ADMISSIBILITY AND BAYES ESTIMATION IN SAMPLING FINITE POPULATIONS. I

By V. P. Godambe[1] and V. M. Joshi[2]

The Johns Hopkins University and University of North Carolina, Chapel Hill

1. Introduction. The concept of linear estimation for a finite population total was generally defined by Godambe (1955). Since then following results have been established concerning the class of *linear* unbiased estimates:

(i) non-existence of a uniformly least variance estimate [Godambe (1955)],

(ii) non-existence of uniformly least variance estimates in subclasses [Koop (1957), Prabhu Ajagaonkar (1962)],

(iii) admissibility of a certain estimate in common use [Godambe (1960), Roy and Chakravarti (1960)],

(iv) the Bayesness of the above estimate [Godambe (1955), Hájek (1959)].
The results (i), (iii), (iv) above are, in the present article, extended to entire class of unbiased estimates, removing the restriction of linearity.

It is interesting that though unbiased non-linear estimates of the population total could be easily constructed, (Section 3), none of them have been proposed in the literature. The concept of *invariance*, namely that an estimate of population mean should be independent of the scale and origin, discussed by Roy and Chakravarti (1960), does not enable us to restrict ourselves exclusively to linear estimates. Neither does, as shown in Section 3, the intuitive feeling that for every nonlinear unbiased estimate of the population total, there exists an unbiased linear estimate with uniformly smaller variance, is true.

In the present paper, after extending the admissibility of the estimate referred to in (iii) above, we have also investigated the admissibility of the estimates for its variance, one proposed by Horvitz and Thomson (1952) and the other by Yates and Grundy (1953). The Horvitz and Thomson's estimator for variance is proved to be admissible.

Finally we have presented some results, in connection with estimation of the population total, where the usual restriction of unbiasedness is removed, especially (iii) is established without the restriction of unbiasedness.

2. Definitions and preliminaries. Let U denote a finite population (set) of units (elements) u, i.e. $U = \{u\}$. On U is defined a real variate (function) x, $x(u)$ being its value for the unit u. If $\mathfrak{X}$ is the class of all real variates x defined on U and G a real valued function on $\mathfrak{X}$ then the general problem in sampling is to estimate $G(x)$ by observing the values $x(u)$ for just those units u belonging to specified subset s of U, given the sampling design (Definition 2.3). In the usual notation $s \subset U$. More often however we particularly are concerned with estimating a special type of function G, conventionally called population total.

Received 13 March 1964; revised 24 May 1965.

[1] On leave from the Institute of Science, Bombay.

[2] On leave from the Maharashtra Government, Bombay.

DEFINITION 2.1. A function T on $\mathfrak{X}$ is called the population total if for every $x \varepsilon \mathfrak{X}$,

$$T(x) = \sum_{u \varepsilon U} x(u). \tag{1}$$

Thus here the problem is to estimate $T(x)$ by observing just those values $x(u)$ for which $u \varepsilon s$ where $s \subset U$.

DEFINITION 2.2. A subset s of U, $s \subset U$, is called a *sample*.

Let S denote the set of all possible samples (subsets of U) s. On S is defined a function p such that $1 \geqq p(s) \geqq 0$, for all $s \varepsilon S$ and $\sum_{s \varepsilon S} p(s) = 1$.

DEFINITION 2.3. A *sampling design* $d = (S, p)$.

It is easy to see that all the known sample survey designs are special cases of $d = (S, p)$ in the above definition. Of course it is almost never practicable to form all possible samples S and then choose one from them with prescribed probabilities p. Instead, to implement a sampling design $d = (S, p)$, some one of the easily managable *sampling procedures* such as stratification, subsampling, drawing units one after another with varying probabilities, etc., is adopted. Thus to implement a given sampling design $d = (S, p)$, more than one sampling procedures may be possible. One procedure is always there, namely, forming all possible samples s, and then choosing one from them with prescribed probabilities p. In this connection some interesting results are due to Hanurav (1962). We have discussed this topic in Section 5.

DEFINITION 2.4. $n(s)$ the total number of units u in the sample s is called the *sample size* of s.

Let, for a given sampling design $d = (S, p)$,

$$\sum_{s \varepsilon u} p(s) = \pi(u) \tag{2}$$

where $s \varepsilon u$ denote all samples s having the unit u. Then

DEFINITION 2.5. For a sampling design $d = (S, p)$ the *inclusion probability* for the unit u is $\pi(u)$, as given by (2), for all $u \varepsilon U$.

Due to one of the authors (1955) is the following central relation between the inclusion probabilities $\pi(u)$ and the sample sizes $n(s)$:

$$E(n) = \sum_{s \varepsilon S} n(s)p(s) = \sum_{u \varepsilon U} \pi(u), \tag{3}$$

whatever the sampling design $d = (S, p)$ may be. Thus it would be clear that as soon as the inclusion probabilities $\pi(u)$ are specified the *average sample size* $E(n)$ is automatically fixed for the sampling design. In this connection we have some important results due to Hanurav (1962), regarding a particular estimate. Additional relevent results are in Section 7.

DEFINITION 2.6. An *estimate* $e(s, x)$ is a function on $S \times \mathfrak{X}$ (see Definitions 2.1, 2.3), depending on x only through those $x(u)$ for which $u \varepsilon s$. That is for any two x, x' such that $x(u) = x'(u)$ for all $u \varepsilon s$, $e(s, x) = e(s, x')$.

From practical considerations it is evident that the estimate $e(s, x)$ need not be defined at all for samples s for which $p(s) = 0$. Now it is true that the above definition of an estimate is not most general, in the sense that there are estimates in common use and in the literature which are not special cases of $e(s, x)$ in (2.6). However in Section 5, we would present an argument based on the principle of sufficiency to show that we may restrict to Definition 2.6 without any loss of generality in our search for an optimum estimate. Next the estimate $e(s, x)$ is said to be *linear* if

$$e(s, x) = \sum_{u \varepsilon s} \beta(s, u)x(u), \tag{4}$$

where β is a function on $S \times U$ such that $\beta(s, u) = 0$ if $u \not\varepsilon s$.

DEFINITION 2.7. For a given sampling design $d = (S, p)$, an estimate $e(s, x)$ is said be *unbiased* for the population total $T(x)$ in (1) if

$$\sum_{s \varepsilon S} e(s, x)p(s) = T(x) \tag{5}$$

for all $x \varepsilon \mathfrak{X}$.

For a given sampling design d, the class of all unbiased estimates e for the population total T would be denoted by D, i.e.,

$$D = \{e: \sum_{s \varepsilon S} e(s, x)p(s) = T(x), \qquad \text{for all} \quad x \varepsilon \mathfrak{X}\}. \tag{6}$$

For any estimate $e \varepsilon D$ we have its variance $V(e, x)$ where

$$V(e, x) = \sum_{s \varepsilon S} [e(s, x) - T(x)]^2 p(s) \tag{7}$$

for every $x \varepsilon \mathfrak{X}$.

