The Annals of Mathematical Statistics
1968, Vol. 39, No. 2, 621–642

HYPER-ADMISSIBILITY AND OPTIMUM ESTIMATORS FOR SAMPLING FINITE POPULATIONS[1]

By T. V. Hanurav[2]

University of Sheffield and Michigan State University

1. Introduction. The technique of varying probability sampling was partially generalized by Horvitz and Thompson (1952), who furnished an unbiased estimator of the population total. Their estimator, which we shall call the H.T. estimator for brevity, is in fact applicable in a very general set-up and turned out to be an admissible estimator. Godambe (1955) generalized the concepts of sampling design and linear estimators. The important negative result that emerged from his investigation is the nonexistence of a uniformly minimum variance (umv, for brevity) estimator among homogeneous linear unbiased estimators of the population total. While Godambe supposed this to be true for all sampling designs, there are nontrivial exceptions to this result which are completely characterized by the author (1966a) as 'uni-cluster designs.'[3] These, however, have the serious drawback that unbiased variance estimators do not exist for them. Barring therefore these unicluster designs, we have only some negative results, of which Basu's (1958) result concerning the inadmissibility of estimators that depend on the order or repetitions of units in a sample, is the important one. Koop[4] (1957) and Prabhu Ajgaoukar (1962) proved the nonexistence of a umvue even in certain subclasses of linear unbiased estimators. Roy and Chakravorty (1962) proposed the additional criterion of linear invariance but even this did not give a umvue. Godambe (1955) and Hájek (1959) followed the Bayesian approach to the problem and obtained "best strategies" for some important practical situations when a particular type of auxiliary information is available. In all other situations the problem of an 'optimum' estimator still remains unsolved.

In Section 3 of this paper we propose a new criterion which we name *hyper-admissibility* (*h*-admissibility for short). This criterion gave a unique optimum estimator which is the H.T. estimator, in a very wide class of unbiased estimators and for all non-unicluster designs. Even for unicluster designs the H.T. estimator

Received 23 May 1966; revised 14 September 1967.

[1] Research supported by the U.S. Public Health Service (Research Grant NIH-GM 13138-01 from the National Institute of General Medical Sciences. Final stage supported by NIH-GM 14213).

[2] This paper is based on the Ph.D. thesis of the author submitted to the Indian Statistical Institute in June 1965.

[3] In his later paper (1965) Godambe noted this exception to his earlier result.

[4] Koop introduced seven classes of homogeneous linear estimators, identified Godambe's class as his T_5-class and claims that his T_7-class is wider than Godambe's class. However, recalling that Godambe defines his sample as an *ordered* sequence, it can be seen that the two classes are, in fact, identical.

forms the effective component of any h-admissible estimator. In Section 7 we discuss the consequent final reduction of the central problem of estimation.

After these results were obtained by the author [cf. Hanurav (1965)] Godambe and Joshi (1965) and Joshi [(1965a) and (1965b)] published some new results and these are also discussed in the sequel.

2. Definitions and preliminaries. For a full exposition of the basic concepts involved, we refer to Godambe (1955) and Hanurav (1966a). Here we shall give only a brief outline.

A *simple finite population* ('finite population' for brevity) $\mathfrak{U}$ consists of a known number N of distinguishable units $U_i(1 \leqq i \leqq N)$. A *sample* 's' from $\mathfrak{U}$ is a finite ordered sequence of units, not necessarily distinct, from $\mathfrak{U}$. S, the collection of all possible samples, is the *basic sample space*. A probability measure P on S defines a general *sampling design* ('design' for brevity) $D(\mathfrak{U}, S, P)$. Where only a single population is under consideration as in this paper, we often refer to $D(\mathfrak{U}, S, P)$ as $D(S, P)$ or simply as the 'design P.' Any design P is equivalent to a *unit drawing mechanism* biuniquely [Hanurav (1962) and (1966a)].

Let $\mathcal{Y}$ be a real-valued variable defined over $\mathfrak{U}$, taking the value Y_i on $U_i(1 \leqq i \leqq N)$. The vector $\mathbf{Y} = (Y_1, \cdots, Y_N)$ is treated as a parameter in R^N. Single-valued functions $f(\mathbf{Y})$ of $\mathbf{Y}$ are parametric functions of which the most important is the population total

$$Y = \textstyle\sum_{i=1} Y_i. \tag{2.1}$$

A *statistic* T defined over S is a single-valued function of the $\mathcal{Y}$-values of only those units that appear in 's.' Tautologically, T is an *estimator* of $f(\mathbf{Y})$ if it is used to estimate $f(\mathbf{Y})$. The pair (P, T) is called a *sampling strategy* (or simply 'strategy') and is denoted by $H(P, T)$. The problem is to choose an 'optimum' strategy, optimality being defined in a reasonable way. The concepts of cost function $C(H)$ and loss function $\delta(H)$ now enter into the picture. For a meaningful interpretation, we should choose from the class of all strategies with $C(H) \leqq C_0$, a given budget, those that minimize $\delta(H)$. The inequality $C(H) \leqq C_0$ can be replaced without loss of generality, by the equation $C(H) = C_0$ [Hanurav (1966a)]. Throughout this paper we shall take the mean-square error (mse) to be our loss function.

For a sample s, ν_s, the *effective size* of s, denotes the number of 'distinct' units in s. For a given design P

$$\nu(P) = \textstyle\sum_{s\varepsilon S} \nu_s P_s \tag{2.2}$$

is the *expected effective sample size* in P. In a number of situations in practice $C(H)$ depends on H only through P and can be approximated by $\nu(P)$ by a proper choice of origin and scale of measurement of the cost; and in some situations where this is not so, as for example in stratified or multistage sampling, it can be reduced to this form by a proper splitting up of the problem. We can thus investigate the problem of choosing optimum strategy $H_0(P_0, T_0)$ from the class $\mathcal{H}(\nu)$ of strategies $H(P, T)$ satisfying $\nu(P) = \nu$, a given number. The problem can be broken up into two steps: first we shall find optimum T's for a given P and then choose the optimum of these optimum estimators for P varying over the above class. We shall consider in detail only the first of these in this paper and comment briefly on the second step towards the end. We shall also confine our attention to the particular parametric function Y defined by (2.1). The problem of estimation of any polynomial parametric function can be reduced to this case by a suitable definition of a new variable or population or both [Hanurav (1966a)].

A general linear estimator (gle, for brevity) is of the form

$$T:\{T_s = K_s + \sum_{i\varepsilon s} \beta_{si} Y_i\},$$

where K_s and β_{si} do not depend on $\mathbf{Y}$ and the sum in the braces is over all 'distinct' units U_i belonging to the sample. If $K_s \equiv 0$ in the above, we have a ghle ('h' for homogeneous). If further the estimator is unbiased for Y we have a glue or a ghlue (u for unbiasedness) as the case may be. Higher order polynomial estimators are defined similarly. A general polynomial estimator of nth degree (gne) is of the form

$$T = T^{(0)} + T^{(1)} + \cdots + T^{(n)},$$

where $T_s^{(r)}$ is a homogeneous polynomial of degree r (which may vanish identically for some or all s) in its arguments which are the $\mathcal{Y}$-values of only those units that occur in s; and $T_s^{(n)}$ has degree n for at least one sample s for which $P_s > 0$.

A gne which is unbiased for Y is termed as a gnue.

Let

$\mathfrak{I}(P)$ be the class of all estimators of Y,
$\mathfrak{L}(P)$ be the class of all gle's of Y,
$\mathfrak{L}_0(P)$ be the class of all ghle's of Y,
$\mathfrak{M}_n(P)$ be the class of all gne's of Y,

and $\mathfrak{M}(P) = \bigcup_{n=1}^{\infty} \mathfrak{M}_n(P)$, the class of all polynomial estimators. Further let $\mathfrak{I}^*(P)$, $\mathfrak{L}^*(P)$, $\mathfrak{L}_0^*(P)$, $\mathfrak{M}_n^*(P)$ and $\mathfrak{M}^*(P)$ denote the corresponding classes of unbiased estimators.

In the literature on this subject attention is mainly focussed on $\mathfrak{L}_0^*(P)$, and variance is taken to be the loss function. The condition of unbiasedness is relaxed in some cases (like ratio-estimators and regression estimators) for obtaining simpler estimators but only when we are satisfied that the biases of the estimators are negligible. While the criterion of unbiasedness has a pertinent interpretation in theory as well as in practice, the same cannot be said of the condition of linearity imposed on the estimator, at any rate not so from the purely theoretical point of view. We shall comment further on this aspect later in Section 5 and refer the reader for more discussion, to Godambe and Joshi (1965) and Hanurav (1966a).

Even restricting to $\mathfrak{L}_0^*(P)$ does not settle our problem. Godambe (1955) proved that there does not exist a 'best' (i.e. uniformly minimum variance estimator—uniformly for all $\mathbf{Y}\ \varepsilon\ R^N$) in $\mathfrak{L}_0^*(P)$ for any P. There however exist designs P, termed *unicluster* designs, for which this is not true [Hanurav (1966a)]. But these designs are undesirable from another important consideration viz. the estimability of the variance of any unbiased estimator of Y.

