SOME ASPECTS OF UNIFIED SAMPLING THEORY

By T. V. HANURAV*

Indian Statistical Institute

SUMMARY. Starting with the basic concepts in sampling theory for finite populations we proceed to the fundamental problem of optimum estimation procedures to estimate the population total (or equivalently the mean) from a unified approach.

The intrinsic properties of sampling designs are dealt with and the problem of estimation is clearly formulated. Estimability of parametric functions is discussed. A complete characterisation of designs admitting best estimators, is given. Various criteria for the reduction of minimal complete class of estimators are discussed including the latest criterion of 'hyper admissibility' which is due to the author. A number of unsolved problems are posed in the sequel and some new terminology is introduced in the hope of standardising the same. The final reduction of the problem and some open problems are presented.

1. Introduction

Sampling theory for finite populations—often called sample surveys—has seen some significant developments during the last ten or twelve years. While the earlier development of the subject had been guided by intuitive considerations (which no doubt were quite powerful) to obtain unbiased estimators, it is only during the past few years that attempts are being made to formalise the theory and to consider the purely mathematical aspects of the theory. The first attempts in this direction can be found in the work of Horvitz and Thompson (1952) and the first formalisation is due to Godambe (1955) who generalised the concepts of sampling design and linear estimators and proved the important result that for no sampling design does there exist a uniformly minimum variance unbiased estimator of the population total (or equivalently, the mean). As will be seen in Section 4, this result has some exceptions. Barring these unhappy exceptions (completely characterised in Section 4), this result pointed to the inadequacy of the then existing method of applying the famous Markov's theorem on least squares to derive best linear unbiased estimators of the population total, and made clear the main difference between the classical theory of estimation for theoretical populations and the theory for finite populations. We can briefly describe this as the *identifiability* of units that exists in the latter theory.

In this paper we approach the problem of estimation for finite populations, in an orderly fashion. While inevitably discussing the developments in this field to date, we shall also give some results that are necessary to fill the gaps existing now. The paper is neither a purely review paper nor a purely original paper but is a mixture of both. However, there is no scope for any confusion about the points of fusion between the two, since the works of other authors are clearly annotated. While we indulge fully in the historical aspects of the problems considered here no pretence is made to claim that all the problems treated by earlier authors are covered here, and we confine ourselves only to specific lines of development of the theory.

* The paper was based on a thesis submitted by the author to the Indian Statistical Institute. The final draft of the paper was prepared when he was a Visiting Lecturer at the University of Sheffield, England.

A 'simple finite population' $\mathcal{U}$ is a population of *known* number N of *identifiable* units

$$U_1, U_2, \ldots, U_N. \qquad \ldots \quad (1.1)$$

This definition excludes such known finite populations like the fish in a lake for which *a priori* neither N is known nor the units are identifiable (in fact the problem of main interest in these cases is to estimate the unknown N); or a box of 'unmarked' bolts produced by a machine because for this though N is known, the units, i.e., the bolts, are not identifiable *a priori*. The reason for the exclusion of such populations is that it is not possible to draw probability samples (as explained below) from such populations. Samples drawn from such populations are assumed to be probability samples by means of a reasoning running like "because there is no reason to believe that the sample is not a random sample...". In these cases there is no way of testing (say with the help of tested tables of random numbers) the validity or otherwise of the nature of randomness of the sample.

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

$$s = \{U_{i_1}, U_{i_2} \ldots, U_{i_{n(s)}}\}, \; n_{(s)} < \infty \qquad \ldots \quad (1.2)$$

where $1 \leqslant i_t \leqslant N$ for $1 \leqslant t \leqslant n(s)$. The i_t's need not necessarily be distinct but the interchange of U_{i_t} and $U_{i_{t'}}$ for $i_t \neq i_{t'}$ results in a new sample. $n(s)$ is the size of s, and $\nu(s)$, the number of *distinct* units of s, is the effective size of s. While $n(s)$ can even exceed N (because repetitions are allowed), $\nu(s) \leqslant N$.

While any specific sample has to be of finite size only, there is no reason, *a priori*, to restrict ourselves to samples of a fixed size only (i.e. $n(s) = n$) nor is it obviously justified to restrict to samples of size less than a given number M say. Accordingly we define $\mathcal{S}$, the collection of all possible sample s from $\mathcal{U}$ as our basic sample space :

$$\mathcal{S} = \{s\}. \qquad \ldots \quad (1.3)$$

Evidently $\mathcal{S}$ contains a countably infinite number of samples, and

$$\sup_{s \in \mathcal{S}} n(s) = \infty.$$

A simple sampling design $D = D(\mathcal{U}, \mathcal{S}, P)$, briefly called the design P, is a probability measure P defined on $\mathcal{S}$,

$$P_s \geqslant 0 \quad \text{and} \; \sum_{s \in \mathcal{S}} P_s = 1. \qquad \ldots \quad (1.4)$$

The above definition excludes designs such as those obtained thus; continue simple random sampling with replacement until the sample variance is less than 10 per cent of the sample mean'. The reason for the exclusion of such not uninteresting designs is the resulting simplification in the theory.

In practice, however, samples are not drawn by listing the P_s's for all possible samples. Instead, they are drawn by what can be termed as 'sampling methods.' Any sampling method in which the samples are all ordered samples as in (1.2) gives rise to a unique design. Of particular interest among these sampling methods are the unit drawing mechanisms which are methods of drawing the sample by means of selecting the units from $\mathcal{U}$ one by one and with replacement. In its most general form a unit drawing mechanism can be rigorously defined as an algorithm

$$A = A\{q_1(U_i);\ q_2(s);\ q_3(s,\ U_i)\} \qquad \ldots \quad (1.5)$$

where

(1) q_1 is a probability measure on $\mathcal{U}$ so that

$$q_1(U_i) > 0 \quad \text{for} \quad 1 \leq i \leq N \quad \text{and} \quad \sum_{i=1}^{N} q_1(U_i) = 1, \qquad \ldots \quad (1.6)$$

(2) $q_2(s)$ defined for any sample $s \epsilon \mathcal{S}$ is a number in (0, 1)

$$0 \leq q_2(s) \leq 1 \quad \text{for} \quad s \epsilon \mathcal{S} \qquad \ldots \quad (1.7)$$

and

(3) $q_3(s,\ U_i)$, defined only for those s for which $q_2(s) \neq 0$, is a probability measure on $\mathcal{U}$:

$$q_3(s, U_i) \geq 0 \quad \text{for} \quad 1 \leq i \leq N \quad \text{if} \quad q_2(s) \neq 0 \quad \text{and} \quad \sum_{i=1}^{N} q_3(s, U_i) = 1. \quad \ldots \quad (1.8)$$

The sampling method using the algorithm is as follows : Draw the first unit using the measure q_1. If the sample thus obtained is denoted by $s_{(1)}$, impute $s_{(1)}$ in q_2. If $q_2(s_{(1)}) = 0$ sampling is terminated. Otherwise, a binomial trial with probability of success as $q_2(s_{(1)})$ is performed and sampling is terminated if the trial results in a failure. If the trial results in a success, a second unit is drawn using the probability measure $q_3(s_{(1)}, U_i)$ and the resulting sample (which is $s_{(1)}$ followed by the unit now selected) is denoted by $s_{(2)}$. The operations of imputing $s_{(2)}$ in q_2 and using $q_3(s_{(2)}, U_i)$ etc., are repeated until a sample $s_{(k)}$, of size k say, is reached for which $q_2(s_{(k)}) = 0$ and $s_{(k)}$ is then accepted as the final sample.

It is easy to see that the various customary methods of sampling are particular cases of the above general method. Where a method of sampling does not specify the order of the units in the samples, for each method of ordering of the units in the sample, there is an algorithm of the above type.

The advantages of the definition of the design as given by (1.4) is that it supplies a unified framework within which to work for a search for optimum estimators of a given parametric function. It is not possible otherwise to discuss the apparently diverse methods of sampling all in a single framework. Since samples are drawn, in practice, by methods like unit drawing mechanisms and not from the design, a question of primary interest then arises—can every design be generated by a suitable sampling method ? The answer to this is given by the following theorem.

Theorem 1.1 : *To any given design $D(\mathcal{U}, \mathcal{S}, P)$ there corresponds a unique unit drawing mechanism $A(q_1, q_2, q_3)$, such that sampling according to A results in the design P, and conversely.*

The result assures us that we can work within the unified framework of designs, for a search for optimum estimation procedures, and can then generate suitable sampling methods as unit drawing mechanisms, to achieve these optimum designs. The proof of this result (which was given in a less general form earlier (Hanurav, 1962a)) runs thus :

The second part of the theorem is evident. For, if $A = A(q_1, q_2, q_3)$ be the algorithm, then for any sample

$$s = \{U_{i_1}, U_{i_2}, \ldots, U_{i_{n(s)}}\} = s(i_1, i_2, \ldots, i_{n(s)}) \text{ say,}$$

we have for the final probability of the sample, the unique value

$$Q_s = \Pr(s|A) = q_1(U_{i_1}) \prod_{k=1}^{n(s)-1} q_2(\{U_{i_1}, U_{i_2}, \ldots, U_{i_k}\})$$

$$\times \prod_{k=1}^{n(s)-1} q_3(\{U_{i_1}, \ldots U_{i_k}\}, \; U_{i_{k+1}})[1-q_2(\{U_{i_1}, U_{i_2}, \ldots, U_{i_{n(s)}}\})]. \quad \ldots \quad (1.9)$$

It is easy to verify that $\sum_{s \in \mathcal{S}} Q_s = 1$.

