to that of Kathuria's (1975) strategies E and F when M_i and m_i's are the same for each i. Kathuria' estimators for $\bar{Y}$ are

$$\hat{\bar{Y}}_5 = k\,(\bar{t}_1 + \bar{t}_{12} - \bar{t}_{\lambda 1}) + (1-k)\,\bar{t}_2'''$$

$$\hat{\bar{Y}}_6 = P\,\{\bar{t}_2 + Q(\bar{t}_1 - \bar{t}_{\lambda 1})\} + (1-P)\,\bar{t}_2'''$$

using the lemma (5.1) one has

$$V(N\,\bar{t}_{12}) = v_{13}(\frac{1}{n\lambda} \sum_r \sum_i L_i'(r) \frac{Y_i}{P_i})$$

$$+ E_{13} \frac{1}{(n\lambda)^2} \sum_r \sum_i \frac{L_i'^2(r)}{P_i^2} v_{24} [Q\{ (\bar{x}_i(r) - \bar{x}_i'(r)$$

$$+ \bar{y}_i(r)\} + (1-Q)\,\bar{y}_i''(r)]$$

$$\geq v_{13} (\frac{1}{n\lambda} \sum_r \sum_i L_i'(r) \frac{Y_i}{P_i})$$

$$+ E_1 E_3 \frac{1}{(n\lambda)^2} \sum_r \sum_i \frac{L_i^2(r)}{P_i^2} v_{24}(t_{io}(r))$$

$$= V(\frac{T_1^{**}}{M_o}).$$

If $M_i = M$ and $m_i = m \quad \forall\, i$, then

$$\bar{t}_1 - \bar{t}_{\lambda 1} = -\frac{T_3^*}{M_o}, \qquad \bar{t}_2''' = \frac{T_2^*}{M_o}$$

and

$$\mathrm{Cov}\,(\frac{T_1^{**}}{M_o}, \frac{T_3^*}{M_o}) = \mathrm{Cov}\,(\bar{t}_{12}, \bar{t}_{\lambda 1} - \bar{t}_1)$$

$$\mathrm{Cov}\,(\bar{t}_1 - \bar{t}_{\lambda 1}, \bar{t}_2''') = 0.$$

ence, the strategy III is superior to strategy E due to athuria (1975).

ow $\hat{\bar{Y}}_6$ can be written in the form

$$\hat{\bar{Y}}_6 = PZ^* - QP \frac{T_3^*}{M_o} + (1-P) \frac{T_2^*}{M_o}$$

where

$$Z^* = \frac{1}{Nn\lambda} \sum_i \sum_r \frac{L_i(r)}{P_i} Z_i(r)$$

with

$$Z_i(r) = g\, \bar{y}_i(r) + hg\, \{\bar{x}_i(r) - \bar{x}_i'(r)\} + (1-g)\, \bar{y}_i'(r)$$

ow $t_{io}(r)$ in T_1^{**} reduces to $Z_i(r)$ in case $a_i = g$ and $_i = hg \;\forall\; i$.

So $\quad \text{Min } V(Z^*) \geq V_{min}(T_1^{**})$

ince $\quad Cov\left(\frac{T_1^{**}}{M_o}, \frac{T_3^*}{M_o}\right) = Cov\left(Z^*, \frac{T_3^*}{M_o}\right)$, it follows that

$$V_{min}(N\, \hat{\bar{Y}}_6) > V_{min}(T^{**})$$

hus, the strategy III excels Kathuria's (1975) with either stimator $\hat{\bar{Y}}_5$ or $\hat{\bar{Y}}_6$, in the sense of yielding smaller ariances.

STRATEGY IV

Here S_1, S_{1j}'s and S_1' are selected as in strategy III. But S_{2mj} and S_{2uj}'s are selected as in strategy II. Finally, the un-matched f.s.u's and the s.s.u's therefrom are selected as in strategy III.

For Y the estimator is

$$T^{***} = \phi\, T_1^{***} + (1-\phi)\, T_2^* + \psi\, T_3^*$$

with

$$T_1^{***} = \frac{1}{n\lambda} \sum_r \sum_i \frac{L_i(r)}{P_i}\, t'_{io}(r)$$

where

$$\left.\begin{aligned} L_i'(r) &= 1 \\ &= 0 \end{aligned}\right\} \begin{aligned} &\text{if the rth draw produces ith f.s.u in } S_1' \\ &\text{otherwise,} \end{aligned}$$

$t'_{io}(r)$ is an estimator of the form t'_{io} defined in (5.5.13) when rth draw produces ith f.s.u in S_1' and T_2^*, T_3^* are as in strategy IIf.

Then,

$$V(T_1^{***}) = V_1^* \text{ (say)} = \frac{1}{n\lambda} \left[\sum_i p_i \left(\frac{Y_i}{P_i} - Y\right)^2 + \sum_i \frac{\sigma_{io}^{*2}}{P_i}\right]$$

(using lemma 5.1)

with

$$\sigma_{io}^{*2} = M_i^2\, S_{yi}^2 \left(\frac{1 + \delta_i}{2m_i} - \frac{1}{M_i}\right)$$

and

$$\text{Cov}\,(T_1{}^{***},\ T_3{}^{*}) = \text{Cov}\,(T_1{}^{**},\ T_3{}^{*}) = V_{13}{}'$$

The optimum value of μ comes out as

$$\mu_o = \frac{R''' - \{P''\,(R'''-Q')\}^{1/2}}{Q'} \quad (\text{assuming } R''' > Q') \tag{5.5.25}$$

with

$$R''' = n\lambda V_1{}^*, \quad P'' = n\mu V_2{}' \text{ and } Q' = \left(\frac{1}{n\lambda} - \frac{1}{n}\right)^{-1} \frac{V_{13}'^{2}}{V_3}$$

Finally we get the minimum variance of T*** as

$$V_{min}(T^{***}) = \frac{P''Q'}{n}\left[\,P'' + R''' - 2\sqrt{P''(R'''-Q')}\,\right]^{-1}$$

where

$$R''' = n\lambda V_1{}^* \quad (\text{assuming } R''' > Q')$$

STRATEGY V

Here S_1 and S_{2u} are selected (the latter from the entire population in an independent manner) following suitable πps schemes using normed size-measures p_i's. The sub-samples are all selected as in strategy I.

For Y the estimator considered is

$$T^V = \phi\, T_1{}^V + (1-\phi)\, T_2{}^V + \psi\, T_3{}^V$$

with

$$T_1^{v} = \frac{1}{\lambda} \sum_{S_1'} \frac{t_{io}}{\pi_i}$$

$$T_2^{v} = \sum_{S_{2u}} \frac{M_i \bar{y}_i^{*\prime}}{\pi_i^{*}}$$

and $$T_3^{v} = \frac{1}{\lambda} \sum_{S_1'} \frac{M_i \bar{x}_i}{\pi_i} - \sum_{S_1} \frac{M_i \bar{x}_i}{\pi_i}$$

here π_i $(= nP_i)$ and π_i^{*} $(= uP_i)$ are the inclusion-probabilities of the ith unit in selecting the samples S_1 and S_{2u} respectively, and t_{io} is as defined in (5.5.2).

Then using lemma (5.1) one has

$$V(T_1^{v}) = V_1 \left(\sum_{S_1} \frac{Y_i}{\pi_i}\right) + \frac{n^2}{n-1} \frac{1}{n} \left(\frac{1}{\lambda} - 1\right) \left[\sum_i \frac{Y_i^2}{\pi_i} - \frac{V_1 \left(\sum_{i \in S_1} \frac{Y_i}{\pi_i}\right) + Y^2}{n} \right] + \frac{1}{\lambda} \sum_i \frac{\sigma_{io}^2}{\pi_i}$$

$$= V_{\pi ps}(y, n) + \frac{1}{(n-1)} \left(\frac{1}{\lambda} - 1\right) [V_{pps}(y) - V_{\pi ps}(y, n)]$$

$$+ \frac{1}{n\lambda} \sum_i \frac{\sigma_{io}^2}{P_i}$$

and

$$V(T_2^{v}) = \frac{1}{n\mu} \sum_i \frac{M_i^2}{P_i} \left(\frac{1}{m_i} - \frac{1}{M_i}\right) S_{yi}^2 + V_{\pi ps}(y, n\mu)$$

Writing

$$V_{pps}(y) = \sum_i P_i \left(\frac{Y_i}{P_i} - Y\right)^2, \quad V_{\pi ps}(y,n) = V_1\left(\sum_{S_1} \frac{Y_i}{\pi_i}\right)$$

$$\text{and} \quad V_{\pi ps}(y, n\mu) = V_1\left(\sum_{S_{2u}} \frac{Y_i}{\pi_i^*}\right)$$

one gets

$$V(T_3^*) = \left(\frac{1}{n\lambda} - \frac{1}{n}\right) \left[\sum_i M_i^2 \left(\frac{1}{m_i} - \frac{1}{M_i}\right) \frac{S_{xi}^2}{P_i} + \frac{n}{n-1} \left(V_{pps}(x) - V_{\pi ps}(x,n)\right)\right]$$

$$= V_3^V \quad \text{(say)}$$

where

$$V_{\pi ps}(x,n) = V\left(\sum_{S_1} \frac{X_i}{\pi_i}\right)$$

$$\text{Cov}\,(T_1^V, T_3^V) = \left(\frac{1}{n\lambda} - \frac{1}{n}\right) \left[\sum_i \left(\frac{1}{m_i} - \frac{1}{M_i}\right) \frac{M_i^2 \rho_i S_{xi} S_{yi}}{P_i} + \sum_i \frac{X_i Y_i}{P_i} - \frac{1}{n(n-1)} \sum_{i \neq j}\sum \frac{X_i}{P_i} \frac{Y_j}{P_j} \pi_{ij}\right]$$

$$= V_{13}^V \quad \text{(say)}$$

Clearly, $\text{Cov}\,(T_1^V, T_2^V) = \text{Cov}\,(T_2^{IV}, T_3^{IV}) = 0.$

The optimum value of μ and the corresponding $V_{min}(T^V)$ can be obtained for a specific πps sampling scheme. But an expression for $V_{min}(T^V)$ for a particular scheme is complicated in general and hence we do not present one. It will be seen nevertheless that the strategy V is very often superior to strategy III.

5.6 COMPARISON AMONG THE DIFFERENT STRATEGIES

5.6.1 Relative efficiencies of the strategies I vs II and III vs IV

Since in the strategies I and II the estimators are $T = \phi\, T_1 + (1-\phi)\, T_2 + \psi\, T_3$ and $T^* = \phi\, T_1^* + (1-\phi)\, T_2 + \psi\, T_3$ respectively with $\text{Cov}\,(T_1, T_2) = \text{Cov}\,(T_1^*, T_2)$ and $\text{Cov}\,(T_1, T_3) = \text{Cov}\,(T_1^*, T_3)$.

$$V_{min}(T) \gtreqless V_{min}(T^*) \text{ according as } V(T) \gtreqless V(T_1^*)$$

$$\text{i.e. according as } \sigma_{io}^2 \gtreqless \sigma_{io}^{*2} \quad \forall\ i = 1, \ldots, N \qquad (5.6.1)$$

Similarly, the condition (5.6.1) determines the relative efficiencies of strategies III vs IV.

We shall now assume the following simple and straight-forward extension of the model due to Avadhani and Sukhatme (1970), denoted as M_2' viz.,

$$\underline{\text{Model}}\ M_2':\ y_{ij} = \beta_i\, x_{ij} + e_{ij}, \qquad i = 1, \ldots, N$$
$$j = 1, \ldots, M_i$$

$$\sum_{j=1}^{M_i} e_{ij} = \sum_{j=1}^{M_i} e_{ij}\, x_{ij} = 0 \qquad \forall\, i$$

and

$$e_{ij}^{2} = \sigma_i^{2}\, x_{ij}^{g} \qquad \forall\, ij$$

Under this model M_2' one may note that

$$\sigma_{io}^{2} - \sigma_{io}^{*2} = \frac{M_i^{2}}{2m_i}\left[\sqrt{1-\;_i^{2}} - \sqrt{\delta_i}\right] S_{yi}^{2} \gtreqless 0$$

according as

$$\frac{M_i^{2}}{2m_i}\, S_{yi}^{2}\, \sigma_i^{2} \left[\sum_{j=1}^{M_i} x_{ij}^{g} - \bar{X}_i \sum_{j=1}^{M_i} x_{ij}^{g-1}\right] \gtreqless 0$$

i.e., according as $g \gtreqless 1$.

Thus if the model M_2' holds, the strategy I is superior or inferior to strategy II as well as strategy III is superior or inferior to strategy IV according as g is less than or greater than 1.

5.6.2 <u>Comparison between strategies III and V</u>

The strategy V will be superior to strategy III if the following conditions are satisfied viz.,

(i) $V_{\pi ps}(y,m) < \frac{V_{pps}(y)}{m}$ for m = n and nμ, and

(ii) $$\frac{Cov^{2}\,(T_1^{v},\, T_3^{v})}{V(T_3^{v})} > \frac{Cov^{2}\,(T_1^{**},\, T_3^{*})}{V(T_3^{*})}$$

The condition (i) may be realized if the πps scheme satisfies [vide Raj (1968)] the condition

(iii) $\pi_{ij} > (m-1)m\, p_i p_j$ for $m = n$ and $n\mu$ and $i \neq j = (1, \ldots, N)$

or if the scheme is Midzuno's (1952) modified by Chaudhuri (1974) such that

$$p_i > \frac{m-1}{N-1} \quad \text{for } m = n \text{ and } n\mu \text{ and } i = 1, \ldots, N$$

But it is difficult to check whether (ii) is satisfied in practice. So, no categorical comparison between these two strategies is available. The condition (i) is satisfied approximately if the πps scheme is based on Sampford's (1967) scheme with N sufficiently large [vide Asok and Sukhatme (1977)]

5.6.3 <u>Comparison between strategies IV and V</u>

The strategy V will be superior to strategy IV if in conjunction with the conditions (i) and (ii) above the condition $\sigma_{io}^2 < \sigma^{*2}_{io}$ is also true. So, here also no simple comparison is available.

APPENDIX 5.1

Consider the following sampling scheme over two occasions

(i) On the first occasion a sample S_1 of size n is selected by SRSWOR from a population of N units.

(ii) On the second occasion a sub-sample S_{21} of size m (= nλ) is selected from S_1 by SRSWOR and a sample S_{22} of size u (= nμ, μ = 1-λ) is selected from the remaining N-n units which are not selected on the first occasion.

Let y_i, x_i be the value of the character under study measured on ith unit on the second and the first occasion respectively and

$$\bar{x}_m = \frac{1}{m} \sum_{S_{21}} x_i , \qquad \bar{y}_m = \frac{1}{m} \sum_{S_{21}} y_i ,$$

$$\bar{x}_n = \frac{1}{n} \sum_{S_1} x_i \qquad \bar{y}_u = \frac{1}{u} \sum_{S_{22}} y_i .$$

Consider the following estimator for $\bar{Y}$ ($= \frac{1}{N} \sum_i y_i$), the population mean of y as

$$\hat{\bar{Y}} = a\, \bar{y}_m + b\, \bar{y}_u + c\, \bar{x}_m + d\, \bar{x}_n \qquad (1)$$

Here a, b, c, d are constants to be determined so that $\hat{\bar{Y}}$ becomes unbiased for $\bar{Y}$ and have the minimum variance.