DEFINITION 2.8. For a given sampling design d, an unbiased estimate $\bar{e}$ (i.e. $\bar{e} \varepsilon D$, the class of all unbiased estimates given by (6)) is said to be *admissible* in D if for no other estimate $e \varepsilon D$,

$$V(e, x) \leqq V(\bar{e}, x) \tag{8}$$

for all $x \varepsilon \mathfrak{X}$, strict inequality being true for at least some $x \varepsilon \mathfrak{X}$. The admissibility for any subclass of D is defined similarly.

3. Comments on unbiasedness and linearity. The criteria of unbiased estimation, i.e. estimation satisfying (5), seems to have been taken for granted in the field of sample survey. Here, apart from its intuitive appeal, its additive property is most important. If e_i is an unbiased estimate of some population total T_i then $\sum e_i$ is an unbiased estimate of $\sum T_i$. Further if all e_i are independent by law of large number, $(\sum e_i / \sum T_i) \to 1$, in probability, as more and more estimates are added together. And in a certain sense the convergence is faster, the smaller the variances of e_i. Indeed this provides a very good justification for preferring an unbiased estimate with smaller variance, in sample surveys, where more often the primary aim is to estimate some grand total $\sum T_i$. However in the last section we would investigate the criteria of unbiasedness further.

Now, though in literature non-linear unbiased estimates of the population total have never been proposed, they can be constructed quite easily. Consider an artificial but a very simple illustrative example. The population $U = \{u_1, u_2, u_3\}$ consists of only three units. The sampling design $d = (S, p)$, (Definition 2.3), is such that for the samples $s = (u_1, u_2)$, $s' = (u_1, u_3)$ we have

$p(s) = p(s') = \frac{1}{2}$. For the rest of the samples the probability is zero. Now consider the *non-linear* estimate e given by $e(s, x) = \bar{e}(s, x) + x^2(u_1)$ and $e(s', x) = \bar{e}(s', x) - x^2(u_1)$ where $\bar{e}(s, x) = x(u_1) + 2x(u_2)$ and $\bar{e}(s', x) = x(u_1) + 2x(u_3)$. Clearly e is an unbiased estimate for the population total. Moreover, contrary to ones intuitive feeling it is impossible to construct an unbiased *linear* estimate $l(s, x)$ whose variance (as in (7)) $V(l, x)$ is smaller than or equal to $V(e, x)$ for all $x \varepsilon \mathfrak{X}$. For if $V(l, x) \leqq V(e, x)$ for all $x \varepsilon \mathfrak{X}$, where $V(e, x) = V(\bar{e}, x) + x^4(u_1) + 2x^2(u_1)(x(u_2) - x(u_3))$, considering the magnitudes of terms in $V(l, x)$ and $V(e, x)$ it follows that $V(l, x) \leqq V(\bar{e}, x)$, if x is such that $|x(u)|$ for all $u \varepsilon U$ is less than a certain ϵ. But since in the present case $\bar{e}$ is nothing but an estimate in (9) of Section 4, it follows from Theorem 4.1 and the subsequent Remark 4.1, that $l = \bar{e}$. Next since $V(e, x) = V(\bar{e}, x) + x^4(u_1) + 2x^2(u_1)(x(u_2) - x(u_3))$, it is evident that for some x, $V(e, x) < V(\bar{e}, x)$, thus proving that a linear unbiased estimate l, for which $V(l, x) \leqq V(e, x)$ for all $x \varepsilon \mathfrak{X}$ does *not* exist.

Next, even the principle of *invariance*, discussed by Roy and Chakravarti (1960), stating that an estimate of population mean should be invariant relative to the scale and origin of measurement does not restrict us exclusively to linear estimates. Indeed along with unbiasedness, if some *additional* assumptions such as continuity etc. are made the linearity of estimation may follow. But these additional assumptions appear to be far from necessary or even plausible.

Yet in the literature on sample surveys only two authors seem to have considered the possibility of removing the restriction of linearity. One of them is Aggarwal (1959). Without restricting to unbiased or linear estimates he has established the *minimaxity* of certain estimates over a subset of $\mathfrak{X}$, defined by $\sum_{u \varepsilon U} (x(u) - T(x)/N)^2 \leqq N\sigma^2$, where N is the total number of units in U. In our opinion this is not a very practically useful approach as it is not easy to guess the value of σ^2 of the variate under study. (In Section 8 we demonstrate the failure of "minimaxity" over a practically interesting subset of $\mathfrak{X}$.) To us it seems that practically most fruitful subset of $\mathfrak{X}$ for investigating admissibility, minimaxity, etc. is given by $\alpha_1(u) \leqq x(u) \leqq \alpha_2(u)$, $u \varepsilon U$. Indeed the sampler usually will have such knowledge of α's about the variate under study, from his past experience or from his knowledge of the values $y(u)$, $u \varepsilon U$, of some other correlated variate y. Next Das (1962) has attempted to prove some results without the restriction of linearity. But his results are false. His fallacy rests in his assumption of the existence of a best (i.e. least variance for all $x \varepsilon \mathfrak{X}$) estimate in the entire class of unbiased estimates D, in (6). Then he claims to have proved that the *linear* estimate $\bar{e}$ in (9) (to follow) is the best estimate in D, which implies that $\bar{e}$ is also the best in the subclass of D, consisting of all the linear estimates in D. This straightaway contradicts a well known result by Godambe (1955) namely the non-existence of a best estimate in the class of all unbiased linear estimates, about which Das seems to be completely unaware. He has committed a similar mistake, while claiming to have proved the bestness of a certain unbiased estimate of the variance of $\bar{e}$ in (9).

4. An admissible estimate in D. Though in the last section we would investigate to some extent the criteria of unbiasedness, until then we would take it for granted. Now it is easy to see that *unless* the sampling design $d = (S, p)$, (Definition 2.3), is such that the inclusion probabilities $\pi(u)$, $u \varepsilon U$, in (2) are *all* non-zero, no estimate of T in (1) can be unbiased. That is the class D in (6) is empty.

THEOREM 4.1. *For any sampling design d, having the inclusion probabilities* $\pi(u) > 0$ *for all* $u \varepsilon U$, *the estimate* $\bar{e}$ *given by*

$$\bar{e}(s, x) = \textstyle\sum_{u \varepsilon s} x(u)/\pi(u) \tag{9}$$

is admissible in D, according to the Definition 2.8.