To prove the first part, for $1 \leqslant i_1, i_2, \ldots \leqslant N$, let

$$S_i = \{s : i_1 = i\} \qquad S_{ij} = \{s : i_1 = i, \; i_2 = j\} \quad \text{etc.}$$

$$\alpha_i = \sum_{s \in S_i} P_s \qquad \alpha_{ij} = \sum_{s \in S_{ij}} P_s \text{ etc.}$$

$$s(i) = \{U_i\}, \qquad s(i,j) = \{U_i, U_j\} \quad \text{etc.,}$$

and $$\beta_i = P_{s(i)}, \qquad \beta_{ij} = P_{s(i,j)} \quad \text{etc.}$$

Clearly we have

$$\mathcal{S} = \bigcup_{i=1}^{N} S_i$$

$$S_i = \bigcup_{j=1}^{N} S_{ij} \bigcup s(i)$$

$$S_{ij} = \bigcup_{k=1}^{N} S_{ijk} \bigcup s(i,j) \text{ etc.,}$$

that

$$\sum_{1}^{N} \alpha_i = 1$$

$$\beta_i + \sum_j \alpha_{ij} = \alpha_i$$

$$\beta_{ij} + \sum_k \alpha_{ijk} = \alpha_{ij} \text{ etc.}$$

Defining $A(q_1, q_2, q_3)$ by the equations

$$q_1(U_i) = \alpha_i$$

$$q_2(s(i_1, i_2, \ldots, i_k)) = \begin{cases} 1 - \dfrac{\beta_{i_1 i_2 \cdots i_k}}{\alpha_{i_1 i_2 \cdots i_k}} & \text{if } \alpha_{i_1 i_2 \cdots i_k} \neq 0 \\ 0 & \text{otherwise} \end{cases} \qquad \ldots \quad (1.10)$$

and $$q_3(s(i_1, \ldots, i_k), U_i) = \frac{\alpha_{i_1 i_2 \cdots i_k i}}{\alpha_{i_1 \cdots i_k} - \beta_{i_1 \cdots i_k}} \qquad \text{if} \qquad q_2(s(i_1 \ldots i_k)) \neq 0.$$

It is easy to check that A satisfies all the conditions (1.6), (1.7) and (1.8) of a unit drawing mechanism. Further, from (1.9) (1.10) we have, for any sample s, the probability $\Pr(s \mid A)$ given to s by the algorithm A is given by

$$\begin{aligned} \Pr(s \mid A) &= \alpha_{i_1} \left\{ 1 - \frac{\beta_{i_1}}{\alpha_{i_1}} \right\} \cdot \frac{\alpha_{i_1 i_2}}{\alpha_{i_1} - \beta_{i_1}} \cdot \left\{ 1 - \frac{\beta_{i_1 i_2}}{\alpha_{i_1 i_2}} \right\} \\ &\quad \times \frac{\alpha_{i_1 i_2 i_3}}{\alpha_{i_1 i_2} - \beta_{i_1 i_2}} \cdots \left\{ 1 - \frac{\beta_{i_1 \cdots i_{n(s)-1}}}{\alpha_{i_1 \cdots i_{n(s)-1}}} \right\} \cdot \frac{\alpha_{i_1 \cdots i_{n(s)}}}{\alpha_{i_1 \cdots i_{n(s)-1}} - \beta_{i_1 \cdots i_{n(s)-1}}} \cdot \frac{\beta_{i_1 i_2 \cdots i_{n(s)}}}{\alpha_{i_1 i_2 \cdots i_{n(s)}}} \\ &= \beta_{i_1 \ldots i_{n(s)}} \\ &= P_s. \end{aligned}$$

This proves that sampling according to A given by (1.10) generates the given design. That in fact A is unique can be proved by retracing the above argument and using (1.9). This completes the proof of the theorem.

The proof of the above theorem is constructive and enables one to derive A from P. In several situations we have the design P, only partially specified. In such cases, corresponding to any further consistent specifications that completely specify the design, we have a unit drawing mechanism generating a design with the specifications given. The simplicity of the resulting mechanism depends on a clever choice of these further specifications. An example will make this point clear.

In the theory of ratio-estimators, we have the values X_i and U_i $(1 \leqslant i \leqslant N)$ of an auxiliary variate $\mathscr{X}$, completely beforehand. A problem considered and solved, independently, by Midzuno (1952) and Sen (1952) is as follows. Given a positive integer n and given that the design P satisfies,

$$P_s = 0 \text{ if } n(s) \neq \nu(s) \text{ or } n(s) = \nu(s) \neq n, \qquad \ldots \quad (1.11)$$

what should be the sampling method to ensure the estimator

$$\hat{Y}_{ms} = \frac{\bar{y}_s}{\bar{x}_s} X \qquad \ldots \quad (1.12)$$

is unbiased for Y? In the above, $\bar{x}_s$ and $\bar{y}_s$ are the sample means of $\mathscr{X}$ and $\mathscr{Y}$, and $X = \sum_{1}^{N} X_i$. It can be seen that a set of necessary and sufficient conditions is that P satisfies, in addition to (1.11)

$$\sum_{s \sim s_0} P_s = \frac{1}{X\binom{N-1}{n-1}} n\bar{x}_{s_0} \quad . \tag{1.13}$$

for any sample s_0 and where the sum on the l.h.s. is over all samples s that are permutations of s_0. (1.13) remains unaltered if s_0 is replaced by any sample which is a permutation of s_0. These conditions do not completely pin down the design P and we need further specifications to allocate the total probability on the r.h.s. of (1.13) to all the individual samples that are permutations of s_0. Considering the simple allocation of equal probabilities to all samples that are permutations of one another, we have a fully specified design P' given by

$$P'_s = \frac{1}{n!} \frac{1}{X\binom{N-1}{n-1}} \cdot n\bar{x}_s. \qquad \ldots \quad (1.14)$$

However, this results is an inconveniently complicated algorithm $A' = A'(q'_1, q'_2, q'_3)$ as can be easily seen even for the simple case of $N = 3$, $n = 2$. Following the proof of Theorem (1.1) it can be verified that

$$q'_1(U_i) = \frac{X+X_i}{4X}, \qquad i = 1, 2, 3$$

$$q'_2(s(i)) = 1 \quad \text{and } q_2(s(i, j)) = 0$$

$$q'_3(s(i), U_j) = \begin{cases} \dfrac{X_i+X_j}{X+X_i} & \text{if } j \neq i \\ 0 & \text{if } j = i. \end{cases} \qquad \ldots \quad (1.15)$$

Thus not only the initial probabilities q'_1's but also the conditional probabilities q'_3's have to be calculated afresh from the X_i's for each draw. For larger values of N and n this will become more laborious.

Considering now an alternative allocation proportional to the $\mathscr{X}$-value of the first unit in the sample we get another design P'' also satisfying (1.11) and (1.13) given by

$$P''_s = \frac{X_{i_1}}{n\bar{x}_s} \cdot \frac{1}{X\binom{N-1}{n-1}} n\bar{x}_s$$

$$= \frac{X_{i_1}}{X} \cdot \frac{1}{\binom{N-1}{n-1}} \quad . \qquad \ldots \quad (1.16)$$

It can be verified that the corresponding algorithm $A^*(q_1^*, q_2^*, q_3^*)$ is given by

$$q_1^*(U_i) = \frac{X_i}{X}$$

$$q_2^*(s(i_1)) = q_2^*(s(i_1 i_2)) = \ldots = q_2^*(s(i_1 i_2 \ldots i_{n-1})) = 1$$

$$q_3^*(s(i_1 i_2 \ldots i_k),\ U_j) = \begin{cases} \dfrac{1}{N-k} & \text{if } j \neq i_1, i_2, \ldots, i_k \\ 0 & \text{otherwise} \end{cases} \quad \ldots \quad (1.17)$$

$$q_2^*(s(i_1, i_2, \ldots, i_n)) = 0$$

for $1 \leqslant i_1, i_2, \ldots, i_{n-1} \leqslant N$ and for $1 \leqslant k \leqslant n-1$.

The algorithm thus gives a simple unit drawing mechanism described thus, 'select the first unit with probabilities proportional to the original size measures X_i's. Omitting the selected unit, from the remaining $(N-1)$ units draw a simple random sample of size $(n-1)$ without replacement'. This is in fact the solution obtained by Midzuno and Sen though by other methods.

We see thus that the above theorem enables us to systematically investigate the possible unit drawings mechanism that generates the design with the required properties viz (1.13) and to discover incidentally the simplest solution.

2. The design

We return to the unified framework of the basic simple space $\mathcal{S}$ and a probability measure P on $\mathcal{S}$.

Given a design P, let

$$\pi_i = \pi_i(P) = \sum_{s \supset i} P_s, \quad 1 \leqslant i \leqslant N \quad \ldots \quad (2.1)$$

and

$$\pi_{ij} = \pi_{ij}(P) = \sum_{s \supset i, j} P_s, \quad 1 \leqslant i \neq j \leqslant N \quad \ldots \quad (2.2)$$

where in (2.1) the sum on the r.h.s. is over all samples that contain U_i and in (2.2) the sum is over all samples that contain U_i and U_j. π_i is the probability that a random sample contains U_i and π_{ij} gives the probability that a random sample contains U_i and U_j. The π_i's and π_{ij}'s can be called the first and second order inclusion probabilities respectively. Higher order inclusion probabilities can be defined similarly but are not of immediate interest to us. These *structure constants* π_i's and π_{ij}'s play an important role in our theory.

It follows from the definitions that

$$0 \leqslant \pi_i \leqslant 1 \quad \ldots \quad (2.3a)$$

and

$$0 \leqslant \pi_{ij} \leqslant \min(\pi_i, \pi_j) \quad \ldots \quad (2.3b)$$

for $1 \leqslant i \neq j \leqslant N$, where min (α, β) denotes as usual the smaller of α and β.

Let ν be the expected effective sample size of a design P

$$\nu = \nu(P) = \sum_{s \in \mathcal{S}} \nu(s) \cdot P_s. \qquad \ldots \quad (2.4)$$

Three important formulae connect the π_i's and π_{ij}'s and ν_s's.

Theorem 2.1 (Godambe, 1955): *For any design P*

$$\sum_{i=1}^{N} \pi_i = \nu. \qquad \ldots \quad (2.5)$$

Theorem 2.2 (Hanurav, 1962a): *For any design P*

$$\sum_{i \neq j} \sum \pi_{ij} = \nu(\nu-1) + V(\nu(s)). \qquad \ldots \quad (2.6a)$$

Observing that for any design

$$1 \leqslant \nu(s) \leqslant N$$

if $\nu = [\nu]+\theta$ where $0 \leqslant \theta < 1$ i.e. θ is the fractional part of ν one can show that

$$\theta(1-\theta) \leqslant V\nu(s) \leqslant (N-\nu)(\nu-1)$$

and from (2.6a) follows that

$$\nu(\nu-1)+\theta(1-\theta) \leqslant \sum_{i \neq j} \sum \pi_{ij} \leqslant N(\nu-1). \qquad \ldots \quad (2.6b)$$

Theorem 2.3 (Yates and Grundy, 1953): *For a design P for which*

$$P_s > 0 \rightarrow \nu(s) = \nu \quad \textit{for all } s \in \mathcal{S} \qquad \ldots \quad (2.7)$$

we have, for any

$$\sum_{j \neq i} \pi_{ij} = (\nu-1)\pi_i \qquad \ldots \quad (2.8a)$$

and hence or otherwise from (2.6a)

$$\sum_{i \neq j} \sum \pi_{ij} = \nu(\nu-1). \qquad \ldots \quad (2.8b)$$

Some interesting questions of internal consistency now arise. The answers to these have a direct bearing on problems of estimation, as will be seen in Section 7.

(a) Given a set of numbers $\{\pi_i\}$, $1 \leqslant i \leqslant N$ satisfying (2.3a) does there exist a design P such that $\pi_i(P) = \pi_i$ for all i? The answer is 'yes', as is easily seen from the design obtained thus: Conduct N independent binomial trials with probability of success for the i-th trial being equal to π_i. If and only if the i-th trial results in a success do we include U_i in the sample which is now made to consist of the units selected, arranged in the increasing order of their indices i. Evidently for this design $\pi_i(P) = \pi_i$ for all i. Some more solutions were given earlier (Hanurav, 1962b).