Two courses are open to progress towards the choice of an optimum estimator. We can either restrict ourselves to some subclasses of $\mathcal{L}_0^*(P)$ by means of additional (yet meaningful) criteria, or weaken our criterion of uniform minimization of the variance (umv) as the criterion of optimality. Along the former course Koop (1957) and Prabhu Ajgaonkar (1962) restricted themselves to some subclasses of $\mathcal{L}_0^*(P)$ for the sake of simplicity; Roy and Chakravorty (1960) considered the subclass of linearly invarient estimators. These methods did not yield a umv estimator in these subclasses. Roy and Chakravorty restricted themselves further to 'regular estimators' and obtained a best in the class of all 'balanced designs'. However their restrictions are unnatural and severe. Along the other approach Godambe (1955) and Hájek (1959) took up the Bayesian approach and proved the existence of optimum estimators in $\mathcal{L}_0^*(P)$ (and in fact the existence of optimum designs P too) for a very special, though important, case of the availability of a suitable auxiliary information. For further problems that arise in this connection cf. Hanurav (1966b). Aggarwal (1959) introduced minimax criterion and obtained a class of optimum estimators in the class $\mathfrak{I}(P)$ but this criterion is not a meaningful one in our context. [cf. Godambe and Joshi (1965)].

Along the second course we consider the criterion of admissibility which is weaker than the umv criterion. In a class $\mathcal{C}(P)$ of estimators, a member T_1 of $\mathcal{C}(P)$ is *admissible*, with respect to a given loss function δ, if there does not exist a $T_2 \in \mathcal{C}(P)$ for which

$$\delta(T_2) \leqq \delta(T_1), \qquad \text{a.s.} \quad (P)$$

(i.e. except for a set of samples with P-measure equal to zero) and for all $\mathbf{Y} \in R^N$, inequality holding for at least one $\mathbf{Y}$. Obviously admissibility is the minimum to be demanded from an estimator.

A subclass $\mathcal{C}_1(P) \subseteq \mathcal{C}(P)$ is said to be *complete* in $\mathcal{C}(P)$, with respect to the loss function δ, iff every member of $\mathcal{C}(P) - \mathcal{C}_1(P)$ is inadmissible. A subclass $\mathcal{C}_0(P)$ is said to be the *minimal complete class* in $\mathcal{C}(P)$ iff every member of $\mathcal{C}_0(P)$ is admissible in $\mathcal{C}(P)$. Our aim is to obtain a small enough minimal complete class and ideally to get one consisting of a single member.

The criterion of admissibility succeeded in eliminating a number of members of $\mathcal{L}_0^*(P)$—and in fact in $\mathcal{L}^*(P)$—but the complete class of estimators left, even in $\mathcal{L}_0^*(P)$, is too wide to satisfy us. Two samples s_1 and s_2 are *effectively equivalent*—in symbols $s_1 \sim s_2$—iff every unit belonging to s_1 belongs to s_2 and conversely. Basu (1958) introduced the fruitful concept of sufficiency in this theory, and proved.

THEOREM (2.1). (Basu) *Given a design P in which Y is estimable, if T is an unbiased estimator of Y, then the estimator T^* defined by*

$$\begin{aligned} T_s^* &= \textstyle\sum_{s_0 \sim s} T_{s_0} P_{s_0} / \sum_{s_0 \sim s} P_{s_0}, \quad \textit{if} \quad \sum_{s_0 \sim s} P_s > 0 \\ &= 0, \quad \textit{otherwise}, \end{aligned}$$

(*where the sums occurring are over all samples* s_0 *effectively equivalent to* s) *is also unbiased for* Y *and for any convex loss function* δ

$$\delta(T^*) \leqq \delta(T), \quad \textit{for all} \quad \mathbf{Y} \varepsilon R^N,$$

with the strict inequality holding for at least one value of $\mathbf{Y}$ iff

$$P\{s_1 \sim s_2, T_{s_1} \not\equiv T_{s_2}\} > 0.$$

In fact the above theorem remains true if Y is replaced by any estimable parametric function. It follows that T is an inadmissible estimator in any class of unbiased estimators containing the corresponding T^* also, unless

$$s_1 \sim s_2 \Rightarrow T_{s_1} \equiv T_{s_2}, \quad \text{a.s.} \quad (P). \tag{2.3}$$

Though Basu restricted himself to simple classes of designs like those arising out of simple random sampling and the customary probability proportional to size sampling, his result remains true with no modifications in the proof for any general sampling design. This is evident from the works of Roy and Chakravorty (1960) and of Takeuchi (1961). It may be mentioned that Hájek (1959), independently, makes a passing remark conjecturing Theorem (2.1). Pathak had systematically applied Basu's theorem to a number of examples. However, his (1964) 'expository' account of this main idea of 'considerable potential value' (wrongly accredited to him by some workers—for example cf. Chernoff (1965) and other reviews) where he attempted to formalise the result did not throw any further light on the problem as he *postulates* the sufficiency of the relevant statistic—a result *proved* by Basu (cf. Hájek (1965) and Hanurav (1966a)).

Before we pass on to our main result in Section 3 we give some results that are needed in the sequel. For proofs and discussion of these we refer to Hanurav (1966a).

Given a design P, the first and second order inclusion probabilities π_i's and π_{ij}'s are defined by

$$\pi_i = \pi_i(P) = \textstyle\sum_{s \supset i} P_s,$$
$$\pi_{ij} = \pi_{ij}(P) = \textstyle\sum_{s \supset i,j} P_s,$$

for $1 \leqq i \neq j \leqq N$. (In the above, the first sum is over all samples that contain U_i and the second is over all samples that contain U_i and U_j.) A set of necessary and sufficient conditions for the estimability of Y (i.e. for the nonemptiness of $\mathfrak{J}^*(P)$) is that (Godambe, 1955)

$$\pi_i > 0, \quad \text{for} \quad 1 \leqq i \leqq N. \tag{2.4}$$

When (2.4) holds good and T^* is an unbiased estimator of Y, a set of necessary and sufficient conditions for the estimability of $V(T^*)$ is that [Hanurav (1966a)]

$$\pi_{ij} > 0, \quad \text{for} \quad 1 \leqq i \neq j \leqq N. \tag{2.5}$$

When (2.4) holds good, the Horvitz and Thompson (1952) estimator of Y is

$$\hat{Y}_{\mathrm{HT}} = \textstyle\sum_{i \varepsilon s} Y_i/\pi_i, \tag{2.6}$$

where the sum is over all *distinct* units U_i belonging to s.

Definition. A design P is a *unicluster design* iff any two samples with positive probabilities are either disjoint (i.e. have no common unit) or are effectively equivalent i.e.

$$s_1 \cap s_2 = \emptyset, \quad \text{or} \quad s_1 \sim s_2, \quad \text{a.s.} \quad (P). \tag{2.7}$$

The above terminology is derived from the fact that these designs are effectively equivalent to designs obtained by dividing the population into clusters of units and then choosing just one cluster from among them.

In Section 3 we introduce a new criterion of optimality for estimators and prove the existence of a unique optimum estimator of Y in $\mathfrak{M}^*(P)$, for any non-unicluster design. For the unicluster designs we give a complete characterization of all optimum estimators.

3. The criterion of hyper-admissibility and the main theorem. We recall the definition of admissibility. In a class $\mathcal{C}(P)$ of unbiased estimators of Y, $T_1 \varepsilon \mathcal{C}_1(P)$ is admissible iff

$$T_2 \varepsilon \mathcal{C}(P) \quad \text{implies that there exists} \quad \mathbf{Y}^{(0)} = \mathbf{Y}^{(0)}(T_1, T_2) \varepsilon R^N \quad \text{such that} \quad V(T_1)|_{\mathbf{Y}^{(0)}} < V(T_2)|_{\mathbf{Y}^{(0)}}, \tag{3.1}$$

where the variances are evaluated at the point $\mathbf{Y}^{(0)}$, which, possibly, depends on T_1 and T_2. It is clear that if we restrict the parameter space to some given subsets, an estimator T_1 which is admissible for the whole parameter-space may cease to be admissible in the restricted parameter space.

Let R^{N-} be any principle hypersurface (phs) in R^N. There are $\binom{N}{r}$ such phs's of dimension r, for $1 \leqq r \leqq N$, and in all there are $(2^N - 1)$ phs's the totality of which we denote by $\mathcal{R}^{N-}$.

Definition (3.1). In a class $\mathcal{C}(P)$ of unbiased estimators of Y, $T_1 \varepsilon \mathcal{C}(P)$ is *hyper-admissible* (h-admissible, for brevity) iff it is admissible when the parameter $\mathbf{Y}$ is restricted to the 'interior' of any phs R^{N-} of $\mathcal{R}^{N-}$. (By the interior of phs R^{N-} we shall mean all those points of the phs that do not lie in any phs of smaller dimension that lies entirely in R^{N-}.)

Thus $T_1 \varepsilon \mathcal{C}(P)$ is h-admissible iff

$$T_2 \varepsilon \mathcal{C}(P),\ R^{N-} \varepsilon \mathcal{R}^{N-} \quad \text{implies that there exists} \quad \mathbf{Y}' = \mathbf{Y}'(T_1, T_2) \varepsilon \underline{R}^{N-} \quad \text{such that} \quad V(T_1)|_{\mathbf{Y}'} < V(T_2)|_{\mathbf{Y}'}, \tag{3.2}$$

where $\underline{R}^{N-}$ denotes the interior of R^{N-}.

