The Annals of Statistics
1974, Vol. 2, No. 1, 205–215

HYPERADMISSIBILITY IN SURVEY SAMPLING

BY JAN LANKE AND M. K. RAMAKRISHNAN

University of Lund, Sweden and University of Sheffield

Two different definitions of hyperadmissibility are considered. The relation between hyperadmissibility and admissibility is discussed. Special attention is given to the hyperadmissibility of the Horvitz–Thompson estimator.

1. Introduction. It is well known that most sampling designs do not admit uniformly minimal variance estimators in reasonable classes of unbiased estimators of a finite population total. As a result several other optimality criteria have been proposed in the literature. In this paper we deal with one of these, viz. the criterion of hyperadmissibility due to Hanurav. Two slightly different definitions of hyperadmissibility have been proposed. In Section 3 we discuss these two definitions, their interrelation and their relations to the definition of admissibility. Section 4 is devoted to a detailed study of hyperadmissibility in one of the two senses and Section 5 to that of the other.

2. Notations and definitions. Let U denote the finite population consisting of the units $1, 2, \cdots, N$. A subset s of U is called a sample and a nonnegative function p on the set S of all possible samples is called a (sampling) design if $\sum_s p(s) = 1$. For a given design p, define $\pi_i = \sum_{s \ni i} p(s)$. A design p is said to be a unicluster design if

$$s_1 \in \bar{S},\ s_2 \in \bar{S},\ s_1 \neq s_2 \Rightarrow s_1 \cap s_2 = \emptyset$$

where $\bar{S}$ denotes the set of samples s for which $p(s) > 0$. For a design p, we denote by $\bar{S}_i$ the set of samples s in $\bar{S}$ such that $i \in s$ and by $\bar{S}_i^c$ the complement of $\bar{S}_i$ with respect to $\bar{S}$.

With each unit i is associated a variate value y_i and $\mathbf{y} = (y_1, y_2, \cdots, y_N)$ is considered as a point in the N-dimensional Euclidean space R_N which thus acts as parameter space. Any real-valued function on R_N is called a parametric function and the particular function T given by $T = \sum_1^N y_i$ is called the population total. The conventional problem in survey sampling is to estimate T by observing the values of those y_i for which $i \in s$ where s is a sample drawn according to a design p. An estimator e is a real-valued function on $S \times R_N$ such that $e(s, \mathbf{y})$ depends upon $\mathbf{y}$ only through those y_i for which $i \in s$, that is, $e(s, \mathbf{y}) = e(s, \mathbf{y}')$ for any two $\mathbf{y}$ and $\mathbf{y}'$ such that $y_i = y_i'$ for all $i \in s$. An estimator e is said to be a polynomial, to vanish at the origin, $\cdots$ if, for each $s \in S$, the function $\mathbf{y} \to e(s, \mathbf{y})$ is a polynomial, vanishes at the origin, $\cdots$. The expectation and the variance of an estimator e in a design p are denoted by $E(e, \mathbf{y})$

Received April 1972; revised April 1973.
Key words and phrases. Hyperadmissibility, survey sampling, Horvitz–Thompson.

and $V(e, \mathbf{y})$, respectively. For any design p for which $\pi_i > 0$ for $i = 1, 2, \cdots, N$, the Horvitz–Thompson estimator of T is defined by $\bar{e}(s, \mathbf{y}) = \sum_{i \in s} y_i/\pi_i$.

Given a class C of unbiased estimators of a certain parametric function, an estimator $e_0 \in C$ is said to be admissible in C if for no other estimator $e \in C$ we have $V(e, \mathbf{y}) \leqq V(e_0, \mathbf{y})$ for all $\mathbf{y} \in R_N$ with strict inequality for at least one $\mathbf{y} \in R_N$.

The principal hypersurface (phs, for short) $R(i_1, i_2, \cdots, i_k)$ is defined as the set of points $\mathbf{y} \in R_N$ such that $y_j = 0$ for $j \neq i_1, i_2, \cdots, i_k$ where $\{i_1, i_2, \cdots, i_k\}$ is an arbitrary non-empty subset of U. Clearly the whole space R_N can be regarded as a phs by taking $k = N$ and there are in all $2^N - 1$ phs's in R_N. The interior $R^0(i_1, i_2, \cdots, i_k)$ of $R(i_1, i_2, \cdots, i_k)$ is the set of all $\mathbf{y} \in R(i_1, i_2, \cdots, i_k)$ which do not lie in any phs of lower dimension.

3. Various definitions of hyperadmissibility. The first mention of hyperadmissibility occurs in Hanurav (1965) where the following definition is given.