(b) Given any set of numbers π_i, $1 \leqslant i \leqslant N$ satisfying (2.3a) and such that $\nu = \sum_{1}^{N} \pi_i$ is a positive integer, does there exist a design P which satisfies (2.7) and for which $\pi_i(P) = \pi_i$ for all i? The answer to this is also 'yes' as can be seen by the design obtained by the 'pps systematic sampling' method of Goodman and Kish (1950). However, if in addition to (2.8) we impose the further conditions

$$0 < \pi_{ij} \lneq \pi_i \pi_j \quad \text{for} \quad 1 \leqslant i \neq j \leqslant N \qquad \dots \quad (2.9)$$

which ensure the existence of non-negative unbiased estimator of the variance of an important estimator (cf. Section 7), then no answer is known for $\nu > 2$. For $\nu = 2$ the author solved this problem satisfactorily (Hanurav, 1965; 1966). Constructive solutions for higher values of ν are of importance in our estimation theory.*

Perhaps more complicated is the question of internal consistency of π_{ij}'s and π_i's. (2.3a) and (2.3b) give a set of necessary conditions. It can be easily seen that they are not sufficient. For example for $N = 3$ there cannot exist a design with

$$\pi_1 = .7, \qquad \pi_2 = .8, \qquad \pi_3 = .8$$

$$\pi_{12} = .4, \qquad \pi_{23} = .7 \quad \text{and} \quad \pi_{31} = .6$$

which clearly satisfies (2.3a) and (2.3b). It can be proved that for any design P

$$\pi_{ij}(P) \geqslant \pi_i(P) + \pi_j(P) - 1 \qquad \dots \quad (2.10)$$

for $1 \leqslant i \neq j \leqslant N$. To prove the above we need consider probability of the event that neither U_i nor U_j are included in a random sample, and express the condition that this quantity is nonnegative. However, even (2.3a), (2.3b) and (2.10) do not constitute a set of sufficient conditions for the existence of a design P with these as its $\pi_i(P)$'s and $\pi_{ij}(P)$'s. For example there cannot exist a design with

$$\pi_1 = .4, \qquad \pi_2 = .6, \qquad \pi_3 = .4$$

$$\pi_{12} = .05, \quad \pi_{23} = .05, \quad \pi_{13} = .05.$$

A compact set of sufficient conditions on π_i's and π_{ij}'s are of some interest. Of course a complete set of necessary and sufficient conditions are provided by considering all possible $2^N - 1$ events that specify the units belonging to a sample and express the conditions that their probabilities lie between 0 and 1. But these conditions involve the higher order inclusion probabilities also.

Another problem that arises is as follows. Given a design P_1 with a set of $\pi_i(P_1)$'s and $\pi_{ij}(P_1)$'s, does there exist a design P_2 such that

$$\pi_i(P_2) = \pi_i(P_1) \quad \text{for} \quad 1 \leqslant i \leqslant N$$

and

$$\pi_{ij}(P_2) \leqslant \pi_{ij}(P_1) \quad \text{for} \quad 1 \leqslant i \neq j \leqslant N. \qquad \dots \quad (2.11)$$

More important is the question of existence of P_2 which further satisfies

$$\sum_{i \neq j} \sum \pi_{ij}(P_2) = \nu(\nu-1) + \theta(1-\theta) \qquad \dots \quad (2.12)$$

where θ, as given in (2.6b), is the fractional part of ν. These questions play a crucial role in the choice of optimum strategies, as will be explained later in Section 7.

* Recently the author solved this problem fully for all integral values of ν. The solution was read at the 29th annual meeting of the IMS at New Brunswick and will be published shortly.

3. Estimation

Consider now a variable $\mathcal{Y}$ defined on the units (1.1) of $\mathcal{U}$ taking the value Y_i on $U_i (1 \leqslant i \leqslant N)$. The Y_i's can be vectors but no essential feature is lost by restricting as we shall do in the sequel, to the case that $\mathcal{Y}$ is a real-valued function. The vector

$$Y = (Y_1, Y_2, ..., Y_N) \qquad ... \quad (3.1)$$

is unknown *a priori* and is treated as a parameter and R^N, the N-dimensional Eucledian space is the parameter space. Any single-valued function $f(Y)$ of Y is called a parametric function. The problem is to estimate certain parametric functions that are of interest to us. Of particular interest is the population total

$$Y = \sum_1^N Y_i \qquad ... \quad (3.2)$$

and a number of other interesting problems can be boiled down to the problem of estimation of Y, as we shall see, by a suitable redefinition of the variable or the population or both. For example the estimation of any linear parametric function

$$l_0 + \sum_1^N l_i Y_i$$

is the same as the estimation of the population total $Z = \sum_1^N Z_i$ of a new variable $\mathcal{Z}$ defined by $Z_i = \frac{l_0}{N} + l_i Y_i$. Similarly, to estimate a quadratic

$$l_0 + \Sigma\, l_i Y_i + \Sigma q_{ii} Y_i^2 + \sum_{i \neq j} \Sigma q_{ij} Y_i Y_j$$

we can estimate the first two terms as explained above. The estimation of the third term is the same as the estimation of the total $Q = \sum_1^N Q_i$ of the new variable $\mathcal{Q}$ defined by $Q_i = q_{ii} Y_i^2$. For the estimation of the last term we consider the new population $\mathcal{U}'$ whose elements are ordered pairs of units of $\mathcal{U}$, and define a new variable $\mathcal{Q}'$ which takes the value $Q'(U_i, U_j) = Q'(U_j, U_i) = \frac{1}{2} q_{ij} Y_i Y_j$ if $i \neq j$ and $Q'(U_i, U_i) = 0$. Clearly the last term equals the total $Q' = \Sigma\Sigma Q'(U_i, U_j)$, and the estimation of the last term is the same as the estimation of Q' defined over $\mathcal{U}'$. Now from samples s of $\mathcal{U}$ we construct samples s' of $\mathcal{U}'$, as follows. If

$$s = \left\{ U_{i_1}, U_{i_2}, ..., U_{i_{n(s)}} \right\}$$

then

$$s' = \Big\{ \left(U_{i_1}, U_{i_1}\right), \left(U_{i_1}, U_{i_2}\right), ..., \left(U_{i_1}, U_{i_{n(s)}}\right),$$

$$\left(U_{i_2}, U_{i_2}\right), ... \left(U_{i_2}, U_{i_{n(s)}}\right), ..., \left(U_{i_{n(s)-1}}, U_{i_{n(s)}}\right), \left(U_{i_{n(s)}}, U_{i_{n(s)}}\right) \Big\}.$$

The set of all such s''s is a subset of the basic sample space $\mathcal{S}'$ of $\mathcal{U}'$. From the given design $D(\mathcal{S}, P)$ over $\mathcal{U}$ we construct a design $D'(\mathcal{S}', P')$ on $\mathcal{U}'$ thus :

$$P'_{s'} = \begin{cases} P_s & \text{if } s' \text{ is one generated as above, by a sample } s \text{ of } \mathcal{S} \\ 0 & \text{otherwise.} \end{cases}$$

The problem of estimation of $\sum\sum_{i \neq j} q_{ij} Y_i Y_j$ now reduces to the problem of estimation the total of the variable $\mathcal{Q}'$ defined over $\mathcal{U}'$ from the design $D'(\mathcal{U}', \mathcal{S}', P')$. Estimation of all polynomial parametric functions of Y on $\mathcal{U}'$ in $D(S, P)$ can be similarly reduced to the estimation of the totals of some variables on some populations. Thus it is the estimation of Y defined by (3.2) that need be of central interest to us.

A statistic T defined for $s \in \mathcal{S}$, at any rate for those samples s for which $P_s > 0$, is a single-valued function of the $\mathcal{Y}$-values of the units belonging to s,

$$T_s = T_s\left(Y_{i_1}, Y_{i_2}, ..., Y_{i_{n(s)}}\right). \quad ... \quad (3.3)$$

Tautologically, T is said to be an estimator of a given parametric function $f(Y)$ if from a sample s we estimate $f(Y)$ by T. Recognising T as a random variable defined for $s \in \mathcal{S}$ with the given measure P we can talk of the expectation, mean square error, variance etc., of T. It is said to be an unbiased estimator of $f(Y)$ iff

$$E(T) = \sum_{s \in \mathcal{S}} T_s P_s = f(Y), \; Y \in R^N. \quad ... \quad (3.4)$$

A given parametric function $f(Y)$ is *estimable* in a given design P iff there exists a statistic T such that (3.4) holds. A sampling design P together with an estimator T of $f(Y)$ constitute a sampling strategy or simply strategy, and is denoted by $H(\mathcal{S}, P, T)$. The mean, variance etc., of H are defined to be the corresponding quantities of T with respect to the measure P. The problem is to find 'optimum' sampling strategies for the estimation of a given parametric function $f(Y)$, optimality being defined in a reasonable way.

We first define some classes of linear estimators. The concept of 'linear estimator' has been generalised by Godambe (1955) in this context. A general homogeneous linear estimator (g.h.l.e., for brevity) is of the form

$$T : \{T_s = \sum_{\lambda \in s} \beta_{s\lambda} Y_\lambda\}. \quad ... \quad (3.5)$$

Higher order polynomial estimators can be defined similarly. For example a general homogeneous quadratic estimator (g.h.q.e.) is of the form

$$T : \{T_s = \sum_{\lambda \in s} q_{s\lambda\lambda} Y_\lambda^2 + \sum_{\lambda \neq \lambda' \in s}\sum q_{s\lambda\lambda'} Y_\lambda Y_{\lambda'}\}' \quad ... \quad (3.6)$$

and a g.q.e. is of the form

$$T : \{T_s = \alpha_s + \sum_{\lambda \in s} \beta_{s\lambda} Y_\lambda + \sum_{\lambda \in s} \gamma_{s\lambda\lambda} Y_\lambda^2 + \sum_{\lambda \neq \lambda' \in s} \gamma_{s\lambda\lambda'} Y_\lambda Y_{\lambda'}\}'. \quad ... \quad (3.7)$$

A g.h.l.e. which is unbiased (w.r.t. a given design), say, for Y, is called a g.h.l.u.e. of Y, and similarly for the other. Let $L(P)$, $Q(P)$ and $M_r(P)$ denote the classes of g.l.e.'s, g.q.e.'s and general polynomial estimators of r-th degree respectively and let $L^*(P)$, $Q^*(P)$, $M_r^*(P)$ denote the corresponding classes of estimators unbiased for Y; $M(P) = \bigcup_{r=1}^{\infty} M_r(P)$, and $M^*(P) = \bigcup_{r=1}^{\infty} M_r^*(P)$; and let $L_0^*(P)$ denote the subclass of $L^*(P)$ consisting only of homogeneous linear estimators. Corresponding $Q_0^*(P)$ and $M_r^*(P)$ can be shown to be empty.

The first question that is of interest is regarding the estimability of a given parametric function $f(Y)$ in a given design P. Godambe (1955) proved the following theorem.

Theorem 3.1: *A set of necessary and sufficient conditions* (n.s.c.'s, *for brevity*) *for the estimability of Y in a given design P is that*

$$\pi_i(P) > 0, \quad 1 \leqslant i \leqslant N. \qquad \ldots \quad (3.8)$$

Godambe proved this by restricting to $L_0(P)$ but the proof can be easily carried through for the class of all estimators, as in the proof of Theorem 3.2.

An unbiased estimator of Y, when (3.8) hold good, is the Horvitz and Thompson estimator of Y (Horvitz and Thompson, 1952),

$$\hat{Y}_{HT} = \sum_{\lambda \epsilon s} \frac{Y_\lambda}{\pi_\lambda} \qquad \ldots \quad (3.9)$$

where the sum on the r.h.s. is over all *distinct* units U_λ that belong to s.