After the non-existence of a uniformly (i.e. for all $x \varepsilon \mathfrak{X}$) least variance estimate was demonstrated for the class of all *linear* unbiased estimates of the population total, for the same class the $\bar{e}$ in (9) was proved to be admissible, according to the Definition 2.8, by Godambe (1960) and Roy and Chakravarti (1960), independently. We now propose to take the reverse way round. First the admissibility of $\bar{e}$ in (9) would be established for the entire class of unbiased estimates, D in (6). And we then, as a logical consequence, deduce the non-existence of a uniformly (i.e. for all $x \varepsilon \mathfrak{X}$) least variance estimate in D.

PROOF OF THEOREM 4.1. Let $\bar{e}$ in (9) be not admissible. Then by Definition 2.6, there would be an estimate

$$\begin{aligned} e(s, x) &= \bar{e}(s, x) + [e(s, x) - \bar{e}(s, x)] \\ &= \bar{e}(s, x) + h(s, x), \end{aligned} \tag{10}$$

such that $e \varepsilon D$, i.e.

$$\textstyle\sum_{s \varepsilon S} h(s, x)p(s) = 0 \tag{11}$$

and $\sum_{s \varepsilon S} e^2(s, x)p(s) \leqq \sum_{s \varepsilon S} \bar{e}^2(s, x)p(s)$ for all $x \varepsilon \mathfrak{X}$. That is

$$\textstyle\sum_{s \varepsilon S} h^2(s, x)p(s) \leqq -2 \sum_{s \varepsilon S} h(s, x)\bar{e}(s, x)p(s) \tag{12}$$

for all $x \varepsilon \mathfrak{X}$. We note here that $h(s, x)$ is a function defined on $S \times \mathfrak{X}$, depending on $\mathfrak{X}$ only through those values $x(u)$ of x for which $u \varepsilon s$ (see Definition 2.6). Now let $\mathfrak{X}_k \subset \mathfrak{X}$ such that if $x \varepsilon \mathfrak{X}_k$ then *just* k values of x are non-zero (that is $x(u) \neq 0$ for just some k units $u \varepsilon U$). We now first establish the

LEMMA. *If* $h(s, x)p(s) = 0$ *for all* (s, x) *such that* $s \varepsilon S$, $x \varepsilon \mathfrak{X}_k$, *then* $h(s, x)p(s) = 0$ *for all* (s, x) *such that* $s \varepsilon S$, $x \varepsilon \mathfrak{X}_{k+1}$.

Let $x' \varepsilon \mathfrak{X}_{k+1}$ then from (11) and (12) we have

$$\textstyle\sum_{i=0}^{k+1} \sum_{s \varepsilon Si} h(s, x')p(s) = 0 \tag{13}$$

and

$$\textstyle\sum_{i=0}^{k+1} \sum_{s \varepsilon Si} h^2(s, x')p(s) \leqq -2 \sum_{i=0}^{k+1} \sum_{s \varepsilon Si} h(s, x')\bar{e}(s, x')p(s), \tag{14}$$

where $Si \subset S$ such that a sample $s \varepsilon Si$ if and only if *just* for i number of units u in s, $x'(u) \neq 0$. Now $h(s, x)p(s) = 0$ for all (s, x) such that $s \varepsilon S$, $x \varepsilon \mathfrak{X}_k$, implies $h(s, x')p(s) = 0$ for all $s \varepsilon Si$ for $i = 1, \cdots, k$. Thus from (13) and (14) we have

(15) $$\sum_{s \varepsilon S_{k+1}} h(s, x')p(s) = 0$$

and

(16) $$\sum_{s \varepsilon S_{k+1}} h^2(s, x')p(s) \leqq -2 \sum_{s \varepsilon S_{k+1}} h(s, x')\bar{e}(s, x')p(s).$$

However in (16)

(17) $\bar{e}(s, x')$ for all samples $s \varepsilon S_{k+1}$ is constant,

namely $\sum_{u \varepsilon U} x'(u)/\pi(u)$. Thus from (16) we have

(18) $$\sum_{s \varepsilon S_{k+1}} h^2(s, x')p(s) \leqq -2(\sum_{u \varepsilon U} x'(u)/\pi(u)) \sum_{s \varepsilon S_{k+1}} h(s, x')p(s).$$

Now (15) and (18) imply $h(s, x')p(s) = 0$ for all $s \varepsilon S_{k+1}$. Noting $\sum_{i=0}^{k+1} Si = S$, we have $h(s, x')p(s) = 0$ for all $s \varepsilon S$. Thus the Lemma is proved.

Further from (12) it follows that

(19) $h(s, x)p(s) = 0$ for all (s, x) such that $s \varepsilon S, x \varepsilon \mathfrak{X}_0$.

It follows from (19) and the above Lemma that

(20) $h(s, x)p(s) = 0$ for all (s, x) such that $s \varepsilon S, x \varepsilon \mathfrak{X}$.

And (10) and (20) imply the Theorem 4.1.

Once the admissibility of the estimate $\bar{e}$ in (9) is proved an example of an inadmissible estimate may be of interest. Consider the following simple though artificial illustration. The population $U = \{u_1, u_2, u_3\}$ consists of three units. Let the samples $s^1 = (u_1)$, $s^2 = (u_2)$, $s^3 = (u_3)$, $s^4 = (u_1, u_2)$, $s^5 = (u_1, u_3)$ and $s^6 = (u_2, u_3)$. Now the sampling design $d = (S, p)$ is such that $p(s^i) = \frac{1}{6}$ for $i = 1, \cdots, 6$, the probability for other samples being zero. Next consider an unbiased linear estimate $l(s, x)$ for the population total. As in (4), $l(s, x) = \sum_{u \varepsilon s} \beta(s, u)x(u)$. For unbiasedness, let $\beta(s^4, u_2) = \beta(s^5, u_1) = \beta(s^6, u_3) = 0$ and $\beta(s^4, u_1) = 6 - \beta(s^1, u_1)$, $\beta(s^5, u_3) = 6 - \beta(s^3, u_3)$, $\beta(s^6, u_2) = 6 - \beta(s^2, u_2)$. Then the variance of $l(s, x)$ as in (7) is given by,

$$V(l, x) = \tfrac{1}{6} \sum_{i=1}^{3} x^2(u_i)(\beta^2(s^i, u_i) + (6 - \beta(s^i, u_i))^2) - T^2(x)$$

where $T(x) = \sum_{i=1}^{3} x(u_i)$. Clearly for the variations of $\beta(s^i, u_i)$, $i = 1, 2, 3$, $V(l, x)$ is minimized for $\beta(s^i, u_i) = 3$ for $i = 1, 2, 3$, whatever x may be. And evidently for all other values of $\beta(s^i, u_i)$, $i = 1, 2, 3$, the estimate $l(s, x)$ is inadmissible. This simple example also shows that the complete class of linear unbiased estimates as characterized by Roy and Chakravarti (1960) is in fact *not* minimal complete.

