7907233

BIYANI, SHRIRAM HARIKISHAN
ON SOME ESTIMATION PROBLEMS IN FINITE
POPULATION SAMPLING.

IOWA STATE UNIVERSITY, PH.D., 1978

University
Microfilms
International 300 N. ZEEB ROAD, ANN ARBOR, MI 48106

On some estimation problems in finite population sampling

by

Shriram Harikishan Biyani

A Dissertation Submitted to the

Graduate Faculty in Partial Fulfillment of

The Requirements for the Degree of

DOCTOR OF PHILOSOPHY

Major: Statistics

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

For the Major Department

Signature was redacted for privacy.

For the Graduate College

Iowa State University
Ames, Iowa

1978

TABLE OF CONTENTS

CHAPTER I. INTRODUCTION AND REVIEW

Introduction

Survey sampling is the most easily understood branch of statistics for the layman, but perhaps the most challenging one for the theoretical statistician. It is often called, and aptly, the game of "inferring about the whole by observing a part." In a less literal sense this description may be applied to statistics itself, but in sample surveys one deals with "real" populations as opposed to the hypothetical ones considered in most other statistical areas.

Formally, a finite population is defined as a finite collection of units, distinguishable from each other by means of certain labels. A sample surveyor attempts to infer about some function of the values of a characteristic of interest for all units in the population. In the traditional approach to a sample survey, these values are treated as constants, a sample is selected by a randomized mechanism, and the statistical inferences are based on the sampling distribution generated by this mechanism.

The inferential problems with this approach began to draw considerable attention since the first major attempt by Godambe (1955) to give a unified theory of finite population sampling. An excellent discussion of the subsequent developments in the foundations of sample surveys is given by Cassel, Särndal and Wretman (1977).

It has been recognized that when the population units are regarded as distinguishable, yet the characteristic values associated with them are considered to be constants, the likelihood function is generated only by the "man-made" randomization used to draw the samples, and is "flat" or uninformative. If one is to follow the likelihood principle, no non-trivial inference can be derived. This is not surprising. One cannot say much about one set of constants (the unobserved values) on the basis of the knowledge of a different set of constants (the observed values), when no relationship between them is specified.

One suggested way to overcome this difficulty is the "Scale-Load approach" which ignores the unit labels. This has been used by Royall (1968), Hartley and Rao (1969) and an extension of it by Särndal (1976). Although it has succeeded in the simplest problems, it is obviously inefficient when the unit labels carry relevant information, (i.e., are related with the values of the characteristic of interest). When the function to be inferred about is itself dependent on the labels, i.e., not symmetric with respect to all units, then the Scale-Load approach is inapplicable.

The alternative that has gained considerable popularity is the "superpopulation" or prediction approach, which provides a "logical link" between the observed and unobserved values by regarding them as realized values of random variables with distributions involving one or more common unknown parameters. The observed values are used to implicity or explicitly estimate these parameters and hence to predict the unobserved values.

For those who are reluctant to accept the philosophy of this method, Royall (1970) suggests that if the uncertainty about an event in one direction of time is considered to be within the domain of probability, then the same event, placed in another direction of time should be also treated likewise.

"It seems frequently to be true that at some time before $y_1,\ldots y_n$ are fixed, it is natural and generally acceptable to consider these numbers as values to be realized of random variables $Y_1,\ldots Y_N$. If such a model is appropriate before the y's are realized, it seems equally appropriate after they are fixed, but unobserved."

An extension of the prediction approach is the full-scale Bayes approach, as used by Ericson (1965). This involves a further tier of prior probability distributions for the superpopulation parameters.

The prediction approach is used in this dissertation to obtain a unified method of constructing "optimal estimators" under a general symmetric model, for a wide class of functions. This, of course, includes the population total or the mean, to which a large part of the literature in Sampling Theory is devoted. A related problem, specifically investigated here, is the assessment of the precision of a well-known estimator of the population total due to Horvitz and Thompson (1952).

Of course, there are many other functions of interest. Although ad hoc estimators have been suggested for most functions of interest, some of them appear to be contrived on a purely algebraic basis, to satisfy the hallowed property of "unbiasedness." It seems worthwhile to examine their behaviour and find better alternatives, if necessary. This has been the primary motivation behind this investigation.

Sample Survey Model

Consider a finite population U, consisting of a known number N of units, labelled 1, 2....N. For any characteristic y of interest, let y_i denote the value of the characteristic for the unit i.

<u>Definition</u> The vector $\underline{y} = (y_1, \ldots y_N)'$ will be called the <u>population vector</u> of y values.

All components of $\underline{y}$ are initially assumed to be unknown. The range of possible values of $\underline{y}$ will be denoted by $\mathcal{Y}$, and is usually taken to be the N-dimensional Euclidean space R^N. When several characteristics are simultaneously under study, each y_i itself may be regarded as a vector, and $\mathcal{Y}$ then changes accordingly.

<u>Definition</u> (sample) A sample is a subset of the population, and will be denoted by s.

A sample survey is the process of selecting a sample s from U and observing the values y_i for units $i \varepsilon s$.

Sampling Design

For a given survey, let S denote the collection of possible samples.

<u>Definition</u> (sampling design) A sampling design is a probability measure p on S, such that for any $s \varepsilon S$, p(s) is the probability of selecting the sample s.

Although it is possible to define a "sequential design", which is a design depending on $\underline{y}$, such designs are not commonly used because of difficulties in their implementation. We will only consider "non-sequential designs" which do not depend on $\underline{y}$. For such designs, defining a sample as a set, rather than as a sequence, as defined by some authors, entails no essential loss of generality. It can be easily seen that the set of labels and y-values for the units in the sample is a <u>sufficient statistic</u>, and any other information such as the order in which the units are drawn is irrelevant. A formal proof is given by Godambe and Joshi (1965), quoting Hájek (1959). In view of this, the data from a sample survey may be summarized as $d = \{(i, y_i), i \varepsilon s\}$. We may add here that any auxiliary information can be regarded as a function of the unit labels and is therefore included in 'd'.

<u>Definition</u> (sample size) The size of a sample s is the number of distinct units in s, and is denoted by n(s).

<u>Definition</u> A <u>fixed sample size design</u> is a design p which gives positive probability only to samples of a fixed size n.

<u>Definition</u> <u>Simple random sampling</u> (SRS) is the design which gives probability $1/\binom{N}{n}$ to every sample of size n.

Estimation

<u>Definition</u> (estimator) An estimator e is a function defined on $S \times \mathcal{Y}$, such that $e\,(s, \underline{y})$ depends on $\underline{y}$ only through y_i, $i \varepsilon s$.

<u>Definition</u> A <u>sampling strategy</u> is a pair (p, e) consisting of a design p and an estimator e.

<u>Definition</u> <u>Expectation</u> of an estimator e with respect to a design p is defined as,

$$E\,(e, \underline{y}) = \sum_{s \varepsilon S} p(s)\, e\,(s, \underline{y})$$

<u>Definition</u> (design-unbiasedness) An estimator e is said to be p-unbiased (or design-unbiased) for F with respect to a design p, if $\sum_{s \in S} p(s)\, e\,(s, \underline{y}) = F(\underline{y})$ for all $\underline{y} \in \mathcal{Y}$.

We will refer to design-unbiasedness as simply unbiasedness. Certain other types of unbiasedness are defined in the section on "Superpopulation Models." A function is said to be <u>estimable</u> if an unbiased estimator for F exists.

<u>Definition</u> <u>Variance</u> of an estimator e is,

$$V\,(e, \underline{y}) = \sum_{s \in S} p(s)\,[e\,(s, \underline{y}) - E\,(e, \underline{y})]^2$$

<u>Definition</u> <u>Mean Square Error (MSE)</u> of an estimator e for a function F is defined as,

$$MSE\,(e, \underline{y}) = \sum_{s \in S} p(s)\,[e\,(s, \underline{y}) - F\,(\underline{y})]^2$$

For an unbiased estimator, variance and MSE are identical.

<u>Definition</u> A <u>linear estimator</u> (Godambe, 1955) is an estimator of the form

$$e\,(s, \underline{y}) = \sum_{i \in s} b_{si}\, y_i$$

where

b_{si} $i \in s$, $s \in S$ are fixed coefficients.

<u>Definition</u> An <u>affine linear estimator</u> is of the form

$$e\,(s, \underline{y}) = b_{so} + \sum_{i \in s} b_{si}\, y_i$$

where b_{so}, b_{si} $i\varepsilon s$; $s\varepsilon S$ are fixed coefficients.

A note on notation For brevity, we will omit the population vector $\underline{y}$ from the notation of estimators, variance, MSE and other population functions, when there is no ambiguity.

Unequal Probability Sampling

The labels of a population may or may not contain any relevant information about the characteristic of interest. A typical case of informative labels is when we have the values of an auxiliary variable x, related with y, available for all population units. The information in the labels may be utilized in the design of the survey as well as in the estimation. At the design stage, one way to use this information is to group the units into different strata based on the x-values. Another way is to allow different probabilities of selection to different units, depending on their x-values.

When y is expected to be approximately proportional to x, it may be advantageous to use a design which gives each unit a probability of selection proportional to its x-value. Such designs are often called "probability proportional to size" (p.p.s.) designs, since the variable x is typically a measure of the size of the units.

Definition Inclusion probability of a unit i, denoted by π_i, is the probability that i will be included in the sample.

$$\pi_i = \Pr\ (i\varepsilon s) = \sum_{s \ni i} p(s)$$

where the sum is over all samples containing the unit i.

Joint inclusion probability of two units i and j is

$$\pi_{ij} = \Pr\ (i \text{ and } j \varepsilon s) = \sum_{s \ni i,j} p(s)$$

Definition Horvitz-Thompson estimator of the population total $T(\underset{\sim}{y}) = \sum_{i=1}^{N} y_i$ is

$$e_{HT}(s) = \sum_{i \varepsilon s} y_i/\pi_i$$

Defining $z_i = y_i/\pi_i$, $e_{HT}(s) = \sum_{i \varepsilon s} z_i$ (1.1)

If $\pi_i > o$ for all i, then e_{HT} is unbiased for T. (If $\pi_i = o$ for some unit, then T is not estimable).

Variance of the Horvitz-Thompson estimator, denoted by V_{HT}, is

$$V_{HT} = \sum_{i=1}^{N} \pi_i(1-\pi_i)z_i^{\,2} + \sum_{i \neq j}^{N} (\pi_{ij}-\pi_i\pi_j)z_i z_j$$

For fixed sample size designs, an alternative (but equivalent) expression for V_{HT}, due to Yates and Grundy (1953) is,

$$V_{HT} = \sum_{1 \le i < j \le N} (\pi_i\pi_j-\pi_{ij})(z_i-z_j)^2 \qquad (1.2)$$

It is easy to show that V_{HT} is estimable, if and only if $\pi_{ij} > o$

for every pair (i, j).