DEFINITION 3.1. An estimator e in a class C of unbiased estimators of a certain parametric function is said to be hyperadmissible in C if for any principal hypersurface $R(i_1, i_2 \cdots, i_k)$ in R_N, the estimator e is admissible in C when the parameter $\mathbf{y}$ is restricted to $R(i_1, i_2, \cdots, i_k)$.

In Hanurav (1968) the definition has been changed into

DEFINITION 3.2. An estimator e in a class C of unbiased estimators of a certain parametric function is said to be hyperadmissible in C if for any principal hypersurface $R(i_1, i_2, \cdots, i_k)$ in R_N, the estimator e is admissible in C when the parameter $\mathbf{y}$ is restricted to the interior $R^0(i_1, i_2, \cdots, i_k)$ of $R(i_1, i_2, \cdots, i_k)$.

In fact, the two definitions above do not exactly reproduce what Hanurav has written but rather what Hanurav probably intended to write; a discussion on this point is given in Ramakrishnan (1970, page 78) and in Basu (1971, page 228).

It is not clear why Hanurav changed his original definition, especially since he considers only polynomial estimators in which case the two definitions are equivalent; cf. statement (iii) in Theorem 3.3 below.

As for the rationale behind hyperadmissibility, Section 4 of Hanurav (1968) gives a very convincing argument in support of this concept. However, the reader can easily check that Hanurav's argument in fact justifies hyperadmissibility in the sense of Definition 3.1 rather than in the sense of Definition 3.2.

At first sight hyperadmissibility in the sense of Definition 3.2 may appear to be a stronger property than hyperadmissibility in the sense of Definition 3.1. As will be shown below, this impression is, however, false; one should remember the peculiar role which the point $\mathbf{0}$ plays in Definition 3.2: $\mathbf{0}$ does not belong to any $R^0(i_1, i_2, \cdots, i_k)$ and the definition thus imposes no condition at all upon the variance at the point $\mathbf{0}$.

The following theorem summarizes some relations between the two concepts of hyperadmissibility.

THEOREM 3.3. (i) *Hyperadmissibility in the sense of Definition* 3.2 *does not imply hyperadmissibility in the sense of Definition* 3.1.

(ii) *If all estimators in* C *vanish at the origin, hyperadmissibility in the sense of Definition* 3.2 *implies hyperadmissibility in the sense of Definition* 3.1.

(iii) *If the variance* $V(e, \mathbf{y})$ *of every estimator* e *in* C *is continuous, the two concepts are equivalent. In particular, if all estimators in* C *are continuous, the two concepts are equivalent.*

PROOF. For the proof of (i) we ask the reader to look at Example 5.2 in Joshi (1971). As stated by Joshi, the estimator given in this example is hyperadmissible (in the sense of our Definition 3.2) in the class A^* of all unbiased estimators of the population total—although the present authors do not entirely agree that this fact is *easily* verified. That the estimator in question is not hyperadmissible in the sense of Definition 3.1 follows from Theorem 4.4 (to come): the design is obviously not a unicluster design and thus the Horvitz–Thompson estimator is the only hyperadmissible (Definition 3.1) estimator in A^*. To prove statement (ii), suppose e is hyperadmissible in some class C in the sense of Definition 3.2 and suppose that for a certain $e' \in C$ we have $V(e', \mathbf{y}) \leqq V(e, \mathbf{y})$ for all $\mathbf{y}$ in a certain phs R. Then we have $V(e', \mathbf{y}) \leqq V(e, \mathbf{y})$ in every open phs $R^0 \subseteqq R$ and hence $V(e', \mathbf{y}) = V(e, \mathbf{y})$ for every $\mathbf{y}$ which belongs to some open phs in R, i.e., for $\mathbf{0} \neq \mathbf{y} \in R$. But since all estimators in C have zero variance at $\mathbf{y} = \mathbf{0}$, we have $V(e', \mathbf{y}) = V(e, \mathbf{y})$ for $\mathbf{y} = \mathbf{0}$ as well and the hyperadmissibility of e in the sense of Definition 3.1 follows. Finally, the statement (iii) is an immediate consequence of the fact that any weak inequality between two functions in an open phs holds in the corresponding full phs as well if the functions in question are continuous. □

The prefix in "hyperadmissibility" suggests that hyperadmissibility is a stronger property than admissibility. In case we use Definition 3.1 it is obvious, putting $k = N$ in the definition, that this suggestion is true. Using Definition 3.2 and defining "sample" as "sequence," Hanurav (1968) states that hyperadmissibility implies admissibility. This is, however, incorrect, and the following example illustrates that in the setup used by Hanurav (1968) and by Joshi (1971), hyperadmissibility does *not* imply admissibility.