REMARK 4.1. It may be seen that the Theorem 4.1 is valid for any subset of $\mathfrak{X}$ defined by $-\alpha_1(u) \leqq x(u) \leqq \alpha_2(u)$ where $\alpha_1(u) \geqq 0$, $\alpha_2(u) > 0$ for all $u \varepsilon U$.

This is practically very important for as said before in Section 3, the sampler will usually have some knowledge (from previous investigations or otherwise) concerning each unit $u \varepsilon U$ which enables him to specify some range as

$\alpha_1(u) \leqq x(u) \leqq \alpha_2(u)$, for the value of the variate x associated with the unit u, for $u \varepsilon U$. Note Remark 4.1 covers all the positive intervals of $\mathfrak{X}$, $0 \leqq x(u) < \alpha(u)$, $u \varepsilon U$, which are more often relevent.

REMARK 4.2. The Theorem 4.1 is also valid for any subset of $\mathfrak{X}$ defined by $x(u) = c(u)$ or $x(u) = 0$ for $u \varepsilon U$, $c(u)$, $u \varepsilon U$ being some fixed numbers. Now substituting in the above remark $c(u) = 1$, $u \varepsilon U$ we derive the admissibility for the case of sampling attributes, i.e. zero-one variables. (The Remark 4.2 and its implication is due to the referee.)

REMARK 4.3. The only property of $\bar{e}$ in (9) (in addition to its unbiasedness) that is utilized in proving its *admissibility* is given by (17). It can be more clearly stated as: the estimate $\bar{e}(s, x)$ in (9) depends on s only through those units $u \varepsilon s$ for which $x(u) \neq 0$. That is for any two samples s_1, s_2 and a given x *if* for all $u \varepsilon s_1 + s_2 - s_1 s_2$, $x(u) = 0$, *then* $e(s_1, x) = e(s_2, x)$. Thus over previous proof of admissibility in fact establishes the more general

THEOREM 4.2. *Any function $g(s, x)$ defined on $S \times \mathfrak{X}$ which depends on (s, x) only through those $(u, x(u))$ for which $u \varepsilon s$ and $x(u) \neq 0$ is an unbiased admissible estimate of $G(x) = \sum_{s \varepsilon S} g(s, x)p(s)$, according to the Definition 2.8 modified by replacing $T(x)$ by $G(x)$ in (5) and (7).*

(Note: $G(x)$ here may depend on $p(s)$, $s \varepsilon S$, in contradistinction to $T(x)$.)

It is easy to see that Theorem 4.1 is a special case of Theorem 4.2. Next since the estimate $\bar{e}$ in (9) is a *linear* (as in (4)) unbiased estimate of the population total and since the non-existence of a uniformly (i.e. for all $x \varepsilon \mathfrak{X}$) least variance estimate, in the class of all *linear* unbiased estimates, is already established by one of the authors (1955), from Theorem 4.1 we have,

COROLLARY 4.1. *In the entire class D in (6) of the unbiased estimates of the population total, a uniformly (i.e. for all $x \varepsilon \mathfrak{X}$) least variance estimate does not exist.*

Now two unbiased estimates have been proposed in the literature for the variance of $\bar{e}$ in (9), namely

$$(21) \quad v_1 = \sum_{u \varepsilon s} (1 - \pi(u))x^2(u)/\pi^2(u) + \sum_{u,u' \varepsilon s, u \neq u'} (1 - \pi(u)\pi(u')/\pi(u, u'))x(u)x(u')/\pi(u)\pi(u'),$$

$$(22) \quad v_2 = \sum_{u,u' \varepsilon s} (\pi(u)\pi(u')/\pi(u, u') - 1)(x(u)/\pi(u) - x(u')/\pi(u'))^2,$$

where $\pi(u, u') = \sum_{s \ni u,u'} p(s)$, i.e. summation of $p(s)$ over all samples containing u, u'. Similarly $\sum_{u,u' \varepsilon s}$ denotes summation over all pairs of units u, u' in s. The estimate v_1 has been proposed by Horvitz and Thomson (1952) and v_2 by Yates and Grundy (1953). Yates and Grundy (1953) have vigorously rejected v_1 in (21), preferring v_2 in (22) on the considerations of the sampling fluctuations, mostly based on illustrative examples. However from Theorem 4.2 we have

THEOREM 4.3. *In the entire class of unbiased estimates of $V(\bar{e}, x)$ the variance of $\bar{e}$ in (9), the estimate v_1, in (21) is admissible.*

It is interesting that the Theorem 4.2 does *not* say whether the estimate v_2 in (22) is admissible or not. Now the result that in the class of all unbiased *quadratic* estimates of $V(\bar{e}, x)$, no uniformly (i.e. for all $x \varepsilon \mathfrak{X}$) least variance estimate exists, can be established on the same lines as (i) in the Section 1 without any difficulty. Next, following the arguments leading to Corollary 4.1, we have from Theorem 4.3.

COROLLARY 4.2. *In the entire class of unbiased estimates of* $V(\bar{e}, x)$, *no uniformly* (*i.e. for all* $x \varepsilon \mathfrak{X}$) *least variance estimate exists.*

5. The principle of sufficiency. (Here we reproduce some ideas due to Hájek (1959) for continuity of presentation.) In the comments following the Definition 2.6 of an estimate, we already have said that the Definition 2.6 is not general enough to cover all the known or possible estimates. This can be expressed more precisely as follows: As it has been pointed out following the Definition 2.3 of a sampling design $d = (S, p)$, that to implement $d = (S, p)$ in practice one or more *sampling procedures* are possible. And usually the outcome of the application of a sampling procedure would not be *just* a sample s but something more. For instance, consider the sampling procedure of drawing units one after another. Here we not only have a sample s but also know the order in which the units of s have been drawn. Or consider a more complicated sampling procedure given by Rao, Hartley and Cochran (1962), where first the population is stratified by some random device and then from each stratum some units are drawn. Here in addition to a sample s we also observe the strata to which the different units of s belong. Further the order in which the units from different strata are drawn is also provided. All this may be denoted by $\bar{s}$. (Here one must note that $\bar{s}$ does *not* include any variate values $x(u)$ at all.) Thus if $\bar{S}$ denotes all possible outcomes $\bar{s}$ for a sampling procedure, our sample s (Definition 2.2) can be considered as a function on $\bar{S}$. There may be a number of points $\bar{s} \varepsilon \bar{S}$ such that

$$s(\bar{s}) = s. \tag{23}$$

Of course the sampling procedure would also uniquely define the probabilities $q(\bar{s})$ of the outcome of $\bar{s}$ for all $\bar{s} \varepsilon \bar{S}$. Hence we can denote a sampling procedure by $(\bar{S}, q)$. And if the resulting sampling design is (S, p) we have