When V_{HT} is estimable, two well-known unbiased estimators of it, derived from the two variance expressions by Horvitz and Thompson (1952), and Yates and Grundy (1953) respectively, are

$$v_{HT}(s) = \sum_{i\varepsilon s} (1-\pi_i)z_i^2 + 2\sum_{i<j\varepsilon s} (\pi_{ij}-\pi_i\pi_j)/\pi_{ij} \cdot z_i z_j \tag{1.3}$$

and

$$v_{YG}(s) = \sum_{i<j\varepsilon s} (\pi_i\pi_j-\pi_{ij})/\pi_{ij} \cdot (z_i-z_j)^2 \tag{1.4}$$

A generalization of (1.2) for the variance of a more general functional form is given in Appendix I, from which an unbiased estimator similar to (1.4) may be derived. This form can be expected to be computationally more stable than the original expression.

It may be noted that e_{HT} as well as its two variance estimators are derived by dividing each term in the respective population expressions (T and the two expressions for V_{HT}) by the corresponding inclusion probability and restricting the sum to the sample instead of the population.

Although this has become a common technique for constructing unbiased estimators, the result may not always be very useful. This is one of the morals of a humorous story by Basu (1971) involving the Horvitz-Thompson estimator. Examples in Chapter II involving variance estimators also serve to emphasize the point.

Uniformly Best Estimation

The choice of estimator in most sample survey situations, as in other areas of statistics, is by no means simple, because estimators which are best for all $\underline{y}$, where "bestness" is defined in some reasonable sense, usually do not exist. One might for example, like to minimize the error, as measured by the MSE. But it is easy to see that no estimator minimizing MSE for all $\underline{y}$ can exist, except in the trivial case where we observe (with probability one) every y_i on which the function F of interest depends.

We may impose the restriction of unbiasedness, a property considered desirable, or even essential by many statisticians. Some arguments in its favour are given by Godambe and Joshi (1965, p. 1709) and Godambe and Thompson (1973). However, a uniformly best estimator within the class of unbiased estimators also does not exist and a further restriction of linearity is also of not much help.

Definition An unbiased estimator e (for a given function) is said to be Uniformly Minimum Variance Unbiased Estimator (UMVUE) in a class C, if

$$V\,(e,\ \underline{y}) \le V\,(e',\ \underline{y})$$

for every unbiased estimator $e' \varepsilon C$ and all $\underline{y} \varepsilon \mathcal{Y}$.

A Uniformly Minimum MSE estimator in a class can be similarly defined. The following is a brief review of the results on the non-existence of uniformly best estimators in sample surveys.

Godambe (1955) gave a proof of the non-existence of a UMVUE of the population total among linear unbiased estimators.

Hanurav (1966) pointed out that Godambe's proof fails for a small class of designs, called Unicluster Designs. These are designs for which any two possible samples are mutually disjoint. For such designs, there is only one linear unbiased estimator of the population total, which is trivially the UMVUE.

Basu (1971) gave an elementary proof of a very general result. He showed that for any estimable function F and any point $\underline{y}_o$ in $\mathcal{Y}$, we can construct an unbiased estimator e_o which coincides with F for $\underline{y} = \underline{y}_o$, i.e., $e_o(s, \underline{y}_o) = F(\underline{y}_o)$ for all $s \in S$.

Thus, the variance (MSE) of e_o at $\underline{y}_o$ is zero. If a UMVUE of F exists, it must also have zero variance at $\underline{y}_o$, and $\underline{y}_o$ being arbitrary, it must have zero variance for all $\underline{y}$. This is impossible, except in the trivial case mentioned above.

The following results follow from Basu's proof, for any function F and any design for which it is estimable, (except for the trivial case when all relevant y_i's are observed).

No Uniformly Minimum MSE estimator exists in the classes of

1. all estimators
2. all affine linear estimators
3. all unbiased estimators
4. affine linear unbiased estimators

Basu's proof actually holds in a more general set-up, where "squared error" is replaced by a more general "loss function", satisfying only the condition that it vanishes when the estimator equals the function value, but is not identically zero.

Admissibility

The non-existence of uniformly best estimators in most practical situations led to a look at weaker optimality criteria. One criterion which has received considerable attention is admissibility.

Definition (uniformly better) An estimator e' is said to be uniformly better (for a given design) than e, if MSE (e', $\underline{y}$) ≤ MSE (e, $\underline{y}$) for all $\underline{y}$ in $\mathcal{Y}$, with strict inequality for some $\underline{y}$.

Definition (admissibility) An estimator e is said to be admissible (for a given design) in a class C of estimators, if there is no estimator in C which is uniformly better than e.

If e is admissible in the class of all estimators, then e is said to be admissible.

When C is a convex class, then the definition of admissibility can be reduced to the following:

"e is admissible in C, if there is no estimator e' ≠ e in C such that

$$\text{MSE } (e', \underline{y}) \le \text{MSE } (e, \underline{y}) \text{ for all } \underline{y}."$$

Remarks

(a) Although the simpler admissibility definition above is defined in the context of finite populations, it is more generally applicable, provided $e' \neq e$ is interpreted as $\Pr(e' \neq e) > o$.

(b) Most classes of estimators of interest- all estimators, unbiased, polynomial, non-negative, or continuous estimators or any combination of these - are convex.

For the estimation of the population total, Joshi (1965a, 1966) proved the admissibility of several well-known estimators, for any sampling design. These include

1. The "expansion estimator" $N\,\bar{y}_s$, (1.5)

where

$$\bar{y}_s = \sum_{i\varepsilon s} y_i/n(s),$$

2. The ratio estimator $\bar{y}_s/\bar{x}_s \cdot \sum_{i=1}^{N} x_i$, (1.6)

3. The regression estimator

$$N\,[\bar{y}_s + b_s\,(\bar{x} - \bar{x}_s)] \tag{1.7}$$

where

$$\bar{x} = \sum_{i=1}^{N} x_i/N \text{ and}$$

$$b_s = \sum_{i\varepsilon s} (x_i-\bar{x}_s)(y_i-\bar{y}_s) / \sum_{i\varepsilon s} (x_i-\bar{x}_s)^2$$

In addition, for fixed sample size designs, Joshi (1965b) proved the following result, which in particular implies the admissibility of the Horvitz-Thompson estimator.

An estimator e_b of the form $e_b(s) = \sum_{i\varepsilon s} b_i\, y_i$, (where b_i i=1...N are known constants) is admissible for the population total, if

(i) $b_i \geq 1 \quad i = 1, 2...N$

and

(ii) $\sum_{i=1}^{N} b_i^{-1} \geq n.$

Since there is a multitude of choices for b_i's, there are more than one admissible estimators of the above form. This implies the non-existence of a Uniformly Minimum MSE linear estimator for designs with fixed sample size $n < N$.

From the preceding results we can see that the criterion of admissibility does not substantially limit our search for an optimal estimator. This can be further seen from the following characterization of linear admissible estimators of population total for samples of size one, proved in Appendix II.

<u>Theorem 1.1</u> For a population with $N > 1$, and a design with fixed sample size 1 and inclusion probabilities $p_i > o$, $i = 1,2...N$, the estimator e_b defined by $e_b(\{i\}) = b_i\, y_i$, $i = 1, 2...N$ is admissible in the class of linear estimators of the population total if and only if one of the following holds:

(i) $\sum_{i\varepsilon A} b_i^{-1} \geq 1$ where $A = \{i\varepsilon U\colon b_i \geq 1\}$

(ii) $o \leq b_i < 1$ for at least two units i

(iii) $o < b_i < 1$ for some i and $\sum_{i\varepsilon B} b_i^{-1} \leq 1$

where $B = \{i\varepsilon U: b_i < 1\}$

From (i) we may observe that choosing $b_i = 1$ for any one unit i is sufficient to ensure admissibility, irrespective of the choice of other b_i's.

Hyperadmissibility

This is a criterion stronger than admissibility and requires admissibility in certain subspaces of R^N. It was defined in its original form by Hanurav (1965). It has been criticized as lacking intuitive appeal, involving an element of arbitrariness and being tailor-made to establish the optimality of the Horvitz-Thompson estimator (Basu, 1971).

Uniform admissibility

This refers to the admissibility of a strategy rather than of an estimator, for a given design. Some major results are given by Joshi (1966, 1969), Sekkappan and Thompson (1975) and others.

Although formally stronger than admissibility, it does not greatly reduce the choice of strategies.

Superpopulation Models

The basic idea of this approach is to regard the finite population itself as a random sample from another population (the "superpopulation"). Under this model, the population vector $\underline{y}$ is no longer a fixed vector, but the realization of a random vector $\underline{Y} = (Y_1 \ldots . Y_N)'$ with an assumed distribution ξ, sometimes called a prior.

The problem of estimating $F(\underline{y})$ now becomes a prediction problem. The optimality criterion used in most cases is the minimization of the average of the MSE with respect to ξ, often under further restrictions such as unbiasedness and linearity.

In general, it is not necessary to specify ξ completely. A few assumptions about certain features of it are often enough to give us the optimal estimator. The optimality thus holds over a fairly wide class of distributions ξ.

<u>Definition</u> A distribution ξ of $\underline{X} = (X_1 \ldots X_N)'$ is said to be <u>exchangeable</u>, if any permutation of $\underline{X}$ has the same distribution as $\underline{X}$.

An exchangeable prior distribution is appropriate when the prior knowledge about the vector is believed to be symmetric in its components.

According to Cochran (1977), the idea of a superpopulation dates back to Laplace in early 1800's. Its first formal use for comparison of estimators seems to belong to Cochran (1946).

Royall (1970) and Royall and Herson (1973) have used this approach to obtain optimal strategies under certain linear regression models.

Godambe and Thompson (1973) have shown that under a prior exchangeable in $\underline{Z} = (Z_1, \ldots . Z_N)'$, where $Z_i = Y_i/\pi_i$, the Horvitz-Thompson estimator minimizes the expected variance in the class of unbiased estimators.

Definition (model-unbiasedness) An estimator e is said to be ξ - unbiased, if $E_\xi\ [e(s, \underline{Y}) - F\ (\underline{Y})] = o$ for all $s \varepsilon S$.

Still another type of unbiasedness is "joint unbiasedness" with respect to a design p and a prior ξ.

