OPTIMUM SAMPLING STRATEGIES

by

T. V. HANURAV

RESEARCH AND TRAINING SCHOOL
INDIAN STATISTICAL INSTITUTE
203, B. T. ROAD
CALCUTTA-35
INDIA

ERRATA

Page	Para	Line	Instead of	Read
(ii)	2	3	rearch this	research. His
	3	2	Sri K.	Sri P.
(iii)	§2.5		problem	problem of Chaps. IV and V
(iv)	§5.4		$\overset{*}{Y}_{HT}$	$V(\overset{*}{Y}_{HT})$
2	2	7	Hurmitz	Hurwitz
	3	4	Harvitz	Horvitz
4	2	11	sense sampling	sense. Sampling
6	2	1 and 3	PS	π PS
9	1	1	$\sum_{m=1}^{\infty} \sum$	$\sum_{m=1}^{\infty} \left(\sum_{s \in S_m} P_s \right)$
→12	2	3	t_2	T_2
→13	2	6	if	if and only if
14	2	1	§4.1 of Chapter IV	§7.3 of Chapter VII
→16	3	5	(2.2.14)	(2.2.15)
		9	U_i	u_i
→17	1	3	β_2	β_s
21	4	11	for all	for all y
		12	one	one y
22	1	2	for all	for all y
		3	one	one y
26	3	5	h.l.g.u.e.	g.h.l.u.e.

Page	Para	Line	Instead of	Read
27	1	4,6,8,9	H_{μ}, $H_{(\mu)}$	$H_{e\mu}$
28	1	4		
	2	3		
	2	6	$\alpha\ X_i$	$a\ X_i$
	2	11, 12	(delete primes on H, S, P and T)	
		13	(delete the line)	
29	1	3	T_0	$T_{0.1}$
30	3	7	if	if and only if
33	1	2	(2.3.8)	(2.3.12)
			$D^{(o)}$	$D_{e\mu}$
		3	(2.2.13)	(2.2.9)
		5	(2.2.23)	(2.3.5)
	2	1	(iii)	(ii)
		2	(2.3.13)	(ii) of (2.3.1)
	3	8 and 9	§5.3 of Chap. V	§4.5 of Chap. IV.
	4	7	(2.3.8)	(2.3.12)
34	1	4	works of	works of Goodman and Kish [13],
35	1	11	§ 7.2	§ 7.4
		19	denominator	denominator of (2.4.1)
36	1	3	∂_1	$\partial_1^{(0)}$
		5	μ_s	μ_s (c.f. § 5.2.1)
		8	(2.3.8)	(2.3.12)

Page	Para	Line	Instead of	Read
36	2	2	$\Sigma\Sigma \frac{y_\lambda y_{\lambda'}}{\pi_\lambda \pi_{\lambda'}}$	$\Sigma\Sigma \frac{y_\lambda y_{\lambda'}}{\pi_\lambda \pi_{\lambda'}} \cdot \pi_{\lambda\lambda'}$
37	2	7	class of	class $D^{(\mu)}$ of
39	1	1	forwards	towards
		3,11,18	(2.3.8)	(2.3.12)
39	1	13	are	are (c.f. § 7.3)
40	1	7 and 8	D	$D \in D^{(\mu)}$
		10	π_i	π_λ
42	2	7,9,11	i	λ
53	equation (3.2.3)		$0 \leq q_2{}^{s_{(k)}} \leq 1$	$0 \leq q_2(s_{(k)}) \leq 1$
57	2	4	probability	probability 1
58	2	7	respectively,	respectively, is better the estimators which ig the information on
		6 and 8	(3.3.1)	(3.3.1')
61			q_0	q_1
62			q_1	q_2
			$q_3(2, s(i), U_j)$	$q_3(s(i), U_j)$
			$q_3(K+1, s_{(k)}, U_j)$	$q_3 \cdot (s_{(k)}, U_j)$
63	2	3	(3.1)	(3.2.1)
		8	Chapter V	Chapter IV.
64	2	11	time	true

Page	Para	Line	Instead of	Read
68	1	5	$\frac{P_N^2}{2^{r-1}} = \frac{S_{(r)}}{P_N^{2^r}}$	$\frac{P_N^2}{2^{r-1}} \cdot \frac{S_{(r)}}{P_N^{2^r}}$
69	1	12	$\sum_{k=1}^{2^j}$	$\sum_{k=1}^{2^m}$
72	1	12	$2P_N^{2k}$	$2P_N^{2^k}$
		14	$\pi_i \pi_j$	$\pi_i \pi_j$ (4.1.8)
73	1	2	$> \frac{1}{2}$	$> \frac{1}{2}$, ... (4.1.9)
	3	14	(0.1.6)	(4.1.7)
75	4	5	$1 - \frac{}{2(1 - P_N)}$	$1 - \frac{\partial}{2(1 - P_N)}$
		8	(6.1.6) of Chp. VI.	(4.1.6) of Chap. IV.
76	1	14	$j \leq N - 1$	$j \leq N$
78	1	9	(6.2.6) and (6.2.7)	(4.2.6) and (4.2.7)
		last	a_N	∂
80	last but one		$P_{N.1}$	P_{N-1}
81	1	8	scheme	scheme B
82	3	2	later	later, in § 4.6,
88	1	1,2	4	5
90	1	3	R + K	R + 1
93	throughout		'change y's to Y's'	
102	2	7	(2.3.7)	(2.3.8)

Page	Para	Line	Instead of	Read
102	2	7	(2.3.7)	(2.3.8)
105	2	4	(2.3.7)	(2.3.8)
108	1	9	$(1 - p_i^{(r)} \cdot p_j^{(r)}$	$(1 - p_i^{(r)} - p_j^{(r)})$
110	2	2	last draw	last unit
111	2	3	not strictly	slightly
116	1	2	units	events
118	1	2	$(1 - \pi_i - \pi_j)$	$(1 - \pi_i - \pi_j) \cdot$ (5.4.4)
120	1	4	undecided	undecided or
126	1	9	π_{ij}^{n}	$\pi_{ij}^{(n)}$
		10	$p^{(r)}$	$p_i^{(r)}$
		11	$\frac{\pi}{n}$	$\frac{\pi}{r}$ $1 - \frac{(K_r - K_{r-1})(P_i + P_j) - 2K}{(1 - K_{r-1}P_i)(1-K}$
129	1	1	§ 5.6	§ 5.7
132	1	7	stratefies	strategies
	2	3	gain	∂_1 - gain
			loss	∂_1 - loss
136	1	6	π_{ij}	$\pi_{i_1 j}$
	2	1	chapter IV	chapter V
		4	chapter V	chapter IV
140	2	add at the end:	Also, when this condition is satisfied, the optimum allocation given by (6.2.2) reduces allocation proportional to strata-totals of	
142	1	3	and $\neq$ a π PS	and a π PS

Page	Para	Line	Instead of	Read
142	1	11	$\pi_{\cdot} \frac{X_i}{X}$	$\mu_{\cdot} \frac{X_i}{X}$
145	4	6	(6.5.4)	(6.4.4)
136	1	last but one	$\frac{\sigma_k^2}{N_k}$	$\frac{\sigma_k^2}{N_k^2}$
147	1	3,4	$\bar{X}_k \overset{**}{\sigma}{}_k^2$	$\frac{\bar{X}_k \overset{**}{\sigma}{}_k^2}{N_k}$
		3	'after $\frac{1}{N_k^2}$ insert braces and close at the end'	
152	1	3	criterion	condition
153		last line	H	H_1
157	1	7	on what point	on y , a point
		8 and 9	delete 'these probabilities are situated'.	
		12	point	vector
		13	$y(s) = (y_{i_1}, \ldots, y_{i_{n_s}})$	$y(s) = ((U_{i_1}, y_{i_1}), \ldots, (U_{i_{n_s}}, y_{i_{n_s}}))$
		14	delete 'of R^N'	
		18	s'	s', containing the same set of distinct units as s, i.e.
158	2	13	P_s in R^N	P_s
160	1	3	y_5^2	y_3^2
		9	s_1 and s_2	s_1 and s_2 where $s_1 \neq s_2$

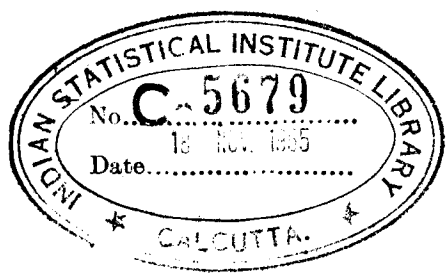
Page	Para	Line	Instead of	Read
162	2	add at the end:		The use of 'difference estimator' of the fo (7.2.18) also indicates that adherence to physical units is not always done.
163	throught		C	K
	1	12	delete: (where $\frac{L_i y_i}{\pi_i}$ is to be set equal to zero when L_i =	
164	equation (7.3.6)		delete	'$L_i \neq 0$' and '$q_{ij} \neq 0$.
170	1	10	possible	impossible
		12	no unit	a unit not
173	3	8	alone.	alone, and is a homogeneous polynomial of degree i in y_i's.
174	1	1	$T_{2,s} \neq 0$	$T_{2,s} = \sum_{i \in s} \beta_{sii}\, y_i^2 + \sum\sum_{\substack{i \neq j \\ \in s}} \beta_{sij} y_i$
	1	6	unordered samples	unordered samples with distin units
179	1	8	follows that	follows that (c.f. (7.2.20))
180	1	5	$\beta_{si\ldots i}\, y_i^2$	$p_{si\ldots i}\, y_i^r$
181	1	10	$\sum_{s\ i} P_s$	$\sum_{s\ i} P_s$
	1	15 and 17	$s \quad U_i$	$s \supset U_i$
190	1	5	$(T''^2) - (T'''^2)$	$E(T''^2) - E(T'''^2)$
193	1	1	only	any
		11	π_λ	$\pi_\lambda^{(0)}$

OPTIMUM SAMPLING STRATEGIES AND SOME RELATED PROBLEMS

By

T. V. HANURAV

Dissertation submitted to the Indian Statistical Institute in partial fulfillment for the award of the degree of Doctor of Philosophy



Research and Training School
Indian Statistical Institute
203, B. T. Road,
Calcutta-35,
INDIA

P R O L O G U E

This thesis is being submitted in partial requirement for the degree of Doctor of Philosophy of the Indian Statistical Institute and presents the work done by the author at the Indian Statistical Institute during the past few years in the field of sampling from finite populations. Some of the results given herein have earlier appeared in published from ([18], [19], [20])

The main contributions in this thesis are (1) Proof of optimality of Horvitz and Thompson estimator of the population total (2) development of a one-by-one drawing mechanism to result in a given sampling design and (3) optimum utilisation of auxiliary information of a type commonly met with in practice.

A new criterian of "hyper admissibility" of an estimator is introduced and the optimality of Horvitz and Thompson estimator of the population total in the class of all polynomial unbiased estimators is demonstrated in Chapter VII. A method of drawing units of a population one-by-one and with replacement to result in any given sampling design is given in Chapter III. Sampling methods that result in sampling designs with assigned values of inclusion probabilities for the units, for the optimum utilisation of auxiliary information, are obtained in Chapter IV and V. Some miscellaneous results regarding optimum utilisation of auxiliary information are given in Chapter VI. Chapter I is devoted to introduction. In Chapter II the problems considered in this thesis are clearly formulated and a brief review of the work of earlier authors

is given. For every Chapter a summary of the contents is first given. Towards the end a brief indication of the present position is given and some further problems for research in this field are given. A bibliography of related earlier work is provided.

The author wishes to record his gratitude to Prof. C. R. Rao, Director of the Research and Training School of the Indian Statistical Institute for providing facilities to carry out this research. His sincere thanks are due to Prof. J. Roy, Head of the Data Processing unit of the Indian Statistical Institute, for his valuable guidance and many critical comments that helped to considerably improve the presentation of the material as also for kindly providing facilities of the Electronic Data Processing Machines of his unit. Thanks are due to Prof. V.P.Godambe who initially set the author working in this field.

The author records with thanks the help he received from a number of his colleqgues. Sri P. Unnikrishnan has kindly programmed a number of complicated programmes for the author on IBM 1401, Sri K. Vijayan participated in many valuable discussions, Sri T. J. Rao rendered valuable help in proof-reading and Sri C. Parthasarathy corrected a number of mistakes regarding the language and offered many critical comments. Thanks are due to Sri G.M.Das for his very patient and elegand typing of the thesis.

TABLE OF CONTENTS

TABLE OF CONTENTS

CHAPTER 1

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

The advantages of sample surveys over complete censuses are well known and seem to be fully appreciated as is evidenced by the increasing use of sample surveys now a days as a means of collecting information.

The use of probability theory to make rigorous inductive inferences has been well recognised for a long time. Such inferences can be made only when observations which form the basis of the inference are generated by some chance mechanism. In traditional applications, the statistician usually assumes or takes for granted some kind of chance mechanism behind the observations, where as in sample surveys or planned experiments the statistician consciously introduces the chance element by having recourse to the mechanism of randomisation, introduced by R.A.Fisher [B5]. This has the advantage that the validity of the inferences does not depend on any extraneous assumptions. He demonstrated that a deliberately introduced randomisation in the selection of a part from the whole itself provides a valid method of obtaining a rigorous expression to the amount of error committed while arguing from a part to the whole.

Another important concept introduced by P. C. Mahalanobis [24a] in this field, is the cost function. While the efficiency of a sample survey as measured by the precision of the estimate - is important, it has to be delicately balanced against the cost of the survey to

have a meaningful application of these techniques in practice.

The earlier developements in sampling theory of finite populations relate mainly to a number of techniques of sampling appropriate to varians situations in practice, to estimate the total of real valued character defined for units of the population (briefly referred to as population total . Significant advances in this direction are stratified sampling first studied by Neyman [29], multi stage sampling and use of auxiliary information first studied by Hanson and Hurmitz [20a], various methods in current practice of using auxiliary information, are the ratio and regression methods of estimation and probability proportional to size sampling. While for the first two methods it is not necessary to have the auxiliary information completely beforehand (and can even be collected along with the main information) but only its total for the entire population, for the latter, it is necessary. However, the first two methods do not have an exact small sample theory while for the last we have an exact theory.

The earlier developments have been guided usually on heuristic considerations and attention limited to get at some unbiased estimators. It is only recently that a systematic investigation of sampling from finite populations has begun with the works of Harvitz and Thompson [22] and Godambe [11].

The main contributions in this thesis relate to (1) a search for reasonable criteria of optimality and sampling strategies that

one optimum under these criteria; (2) development of a unified operational method drawing samples when a sampling design is partly or fully specified and (3) optimum utilisation of auxiliary information of a type commonly met with in practice.

Godambe has stated [11] that for any sampling design there does not exist a uniformly minimum variance unbiased estimator of the population total in the class of all homogeneous linear unbiased estimators. (As shown in this thesis, there are certain designs for which this result does not hold which form the class of what we call "uni-cluster" designs.)

The choice of an estimator therefore has to be made from a rather wide class of admissible unbiased estimator. Moreover there is no mathematical reason for the exclusion of all nonlinear unbiased estimators. We therefore introduce the criterion of "hyper admissibility" which has meaningful practical interpretation. This criterion yields, for any non uni-cluster sampling design, a unique estimator in the class of all polynomial unbiased estimators of the population total, and can therefore be called the 'best' estimator in this class under this criterion. This best estimator is the well known Harvitz and Thompson estimator. This result given in § 7.4 and § 7.5 of Cha Chapter VII removes the difficulties of choice of an estimator.

As Godambe has shown, a sampling design is completely characterised by the specification of a set of samples and defining a

probability measure on it. However for practical application an operational procedure of drawing the sample one by one is very often required. We show in Chapter III, how, given any sampling design, one can realise it by a definite one by one drawing mechanism. Any partially specified sampling design can then be obtained by a simple drawing mechanism by a proper choice of further specifications, as illustrated in § 3.3. The results of this Chapter were published earlier [18].

It is latent in the work of Godambe and is explicitly spelt here (in Ch. II) that when information on a positive valued auxiliary variable is available before-hand on all the units of the population, then in the class of all sampling strategies with given expected number μ of distinct units, any strategy such that

i) every sample has the same number of distinct unit

ii) probability of including any unit is proportional to the auxiliary information on the unit

and iii) the estimator used is the corresponding Horvitz and Thompson estimator

is 'best' in a well defined Bayesian sense. Sampling designs satisfying (ii) are referred to as π PS design. The problem of constructing designs satisfying (i) and (ii) is then an important problem that is engaging the attention of a number of many workers in this field of whom we mention Goodman and Kish [13]. Horvitz and Thompson [22], Yates and Grundy [40], [14], Durbin [9], Stevens [39],

Felligi [10], DesRaj [7] and Rao, Hartley and Cochran [35], besides the author himself [19].

Though conditions (i) and (ii) ensure the above criterion of optimality, they do not completely specify the design. We therefore add further desirable conditions and obtain a complete solution for the important practical case of sampling two units from a stratum. This while satisfying (i) and (ii), also allows a stable non-negative unbiased estimator of the sample variance of the estimator. The method compares favourably with two other method in current use of utilisation of auxiliary information. These results are given in Chapter IV. For the case of general values of μ, some near-optimum solutions strictly satisfying (ii) are given in Chapter V. The results of this chapter were obtained by the author towards the end of 1959 and were later published [19], [20].

The thesis is divided into seven chapters. Chapter I (the present one) is introductory giving a broad summary of the major results of the thesis.

Chapter II gives the necessary definitions and concepts, focusses the attention on the problems considered in this thesis and clearly formulates them and gives a review of the related work of the earlier authors in this field. The rest of the thesis contains the author's contributions.

In Chapter III the equivalence of sampling designs and one by one drawing mechanisms is proved and the method used to illustrate the convenient way of choosing further specifications of a partially specified design to arrive at simple drawing mechanisms.

Chapters IV and V are devoted to the construction of π PS sampling designs Chapter VI gives some results concerning the optimum choice of multistage PS designs, validity of prior estimates of parameters in stratified sampling and integration of two different surveys on the same population. The methods of these chapters are illustrated by means of emperical examples and are compared with some other methods in current use of the use of auxiliary information.

Finally, Chapter VII deals with the general problem of estimation and gives necessary and sufficient conditions on a design for the estimability of a class of parametric functions, the exceptional sampling designs admitting a uniformly minimum variance estimator in the class of general homogeneous linear unbiased estimators of the population total and ends up with the proof of 'bestness' (under the criterion of hyperadmissibility) of the Horvitz and Thompson estimator.

In the epilogue we take stock of the present position and indicate some further lines of research on problems related to those considered in this thesis.

A bibliography giving mainly earlier work on the problems considered in this thesis, is given at the end.

CHAPTER II

DEFINITIONS AND CONCEPTS

In this chapter we shall give some definitions and formulate the concepts. A broad review of the felated work of earlier authors is also given. The review thus relates to the general aspects of estimation and the optimum utilisation of auxiliary information. The central problem throughout is the estimation of total (or equivalently the mean) of a finite population of known size, though occasionally we shall formulate our concepts for a wider area.

§ 2.1 Sampling design and drawing mechanism

We shall give some definitions.

A '<u>finite population</u>', $\mathcal{U}$ in this thesis refers to a collection of a <u>known</u> number N, of '<u>sampling units</u>' ('units for brevity henceforth)

$$U_1, U_2, \ldots, U_N \qquad \cdots(2.1.1)$$

such that the units are distinguishable and a list like (2.1) is possible, at least conceptually. Such a list is called a '<u>sampling frame</u>', and 'N' is called the '<u>population size</u>'.

A 'sample' s from $\mathcal{U}$ is a finite ordered sequence of units from $\mathcal{U}$:

$$s = (u_1, u_2, \cdots, u_{n_s}) \qquad \cdots \qquad (2.1.2)$$

where each u_i belongs to $\mathcal{U}$. Equivalently we can also write

$$s = (U_{i_1}, U_{i_2}, \cdots, U_{i_{n_s}}) \qquad \cdots \qquad (2.1.2)$$

where $1 \leq i_t \leq N$ for $1 \leq t \leq n_s$. In the above the units need not be distinct from each other. n_s is called the sample size. The number of distinct units in s is called the 'effective sample size'.and we shall denote it throughout by μ_s. (The justification for this terminology will be clear latter in § 2.2).

A 'sampling design' (sometimes briefly referred to as 'design') D, over $\mathcal{U}$, is a collection (i.e. unordered set) S of samples s from $\mathcal{U}$ with a probability measure P defined on it, i.e. to each $s \in S$ is attached a probability P_s. Then we can write

$$D = D(S, P) \qquad (2.1.3)$$

where $\sum_{s \in S} P_s = 1$

We note that though each sample is a finite sequence of elements from $\mathcal{U}$, the sample sizes need not be uniformly bounded in S. For example we may take for S the collection of all possible samples obtained by sampling with replacement until two distinct units are obtained. However large 'm' may be there are always samples in the above collection whose size exceeds 'm'.

Since the set of all finite sequences from $\mathcal{U}$ is a countable set, S is a countable collection of samples. Further, writing

$$S = \bigcup_{m=1}^{\infty} S_m$$

where S_m is the set of all samples s in S whose size is m (S_m is a finite set having utmost N^m elements), we have

$$1 = \sum_{s \in S} P_s = \sum_{m=1}^{\infty} \left(\sum_{s \in S_m} P_s \right) = \sum_{m=1}^{\infty} q_m \quad \text{say}$$

so that $\sum_{m=r}^{\infty} q_m \to 0$ as $r \to \infty$. Hence the probability of obtaining a sample of size greater than or equal to r tends to zero as $r \to \infty$.

The above definition of a sampling design covers all conceivable methods of drawing a sample from a finite population of known size, because each such method generates a collection of possible samples and defines a probability distribution over such samples. On the other hand, corresponding to any given sampling design, there exist various ways of drawing the sample. The obvious method is to prepare a list of all samples belonging to the design and then to choose one from the list with probabilities assigned by the design.

However, in practice it is not always convenient or possible to draw a sample from a population by first listing down all possible samples and selecting one of them with the preassigned probabilities. For example consider a population $\mathcal{U}$ of 100 units and S be the collection of all ordered pairs of units from $\mathcal{U}$ and let $P_s = 10^{-4}$ for each $s \in S$. Here there are 10^4 possible pairs. In this case one does not list down all these 10^4 pairs and then select one of the pairs with a probability 10^{-4} to each pair. Instead, one draws one unit from $\mathcal{U}$ with probability 10^{-2} to each unit of $\mathcal{U}$ and repeats the procedure once more. This is adopted firstly because it is convenient to stick to a single sampling frame all through without preparing a fresh another much bigger frame and secondly one can work with a smaller number of choices. How vastly difficult it is in practice to

list down all possible samples with their associated probabilities can easily be seen if we consider samples of larger sizes say 10 each. This is a matter of practical convenience only and in the example discussed above both the methods are theoretically equivalent.

We show in chapter III how a special method of sampling, called one-by-one drawing mechanism, can be adopted to result in any given sampling design. Such a drawing mechanism is roughly defined as a method of drawing units from $\mathcal{U}$ one-by-one and with replacement with probabilities at any stage depending possibly on the sample of units already selected up to that draw. A more rigorous definition of this is given in § 3.1 of chapter III.

We remark that (2.1) explicitly excludes such populations like the population of fish in a pond because in the latter, the population size is not known and in fact the problem of interest in such cases is to estimate N. However, in these problems, the selection of a sample can not be rigorously made probabilistic and some physical methods of selection are approximately taken to correspond to certain probability selections on intuitive grounds (which often amount to a reasoning like 'because there is no reason to belive that ....').

§ 2.2 Estimation

We now turn our attention to the problem of estimation.

Consider a real-valued character $\mathcal{Y}$ which takes the value Y_i on the unit U_i of $\mathcal{U}$ $(1 \leq i \leq N)$. Let

$$\underset{\sim}{Y} = (Y_1, Y_2, \cdots, Y_N). \qquad \cdots \quad (2.2.1')$$

A 'parametric function' (p.f., for brevity) θ is a function of the arguments $Y_1, Y_2, \cdots, Y_N$ i.e.,

$$\theta = \theta(\underset{\sim}{Y}) = \theta\ (Y_1, Y_2, \cdots, Y_N) \qquad \cdots (2.2.1)$$

A 'statistic' T defined over a design $D = D\ (S, P)$ is a function defined over the samples s of D such that if $s \in D$, the value T_s of T is a function of the $\mathcal{Y}$-values of only those units that occur in s. A statistic when used to estimate a parametric function $\theta\ (\underset{\sim}{Y})$ is called an 'estimator' of θ. It is called an 'unbiased estimator' of θ ifand only if (iff, for brevity)

$$E(T) = \sum_{s \in S} T_s P_s = \theta\ (\underset{\sim}{Y}), \quad \text{for all values of } \underset{\sim}{Y}. \qquad \cdots (2.2.2)$$

An estimator of θ which is not unbiased for θ is called a biased estimator.

In estimating a p.f. θ by an estimator T defined over a design D, $(T_s - \theta)$ is called the 'error' resulting, for the sample s, in the estimation of θ. A convex function G of $(T_s - \theta)$ is usually taken as the 'loss function for the estimation, and $E(G) = \sum_{s \in S} G_s P_s$ is called the 'expected loss'. A commonly used loss function is the mean-square error (defined below). We shall formulate our problems in terms of this loss function only though most of the definitions and conclusions given here apply broadly to any convex loss function.

The 'mean-square error' (m.s.e. for brevity) of an estimator T of θ is

$$M(T) = E(T - \theta)^2 = \sum_s T_s^2 P_s - 2\theta \sum_s T_s P_s + \theta^2 \qquad \ldots(2.2.3)$$

If T is an unbiased estimator, (2.2.3) is called the 'variance' of T, and can be written as

$$V(T) = \sum_s T_s^2 P_s - \theta^2 \qquad (2.2.4)$$

Given a design D and estimators T_1 and T_2 of the same p.f. θ, both defined over D, T_1 is said to be 'uniformly better' or simply better than T_2 if and only if

$$M(T_1) \leq M(T_2), \text{ for all } \underset{\sim}{Y} \qquad \ldots(2.2.5)$$

with strict inequality holding good for at least one value of $\underset{\sim}{Y}$. Similarly for a given design D and a class L of estimators of a p.f. θ, all defined over D, a member T_1 is called the 'best' in L if it is better than every other member of L, i.e.

$$M(T_1) \leq M(T_2), \text{ for all } T_2 (\neq T_1) \in L, \text{ and for all } \underset{\sim}{Y} \qquad \ldots(2.2.6)$$

A design D together with an estimator T of θ defined over the design, is called a 'strategy' H for the estimation of θ. Thus

$$H = H(D, T) = H(S, P, T) \qquad \ldots(2.2.7)$$

A strategy $H\ (S, P, T)$ for the estimation of θ is said to be unbiased if T is an unbiased estimator of θ. Otherwise it is called a biased one.

The m.s.e. or variance of a strategy H, denoted by $M(H)$ or $V(H)$ as the case may be, is defined to be the m.s.e. or variance of the corresponding estimator. Analogous to (2.2.5) and (2.2.6), a strategy H_1 is said to be better than another strategy H_2 cost being ignored, if

$$M(H_1) \leq M(H_2), \quad \text{for all } \underset{\sim}{Y} \qquad \cdots(2.2.8)$$

and H_1 is said to be best in a class Ω if

$$M(H_1) \leq M(H_2), \text{ for all } H_2 \in \Omega \text{ and for all } \underset{\sim}{Y} \qquad \cdots(2.2.9)$$

With respect to a design D, a p.f. θ is estimable if there exists a T unbiased for θ. For any sampling design $D=D(S,P)$ we define the 'inclusion probability' of U_i by

$$\pi_i = \sum_{s \supset i} P_s \qquad \cdots(2.2.10)$$

where the sum on the r.h.s. is taken over all samples that contain U_i at least once. Similarly the 'joint inclusion probability' of a pair (U_i, U_j), $i \neq j$, is defined by

$$\pi_{ij} = \sum_{s \supset i,j} P_s \qquad \cdots(2.2.11)$$

where the sum on the r.h.s. is taken over all samples that contain both U_i and U_j. It seems to be well-known that for the estimability

of the population total

$$Y = \sum_{1}^{N} Y_i \qquad \cdots(2.2.12)$$

or, equivalently, of the population mean

$$\bar{Y} = \frac{1}{N} \sum_{1}^{N} Y_i \qquad \cdots(2.2.13)$$

a set of necessary and sufficient conditions (n.s.c. for brevity) on $\mathcal{D}$ is that

$$\pi_i > 0 \quad \text{for} \quad 1 \leq i \leq N \qquad \cdots(2.2.14)$$

It is also assumed in the literature that a set of n.s.c's for the estimability of the variance of an unbiased estimator T of Y is that

$$\pi_{ij} > 0 \quad \text{for} \quad 1 \leq i \neq j \leq N \qquad \cdots(2.2.15)$$

The author is not aware of formal proof of this later contention in a general set up.

In § 4. 1 of chapter IV we give formal proofs of both the results given by the conditions (2.2.14) and (2.2.15).

We shall now turn our attention to what are called linear estimators.

In the classical theory of estimation from infinite populations, a linear estimator for a sample of size n is defined as

$$T_c = \sum_1^n c_i y_i \qquad \ldots \quad (2.2.15')$$

where $y_1, y_2, \ldots, y_n$ are independent observations on y and $c_1, c_2, \ldots, c_n$ are constants which are independent of the sample. In this case it is known that for any given sample size n, in the class of all linear unbiased estimators of $E(y)$, the estimator

$$\bar{y} = \frac{1}{n} \sum_1^n y_i \qquad (2.2.15'')$$

has minimum variance.

However, there is an important difference between the classical theory for infinite populations and the theory for finite populations, which we can call the distinguishability of units in the case of finite populations. Where as in the classical theory, two observations having the same value for y are both taken into account, in the theory of finite populations it does matter whether these identical values of y are observed for the same unit or for different units of the population. In the former case no new information regarding a parametric function θ is supplied by the second observation while in the latter case it is.

This fact first came to light from the results of Des Raj and Khamis [8] and Raja Rao (c.f., foot note on p. 294 of [1]), Basu [1] which are given below as

Theorem (2.2.1): For simple random sampling with replacement, of a

fixed size n of units from $\mathcal{U}$, the mean over the distinct units in the sample is a uniformly better estimator of the population mean $\bar{Y}$ than the mean over the entire sample.

We may stress here that this departure from the results of classical theory is due only to the distinguishability of the units and not due to finiteness of the population.

It is evident that the estimator

$$\bar{y}' = \frac{1}{\mu_s} \Sigma' y_i' \qquad \cdots(2.2.16)$$

where μ_s is the number of distinct units in the sample and the summation is over all distinct units of the sample, is not of the type (2.2.15). In (2.2.16) the coefficient c_i attached to the ith selected unit is equal to $\frac{1}{\mu_s}$ or 0 thus depending on the sample in two ways:- It depends (i) on μ_s, the effective sample size of the sample s and (ii) on all the units selected in the first (i-1) draws, being $\frac{1}{\mu_s}$ if U_i is not one of them and 0 otherwise. Thus there is a breach between the classical theory and the theory for finite populations and hence the need for defining afresh linear estimators for the latter case.

The first attempt in this direction was due to Horvitz and Thompson [22] who defined 3 classes of linear estimators. Typical members of these three classes are given by

$$T_1 = \sum_{\lambda \in s} \beta_\lambda Y_\lambda \qquad \ldots(2.2.17)$$

$$T_2 = \sum_{i=1}^{n} c_i y_i \qquad \ldots(2.2.18)$$

and

$$T_3 = \beta_s \sum_{\lambda \in s} Y_\lambda \quad . \qquad \ldots(2.2.19)$$

Members of the first class have coefficients depending only on the U_λ's to whose variate-values they are attached. Members of the second class have coefficients β_i for the variate-value of the ith selected unit where $\beta_1, \ldots, \beta_n$ are given in advance for sampling n units. Members of the last class have a single coefficient (which can be called the inflation factor for the sample). A typical example of T_3 is the ratio-estimator.

The next attempt was by Godambe [11] who generalised these concepts and defined the 'general linear estimator' in a design $D = D(S,P)$ by

$$T_s = \sum_{\lambda \in s} \beta_{s\lambda} Y_\lambda \qquad \ldots(2.2.20)$$

where the suffix 's' denotes the sample. The coefficients β's to be attached to the variate values depend both on the sample as well the units to which they are attached, but not on the variate values Y_i's. They are thus given a-priori. Evidently (2.2.17), (2.2.18) and (2.2.19) are special cases of (2.2.20). We shall call (2.2.20) the 'general

homogeneous linear estimator', (g.h.l.e, for brevity) since (2.2.20) is a homogeneous linear function in Y_i's. Recently Koop [24], proceding from an axiomatic set-up, defined seven classes of linear estimators (these are also homogeneous linear estimators). However all of these are again only special cases of (2.2.20). A g.h.l.e. is thus characterised by a double sequence $\{\{\beta_{s\lambda}\}\}$ of coefficients with $1 \leq \lambda \leq N$ and $s \subset D$.

Several interesting consequences followed the definition of g.h.l.e. as given in (2.2.20). It can be easily seen that the conditions for T_s of (2.2.20) to be an unbiased estimator (g.h.l.u.e., for brevity) of the population total Y are given by

$$\sum_{s \supset \lambda} \beta_{s\lambda} P_s = 1, \quad 1 \leq \lambda \leq N \qquad \ldots (2.2.21)$$

where the sum on the l.h.s. is over all samplescontaining the unit U_λ. Further for the variance of T_s we have

$$V(T_s) = \sum_{\lambda=1}^{N} Y_\lambda^2 \left(\sum_{s \supset \lambda} \beta_{s\lambda}^2 P_s - 1 \right) + \sum_{\lambda \neq \lambda'}^{N} \sum Y_\lambda Y_{\lambda'} \left(\sum_{s \supset \lambda, \lambda'} \beta_{s\lambda} \beta_{s\lambda'} P_s - 1 \right) \qquad \ldots (2.2.22)$$

Godambe [11] then proved the

Theorem (2.2.2):- For any general sampling design D, in the class L_u of g.h.l.u.e.'s there does not exist one which is best.

This theorem as given by Godambe is not strictly correct because there exist nontrivial sampling designs for which there exists a best estimator in the class L_u given above. The theorem (2.2.2) thus needs a slight modification. This modification, clearly characterising the exceptional cases, is given in § 7.4 of Chapter VII.

The nonexistence (in most situations) of a best estimator in the class of g.h.l.u.e's does not seem to have been noticed by some authors, since in a number of recent papers dealing with estimation by sampling from finite populations we find statements like '... is the best linear estimator ...' . These statements are thus incorrect unless one starts with some other definition of bestness.

Note :- We shall henceforth concern ourselves exclusively with unbiased estimators only. For the case of estimating Y (or any linear function of y_i's) we shall further restrict ourselves only to L_u, the class of g.h.l.u.e's.

Having thus noted the non-existence of a best estimator in the class L_u of g.h.l.u.e's of y for most sampling designs, the next logical step is to weed out 'bad' estimators.

With respect to a given design D, an estimator T of an estimable p.f. θ is said to be 'admissible' if and only if there does not exist another estimator T' of θ which is better than T, i.e. given any $T' (\neq T) \in L_u$, there exists at least one value $\underset{\sim}{Y}'$ (depending on D and T and possibly on T' also) of $\underset{\sim}{Y}$ such that

$$V(T) \Big|_{\underset{\sim}{Y}'} < V(T') \Big|_{\underset{\sim}{Y}'}$$

where both the variances in the above are evaluated at $Y = Y'$. Otherwise T is said to be 'inadmissible' (i.e., 'bad').