$$\sum_{\bar{s}:s(\bar{s})=s} q(\bar{s}) = p(s). \tag{24}$$

Thus by now it should be clear that given a sampling procedure, generally we can construct estimates $e(\bar{s}, x)$, which are *not* special cases of the estimate $e(s, x)$ in Definition 2.6. Again we may have, for example, one estimate $e(\bar{s}, x)$ given by Rao, Hartley and Cochran (1962), Equation (1). But this does *not* at all affect the generality of our results in the earlier or later sections for the following considerations. Let the units of a sample s be denoted by $u_1^s, \cdots, u_{n(s)}^s$. Now the total outcome *including* the variate values $x(u)$, of a sampling procedure can be denoted by $(\bar{s}, x(u_1^s), \cdots, x(u_{n(s)}^s))$, as in (23), $s(\bar{s}) = s$ and $n(s)$ as in Definition 2.4. Considering $\mathfrak{X}$ as the parameter space we may denote by

$$\text{Prob } (\bar{s}, x(u_1^s), \cdots, x(u_{n(s)}^s) \mid x'), \tag{25}$$

the corresponding probability for any specified point $x \varepsilon \mathfrak{X}$. Then for all $\bar{s}$ and s satisfying (23) we have

(26) $$\text{Prob}\,(\bar{s}, x(u_1^s), \cdots, x(u_{n(s)}^s) \mid x') = q(\bar{s}) \quad \text{if } x'(u) = x(u),\ u \varepsilon s$$
$$= 0 \quad \text{otherwise.}$$

Thus (26) clearly expresses that the probability of the outcome $\bar{s}$, $x(u_1^s), \cdots, x(u_{n(s)}^s)$ depends on x', the unknown parameter, only through s. Hence by definition of a *sufficient statistic* we can say that $(s, x(u_1^s), \cdots, x(u_{n(s)}^s))$ *is sufficient for* $\mathfrak{X}$. It then follows that if $g(\bar{s}, x(u_1^s), \cdots, x(u_{n(s)}^s))$ is an unbiased estimate of $G(x)$, a function defined on $\mathfrak{X}$, we can construct by Rao-Blackwellization of g another unbiased estimate $e(s, x(u_1^s), \cdots, x(u_{n(s)}^s))$ of $G(x)$ such that the variance of e is smaller than or equal to that of g everywhere in $\mathfrak{X}$. This shows that our Definition 2.6 of an estimate, does *not* make any of our results less general.

6. Bayes approach. The admissibility of the estimate $\bar{e}$ in (9) being established for the class D in (6) of all the unbiased estimates for the population total, it may be of interest to see if $\bar{e}$ is a Bayes solution in D, with respect to some prior distribution α on $\mathfrak{X}$. This has its own importance for people who think that all our prior knowledge, in the present case about the population or the variate under study, can be formulated in some sort of a prior distribution α. But it is necessary to note that our Bayes approach is *partial* in the sense that *we have preserved the criteria of unbiasedness* in (5). In fact as would be clear later, our investigation in terms of prior distributions on $\mathfrak{X}$ is for choosing between different sampling designs and between different estimates. But our ultimate inference about the variate x would exclusively depend on the observed sample s and the variate values $x(u)$ for $u \varepsilon s$. Especially *the inference would be independent of the assumption of any prior distribution.* This approach was first adopted by one of the authors (1955) and later on by Hájek (1959) and Aggarwal (1959).

Our criteria of judgement in terms of a prior distribution would exclusively be the expected variance of an estimate as defined in (30), *to follow.*

Now with respect to a given prior distribution α on $\mathfrak{X}$ we define

(27) $$\mathcal{E}(x(u)) = \int_{\mathfrak{X}} x(u)\,d\alpha,$$

(28) $$\sigma^2(x(u)) = \int_{\mathfrak{X}} (x(u) - \mathcal{E}(x(u)))^2\,d\alpha.$$

Similarly

(29) $\mathcal{E}(e(s, x))$, Var $(e(s, x))$, Cov $(e(s, x), h(s, x))$, etc. would denote the corresponding expectation, variance and covariance, etc. wrt α, *s being held fixed.*

And if as in (7), $V(e, x)$ denotes the variance of an estimate e, the *expected variance* of e wrt α is

(30) $$\mathcal{E}V(e) = \int_{\mathfrak{X}} V(e, x)\,d\alpha.$$

THEOREM 6.1. *For every prior distribution α on $\mathfrak{X}$ such that*

(31) *the variates $x(u)$, $u \varepsilon U$ are distributed independently* wrt α

and any sampling design $d = (S, p)$ *for which the inclusion probabilities* (*Definition* 2.5)

$$\pi(u) > 0 \quad \text{for all } u \in U, \tag{32}$$

the expected variance (30),

$$\mathcal{E}V(e) \geqq \sum_{u \in U} \sigma^2(x(u))/\pi(u) - \sum_{u \in U} \sigma^2(x(u)) \tag{33}$$

for every unbiased estimate e i.e. $e \in D$ *in* (6), $\sigma^2(x(u))$ *defined as in* (28).

Earlier the inequality (33) was established for the class of all linear (as in (4)) unbiased estimates under the same restrictions as (31), and (32) by one of the authors (1955). The following proof generalizes it to the entire class D, in (6), of unbiased estimates.

Proof of Theorem 6.1. Let, as in (9),

$$\bar{e}(s, x) = \sum_{u \in s} x(u)/\pi(u). \tag{34}$$

Then as in (10) we can express any other estimate

$$e(s, x) = \bar{e}(s, x) + h(s, x). \tag{35}$$

Then using the notation (29), we have from (35)

$$\operatorname{Var} e(s, x) = \operatorname{Var} \bar{e}(s, x) + \operatorname{Var} h(s, x) + 2 \operatorname{Cov} (\bar{e}(s, x), h(s, x)). \tag{36}$$

Multiplying (36) by $p(s)$ and summing over S, we get

$$\begin{aligned} \sum_{s \in S} p(s) \operatorname{Var} e(s, x) = \sum_{s \in S} p(s) \operatorname{Var} \bar{e}(s, x) &+ \sum_{s \in S} p(s) \operatorname{Var} h(s, x) \\ &+ 2 \sum_{s \in S} p(s) \operatorname{Cov} (\bar{e}(s, x), h(s, x)) \end{aligned} \tag{37}$$