EXAMPLE 3.4. In this example (but nowhere else) we follow Hanurav and Joshi in defining a sample as a *sequence* $(i_1, i_2, \cdots, i_n)$ of units from U. Consider $U = \{1, 2\}$ and the design p which gives probability $\frac{1}{4}$ to each of the four samples $s_1 = (1)$, $s_2 = (2)$, $s_3 = (1, 2)$, $s_4 = (2, 1)$ and let e be the estimator

$$\begin{aligned} e(s, \mathbf{y}) &= \bar{e}(s, \mathbf{y}) && \text{for} \quad s = s_1, s_2 \\ &= \bar{e}(s, \mathbf{y}) + f(\mathbf{y}) && \text{for} \quad s = s_3 \\ &= \bar{e}(s, \mathbf{y}) - f(\mathbf{y}) && \text{for} \quad s = s_4 \end{aligned}$$

where $\bar{e}$ is the Horvitz–Thompson estimator and where

$$f(\mathbf{y}) = 0 \quad \text{for} \quad \mathbf{y} \neq \mathbf{0}$$
$$= 1 \quad \text{for} \quad \mathbf{y} = \mathbf{0}\,.$$

This estimator is easily seen to be unbiased for the population total. Further we have

$$E(e^2 - \bar{e}^2, \mathbf{y}) = 0 \quad \text{for} \quad \mathbf{y} \neq \mathbf{0}$$
$$> 0 \quad \text{for} \quad \mathbf{y} = \mathbf{0}$$

whence e is hyperadmissible (Definition 3.2) in the class $C = \{e, \bar{e}\}$ but not admissible in C. To justify part of Remark 3.8 (to come), observe that by virtue of Theorem 5.1, the same conclusions hold if we replace the class $\{e, \bar{e}\}$ by the class A^* of all unbiased estimators of the population total.

In view of the above example, it would be of interest to seek conditions under which hyperadmissibility in the sense of Definition 3.2 implies admissibility. The following two theorems furnish such conditions.

THEOREM 3.5. *If the design p gives zero probability to the sample consisting of the whole population and if the estimator e is hyperadmissible* (Definition 3.2) *in a class C of unbiased estimators of the population total, then e is admissible in C.*

THEOREM 3.6. *If the estimator e is hyperadmissible* (Definition 3.2) *in a class C of unbiased estimators of the population total, and if C is convex* (i.e., *contains* $(e_1 + e_2)/2$ *whenever it contains e_1 and e_2*), *then e is admissible in C.*

LEMMA 3.7. *Let $F(\mathbf{y}) = \sum_1^N f_i(\mathbf{y})$ where $f_i(\mathbf{y})$ is independent of y_i for $i = 1, 2, \cdots, N$. If $F(\mathbf{y}) = 0$ for all $\mathbf{y} \neq \mathbf{0}$, then $F(\mathbf{0}) = 0$.*

PROOF. Let $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_N)$ where each ε_i is 0 or 1 and put $|\boldsymbol{\varepsilon}| = \sum_1^N \varepsilon_i$. Then it is easily seen that any function $f = f(\mathbf{y})$ which is independent of at least one y_i satisfies

$$\sum_{\boldsymbol{\varepsilon}} (-1)^{|\boldsymbol{\varepsilon}|} f(\boldsymbol{\varepsilon}) = 0 \tag{3.1}$$

where the sum is over all the 2^N possible values of $\boldsymbol{\varepsilon}$. Since each term in $\sum_1^N f_i(\mathbf{y})$ satisfies (3.1), so does the sum $F(\mathbf{y})$:

$$\sum_{\boldsymbol{\varepsilon}} (-1)^{|\boldsymbol{\varepsilon}|} F(\boldsymbol{\varepsilon}) = 0\,.$$

In this sum we know that the $2^N - 1$ terms corresponding to $\boldsymbol{\varepsilon} \neq \mathbf{0}$ are zero and thus the remaining term $F(\mathbf{0})$ must be zero as well. □

PROOF OF THEOREM 3.5. Given any $e_1 \in C$ with

$$V(e_1, \mathbf{y}) \leqq V(e, \mathbf{y}) \qquad \text{for} \quad \mathbf{y} \in R_N \tag{3.2}$$

we shall show that

$$V(e_1, \mathbf{y}) = V(e, \mathbf{y}) \qquad \text{for} \quad \mathbf{y} \in R_N\,. \tag{3.3}$$

Now let R denote any phs and R^0 its interior. Then the restriction of (3.2) to

R^0 and the hyperadmissibility of e give $V(e_1, \mathbf{y}) = V(e, \mathbf{y})$ for $\mathbf{y} \in R^0$. Thus $V(e_1, \mathbf{y}) = V(e, \mathbf{y})$ for every $\mathbf{y}$ which belongs to the interior of some phs, i.e., for $\mathbf{y} \neq \mathbf{0}$. Both e and e_1 being unbiased, we thus have

$$\sum_s p(s)(e_1^2(s, \mathbf{y}) - e^2(s, \mathbf{y})) = 0 \qquad \text{for } \mathbf{y} \neq \mathbf{0}\,. \tag{3.4}$$