Towards the elimination of these inadmissible estimators there are important advances recently. Theorem (2.2.1) of Des Raj, Khamis, Raja Rao and Basu proves that for designs obtained by simple random sampling with replacement with a fixed sample size, the unweighted sample mean (which is the conventional estimator) is inadmissible being inferior to the unweighted mean over the distinct units of the sample. Murthy [25] proved that for PPS sampling design D an estimator which depends on the order of selection of units - an ordered estimator, for brevity - is inadmissible. Finally Roy and Chakravarthy [36] proved the

Theorem (2.2.3): For any sampling design D, an unbiased estimator of $\bar{Y}$ which is either an ordered estimator or depends on repetition of a unit in a sample, is inadmissible.

Des Raj and Khamis obtained their result by direct calculation of the variances of the two estimators concerned. The proofs of Murthy and Roy and Chakravarthy are constructive in the sense that given any unbiased estimator T (of θ) which is either ordered or depends on repetitions, they showed how to construct a better estimator T' which is unordered and is independent of the repetitions of units in the sample. Basu and Raja Rao proved that the unordered set of distinct units of the sample - the effective sample, for brevity - forms a sufficient statistic for the estimation of a p.f. θ and deduced their result from Rao-Blackwell theorem ([32], [2]). The proofs of Murthy and Roy and Chakravarthy

while not bringing in the notion of sufficient statistic explicitly, use the same techniques as used by Rao and Blackwell.

It may be remarked that Hájek [16] seems to have anticipated theorem (2.2.3)! Some of the essential features of this theorem are also traceable, though in a disguised form, in a much earlier work of Halmos [17].

Examples of direct application of theorem (2.2.3) to a number of particular cases are worked out by some authors recently.

It may be mentioned here that theorem (2.2.3) furnishes a comparison among various estimators of a p.f. θ all of which are defined over the same design D (in which θ is estimable). It does not furnish a comparison among various designs themselves. For a comparison between two designs D_1 and D_2 (in both of which θ is estimable) we should first compare their costs in terms of a reasonable cost function. (This point was over looked for sometime while making comparison between 'with replacement' designs and 'without replacement' designs). After equalising the costs if 'best' estimators T_1' and T_2' exist for D_1 and D_2 then one can say that D_1 is better than D_2 iff

$$V\left\{H_1(D_1, T_1')\right\} \leq V\left\{H_2(D_2, T_2')\right\}, \text{ for all } \gamma$$

with strictly inequality occurring for at least one γ. Unfortunately there are not many designs in which a 'best' estimator exists. A natural definition of betterness of D_1 over D_2 should then be as follows: D_1 is said to be better than D_2 iff given any T_2 defined over D_2 there exists a T_1 defined over D_1 (and depending possibly on T_2) such

that

$$V\left\{H_1\ (D_1,\ T_1)\right\} \leq V\left\{H_2\ (D_2,\ T_2)\right\} \text{ for all } \underset{\sim}{Y} \qquad (2\cdot2\cdot23)$$

with strict inequality occurring for at least one $\underset{\sim}{Y}$.

This point does not seem to have been noticed by a number of authors who remark '... it is well known that sampling without replacement is better than sampling with replacement ... '. They base their conclusion usually by comparing some two estimators T_1 and T_2 for which (2·2·23) holds good. This is clearly illogical. It is for this reason that we stress that estimator and design should always be considered together as one 'strategy'. We are not aware of a proof that sampling without replacement is better than sampling with replacement even for simple classes of designs like those generated by simple random sampling.

A cost function which is reasonable one in most of the cases of unistage sampling is given by

$$C_s = A\ \mu_s + B \qquad \ldots (2\cdot2\cdot24)$$

where A and B are constants independent of the sample $s \in D$. The cost of a strategy H (S, P, T) can then be defined as the expected cost of a random sample from H thus:

$$C(H) = \sum_{s \in S} C_s P_s = A\,\mu(H) + B\,.$$

Two strategies of same cost are then comparable from the point of view of their variances. Between two strategies of same cost (i.e. same value of μ) and with same variance we prefer, from a practical point of view, the one in which the variation of the cost over different samples is smaller i.e. for which $V(\mu_s)$ is smaller.

Given a design D and a class L_1 of estimators (of an estimable p.f. θ) defined over D, a subclass L_0 is said to be 'complete' in L_1 for θ iff

$$T_1 \in L_1;\ T_1 \notin L_0 \Rightarrow \exists\ T_0 \in L_0 \ni V(T_0) \leq V(T_1) \text{ for all } Y.$$

Obviously for the estimation of θ by means of members of L_1 we need restrict our search only to the subclass L_0. Since it is known that in the class L_u there doesnot exist a best to estimate Y, interest centers round in finding a narrow enough complete class in L_u and we may then enforce some further secondary criteria of bestness to get hold of a small class of optimum estimators. It can be seen that the class A_u of admissible estimators of Y forms the minimal complete class in L_u. However, a complete characterisation of members of A_u is lacking. Theorem (2.2.3) gives necessary conditions for an estimator to be admissible. It seems quite plausible that they form a set of sufficient conditions also.

In § 7.2, § 7.4 and § 7.5 of Ch.VII we make a systematic investigationtion of the problem of estimating linear parametric functions with special reference to the estimation of the population total, We show that the criterion of smaller mean square error does not automatically exclude

biased estimators, at any rate not so in preference to the class L_u. We then prove the existence in most designs (specifically for all non-unicluster designs - c.f. § 7.4) of non-linear unbiased estimators and non-homogeneous linear unbiased estimators of linear p.f.s. Further these estimators which do not belong to L_u can be chosen to be functions of order statistic alone so that there is no known method by which they can be proved to be inadmissible. In fact as remarked towards the end of the last paragraph it seems likely that such estimators are also admissible. Finally in § 7.5 we introduce a new criterion of 'hyper-admissibility' and prove that for any non-unicluster designs not only in the class L_u but in the entire class L_p of polynomial unbiased estimators of Y, there is just one which is hyper-admissible and which is therefore the 'best' in L_p under this criterion. This best estimator is the Horvitz and Thompson estimator [22] corresponding to the given design and estimates admissibly not only Y but all estimable linear p.f. In § 7.4 we also completely characterise the class of all admissible estimators for any unicluster design and notice the important role of Horvitz and Thompson estimator in this case also.

§ 2.3 Optimum utilisation of auxiliary information and Bayes' solution

Throughout this section we shall restrict our attention only to the estimation the population total Y. We also take (2.2.24) to be our cost function.

We have seen in § 2.2 that with the criteria of unbiasedness and minimum variance alone we do not have a best estimator nor do we have a narrow enough class of strategies to which only, we need restrict our attention. However, this desperate state of affairs prevails in the situation when no apriori knowledge, however imprecise, is available about the variate values $Y_1, \ldots, Y_N$. When an auxiliary information on a variate $\mathcal{X}$ which takes the values X_i on U_i $(1 \leq i \leq N)$ is available beforehand by slightly diluting the criterion of bestness this information can at times be utilised to get at an optimum sampling strategy to estimate Y. The nature of this utilisation depends on the relationship between $\mathcal{X}$ and $\mathcal{Y}$.

When information on $\mathcal{X}$ as available completely beforehand it is at times possible to make some plausible guesses about the relative magnitudes of the variate values Y_i's. Whether these guesses can always be formulated as apriori distributions is a controversial point about which we do not propose to enter into a discussion here. There are, however, a few situations in practical problems where such formulation in terms of an apriori distribution is plausible, at least partially. This concept, borrowed from Bayesian inference, is first

introduced into the sampling theory of finite populations by Cochran [3] who called it the super-population concept. In this set up the actual (though unknown before the sampling strategy is decided upon) value of the vector

$$\underset{\sim}{Y} = (Y_1, Y_2, \cdots, Y_N)$$

is supposed to be the realisation of a N length random vector having a certain distribution. We shall then treat $\underset{\sim}{Y}$ as a random vector and accordingly consider various quantities like expectations, variances and covariances of components of $\underset{\sim}{Y}$. The distribution of $\underset{\sim}{Y}$ depends on

$$\underset{\sim}{X} = (X_1, X_2, \cdots, X_N)$$

besides possibly on some unknown parameters. We shall call the distribution of $\underset{\sim}{Y}$ given $\underset{\sim}{X}$ as an 'apriori distribution' and denote it by ∂. We denote the expectations, variances and covariances taken over ∂ by a suffix '∂' e.g. E_∂, V_∂, Cov_∂ etc.

We shall now weaken our criterion of bestness thus:

For a given ceiling C on the cost (which is equivalent to a ceiling on the expected effective sample size μ, c.f (2.2.24) we consider the class $H_{e\mu}$ of sampling strategies such that if $H = H(S, P, T) \in H_{e\mu}$, then

$$\left.\begin{array}{ll} \text{i)} & T \text{ is a h.l.u.e of } Y \\ \text{ii)} & C(H), \text{ the cost of the strategy } = C, \\ & \text{(i.e.) } \mu(H) = \mu \end{array}\right\} \quad (2.3.1)$$

For any strategy $H = H(S,P,T)$ in which the estimator T is unbiased for Y, we consider

$$E_{\partial} V(H)$$

to be our loss function. If in the class of strategies $H_{(\mu)}$ of the strategies there exists one H_{opt} such that

$$E_{\partial} V(H_{opt}) \leq E_{\partial} V(H_1) \text{ for all } H_1 \in H_{(\mu)} \text{ and}$$

$$\text{for all } \underset{\sim}{X} \qquad \ldots(2.3.2)$$

then we call H_{opt} to the <u>∂-best sampling strategy</u> in H_{μ}, or that H_{opt} is ∂-optimum in H_{μ}. We thus consider the expected variance of H over the apriori distribution ∂. It is important to note that the assumption of the existence of a prior distribution ∂ is relevent only for a proper choice of the strategy and that the ultimate statement of the inference of about Y by means of an estimate of Y together with an estimated variance of this estimator of Y depend completely on the basis of the observations, as they hold good for any sampling strategy. It is only the validity of the optimality of the strategy adopted that is dependent on the prior distribution ∂. This optimality which is interms of minimum expected variance can in its turn be given as a rigorous probabilistic statement, according to some statisticians while some others dispute it and here we shall not enter into a discussion on this point.

If (2.3.2) does not hold good for all $\underset{\sim}{X}$ we may then have to assume some further models for $\underset{\sim}{X}$ which in their turn will restrict θ to a narrower class. If however (2.3.2) is satisfied for some $H_{opt} \in H_{e\mu(\mu)}$ (there can be more than one such in the class $H_{e\mu(\mu)}$) we choose one such as our strategy. It is important to note that we demand (2.3.2) to be satisfied not only for all $\underset{\sim}{X}$ but also for all possible values of other parameters, if any, that enter into the apriori distribution.

It may not be necessary to formulate completely the prior distribution θ, for the existence of H_{opt} in $H_{e\mu(\mu)}$. Two important results in this direction are due to Godambe [11] and Hájek [15] we shall describe them below. Let Δ_1 be the class of prior distributions θ_1 for which

$$\left.\begin{aligned} &\text{i)}\quad E_{\theta_1}(Y_i \mid X_i) = \alpha X_i \\ &\text{ii)}\quad V_{\theta_1}(Y_i \mid X_i) = \sigma^2 X_i^2 \\ &\text{and}\quad \text{iii)}\quad \mathrm{Cov}_{\theta_1}(Y_i, Y_j \mid X_i, X_j) = 0 \end{aligned}\right\} \quad \cdots (2.3.3)$$

Δ_1 is thus a family of distributions each depending on two unknown parameters α and σ^2. Godambe then proved that in the class $H'_{(\mu)} = H'_{(\mu)}(S', P', T')$ of sampling strategies satisfying

$$\left.\begin{aligned} &\text{i)}\quad T' \in \mathcal{L}_u, \text{ the class of g.h.l.u.e's} \\ &\text{ii)}\quad C(H'_{(\mu)}) \leq C \\ &\text{and}\quad \text{iii)}\quad \mu_s = \mu \text{ for every sample } s \in S' \end{aligned}\right\} \quad \cdots (2.3.4)$$

any $H_{0.1} = H_{0.1}(S_{0.1}, P_{0.1}, T_{0.1}) \in \mathbb{H}_{(\mu)}$ and satisfying

$$\left.\begin{aligned} &\text{i)}\quad \pi_i(H_{0.1}) = \frac{X_i}{X}\,\mu \\ &\text{and}\quad \text{ii)}\quad T_{0.1} = \overset{*}{Y}_{HT} \\ &\qquad\qquad = \sum_{\lambda \in s} \frac{Y_\lambda}{\pi_\lambda} \end{aligned}\right\} \qquad (9.3.5)$$

is ∂_1-best, with respect to any prior distribution $\partial_1 \in \Delta_1$

We briefly say that $H_{0.1}$ is Δ_1-best in $\mathbb{H}_{(\mu)}$.

Hájek considers the class Δ_2 of prior distributions ∂_2 which are more general than (9.3.3) and for which

$$\left.\begin{aligned} &\text{i)}\ E_{\partial_2}(Y_i \mid X_i) = a.X_i \\ &\text{ii)}\ V_{\partial_2}(Y_i \mid X_i) = \sigma^2 X_i^2 \\ &\text{and}\quad \text{iii)}\ \mathrm{Cov}_{\partial_2}(Y_i, Y_j \mid X_i, X_j) = w(|j-i|) \end{aligned}\right\} \quad (1 \le i+j \le N) \qquad (9.3.6)$$

where w is a convex function of $|\, j-i \,|$. Thus the covariance depends only on $|\, j-i \,|$. Clearly Δ_1 is a subclass of Δ_2. Hájek proved that with respect to any $\partial_2 \in \Delta_2$, in the class $\mathbb{H}_{(\mu)}$ of strategies defined by (9.3.4) there is only one $H_{0.2} = H_{0.2}(S_{0.2}, P_{0.2}, T_{0.2})$ which is ∂_2-optimum and that it is specified by

$$\text{i)}\quad \pi_i(H_{0.2}) = \frac{X_i}{X}.\mu$$

$$\text{ii)}\quad T_{0.2} = \sum_{\lambda \in s} \frac{Y_\lambda}{\pi_\lambda}$$

and iii) The design $D_{0.2}$ is obtained by systematic sampling to achieve (i). Clearly the strategy $H_{0.2}$ is a member of those satisfying

(2.3.5) as it should be since Δ_1 is a proper subclass of Δ_2. In this case there is just one Δ_2- optimum strategy which is completely specified. Not only the sampling design but also the corresponding sampling method is fully specified. Hence we shall not discussthe result of Hájek further.

Returning to Godambe's result, we shall indicate the various steps that lead to the final result to clarify the problem further.

With respect to a given prior distribution ∂ for any design D, and in the class L_u of g.h.l.u.e's defined over D, any T_{opt} (if one such exists) is said to be the (D, ∂) optimum iff

$$\bar{E}_{\partial_3} V(D) = E_{\partial_3} V(D, T_{opt}) \leq E_{\partial_3} V(D,T) \quad \text{for any } T \in L_u \text{ and for all values of } \underset{\sim}{X} \qquad \ldots (2.3.7)$$

Of two designs D_1 and D_2, D_1 is said to be better than D_2 (cost being ignored) iff

$\bar{E}_{\partial} V(D_1) \leq \bar{E}_{\partial} V(D_2)$ for all $\underset{\sim}{X}$. Similarly the ∂-bestness of D_1 in a class of designs, the Δ -bestness of D_1 over D_2 and the Δ -bestness of D_1 in a class of designs are defined for any class Δ of ∂'s.

Let now Δ_3 be the class of prior distribution ∂_3 for which

$$\text{i) } E_{\partial_3}(Y_i \mid X_i) = a X_i$$

$$\text{ii) } V_{\partial_3}(Y_i \mid X_i) = \sigma_i^2 \qquad (2.3.8)$$

$$\text{iii) } Cov_{\partial_3}(Y_i, Y_j \mid X_i, X_j) = 0$$

Δ_3 is thus a wider class than Δ_1 where we do not assume anything about the conditional variances.

For a given cost

$$C = A\mu + B$$

i.e. for a given integer μ, let $D_{e\mu}$, be the class of designs such that

$$E(\mu_s) = \sum_1^N \pi_i = \mu, \qquad (2.3.9)$$

$D^{(\mu)}$ be the class of designs such that

$$\mu_s = \mu \text{ for every sample } s, \qquad (2.3.10)$$

$D^{(\pi)}$ be the class of designs such that

$$\pi_\lambda = \mu \frac{X_\lambda}{X} \qquad (2.3.11)$$

and finally $D^{(\mu,\pi)}$ be the class of designs satisfying

$$\left.\begin{aligned} &\mu_s = \mu \quad \text{for every} \quad s \\ \text{and} \quad &\pi_\lambda = \mu \cdot \frac{X_\lambda}{X} \end{aligned}\right\} \qquad (2.3.12)$$

clearly

$$\left.\begin{aligned} &D_{e\mu} \supseteq D^{(\mu)} \supseteq D^{(\mu,\pi)} \\ \text{and} \quad &D_{e\mu} \supseteq D^{(\pi)} \supseteq D^{(\mu,\pi)} \end{aligned}\right\} \qquad (2.3.13)$$

Godambe then proves that for any $D \in D^{(\mu,\pi)}$, T_{opt} as defined by (2.3.7) is given by $\bar{Y}^*_{HT}$ of (2.3.5) and that

$$\bar{E}_{\partial_3} V(D) = \sum_1^N \sigma_\lambda^2 \left(\frac{1}{\pi_\lambda} - 1\right)$$

for any general sampling design D he then proves that

$$\bar{E}_{\partial_3} V(D) \geq \sum_1^N \sigma_\lambda^2 \left(\frac{1}{\pi_\lambda} - 1\right) \qquad (2.3.14)$$

He then concludes that the lower bound for $E_{\partial_3} V(D, T)$ taken over

$T \in L_u$ is in fact attained for designs $D \in D^{(\mu,\pi)}$. Further, minimising the bound on the r.h.s. of (2.3.14) subject only to the condition

$$\sum_{1}^{N} \pi_\lambda = \mu$$

he concludes that if $\partial_3 \in \Delta_1$ i.e. if it further satisfies (ii) of (2.3.3), then for whatever sampling design $D' \in D^{(\mu)}$

$$\bar{E}_{\partial_1} V(D \in D^{(\mu,\pi)}) \leq \bar{E}_{\partial_1} V(D') \qquad (2.3.15)$$

The author [2 0] has proved that while $\overset{*}{Y}_{HT}$ is Δ_3- optimum in the class $D^{(\mu,\pi)}$ of designs (as proved by Godambe), the class $D^{(\mu,\pi)}$ is Δ_3-optimum in the class $D^{(\pi)}$, for the use of $\overset{*}{Y}_{HT}$, i.e.

$$E_{\partial_3} V(D_1, \overset{*}{Y}_{HT}) \leq E_{\partial_3} V(D_2, \overset{*}{Y}_{HT}) \qquad (2.3.16)$$

for any $D_1 \in D^{(\mu,\pi)}$ and $D_2 \in D^{(\pi)}$. This result is given in § 5.2 of chapter V.

While Godambe proved the Δ_1-optimality of $D^{(\mu,\pi)}$ in the class $D^{(\mu)}$, it can be seen from the minimisation that lead to (2.3.15) from (2.3.14) that he has in fact proved the Δ_1-optimality of $D^{(\mu,\pi)}$ in the wider class $D_{e\mu}$ of designs for which the expected cost is C. This fact was not clearly noted by Godambe in his work. We thus have

Theorem 2.3.1: For any $\theta_1 \in \Delta_1$, the class $D^{(\mu,\pi)}$ of designs satisfying (2.3.12) is θ_1-optimum in the wider class $D_{e\mu}^{(0)}$ satisfying (2.3.9) which are chafacterised by an expected cost C. The best estimator in this optimum class of designs $D^{(\mu,\pi)}$ is given by $\overset{*}{Y}_{HT}$ as given by (2.3.5).

Thus we can actually replace the condition (iii) of (2.3.4) by a weaker condition viz. (ii) of (2.3.13).

The author [20] has given a direct algebraic proof of a special case of this theorem by proving that for $\mu = 2$, the strategy $H(D_1, \overset{*}{Y}_{HT})$ where $D_1 \in D^{(\mu,\pi)}$ is better than $H(D_2, \overset{*}{Y}_{SD})$ where D_2 is the sampling design obtained by taking two units from the population with probabilities proportional to X_i's and without replacement and $\overset{*}{Y}_{SD}$ is the symmetrised estimator of Des Raj [6] as given by Murthy [26]. The magnitude of this difference is given there in. This result for its own algebraic interest is given in § 5.3 of chapter IV.

It is of interest to find classes of plausible prior distributions like Δ_1 and Δ_2 for which optimum strategies exist. Another direction of research from the practical point of view, is to generate these optimum strategies in the cases where they are proved (as in the results of Godambe and Hájek) to exist. While Hájek's optimum strategy is completely identified, that of Godambe is only partially specified by (2.3.12). It is therefore of interest to find suitable sampling

designs which satisfy (2.3.12) for given values of $\mu, X_1, X_2, \ldots, X_N$ and which can be obtained in simple ways. This problem has in its own way attracted the attention of several authors even prior to the publication of Godambe's results. We mention the works of Horvitz and Thompson [22], Yates and Grundy [40], Durbin [9], Grundy [14], DesRaj [7], Stevens [39], Rao, Hartely and Cochran [35] and Felligi [10]. Goodman & Kish [13]

The author's contributions to this problem are given in Chapters IV and V and some related problems are considered in Chapter VI.

§ 2.4 The Horvitz and Thompson's estimator

Before reviewing the work of the earlier authors and giving our results we shall pause for a while to examine the estimator

$$\overset{*}{Y}_{HT} = \sum_{\lambda \in s} \frac{Y_\lambda}{\pi_\lambda} \qquad \cdots (2.4.1)$$

of the population total Y, that plays such a central role in Godambe's as well as Hájek's results. The estimator is based only on the distinct units in the sample and is independent of the order of the units in the sample and hence can not be uniformly improved upon by the methods of Murthy or Roy and Chakravorty. In fact it is proved by Roy and Chakravorty and independently by Godambe [12] that it is an admissible estimator and hence can not be uniformly improved upon. In fact as we shall see in § 7.2 of chapter VII if for any sampling design a best estimator exists it can only be $\overset{*}{Y}_{HT}$ and that there exist sampling designs in which it is the best. It is an estimation procedure (an estimator having to be defined over a design D) which gives an estimator applicable in any sampling design D in which Y is estimable. The condition for the later is that

$$\pi_\lambda > 0 \quad \text{for all} \quad \lambda$$

(This later condition, however, is not due to the fact that π_λ occurs in the denominator, because it does not occur so, unless there is a sample s containing U_λ in which case $\pi_\lambda > 0$). It is criticised

sometimes that it is not a 'reasonable' estimator in the sense that even when Y_λ's are exactly proportional to π_λ's (which seems to be the ideal situation and is a special case of a $\partial_1^{(0)} \in \triangle_1$ for which σ^2 of (2.3.6) is zero) it does not have zero variance having a component of variance due to variations in μ_s. c.f. § 5.2.1 It should however be noted that the $\triangle_1$-optimality of this estimator is not proved in its own way but only in conjunction with a specific class of sampling designs $D(\mu,\pi)$ satisfying (2.3.12) in which designs it has zero variance since $\mu_s = \mu$ for all samples 's' of those designs. It is for this reason that we stress that an estimator or a design should not be taken as separate pieces but should always be considered together as a strategy. The only logical way of defining the supremacy of an estimator T_1 over another estimator T_2 is to establish that the estimator T_1 applied in the design $D(T_1)$ which is optimum for T_1 is superior to the estimator T_2 applied in its own optimum design $D(T_2)$, by putting constraints on $D(T_1)$ and $D(T_2)$ through a cost function. A similar argument holds good for a comparison of designs also.

We now turn to the variance of $\overset{*}{Y}_{HT}$. This is given by

$$V(\overset{*}{Y}_{HT}) = \sum_{\lambda=1}^{N} \frac{Y_\lambda^2}{\pi_\lambda} + \sum_{\lambda \neq \lambda'}^{N} \sum \frac{Y_\lambda Y_{\lambda'}}{\pi_\lambda \pi_{\lambda'}} - Y^2$$

$$= \sum_{\lambda=1}^{N} Y_\lambda^2 \left(\frac{1-\pi_\lambda}{\pi_\lambda}\right) + \sum_{\lambda \neq \lambda'}^{N} \sum \frac{Y_\lambda Y_{\lambda'}}{\pi_\lambda \pi_{\lambda'}} (\pi_{\lambda\lambda'} - \pi_\lambda \pi_{\lambda'}) \quad \ldots (2.4.2)$$

For an unbiased estimator of $V(\overset{*}{Y}_{HT})$ we have

$$\overset{*}{V}_{HT}(\overset{*}{Y}_{HT}) = \sum_{\lambda \in s} Y_\lambda^2 \frac{1-\pi_\lambda}{\pi_\lambda^2} + \sum_{\substack{\lambda \neq \lambda' \\ \lambda,\lambda' \in s}} \sum \frac{Y_\lambda Y_{\lambda'}}{\pi_\lambda \pi_{\lambda'}} \frac{(\pi_{\lambda\lambda'} - \pi_\lambda \pi_{\lambda'})}{\pi_{\lambda\lambda'}} \qquad \ldots (2.4.3)$$

The derivation of this follow straight away from the unbiasedness of $\overset{*}{Y}_{ht}$ by merely arguing for different populations, and this is also due to Horvitz and Thompson. This has been criticised by Yates and Grundy [4 9] on three grounds viz., (i) it is not reducible to a linear function of $(\frac{Y_i}{\pi_i} - \frac{Y_j}{\pi_j})^2$'s (ii) it does not vanish when $\frac{Y_i}{\pi_i} = \frac{Y_j}{\pi_j}$ for all $i \neq j$ and (iii) it may assume negative values. They proposed an alternative estimator of (2.4.2) for the class $D^{(u)}$ of sampling designs in which

$$\mu_s = \mu$$

and this is given by

$$\overset{*}{V}_{YG}(\overset{*}{Y}_{HT}) = \sum_{\substack{\lambda \neq \lambda' \\ \lambda,\lambda' \in s}} \sum \frac{\pi_\lambda \pi_{\lambda'} - \pi_{\lambda\lambda'}}{\pi_{\lambda\lambda'}} \left(\frac{Y_\lambda}{\pi_\lambda} - \frac{Y_{\lambda'}}{\pi_{\lambda'}}\right)^2 \qquad \ldots (2.4.4)$$

of the three grounds on which (2.4.3) is criticised, the justification for demanding (i) is not clear. Regarding (ii) we remark that while it is true that (2.4.3) does not in general vanish even when $\frac{Y_i}{\pi_i}$'s are equal to a constant

$$\overset{*}{V}_{HT}\left(\overset{*}{Y}_{HT}\right) = \sum_{\lambda \neq \lambda'} \sum \frac{\pi_{\lambda\lambda'}}{\pi_\lambda \pi_{\lambda'}} - \mu \qquad \ldots(2.4.6)$$

so that for sampling designs satisfying

$$\sum_{\lambda \neq \lambda'} \sum \frac{\pi_{\lambda\lambda'}}{\pi_\lambda \pi_{\lambda'}} = \mu \qquad \ldots(2.4.7)$$

(2.4.3) vanishes identically. However (2.4.4) vanishes identically for any sampling design in this case. However, (2.4.3) is applicable in any design in which $V\left(\overset{*}{Y}_{HT}\right)$ is at all estimable, while (2.4.4) is applicable in a restricted though useful class for which $\mu_s = \mu$ for all s.

§ 2.5 Formulation of the main problem

We have already formulated our problem ~~forwards~~ towards the end of § 2.3 as that of finding sampling methods that generate designs satisfying (2.3.12), for given values of $\pi_1, \ldots, \pi_N$ which are proportional to known auxiliary information $X_1, \ldots, X_N$, the constants of proportionality being equal to the expected sample size which we wish to have. This is so because we know that these designs together with the estimator (2.4.1) defined over them constitute Δ-optimum sampling strategies. There are however some further conditions that we may like to demand. Usually it is not only of interest to estimate Y (which we are doing in a Δ_1-optimal way by selecting designs satisfying (2.3.12), and using $\overset{*}{Y}_{HT}$) but also of interest to estimate the variance (2.4.2) of our estimator. For this a necessary and sufficient set of conditions are (cf. §7.3)

$$\pi_{\lambda\lambda'} > 0 \quad \text{for} \quad \lambda = \lambda' \qquad \ldots(2.5.1)$$

Again since (2.4.2) being a variance is a non-negative function it is reasonable to demand that an estimate of the same be found which is always non-negative. If the class of designs we are choosing are any way satisfying (2.3.12) then (2.4.4) can be used and the conditions for the non-negativeness of (2.4.4) can be easily expressed as

$$\pi_{\lambda\lambda'} \leq \pi_\lambda \pi_{\lambda'} \quad \text{for all} \quad \lambda \neq \lambda' \qquad \ldots(2.5.2)$$

The conditions can not be written in easy form if we wish to use (2.4.3). A set of sufficient conditions for (2.4.3) to be non-negative for all positive values of $Y_1, \ldots, Y_N$ (in fact these alone can be written down as are of interest to us)

$$\pi_{\lambda\lambda'} \geq \pi_\lambda \pi_{\lambda'} \quad \text{for all} \quad \lambda \neq \lambda' \tag{2.5.3}$$

but these are not a consistent set of conditions if strict inequality occurs at least once. For, it can be easily shown that in any design D having a constant effective sample size of μ, we have

$D \in D(\mu)$

$$\left.\begin{aligned} \Sigma\, \pi_\lambda &= \mu \\ \sum_{\lambda' \neq \lambda} \pi_{\lambda\lambda'} &= (\mu - 1)\, \pi_\lambda \\ \text{and} \qquad \sum_{\lambda \neq \lambda'} \sum \pi_{\lambda\lambda'} &= \mu\,(\mu - 1) \end{aligned}\right\} \quad \ldots (2.5.4)$$

If $\pi_{\lambda\lambda'} \geq \pi_\lambda \pi_{\lambda'}$ for all $\lambda \neq \lambda'$, and strict inequality occurs at least once we have

$$\sum_{\lambda \neq \lambda'} \sum \pi_{\lambda\lambda'} > \sum_\lambda \pi_\lambda\, (\mu - \pi_\lambda) = \mu^2 - \Sigma\, \pi_\lambda^2$$

which together with the last equation of (2.5.4) yields

$$\Sigma\, \pi_\lambda^2 > \mu$$

which is clearly impossible since $\Sigma\, \pi_\lambda^2 \leq \Sigma\, \pi_\lambda = \mu$, each π_λ being

a positive number not exceeding 1. We shall therefore use (2.4.4) to extimate $V(\overset{*}{Y}_{HT})$.

While getting an unbiased estimator of $V(\overset{*}{Y}_{HT})$ we would like to have one which itself has a small variance so that our estimated error is a reliable one. For this it can be seen intuitively that $\pi_{\lambda\lambda'}$'s should not be too small as they occur in the denominators in both (2.4.3) and (2.4.4). If there are some $\pi_{\lambda\lambda'}$'s which are very small (compared to $\pi_\lambda \pi_{\lambda'}$) then the corresponding terms in (2.4.3) and (2.4.4) will be very large. But since (2.4.3) and (2.4.4) are unbiased estimators of the variance, they have to take small values also for some other pairs (i,j) so that the estimators (2.4.3) and (2.4.4) of the variance (2.4.2) will be highly variable. It is thus desirable to have $\pi_{\lambda\lambda'}$'s while being less than $\pi_\lambda \pi_{\lambda'}$ (if it can be helped) should not be too small i.e.

$$\frac{\pi_{\lambda\lambda'}}{\pi_\lambda \pi_{\lambda'}} \quad \text{should not be too small} \quad \lambda \neq \lambda' \qquad \cdots(2.5.5)$$

For the estimation of $V(\overset{*}{Y}_{HT})$, using either (2.4.3) or (2.4.4), we need the values of $\pi_{\lambda\lambda'}$'s. It is therefore necessary that the sampling design (or equivalently the sampling method) which we are in search of should not only satisfy (2.5.1) but also yield formulae for $\pi_{\lambda\lambda'}$'s which are easily calculable given the sampling method (unless one thinks of listing down $\pi_{\lambda\lambda'}$'s for all values of λ and λ' !).

Finally, as a matter of practical convenience we may add that the method that we find should be neat and tractable and should not involve too heavy computations.

We shall now formulate our main problem thus:

Given a population $\mathcal{U}$ of N units

$$U_1, U_2, \ldots, U_N$$

on each of which are defined two characters $\mathcal{X}$ and $\mathcal{Y}$ which take the values

$$X_1, X_2, \ldots, X_N$$

and

$$Y_1, Y_2, \ldots, Y_N$$

respectively, all X_i's being positive, and completely known in advance, and given a positive integer μ such that

$$\frac{\mu X_i}{X} \leq 1 \quad \text{for all } i$$

(this is necessary since $\pi_i = \frac{\mu X_i}{X}$), to find a sampling method for which the corresponding design $D = D(S, P)$ satisfies

$$
\left.
\begin{array}{lll}
\text{i)}\ \pi_\lambda = \dfrac{\mu X_\lambda}{X} & (1 \le \lambda \le N) & \\
\text{ii)}\ \mu_s = \mu & \text{for all } s \in S & \\
\text{iii)}\ \pi_{\lambda\lambda'} > 0 & (\text{for } 1 \le \lambda \ne \lambda' \le N) & \\
\text{iv)}\ \pi_{\lambda\lambda'} \le \pi_\lambda \pi_{\lambda'} & (\quad '' \quad '' \quad) & \\
\text{v)}\ \dfrac{\pi_{\lambda\lambda'}}{\pi_\lambda \pi_{\lambda'}} \text{ is not too small} & (\quad '' \quad '' \quad) & \\
\text{vi)}\ \pi_{\lambda\lambda'} \text{ is easily computable from a simple formula} & & \\
\text{and vii) The computations involved in the method are not very heavy.} & &
\end{array}
\right\} \quad (2.5.6)
$$

We have listed down the properties of the design in the order of their importance.

Conditions (i) and (ii) given above are the crucial conditions. In view of the repeated reference to the condition (i) we shall call a sampling design satisfying that condition as a 'πPS sampling design (i.e. the π_i's are proportional to the sizes of the units) in analogy with and distinct from the 'PPS sampling' (probability proportional to the size) designs that are now commonly used.

§ 2.6. Review of previous work in the chronological order:

In this section we shall briefly review the various solutions of other authors to the problem of § 2.5. In what follows, (i), (ii),..., (vii) refer to the conditions of (2.5.6).

Goodman and Kish [13] gave a method analogous to systematic sampling to achieve given values of π_λ's. This consists of arranging the units in some order, cumulate their π values and then take a linear systematic sample by choosing a random number between 0 and 1 and with an interval of 1. This method is valid for general values of π_i's, satisfies (i) and (ii) is satisfied approximately. However (iii) is not satisfied so that the sampling variance can not be estimated. This difficulty can be overcome by modifications similar to those given later in § 5.4. But the major draw back is that there does not exist a neat way of calculating $\pi_{\lambda\lambda'}$'s. Recently, Hartley and Rao [21] considered a modification of the above procedure by randomly arranging the units before drawing the sample, and obtained compact approximations to $\pi_{\lambda\lambda'}$'s using an asymptotic theory, for large N, on the assumption that, there exists a constant B such that $\pi_\lambda \leq \frac{B}{N}$ for all λ.