To show that the last term in the right hand side of (37) vanishes, we have

$$\begin{aligned} \operatorname{Cov} (\bar{e}(s, x), h(s, x)) &= \mathcal{E}[(\bar{e}(s, x) - \mathcal{E}(\bar{e}(s, x)))(h(s, x) - \mathcal{E}(h(s, x)))] \\ &= \mathcal{E}[(\bar{e}(s, x) - \mathcal{E}(\bar{e}(s, x)))h(s, x)] \\ &= \sum_{u \in s} \mathcal{E}\{[(x(u) - \mathcal{E}(x(u)))/\pi(u)]h(s, x)\} \end{aligned} \tag{38}$$

due to (34). Multiplying (38) by $p(s)$ and summing, we get,

$$\begin{aligned} &\sum_{s \in S} p(s) \operatorname{Cov} (\bar{e}(s, x), h(s, x)) \\ &\qquad = \sum_{u \in U} \mathcal{E}\{[(x(u) - \mathcal{E}(x(u)))/\pi(u)] \sum_{s \ni u} p(s)h(s, x)\}, \end{aligned} \tag{39}$$

$s \ni u$ denoting all s which include the unit u. Now if estimate e in (35) is unbiased, i.e. if $e \in D$ in (6), then from (34), (35) we have

$$\sum_{s \in S} p(s)h(s, x) = 0 \tag{40}$$

for all $x \in \mathfrak{X}$. That is

$$\sum_{s \ni u} p(s)h(s, x) = -\sum_{s \not\ni u} p(s)h(s, x), \tag{41}$$

$s \not\ni u$ denoting all samples s which do *not* include the unit u. It follows from (31) and (41) that the two factors $(x(u) - \mathcal{E}(x(u)))/\pi(u)$ and $\sum_{s \ni u} p(s)h(s, x)$ are distributed independently wrt the prior distribution α. Hence we have in (39),

$$(42) \quad \mathcal{E}\{[(x(u) - \mathcal{E}(x(u)))/\pi(u)] \textstyle\sum_{s \ni u} p(s)h(s, x)\} = \mathcal{E}[(x(u) - \mathcal{E}(x(u)))/\pi(u)]\mathcal{E}(\textstyle\sum_{s \ni u} p(s)h(s, x)) = 0.$$

Thus from (39) and (42) we have

$$(43) \quad \textstyle\sum_{s \varepsilon S} p(s) \operatorname{Cov}(\bar{e}(s, x), h(s, x)) = 0.$$

And from (37) and (43) we have

$$(44) \quad \textstyle\sum_{s \varepsilon S} p(s) \operatorname{Var} e(s, x) \geqq \textstyle\sum_{s \varepsilon S} p(s) \operatorname{Var} \bar{e}(s, x).$$

Using (28) and (34) we have from (44),

$$(45) \quad \textstyle\sum_{s \varepsilon S} p(s) \operatorname{Var} e(s, x) \geqq \textstyle\sum_{u \varepsilon U} \sigma^2(x(u))/\pi(u).$$

Now for the estimate e the expected variance (30)

$$\begin{aligned} \mathcal{E}V(e) &= \mathcal{E}(\textstyle\sum_{s \varepsilon S} p(s)(e(s, x) - T(x))^2) \\ (46) \qquad &= \textstyle\sum_{s \varepsilon S} p(s) \operatorname{Var} e(s, x) + \textstyle\sum_{s \varepsilon S} p(s)[\mathcal{E}(e(s, x)) - \mathcal{E}(T(x))]^2 \\ &\qquad - \textstyle\sum_{u \varepsilon U} \sigma^2(x(u)). \end{aligned}$$

(45) and (46) give the required result

$$\mathcal{E}V(e) \geqq \textstyle\sum_{u \varepsilon U} \sigma^2(x(u))/\pi(u) - \textstyle\sum_{u \varepsilon U} \sigma^2(x(u)).$$

Actually (46) enables us to prove the more interesting

THEOREM 6.2. *For any sampling design d, for which in addition to* (32), *the sample size* (*Definition* 2.4)

$$(47) \quad (n(s) \neq n) \to p(s) = 0$$

(i.e. *d is a fixed sample size* ($= n$) *design*), *and for any prior distribution α which in addition to* (31) *has*

$$(48) \quad \mathcal{E}(x(u)) = [\mathcal{E}(T(x))/n]\pi(u) \qquad \textit{for all } u \varepsilon U,$$

(i.e. $\mathcal{E}(x(u))$ *proportional to* $\pi(u)$),

$$(49) \quad \mathcal{E}V(e) \geqq \mathcal{E}V(\bar{e})$$

for any unbiased estimate e, i.e. $e \varepsilon D$ *in* (6), *$\bar{e}$ being given by* (34).

That is for any sampling design d satisfying (32) and (47), the estimate $\bar{e}$ in (34), is a *Bayes solution*, in the class of unbiased estimates D in (6), wrt a prior distribution α, satisfying (31) and (48).

PROOF. Due to (47) and (48), the second term in the right hand side of (46) vanishes, for the estimate $\bar{e}$ in (34). Hence we have from (46),

$$(50) \quad \mathcal{E}V(\bar{e}) = \textstyle\sum_{u \varepsilon U} \sigma^2(x(u))/\pi(u) - \textstyle\sum_{u \varepsilon U} \sigma^2(x(u)).$$

The Theorem 6.2 follows from (50) and (33).

The practical significance of Theorem 6.2 is as follows: Often our prior knowledge about a population would be such that we may know the expected values $\mathcal{E}(x(u))$ for different units $u \varepsilon U$, and condition (31) is satisfied. Under these circumstances the Theorem 6.2 states that for a sampling design d satisfying (47) and having the inclusion probabilities $\pi(u)$ proportional to $\mathcal{E}(x(u))$, the estimate $\bar{e}$ in (34) is optimum (unbiased Bayes). If in addition we assume $\sigma^2(x(u))$ to be proportional to $\mathcal{E}^2(x(u))$, $u \varepsilon U$, the right hand side of (33) is minimized for fixed sample size ($n = \sum_{u\varepsilon U} \pi(u)$) designs, when $\pi(u)$ is proportional to $\mathcal{E}(x(u))$, $u \varepsilon U$, as shown by Godambe (1955). Hence the optimality of sampling design d, satisfying (47), (48) and the estimate $\bar{e}$. However, concerning this question of sampling designs a few more comments would be found in the next section.

7. Fixed sample size designs. We have already seen that the estimate $\bar{e}$ in (9) has some very desirable properties, such as admissibility or Bayesness. Now admissibility of (9) is valid regardless of the condition (47) of fixed sample size. However Bayesness of (9) is proved only for fixed sample designs. Following is an additional argument in favour of fixed sample size designs in connection with the estimate $\bar{e}$ in (9). The variance of $\bar{e}$,

$$(51)\quad V(\bar{e}, x) = \sum_{u\varepsilon U} x^2(u)/\pi(u) + \sum_{u,u'\varepsilon U, u\neq u'} [\pi(u, u')/\pi(u)\pi(u')]x(u)x(u') - T^2(x),$$

$\pi(u, u')$ being the same as in (21).