Definition An estimator e is p - unbiased, if

$$\underset{\xi}{E} \sum_{s \varepsilon S} p(s)\ [e(s, \underline{y}) - F(\underline{y})] = o$$

Minimizing the expected MSE under a prior without imposing any type of unbiasedness restriction, or under model-unbiasedness, can be easily seen to be equivalent to separately minimizing $E_\xi\ [e(s, \underline{Y}) - F\ (\underline{Y})]^2$ for each sample $s \varepsilon S$. The optimal estimator in this case is the same for all designs. In other words, design plays no role in the estimation.

The superpopulation approach has been very helpful in pinpointing optimal estimators and strategies for given types of population, formally described by the models. But its greatest benefit seems to be the insight it can provide into the behaviour of various sampling strategies for different kinds of populations.

CHAPTER II. ON ESTIMATION OF VARIANCE IN UNEQUAL PROBABILITY SAMPLING

Comparison of Two Estimators

Two estimators v_{HT} and v_{YG} for the variance (V_{HT}) of the Horvitz-Thompson estimator were given in (1.3) and (1.4). Some of their properties are discussed below.

The range of the population vector $\underline{y}$ will be assumed to be R^N and the sampling design will be assumed to be of fixed sample size n throughout this chapter, unless otherwise stated.

Design-unbiasedness

The estimator v_{HT} is unbiased for all designs for which V_{HT} is estimable; v_{YG} is unbiased for all fixed sample size designs for which V_{HT} is estimable.

Non-negativity

Both v_{HT} and v_{YG} can take negative values depending on the design and the data. But whereas v_{YG} fails to be non-negative only for a limited class of designs, v_{HT} takes negative values for all of these designs and many more, as shown in theorem 2.2. However, for a particular dataset, it is possible to get a negative value of v_{YG} and a positive value of v_{HT}.

Behavior when $y_i \propto \pi_i$

If $z_1 = z_2 = \ldots z_N$ (where $z_i = y_i/\pi_i$), then it is clear from (1.2) that $V_{HT} = 0$. For such populations, v_{YG} also vanishes, but v_{HT} does not, in general. This property of v_{YG} is intuitively appealing, since we might expect the estimator to be accurate when there is no variability in the population.

Invariance

Consider the group of transformations $G = \{g_c : g_c (z_1, \ldots . z_N) = (z_1 + c, \ldots . z_N + c), \text{ c real}\}$.

It is evident from (1.2) that V_{HT} does not change under any of these transformations. From (1.4), v_{YG} can be seen to be invariant under this group of transformations, but v_{HT} is not invariant, unless the design is such that v_{HT} and v_{YG} coincide.

Admissibility

Godambe and Joshi (1965) proved the admissibility of v_{HT} in the class of unbiased estimators. Of course, it is not admissible among all estimators as it can take negative values. An estimator obtained by truncating v_{HT} to the left, at zero for instance, would be uniformly better than v_{HT}.

Since v_{YG} seems to have generally much better characteristics than v_{HT}, one might expect that it should be easier to prove its admissibility, at least in the class of unbiased estimators. But the only such result available is for sample size two, due to Joshi (1970).

Example 2.3 shows that for any higher sample size there are designs for which v_{YG} is inadmissible even in the more restricted class of non-negative unbiased quadratic estimators.

It should be also noted that for some designs, (e.g., simple random sampling), v_{YG} coincides with v_{HT}, and is therefore admissible among unbiased estimators. It is an open question whether there is a design for a sample size greater than two, for which v_{YG} does not coincide with v_{HT}, but is admissible in the class of unbiased estimators (or quadratic and/or non-negative unbiased estimators).

Non-Negative Variance Estimation

A necessary condition for non-negativity

We have noted earlier that both v_{HT} and v_{YG} can take negative values. It is obvious from (1.4) that a sufficient condition for v_{YG} to be non-negative definite (NND) is,

$$\pi_{ij} \leq \pi_i \pi_j \text{ for all } i, j \in U \; i \neq j \tag{2.1}$$

It is also clearly, a necessary condition for sample size two. For higher sample sizes it is not necessary as shown in example 2.1. However, designs satisfying (2.1) have been given by Fellegi (1963), Hajek (1964), Sampford (1967) and others.

A necessary condition for an unbiased polynomial estimator of V_{HT} to be NND was given by Vijayan (1975). His results are presented in lemma 2.1 and theorem 2.1 along with a simplified proof of the former.

Lemma 2.1

Let $Q(\underline{x}) = \underline{x}' A \underline{x}$ be a NND quadratic form, where $A = (a_{ij})$ is a mxm symmetric matrix and $\underline{x} = (x_1, \ldots x_m)'$.

If $Q(\underline{x}_o) = 0$ for $\underline{x}_o = (1, 1 \ldots 1)'$,

then

$$Q(\underline{x}) = \sum_{1 \le i < j \le m} (-a_{ij})(x_i - x_j)^2 \tag{2.2}$$

Proof $Q(\underline{x}_o) = \underline{x}_o' A \underline{x}_o = 0$ implies $A\underline{x}_o = 0$, since A is NND. That is,

$$\sum_{j=1}^{m} a_{ij} = 0 \qquad i = 1, 2 \ldots m \tag{2.3}$$

$$Q(\underline{x}) = \sum_{i=1}^{m} \sum_{j=1}^{m} a_{ij} x_i x_j$$

$$= \Sigma \; \Sigma \; a_{ij} [x_i^2 + x_j^2 - (x_i - x_j)^2]/2$$

We can split l.h.s. into three sums, first two of which vanish by (2.3) and the symmetry of A, and we have,

$$Q(\underline{x}) = 1/2 \sum_{i=1}^{m} \sum_{j=1}^{m} (-a_{ij})(x_i - x_j)^2$$

which reduces to (2.2) using the symmetry of A and noting that the terms corresponding to $i = j$ vanish.

For $\underline{z} = (z_1, \ldots z_N)$, define

$$\phi_{ij}\ (\underline{z}) = (z_i - z_j)^2 \quad i, j = 1, \ldots .N.$$

For brevity, we will write ϕ_{ij} for $\phi_{ij}\ (\underline{z})$.

Theorem 2.1

If e is an estimator belonging to either of the classes -

(1) quadratic estimators vanishing when all observed components of $\underline{z}$ (z_i $i \varepsilon s$) are equal

(2) polynomial estimators unbiased for a NND quadratic form $Q\ (\underline{z})$ vanishing when $z_1 = z_2 = \ldots z_N$, then a necessary condition for e to be NND is,

$$e\ (s) = \sum_{i<j\varepsilon s} c_{sij}\ \phi_{ij} \quad \text{for all } s \varepsilon S \tag{2.4}$$

where

c_{sij} $(1 \le i < j \le N,\ s \varepsilon S)$ are constants.

The first part of the theorem is a direct consequence of lemma 2.1. For the proof of the second part, see theorems 2.1 and 2.2 of Vijayan.

Theorem 2.1 provides a justification for restricting the search for estimators of variance to those of the form (2.4). We will consider the following classes of estimators.

$$C_{LN} = \{e\colon \text{e is of the form (2.4) and NND}\} \tag{2.5}$$

$$C_{LNU} = \{e\colon e\varepsilon C_{LN} \text{ and is unbiased}\} \tag{2.6}$$

Because of theorem 2.1, C_{LNU} is identical with the class of NND unbiased quadratic estimators.

Non-negativity of v_{HT}

As a corollary to the above theorem, Vijayan has claimed, without proof, that v_{HT} can take negative values for any design. This is, however, not true for most of the well-known equal probability designs, as well as stratified designs with equal probabilities within the strata. The following is a partial characterization of designs for which v_{HT} is NND.

Theorem 2.2

A necessary condition for v_{HT} to be NND for a design is, $v_{HT} = v_{YG}$.

Proof By theorem 2.1 v_{HT} can be NND only if

$$v_{HT}\,(s) = \sum_{i<j\varepsilon s} C_{sij}\,(z_i - z_j)^2 \tag{2.7}$$

for some C_{sij} $i < j \varepsilon s$, $s\varepsilon S$.

Equating the coefficient of $z_i\,z_j$ in (1.3) and (2.7),

$$2\,(\pi_{ij} - \pi_i\,\pi_j)\,/\,\pi_{ij} = -2\,C_{sij}$$

or

$$C_{sij} = (\pi_i\,\pi_j - \pi_{ij})\,/\,\pi_{ij}$$

Substituting for C_{sij} in (2.7), we find that v_{HT} reduces to v_{YG}.

Remark Theorem 2.2 tells us that for any design for which v_{YG} is not NND, neither can be v_{HT}.

The following lemma is useful in the construction of various examples in this chapter.

Lemma 2.2

Let $C_1, \ldots C_m$ be constants such that $\sum_{i=1}^{m} C_i = 0$. Then,

(i) $\sum_{1 \le i < j \le m} C_i C_j \phi_{ij} \le 0$

(ii) $\phi_{ij} \le 2 (\phi_{ik} + \phi_{jk})$ (2.8)

Proof Applying lemma 2.1 to $Q(\underline{z}) = \left(\sum_{i=1}^{m} C_i z_i \right)^2$,

we get

$$\sum_{1 \le i < j \le m} (-C_i C_j) \phi_{ij} = \left(\sum_{i=1}^{m} C_i z_i \right)^2 \ge 0 \quad \text{which gives (i).}$$

The second part is a special case of (i) with $C_i = C_j = 1$, $C_k = -2$ and $C_r = 0$ for $r \ne i, j, k$.

Example 2.1 (Non-necessity of the condition $\pi_{ij} \le \pi_i \pi_j$)

The following example shows that condition (2.1) is not necessary for v_{YG} to be NND.

Let $N = 5$, $n = 3$

$p(s) = 1/4 \quad s = (1\ 2\ 3),\ (1\ 2\ 4),\ (1\ 2\ 5)$ or $(3\ 4\ 5)$

$\quad = 0$ for all other samples.

The inclusion probabilities are $\pi_1 = \pi_2 = 3/4$, $\pi_3 = \pi_4 = \pi_5 = 1/2$, *and the joint inclusion probabilities are*

$$\pi_{ij} = 3/4 \quad (i,j) = (1,2)$$

$$= 1/4 \text{ otherwise.}$$

$$\pi_i \pi_j - \pi_{ij} = -3/16 \quad (i,j) = (1,2)$$

$$= 1/4 \quad i = 1,2 \quad j = 3,\ 4,\ 5$$

$$= 0 \quad (i,j) = (3,4),\ (3,5) \text{ or } (4.5).$$

Thus, inequality (2.1) does not hold for the pair (1,2). However, the value of v_{YG} for $s = (1\ 2\ 3)$ is,

$$-\ 1/4\ \phi_{12} + 1/2\ (\phi_{13} + \phi_{23})$$

which is non-negative by lemma 2.2 (ii). The non-negativity for other samples can be similarly verified.