Now

$$E\,\hat{\bar{Y}} = (a+b)\,\bar{Y} + (c+d)\,\bar{X} = \bar{Y}$$

$$\Leftrightarrow \quad c+d = 0 \quad \text{and} \quad a+b = 1 \tag{2}$$

Using (2) we get

$$\hat{\bar{Y}} = a\,\bar{y}_m + (1-a)\,\bar{y}_u + c(\bar{x}_m - \bar{x}_n) \tag{3}$$

Now,

$$V(\hat{\bar{Y}}) = a^2 \left(\frac{1}{m} - \frac{1}{n}\right) S_y^{\,2} + (1-a)^2 \left(\frac{1}{u} - \frac{1}{N}\right) S_y^{\,2}$$

$$+ c^2 \left(\frac{1}{m} - \frac{1}{n}\right) S_x^{\,2} - 2a\,(1-a)\,\frac{S_y^{\,2}}{N}$$

$$+ 2ac \left(\frac{1}{m} - \frac{1}{n}\right) \rho\, S_x\, S_y \tag{4}$$

where

$$S_x^{\,2} = \frac{1}{N-1} \sum_i (x_i - \bar{X})^2, \quad S_y^{\,2} = \frac{1}{N} \sum_i (y_i - \bar{Y})^2$$

$$\rho\, S_x\, S_y = \frac{1}{N-1} \sum (x_i - \bar{X})(y_i - \bar{Y}), \quad \bar{X} = \frac{1}{N} \sum_i x_i$$

Solving $\frac{\partial}{\partial a} V(\hat{\bar{Y}}) = 0$ and $\frac{\partial}{\partial c} V(\hat{\bar{Y}}) = 0$ we get

$$a = \frac{\lambda}{1 - \mu^2 \rho^2} \quad \text{and} \quad c = \frac{-\lambda}{1 - \mu^2 \rho^2} \frac{S_y}{S_x} \tag{5}$$

Now putting the optimum values of a, c in (4) we get the optimum valu of $V(\hat{\bar{Y}})$ for given μ as

$$V_{opt}(\bar{Y}|\mu) = \left[\frac{1 - \mu\rho^2}{n(1 - \mu^2\rho^2)} - \frac{1}{N}\right] S_y^{\,2} \tag{6}$$

Minimizing $V_{opt}(\hat{\bar{Y}}|\mu)$ w.r.t. μ we get the optimum value of μ as

$$\mu_o \text{ (say)} = \frac{1}{\text{[illegible]}}$$

Finally putting $\mu = \mu_o$ in (5) and (6) we get the optimum values of $a = a_o$ (say) $= \frac{1}{2}$, $c = c_o$ (say) $= -\frac{1}{2}\rho\frac{S_y}{S_x}$ and the minimum value of $V(\hat{\bar{Y}})$ as

$$V_{min}(\hat{\bar{Y}}) \text{ (say)} = \left[\frac{1+\sqrt{1-\rho^2}}{2n} - \frac{1}{N}\right] S_y^2 .$$

CHAPTER - 6

SURVEYING A FINITE POPULATION ON MORE THAN TWO OCCASIONS

6.1 SUMMARY

Among strategies for sampling a finite population on two successive occasions, choosing units with varying probabilities, those proposed by Raj (1965), Ghangurde and Rao (1969) and Chotai (1974) are well-known and have been discussed in great details in the chapter 4. These are generalized here when sampling is performed on h (> 2) occasions. Relative efficiencies of these extended strategies are investigated.

6.2 INTRODUCTION

A finite population is required to be sampled on h ($\geq$ 2) successive occasions employing suitable sampling strategies with varying probabilities of selection of its N units. Writing the population total $Y_j = \sum_{i=1}^{N} y_{ji}$ of a real variate y assuming the value y_{ji} on its ith unit on the jth occasion for $j = 1, \ldots, h$, our problem is to estimate Y_h, the total on the current occasion. It is supposed that the known normed sizes p_i's for the units are available on every occasion to be utilized in sample-selection (throughout this chapter as 'size' measures these p_i's are implied). For the sake of simplicity it is assumed that there exist two known numbers ρ and V_o such that

$$\sum_{i=1}^{N} p_i \left(\frac{y_{j,i}}{p_i} - Y_j\right) \left(\frac{y_{k,i}}{p_i} - Y_k\right) = \rho^{|j-k|} V_o$$

$$\text{for every } j,k = 1,\ldots,h \qquad (6.2.1)$$

In case $p_i = \frac{1}{N}$ $\forall$ i, we write ρ as [illegible] and V_o as [illegible]. We propose the following sampling strategies I, II, III as generalizations respectively to those due to Raj (1965), Chotai (1974) and Ghangurde and Rao (1969).

6.3 SAMPLING STRATEGIES

6.3.1 Strategy I

On the first occasion a sample S_{11} of size m_{11} is selected following the scheme of selection with probabilities proportional to sizes with replacement (PPSWR). On the second occasion a simple random (sub-) sample (SRSWOR) S_{21} of size m_{21} is chosen without replacement from S_{11} and an independent sample S_{22} of size m_{22} $(= m_{11} - m_{21})$ is chosen from the whole population by PPSWR method. The value of the total sample size m_{11} on each occasion is supposed to be fixed and determined from cost considerations. In general on jth occasion a sample S_{jk} of size m_{jk} is selected by SRSWOR from $S_{j-1,k}$ independently for each $k = 1,\ldots,j-1$ and finally another sample of size m_{jj} $(= m_{11} - \sum_{k=1}^{j-1} m_{jk})$ is selected by PPSWR method independently from the entire population, this process being repeated for $j = 3,4,\ldots,h$. Our proposed estimator for Y_h on the hth occasion is of the form

$$\hat{Y}_h = \sum_{j=1}^{h} c_j \hat{Y}_h(j) \tag{6.3.1}$$

with

$$\hat{Y}_h(j) = \frac{1}{m_{hj}} \sum_{S_{hj}} \frac{y_{hi}}{p_i} - \beta_h(j) \left\{ \frac{1}{m_{hj}} \sum_{S_{hj}} \frac{y_{h-1,i}}{p_i} - \hat{Y}_{h-1}(j) \right\}$$

$$\text{for } j = 1,\ldots,h-1$$

$$= \frac{1}{m_{hh}} \sum_{S_{hh}} \frac{y_{hi}}{p_i} \qquad \text{for } j = h.$$

where $\beta_h(j)$'s and c_j's are constants rendering $\hat{Y}_h$ *unbiased* for Y_h with the least variance.

For this strategy I we have the following theorems.

Theorem 6.1

The variance $V(\hat{Y}_k(j))$ attains its minimum value when $\beta_k(j)$ equals ρ for every $j = 1,\dots,h-1$; $k = 1,2,\dots,h$ and the corresponding minimum variance is

$$V(\hat{Y}_h(j)) \begin{cases} = [\frac{1}{m_{hj}} - \sum_{k=1}^{h-j} (\frac{1}{m_{h+1-k,\ j}} - \frac{1}{m_{h-k,j}})(\rho^2)^k] & \text{for } j = 1,\dots h-1 \\ = \frac{V_o}{m_{hh}} & \text{for } j = h \end{cases}$$

Proof

We shall prove the theorem for $j = 1$ by induction over h and note that the result readily applies $\forall\ j = 2,\dots,h$. By actual calculation the theorem can be verified for $h = 2$ and 3. Let us suppose that the theorem holds for $h = t$, then we have show that the theorem holds for $h = t+1$.

Now

$$\hat{Y}_{t+1}(1) = \frac{1}{m_{t+1,1}} \sum_{S_{t+1,1}} \frac{y_{t+1,i}}{p_i} - \beta_{t+1}(1) \left(\frac{1}{m_{t+1,1}} \sum_{S_{t+1,1}} \frac{y_{ti}}{p_i} - \hat{Y}_t(1)\right)$$

and

$$\hat{Y}_t(1) = \frac{1}{m_{t,1}} \sum_{S_{t1}} \frac{y_{ti}}{p_i} - \rho \left(\frac{1}{m_{t1}} \sum_{S_{t1}} \frac{y_{t-1,i}}{p_i} - \hat{Y}_{t-1}(1)\right)$$

(since $\beta_h(1) = \rho$ for $h \leq t$)

$$= \sum_{j=0}^{t-2} \frac{(\rho)^j}{m_{t-j,1}} \left(\sum_{s_{t-j,1}} \frac{y_{t-j,i}}{p_i} - \rho \sum_{s_{t-j,1}} \frac{y_{t-j-1,i}}{p_i} \right)$$

$$+ \rho^{t-1} \hat{Y}_1(1)$$

$$V\left(\frac{1}{m_{t+1,1}} \sum_{s_{t+1,1}} \frac{y_{ti}}{p_i} - \hat{Y}_t(1)\right)$$

$$= V\left(\frac{1}{m_{t+1,1}} \sum_{s_{t+1,1}} \frac{y_{ti}}{p_i}\right) + V(\hat{Y}_t(1))$$

$$- 2\mathrm{Cov}\left(\frac{1}{m_{t+1,1}} \sum_{s_{t+1,1}} \frac{y_{ti}}{p_i}, \hat{Y}_t(1)\right)$$

$$= \Big[\frac{1}{m_{t+1,1}} + \frac{1}{m_{t1}} - \sum_{j=1}^{t-1} \left(\frac{1}{m_{t-j+1,1}} - \frac{1}{m_{t-j,1}}\right) (\rho^2)^j$$

$$- 2 \left\{ \frac{1}{m_{t1}} - \sum_{j=1}^{t-1} \left(\frac{1}{m_{t-j+1,1}} - \frac{1}{m_{t-j,1}}\right) (\rho^2)^j \right\} \Big] V_o$$

$$= \Big[\left(\frac{1}{m_{t+1,1}} - \frac{1}{m_{t1}}\right) + \sum_{j=1}^{t-1} \left(\frac{1}{m_{t-j+1,1}} - \frac{1}{m_{t-j,1}}\right) (\rho^2)^j \Big] V_o$$

(6.3.2)

Since it is assumed

$$V(\hat{Y}_t(1)) = \Big[\frac{1}{m_{t1}} - \sum_{j=1}^{t-1} \left(\frac{1}{m_{t-j+1,1}} - \frac{1}{m_{t-j,1}}\right) (\rho^2)^j\Big] V_o$$

and

$$\text{Cov}\left(\frac{1}{m_{t+1,1}} \sum_{S_{t+1,1}} \frac{y_{ti}}{p_i}, \hat{Y}_t(1)\right)$$

$$= \sum_{j=0}^{t-2} \rho^j \left[\text{Cov}\left(\frac{1}{m_{t+1,1}} \sum_{S_{t+1,1}} \frac{y_{ti}}{p_i}, \frac{1}{m_{t-j,1}} \sum_{S_{t-j,1}} \frac{y_{t-j,i}}{p_i}\right)\right.$$

$$\left. - \rho \text{ Cov}\left(\frac{1}{m_{t+1}} \sum_{S_{t+1,1}} \frac{y_{ti}}{p_i}, \frac{1}{m_{t-j,1}} \sum_{S_{t-j,1}} \frac{y_{t-1-j,i}}{p_i}\right)\right]$$

$$+ \rho^{t-1} \text{ Cov}\left(\frac{1}{m_{t+1,1}} \sum_{S_{t+1,1}} \frac{y_{ti}}{p_i}, \frac{1}{m_{11}} \sum_{S_{11}} \frac{y_{1i}}{p_i}\right)$$

$$= \left[\frac{1}{m_{t1}} - \sum_{j=1}^{t-1} \left(\frac{1}{m_{t-j+1,1}} - \frac{1}{m_{t-j,1}}\right) (\rho^2)^j\right] V_o$$

Similarly it can be shown that

$$\text{Cov}\left(\frac{1}{m_{t+1,1}} \sum_{S_{t+1,1}} \frac{y_{t+1,i}}{p_i}, \hat{Y}_t(1)\right)$$

$$= \rho \quad \frac{1}{m_{t1}} - \sum_{j=1}^{t-1} \left(\frac{1}{m_{t-j+1,1}} - \frac{1}{m_{t-j,1}}\right) (\rho^2)^j \quad V_o$$

(6.3.3)

Using (6.3.3) we get

$$\text{Cov}\left(\frac{1}{m_{t+1,1}} \sum_{S_{t+1,1}} \frac{y_{t+1,i}}{p_i}, \frac{1}{m_{t+1,1}} \sum_{S_{t1}} \frac{y_{ti}}{p_i} - \hat{Y}_t(1)\right)$$

$$= \rho \left[\left(\frac{1}{m_{t+1,1}} - \frac{1}{m_{t1}}\right) + \sum_{j=1}^{t-1} \left(\frac{1}{m_{t-j+1,1}} - \frac{1}{m_{t-j,1}}\right) (\rho^2)^j \right] V_o$$

(6.3.4)

From (6.3.2) and (6.3.4) we get the optimum value of $\beta_{t+1}(1)$ as ρ and the corresponding variance of $\hat{Y}_{t+1}(1)$ as

$$V(\hat{Y}_{t+1}(1))$$

$$= \left[\frac{1}{m_{t+1,1}} - \sum_{j=1}^{t}\left(\frac{1}{m_{t+2-j,1}} - \frac{1}{m_{t+1-j,1}}\right)(\rho^2)^j\right] V_o$$

Q.E.D.

Now taking $\beta_h(j)$ as ρ for every $j = 1,\ldots,h-1$, the optimum value of $V(\hat{Y}_h)$ with variation of c_j's $\forall\ j = 1,\ldots,h$ can be seen to equal

$$V_{opt}(\hat{Y}_h) = \left[\sum_{k=1}^{h-1} \frac{1}{V(\hat{Y}_h(k))} + \frac{1}{V(\hat{Y}_h(h))}\right]^{-1}$$

$$= \frac{V_o}{m_{11}}\left[\sum_{k=1}^{h-1} \frac{1}{\frac{1}{\lambda_{hk}} - \sum_{j=1}^{h-k}\left(\frac{1}{\lambda_{h+1-j,k}} - \frac{1}{\lambda_{h-j,k}}\right)(\rho^2)^j} + \lambda_{hh}\right]^{-1}$$

$$= \frac{V_o}{m_{11}}\left[\sum_{k=1}^{h} \frac{1}{\frac{1-\rho^2}{\lambda_{hk}} + \rho^2\, \phi_{h-1}(k)}\right]^{-1} \qquad (6.3.5)$$

on writing

$$\lambda_{hj} = \frac{m_{hj}}{m_{11}}, \qquad j = 1,\dots,h$$

and

$$\phi_h(k) \begin{cases} = \dfrac{1}{\lambda_{hk}} - \sum_{j=1}^{h-k} \left(\dfrac{1}{\lambda_{h+1-j,k}} - \dfrac{1}{\lambda_{h-j,k}}\right)(\rho^2)^j \\ \qquad\qquad k = 1,\dots,h-1 \\ = \dfrac{1}{\lambda_{hh}} \quad \text{for } k = h \end{cases} \tag{6.3.6}$$

Remembering that the m_{hj}'s are actually not pre-assigned but are to be determined optimally one has to minimize (6.3.5) with respect to λ_{hj}'s. The optimizing λ_{hj}'s work out as

$$\lambda_{hj(opt)} = \frac{\sqrt{1-\rho^2}}{1+\sqrt{1-\rho^2}} \quad \frac{1}{\phi_{(h-1)opt}^{(j)}}$$

for $j = 1,\dots,h-1$

and

$$\lambda_{hh(opt)} = 1 - \frac{\sqrt{1-\rho^2}}{1+\sqrt{1-\rho^2}} \quad \frac{1}{g_{(h-1)\ opt}}$$

where

$$g_{(h)opt} = \sum_{j=1}^{h} \frac{1}{\phi_{(h)opt}^{(j)}}$$

$$= \frac{(1+\sqrt{1-\rho^2})\ g_{(h-1)opt}}{(1-\sqrt{1-\rho^2}) + (1+\sqrt{1-\rho^2})\ g_{(h-1)opt}}$$

$$g_{(1)opt} = 1 \quad \text{and} \quad \phi_{(1)opt}^{(1)} = 1$$

$\phi_{(t)\,opt}(j)$'s being determined on substituting $\lambda_{kj(opt)}$ for λ_{kj} $(= m_{kj}/m_{11})$ $(k = j,\ j+1,\ \ldots,\ t)$ in $\phi_t(j)$ for $j = 1, \ldots, t,$ $t = 1, \ldots, h$. It follows that the minimum value of $V_{opt}(\hat{Y}_h)$ equals $V_{min}(\hat{Y}_h) = \frac{V_o}{m_{11}} g_{(h)\ opt}$ and the corresponding values of c_j's turn out to be

$$c_{j\ opt} = \frac{g_{(h)\,opt}}{\phi_{(h)\ opt}(j)} \qquad \text{for } j = 1, \ldots, h$$

For the hth occasion the optimum proportion of units to be matched is given by

$$\lambda_{h(opt)} = \sum_{j=1}^{h-1} \lambda_{hj(opt)} = 1 - \lambda_{hh(opt)}$$

$$= \frac{\sqrt{1-\rho^2}}{1+\sqrt{1-\rho^2}} \quad \frac{1}{g_{(h-1)\,opt}}$$

REMARK

Recently, Tripathi and Srivastava (1978) considered an extension which is virtually equivalent to our strategy I in the sense that both the strategies yield estimators with the same variance and requiring the same optimal proportion of matched units. We propose our slightly different sampling scheme because this suggests a direct procedure to get extensions of Ghangurde-Rao's (1969) and Chotai's (1974) schemes involving selection with varying probabilities 'without' replacement as we discuss next, whereas theirs is not so available. However, we are unable to get a straight-forward extension to Chaudhuri and Arnab's (1977) scheme to the case h > 2, even with our modifications.