As a simple corollary we see that a linear parametric function $l_0 + \sum_1^N l_i Y_i$ is estimable in P iff

$$l_i \neq 0 \Longrightarrow \pi_i(P) > 0, \quad \text{for} \quad 1 \leqslant i \leqslant N. \qquad \ldots \quad (3.10)$$

The above theorem seems intuitively true because if for some i we have $\pi_i(P) = 0$ then the corresponding Y_i is never observable from the design P and hence any parametric function that depends on Y_i cannot be estimated unbiasedly. However, the following theorem though frequently used in a special form in the literature is not equally obvious—at any rate not to the author—and we shall give a direct formal proof of the same.

Theorem 3.2: *A set of* n.s.c.'s *for the estimability of the quadratic parametric function*

$$Q = l_0 + \sum_{i=1}^N l_i Y_i + \sum_i q_{ii} Y_i^2 + \sum_{i \neq j}\sum q_{ij} Y_i Y_j \qquad \ldots \quad (3.11)$$

in a design P is given by

$$\left.\begin{array}{ll} \text{(a)} & \pi_i(P) > 0 \quad \textit{if} \quad l_i^2 + q_{ii}^2 > 0 \\ \textit{and} \quad \text{(b)} & \pi_{ij}(P) > 0 \quad \textit{if} \quad q_{ij} + q_{ji} \neq 0. \end{array}\right\} \qquad \ldots \quad (3.12)$$

Proof: It can be easily verified that (3.2) constitutes a set of sufficient conditions, for, when (3.12) holds good the estimator

$$\hat{Q}_{HT} = l_0 + \sum_{\lambda \epsilon s} \frac{l_\lambda \cdot Y_\lambda}{\pi_\lambda} + \sum_{\lambda \epsilon s} \frac{q_{\lambda\lambda} \cdot Y_\lambda^2}{\pi_\lambda} + \sum_{\lambda \neq \lambda' \epsilon s} \sum \frac{q_{\lambda\lambda'} Y_\lambda Y_{\lambda'}}{\pi_{\lambda\lambda'}} \qquad \ldots \quad (3.13)$$

is unbiased for Q. To prove that the conditions (3.12) are necessary, let G be a statistic unbiased (the design P w.r.t.) for Q so that

$$\sum_{s \epsilon \mathcal{S}} G_s P_s = l_0 + \sum_1^N l_i Y_i + \sum_1^N q_{ii} Y_i^2 + \sum_{\substack{i \neq j \\ 1}}^N \sum q_{ij} Y_i Y_j. \qquad \ldots \quad (3.14)$$

To prove that (a) of (3.12) are necessary, let there exist a $k (1 \leqslant k \leqslant N)$ for which $l_k^2 + q_{kk}^2 > 0$ and $\pi_k(P) = 0$. Since (3.14) is an identity in Y, setting $Y_i = 0$ for $i \neq k$, we see that the r.h.s. of (3.14) equals

$$l_0 + l_k Y_k + q_{kk} Y_k^2$$

which depends on Y_k. But since $\pi_k(P) = 0$ there is no sample s with $P_s > 0$ and containing U_k. Since G_s, being a statistic, can depend only on the $\mathcal{Y}$-values of the units in the sample, the l.h.s. is independent of Y_k. This leads to a contradiction and hence the necessity of (a) of (3.12).

To prove the necessity of (b) of (3.12) (which form the crucial set in this context), as before let there exist k and $k' (1 \leqslant k \neq k' \leqslant N)$, such that $\pi_{kk'}(P) = 0$ and $q_{kk'} \neq 0$. The sum on the l.h.s. of (3.14) can be written as

$$\sum_{s \epsilon S_1} G_s P_s + \sum_{s \epsilon S_2} G_s P_s + \sum_{s \epsilon S_3} G_s P_s + \sum_{s \epsilon S_4} G_s P_s \qquad \ldots \quad (3.15)$$

where

$$S_1 = \{s : U_k \epsilon s, \quad U_{k'} \notin s\}$$

$$S_2 = \{s : U_k \notin s, \quad U_{k'} \epsilon s\}$$

$$S_3 = \{s : U_k \notin s, \quad U_{k'} \notin s\}$$

and

$$S_4 = \{s : U_k \epsilon s, \quad U_{k'} \epsilon s\}.$$

Since $\pi_{kk'}(P) = 0$, S_4 carries zero probability and hence can be omitted. Substituting (3.15) in (3.14) and setting $Y_i = 0$ for $i \neq k, k'$ we have

$$\sum_{s \epsilon S_1} G_s P_s + \sum_{s \epsilon S_2} G_s P_s + \sum_{s \epsilon S_3} G_s P_s$$

$$= l_0 + (l_k Y_k + q_{kk} Y_k^2) + (l_{k'} Y_{k'} + q_{k'k'} Y_{k'}^2) + (q_{kk'} + q_{k'k}) Y_k Y_{k'}. \qquad \ldots \quad (3.16)$$

The l.h.s. of the above can be written as

$$\alpha_1(Y_k) + \alpha_2(Y_{k'}) + \alpha_3$$

where α_1 is a function of Y_k only, α_2 is a function of $Y_{k'}$ only and α_3 is a constant, so that (3.16) can be recast as

$$\phi_1(Y_k) + \phi_2(Y_{k'}) + \phi_3 = (q_{kk'} + q_{k'k}) Y_k Y_{k'} \qquad \ldots \quad (3.17)$$

where ϕ_1, ϕ_2 are functions of Y_k and $Y_{k'}$ respectively and ϕ_3 is a constant. A relation like (3.17) is then clearly impossible and hence the necessity of (b) of (3.12).

This completes the proof of the theorem.

For a given design P let (3.8) hold good and let $T \epsilon L(P)$ be an unbiased estimator of Y. If $\pi_{kk'}(P) = 0$ for some integers k and k' $(1 \leqslant k \neq k' \leqslant N)$, then in the expression

$$V(T) = \sum_{s \in S} T_s^2 P_s - Y^2 \qquad \ldots \quad (3.18)$$

the coefficient of $Y_k Y_{k'}$ equals -2. Hence the following corollary from the above theorem.

Corollary : *$V(T)$ is not estimable in P, and the* n.s.c.'s *are given by*

$$\pi_{ij}(P) > 0 \quad \textit{for} \quad 1 \leqslant i \neq j \leqslant N. \qquad \ldots \quad (3.19)$$

Remark : If (a) of (3.12) hold good then from Theorem 3.1 and the discussion given at the beginning of this section (regarding the change of the variable) it follows that the first three terms of (3.11) are estimable in $D(S, P)$. Hence if Q itself is estimable it follows that the last term of (3.11) also is estimable. Referring again to the beginning of this section regarding the estimability of $Q = \Sigma\ \Sigma\ q_{ij} Y_i Y_j$ it follows from an application of Theorem 3.1 that Q is estimable in $D'(\mathcal{U}', S', P')$ iff

$$\pi((i,j),\ P') > 0 \quad \text{for all} \quad i, j \ \text{ such that } \ q_{ij} + q_{ji} \neq 0$$

where $\pi(i, j), P')$ is the probability of including the unit (U_i, U_j) of $\mathcal{U}'$ as given by the design $D'(\mathcal{U}, S', P')$. From the definition of P' it follows that

$$\pi((i, j), P') + \pi((j, i), P') = \pi_{ij}(P)$$

and from the above we see that Q is estimable in $D'(\mathcal{U}', S', P')$ iff

$$\pi_{ij}(P) > 0 \quad \text{for all} \quad i, j \text{ such that } q_{ij} + q_{ji} \neq 0.$$

However this argument does not suffice to form a rigorous proof of Theorem 3.2 for two reasons. Firstly it has to be established that the necessary conditions for the estimability of Q in $D'(\mathcal{U}', S', P')$ are the same as those for its estimability in $D(\mathcal{U}, S, P)$, and secondly the necessity of (b) of (3.12) given (a) of (3.12) does not mean the necessity of (a) and (b) put together. Because of these logical intricacies, which however can be settled easily, we preferred a direct proof as given above.

Theorems 3.1 and 3.2 can obviously be generalised along the lines of the proof of Theorem 3.2, to be estimability of any parametric function which is a polynomial in $Y_1, Y_2, \ldots, Y_N$.

We can now pass on to the problem of 'optimum' strategies for the estimation of linear parametric functions. Since any such function can be reduced to the form (3.2) by a change of variable, we shall henceforth consider the population total as our parametric function. We shall take, as usual, the squared error as our loss function, and attempt to minimise the expected loss i.e. the mean square error (m.s.e.). Thus the problem is to choose $H(S, P, T)$ such that

$$\text{m.s.e.}\ (H) = \sum_{s \in S} (T_s - Y)^2 P_s \qquad \ldots \quad (3.20)$$

is minimum for all $Y \epsilon R^N$. To avoid trivialities, as also to have a meaningful practical interpretation, we have to restrict H to a class $\mathcal{H}$ of strategies members of which are equally preferable in all respects other than their m.s.e.'s. The concept of cost function enters here and if $C(H)$ denotes the cost of the strategy (in some units) and if C_0 is a given budget, then we restrict H to

$$\mathcal{H}(C_0) = \{H : C(H) = C_0\} \qquad \ldots \quad (3.21)$$

and proceed to find the 'best' strategy $H_0 = H_0(\mathcal{S}, P_0, T_0)$ in $\mathcal{H}(C_0)$ for which

$$\text{m.s.e.}(H_0) \leqslant \text{m.s.e.}(H), \ \forall\ H \epsilon \mathcal{H}(C_0) \text{ and } \forall Y \epsilon R^N.$$

We assume that $C(H)$ depends on H only through P so that $C(H) = C(P)$. At this stage we break up our problem into two steps. First, for a given P we shall choose the P-optimum estimator $T_0(P)$, if it exists, for which

$$\text{m.s.e. } (T_0(P)) \leqslant \text{m.s.e.}(T_1(P)) \qquad \ldots \quad (3.22)$$

for all $Y \epsilon R^N$ and for all $T_1(P)$ belonging to a prescribed class of estimators. We then choose the optimum P, say P_0, for which

$$\text{m.s.e.}(T_0(P_0)) \leqslant \text{m.s.e.}(T_0(P_1))$$

for all P_1 such that $C(P_1) \leqslant C_0$ and for all $Y \epsilon R^N$. We shall first take up the first step in Sections 4 and 5.

4. Unbiasedness, linearity and unicluster designs

If in (3.22) we allow T_1 to vary over the class $\mathcal{T}(P)$ of all statistics then it is evident that there does not exist a T_0 which is optimum in $\mathcal{T}$. This is evident by considering the estimator T_1 which is identically equal to a constant a say. (3.22) then requires that

$$\sum_{s \epsilon \mathcal{S}} T_{0s}^2 P_s - Y^2$$

vanish for all Y such that $Y = a$. If T_0 is the best in $\mathcal{T}$ this should hold good for all such T_1's obtained by varying a and this is clearly impossible.