Evidently h-admissibility is a stronger requirement than admissibility but is weaker than the umv criterion.

Theorem (3.1). *For any nonunicluster design P that satisfies* (2.4), (*so that Y is estimable*) *the class $\mathfrak{M}^*(P)$ contains a unique hyperadmissible estimator which is given by $\hat{Y}_{\text{HT}}$ of* (2.6).

Proof. Let m^* be a member of $\mathfrak{M}^*(P)$, which is h-admissible, in $\mathfrak{M}^*(P)$. For each s, m_s^* is a polynomial in its arguments (which are the $\mathcal{Y}$-values of the units occurring in s), of degree r_s, say.

Since m^* is h-admissible it is, in particular, admissible. From Theorem (2.1) it then follows that

$$s_1 \sim s_2 \Rightarrow m^*_{s_1} \equiv m^*_{s_2}, \quad \text{a.s.} \quad (P)$$

and hence in particular

$$r_{s_1} = r_{s_2}, \quad \text{if} \quad s_1 \sim s_2, \quad \text{a.s.} \quad (P). \tag{3.3}$$

Even though S may contain an infinite number of samples with positive probabilities, since there are only a finite number $(2^N - 1)$ of equivalence classes of samples, from (3.3), r_s is essentially bounded for s varying over S. Let then

$$\sup_{s \varepsilon S} r_s = r, \quad \text{a.s.} \quad (P),$$

and

$$m^* = T^{(0)} + T^{(1)} + \cdots + T^{(r)}, \quad \text{a.s.} \quad (P), \tag{3.4}$$

where $T_s^{(q)}$ is a homogeneous polynomial of degree q in its arguments (which are the $\mathcal{Y}$-values of the units belonging to s). In (3.4) $T^{(q)}$ may be identically zero for some or all samples for $1 \leqq q \leqq r - 1$, but, for at least one sample s with positive probability we shall have $T_s^{(r)} \not\equiv 0$.

Let

$$\begin{cases} T_s^{(0)} = K_s, \\ T_s^{(1)} = \sum_{i \varepsilon s} \beta_{si} Y_i, \\ T_s^{(2)} = \sum_{i \varepsilon s} \beta_{sii} Y_i^2 + \sum \sum_{i \neq j \varepsilon s} \beta_{sij} Y_i Y_j, \\ T_s^{(3)} = \sum_{i \varepsilon s} \beta_{siii} Y_i^3 + \sum \sum_{i \neq j \varepsilon s} \beta_{siij} Y_i^2 Y_j + \sum \sum \sum_{i \neq j \neq k \varepsilon s} \beta_{sijk} Y_i Y_j Y_k \end{cases} \tag{3.5}$$

etc., where the coefficients β's do not depend on $\mathbf{Y}$. (3.5) gives the most general form of m^* (which is admissible in $\mathfrak{M}^*(P)$). Since

$$E(m^*) \equiv Y, \quad \text{for all} \quad \mathbf{Y} \varepsilon R^N,$$

it follows that

$$E(T_1) \equiv Y \quad \text{and} \quad E(T^{(q)}) \equiv 0, \quad \text{for} \quad q \neq 1. \tag{3.6}$$

Since m^* is hyperadmissible it is, in particular, admissible on each of the coordinate axes of R^N. For any given i $(1 \leqq i \leqq N)$, considering the ith coordinate axis on which $Y_j = 0$, for $j \neq i$, m_s^* reduces to

$$\begin{aligned} m_s^*(i) &= K_s + \beta_{si} Y_i + \beta_{sii} Y_i^2 + \cdots + \beta_{si \cdots i} Y_i^r, \quad &\text{if} \quad U_i \varepsilon s \\ &= K_s, &\text{if} \quad U_i \not\varepsilon s. \end{aligned} \tag{3.7}$$

If

$$S = S_i \cup S_i^*$$

where

$$S_i = \{s : U_i \varepsilon s\},$$

it follows that for m^* to be admissible on the ith axis, it is necessary that a.s. $[P]$ we have, identically for all values of Y_i,

$$m^*_{s_1}(i) \equiv m^*_{s_2}, \quad \text{if} \quad s_1, s_2 \varepsilon S_i$$

and

$$m^*_{s_3}(i) \equiv m^*_{s_4}, \quad \text{if} \quad s_3, s_4 \varepsilon S_i^*. \tag{3.8}$$

To see this we generate new samples s' and a new design $D'(\mathcal{U}', S', P')$ from the original design as follows: $\mathcal{U}'$ is the population consisting of the original unit U_i and a null-unit U_{N+1} for which Y_{N+1} is known to be equal to zero; from any sample $s \varepsilon S$ we generate a new sample s' by replacing every unit occurring in s, other than the U_i, by U_{N+1} and retaining U_i as such if it occurs in s; for the new probability measure P' we set $P'_{s'} = P_s$ where s is the original sample from which s' is generated. We note that the behaviour of any estimator m^* on the ith coordinate axis is that of the estimator $m^{**}_{s'}$ defined by

$$m^{**}_{s'} = m_s^*(i),$$

over $D'(\mathcal{U}', S', P')$. The desired result now follows from Theorem (2.1) by noting that the classes S_i' and $(S_i')^*$ now represent classes of effectively equivalent samples of $\mathcal{U}'$.

From (3.7) and (3.8) we have

$$\begin{aligned} m_s^*(i) &= K_1(i) + \beta_i Y_i + \beta_{ii} Y_i^2 + \cdots + \beta_{i\dots i} Y_i^r, \quad \text{if} \quad s \varepsilon S_i \\ &= K_2(i), \qquad\qquad\qquad\qquad\qquad\qquad \text{if} \quad s \varepsilon S_i^*, \end{aligned} \tag{3.9}$$

where $K_1(i)$, $K_2(i)$ and the β's do not depend on s (or any $\mathbf{Y}$). From (3.6) we have

$$E(m_s^*(i)) = Y_i$$

and hence from (3.9)

$$K_1(i)\pi_i + K_2(i)(1 - \pi_i) = 0, \qquad \beta_i \pi_i = 1$$

and $\beta_{ii}\pi_i = \beta_{iii}\pi_i = \cdots = \beta_{i\dots i}\pi_i = 0$. Since Y is estimable over $D(\mathcal{U}, S, P)$, we have $\pi_i > 0$, so that a.s. (P),

$$\begin{aligned} \beta_{si} &= \beta_i = \pi_i^{-1} \quad \text{and} \\ \beta_{sii} &= \beta_{siii} = \cdots = \beta_{si\dots i} = 0, \end{aligned} \tag{3.10}$$

for all samples $s \varepsilon S_i$. (For all other samples these β's vanish by the definition of a statistic.)

Since i is arbitrary, (3.10) is in fact true for $1 \leqq i \leqq N$.

To show that

$$K_1(i) = K_2(i) = 0, \quad \text{for} \quad 1 \leqq i \leqq N$$

we invoke our hypothesis that P is a nonunicluster design. It is easy to see that this implies that there exist at least one pair (i', j') such that

$$0 < \pi_{i'j'} < \pi_{i'} . \tag{3.11}$$

From (3.9) we see that for all samples s with $P_s > 0$,

$$\begin{aligned} K_s &= K_1(i), \quad \text{if} \quad s \,\varepsilon\, S_i \\ &= K_2(i), \quad \text{if} \quad s \,\varepsilon\, S_i^* \end{aligned}$$

and similarly

$$\begin{aligned} K_s &= K_1(j), \quad \text{if} \quad s \,\varepsilon\, S_j \\ &= K_2(j), \quad \text{if} \quad s \,\varepsilon\, S_j^*. \end{aligned}$$

The left hand side of (3.11) implies that there is a sample $s \,\varepsilon\, S_iS_j$, for which $P_s > 0$, which gives, from the above,

$$K_1(i) = K_1(j).$$

However, the right hand side of (3.11) implies that there is at least one sample $s \,\varepsilon\, S_iS_j^*$, for which $P_s > 0$, and this gives

$$K_1(i) = K_2(j),$$

and hence finally we see that there exists a j for which

$$K_1(j) = K_2(j).$$

This implies that

$$K_s = K, \quad \text{say, a.s.} \quad (P),$$

and from (3.6) now follows that

$$K_s \equiv 0, \quad \text{a.s.} \quad (P). \tag{3.12}$$

We now consider the admissibility of m^* in the interior of the (i, j) plane, for a pair of numbers (i, j). Setting $Y_t = 0$ for $t \neq i, j$, and from (3.10) and (3.12), we see that in this region m^* reduces to

$$\begin{aligned} m_s^*(i, j) &= Y_i/\pi_i + Y_j/\pi_j + \beta_{sij} Y_i Y_j + (\beta_{siij} Y_i^2 Y_j \\ &\quad + \beta_{sjji} Y_j^2 Y_i) + \cdots , && \text{for} \quad s \,\varepsilon\, S_iS_j \\ &= Y_i/\pi_i , && \text{for} \quad s \,\varepsilon\, S_iS_j^* \\ &= Y_j/\pi_j , && \text{for} \quad s \,\varepsilon\, S_i^*S_j \\ &= 0, && \text{for} \quad s \,\varepsilon\, S_i^*S_j^*. \end{aligned} \tag{3.13}$$