Horvitz and Thompson [22] gave two solutions to the problem. The first of them, meant for all integral values of μ, is that given by Midzuno [25] and Sen [38] in a different context and consists of selecting one unit with probability p_λ to U_λ $(1 \leq \lambda \leq N)$ and selecting $(\mu-1)$ more from the remaining $(N-1)$ by simple random sampling without replacement. They then proceed to find the p_λ's so that

the scheme results in a given set of π_λ's. This scheme is valid only if $\pi_\lambda \geq \frac{\mu - 1}{N - 1}$ for all λ and since $\Sigma \pi_\lambda = \mu$ this implies that the variation in the sizes is small. In such a situation, however, varying probability sampling is of little use. The other solution given for case $\mu = 2$, is valid only if $\pi_\lambda \leq 0.25$ for all λ, and is based on the approximation that in sampling with varying probabilities the expressions for the π_λ's are the same for sampling with or without replacement. Unless the sampling fraction is low this is not a good approximation. Thus both their schemes are of little use.

Yates and Grundy [40] gave a solution for $\mu = 2$. This also consists in selecting two units with probability p_λ for U_λ and without replacement. The p_λ's are to be found by equating the resulting inclusion probabilities to the given π_λ's. Thus their approach closely resembles that of Horvitz and Thompson for this case but goes a step further in that it does not assume that the sampling fraction is low. However, the solution for p_λ's can only be obtained by a cumbersome iteration procedure and the question of its convergence is not dealt with. Their method thus satisfies (i), (ii), (iii) and (vi)

Narain [28] considers two-stage sampling in which 2 primary units are to be selected with varying probabilities, θ_i's say, without replacement, n second stage unit are to be selected from each of the selected primary units by simple random sampling without replacement. He then evaluates θ_i's so that the usual sample mean per second stage unit is an unbiased estimator of the corresponding population mean.

This work, however, does not have any direct bearing on the present study.

Durbin [9] gave two procedures. The first consists of selecting μ units with probability $\frac{\pi_i}{\mu}$ for U_i and with replacement. If there are repetitions, the sample is rejected and another fresh sample of size μ is selected. The procedure is continued until an acceptable sample is obtained. This method does not ensure (i). A biased estimator is considered and a biased estimate of its mean square error is given. These biases may be small if π_i's are nearly equal to each other, as he remarks, but this situation is not of much interest in practice. He gave another procedure which consists of splitting the population into μ strata in any manner we like and then to select one unit from each stratum with probability proportional to π_i's. This does not result in final inclusion probabilities equal to π_i's but an unbiased estimate of population total is provided with a conservative non-negative estimate of its variance. The method though given for $\mu = 2$ can be generalised directly.

Grundy [14] gave a procedure which is a variation of that of Goodman and Kish.

Des Raj [7] considers the case $\mu = 2$ and obtains the π_{ij}'s $(1 \leq i \neq j \leq N)$ such that $\pi_{ij} \geq 0$ and $\sum_{j \neq i} \pi_{ij} = \pi_i$ $(1 \leq i \neq j \leq N)$ such that

$$\pi_{ij} \geq 0 \quad \text{and} \quad \sum_{j \neq i} \pi_{ij} = \pi_i \qquad (1 \leq i \neq j \leq N) \qquad (2.6.1)$$

by linear programming. The method as also the selection of the sample thereupon are very laborious even for moderately large N. It is of some interest to make a few remarks on this work. Des Raj obtained the set of π_{ij}'s not only satisfying (2.6.1) but also to

$$\text{minimise} \sum_{i \neq j} \sum \frac{\pi_{ij}}{\pi_i \pi_j} \text{, subject to (2.6.1)} \qquad \ldots (2.6.2)$$

He considers $V(\overset{*}{Y}_{HT})$ given by (2.4.2) and argues that for given values of π_i's it is minimised for the set of π_{ij}'s satisfying (2.6.1) and minimising

$$\sum_{i \neq j} \sum \frac{Y_i Y_j}{\pi_i \pi_j} \cdot \pi_{ij} \qquad (2.6.3)$$

Since the Y_i's themselves are unknown he assumes a mathematical model thus:

$$Y_i = a X_i + b \qquad (1 \leq i \leq N) \qquad (2.6.4)$$

where X_i's are known and a and b are unknown parameters. He then substitutes (2.6.4) in (2.6.3) which gives (2.6.2). However, if one assumes the mathematical model (2.6.4) then any two units U_i and U_j with $X_i \neq X_j$ can be selected with probability 1, a and b can then be found from (2.6.4)

$$a = \frac{Y_j - Y_i}{X_j - X_i}$$

$$b = (Y_i X_j - Y_j X_i)/X_j - X_i$$

and Y can be estimated with zero variance, so that the problem is trivially solved. If however he meant (2.6.4) to be a stochastic model

i.e.

$$E(Y_i \mid X_i) = a X_i + b \qquad (2.6.5)$$

then even under the simple assumptions that

$$V(Y_i \mid X_i) = \sigma^2 X_i^2$$
$$\text{and} \qquad \text{Cov}(Y_i, Y_j \mid X_i, X_j) = 0 \qquad (2.6.6)$$

$V(\overset{*}{Y}_{HT})$ is minimised, for given values of π_i's and X_i's, uniformly for all values of a and b, if and only if

$$\sum\sum_{i\neq j} \frac{X_i X_j}{\pi_i \pi_j} \pi_{ij}; \quad \sum\sum_{i\neq j} \frac{X_i + X_j}{\pi_i \pi_j} \pi_{ij} \quad \text{and} \quad \sum\sum_{i\neq j} \frac{\pi_{ij}}{\pi_i \pi_j}$$

are all simultaneously minimised. If π_i is taken proportional to X_i, minimisation of the first of there is achived (c.f. lemma 5.2.1) by having μ_s same (i.e. $\mu_s = 2$ in this case) for all samples, which is achieved in this case. However one still has to minimise the last two terms simultaneously which is not what Des Raj has done. If, however, one likes to take a simpler model with $b = 0$ in (2.6.5) then the last two terms are irrelevant (as can also be seen from theorem (2.2.3)) in which case Des Raj's solution is redundant. Thus his solution in so far as it is gives a method for $\mu = 2$ is valid (though is unwieldly) but the optimality properties claimed therein are not valid.

Stevens [39] gave a solution for any integral μ under the assumption that the π_i's can be divided into groups of equal values, each group consisting of not less than μ elements. This implies some sort of rounding off of the X values. His method than consists of selecting μ groups at random with probability proportional to the

total of the π values in the group and with replacement, if a group is selected k times, k units from that group are to be choosen by simple random sampling without replacement. He gave an estimator of the sampling variance which again is not positive definite. His solution thus satisfies (i), (ii), (iii) and (vii), under some restrictions on Z_i's.

Hájek [16] gave a 'simple' solution to achieve (i) which he terms as 'Poisson sampling'. This consists of conducting N independent binomial trials, the ith one with a probability of success π_i and to include all the units for which the corresponding trials resulted in a success. He arrives at this solution through an optimality criterion which is entirely different from ours. In this case $\pi_{\lambda\lambda'} = \pi_\lambda \pi_{\lambda'}$ so that the variance estimator (2.4.3) due to Harvitz and Thompson is non-negative. All the conditions except (ii) are satisfied most satisfactorily but the deviation from (ii) is so great as to make the procedure almost useless in practice where the variation in actual cost has to be within reasonable limits, and also because (ii) is also a basic requirement. The procedure yields samples of sizes 1 to N and even a sample of size 0 can be obtained, when we have to estimate the population total as 0.

Rao, Hartley and Cochran [35] gave a procedure which is essentially the same as the latter of the two methods given by Durbin [9], while Durbin advocated that the splitting of the population

into μ strata can be done in any manner we like bringing in any additional advantages that might result in, these authors suggest the population be divided at random into μ strata of approximately equal sizes. The estimator considered is the same as that given by Durbin but now they are able to get a non-negative unbiased estimator of its variance instead of a conservative one given by Durbin. Their procedure thus does not aim at satisfying (i) but satisfies (ii), (iii) and gives a non-negative unbiased estimator of the sampling variance.

Felligi [10] gave a procedure which is suitable to rotation samplings also. This consists of selecting μ systems of probabilities $\{ p_i^{(k)}, \; i = 1, \ldots, N \}$ for $k = 1, 2, \ldots, \mu$ and to take a sample of size μ without replacement using these probabilities. The probability systems are then sought in a manner such that the unconditional probabilities of selection of the ith unit equals π_λ / μ for all λ and for each of the μ draws made. The computations involved look prohibitive and the question of convergence of the iteration adopted was not investigated. There does not seem to be a manageable method to compute the $\pi_{\lambda\lambda'}$'s. The method thus satisfies (i) approximately, and (ii) and (iii). For the case $\mu = 2$ his procedure is slightly simpler under an additional assumption that $\pi_\lambda \leq \frac{2}{3}$ for all i - and of course this is not a serious practical limitation - and reduces to an iterative solution similar to that given by Yates and Grundy. In this case the calculation of $\pi_{\lambda\lambda'}$'s is relatively simple but it is not known whether (iv) and (v) are satisfied. The author's contributions to this problem mentioned briefly in Chapter I are given in Chapters IV and V.

choice of an optimum probability measure on the class $\mathcal{S}_0$ of subsets of $\mathcal{U}$, and a corresponding optimum estimator defined over the design thus obtained. Lastly, after obtaining an optimum strategy on the basis of some optimality criteria if the design of this optimum strategy is not fully specified, the result assures us that corresponding to any further specifications (consistent with the original ones) that completely specify the design, there is a mechanism of drawing units that gives rise to a design with the given specifications. By choosing these further specifications in a number of ways we can methodically get a number of drawing mechanisms all giving rise to designs satisfying the original requirements. By properly choosing these further specifications we may get a simple drawing mechanism that yields the required type of design. This point is illustrated fully in § 3.3 by means of an example from the theory of ratio-estimation.

§ 3.2. Drawing mechanisms and the main theorem.

We now define a drawing mechanism as an algorithm A given by three functions q_1, q_2, q_3 thus:

$$A = A\left\{ q_1(U_i),\ q_2(s_{(k)}),\ q_3\ (s_{(k)},\ U_i) \right\} \quad \cdots \qquad (3.2.1)$$

In the above, q_1 is a probability measure on $\mathcal{U}$, so that

$$\left.\begin{aligned} & q_1\ (U_i) \geq 0 \qquad (1 \leq i \leq N) \\ \text{and} \quad & \sum_{i=1}^{N} q_1\ (U_i) = 1 \end{aligned}\right\} \qquad (3.2.2)$$

$q_2\ (s_{(k)})$ defined for a typical sample $s_{(k)}$ of size k, is a number lying in $(0, 1)$. i.e.,

$$0 \leq q_2\ s_{(k)} \leq 1 \qquad \cdots \qquad (3.2.3)$$

Finally, q_3, defined only for $s_{(k)}$'s such that $q_2(s_{(k)}) \neq 0$, is a probability measure on $\mathcal{U}$. i.e.,

$$\left.\begin{aligned} & q_3\ (s_{(k)},\ U_i) \geq 0 \\ \text{and} \quad & \sum_{i=1}^{N} q_3\ (s_{(k)},\ U_i) = 1. \end{aligned}\right\} \quad \text{if. } q_2(s_{(k)}) \neq 0 \qquad (3.2.4)$$

The sample is then drawn using the algorithm A as follows. First a unit is selected from $\mathcal{U}$ according to the measure q_1. If the sample thus obtained is $s_{(1)} = \{U_{i_1}\}$, $s_{(1)}$ is then imputed in q_2, and the valued noted. Next, a binomial trial is conducted with probability of success $q_2\ (s_{(1)})$. If the trial results in a failure the drawing is terminated and $s_{(1)}$ is taken as the sample. Otherwise,

a second unit is drawn from $\mathcal{U}$ according to the measure $q_3(s_{(1)}, U_i)$. If the realised sample of size two thus obtained is $s_{(2)} = \{U_{i_1}, U_{i_2}\}$, this $s_{(2)}$ is imputed in q_2 and the value noted. Again a binomial trial is conducted with probability of success $q_2(s_{(2)})$. If the trial results in a failure the drawing is terminated and $s_{(2)}$ is taken as the sample. Otherwise a third unit is drawn from $\mathcal{U}$ using the measure $q_3(s_{(2)}, U_i)$, and so on. The cycle of operations continues until a sample $s_{(m)}$ is obtained for which $q_2(s_{(m)}) = 0$.

Thus, drawing a sample of fixed size n is achieved by a q_2 with

$$q_2(s_{(n)}) = 0 \quad \text{for every sample } s_{(n)} \text{ of size } n;$$

sampling with equal probabilities correspond to uniform measures q_3 in which all nonzero probabilities are equal to each other (for any fixed $s_{(sk)}$); and sampling without replacement are characterised by q_3's such that

$$q_3(s_{(k)}, U_i) = 0 \quad \text{if } s_{(k)} \text{ contains } U_i.$$

We now prove our main

Theorem (3.2.1): There is (1,1) correspondence between sampling designs and drawing mechanisms such that if D is a design and $A(D)$ is the corresponding drawing mechanism, then sampling according to $A(D)$ gives D.

Proof: That any given drawing mechanism A gives rise to a unique design $D(A)$, is evident. To prove the converse, let $D = D(S,P)$ be any given design. Let S_{i_1} $(1 \le i_1 \le N)$ be the set of all samples s of S, whose first element is U_{i_1}, $S_{i_1 i_2}$ $(1 \le i_1, i_2 \le N)$ be the set of all samples of S whose first element is U_{i_1} and second element is

U_{i_2}. Similarly $S_{i_1 i_2 i_3}$ etc., are defined. Note that at least one of S_{i_1}'s has to be nonempty. Let further

$$s(i_1) = (U_{i_1})$$

$$s(i_1, i_2) = (U_{i_1}, U_{i_2}) \qquad (1 \leq i_1, i_2, \cdots, \leq N)$$

etc., and let their corresponding probabilities be $P(i_1)$, $P(i_1, i_2)$ etc. Clearly

$$S = \bigcup_{i_1} S_{i_1}$$

$$S_{i_1} = \left(\bigcup_{i_2} S_{i_1 i_2}\right) \cup s(i_i)$$

$$S_{i_1 i_2} = \left(\bigcup_{i_3} S_{i_1 i_2 i_3}\right) \cup s(i_1 i_2)$$

etc. For brevity let

$$\alpha_{i_1} = \sum_{s \in S_{i_1}} P_s, \qquad \alpha_{i_1 i_2} = \sum_{s \in S_{i_1 i_2}} P_s \quad \cdots \quad (3.2.5)$$

etc., so that we have

$$\left.\begin{aligned} 1 &= \sum_{i_1} \alpha_{i_1} \\ \alpha_{i_1} &= \sum_{i_2} \alpha_{i_1 i_2} + P(i_1) \\ \alpha_{i_1 i_2} &= \sum_{i_3} \alpha_{i_1 i_2 i_3} + P(i_1, i_2) \end{aligned}\right\} \cdots \quad (3.2.6)$$

etc. We now define q_1 by

$$q_1(U_{i_1}) = \alpha_{i_1} \quad \cdots \quad (3.2.7)$$

From (3.2.5) and (3.2.6), it follows that q_1 is a probability measure on $\mathcal{U}$. For samples $s_{(i)} = s(i_1)$ of size 1, define

$$q_2(s(i_1)) = 1 - \frac{P(i_1)}{\alpha_{i_1}} \quad \text{if } \alpha_{i_1} \neq 0 \quad \cdots \quad (3.2.8)$$

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

From (3.2.6) it is clear that $0 \leq q_2(s(i_1)) \leq 1$. Finally define

$$q_3(s(i_1), U_{i_2}) = \frac{\alpha_{i_1 i_2}}{\alpha_{i_1} - P(i_1)} \quad \text{if } q_2(s(i_1)) \neq 0 \quad \cdots (3.2.9)$$

for $1 \leq i_2 \leq N$. Again from (3.2.6), it follows that q_3, if defined, is a probability measure on $\mathcal{U}$.

Analagously to (3.2.8) and (3.2.9) we define q_2 and q_3 successively for samples of sizes 2, 3, ... such that for any sample

$$s(i_1, \cdots, i_k) = (U_{i_1}, \cdots, U_{i_k})$$

of S, we have

$$q_2(s(i_1, \cdots, i_k)) = 1 - \frac{P(i_1, \cdots, i_k)}{\alpha_{i_1, \cdots, i_k}}$$

$$\text{if } \alpha_{i_1, \cdots i_k} \neq 0 \quad \cdots \quad (3.2.10)$$

$$\text{and } q_3(s(i_1, \cdots, i_k), U_{i_{k+1}}) = \frac{\alpha_{i_1, \cdots, i_k i_{k+1}}}{\alpha_{i_1, \cdots, i_k} - P(i_1, \cdots, i_k)}$$

$$\text{if } q_2(s(i_1, \cdots, i_k)) \neq 0 \quad \cdots (3.2.11)$$

From (3.2.6), (3.2.10) and (3.2.11) it followsthat q_2, lies in (0,1) and that q_3 whenever defined is a probability measure on $\mathcal{U}$. Thus these q's satisfy (3.2.2), (3.2.3) and (3.2.4). If by successively applying the algorithm we come across a sample $s(i_1, \cdots, i_k)$ for which $q_2 = 0$ we then terminate the sampling and q_3 $(s(i_1, \cdots, i_k), U_i)$'s are not defined nor are the subsequent q_2's and q_3's.

Further since S is a countable set, it follows (as remarked in § 2.1) that the total probability of samples of size greater than or equal to m tends to zero as m tends to infinity so that with probability 1 the algorithm given above terminates after a finite number of steps.

From the construction of the q's as given above, it is clear that drawing units from $\mathcal{U}$, one by one with replacement, according to the algorithm with the above q's, gives the design D. Further, any algorithm that gives rise to D has to satisfy equations (3.2.6) to (3.2.11) and hence the algorithm is unique.

The theorem is proved.

We shall say that A (D) is the generating mechanism of the design D and D(A) is the design generated by the drawing mechanism A.

§ 3.3. Application to a problem in ratio-estimation.

We have seen that any design D is generated by a drawing mechanism A(D). If the design D is partially specified, corresponding to any further specifications which are consistent with the given specifications and which together with them completely specify the design D, we have a drawing mechanism $A(D)$ generating a design satisfying the original specifications. A good deal of ingenuity lies in so choosing these additional specifications that the resulting mechanism A is a simple one. We shall illustrate this point by means of a problem from the theory of ratio-estimation.

We consider two real-valued characters $\mathcal{X}$ and $\mathcal{Y}$ defined for each unit U_i of $\mathcal{U}$. The X_i's $(1 \leq i \leq N)$ are known completely in advance. The problem is to estimate Y, on the basis of a design D. In some situation when the value of $\mathcal{Y}$ on U_i can be taken to be an unknown multiple of X_i, (but for random errors) it is felt that the estimator

$$\hat{Y}_{ratio} = \frac{\overset{*}{Y}}{\overset{*}{X}} X \qquad \ldots \qquad (3.3.1')$$

where $\overset{*}{Y}$ and $\overset{*}{X}$ are unbiased estimators of Y and X respectively. The estimator (3.3.1) is not, in general, unbiased and the problem is to find an unbiased estimator of Y, using the information on X also. Midzuno [25] and Sen [38] considered the problem from a slightly different angle thus. They did not insist that $\overset{*}{Y}$ and $\overset{*}{X}$ of (3.3.1) should by themselves be unbiased estimates of Y and X. They specified the design $D = D(S, P)$ partially by demanding that every $s \in S$ should consist of n distinct units only. They then enquired as to what should be the probability measure P such that the simple estimator

$$\overset{*}{Y}_{M.S} = \frac{\bar{y}}{\bar{x}} \cdot X \qquad \ldots \qquad (3.3.1)$$

where

$$\bar{x} = \frac{1}{n} \sum_{i \in s} X_i$$

and

$$\bar{y} = \frac{1}{n} \sum_{i \in s} Y_i$$

is unbiased for Y. It can be seen that a necessary and sufficient condition for this is that

$$Q_{s_o} = \sum_{s \sim s_o} P_s = \alpha \bar{x}_{s_o} \quad \text{for} \quad s_o \in S \qquad \cdots (3.3.2)$$

where $s \sim s_o$ means that s is a permutation of the elements of s_o and α is a constant. The above relation is independent of the choice of s_o out of all samples s for which $s \sim s_o$. It can be easily seen that α should equal $1/\binom{N-1}{n-1}X$. The conditions (3.3.2) give only a partial specification of the probability measure and hence of the design. Corresponding to any further specification consistent with (3.3.2) and which completely pin down the design D we have a drawing mechanism generating a design D which satisfies (3.3.2).

To make matters clear, let us consider the case of $N = 3$ and $n = 2$. We have 6 possible samples with corresponding probabilities thus:

Sample	Probability
(U_1, U_2)	P_1
(U_2, U_1)	P_2
(U_1, U_3)	P_3
(U_3, U_1)	P_4
(U_2, U_3)	P_5
and (U_3, U_2)	P_6

conditions (3.3.2) can be written as

$$P_1 + P_2 = (X_1 + X_2) / 2X$$
$$P_3 + P_4 = (X_1 + X_3) / 2X \qquad \cdots (3.3.3)$$
and $$P_5 + P_6 = (X_2 + X_3) / 2X$$

Let us now have

$$P_1 = P_2, \quad P_3 = P_4 \quad \text{and} \quad P_5 = P_6$$

so that

$$P_1 = P_2 = \frac{X_1 + X_2}{4X}$$

$$P_3 = P_4 = \frac{X_1 + X_3}{4X}$$

and

$$P_5 = P_6 = \frac{X_2 + X_3}{4X}$$

This completely defines D_1 consistent with (3.3.3). The corresponding drawing mechanism that generates D_1 can be described thus:

$$q_0(U_i) = \frac{X + X_i}{4X}$$

$$q_1(s(i)) = 1$$

$$q_3(2, s(i), U_j) = \begin{cases} \dfrac{X_i + X_j}{X_i + X_k} & \text{for } j \neq i \text{ and where } k \text{ is the positive integer } \leq 3 \text{ which is } \neq i \text{ and } \neq j. \\ 0 & \text{if } j = i \end{cases}$$

$$q_1(s(i,j)) = 0 \quad \text{for} \quad i \neq j \qquad \ldots (3.3.4)$$

It can be seen that not only the initial probabilities are to be calculated afresh (given by q_0) but also the conditional probabilities (given by q_3) depend on the units already chosen. For larger n this method gives a cumbersome drawing mechanism. However by putting the further specifications as

$$P_1 : P_2 = X_1 : X_2$$

$$P_3 : P_4 = X_1 : X_3 \qquad \ldots (3.3.5)$$

and

$$P_5 : P_6 = X_2 : X_3$$

we obtain a design D_2 consistent with (3.3.3) and for which the corresponding drawing mechanism is described thus:

$$\left.\begin{aligned} q_0 &= \frac{X_i}{X} \\ q_1\ (s(i)) &= 1 \\ q_3\ (2,\ s(i),\ U_j) &= \frac{1}{2} \quad \text{if} \quad j \neq i \end{aligned}\right\} \qquad \ldots (3.3.6)$$

In this drawing mechanism only the initial probabilities are unequal but they need not be calculated a fresh. The mechanism consists of selecting one unit with probability proportional to X_i's and then to select one more unit out of the remaining 2 units by simple random sampling. For the general case of N and n, we set

$$P_{s_1} : P_{s_2} : \ldots : P_{s_M} = X_{j1} : X_{j2} : \ldots : X_{jM} \qquad \ldots (3.3.7)$$

where $s_1, s_2, \ldots, s_M$ are the $M = n!$ samples which are permutations of the elements of a subset of size n, and $(\ )_{j_1} \ldots, (\ )_{j_M}$ are their respective first elements. It can be seen that this method of allocation gives the following simple drawing mechanism.

$$\left.\begin{aligned} q_0 &= \frac{X_i}{X} \\ q_1\,(s(i_1)) &= q_1\,(s(i_1 i_2)) = \ldots = q_1\,(s(i_1 i_2 \ldots i_{n-1})) = 1 \\ q_1\,(s(i_1, i_2, \ldots, i_n)) &= 0 \end{aligned}\right\} \qquad (3.3.8)$$

$$q_3\,(K+1, s_{(k)}, U_i) = \begin{cases} \dfrac{1}{N-K} & \text{if} \quad i \neq s_{(k)} \\ 0 & \text{otherwise} \end{cases}$$

Verbally this means that the mechanism consists of drawing one unit with probabilities proportional to X_i's and then to draw (n-1) units out of the remaining (N-1) units by simple random sampling without replacement. The merit of the allocation (3.3.7) thus lies in the simplicity of the corresponding drawing mechanism (3.3.8), which incidentally was first given by Midzuno [25] and Sen [37]

We thus see that the problem solved by Midzuno and Sen has many other solutions all of which can be obtained methodically, following the proof of theorem (3.1). This is an example of characterising the design D first (to have same optimum property) and then search for a suitable drawing mechanism to generate it. Another important illustration of such a method of going from the design (at times partially specified) to the generating drawing mechanism is fully furnished by the problem considered in chapter V.

OPTIMUM UTILISATION OF AUXILIARY INFORMATION - COMPLETE SOLUTION FOR $\mu = 2$

We shall now turn our attention to the problem posed in § 2.5 in this and the next two chapters. In this chapter we give a complete solution, satisfying all the conditions of (2.5.6), for the important practical case of sampling two units from a stratum i.e., $\mu = 2$.

In § 4.1 we present a solution under a mild restriction on X_i's. In § 4.2 we modify this method to hold good for general values of X_i's (subject to $0 < X_i \leq \frac{X}{2}$ for all i, these being inherent in (2.5.6)). In § 4.3 we make some remarks regarding our solution. Some illustrative examples are given. In § 4.4 we compare the method with that of Rao, Hartley and Cochran and prove that our strategy is superior to theirs for a wide class Δ_4 of prior distributions that include Δ_1. In § 4.5 we compare our strategy with that of an improvement [26] of Des Raj's [6] and give a direct proof of the Δ_1 superiority of our strategy. In § 4.6 we compare these three strategies in our illustrative examples, both with respect to the true variances and expected variances over Δ_1. In § 4.7 we present two simple solutions of our problem, which satisfy (i), (ii), (iii), (iv) and (vii) of (2.5.6).

§ 4.1. The case $X_{N-1} = X_N$

Let $P_k = \frac{X_k}{X}$ $\quad (1 \leq k \leq N)$

and assumed without loss of generality that

$$0 < P_1 \leq P_2 \leq \cdots \leq P_{N-1} \leq P_N < \frac{1}{2} \qquad (4.1.1)$$

When under the sole restriction that $P_{N-1} = P_N$, all the conditions of (2.5.6) are satisfied by the

Sampling Scheme (A) :- Step 1. Select two units from the population with probability P_k for the kth unit U_k, and with replacement. If the sample thus obtained consists of distinct units U_i and U_j, accept the sample. Otherwise, reject the sample and proceed to,

Step II : Select two units from $\mathcal{U}$ with probabilities proportional to

$$P_1^2, P_2^2, \ldots, P_N^2$$

If the sample thus obtained consists of distinct units, accept it, Otherwise, proceed to

Step III : Select two units from $\mathcal{U}$ with probabilities proportional to

$$P_1^4, P_2^4, \ldots, P_N^4$$

If the sample consists of two distinct units, accept it. Otherwise continue the process.

In general, if the 1st, 2nd, ..., (r-1)th steps result in rejections, at the rth step, the units are drawn with probabilities proportional to

$$P_1^{2^{r-1}}, P_2^{2^{r-1}}, \ldots, P_N^{2^{r-1}}$$

We shall first prove.

Lemma 1 : The sequential sampling scheme given above, terminates after a finite number of steps, with probability 1.

Proof : Let $S_r = \sum_{k=1}^{N} p_k^r$, $r = 1, 2, \ldots$ } ...(4.1.)

and $S_{(r)} = S_{2^r}$.

In particular, we have $S_{(0)} = S_1 = 1$

Let Q_r = Probability that the scheme terminates at or before the r^{th} step clearly Q_r increases with r. We need show that

$$Q_r \to 1 \quad \text{as} \quad r \to \infty.$$

We have

$$Q_1 = \sum_{i \neq j}^{N} \sum P_i P_j = 1 - \sum_{1}^{N} P_k^2 = 1 - S_2 = 1 - S_{(1)}.$$

$$Q_2 = Q_1 + (1 - Q_1) \cdot \sum_{i \neq j}^{N} \sum \frac{P_i^2}{S_2} \cdot \frac{P_j^2}{S_2}$$

$$= Q_1 + \frac{S_2}{S_2^2} (S_2^2 - S_4)$$

$$= (Q_1 + S_2) - \frac{S_2 S_4}{S_2^2}$$

$$= 1 - \frac{S_{(2)}}{S_{(1)}} .$$

The above derivation is clearly explained. The probability Q_2 of the scheme terminating at or before the 2nd step is the probability Q_1 of its terminating at the 1st step plus the probability of its not terminating at the 1st step, which is $(1-Q_1)$, multiplied by the conditional probability of its terminating at the 2nd step, which is $\sum_{i \neq j}^{N} \sum \frac{p_i^2}{S_1} \frac{p_j^2}{S_1}$. Similarly

$$Q_3 = Q_2 + (1 - Q_2) \cdot \sum_{i \neq j} \sum \frac{p_i^4 \, p_j^4}{S_4^2}$$

$$= \left(1 - \frac{S_4}{S_2}\right) + \frac{S_4}{S_2} \cdot \frac{1}{S_4^2} \left(S_4^2 - S_8\right)$$

$$= 1 - \frac{S_8}{S_2 \, S_4} = 1 - \frac{S_{(3)}}{S_{(1)} S_{(2)}}$$

and in general it can be easily shown by induction, that

$$Q_r = 1 - \frac{S_{(r)}}{S_{(1)} \, S_{(2)} \cdots S_{(r-1)}} \qquad \text{for } r \geq 2 \ldots \qquad (4.1.3)$$

To prove the lemma, we need only show that

$$L_r = \frac{S_{(r)}}{S_{(1)} \, S_{(2)} \cdots S_{(r-1)}} \to 0 \qquad \text{as } r \to \infty.$$

From (4.1.1), $S_{(i)} = \sum_{k}^{N} P_k^{2^i} > 2P_N^{2^i}$

$$S_{(1)}\, S_{(2)} \cdots S_{(r-1)} > 2^{r-1} \cdot P_N^{(2 + 2^2 + \cdots + 2^{r-1})}$$

$$= 2^{r-1} \cdot P_N^{2^r - 2}$$

Hence

$$L_r = \frac{S_{(r)}}{S_{(1)}\, S_{(2)} \cdots S_{(r-1)}} < \frac{P_N^2}{2^{r-1}} \times \frac{S_{(r)}}{P_N^{2^r}} = P_N^2 \sum_{k=1}^{N} \frac{P_k^{2^r}}{P_N^{2^r}} \cdot \frac{1}{2^{r-1}}$$

As $r \to \infty$, each of the N terms of the r.h.s. tends to zero so that $L_r \to 0$. Hence the lemma.

Note :- The assumption that $P_N = P_{N-1}$ is necessary for L_r to tend to zero, i.e. for Q_r to tend to 1. Otherwise, it can be shown that although Q_r strictly increases with r, it tends to a limit which is less than 1. For, if $P_N \neq P_{N-1}$, we still have

$$S_i > P_N^i ,$$

so that

$$\frac{P_k^{2^r}}{S_{(1)} S_{(2)} \cdots S_{(r-1)}} < \frac{P_k^{2^r}}{P_N^{2^r}} \cdot P_N^2$$

and the r.h.s. tends to zero as $r \to \infty$ for $1 \leq k \leq N-1$. For L_r to tend to 0 it is therefore necessary that

$$M_r = \frac{P_N^{2^r}}{S_{(1)} S_{(2)} \cdots S_{(r-1)}} \to 0 \quad \text{as} \quad r \to \infty$$

We can write

$$M_r = P_N^2 \cdot \frac{P_N^2}{S_{(1)}} \cdot \frac{P_N^4}{S_{(2)}} \cdots \frac{P_N^{2^{r-1}}}{S_{(r-1)}} = P_N^2 \cdot \prod_{j=1}^{r-1} (1 - \theta_j)$$

where

$$\theta_j = 1 - \frac{P_N^{2^j}}{S_{(j)}}$$

However, $\lim_{r \to \infty} M_r = P_N^2 \cdot \prod_{j=1}^{\infty} (1 - \theta_j)$, and the later infinite product diverges to zero if and only if the infinite series

$$\sum_{j=1}^{\infty} \theta_j$$

diverges to ∞. But,

$$\sum_{1}^{r} \theta_j = \sum_{j=1}^{r} \left(1 - \frac{P_N^{2^j}}{S_{(j)}}\right) < \sum_{k=1}^{2^j} \left(1 - \frac{P_N^k}{S_k}\right) < \sum_{k=1}^{m} \left(1 - \frac{P_N^k}{S_k}\right)$$

if $m > 2^r$. The last sum on the r.h.s. can be written as

$$\sum_{k=1}^{m} \frac{1}{1 + \left(\frac{P_N^k}{S_k - P_N^k}\right)} \leq \sum_{1}^{m} \frac{1}{\left[1 + \left(\frac{P_N}{P_{N-1}}\right)^k\right] \cdot \frac{1}{N-1}} = (N-1) \sum_{1}^{m} z_k$$

where

$$z_k = \frac{1}{(N-1) + \varrho^k}$$

and $\varrho = \frac{P_N}{P_{N-1}}$. Since $P_N \neq P_{N-1}$, $\varrho > 1$. However,

$$\frac{z_{k+1}}{z_k} = \frac{(N-1) + \varrho^k}{(N-1) + \varrho^{k+1}} \to \frac{1}{\varrho} < 1 \quad \text{as} \quad k \to \infty.$$

Hence the series $\sum_{1}^{\infty} z_k$ converges and hence also the series $\sum_{1}^{\infty} \delta_j$ so that M_r tends to a positive limit as $r \to \infty$.

Lemma 2 :- Let

$\pi_i^{(r)}$ = Probability of selecting the ith unit at or before the rth step.

and $\pi_{ij}^{(r)}$ = Probability of selecting the unordered sample (u_i, u_j) at or before the rth step.

Then

$$\pi_i^{(r)} = 2 P_i - \frac{2 P_i^{2^r}}{S_2 S_4, \ldots S_{2^{r-1}}} \qquad \ldots (4.1.4)$$

and $$\pi_{ij}^{(r)} = 2P_iP_j\left\{1 + \frac{P_iP_j}{S_2} + \frac{P_i^3P_j^3}{S_2S_4} + \cdots + \frac{(P_iP_j)^{2^{r-1}}}{S_2S_4\cdots S_{2^{r-1}}}\right\} \quad \ldots(4.1.5)$$

Proof:- We have

$$\pi_{ij}^{(r)} = 2P_iP_j + (1 - Q_1)\cdot\frac{2P_i^2P_j^2}{S_{(1)}^2} + (1 - Q_2)\frac{2P_i^4P_j^4}{S_{(2)}^2} + \ldots + (1-Q_{r-1})\frac{2P_i^{2^{r-1}}P^{2}}{S_{(r-1)}^2}$$

$$= 2P_iP_j + 2.S_{(1)}\frac{P_i^2P_j^2}{S_{(1)}^2} + 2\frac{S_{(2)}}{S_{(1)}}\frac{P_i^4P_j^4}{S_{(2)}^2} + \ldots + 2.\frac{S_{(r-1)}(P_iP_j)^{2^{r-1}}}{S_{(1)}S_{(2)}\cdots S_{(r-2)}S_{(r-1)}^2}$$

$$= 2P_iP_j\left\{1 + \frac{P_iP_j}{S_{(1)}} + \frac{P_i^3P_j^3}{S_{(1)}S_{(2)}} + \cdots + \frac{(P_iP_j)^{2^{r-1}}}{S_{(1)}S_{(2)}\cdots S_{(r-1)}}\right\},$$

which proves (4.1.5).