For any class $\mathcal{C}(P)$ of estimators we define $T_1 \epsilon \mathcal{C}(P)$ to be admissible iff

$$\{T_2 \epsilon \mathcal{C}(P),\ T_2 \neq T_1\} \Longrightarrow \{\exists\, Y^{(0)} = Y^{(0)}(T_1, T_2) \epsilon R^N \ni \text{m.s.e.}(T_1)|_{Y^{(0)}} < \text{m.s.e.}(T_2)|_{Y^{(0)}}\} \qquad \ldots \quad (4.1)$$

where the m.s.e.'s are evaluated at the particular point $Y^{(0)}$. A subclass $\mathcal{C}_1(P) \subset \mathcal{C}(P)$ is said to be complete in $\mathcal{C}(P)$ iff

$$T_2 \epsilon \mathcal{C} - \mathcal{C}_1 \Longrightarrow \{\exists\, T_1 \epsilon \mathcal{C}_1, \ \ni \text{m.s.e.}(T_1) \leqslant \text{m.s.e.}(T_2), \ \ \forall Y \epsilon R^N\}. \qquad \ldots \quad (4.2)$$

The intersection of all complete classes in $\mathcal{C}(P)$, if it exists, is called the minimal complete class in $\mathcal{C}(P)$ and coincides then with the class $\mathcal{C}_a(P)$ of all admissible estimators in $\mathcal{C}(P)$. For an investigation of P-optimum estimators in $\mathcal{C}(P)$ we need restrict ourselves only to a complete class in $\mathcal{C}(P)$. Since a 'best' estimator does not exist in the class $\mathcal{T}(P)$ of all estimators, we should proceed to consider some reasonable subclass of $\mathcal{T}(P)$ and see if a best exists in that subclass.

As a first but a big jump, we shall restrict to the class $L_0^*(P)$ of g.h.l.u.e.'s of Y. It should be noted here that the criterion of unbiasedness is taken for its well known statistical interpretability and mathematical simplicity and that the exclusion of the class B of biased estimators in preference to $L_0^*(P)$ is not a consequence of the completeness of $L_0^*(P)$ in $B \cup L_0^*(P)$. On the other hand we have the following lemma.

Lemma 4.1 : *The class $L_0^*(P)$ is complete in $B \cup L_0^*(P)$ if and only if for the design P we have $\pi_i(P) = 1$ for $1 \leqslant i \leqslant N$.*

Proof: If possible let $L_0^*(P)$ be complete in $B \cup L_0^*(P)$. Consider $T_1 \epsilon B$, given by

$$T_1 = \partial, \text{ a constant not equal to zero.}$$

From hypothesis, there exists a $T_0 \epsilon L_0^*(P)$ such that

$$V(T_0) \leqslant \text{m.s.e. } (T_1) = (\partial - Y)^2$$

for all $Y \epsilon R^N$.

If
$$T_0 = \sum_{\lambda \epsilon s} \beta_{s\lambda} Y_\lambda$$

then
$$V(T_0) = \sum_{\lambda=1}^{N} \Big\{ \sum_{s \supset \lambda} \beta_{s\lambda}^2 P_s - 1 \Big\} Y_\lambda^2 + \sum\sum_{\lambda \neq \lambda'}^{N} \Big\{ \sum_{s \supset \lambda\lambda'} \beta_{s\lambda} \beta_{s\lambda'} P_s - 1 \Big\} Y_\lambda Y_{\lambda'} .$$

For $Y^{(i)} = (0, \ldots, 0, \partial, 0, \ldots, 0)$ where $Y_i = \partial$, we have m.s.e. $(T_1)|_{Y^{(i)}} = 0$ so that $V(T_0)|_{Y^{(i)}}$ should also vanish. Minimising $V(T_0)$ given above for variations of Y we obtain the minimising equations to be

$$\sum_{\lambda=1}^{N} b_{\lambda k} Y_k = 0 \quad \text{for } k = 1, \ldots, N$$

where

$$b_{\lambda k} = \begin{cases} \sum\limits_{s \supset \lambda} \beta_{s\lambda}^2 P_s - 1 & \text{if } k = \lambda \\ \sum\limits_{s \supset \lambda} \beta_{s\lambda} \beta_{sk} P_s - 1 & \text{if } k \neq \lambda. \end{cases}$$

Since the above set of minimising equations has to be satisfied for $Y = Y^{(i)}$ it follows that

$$b_{ii} = \sum_{s \supset i} \beta_{si}^2 P_s - 1 = 0.$$

But from the conditions of unbiasedness of T_0 it can be seen that

$$\sum_{s \supset i} \beta_{si}^2 P_s \geqslant \frac{1}{\pi_i(P)}$$

so that we have $\pi_i(P) = 1$. Considering the vectors $Y^{(i)}$ for $i = 1, 2 \ldots, N$ (since at all these vectors m.s.e. $(T_1) = 0$ and hence $V(T_0) = 0$) we have

$$\pi_i(P) = 1 \quad \text{for} \quad 1 \leqslant i \leqslant N.$$

If these conditions hold, clearly $L_0^*(P)$ has a member with $V(T_0) = 0$. Hence the lemma.

It suffices only to add that the above situation i.e. $\pi_i(P) = 1$ (i.e. a complete census) is utterly uninteresting from the point of view of sampling theory and we conclude that $L_0^*(P)$ is not complete in $B \bigcup L_0^*(P)$ in situations of any interest to us. Thus the restriction to $L_0^*(P)$ is purely arbitrary. Though the inclusion of the criterion of unbiasedness has some important statistical significance and interpretations, no such defence can be put forward for demanding the linearity of the estimators. We shall comment more on these aspects, in Section 5.

Godambe (1955) proved that even when we restrict ourselves to $L_0^*(P)$ there does not exist a best, whatever may be the design P. Later (1965) he excludes the 'trivial design' P for which every sample of positive probability contains the whole population, from this result. However, this is not all the truth and we give a complete characterisation of all sampling designs P that admit a best in $L_0^*(P)$.

Theorem 4.1 : *A set of* n.s.c.'s *for a design P to admit of a best member in $L_0^*(P)$ is that*

$$\left.\begin{array}{ll} (1) & \pi_i(P) > 0, \qquad 1 \leqslant i \leqslant N \\ (2) & P_{s_1} > 0,\ P_{s_2} > 0 \Longrightarrow s_1 \bigcap s_2 = \phi \text{ or } s_1 \sim s_2 \end{array}\right\} \qquad \ldots \quad (4.3)$$

where ϕ denotes the null set and $s_1 \sim s_2$ (in words s_1 and s_2 are effectively equivalent) implies that every unit belonging to s_1 also belongs to s_2 and vice versa.

The best estimator in $L_0^*(P)$ for P satisfying (4.3) *is the corresponding Horvitz and Thompson estimator defined by* (3.9).

Proof: From Theorem 3.1 it follows that (1) of (4.3) are n.s.c.'s for $L_0^*(P)$ to be nonempty. We need verify (2) of (4.3).

Let
$$T : \left\{ T_s = \sum_{\lambda \epsilon s} \beta_{s\lambda} Y_\lambda \right\} \qquad \ldots \quad (4.4)$$

be any member of $L_0^*(P)$. The conditions for unbiasedness of T, as can be easily verified, are given by

$$\sum_{s \supset \lambda} \beta_{s\lambda} P_s = 1, \quad 1 \leqslant \lambda \leqslant N \qquad \ldots \quad (4.5)$$

where the sum on the l.h.s. of (4.5) is over all samples that contain U_λ. For the variance of T we have

$$V(T) = \sum_{s \epsilon \mathcal{S}} T_s^2 P_s - Y^2. \qquad \ldots \quad (4.6)$$

If there exists a best in $L_0^*(P)$, say

$$T_0 : \left\{ T_{0,s} = \sum_{\lambda \epsilon s} \bar{\beta}_{s\lambda} Y_\lambda \right\},$$

it is obtained by minimising (4.6) for variations of $\beta_{s\lambda}$'s subject only to (4.5). Introducing the Lagrangian multipliers $\alpha_1, \ldots, \alpha_N$, we seek to minimise

$$\phi = \sum_{s \epsilon \mathcal{S}} T_s^2 P_s - Y^2 - \sum_{\lambda=1}^{N} \alpha_\lambda \Big(\sum_{s \supset \lambda} \beta_{s\lambda} P_s - 1 \Big) \qquad \ldots \quad (4.7)$$

with respect to $\beta_{s\lambda}$'s and α_λ's. The minimising equations are given by (4.5) and

$$\frac{\partial}{\partial \beta_{s\lambda}} \{T_s^2 P_s\}\,|_{T=T_0} - \alpha_\lambda P_s = 0$$

for $1 \leqslant \lambda \leqslant N$ and for every $s \supset \lambda$. From (4.4) this gives

$$2T_{0s} P_s Y_\lambda - \alpha_\lambda P_s = 0\,, \quad 1 \leqslant \lambda \leqslant N,\ s \supset U_\lambda$$

and if $P_s \neq 0$

$$T_{0s} = \frac{\alpha_\lambda}{2Y_\lambda}.$$

This implies that for any two samples s_1 and s_2 that have a unit in common i.e. $s_1 \cap s_2 \neq \phi$, and for which $P_{s_1} > 0$ and $P_{s_2} > 0$, we should have

$$T_{0s_1} = T_{0s_2} \ \forall\ Y \in R^N. \qquad \ldots \quad (4.8)$$

At this point Godambe closes his argument saying that this is clearly impossible. However, we shall carry the argument further.

To prove that (2) of (4.3) are necessary we need observe (see below) that T_{0s} has to take into account all the $\mathcal{Y}$-values of all the units that belong to s. If then for two samples s_1 and s_2 (2) of (4.3) is violated but (4.8) is satisfied, we have $P_{s_1} > 0$, $P_{s_2} > 0$, $s_1 \cap s_2 \neq \phi$ and s_1 is not effectively equivalent to s_2, so that there is a U_{λ_0} belonging to, s_1 say, but not belonging to s_2. In this case T_{0s_1} does depend on $Y_{\lambda 0}$ while T_{0s_2} cannot depend on $Y_{\lambda 0}$, which contradicts (4.8). Thus (2) of (4.3) are necessary.