Since $p(U) = 0$, every term in this sum is independent of at least one y_i and so Lemma 3.7 shows that the equality in (3.4) in fact holds for $\mathbf{y} = \mathbf{0}$ also. □

PROOF OF THEOREM 3.6. The proof runs exactly parallel to the proof of Theorem 3.5 until we reach

$$V(e_1, \mathbf{y}) = V(e, \mathbf{y}) \qquad \text{for } \mathbf{y} \neq \mathbf{0}\,. \tag{3.5}$$

Let $\rho(\mathbf{y})$ denote the correlation coefficient between e and e_1 and introduce $e_2 = (e + e_1)/2$. Since C is convex, $e_2 \in C$. Using (3.5) we get

$$V(e_2, \mathbf{y}) = \tfrac{1}{2}(1 + \rho(\mathbf{y}))V(e, \mathbf{y}) \qquad \text{for } \mathbf{y} \neq \mathbf{0}$$

and since e is hyperadmissible we have $\rho(\mathbf{y}) = 1$ for $\mathbf{y} \neq \mathbf{0}$ which means that there are two functions $A(\mathbf{y})$ ($\geqq 0$) and $B(\mathbf{y})$ such that

$$e_1(s, \mathbf{y}) = A(\mathbf{y})e(s, \mathbf{y}) + B(\mathbf{y}) \qquad \text{for } \mathbf{y} \neq \mathbf{0},\ s \in \bar{S}\,. \tag{3.6}$$

Taking variances on both sides in (3.6) gives $A^2(\mathbf{y}) = 1$ for $\mathbf{y} \neq \mathbf{0}$ and hence $A(\mathbf{y}) = 1$ for $\mathbf{y} \neq \mathbf{0}$. Taking expectations on both sides in (3.6) gives $B(\mathbf{y}) = 0$ for $\mathbf{y} \neq \mathbf{0}$:

$$e_1(s, \mathbf{y}) = e(s, \mathbf{y}) \qquad \text{for } \mathbf{y} \neq \mathbf{0},\ s \in \bar{S}\,. \tag{3.7}$$

Since $e(s, \mathbf{y})$ and $e_1(s, \mathbf{y})$ depend upon $\mathbf{y}$ only through those y_i for which $i \in s$, (3.7) implies

$$e_1(s, \mathbf{0}) = e(s, \mathbf{0}) \qquad \text{for } s \neq U,\ s \in \bar{S}\,. \tag{3.8}$$

Finally, since e and e_1 have the same expectation, (3.8) implies

$$e_1(U, \mathbf{0}) = e(U, \mathbf{0}) \tag{3.9}$$

in case $U \in \bar{S}$. Now (3.7) and (3.8) together with (3.9), if relevant, give

$$e_1(s, \mathbf{y}) = e(s, \mathbf{y}) \qquad \text{for } \mathbf{y} \in R_N,\ s \in \bar{S}$$

from which (3.3) obviously follows. □

REMARK 3.8. An examination of the proof of Theorem 3.5 reveals that we have never used the fact that "sample" = "set" and thus the theorem remains true also if we interpret "sample" as "sequence". In contrast to this, Theorem 3.6 is true only with the set-definition: the proof breaks down between equations (3.8) and (3.9) if we use the sequence-definition and Example 3.4 with $C = A^*$ shows that no other proof can be found in this case.

The reader will no doubt observe that one case is left out: if $p(U) > 0$, if C is not convex and if "sample" = "set", does then hyperadmissibility in the sense of Definition 3.2 imply admissibility? The present authors have been entirely unsucessful in answering this question.

4. Hyperadmissibility in the sense of Definition 3.1. In this section hyperadmissibility is to be understood in the sense of Definition 3.1.

It is well known that the Horvitz–Thompson estimator $\sum_{i\in s} y_i/\pi_i$, if unbiased, is admissible in the class A^* of all unbiased estimators of the population total. In view of the results of the preceding section, the following theorem is a generalization of this fact.