6.3.2 Strategy II

On the first occasion a sample S_{11} of size m_{11} is chosen by Rao-Hartley-Cochran (1962) scheme, using size-measures p_i's. On the second occasion from S_{11} a sub-sample S_{21} of size m_{21} is selected by RHC scheme using size-measures ${}_1Q_i(2)$'s. Here ${}_1Q_i(2)$ is the sum of the p_i values of the group containing the ith unit formed in selecting S_{11}. Also an independent sample of size m_{22} $(= m_{11} - m_{21})$ is selected again from the entire population by RHC method using p_i's as size-measures. In general, on the jth occasion a sample S_{jk} of size m_{jk} is selected from $S_{j-1,k}$ independently for each $k = 1,2,\ldots,j-1$, following RHC scheme for each using normed size-measures ${}_kQ_i(j)$'s. Here ${}_kQ_i(j)$ is the sum of ${}_kQ_i(j-1)$ values of the group containing the ith unit in selecting the sample $S_{j-1,k}$

for $k = 1,2,\ldots,j-1$ and another independent sample S_{jj} is selected from the entire population using p_i's as normed size-measures (we shall use the notation ${}_kQ_i(k) = p_i$ for every $i = 1,2,\ldots,N$, $k = 2,\ldots,h$). This procedure is to be continued for $j = 2,\ldots,h$.

The estimator for Y_h proposed is

$$\hat{Y}'_h = \sum_{j=1}^{h} c'_j \, \hat{Y}'_h(j)$$

with

$$\hat{Y}'_h(j) = \sum_{S_{hj}} \frac{y_{hi}}{p_i} \; {}_jQ_i(h+1) - \beta'_h(j)\Big[\sum_{S_{hj}} \frac{y_{h-1,i}}{p_i} \; {}_jQ_i(h+1) - \hat{Y}'_{h-1}(j)\Big] \quad \text{for } j = 1,\ldots,h-1$$

$$= \sum_{S_{hh}} \frac{y_{hi}}{p_i} \; {}_hQ_i(h+1) \quad \text{for } j = h$$

where c'_j and $\beta'_h(j)$'s are constants chosen to make $\hat{Y}'_h$ and $\hat{Y}'_h(j)$ unbiased with least variances.

We have the following results .

Theorem 6.3

The value of $\beta'_h(j)$ that minimizes $V(\hat{Y}'_h(j))$ is ρ and the corresponding *minimum variance is*

$$V(\hat{Y}_h'(j)) = \frac{N}{N-1} \; V_o \, [\, (\frac{1}{m_{hj}} - \frac{1}{N}) - \sum_{i=1}^{h-j} (\frac{1}{m_{h+1-i,j}} - \frac{1}{m_{h-i,j}}) \, (\rho^2)^i \,]$$

for $j = 1, \ldots, h-1$

$$= \frac{N}{N-1} \; V_o \, (\frac{1}{m_{hh}} - \frac{1}{N}) \quad \text{for } j = h.$$

PROOF

We shall prove the theorem for $j = 1$ by induction over h and note that the result straightforwardway applies $\forall \, j = 2, \ldots, h-1$.

Now

$$\hat{Y}_h'(1) = \sum_{S_{h1}} \frac{y_{hi}}{P_i} \; {}_1Q_i(h+1) - \beta_h'(1) \; \sum_{S_{h1}} \frac{y_{h-1,i}}{P_i} \; {}_1Q_i(h+1) - \hat{Y}_{h-1}'(1)) \, .$$

For $h = 2$ and 3 we can easily verify the theorem. Let us suppose that the theorem is true for $h = t$ i.e. the optimum value $\beta_t'(1)$ is ρ and

$$V(\hat{Y}_t'(1 \;) = \frac{N}{N-1} \; V_o \, [\, (\frac{1}{m_{t1}} - \frac{1}{N}) - \sum_{i=1}^{t-1} (\frac{1}{m_{t+1-i,1}} - \frac{1}{m_{t-i,1}}) \, (\rho^2)^i]$$

we shall prove that the theorem holds for h = t+1.

$$\hat{Y}'_{t+1}(1) = \sum_{S_{t+1,1}} \frac{y_{t+1,i}}{p_i} \; {}_1Q_i(t+2)$$

$$- \beta'_{t+1}(1) \; (\sum_{S_{t+1,1}} \frac{y_{ti}}{p_i} \; {}_1Q_i(t+2) - \hat{Y}'_t(1) \;)$$

$$\hat{Y}'_t(1) = \sum_{k=0}^{t-2} \rho^k \Big[\sum_{S_{t-k,1}} \frac{y_{t-k,i}}{p_i} \; {}_1Q_i(t-k+1)$$

$$- \rho \sum_{S_{t-k,1}} \frac{y_{t-k-1,i}}{p_i} \; {}_1Q_i(t-k+1) \Big]$$

$$+ \rho^{t-1} \sum_{S_{11}} \frac{y_{1i}}{p_i} \; {}_1Q_i(2)$$

can be verified that

$$\mathrm{Cov} \Big(\sum_{S_{t+1,1}} \frac{y_{ti}}{p_i} \; {}_1Q_i(t+2), \; \hat{Y}'_t(1) \Big)$$

$$= \frac{N}{N-1} \Big[\Big(\frac{1}{m_{t1}} - \frac{1}{N} \Big) - \sum_{k=1}^{t-1} \Big(\frac{1}{m_{t+1-k,1}} - \frac{1}{m_{t-k,1}} \Big) (\rho^2)^k \Big] \tag{6.3.7}$$

nd

$$\mathrm{Cov} \Big(\sum_{S_{t+1,1}} \frac{y_{t+1,i}}{p_i} \; {}_1Q_i(t+2), \; \hat{Y}'_t(1) \Big)$$

$$= \rho \; \mathrm{Cov} \Big(\sum_{S_{t+1,1}} \frac{y_{ti}}{p_i} \; {}_1Q_i(t+2), \; \hat{Y}'_t(1) \Big) \tag{6.3.8}$$

Using (6.3.7) and (6.3.8) we get

$$V\left(\sum_{S_{t+1,1}} \frac{y_{ti}}{p_i} \; {}_1Q_i(t+2) - \hat{Y}'_t(1) \right)$$

$$= \frac{N}{N-1} \left[\left(\frac{1}{m_{t+1,1}} - \frac{1}{m_{t1}} \right) + \sum_{i=1}^{t-1} (\rho^2)^i \left(\frac{1}{m_{t+1-i,1}} - \frac{1}{m_{t-i,1}} \right) \right] V_o \quad (6.3.9)$$

and

$$\text{Cov} \left(\sum_{S_{t+1,1}} \frac{y_{t+1,i}}{p_i} \; {}_1Q_i(t+2) , \sum_{S_{t+1,i}} \frac{y_{ti}}{p_i} \; {}_1Q_i(t+2) - \hat{Y}'_t(1) \right)$$

$$= \rho \, V \left(\sum_{S_{t+1}} \frac{y_{ti}}{p_i} \; {}_1Q_i(t+2) - \hat{Y}'_t(1) \right) \quad (6.3.10)$$

From (6.3.9) and (6.3.10) we get the optimum value of $\beta'_{t+1}(1)$ as ρ and the corresponding variance of $\hat{Y}'_{t+1}(1)$ as

$$V(\hat{Y}'_{t+1}(1)) = \frac{N}{N-1} \left[\left(\frac{1}{m_{t+1,1}} - \frac{1}{N} \right) - \sum_{i=1}^{t} (\rho^2)^i \left(\frac{1}{m_{t+2-i,1}} - \frac{1}{m_{t+1-i,1}} \right) \right] V_o$$

Q.E.D.

s S_{jj} (1,...,h)'s are independent samples from the whole population the $\hat{Y}'_h(j)$'s are independent for different j's j = 1,...,h). So the optimum value of the variance of $\hat{Y}'_h$ with respect to c'_j is given by

$$V_{opt}(\hat{Y}'_h) = [\sum_{j=1}^{h} \frac{1}{V(\hat{Y}'_h(j))}]^{-1}$$

$$= \frac{N}{N-1} V_o [\sum_{j=1}^{h} \frac{1}{\phi'_h(j)}]^{-1} \qquad (6.3.11)$$

where

$$\phi'_h(j) \begin{cases} = (\frac{1}{m_{hj}} - \frac{1}{N}) - \sum_{i=1}^{h-j} (\frac{1}{m_{h+1-i,j}} - \frac{1}{m_{h-i,j}}) (\rho^2)^i & \text{for } j = 1,\dots,h-1 \\ = (\frac{1}{m_{hh}} - \frac{1}{N}) & \text{for } j = h. \end{cases}$$

Clearly

$$\phi'_1(1) = (\frac{1}{m_{11}} - \frac{1}{N})$$

Now minimizing $V_{opt}(\hat{Y}'_h)$ with respect to $\lambda_{hj} = \frac{m_{hj}}{m_{11}}$ for = 1,...,h, subject to the restriction $\sum_{j=1}^{h} \lambda_{hj} = 1$ we get the optimum value of λ_{hj}'s as

$$\lambda_{hj(opt)} = \frac{1}{m_{11}} \cdot \frac{\sqrt{1-\rho^2}\ \{N(1-\sqrt{1-\rho^2}) - m_{11}\ \lambda_{hh(opt)}\}}{N\ \rho^2\ \phi'_{(h-1)}(j)_{opt} - (1-\rho^2)}$$

$$\text{for } j = 1,\dots,h-1 \qquad (6.3.12)$$

and

$$\lambda_{hh(opt)} = \frac{1}{m_{11}} \cdot \frac{m_{11} - N\sqrt{1-\rho^2}\,(1-\sqrt{1-\rho^2})\;B(h-1)}{1-\sqrt{1-\rho^2}\;B(h-1)}$$

where

$$B(h-1) = \sum_{j=1}^{h-1} \frac{}{\rho^2\, N\, \phi'_{(h-1)\,opt}(j) - (1-\rho^2)}$$

For h = 2 we get

$$\lambda_{21(opt)} = \frac{\sqrt{1-\rho^2}}{1+\sqrt{1-\rho^2}} \quad \text{and} \quad \lambda_{22(opt)} = \frac{1}{1+\sqrt{1-\rho^2}}$$

In general $\phi'_{(t)\,opt}(j)$'s will be determined on substituting $\lambda_{kj(opt)}$ for λ_{kj} ($= m_{kj}/m_{11}$) ($k = j, j+1, \ldots, t$) in $\phi'_{(t)}(j)$ for $j = 1, \ldots, t,$ $t = 1, 2, \ldots, h.$

Finally, putting the optimum values of λ_{hj}'s in (6.3.11) we get the minimum variance of $\hat{Y}'_h$ as

$$V_{min}(\hat{Y}'_h) = \frac{N}{N-1}\; V\; \left(\frac{1}{m_{11}\;\lambda_{hh(opt)}} - \frac{1}{N}\right)$$

$$\left[1 - \frac{1}{\sqrt{1-\rho^2}}\left(1 - \frac{1}{\lambda_{hh(opt)}}\right)\right]^{-1} \qquad (6.3.13$$

and the corresponding values of c'_j 's as

$$c'_{j(opt)} = \frac{1}{\phi_{(h)\,opt}(j)} \cdot \left(\frac{1}{m_{11}\;\lambda_{hh(opt)}} - \frac{1}{N}\right)$$

$$\left[1 - \frac{1}{\sqrt{1-\rho^2}}\left(1 - \frac{1}{\lambda_{hh(opt)}}\right)\right]^{-1}$$

for $j = 1, \ldots, h-1$

$$= \left[1 - \frac{}{\sqrt{1-\rho^2}}\left(1 - \frac{1}{\lambda_{hh(opt)}}\right)\right]^{-1} \quad \text{for } j = h$$

6.3.3 STRATEGY III

On the first occasion we choose S_{11} of size m_{11} by RHC method as in strategy II. On the second occasion a simple random sub-sample S_{21} of size m_{21} from S_{11} is selected and an independent sample S_{22} of size m_{22} $(= m_{11}-m_{21})$ is selected from the entire population by RHC scheme using p_i's. Then, in general, on the jth occasion sub-samples S_{jk}'s of sizes m_{jk}'s are selected by SRSWOR from $S_{j-1,k}$ $(k = 1, \ldots, j-1)$ and an independent sample of size m_{jj} $(=, m_{11} - \sum_{k=1}^{j-1} m_{jk})$ from the entire population by RHC method using p_i's. This is repeated for all $j = 2, \ldots, h$.

The proposed estimator for Y_h is

$$\hat{Y}_h'' = \sum_{j=1}^{h} c_j'' \; \hat{Y}_h'(j) \qquad (6.3.14)$$

with

$$\hat{Y}_h''(j) = \frac{m_{jj}}{m_{hj}} \sum_{S_{hj}} \frac{y_{hi}}{p_i} \; {}_jQ_i(j+1)$$

$$- \left\{ \frac{m_{jj}}{m_{hj}} \sum_{S_{hj}} \frac{y_{h-1,i}}{p_i} \; {}_jQ_i(j+1) - \hat{Y}_{h-1}''(j) \right\}$$

$$\text{for } j = 1, \ldots, h-1$$

$$\doteq \sum_{S_{hh}} \frac{y_{hi}}{p_i} \; {}_hQ_i(h+1) \text{ for } j = h.$$

Here c''_j 's are constants to make $\hat{Y}''_h$ unbiased with the least variance, the quantities ${}_jQ_i(j+1)$'s are same as those in strategy II.

We have then the following results

<u>Theorem 6.4</u>

$$V(\hat{Y}''_h(1)) = \frac{N}{N-1}\left(\frac{1}{m_{11}} - \frac{1}{N}\right) V_o$$

$$+ \frac{2}{N-1}\left[\sum_{j=1}^{h-1}\left(\frac{1}{m_{h+1-j,1}} - \frac{1}{m_{h-j,1}}\right)\right.$$

$$\left.(1-\rho^j)(N-m_{11})V_o + (1-\delta^j)\, m_{11}\, \bar{v}\right]$$

<u>Proof</u>

By actual calculation it can be easily verified that the theorem is true for h = 1,2. Next we apply inductio First we assume that the theorem is true for h = r, to show that it extends to h = r+1, let us note the following :

$$\hat{Y}''_{r+1}(1) = \frac{m_{11}}{m_{r+1,1}} \sum_{S_{r+1,1}} \frac{y_{r+1,i}}{p_i}\ {}_1Q_i(2)$$

$$-\left\{\frac{m_{11}}{m_{r+1,1}} \sum_{S_{r+1,1}} \frac{y_{ri}}{p_i}\ {}_1Q_i(2) - \hat{Y}''_r(1)\right\}$$

$$\text{ov}\left(\frac{m_{11}}{m_{r+1,1}} \sum_{S_{r+1,1}} \frac{y_{ri}}{p_i} \; {}_1Q_i(2),\ \hat{Y}''_r(1)\right)$$

$$= \sum_{j=0}^{r-2} \text{Cov}\left[\frac{m_{11}}{m_{r-j,1}} \sum_{S_{r-j,1}} \frac{y_{ri}}{p_i} \; {}_1Q_i(2) \right. ,$$

$$\left\{ \frac{m_{11}}{m_{r-j,1}} \sum_{S_{r-j,1}} \frac{y_{r-j,1}}{p_i} \; {}_1Q_i(2) \right.$$

$$\left. \left. - \frac{m_{11}}{m_{r-j,1}} \sum_{S_{r-j,1}} \frac{y_{r-j-1,i}}{p_i} \; {}_1Q_i(2) \right\} \right]$$

$$+ \text{Cov}\left\{ \sum_{S_{11}} \frac{y_{ri}}{p_i} \; {}_1Q_i(2),\ \sum_{S_{11}} \frac{y_{1i}}{p_i} \; {}_1Q_i(2) \right\}$$

$$= \sum_{j=0}^{r-2} \left[\frac{m_{11}}{N-1} \left(\frac{1}{m_{r-j,1}} - \frac{1}{m_{11}}\right) (\delta^j - \delta^{j+1})\ \bar{V} \right.$$

$$\left. + \frac{N-m_{11}}{N-1} \frac{(\rho^j - \rho^{j+1}) V_o}{m_{r-j,1}} \right] + \frac{N-m_{11}}{N-1} \frac{\rho^{r-1}}{m_{11}} V_o$$

ince

$$\text{Cov}\left\{ \frac{m_{11}}{m_{r-j,1}} \sum_{S_{r-j,1}} \frac{y_{ri}}{p_i} \; {}_1Q_i(2),\ \frac{m_{11}}{m_{r-j,1}} \sum_{S_{r-j,1}} \frac{y_{r-j,i}}{p_i} \; {}_1Q_i(2) \right\}$$

$$= \frac{m_{11}}{N-1} \left(\frac{1}{m_{r-j,1}} - \frac{1}{m_{11}}\right) \bar{V}\ \delta^j + \frac{N-m_{11}}{N-1}\ V_o \frac{\rho^j}{m_{r-j,1}}$$

$$= \frac{1}{N-1} \left[\frac{(N-m_{11})\ V_o}{m_{r1}} + m_{11} \left(\frac{1}{m_{r1}} - \frac{1}{m_{11}}\right) \bar{V} \right.$$

$$\left. + \sum_{j=1}^{r-1} \left(\frac{1}{m_{r-j,1}} - \frac{1}{m_{r-j+1,1}}\right) (\delta^j\ \bar{v}\ m_{11} + (N-m_{11})\ \rho^j\ V_o) \right]$$