To prove the sufficiency part we use the fact (cf. Theorem 5.1) that every admissible estimator in L_0^* (and in fact in the class $\mathcal{T}^*$ of all unbiased estimators of Y) must satisfy the condition

$$T_s \equiv T_{s'}$$

for all s and s' for which $\qquad P_s > 0, \quad P_{s'} > 0 \quad \text{and} \quad s \sim s'. \qquad \ldots \quad (4.9)$

Thus in $L_0^*(P)$, T given by (4.4) is admissible only if

$$\beta_{s\lambda} = \beta_{s'\lambda}$$

for all s and s' for which $\qquad P_s > 0, \quad P_{s'} > 0 \quad \text{and} \quad s \sim s'. \qquad \ldots \quad (4.10)$

If T is an admissible member of L_0^*, we then have from (4.5)

$$1 = \sum_{s \supset \lambda} \beta_{s\lambda} P_s = \sum_{s \sim s_0} \beta_{s_0\lambda} P_s \text{ for any } s_0 \supset \lambda$$

(because, from (2) of (4.3), since $s_0 \supset \lambda$ and $s \supset \lambda, s \sim s_0$)

$$= \beta_{s_0\lambda} \sum_{s \sim s_0} P_s \quad \text{from} \quad (4.10)$$

$$= \beta_{s_0\lambda}\, \pi_\lambda$$

so that for any $s_0 \supset \lambda$, we have

$$\beta_{s_0 \lambda} = \frac{1}{\pi_\lambda}. \qquad \dots \quad (4.11)$$

This shows that the only possible admissible estimator in $L_0^*(P)$ is given by

$$\hat{Y}_{HT} = \sum_{\lambda \epsilon s} \frac{Y_\lambda}{\pi_\lambda}.$$

But it is known (Godambe, 1960 ; Roy and Chakravorty, 1960) that $\hat{Y}_{HT}$ is admissible in $L_0^*(P)$ for any P satisfying (1) of (4.3). Hence $\hat{Y}_{HT}$ is the best, i.e. uniformly minimum variance estimator, in $L_0^*(P)$. This proves that (2) of (4.3) are also sufficient. Hence the theorem.

We shall term sampling designs satisfying (4.3) as unicluster designs. The term is derived from the analogy of those designs to designs obtained by cluster sampling with one cluster being chosen. With a general definition of cluster as a subset of the population a design satisfying (2) of (4.3) can be looked upon as one obtained by sampling one cluster from $\mathcal{U}$, if one treats all samples that are effectively equivalent, as a single sample.

It can be easily verified that, except for trivial designs for which $\pi_i(P) = 1$ for all i, there exist pairs i and j such that $\pi_{ij} = 0$.

For unicluster designs we can in fact prove a stronger result fully characterising the class of all admissible estimators in the class $M^*(P)$ of all polynomial unbiased estimators of Y, which is much wider than $L_0^*(P)$. We have the following theorem.

Theorem 4.2 : *For any unicluster design any estimator T_0 in $M^*(P)$ is admissible in $M^*(P)$ iff*

$$T_0 = K_s + \hat{Y}_{HT}(P) \qquad \dots \quad (4.12)$$

where K_s's are constants (i.e. *independent of Y*) *satisfying*

$$\sum_{s \epsilon \mathcal{S}} K_s P_s = 0.$$

Proof: From Theorem 5.1 of Section 5 we know that every admissible estimator T is of the form (4.9).

*This result is also obtained independently by Vijaya Hege (1965).

Let T be any admissible estimator in $M^*(P)$. Since there are only (2^N-1) possible equivalence classes of samples (i.e. $s \sim s'$ for s and s' belonging to the same equivalence class) T, which is a polynomial in $Y_{i_1}, \ldots, Y_{i_{n(s)}}$ for every s, has a finite upper bound for the degree of T_s. Let this upper bound be r and let

$$T : T_s = T_{0s}+T_{1s} = \ldots +T_{rs} \qquad \ldots \quad (4.13)$$

where T_{0s} is a constant (independent of Y) and T_{ks} is a homogeneous polynomial of degree k in the y-values of the units belonging to the sample. From the unbiasedness (for Y) of T follows that

$$E(T_1) = Y$$

and

$$E(T_k) = 0 \quad \text{for} \quad k \neq 1 \qquad \ldots \quad (4.14)$$

we shall prove that

$$T_k \equiv 0 \quad \text{for} \quad k \neq 0, 1.$$

Considering the case $k = 2$, let

$$T_{2s} = \sum_{\lambda \epsilon s} \gamma_{s\lambda\lambda} Y_\lambda^2 + \sum_{\lambda \neq \lambda'} \sum \gamma_{s\lambda\lambda'} Y_\lambda Y_{\lambda'}$$

and if possible let $T_2 \not\equiv 0$ so that there exists a sample s with $P_s > 0$ for which

$$\gamma_{s\lambda_0\lambda_0} \neq 0 \quad \text{for some} \quad U_{\lambda_0} \epsilon s \qquad \ldots \quad (4.14a)$$

or

$$\gamma_{s\lambda\lambda'} \neq 0 \quad \text{for some} \quad U_\lambda, U_{\lambda'} \epsilon s. \qquad \ldots \quad (4.14b)$$

Since the design P is assumed to be a unicluster design satisfying (1) and (2) of (4.3) any other sample s' (with $P_{s'} > 0$) for which $s \cap s' \neq 0$ is such that $s \sim s'$. However, T being admissible, from (4.9) we have

$$T_s \equiv T_{s'}$$

so that

$$T_{k,s} \equiv T_{k,s'} \quad \text{for } 0 \leqslant k \leqslant r,$$

and in particular

$$T_{2s} \equiv T_{2s'}$$

so that

$$\gamma_{s\lambda_0\lambda_0} = \gamma_{s'\lambda_0\lambda_0} \text{ for any } \quad s' \supset \lambda_0 \qquad \text{if (4.14a) holds}$$

and

$$\gamma_{s\lambda\lambda'} = \gamma_{s'\lambda\lambda'} \quad \text{for any} \quad s' \supset \lambda, \lambda' \quad \text{if (4.14b) holds.}$$

The term containing Y_λ^2 in $E(T_2)$ is

$$\gamma_{s\lambda_0\lambda_0} Y_{\lambda_0}^2 \pi_{\lambda_0} \neq 0$$

or that for $Y_\lambda Y_{\lambda'}$ is

$$\gamma_{s\lambda\lambda'} Y_\lambda Y_{\lambda'} \pi_{\lambda\lambda'} \neq 0.$$

In either case this contradicts (4.14). Thus $T_2 \equiv 0$ and similarly it can be proved that

$$T_k \equiv 0 \quad \text{for} \quad k \neq 0, 1.$$

As already contained in the proof of Theorem 4.1, it follows that

$$T_1 \equiv \hat{Y}_{HT}(P) = \sum_{\lambda \epsilon s} \frac{Y_\lambda}{\pi_\lambda}.$$

This shows that the only possible admissible estimators of Y in $M^*(P)$ are of the form (4.12). That every such estimator is admissible in $M^*(P)$ can easily be proved by comparing, without loss of generality, any two members of the form (4.12). Hence the theorem.

Even though restriction to unicluster designs solves our first step mentioned at the end of Section 3, at any rate when we restrict ourselves to $L_0^*(P)$, these designs have a serious draw-back. In practice we not only aim to estimate Y by means of an estimator $\hat{Y}$ say, in an optimum way if we can help it, but would also like to have an estimate of $V(\hat{Y})$ to be able to know the precision of our estimate $\hat{Y}$. From corollary of Theorem 3.2 and from Theorem 4.1 we see that this is not possible for unicluster designs. This being a serious limitation, we turn our attention to designs for which

$$(1) \quad \pi_i > 0 \quad \text{for} \quad 1 \leqslant i \leqslant N$$

and

$$(2) \quad \pi_{ij} > 0 \quad \text{for} \quad 1 \leqslant i \neq j \leqslant N, \qquad \ldots \quad (4.15)$$

to see what we can do in these designs.

5. Sufficiency of the effective sample

Since we noted from Theorem 4.1 that for designs satisfying (4.15) there is no best even in the restricted class $L_0^*(P)$ attention should now be focussed on finding the minimal complete class in $L_0^*(P)$ to see if a 'best' exists in some of these subclasses. Apart from the unnaturalness (from mathematical point of view) such attempts did not yield fruitful results. In fact restricting ourselves to such a narrow class as $L_0^*(P)$ itself is not a justifiable course and is an unnatural mathematical restriction, but we shall reserve our comments on this aspect for a latter section. Another alternative is to lay down some new criteria of optimality other than that of uniform (in Y) minimisation of variance, which are both reasonable and fruitful. This aspect we shall discuss in Section 6.

Steps towards characterising inadmissible estimators were started by various authors at more or less the same time. Murthy (1957) proved that when the design is one generated by the customary 'probability proportional to size (pps)' sampling without replacement, estimators that take into account the order in which the units occur in the sample are inadmissible. He also furnished a method of getting uniformly better estimators in such cases. Sometime about that time Raja Rao noticed, first through empirical evidences, that for designs generated by simple random sampling with replacement, the sample mean is inadmissible as an estimator of the population mean and that it is inferior to the mean over the *distinct* units of the sample. Later, in collaboration with Basu (1958), he proved the result and extended it to prove that for designs generated by the customary pps sampling with replacement, the conventional estimator of the population total (which takes into account the number of repetitions of the units in the sample) is inadmissible. For the case of simple random sampling this result is also proved by Des Raj and Khamis (1958). Roy and Chakravorty (1960) proved that for any sampling design in which Y is estimable, admissible members in $L_0^*(P)$ must satisfy (4.9). Basu (1958) gave the first clues to the generality when he introduced the fruitful notion of a sufficient statistic, in this field and proved that the 'effective sample' by which we mean the unordered set of distinct units contained in the sample, together with the corresponding $\mathcal{Y}$-values, forms a sufficient statistic. (Basu terms this as the 'order—statistic' but since this confuses with another popular meaning that the term has in statistics, we shall avoid it. Moreover the 'order' in Basu's 'order-statistic' is not really relevant). The main ideas behind this important result can be briefly explained thus. The variable $\mathcal{Y}$ *operating* on a sample

$$s = \{U_{i_1}, U_{i_2}, \ldots, U_{i_{n(s)}}\}$$

gives rise to the sampley

$$(s, Y) = \{(U_{i_1}, Y_{i_1}); (U_{i_2}, U_{i_2}); \ldots; (U_{i_{n(s)}}, Y_{i_{n(s)}})\}. \qquad \ldots \quad (5.1)$$

Godambe (1965) uses the term 'observable' to denote this, but we shall avoid that terminology because of its much deeper meaning in theoretical physics. The above terminology is simple and being new does not confuse with any existing ones. Moreover it seems convenient when we consider more than one variable. For example with the variables $\mathcal{X}$ and $\mathcal{Z}$ we can have 'samplex' and 'samplez' respectively).

The basic sample space gives rise to the basic sampley space

$$(\mathcal{S}, \mathcal{Y}) = \{(s, Y) : s \epsilon \mathcal{S}, Y \epsilon R^N\}. \qquad \ldots \quad (5.2)$$

From the given probability measure P on S i.e. the given design $D(S, P)$ is now generated the family $\mathcal{P}_Y$ of probability measures on the basic sampley space $(\mathcal{S}, \mathcal{Y})$ with parameter Y belonging to the parameter space R^N. For a point $Y^{(0)} = (Y_1^{(0)}, \ldots, Y_N^{(0)})$ in R^N this measure assigns the probabilities thus giving the *likelihood* (Takeuchi, 1961):

$$\mathcal{P}_{Y^{(0)}}(s, Y) = \begin{cases} P_s & \text{if} \quad Y_{i_k} = Y_{i_k}^{(0)} \text{ for } 1 \leqslant k \leqslant n(s) \\ \\ 0 & \text{otherwise.} \end{cases} \qquad \ldots \quad (5.3)$$

Defining the effective sampley corresponding to the sampley (5.1) by

$$(s, Y)_d = [(U_{j_1}, Y_{j_1}), (U_{j_2}, Y_{j_2}), \ldots, (U_{j_{\nu(s)}}, Y_{j_{\nu(s)}})] \qquad \ldots \quad (5.4)$$

where $\nu(s)$ is the effective size of s (as defined in Section 1 and $U_{j_1}, \ldots, U_{j_{\nu(s)}}$ are the distinct units in s arranged in any order, it can now be verified, by the usual factorization criterion of Neyman that the statistic $(s, Y)_d$ is a sufficient statistic for the family of probability distributions $\mathcal{P}_Y$. These ideas are transparent in Basu's work though he limited his exposition to pps designs. The generality of his results was quickly pointed out by Takeuchi (1961) and is also evident from the work of Roy and Chakravorti (1960). However, there seems to exist some confusion among some regarding these concepts as when Pathak (1964) set out to rigorise these concepts but is no clearer (Hájek, 1965).