THEOREM 4.1. *Let p be a design such that $\pi_i > 0$ for all i. Then the Horvitz–Thompson estimator $\bar{e}$ is hyperadmissible in the class A^*.*

PROOF. Given any $e \in A^*$ such that $V(e, \mathbf{y}) \leqq V(\bar{e}, \mathbf{y})$ in a certain phs $R(i_1, i_2, \cdots, i_k)$, we shall show that $V(e, \mathbf{y}) = V(\bar{e}, \mathbf{y})$ in that phs. For the case $k = N$, the result is well known from Godambe and Joshi (1965). The essential idea in their proof is to use induction on n to show for every n that $e(s, \mathbf{y}) = \bar{e}(s, \mathbf{y})$ in that part of R_N where at most n coordinates have nonzero values. It is, however, easily seen that exactly the same idea can be used to show the corresponding result also in the case $k < N$; here the induction should be performed with respect to the number of nonzero elements among the coordinates with numbers $i_1, i_2, \cdots, i_k$. □

Now we will investigate whether the Horvitz–Thompson estimator is the only hyperadmissible estimator. Let A_0^* denote the class of estimators in A^* which vanish at the origin.

THEOREM 4.2. *If an estimator e is hyperadmissible in the class A_0^*, then $e = \bar{e}$.*

PROOF. Let $R_N^{(k)}$ denote the part of R_N where at most k coordinates are nonzero and consider the following statement:

(4.1.k) if for every phs of dimension at most k, the estimator e is admissible in A_0^* when the parameter is restricted to that phs, then $e(s, \mathbf{y}) = \bar{e}(s, \mathbf{y})$ for all $s \in \bar{S}$ and all $\mathbf{y} \in R_N^{(k)}$.

Proceeding by induction we shall show that (4.1.k) is true for $k = 1, 2, \cdots, N$; then the case $k = N$ obviously gives the theorem. First consider the case $k = 1$. It is easily seen that

$$E(e\bar{e}, \mathbf{y}) = E(\bar{e}^2, \mathbf{y}) \qquad \text{for } \mathbf{y} \in R_N^{(1)}$$

whence

$$V(e, \mathbf{y}) - V(\bar{e}, \mathbf{y}) = E((e - \bar{e})^2, \mathbf{y}) \qquad \text{for } \mathbf{y} \in R_N^{(1)}. \tag{4.2}$$

Considering the one-dimensional phs $R(i)$, we have

$$V(\bar{e}, \mathbf{y}) \leqq V(e, \mathbf{y}) \qquad \text{for } \mathbf{y} \in R(i)$$

and hence, e being admissible when $\mathbf{y}$ is restricted to $R(i)$,

$$V(e, \mathbf{y}) = V(\bar{e}, \mathbf{y}) \qquad \text{for } \mathbf{y} \in R(i)$$

which in view of (4.2) gives

$$e(s, \mathbf{y}) = \bar{e}(s, \mathbf{y}) \qquad \text{for} \quad s \in \bar{S},\ \mathbf{y} \in R(i)\,.$$

This being true for every i, (4.1.1) follows. Now suppose we know that (4.1.k-1) is true and suppose e is admissible when $\mathbf{y}$ is restricted to any phs of dimension at most k. Since $e(s, \mathbf{y})$ depends upon $\mathbf{y}$ only through those coordinates y_i for which $i \in s$, we know that

$$e(s, \mathbf{y}) = \bar{e}(s, \mathbf{y})$$

holds not only for all $\mathbf{y} \in R_N^{(k-1)}$ and all $s \in \bar{S}$ but also for all $\mathbf{y} \in R_N^{(k)}$ and those $s \in \bar{S}$ in which at least one of the units with nonzero y-value is missing. Now let $\mathbf{y}$ be any vector with k nonzero components and let $\bar{S}'$ denote the set of samples s in $\bar{S}$ which contain all the k units with nonzero y-values. Both e and $\bar{e}$ being unbiased we have

$$\textstyle\sum_{s \in \bar{S}} p(s) e(s, \mathbf{y}) = \sum_{s \in \bar{S}} p(s) \bar{e}(s, \mathbf{y})$$

and since $e(s, \mathbf{y}) = \bar{e}(s, \mathbf{y})$ for $s \notin \bar{S}'$, we have

$$\textstyle\sum_{s \in \bar{S}'} p(s) e(s, \mathbf{y}) = \sum_{s \in \bar{S}'} p(s) \bar{e}(s, \mathbf{y})\,.$$

But $\bar{e}(s, \mathbf{y})$ has the same value for all $s \in \bar{S}'$ whence

$$\textstyle\sum_{s \in \bar{S}'} p(s) e(s, \mathbf{y}) \bar{e}(s, \mathbf{y}) = \sum_{s \in \bar{S}'} p(s) (\bar{e}(s, \mathbf{y}))^2$$

and

$$E(e\bar{e}, \mathbf{y}) = E(\bar{e}^2, \mathbf{y})\,.$$

As before, this gives (4.2) for $\mathbf{y} \in R_N^{(k)}$ and the rest of the argument is identical with that in the case $k = 1$. $\square$

The following lemma is the essential step in the investigation of the uniqueness of $\bar{e}$ as hyperadmissible estimator in A^*.