Arguing as before we see that the classes S_iS_j, $S_iS_j^*$, $S_i^*S_j$ and $S_i^*S_j^*$ of samples, form a partition of S into classes of effectively equivalent samples and that the admissibility of m^* in the interior of the (i, j) plane demands that $m^*(i, j)$ be constant over these equivalence classes, except possibly for samples of total

P-measure zero. This condition is met with for the later three equivalence classes. If $\pi_{ij} = 0$ so that $P(S_iS_j) = 0$ then the condition is trivially satisfied over the first equivalence class also. If $\pi_{ij} > 0$, then the requirement is met with iff the β coefficients attached to the terms of the form $Y_iY_j^t$ and $Y_i^tY_j$, for $i \leqq t \leqq r - 1$, that occur in the first member on the right of (3.13), remain independent of s for $s \varepsilon S_iS_j$, a.s. $[P]$. This, together with (3.6) and the assumption that $\pi_{ij} > 0$, yields, as in (3.10), that all these β coefficients vanish identically—and this is true for any pair (i, j). Thus

$$(3.14) \qquad \beta_{sij} \equiv \beta_{siij} \equiv \beta_{siiij} \equiv \cdots \equiv \beta_{si\cdots ij} \equiv 0, \quad \text{a.s.} \quad (P),$$

for $1 \leqq i \neq j \leqq N$.

Arguing similarly for the interiors of the higher order hyperplanes it can be shown that all the β coefficients occurring in $T^{(2)}, T^{(3)}, \cdots, T^{(r)}$ of m^* vanish identically and we have

$$m^*: \{m_s^* = \textstyle\sum_{i\varepsilon s} Y_i/\pi_i\}$$

as the only possible h-admissible estimator in $\mathfrak{M}^*(P)$. This is the Horvitz and Thompson (H.T.) estimator $\hat{Y}_{\mathrm{HT}}$, given by (2.6).

That in fact (2.6) is h-admissible in $\mathfrak{M}^*(P)$ can be seen thus:

The derived estimators $m^*(i)$, $m^*(i, j)$ etc. defined above are each the corresponding H.T. estimators of the totals of the corresponding subpopulations, and defined over the respective reduced sampling designs and with the corresponding shrunken parameter spaces. It was shown earlier by Godambe (1960) and Roy and Chakravorty (1960) that for any population $\mathfrak{U}'$ of N' units and any design $D'(\mathfrak{U}', S', P')$ in which the total of $\mathfrak{U}'$ is estimable, the corresponding H.T. estimator $\hat{Y}'_{\mathrm{HT}}$ of the population total Y' of $\mathfrak{U}'$ is admissible in $\mathfrak{L}_0^*(P')$. Their proofs can be easily extended to prove the admissibility of $\hat{Y}'_{\mathrm{HT}}$ in $\mathfrak{M}_0^*(P')$. (In fact recently Godambe and Joshi (1965) extended this to $\mathfrak{I}^*(P')$ itself.) Though these proofs showed the admissibility of $\hat{Y}'_{\mathrm{HT}}$ in the whole parameter space $R^{N'}$—and in fact by exhibiting, corresponding to any other estimator $\hat{Y}'$, a point $\mathbf{Y}'$ on one of the coordinate axes, at which $V(\hat{Y}'_{\mathrm{HT}}) < V(\hat{Y}')$—they can be trivially modified to prove the admissibility of $\hat{Y}'_{\mathrm{HT}}$ in the interior of $R^{N'}$. The easiest way to see this is by noting that the function $V(\hat{Y}'_{\mathrm{HT}}) - V(\hat{Y}')$ is analytic in its N' arguments so that its negativeness at $\mathbf{Y}'$ implies the negativeness throughout a neighborhood of $\mathbf{Y}'$ which certainly contains points belonging to the interior of $R^{N'}$.

Thus the derived estimators $m_s^*(i)$, $m_s^*(i, j)$ of (2.6) not only satisfy (2.3) of Theorem (2.1), which is necessary for their admissibility, but are in fact admissible in their corresponding classes of $\mathfrak{M}^*$'s and in the interiors of the corresponding parameter spaces. Since this is all that is required for the h-admissibility of (2.6) we conclude that (2.6) is the unique h-admissible estimator in $\mathfrak{M}^*(P)$.

This completes the proof of our theorem.

REMARKS:

(1) Referring to the representation (3.4) of any admissible estimator m^* of $\mathfrak{M}^*(P)$, we see that the nonuniclusterness of P and the admissibility of m^* on the coordinate axes implies that $T_0 \equiv 0$; admissibility on the axes alone implied that $T_1 \equiv \hat{Y}_{HT}$; admissibility in the interiors of the axes and planes implied that $T_2 \equiv 0$ etc.

(2) Though we proved that the estimator (2.6) is admissible in the interior of every principal hyperplane of R^N, the proof can easily be modified to show that (2.6) is admissible in the interior of every signed quadrant of every principal hyperplane obtained by restricting the nonzero coordinates that occur, to any combination of positive and negative signs. In particular this implies the h-admissibility of (2.6) when $\mathbf{Y}$ is restricted to nonnegative values for all its N coordinates.

(3) It is easy to satisfy ourselves that (2.6) is admissible in the interior of every hyperplane. (i.e., in our terminology, in the subset of R^N obtained by fixing the values of some of the Y_i''s to be any quantities, not necessarily equal to zero) and that indeed (2.6) is the only member of $\mathfrak{M}^*(P)$ with this property.

4. Justification of the criterion of h-admissibility. Given a design P, when the conditions (2.4) hold good, so that $\mathfrak{M}^*(P)$ is non-empty, it can be seen that every linear parametric function (lpf, for brevity)

$$L(\mathbf{Y}) = \sum_{i=1}^{N} L_i Y_i \tag{4.1}$$

is estimable. In fact if T is any member of $\mathfrak{M}^*(P)$ then an unbiased estimator of $L(\mathbf{Y})$ is obtained by replacing the Y_i's occurring in T by the corresponding L_iY_i's. Any unbiased estimator T of Y, i.e., any given sequence of β coefficients, can therefore be looked upon not merely as an estimator of Y but as a method of estimation to unbiasedly estimate all lpf's.

In practical sample surveys often one is interested in estimating not only Y but also of a number of lpf's of the type (4.1). For example, totals and means of several subpopulations and contrasts between these totals and means are very often of interest in stratified and multistage sampling designs. Even when a prior stratification is not done a post-stratification of the population may arise and one may need a breakdown of the gross estimate for these strata. It is natural to make the minimum demand that any estimator considered should be admissible. If then one builds up these estimates from the original estimator T by the method described above then the demand is met with, if T is chosen to be a h-admissible estimator. As it is difficult to envisage beforehand all the lpf's that will be of interest to us, it is a safe precaution to have T as a h-admissible estimator in which case the derived estimator of *any* lpf will remain admissible. The crux of the matter is that in the estimation of lpf the form (4.1) in which some of the L_i are known in advance to be zero, the parameter space of relevance should no longer be the whole space R^N and admissibility in the whole of R^N is no consolation. On the other hand, since (4.1) remains unchanged for any values of those Y_i's for which the L_i's in (4.1) vanish, what is required is that for any fixed set of values of these Y_i's the estimator should remain admissible in the subset of R^N obtained by fixing these coordinates. From remark (3) of the previous section it is clear that only a h-admissible estimator will satisfy this condition and the choice of any other estimator for Y will necessitate a fresh construction of a new estimator if some lpf other than Y also becomes of interest at a later date, which indeed is very undesirable, especially from the point of view of large scale mechanized computations. These considerations lie at the bottom of the mathematical criterion given in Section 3.

There is also an interesting interplay between the equivalence of samples of a design $D(\mathcal{U}, S, P)$ and the problem of estimation of a given lpf, $\sum L_i Y_i$. Samples of D which are not effectively equivalent for the estimation of Y may become so for the estimation of $L(\mathbf{Y})$, if they contain some or all of only those units U_i, for which $L_i \neq 0$. Thus the effective equivalence of samples is not an intrinsic property of the design but is contingent upon the parametric function of interest to us, unless all lpf's are of possible interest to us.