An application of Rao-Blackwell theorem now yields Basu's theorem.

Theorem 5.1. : *Given a design P and an estimable parametric function $f(Y)$, if T is any unbiased estimator* (w.r.t. P) *of $f(Y)$, which violates* (4.9) *then the estimator*

$$T^* = E\,(T\,|\,(s, Y)_d)$$

is also unbiased for $f(Y)$ and

$$V(T^*) \leq V(T) \;\; \forall\, Y \epsilon R^N$$

with strict inequality at least once. The results of Murthy, Raja Rao, Des Raj and Khamis and Roy and Chakravorty are special cases of the above result.

It may be noted that Hajék (1959) seems to have intuitively felt the truth of the above theorem at about the same time but only makes a passing mention of it and at any rate does not indicate any lines of proof. Some of the essential features of the above theorem are also traceable, though in a disguised form, in a much earlier work of Halmos (1946) for the special case of pps designs.

It can be seen that the statistic $(s, Y)_d$ is not only sufficient for the family $\mathcal{P}_Y$ but is in fact the minimal sufficient statistic. It seems safe to conjecture from this that any non-zero function of the minimal sufficient statistic $(s, Y)_d$ is an admissible estimator of its own expectation. The validity or otherwise of this result in a general context seems to be an interesting question which does not seem to have been answered. If this conjecture is true—at any rate in our setup—then we not only have the complete characterisation of all admissible estimators but also recognise that the class $L_0^*(P)$ does not enjoy any special privilege in the wider class $M^*(P)$, as it is easy to construct members of $M^*(P)-L_0^*(P)$ that are functions of $(s, Y)_d$ only and hence are admissible in $T^*(P)$, the class of all unbiased estimators of Y. Further, if this conjecture is true, even for the class $L_0^*(P)$, Theorem 5.1 gives us a big minimal complete class of estimators (those satisfying (4.9)). We note that there are (2^N-1) possible equivalence classes of samples and for a sample with $\nu(s) = r$ we have r coefficients $\beta_{s i}$'s to be defined. These give $N.2^{N-1}$ coefficients $\left(\sum_{r=1}^{N}\binom{N}{r}.\, r\right)$ to be determined subject to the N conditions given by (4.5). Thus defining optimality as the uniform minimisation

13

of the variance does not lead us to a narrow complete class of estimators. Restriction to a suitable subclass of $L_0^*(P)$ is, as remarked earlier, neither desirable nor fruitful. The alternative therefore is to explore some other criteria of optimality.

6. Some alternative criteria of optimality

(a) *Bayesian approach.* In a number of practical situations we are not totally in the dark about the value of Y. The value of $X = (X_1, X_2, \ldots, X_N)$ of a positive character $\mathscr{X}$ on $\mathscr{U}$, which is well correlated with Y is available beforehand in these cases. In such situations one can look for stochastic model and assume that the actual value of Y is the realisation of a random variable whose distribution depends on X besides possibly on some unknown parameters. If this *a priori distribution* is denoted by θ, instead of trying to minimise $V(H(S, P, T))$ uniformly in Y, for variations of H over the class $\mathscr{H}(\nu_0)$ defined by (3.20), we can try to minimise the 'expected loss'

$$\int V(T)d\theta \qquad \ldots \quad (6.1)$$

over the distribution θ. If $H_0(S, P_0, T_0)$ is the minimising member in $\mathscr{H}(\nu_0)$, uniformly for all values of $X \epsilon R^N$ and for all values of all the unknown parameters that enter into the *apriori* distribution θ, then H_0 is said to be a *optimun* strategy in $\mathscr{H}(\nu_0)$.

From Godambe (1955) and Hàjek (1959) we have two important situations in which such an optimum strategy exists.

Let ν_0 be a positive integer and let $\mathscr{H}^*(\nu_0)$ denote the class of strategies for which

$$\nu(P) = \nu_0. \qquad \ldots \quad (6.2)$$

When $C(H)$ is a function of P, say $C(P)$, and is a monotonic increasing function of $\nu(P)$ this class $\mathscr{H}^*(\nu_0)$ coincides with $\mathscr{H}(\nu_0)$ defined by (3.21) with suitable units chosen for the cost.

Let Θ_1 be the class of *a priori* distributions θ_1 for which

$$(1) \quad E_{\theta_1}(Y_i | X_i) = aX_i$$

$$(2) \quad V_{\theta_1}(Y_i | X_i) = \sigma^2 X_i^2$$

and
$$(3) \quad \text{cov}_{\theta_1}(Y_i, Y_j | X_i, X_j) = 0 \qquad \ldots \quad (6.3)$$

while Θ_2 is the wider class with (3) of (6.3) replaced by

$$(3') \quad \text{cov}_{\theta_2}(Y_i, Y_j | X_i, X_j) = w(|j-i|) \qquad \ldots \quad (6.4)$$

where w is any convex function of $|j-i|$. In the above, a and σ^2 are unknown parameters.

Godambe then proved (cf. Hanurav, 1965) the following theorem.

Theorem 6.1 : *In* $\mathcal{H}^*(\nu_0)$ *defined by* (6.2), *any strategy* $H_0(S, P_0, T_0)$ *satisfying*

$$\left.\begin{array}{ll} \text{(i)} & \nu_s(P_0) = \nu(P_s) = \nu_0 \text{ for all 's' with } P_s > 0 \\ \text{(ii)} & \pi_i(P_0) = \nu_0 \dfrac{X_i}{X} \\ \text{and (iii)} & T_0 = \hat{Y}_{HT}(P_0) \end{array}\right\} \quad \ldots \quad (6.5)$$

is O_1*-optimum in* $\mathcal{H}^*(\nu_0)$ *for any* $O_1 \epsilon \Theta_1$.

Hájek proved the following theorem.

Theorem 6.2 : *There is just one strategy* $H_0(S, P_0', T_0')$ *which is* O_2*-optimum in* $\mathcal{H}^*(\nu_0)$ *for any* $O_2 \epsilon \Theta_2$ *and that it is given, in addition to* (6.5), *by*

(iv) *the sampling is by means of the pps systematic sampling.* ... (6.6)

The above sampling method was first given by Goodman and Kish (1950) as a generalisation of systematic sampling to the varying probability case.

Designs satisfying (ii) of (6.5) we shall term as 'π ps' designs, in analogy with but distinct from 'pps' designs. The problem of constructing easy sampling methods to achieve (i) and (ii) of (6.5) and to satisfy some other desirable properties like admitting a stable nonnegative unbiased estimator of $V(T_0)$ is an interesting combinatorial problem in itself. A solution to this problem, that has several other desirable properties, is given for the case $\nu = 2$ (Hanurav, 1965). The case of general integral values of ν has also been solved recently (cf. footnote on page 183).

There do not seem to be further examples of realistic families of distributions, Θ_1 and Θ_2, for which O-optimum strategy exists. It can be easily proved that if (2) of (6.3) is replaced by the more general (and realistic) condition

$$V_{\theta_2}(Y_i | X_i) = \sigma^2 X_i^g, \quad 1 \leqslant i \leqslant N \quad \ldots \quad (6.7)$$

for some $g > 0$, then O_2-optimum strategies do not exist for $g \neq 2$.

(b) *Linear invariance.* This concept, discussed by Roy and Chakravorty (1960), requires that an estimator should remain invariant under linear transformations on Y. However, this does not lead to an optimum estimator (i.e. minimum variance estimator in the class of linearly invariant estimators), nor to a narrow enough complete class of estimators even, in $L_0^*(P)$.

(c) *Regular estimators.* This class is also discussed by Roy and Chakravorty, and restricts to the class of estimators T of Y for which

$$V(T) = k \cdot \sigma^2 = k \cdot \left(\frac{1}{N} \sum_1^N Y_i^2 - \bar{Y}^2 \right). \quad \ldots \quad (6.8)$$

This seems to us to be a very unnatural and costly demand. It is only for a narrow class of very specialised designs which they termed as 'balanced designs' that they proved that a 'best' estimator exists in the class of regular estimators.

Other criteria like the minimax principle or the principle of maximum regret etc., also fail to lead us to our goal.

(d) *Hyper admissibility.* While (b) and (c) above try to restrict us to subclasses of $L_0^*(Y)$ in an attempt to get at optimum estimators, in this criterion (Hanurav, 1965) we weaken our criterion of uniform minimum variance (which proved too strong for us). Looking upon an unbiased estimator T of Y as an unbiased estimation procedure that can be used to estimate all linear parametric functions $\Sigma l_i Y_i$(cf. Theorem (3.1)) by replacing Y_i's occurring in T by $l_i Y_i$'s, we demand from T that it should give an admissible estimator not only of Y but of all linear parametric functions. This criterion is thus weaker than uniform minimisation of variance, but is stronger than admissibility. We note that when T is used to estimate $\Sigma l_i Y_i$, by deriving from T the estimator T^* obtained by replacing Y_i's in T by $l_i Y_i$'s, the parameter space shrinks to the principal hyperplane containing the coordinate axes of those Y_i's for which l_i's are non-zero. Thus hyperadmissibility requires T to be admissible not only in the whole of R^N but also in each of its principal hyperplanes. The practical implication of this criterion is that sub-totals, means of subpopulations, contrasts involving such submeans (which all form an important class of parametric functions of interest in practice) should all be admissibly estimable, by means of a single estimator T.

We now have the following theorem.

Theorem 6.3 (Hanurav, 1965) : *For any design P which is not a unicluster design the class $M^*(P)$ of all polynomial unbiased estimators of Y admits just one estimator which is hyperadmissible. This 'optimum' estimator is Horvitz-Thompson estimator $\hat{Y}_{HT}(P)$. For any unicluster design the class of all hyper-admissible estimators is given by* (4.12).