LEMMA 4.3. *If the design p is not a unicluster design with exactly two clusters and if the estimator e is hyperadmissible in the class A^*, then $e \in A_0^*$.*

PROOF. Let $R(i)$ be any one-dimensional phs. Arguing as in Joshi (1971, equations 14–18) we see that the admissibility of e in A^* when $\mathbf{y}$ is restricted to $R(i)$ implies the existence of a function g_i on R_1 and a constant k_i^c such that for $\mathbf{y} \in R(i)$ we have

$$\begin{aligned} e(s, \mathbf{y}) &= g_i(y_i) \qquad &&\text{for} \quad s \in \bar{S}_i \\ &= k_i^c &&\text{for} \quad s \in \bar{S}_i^c\,. \end{aligned}$$

Here we (unlike Joshi) can put $\mathbf{y} = \mathbf{0}$ and, introducing the notations $k(s)$ and k_i for $e(s, \mathbf{0})$ and $g_i(0)$ respectively, we have

$$\begin{aligned} k(s) &= k_i \qquad &&\text{for} \quad s \in \bar{S}_i \\ &= k_i^c &&\text{for} \quad s \in \bar{S}_i^c\,. \end{aligned} \tag{4.3}$$

To show that $k(s) = 0$ for all s in $\bar{S}$, we distinguish three cases: p is a nonunicluster design, p is a unicluster design with at least three clusters and p is a unicluster design with just one cluster, i.e., a census. In the first case there exist two units, say 1 and 2, and two samples in $\bar{S}$, say s_1 and s_2, such that

$$s_1 \ni 1 , \qquad s_1 \ni 2 , \qquad s_2 \ni 1 , \qquad s_2 \not\ni 2 .$$

This implies

$$k(s_1) = k_1 , \qquad k(s_1) = k_2 , \qquad k(s_2) = k_1 , \qquad k(s_2) = k_2^c$$

and hence in particular

$$k_2 = k_2^c . \tag{4.4}$$

But the unbiasedness of e for $\mathbf{y} = \mathbf{0}$ gives $\sum_s p(s)k(s) = 0$ whence

$$k_2 \pi_2 + k_2^c(1 - \pi_2) = 0 . \tag{4.5}$$

Now (4.4) and (4.5) obviously give $k_2 = k_2^c = 0$ and thus (4.3) shows that $k(s) = 0$ for all $s \in \bar{S}$.

In case p is a unicluster design with at least three clusters, it is obvious that, given any two samples s_1 and s_2 in $\bar{S}$, there is a unit, say i, which occurs in neither s_1 nor s_2 and thus

$$k(s_1) = k_i^c , \qquad k(s_2) = k_i^c .$$

This shows that k is constant over $\bar{S}$ and since $E(k) = 0$, this constant value can be nothing but zero.

Finally, if p is a census, A^* contains just one estimator, so does A_0^*, and the proof is complete. □

REMARK 4.4. That the condition "not unicluster with exactly two clusters" is necessary in Lemma 4.3. will be seen in the proof of Theorem 4.5.

Now we are in a position to give a complete description of all hyperadmissible estimators in A^*.

THEOREM 4.5. *If the design p has $\pi_i > 0$ for all i, then either*

$$\bar{e} \quad \textit{is the only estimator which is hyperadmissible in} \quad A^*$$

or

$$\textit{every} \quad e \in A^* \quad \textit{is hyperadmissible in} \quad A^* .$$

When p is a non-unicluster design or a unicluster design with at least three clusters, the first case occurs; when p is a unicluster design with exactly two clusters, the second case occurs; when p is a unicluster design with just one cluster, i.e., a census, both (!) cases occur.

PROOF. If p is a non-unicluster design or a unicluster design with at least three clusters, the result follows from Lemma 4.3, Theorem 4.2 and Theorem 4.1 if we observe that an estimator $e \in A_0^*$ which is hyperadmissible in A^* also is hyperadmissible in A_0^*. If p is a unicluster design with exactly two clusters, it is easily seen that A^* coincides with the class of estimators of the form $\bar{e}(s, \mathbf{y}) + k(s)$ where $\sum_s p(s)k(s) = 0$ and that every such estimator is hyperadmissible. Finally, when p is a census, $A^* = \{\bar{e}\}$ and the result follows. □

REMARK 4.6. The part of Theorem 4.5 which deals with non-unicluster designs was given in Ramakrishnan (1970) with another (and much longer) proof.

5. Hyperadmissibility in the sense of Definition 3.2. In this section hyperadmissibility, unless otherwise stated, is to be understood in the sense of Definition 3.2.