THEOREM 7.1. *For any given fixed sample size design d (i.e. d satisfies (47)) with inclusion probabilities (Definition 2.5) equal to $\pi(u)$, $u \varepsilon U$, it is impossible to construct a varying sample size design d^* (i.e. d^* does not satisfy (47)) with the same inclusion probabilities $\pi(u)$, $u \varepsilon U$, such that the variance in (51) for d^* is smaller than or equal to that for d, for all $x \varepsilon \mathfrak{X}$.*

PROOF. For a sampling design with inclusion probabilities $\pi(u)$, $u \varepsilon U$, the average sample size $E(n(s))$ as in (3) is given by $E(n(s)) = \sum_{u\varepsilon U} \pi(u)$. Let further Var $n(s)$ denote the variance of the sample size $n(s)$, for the sampling design d^*. Then we have due to Hanurav (1962), the equation

$$(52)\quad \sum_{u,u'\varepsilon U, u\neq u'} \pi^*(u, u') = E^2(n(s)) - E(n(s)) + \text{Var } n(s),$$

$\pi^*(u, u')$ being the probabilities for d^*, corresponding to $\pi(u, u')$ in (51). Now the Theorem 7.1 is true due to the fact that if it was false, putting in (51), x: $x(u) = \pi(u)$, $u \varepsilon U$, we would get, for the fixed sample size design d, $\sum_{u,u'\varepsilon U, u\neq u'} \pi(u, u') \geqq \sum_{u,u'\varepsilon U, u\neq u'} \pi^*(u, u')$, contradicting (52).

Hanurav (1962) has given a different justification for fixed sample size designs, in terms of some prior distributions.

8. On relaxing the criteria of unbiasedness. Now as said in the Section 3, in the field of sample surveys the criteria of unbiasedness as defined in (5) is very appealing. And nearly always it is taken for granted. Actually without this criteria of unbiasedness, mere considerations of admissibility or minimaxity may lead us astray. For instance it is easy to see that every 'constant' is an admissible minimax estimate, of the population total, over the entire $\mathfrak{X}$. One can also construct examples to show that generally the estimate $\bar{e}$ in (9), which we have been primarily investigating, is *not* minimax over intervals $\mathfrak{X}$, x: $\alpha_1(u) \leqq x(u) \leqq \alpha_2(u)$, $u \varepsilon U$, which are of practical importance as said in Remark 4.1. Consider the following simple though artificial illustration: The population $U = \{u_1, u_2\}$ consists of two units. Let the sample $s_1 = (u_1)$ and $s_2 = (u_2)$. Assume the sampling design $d = (S, p)$ to be such that $p(s_1) = p(s_2) = \frac{1}{2}$. Now for the subset of $\mathfrak{X}$ given by x: $|x(u_1)| \leqq C$, $|x(u_2)| \leqq C$, the maximum variance of the estimate $\bar{e}$ in (9) is $4C^2$ while the maximum value of $\sum_{s \varepsilon S} (e(s, x) - T(x))^2 p(s)$ for the estimate e given by $e(s_1, u_1) = x(u_1)$, $e(s_2, u_2) = x(u_2)$, is C^2. Here clearly $\bar{e}$ is not minimax. All these comments are relevent in connection with Aggarwal's work (1959) attempting to justify the estimate (9), without appealing to the concept of unbiasedness. Finally in this connection we prove the following Theorem 8.1 relating to all linear estimates (as in (4)) $e_1(s, x) = \sum_{u \varepsilon s} \beta(s, u) x(u)$ such that for any given unit u, $u \varepsilon U$, $\beta(s, u)$ is the same for all samples s which include u, so that the estimate $e_1(s, x)$ can be written as

$$e_1(s, x) = \sum_{u \varepsilon s} b(u) x(u), \tag{53}$$

and where further $b(u)$, $u \varepsilon U$ satisfy

$$b(u) \geqq 1,\ u \varepsilon U \quad \text{and} \quad \sum_{u \varepsilon U} (1/b(u)) = n. \tag{54}$$

THEOREM 8.1. *For any fixed sample size design (that is one satisfying (47)), every estimate $e_1(s, x)$ in (53) which satisfies (54) with n equal to the sample size, is admissible in the class of all linear estimates of the population total $T(x)$, in the sense there does not exist any linear estimate $e(s, x)$ such that*

$$\sum_{s \varepsilon S} (e(s, x) - T(x))^2 p(s) \leqq \sum_{s \varepsilon S} (e_1(s, x) - T(x))^2 p(s) \tag{55}$$

for all $x \varepsilon \mathfrak{X}$, strict inequality holding for at least one x.

PROOF. Let the linear estimate satisfying (55) be

$$e(s, x) = \sum_{u \varepsilon s} \beta(s, u) x(u). \tag{56}$$

Now $h(s, x)$ is defined from (53) and (56) by

$$e(s, x) = e_1(s, x) + h(s, x), \tag{57}$$

so that

$$\begin{aligned} h(s, x) &= \sum_{u \varepsilon s} (\beta(s, u) - b(u)) x(u) \\ &= \sum_{u \varepsilon s} \alpha(s, u) x(u), \end{aligned} \tag{58}$$

where

$$\alpha(s, u) = \beta(s, u) - b(u), \qquad s \varepsilon S,\ u \varepsilon s. \tag{59}$$

On substituting (57) in (55) and simplifying we get

(60) $\sum_{s\varepsilon S} p(s)h^2(s, x) + \sum_{s\varepsilon S} p(s)h(s, x)(e_1(s, x) - T(x)) \leqq 0.$

As the left hand side of (60) is a semi-negative definite quadratic form, the coefficient in it of $x^2(u)$, $u \varepsilon U$, must be non-positive. Hence using (58), we get

(61) $$\sum_{s \ni u} p(s)\alpha^2(s, u) + \sum_{s \ni u} p(s)\alpha(s, u)(b(u) - 1) \leqq 0,$$

where $s \ni u$ denotes all samples s, which include the unit u, $u \varepsilon U$. Now let,

(62) $$\sum_{s \ni u} p(s)\alpha(s, u) = \delta(u).$$

From (61) and (62), we get,

(63) $$\sum_{s \ni u} p(s)(\alpha(s, u) - \delta(u))^2 + \pi(u)\delta^2(u) + \delta(u)(b(u) - 1) \leqq 0,$$

where $\pi(u) = \sum_{s \ni u} p(s)$ as defined in (2). Now from (54) and (63) we have

(64) $$\delta(u) \leqq 0, \quad \text{all } u \varepsilon U.$$

Next for $\bar{x} \varepsilon \mathfrak{X}$ given by $\bar{x}(u) = 1/b(u)$, $u \varepsilon U$ in (53),

(65) $$e_1(s, \bar{x}) = \sum_{u\varepsilon s} b(u)(1/b(u)) = n(s) = n$$

for $p(s) \neq 0$ by (47). Again by (54)