As we have remarked earlier, the restriction to the class $L_0^*(P)$ of g.h.l.u.e.'s, for the estimation of Y, is an unnatural mathematical restriction especially since $L_0^*(P)$ is not complete in the wider class $M^*(P)$. While the criterion of unbiasedess can be retained owing to the simple and meaningful interpretations that can be given to it, no such reason can be put forward for the criterion of homogeneous linearity. One reason (and perhaps the only one carrying some weight) that is often advanced in favour of this criterion is based on the units of measurement. If for example $\mathcal{Y}$ is variable like income measured in rupees say, it is clearly difficult to interpret a quadratic estimator to estimate Y for the former is a sum of a constant, rupees and (rupees)^2 while Y is in rupees. However, once the units are chosen for the measurement of $\mathcal{Y}$ the problem should be considered as a purely mathematical problem. Moreover, the use of the well-known 'difference estimator'

$$\hat{Y}_{dif} = N(\bar{y}+k(\bar{x}-\bar{X})) \qquad \dots \quad (6.6)$$

in simple random sampling, where $\bar{x}$ and $\bar{y}$ are the sample means of $\mathscr{X}$ and $\mathscr{Y}$, $\bar{X}$ is the population mean of $\mathscr{X}$ and k is a constant, shows that this adherence to units of measurement is not as pious as one makes it to be when restricting to $L_0^*(P)$: For, in (6.9), $\mathscr{X}$ and $\mathscr{Y}$ can be variables which are not measurable in the same units as for example $\mathscr{Y}$ is income and $\mathscr{X}$ is population size.

Theorem 6.3 not only eliminates through a purely mathematical criterion these non-linear estimators, but in fact pins down the optimum estimator uniquely for all designs which are of interest. Thus it completely accomplishes the first step in the choice of an optimum strategy.

It is interesting to note that the estimator $\hat{Y}_{HT}(P)$ plays a crucial role throughout our theory as is evident from Theorems 4.1, 4.2, 6.1, 6.2 and 6.3.

7. Choice of optimum design

The second (and in fact last) step in the problem of the choice of optimum strategy, is that of the choice of optimum designs. With the criteria of unbiasedness and minimum variance only, the only logical way of asserting that a design $D_1(S, P_1)$ is better than another, $D_2(S, P_2)$ is to establish that

$$V(T_1) \leqslant V(T_2) \forall Y \epsilon R^N \quad \dots \quad (7.1)$$

for *any* estimators T_1 and T_2 that are unbiased (w.r.t. P_2 and P_1 respectively) for Y, or by establishing a weaker but perhaps more meaningful result like

$$V(T_1) \leqslant V(T_2) \forall Y \epsilon R^N \quad \dots \quad (7.2)$$

where T_1 and T_2 are *any* admissible estimators. This is the correct formulation as we can always pick up bad estimators in any design. One can perhaps restrict T_1 and T_2 to some classes like, say, $L_0^*(P_1)$ and $L_0^*(P_2)$. We are not aware of any result that is anywhere near such a logical method even for simple types of designs P_2 and P_1, but a number of authors freely make statements like "... it is evident that sampling without replacement is better than sampling with replacement ...". What such authors normally do is to establish (7.1) for *some* estimators T_1 and T_2 that are perhaps commonly used (of which T_2 is invariably an inadmissible estimator in $L_0^*(P_2)$) and then jump to statements like these made above.

By far the only criterion that gives optimum designs also is the one discussed in Section 6(a), through a Bayesian approach. But even this gives the optimum designs in rather specialised situations which are given in Theorems 6.1 and 6.2.

If instead of unbiasedness and minimum variance, we take unbiasedness and hyperadmissibility, the problem is considerably simplified. For, in this case we need not prove such strong results like (7.2) but need only prove that

$$V(\hat{Y}_{HT}(P_1)) \leqslant V(\hat{Y}_{HT}(P_2)), \ \forall Y \epsilon R^N \quad \dots \quad (7.3)$$

to establish the superiority of P_2 over P_1. For the variance of $\hat{Y}_{HT}(P)$ we have

$$V(\hat{Y}_{HT}) = \sum_{1}^{N} \frac{Y_i^2}{\pi_i} + \sum_{i \neq j}\sum \frac{Y_i Y_j}{\pi_i \pi_j} \pi_{ij} - Y^2. \quad \ldots \quad (7.4)$$

(7.3) then gives the required condition as

$$\sum_{i=1}^{N} Y_i^2 \left(\frac{1}{\pi_i(P_1)} - \frac{1}{\pi_i(P_2)} \right) + \sum_{i \neq j}\sum Y_i Y_j \left(\frac{\pi_{ij}(P_1)}{\pi_i(P_1)\, \pi_j(P_1)} - \frac{\pi_{ij}(P_2)}{\pi_i(P_2)\, \pi_j(P_2)} \right) \leqslant 0 \;\forall\; Y. \quad \ldots \quad (7.5)$$

A set of necessary conditions for (7.5) to hold good are given by

$$\pi_i(P_1) \geqslant \pi_i(P_2), \quad 1 \leqslant i \leqslant N. \quad \ldots \quad (7.6)$$

Thus it is not sufficient, as perhaps is intuitively felt, that ν_1 should be larger than ν_2.

However, for a fair comparison between P_2 and P_1 regarding the variances of $\hat{Y}_{HT}$'s in these designs, it is necessary to ensure that P_2 and P_1 are equally preferrable in other respects and in particular that they are equally costly. If the cost $C(P)$ of a design is taken to be a monotonic increasing function of $(\pi_1, \pi_2, \ldots, \pi_N)$, such as the expected effective size ν, then the condition

$$C(P_2) = C(P_1)$$

together with (7.6) yields

$$\pi_i(P_1) = \pi_i(P_2) = \pi_i \quad \text{say,} \quad \text{for} \quad 1 \leqslant i \leqslant N \quad \ldots \quad (7.7)$$

as a set of necessary conditions for either P_2 to be superior to P_2 or P_1 to be superior to P_1. From (7.5) and (7.7) we then have

$$\sum_{i \neq j}\sum \frac{Y_i Y_j}{\pi_i \pi_j} (\pi_{ij}(P_1) - \pi_{ij}(P_2)) \leqslant 0 \;\forall\; Y \quad \ldots \quad (7.8)$$

as a set of necessary and sufficient conditions for P_2 to be superior to P_1. If we restrict to $Y \geqslant 0$ —often realistically—a set of sufficient conditions for (7.8) to hold good is given by

$$\pi_{ij}(P_1) \leqslant \pi_{ij}(P_2), \quad \text{for} \quad 1 \leqslant i \neq j \leqslant N. \quad \ldots \quad (7.9)$$

Given the design $D_2(S, P_2)$ if we can construct a $D_1(S, P_1)$ with the same π_i's but with uniformly smaller (or equal) π_{ij}'s then such a P_2 is superior to P_1. Referring to (2.6a) and (2.6b) we see that this is not possible if

$$\sum_{i \neq j}\sum \pi_{ij}(P_2)$$

attains its lower bound given in (2.6b) viz., $\nu(\nu-1)+\theta(1-\theta)$. Whether this is possible if $\sum_{i \neq j}\sum \pi_{ij}(P_2)$ exceeds its lower bound is an open problem, but a plausible conjecture is that it is possible. If so this provides a valid justification for restricting to 'without replacement' designs in preference to the 'with replacement' designs, in a very general sense. (Of course the 'without replacement' designs possess the practical advantage of a relatively stable cost of sampling, in comparison to the with replacement designs).

The above discussion as also Theorem 6.1 should now provide ample proof of the importance of the problems of existence and construction of designs that are considered at the end of Section 2.

From (7.6) we see that a given class $\mathcal{H}$ of strategies H for which

$$C(H) = C_0 \quad \text{for} \quad H \epsilon \mathcal{H} \qquad \ldots \quad (7.10)$$

where C_0 is a given number and $C(H)$ is a given cost function which is a monotonic increasing function of $(\pi_1, \pi_2, \ldots, \pi_N)$, then we can only break up $\mathcal{H}$ into equivalence classes with members in the same class having the same values of π_i's which satisfy (7.10). No member of a class is better than a member of another class so that no optimum strategy exists. If the conjecture given in the last para proves true then we have optimum strategies within each subclass. The existence of an optimum strategy in $\mathcal{H}$ as a whole implies the existence of an optimum set of π_i's. In absence of any auxiliary information there does not seem to exist such a set under the present definitions of optimality. In such cases the only possible optimum set, with any reasonable definition of optimality, seems to be the equal values of π_i's. Coupled with the earlier conjecture this gives rise to the simple solution of 'simple random sampling without replacement' design with the estimator $N\bar{y}$, $\bar{y}$ being the sample mean, as the optimum strategy. A third step that now emerges in our investigation is to discover various realistic *a priori* distributions in terms of auxiliary informations which give rise to optimum values of π_i's.

References

Basu, D. (1958): On sampling with and without replacement. *Sankhyā*, 20, 287-294.

Des Raj and Khamis, S. H. (1958): Some remarks on sampling with replacement. *Ann. Math. Stat.*, 29, 550-557.

Godambe, V. P. (1955): A unified theory of sampling from finite populations. *J. Roy. Stat., Soc.* B, 17, 269-277.

——— (1960): An admissible estimator for any sampling design. *Sankhyā*, 22, 285-288.

——— (1965): A review of the contributions towards a unified theory of sampling from finite populations. *Rev. Int. Stat. Inst.*, 33, 242-258.

Goodman, L. A. and Kish, L. (1950): Controlled selection–a technique in probability sampling. *Jour. Amer. Stat. Assoc.*, 45, 350-372.

Hájek, J. (1959): Optimum strategy and other problems in probability sampling. *Casopis Pest Math.*, 84, 387-423.

——— (1965): *Math. Reviews*, 29, 142.

HALMOS, P. R. (1946) : The theory of unbiased estimation. *Ann. Math. Stat.*, **17**, 34-43.

HANURAV, T. V. (1962a) : An existence theorem in sampling. *Sankhyā*, Series A, **24**, 227, 230.

——— (1962b) : Some some sampling schemes in probability sampling. *Sankhyā*, Series A, **24**, 421–328.

——— (1962c) : On Horvitz and Thompson estimator. *Sankhyā*, Series A, **24**, 429-436.

——— (1965) : Optimum sampling strategies. Unpublished Ph.D. thesis.

——— (1966) : Optimum utilisation of auxiliary information. To appear in *J. Roy. Stat. Soc.*, B.

HEGE, V. (1965) : A note on UMV estimator of mean of finite population. *Bull. Cal. Stat. Assoc.*

——— (1965a) : Sampling designs which admit uniformly minimum variance. *Bull. cal. Stat. Assoc.*, **14**, 160-162.

HORVITZ, D. G. and THOMPSON, D. J. (1952) : A generalisation of sampling without replacement from a finite universe. *Jour. Amer. Stat. Assoc.*, **47**, 663-685.

MIDZUNO, H. (1952) : On the sampling system with probability proportional to the sum of sizes. *Ann. Inst. Stat. Math.*, **3**, 99-108.

PATHAK, P. K. (1964) : Sufficiency in sampling theory. *Ann. Math. Stat.*, **35**, 705-808.

ROY, J. and CHAKRAVORTY, I. M. (1960) : Estimating the mean of a finite population. *Ann. Math. Stat.*, **31**, 392-298.

SEN, A. R. (1952) : Present status of probability sampling and its use in estimation of characteristics (abstract). *Econometrica*, **20**, 103.

TAKEUCHI, K. (1961) : Some remarks about unbiased estimation in sampling from finite populations. *Jour. Union of Scientists and Engineers* (Japan), **8**, 35-38.

VIJAYAN, K. (1966) : On an exact π PS sampling scheme. *Indian Statistical Institute Tech. Report* 6/66.

Paper received : February, 1966.