(6.3.15)

Similarly we get

$$\mathrm{Cov}\left(\frac{m_{11}}{m_{r+1,1}} \sum_{S_{r+1,1}} \frac{y_{r+1,i}}{p_i} \; {}_1Q_i(2),\ \hat{Y}''_r\right)$$

$$= \frac{1}{N-1} \left[\frac{N-m_{11}}{m_{r1}} \rho V_o + m_{11} \left(\frac{1}{m_{r1}} - \frac{1}{m_{11}}\right) \bar{v}\, \delta\right.$$

$$+ \sum_{j=1}^{r-1} \left(\frac{1}{m_{r-j,1}} - \frac{1}{m_{r-j+1,1}}\right) (\delta^{j+1} \bar{v}\, m_{11}$$

$$\left. + (N-m_{11})\ \rho^{j+1} V_o\right) \qquad (6.3.16)$$

Now using (6.3.15) and (6.3.16) and remembering

$$V(\hat{Y}''_r(1)) = \frac{N}{N-1} \left(\frac{1}{m_{11}} - \frac{1}{N}\right) V_o + \frac{2}{N-1} \sum_{j=1}^{r-1} \left(\frac{1}{m_{r+1-j,1}}\right.$$

$$\left. - \frac{1}{m_{r-j,1}}\right) \{(1-\rho^j)(N-m_{11})V_o + (1-\delta^j)m_{11}\ \bar{v}\}$$

we get

$$V(\hat{Y}'_{r+1}(1)) = \frac{N}{N-1} \left(\frac{1}{m_{11}} - \frac{1}{N}\right) V_o + \frac{2}{N-1} \sum_{j=1}^{r} \left(\frac{1}{m_{r+2-j,1}}\right.$$

$$\left. - \frac{1}{m_{r+1-j}}\right) \{(1-\rho^j)(N-m_{11})V_o + (1-\delta^j)m_{11}\ \bar{v}\}$$

Q.E.D.

similarly we can show that

$$V(\hat{Y}''_h(j)) = \frac{N}{N-1} \left(\frac{1}{m_{jj}} - \frac{1}{N}\right) V_o + \frac{2}{N-1} \sum_{i=1}^{h-j} \left(\frac{1}{m_{h+1-i,j}}\right.$$

$$\left. - \frac{1}{m_{h-i,j}}\right) \{(1-\rho^i)(N-m_{11})V_o + (1-\delta^i)m_{11}\ \bar{v}$$

$$= \frac{N}{(N-1)} \frac{V_o}{m_{11}} \left[\frac{2(1-\rho)(1+\gamma_1 f)}{\lambda_{hj}} + B'(j) \right]$$

for $j = 1, \ldots, h-1$

Since S_{hh} are selected from the entire population by following RHC scheme of sampling we have

$$v(\hat{Y}_h^{\prime}(h)) = \frac{N}{N-1} \frac{V_o}{m_{11}} \left(\frac{1 - f \lambda_{hh}}{\lambda_{hh}} \right)$$

where

$$\lambda_{hj} = \frac{m_{hj}}{m_{11}} \quad \text{for } j = 1, \ldots, h$$

$$\gamma_j = \frac{(1-\delta^j)}{(1-\rho^j)} \frac{\bar{v}}{v_o} - 1 \quad \text{for } j = 1, \ldots, h-1$$

$$f = \frac{m_{11}}{N}$$

$$B'(j) = \left(\frac{1}{\lambda_{jj}} - f\right) - \frac{2}{\lambda_{h-1,j}} (1-\rho)(1+\gamma_1 f)$$

$$+ 2 \sum_{i=2}^{h-j} \left(\frac{1}{\lambda_{h+1-i,j}} - \frac{1}{\lambda_{h-i,j}} \right) (1-\rho^i)(1+\gamma_i f)$$

The optimum value of $v(\hat{Y}_h^{\prime\prime})$ with respect to $c_i^{\prime\prime}$ is given by

$$V_{opt}(\hat{Y}_h^{\prime\prime}) = \left[\sum_{j=1}^{h-1} \frac{1}{\frac{2(1-\rho)(1+\gamma_1 f)}{\lambda_{hj}} + B'(j)} + \frac{1}{\left(\frac{1}{\lambda_{hh}} - f\right)} \right]^{-1} \frac{N}{(N-1)m_{11}} V_o \qquad (6.3.17)$$

Now minimizing $V_{opt}(\hat{Y}''_h)$ with respect to λ_{hj}'s subject to the condition $\sum_{j=1}^{h} \lambda_{hj} = 1$, we get the optimum value of λ_{hj}'s a

$$\lambda_{hj(opt)} = \frac{\sqrt{2(1-\rho)(1+\gamma_1 f)}\ \{1-f-\sqrt{2(1-\rho)(1+\gamma_1 f)}\}}{1 - f\sqrt{2(1-\rho)(1+\gamma_1 f)}\ B'} \cdot \frac{1}{B'}$$

for $j = 1, \ldots, h-1$

$$= \frac{1 - \sqrt{2(1-\rho)(1+\gamma_1 f)}\{1 - 2(1-\rho)(1+\gamma_1 f)\}\ B'}{1 - f\sqrt{2(1-\rho)(1+\gamma_1 f)}\ B'}$$

for $j = h$

where

$$B' = \sum_{j=1}^{h-1} \frac{1}{B'(j)}$$

For $h = 2$ we get

$$\lambda_{21(opt)} = \frac{\sqrt{2(1-\rho)(1+\gamma_1 f)}}{1 + \sqrt{2(1-\rho)(1+\gamma_1 f)}},$$

$$\lambda_{22(opt)} = \frac{1}{1 + \sqrt{2(1-\rho)(1+\gamma_1 f)}}$$

the same result as was obtained by Chotai (1974).

Putting the optimum values $\lambda'_{hj(opt)}$s in (6.3.17) we get the minimum variance of $\hat{Y}''_h$ as

$$V_{min}(\hat{Y}_h'') = \frac{N\ V_o}{(N-1)m_{11}} \left(\frac{1}{\lambda_{hh(opt)}} - f\right) .$$

$$\left[1 + \frac{1 - \lambda_{hh(opt)}}{\lambda_{hh(opt)}\sqrt{2(1-\rho)(1+\gamma_1 f)}}\right]^{-1}.$$

corresponding optimum values of c_j^{*} as

$$c_{j(opt)}'' = \frac{1}{\frac{2(1-\rho)(1+\gamma_1 f)}{\lambda_{hj(opt)}} + B'(j)} \left(\frac{1}{\lambda_{hh(opt)}} - f\right)$$

$$.\left[1 + \frac{1 - \lambda_{hh(opt)}}{\lambda_{hh(opt)}\sqrt{2(1-\rho)(1+\gamma_1 f)}}\right]^{-1}$$

for $j = 1, \ldots, h-1$

$$= \left[1 + \frac{1 - \lambda_{hh(opt)}}{\lambda_{hh(opt)}\sqrt{2(1-\rho)(1+\gamma_1 f)}}\right]^{-1}$$

for $j = h$.

6.4 RELATIVE EFFICIENCIES OF THE PROPOSED STRATEGIES

6.4.1 Strategy I Vs Strategy II

From theorems 6.2 and 6.3 we have

$$V(\hat{Y}_h(j)) - V(\hat{Y}_h'(j))$$

$$= \left[\frac{1}{m_{hj}} - \sum_{k=1}^{h-j}\left(\frac{1}{m_{h+1-k,j}} - \frac{1}{m_{h-k,j}}\right)(\rho^2)^k\right] V_o$$

$$- \frac{N}{N-1}\left[\left(\frac{1}{m_{hj}} - \frac{1}{N}\right) - \sum_{k=1}^{h-j}\left(\frac{1}{m_{h+1-k,j}} - \frac{1}{m_{h-k,j}}\right)(\rho^2)^k\right] V_o$$

≥ 0 for $j = 1, \ldots, h-1$

and also $V(\hat{Y}_h(h)) - V(\hat{Y}'_h(h)) \geq 0$

So $V_{min}(\hat{Y}_h) \geq V_{min}(\hat{Y}'_h)$.

Thus the strategy II is superior to the strategy I in the sense of having the smaller minimum variance for the estimator employed. the strategy I is as efficient as Tripathi and Srivastava (1978)'s strategy it follows that the strategy II is also superior to that of Tripathi and Srivastava.

6.4.2 Strategie II Vs strategy III

A comparison is not available between the strategies II and III. However we may attempt a comparison if in we use the difference type estimator $\hat{Y}'^{*}_h(j)$ as below in lieu of the regression type estimator $\hat{Y}'_h(j)$.

Let $\hat{Y}'^{*}_h(j)$ equal

$$\sum_{S_{hj}} \frac{y_{hi}}{p_i} \; {}_jQ_i(h+1) - \{ \sum_{S_{hj}} \frac{y_{h-1,i}}{p_i} \; {}_jQ_i(h+1) - \hat{Y}'^{*}_{h-1}(j) \} \quad \text{for } j = 1, \ldots, h-1$$

and equal

$$\sum_{S_{hh}} \frac{y_{hi}}{p_i} \; {}_hQ_i(h+1) \qquad \text{for } j = h$$

the estimator for Y_h be taken as $\hat{Y}'^{*}_{h} = \sum_{j} c^{*}_{j} \hat{Y}'^{*}_{h}$'s are constants chosen to make $\hat{Y}'^{*}_{h}$ unbiased with the imum variance). Now we may compare this resulting ategy II* (say) with the strategy III as below.

It can be checked that

$$V(\hat{Y}'^{*}_{h}(j)) = \frac{N}{m_{11}(N-1)} V_o \left[\left(\frac{1}{\lambda_{jj}} - f\right) + 2 \sum_{i=1}^{h-j} \left(\frac{1}{\lambda_{h-i+1,j}} - \frac{1}{\lambda_{h-i,j}}\right)(1-\rho^i)\right]$$

$$\text{for } j = 1,\ldots,h-1$$

$$= \frac{N}{m_{11}(N-1)} \left(\frac{1 - f\lambda_{hh}}{\lambda_{hh}}\right) V_o \quad \text{for } j = h$$

$$V(\hat{Y}''_{h}(j)) \geq, = \text{ or } < V(\hat{Y}'^{*}_{h}(j))$$

if

$$\frac{\bar{V}}{V_o} \geq, = \text{ or } < \frac{1-\rho^i}{1-\delta^i} \qquad \text{for } i = 1,\ldots,h-j$$

is a sufficient condition for the proposed strategy II* have a smaller variance than that based on III, is that

$$\frac{\bar{V}}{V_o} \geq, = \text{ or } < \frac{1-\rho^i}{1-\delta^i} \qquad \text{for } i = 1,\ldots,h-1$$

is verified that in case of h = 2, our conclusion ncides with that reached earlier by Chotai (1974). To er the case h > 2 we first prove the following lemma.

Lemma 6.1

If $1 > \rho > \delta > 0$, then $\dfrac{1-\rho^i}{1-\delta^i}$ increases monotonically with $i \geq 0$.

Proof

Let $f(x) = \dfrac{1-\rho^x}{1-\delta^x}$,

then

$$f'(x) = \frac{(1-\delta^x)(-\rho^x \log\rho) + (1-\rho^x)\ \delta^x \log \delta}{(1-\delta^x)^2}$$

$$= \frac{(1-\rho^x)(1-\delta^x)}{(1-\delta^x)^2} \left\{ - \frac{\rho^x \log \rho}{1-\rho^x} + \frac{\delta^x \log \delta}{1-\delta^x} \right\}$$

$$\geq 0.$$

This is because

$$\frac{\partial}{\partial t} \quad \frac{t^x \log t}{1-t^x}$$

$$= \frac{(1-t^x)(x\, t^{x-1} \log t + t^{x-1}) + x\, t^{2x-1} \log t}{(1-t^x)^2}$$

$$= \frac{t^{x-1}}{(1-t^x)^2} (x \log t + 1 - t^x) \leq 0$$

for $x \geq 0$ and $0 < t < 1$.

because

$$\frac{\partial}{\partial x} (x \log t + 1 - t^x)' = (1-t^x) \log t$$

and $\log t < 0$.

This above lemma may be used to yield the ollowing conclusions

(i) If Chotai's (1974) strategy fares better than hangurde and Rao's (1969) in case $h = 2$ it follows that ur strategy II* fares better than III in case $1 > \delta > \rho > 0$.

(ii) If Chotai's strategy fares worse than Ghangurde nd Rao (1969)'s in case $h = 2$ then our II* fares worse than III in case $1 > \rho > \delta > 0$.

CHAPTER - 7

ON SCHEMES FOR RETAINING SAMPLES FOR SUBSEQUENT SURVEYS UNDER DESIGN-CONSTRAINTS

7.1 INTRODUCTION AND SUMMARY

So far we have discussed various ways of controlling the magnitudes of errors in estimating population parameters. Our efforts have been directed towards reducing the variances of estimators for sampling strategies adopted on one or two occasions. Once such appropriate strategies are decided upon ensuring optimality of the variances, efforts may be directed towards controlling the expenditures involved in sampling also, particularly in repeated surveys. We shall be specially interested in maximizing the number of common units on an average in samples chosen on two occasions so that the average cost of surveys is reduced to the minimum at the same time variances for the estimators for the two occasions having also been kept under control in appropriate manners. Some works alsong this line are available in the literature as described below. To this we shall add a few more in what follows next.

Keyfitz (1951) gave a method to maximize the bability of retaining a unit already chosen for a latter vey such that its selection-probabilities on the two casions are proportional to size-measures permitted to different for the two occasions. *Fellegi* (1966) extended is to produce a method of selecting two units on two casions such that the inclusion probabilities of units d pairs of units stipulated to vary over two occasions e exactly realised and observed (on empirical evidence) s method often to approximately achieve a maximal inter-ection between the samples on an average for the two ccasions. Lanke (1975) extended Keyfitz's ideas further o devise a method of maximizing the expected size of the ntersection of the two samples each of size n ($\geq$ 2) for wo designs such that for either design their selection robabilities are proportional to the sums of the size-easures of the units in them on the respective occasions. In this chapter we suggest an analogous but slightly more eneral and flexible scheme than Fellegi's. Following Lanke (1975) we obtain the expressions for the expected size of the intersection in samples of two units chosen in two consecutive surveys for this scheme. We also note that Fellegi's result follows as a special case. We also examine its straight-forward extension to samples of arbitrary sizes. Incidentally, we also suggest a procedure simpler to execute than Fellegi's and having an identical expectation for the size of the intersection. We also offer a third scheme based on Keyfitz's ideas but applicable not for sampling on two occasions but for sampling a given population simultaneously to estimate parameters of two variates with different appropriate size-measures. It may then be desirable to choose two samples on the same occasion (with no time-lag in actual field investigation) each appropriate to estimate a separate single parameter but to have the maximal overlap between the samples to ensure the greatest cut in the budget. We use the symbols B_1 and B_2 to denote the designs for chosing the two samples for which selection-probabilities are appropriately fixed from size-measure considerations but we wish to adopt a scheme for which subject to these design-constraints determined from efficiency considerations the cost is suitably reduced by ensuring substantial intersection in the samples. Our proposed scheme (for $n = 2$, but also amenable to extension for $n > 2$, obviously following Lanke's procedure) is slightly complicated but the average size of intersection s worked out. Obviously the previously discussed schemes lso may well be applied to surve the present purpose of stimating two parameters on a single occasion alone from wo surveys on a given population. But we suggest our lternative because the expected value here works out ifferently and hence may well be considered for comparison ith the others. Detailed discussions are to follow.