Further, since the sample size is two for all the samples,

$$\pi_i^{(r)} = \sum_{j\neq i}^{N}\pi_{ij}^{(r)}$$

$$= 2P_i(1 - P_i) + \frac{2P_i^2}{S_{(1)}}(S_{(1)} - P_i^2) + \frac{2P_i^4}{S_{(1)}S_{(2)}}(S_{(2)} - P_i^4)$$

$$+ \cdots + \frac{2P_i^{2^{r-1}}}{S_{(1)}S_{(2)}\cdots S_{(r-1)}}(S_{(r-1)} - P_i^{2^{r-1}})$$

$$= 2P_i - \frac{2P_i^{2^r}}{S_{(1)}S_{(2)}\cdots S_{(r-1)}}$$

which proves (4.1.4). Hence the lemma.

We shall now prove the main theorem (A) :- The sampling scheme (A) satisfies all the conditions of (2.5.6).

Proof :- Clearly (ii) of (2.5.6) is satisfied. From the assumption that $P_N = P_{N-1}$, it follows as in lemma 1,

$$\frac{P_i^{2^r}}{S_{(1)} S_{(2)} \cdots S_{(r-1)}} \to 0 \quad \text{as} \quad r \to \infty$$

so that

$$\pi_i = \lim_{r \to \infty} \pi_i^{(r)} = 2P_i = 2\frac{X_i}{X} \qquad \ldots(4.1.6)$$

which satisfies (i) of (2.5.6). Further, we have

$$\pi_{ij} = \lim_{r \to \infty} \pi_{ij}^{(r)} = 2P_iP_j \left\{ 1 + \frac{P_iP_j}{S_{(1)}} + \frac{P_i^3 P_j^3}{S_{(1)} S_{(2)}} + \ldots. \right\} \qquad \ldots(4.1.7)$$

$$= 2P_iP_j \left(1 + \sum_1^\infty w_k\right)$$

where

$$w_k = \frac{(P_i P_j)^{2^k - 1}}{S_{(1)} S_{(2)} \cdots S_{(k)}} < \frac{(P_N^2)^{2^k - 1}}{(2P_N^2)(2P_N^4)\ldots(2P_N^{2k})} = \frac{1}{2^k}$$

so that

$$0 < \pi_{ij} < 2P_iP_j \left(1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \ldots\right) = 4 P_i P_j = \pi_i \pi_j.$$

which satisfies (iii) and (iv) of (2.5.6).

Also $\pi_{ij} > 2P_iP_j = \frac{\pi_i \pi_j}{2}$

so that $\frac{\pi_{ij}}{\pi_i \pi_j} > \frac{1}{2}$,

which satisfies (v) of (2.5.6). (vi) and (vii) are also satisfied fairly well.

The theorem is proved.

The sampling scheme given above thus satisfies all the conditions given in (2.5.6). In fact π_{ij} is substantially greater than $\frac{\pi_i \pi_j}{2}$, as is evident from lemma 2, so that we have a non-negative unbiased estimator of the variance of our optimum estimator, which is itself very stable. In practice it is not necessary to compute more than the first few of the S's. If only step I results in a rejection, need we proceed to find P_k^2's and S_2 to make the selection. Unless the values $P_1, \ldots, P_N$ are all too equal, the scheme results in the selection of a sample in the first two or three steps. After the sample is obtained we need calculate S_i's for the evaluation of π_{ij} from the formula (4.1.6) term by term. The calculation can be stopped after a required degree of accuracy in the calculation of π_{ij} is attained. Unless the sample happens to contain the units with large values of P_i and P_j, this series given in (4.1.6) converges quite rapidly. At any rate it converges at least as fast as the geometric series $\sum \frac{1}{2^k}$.

§ 4.2 The case of general X_i's.

We shall now drop the restriction that $P_N = P_{N-1}$ and modify the method of § 4.1. The lower bound for $\pi_{ij}/\pi_i\pi_j$ will then be slightly less than $1/2$. Consider the

Sampling Scheme (B) :-

Let

$$P_1 \leq P_2 \leq \cdots \leq P_{N-1} \leq P_N < \frac{1}{2}$$

be the original size measures. Let

$$\partial = \frac{2(1 - P_N)(P_N - P_{N-1})}{(1 - P_N - P_{N-1})} \qquad \cdots (4.2.1)$$

Step I : Conduct a binomial trial with probability of success equal to ∂. If the trial results in a success proceed to

Step II : Select one of the units

$$U_1, U_2, \cdots, U_{N-1}$$

with probabilities proportional to the original size measures. If U_j be the unit thus selected, accept (U_N, U_j) as the (unordered) sample

Step III : If the binomial trial in step I results in a failure, proceed for the sampling scheme A, given in § 4.1, with P_N replaced by P_{N-1} normalising the new set of P_k's.

We shall now prove our general result.

Theorem (B) :- The sampling schemed (B) given above satisfies the conditions of (2.5.6).

Proof :- From lemma 1 of § 4.1, we see that with probability 1 the scheme (B) terminates after a finite number of steps. (ii) and (iii) of (2.5.6) are obviously satisfied, so that we need only prove (i), (iv) and (v).

Let ϕ_i and ϕ_{ij} be the conditional inclusion probabilities under the condition that the trial at step 1 resulted in a failure, and P_i^*'s $(1 \leq i \leq N)$ be the probabilities which we start with at step 3. Clearly

$$P_i^* = \frac{P_i}{1 - P_N + P_{N-1}} = P_i \frac{1 - \frac{\delta}{2(1-P_N)}}{(1 - \delta)} \quad \text{for } i \neq N$$

and (4.2.2)

$$P_N^* = P_{N-1}^* = \frac{(P_N - \frac{\delta}{2})}{(1 - \delta)} .$$

from (6.1.6) and (6.1.7) we then have

$$\phi_i = 2 P_i^*$$

and $$\phi_{ij} = 2 P_i^* P_j^* \left\{ 1 + \frac{P_i^* P_j^*}{S_2^*} + \frac{P_i^{*3} P_j^{*3}}{S_2^* S_4^*} + \cdots\cdots \right\} \quad \ldots (4.2.3)$$

where

$$S_r^* = \sum_{i=1}^{N} P_i^{*r} .$$

Let π_i and π_{ij} be the over-all inclusion probabilities resulting from the scheme. Clearly

$$\left.\begin{aligned} \pi_i &= \partial \cdot \frac{P_i}{1 - P_N} + (1 - \partial)\,\phi_i && \text{for } i \neq N \\ \pi_N &= \partial + (1 - \partial)\,\phi_N \\ \pi_{ij} &= (1 - \partial)\,\phi_{ij} && \text{for } 1 \leq i \neq \leq N-1 \\ \pi_{Nj} &= \partial \cdot \frac{P_j}{1 - P_N} + (1 - \partial)\,\phi_{Nj} \end{aligned}\right\} \quad \ldots(4.2.4)$$

Substituting from (4.2.1), (4.2.2) and (4.2.3), we have

$$\left.\begin{aligned} \pi_i &= P_i \partial \cdot \left(\frac{1}{1 - P_N} - \frac{2}{1 - P_N + P_{N-1}}\right) + \frac{2P_i}{1 - P_N + P_{N-1}} \\ &= 2P_i && \text{for } i \neq N \end{aligned}\right\} \quad \ldots(4.2.5)$$

and

$$\pi_N = \partial + (1 - \partial) \cdot \frac{(2\,P_N - \partial)}{1 - \partial} = 2\,P_N$$

which proves (i) of (2.5.6). To prove (iv), note that from (4.1.8) we have

$$\phi_{ij} < \phi_i\,\phi_j \qquad \text{for } 1 \leq i \neq j \leq N-1$$

Hence for $1 \leq i \neq j \leq N-1$,

$$\pi_{ij} = (1-\partial)\,\phi_{ij}$$

$$< (1-\partial)\,\phi_i\,\phi_j$$

$$= 4(1-\partial).\,P_i^*\,P_j^*$$

$$= 4P_iP_j\,\frac{\left\{1 - \frac{\partial}{2(1-P_N)}\right\}^2}{(1-\partial)}$$

To prove (iv) for this case we need only show that

$$\left\{1 - \frac{\partial}{2(1-P_N)}\right\}^2 < 1-\partial$$

i. e. $$\partial < 4P_N(1-P_N)$$

But $$\partial - 4P_N(1-P_N) = \frac{2(1-P_N)\,(1-2P_N)\,(P_N + P_{N-1})}{(1-P_N-P_{N-1})} > 0$$

since $P_N < \frac{1}{2}$, and hence

$$\pi_{ij} < \pi_i\,\pi_j \qquad \text{for } (1 \leq i \neq j \leq N-1) \qquad \ldots (4.2,6)$$

Also

$$\pi_{Nj} < \partial.\,\frac{P_j}{1-P_N} + (1-\partial)\,\phi_N\,\phi_j$$

$$= \frac{\partial P_j}{1-P_N} + 4P_j\left(1 - \frac{\partial}{2(1-P_N)}\right).\left(P_N - \frac{\partial}{2}\right) / 1-\partial$$

so that

$$\pi_{Nj} < \pi_N \pi_j = 4 P_N P_j$$

if

$$\frac{\partial}{1 - P_N} + \frac{4\left(1 - \frac{\partial}{2(1-P_N)}\right)\left(P_N - \frac{\partial}{2}\right)}{1 - \partial} - 4 P_N < 0$$

After simplification, the l.h.s. turns out to be equal to

$$- \partial(1 - 2P_N)^2$$

and hence follows that

$$\pi_{Nj} < \pi_N \pi_j \qquad \ldots(4.2.7)$$

(6.2.6) and (6.2.7) prove (iv) of (2.5.6).

To verify (v), we have

$$\pi_{ij} = (1 - \partial)\, \phi_{ij} \qquad (1 \leq i \neq j \leq N - 1)$$

$$> (1 - \partial)\, \frac{\phi_i \phi_j}{2} \qquad \text{from (4.1.9)}$$

$$= \frac{\pi_i \pi_j}{2} \cdot \frac{\left\{1 - \frac{\partial}{2(1 - P_N)}\right\}^2}{(1 - \partial)} \qquad \ldots(4.2.8)$$

and

$$\pi_{Nj} > \frac{\partial P_j}{1 - P_N} + (1 - \partial)\, \frac{\phi_N \phi_j}{2}$$

$$= \partial \cdot \frac{P_j}{1 - P_N} + \frac{(2P_N - a_N)\left(2P_j - \frac{\partial P_j}{1 - P_N}\right)}{2(1 - \partial)}$$

$$= \frac{\partial P_j}{1 - P_N} + \frac{\pi_N \pi_j}{2(1-\partial)} \Big\} 1 - \frac{\partial}{2(1-P_N)} \Big\{ 1 - \frac{\partial}{2P_N} \Big\} \qquad (4.2.9)$$

Since P_N and P_{N-1} usually do not differ much the above bound in general is not small. The examples given after § 4.3 will illustrate this point.

We may note here that for both the schemes A and B

$$\frac{\pi_{ij}}{\pi_i \pi_j}$$

is a non-decreasing function of j for any given This is easy to see from (4.1.7) and (4.2.4). We thus have

$$\frac{\pi_{1,2}}{\pi_1 \pi_2} \leq \frac{\pi_{ij}}{\pi_i \pi_j} \leq \frac{\pi_{N-1,N}}{\pi_{N-1} \pi_N} \leq 1 \qquad \cdots \qquad (4.2.10)$$

and when $P_N = P_{N-1}$ we have

$$\frac{1}{2} \leq \frac{\pi_{1,2}}{\pi_1 \pi_2} \leq \frac{\pi_{ij}}{\pi_i \pi_j} \leq \frac{\pi_{N-1,N}}{\pi_{N-1} \pi_N} \leq 1 \qquad \cdots \qquad (4.2.11)$$

§ 4.3. Some remarks on the schemes A and B.

We have seen that the sampling schemes given in § 4.1 and § 4.2 terminate after a finite number of steps, with probability 1. It is clear that the scheme A terminates fastest when $P_i = \frac{1}{N}$ for all i and that it terminates slowest when $P_{N-1} = P_N = \frac{1}{2}$ and $P_i = 0$ for $1 \leq i \leq N-2$. Though in our problem $P_i > 0$ for all i, since values as near to zero as we please are permissible for P_i's, the upper bound for the expected number $E_A(n)$ of steps required to select the sample will be that obtained by setting $P_{N-1} = P_N = \frac{1}{2}$ and $P_i = 0$ for $1 \leq i \leq N-2$, and this bound is attained only in the limit. These two cases thus give the lower and upper bounds for the expected number of steps required for the selection of a sample by the sampling scheme A. If

$$P_i = \frac{1}{N} \qquad \text{for } 1 \leq i \leq N$$

the probability λ_t for the sampling scheme A to terminate at the r^{th} step is given by

$$\lambda_t = N^{-(t-1)} - N^{-t} \qquad t \geq 1$$

Hence

$$E_A(n) = \sum_{t=1}^{\infty} t\lambda_t = \sum_{t=0}^{\infty} N^{-t} = \frac{N}{N-1}.$$

Further, if

$$P_{N-1} = P_N = \frac{1}{2}$$

and

$$P_i = 0 \qquad \text{for } 1 \leq i \leq N-2$$

the probability μ_t for the scheme A to terminate at the rth step is given by

$$\mu_t = \lambda^{-t} \qquad t \geq 1$$

so that in this case

$$E_A(n) = \sum_{t=1}^{\infty} t\,\mu_t = 2$$

Thus we have

$$\frac{N}{N-1} \leq E_A(n) < 2 \quad \ldots \qquad (4.3.1)$$

The scheme B terminates in one step if the binomial trial in step I results in a success, for which the probability θ is given by (4.2.1). If the trial results in a failure then we have a scheme of the type A. Denoting by $E_B(n)$ the expected number of steps required for the selection of the sample according to the scheme B, we thus have

$$E_B(n) = \theta + (1-\theta)\,E_A(n)$$

where A is the corresponding scheme in step III of § 4.2. From (4.3.1) it follows that

$$1 \leq \frac{N-\theta}{N-1} \leq E_B(n) < 2-\theta \leq 2 \qquad (4.3.2)$$

we shall now illustrate our sampling schemes by means of some illustrative examples.

We consider 5 populations one of which viz., the 'Yates population' (c.f., page 256 of [40]) is purely fictitious and the other four are taken from live data. These examples will be used repeatedly in this thesis for purposes of illustration and emperical comparisons. Though the estimated variances and expected variances as given by a selected sample can be easily computed using only a desk calculator, since we compared the actual variances in all the cases, the computations are organised in the electronic machine IBM 1401. The Yates population has been mainly used to check up these programmings by direct manual calculations.

The data for the four live examples considered here consist of the district-wise population figures for four of the Indian States of Andhra Pradesh, Punjab, West Bengal and Gujerat coming respectively from the Southern, Northern, Eastern and Western parts of India. The population as per the 1961 census is taken to be the Y-variable and that as per the 1951 census, measured in thousands, is taken to be the X-variable. In practice too the auxiliary information should not (or rather need not) be taken as acurately as the main variable and can be rounded off to a reasonable extent We remember that only the proportionate values of X_i's are of interest to us.

The object is to estimate the total population in 1961 for each of the states, by taking a sample of two districts from each state. We shall later compare the true variances as well the expected variances under the model (2.3.3), of various estimation methods. In practice one resorts to a sub-sampling in each selected district and in this case our results concern the component of sampling variance due to first stage sampling only and the other components can be dealt with simillarly.

1. Yates Population

serial no. of the unit (i)	X_i	Y_i
1	1	5
2	2	12
3	3	21
4	4	32

2. Andhra Pradesh

serial no. of district (i)	name of the district	X_i	Y_i
1	Srikakulam	2123	2342291
2	Visakhapatnam	2072	2288976
3	East Godavari	2301	2609311
4	West Godavari	1697	1978434
5	Krishna	1736	2076103
6	Guntur	2560	3009997
7	Nellore	1794	2033963
8	Chittoor	1666	1913169
9	Cuddapah	1628	1342140
10	Anantapur	1483	1764223
11	Kurnool	1617	1909644
12	Mahbubnagar	1447	1590689
13	Hyderabad	1821	2063601
14	Medak	1109	1226465
15	Nizamabad	835	1021503
16	Adilabad	831	1009301
17	Karimnagar	1428	1620417
18	Warangal	1329	1545750
19	Khammam	808	1057225
20	Nalgonda	1287	1574797

3. Punjab

serial no. of district (i)	name of the district	X_1	Y_1
1	Hissar	1045	1545887
2	Rohtak	1122	1416915
3	Gurgaon	967	1238128
4	Karnal	1007	1489679
5	Ambala	1017	1372193
6	Simla	106	111256
7	Kangra	921	1057066
8	Lahaul and Spiti	12	20478
9	Hoshiarpur	1094	1229473
10	Jullundur	1055	1224434
11	Ludhiana	807	1021190
12	Ferozepur	1275	1620389
13	Amritsar	1367	1547241
14	Gurdaspur	851	984152
15	Kapurtyala	295	343775
16	Bhatinda	786	1056033
17	Sangrur	1111	1425261
18	Patiala	777	1047437
19	Mohindergarh	443	547164

4. West Bengal

serial no. of district (i)	name of the district	X_i	Y_i
1	Darjeeling	459	624879
2	Jalpaiguri	914	1360110
3	Cooch Behar	671	1019747
4	West Dinajpur	979	1330346
5	Malda	937	1220491
6	Murshidabad	1715	2293074
7	Nadia	1144	1715068
8	24 Parganas	4459	6293758
9	Calcutta	2698	2926498
10	Howrah	1611	2043225
11	Hooghly	1604	2233798
12	Burdwan	2191	3083564
13	Birbhum	1066	1447638
14	Bankura	1319	1667527
15	Midnapur	3359	4349069
16	Purulia	1169	1358842

5. Gujarat

serial no. of district (i)	name of the district	X_i	Y_i
1	Jamnagar	616	824725
2	Rajkot	930	1209004
3	Surendranagar	506	662308
4	Bhavnagar	885	1116206
5	Amreli	538	667387
6	Junagadh	987	1244086
7	Kutch	567	695704
8	Banaskantha	773	905980
9	Sabarkatna	684	917809
10	Mehsana	1393	1685090
11	Ahmedabad	1675	2231534
12	Kaira	1612	1874351
13	Panchmahals	1131	1467485
14	Baroda	1211	1527044
15	Broach	717	892241
16	Surat	1982	2438740
17	Dangs	47	71589

The values of (i) ∂, the probability with which the binomial trial of Step I of Scheme (B) is to be conducted, given by (4.2.1); (ii) the first few L_r's where $L_r = 1 - Q_r$ is the probability that sampling in Step III of the Scheme (B) continues beyond r steps (i.e. the first r pairs obtained getting rejected), where Q_r is given by (4.1.3) and (iii) $\frac{\pi_{12}}{\pi_1 \pi_2}$ and $\frac{\pi_{N-1,N}}{\pi_{N-1} \pi_N}$ (after arranging the units in the ascending order of π's) giving the lower and upper bounds of $\frac{\pi_{ij}}{\pi_i \pi_j}$'s (c.f. (4.2.10) are given for all the five populations considered, in the following table. Though all these values are found to 16 significant figures, here we give them to 5 places of decimals only.

Table 4.3

Values of ∂, L_r's, $\frac{\pi_{12}}{\pi_1 \pi_2}$ and $\frac{\pi_{N-1,N}}{\pi_{N-1} \pi_N}$

Population	∂	L_1	L_2	L_3	L_4	L_5	$\frac{\pi_{12}}{\pi_1 \pi_2} = \min \frac{\pi_{ij}}{\pi_i \pi_j}$	$\frac{\pi_{N-1,N}}{\pi_{N-1} \pi_N} = \max \frac{\pi_{ij}}{\pi_i \pi_j}$
Yates'	.40000	.28395	.00608	.04012	.01931	.00964	.40330	.89452
Andhra Pradesh	.01782	.05394	.00342	.00032	.00006	.00002	.50560	.56356
Punjab	.01255	.06226	.00437	.00039	.00006	.00002	.49947	.56378
W.Bengal	.09888	.08135	.01108	.00302	.00125	.00060	.49369	.67652
Gujarat	.04280	.07214	.00720	.00123	.00033	.00011	.49785	.60693

It is seen from the above table that in all the live examples considered the probability is less than 1 in thousand that Step III of

the sampling scheme B need we have to go beyond the 4th step i.e. reject the sample more than 4 times. The scheme given thus results in the selection of the sample quite rapidly (of course it converges at least as fast as the the series $\Sigma 2^{-r}$, but we see that in practice the convergence is much faster).

The variability of $\frac{\pi_{ij}}{\pi_i \pi_j}$'s is also very small, as can be seen from the last two columns of the above table, considering especially that they represent the bounds for about 200 values in each case. The lower bound is greater than $\frac{1}{2}$ in one case and in the other cases it is not much different from $\frac{1}{2}$. The high stability of the values of $\frac{\pi_{ij}}{\pi_i \pi_j}$'s is helpful to keep the strategy robust from slight deviations from the model assumed.

For the selection of the sample, as remarked earlier, we need go to the computation of squares, fourth powers etc., of P_i's only as and when the need arises. After the sample is drawn we need calculate just the relevent π_{ij} only, by (4.2.3) and (4.2.4). While using (4.2.3) we need calculate the first few terms only as the convergence of the series is quite rapid, being at least as fast as the series $\Sigma 2^{-r}$. In the examples considered here it is found that in none of the cases need we find more than the first five terms to compute π_{ij} to within an error of .01 per cent.

§ 4.4 Comparison with the method of Rao, Hartely and Cochran.

In this section we shall compare the sampling strategy of using $\overset{*}{Y}_{HT}$ in the design generated by the sampling procedure given in § 4.2 with another recent one due Rao, Hartely and Cochran [35] (which we shall briefly refer to as the RHC strategy). This method is briefly described thus, for the case of any integer μ.

a) Split the population at random into μ groups of sizes $N_1, N_2, \cdots, N_\mu$ such that $\sum_{1}^{\mu} N_i = N$.

b) Draw a sample of μ distinct units by selecting one unitfrom each of the μ groups, with probability proportional to X_i's.

c) Use the estimator

$$\overset{*}{Y}_{RHC} = \sum_{i \in s} \frac{Y_i}{X_i / \phi_i}$$

where

$$\phi_i = \sum_{j \supset i} X_j$$

the summation being over all the units U_j that fall in the same group as i.

It follows from theorem (2.3.1) that the RHC strategy is sub-optimum in Δ_1-sense. We shall show that it is sub-optimum in Δ_4 sense where Δ_4 is a class of prior distributions ∂_4, which is wider than Δ_1 but is somewhat narrower than Δ_3. This Δ_4 is defined later. We give exact expressions for the efficiencies of the RHC strategy relative to our strategy, in the Δ_1-sense and Δ_3 sense. In this connection we use some results which are proved later in § 5.2.

For the RHC strategy the optimum values of the N_i's are given by

$$N_1 = N_2 = \cdots = N_\mu = \frac{N}{\mu}$$

if μ is a divisor of N, and

$$N_1 = N_2 = \cdots = N_K = R + K$$

and $$N_{K+1} = N_{K+2} = \cdots = N_\mu = R$$

if $$N = R\mu + K \qquad 0 < K < \mu .$$

The variance of (4.4.1) for the above given optimum values of N_i's is

$$V(\overset{*}{Y}_{RHC}) = \left(1 - \frac{\mu-1}{N-1}\right) \left\{ \frac{X}{\mu} \sum_1^N \frac{Y_i^2}{X_i} - \frac{Y^2}{\mu} \right\}$$

if N is a multiple of μ, and

$$= \left\{ 1 - \frac{\mu-1}{N-1} + \frac{K(\mu-K)}{N(N-1)} \right\} \left\{ \frac{X}{\mu} \sum_1^N \frac{Y_i^2}{X_i} - \frac{Y^2}{\mu} \right\}$$

if $N = R\mu + K \qquad 0 < K < \mu$ ... (4.4.2)

It can be easily seen that if μ divides N

$$E_{\partial_3} V(\overset{*}{Y}_{RHC}) = \left(1 - \frac{\mu-1}{N-1}\right) \left\{ \frac{X}{\mu} \sum_1^N \frac{(a^2 X_i^2 + \sigma_i^2)}{X_i} - \frac{(a^2 X^2 + \sum_1^N \sigma_i^2)}{\mu} \right\}$$

$$= \left(1 - \frac{\mu-1}{N-1}\right) \left\{ \frac{X}{\mu} \sum_1^N \frac{\sigma_i^2}{X_i} - \frac{\sum_1^N \sigma_i^2}{\mu} \right\} \qquad (4.4.3)$$

and

$$E_{\partial_1} (\overset{*}{Y}_{RHC}) = \left(1 - \frac{\mu-1}{N-1}\right) \left\{ \frac{X^2}{\mu} - \frac{\sum X_i^2}{\mu} \right\} \sigma^2 \qquad (4.4.4)$$

When $N = R\mu + K$ the factor $\left(1 - \frac{\mu-1}{N-1}\right)$ on the r.h.s's of (4.4.3) and (4.4.4) has to be replaced by $\left\{ 1 - \frac{\mu-1}{N-1} + \frac{k(\mu-k)}{N(N-1)} \right\}$.

It is shown later in (5.2.8) and (5.2.11) that for our Δ_1-optimum strategy consisting of a design $D \in D^{(\mu,\pi)}$ as defined by (2.3.12) and

the estimator $\overset{*}{Y}_{HT}$,

$$E_{\partial_3} V(D \in D^{(\mu,\pi)}, \overset{*}{Y}_{HT}) = \sum_1^N \sigma_i^2 \left(\frac{X}{\mu X_i} - 1 \right) \qquad \ldots (4.4.5)$$

and

$$E_{\partial_1} V(D \in D^{(\mu,\pi)}, \overset{*}{Y}_{HT}) = \sigma^2 \left(\frac{X^2}{\mu} - \sum_1^N X_i^2 \right) \qquad \ldots (4.4.6)$$

Let now Δ_4 be the class of prior distributions ∂_4 satisfying

$$\left.\begin{aligned} &\text{i)}\ E_{\partial_4}(Y_i \mid X_i) = a X_i \\ &\text{ii)}\ V_{\partial_4}(Y_i \mid X_i) = \sigma^2 X_i^g, \quad g \geq 1 \\ &\text{and}\quad \text{iii)}\ \mathrm{Cov}_{\partial_4}(Y_i, Y_j \mid X_i, X_j) = 0 \end{aligned}\right\} \qquad \ldots (4.4.7)$$

Thus Δ_4 is a wider class than Δ_1 but is somewhat narrower than Δ_3. It may be mentioned however that in most situations met with in practice the parameter 'g' is found to lie between 1 and 2. We now prove

Lemma (4.4.1) :- The strategy $H_0 = H_0(D \in D^{(\mu,\pi)}, \overset{*}{Y}_{HT})$ is superior to the strategy $H_1 = H(RHC)$ in the Δ_4-sense.

Proof :- From (4.4.3) and (4.4.5) we have

$$E_{\partial_3}(H_1) - E_{\partial_3}(H_0) = \frac{1}{\mu}\left(\sum \sigma_i^2\right)(\mu - 1 + \frac{\mu-1}{N-1}) - \frac{X}{\mu}\,\frac{\mu-1}{N-1}\sum \frac{\sigma_i^2}{X_i}$$

$$= \frac{(\mu-1)}{\mu(N-1)}\left\{ N \sum \sigma_i^2 - X \sum \frac{\sigma_i^2}{X_i} \right\}$$

Setting

$$\sigma_i^2 = \sigma^2 X_i^g \qquad \ldots (4.4.8)$$

in the above we have

$$E_{\partial}(H_1) - E_{\partial}(H_0) = \frac{(\mu-1)\sigma^2}{\mu(N-1)}\left\{ N \sum X_i^g - X \sum X_i^{g-1} \right\}$$

where ∂ is the apriori distribution obtained by replacing (ii) of

(9.3.8) by (4.4.8). Since all the X_i's are positive from an elementary inequality we know that

$$\Sigma X_i^{g-1} \Sigma X_i \gtrless N \Sigma X_i^g$$

according as $g-1$ and 1 are of the same or opposite signs. Hence

$$E_\partial (H_1) \gtrless E_\partial (H_0)$$

according as $g > 1$ or < 1, there being equality when $g=1$. It follows that H_0 is superior to H_1 in the Δ_4-sense and that it is inferior to H_1 in the Δ_5-sense where Δ_5 is the class of prior distributions ∂_5 with the inequality for g reversed in (ii) of (4.4.7).

Incidentally the above lemma also gives an alternate proof that H_0 is superior to H_1 in the Δ_1-sense since Δ_1 is narrower than Δ_4. This result, however, is also a consequence of theorem (9.3.1). The relative efficiencies of H_1 with respect to H_0 can be directly obtained from (4.4.3), (4.4.4), (4.4.5) and (4.4.6) and need not be given separately.

§ 4.5. Comparison with unordered DesRaj's estimator.

In this section we shall compare, in the Δ_1-sense, our sampling strategy with that of Murthy's [26] improved estimator of Des Raj [6], for the case $\mu = 2$. This latter is briefly described thus:

The sampling design $D_{d,2}$ is obtained by selecting μ units with probabilities proportional to $P_i = \frac{X_i}{X}$ for U_i and without replacement.

An unbiased estimator of Y for the above sampling design, as given by Des Raj [6] is

$$\overset{*}{Y}_{d,\text{usym}} = \frac{T}{\mu} \sum_{k=1}^{\mu} t_k$$

where

$$t_k = y_{i_1} + y_{i_2} + \cdots + y_{i_{k-1}} + \frac{y_{i_k}}{P_{i_k}} (1 - P_{i_1} - P_{i_2} - \cdots - P_{i_k}) \qquad (4.5.1)$$

Murthy [26] improved the above estimator using the techniques discussed in § 2.2 by taking the conditional expectation of (4.5.1) given the distinct units in the sample. For the case $\mu = 2$ this turns out to be

$$\overset{*}{Y}_{d,\text{sym}} = \left\{ \frac{y_{i_1}}{P_{i_1}} (1-P_{i_2}) + \frac{y_{i_2}}{P_{i_2}} (1-P_{i_1}) \right\} \cdot \frac{1}{2-P_{i_1}-P_{i_2}} \qquad (4.5.2)$$

which is symmetric in i_1 and i_2, as it should be. We shall now prove

Theorem (4.5.1):- The strategy $(D \in \mathcal{D}^{(2,\pi)}, \overset{*}{Y}_{HT})$ is uniformly superior to $(D_{d,2}, \overset{*}{Y}_{d,\text{sym}})$, in the Δ_1- sense.

Proof: For the variance of (4.5.2) we have

$$V(\overset{*}{Y}_{d,sym}) = \frac{1}{2}\sum\sum_{i\neq j} P_iP_j\left(\frac{1-P_i-P_j}{2-P_i-P_j}\right)\cdot\left(\frac{Y_i}{P_i}-\frac{Y_j}{P_j}\right)^2$$

$$= \sum_{i=1}^{N}\frac{Y_i^2}{P_i}\left\{\sum_{j\neq i}P_j\left(\frac{1-P_i-P_j}{2-P_i-P_j}\right)\right\} - \sum\sum_{i\neq j} Y_iY_j\left(\frac{1-P_i-P_j}{2-P_i-P_j}\right)$$

$$= \sum_{i=1}^{N}\frac{Y_i^2}{X_i}\left\{\sum_{j\neq i}X_j\left(\frac{X-X_i-X_j}{2X-X_i-X_j}\right)\right\} - \sum\sum_{i\neq j} Y_iY_j\left(\frac{X-X_i-X_j}{2X-X_i-X_j}\right) \qquad (4.5.3)$$

so that taking the expectation of (4.5.3) over ∂_1 given by (2.3.3)

$$E_{\partial_1}V(Y_{d,sym}) = \Sigma\left(\frac{a^2X_i^2+\sigma^2X_i^2}{X_i}\right)\sum_{j\neq i}X_j\left(\frac{X-X_i-X_j}{2X-X_i-X_j}\right) - \sum\sum_{i\neq j}a^2X_iX_j\left(\frac{X-X_i-X_j}{2X-X_i-X_j}\right)$$

$$= \sigma^2\sum\sum_{i\neq j}X_iX_j\left(\frac{X-X_i-X_j}{2X-X_i-X_j}\right) \qquad (4.5.4)$$

Substituting from (5.2.11) with $\mu = 2$, we then have

$$G = E_{\partial_1}V(\overset{*}{Y}_{d,sym}) - E_{\partial_1}V(D\epsilon D^{(2,\pi)}, \overset{*}{Y}_{HT})$$

$$= \sigma^2\left[\sum\sum_{i\neq j}X_iX_j\left(\frac{X-X_i-X_j}{2X-X_i-X_j}\right) - \left(\frac{X^2}{2}-\Sigma X_i^2\right)\right]$$

$$= \sigma^2\left[X^2 - X\sum\sum_{i\neq j}\frac{X_iX_j}{2X-X_i-X_j} - \frac{X^2}{2}\right]$$

Thus

$$G = \frac{X^2\sigma^2}{2}\left[1 - 2\sum\sum_{i\neq j}\frac{P_iP_j}{2-P_i-P_j}\right] \qquad (4.5.5)$$

To prove that $H(D\epsilon D^{(2,\pi)}, \overset{*}{Y}_{HT})$ is superior to $H(D_{d,2}, \overset{*}{Y}_{d,sym})$ in the Δ_1-sense we need only prove that the r.h.s. of (4.5.6) is ≥ 0. From the inequality between the arithmetic mean and harmonic mean we have

for $0 < P_i, P_j < 1$,

$$\left\{\frac{(1-P_i) + (1-P_j)}{2}\right\}^{-1} \leq \frac{\frac{1}{1-P_i} + \frac{1}{1-P_j}}{2}$$

i.e.

$$\frac{1}{2 - P_i - P_j} \leq \frac{1}{4}\left(\frac{1}{1 - P_i} + \frac{1}{1 - P_j}\right).$$

Substituting in r.h.s. of (4.5.6) we have

$$\begin{aligned} G &\geq \frac{X^2\sigma^2}{2}\left[1 - \frac{2}{4}\sum\sum_{i\neq j} P_iP_j\left(\frac{1}{1-P_i} + \frac{1}{1-P_j}\right)\right] \\ &\geq \frac{X^2\sigma^2}{2}\left[1 - \frac{1}{2}\left\{\sum_i P_i\left(\sum_{j\neq i}\frac{P_j}{1-P_i}\right) + \sum_j P_j\left(\sum_{i\neq j}\frac{P_i}{1-P_j}\right)\right\}\right] \\ &\geq \frac{X^2\sigma^2}{2}\left[1 - \frac{1}{2}(1+1)\right] \\ &\geq 0 \end{aligned}$$

which proves the required result. (4.5.6) gives an exact expression to the increase in the expected variance in using $H(D_{d,2}, \overset{*}{Y}_{d,sym})$ instead of $H(D_{\pi}^{(2,\pi)}, \overset{*}{Y}_{HT})$. The ratio of (4.5.6) to (5.2.11) gives the relative increase in the expected variance.