Example 2.2 (Existence of NND unbiased estimator when v_{YG} is not NND)

The sampling design and the values of v_{YG} and another unbiased estimator v_T are given in Table 2.1.

The inclusion probabilities and joint inclusion probabilities are

$$\pi_1 = \pi_2 = 5/6,\ \pi_3 = 1/3,\ \pi_4 = \pi_5 = 1/2;$$

$$\pi_{12} = 5/6,\ \pi_{13} = \pi_{23} = \pi_{34} = \pi_{35} = \pi_{45} = 1/6,$$

and

$$\pi_{14} = \pi_{15} = \pi_{24} = \pi_{25} = 1/3.$$

The unbiasedness of v_T is easily verified by observing that $E\ (v_{YG} - v_T) = 0$ and v_{YG} is unbiased.

Table 2.1. Example 2.2

Sample s	Prob. p(s)	v_{YG} (s)	v_T (s)
1 2 3	1/6	$-1/6\ \phi_{12} + 2/3\ (\phi_{13} + \phi_{23})$	$-1/3\ \phi_{12} + 2/3\ (\phi_{13} + \phi_{23})$
1 2 4	1/3	$-1/6\ \phi_{12} + 1/4\ (\phi_{14} + \phi_{24})$	$-1/8\ \phi_{12} + 1/4\ (\phi_{14} + \phi_{24})$
1 2 5	1/3	$-1/6\ \phi_{12} + 1/4\ (\phi_{15} + \phi_{25})$	$-1/8\ \phi_{12} + 1/4\ (\phi_{15} + \phi_{25})$
3 4 5	1/6	$1/2\ \phi_{45}$	$1/2\ \phi_{45}$

Non-negativity of v_T can be seen from lemma 2.2 (ii). However, for s = (1 2 4) and $z_1 = 0,\ z_2 = 2,\ z_4 = 1$, v_{YG} has the value

$$-1/6\ (0-2)^2 + 1/4\ [(0-1)^2 + (2-1)^2] = -1/6$$

Inadmissibility Examples for the Yates-Grundy Estimator for $n > 2$

Example 2.3

Let $n > 2$ be arbitrary, and $N > 2n$, then for the sampling design described below, v_{YG} is inadmissible in the class C_{LNU} defined in (2.6).

Let $p(s) = \alpha$, if at least one of the units $1, 2....n \varepsilon s$

$= \beta$ otherwise

where $\alpha, \beta > o$ are to be chosen suitably.

Since selection probabilities must add to unity,

$$\left\{\binom{N}{n} - \binom{N-n}{n}\right\} \alpha + \binom{N-n}{n} \beta = 1 \tag{2.9}$$

The inclusion probabilities are easily seen to be,

$$\begin{aligned}\pi_i &= \binom{N-1}{n-1} \alpha & i = 1, 2, \ldots n \\ &= \binom{N-1}{n-1}\alpha + \binom{N-n-1}{n-1} (\beta - \alpha) & i = n+1, \ldots N \end{aligned} \tag{2.10}$$

and

$$\begin{aligned}\pi_{ij} &= \binom{N-2}{n-2} \alpha & i \text{ or } j \leq n \\ &= \binom{N-2}{n-2} \alpha + \binom{N-n-2}{n-2} (\beta - \alpha) & i \text{ and } j > n. \end{aligned} \tag{2.11}$$

Let $s_o = (1, 2...n)$ and $s_1 = (1, 2, n+1, ...2n-2)$.

For $h > o$, let v_h be defined as,

$$v_h\ (s) = v_{YG}\ (s) + h\ \phi_{12} \qquad s = s_o$$

$$= v_{YG}\ (s) - h\ \phi_{12} \qquad s = s_1$$

$$= v_{YG}\ (s) \quad \text{otherwise.}$$

Since $p\ (s_o) = p\ (s_1) = \alpha$, unbiasedness of v_h is easily verified from $E\ (v_h - v_{YG}) = 0$.

Now, choose α such that, $\pi_i\ \pi_j - \pi_{ij} = o \quad 1 \le i < j \le n$ (2.12)

i.e., $$\left[\binom{N-1}{n-1} \alpha\right]^2 - \binom{N-2}{n-2} \cdot \alpha = o \tag{2.13}$$

or, $$\binom{N-1}{n-1} \alpha = (n - 1)/(N-1)$$

β is determined from (2.9), and is given by,

$$\binom{N-n}{n} (\beta - \alpha) = 1 - \binom{N}{n} \alpha$$

$$= 1 - (N/n)(n - 1)/(N - 1)$$

$$= (N - n)/[n\ (n - 1)]$$

We may observe that $\beta - \alpha > o$.

The inequality $\pi_{ij} \le \pi_i\ \pi_j$ can be easily verified for all (i, j). For $i \le n$, $j > n$

$$(\pi_i \; \pi_j - \pi_{ij})/\pi_{ij} = \left[\binom{N-1}{n-1} \alpha \left\{\binom{N-1}{n-1} \alpha + \binom{N-n-1}{n-1} (\beta - \alpha)\right\} / \binom{N-2}{n-2} \alpha\right] - 1$$

[by (2.13)]

$$= \frac{N-1}{n-1} \binom{N-n-1}{n-1} (\beta - \alpha)$$

$$= k \text{ (say)} > 0.$$

From (2.12) we can see that $v_{YG}(s_o) = 0$.

$$v_{YG}(s_1) = \sum_{i=1}^{2} \sum_{j=n+1}^{2n-2} \left[(\pi_i \; \pi_j - \pi_{ij})/ \pi_{ij}\right] \phi_{ij}$$

$$= k \; (\phi_{1,n+1} + \phi_{2,n+1}) + \text{positive terms.}$$

Applying lemma 2.2(ii),

$$v_{YG}(s_1) \geq k \; \phi_{12}/2 \tag{2.14}$$

$$V(v_h) - V(v_{YG}) = E\,(v_h^{\;2} - v_{YG}^{\;2})$$

$$= \sum_{s=s_o,s_1} p(s) \left[v_h(s) - v_{YG}(s)\right]\left[v_h(s) + v_{YG}(s)\right]$$

$$= \alpha h\phi_{12}\ (h\ \phi_{12}) - \alpha h\ \phi_{12}\ [2\ v_{YG}(s_1) - h\ \phi_{12}]$$

$$= 2\alpha h\ \phi_{12}\ [h\ \phi_{12} - v_{YG}\ (s_1)]$$

$$\leq 0 \text{ for } 0 < h < k/2, \text{ using (2.14).}$$

This proves the inadmissibility of v_{YG} for the design in the example.

Remark In the preceding example, we had the relation $\pi_{ij} = \pi_i\ \pi_j$ for several pairs of units, which is unlikely to be true for most unequal probability designs. This relation is, however, not essential for the inadmissibility of v_{YG}, it merely makes the example easier to construct. The condition $N > 2n$ was also only a convenience. In the next example, v_{YG} is seen to be inadmissible even without these limitations. In this, we have $N<2n$ and the strict inequality $\pi_{ij} < \pi_i\ \pi_j$ holds for all pairs (i, j). With some trial and error it seems possible to construct similar examples for any combination of N and n $(2 < n < N)$, except possibly for $N = 4$ and $n = 3$.

Example 2.4

Table 2.2 gives the design and the values of v_{YG} and another estimator v_h (with h to be suitably chosen).

The inclusion probabilities are:

$$\pi_i = 7/8 \qquad i = 1, 2, 3, 4$$

$$= 1/2 \qquad i = 5$$

and

$$\pi_{ij} = 3/4 \qquad i < j \le 4$$

$$= 3/8 \qquad i \le 4,\ j = 5.$$

$$(\pi_i \pi_j - \pi_{ij})/ \pi_{ij} = 1/48 \qquad i < j \le 4$$

$$= 1/6 \qquad i \le 4,\ j = 5.$$

Table 2.2 Example 2.4

Sample s	Prob. p(s)	$v_{YG}(s)$	v_h (s)
1 2 3 4	1/2	$1/48(\phi_{12}+\phi_{13}+\phi_{23}+\phi_{14}+\phi_{24}+\phi_{34})$	$v_{YG}(s) + h\ \phi_{12}$
1 2 3 5	1/8	$1/48(\phi_{12}+\phi_{13}+\phi_{23})+1/6(\phi_{15}+\phi_{25}+\phi_{35})$	$v_{YG}(s) - 2h\ \phi_{12}$
1 2 4 5	1/8	$1/48(\phi_{12}+\phi_{14}+\phi_{24})+1/6(\phi_{15}+\phi_{25}+\phi_{45})$	$v_{YG}(s) - 2h\ \phi_{12}$
1 3 4 5	1/8		same as $v_{YG}(s)$
2 3 4 5	1/8		same as $v_{YG}(s)$

It is easy to verify $E(v_{YG} - v_h) = 0$, showing v_h is unbiased.

$$V(v_h) - V(v_{YG}) = \sum_s p(s)\,[v_h^{\,2}(s) - v_{YG}(s)]^2$$

$$= 1/2\ h\ \phi_{12}\ [h\ \phi_{12} + 2\ v_{YG}\ (1\ 2\ 3\ 4)]$$

$$+ 1/8\ (-2h\ \phi_{12})[-4h\ \phi_{12} + 2v_{YG}\ (1\ 2\ 3\ 5) + 2v_{YG}\ (1\ 2\ 4\ 5)]$$

$$= 1/2\ h\ \phi_{12}\ [3h\ \phi_{12} + 2\ v_{YG}\ (1\ 2\ 3\ 4) - v_{YG}\ (1\ 2\ 3\ 5) - v_{YG}\ (1\ 2\ 4\ 5)]$$

$$= 1/2\ h\ \phi_{12}\ [3h\ \phi_{12} + 1/48\ (\phi_{13} + \phi_{23} + \phi_{14} + \phi_{24} + 2\phi_{34}) - 1/6\ (2\phi_{15} + 2\phi_{25} + \phi_{35} + \phi_{45})] \qquad (2.15)$$

Lemma 2.2 (ii) with $c_1 = c_2 = c_3 = c_4 = 1$, $c_5 = -4$ gives

$$\phi_{12} + \phi_{13} + \phi_{23} + \phi_{14} + \phi_{24} + \phi_{34} - 4\ (\phi_{15} + \phi_{25} + \phi_{35} + \phi_{45}) \le 0 \qquad (2.16)$$

Putting h = 1/72 and using 1/24 times (2.16) in (2.15), we get,

$$V\ (v_h) - V\ (v_{YG}) \leq 0,$$

which shows the inadmissibility of v_{YG}.