1.2 KEYFITZ (1951) PRINCIPLE FOR CHOOSING A UNIT WITH VARYING PROBABILITIES ON TWO OCCASIONS

Let p_i and p_i' $(i = 1, \ldots, N)$ denote the normed size-measures for the ith unit of a given finite universe U of N units on the 1st and the 2nd occasion. To choose a unit with these probabilities on the two occasions Keyfitz's method is as follows.

$$\text{Let} \quad I = \{i \mid p_i \leq p_i'\}$$

$$\text{and} \quad D = \{i \mid p_i > p_i'\}.$$

On the 1st occasion a unit j (say) is chosen with probability p_j. If $j \varepsilon I$, then it is retained with probability 1 for the second occasion. If $j \varepsilon D$, then it is retained with probability p_i'/p_i. If it is discarded, then another unit, say, k is chosen from I on the second occasion with a probability proportional to $(p_k' - p_k)$. For this scheme the ith unit has the selection-probability as $p_i(p_i')$ for the 1st (2nd) occasion $(i = 1, \ldots, N)$. With this scheme the expected number of common units equals $\sum_i \text{Min}\,(p_i, p_i')$, as has been shown by Lanke (1975)

7.3 LANKE'S (1975) EXTENSIONS TO KEYFITZ PRINCIPLE.

Lanke's extension to choose samples of arbitrary sizes $n\ (> 1)$ is as follows. On the first occasion the ith unit is chosen with probability p_i and then an SRSWOR sample S_o' (say) of size $(n-1)$ is chosen from the population dropping the ith unit. The overall sample is $S_1 = i \cup S_o'$. On the second occasion, on the 1st draw a jth unit is select. following Keyfitz's principle on noting the selection of the unit on the first draw on the 1st occasion. The overall sample is then

$$\left.\begin{aligned} S_2 &= S_1 \\ &= j \cup S_o' \end{aligned}\right\} \begin{aligned} &\text{if } j \in S_1 \\ &\text{if } j \notin S_1 . \end{aligned}$$

Lanke (1975) proved that S_2/j is an SRSWOR from U/j. He also observed that the expected number of common units in the two samples, written as

$$D\,(S_1 \cap S_2) \quad \text{equals} \quad \sum_i \text{Min}\,(\pi_i, \pi_i'),$$

where π_i, π'_i are the inclusion-probabilities, which satisfy Midzuno-Sen formulae

$$\pi^*_i = \frac{n-1}{N-1} + \frac{N-n}{N-1} p^*_i$$

(where $\pi^*_i = \pi_i$ when $p^*_i = p_i$

and $\pi^*_i = \pi'_i$ when $p^*_i = p'_i$)

7.4 PROPOSED SAMPLING SCHEMES

We shall consider the following three sampling schemes :

7.4.1 Scheme I

The Scheme I is quite analogous to and is a slight extension of Fellegi's (1966) as described below :

On the first occasion following Chaudhuri (1971) the first unit i_1 is selected with probability $\pi_{i_1}/2$ and the second unit i_2 is selected from U/i_1 with probability $\pi_{i_1 i_2}/\pi_{i_1}$ - the sample drawn being an ordered one viz (i_1, i_2)

On the second occasion we select an ordered sample (j_1, j_2) say, of size 2 by using the following rules.

Rule 1

Divide the units of the population U in two groups viz I and D as

$$I = \{i \mid \frac{\pi_i}{2} \leq \frac{\pi'_i}{2}\}, \quad D = \{i \mid \frac{\pi_i}{2} > \frac{\pi'_i}{2}\}$$

and apply Keyfitz's rule to select the units on the first draw

Rule 2

If the first unit i_1 is retained for the second occasion i.e. if $i_1 = j_1$, we divide the units of U/i_1 into

$$I_{i_1} = \{j \mid \frac{\pi_{i_1 j}}{\pi_{i_1}} \leq \frac{\pi'_{i_1 j}}{\pi'_{i_1}}\} \text{ and } D_{i_1} = \{j \mid \frac{\pi_{i_1 j}}{\pi_{i_1}} > \frac{\pi'_{i_1 j}}{\pi'_{i_1}}\}$$

and use these groups to choose j_2 using Keyfitz's principle in terms of $\pi_{i_1 j}/\pi_{i_1}$ and $\pi'_{i_1 j}/\pi'_{i_1}$'s.

Rule 3

If the first unit is not retained i.e. $i_1 \neq j_1$ then for $j_1 \neq i_2$ we shall form the following groups as

$$I_{i_1 j_1} = \{i \neq i_1 \neq j_1 \mid \frac{\dfrac{\pi_{i_1 i}}{\pi_{i_1}}}{1 - \dfrac{\pi_{i_1 j_1}}{\pi_{i_1}}} \leq \frac{\pi'_{j_1 i}}{\pi'_{j_1}}\}$$

$$D_{i_1 j_1} = \{ i \neq i_1 \neq j_1 \mid \frac{\frac{\pi_{i_1 i}}{\pi_{i_1}}}{1 - \frac{\pi_{i_1 j_1}}{\pi_{i_1}}} > \frac{\pi'_{j_1 i}}{\pi'_{j_1}} \}$$

for $\forall\, j \neq i_2 \neq i_1$.

ow

a) if $i_2 \in I_{i_1 j_1}$, we retain it for the seco occasion with probability 1.

b) if $i_2 \in D_{i_1 j_1}$, we retain it with probability

$$\frac{\pi'_{j_1 i_2}}{\pi'_{j_1}} \Big/ \frac{\pi_{i_1 i_2}}{\pi_{i_1} - \pi_{i_1 j_1}} ,$$

f it is discarded, we select i_1 for the second occasion ith probability

$$\frac{\pi'_{j_1 i_1} / \pi'_{j_1}}{\sum\limits_{j \in D_{i_1 j_1}} \left(\frac{\pi_{i_1 j}}{\pi_{i_1} - \pi_{i_1 j_1}} - \frac{\pi'_{j_1 j}}{\pi'_{j_1}} \right)}$$

f i_1 is not selected we select a unit j_2 from $I_{i_1 j_1}$ with robability

$$\frac{1}{S_{i_1 j_1}} \left(\frac{\pi'_{j_1 j_2}}{\pi'_{j_1}} - \frac{\pi_{i_1 j_2}}{\pi_{i_1} - \pi_{i_1 j_2}} \right)$$

where

$$S_{i_1 j_1} = \sum_{j \varepsilon I_{i_1 j_1}} \left(\frac{\pi'_{j_1 j}}{\pi'_{j_1}} - \frac{\pi_{i_1 j}}{\pi_{i_1} - \pi_{i_1 j}} \right)$$

Rule 4

In the case where $j_1 \neq i_1$ but $j_1 = i_2$, then we select j_2 from U/i_2 with probability $\pi'_{i_2 j_2}/\pi'_{i_2}$.

Now we shall prove the following theorem :

Theorem 7.1

The selection probability of the ordered sample (j_1, j_2) is $\pi'_{j_1 j_2}/2$.

Proof

By Keyfitz's principle we note that the probability of selecting j_1 in the first draw is $\pi'_{j_1}/2$. To select j_2 on the second occasion we consider the following possibilities

i) $j_1 = i_1$

ii) $j_1 \neq i_1$, but $j_1 = i_2$

iii) $j_1 \neq i_1 \neq i_2$;

(i) when $j_1 = i_1$, the probability of selecting j_2 $(\neq j_1)$ is $\pi'_{j_1 j_2}/\pi'_{j_1}$ (by applying Keyfitz's principle)

i) when $j_1 \neq i_1$ but $j_1 = i_2$, then the probability is

$_2/\pi'_{j_1}$ for $j_2 \neq i_2$.

i) on the 1st occasion, the conditional probability

electing i_2 $(\neq j_1)$ when i_1 $(\neq j_1)$ is already selected, is

$$\frac{\pi_{i_1 i_2}/\pi_{i_1}}{\sum_{j \neq i_1} \frac{\pi_{i_1 j}}{\pi_{i_1}}} = \frac{\pi_{i_1 i_2}}{\pi_{i_1} - \pi_{i_1 j_1}}$$

ppose $j_2 \in I_{i_1 j_1}$, then on the second occasion the

obability of selecting j_2 given that j_1 is already

lected is

$$\frac{\pi_{i_1 j_2}}{\pi_{i_1} - \pi_{i_1 j_1}} + \sum_{i_2 \in D_{i_1 j_1}} \frac{\pi_{i_1 i_2}}{\pi_{i_1} - \pi_{i_1 j_1}}$$

$$\left(1 - \frac{\pi'_{j_1 i_2}/\pi'_{j_1}}{\pi_{i_1 i_2}/(\pi_{i_1} - \pi_{i_1 j_1})}\right)$$

$$\left(1 - \frac{\pi'_{j_1 i_1}/\pi'_{j_1}}{\sum_{j \in D_{i_1 j_1}} \left(\frac{\pi_{i_1 j}}{\pi_{i_1} - \pi_{i_1 j_1}} - \frac{\pi'_{j_1 j}}{\pi'_{j_1}}\right)}\right)$$

$$\cdot \left(\frac{\pi'_{j_1 j_2}}{\pi'_{j_1}} - \frac{\pi_{i_1 j_2}}{\pi_{i_1} - \pi_{i_1 j_1}}\right) \frac{1}{S_{i_1 j_1}}$$

$$= \frac{\pi_{i_1 j_2}}{\pi_{i_1} - \pi_{i_1 j_1}} + \sum_{i_2 \in D_{i_1 j_1}} \left(\frac{\pi_{i_1 i_2}}{\pi_{i_1} - \pi_{i_1 j_1}} - \frac{\pi'_{j_1 i_2}}{\pi'_{j_1}} \right) \cdot$$

$$\left[\sum_{j \in D_{i_1 j_1}} \left(\frac{\pi_{i_1 j}}{\pi_{i_1} - \pi_{i_1 j_1}} - \frac{\pi'_{j_1 j}}{\pi'_{j_1}} \right) - \frac{\pi'_{j_1 i_1}}{\pi'_{j_1}} \right]$$

$$\cdot \frac{\dfrac{\pi'_{j_1 j_2}}{\pi'_{j_1}} - \dfrac{\pi_{i_1 j_2}}{\pi_{i_1} - \pi_{i_1 j_2}}}{S_{i_1 j_1}} \cdot \frac{1}{\sum_{j \in D_{i_1 j_1}} \left(\dfrac{\pi_{i_1 j}}{\pi_{i_1} - \pi_{i_1 j_1}} - \dfrac{\pi'_{j_1 j}}{\pi'_{j_1}} \right)}$$

$$= \frac{\pi'_{j_1 j_2}}{\pi'_{j_1}}$$

[Since $\sum_{j \in D_{i_1 j_1}} \left(\frac{\pi_{i_1 j}}{\pi_{i_1} - \pi_{i_1 j_1}} - \frac{\pi'_{j_1 j}}{\pi'_{j_1}} \right)$

$$= - S_{i_1 j_1} + \frac{\pi'_{j_1 i_1}}{\pi'_{j_1}}]$$

When $j_2 \in D_{i_1 j_1}$ the probability of retaining j_2 given that i_1 ($\neq j_1$) already selected for first occasion is

$$\frac{\pi_{i_1 j_2}}{\pi_{i_1} - \pi_{i_1 j_1}} \cdot \frac{\pi'_{j_1 j_2} / \pi'_{j_1}}{\pi_{i_1 j_2} / (\pi_{i_1} - \pi_{i_1 j_1})} = \frac{\pi'_{j_1 j_2}}{\pi'_{j_1}} .$$

When $j_2 = i_1$, then probability of selecting j_2 when j_1 is already selected is

$$\sum_{i_2 \varepsilon D_{i_1 j_1}} \frac{\pi_{i_1 i_2}}{\pi_{i_1} - \pi_{i_1 j_1}} \left(1 - \frac{\pi'_{j_1 i_2}/\pi'_{j_1}}{\pi_{i_1 i_2}/(\pi_{i_1} - \pi_{i_1 j_1})}\right).$$

$$\frac{\pi'_{j_1 i_1}/\pi'_{j_1}}{\sum_{j \varepsilon D_{i_1 j_1}} \left(\frac{\pi_{i_1 j}}{\pi_{i_1} - \pi_{i_1 j_1}} - \frac{\pi'_{j_1 j}}{\pi'_{j_1}}\right)}$$

$$= \frac{\pi'_{j_1 i_1}}{\pi'_{j_1}} \qquad \text{Q.E.D.}$$

<u>Remark</u>

In the above π_i, π'_i 's are arbitrary positive quantities subject to

$$\sum_i \pi_i = 2 = \sum_i \pi'_i$$

and π_{ij}, π'_{ij}'s are arbitrary positive quantities subject to the restrictions

$$\sum_{j \neq i} \pi_{ij} = \pi_i \quad \text{and} \quad \sum_{j \neq i} \pi'_{ij} = \pi'_i \qquad \forall\ i = 1, \ldots, N.$$

When in particular,

$$\pi_i = 2p_i \quad \text{and} \quad \pi'_i = 2q_i$$

and

$$\pi_{ij} = \frac{1}{2}\ p_i\ \frac{q_j}{1-q_i} = \frac{1}{2}\ q_i\ \frac{p_j}{1-p_i}$$

$$\pi'_{ij} = \frac{1}{2}\ P_i\ \frac{Q_j}{1-Q_i} = \frac{1}{2}\ Q_i\ \frac{P_j}{1-P_i}$$

for

$$0 < p_i,\quad q_i < 1,\qquad \Sigma\ p_i = \Sigma\ q_i = 1,$$

$$0 < P_i,\quad Q_i < 1,\qquad \Sigma\ P_i = \Sigma\ Q_i = 1,$$

the above SCHEME-I reduces to Fellegi's (1966) scheme, ours being a slight generalization over the latter introduced with a view to having more flexibility with the possibility of enabling us to have a greater control on the sampling scheme which may lead to an increase in the average overlap for the two occasions. To calculate the expected number of common units between the samples (i_1, i_2) and (j_1, j_2) we first find the following probabilities using obvious notations following Lanke (1975).

$$P_1 = P\ \{i_1 = j_1,\quad i_2 = j_2\}$$
$$P_2 = P\ \{i_1 = j_1,\quad i_2 \neq j_2\}$$
$$P_3 = P\ \{i_1 \neq j_1,\quad i_2 = j_1,\quad j_2 = i_1\}$$
$$P_4 = P\ \{i_1 \neq j_1,\quad i_2 = j_1,\quad j_2 \neq i_1\}$$
$$P_5 = P\ \{i_1 \neq j_1,\quad i_2 \neq j_1,\quad j_2 = i_1\}$$
$$P_6 = P\ \{i_1 \neq j_1,\quad i_2 \neq j_1,\quad i_2 = j_2\}$$

Calculation of P_1

Here we have the following possibilities

i) $i_1 \varepsilon I$, $i_2 \varepsilon I_{i_1}$

ii) $i_1 \varepsilon I$, $i_2 \varepsilon D_{i_1}$

iii) $i_1 \varepsilon D$, $i_2 \varepsilon I_{i_1}$

iv) $i_1 \varepsilon D$, $i_2 \varepsilon D_{i_1}$

Hence,

$$P_1 = \frac{1}{2} \sum_{i_1 \varepsilon I} \pi_{i_1} \left\{ \sum_{i_2 \varepsilon I_{i_1}} \frac{\pi_{i_1 i_2}}{\pi_{i_1}} + \sum_{i_2 \varepsilon D_{i_1}} \frac{\pi'_{i_1 i_2}}{\pi'_{i_1}} \right\}$$

$$+ \frac{1}{2} \sum_{i_1 \varepsilon D} \pi'_{i_1} \left\{ \sum_{i_2 \varepsilon I_{i_1}} \frac{\pi_{i_1 i_2}}{\pi_{i_1}} + \sum_{i_2 \varepsilon D_{i_1}} \frac{\pi'_{i_1 i_2}}{\pi_{i_1}} \right\}$$

$$= \frac{1}{2} \sum_{i} \text{Min} (\pi_i, \pi'_i) \sum_{j \neq i} \text{Min} \left(\frac{\pi_{ij}}{\pi_i} , \frac{\pi'_{ij}}{\pi'_i} \right)$$