5. The unicluster designs. As seen from Theorem 3.1, the unicluster designs P_{uc} form an exceptional class. A complete characterization of admissible and h-admissible estimators in $\mathfrak{M}^*(P_{uc})$ for these designs is given in the following:

THEOREM (5.1). *For any unicluster design* $P_{uc} = P$, say, *which satisfies* (2.4) *any estimator* $T \varepsilon \mathfrak{M}^*(P)$ *is admissible* iff T *is of the form*

$$T:\{T_s = K_s + \textstyle\sum_{i\varepsilon s} Y_i/\pi_i\} \tag{5.1}$$

where K_s's *are constants* (i.e., *independent of* $\mathbf{Y}$) *satisfying*

$$\text{(5.2)}\quad \begin{aligned} &\text{(i) } K_{s_1} = K_{s_2}, \quad \text{if} \quad s_1 \sim s_2 \text{; and} \\ &\text{(ii) } \textstyle\sum_{s\varepsilon S} K_s P_s = 0. \end{aligned}$$

Further, every T *satisfying* (5.1) *and* (5.2) *is h-admissible.*

PROOF. As proved in Theorem (3.1) any admissible estimator $T \varepsilon \mathfrak{M}^*(P)$ can be represented by (3.4) and (3.5) and the unbiasedness of T implies (3.6). We shall first prove that for unicluster designs these imply the stronger relations

$$T^{(q)} \equiv 0, \quad \text{for} \quad 1 < q \leqq r. \tag{5.3}$$

To prove that $T^{(2)} \equiv 0$, if possible let the contrary be true so that there exists a sample s, with $P_s > 0$ such that

$$\text{(5.4)}\quad \begin{aligned} &\text{(a) } \beta_{sii} \neq 0, \quad \text{for some} \quad i, \quad \text{for which} \quad U_i \varepsilon s \quad \text{or} \\ &\text{(b) } \beta_{sij} \neq 0, \quad \text{for some} \quad (i, j), \quad \text{for which} \quad U_i, U_j \varepsilon s. \end{aligned}$$

If, for example, (a) holds good then for any other sample s' with $P_{s'} > 0$ if we have $U_i \varepsilon s'$ also, then the uniclusterness of P implies, from (2.7) that $s' \sim s$. In this case the admissibility of T implies, from Theorem (2.1), that $T_s \equiv T_{s'}$ and in particular that $\beta_{sii} = \beta_{s'ii}$. Thus the coefficient of Y_i^2 in $E(T^{(2)})$ equals only $\beta_{sii}\,\pi_i$, which can vanish (as it should from (5.3)), from (2.4), iff $\beta_{sii} = 0$. If no such s' exists then evidently $\beta_{sii}\,\pi_i$ is the coefficient of Y_i^2 in $E(T^{(2)})$ hence also $\beta_{sii} = 0$. Thus in either case $\beta_{sii} = 0$. If (b) of (5.4) holds good then a similar reason holds for the coefficient of Y_iY_j in $E(T^{(2)})$ and we shall have $\beta_{sij} = 0$. Thus we have a contradiction and hence

$$T^{(2)} \equiv 0 \tag{5.5}$$

and the proof for $T^{(3)}, T^{(4)}, \cdots, T^{(r)}$ is exactly similar.

From the first equation of (3.6) we have

$$\textstyle\sum_{s \ni i} \beta_{si} P_s = 1, \quad \text{for } 1 \leqq i \leqq N. \tag{5.6}$$

For any fixed i since $\pi_i > 0$, there exists a s_0 such that $P_{s_0} > 0$ and $U_i \,\varepsilon\, s_0$. For any other sample s such that $P_s > 0$ and $U_i \,\varepsilon\, s$, the uniclusteredness of P implies that $s \sim s_0$ in which case the admissibility of T in $\mathfrak{M}^*(P)$ implies that $\beta_{si} = \beta_{s_0 i}$. From (5.6) we then have

$$\textstyle\sum_{s \ni i} \beta_{si} P_s = \beta_{s_0 i} \sum_{s \ni i} P_s = \beta_{s_0 i}\pi_i = 1$$

so that

$$\beta_{si} = \pi_i^{-1}, \quad \text{if } U_i \,\varepsilon\, s,$$

which proves that

$$T^{(1)} = \textstyle\sum_{i \varepsilon s} Y_i/\pi_i. \tag{5.7}$$

This shows that any admissible member of $\mathfrak{M}^*(P)$ is necessarily of the form given in Theorem (5.1). To prove that every such estimator is in fact admissible we need only compare estimators satisfying (5.1) and (5.2). Observe that since the design P is a unicluster one, $U_i, U_j \,\varepsilon\, s$ for some s (for which $P_s > 0$) implies that they always occur together in any sample, so that $\pi_i = \pi_j$. Hence (5.1) can in fact be written as

$$T{:}\{T_s = K_s + \textstyle\sum_{i\varepsilon s} Y_i/\pi(s) = K_s + Z_s/\pi(s), \text{ say}\}$$

where

$$\pi(s) = \textstyle\sum_{s' \sim s} P_{s'} \qquad \text{and} \qquad Z_s = \sum_{i \varepsilon s} Y_i. \tag{5.8}$$

If T_1 and T_2 be any two estimators of the form (5.8) and (5.2), which therefore differ only in their K_s's, we have

$$\begin{aligned} V(T_1) - V(T_2) &= E(T_1^2) - E(T_2^2) \\ &= \textstyle\sum_{s\varepsilon S} (K_{1,s}^2 - K_{2,s}^2)P_s \\ &\quad - \textstyle\sum_{s\varepsilon S} (K_{1,s} - K_{2,s})Z_s/\pi(s)\cdot P_s \\ &= \textstyle\sum'(K_{1,s}^2 - K_{2,s}^2)\pi(s) - \sum'(K_{1,s} - K_{2,s})Z_s, \end{aligned} \tag{5.9}$$

where the sums $\sum'$ are over one member each of the classes of effectively equivalent samples, since K_s's are and Z_s's have to remain constant over these classes.

It is now evident that for any given set of $K_{1,s}$'s and $K_{2,s}$'s which satisfy (5.2) and $K_{1,s} \neq K_{2,s}$ for at least some samples s (with $P_s > 0$), (5.9) can be made positive or negative by a proper choice of Z_s's. Since, however, Z_{s_1} and Z_{s_2} contain sums over disjoint sets of Y_i's for $s_1 \nsim s_2$, this implies that (5.9) can be made positive or negative by a proper choice of Y_i's and this proves that every estimator of the form (5.1) and (5.2) is in fact admissible in $\mathfrak{M}^*(P)$ and conversely.

From the proof of Theorem (3.1) and the remark (1) of Section 3 it is clear that every estimator of the form (5.1) and (5.2) is in fact h-admissible.

This completes the proof of our theorem.

REMARKS.

(1) For any unicluster design it is easy to see, from the definition, that there exists at least one pair (i, j) such that $\pi_{ij} = 0$ unless of course the design is a trivial one in which every sample (with positive probability) contains every unit of the population (in which case the design is effectively equivalent to complete census). From (2.5) it then follows that in any nontrivial unicluster design, $V(\hat{Y})$ is not estimable for any unbiased estimator $\hat{Y}$ of Y. This is a serious drawback in practice where one does not relish to be totally in the dark regarding the precision of the estimate $\hat{Y}$ of Y that he makes.

(2) It is interesting to note that while under the criterion of umv it is unfortunate that the undesirable unicluster designs are the only designs for which there exists an unambiguous 'best' in $\mathfrak{L}_0^*$ yet, under the criterion of h-admissibility, fortunately, these are precisely the designs barring which there exists an unambiguous best in even a wider class $\mathfrak{M}^*$.

6. Some general comments. Since the 'optimality' of (2.6) is established in the class $\mathfrak{M}^*(P)$ of polynomial unbiased estimators which is wider than $\mathfrak{L}_0^*(P)$ it is only proper to justify the widening of the area of search. We have accepted the mean square error as the loss function as is customary as also because the ultimate choice has to be arbitrary in some way if we are to have worthwhile results. The customary criterion of unbiasedness is also accepted for its intuitive content, statistical interpretability in theory as well as in practice and also for the sake of mathematical simplicity. Regarding the restriction of linearity the only justification is mathematical simplicity. One further justification often advanced by some workers is based on the considerations of dimensionality. If Y_i's are the measurements of some physical characteristic like income or area, the use of a nonlinear estimator of Y is, according to these authors, incompatible with the dimensionalities. This is purely a nonmathematical reason and once the units of measurement are chosen the Y_i's are to be regarded as pure numbers and the problem of estimation should be looked upon solely as a mathematical one. If, in fact, this argument is to be regarded as very sacred then it has to be carried to its further logical consequence that the estimator of Y should be a linearly invariant one; i.e., it should remain independent of the origin and choice of the unit of measurement. Though in some situations we have a natural choice of origin and unit of measurement, there are a number of situations in practice in which the unit of measurement is purely arbitrary and in some of these cases the origin of measurement is also arbitrary. For the theory to cover all these important cases, the argument based on dimensionality demands the choice of only linearly invariant estimators. However, barring some of the simplest estimators, none of the estimators used in practice are linearly invariant estimators. To this class belong almost all the estimators used in the various unequal probability sampling methods. Even for equal probability sampling (with or without replacement) the well-known 'difference-estimator'

$$\hat{Y}_{\text{dif}} = N\bar{y} + k(\bar{x} - \bar{X}),$$

(where $\mathfrak{X}$ is an auxiliary characteristic, $\bar{x}$ and $\bar{y}$ are the sample means, $\bar{X}$ is the known population mean of $\mathfrak{X}$ and k is any constant) does not satisfy even the compatibility of dimensions if $\mathfrak{X}$ and $\mathcal{Y}$ are two entirely different characteristics as is often the case in practice.