§ 4.6. Some emperical comparisons

We now compare our strategy of using sampling scheme B of § 4.2 and using the Horvitz and Thompson estimator, with those of § 4.4 and § 4.5 due to Rao, Hartely and Cochran and Des Raj, for the five populations considered in § 4.3.

We denote the variances by suffixes H, RHC and D respectively for the above three cases. We shall compare not only the expected variances over ∂_1 i.e. $E_{\partial_1} V(\overset{*}{Y})$'s but also the actual variances $V(\overset{*}{Y})$'s in each case. The computations are performed on the electronic machine IBM 1401 because the computation of $V(\overset{*}{Y}_H)$, $V(\overset{*}{Y}_D)$ and $E_{\partial_1} V(\overset{*}{Y}_D)$ are rather laborious, though the other quantities can easily be found with a desk-calculator. (Note, however, that in practice we cannot find these variances but only their estimates from the sample and these are not difficult to compute on a desk-calculator). The results are given in the following table, along with the true value of Y for comparison with the corresponding standard errors.

Table giving emperical comparisons of $\overset{*}{Y}_N$, $\overset{*}{Y}_{RHC}$ and $\overset{*}{Y}_D$

Population	Y	$V(\overset{*}{Y}_H)$	$V(\overset{*}{Y}_{RHC})$	$V(\overset{*}{Y}_D)$	$E_{\partial_1} V(\overset{*}{Y}_H)/\sigma^2$	$E_{\partial_1}(\overset{*}{Y}_{RHC})/\sigma^2$	$E_{\partial_1}(\overset{*}{Y}_D)/\sigma^2$	e_{RHC} (3) ÷ (4)	$\bar{e}_{RHC}$ (6) ÷ (7)	$\bar{e}\ e_D$ (3) ÷ (5)	$\bar{e}_D$ (6) ÷ (8)
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Yates	70	20.8811	33.3333	31.2356	20.0000	23.3333	20.6736	.3264	.8571	.6685	.9674
Anghra Pradesh	35977999	10^8. 34567	10^8.34534	10^8.34483	10^4.44425	10^4.44652	10^4.44431	1.0010	.9949	1.0024	.9998
Punjab	20293151	10^8.15194	10^8.15449	10^8.15449	10^4.15219	10^4.11282	10^4.11454	.9835	.9850	.9984	1.0000
West Bengal	34967634	10^8.39022	10^8.42072	10^8.40805	10^4.29547	10^4.29455	10^4.28602	.9275	.9692	.9563	.9981
Gujarat	20621283	10^8.2327	10^8. 2475	10^8. 2423	10^4.11263	10^4.11516	10^4.11270	.9402	.9780	.9604	.9994

Columns (9) and (11) give the true relative efficiences of the methods of Rao, Hartley and Cochran and DesRaj and columns (10) and (12) give the corresponding expected relative efficiencies under the model (2.3.3). (The entries in columns (10) and (12) are utmost unity, as they should be according to our theoretical results).

It is seen from columns (9) and (11) that in all but one of the five examples considered, our method is superior to those of Rao, Hartely and Cochran and Des Raj. Further, the close agreements between the corresponding entries of columns (9) and (10) and of (11) and (12), in all the four live examples considered, indicates the appropriateness of the model (2.3.3) in these cases. The Yates population is in fact constructed with a significant departure from this model. However the departure from the model further increased the relative superiority of our method which is perhaps an indication of its robustness.

The relative efficiencies of the Rao, Hartely and Cochran strategy and Des Raj strategy, though are generally below unity, are not much below unity in all the four live examples considered. This is due to the fact that in these examples the variation in the auxiliary information is not much as in practice districts are formed so as to have nearly equal population in each district. Districts that become too large will be bifurcated and those that are too small will be amalgamated for the sake of administrative convenience. Where the variation in the auxiliary information is considerable, the gain in efficiency due to our method can be expected to be considerable. This is indicated by the entries in columns (10) and (12) corresponding to the Yates population, where the expected relative efficiencies of the Rao, Hartely and Cochran strategy and Des Raj's strategy are considerably smaller than unity.

§ 4.7. Two more 'π PS' sampling schemes

In this section we give two more 'π PS' sampling schemes to achieve the main conditions viz., (i) and (ii) of (2.5.6). They satisfy (iii), (vi) and (vii) also. Though the schemes given in § 4.1 and § 4.2 are fully optimum and and hence should be preferred to the schemes of this section, these latter are being given mainly to point out that if just (i) and (ii) of (2.5.6) need be satisfied the problem admits a number of very simple solutions. (We may recall that the solutions of the earlier authors to this problem involve laborious iterations and linear programmings that are not suitable even for moderately large N). We shall briefly give the solutions, omitting all details.

Without loss of generality, suppose that

$$X_1 \leq X_2 \leq \cdots \leq X_N \qquad (4.7.1)$$

and

$$P_i = \frac{X_i}{\sum_{i}^{N} X_j} = \frac{X_i}{X}$$

We shall further assume as in § 4.1, that

$$P_{N-1} = P_N \qquad \cdots \qquad (4.7.2)$$

and give sampling schemes to satisfy (i) and (ii) of (2.5.6) under the condition (4.7.2). The case of general values of P_i's can then be dealt with exactly as it is done in Steps I and II of § 4.2.

Sampling Scheme (c) :-

Let $\partial_1, \partial_2, \cdots, \partial_N$ be the successive differences of the P_i's i.e.

$$\left.\begin{aligned} \partial_i &= P_1 && \text{if } i = 1 \\ &= (P_i - P_{i-1}) && \text{if } i \geq 2 \end{aligned}\right\} \qquad (4.7.2)$$

Let

$$\phi_i = \frac{\partial_1}{N-1} + \frac{\partial_2}{N-2} + \cdots + \frac{\partial_i}{N-i}, \qquad 1 \leq i \leq (N-1). \qquad (4.7.3)$$

The sampling scheme is as follows. Choose one unit from the first $(N-1)$ units with probability ϕ_i for the i^{th} unit. If U_λ be the unit thus chosen, choose one out of the units

$$U_{\lambda+1}, U_{\lambda+2}, \cdots, U_N$$

by simple random sampling. If $U_{\lambda'}$ be the unit thus selected, accept $(U_\lambda, U_{\lambda'})$ to be the unordered sample.

This scheme satisfies (i), (ii), (iii), (vi) and (vii) of (2.5.6). It can be shown by easy counter examples that (iv) is not always satisfied.

We omit the proofs of our assertions.

Sampling Scheme (D) :-

Let constants $\beta_1, \beta_2, \beta_{N-1}$ be defined as follows:

$$\beta_1 = P_1$$

$$\beta_2 = P_2 \left(1 - \frac{\beta_1}{1-P_1}\right)$$

$$\beta_3 = P_3 \left(1 - \frac{\beta_2}{1-P_1-P_2} - \frac{\beta_1}{1-P_1}\right)$$

$$\beta_{N-1} = P_{N-1} \left(1 - \frac{\beta_{N-2}}{1-P_1-P_2-\cdots-P_{N-2}} - \cdots - \frac{\beta_2}{1-P_1-P_2} - \frac{\beta_1}{1-P_1}\right)$$

(4.6.4)

The sampling scheme in this case is as follows:

Choose one unit from the first (N-1) units with probability β_i for the i^{th} unit. If U_λ be the unit thus chosen, choose one out of

$$U_{\lambda+1}, U_{\lambda+2}, \cdots, U_N$$

with probabilities proportional to the original probabilities P_i's. If $U_{\lambda'}$ be the unit thus chosen, accept $(U_\lambda, U_{\lambda'})$ as the unordered sample.

This scheme too satisfies (i), (ii), (iii), (iv) and (vi) of (2.5.6).*

We omit the proofs of our assertions.

---

* This scheme is also independently found by the author's colleague Mr. K.Vijayan who proves that the scheme satisfies (iv) also thus giving a non-negative variance estimator. Nothing is known about (v).

CHAPTER V

OPTIMUM UTILISATION OF AUXILIARY INFORMATION - NEAR-OPTIMUM πPS DESIGNS FOR ANY VALUE OF μ

§ 5.1 Summary

In this chapter we shall give some near optimum solutions to our problem posed in § 2.5 for the case of general (even non-integral) values of μ and general values of X_i's subject of course to the condition

$$0 < X_i \leq \frac{X}{\mu} \quad \text{for} \quad 1 \leq i \leq N \qquad (5.1.1)$$

In all these solutions, the basic condition viz., (i) of (2.5.6) is satisfied for the general values of X_i's and hence they are all 'πPS sampling schemes'. The other conditions of (2.5.6) are satisfied to varying degrees.

We shall first prove in § 5.2 that for integral values of μ, the class $D^{(\mu,\pi)}$ of designs as defined by (2.3.12) forms a ∂_3-optimum subclass in the wider class $D^{(\pi)}$ defined by (2.3.11) for the use of $\overset{*}{Y}_{HT}$. Expressions were then given for the expected variances of $\overset{*}{Y}_{HT}$ when applied in designs belonging to $D^{(\pi)}$ or $D^{(\mu,\pi)}$ over the two classes of super-populations Δ_1 and Δ_3 as defined by (2.3.3) and (2.3.7). Some of these expressions, for the special case $\mu = 2$, have already been used in § 4.4 and § 4.5.

In § 5.3 we give a solution, for general values of μ and π_i's, which satisfies (i), (vi) and (vii) of (2.5.6). For this method

it is not possible to estimate the variance of our estimator, since some of the π_{ij}'s are zero. In § 5.4 we present two modifications to overcome this drawback and thus satisfy (iii) of (2.5.6). In §5.5 we present a sequential scheme with the property that at any stage of termination, the inclusion probabilities attained by that time are proportional to a given set of X_i's $(1 \leq i \leq N)$. It is expected that this solution will be of some use in practice when one wants to have a flexible sample size. In each of the solutions given there is thus some departure from the optimum strategy of (2.5.6), though all satisfy (i) of (2.5.6) and hence are 'πPS sampling schemes'. The degree of departure (in terms of the increase in expected variance over ∂_1) can be easily found in any given problem, since the π_{ij}'s are easily calculable in all these cases. Further, for the schemes of § 5.4 and § 5.5 this departure can also be estimated on the basis of a sample, since π_{ij}'s are strictly greater than zero in this cases.

It may be mentioned here that the results given in this chapter were obtained by the author towards the end of 1959 and were read before the Indian Science Congress in January 1960. (Subsequently they were published [19]). At that time the only πPS sampling schemes known (as far as the author is aware) were those of Goodman and Kish [13], Grundy [14], Stevens [39], and Majek [16] besides those of Horvitz and Thompson [22] and Yates and Grundy [40]. All these methods were described briefly in § 2.6.

We illustrate our schemes of § 5.3, § 5.4 and § 5.5 by means of the emperical example considered in §4.3

§ 5.2 Optimum designs for the use of $\overset{*}{Y}_{HT}$

We prove a

Lemma (5.2.1):- For any general sampling design $D = D(S, P)$ we have

$$\sum_{i \neq j} \sum \pi_{ij} = \mu^2 - \mu + V(\mu_s) \qquad \cdots(5.2.1)$$

Proof:- For every $s \in D$ and $1 \leq i \leq N$, define random variables R_{si} thus:

$$R_{si} = \begin{cases} 1 & \text{if} \quad U_i \in s \\ 0 & \text{otherwise} \end{cases}$$

clearly,

$$\pi_i = \sum_{s \in S} R_{si} P_s$$

$$\pi_{ij} = \sum_{s \in S} R_{si} R_{sj} P_s$$

and $$\mu_s = \sum_{i=1}^{N} R_{si} = \sum_{i=1}^{N} R_{si}^2$$

We then have

$$\sum_{\substack{i \neq j \\ 1}}^{N} \sum \pi_{ij} = \sum_{\substack{i \neq j \\ 1}}^{N} \sum \left\{ \sum_{s \in S} R_{si} R_{sj} P_s \right\}$$

$$= \sum_{s \in S} \left\{ \sum_{\substack{i \neq j \\ 1}}^{N} \sum R_{si} R_{sj} P_s \right\}$$

$R_{si}\, R_{sj}$ equals 1 if s contains both U_i and U_j and is zero otherwise. Since a sample consisting of μ_s distinct units gives rise to exactly $\mu_s(\mu_s-1)$ unordered pairs of units of the population both members of each of which belong to s, we have

$$\sum_{i\neq j}^{N}\sum \pi_{ij} = \sum_{s\in S} \mu_s\,(\mu_s - 1)\, P_s$$

$$= \mu^2 - \mu + V(\mu_s)$$

and hence the lemma.

We shall now evaluate the expected variance of the estimator $\overset{*}{Y}_{HT}$ defined by (2.4.1) applied in a given design D belonging to the class $D^{(\pi)}$ as defined by (2.3.11), the expectation being taken over a super-population ∂_3 belonging to the class Δ_3 as definedby (2.3.8). We prove

Lemma (5.2.2) :- For any $D\in D^{(\pi)}$ and $\partial_3 \in \Delta_3$

$$E_{\partial_3} V\,(D,\ \overset{*}{Y}_{HT}) = \frac{a^2X^2}{\mu^2} V(\mu_s) + \sum_{1}^{N} \sigma_i^2 \left(\frac{1-\pi_i}{\pi_i}\right) \qquad \ldots (5.2.6)$$

Proof: For the variance of $\overset{*}{Y}_{HT}$ we have, as remarked in (2.4.2)

$$V(\overset{*}{Y}_{HT}) = \sum_{1}^{N} \frac{Y_i^2}{\pi_i} + \sum_{i\neq j}^{N}\sum \frac{Y_i Y_j}{\pi_i \pi_j}\ \pi_{ij} - Y^2$$

Observing that

$$E_{\partial_3}\ (Y_i^2 \mid X_i) = a^2 X_i^2 + \sigma_i^2$$

and

$$E_{\partial_3}\ (Y_i Y_j \mid X_i X_j) = a^2 X_i X_j$$

we have

$$E_{\partial_3} V(\overset{*}{Y}_{HT}) = \sum_1^N \frac{(a^2X_i^2 + \sigma_i^2)}{\pi_i} + \sum_{\substack{1 \\ i\neq j}}^{N}\sum \frac{a^2 X_i X_j}{\pi_i \pi_j}\pi_{ij} - (a^2X^2 + \sum_1^N \sigma_i^2)$$

$$= a^2X^2\ \left(\frac{1}{\mu} + \frac{1}{\mu^2}\sum_{\substack{1 \\ i\neq j}}^{N}\sum \pi_{ij} - 1\right) + \sum \sigma_i^2\ \frac{(1-\pi_i)}{\pi_i}$$

The result now follows from lemma (5.2.1).

From lemma (5.2.2) we now have

Theorem 5.2. In the class $D^{(\pi)}$ of sampling designs as defined by (2.3.11) the designs that are Δ_5-optimum for the use of $\overset{*}{Y}_{HT}$ are those for which the variance of μ_s is least.

In particular if μ isan integer these Δ_5-optimum designs are the designs of the class $D^{(\mu,\pi)}$ as defined by (2.3.12).

If μ is an integer the minimum (which may not be attainable) value of $V(\mu_s)$ is zero, while if

$$\mu = [\mu] + f, \qquad 0 < f < 1$$

$$\min\ V(\mu_s) = f(1-f)$$

which is attained when the sampling design is such that

$$\mu_s = \begin{cases} \lfloor \mu \rfloor & \text{with probability } f \\ \lfloor \mu \rfloor + 1 & \text{with probability } 1-f \end{cases}$$

If then a design $D'_{opt} \in D^{(\pi)}$ can be found in which the effective sample size μ_s iseither $\lfloor \mu \rfloor$ or $\lfloor \mu \rfloor + 1$, then

$$E_{\partial_3} V(D'_{opt}, \overset{*}{Y}_{HT}) = \frac{a^2 X^2}{\mu^2} f(1-f) + \sum_1^N \sigma_i^2 \frac{(1-\pi_i)}{\pi_i} \qquad (5.2.7)$$

In particular if μ is a positive integer we have for a $D_{opt} \in D^{(\mu,\pi)}$,

$$E_{\partial_3} V(D_{opt}, \overset{*}{Y}_{HT}) = \sum_1^N \sigma_i^2 \frac{(1-\pi_i)}{\pi_i}$$

$$= \sum_1^N \sigma_i^2 \left(\frac{X}{\mu X_i} - 1 \right) \qquad (5.2.8)$$

If further

$$\sigma_i^2 = \sigma^2 X_i^2, \quad 1 \le i \le N$$

then the corresponding apriori distribution is a ∂_1 as defined by (2.3.3), and we have from (5.2.6), (5.2.7) and (5.2.8)

$$E_{\partial_1} V(D \in D^{(\pi)}, \overset{*}{Y}_{HT}) = \frac{a^2 X^2}{\mu^2} V(\mu_s) + \sigma^2 \cdot \left(\frac{X^2}{\mu} - \sum_1^N X_i^2 \right) \qquad (5.2.9)$$

$$E_{\partial_1} V(D_{opt} \in D^{(\pi)}, \overset{*}{Y}_{HT}) = \frac{a^2 X^2}{\mu^2} f(1-f) + \sigma^2 \left(\frac{X^2}{\mu} - \sum_1^N X_i^2 \right) \qquad (5.2.10)$$

and

$$E_{\partial_1} V(D \in D^{(\mu,\pi)}, \overset{*}{Y}_{HT}) = \sigma^2 \left(\frac{X^2}{\mu} - \sum_1^N X_i^2 \right) \qquad (5.2.11)$$

§ 5.3 A solution to the problem for general values of X_i's and μ.

In this section we give a ' π P S' sampling scheme, valid for general values of X_i's and μ. The scheme consists of independent draws of units from the population. If $p_i^{(r)}$ then denotes the probability of drawing U_i in the r-th draw, for any such design, we have for the inclusion probabilities the following simple formulae

$$\pi_i = 1 - \prod_{r=1}^{n} \left(1 - p_i^{(r)}\right) \qquad \cdots (5.3.1)$$

and

$$\pi_{ij} = 1 - \prod_{r=1}^{n} \left(1 - p_i^{(r)}\right) - \prod_{r=1}^{n} \left(1 - p_j^{(r)}\right) + \prod_{r=1}^{n} \left(1 - p_i^{(r)} - p_j^{(r)}\right)$$

$$= \pi_i + \pi_j + \prod_{r=1}^{n} \left(1 - p_i^{(r)} \cdot p_j^{(r)}\right) - 1 \qquad \cdots (5.3.2)$$

where n is the total number of draws made. We now consider the sampling scheme (A).

Suppose, without loss of generality, that

$$\pi_1 \geq \pi_2 \geq \cdots \geq \pi_N \qquad (5.3.3)$$

(If this condition is not satisfied, we need only rearrange the units in the population). Let K_1 be the positive integer such that

$$S_1 = \sum_{1}^{K_1} \pi_i \leq 1 < S_1 + \pi_{K_1+1} \qquad (5.3.4)$$

Denoting by $p_i^{(r)}$ the probability of selecting U_i in the r^{th} draw (this depending only on i and r, since the draws are independent, we choose a **unit** from $\mathcal{U}$ in the first draw, with probabilities

$$p_i^{(1)} = \begin{cases} \pi_i & \text{if} \quad i \leq K_1 \\ 1 - S_1 & '' \quad i = K_1 + 1 \\ 0 & '' \quad i \geq K_1 + 2 \end{cases} \qquad \cdots (5.3.5)$$

Let

$$\delta_1 = \frac{\pi_{K_1+1} - (1 - S_1)}{S_1} \qquad \cdots (5.3.6)$$

From (5.3.4) it follows that $0 < \delta_1 < \pi_{K_1+1}$. Let K_2 be the positive integer such that

$$S_2 = \delta_1 + \sum_{i=K_1+2}^{K_1+K_2} \pi_i \leq 1 < S_2 + \pi_{K_1+K_2+1} \qquad \cdots (5.3.7)$$

for the probabilities at the 2nd draw we shall have

$$p_i^{(2)} = \begin{cases} 0 & \text{if} \quad i \leq K_1 \\ \delta_1 & \text{if} \quad i = K_1 + 1 \\ \pi_i & \text{if} \quad K_1 + 2 \leq i \leq K_1 + K_2 \\ 1 - S_2 & \text{if} \quad i = K_1 + K_2 + 1 \\ 0 & \text{if} \quad i > K_1 + K_2 + 1 \end{cases} \qquad \cdots (5.3.8)$$

Similarly we let

$$\delta_2 = \frac{\pi_{K_1+K_2+1} - (1 - S_2)}{S_2}$$

and K_3 be the positive integer such that

$$S_3 = \delta_2 + \sum_{i=K_1+K_2+2}^{K_1+K_2+K_3} \pi_i \leq 1 < S_3 + \pi_{K_1+K_2+K_3+1}$$

and choose units in the third draw with probabilities

$$p_i^{(3)} = \begin{cases} 0 & \text{if} \quad i \leq K_1+K_2 \\ \delta_2 & \text{if} \quad i = K_1+K_2+1 \\ \pi_i & \text{if} \quad K_1+K_2+2 \leq i \leq K_1+K_2+K_3 \\ 1 - S_3 & \quad i = K_1+K_2+K_3+1 \\ 0 & \quad i > K_1+K_2+K_3+1 \end{cases}$$

The process continues until a partial sum (like S_2 and S_3) absorbs the last draw. If the total number of draws required is n, and

$$K = K_1+K_2+ \ldots + K_{n-1}$$

there is a probability

$$\delta_n = 1 - (\delta_{n-1} + \pi_{K+1} + \pi_{K+2} + \ldots + \pi_N)$$

for not selecting any unit in the n-th draw. At every other draw, the probabilities assigned add up to 1.

We can now easily verify that the scheme given above satisfies (i) of (2.5.6). This is obvious for all the units except U_{K_1+1}, $U_{K_1+K_2+1}$, $\cdots$, $U_{K_1+K_2+\ldots+K_{n-1}+1}$ because each of these units can be selected in only one specific draw and that too with the probability equal to the corresponding π_i. Each of the units U_{K_1+1}, $U_{K_1+K_2+1}$ etc., can be selected in one or other of two specific draws, $U_{K_1+\ldots+K_r+1}$ being selectable in the r^{th} or $(r+1)^{th}$ draw but in no other draw. The δ's are so constructed for these that the total inclusion probability is the corresponding π_i. For example, the total inclusion probability for U_{K_1+1} is equal to

$$(1 - S_1) + S_1 \cdot \delta_1 = \pi_{K_1+1}$$

from (5.3.6). Similarly for the other units.

What is essentially done in the above scheme is that we have divided the population into a number of homogeneous strata with however are not strictly overlapping, two adjacent strata having possibly one unit in common (which we call the junctional units) We then selected one unit from each of the stratum with probabilities equal to the original π_i's, except for the junctional units where these probabilities are properly adjusted. The sampling in one stratum is independent of that in another stratum.

The condition (ii) of (2.5.6) is not strictly satisfied but since the only possible repetitions are restricted to a few junctional units, the variance of μ_s will be quite low (μ_s in fact hasto be between n and $\frac{n}{2}$). The method given is thus a good approximation to the optimum method required. It is not possible to estimate the variance of the estimator of the total since some of the π_{ij}'s are zero. For U_i and U_j belonging to the same stratum and neither of which is a junctional unit we have $\pi_{ij} = 0$. Thus (iii) of (2.5.6) is not satisfied. We shall however give in § 5.4 two modification of this procedure which satisfy (iii) and which involve the calculation of π_{ij}'s for a scheme of this section with revised values of π_i's. We therefore remark here that the calculation of π_{ij}'s is very easy for our scheme since the draws are independent and thus satisfies the requirement (v) of (2.5.6). Referring to (5.3.2) we see that the calculation of a π_{ij} needs the calculation of

$$\prod_{r=1}^{n} \left\{ 1 - p_i^{(r)} - p_j^{(r)} \right\}$$

However large n may be, this product contains utmost four terms, since $p_i^{(r)}$, for any given p_i, is zero for all but utmost four values of r. The method does not need heavy calculations anywhere since the given π_i's are utilised very liberally. All that is necessary is to go on adding up the π_i's (the cumulated sums need not be noted down) until the sum is about to exceed 1 and at that stage note down the S_1

and calculate ∂_1. Again ∂_1, π_{K_1+2} π_{K_1+3}, etc., are added up until the sum is about to exceed 1 and note down the S_2 and calculate ∂_2 etc.

The calculation of π_i's from the X_i's itself is not necessary since the formulae for S_i's and ∂_i's can be translated to those given in terms of X_i's.

The method is illustrated for the two populations 'Anthra Pradesh' and 'Punjab' considered in § 4.3 .

We consider for illustration, the problem of selecting an expected effective sample of size 4 districts from each of the populations of districts of 'Andhra Pradesh' and 'Punjab' considered in §4.3.

The first step is to arrange the units in the ascending order of X_i's. (This is not actually necessary for the validity of the method given in this section but is proposed for robustness of the method against small departures from the model assumed).

For 'Andhra Pradesh' the arrangement is at follows, where we give the corresponding original serial number in brackets.

1	(19)	6 (18)	11 (9)	16 (13)
2	(16)	7 (17)	12 (8)	17 (2)
3	(15)	8 (12)	13 (4)	18 (1)
4	(14)	9 (10)	14 (5)	19 (3)
5	(20)	10 (11)	15 (7)	20 (6)

We find that

$K_1 = 7, \quad S_1 = .9663; \quad \partial_1 = .1548$

$K_2 = 12, \quad S_2 = .9649, \quad \partial_2 = .1865$

$K_3 = 16, \quad S_3 = .8644, \quad \partial_3 = .1468$

$K_4 = 19, \quad S_4 = .7073, \quad \partial_4 = .0448$

$\epsilon = .9552$

We thus need make 5 draws in all to get an expected size of 4 distinct units. The probability is .9552 that in the 5th draw no unit of the population is selected. It is evident that $V(\mu_s)$ will be exceedingly small in this case.

For 'Punjab' the arrangement of the units is as follows:

1 (8)	6 (16)	11 (4)	16 (17)
2 (6)	7 (11)	12 (5)	17 (2)
3 (15)	8 (14)	13 (1)	18 (12)
4 (19)	9 (7)	14 (10)	19 (13)
5 (18)	10 (3)	15 (9)	

We find in this case that

$K_1 = 7, \quad S_1 = .8036, \quad \partial_1 = .0192$

$K_2 = 12, \quad S_2 = .9936, \quad \partial_2 = .2555$

$K_3 = 15, \quad S_3 = .7907, \quad \partial_3 = .0853$

$K_4 = 18, \quad S_4 = .6823, \quad \partial_4 = .0334$

and $\epsilon = .9666.$

In this case also 5 draws are required to get on the average 4 distinct units. The probability of not getting a unit in the 5th draw is .9666 and it is evident that $V(\mu_s)$ will be exceedingly small in this case also.

§ 5.4 Two modifications to suit the estimability of $V(\overset{*}{Y}_{HT})$

We shall now present two modifications to make $\pi_{ij} > 0$ for all i and j.

Modification (A):- Let

$$\left.\begin{aligned} p_i^{(o)} &= \frac{\pi_i}{\mu} \\ \text{and} \qquad \pi_i' &= \frac{\pi_i(\mu - 1)}{(\mu - \pi_i)} \end{aligned}\right\} \qquad \ldots (5.4.1)$$

Step I:- Select a sample by the method of § 5.3, replacing π_i's therein by π_i's as given by (5.4.1).

Step II:: Now choose a unit with probability $p_i^{(o)}$ for the i-th unit. Add thus selected unit to the sample already chosen.

For the over-all probability of including U_i in the sample we argue as follows. The unit U_i is either selected in Step I or is not selected in step I but is selected in the II step. The overall probability is thus equal to

$$\begin{aligned} &\pi_i' + (1 - \pi_i')\frac{\pi_i}{\mu} \\ &= \frac{\pi_i}{\mu} + \pi_i'\left(1 - \frac{\pi_i}{\mu}\right) \\ &= \pi_i \end{aligned}$$

which is the required inclusion probability. For the value of π_{ij} we have the event

e_{ij} : U_i and U_j are included in the sample.

split up into three mutually disjoint units $e_{ij}^{(1)}$, $e_{ij}^{(2)}$, $e_{ij}^{(3)}$ thus:

$e_{ij}^{(1)}$: U_i and U_j are both selected in step I

$e_{ij}^{(2)}$: U_i alone is selected in step I and U_j is selected in step II

$e_{ij}^{(3)}$: U_j alone is selected in step I and U_i is selected in step II

Clearly

$$P(e_{ij}^{(1)}) = \pi'_{ij}$$

$$P(e_{ij}^{(2)}) = (\pi_i - \pi'_{ij}) \frac{\pi_j}{\mu}$$

and

$$P(e_{ij}^{(3)}) = (\pi_j - \pi'_{ij}) \frac{\pi_i}{\mu}$$

where π'_{ij}'s are the joint inclusion probabilities of step I, so that for the over-all inclusion probabilities we have

$$\pi_{ij} = \frac{2\,\pi_i \pi_j}{\mu} + \frac{\pi'_{ij}}{\mu}(\mu - \pi_i - \pi_j) \qquad \ldots(5.4.2)$$

The calculation of π_{ij} thus needs that of π'_{ij}. These are calculated as indicated in § 5.3.

The method thus satisfies an important requirement that the π_{ij}'s should be easily calculable for the estimation of $V(\hat{Y}^*_{HT})$. In this respect the method given in § 5.3 compares favourably with that

of Goodman and Kish which also has the same drawback of not satisfying (iii) of (2.5.6) and which can also be modified as above. However, the calculation of π'_{ij}'s for their proceedure is not facilitated by a neat algebraic form as remarked in § 2.4.

With the modification (A) given above (i), (iii), (vi) and (vii) of (2.5.6) are thus satisfied but there is a slightly greater departure from (ii) since we allowed repetitions to occur more often. Also this method may not give a reliable estimate of $V(\overset{*}{Y}_{HT})$ since some of the π_{ij}'s may be rather smaller. If it desired to reduce this disadvantage we may adopt, instead, the

Modification (B):- Select two independent and equivalent subsamples s_1 and s_2 with inclusion probabilities π_i^*'s for each where

$$\pi_i^* = 1 - \sqrt{1-\pi_i}, \qquad 1 \leq i \leq N \tag{5.4.3}$$

For the overall inclusion probabilities we have

$$P(U_i \in s_1 \cup s_2) = P(U_i \in s_1) + P(U_i \in s_2) - P(U_i \in s_1 \text{ and } s_2)$$
$$= 2\pi_i^* - \pi_i^{*2} = \pi_i$$

by (5.4.3), which thus satisfies (i) of (2.5.6). For the overall joint inclusion probabilities we have

$$P(U_i \notin s_1 \cup s_2,\ U_j \notin s_1 \cup s_2) = P(U_i \notin s_1, U_j \notin s_1) \cdot P(U_i \notin s_2, U_j \notin s_2)$$

Further $P(U_i \notin s_1,\ U_j \notin s_1) = 1 - P(U_i \in s_1 \text{ or } U_j \in s_1)$

$$= 1 - \pi_i^* - \pi_j^* + \pi_{ij}^*$$

where π_{ij}^* is the joint inclusion probability of U_i and U_j for a given subsample.

Similarly for s_2 and for $s = s_1 \cup s_2$

$$P(U_i \notin s,\ U_j \notin s) = (1 - \pi_i - \pi_j + \pi_{ij}).$$

Hence

$$\pi_{ij} = (1 - \pi_i' - \pi_j' + \pi_{ij}')^2 - (1 - \pi_i - \pi_j)$$

This modification, however, deviates much farther from (ii) of (2.5.6). Whether so much sacrifice of accuracy in the estimation of the population total need be made just to get a reliable estimate of the variance of that estimator is left to the individual case. However where we expect nonsampling errors to operate heavily, and thus prefer to have a check by having two independent subsamples the modification given above can profitably used. It may be noted that having obtained two independent subsamples we can have the individual estimates

$$\left.\begin{aligned} \overset{*}{Y}_{HT,1} &= \sum_{i \in s_1} \frac{Y_i}{\pi_i} \\ \text{and} \quad \overset{*}{Y}_{HT,2} &= \sum_{i \in s_2} \frac{Y_i}{\pi_i} \end{aligned}\right\} \qquad \cdots (5.4.5)$$

and take for our pooled estimate

$$\overset{*}{Y}_{HT,p} = \frac{1}{2} (\overset{*}{Y}_{HT,1} + \overset{*}{Y}_{HT,2}) \qquad \cdots (5.4.6)$$

This is not the same as our optimum estimator (5.2.2) and is in fact inadmissible being dependent on the number of times a unit is repeated in the sample, to some extent. While repetetions within a subsample

are ignored if takes note of repetitions between subsamples. A quick unbiased estimate of the variance of (5.4.6) is given by

$$\overset{*}{V}(\overset{*}{Y}_{HT,P}) = \frac{1}{4}\ (\overset{*}{Y}_{HT,1} - \overset{*}{Y}_{HT,2})^2 \qquad \ldots (5.4.7)$$

Since $V(\overset{*}{Y}_{HT})$ is lessthan $V(\overset{*}{Y}_{HT,P})$, the r.h.s. of (5.4.7) gives a quick over estimate (unbiasedly only) of the variance of $V(\overset{*}{Y}_{HT})$

§ 5.5 A sequential scheme

We shall now give a sampling scheme which answers an important problem that often arises in practice. This is the situation where we would like to have a certain amount of flexibility in the sample size to be taken. This may occur either when our budget is undecided when we find at a later stage that our original estimates (of the cost per unit chosen in the sample) are grossly misleading. In such cases it is of interest to find a sampling method which consists of selecting units one by one such that whatever may be number of draws made, the inclusion probability π_i attained is exactly proportional to the given auxiliary information. We now give a method of sampling which solves this problem in almost all practical situations. This scheme consists of independent draws such that at the end of every draw (up to a certain stage), the inclusion probabilities attained till then, are proportional to a given set of values $P_i = \frac{X_i}{X}$ $(1 \leq i \leq n)$ say. Satisfying these properties, there is a unique sampling method, which we shall now give.