Calculation of P_2

Here we have the following possibilities

$i_1 \varepsilon I$, $i_2 \varepsilon D_{i_1}$, $j_2 \varepsilon I_{i_1}$

$i_1 \varepsilon D$, $i_2 \varepsilon D_{i_1}$, $j_2 \varepsilon I_{i_1}$

Hence,

$$P_2 = \frac{1}{2} \sum_{i_1 \varepsilon I} \pi_{i_1} \sum_{i_2 \varepsilon D_{i_1}} \frac{\pi_{i_1 i_2}}{\pi_{i_1}} \left(1 - \frac{\pi'_{i_1 i_2}/\pi'_{i_1}}{\dfrac{\pi_{i_1 i_2}}{\pi_{i_1}}}\right)$$

$$\sum_{j_2 \varepsilon I_{i_1}} \left(\frac{\pi'_{i_1 j_2}}{\pi'_{i_1}} - \frac{\pi_{i_1 j_2}}{\pi_{i_1}}\right) \Big/ \sum_{j \varepsilon D_{i_1}} \left(\frac{\pi'_{i_1 j}}{\pi'_{i_1}} - \frac{\pi_{i_1 j}}{\pi_{i_1}}\right)$$

$$+ \frac{1}{2} \sum_{i_1 \varepsilon D} \pi'_{i_1} \sum_{i_2 \varepsilon D_{i_1}} \frac{\pi_{i_1 i_2}}{\pi_{i_1}} \left(1 - \frac{\pi'_{i_1 i_2}/\pi'_{i_1}}{\pi_{i_1 i_2}/\pi_{i_1}}\right)$$

$$\sum_{j_2 \varepsilon I_{i_1}} \left(\frac{\pi'_{i_1 j_2}}{\pi'_{i_1}} - \frac{\pi_{i_1 j_2}}{\pi_{i_1}}\right) \Big/ \sum_{j \varepsilon D_{i_1}} \left(\frac{\pi'_{i_1 j}}{\pi'_{i_1}} - \frac{\pi_{i_1 j}}{\pi_{i_1}}\right)$$

$$= \frac{1}{2} \sum_{i_1 \varepsilon I} \pi_{i_1} \sum_{i_2 \varepsilon D_{i_1}} \left(\frac{\pi_{i_1 i_2}}{\pi_{i_1}} - \frac{\pi'_{i_1 i_2}}{\pi'_{i_1}}\right)$$

$$+ \frac{1}{2} \sum_{i_1 \varepsilon D} \pi'_{i_1} \sum_{i_2 \varepsilon D_{i_1}} \left(\frac{\pi_{i_1 i_2}}{\pi_{i_1}} - \frac{\pi'_{i_1 i_2}}{\pi'_{i_1}}\right)$$

$$= \frac{1}{2} \sum_{i} \text{Min}\,(\pi_i, \pi'_i) \sum_{j \varepsilon D_i} \left(\frac{\pi_{ij}}{\pi_i} - \frac{\pi'_{ij}}{\pi'_i}\right)$$

Calculation of P_3

Here $i_1 \varepsilon D$, $j_1 \varepsilon I$.

$$P_3 = \sum_{i_1 \varepsilon D} \frac{\pi_{i_1}}{2} \left(1 - \frac{\pi'_{i_1}}{\pi_{i_1}}\right) \sum_{j_1 \varepsilon I} \frac{\pi'_{j_1} - \pi_{j_1}}{A} \frac{\pi_{i_1 j_1}}{\pi_{i_1}} \frac{\pi'_{j_1 i_1}}{\pi'_{j_1}}$$

$$\left[\text{where } 2A = \sum_{i \varepsilon I} (\pi'_i - \pi_i)\right]$$

$$= \frac{1}{2} \sum_{i \varepsilon D} \frac{(\pi_i - \pi'_i)}{\pi_i} \sum_{j \varepsilon I} \pi_{ij} \frac{(\pi'_j - \pi_j)}{A} \frac{\pi'_{ji}}{\pi'_j}$$

Calculation of P_4

Here $i_1 \varepsilon D$, $j_1 = i_2 \varepsilon I$

and

$$P_4 = \frac{1}{2} \sum_{i_1 \varepsilon D} \pi_{i_1} \left(1 - \frac{\pi'_{i_1}}{\pi_{i_1}}\right) \sum_{i_2 \varepsilon I} \frac{\pi'_{i_2} - \pi_{i_2}}{A} \frac{\pi_{i_1 i_2}}{\pi_{i_1}} \left(1 - \frac{\pi'_{i_2 i_1}}{\pi'_{i_2}}\right)$$

$$= \frac{1}{2} \sum_{i \varepsilon D} \frac{\pi_i - \pi'_i}{\pi_i} \sum_{j \varepsilon I} \frac{\pi_{ij}}{\pi'_j} \frac{(\pi'_j - \pi_j)}{A} (\pi'_j - \pi'_{ji})$$

Calculation of P_5

Here $i_1 \varepsilon D$, $j_1 \varepsilon I$, $i_2 \varepsilon D_{i_1}$, $j_2 = i_1$

and

$$P_5 = \frac{1}{2} \sum_{i_1 \varepsilon D} \pi_{i_1} \left(1 - \frac{\pi'_{i_1}}{\pi_{i_1}}\right) \sum_{j_1 \varepsilon I} \frac{\pi'_{j_1} - \pi_{j_1}}{A}$$

$$\sum_{i_2 \varepsilon D_{i_1}} \frac{\pi_{i_1 i_2}}{\pi_{i_1}} \left(1 - \frac{\pi'_{j_1 i_2}/\pi'_{j_1}}{\pi_{i_1 i_2}/(\pi_{i_1} - \pi_{i_1 j_1})}\right)$$

$$\cdot \frac{\pi'_{j_1 i_1}/\pi'_{j_1}}{\sum_{j \varepsilon D_{i_1 j_1}} \left(\frac{\pi_{i_1 j}}{\pi_{i_1} - \pi_{i_1 j}} - \frac{\pi'_{j_1 j}}{\pi'_{j_1}} \right)}$$

$$= \frac{1}{2} \sum_{i \varepsilon D} \frac{(\pi_i - \pi'_i)}{\pi_i} \sum_{j \varepsilon I} \frac{\pi'_j - \pi_j}{A} \frac{\pi'_{ji}}{\pi'_j} (\pi_i - \pi_{ij})$$

Calculation of P_6 .

Here we have two possibilities viz

i) $i_1 \varepsilon D, \quad j_1 \varepsilon I, \quad i_2 \varepsilon I_i, \quad j_2 = i_2$

ii) $i_1 \varepsilon D, \quad j_1 \varepsilon I, \quad i_2 \varepsilon D_{i_1}, \quad j_2 = i_2$

So,

$$P_6 = \frac{1}{2} \sum_{i_1 \varepsilon D} \pi_{i_1} \left(1 - \frac{\pi'_{i_1}}{\pi_{i_1}}\right) \sum_{j_1 \varepsilon I} \frac{\pi'_{j_1} - \pi_{j_1}}{A} \sum_{i_2 \varepsilon I_{i_1 j_1}} \frac{\pi_{i_1 i_2}}{\pi_{i_1}}$$

$$+ \frac{1}{2} \sum_{i_1 \varepsilon D} \pi_{i_1} \left(1 - \frac{\pi'_{i_1}}{\pi_{i_1}}\right) \sum_{j_1 \varepsilon I} \frac{\pi'_{j_1} - \pi_{j_1}}{A}$$

$$\sum_{i_2 \varepsilon D_{i_1 j_1}} \frac{\pi_{i_1 i_2}}{\pi_{i_1}} \frac{\pi'_{j_1 i_2}/\pi'_{j_1}}{\pi_{i_1 i_2}/(\pi_{i_1} - \pi_{i_1 j_1})}$$

$$= \frac{1}{2} \sum_{i \varepsilon D} (\pi_i - \pi'_i) \sum_{j \varepsilon I} \frac{\pi'_j - \pi_j}{A}$$

$$\left\{ \sum_{k \varepsilon I_{ij}} \frac{\pi_{ik}}{\pi_i} + \sum_{k \varepsilon D_{ij}} \frac{\pi_i - \pi_{ij}}{\pi_i} \cdot \frac{\pi'_{kj}}{\pi'_j} \right\}$$

$$= \frac{1}{2} \sum_{i \varepsilon D} (\pi_i - \pi_i') \sum_{j \varepsilon I} \frac{\pi_j' - \pi_j}{A} \left(1 - \frac{\pi_{ij}}{\pi_i}\right)$$

$$\sum_{k \neq i \neq j} \text{Min}\left(\frac{\pi_{ik}}{\pi_i - \pi_{ij}}, \frac{\pi_{jk}'}{\pi_j'}\right)$$

So, expected number of common units between the first and second occasion for the SCHEME-I is

$$\nu_1 = 2(P_1+P_3) + P_2 + P_4 + P_5 + P_6$$

$$= \sum_i \text{Min}(\pi_i, \pi_i') \sum_{j \neq i} \text{Min}\left(\frac{\pi_{ij}}{\pi_i}, \frac{\pi_{ij}'}{\pi_i}\right)$$

$$+ \sum_{i \varepsilon D} \frac{\pi_i - \pi_i'}{\pi_i} \sum_{j \varepsilon I} \pi_{ij} \frac{\pi_j' - \pi_j}{A} \frac{\pi_{ij}'}{\pi_j'}$$

$$+ \frac{1}{2} \sum_i \text{Min}(\pi_i, \pi_i') \sum_{j \varepsilon D_i} \left(\frac{\pi_{ij}}{\pi_i} - \frac{\pi_{ij}'}{\pi_i'}\right)$$

$$+ \frac{1}{2} \sum_{i \varepsilon D} \frac{\pi_i - \pi_i'}{\pi_i} \sum_{j \varepsilon I} \frac{(\pi_j' - \pi_j)}{A} \frac{\pi_{ij}}{\pi_j'} (\pi_j' - \pi_{ji}')$$

$$+ \frac{1}{2} \sum_{i \varepsilon D} \frac{(\pi_i - \pi_i')}{\pi_i} \sum_{j \varepsilon I} \frac{\pi_j' - \pi_j}{A} \frac{\pi_{ji}'}{\pi_j'} (\pi_i - \pi_{ij})$$

$$+ \frac{1}{2} \sum_{i \varepsilon D} (\pi_i - \pi_i') \sum_{j \varepsilon I} \frac{\pi_j' - \pi_j}{A} \left(1 - \frac{\pi_{ij}}{\pi_i}\right)$$

$$\sum_{k \neq i \neq j} \text{Min}\left(\frac{\pi_{ik}}{\pi_i - \pi_{ij}}, \frac{\pi_{jk}'}{\pi_j'}\right)$$

$$= \sum_{i} \text{Min}\,(\pi_i, \pi'_i) \sum_{j \neq i} \text{Min}\,(\frac{\pi_{ij}}{\pi_i}, \frac{\pi'_{ij}}{\pi'_i})$$

$$+ \frac{1}{2} \sum_{i} \text{Min}\,(\pi_i, \pi'_i) \sum_{j \varepsilon D_i} (\frac{\pi_{ij}}{\pi_i} - \frac{\pi'_{ij}}{\pi'_i})$$

$$+ \frac{1}{2} \sum_{i \varepsilon D} (\pi_i - \pi'_i) \sum_{j \varepsilon I} \frac{\pi'_j - \pi_j}{A} (\frac{\pi_{ij}}{\pi_i} + \frac{\pi'_{ij}}{\pi'_j})$$

$$+ \frac{1}{2} \sum_{i \varepsilon D} (\pi_i - \pi'_i) \sum_{j \varepsilon I} \frac{\pi'_j - \pi_j}{A} (1 - \frac{\pi_{ij}}{\pi_i})$$

$$\sum_{k \neq i \neq j} \text{Min}\,(\frac{\pi_{ik}}{\pi_i - \pi_{ij}} \quad \frac{\pi'_{jk}}{\pi'_j})$$

7.4.2 SCHEME - II

The ordered sample (i_1, i_2) on the first occasion is chosen in the same manner as in SCHEME-I.

On the second occasion, of the ordered pair (j_1, j_2), the first unit j_1 is also chosen in the same way as in the SCHEME-1. The unit j_2 is chosen in the same manner as in the SCHEME-1 either when the first unit i_1 of first occasion is retained (i.e., $i_1 = j_1$) on the 2nd occasion or when $j_1 = i_2$. But if neither happens, i.e. if $i_1 \neq j_1$ and $j_1 \neq i_2$, we proceed as follows.

We define two sets of numbers r_j and r'_j as follows:

$$r_j = \frac{\pi_{i_1 j}}{\pi_{i_1} - \pi_{i_1 j_1}} \quad \text{for } j \neq i_1 \neq j_1$$
$$= 0 \quad \text{for } j = i_1, j_1$$

$$r'_j = \frac{\pi'_{i_1 j}}{\pi'_{j_1}} \quad \text{for } j \neq j_1$$
$$= 0 \quad \text{for } j = j_1$$

and form the two following groups.

$$I'_{i_1 j_1} = \{ j \mid r_j \leq r'_j \}$$

$$D'_{i_1 j_1} = \{ j \mid r_j > r'_j \} .$$

Now, we use these groups to apply Keyfitz's principle to choose j_2 with probability r'_{j_2} on the second draw of second occasion given that i_2 was selected on the second draw of first occasion. Here also it is easy to show that the probability of selecting the ordered sample (j_1, j_2) is $\pi'_{j_1 j_2}/2$.

To calculate the expected number of common units we are to find the following probabilities.

$$P'_1 = P\{i_1 = j_1, \quad i_2 = j_2\}$$

$$P_2 = P\{i_1 = j_1, \quad i_2 \neq j_2\}$$

$$P'_3 = P\{i_1 \neq j_1, \quad i_2 = j_1, \quad j_2 = i_1\}$$

$$P'_4 = P\{i_1 \neq j_1, \quad i_2 = j_1, \quad j_2 \neq i_1\}$$

$$P'_5 = P\{i_1 \neq j_1, \quad i_2 \neq j_1, \quad j_2 = i_1\}$$

$$P'_6 = P\{i_1 \neq j_1, \quad i_2 \neq i_1, \quad j_2 = i_2\}$$

Here $P'_i = P_i$ for $i = 1,2,3,4$ because in these cases samples were selected in the same way as in SCHEME-1. Further we note that

$$P'_5 = \frac{1}{2} \sum_{i_1 \in D} \pi_{i_1} \left(1 - \frac{\pi'_{i_1}}{\pi_{i_1}}\right) \sum_{j_1 \in I} \frac{\pi'_{j_1} - \pi_{j_1}}{A}$$

$$\sum_{i_2 \in D'_{i_1 j_1}} \frac{\pi_{i_1 i_2}}{\pi_{i_1}} \left(1 - \frac{\pi'_{j_1 i_2}/\pi'_{j_1}}{\dfrac{\pi_{i_1 i_2}}{\pi_{i_1} - \pi_{i_1 j_1}}}\right)$$

$$\frac{\pi'_{j_1 i_1}/\pi'_{j_1}}{\sum\limits_{j \in D'_{i_1 j_1}} \left(\dfrac{\pi'_{j_1 j}}{\pi'_{j_1}} - \dfrac{\pi_{i_1 j}}{\pi_{i_1} - \pi_{i_1 j_1}}\right)}$$

$$= \frac{1}{2} \sum_{i \in D} \frac{\pi_i - \pi'_i}{\pi_i} \sum_{j \in I} (\pi_i - \pi_{ij}) \frac{\pi'_{ji}}{\pi'_j} \frac{\pi'_j - \pi_j}{A}$$

$$= P_5$$

nd $$P'_6 = \frac{1}{2} \sum_{i_1 \in D} \pi_{i_1} \left(1 - \frac{\pi'_{i_1}}{\pi_{i_1}}\right) \sum_{j_1 \in I'} \frac{\pi'_{j_1} - \pi_{j_1}}{A} \sum_{i_2 \in I'_{i_1 j_1}} \frac{\pi_{i_1 i_2}}{\pi_{i_1}}$$

$$+ \frac{1}{2} \sum_{i_1 \in D} \pi_{i_1} \left(1 - \frac{\pi'_{i_1}}{\pi_{i_1}}\right) \sum_{j_1 \in I} \frac{\pi'_{j_1} - \pi_{j_1}}{A}$$

$$\cdot \sum_{j_2 \in D'_{i_1 j_1}} \frac{\pi_{i_1 i_2}}{\pi_{i_1}} \cdot \frac{\dfrac{\pi'_{j_1 i_2} / \pi'_{j_1}}{\pi_{i_1 i_2}}}{\pi_{i_1} - \pi_{i_1 j_1}}$$

$$= \frac{1}{2} \sum_{i \in D} (\pi_i - \pi'_i) \sum_{j \in I} \frac{\pi'_j - \pi_j}{A} \sum_{k \neq i \neq j} \left(\frac{\pi_i - \pi_{ij}}{\pi_i}\right)$$

$$\text{Min}\left(\frac{\pi_{ik}}{\pi_i - \pi_{ij}}, \frac{\pi'_{jk}}{\pi'_j}\right)$$

$$= P_6$$

Hence SCHEME-I and SCHEME-II are equivalent in the sense of yielding the same expected number of common units.

7.4.3 SCHEME - III

Here we select samples as follows .

Rule 1

We select i_1 for the design B_1 with probability $\pi_{i_1}/2$, then we select a unit j_1 for the design B_2 by applying Keyfitz's rule with probability $\pi'_{j_1}/2$.

Rule 2

(a) If i_1 is retained i.e. $i_1 = j_1$, select another unit i_2 ($\neq i_1$) from U/i_1 with probability $\pi_{i_1 i_2}/\pi_{i_1}$ for B_1 and apply Keyfitz rule again to select j_2 for B_2 with probability $\pi'_{j_1 j_2}/\pi'_{j_1}$, $(j_1 = i_1)$.