(66) $$T(\bar{x}) = n.$$

Now substituting (65) and (66) in (55) we have from (57) for all samples s with $p(s) \neq 0$,

(67) $$h(s, \bar{x}) = 0.$$

Since from (58) and (62),

$$\sum_{s\varepsilon S} p(s)h(s, \bar{x}) = \sum_{u\varepsilon U} \delta(u)\bar{x}(u) = \sum_{u\varepsilon U} \delta(u)/b(u),$$

it follows from (67) that

(68) $$\sum_{u\varepsilon U} \delta(u)/b(u) = 0.$$

Now noting from (54) that $b(u) > 0$, $u \varepsilon U$, we have from (64) and (68), $\delta(u) = 0$, $u \varepsilon U$, which from (63) means for all samples s for which $p(s) \neq 0$, $\alpha(s, u) = 0$, $u \varepsilon U$, so that from (58), $h(s, x) = 0$, which further with (57) and (59) gives

(69) $$e(s, x) = e_1(s, x),$$

for all s with $p(s) \neq 0$. As (69) implies that strict inequality in (55) cannot be satisfied for any $x \varepsilon \mathfrak{X}$, the Theorem 8.1 is proved.

Since $\bar{e}(s, x)$ in (9) is a special case of $e_1(s, x)$ in (53) we have from Theorem 8.1 the

COROLLARY 8.1. *For any fixed sample size design the estimate $\bar{e}(s, x)$ in* (9) *is admissible in the class of all linear estimates of the population total $T(x)$.*

9. Necessity of the condition of fixed sample size in the Theorem 8.1. If the sampling design does not satisfy the condition of fixed sample size namely (47), the estimate $e_1(s, x)$ in Theorem 8.1 is inadmissible excepting trivial cases. This we demonstrate in the case of the practically interesting estimate $\bar{e}(s, x)$ in (9) by constructing a uniformly superior linear estimate. Put

(70) $$e'(s, x) = (1 - k)\bar{e}(s, x),$$

where k is a constant such that

(71) $$0 < k < 1.$$

From (70)

$$\textstyle\sum_{s \varepsilon S} p(s)(e'(s, x) - T(x))^2 < \sum_{s \varepsilon S} p(s)(\bar{e}(s, x) - T(x))^2$$

for all $x \varepsilon \mathfrak{X}$, if for all $x \varepsilon \mathfrak{X}$,

$$k^2 \textstyle\sum_{s \varepsilon S} p(s)\bar{e}^2(s, x) - 2k(\sum_{s \varepsilon S} p(s)\bar{e}^2(s, x) - T^2(x)) < 0,$$

which since $k > 0$, implies $k \leqq 2(1 - T^2(x)/\sum_{s \varepsilon S} p(s)\bar{e}^2(s, x))$, that is

(72) $$k < 2V(\bar{e}(s, x))/[T^2(x) + V(\bar{e}(s, x))]$$

where $V(\bar{e}(s, x))$ is the variance as defined in (7). If the sampling design is one with fixed sample size (i.e. (47) holds) then the minimum value of (72) for the variation of x in $\mathfrak{X}$ is zero, as $V(\bar{e}(s, \bar{x})) = 0$ for $\bar{x}$ given by $\bar{x}(u) = \pi(u)$, $u \varepsilon U$ as defined in (2). Hence no k satisfying (71) and (72) exists. However if sampling design is one with varying sample size, $V(\bar{e}(s, x))$ would not vanish for any $x \varepsilon \mathfrak{X}$, except in the trivial cases when the linear simultaneous equations $\bar{e}(s, x) = T(x)$ for all $s \varepsilon S$ with $p(s) \neq 0$, reduce to N (the number of units in U) or less independent equations. Excluding such cases the right hand side of (72) will have a minimum $K_0 > 0$. Hence assigning to k in (70) a value such that $0 < k <$ minimum of $(K_0, 1)$, we get a linear estimate $e'(s, x)$ which is uniformly superior to $\bar{e}(s, x)$.

Illustration. We take a simple though rather artificial example. Let the population consist of 3 units u_1, u_2, u_3, and let the sampling design assign a probability $\frac{1}{6}$ to each of the samples (u_1), (u_2), (u_3), (u_1, u_2), (u_2, u_3) and (u_3, u_1). Then denoting $x(u_r)$ by x_r, $r = 1, 2, 3$,

(73) $$\textstyle\sum_{s \varepsilon S} p(s)\bar{e}^2(s, x) = 2(x_1^2 + x_2^2 + x_3^2) + \frac{4}{3}(x_1x_2 + x_2x_3 + x_3x_1).$$

We minimize the right hand side of (73) subject to a fixed value of $T(x)$ say T. By using Lagrange's multipliers this minimum value is easily found to be $\frac{10}{9}T^2$. Hence clearly K_0, the minimum value of right hand side of (72) $= 2(1 - \frac{9}{10}) = \frac{1}{5}$. Hence taking $k = \frac{1}{10}$ say we get from (70) the estimate $e'(s, x) = \frac{9}{10}\bar{e}(s, x)$, which is uniformly superior to $\bar{e}(s, x)$.

Acknowledgment. The authors are indebted to Dr. Kale, Dr. Patil, and Dr. J. Hájek for many helpful suggestions, to the referee for a thorough going scrutiny of this paper and many helpful suggestions, and to Miss Patkar for doing the necessary typing very carefully.

REFERENCES

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

DAS, A. C. (1962). On MVU estimates of parameters of finite population. *Calcutta Statist. Assoc. Bull.* **2.**

GODAMBE, V. P. (1955). A unified theory of sampling from finite populations. *J. Roy. Statist. Soc. Ser. B* **17** 268–278.

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

HÁJEK, J. (1959). Optimum strategy and other problems in probability sampling. *Časopis Pest Mat.* **84** 387–423.

HANURAV (1962). On Horvitz and Thomson estimates. *Sankhyā* **24** 421–428.

HORVITZ, D. G. and THOMSON, D. J. (1952). A generalization of sampling without replacement from a finite universe. *J. Amer. Statist. Assoc.* **47** 663–685.

KOOP, J. C. (1957). Contributions to the general theory of sampling finite populations without replacement and with unequal probabilities. Mimeograph Series 296, Institute of Statistics, Univ. of North Carolina.

PRABHU AJGAONKAR, S. G. (1962). Thesis submitted to Karnâtak University for the Ph.D. degree.

RAO, J. N. K., HARTLEY, H. O., and COCHRAN, W. G. (1962). On a sample procedure of unequal probability sampling without replacement. *J. Roy. Statist. Soc. Ser. B* **24** 482–591.

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

YATES, F. and GRUNDY, P. M. (1953). Selection without replacement from within strata with probability proportional to size. *J. Roy. Statist. Soc. Ser. B* **15** 253–261.