The above argument shows that restriction to the class $\mathcal{L}_0^*(P)$ of homogeneous linear unbiased estimators of Y stems purely from considerations of theoretical convenience, unless one proves that from the point of view of some loss function like the variance, the class $\mathcal{L}_0^*(P)$ is complete in a wider class like, say $\mathfrak{M}^*(P)$, that includes nonlinear unbiased estimators also. We are not aware of any such result. In fact Godambe and Joshi (1965) give an example of a design P (for $N = 3$) and a member of $\mathfrak{M}^*(P) = \mathcal{L}_0^*(P)$, for which there does not exist a member of $\mathcal{L}_0^*(P)$, which has uniformly smaller variance. To indicate more general results in this direction we shall digress for a discussion on the sufficiency of the "effective sample". For further discussion of this aspect we refer to Hanurav (1966a), where a detailed bibliographical discussion is also given.

Given a design P and a sample

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

on each of whose units $\mathcal{Y}$ is observed, we have the '*sampley*'

$$(6.1) \qquad (s, \mathbf{Y}) = \{(U_{i_1}, Y_{i_1}), (U_{i_2}, Y_{i_2}), \cdots, (U_{i_{n(s)}}, Y_{i_{n(s)}})\}$$

which denotes the totality of the observations made. If the set of 'distinct' units belonging to s and their corresponding $\mathcal{Y}$-values are given in any order whatsoever, say for the sake of definiteness, in the increasing order of the indices of the units that occur, we have the '*effective sampley*' denoted by

$$(6.2) \qquad (s, \mathbf{Y})_d = \{(U_{j_1}, Y_{j_1}), (U_{j_2}, Y_{j_2}), \cdots, (U_{j_{\nu_s}}, Y_{j_{\nu_s}})\}.$$

The basic sample space S gives rise to the '*basic sampley space*'

$$(6.3) \qquad (\mathcal{S}, \mathcal{Y}) = \{(s, \mathbf{Y}) : s \,\varepsilon\, S, \mathbf{Y} \,\varepsilon\, R^N\}.$$

The given design P on S now generates a N-parameter family $\mathcal{P}_{\mathbf{Y}}$ of probability distributions on $(S, \mathcal{Y})$ defined by

$$(6.4) \qquad \mathcal{P}_{\mathbf{Y}^{(0)}}(s, \mathbf{Y}) = P_s, \quad \text{if } Y_{i_k} = Y_{i_k}^{(0)}, \quad \text{for } 1 \leqq k \leqq n(s)$$
$$= 0, \quad \text{otherwise,}$$

for any given point $\mathbf{Y}^{(0)}$ of R^N. With this terminology it is seen [Takeuchi (1961)] that the effective sampley is a sufficient statistic for the family $\mathcal{P}_{\mathbf{Y}}$. It is, however, not a complete sufficient statistic. To see this let $\mathcal{U} = (U_1, U_2, U_3)$ and let

$$s_1 = (U_1, U_2); \quad s_2 = (U_2, U_3); \quad s_3 = (U_3, U_1);$$
$$s_4 = (U_2, U_1); \quad s_5 = (U_3, U_2) \quad \text{and} \quad s_6 = (U_1, U_3).$$

Let P be defined as $P_{s_k} = \frac{1}{6}$, for $1 \leqq k \leqq 6$, and $P_s = 0$, for all other samples $s \varepsilon S$. Consider the statistic T defined by

$$T_{s_1} = T_{s_4} = Y_1^2 - Y_2^2,$$
$$T_{s_2} = T_{s_5} = Y_2^2 - Y_3^2,$$
$$T_{s_3} = T_{s_6} = Y_3^2 - Y_1^2,$$

and $T_s = 0$, for all other $s \varepsilon S$. Clearly T is a function of $(s, \mathbf{Y})_d$ alone, $T \not\equiv 0$ and $E(T) \equiv 0$, for all $\mathbf{Y} \varepsilon R^N$, which shows that $(s, \mathbf{Y})_d$ is not complete and the existence of a umv unbiased estimator is not guaranteed.

However, what is more interesting to note is that $(s, \mathbf{Y})_d$ is not just a sufficient statistics for $\mathcal{P}_{\mathbf{Y}}$ but is in fact the minimal sufficient statistics so that a statistic which is unbiased for Y and which depends on $(s, \mathbf{Y})$ only through $(s, \mathbf{Y})_d$ fully, cannot be improved upon by Rao-Blackwellisation. It is now easy to construct members of $\mathfrak{M}^*(P) - \mathfrak{L}_0^*(P)$ that are thus not obviously inadmissible. If T_1 is any member of $\mathfrak{L}_0^*(P)$, which is a function of $(s, \mathbf{Y})_d$ alone then, in the above example, the estimator $T_1^* = T_1 + T$ is one such. In general, given any non-unicluster design P we have at least two samples s_1 and s_2 with $P_{s_1} > 0$, $P_{s_2} > 0$ which are neither disjoint, so that for some i, $U_i \varepsilon s_1$, $U_i \varepsilon s_2$, nor are effectively equivalent. If T is any member of $\mathfrak{L}_0^*(P)$, which is a function of $(s, \mathbf{Y})_d$, then T^* defined by

$$T_{s_3}^* = T_{s_1}^* = T_{s_1} + Y_1^2/\pi(s_1), \quad \text{for} \quad s_3 \sim s_1,$$
$$T_{s_4}^* = T_{s_2}^* = T_{s_2} - Y_1^2/\pi(s_2), \quad \text{for} \quad s_4 \sim s_2,$$
$$\text{and} \qquad T_s^* \equiv T_s, \qquad \text{for all other } s \varepsilon S,$$

(where $\pi(s)$ is as defined in (5.8)), is a member of $\mathfrak{M}^*(P) - \mathfrak{L}_0^*(P)$ which is admissible, showing that $\mathfrak{L}_0^*(P)$ is not complete in $\mathfrak{M}^*(P)$ for any non-unicluster design P. For unicluster designs though $\mathfrak{L}_0^*(P)$ is not complete in $\mathfrak{M}^*(P)$, the slightly enlarged class $\mathfrak{L}^*(P)$ is complete in $\mathfrak{M}^*(P)$ as shown in Theorem (5.1).

It may be noted and is easily verified that for a unicluster design P the class $\mathfrak{M}_0^*(P)$, which consists of all polynomial unbiased estimators of Y with the 'constant part' K_s in each estimator set to zero, contains just one estimator viz the H.T. estimator which is therefore the best. If the definition of an estimator is slightly modified to exclude the 'constant part' (which is a randomized component that has nothing to do with the parameter) then it follows that for a unicluster design there do not exist nontrivial zero functions of the effective sampley (which as observed above is not the case with nonunicluster designs) which therefore forms a complete sufficient statistic thus guaranteeing the existence of a umvue as indeed it has.

From the above discussion it should be apparent that there does not seem to be any theoretical justification for restricting ourselves to $\mathfrak{L}_0^*(P)$ only, for the estimation of Y. In the above we have taken $\mathfrak{M}^*(P)$, which is much wider than $\mathfrak{L}_0^*(P)$ to be our area of search for an optimum estimator. Even though this also is a less ambitious plan than searching the entire class $\mathfrak{I}^*(P)$ we can content ourselves with it firstly because we have just one optimum estimator in $\mathfrak{M}^*(P)$ and secondly because $\mathfrak{M}^*(P)$ is dense in the important class $\mathfrak{C}^*(P)$ of continuous unbiased estimators of Y.

From the fact that the effective sampley $(s, \mathbf{Y})_d$ is the minimal sufficient statistic a plausible conjecture of the author is that any function $f((s, \mathbf{Y})_d)$ which depends on $(s, \mathbf{Y})_d$ fully (i.e., is not independent of any of the arguments that occur in $(s, \mathbf{Y})_d$) is an admissible unbiased estimator of $E(f)$ so that the class $\mathfrak{I}_e^*(f)$ of all such functions f for which $E(f) \equiv Y$ constitute the minimal complete class in $\mathfrak{I}^*(P)$. The validity or otherwise of this conjecture in a general set-up seems to be an open problem of considerable interest but it seems like to be true at least in our present set-up. The example of Godambe and Joshi (1965), page 1712, does not, therefore, provide a valid counter-example because the function f chosen by them does not depend on $(s, \mathbf{Y})_d$ fully. By letting $\beta(s^4, U_2) = \beta(s^5, U_1) = \beta(s^6, U_3) = 0$, as they did in their example, they have in fact made samples that were hitherto not effectively equivalent as effectively equivalent samples but let their estimator to vary over such samples in which case the statistic obviously gives an inadmissible estimator.