The following condition upon a design p is used by Joshi (1971).

CONDITION 5.1. There is a sequence $i_1, i_2, \cdots, i_k$ of units such that

$$\bar{S} = \bigcup_{j=1}^{k} \bar{S}_{i_j}^c$$

$$\bar{S}_{i_j}^c \cap \left(\bigcup_{r=1}^{j-1} \bar{S}_{i_r}^c\right) \neq \varnothing \qquad \text{for } j = 2, 3, \cdots, k .$$

Joshi (1971) states the following theorem.

THEOREM 5.2. *Let p be a design such that $\pi_i > 0$ for all i and such that Condition 5.1 is fulfilled. Then the Horvitz–Thompson estimator $\bar{e}$ is the only hyperadmissible estimator in the class A^* of all unbiased estimators of the population total.*

This theorem obviously asserts two things:

$$\bar{e} \text{ is hyperadmissible in } A^* , \tag{5.1}$$

$$\text{no other estimator is hyperadmissible in } A^* . \tag{5.2}$$

An examination of Joshi's proof reveals that he only proves (5.2), completely forgetting (5.1). (It should be observed that neither (5.1) nor (5.2) follows from the corresponding part of Hanurav's (1968) theorem where, in addition to some other changes, A^* is replaced by the class M^* of all polynomial unbiased estimators.) The following theorem obviously fills up the gap left by Joshi.

THEOREM 5.3. *Let p be a design such that $\pi_i > 0$ for all i. Then the Horvitz–Thompson estimator $\bar{e}$ is hyperadmissible in the class A^* of all unbiased estimators of the population total.*

PROOF. Given any $e \in A^*$ such that $V(e, \mathbf{y}) \leqq V(\bar{e}, \mathbf{y})$ in the interior $R^0(i_1, i_2, \cdots, i_k)$ of a certain phs, we shall show that $V(e, \mathbf{y}) = V(\bar{e}, \mathbf{y})$ in $R^0(i_1, i_2, \cdots, i_k)$. For notational simplicity we assume that the phs in question is $R(1, 2, \cdots, k)$. Putting

$$h(s, \mathbf{y}) = e(s, \mathbf{y}) - \bar{e}(s, \mathbf{y})$$

it is sufficient to show that

$$h(s, \mathbf{y}) = 0 \qquad \text{for } s \in \bar{S},\ \mathbf{y} \in R^0(1, 2, \cdots, k) . \tag{5.3}$$

To achieve this, we introduce a discrete probability measure α on R_N such that if $\mathbf{y}$ is a random variable with distribution given by α, then

$$y_1, y_2, \cdots, y_N \text{ are independent,} \tag{5.4}$$

$$\begin{aligned} P_\alpha(y_i = 0) &= 0 \qquad \text{for } i = 1, 2, \cdots, k \\ &= 1 \qquad \text{for } i = k+1, k+2, \cdots, N . \end{aligned} \tag{5.5}$$

Expectation, variance and covariance with respect to α will be denoted by E_α, V_α and C_α, respectively. Now we will use some results from Section 6 of Godambe and Joshi (1965). Their equations (37) and (43) give

$$E_p V_\alpha(e) = E_p V_\alpha(\bar{e}) + E_p V_\alpha(h) . \tag{5.6}$$

Their equation (46) gives

$$E_\alpha V_p(e) = E_p V_\alpha(e) + E_p(E_\alpha(e) - E_\alpha(T))^2 - V_\alpha(T) \tag{5.7}$$

and for $e = \bar{e}$ this simplifies into

$$E_\alpha V_p(\bar{e}) = E_p V_\alpha(\bar{e}) + V_p(\bar{e}, \mathbf{x}) - V_\alpha(T) \tag{5.8}$$

where $\mathbf{x} = (x_1, x_2, \cdots, x_N) = (E_\alpha y_1, E_\alpha y_2, \cdots, E_\alpha y_N)$. Combining these three relations and skipping one nonnegative term we get

$$E_p V_\alpha(h) \leqq V_p(\bar{e}, \mathbf{x}) + E_\alpha V_p(e) - E_\alpha V_p(\bar{e}) . \tag{5.9}$$