Other Inadmissibility Examples

Ajgaonkar's estimators

Two further unbiased estimators of V_{HT}, due to Ajgaonkar (1967) are

$$v_{A1}(s) = \left[\sum_{i \varepsilon s} \pi_i\ (1-\pi_i)\ z_i^{\,2} / \binom{N-1}{n-1} + 2 \sum_{i<j \varepsilon s} (\pi_{ij} - \pi_i\ \pi_j) \cdot z_i\ z_j / \binom{N-2}{n-2} \right] / p(s)$$

and

$$v_{A2}(s) = \sum_{i<j \varepsilon s} (\pi_i\ \pi_j - \pi_{ij})(z_i - z_j)^2 / \left[\binom{N-2}{n-2}\ p(s) \right]$$

v_{A1}, like v_{HT}, takes negative values for most designs.

v_{A2} reduces to v_{YG} for n = 2. For n > 2, like v_{YG}, it can be inadmissible among unbiased estimators, or even in the more restricted class C_{LNU}, as shown in the next example.

Example 2.5

The design used is the same as in example 2.4. The values of v_{A2}

and another estimator v_h are given in Table 2.3.

Table 2.3. Example 2.5

Sample s	$v_{A2}(s)$	v_h (s)
1 2 3 4	$1/96\ (\phi_{12} + \phi_{13} + \phi_{23} + \phi_{14} + \phi_{24} + \phi_{34})$	$v_{A2}(s) + h\ \phi_{12}$
1 2 3 5	$1/24\ (\phi_{12} + \phi_{13} + \phi_{23}) + 1/6\ (\phi_{15} + \phi_{25} + \phi_{35})$	$v_{A2}(s) - 2h\ \phi_{12}$
1 2 4 5	$1/24\ (\phi_{12} + \phi_{14} + \phi_{24}) + 1/6\ (\phi_{15} + \phi_{25} + \phi_{45})$	$v_{A2}(s) - 2h\ \phi_{12}$
1 3 4 5		same as $v_{A2}(s)$
2 3 4 5		same as $v_{A2}(s)$

Along the lines of example 2.4, we can show

$$V(v_h) - V\ (v_{A2}) = 1/2\ h\ \phi_{12}\ [3h\ \phi_{12} + 1/96\ (\phi_{13} + \phi_{23} + \phi_{14} + \phi_{24} + \phi_{34}) - 1/6\ (2\ \phi_{15} + 2\ \phi_{25} + \phi_{35} + \phi_{45})]$$

$$\leq 0 \qquad \text{for } h = 1/72$$

which shows the inadmissibility of v_{A2}.

<u>Unbiased estimator of population variance in unequal probability sampling</u>

$$\text{Let } S^2 = \sum_{i=1}^{N} (y_i - Y)^2 \ / (N - 1)$$

where

$$\bar{y} = \sum_{i=1}^{N} y_i/N$$

Consider the following multiple of the population variance.

$$Q = N\ (N - 1)\ S^2 = \sum_{1 \le i < j \le N} \phi_{ij}$$

where

$$\phi_{ij} = (y_i - y_j)^2.$$

A "Yates-Grundy type" unbiased estimator of Q is given by

$$e_\sigma(s) = \sum_{i<j \epsilon s} \phi_{ij}/\pi_{ij}$$

provided, $\pi_{ij} > 0$ for all (i, j).

The following example shows that e_σ can be inadmissible in the class of non-negative unbiased quadratic estimators.

Example 2.6

The design and the values of e_σ and another estimator e_h are shown in Table 2.4.

Table 2.4. Example 2.6

Sample s	Prob. p(s)	e_σ (s)	e_h (s)
1 2 3 4	7/8	$16/15\ (\phi_{12} + \phi_{13} + \phi_{23} + \phi_{14} + \phi_{24} + \phi_{34})$	$e_\sigma(s) + h\phi_{12}$
1 2 3 5	1/32	$16/15\ (\phi_{12} + \phi_{13} + \phi_{23}) + 32/3(\phi_{15} + \phi_{25} + \phi_{35})$	$e_\sigma(s) - 14h\phi_{12}$
1 2 4 5	1/32	$16/15\ (\phi_{12} + \phi_{14} + \phi_{24}) + 32/3(\phi_{15} + \phi_{25} + \phi_{45})$	$e_\sigma(s) - 14h\phi_{12}$
1 3 4 5	1/32		same as $e_\sigma(s)$
2 3 4 5	1/32		same as $e_\sigma(s)$

Unbiasedness of e_h follows from $E\ (e_\sigma - e_h) = 0$.

$$V\ (e_h) - V\ (e_\sigma) = 7/8h\phi_{12}\ [h\phi_{12} + 2e_\sigma\ (1\ 2\ 3\ 4)]$$

$$+ 1/32\ (-\ 14h\phi_{12})[28h\phi_{12} + 2e_\sigma\ (1\ 2\ 3\ 5) + 2e_\sigma\ (1\ 2\ 4\ 5)]$$

$$= 7/8\ h\phi_{12}\ [15h\phi_{12} + 16/15\ (\phi_{13} + \phi_{23} + \phi_{14} + \phi_{24} + 2\ \phi_{34})$$

$$-\ 32/3\ (2\phi_{15} + 2\ \phi_{25} + \phi_{35} + \phi_{45})]$$

$$\le 0 \text{ for } h = 1/15.$$

Hence e_σ is inadmissible in C_{LNU}.

A Lower Bound For Non-Negative Definite Quadratic Forms

When a negative estimate is encountered for a function such as a variance, known to be non-negative, a solution sometimes suggested is to use zero as the estimate. But in finite population sampling, this would be usually known to be too low because of the observed variability in the sample. The idea is illustrated in the following example, which shows that even some positive estimates could be known with certainty to be too low.

Example 2.7 (Irrationality of v_{YG} for $n = 2$)

Although v_{YG} is known to be admissible for $n = 2$, in the class of unbiased estimators, it can be inadmissible when we drop the unbiasedness restriction, as shown below for the simplest possible case.

Let $N = 3$ and the design be as given below.

s	p (s)
(1 2)	0.6
(1 3)	0.2
(2 3)	0.2

We have, $\pi_1 = \pi_2 = .8$, $\pi_3 = .4$, $\pi_{12} = .6$, and $\pi_{13} = \pi_{23} = .2$

The true variance of Horvitz-Thompson estimator is,

$$V_{HT} = \sum_{1 \le i < j \le 3} (\pi_i \pi_j - \pi_{ij}) \phi_{ij} = .04\phi_{12} + .12 (\phi_{13} + \phi_{23})$$

(by lemma 2.2) $$\ge .04 \phi_{12} + .06 \phi_{12} = \phi_{12}/10 \qquad (2.17)$$

Now, the Yates-Grundy estimate for s = (1 2) is

$$v_{YG}\,(1\ 2) = (\pi_1\,\pi_2 - \pi_{12})\,\pi_{12}^{-1}\,\phi_{12} = \phi_{12}/15$$

We observe that v_{YG} gives a value smaller than the lower bound (2.17), which is known when the sample (1 2) is drawn. (And we get this impossibly low estimate from the sample with the largest selection probability!)

The estimator obtained by replacing v_{YG} by the lower bound (2.17) for s = (1, 2) and leaving it unchanged for other samples would be uniformly better than v_{YG}. This shows the inadmissibility of v_{YG} in the class C_{LN} defined in (2.5).

Remark Lanke (1974) has shown that for n = 2, v_{YG} is the only non-negative unbiased estimator, when one exists. It follows that no reasonable unbiased estimator of V_{HT} exists for the above design. This suggests that the design-unbiasedness criterion may not be a reasonable one for the estimation of variance in unequal probability sampling.

A general lower bound

The bound (2.17) is a special case of the following:

Let $\underline{y}'\,A\,\underline{y}$ be a non-negative definite quadratic form, where A is a symmetric N x N matrix.

Let $\underline{y} = (\underline{y}_1, \underline{y}_2)$ and A be correspondingly partitioned as

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

Let A_{22}^* be a symmetric matrix such that

$$A_{22} A_{22}^* A_{22} = A_{22}$$

and

$$A_{22}^* A_{22} A_{22}^* = A_{22}^*$$

(We may, for instance, take the Moore-Penrose inverse of A_{22}.) For non-singular A_{22}, of course, $A_{22}^* = A_{22}^{-1}$.

It is easy to verify that

$$\underline{y}' A \underline{y} = \underline{y}_1' (A_{11} - A_{12} A_{22}^* A_{21}) \underline{y}_1 + \underline{y}'_{2.1} A_{22} \underline{y}_{2.1}$$

where

$$\underline{y}_{2.1} = A_{22}^* A_{22} \underline{y}_2 - A_{22}^* A_{21} \underline{y}_1$$

Since A_{22} is NND,

$$\underline{y}' A \underline{y} \geq \underline{y}_1' (A_{11} - A_{12} A_{22}^* A_{21}) \underline{y}_1 \tag{2.18}$$

For the problem of estimating $\underline{y}' A \underline{y}$ when $\underline{y}_1$ is observed, it may be reasonable to expect an estimator to be not less than the lower bound (2.18). However, it may not always be computationally feasible to verify this, particularly for large N.

Bound (2.18) is the best possible, based on $\underline{y}_1$, because it is attained when $\underline{y}_{2.1} = 0$ and this is achieved for $\underline{y}_2 = A_{22}^* A_{21} \underline{y}_1$.

CHAPTER III. ESTIMATION OF A CLASS OF FUNCTIONS

We will consider a class of functions which can be considered as a generalization of either the linear functions or of Hoeffding's (1948) U-statistics. We will discuss the optimal estimation of such a function under a general superpopulation model, after showing the non-existence of a UMVUE, in general, in a class of estimators having the same form as the function to be estimated.

Linear Combination of Symmetric Kernels

(L-functions)

Definitions and notation

Definition (symmetric function) A function $f = f(x_1, \ldots x_m)$ is said to be symmetric, if, for every permutation $(i_1, \ldots i_m)$ of $(1, 2, \ldots m)$,

$$f(x_{i_1}, \ldots x_{i_m}) = f(x_1, \ldots x_m)$$

Definition A set of m elements will be called an m-tuple.

The collection of all m-tuples contained in any set A will be denoted by A(m). Thus, U(m) = {all m-tuples contained in U} and a similar definition holds for s(m).

Definition (symmetric kernel) For a population U, a symmetric function f and $I = \{i_1, \ldots i_m\} \in U(m)$, the function

$$f_I(\underline{y}) = f_{i_1, \ldots i_m}(\underline{y}) = f(y_{i_1}, \ldots y_{i_m}) \tag{3.1}$$

will be called a symmetric kernel.