Let $p_i^{(n)}$ be the probability of selecting U_i in n-th draw, and $\pi_i^{(n)}$ be the inclusion probability of U_i at the end of n-th draw. We want to have

$$\pi_i^{(n)} = K_n P_i \quad i = 1,2,\ldots, N, \quad n = 1,2,\ldots \qquad (5.5.1)$$

where K_n's are constants. At the end of n draws we have

$$\pi_i^{(n)} = 1 - (1 - p_i^{(1)})(1 - p_i^{(2)}) \ldots (1 - p_i^{(n)}) \qquad \ldots (5.5.2)$$

Further, we have the conditions

$$\sum_{i=1}^{N} p_i^{(n)} = 1, \quad n = 1,2,\ldots \qquad \ldots (5.5.3)$$

since the probabilities should add up to 1 at each draw. Setting n=1, we have from (5.5.2 and 5.5.3)

$$K_1 = 1 \text{ and } p_i^{(1)} = P_i \quad (1 \leq i \leq N) \qquad \ldots (5.5.4)$$

Setting $n = 2$ in (5.5.2), we have for all 'i'

$$(1 - p_i^{(1)})(1 - p_i^{(2)}) = 1 - K_2 P_i$$

giving $$p_i^{(2)} = P_i (K_2 - K_1)/(1 - K_1 P_i). \qquad \ldots (5.5.5)$$

We then have from (5.5.3)

$$K_2 = K_1 + \left\{ \sum_1^N \frac{P_i}{1 - K_1 P_i} \right\}^{-1} \qquad \cdots (5.5.6)$$

Similarly, setting $n = 3$ we have for $1 \leq i \leq N$,

$$(1 - p_i^{(1)})(1 - p_i^{(2)})(1 - p_i^{(3)}) = 1 - K_3 P_i$$

giving

$$p_i^{(3)} = \frac{P_i (K_3 - K_2)}{1 - K_2 P_i} \qquad \cdots (5.5.7)$$

and (5.5.3) then gives

$$K_3 = K_2 + \left\{ \Sigma \frac{P_i}{1 - K_2 P_i} \right\}^{-1} \qquad \cdots (5.5.8)$$

In general (writing $K_o = 0$ conventionally), it can be shown that

$$K_{n+1} = K_n + \left\{ \Sigma \frac{P_i}{1 - K_n P_i} \right\}^{-1} \qquad \cdots (5.5.9)$$

and

$$p_i^{(n)} = \frac{P_i (K_n - K_{n-1})}{1 - K_{n-1} P_i} \qquad \cdots (5.5.10)$$

for $1 \leq i \leq N$ and for n upto a certain limit which we shall presently investigate. After the sample is drawn π_{ij}'s for sampled units can be calculated using (5.3.2).

From (5.5.9) we see that

$$K_{n+1} > K_n$$

surely if

$$1 - K_n P_i > 0 \qquad (1 \leq i \leq N)$$

which condition can be written compactly as

$$\max_i P_i = P_{max} < \frac{1}{K_n} \quad \ldots \quad \ldots \quad (5.5.11)$$

Further if in the succession of values of K_n's calculated from the recurrence formula (5.5.9) there is one K_{n_o} say for which (5.5.11) does not hold good, we see from (5.5.10) that the value of $p_i^{(n_o+1)}$ for the unit U_i corresponding to P_{max} turns out to be negative and hence untenable since $p_i^{(n_o+1)}$'s are to be used as probabilities in the (n_o+1)th draw. Thus, if $(n'-1)$ is the maximum value of n for which (5.5.11) holds good the scheme given above remains valid upto n'th draw only.

This, however, is not a serious limitation in practice. From (5.5.1) we have

$$\sum_{i=1}^{N} \pi_i^{(n)} = K_n$$

where $\pi_i^{(n)}$ can be interpreted as the inclusion probabilities of U_i upto the nth draw, if $n \leq n'$. Hence K_n is the expected effective sample size upto the nth draw for all $n \leq n'$. The sequence of formulae

inclusion probabilities attained are exactly proportional to the given values of the auxiliary character. The only condition the values X_i's are that

$$0 < X_i \leq \frac{X}{Nf}, \qquad (1 \leq i \leq N) \qquad (5.5.13)$$

where f is the sampling fraction aimed at. Further such a scheme of independent draws is unique since the values of $p_i^{(n)}$'s are uniquely determined. If after drawing a sample of size n we want to increase the sample size, all we need to do is to perform the (n+1)th, (n+2)th,... draws with the probabilities as given by (5.5.9) and (5.5.10) until we get the required sample size. Similarly, in cases where we are not quite sure of the reliability of our estimates of the parameters of our cost function or of our budget allocations, we can draw a sample of sufficiently big size at first but then inspect the units in the sample one after another in the order in which they are drawn - at least on and after a critical stage. If at any stage we find that our total cost is about to exceed the budget allocation or that sufficient accuracy of the estimates is expected to have been already attained, we terminate inspection at that stage and note the value of 'n', the number of draws made in all. The population total can then be estimated as also an unbiased estimate of the variance of the estimator, as usual, using (2.4.1) and (2.4.6).

If n draws are made in all, denoting by a suffix 'seq' the method given in this section, we have from (5.3.2), (5.5.1) and (5.5.10),

$$\overset{*}{Y}_{HT,seq} = \frac{1}{K_n} \sum_{i \in s} \frac{Y_i}{P_i} \tag{5.5.14}$$

$$V(\overset{*}{Y}_{HT,seq}) = \frac{1}{K_n} \sum_{i=1}^{N} \frac{Y_i^2}{P_i} + \frac{1}{K_n^2} \sum\sum_{i \neq j} \frac{Y_i Y_j}{P_i P_j} \pi_{ij}^{(n)} - Y^2 \tag{5.5.15}$$

and

$$\begin{aligned} \overset{*}{V}(\overset{*}{Y}_{HT,seq}) &= \left(\frac{1}{K_n^2} \sum_{i \in s} \frac{Y_i^2}{P_i^2} - \frac{1}{K_n} \sum_{i \in s} \frac{Y_i^2}{P_i}\right) + \\ &\quad + \left(\frac{1}{K_n^2} \sum\sum_{\substack{i \neq j \\ \in s}} \frac{Y_i Y_j}{P_i P_j} - \sum\sum_{i \neq j} \frac{Y_i Y_j}{\pi_{ij}^{(n)}}\right) \\ &= \overset{*}{Y}_{HT}^2 - \left(\frac{1}{K_n} \sum \frac{Y_i^2}{P_i} + \sum\sum_{i \neq j} \frac{Y_i Y_j}{\pi_{ij}^{(n)}}\right) \quad \ldots \end{aligned} \tag{5.5.16}$$

where $\pi_{ij}^{(n)} = K_n (P_i + P_j) - 1 + \prod_{r=1}^{n} (1 - p_i^{(r)} - p_j^{(r)})$

$$\begin{aligned} &= K_n (P_i + P_j) - 1 + \overline{\prod_{n}} \frac{1 - (K_r - K_{r-1})(P_i + P_j)}{(1 - K_{r-1} + P_i)(1 - K_{r-1} P_j)} \\ &= K_n (P_i + P_j) - 1 + \frac{\overline{\prod} \; 1 - K_r (P_i + P_j) + K_{r-1}^2 P_i P_j}{\overline{\prod} (1 - K_{r-1} P_i) \cdot \overline{\prod} (1 - K_{r-1} P_j)} \end{aligned} \tag{5.5.17}$$

After drawing the sample the products

$$\overline{\prod_{r=1}^{n}} (1 - K_{r-1} P_i)$$

have to be calculated for each selected unit. For every pair of units selected we need then find

$$\overline{\prod} \; 1 - K_r (P_i + P_j) + K_{r-1}^2 P_i P_j$$

Making use of these products $\pi_{ij}^{(n)}$'s can be calculated from (5.5.17).

§ 5.6. Illustrations

We shall now illustrate the sampling scheme of § 5.5 for the four live examples considered in § 4.3.

The first step is to calculate K_i's using the recurrence formula (5.5.9), until (5.5.11) breaks down. These are given below for each of the populations. The computations are made on IBM 1401 giving K_n's correct to 16 significant figures but we shall present them here only to 4 decimal places, for simplicity.

Table 5.6α. Table of K_i's

Order of the draw (i)	Andhra Pradesh	West Bengal	Punjab	Gujarat
1	1.0000	1.0000	1.0000	1.0000
2	1.9455	1.9103	1.9373	1.9254
3	2.8389	2.7331	2.8154	2.7798
4	3.6828	3.4696	3.6376	3.5667
5	4.4793	4.1195	4.4070	4.2887
6	5.2305	4.6806	5.1266	4.9487
7	5.9385	5.1473	5.7990	5.5487
8	6.6051	5.5096	6.4269	6.0905
9	7.2319	5.7530	7.0125	6.5752
10	7.8207	5.8699	7.5582	7.0029
11	8.3728	5.8959	8.0661	7.3729
12	8.8896	5.8971	8.5379	7.6822
13	9.3721		8.9754	7.9251
14	9.8214		9.3802	8.0925
15	10.2380		9.7536	8.1779
16	10.6225		10.0967	8.1995
17	10.9748		10.4103	8.2008
18	11.2945		10.6949	
19	11.5800		10.9504	
20	11.8286		11.1761	

contd.

(contd. from previous page)

Order of the draw (i)	Andhra Pradesh	West Bengal	Punjab	Gujarat
21	12.0354		11.3702	
22	12.1924		11.5292	
23	12.2899		11.6474	
24	12.3276		11.7182	
25	12.3327		11.7438	
26	12.3328		11.7468	
27			11.7469	

The method breaks down after the 26th, 12th, 27th and 19th draws for the populations of 'Andhra Pradesh', 'West Bengal', 'Punjab' and 'Gujarat' respectively. The expected effective sample sizes attainable (given by the value of K_n last recorded, (c.f.) (5.5.1)) are 12.3328; 5.8071, 11.7469 and 8.2008 respectively which give the attainable expected effective sampling fractions to be about 62 per cent, 37 per cent, 62 per cent and 48 per cent respectively. These are much larger than the sampling fractions that we need in practice and thus the method is usuable in all most all practical situations.

§ 5.6. Some concluding remarks:-

The methods given in this chapter can be mixed with one another as also with the sampling schemes (A) and (B) of Chapter IV. For example the sampling scheme (A) of § 5.3 can be modified by a mixture with scheme (B) of § 4.2 as follows: We cumulate π_i's until their sum just falls below 2 (instead of 1 as in § 5.3) and find θ's suitably, which are obtained by replacing 1 on the r.h.s. of (5.3.6) etc. by 2. We thus break up the population into slightly overlapping strata with sum of π_i's equal to 2 in each stratum. In fact we can manipulate to keep this overlapping quite low by grouping the π_i's so as to have their sum very nearly 2 in each group and then apply the method of § 5.3. We can then apply the scheme (B) of § 4.2 and draw 2 distinct units from each stratum. The strata totals can be estimated by using (2.4.1) with π_i's of the junctional units modified as $(2-S_k)$'s or θ_k's as the case may be. The sum of these strata estimates gives an estimate of Y which is not admissible since the overall sample can ce contain repetitions (due to junctional units), though the sample within each stratum consists of distinct units. We can then use the estimator (2.4.1) with the original values of the π_i's themselves. However, the calculation of π_{ij}'s needs a slight modification over that given in § 4.1 and § 4.2 which can be carried out as in § 5.3. Instead of π_{ij} being zero for units belonging to the stratum (and neither of which is a junctional unit) we now have π_{ij} given by an expression corresponding to (4.2.4). All π_{ij}'s are strictly positive so that the

variance of the estimator can also be estimated. This method is likely to be better than both the modifications given in § 5.4 (which were given much earlier to obtaining of the results of Chapter IV, by the author).

Similarly in situations where there is expected to be a correlation between the variations in adjecent units,

$$\mathrm{Cov}_{\theta_1} (Y_i, Y_j \mid X_i, X_j) \neq 0$$

in general for units U_i and U_j that are geographically near to each other. In such cases we may first stratify the population so as to keep adjacent or near adjacent units in different strata and then apply our methods to each stratum seperately and add up the estimates of the various strata totals.

CHAPTER VI

SOME MISCELLANEOUS RESULTS IN 'πPS' SAMPLING

§ 6.0. Summary

In this chapter we shall consider some miscellaneous results in 'πPS' sampling. In § 6.1 we consider multistage πPS sampling schemes and derive the optimum (in the sense of lowest cost, for a simple cost function) multistage designs in the class of all multistage designs with given values of inclusion probabilities and with a fixed number of kth-stage units being drawn from every selected (k-1)th stage unit. In § 6.2 we probe into the method of substituting prior estimates of parameters in the derivation of optimum allocation of sample size in stratified sampling and remark that the conventional method is justifiable only if the coefficients of variation are nearly equal for all the strata. In § 6.3 we give a method of amalgamating two different πPS sample surveys on the same population, so as to have a large number of common units (to reduce travelling costs). The method is analogous to that of Keyfitz [23]. Finally in § 6.4 we deal with the problem of testing the validity of the basic model (2.3.3), on the basis of a sample drawn, which leads to a test for the validity of the model used.

§ 3.1. Optimum allocation in multistage sampling.

We have seen in § 2.3 , that with respect to the class Δ_1 of prior distributions defined by (2.3.3), the class of sampling strategies H_o for which

$$\left.\begin{aligned} \pi_i &= \mu \frac{X_i}{X} \\ \mu_s &= \mu \qquad \text{for all samples} \end{aligned}\right\} \qquad (6.1.1)$$

and $$\overset{*}{Y} = \overset{*}{Y}_{HT} = \sum_{i \in s} \frac{Y_i}{\pi_i}$$

are optimum in the class of all stratefies for which

$$\sum_{1}^{N} \pi_i = \mu \qquad (6.1.2)$$

Further, all strategies satisfying (6.1.1) are δ_1-equivalent, in whatever other respects they may differ from each other. This means that there is no such thing like gain due to stratification, loss due to multistage sampling etc. so long as the designs all satisfy (6.1.1), and that all are δ_1 - best. This has an immediate application in our choice of the design to reduce our costs. For unistage sampling with the simple cost function

$$C = C_o + B.\mu_s \qquad (6.1.3)$$

which because of (6.1.1) gives

$$C = C_o + B\mu$$

all designs of H_o are equally costly. Let us consider a multistage sampling with the simple cost function

$$c = c_o + B_1 \mu_{1s} + B_2 \mu_{2s} + \cdots B_k \mu_{ks} \qquad (6.1.4)$$

where the design is a k-staged one and B_t is the cost of inspection per t-th stage unit, μ_{ts} the number of t-th stage units chosen. We have the restriction

$$n = \mu_{k,s} = \mu \qquad (6.1.5)$$

for all samples. Here we are considering that the auxiliary information is available upto the last stage, and the validity of a prior distribution $\partial_1 \in \Delta_1$ is also assumed. From the remarks made earlier, all designs satisfying (6.1.5) are all Δ_1-equivalent. We shall therefore try to minimise our cost.

From (6.1.4) we find that the expected cost for samples from a given multistage design satisfying (6.1.5) is given by

$$E(c) = c_o + B_1 \mu_1 + B_2 \mu_2 + \cdots + B_k \mu_k \qquad (6.1.4')$$

we have to minimise the above by a proper choice of the μ_t's for $1 \leq t \leq k-1$, and remembering that $\mu_k = \mu$, a given constant. We shall assume, as is the case in multistage sampling, that

$$B_1 \geq B_2 \geq \cdots \geq B_{k-1} \geq B_k \qquad (6.1.6)$$

Then (6.1.4') is minimised by choosing for $\mu_1, \mu_2, \cdots, \mu_{k-1}$ the smallest possible values.

We shall now restrict our attention to the subclass of multistage designs which besides satisfying (6.1.5) also satisfy

$$\mu_{t,s} = \mu_t \quad \text{for} \quad 1 \leq t \leq k-1 \text{ and for all samples } s \qquad (6.1.7)$$

This is so because the construction of the derived optimum design is relatively simple for these designs.

We can take

$$\mu_{t,opt} = 1 \quad \text{for} \quad 1 \leq t \leq k-1 \qquad (6.1.8)$$

as this minimises the cost. But this is a feasible solution only if every $(k-1)$th-stage unit contains at least $\mu_k = \mu$ k^{th}-stage units because the total number of k^{th}-stage units to be chosen is μ, a given number. If this condition does not hold good we consider the $(k-1)$th-stage unit containing the smallest number of the k^{th}-stage units, and if this number is denoted by ϕ_{k-1}, we should take, in virtue of (6.1.6)

$$\left.\begin{aligned} \mu_{k-1,opt} &= \frac{\mu}{\phi_{k-1}} \\ \mu_{t,opt} &= 1 \qquad \text{for} \quad 1 \leq t \leq k-2 \end{aligned}\right\} \qquad (6.1.9)$$

Again, the second condition of (6.1.9) is a consistent condition only if every $(k-2)^{th}$-stage unit contains at least $\frac{\mu}{\phi_{k-1}}$ distinct $(k-1)^{th}$-stage units.

Otherwise we set

$$\mu_{k-2,opt} = \frac{\mu_{l-1,opt}}{\phi_{k-2}} \qquad (6.1.9)$$

where ϕ_{k-2} is the smallest number of $(k-1)^{th}$-stage units contained in a $(k-2)^{th}$-stage unit, again in virtue of (6.1.6). By successively applying (6.1.6) thus, k times, we can find the $\mu_{t,opt's}$.

Let now

$$n_{t,opt} = \frac{\mu_{t+1,opt}}{\mu_{t,\ opt}} \qquad (1 \leq t \leq k-1) \qquad (6.1.10)$$

$n_{t,opt}$ can thus be interpreted as the optimum number of the $(t+1)^{th}$-stage units to be chosen from each selected t^{th}-stage unit.

The $n_{t,opt}$'s may not be integers. In that case we have to take integral values for them near the optimum solution. In practice n is not rigidly fixed so that there will be room for some manouvre. Further details like deriving the loss due to such deviation from the optimum values will not be discussed here as they are straight forward and well-known methods. At any rate this loss is likely to be quite small when $n_{i,\ opt}$'s are even moderately large sampling can be realised. Let N be the total number of f s u's N_i the number of s s u's in the i^{th} f s u, N_{ij} the number of t s u's in the j^{th} s s u of the i^{th} f s u-briefly the (i,j)th s s u - etc. If X_i is the total of $\mathcal{X}$ character for the i^{th} f s u and X the population total of $\mathcal{X}$, we choose a πPS sampling scheme for the selection of the f s u's with the inclusion probability

$$\pi_i = \frac{X_i}{X}\ n_{i,opt} \qquad (6.1.11)$$

for the i^{th} fsu ($1 \leq i \leq N$) and with the number of fsu's being the same in every sample. If i_1-th fsu gets selected in the sample, the ssu's are selected from it by a πPS sampling scheme consisting of $n_{2,opt}$ distinct units in each sample such that the inclusion probability for the j^{th} ssu is given by

$$\pi_{ij} = \frac{X_{i_1 j}}{X_{i_1}} \cdot n_{2,opt} \qquad \ldots (6.1.12)$$

Similarly the procedure is carried on upto the k-th stage units. Clearly the over-all inclusion probability for the $(i_1\ i_2\ \ldots\ i_k)$ th k-th stage unit is given by

$$\pi_{i_1 i_2 \ldots i_k} = \frac{X_{i_1}}{X} \cdot n_{1.opt} \cdot \frac{X_{i_1 i_2}}{X_{i_1}} \cdot n_{2,opt} \ldots \frac{X_{i_1 \ldots i_k}}{X_{i_1 \ldots i_{k-1}}} \cdot n_{k-1,op}$$

$$= \frac{X_{i_1 \ldots i_k}}{X} \cdot n_{1,opt} \cdot \ldots \cdot n_{(k-1),opt}$$

$$= n \cdot \frac{X_{i_1 \ldots i_k}}{X} \qquad \ldots (6.1.13)$$

which is what is required by the optimum strategy.

In Chapter IV some near-optimum solutions are given for the problem of selecting n units by a unistage sampling scheme. These methods can be applied in the above problem for the selections at various stages, until such a time when an exact method, like the one given in cahpter V for the case n=2, is developed for general values of n.

For an application of a πPS sampling method of the type given above it is of course necessary that for the given $n = n_k$, we have

$$X_{i_1} \leq \frac{X}{n_1,opt} \qquad (1 \leq i_1 \leq N)$$

$$X_{i_1 i_2} \leq \frac{X_{i_1}}{n_{2,opt}} \qquad (1 \leq i_2 \leq N_{i_1})$$

$$X_{i_1 i_2 i_3} \leq \frac{X_{i_1 i_2}}{n_{3,opt}} \qquad (1 \leq i_3 \leq N_{i_1 i_2})$$

.....

$$X_{i_1 i_2 \ldots . i_k} \leq \frac{X_{i_1 i_2 \ldots . i_{k-1}}}{n_k}, \quad (1 \leq i_k \leq N_{i_1 \ldots i_{k-1}})$$

.....(6.1.14)

which are slightly stronger requirements than the one in unistage sampling for which we should have simply

$$X_{i_1 \ldots . i_k} \leq \frac{X}{n} \qquad \ldots (6.1.15)$$

However in general (6.1.14) will be satisfied. The modifications necessary to meet any violation of some of the conditions of (6.1.14) are not likely to present much difficulty.

§ 6.2 On optimum allocation in stratified simple random sampling.

For the case of simple random sampling with or without replacement, optimum allocations of the sample size to the various strata were derived first by Neyman [29] for simple cost function given by

$$C = C_o + C_1 n \qquad \cdots (6.2.1)$$

Later the method was extended to cover the case of unequal sampling costs per unit for the strata. It may be noted that this optimality is slightly different from the optimality discussed in § 6.1. Here the optimality is for the minimization of the actual variance (not its expectation with respect to some prior distribution) of a particular (conventional) estimation procedure only.

We need not go into the details of these wellknown methods. For the case of simple random sampling with replacement, with the cost function (6.2.1), the optimum values of n_i are given by

$$n_{i,opt} = n. \frac{N_i \sigma_i}{\Sigma N_i \sigma_i} \qquad \cdots (6.2.2)$$

while for the case of 'without replacement' sampling

$$n_{i,opt} = n. \frac{N_i S_i}{\Sigma N_i S_i} \qquad \cdots (6.2.3)$$

where N_i is the size, σ_i^2 the variance and $S_i^2 = \frac{N_i}{N_i-1}\sigma_i^2$ for the i-th stratum. Similar expressions can be obtained for the case of unequal costs over strata.

For the calculation of $n_{i,opt}$'s given by (6.2.2) and (6.2.3) one needs at least the proportionate values of σ_i^2's, which are unknown. In practice one substitutes for them plausible estimates α_i^2's which are usually nothing but the σ_i^2's of some auxilary information (usually the values of the same character studied in a recent epoch of time). The justification for the assumption that the unknown proportionate values of σ_i^2's are usually not far from the known proportionate values of the known α_i^2's, can be examined in the light of an apriori distribution $\delta_1 \in \Delta$, as defined by (2.3.3). We have

$$\sigma_i^2 = \frac{1}{N_i}\left\{\sum_{j=1}^{N_i} Y_{ij}^2 - \frac{Y_i^2}{N_i}\right\}$$

where Y_{ij} is the y-value on the j-th unit of the i-th stratum.

Hence

$$E_{\delta_1}(\sigma_i^2) = \frac{1}{N_i}\left\{\sum_j (a^2+\sigma^2) X_{ij}^2 - \frac{1}{N_i}\left(a^2 X_i^2 + \sigma^2 \sum_j X_{ij}^2\right)\right\}$$

$$= \frac{1}{N_i}\left\{a^2\left(\sum_j X_{ij}^2 - \frac{X_i^2}{N_i}\right) + \sigma^2\left(1 - \frac{1}{N_i}\right)\sum_j X_{ij}^2\right\}$$

$$= a^2 \alpha_i^2 + \sigma^2 \cdot \frac{N_i-1}{N_i}\left(\bar{X}_i^2 + \alpha_i^2\right).$$

$$= a^2 \alpha_i^2 + \sigma^2 \left(\bar{X}_i^2 + \alpha_i^2\right),$$

assuming $\frac{N_i-1}{N_i}$'s to be nearly equal. This is the case of N_i's are large enough or are nearly equal. In the latter case the constant value of $\frac{N_i-1}{N_i}$ can be absorbed into σ^2. Hence we can write

$$E_{\delta_1}(\sigma_i^2) = (a^2 + \sigma^2)\,\alpha_i^2 + \sigma^2\,\bar{X}_i^2 \qquad \cdots (6.2.4)$$

(In the above derivation σ^2 should not be confused with population variance. It is the σ^2 of (2.3.3)).

Thus the σ_i^2's can be expected to be in the same proportion as α_i^2's only if $\bar{X}_i^2$'s are proportional to α_i^2's; i.e. if the coeffecients of variation for the x-character are more or less equal for all the strata. If this condition is far from being fulfilled there is not much point in using the relative values of α_i^2's instead of those of σ_i^2's in (6.2.2) or (6.2.3) or similar such expressions.

No results analagous to optimum allocation of sample size of the type known in literature are applicable in our set-up. This is so because the total expected effective sample size μ gives

$$\pi_{it} = \frac{\mu\ X_{it}}{X}$$

for the i-th unit of t-th stratum so that the expected effective sample size μ_t for the t-th stratum automatically gets fixed by the relation

$$\mu_t = \sum_i \pi_{it} = \mu\ \frac{X_t}{X} \qquad \cdots (6.2.5)$$

where X_t is the size of the t-th stratum with the cost function

$$C = C_o + \Sigma C_t \mu_t \qquad \cdots (6.2.6)$$

the total cost thus turns out to be

$$C = C_o + \Sigma C_t . \mu . \frac{X_t}{X}$$

$$= C_o + \frac{\mu}{X} \sum_t X_t \mu_t \qquad \cdots (6.2.7)$$

being completely determined by C_o, C_t's and μ and X.

§ 6.3 Integration of surveys

Suppose it is intended to conduct two surveys on the same population totals of the characters, y and y', which need not be distinct. Let a πPS sample of expected effective size μ be need for the survey for y and a πPS sample of expected effective size μ' be needed for the other survey. Let the auxilary characters to be used for the surveys be x and x' respectively, and these also need not be distinct. It is intended to draw two samples for the two surveys such that the number of common units is maximized. Let π_i be the inclusion probability required for U_i for the survey for y and π_i' be the inclusion probability for the survey for y' for the same unit. $(1 \leq i \leq N)$. Clearly

$$\pi_i = \mu \cdot \frac{X_i}{X}$$

$$\text{and} \quad \pi_i' = \mu' \cdot \frac{X_i'}{X'} \qquad \ldots\ldots(6.3.2)$$

Analagously to Keyfitz's [23a] procedure for pps sampling, let

$$\partial_i = \min (\pi_i, \pi_i') \qquad \ldots(6.3.2)$$

Further, let

$$\left.\begin{aligned} \lambda_i &= \frac{\pi_i - \partial_i}{1 - \partial_i} \\ \text{and} \quad \lambda_i' &= \frac{\pi_i' - \partial_i}{1 - \partial_i} \end{aligned}\right\} \qquad \ldots(6.3.3)$$

A procedure that can be expected to give maximum expected effective number of common units is then given by choosing a sample s_1 with a πPS sampling scheme in which the inclusion probabilities are ∂_i's

given by (6.3.2) and earmark it for both the surveys. A further πPS sample s_2 is chosen with inclusion probabilities λ_i's given in (6.3.3) and is taken, along with s_1, for the survey for y. Similarly a sample s_3 is chosen by a πPS method with inclusion probabilities λ_i's given in (6.3.3) and is taken, along with s_1, for the survey for y'.

It is easy to verify that the required inclusion probabilities are in fact attained. The expected number of common distinct units is at least equal to

$$\sum_{1}^{N} \theta_i \qquad \ldots(6.3.4)$$

§ 6.4 Testing the validity of the basic model (2.3.3)

We have seen that when the model (2.3.3) holds good, then in a certain well defined sense, we have an optimum sampling strategy to estimate the population total Y. It is therefore of interest to know if the model (2.3.3) is valid for the population. One way of partly achieving this is to stratify the population in some way and test for the equality of the 'a's and σ^2's applicable to the different strata. A first step towards this is to estimate these parameters for the different strata, on the basis of the samples drawn therefrom.

Let

$$U_{k1}, U_{k2}, \cdots, U_{k,N_k} \qquad \cdots \qquad (6.4.1)$$

be the units of the kth stratum with variate values X_{ki} and Y_{ki} for the $\mathcal{X}$ and $\mathcal{Y}$ variables for U_{ki} $(1 \leq i \leq N_k)$. Let us assume a model like (2.3.3) for the kth stratum thus:

$$\left.\begin{aligned} E_{\partial_1}(Y_{ki} \mid X_{ki}) &= a_k \cdot X_{ki} \\ V_{\partial_1}(Y_{ki} \mid X_{ki}) &= \sigma_k^2 \cdot X_{ki}^2 \\ \text{and} \quad Cov_{\partial_1}(Y_{ki}, Y_{kj} \mid X_{ki}, X_{kj}) &= 0 \end{aligned}\right\} \quad (1 \leq i \neq j \leq N_k) \qquad (6.4.2)$$

Under (6.4.2), the usual least-square estimates of a_k and σ_k^2 are given by

$$\overset{*}{a}_k = \frac{1}{N_k} \sum_{i=1}^{N_k} \frac{Y_{ki}}{X_{ki}} = \frac{1}{N_k} \sum_i Z_{ki} = \bar{Z}_k \text{ say} \qquad (6.4.3)$$

and

$$\overset{*2}{\sigma}_k = \frac{1}{N_k - 1} \left(\sum_i Z_{ki}^2 - N_k \bar{Z}_k^2 \right) \qquad (6.4.4)$$

where $Z_{ki} = \frac{Y_{ki}}{X_{ki}}$.

These are the best estimates of a_k and σ_k^2 if the prior distribution ∂_1 under consideration is in fact a N_k-variate normal distribution.

These estimates of the parameters of the prior distribution are themselves parameters of the main population (6.4.1) and hence have to be estimated unbiasedly from the sample drawn. For an unbiased estimates of $\overset{*}{a}_k$ we have

$$\overset{**}{a}_k = \overset{*}{\bar{Z}}_k = \frac{1}{N_k} \sum_{i \in s} \frac{Z_{ki}}{\pi_{ki}} = \frac{1}{N_k} \sum_{i \in s} \frac{Y_{ki}}{X_{ki} \cdot \pi_{ki}} \qquad \ldots (6.4.5)$$

To obtain an unbiased estimate of $\overset{*}{\sigma}_k^2$ we observe that an unbiased estimate of population variance

$$\underset{\sim}{\sigma^2}(Y) = \frac{1}{N} \left(\sum Y_i^2 - N \bar{Y}^2 \right) \qquad (6.4.6)$$

is given by applying (2.4.1) for the two terms on the r.h.s. of (6.4.6), thus

$$\underset{\sim}{\overset{*}{\sigma}^2}(Y) = \frac{1}{N} \left(1 - \frac{1}{N}\right) \sum_{i \in s} \frac{Y_i^2}{\pi_i} - \frac{1}{N^2} \sum_{\substack{i \neq j \\ \in s}} \sum \frac{Y_i Y_j}{\pi_{ij}} \qquad \ldots (6.4.7)$$

Hence an unbiased estimate of $\overset{*}{\sigma}_k^2$ of (6.5.4) is given by

$$\overset{**}{\sigma}_k^2 = \frac{1}{N_k (N_k - 1)} \left\{ (N_k - 1) \sum_{i \in s} \frac{Z_{ki}^2}{\pi_{ki}} - \sum_{\substack{i \neq j \\ \in s}} \sum \frac{Z_{ki} Z_{kj}}{\pi_{kij}} \right\} \qquad (6.4.8)$$

(6.4.5) and (6.4.8) thus give the required estimates. These can be obtained for the various strata and can then be compared. Since

sampling is done independently in the various strata the variances of the estimates of contrasts like

$$(a_k - a_{k'}),\ (\sigma_k^2 - \sigma_{k'}^2), \quad k \neq k' \qquad (6.4.9)$$

can be obtained as

$$V\left(\overset{**}{a}_k\right) + V\left(\overset{**}{a}_{k'}\right)$$

and

$$V\left(\overset{**}{\sigma}{}^2_k\right) + V\left(\overset{**}{\sigma}{}^2_{k'}\right)$$

respectively. For the variance of, say, $\overset{**}{a}_k$, we have

$$V\left(\overset{**}{a}_k\right) = V_{\partial_1} E_u \left(\overset{**}{a}_k\right) + E_{\partial_1} V_u \left(\overset{**}{a}_k\right) \qquad \ldots \qquad (6.4.10)$$

where the suffixes u and ∂_1 denote that the corresponding expectations and variances are taken over the various samples drawn from the actual population and various actual populations that can be obtained from the super population ∂_1. For an unbiased estimate of

$$E_{\partial_1} V_u \left(\overset{**}{a}_k\right)$$

we have

$$V_u \left(\overset{**}{a}_k\right)$$

which itself can be unbiased estimated from the sample from the formula (4.2.3) giving the variance-estimator of Horvitz and Thompson estimator. Regarding the first term we have

$$V_{\partial_1} E_u \left(\overset{**}{a}_k\right) = V_{\partial_1} \left(\overset{*}{a}_k\right) = \frac{\sigma_k^2}{N_k} \sum_{i=1}^{N_k} X_{ki} = \sigma_k^2 \bar{X}_k / N_k \qquad \ldots (6.4.11)$$

where $\bar{X}_k$ is the mean of X in the kth stratum. An unbiased estimate

of (6.4.11) is given by (since the X_{ki} - values are completely known)

$$\overset{**}{\sigma}{}^{2}_{k} \quad \bar{X}'_{k} / N_{k}$$

For an unbiased estimate of (6.4.10) we thus have

$$\overset{**}{V}\left(\overset{**}{a}_{k}\right) = \bar{X}_{k} \overset{**}{\sigma}{}^{2}_{k} + \frac{1}{N_{k}^{2}} \sum_{i \in s} \frac{Y_{ki}^{2}}{X_{ki}^{2}} \frac{(1 - \pi_{ki})}{\pi_{ki}^{2}}$$

$$+ \sum_{\substack{i \neq j \\ \in s}} \sum \frac{Y_{ki}\, Y_{kj}}{X_{ki}\, X_{kj}} \frac{\pi_{kij} - \pi_{ki}\pi_{kj}}{\pi_{ki}\pi_{kj}\pi_{kij}}$$

$$= \bar{X}_{k} \overset{**}{\sigma}{}^{2}_{k} + \frac{1}{N^{2}} \left\{ \sum_{i} W_{ki}^{2} (1 - \pi_{ki}) + \sum_{i \neq j} \sum W_{ki}\, W_{kj} \cdot \frac{\pi_{kij} - \pi_{ki}\pi_{kj}}{\pi_{kij}} \right\} \qquad (6.4.12)$$

where $W_{ki} = \dfrac{Y_{ki}}{X_{ki}\, \pi_{ki}}$

Similarly $\overset{**}{V}\left(\overset{**}{\sigma}{}^{2}_{k}\right)$ can be found. Using these expressions one can obtain estimates of the variances of the estimates of contrasts of (6.4.9) and can then test for their significance assuming that the estimates of the contrasts are normally distributed. A pooled test to test for the significance of all the contrasts can also be constructed Should this reveal that there are no significant differences among a_k's and among σ_k^2's the model (2.3.3) can be taken to be correct enough. Otherwise the population has to be stratified to put homogeneous or adjacent units into the same stratum and the strata totals should be estimated by individual optimum strategies.

CHAPTER VII

SOME GENERAL ASPECTS OF ESTIMATION AND THE OPTIMALITY OF THE HORVITZ AND THOMPSON ESTIMATOR

§ 7.1 Summary

This chapter is devoted to some general aspects of estimation pertinent to sampling from finite population, with special reference to the estimation of population totals and means. The aim is to approach in a unified way for the search of 'best' (in some well defined sense explained in § 7.2) sampling strategy. Towards this a search is first made towards finding the 'best' estimator of a given parametric function in the class of all estimators defined over a given design D. The fundamental result in this direction is given by theorem (7.5.1) of § 7.5 where a new criterion of hyper admissibility of an estimator is introduced and the bestness of the Horvitz and thompson estimator Y^{*}_{HT} established.