(b) If i_1 is not retained i.e. $i_1 \neq j_1$, then $i_1 \in D$ and $j_1 \in I$ [where $I = (i | \frac{\pi_i}{2} \leq \frac{\pi'_i}{2})$, $D = (i | \frac{\pi_i}{2} > \frac{\pi'_i}{2})$, $A = \sum_{i \in I} \frac{(\pi'_i - \pi_i)}{2}$] , then we define

$$I'_{i_1} = \{ j \mid \frac{\pi'_j - \pi_j}{A} = r'_j \leq \frac{\pi_{i_1 j}}{\pi_{i_1}} = r_j(i_1) \}$$

$$D'_{i_1} = \{ j \mid \frac{\pi'_j - \pi_j}{A} = r'_j > \frac{\pi_{i_1 j}}{\pi_{i_1}} = r_j(i_1) \}$$

(assuming $r'_j = 0$ for $j \in D$ and $r_j(i_1) = 0$ for $j = i_1$

If $j_1 \in I'_{i_1}$, then we retain j_1 as the second unit of B_1 (i.e. $i_2 = j_1$), on the other hand if $j_1 \in D'_{i_1}$ it is retained with probability $r_{j_1}(i)/r'_{j_1}$ and if it is rejected we select i_2 for B_1 from I'_{i_1} with probability

$$\frac{r_{i_2}(i_1) - r'_{i_2}}{A'_{i_1} = \sum_{i_2 \in I'_{i_1}} (r_{i_2}(i_1) - r'_{i_2})}$$

Rule 3

(a) If j_1 is retained i.e. if $j_1 = i_2$, then we select j_2 ($\neq j_1$) for B_2 with probability $\pi'_{j_1 j_2}/\pi'_{j_1}$

(b) If j_1 is not retained, then clearly $j_1 \in D'_{i_1}$ and $i_2 \in I'_{i_1}$ and we form two groups as

$$I''_{i_1 j_1} = \{j \mid \frac{r_j(i) - r'_j}{A'_{i_1}} = p_g(i_1) \leq \frac{\pi'_{j_1 j}}{\pi'_{j_1}} = p'_j(j_1)\}$$

$$D''_{i_1 j_1} = \{j \mid \frac{r_j(i) - r'_j}{A'_{i_1}} = p_j(i_1) > \frac{\pi'_{j_1 j}}{\pi'_{j_1}} = p'_j(j_1)\}$$

[Here we assume $p_j(i_1) = 0$ for $j \in D'_{i_1}$ and $p'_j(j_1) = 0$ for $j = j_1$]

Now if (i) $i_2 \in I''_{i_1 j_1}$, then we take j_2 as retained for B i.e. $j_2 = i_2$ with probability 1.

if (ii) $i_2 \varepsilon D''_{i_1 j_1}$, then i_2 is retained with probabilit $p'_{i_2}(j_1)/p_{i_2}(i_1)$.

if (iii) $i_2 \varepsilon D''_{i_1 j_1}$, and if itis not retained then we select j_2 from $I''_{i_1 j_1}$ with probability

$$\frac{p'_{j_2}(j_1) - p_{j_2}(i_1)}{A'(i_1, j_1)}$$

with

$$A'(i_1, j_1) = \sum_{j \varepsilon I''_{i_1 j_1}} (p'_j(j_1) - p_j(i_1))$$

Here we can check that the selection probability of the ordered samples (i_1, i_2) for the design B_1 and (j_1, j_2) for the design B_2 are $\pi_{i_1 i_2}/2$ and $\pi'_{j_1 j_2}/2$ respectively.

To calculate the expected number of common units between (i_1, i_2) and (j_1, j_2) we are to find out the following probabilities.

$$P''_1 = P\{i_1 = j_1, \quad i_2 = j_2\}$$

$$P''_2 = P\{i_1 = j_1, \quad i_2 \neq j_2\}$$

$$P''_3 = P\{i_1 \neq j_1, \quad j_1 = i_2, \quad j_2 = i_1\}$$

$$P''_4 = P\{i_1 \neq j_1, \quad j_1 = i_2, \quad j_2 \neq i_1\}$$

$$P''_5 = P\{i_1 \neq j_1, \quad j_1 \neq i_2, \quad j_2 = i_1\}$$

$$P''_6 = P\{i_1 \neq j_1, \quad j_1 \neq i_2, \quad j_2 = i_2\}$$

$$P_1^* = \frac{1}{2} \sum_{i_1 \in I} \pi_{i_1} \left\{ \sum_{i_2 \in I'_{i_1}} \frac{\pi_{i_1 i_2}}{\pi_{i_1}} + \sum_{i_2 \in D'_{i_1}} \frac{\pi'_{i_1 i_2}}{\pi'_{i_1}} \right\}$$

$$+ \frac{1}{2} \sum_{i_1 \in D} \pi'_{i_1} \left\{ \sum_{i_2 \in I'_{i_1}} \frac{\pi_{i_1 i_2}}{\pi_{i_1}} + \sum_{i_2 \in D'_{i_1}} \frac{\pi'_{i_1 i_2}}{\pi'_{i_1}} \right\}$$

$$+ \frac{1}{2} \sum_{i} \mathrm{Min}\,(\pi_i, \pi'_i) \sum_{j \neq i} \mathrm{Min}\left(\frac{\pi_{ij}}{\pi_i}, \frac{\pi'_{ij}}{\pi'_i}\right)$$

$$= P_1$$

$$P_2'' = \frac{1}{2} \sum_{i_1 \in I} \frac{\pi_{i_1}}{2} \sum_{i_2 \in D'_{i_1}} \frac{\pi_{i_1 i_2}}{\pi_{i_1}} \left(1 - \frac{\pi'_{i_1 i_2}}{\pi'_{i_1}} \Big/ \frac{\pi_{i_1 i_2}}{\pi_{i_1}}\right)$$

$$\cdot \frac{\sum_{j_2 \in I'_{i_1}} \left(\frac{\pi'_{i_1 j_2}}{\pi'_{i_1}} - \frac{\pi_{i_1 j_2}}{\pi_{i_1}}\right)}{A'_{i_1}}$$

$$+ \frac{1}{2} \sum_{i_1 \in D} \pi_{i_1} \frac{\pi'_{i_1}}{\pi_{i_1}} \sum_{i_2 \in D'_{i_1}} \frac{\pi_{i_1 i_2}}{\pi_{i_1}} \left(1 - \frac{\pi'_{i_1 i_2}}{\pi'_{i_1}} \Big/ \frac{\pi_{i_1 i_2}}{\pi_{i_1}}\right)$$

$$\cdot \frac{\sum_{j_2 \in I'_{i_1}} \left(\frac{\pi'_{i_1 j_2}}{\pi'_{i_1}} - \frac{\pi_{i_1 j_2}}{\pi_{i_1}}\right)}{A'_{i_1}}$$

$$= \frac{1}{2} \sum_{i} \mathrm{Min}\,(\pi_i, \pi'_i) \sum_{j \in D_i} \left(\frac{\pi_{ij}}{\pi_i} - \frac{\pi'_{ij}}{\pi'_i}\right)$$

$$= P_2 \,.$$

$$P_3'' = \frac{1}{2} \sum_{i_1 \varepsilon D} \pi_{i_1} \left(1 - \frac{\pi'_{i_1}}{\pi_{i_1}}\right) \sum_{j_1 \varepsilon I'_{i_1}} \frac{\pi'_{j_1} - \pi_{j_1}}{A} \frac{\pi'_{j_1 i_1}}{\pi'_{j_1}}$$

$$+ \frac{1}{2} \sum_{i_1 \varepsilon D} \pi_{i_1} \left(1 - \frac{\pi'_{i_1}}{\pi_{i_1}}\right) \sum_{j_1 \varepsilon D'_{i_1}} \frac{\pi'_{j_1} - \pi_{j_1}}{A}$$

$$\frac{\pi_{i_1 j_1}/\pi_{i_1}}{\frac{\pi'_{j_1} - \pi_{j_1}}{A}} \quad \frac{\pi'_{j_1 i_1}}{\pi'_{j_1}}$$

$$= \frac{1}{2} \sum_{i \varepsilon D} (\pi_i - \pi'_i) \sum_{j \neq i} \mathrm{Min}\left(\frac{\pi'_j - \pi_j}{A}, \frac{\pi_{ij}}{\pi_i}\right) \frac{\pi'_{ji}}{\pi'_j}$$

$$P_4'' = \frac{1}{2} \sum_{i_1 \varepsilon D} \pi_{i_1} \left(1 - \frac{\pi'_{i_1}}{\pi_{i_1}}\right) \sum_{j_1 \varepsilon I'_{i_1}} \frac{\pi'_{j_1} - \pi_{j_1}}{A} \left(1 - \frac{\pi'_{j_1 i_1}}{\pi'_{j_1}}\right)$$

$$+ \frac{1}{2} \sum_{i_1 \varepsilon D} \pi_{i_1} \left(1 - \frac{\pi'_{i_1}}{\pi_{i_1}}\right) \sum_{j_1 \varepsilon D'_{i_1}} \frac{\pi_{i_1 j_1}}{\pi_{i_1}} \left(1 - \frac{\pi'_{j_1 i_1}}{\pi'_{j_1}}\right)$$

$$= \frac{1}{2} \sum_{i \varepsilon D} (\pi_i - \pi'_i) \sum_{j \neq i} \mathrm{Min}\left(\frac{\pi'_j - \pi_j}{A}, \frac{\pi_{ij}}{\pi_i}\right) \left(1 - \frac{\pi'_{ji}}{\pi'_j}\right)$$

$$P_5'' = \frac{1}{2} \sum_{i_1 \varepsilon D} \pi_{i_1} \left(1 - \frac{\pi'_{i_1}}{\pi_{i_1}}\right) \sum_{j \varepsilon D'_{i_1}} \frac{\pi'_{j_1} - \pi_{j_1}}{A}$$

$$\left(1 - \frac{\pi_{i_1 j_1}/\pi_{i_1}}{(\pi'_{j_1} - \pi_{j_1})/A}\right) \sum_{i \; \varepsilon D''_{i_1 j_1}} \left(\frac{\pi_{i_1 i_2}}{\pi_{i_1}} - \frac{\pi'_{i_2} - \pi_{i_2}}{A}\right) \frac{1}{A'_{i_1}}$$

$$\left(1 - \frac{\pi'_{j_1 i_2}/\pi'_{j_1}}{\dfrac{\dfrac{\pi_{i_1 i_2}}{\pi_{i_1}} - \dfrac{\pi'_{i_2} - \pi_{i_2}}{A}}{A'_{i_1}}}\right) \frac{\pi'_{j_1 i_1}/\pi'_{j_1}}{A'(i_1 j_1)}$$

$$= \frac{1}{2} \sum_{i \in D} (\pi_i - \pi'_i) \sum_{j \in D'_i} \left(\frac{\pi'_j - \pi_j}{A} - \frac{\pi_{ij}}{\pi_i}\right) \frac{\pi'_{ji}}{\pi'_j}$$

$$P''_6 = \frac{1}{2} \sum_{i_1 \in D} \pi_{i_1} \left(1 - \frac{\pi'_{i_1}}{\pi_{i_1}}\right) \sum_{j_1 \in D'_{i_1}} \frac{\pi'_{j_1} - \pi_{j_1}}{A}$$

$$\cdot \left(1 - \frac{\pi_{i_1 j_1}/\pi_{i_1}}{(\pi'_{j_1} - \pi_{j_1})/A}\right) \sum_{i_2 \in I''_{i_1 j_1}} \frac{\dfrac{\pi_{i_1 i_2}}{\pi_{i_1}} - \dfrac{\pi'_{i_2} - \pi_{i_2}}{A}}{A'_{i_1}}$$

$$+ \frac{1}{2} \sum_{i_1 \in D} \pi_{i_1} \left(1 - \frac{\pi'_{i_1}}{\pi_{i_1}}\right) \sum_{j_1 \in D'_{i_1}} \frac{\pi'_{j_1} - \pi_{j_1}}{A}$$

$$\cdot \left(1 - \frac{\pi_{i_1 j_1}/\pi_{i_1}}{(\pi'_{j_1} - \pi_{j_1})/A}\right) \sum_{i_2 \in D''_{i_1 j_1}} \frac{\pi'_{j_1 i_2}}{\pi'_{j_1}}$$

$$= \frac{1}{2} \sum_{i \in D} (\pi_i - \pi'_i) \sum_{j \in D'_i} \left(\frac{\pi'_j - \pi_j}{A} - \frac{\pi_{ij}}{\pi_i}\right)$$

$$\left\{ \sum_{k \in I''_{ij}} \frac{\dfrac{\pi_{ik}}{\pi_i} - \dfrac{\pi'_k - \pi_k}{A}}{A'_i} + \sum_{k \in D''_{ij}} \frac{\pi'_{jk}}{\pi'_j} \right\}$$

$$= \frac{1}{2} \sum_{i \in D} (\pi_i - \pi'_i) \sum_{j \in D'_i} \left(\frac{\pi'_j - \pi_j}{A} - \frac{\pi_{ij}}{\pi_i} \right)$$

$$\sum_{k \neq i \neq j} \text{Min} \left(\frac{\frac{\pi_{ik}}{\pi_i} - \frac{\pi'_k - \pi_k}{A}}{A'_i} , \frac{\pi'_{jk}}{\pi'_j} \right)$$

The expected number of common units in the SCHEME-3 is

$$\nu_3 = 2(P''_1 + P''_3) + P''_2 + P''_4 + P''_5 + P''_6$$

$$= 2P_1 + P_2 + 2P''_3 + P''_4 + P''_5 + P''_6$$

$$(\text{since } P''_1 = P_1, \quad P''_2 = P_2)$$

Now,

$$2P''_3 + P''_4 + P''_5$$

$$= \frac{1}{2} \sum_{i \in D} (\pi_i - \pi'_i) \Big[2 \sum_{j \neq i} \text{Min} \left(\frac{\pi'_j - \pi_j}{A}, \frac{\pi_{ij}}{\pi_i} \right) \frac{\pi'_{ji}}{\pi'_j}$$

$$+ \sum_{j \neq i} \text{Min} \left(\frac{\pi'_j - \pi_j}{A} , \frac{\pi_{ij}}{\pi_i} \right) \left(1 - \frac{\pi'_{ji}}{\pi'_j} \right)$$

$$+ \sum_{j \in D'_i} \left(\frac{\pi'_j - \pi_j}{A} - \frac{\pi_{ij}}{\pi_i} \right) \frac{\pi'_{ji}}{\pi'_j} \Big]$$

$$= \frac{1}{2} \sum_{i \in D} (\pi_i - \pi'_i) \Big[\sum_{j \neq i} \text{Min} \left(\frac{\pi'_j - \pi_j}{A}, \frac{\pi_{ij}}{\pi_i} \right) \left(1 + \frac{\pi'_{ji}}{\pi'_j} \right)$$

$$+ \sum_{j \in D'_i} \left(\frac{\pi'_j - \pi_j}{A} - \frac{\pi_{ij}}{\pi_i} \right) \frac{\pi'_{ji}}{\pi'_j} \Big]$$

$$= \frac{1}{2} \sum_{i \in D} (\pi_i - \pi_i') \left[\sum_{j \in I_i'} \frac{\pi_j' - \pi_j}{A} \left(1 + \frac{\pi_{ij}'}{\pi_j'}\right) \right.$$

$$+ \sum_{j \in D_i'} \frac{\pi_{ij}}{\pi_i} \left(1 + \frac{\pi_{ij}'}{\pi_j'}\right) + \left. \sum_{j \in D_i'} \left(\frac{\pi_j' - \pi_j}{A} - \frac{\pi_{ij}}{\pi_i}\right) \frac{\pi_{ij}'}{\pi_j'} \right]$$

$$= \frac{1}{2} \sum_{i \in D} (\pi_i - \pi_i') \left[\sum_{j \in I_i'} \frac{\pi_j' - \pi_j}{A} \left(\frac{\pi_{ij}}{\pi_i} + \frac{\pi_{ij}'}{\pi_j'}\right) \right.$$

$$+ \sum_{j \in D_i'} \frac{\pi_j' - \pi_j}{A} \left(\frac{\pi_{ij}'}{\pi_j'} + \frac{\pi_{ij}}{\pi_i}\right)$$

$$+ \sum_{j \in I_i'} \frac{\pi_j' - \pi_j}{A} \left(1 - \frac{\pi_{ij}}{\pi_i}\right) + \left. \sum_{j \in D_i'} \frac{\pi_{ij}}{\pi_i} \left(1 - \frac{\pi_j' - \pi_j}{A}\right) \right]$$

$$= \frac{1}{2} \sum_{i \in D} (\pi_i - \pi_i') \left[\sum_{j \in I} \frac{\pi_j' - \pi_j}{A} \left(\frac{\pi_{ij}}{\pi_i} + \frac{\pi_{ij}'}{\pi_j'}\right) \right.$$

$$+ \sum_{j \in D_i'} \frac{\pi_{ij}}{\pi_i} \left(1 - \frac{\pi_j' - \pi_j}{A}\right) + \left. \sum_{j \in I_i'} \frac{\pi_j' - \pi_j}{A} \left(1 - \frac{\pi_{ij}}{\pi_i}\right) \right]$$

$$= 2P_3 + P_4 + P_5 + \frac{1}{2} \sum_{i \in D} (\pi_i - \pi_i')$$

$$\left[\sum_{j \in I_i'} \frac{\pi_j' - \pi_j}{A} \left(1 - \frac{\pi_{ij}}{\pi_i}\right) + \sum_{j \in D_i'} \frac{\pi_{ij}}{\pi_i} \left(1 - \frac{\pi_j' - \pi_j}{A}\right) \right]$$

Hence $v_3 = v_1 + K$

where

$$K = \frac{1}{2} \sum_{i \in D} (\pi_i - \pi_i') \left[\sum_{j \in I_i'} \frac{\pi_j' - \pi_j}{A} \left(1 - \frac{\pi_{ij}}{\pi_i}\right) + \sum_{j \in D_i'} \frac{\pi_{ij}}{\pi_i} \left(1 - \frac{\pi_j' - \pi_j}{A}\right) \right] + P_6^* - P_6$$

Now the sign of K cannot be claimed to be positive in general. If for certain π_i, π_i', π_{ij} and π_{ij}' 's the sign turns out to be positive then our SCHEME-III fares better than the others, otherwise, the claim will be the reverse and in practice the appropriate course may be followed in the light of these findings.