We will abbreviate $f_I(\underline{y})$ to f_I or $f_{i_1 \ldots i_m}$ when there is no ambiguity.

Definition (L-function) A linear combination of symmetric kernels or, in brief an L-function is a function of the form

$$F(\underline{y}) = \sum_{I\varepsilon U(m)} C_I f_I(\underline{y}) \tag{3.2}$$

where C_I, $I\varepsilon$ $U(m)$ are constants and f_I, $I\varepsilon U(m)$ are symmetric kernels.

Definition (L-estimator) An estimator e of the form

$$e(s) = \sum_{I\varepsilon s(m)} b_{sI} f_I, \tag{3.3}$$

will be called an L-estimator, where $b_{\varepsilon I}$, $I\varepsilon s(m)$, $s\varepsilon S$ are constants.

Examples of L-functions

For $m = 1$ and $f(x) = x$, we get $F(\underline{y}) = \sum_{i\varepsilon U} C_i y_i$.

Thus all linear functions of the population vector, and in particular, the population total are L-functions.

For $m = 2$ and $f(x_1,x_2) = (x_1-x_2)^2$,

$$F(\underline{y}) = \sum_{\{i,j\}\varepsilon U(2)} C_{ij}(y_i-y_j)^2.$$

With $C_{ij} = 1/[N(N-1)]$, this gives the population mean square S^2. With $C_{ij} = \pi_i\pi_j - \pi_{ij}$, and $\underline{y}$ replaced by $\underline{z}$ (where $z_i = y_i/\pi_i$), we get $F(\underline{z}) = V_{HT}$ defined in (1.2).

In general, when C_I, $I\varepsilon U(m)$ are all equal, we get the U-statistics for a finite population. These include all population moments. If, in addition to the C_I's being equal, y_i's are vectors, we get the generalized U-statistics, which include the covariance and other product moments.

Linear independence of symmetric kernels

Definition (linear dependence) Symmetric kernels f_I $I \varepsilon U(m)$ are said to be linearly dependent, if there exist constants C_I, not all zero, such that

$$\sum_{I \varepsilon U(m)} C_I f_I(\underline{y}) = 0 \qquad \text{for all } \underline{y} \varepsilon \mathcal{Y}.$$

If no such constants exist, then the symmetric kernels are said to be linearly independent. Linear independence is desirable for unique representation of an L-function.

A simple example of linearly dependent symmetric kernels is provided by $f_{ij} = y_i + y_j$, $1 \leq i < j \leq 4$. It is trivial to see that $f_{12} + f_{34} - f_{13} - f_{24} = 0$.

However, in the above example, the linear dependence creates no problems, as any linear combination of f_{ij}'s can be reduced to a linear combination of y_i's, which are linearly independent. The following theorem, proved in Appendix III, shows that a similar reduction is possible for other linearly dependent kernels under a minor condition on the range of $\underline{y}$.

Theorem 3.1

Let f_I, $I \varepsilon U(m)$ be symmetric kernels of $m > 1$ arguments defined over $\mathcal{Y} = A^N$, where A is any set. If f_I, $I \varepsilon U(m)$ are linearly dependent, then there exists a symmetric function g of m-1 arguments, such that

$$f_I(\underline{y}) = \sum_{j \varepsilon I(m-1)} g(y_{j_1},\ldots y_{j_{m-1}}) \text{ for all } \underline{y}\varepsilon A^N, \tag{3.4}$$

where $J = \{j_1,\ldots j_{m-1}\}$ and $I(m-1) = \{$subsets of I containing m-1 elements$\}$.

For example, for m = 3, (3.4) is of the form $f(y_1,y_2,y_3) = g(y_1,y_2) + g(y_1,y_3) + g(y_2,y_3)$.

As a consequence of the theorem, the symmetric kernels composing an L-function can be assumed, without loss of generality, to be linearly independent. If they are not, they can be replaced by symmetric kernels of fewer arguments, for which (3.4) holds. Any linear combination of the original kernels f_I is also a linear combination of the new kernels g_J. The process can be repeated, if necessary, until we get linearly independent kernels. However, the following example shows that the theorem may not hold when the range of $\underline{y}$ is not a Cartesian product.

Example 3.1

Let N = 4, $\mathcal{Y}$ = {permutations of (1, 1, - 1, - 1)}, and $f_{ij} = y_i y_j$.

Evidently, $y_1y_2 - y_3y_4 = 0$ for all $\underline{y}\varepsilon\mathcal{Y}$. Thus f_{ij}'s are linearly dependent. But it can be quickly verified that no function g exists, such that $f_{ij} = g(y_i) + g(y_j)$ for all $\underline{y}\varepsilon\mathcal{Y}$.

Non-existence of UMVUE

Consider the problem of estimating the function F in (3.2) with f_I,

$I \varepsilon U(m)$ linearly independent. Let the sampling design be p, with (joint) inclusion probabilities $\pi_I = \Pr(I \varepsilon s(m))$, $I \varepsilon U(m)$. It is clear that F is estimable, if and only if $\pi_I \neq 0$ for all I, for which $C_I \neq 0$.

Let C_{Lp} be the class of L-estimators [defined in (3.3)], which are p-unbiased for F. The following theorem shows that no UMVUE of F exists in C_{Lp} for most designs of practical interest. This is a direct extension of Godambe's (1955) theorem, referred to in Chapter I. The exceptional class of designs in the present case is slightly wider, and is defined below.

For any sample s, define the set of "relevant m-tuples"

$$s_F = \{I \varepsilon s(m): \; C_I \neq 0\} \tag{3.5}$$

Let E be the class of designs p for which F is estimable and for any two samples with $p(s_1), p(s_2) > 0$, either $s_1(m)$ and $s_2(m)$ are disjoint, or $s_{1F} = s_{2F}$. (3.6)

Theorem 3.2

With C_{Lp} and E as defined above, and f_I, $I \varepsilon U(m)$ linearly independent,

(1) No UMVUE of F exists in C_{Lp}, if $p \not\varepsilon E$.

(2) If $p \varepsilon E$, the UMVUE of F is given by

$$e_*(s) = \sum_{I \varepsilon s(m)} C_I f_I / \pi_I \tag{3.7}$$

Proof An estimator $e(s) = \Sigma\, b_{sI} f_I$ is p-unbiased for F if and only if

$$\sum_{s \supset I} p(s) b_{sI} = C_I \text{ for all } I \varepsilon U(m) \tag{3.8}$$

Suppose $e_1(s) = \Sigma\ d_{sI} f_I$ is UMVUE of F in C_{Lp}. We will show $p \varepsilon E$, to prove the first part of the theorem.

Since e_1 is unbiased, (3.8) holds for $b_{sI} = d_{sI}$, $I \varepsilon s(m)$, $s \varepsilon S$. Further, these values minimize V(e) subject to (3.8), or equivalently they minimize

$$\Phi = \sum_{s \varepsilon S} p(s)[e(s)]^2 + \sum_{I \varepsilon U(m)} \lambda_I \left(\sum_{s \supset I} p_s b_{sI} - C_I \right),$$

where λ_I's are the Lagrangian multipliers. Hence, the partial derivatives of Φ with respect to b_{sI} must vanish at $b_{sI} = d_{sI}$, and we get

$$p(s) e_1(s) f_I + \lambda_I p(s) = 0,$$

or, for a sample s with p(s)>0,

$$e_1(s) f_I + \lambda_I = 0 \qquad I \varepsilon U(m), s \varepsilon S \tag{3.9}$$

If s_1 and s_2 are any two samples, such that $s_1(m)$ and $s_2(m)$ are not disjoint, let $I \varepsilon s_1(m)$ and $s_2(m)$. By (3.9),

$$e_1(s_1) = e_1(s_2) \tag{3.10}$$

Because of the assumption of linear independence, the coefficient of any f_J, $J \varepsilon U(m)$, must be the same on both sides of (3.10). Suppose, $J \varepsilon s_1(m)$, but $\not\varepsilon s_2(m)$. Then $d_{s_1} J = 0$.

If s_3 is any sample containing J, then applying (3.9) to (s_1,J) and (s_3,J), $e_1(s_1) = e_1(s_3)$ and hence $d_{s_3J} = 0$ for all s_3 containing J. Hence, by (3.8),

$$C_J = 0 \text{ for all } J\varepsilon s_1(m) \text{ and } \not\varepsilon\ s_2(m).$$

Similarly,

$$C_J = 0 \text{ for all } J\varepsilon s_2(m) \text{ and } \not\varepsilon\ s_1(m).$$

Hence, from (3.5), $s_{1F} = s_{2F}$. Since this holds whenever $s_1(m)$ and $s_2(m)$ are not disjoint, $p\varepsilon E$.

To prove the second part, it is sufficient to show that, for $p\varepsilon E$, (3.9) holds when $e_1(s)$ is replaced by $e_*(s)$ defined in (3.7), and λ_I's are suitably chosen.

Now, $e_*(s) = \sum_{I\varepsilon s_F} C_I f_I/\pi_I$.

For $p\varepsilon E$, s_F is the same for all s containing 'I', and hence the ℓ.h.s. of (3.9) becomes identical for all such samples. Thus, (3.9) is satisfied when λ_I is chosen to be this common value.

Remark The class E of exceptional designs does not depend on the particular symmetric function f, but depends on the set of 'I' for which C_I's are non-zero. When all C_I's are non-zero, E is the class of designs, such that $s_1(m)$ and $s_2(m)$ are disjoint for any two samples s_1 and s_2 with positive selection probabilities.

The following example shows that in the exceptional case, a UMVUE may exist in a non-trivial sense, i.e., several unbiased estimators may exist, unless all C_I's are non-zero.

Example 3.2

Let $N = 5$, $m = 2$ and $F = f_{12}+f_{34}+f_{35}+f_{45}$. Let p be the design for which

$$p(s) = 1/3, \text{ if } s = (1\ 2\ 3),\ (1\ 2\ 4) \text{ or } (3,\ 4,\ 5)$$
$$= 0 \text{ otherwise.}$$

Any unbiased estimator of F must be of the form

$$e_a(s) = (3/2+a)f_{12}, \text{ if } s = (1\ 2\ 3)$$

$$= (3/2-a)f_{12}, \text{ if } s = (1\ 2\ 4)$$

$$= 3(f_{34}+f_{35}+f_{45}), \text{ if } s = (3\ 4\ 5)$$

where 'a' is an arbitrary constant.

It is easy to see that $V(e_a)$ is uniformly minimum for $a = 0$.