We shall now compare our results with the recent results of Godambe and Joshi (1965) and Joshi (1965a, b). Godambe and Joshi proved the admissibility of the H.T. estimator $\hat{Y}_{\mathrm{HT}}$ in $\mathfrak{I}^*(P)$ when R^N is the parameter space. They further remarked (cf. remark (4.1) of their paper) that $\hat{Y}_{\mathrm{HT}}$ is admissible in $\mathfrak{I}^*(P)$ even when the parameter space is restricted to any interval round the origin of R^N. This, however, only implies that given any estimator $T \varepsilon \mathfrak{I}^*(P)$ other than $\hat{Y}_{\mathrm{HT}}$ there is a dense set of points $\mathbf{Y}$ around the origin of R^N at which $V(\hat{Y}_{\mathrm{HT}}) < V(T)$. If parameter spaces like intervals, defined by $\alpha_i \leqq Y_i \leqq \alpha_i'$, $(1 \leqq i \leqq N)$, where α_i's and α_i''s are some constants, are of any interest to us those that are of real interest are those for which $\alpha_i > 0$, $\alpha_i' > 0$ and one is rarely interested in the behaviour of an estimator around the origin of R^N. However, it is easy to see that no estimator can be admissible even in the smaller class $\mathfrak{L}_0^*(P)$ in all intervals of the parameter space that are of the form $\alpha_i \leqq Y_i \leqq \alpha_i'$ where $\alpha_i > 0$, $\alpha_i' > 0$. For, if there is one such, say T_0, then corresponding to any fixed point $\mathbf{Y}^{(0)}$ of R^N for which $Y_i^{(0)} > 0$, for all i, we can choose a sequence of intervals of the above type which converge to $\mathbf{Y}_i^{(0)}$. Given any other estimator $T_1 \varepsilon \mathfrak{L}_0^*(P)$ we should then have $V(T_0) \leqq V(T_1)$ at $\mathbf{Y}^{(0)}$, in view of the continuity of

$V(T_0) - V(T_1)$ in all its arguments. Since $\mathbf{Y}^{(0)}$ is arbitrary this implies that $V(T_0) \leqq V(T_1)$, for all $\mathbf{Y}$ in the interior of the positive quadrant of R^N and since T_1 is any member of $\mathfrak{L}_0^*(P)$ this clearly is impossible.

While the results of these authors are more general in one direction (viz consideration of the bigger class $\mathfrak{J}^*(P)$ as against our $\mathfrak{M}^*(P)$) and point out some of the good properties of the H.T. estimator, they are less conclusive than ours as their considerations did not preclude any other estimator from possessing those good properties. In fact it seems very likely that our main result, i.e., Theorem (3.1) will remain true if $\mathfrak{M}^*(P)$ is replaced by the larger class $\mathfrak{J}^*(P)$ but we shall not bother to attempt this generalization as there are no competing estimators left, save the H.T. estimator, after the first round of competition.

Relaxing the criterion of unbiasedness and restricting to $\mathfrak{L}_0^*(P)$ these authors proved that even for ν-invariant designs, i.e., for which

$$\nu(s) = \nu, \quad \text{a.s. } [P] \tag{6.5}$$

there are a large number of admissible estimators, though again the H.T. estimator too is one among them. Thus the relaxation of the criterion of unbiasedness does not help us to pin down a unique optimum estimator. Besides, as yet, there is not enough mathematical justification to restrict ourselves to ν-invariant designs alone. More discussion on this aspect is given in the next section.

7. Final reduction and further problems. We shall now pick up our discussion of Section 2. The central problem is to choose an 'optimum' strategy $H_0(P_0, T_0)$ from the class $\mathfrak{H}(\nu)$ of all strategies $H(P, T)$ for which

$$\nu(P) = \nu, \tag{7.1}$$

where ν is a given positive number. Since unicluster designs have a serious drawback as explained in Remark (1) of Section 5, we shall confine our attention to the non-unicluster designs only. Restricting T in $H(P, T)$ to $\mathfrak{M}^*(P)$ and accepting the criteria of unbiasedness and h-admissibility (wrt to variance as the loss function) we then have accomplished the first step of the problem viz, the choice of an optimum $T_0 = T_0(P)$, for any fixed P which is not unicluster design and which satisfies (2.4). This optimum $T_0(P)$ is the $\hat{Y}_{\mathrm{HT}}(P)$ of (2.6). We can thus proceed with the second step which is the choice of 'optimum' P_0 from among those P's satisfying (7.1).

If uniform minimum variance is taken to be the criterion of optimality now, for any two designs P_1 and P_2 belonging to the class $\mathcal{P}(\nu)$ of designs satisfying (7.1), P_1 is superior to P_2, ($P_1 \succ P_2$, in symbols) iff

$$V(\hat{Y}_{\mathrm{HT}}(P_1)) \leqq V(\hat{Y}_{\mathrm{HT}}(P_2)), \quad \text{for all } \mathbf{Y} \,\varepsilon\, R^N. \tag{7.2}$$

For the variance of $\hat{Y}_{\mathrm{HT}}(P)$ we have

$$V(\hat{Y}_{\mathrm{HT}}(P)) = \sum_{i=1}^N Y_i^2/\pi_i + \sum\sum_{i\neq j}^N (Y_iY_j/\pi_i\pi_j)\pi_{ij} - Y^2 \tag{7.3}$$

so that (7.2) requires

$$\sum Y_i^2\{(\pi_i(P_1))^{-1} - (\pi_i(P_2))^{-1}\}$$

$$(7.4) \qquad + \sum \sum Y_i Y_j \{\pi_{ij}(P_1)/\pi_i(P_1)\pi_j(P_1) - \pi_{ij}(P_2)/\pi_i(P_2)\pi_j(P_2)\} \leqq 0, \quad \text{for all } \mathbf{Y} \varepsilon R^N.$$

A set of necessary conditions for (7.4) to hold good are

$$(7.5) \qquad \pi_i(P_1) \geqq \pi_i(P_2), \quad \text{for} \quad 1 \leqq i \leqq N$$

and since

$$(7.6) \qquad \sum \pi_i(P_1) = \nu(P_1) = \nu = \nu(P_2) = \sum \pi_i(P_2)$$

from (7.5) and (7.6), a set of necessary conditions for (7.4) to hold good are

$$(7.7) \qquad \pi_i(P_1) = \pi_i(P_2), \quad \text{for} \quad 1 \leqq i \leqq N.$$

Thus it is clear that an optimum P_0 does not exist in $\mathcal{P}(\nu)$ unless there is an 'optimum' set of $\pi_i(P_0)$'s. Such an optimum set does not exist under the criterion of uniform minimum variance alone.

When an optimum P_0 does not exist in $\mathcal{P}(\nu)$ it is of interest to characterize fully the minimum complete class, $\mathcal{P}^*(\nu)$ say, of designs in $\mathcal{P}(\nu)$. At this stage it is convenient to break up the problem into two steps. Firstly for a fixed set $\boldsymbol{\pi} = (\pi_1, \pi_2, \cdots, \pi_N)$ of π_i's which satisfy $0 < \pi_i \leqq 1$ and $\sum \pi_i = \nu$ we may investigate for the optimum $P_{0,\boldsymbol{\pi}}$ in the class $\mathcal{P}_{\boldsymbol{\pi}}$ of designs for which $\pi_i(P) = \pi_i$, and then proceed to find the optimum set of π_i's through some other criteria of optimality.

In a given class $\mathcal{P}_{\boldsymbol{\pi}}$ an optimum $P_{0,\boldsymbol{\pi}}$ exists iff for any other $P_{1,\boldsymbol{\pi}} \varepsilon \mathcal{P}_{\boldsymbol{\pi}}$

$$(7.8) \qquad \sum \sum_{i \neq j} Y_i Y_j / \pi_i \pi_j \{\pi_{ij}(P_{0,\boldsymbol{\pi}}) - \pi_{ij}(P_{1,\boldsymbol{\pi}})\} \leqq 0, \quad \text{for all } \mathbf{Y} \varepsilon R^N.$$

If attention is restricted to $\mathbf{Y}$'s, for which $Y_i \geqq 0$, for all i, which case is in fact of main interest in sample survey theory, then a set of sufficient conditions for (7.8) to hold good are

$$(7.9) \qquad \pi_{ij}(P_{0,\boldsymbol{\pi}}) \leqq \pi_{ij}(P_{1,\boldsymbol{\pi}}), \quad \text{for} \quad 1 \leqq i \neq j \leqq N.$$

Even if an optimum $P_{0,\boldsymbol{\pi}}$ does not exist in $\mathcal{P}_{\boldsymbol{\pi}}$ one can still proceed to find the minimal complete class in $\mathcal{P}_{\boldsymbol{\pi}}$. Whether or not a restriction to ν-invariant designs is justified can now be examined through this approach. Recalling a basic relation between π_{ij}'s and ν_s's for any sampling design (Hanurav 1962b) we have

$$(7.10) \qquad \sum \sum_{i \neq j} \pi_{ij} = \nu(\nu - 1) + V(\nu_s)$$

where $V(\nu_s)$ is the variance of ν_s. Observing further that $1 \leqq \nu_s \leqq N$, we have

$$(7.11) \qquad \nu(\nu - 1) + \theta(1 - \theta) \leqq \sum \sum \pi_{ij} \leqq (N - \nu)(\nu - 1)$$

where $\theta = \nu - [\nu]$ is the fractional part of ν. If, given any design $P_{1,\boldsymbol{\pi}} \varepsilon \mathcal{P}_{\boldsymbol{\pi}}$, for which $\sum \sum \pi_{ij}(P_{1,\boldsymbol{\pi}})$ exceeds the bound given on the left of (7.11), we can find a $P_{0,\boldsymbol{\pi}} \varepsilon \mathcal{P}_{\boldsymbol{\pi}}$ for which (7.9) holds good and $\sum \sum \pi_{ij}(P_{0,\boldsymbol{\pi}})$ equals the bound on the left of (7.11) then this provides a complete justification (under the present criterion) to restrict ourselves to designs for which $V(\nu_s)$ is minimum.