7.4.4 EXTENSION OF THE ABOVE SAMPLING SCHEMES TO THE CASE OF ARBITRARY SAMPLE SIZE n (> 2)

Let $S = (i_1, i_2)$ and $S' = (j_1, j_2)$ be the ordered samples each of size 2 selected from the population by following any of the sampling schemes I to III. To generalize the above sampling schemes to those with arbitrary sample size n, we first select an SRSWOR sample S_o of size n-2 from $U/i_1, i_2$. Thus the sample $S_1 = S \cup S_o$ is selected according to Seth's (1966) sampling scheme [since the first two units (i_1, i_2) are selected with probability $\frac{1}{2}\pi_{i_1 i_2}$ and the rest are selected by SRSWOR from $U/i_1\ i_2$]. Then we select S_o' in such a manner that the resulting sample $S_2 = S' \cup S_o'$ is selected according to Seth's (1966) sampling scheme. To do this we select S_o' in the following ways :

(i) if S and S' have both the units in common i.e. $S \simeq S'$ then $S_o' = S_o$.

(ii) if S and S' have one unit in common (say $i_1 = j_1$) and the other unit of S' (say j_2) belongs to S_o then we construct S_o' by replacing j_2 by i_2.

(iii) if S and S' have one unit in common (say $i_1 = j_1$) and the other unit of S' (i.e. j_2) does not belongs to S_o then we take $S_o' = S_o$.

(iv) if S and S' have no units in common and S_o contains j_1, j_2 or both, then S_o' is obtained by replacing j_1 by i_1 and j_2 by i_2.

(v) if S and S' have no units in common and S_o does not contain j_1 and j_2 then we take $S_o' = S_o$.

Following Lanke (1975) it can be shown that S_o' is an SRSWOR from $U/j_1, j_2$.

Writing $r(d)$ as the number of distinct units in a sample d, we have the expected number of common units between S_1 and S_2 as

$$\nu(n) = (n-2)\, P\{r(S \cap S') = 0\} + (n-1)\, P\{r(S \cap S')$$

$$+\, n\, P\{r(S \cap S') = 2\}$$

$$= (n-2) + P\{r(S \cap S') = 1\} + 2\, P\{r(S \cap S') = 2\}$$

$$= (n-2) + \nu(2),$$

where $\nu(2)$ is the expected number of common units between S and S'.

REMARK

If it is decided that Seth's (1966) sampling scheme is to be employed to control the variances of the estimator then for the sake of controlling the budget, our object should be to adopt that sampling scheme for which $\nu(2)$ is the maximum. In general we cannot say whether this maximum is achieved any of the schemes I - III. We are even unable to choose among these three samplings from this consideration in general.

REFERENCES

Abraham, T.P., Khosla, R.K. and Kathuria, O.P., (1969) - Some investigation on the use of successive sampling in pest and disease surveys. Jour. Ind. Soc. Agri. Statist., 21, 43-57.

Arnab, R (   ) - On the relative efficiencies of some sampling strategies under a super-population model accepted for publication in Jour. Ind. Soc. Agri. Statist (to appear in 1979).

Arnab, R (1979) - Surveying a finite population on more than two occasions. Tech. report ASC/16, Indian Statistical Institute.

Asok, C. and Sukhatme, B.V. (1976) - On Sampford's procedure of unequal probability sampling without replacement. Jour. Amer. Statist. Assoc., 71, 912-918.

Avadhani, M.S. and Srivastava, A.K. (1972) - A comparison on Midzuno-Sen scheme with PPS sampling without replacement and its application to successive sampling. Ann. Inst. Statist. Math., 24, 153-164.

Avadhani, M.S. and Sukhatme, B.V. (1970) - A comparison of two sampling procedures with an application to successive sampling. Applied Statistics, 19, 251-259.

Bandyopadhyaya, S., Chattopadhyaya, A.K., and Kundu, S (1977) - On estimation of population total. Sankhyā, Ser C, 39, 28-42.

Brewer, K.R.W., Foreman, E.K., Mellor, R.W. and Trewin, D.J. (1977) - Use of experimental design and population modelling in survey sampling. Presented at the 41st session of the International Statistical Institute held at New Delhi.

Chakrabarty, R.P. (1973) - A note on the small sample theory of the ratio estimator in certain specified populations. Jour. Ind. Soc. Agri. Statist. 25, 49-57.

Chaudhuri, Arijit (1971) - Some sampling schemes to use Horvitz-Thompson estimator in estimating a finite population total. Cal. Statist. Assoc. Bull. 20, 37-66.

Chaudhuri, Arijit (1974) - On some properties of the sampling scheme due to Midzuno. Cal. Statist. Assoc. Bull., 23, 1-19.

Chaudhuri, Arijit (1975) - A simple method of sampling without replacement with inclusion probabilities exactly proportional to size. Metrika, 22, 147-152.

Chaudhuri, Arijit (1976) - A non-negativity criterion for a certain variance estimator, Metrika, 23, 201-205.

Chaudhuri, Arijit (1977) - On some properties of Horvitz-Thompson estimator based on Midzuno's πps sampling scheme. Jour. Ind. Soc. Agri. Statist., 29, 42-52.

Chaudhuri, Arijit and Arnab, R. (1977) - On the relative efficiencies of a few strategies of sampling with varying probabilities on two occasions. Cal. Statist. Assoc. Bull., 26, 25-38.

Chaudhuri, Arijit and Arnab, R. (1977) - On unbiased varian estimation with various multi-stage sampling strategies Tech. report ASC/7. Indian Statistical Institute.

Chaudhuri, Arijit and Arnab, R. (1978) - On the role of sample-size in determining efficiency of Horvitz-Thompson estimator. Sankhyā, Ser C, 40, 104-109.

Chaudhuri, Arijit and Arnab, R. ( ) - On estimating the mean of a finite population sampled on two occasions with varying probabilities. Accepted for publication in Aust. Jour. Statist (to appear in 1979).

Chaudhuri, Arijit and Arnab, R. ( ) - On the relative efficiencies of sampling strategies under a super-population model. Sankhyā, Ser C, accepted for publication in Sankhyā C, 41, (to appear in 1979).

Chotai, J. (1974) - A note on Rao-Hartley-Cochran method for pps sampling over two occasions. Sankhyā 36, Ser C, 173-180.

Cochran, W.G. (1963) - Sampling Techniques, New York, Wiley.

Durbin, J. (1953) - Some results in sampling theory when the units are selected with unequal probabilities. Jour. Roy. Statist. Soc. Ser B, 15, 262-269.

Durbin, J. (1959) - A note on application of Quenouille's method of bias reduction to estimation of ratio's. Biometrika 46, 477-480.

Durbin, J. (1967) - Design of multi-stage surveys for the estimation of sampling errors. Applied Statistics, 16, 152-164.

Eckler, A.R. (1955) - Rotation sampling. Ann. Math. Statist. 26, 664-685.

Fellegi, I.P. (1966) - Changing probabilities of selection when two units are selected with pps without replacement. Proc. Soc. Statist. Sec. Amer. Statist. Assoc. 434-442.

Ghangurde, P.D. and Rao, J.N.K. (1969) - Some results on sampling over two occasions. Sankhyā, A, 31, 463-472.

Goswami, J.N. and Sukhatme, B.V. (1965) - Ratio method of estimation in multiphase sampling with several auxiliary variables. Jour. Ind. Soc. Agri. Statist. 17, 83-103.

Hájek, J. (1964) - Asymptotic theory of rejective sampling with varying probabilities from a finite population. Ann. Math. Statist. 35, 1491-1523.

Hansen, M.H.&Hurwitz, W.N. (1943) - On the theory of sampling from finite populations. Ann. Math. Statist. 35, 1491-1523.

Hansen, M.H., Hurwitz, W.N. and Madow, L.H. (1953) - Sample Survey Methods and Theory, Vol.II, John Wiley and Son. Inc. N.Y. 268-269.

Hanurav, T.V. (1967) - Optimal utilization of auxiliary information : πps sampling of two units from a stratum Jour. Roy. Statist. Soc. Ser B, 29, 374-390.

Horvitz, D.G. and Thompson, D.J. (1952) - A generalization of sampling without replacement from a finite universe. Jour. Amer. Statist. Assoc., 47, 663-685.

Jessen, R.J. (1942) - Statistical investigation of a sample survey for obtaining farm facts. Iowa Agricultural Experimental Station Research Bulletin No.3 (1942).

Kathuria, O.P. (1975) - Some estimates in two-stage sampling on successive occasions with partial matching at both stages. Sankhyā C, 37, 147-162.

Keyfitz, N. (1951) - Sampling with probabilities proportional to size, adjustment for changes in probabilities. Jour. Amer. Statist. Assoc. 46, 105-109.

Kulldorff, G. (1963) - Some problems of optimum allocation for sampling on two occasions. Rev. Int. Statist. Inst. 31, 24-57.

Lanke, J. (1974) - On non-negative variance estimators in survey sampling. Sankhyā C, 37.

Lanke, J. (1975) - Some contributionsto theory of survey sampling - unpublished Ph.D. thesis. University of Lund, Sweden.

Mahalanobis, P.C. (1944) - On large scale sample surveys. Phil. Trans. Roy. Soc. Ser B, 231, 329-451.

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

Mukhopadhyay, P. (1974) - πps sampling scheme to base HTE Cal. Statist. Assoc. Bull, 23, 21-24.

Murthy, M.N. (1957) - Ordered and unordered estimators in sampling without replacement. Sankhyā, 18, 379-390.

Olkin, I. (1958) - Multivariate ratio-estimation for finite populations, Biometrika, 45 154-165.

Quenouille M.H. (1956) Notes on bias in estimation. Biometrika, 43, 353-360.

Pathak, P.K. and Rao, T.J. (1967) - Inadmissibility of customary estimators in sampling over two occasions. Sankhyā, Ser A, 29, 49-54.

Patterson, H.D. (1950) - Sampling on successive occasions with partial replacement of units. Jour. Roy. Statist. Soc. 12, 241-255.

Raj Des (1956) - Some estimators in sampling with varying probabilities without replacement. Jour. Amer. Statist. Assoc. 51, 269-284.

Raj Des (1965) - Sampling over two occasions with probability proportionate to size. Ann. Math. Statist. 36, 327-330.

Raj Des (1968) - Sampling Theory. Mc Graw-Hill Book company, New York, 156-157.

Rao, J.N.K. (1961) - On the estimate of variance in unequal probability sampling. Ann. Inst. Statist. Math. 13, 57-60.

Rao, J.N.K. (1966a) - On the relative efficiency of some estimators in pps sampling for multiple characteristic. Sankhyā, Ser. A, 28, 61-70.

Rao, J.N.K. (1966b) - On the comparison of sampling with and without replacement. Rev. Int. Statist. Inst. 34, 2, 125-138.

Rao, J.N.K. (1973) - On double sampling for stratification and analytical surveys. Biometrika, 60, 125-133.

Rao, J.N.K. and Graham, J.E. (1964) - Rotation designs for sampling on repeated occasions, Jour. Amer. Statist. Assoc, 59, 492-509.

Rao, J.N.K., Hartley, H.O. and Cochran, W.G. (1962) - On a simple procedure of unequal probability sampling without replacement. Jour. Roy. Statist. Soc. B, 24, 482-491.

Rao, J.N.K. and Vijayan, K. (1977) - On estimating the variance in sampling with probability proportional to aggregate size. Jour. Amer. Statist. Assoc. 72 579-584.

Rao, J.N.K. and Webster J.T. (1966) - On two methods of bias reduction in the estimation of ratios. Biometrika 53, 571-577.

Rao, T.J. (1967) - On the choice of strategy for the ratio method of estimation. Jour. Roy. Statist. Soc. Ser B, 29, 392-397.

Rao, T.J. (1967)* - Contribution to the theory of sampling strategies - Unpublished Ph.D. thesis. Indian Statistical Institute.

Rao, T.J. (1972) - On the variance of the r tio-estimator. Metrika, 18, 209-215.

Rao, T.J. (1977) - Estimating the variance of the ratio estimator for the Midzuno-Sen sampling scheme. Metrika, 24, 203-208.

Sampford, M.R. (1967) - On sampling without replacement with unequal probabilities. Biometrika, 54, 499-513.

Sankaranarayanan, K. (1969) - An IPPS sampling scheme using Lahiri's method of selection. Jour. Ind. Soc. Agri. Statist., 21, 58-66.

Sen, A.R. (1953) - On the estimation of variance in sampling with varying probabilities. Jour. Ind. Soc. Agri. Statist. 5, 119-127.

Sen, A.R. (1955) - On the selection of n primary sampling units from a stratum structure (n > 2). Ann. Math. Statist. 26, 744-751.

Sen, A.R. (1973) - Some theory of sampling on successive occasions Aust. Jour. Statist. 15, 1973, 105-110.

Seth, G.R. (1966) - On estimates of variance of estimate of population total in varying probabilities. Jour. Ind. Soc. Agri. Statist. 18, 52-66.

Sharma, S.S. (1970) - On an estimation in T_3 - class of linear estimators in sampling with varying probabilities from a finite population. Ann. Inst. Statist. Math. 22, 495-500.

Shukla, N.D. (1976) - Almost unbiased product- type estimator. Metrika, 23, 127-133.

Singh, D. (1968) - Estimates in successive sampling using a multi-stage design. Jour. Amer. Statist. Assoc. 63, 99-112.

Singh, D. and Kathuria, O.P. (1969) - On two-stage successive sampling. Aust. Jour. Statist. 11, 59-66.

Singh, R. (1972) - On Pathak and Rao's estimate in pps with replacement sampling over two occasions. Sankhyā 34, Ser A, 301-303.

Singh, R. (1975) - A note on the efficiency of ratio estimat with Midzuno's scheme of sampling. Sankhyā, Ser C, 37, 211-214.

Smith, H.F. (1938) - An empirical law governing soil hetero-geneity. Jour. Agri. Soc. 28, 1-23.

Stuart, A. (1963) - Some results on sampling with unequal probabilities. Proc of the 34th sessions, Bull. Int Statist. Inst. 40, 773-780.

Tikkiwal, B.D. (1950) - Estimation by successive sampling - paper presented at 4th annual meeting. Ind. Soc. Agri. Statist.

Tikkiwal, B.D. (1951) - Theory of successive sampling - Unpublished Diploma thesis, 1951. I.A.R.S., New Delhi

Tikkiwal, B.D. (1958) - Theory of successive two-stage sampling. (Abstract). Ann. Math. Statist. 29,1291.

Tikkiwal, B.D. (1964) - A note on two-stage sampling on successive occasions. Sankhyā, Ser A, 26, 97-100.

Tikkiwal, B.D. (1965) - The theory of two-stage sampling on successive occasions. Jour. Ind. Statist. Assoc. 125-136.

Tripathi, T.P. and Srivastava, O.P. (1978) - Estimation on successive occasions using PPSWR sampling. Tech. report. 22/78. Indian Statistical Institute.

Vijayan, K. (1975) - On estimating variance in unequal probability sampling. Jour. Amer. Statist. Assoc. 70, 713-716.

Yates, F. (1949) - Sampling methods for census and surveys. Charges Griffin and Co., London.