In § 7.2 the problem of estimation is posed and a unified approach is made towards a complete class of estimators of a parametric function in the class of all estimators defined over a given design. A simple proof is given in lemma (7.2.1) that the criterion of admissibility with respect to mean square error loss function does not exclude biased estimators of Y and then unbiasedness also is added in the list of criteria. Sufficiency of the order statistic in the special sense relevent to sampling from finite population is explained and its minimality pointed out in a simple way. That it is not a complete sufficient statistic is also pointed. The existence of

non-linear unbiased estimators of Y, and of non-homogeneous linear unbiased estimators of Y, all of which are functions of the minimal sufficient statistic alone, are shown by (7.3.17) and (7.3.18) for a large class of designs. The form of a general polynomial unbiased estimator of Y, is then given by (7.3.20). The inadequacy of the two criteria of unbiasedness and admissibility (with respect to mean square error) is pointed out.

In § 7.3 we derive necessary and sufficient conditions to be satisfied by a given sampling design for the unbiased estimability of a parametric function which is a polynomial in $Y_1, \ldots, Y_N$. While explicit proof for the case of linear parametric functions is well known we are not aware of formal proofs for higher order polynomials. In particular we deduce the conditions for the estimability of the variance of a general homogeneous linear unbiased estimator (g.h.l.u.e., for brevity) of $Y = \sum_1^N Y_i$.

In § 7.4 we point out that Godambe's result concerning the non-existence for any given design, of a uniformly minimum variance (i.e. for all values of $\underset{\sim}{Y}$) unbiased estimator (u.m.v.u.e., for brevity) of Y in the class L_u of all g.h.l,u.e's of Y, is not without some exceptions. In theorem (7.4.1) we completely characterise these exceptions, which we call the 'unicluster' sampling designs and prove that the best estimator (i.e., u.m.v.u.e.) in all these cases is the corresponding Horvitz and Thompson estimator, $\overset{*}{Y}_{HT}$.

In § 7.5 we introduce the criterion of hyperadmissibility of an estimator which is stronger than admissibility. The practical implications of this criterion are pointed out in terms of estimating not only $Y = \sum_{1}^{N} Y_i$ but all linear parametric functions $\sum_{1}^{N} L_i Y_i$ and in particular, all sub-population totals. The important off-shoot of this new criterion is that it not only excludes all but purely homogeneous linear unbiased estimators of Y, but in fact excludes all but one in the class of all unbiased estimators of Y which are polynomials in their arguments. This sole surviving estimator which can therefore be called 'the best estimator' under the criteria of unbiasedness and hyper admissibility, is the corresponding Horvitz and Thompson estimator $\overset{*}{Y}_{HT}$.

In § 7.6 we prove that, to estimate parameters θ that depend on the units of a proper subpopulation alone, corresponding to any given design D and an unbiased estimator T of θ, we can construct a restricted design D' and an estimator T' of θ such that $V(T') \leq V(T)$ for all $\underset{\sim}{Y}$. We point out that though the strategy $H'(D',T')$ has less variance than $H(D,T)$, it is costlier than H. In practice a balance has to be struck depending on cost considerations.

§ 7.2. Some aspects of estimation.

We shall now turn towards a unified approach to the problem of estimation discussed earlier in Chapter II.

Given a parametric function $\underset{\sim}{\theta}(Y)$, the problem of estimation of $\underset{\sim}{\theta}(Y)$ on the basis of a strategy $H(S,P,T)$ has to be approached through two fundamental criteria, viz, bounded cost and minimum loss. A commonly advocated cost function is of the form

$$C = A\,\mu_s + B$$

where A and B are known parameters and μ_s is the effective sample size. The expected cost of a strategy $H(S,P,T)$ is given by

$$C(H) = A\,\mu + B \qquad (7.2.1)$$

where

$$\mu = \sum_{s \in S} \mu_s P_s$$

A commonly advocated loss function is the mean square error defined by

$$M(H) = E(T-\theta)^2 = \sum_{s \in S} T_s^2 P_s - 2\theta \sum_{s \in S} T_s P_s + \theta^2 \qquad (7.2.2)$$

It is not always possible except in some very special cases, to express the loss and cost in same units. The choice of the strategy then has to be made in the class of all strategies for which

$$C(H) \leq C_0 \qquad (7.2.3')$$

where C_0 is a given ceiling on the budget and for which (7.2.2) is least for all possible values of $\underset{\sim}{Y} = (Y_1, Y_2, \cdots, Y_N)$. The condition (7.2.3') can be evidently replaced by a stronger criterion viz.,

$$C(H) = C_0 \tag{7.2.3}$$

because corresponding to any strategy satisfying

$$C(H) < C_0$$

there is always the trivial strategy of taking more units in the sample and ignoring the information supplied by them and this strategy is as good as the original one with respect to the loss function considered. In the sequel we shall therefore take (7.2.3) representing a fixed cost, instead of (7.2.3) representing a bounded cost. When this is not possible then we choose an admissible strategy H satisfying (7.2.3) and for which there is no other strategy H' also satisfying (7.2.3) and for which

$$M(H') \leq M(H) \quad \text{for all } \underset{\sim}{Y} \tag{7.2.4}$$

strict inequality holding good for at least one value of $\underset{\sim}{Y}$. The class of all admissible strategies forms, as is well known, the minimal complete class of strategies. Based only on the two criteria of fixed cost and minimum loss there is nothing to choose between any two admissible strategies.

A search for a complete class of strategies can now be made in two steps. First we can search for a complete class of estimators all defined over a given design D satisfying $\mu = \mu_0$ so that (7.2.3) is satisfied and then we can compare these various complete classes of estimators by varying D over the class of all designs satisfying $\mu = \mu_0$ to get a complete class of complete classes of estimators which gives us a complete class of strategies. In the following sections we shall devote our attention to the first part of our problem and towards the end review the final position.

As we confine our attention usually to unbiased estimators only it is pertinent to enquire whether they form a complete class. Though it is known that in general this is not so, we give a simple proof applicable to our problem that this is not so for the special case of $\Theta = Y$, and if we want to restrict ourselves to the class L_u of g.h.l.u.e's.

Lemma (7.2.1) : Except when $C_0 \geq AN + B$ (c.f. (7.2.1) and (7.2.3) there do exist biased strategies for which there does not exist a uniformly superior g.h.l.u.e..

Proof: Consider a strategy H_0 (S_0, P_0, T_0) satisfying (7.2.3) and in which the estimator is given by

$$T_0 = a \qquad (7.2.5)$$

where a is any given constant not equal to zero. If possible let there exist a strategy H(S, P, T) satisfying (7.2.3), where T is a

member of the class L_u of g.h.l.u.c's defined over $D(S.P)$ and which is uniformly superior to H_o. Since (7.2.3) is satisfied for both the strategies we need only verify (7.2.4). We have

$$M(H_o) = \sum_{s_o \in S_o} (a - Y)^2 P_{s_o} = (a - Y)^2 \qquad (7.2.6)$$

so that $M(H_o) = 0$ forall $\underset{\sim}{Y}$ forwhich $Y = \sum_1^N Y_i = a$, the given constant. Further

$$M(H_1) = \sum_{s \in S} T_s^2 P_s - Y^2$$

$$= \sum_{\lambda=1}^{N} \left\{ \sum_{s \supset \lambda} \beta_{s\lambda}^2 P_s - 1 \right\} Y_\lambda^2 + \sum_{\lambda \neq \lambda'}^{N} \sum_1 \left\{ \sum_{s \supset \lambda\lambda'} \beta_{s\lambda} \beta_{s\lambda'} P_s - 1 \right\} Y_\lambda Y_{\lambda'} \qquad (7.2.7)$$

where

$$T_s = \sum_{\lambda \in s} \beta_{s\lambda} Y_\lambda \qquad (7.2.8)$$

Since

$$0 \leq M(H_1) \leq M(H_o) \qquad \text{for all } \underset{\sim}{Y}$$

it follows that

$$M(H_1) = 0 \qquad \text{for all } \underset{\sim}{Y} \text{ such that } \sum_1^N Y_i = a \qquad (7.2.9)$$

and hence that $M(H_1)$ attains its minimum over $\underset{\sim}{Y}$ at all values $\underset{\sim}{Y}$ for which $\sum_1^N Y_i = a$. Minimising (7.2.7) over $\underset{\sim}{Y}$ we obtain the minimising equations to be

$$b_{\lambda 1} Y_1 + b_{\lambda 2} Y_2 + \cdots + b_{\lambda N} Y_N = 0 \quad (\lambda = 1, 2, \ldots N) \qquad (7.2.10)$$

where

$$b_{\lambda\lambda} = \sum_{s \supset \lambda} \beta^2_{s\lambda} P_s - 1$$

and

$$b_{\lambda\lambda'} = \sum_{s \supset \lambda\lambda'} \beta_{s\lambda}\beta_{s\lambda'} P_s - 1 \qquad \text{for} \quad \lambda \neq \lambda'$$

from (7.2.9) it followsthat the system of equations (7.2.10) should be satisfied by all $\underset{\sim}{Y}$ such that

$$\sum_{1}^{N} Y_i = a \qquad (7.2.11)$$

and hence in particular for

$$\underset{\sim}{Y}^{(1)} = (a, 0, 0, \ldots 0)$$

which implies from (7.2.10) that

$$b_{11} = \sum_{s \supset 1} \beta^2_{s1} P_s - 1 = 0 \qquad (7.2.12)$$

since $a \neq 0$. But from the unbiasedness of (7.2.8) it can be verified that

$$\sum_{s \supset 1} \beta^2_{s1} P_s - 1 \quad \geq \frac{1}{\pi_1} - 1$$

so that (7.2.12) is satisfied only if $\pi_1 = 1$. Similarly arguing for sets of values of $\underset{\sim}{Y}$ given by

$$\underset{\sim}{Y}^{(i)} = (0, 0, \ldots, 0, a, 0, \ldots 0) \qquad (i = 2, 3, \ldots, N)$$

where the non zero element is in the ith position and observing that all these sets satisfy the condition (7.2.11) it follows that it is necessary that

$$\pi_\lambda = 1 \qquad \text{for} \qquad \lambda = 1, 2, \ldots, N \tag{7.2.13}$$

This implies that the sampling design should consist of only such samples which contain every unit of the population, with probability 1. This is possible if and only if

$$C_0 \geq AN + B \tag{7.2.14}$$

where A, B, and C_0 are as defined by (7.2.1) and (7.2.3). It is clear that in this trivial situation we can estimate Y unbiasedly with zero variance so that all biased estimators are inadmissible.

This completes the proof of the lemma.

Since a situation where (7.2.14) holds good is not of interest from sampling point of view we thus see that we can not exclude biased estimators of Y from considerations of bounded cost and minimum loss alone, at any rate **not in preference to g.h.l.u.e.'s.**

We shall thus accept unbiasedness also as a criterion for the estimation of a parametric function.

In the case of sampling from a finite population $\mathcal{U}$ of N distinguishable units the sampling design D(S, P) is at the disposal of the statistician. As remarked in § 2.1 the set S of samples can be taken to be set $\Psi(\mathcal{U})$ of all finite sequences of elements from the

population . The probability measure P then is at our choice so that the probabilities attached to the different samples are completely known in advance, in contrast to sampling from infinite theoretical populations where the probabilities are known in terms of some unknown parameters. However, the distribution of the sampled observations on a real valued variate y depends not only on the probabilities of various sampled units but also on what points of R^N, the N-dimensional Euclidean space, these probability masses are situated. For a sample

$$s = (U_{i_1}, U_{i_2}, \cdots, U_{i_{n_s}}) \qquad (7.2.15)$$

with attached probability P_s this probability mass is attached to t the point

$$\underset{\sim}{Y}(s) = (Y_{i_1}, Y_{i_2}, \cdots, Y_{i_{n_s}}) \qquad (7.2.16)$$

of R^N. Since this point is determined by $\underset{\sim}{Y}$ we see that the distribution of the sampled observations thus depends on $\underset{\sim}{Y}$ which acts as the parameter. If then a sample s is taken and the vector $\underset{\sim}{Y}(s)$ is observed, for any other sample s' such that

$$s \sim s'$$

the corresponding $\underset{\sim}{Y}(s')$ is fully known and is independent of $\underset{\sim}{Y}$. We can thus say that given the order statistic i.e. the unordered set

of distinct units of a sample 's' and the corresponding y-values the condiditional distribution of $\underset{\sim}{Y}(s)$ is independent of the parameter $\underset{\sim}{Y}$ and hence in any sampling design the order statistic is a sufficient statistic for the family of distributions defined for varying values of the parameter $\underset{\sim}{Y}$ in R^N.

This result is well known from the results of Basu [1]. It can be easily seen that the order statistic is also the minimal sufficient statistic in this case. To see this let T be any sufficient statistic and consider the induced partition of S into equivalent samples. This consists of a partition of S which is a countable collection thus:

$$S = \bigcup_{m=1}^{\infty} A_m$$

where the atoms A_m are collections of samples such that given that the sample belongs to a particular atom A_m, the conditional distribution of the sampled observations is independent of $\underset{\sim}{Y}$. It is evident that for this to be true any two samples in the same atom must have the same set of distinct units of the population for otherwise if s and s' two members of an atom such that s contains a unit not belonging to s' then thelocation of the probability mass P_s in R^N can not be determined from that of $P_{s'}$. It follows then that the order statistic is the minimal sufficient statistic in this case. However it is not in general a complete statistic. To see this consider the simple case of the sampling design obtained by taking a simple random sample of size 2 without replacement from a population of 3 units. The possible samples are

$$s_1 = (U_1, U_2), \qquad s_4 = (U_2, U_1)$$
$$s_2 = (U_2, U_3), \qquad s_5 = (U_3, U_2)$$
$$s_3 = (U_3, U_1), \quad \text{and} \quad s_6 = (U_1, U_3)$$

Consider the statistic Z defined by

$$Z_{s_1} = Z_{s_4} = Y_1 - Y_2$$
$$Z_{s_2} = Z_{s_5} = Y_2 - Y_3$$
$$Z_{s_3} = Z_{s_6} = Y_3 - Y_1$$

clearly Z isafunction of the order statistic, $Z \not\equiv 0$, and

$$E(Z) = 0$$

which shows that the order statistic is not complete.

Since attention need be concentrated only on estimators which are functions of the minimal sufficient statistic it is of interest to examine whether the class of all such estimators is sufficiently narrow. We shall indicate below that this is not the case.

To estimate a linear parametric function like say $Y = \sum_1^N Y_i$ it is of interest to note that in general there do exist non linear unbiased estimators of Y which are functions of the minimal sufficient statistic. To see this, consider the example of the previous paragraph and let T be any unbiased estimator of Y which is a function of the order statistic. Also let T^*be defined as

$$T^*_{s_1} = T^*_{s_4} = Y_1^2 - Y_2^2$$

$$T^*_{s_2} = T^*_{s_5} = Y_2^2 - Y_5^2$$

and $$T^*_{s_3} = T^*_{s_6} = Y_3^2 - Y_1^2$$

clearly T^* is a zero function (i.e. one with expectation zero for all values of the parameter) and depends only on the order statistic and $T^* \not\equiv 0$ so that $(T + T^*)$ is a non-linear estimator of Y which is a function of the minimal sufficient statistic. The existence of such non-linear estimators can easily be demonstrated for a design in which Y is estimable and for which there are at least two samples s_1 and s_2 with non-zero probabilities such that s_1 contains a unit not contained in s_2. (Designsin which this conditionis not satisfied are the unicluster designs defined in §§ 7.4). In this case if T is any linear unbiased estimator of Y which depends on the order statistic alone and if U_λ is a unit in s_1 which is not in s_2 then the estimator T' defined by

$$\left.\begin{aligned} T'_s &= T_s \quad \text{if} \quad s \neq s_1,\ s_2 \\ T'_{s_1} &= T_{s_1} + \frac{Y_\lambda^2}{P_{s_1}} \\ T'_{s_2} &= T_{s_2} - \frac{Y_\lambda^2}{P_{s_2}} \end{aligned}\right\} \qquad (7.2.17)$$

is a non-linear unbiased estimator of Y and T' also is a function of the order statistic alone. It will be shown later in theorem (7.4.2)

that for a unicluster design there does not exist any non-linear unbiased estimator which is a function of the order statistic.

Similarly it is easy to see that there exist non-homogeneous linear estimators of homogeneous linear parametric functions. For, if T is a h.l.u.e. of $\sum_{1}^{N} L_i Y_i$ then any T' given by

$$T'_s = K_s + T \qquad (7.2.18)$$

where K_s is a random variable not dependent on $\underset{\sim}{Y}$) such that

$$E(K_s) = 0$$

is one such estimator. In particular such estimators exist for unicluster sampling designs also. In fact it will shown later in theorem (7.4.2) that all estimators of the form (7.2.18) are admissible in this case.

In general it can be easily shown that if T is an unbiased estimator of Y say, and is a polynomial of degree $\leq r$ in the arguments, then T can be written as

$$T = T_0 + T_1 + T_2 + \cdots + T_r \qquad (7.2.19)$$

where T_i is a homogeneous polynomial of degree i in the y-values occurring in the sample, and further

$$\left.\begin{aligned} E(T_i) &= 0 \qquad \text{for } i \neq 1 \\ \text{and } E(T_1) &= Y \end{aligned}\right\} \qquad (7.2.20)$$

This is so because

$$Y = E(T) = E(T_0) + E(T_1) + \cdots + E(T_r)$$

$$= Q_0 + Q_1 + \cdots + Q_r$$

where Q_i is a homogeneous polynomial of degree i, and the result follows from the fact that the above is an identity in $\underset{\sim}{Y}$.

One argument advanced offer for the exclusion of all but homogeneous linear estimators of Y is based on the principle of units of measurement. If, as it happens in practice, the Y_i's are the values of a particular character on the units of the population then all are measured in the same units and Y is also measured in some units so that it is not meaningful to estimate Y in terms of some other units of measurement. Moreover an estimator in which the Y values occur with different exponents can not be interpreted in terms of physical quantities. However, this is purely an argument in terms of physical units and once the units are fixed so that every unit in the population has a specific $\mathcal{Y}$- value the problem should be treated purely as a mathematical one.

The author is unable to exclude all but the g.h.l.u.c's of Y with just the criteria of bounded cost, unbiasedness and admissibility with respect to the mean square error loss function. He is not aware of any further mathematical criteria put forward by earlier authors which achieve this purpose. In § 7.5 we introduce the criterian of 'hyper admissibility' which is stronger than admissibility but weaker than that of uniform minimum variance. This excludes not only all but homogeneous linear estimators of Y but in fact gives rise to a unique estimator of Y.

§ 7.3 Necessary and sufficient conditions for the estimability of parametric functions which are polynomials in the arguments $Y_1, Y_2, \cdots, Y_N$.

Let $D = D(S, P)$ be a given design and let

$$F_i = C + \sum_{i=1}^{N} L_i Y_i \qquad (7.3.1)$$

be a linear p.f., where C and L_i's are constants. We then prove Lemma (7.3.1) :- A set of necessary and sufficient conditions for the estimability of F_i in D is that

$$\pi_i > 0 \quad \text{if} \quad L_i \neq 0 \qquad (1 \leq i \leq N) \qquad (7.3.2)$$

Proof :- When (7.3.2) are satisfied we have

$$\overset{*}{F}_i = C + \sum_{i \in s} L_i \frac{Y_i}{\pi_i}$$

(where $\frac{L_i Y_i}{\pi_i}$ is to be set equal to zero when $L_i = 0$) as an unbiased estimator of (7.3.1) so that the conditions (7.3.2) are sufficient. To prove the necessity part let T be a statistic defined over D and which is unbiased for F_i. Then

$$\sum_{s \in D} T_s P_s = C + \sum_{1}^{N} L_i Y_i, \quad \text{for all } \underset{\sim}{Y} \qquad (7.3.3)$$

If possible let there exist a λ such that $L_\lambda \neq 0$ but $\pi_\lambda = 0$

Then since a statistic (c.f. § 2.1) is a function of the variate values of only those units that are in the sample and since $\pi_\lambda = 0$ so that there is no $s \in D$ containing U_λ, it follows that the l.h.s. of (7.3.3) is independent of Y_λ. But since $L_\lambda \neq 0$, the r.h.s. of (7.3.3) is not independent of Y_λ and hence a contradiction.

Hence the lemma.

Let now

$$G_q = \sum_i \sum_j q_{ij} Y_i Y_j + \sum_i L_i Y_i + c \qquad (7.3.4)$$

be a quadratic p.f., where c, L_i's and q_{ij}'s are constants. Then we prove

Lemma (7.3.2) :- A set of necessary and sufficient conditions that (7.1.4) be estimable in D is that

$$\left.\begin{aligned} &\pi_\lambda > 0 \quad \text{if} \quad L_\lambda \neq 0 \quad \text{and/or} \quad q_{\lambda\lambda} \neq 0 \\ &\pi_{ij} > 0 \quad \text{if} \quad q_{ij} \neq 0 \end{aligned}\right\} (1 \leq i \neq j \leq N) \qquad (7.3.5)$$

Proof: If (7.3.5) are satisfied, clearly

$$\overset{*}{G}_q = c + \sum_{\substack{i \in s \\ L_i \neq 0}} L_i \frac{Y_i}{\pi_i} + \sum_{\substack{i \\ q_{ij} \neq 0}} \sum_j q_{ij} \frac{Y_i Y_j}{\pi_{ij}} \qquad (7.3.6)$$

is an unbiased estimator of (7.3.4) so that the conditions are sufficient To prove the necessity part, let M be a statistic unbiased for G_q. Then

$$\sum_{s \in D} M_s P_s = \mathfrak{G}_q = \sum_i \sum_j q_{ij} Y_i Y_j + \sum_i L_i Y_i + c, \text{ for all } \underset{\sim}{Y} \qquad (7.3.7)$$

Let there exist a λ such that

$$L_\lambda \neq 0 \quad \text{and/or} \quad q_{\lambda\lambda} \neq 0$$

and

$$\pi_\lambda = 0.$$

Then the r.h.s. contains the term

$$q_{\lambda\lambda} Y_\lambda^2 + L_\lambda Y_\lambda$$

and thus depends on Y_λ where as the l.h.s. is independent of Y_λ since there is no sample s containing U_λ and since a statistic can be a function of the variate values of only those units that occur in the sample. This is a contradiction and hence the first set of conditions of (7.3.5) are necessary. To prove that the second set is also necessary if possible let there exist a pair of integers i_o and j_o such that

$$\pi_{i_o j_o} = 0 \quad \text{and} \quad q_{i_o j_o} \neq 0,$$

and let there exist an unbiased estimator, M, of G_q, so that

$$\sum_{s \in S} M_s (Y_1, Y_2, \cdots, Y_N) P_s = G_q, \quad \text{for all } \underset{\sim}{Y}, \qquad (7.3.8)$$

where for a given s, M_s is a function of the $\mathcal{Y}$-values of units of s only. The sum on the l.h.s. of (7.3.8) can be written as

$$\sum_{s_1 \in S_1} M_{s_1} P_{s_1} + \sum_{s_2 \in S_2} M_{s_2} P_{s_2} + \sum_{s_3 \in S_3} M_{s_3} P_{s_3} \qquad (7.3.9)$$

where

$$\left.\begin{aligned} S_1 &= \{s_1 : s_1 \in S,\ s_1 \supset U_{i_o} \text{ and } s_1 \not\supset U_{j_o}\} \\ S_2 &= \{s_2 : s_2 \in S,\ s_2 \supset U_{j_o} \text{ and } s_2 \not\supset U_{i_o}\} \\ \text{and} \quad S_3 &= \{s_3 : s_3 \in S,\ s_3 \not\supset U_{i_o},\ s_3 \not\supset U_{j_o}\} \end{aligned}\right\} . \qquad (7.3.10)$$

Since $\pi_{i_o j_o} = 0$, there does not exist an s (with $P_s > 0$) which contains both U_{i_o} and U_{j_o} so that

$$S = S_1 \cup S_2 \cup S_3, \qquad (S_k \cap S_{k'} = 0, \quad k \neq k')$$

and hence (7.3.9). Setting (7.3.9) in (7.3.8) and recalling that (7.3.8) is an identity in $\underset{\sim}{Y}$. We find by letting $Y_k = 0$ for $k \neq i_o$ and $\neq j_o$,

$$\sum_{s_1 \in S_1} M_{s_1}(Y_{i_o})\, P_{s_1} + \sum_{s_2 \in S_2} M_{s_2}(Y_{j_o})\, P_{s_2} + \sum_{s_3 \in S_3} M_{s_3} P_{s_3}$$

$$= q_{i_o i_o} Y_{i_o}^2 + 2\, q_{i_o j_o} Y_{i_o} Y_{j_o} + q_{j_o j_o} Y_{j_o}^2$$

$$+ L_{i_o} Y_{i_o} + L_{j_o} Y_{j_o} + L \qquad (7.3.11)$$

Since all Y_k's except Y_{i_o} and Y_{j_o} are set equal to constants, from (7.3.10) we see that M_{s_1} is a function of Y_{i_o}, M_{s_2} is a function of Y_{j_o} and M_{s_3} is a constant, so that (7.3.11) can be written as

$$\phi_1(Y_{i_o}) + \phi_2(Y_{j_o}) + \phi_3 = q_{i_o i_o} Y_{i_o}^2 + \ldots + L$$

which can be written as

$$\alpha_1(Y_{i_o}) + \alpha_2(Y_{j_o}) + \alpha_3 = 2 q_{i_o j_o} Y_{i_o} Y_{j_o} \qquad \ldots \qquad (7.3.12)$$

where α_1 and α_2 are functions of Y_{i_o} and Y_{j_o} respectively and

α_3 is a constant. Since (7.3.12) is to be identity in Y_{i_o} and Y_{j_o} a relation like that is clearly impossible since $q_{i_o j_o} \neq 0$, and thus we have a contradiction, which proves that the second set of conditions is also necessary.

This proves the lemma.

Corollary: The variance of an unbiased estimator of Y is not estimable unless $\pi_{ij} > 0$ for all i and j. For, if T is an unbiased estimator of Y, we have

$$V(T) = \sum_{s \in S} T_s^2 P_s - Y^2 \qquad (7.3.13)$$

If now $\pi_{i_o j_o} = 0$, there is no sample containing both U_{i_o} and U_{j_o} so that

$$\sum_{s \in S} T_s^2 P_s$$

can not contain the term $Y_{i_o} Y_{j_o}$. The coefficient of $Y_{i_o} Y_{j_o}$ in (7.3.13) is then equal to -2 which is not zero. From the above lemma it follows that $V(T)$ is not estimable in the given design.

The above corollary is assumed quite freely in the literature but we are not aware of a formal proof of the same.

The results of lemma (7.3.1) and (7.3.2), obviously, can be directly generalised to the case of estimation of higher degree polynomials.

§ 7.4. The existence of umvue in unicluster sampling designs

We have remarked in § 2.2 that theorem (2.2.2) of Godambe regarding the non-existence for any given design, of a uniformly minimum variance unbiased estimator umvue in the class L_u of all general homogeneous linear unbiased estimators (g.h. l.u.e) of the population total Y has some exceptions. We shall completely characterise these exceptions in the following.

Theorem (7.4.1) :- The class of all sampling designs D_{uc} (S,P) for which

$$\left.\begin{aligned} &\text{i)} \quad \pi_i > 0 \qquad (1 \leq i \leq N) \\ &\text{and} \\ &\text{ii)} \quad \text{for any two samples } s_i \text{ and } s_j \text{ of } D_{uc}, \text{ either both contain the same units of the population or they do not have any common units.} \end{aligned}\right\} \quad (7.4.1)$$

are precisely the class of designs admitting a umvue in the class L_u of all g.h.l.u.e's of Y. The best estimator of Y in L_u in any such design is the corresponding Horvitz and Thompson estimator viz.,

$$\overset{*}{Y}_{HT} = \sum_{\lambda \in s} \frac{Y_\lambda}{\pi_\lambda}$$

Proof: In § 2.1. We defined a g.h.l.u.c. over a general design D(S, P) by

$$T_s = \sum_{\lambda=1}^{N} \beta_{s\lambda} Y_\lambda \qquad (7.4.2)$$

where $\beta_{s\lambda} = 0$ if $U_\lambda \notin s$.

The conditions for the unbiasedness for Y of (7.4.2) can be seen to be

$$\sum_{s \in S} \beta_{s\lambda} P_s = 1 \qquad \text{for } 1 \leq \lambda \leq N \qquad \ldots (7.4.3)$$

For the variance of T we have

$$V(T) = \sum_{s \in S} T_s^2 P_s - Y^2 \qquad \ldots (7.4.4)$$

If there exists a unvue in the class L_u of g.h.l.u.e's, let it be

$$T' = \sum_{\lambda} \beta_{s\lambda} Y_\lambda \qquad (7.4.5)$$

T' if it exists can be obtained by minimising (7.4.4) over all possible double sequences of coefficients, $\{\{\beta_{s\lambda}\}\}$ subject only to (7.4.3). Introducing the Lagrangian multipliers $\alpha_1, \alpha_2, \ldots, \alpha_N$ (which are not functions of $\beta_{s\lambda}$'s, but which may depend on Y_λ's) we seek to minimise

$$\phi = \sum_{s \in S} T_s^2 P_s - Y^2 - \sum_{\lambda=1}^{N} \alpha_\lambda \left(\sum_s \beta_{s\lambda} P_s - 1 \right) \qquad (7.4.6)$$

with respect to $\{\{\beta_{s\lambda}\}\}$. Since $\beta_{s\lambda} = 0$ for $s \not\supset U_\lambda$. We need differentiate (7.4.6) with respect to the other β's only which gives the minimising equations to be

$$\frac{\partial}{\partial \beta_{s_o\lambda_o}} (T'_{s_o} P_{s_o}) - \alpha_{\lambda_o} P_{s_o} = 0 \qquad \text{for every } \lambda_o \text{ and every } s_o \supset U_{\lambda_o}$$

i.e. $2\, T'_{s_o} \cdot P_{s_o} \cdot Y_{\lambda_o} - \alpha_{\lambda_o} P_{s_o} = 0, \quad 1 \le \lambda_o \le N, \quad s_o \supset U_{\lambda_o}$

so that for any sample s_o forwhich $P_{s_o} > 0$, we have

$$T'_{s_o} = \frac{\alpha_{\lambda_o}}{2 Y_{\lambda_o}}, \quad 1 \le \lambda_o \le N, \quad s_o \supset U_{\lambda_o} \qquad (7.4.7)$$
$$P_{s_o} > 0.$$

The above condition then implies that for any two samples s_1 and s_2 having aunit in common and for which $P_{s_1} > 0$ and $P_{s_2} > 0$, we should have

$$T'_{s_1} = T'_{s_2}, \qquad \text{for all } \underset{\sim}{Y} \qquad (7.4.8)$$

Godambe then closes his argument saying that (7.4.8) is clearly possible. However we shall carry the argument further. If two samples s_1 and s_2 have a unit in common then (7.4.8) should hold good. But if one of them contains no unit belonging to the other then (7.4.8) is not to be true by the definition of a statistic. Hence a set of necessary conditions for the existence of a true is that any two samples of the design contain the same set of units or they do not have any unit in common, which are precisely the conditions (ii) of (7.4.1) from lemma (7.3.1) conditions (i) of (7.4.1) are necessary and sufficient for the existence of an unbiased estimator of Y. Thus (7.4.1) are a necessary set of conditions. To prove that they are sufficient we recall theorem (2.2.3) according to which we need search for the best estimator of Y only in the class L_u of all g.h.l.u.c's that depend neither on the number of repetitions nor on the order of occurrence of units in a sample.

For any two samples s and s' let us denote by

$$s \sim s'$$

the fact that s and s' contain the same units of the population but for the number of repetitions and orders of occurrence. A complete class of estimators in the class L_u of all g.h.u.c's is then given by estimators T for which

$$T_s \neq T_{s'}, \quad \text{for all} \quad s \sim s'$$

i.e.

$$\beta_{s\lambda} = \beta_{s'\lambda} \quad \text{for} \quad 1 \leq \lambda \leq N, \text{ whenever } s \sim s' \qquad \ldots\ldots (7.4.9)$$

From From (7.4.3) giving the conditions for unbiasedness and from (ii) of (7.4.1) we then have

$$1 = \sum_{s \in S} \beta_{s\lambda} P_s = \sum_{s \supset \lambda} \beta_{s\lambda} P_{s'} \quad \text{i.e.} \quad \beta_{s\lambda} = 0 \quad \text{for} \quad s \not\supset \lambda$$

$$= \sum_{s \sim s_o} \beta_{s\lambda} P_s \quad \text{for any} \quad s_o \supset \lambda$$

$$= \beta_{s_o\lambda} \sum_{s \sim s_o} P_s, \quad \text{from (7.4.9)}$$

$$= \beta_{s_o\lambda} \cdot \pi_\lambda \quad \text{from (i) of (7.4.1)}$$

so that for any $s_o \supset \lambda$,

$$\beta_{s_o\lambda} = \frac{1}{\pi_\lambda} \qquad \ldots\ldots (7.4.10)$$

This shows that our complete class of estimators contains just one estimator which is therefore the best in the class L_u. This completes the proof of the theorem.

Remark: - (1) It may be noted that if a unicluster design contains at least two distinct samples s and s' such that

$$s \sim s'$$

then there are more than one possible unbiased estimators of Y while in the contrary case there is just one unbiased estimator of Y viz $\overset{*}{Y}_{HT}$. This is evident from (7.4.3) or more intuitively by recognising that if (7.4.1) is satisfied and for no two samples s and s' of D_{uc} the relation

$$s \sim s'$$

is true, then the original population of units can be replaced by a new population of clusters which are the samples of the original design, and the corresponding sampling design is the one obtained by the choice of a single cluster with varying probabilities. It is evident that when a sample of just one unit is drawn from a population, the only g.h.l.u.c of Y is

$$\overset{*}{Y} = \frac{y}{p}$$

where y is the sampled value and p is the probability of drawing the unit.

The name 'unicluster designs' is given to the designs satisfying (7.4.1), because, ignoring the repetitions and orders of units in a sample these designs are essentially those generated by grouping the

units of the population into clusters in some arbitrary way and then selecting one of these clusters with varying probabilities if necessary. For unicluster sampling designs we can in fact prove slightly stronger results given by

Theorem (7.4.2): If D_{uc} is a unicluster design then any polynomial unbiased estimator T of Y which is a function of the order statistic alone is necessarily of the form

where

$$\left.\begin{aligned} T_s &= K_s + \sum_{\lambda \in s} \frac{Y_\lambda}{\pi_\lambda} \\ \sum_{s \in S} K_s P_s &= 0 \end{aligned}\right\} \qquad (7.4.11)$$

Further any estimator T of the form (7.4.11) is admissible.

Proof: - Let T be an unbiased estimator of Y which is a polynomial in its arguments, and which depends only on the order statistic. Hence we can consider the sampling design as one consisting of a finite number (utmost N in this case) of unordered samples so that there is a finite upper bound r for the degree of T_s for varying samples s. Let then

$$T = T_0 + T_1 + T_2 + \cdots + T_r$$

so that each T_i is a function of the order statistic alone. Let now

$$T_2 \neq 0$$

so that there is a sample s, consisting of units $U_{i_1}, U_{i_2}, \ldots, U_{i_t}$ for which

$$T_{2,s} \not\equiv 0$$

in which case we have

$$\beta_{sii} \not\equiv 0 \quad \text{for some} \quad U_i \in s$$

or

$$\beta_{sij} \not\equiv 0 \quad \text{for some} \quad U_i, U_j \in s$$

But since the design is a unicluster design and we have started with unordered samples only no other sample contains a unit contained in s, so that

$$E(T_2) \not\equiv 0$$

the l.h.s. containing a non-zero term of the form $\beta_{ii} Y_i^2$ or $\beta_{ij} Y_i Y_j$. This contradicts (7.2.20), so that

$$T_2 \equiv 0$$

Similarly it can be shown that

$$T_i \equiv 0 \quad \text{for} \quad 2 \leq i \leq r$$

As already pointed in remark (1) at the end of theorem (7.4.1), and implicit in the proof of that theorem from the fact that T_1 depends on the order statistic alone it follows in this case that

$$T_{1,s} = \sum_{\lambda \in s} \frac{Y_\lambda}{\pi_\lambda}$$

$$= \frac{1}{P_s} \sum_{\lambda \in s} Y_\lambda = \frac{Z_s}{P_s} \qquad (7.4.12)$$

where Z_s is the cluster total, since U_λ can not belong to more than one unordered sample. This proves the first part of the theorem.