Since by assumption $V_p(e, \mathbf{y}) \leqq V_p(\bar{e}, \mathbf{y})$ in $R^0(1, 2, \cdots, k)$ and since α has its support in $R^0(1, 2, \cdots, k)$ we get

$$E_\alpha V_p(e) \leqq E_\alpha V_p(\bar{e})$$

which, when combined with (5.9), gives

$$E_p V_\alpha(h) \leqq V_p(\bar{e}, \mathbf{x}) . \tag{5.10}$$

Now let $a_1, a_2, \cdots, a_k, b_1, b_2, \cdots, b_k$ be arbitrary numbers with $a_i < 0 < b_i$ for all i, let α have its support on the 2^k points $(c_1, c_2, \cdots, c_k\ 0, 0, \cdots, 0)$ where each c_i equals either a_i or b_i and choose the probabilities $P_\alpha(y_i = a_i)$ and $P_\alpha(y_i = b_i)$ such that $E_\alpha y_i = 0$ for all i, i.e., such that $\mathbf{x} = \mathbf{0}$. For this choice of α, (5.10) gives

$$E_p V_\alpha(h) = 0 ,$$

that is,

$$V_\alpha(h, s) = 0 \qquad \text{for } s \in \bar{S} . \tag{5.11}$$

Until further notice, let s be a fixed sample in $\bar{S}$ and call the function $\mathbf{y} \to h(s, \mathbf{y})$ just "the function h". From (5.11) it obviously follows that the function h is constant on the support of α. For a fixed $i\ (\leqq k)$, let a_i and b_i vary over $(-\infty, 0)$ and $(0, \infty)$, respectively, while all the other a's and b's are constant. Then it is seen that h is constant on each line in $R^0(1, 2, \cdots, k)$ parallel to the ith axis. This being so for all i, it follows that h is constant in the whole of $R^0(1, 2, \cdots, k)$:

$$h(s, \mathbf{y}) = h(s) \qquad \text{for } \mathbf{y} \in R^0(1, 2, \cdots, k) .$$

Now the inequality $V_p(e, \mathbf{y}) \leqq V_p(\bar{e}, \mathbf{y})$ gives

$$V_p(h, \mathbf{y}) + 2C_p(h, \bar{e}, \mathbf{y}) \leqq 0 \qquad \text{for } \mathbf{y} \in R^0(1, 2, \cdots, k) .$$

Letting the nonzero coordinates $y_1, y_2, \cdots, y_k$ tend to zero, $C_p(h, \bar{e}, \mathbf{y})$ tends to zero while $V_p(h, \mathbf{y})$ is constant (since $h(s, \mathbf{y})$ does not depend upon $\mathbf{y}$). Thus

$$V_p(h, \mathbf{y}) = 0$$

and h is constant also as function of s. Since $E_p(h, \mathbf{y}) = 0$, this constant value can be nothing but zero. □

Using some of the results of earlier sections, we now give a short alternative proof of the statement (5.2).

THEOREM 5.4. *Let p be a design such that $\pi_i > 0$ for all i and such that Condition 5.1 is satisfied. If the estimator e is hyperadmissible in the class A^*, then $e = \bar{e}$, the Horvitz–Thompson estimator.*

PROOF. We will first prove that $e \in A_0^*$. Using the fact that e is admissible when $\mathbf{y}$ is restricted to the interior of the one-dimensional phs $R(i)$, Joshi (1971, equations 14–18) shows that $e(s, \mathbf{y})$ must be constant, say k_i^c, for $s \in \bar{S}_i^c$ and $\mathbf{y} \in R(i)$. It follows that $e(s, \mathbf{0}) = k_i^c$ for all $s \in \bar{S}_i^c$. Since i is arbitrary, it is clear that under Condition 5.1, $e(s, \mathbf{0})$ must have the same value, say k^c, for all $s \in \bar{S}$. The unbiasedness of e shows that $k^c = 0$ and hence $e \in A_0^*$. Now it is obvious that e is hyperadmissible not only in A^* but also in A_0^*. From part (ii) of Theorem 3.3 follows that e is hyperadmissible in A_0^* in the sense of Definition 3.1 and finally Theorem 4.2 shows that $e = \bar{e}$. □

REFERENCES

BASU, D. (1971). An essay on the logical foundations of survey sampling, Part I. *Foundations of Statistical Inference* (V. P. Godambe and D. A. Sprott, eds.). Holt, Rinehart and Winston, Toronto. 203–233.

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

HANURAV, T. V. (1965). Optimum sampling strategies. Ph. D. thesis submitted to the Indian Statistical Institute.

HANURAV, T. V. (1968) Hyperadmissibility and optimum estimators for sampling finite populations. *Ann. Math. Statist.* **39** 621–642.

JOSHI, V. M. (1971). Hyperadmissibility of estimators for finite populations. *Ann. Math. Statist.* **42** 680–690.

RAMAKRISHNAN, M. K. (1970). Optimum estimators and strategies in survey sampling. Ph. D. thesis submitted to the Indian Statistical Institute.

DEPARTMENT OF MATHEMATICAL STATISTICS
BOX 725
S-220 07 LUND, SWEDEN

DEPARTMENT OF STATISTICS
UNIVERSITY OF SINGAPORE
SINGAPORE 10