In particular, when ν is an integer this provides a valid justification to restrict ourselves to ν-invariant designs. However, whether the above given result is true or false is still an open problem which seems to be of considerable interest.

We may remark here that Theorem (7.1) of Godambe and Joshi (1965) does not provide enough justification to restrict ourselves to ν-invariant designs, when ν is an integer. Their theorem shows that given any ν-variant design $P_{1,\pi} \in \mathcal{P}_\pi$, there does not exist a non-ν-invariant design $P_{2,\pi} \in \mathcal{P}_\pi$ for which $V(\hat{Y}_{\mathrm{HT}}(P_{2,\pi})) \leqq V(\hat{Y}_{\mathrm{HT}}(P_{1,\pi}))$, for all $\mathbf{Y} \in R^N$. This will only imply that ν-invariant designs are admissible in $\mathcal{P}_\pi$ for any given set of π_i's, but what is needed is something stronger viz. that they constitute the minimum complete class in $\mathcal{P}_\pi$, for any $\boldsymbol{\pi}$.

The problem can be solved to some extent if the criterion for the selection of optimum P_0 in $\mathcal{P}(\nu)$ is slightly diluted. We have seen above that under the criterion of uniform minimum variance no set of optimum π_i's exist so that even if optimum $P_{0,\pi}$'s can be found in $\mathcal{P}_\pi$'s, there still remains the problem of optimum choice of $\boldsymbol{\pi}$. When nothing is known, a priori, of the values Y_i's, there is no reason why the π_i's should be unequal, if we stick to our cost function as $\nu(P)$. Unequal values for the π_i's are justified generally when, a priori, we have guesses about the Y_i's. The situation in practice is that we have the values $\mathbf{x} = (x_1, x_2, \cdots, x_N)$ of an auxiliary positive variate $\mathcal{X}$ which is correlated with $\mathcal{Y}$. The a priori guesses of Y_i's can be formulated in many cases in terms of a priori distribution θ which satisfies

$$\text{(7.12)}\quad \begin{aligned} &\text{(i)} & \mathcal{E}_\theta(Y_i \mid x_i) &= \beta x_i \\ &\text{(ii)} & \mathcal{V}_\theta(Y_i \mid x_i) &= \sigma_i^2 \\ \text{and}\ &\text{(iii)} & \mathcal{C}ov_\theta(Y_i, Y_j \mid x_i, x_j) &= 0 \end{aligned}$$

for $1 \leqq i \neq j \leqq N$, where β and σ_i^2's are unknown positive constants. If the criterion for the selection of optimum $P_{0,\pi}$ is diluted by demanding *θ-optimality* only, i.e.,

$$\text{(7.13)}\quad \begin{gathered} \mathcal{E}_\theta V(\hat{Y}_{\mathrm{HT}}(P_{0,\pi})) \leqq \mathcal{E}_\theta V(\hat{Y}_{\mathrm{HT}}(P_{1,\pi})), \\ \text{for all } \beta \geqq 0, \quad \sigma_i^2 \geqq 0, \quad \mathbf{x} \geqq 0, \end{gathered}$$

then from an earlier result of the author (cf. theorem of Hanurav (1962c)) it follows that in the class $\mathcal{P}_\pi(\nu, x)$ of designs for which

$$\text{(7.14)}\quad \pi_i = \nu x_i / x, \quad \text{for} \quad 1 \leqq i \leqq N$$

the class $\mathcal{P}_\pi^*(\nu, x)$ of designs for which $V(\nu_s)$ is minimum, form the minimal complete class. In particular, when ν is an integer, this minimal complete class is a class of ν-invariant designs. If ν is an integer and the σ_i^2's occurring in (ii) of (7.12) can be expressed as

$$\text{(7.15)}\quad \sigma_i^2 = \sigma^2 \cdot x_i^2,$$

then from an earlier result of Godambe (1955) we have the π_i's given by (7.14) as the θ-optimum set of π_i's, so that the problem is completely solved. [It may be remarked that Godambe proves that in the class of all strategies $H(P, T)$, for which (i) $\nu(s) = \nu$, a.s. [P] and (ii) $T \varepsilon \mathfrak{L}_0^*(P)$ any strategy which satisfies (a) $\pi_i(P) = \nu \cdot x_i/x$ and (b) $T = \hat{Y}_{\text{HT}}(P)$, is θ-optimum where θ is a prior distribution satisfying (7.12) and (7.15). However, from his proof it can be easily seen that (1) above can be replaced by the weaker condition $\sum \pi_i(P) = \nu$ and then the condition $\nu(s) = \nu$, a.s. [P] follows as a necessary condition, in addition to (a) and (b) above, to be satisfied by the optimum P. Thus the ν-invariance need not be stipulated in the hypothesis but follows as a consequence of the optimality of the designs. Note, incidentally, that this result does not need the criterion of h-admissibility but the estimator is restricted to $\mathfrak{L}_0^*(P)$'s.]

It can be verified that if (7.15) does not hold good, then even if the σ_i^2's can be expressed as

$$\sigma_i^2 = \sigma^2 \cdot x_i^\gamma, \qquad \gamma \geqq 0 \tag{7.16}$$

which, incidentally, is a common situation met with in practice (with γ usually lying between 1 and 2) then no θ-optimum set of π_i's exist if $\gamma \neq 2$. When (7.15) holds good the author (Hanurav (1966b)) has considered the problem for $\nu = 2$ in detail and obtained an optimum P_0 in $\mathcal{P}(\nu)$ which possesses many more desirable properties and can be reasonably termed as 'fully optimum.' Later (Hanurav (1966c)) he extended the result to cover general integral values of ν. It is of interest to pursue the problem further when the a priori distribution does not satisfy (7.15) or when it does not satisfy even (7.12).

8. Acknowledgment. It is a pleasure for the author to record his thanks to Professor V. P. Godambe for his stimulating discussions.

REFERENCES

[1] AGGARWAL, O. P. (1959). Bayes and minimax procedures in sampling from finite and infinite populations. *Ann. Math. Statist.* **30** 206–218.

[2] BASU, D. (1958). On sampling with and without replacement. *Sankhyā.* **20** 287–294.

[3] CHERNOFF, H. (1965). *Math. Reviews* **29** 4166.

[4] GODAMBE, V. P. (1955). A unified theory of sampling from finite populations. *J. Roy. Statist. Soc.* **B17** 269–277.

[5] GODAMBE, V. P. (1960). An admissible estimator for any sampling design. *Sankhyā.* **22** 285–288.

[6] GODAMBE, V. P. (1965). Contributions to the unified theory of sampling. *Rev. Int. Statist. Inst.* **33** 242–258.

[7] GODAMBE, V. P., and JOSHI, S. M. (1965). Admissibility and Bayes estimation in sampling finite populations—I. *Ann. Math. Statist.* **36** 1707–1722.

[8] HÁJEK, J. (1959). Optimum strategy and other problems in probability sampling. *Časopis Pěst. Math.* **84** 387–423.

[9] HÁJEK, J. (1965). *Math. Reviews* **29** 715.

[10] HANURAV, T. V. (1962a). An existence theorem in sampling. *Sankhyā.* **A24** 227–230.

[11] HANURAV, T. V. (1962b). Some sampling schemes in probability sampling. *Sankhyā.* **A24** 421–428.

[12] HANURAV, T. V. (1962c). On Horvitz and Thompson estimator. *Sankhyā.* **A24** 429–436.

[13] HANURAV, T. V. (1965). Optimum sampling strategies and some related problems. Ph.D. thesis submitted to the Indian Statistical Institute.

[14] HANURAV, T. V. (1966a). Some aspects of sampling finite populations. *Sankhyā.* **A28** 175–204.

[15] HANURAV, T. V. (1966b). Optimum utilization of auxiliary information—πPS sampling of two units from a stratum. *J. Roy. Statist. Soc.* **B**.

[16] HANURAV, T. V. (1966c). Optimum πPS sampling (Abstract). *Ann. Math. Statist.* **37** 1859.

[17] HORVITZ, D. G., and THOMPSON, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Jour. Amer. Math. Assn.* **47** 663–685.

[18] JOSHI, S. M. (1965a). Admissibility and Bayes estimation in sampling finite populations—II. *Ann. Math. Statist.* **36** 1723–1729.

[19] JOSHI, S. M. (1965b). Admissibility and Bayes estimation in sampling finite populations—III. *Ann. Math. Statist.* **36** 1730–1742.

[20] KOOP, J. C. (1957). Contributions to the general theory of sampling finite populations without replacement. Mimeo 296 Inst. of Stat., Univ. of N. Carolina.

[21] MURTHY, M. N. (1958). On ordered and unordered estimators. *Sankhyā.* **A25** 254–262.

[22] PATHAK, P. K. (1964). Sufficiency in sampling theory. *Ann. Math. Statist.* **35** 795–809.

[23] PRABHU, AJGAONKAR. (1962). Unpublished Ph.D. thesis. Karnatak University, India.

[24] ROY, J. and CHAKARVOTY, I. M. (1960). Estimating the mean of a finite population. *Ann. Math. Statist.* **31** 392–398.

[25] TAKEUCHI, K. (1961). Some remarks about unbiased estimation in sampling from finite populations. *J. Univ. of Scientists and Engineers, Japan* **8** 35–38.