To prove the second part let T_1 and T_2 be two distinct unbiased estimators of Y of the form (7.4.11), so that from (7.4.12)

$$\left.\begin{aligned} T_{1,s} &= K_{1,s} + \frac{Z_s}{P_s} \\ \text{and} \quad T_{2,s} &= K_{2,s} + \frac{Z_s}{P_s} \\ \text{where} \quad \sum_{s \in S} K_{1,s} P_s &= \sum_{s \in S} K_{2,s} P_s = 0 \end{aligned}\right\} \qquad (7.4.13)$$

We have

$$V(T_1) - V(T_2) = E(T_1^2) - E(T_2^2)$$

$$= \sum_{s \in S} (K_{1,s}^2 - K_{2,s}^2) P_s + \sum_{s \in S} (K_{1,s} - K_{2,s}) Z_s \qquad (7.4.14)$$

which can be made positive or negative by a proper choice of Z_s's, i.e. Y_i's, since $(K_{1,s} - K_{2,s})$ is not equal to zero for all s. In fact since from (7.4.13) $(K_{1,s} - K_{2,s})$ is not always of the same sign, (7.4.14) can be made positive or negative by a proper choice of even positive values of Z_s's.

Thus every unbiased estimator of Y of the form (7.4.11) is admissible.

This completes the proof of the theorem.

§ 7.5 The criterion of HYPERADMISSIBILITY and the bestness of $\overset{*}{Y}_{HT}$, the Horvitz and Thompson estimator.

We recall that in a class $\mathcal{C}$ of unbiased estimators of a parametric function θ, an estimator T_0 is said to be admissible if for any other $T_1 \in \mathcal{C}$

$$M(T_0) < M(T_1)$$

at least one point $\underset{\sim}{Y}_0 (T_0, T_1)$ in R^N, this point depending possibly on T_0 and T_1.

An estimator which is admissible in a parameter space may not be admissible in a subset of this parameter space. We now introduce a stronger criterion in

Definition (7.5.1) :- An unbiased estimator T_0 of Y is said to be hyper admissible in a class $\mathcal{C}$ of unbiased estimators of Y, if given any other estimator $T_1 \in \mathcal{C}$, in every principal hyperplane of R^N there is at least one point at which

$$M(T_0) < M(T_1)$$

The principal hyper planes of R^N are those in which some of the co-ordinates are always zero, i.e. the co-ordinate axes, planes etc.

Before proceeding to prove our main result that follows this new criterion we shall point out the practical implications of this criterion. Consider an unbiased estimator $\hat{T}$ of Y which is admissible. If now we wish to have an unbiased estimator of the total

$$Y' = \sum_{k=1}^{M} Y_{i_k} \qquad (7.5.1)$$

of a specified sub-population $\mathcal{U}'$ consisting of the units

$$\mathcal{U}' = (U_{i_1}, U_{i_2}, \cdots, U_{i_M}) \qquad (7.5.2)$$

then the estimator T itself can be used for this purpose by setting

$$Y_\lambda = 0 \quad \text{for} \quad U_\lambda \notin \mathcal{U}' \qquad (7.5.3)$$

because under (7.5.3) the population total for the original population is the same as that for $\mathcal{U}'$. Let this new estimator be denoted by T'. Also the original sampling design $D(S, P)$ gives rise to a new sampling design $D'(S', P')$ obtained by supressing from every sample s of S those elements that do not belong to $\mathcal{U}'$ and keeping the probabilities for the various samples same as those of the corresponding samples of D. This may give rise to some null samples consisting of no units of the population $\mathcal{U}'$ and the estimator T' has to be set equal to zero to retain the condition of unbiasedness. The parameter space relevant for comparison of various estimators of Y', defined over D', is the one in which (7.5.3) is satisfied. T' asobtained above may not be admissible in the class of estimators defined over D' even if T is admissible in the class of estimators defined over D. If however T is hyperadmissible than for every sub-population $\mathcal{U}'$ of $\mathcal{U}$ the corresponding estimator T' is admissible in the corresponding class of estimators defined over D'.

In many practical problems dealing with sampling from finite population one often requires not only an estimator of Y but also of some

sub-populations of interest. If then a hyper admissible estimator is used, it can be used to estimate the total of any sub-population and will still give an admissible estimator of the relevant sub-population total. Hence it is convenient to use a hyper admissible estimator in all such situations so that one does not have to construct new admissible estimator every time one is interested in estimating the total of a sub-population.

We now prove the main result stemming from our new criterion.

Theorem (7.5.1) :- For any given design $D(S, P)$, which is not a unicluster design, in the class L_{pu} of all unbiased estimators of Y defined over D which are polynomials in their arguments there is just one which is hyper admissible and this is given by the Horvitz and Thompson's estimator, $\overset{*}{Y}_{HT}$.

Proof :- Since a hyper admissible estimator has to be an admissible estimator in the first instance, it has to be a function of the order statistic alone. Hence we can consider without loss of generality that S is the collection of all unordered subsets of $\mathcal{U}$, so that S contains a finite number of samples (2^N in all). Further we can omit all samples of this set, which are assigned zero probabilities.

Let now T be an unbiased estimator of Y which for every sample s of S is a polynomial in the $\mathcal{Y}$-values of the units occurring in s. Since there are a finite number of samples in S, there exists a finite upper bound over S for the degree of T. Let then T be a polynomial of degree r and let

$$T = T_0 + T_1 + T_2 + \cdots T_r \tag{7.5.4}$$

where T_m is a homogeneous polynomial of degree m. Let

$$\left.\begin{aligned} T_{0,s} &= K_s \\ T_{1,s} &= \sum_{\lambda \in s} \beta_{s\lambda} Y_\lambda \\ T_{2,s} &= \sum_{\lambda \in s} \beta_{s\lambda\lambda} Y_\lambda^2 + \sum\sum_{\lambda \neq \lambda'} \beta_{s\lambda\lambda'} Y_\lambda Y_{\lambda'} \\ T_{3,s} &= \sum_{\lambda \in s} \beta_{s\lambda\lambda\lambda} Y_\lambda^3 + \sum\sum_{\lambda \neq \lambda'} \beta_{s\lambda\lambda\lambda'} Y_\lambda^2 Y_{\lambda'} \\ &\quad + \sum\sum\sum_{\lambda \neq \lambda' \neq \lambda''} \beta_{s\lambda\lambda'\lambda''} Y_\lambda Y_{\lambda'} Y_{\lambda''} \end{aligned}\right\} \tag{7.5.5}$$

etc. From the unbiasedness of T it follows that

$$\left.\begin{aligned} E(T_1) &= Y \\ \text{and} \quad E(T_m) &= 0 \qquad \text{for } m \neq 1 \end{aligned}\right\} \tag{7.5.6}$$

so that

$$\left.\begin{aligned} &\sum_{s \in S} K_s P_s = 0 \\ &\sum_{s \supset \lambda} \beta_{s\lambda} P_s = 1 \qquad (1 \leq \lambda \leq N) \\ &\sum_{s \supset \lambda} \beta_{s\lambda\lambda} P_s = \sum_{s \supset \lambda, \lambda'} \beta_{s\lambda\lambda'} P_s = 0 \qquad (1 \leq \lambda \neq \lambda' \leq N) \\ &\sum_{s \supset \lambda} \beta_{s\lambda\lambda\lambda} P_s = \sum_{s \supset \lambda, \lambda'} \beta_{s\lambda\lambda\lambda'} P_s = \sum_{s \supset \lambda, \lambda', \lambda''} \beta_{s\lambda\lambda'\lambda''} P_s = 0 \\ &\qquad\qquad (1 \leq \lambda \neq \lambda' \neq \chi' \leq N) \end{aligned}\right\} \tag{7.5.7}$$

etc. Let T be hyyper admissible and consider now the sub-population consisting of the unit U_i alone, for a specific i. Setting $Y_\lambda = 0$ for $\lambda \neq i$ in (7.5.4) and (7.5.5) we obtain the corresponding unbiased estimator of Y_i to be

$$T'_s = K_s + \beta_{si} Y_i + \beta_{sii} Y_i^2 + \beta_{siii} Y_i^3 + \cdots + \beta_{si..i} Y_i^2$$

$$= \begin{cases} \text{if } U_i \in s \\ K_s \quad \text{if } U_i \notin s \end{cases} \quad \cdots (7.5.8)$$

Let

$$S = S_{(i)} \cup S^*_{(i)}$$

where $S_{(i)} = \{ s : s \in S, U_i \in s \}$

and $S^*_{(i)} = \{ s : s \in S, U_i \notin s \}$

The corresponding design $D'(S', P')$ consists of several samples (resulting from s's belonging to $S_{(i)}$) all containing just the unit U_i and several other null samples (resulting from s's belonging to $S^*_{(i)}$). Further the corresponding order statistic which is sufficient is equal to Y_i for all samples s's which result from the original samples of $S_{(i)}$ and is zero for the others. T'_s therefore can not be admissible in the class of unbiased estimators defined over D' unless

$$T'_{s_1} = T'_{s_2} \qquad \text{for all } s_1, s_2 \in S_{(i)}$$

and

$$T'_{s_3} = T'_{s_4} \qquad \text{for all } s_3, s_4 \in S^*_{(i)}$$

Hence T' is admissible only if it is of the form

$$\left.\begin{aligned} T'_s &= K_{il} + \beta_i Y_i + \beta_{ii} Y_i^2 + \cdots + \beta_{ii\ldots i} Y_i^r \\ &\qquad\qquad \text{for all } s \in S_{(i)} \\ &= K_{i2} \qquad \text{for } s \in S^*_{(i)} \end{aligned}\right\} \qquad (7.5.9)$$

From (7.5.7) it follows that

$$K_{il}\pi_1 + K_{i2}(1 - \pi_1) = 0$$

$$\beta_i \pi_i = 1$$

$$\beta_{ii}\pi_i = \beta_{iii}\pi_i = \cdots = \beta_{ii\ldots i}\pi_i = 0$$

where

$$\pi_i = \sum_{s\ i} P_s$$

and since T is unbiased for Y, from lemma (7.2.1) we have $\pi_i > 0$ so that

$$K_s = \begin{cases} K_{il} & \text{if } s \in S_{(i)} \\ K_{i2} & \text{if } s \in S^*_{(i)} \end{cases} \qquad (7.5.10)$$

$$\beta_{si} = \frac{1}{\pi_i} \qquad \text{for all } s \quad U_i$$

and $$\beta_{sii} = \beta_{siii} = \cdots = \beta_{sii\ldots i} = 0$$

$$\text{for all } s \quad U_i$$

and arguing for all single element sub-populations (7.5.10) is true for $1 \leq i \leq N$.

If D is not a unicluster design it can be seen that there exist at least two units U_{i_o} and U_{j_o} such that

$$\pi_{i_o j_o} > 0$$

and

$$S_{(i_o)} \neq S_{(j_o)} \qquad (7.5.11)$$

and consider the decomposition of S thus:

$$S = (S_{(i_o)} \cap S_{(j_o)}) \cup (S_{(i_o)} \cap S^*_{(j_o)}) \cup (S^*_{(i_o)} \cap S_{(j_o)}) \cup (S^*_{(i_o)} \cap S^*_{(j_o)})$$
$$= S^{(1)} \cup S^{(2)} \cup S^{(3)} \cup S^{(4)}$$

say. From (7.4.10)

$$K_s = \begin{cases} K_{i_o 1} & \text{for} \quad s \in S^{(1)} \cup S^{(2)} \\ K_{i_o 2} & \text{for} \quad s \in S^{(3)} \cup S^{(4)} \\ K_{j_o 1} & \text{for} \quad s \in S^{(1)} \cup S^{(3)} \\ K_{j_o 2} & \text{for} \quad s \in S^{(2)} \cup S^{(4)} \end{cases} \qquad (7.5.12)$$

If

$$S^{(3)} \cup S^{(4)} = \emptyset$$

or

$$S^{(2)} \cup S^{(4)} = \emptyset$$

it follows directly from (7.5.12) that

$$K_s \equiv K \qquad \text{for all} \quad s \in S$$

Otherwise, from (7.5.11) we have

$$S^{(1)} \neq \emptyset$$

and

$$S^{(2)} \cup S^{(3)} \neq \emptyset$$

Since $S^{(1)} \neq \emptyset$, from (7.5.12) $K_{i_o 1} = K_{j_o 1}$ so that

$K_s = K_1$, say for all $s \in S^{(1)} \cup S^{(2)} \cup S^{(3)}$

and since $S^{(2)} \cup S^{(3)} \neq \phi$

$$(S^{(2)} \cup S^{(4)}) \cap (S^{(1)} \cup S^{(2)} \cup S^{(3)}) \neq \phi$$

or $$(S^{(3)} \cup S^{(4)}) \cap (S^{(1)} \cup S^{(2)} \cup S^{(3)}) \neq \phi$$

so that in either case from (7.5.12)

$$K_s = K_1 \quad \text{for all} \quad s \in S^{(4)} \text{ also.} \tag{7.5.13}$$

Thus under (7.5.11)

$$K_s = K \quad \text{for all} \quad s \in S$$

and from (7.5.7) follows that

$$K_s = 0 \quad \text{for all} \quad s \in S \tag{7.5.14}$$

if D is not a unicluster design.

We now consider the sub-population consisting of two units, say, U_{i_o} and U_{j_o} only. From (7.5.10) and (7.5.14), the estimator T' in this case reduces to

$$T'_s = \begin{cases} \dfrac{Y_{i_o}}{\pi_{i_o}} & \text{if } U_{i_o} \in s,\ U_{j_o} \notin s \\ \dfrac{Y_{i_o}}{\pi_{j_o}} & \text{if } U_{j_o} \in s,\ U_{i_o} \notin s \\ \dfrac{Y_{i_o}}{\pi_{i_o}} + \dfrac{Y_{j_o}}{\pi_{j_o}} + \beta_{si_oj_o} Y_{i_o} Y_{j_o} + T'(3)\,s & \text{if } U_{i_o} \in s \text{ and } U_{j_o} \in s \\ 0 & \text{if } U_{i_o} \notin s,\ U_{j_o} \notin s \end{cases} \tag{7.5.15}$$

where $T'_{(3)s}$ consists of terms of degree ≥ 3 in Y_{i_o} and Y_{j_o}. Again it can be easily seen that the estimator

$$T''_s = E(T'_s \mid s \supset U_{i_o} \text{ and } U_{j_o}) \quad \text{if } U_{i_o} \in s \text{ and } U_{j_o} \in s$$

$$= T'_s \quad \text{otherwise}$$

is uniformly better than T'_s unless it is identically equal to T'_s. Hence for T'_s to be admissible we should have

$$\beta_{si_oj_o} = \beta_{i_oj_o}$$

$$\beta_{si_oi_oj_o} = \beta_{i_oi_oj_o}, \quad \beta_{sj_oj_oi_o} = \beta_{j_oj_oi_o}$$

etc., for all samples s containing U_{i_o} and U_{j_o}. If there are no samples s (with $P_s > 0$) containing both U_{i_o} and U_{j_o} then all the above β's are zero each by definition, while if $\pi_{i_oj_o} > 0$ it follows from (7.5.7) that they are all identically equal to zero.

Higher order β coefficients can similarly be shown to be equal to zero if T is hyper admissible, by considering the appropriate sub-populations.

It follows that if D is not a unicluster design, then in the class of all polynomial unbiased estimators of Y there is just one viz $\overset{*}{Y}_{HT}$ which is possibly hyper admissible. That in fact it is hyper admissible follows from the fact that if T is the estimator $\overset{*}{Y}_{HT}$ of Y in D then for any sub-population T' is the corresponding estimator

$\overset{*}{Y}_{HT}'$ of Y' in the corresponding design D' and hence is admissible since it is known that for any sampling design the corresponding estimator $\overset{*}{Y}_{HT}$ is an admissible estimator of thepopulation total [12], [36].

This completes the proof of our theorem.

Remarks:-

1) If T is an unbiased estimator $Y = \sum_{1}^{N} Y_i$ then for every linear function

$$L\,(\underset{\sim}{Y}) = \sum_{1}^{N} L_i\, Y_i$$

the unbiased estimator T' of $L\,(\underset{\sim}{Y})$, obtained from T by replacing Y_i by $L_i\, Y_i$ $(1 \leq i \leq N)$ is admissible (for any $L\,(\underset{\sim}{Y})$) if and only if $T = \overset{*}{Y}_{HT}$. Thus $\overset{*}{Y}_{HT}$ admissibly estimates not only all sub-population totals but all linear parametric functions of Y, so that it can be used for the simultaneous estimation of several linear parametric functions.

2) It should be noted that the assumption that D is not a unicluster design is used only to prove that if T is hyper admissible then

$$T_o = 0$$

where T_o (c.f. 7.2.19) is the constant part (i.e. not depending on Y) of T.

3) We see from the proof of the above theorem that admissibility in each of the co-ordinate axes of R^N required that $T_1 = \overset{*}{Y}_{HT}$;

admissibility in each of the co-ordinate axes and co-ordinate planes required that $T_2 = 0$, admissibility further in each of the 3-dimensional co-ordinate hyper planes required that $T_3 = 0$, etc.

4) Though hyper admissibility in each of the principal hyper planes eliminated all but $\overset{*}{Y}_{HT}$ in the class of polynomial unbiased estimators of Y it can be shown that $\overset{*}{Y}_{HT}$ is admissible in a number of other hyper spaces. For example in every (N-1) dimensional hyperplane that cuts all the co-ordinate axes it can be seen that $\overset{*}{Y}_{HT}$ is admissible. For simplicity we shall only prove that it is admissible in the class of all g.h.l.u.e.'s in such a sub-space. For any g.h.l.u.e.

$$T_s = \sum_{\lambda \in s} \beta_{s\lambda} Y_\lambda$$

We have

$$V(T) = \left(\sum_{s \supset \lambda_o} \beta^2_{s\lambda_o} P_s - 1 \right) Y^2_{\lambda_o}$$

at the point

$$\mathcal{Y}_{\lambda_o} = (0, 0, \ldots 0, Y_{\lambda_o}, 0, \ldots, 0)$$

where the hyper plane cuts the Y_{λ_o} axis. Since

$$\sum_{s \; \lambda_o} \beta^2_{s\lambda_o} P_s \geq \frac{1}{\pi_{\lambda_o}}$$

equality if and only if

$$\beta_{s\lambda_o} = \frac{1}{\pi_{\lambda_o}} \quad \text{for all} \quad s \supset U_{\lambda_o}, \qquad (7.5.16)$$

we have

$$V(T) > V(\overset{*}{Y}_{HT}) .$$

at ϕ_{λ_o} unless (7.5.16) holds. It follows that $\overset{*}{Y}_{HT}$ is better than T at at least one of the points ϕ_{λ_o} for $1 \leq \lambda_o \leq N$, unless $T = \overset{*}{Y}_{HT}$ and hence the assertion.

§ 7.6. A remark on estimating parametric functions that depend only on a sub-population.

Let a sampling strategy consisting of a design $D = D(S,P)$ and an unbiased estimator T of Y be given and suppose that we want to make use of this strategy to estimate (perhaps at a later date) a linear p.f.

$$\theta = \sum_{1}^{N} L_i Y_i \qquad (7.6.1)$$

where some of the L_i's are zero. Such a situation can arisen when we already have an optimum or near optimum strategy for Y (like those given in Chapters IV and V) and wish to make use of it later to estimate a p.f. like θ which depends only on a sub-population

$$\mathcal{U}' = U_i : L_i \neq 0 \qquad (7.6.2)$$

An unbiased estimator T' of (7.6.1) is obtained from T by replacing Y_i in T by $L_i Y_i$ for $1 \leq i \leq N$. As remarked in § 7.4, unless D is a unicluster design, the estimator T'' given by

$$T_s'' = \begin{cases} 0 & \text{for } s \subset \overline{\mathcal{U}'} \\ \sum_{\substack{i \in s \\ i \in \mathcal{U}'}} L_i \frac{Y_i}{\pi_i} & \text{otherwise} \end{cases} \qquad (7.6.3)$$

where $s \subset \overline{\mathcal{U}'}$ means that s does not contain any unit of $\mathcal{U}'$, is at least as good as T', to estimate θ.

Consider now the design $D' = D'(S', P')$ which is a restriction of D induced by $\mathcal{U}'$, obtained thus:

$$P'_s = \begin{cases} 0 & \text{for} \quad s \subset \overline{\mathcal{U}'} \\ \dfrac{P_s}{1-q} & \text{for} \quad s \notin \overline{\mathcal{U}'} \end{cases} \tag{7.6.4}$$

where

$$q = \sum_{\substack{s \in S \\ s \subset \overline{\mathcal{U}''}}} P_s \tag{7.6.5}$$

D' thus consists of samples $s \notin \overline{\mathcal{U}'}$ only, neglecting samples with zero probabilities.

Consider now the estimator T''' defined over D' thus

$$T'''_{s'} = T''_{s'}(1 - q) \quad \text{for} \quad s' \in D' \tag{7.6.6}$$

We prove

Lemma (7.6.1):-

$$E(T''') = \theta$$

$$V(T''') \leq V(T')$$

Proof: We have

$$E(T''') = \sum_{s' \in D'} T'''_{s'} P'_{s'}$$

$$= \sum_{s' \in D'} T''_{s'} P_{s'} \qquad \text{from (7.6.4) and (7.6.6)}$$

$$= \sum_{s \in D} T''_s P_s \qquad \text{since } T''_s = 0 \text{ for } s \notin D'$$

$$= \theta,$$

which proves the first part of the lemma. To prove the second part we shall prove that

$$V(T''') \leq V(T'') \qquad (7.6.7)$$

so that the result follows since

$$V(T'') \leq V(T').$$

To prove (7.6.7) we have

$$V(T'') - V(T''') = (T''^2) - (T'''^2)$$

$$= \sum_{s \in D} T_s''^2 P_s - \sum_{s' \in D'} T_{s'}''^2 (1-q)^2 \frac{P_{s'}}{(1-q)}$$

$$= \sum_{s' \in D'} T_{s'}''^2 P_{s'} - (1-q) \sum_{s' \in D'} T_{s'}''^2 P_{s'}$$

since $T_s'' = 0$ for $s \notin D'$

$$= q \cdot \sum_{s' \in D'} T_{s'}''^2 P_{s'} \geq 0 \qquad (7.6.8)$$

equality if and only if

$$q = \sum_{\substack{s \in S \\ s \notin D'}} P_s = 0 \qquad (7.6.9)$$

which proves the required result. Hence the lemma.

From the above we see that from the point of view of variance, the strategy $H_3(D', T''')$, which rejects samples $s \in D'$ of the original design, is better than $H_1(D, T')$ and that it is even better than $H_2(D, T'')$ unless (7.6.9) is satisfied in which case $H_2 \equiv H_3$. Thus even when T is hyper admissible (in which case $H_1 \equiv H_2$), H_3 is better than H_2 from the point of view of variance.

This is not contradictory to the hyper admissibility of T if we note the important point that H_3 and H_2 are strategies involving different designs. In fact the reduction in the variance of H_3 over that of H_2 is achieved only at an extra cost of increasing the expected number of distinct units. For, we have

$$\mu(D') = \sum_{s' \in D'} \mu_{s'} P'_{s'} = \frac{1}{(1-q)} \sum_{s' \in S'} \mu_{s'} P_{s'}$$

$$= \frac{1}{1-q} \cdot \sum_{s \in S} \mu_s P_s, \quad \text{since } \mu_s = 0 \text{ for } s \notin S'$$

$$= \frac{1}{1-q} \cdot \mu(D) \qquad (7.6.10)$$

where μ_s and $\mu_{s'}$ denote the number of distinct units of s and s' that belong to u'. Thus for the estimation of θ. H_3 inflates the expected number of distinct units in the sample by a factor $\alpha = \frac{1}{1-q}$, over that of H_2 and achieves a deflation in the variance given from (7.6.8), thus:

$$V(T''') = V(T'') - q \quad V(T'') + \theta^2 \qquad (7.6.11)$$

$$= V(T'')(1 - q) - q\theta^2$$

Whether H_3 is to be prefered over H_2 or not depends on the cost function as well the relation connecting the variance in terms of loss in the units of the cost, and has to be decided upon taking these into consideration in the individual cases.

EPILOGUE

We now review briefly the present state of affairs and indicate the further lines of research.

For sampling from a finite population $\mathcal{U}$, to estimate a parametric function $\theta(\underset{\sim}{Y})$, to start with we have the basic sample space

$$\mathcal{S} = \mathcal{S}(\mathcal{U})$$

of all finite sequences of elements from $\mathcal{U}$ with a probability measure $\mathcal{P}$ defined on it. An estimator $\mathcal{T}$ defined on this probability space finally gives a sampling strategy

$$\mathcal{H} = \mathcal{H}(\mathcal{S}, \mathcal{P}, \mathcal{T})$$

Any method of sampling results in a measure $\mathcal{P}$, and can be considered as resulting from drawing units from $\mathcal{U}$ one by one with replacement and with probabilities varying if necessary (c.f. Chapter III).

The search for an optimum strategy, subject to some restrictions on $\mathcal{P}$ through considerations of cost, then involves the optimum choice of the pair $(\mathcal{P}, \mathcal{T})$. Theorem (2.3.2) proves that $\mathcal{S}$ can as well be taken to be the class

$$\mathcal{S}_0 = \mathcal{S}_0(\mathcal{U})$$

of all subsets of Λ. Theorem (7.5.1) proves that for only choice of P for unbiased and hyper admissible estimator T should be chosen to be

$$\overset{*}{Y}_{HT} = \sum_{\lambda \in s} \frac{Y_\lambda}{\pi_\lambda}$$

where π_λ's are structure constants of the strategy given by and are

$$\pi_\lambda = \sum_{\substack{s \in \mathcal{S}_0 \\ s \supset \lambda}} P_s$$

This reduces the initial problem of the triple choice of $\mathcal{S}$, P and T to that of the optimum choice of a P_0 on $\mathcal{S}_0$. This problem is still open but the author conjectures that for any P_0 on $\mathcal{S}_0$ subject to the conditions that

$$\sum_{\lambda=1}^{N} \pi_\lambda = \mu$$

there exists a better choice of P_1 on $\mathcal{S}_0$ satisfying

$$\sum_{\lambda=1}^{N} \pi_\lambda^{(1)} = \mu$$

where

$$\pi_\lambda^{(1)} = \sum_{\substack{s \supset \lambda \\ s \in \mathcal{S}_0}} P_{1,s}$$

and which further satisfies

$$P_{1,s} = 0 \quad \text{if} \quad \mu(s) > [\mu] + 1 \quad \text{or} \quad < [\mu].$$

Regarding the optimum use auxiliary information of the type considered in Chapters IV, V and VI and given by (2.3.3) the problem can be considered to have been satisfactorily closed for all practical purposes. Extention of the methods of Chapter IV to general sample sizes will be of some interest. Another and perhaps a more important problem to pursue is that of optimum use of multiple auxiliary information each component of which is of the type given by (2.3.3). Another problem is to find suitable transformations of the auxiliary information to satisfy or approximately satisfy (2.3.3), when it departs from (2.3.3) These searches are made more convenient henceforth if the criterion of hyper admissibility is accepted in which case the problem is reduced from the optimum selection of the pair $(P_o, \mathcal{J})$ to that of P_o alone. If the conjecture given at the end of preceding paragraph is proved true this problem is further reduced to that of finding optimum structure constants $\pi_\lambda^{(1)}$'s, and then searching for an optimum P_1 in a narrow class having these optimum structure constants.

BIBLIOGRAPHY

This bibliography is not exhaustive. For an exhaustive bibliography we refer to three papers of Dalenius [A_1], Murthy [A_2] and Sukhatme [A_3], which together give a fairly exhaustive bibliography. The bibliography given here mostly contains a list of papers in which problems similar to those treated in this thesis are considered.

Expository Papers

A_1. Dalenius, T. — Recent advances in sample survey theory and methods. Ann.Math.Stat.,33(1962),pp.325-349.

A_2. Murthy, M.N. — Some recent advances in sampling theory. Jour.Amer.Stat.Ass.,58(1963),pp.737-755

A_3. Sukhatme, P.V. — Major developments in the theory and application of sampling during the last twenty-five years. Estadistica, 17(1959) pp.652-679.

Books

B_1. Cochran, W.G. — Sampling Techniques. (Second Edition) John Wiley.New York. 1963.

B_2. Deming, W.E. — Some theory of Sampling. John Wiley. New York. 1950.

B_3. Hansen,M.H., Hurwitz,W.N. and Madow,W.G. — Sample Survey Methods and Theory,Vols.I and II. John Wiley. New York. 1953.

B_4. Yates, F. — Sampling Method for Censuses and Surveys. Griffin. London. 1949.

B_5. Fisher, R.A. — The Design of Experiments. Oliver and Boyd, Edinburgh, 19 7.

Papers

1. Basu,D. — On sampling with and without replacement. Sankhyā, 20(1958),pp.287-294.

2. Blackwell, D. — Conditional expectation and unbiased sequential estimation. Ann.Math.Stat., 18(1947),pp.105-110.

3. Cochran, W.G. — Relative accuracies of systematic and stratified random samples for a certain class of population Ann.Math.Stat., 17(1946), pp.164-179.

4. Das, A.C. — On two phase sampling with varying probabilities Bull.Inst.Stat.Inst., 33(1951),pp.105-112.

5. ———— — MVU Estimates of parameters of a finite populations base on varying probability samples. Bull. Calcutta Stat. Assn., 11, pp. 39-48.

6. Des Raj — Some estimators in sampling with varying probabilites without replacement. Jour.Amer. Stat.Ass.,51(1956),pp.269-284.

7. ———— — A note on the determination of optimum probabilities in sampling with replacement. Sankhyā, 17(1956), 197-200.

8. Des Raj and Khamis,S.H. — Some remarks on sampling with replacement. Ann.Math.Stat., 39(1958),pp.550-557.

9. Durbin, J. — Some results in sampling theory when the units are selected with unequal probabilities. Jour.Roy.Stat.Soc.,B, 15(1953), pp.262-269.

10. Fellegi, I.M. — Sampling with varying probabilities without replacement:Rotating and non-rotating samples. Jour.Amer.Stat.Ass.,58(1963),pp.183-201.

11. Godambe, V.P. — A unified theory of sampling from finite populations. Jour. Roy. Stat. Soc., B 17 (1955) pp.269-277.

12. ------ — An admissible estimator for any sampling design. Sankhya, 22 (1960), pp. 285-288.

13. Goodman, L.A., and Kish, L. — Controlled selection-A technique in probability sampling. Jour.Amer.Stat.Ass. 35 (1950), pp. 350-372.

14. Grundy, P.M. — A method of sampling with probability exactly proportional to size. Jour.Roy.Stat.Soc.B, 16 (1954), pp. 236-238.

15. Hajek,J. — Optimum strategy and other problems in probability sampling. Casipis Pest Math. 84 (1959), pp. 387-423.

16. ---------- — Some contributions to the theory of probability sampling. Bull. Int.Stat.Inst. 36 (1958), pp. 127-134

17. Halmos, P.R. — The theory of unbiased estimation. Ann.Math.Stat., 17 (1946), pp. 34-43.

17a. Hansen, M.H. and Hurwitz,W.W. — On the determination of optimum probabilitids in sampling. Ann. Math. Stat. 20 (1949), pp. 426-432

18. Hanurav, T.V. — An existence theorem in sampling. Sankhya, A, 24 (1962), pp. 227 - 230

19. ------- — Some sampling schemes in probability sampling Sankhya, A, 24 (1962), pp. 421-428

20. ------- — On Horvitz and Thompson's estimator. Sankhya, A, 24 (1962), pp. 429-436.

21. Hartley, H.O. and Rao, J.N.K — Sampling with unequal probabilities and without replacement. Ann.Math.Stat., 33 (1962), pp. 350-374

22. Horvitz,D.G. and Thompson, D.J. — A generalisation of sampling without replacement from a finite universe. Jour. Amer. Stat. Ass., 47 (1952), pp. 663-685.

23. Keyfitz, N. — Sampling with probability proportional to size Jour.Amer.Stat.Assoc., 46, (1951), pp.105-109.

24. Koop, J. C. — On the axioms of sample formation and their bearing on the construction of linear estimators in sampling theory for finite populations, Metrika, 7 (1963), 81-114.

24a. Mahalanobis, P.C. — A sample survey of the acreage under jute in Bengal, Sankhya, 4 (1940), pp. 511-530.

25. Midzuno, H. — On the sampling system with probability proportional to the sum of sizes. Ann. Inst. Stat.Math., 3 (1952), pp.99-108.

26. Murthy, M.N. — On ordered and unordered estimators. Sankhya, 18 (1957), pp. 379.390.

27. ..... — Generalised estimation for finite populations. Sankhya, 25A (1963), pp. 245-262

28. Narain, R.D. — On sampling without replacement with envarying probabilities. Jour. Ind.Soc.Agri. Stat., 3 (1951), pp. 169.174.

29. Neyman, J. — On two different aspects of the representative method: the method of stratified sampling and the method of purposive selection. Jour. Roy. Stat. Soc. B. (1938), pp. 558.606.

30. ..... — Contributions to the theory of sampling human populations. Jour. Amer. Stat. Assn. 33 (1938), pp. 101.116.

31. Pathak, P.K. — On sampling schemes providing unbiased ratio estimators. Ann.Math.Stat., 35, pp. 222-231

32. Rao, C.R. — Information and accuracy in the estimation of statistical parameters. Bull.Cal.Stat.Soc., 37 (1945). pp.81-91.

33. Rao, J.N.K. — On three procedures of unequal probability sampling without replacement. Jour.Amer.Stat.Assn., 58 (1963), pp.202-215.

34. ------ — On the estimate of the variance in unequal probability sampling. Ann. Inst. Stat. Math. 13 (1962), p. 57-60.

35. Rao, J.N.K., Hartley, H.O. and Cochran, W.G. — A simple procedure of unequal probability sampling without replacement. Jour.Roy. Stat.Soc., B, 24 (1962), pp. 482-491.

36. Roy, J. and Chakravorty, I.M. — Estimating the mean of a finite population. Ann. Math. Stat., 31, (1960) pp.392-398.

37. Roy Chowdhury, D.K. — Sampling with verying probabilities. Part of the thesis submitted for the Associateship of the Indian Statistical Institute (1956).

38. Sen A. R. — Present status of probability sarpling and its use in estimation of characteristics. (abstract). Econometrica, 20 (1952), p. 103.

39. Stevens, W.L. — Sampling without replacement with probability proportional to size. Jour.Roy.Stat.Soc. B, 20, (1958), pp. 343 - 347.

40. Yates, F. and Grundy, P.M. — Selection without replacement from within strata with probability proportional to size. Jour. Roy. Stat. Soc., B, 15 (1953), pp. 253-261.