Optimal Estimation under Superpopulation Models

We will now consider the problem of estimating an L-function F using the basic optimality criterion of minimization of expected MSE under a "prior" or "superpopulation model". Strictly speaking, the problem now becomes one of prediction, rather than estimation, but we will continue to refer to the predictors as estimators. We will restrict to the "natural" class of L-estimators. Theoretical justification for this in the case of variance estimation has been discussed in Chapter II.

In addition we may require any or none of the different kinds of unbiasedness defined in Chapter I. Accordingly, we get the following classes of estimators for given symmetric kernels and function to be estimated.

C_L = {all L-estimators},

C_{Lp} = {p-unbiased L-estimators (for a given design)},

$C_{L\xi}$ = {ξ-unbiased L-estimators (for a given model)}, and

$C_{Lp\xi}$ = {p ξ- unbiased L-estimators (for a given design and model)}.

Note that C_{Lp} and $C_{L\xi} \subset C_{Lp\xi} \subset C_L$.

The estimators minimizing E MSE within the four classes will be called the B_L - , B_{Lp} -, $B_{L\xi}$ - and $B_{Lp\xi}$ - estimators respectively, and will be also referred to as the "best" or "optimal" estimators in the respective classes.

The sampling design will be assumed to be of fixed sample size n. However, the results for C_L and $C_{L\xi}$ can be extended to variable sample size designs simply by replacing n by n(s).

A symmetric model

Let $\underline{Y} = (Y_1, \ldots Y_N)'$ have a distribution ξ, such that for all $I \varepsilon U(m)$,

$$\left.\begin{array}{l} E_\xi f_I(\underline{Y}) = M \\ \text{and for } I, J \varepsilon U(m) \text{ with } \#(IJ) = t, \\ E_\xi [f_I(\underline{Y}) f_J(\underline{Y})] = a_t, \end{array}\right\} \qquad (3.11)$$

where $\#(IJ)$ = number of elements in IJ (the intersection of I and J), and M, $a_o, \ldots a_m$ are unknown constants. However, in general, it will be necessary to know the ratios $a_o/a_m, \ldots a_{m-1}/a_m$ in order to get an optimal estimator. In certain special cases, this knowledge is not required.

We will further assume that either f is continuous, or ξ is discrete, so that the linear independence of the kernels f_I will imply the non-singularity of the matrix of their second moments $E_\xi(f_I f_J)$.

A sufficient condition for (3.11) to hold is that ξ be exchangeable. An extension of the above model is obtained when $\underline{Y}$ is replaced by a vector of transformed values, such as $Z_i = Y_i/x_i$.

<u>Equations for optimal coefficients</u>

The equations giving the coefficients for the B_L-estimator e are obtained by equating to zero the partial derivatives of $E_\xi MSE(e)$ with respect to b_{sI}. This gives,

$$E_\xi \; [p(s) \; \{e(s,\underline{Y}) - F(\underline{Y})\} \; f_I] = 0$$

or, $$H_{sI} = 0 \qquad I\varepsilon S(m), \; s\varepsilon S \tag{3.12}$$

where

$$H_{sI} = E_\xi \; [\{e(s,\underline{Y})\} - F(\underline{Y})_\xi \;\; f_I]$$

$$= \sum_{J\varepsilon s(m)} b_{sJ} a_{\#(IJ)} - \sum_{J\varepsilon U(m)} C_J a_{\#(IJ)} \tag{3.13}$$

To obtain the B_{Lp}-estimator, we minimize $E_\xi MSE(e)$ subject to the constraints

$$\sum_{s \ni I} p(s) b_{sI} = C_I \qquad I \varepsilon U(m) \tag{3.14}$$

The minimizing equations are $H_{sI} = \lambda_I$, (3.15)

λ_I's being the Lagrangian multipliers.

For the $B_{L\xi}$ - estimator, the constraints are

$$\sum_{I \varepsilon s(m)} b_{sI} = \sum_{I \varepsilon U(m)} C_I, \quad s \varepsilon S \tag{3.16}$$

and the minimizing equations are $H_{sI} = \lambda_s$ (3.17)

For the $B_{Lp\xi}$ - estimator, we get

$$\sum_{s \varepsilon S} p(s) \sum_{I \varepsilon s(m)} b_{sI} = \sum_{I \varepsilon U(m)} C_I \tag{3.18}$$

and

$$H_{sI} = \lambda \tag{3.19}$$

<u>Lemma 3.1</u>

For any $I \varepsilon s(m)$,

$$\sum_{J \varepsilon s(m)} a_{\#(IJ)} = \sum_{k=0}^{m} \binom{m}{k} \binom{n-m}{m-k} a_k$$

<u>Proof</u> Observe that the number of m-tuples $J \varepsilon s(m)$ for which $\#(IJ) = k$ is the coefficient of a_k on the r.h.s..

<u>Relations between the solutions</u>

Solutions for the $B_{L\xi}$ - and $B_{Lp\xi}$ - estimators can be easily obtained from the solution for the B_L-estimator as follows.

Let $b_{sI} = b^o_{sI}$, $I\varepsilon s(m)$, $s\varepsilon S$, be the solutions of $H_{sI} = 0$. Then the solutions of $H_{sI} = \lambda_s$ are $b^\xi_{sI} = b^o_{sI} + \mu_s$, where $\mu_s = \lambda_s/D$, and D is the sum of coefficients of b_{sJ}, $J\varepsilon s(m)$ in H_{sI}. D is given by lemma 3.1, and clearly does not depend on I or s. Putting $b_{sI} = b^\xi_{sI}$ in the constraints (3.16), we can solve for μ_s.

Similarly, a solution of (3.19) is $b^{p\xi}_{sI} = b^o_{sI} + \mu$, where μ is obtained using (3.18).

Optimal estimators for $n = m$

The solutions for $n = m$ can be easily shown to be as follows:

For the unrestricted class C_L, $b^o_{sI} = d_I$, where

$$d_I = E_\xi[F(\underline{Y})f_I(\underline{Y})]/a_m$$

The coefficients for the B_{Lp}-estimator are $b_{sI} = C_I/\pi_I$.

For the $B_{L\xi}$ - estimator, $b^\xi_{sI} = C$,

where $C = \sum_{U(m)} C_I$.

For the $B_{Lp\xi}$ - estimator, $b^{p\xi}_{sI} = C + d_I - \sum_{U(m)} \pi_J d_J$.

Optimal estimators for U-statistics

When $C_I = 1$ for all $I\varepsilon U(m)$, we get the following solutions, easily verified by substituting the coefficients in the respective equations.

Theorem 3.3

Under a model satisfying (3.11), the $B_{Lp\xi}$ - estimator of $F = \sum_{U(m)} f_I$ for any design p is

$$\bar{e}(s) = \binom{N}{m}/\binom{n}{m} \cdot \sum_{s(m)} f_I \tag{3.20}$$

<u>Corollary 3.1</u> The estimator $\bar{e}$ in (3.20) is also the $B_{L\xi}$ - estimator of F.

<u>Corollary 3.2</u> For a design p with $\pi_I = \binom{n}{m}/\binom{N}{m}$ for all $I \epsilon U(m)$, $\bar{e}$ is the B_{Lp} - estimator of F.

The corollaries follow by observing that $C_{L\xi}$ and C_{Lp} are subclasses of $C_{Lp\xi}$, $\bar{e}$ is ξ - unbiased, and is also p-unbiased when all π_I's are equal.

An explicit expression for the B_{Lp} - estimator for an arbitrary design is difficult to obtain.

<u>Theorem 3.4</u>

For a model ξ satisfying (3.11) with a_i/a_m, i = 0, 1,...m-1 known, the B_L-estimator of $F = \sum_{U(m)} f_I$ is

$$e_*(s) = \frac{\sum_{k=o}^{m} \binom{m}{k}\binom{N-m}{m-k} a_k}{\sum_{k=o}^{m} \binom{m}{k}\binom{n-m}{m-k} a_k} \cdot \sum_{s(m)} f_I \tag{3.21}$$

<u>Proof</u> The result follows using lemma 3.1 and a similar result with s(m) replaced by U(m).

Reduced Equations for the Coefficients

We will consider the system (3.12) giving the optimal coefficients in the unrestricted case. Other cases can be dealt with along similar lines with some additional work to eliminate the Lagrangian multipliers.

For each sample s, we have a system of $\binom{n}{m}$ linear equations in b_{sI}, $s\varepsilon S$. This is too large to be solved directly even for moderate n and m = 2. However, as explained below, the structure imposed by the symmetric model enables us to reduce the problem to a system of order (m+1), which can be solved without much difficulty. Readers familiar with Partially Balanced Incomplete Block Designs (PBIBD) may notice that this is analogous to the reduction of the order of the "reduced normal equations" of a PBIBD. (See e.g., Ogawa, 1974, p. 266-68).

Grouping the m-tuples

With respect to any m-tuple 'I', all m-tuples $J\varepsilon U(m)$ can be divided into (m+1) groups according to the number of elements common with 'I'.

Let $G_{Ik} = \{J\varepsilon U(m): \#(IJ) = k\}$ $\quad k = 0, 1\ldots m$

The corresponding groups for the sample m-tuples are,

$$G_{sIk} = G_{Ik} \cap s(m) \qquad k = 0, 1,\ldots m$$

With this notation (3.12) can be written as

$$\sum_{k=o}^{m} a_k \sum_{J\varepsilon G_{sIk}} b_{sJ} = w_I \tag{3.23}$$

where

$$w_I = \sum_{k=0}^{m} a_k \sum_{J\varepsilon G_{Ik}} C_J$$

Define the group totals

$$B_{sIk} = \sum_{J\varepsilon G_{sIk}} b_{sJ} \qquad k = 0,\ 1 \ldots m$$

and

$$W_{sIk} = \sum_{J\varepsilon G_{Ik}} w_I$$

Clearly, $G_{sIm} = G_{Im} = \{I\}$ and $B_{sIm} = b_{sI}$.

We will show that for any $r = 0, 1, \ldots m$, replacing 'I' by H in (3.23) and summing over $H\varepsilon G_{sIr}$ gives an equation in which the ℓ.h.s. is a linear combination of the group totals B_{sIk}, $k = 0, 1 \ldots m$. Thus, we get a system of (m+1) equations in as many unknowns and we can solve it for $B_{sIm} = b_{sI}$ in particular.

Lemma 3.2

$$\sum_{\ell=0}^{m} B_{sI\ell} \sum_{k=0}^{m} a_k \alpha(\ell,r,k) = W_{sIr}, \qquad r = 0, 1, \ldots m \tag{3.24}$$

where

$$\alpha(\ell,r,k) = \sum_{u=0}^{\ell} \binom{\ell}{u}\binom{m-\ell}{r-u}\binom{m-\ell}{k-u}\binom{n-2m+\ell}{m-r-k+u} \tag{3.25}$$

[We interprete $\binom{a}{b}$ as zero when a<b or a<0 or b<0.]

Proof Replacing 'I' by H in (3.23) and summing over $H\varepsilon G_{sIr}$ gives

$$\sum_{k=0}^{m} a_k \sum_{H\varepsilon G_{sIr}} \sum_{J\varepsilon G_{sHk}} b_{sJ} = W_{sIr}. \tag{3.26}$$

To prove (3.24), we must show that the coefficients of $a_k b_{sJ}$ on the left hand sides of (3.24) and (3.26) are equal. That is, for $J \varepsilon G_{sI\ell}$,

$$\alpha(\ell,r,k) = \text{coefficient of } b_{sJ} \text{ in}$$

$$\sum_{H \varepsilon G_{sIr}} \sum_{J \varepsilon G_{sHk}} b_{sJ} = R \text{ (say)}.$$

Now, the coefficient of b_{sJ} in R is the number of ways of choosing $H \varepsilon s(m)$, such that $\#(HI) = r$ and $\#(HJ) = k$.

For $J \varepsilon G_{sI\ell}$, $\#(IJ) = \ell$. If H contains u elements from IJ, it must contain r-u, k-u and m-r-k+u elements respectively from $I\bar{J}$, $\bar{I}J$ and $\bar{I}\bar{J}$, where $\bar{I} = s-I$. The number of ways of doing so is the general term on r.h.s. of (3.25), and the total number of ways is the sum over u. Hence the result.

Solutions for m = 1

Let $F(\underline{y}) = \sum_{i \varepsilon U} C_i y_i$, $E_\xi(Y_i) = M$, $E_\xi(Y_i^2) = a_1$ and $E_\xi(Y_i Y_j) = a_o$ for $i \neq j$. The optimal linear estimators of F are given below without proof.

<u>Theorem 3.5</u>

(1) The B_L - estimator of F, when a_o/a_1 is known, is

$$e_1(s) = \sum_{i \varepsilon s} C_i y_i + q \cdot \sum_{i \not\varepsilon s} C_i \cdot \bar{y}_s \tag{3.27}$$

where $q = na_o/[a_1+(n-1)a_o]$

(2) The $B_{L\xi}$ - estimator of F is

$$e_{1\xi}(s) = \sum_{i\in s} C_i y_i + \sum_{i\notin s} C_i \cdot \bar{y}_s \tag{3.28}$$

(3) When a_o/a_1 is known, the $B_{Lp\xi}$ - estimator of F for a design p with inclusion probabilities π_i, $i\varepsilon U$, is

$$e_{1p\xi}(s) = \sum_{i\in s} C_i y_i = [q \sum_{i\notin s} C_i + (1-q) \sum_{i=1}^{N} (1-\pi_i)C_i]\ \bar{y}_s \tag{3.29}$$

(4) For a design p with $\pi_i = n/N$ for all $i\varepsilon U$, the B_{Lp} - estimator, when a_o/a_1 is known, is

$$e_{1p}(s) = d\cdot[q \sum_{i\in s} C_i y_i + (\frac{n-1}{N-1}C - \sum_{i\in s} C_i) \cdot \bar{y}_s] \tag{3.30}$$

where $C = \sum_{i\in U} C_i$ and $d = N/[q - (N-n)/(N-1)]$

(5) If $C_i = \pi_i$ for all $i\varepsilon U$, then the B_{Lp} - estimator is

$$e_{1p}(s) = \sum_{i\in s} y_i \tag{3.31}$$

<u>Remarks</u> (a) The $B_{L\xi}$ - estimator can be obtained without the knowledge of a_o/a_1. Other optimal estimators require this knowledge, except in the special case $C_i = \pi_i$, when the B_{Lp} - estimator can be also obtained without the knowledge of this ratio.

(b) In e_1, $e_{1\xi}$ and $e_{1p\xi}$, the first part represents the observed terms of F and the second part is the predictor of the rest.

(c) When $C_i = 1$ for all i, $e_{1\xi}$, $e_{1p\xi}$ and e_{1p} (when all π_i are equal) reduce to the "usual" estimator $N\bar{y}_s$. The estimator e_1 shrinks this by a factor approaching unity when either the absolute sample size n becomes large or the sampling fraction n/N approaches unity.

(d) If $Z_i = y_i/\pi_i$, $i\varepsilon U$ are symmetric under ξ[in the sense of (3.11)], then the B_{Lp} - estimator of the population total $\sum_U y_i = \sum_U \pi_i z_i$ is given by part (5) of the theorem as $\sum_{i\varepsilon s} z_i$, which is the Horvitz-Thompson estimator.

Solutions for m = 2

Although lemma 3.2 gives a system of equations of small order, we will use an alternative procedure which goes a step further and gives the explicit solution of (3.12) for m = 2.

A useful operator

For each element x_{ij} of a two-way array we define an operator Σ_1 as follows.

$$\Sigma_1 x_{ij} = \sum_k (x_{ik} + x_{kj}) - x_{ii} - x_{jj}$$

$$= \sum_{\substack{k \\ \neq i,j}} (x_{ik} + x_{kj}) + x_{ij} + x_{ji}$$

In words, Σ_1 gives the row-sum plus the column-sum, excluding the diagonal terms x_{ii}, x_{jj}. Note that $\Sigma_1 x_{ij}$ is defined even when the diagonal terms are not defined.

For m>2, one can analogously define the operator Σ_k ($1 \le k \le m$) giving the sum of "marginal sums over k subscripts at a time", excluding the terms in which some subscript is repeated. Such operators can be useful for obtaining explicit solutions for m>2, but this will not be attempted here.

Lemma 3.3

Let u_i, $i\varepsilon s$ be such that $\sum_{i\varepsilon s} u_i = 0$, and let $x_{ij} = u_i + u_j$. Then,

$$\Sigma_1 x_{ij} = (n-2)x_{ij}$$

Proof $\Sigma_1 x_{ij} = \sum_{k\varepsilon s} (u_i+u_k+u_k+u_j) - 2u_i - 2u_j$

$$= n(u_i+u_j) + 2\sum_k u_k - 2(u_i+u_j)$$

The desired result follows by observing that the middle term vanishes by hypothesis.

Notation

$$\tilde{s} = U - s, \text{ the complement of } s. \qquad (3.32)$$

$U(2) = \{\{k,\ell\}: k, \ell\varepsilon U\}$ and $s(2)$, $\tilde{s}(2)$ defined similarly.

$$C_{\tilde{s}i} = \sum_{j\varepsilon\tilde{s}} C_{ij}$$

$$C_s = \sum_{s(2)} C_{ij}$$

$$C_{s\tilde{s}} = \sum_{\substack{i\varepsilon s\\ k\notin s}} C_{ik} = \sum_{i\varepsilon s} C_{\tilde{s}i}$$

$$C_{\tilde{s}} = \sum_{\tilde{s}(2)} C_{k\ell}$$

$$C = \sum_{U(2)} C_{ij} = C_s + C_{s\tilde{s}} + C_{\tilde{s}}$$

$$C_{s*} = C - C_s = C_{\tilde{s}} + C_{s\tilde{s}}$$

$$D_{si} = C_{\tilde{s}i} - C_{s\tilde{s}}/n$$

$$d_i = a_1 - a_o$$

$$d_2 = a_2 - 2a_1 + a_o$$

$$t_o = \binom{n}{2} a_o + 2(n-1)\, d_1 + d_2$$

$$t_1 = (n-2)\, d_1 + d_2$$

<u>Explicit solution for m = 2</u>

The equations (3.12) are:

$$\sum_{J\varepsilon s(2)} b_{sJ} a_{\#(IJ)} - \sum_{J\varepsilon U(2)} C_J a_{\#(IJ)} = 0$$

Defining $g_J = b_{sJ} - C_J$ (3.33)

$$\sum_{J\varepsilon s(2)} g_J a_{\#(IJ)} = \sum_{J\not\varepsilon s(2)} C_j a_{\#(IJ)}$$

Switching from set subscripts to element subscripts using $I = \{i,J\}$ and

$J = \{i,k\}$ or $\{j,k\}$ when $\#(IJ) = 1$

$= \{k,\ell\}$ when $\#(IJ) = 0$

$$a_o \sum_{k,\ell\varepsilon s-I} g_{k\ell} + a_1 \sum_{k\varepsilon s-I} (g_{ik}+g_{kj}) + a_2 g_{ij}$$

$$= a_o \sum_{\substack{k\varepsilon s-I \\ \ell\notin s}} C_{k\ell} + a_o \sum_{k,\ell\notin s} C_{k\ell} + a_1 \sum_{k\notin s} (C_{ik}+C_{kj})$$

Defining $\bar{g} = \sum_{s(2)} g_{ij}/\binom{n}{2}$, and using the notation (3.32)

$$a_o[\binom{n}{2}\bar{g} - \Sigma_1 g_{ij} + g_{ij}] + a_1(\Sigma_1 g_{ij} - 2g_{ij}) + a_2 g_{ij}$$

$$= a_o(C_{s*} - C_{i\tilde{s}} - C_{j\tilde{s}}) + a_1(C_{i\tilde{s}} + C_{j\tilde{s}})$$

Putting $d_1 = a_1 - a_o$, $d_2 = a_2 - 2a_1 + a_o$,

$$a_o\binom{n}{2}\bar{g} + d_1 \Sigma_1 g_{ij} + d_2 g_{ij} = a_o C_{s*} + d_1(C_{i\tilde{s}} + C_{j\tilde{s}}) \tag{3.34}$$

Averaging (3.34) over $\{i,j\} \varepsilon s(2)$, and noting that $\Sigma_1 g_{ij}$ contains $2(n-1)$ terms,

$$a_o\binom{n}{2}\bar{g} + 2d_1(n-1)\bar{g} + d_2\bar{g} = a_o C_{s*} + 2d_1 C_{s\tilde{s}}/n \tag{3.35}$$

Subtracting (3.35) from (3.34) and putting $h_{ij} = g_{ij} - \bar{g}$ (3.36)

and $D_{si} = C_{\tilde{s}i} - C_{s\tilde{s}}/n$,

$$d_1 \Sigma_1 h_{ij} + d_2 h_{ij} = d_1(D_{si} + D_{sj}). \tag{3.37}$$

Let $h_{i.} = \sum_{\substack{j\varepsilon s \\ j\neq i}} h_{ij}$, $w_{ij} = \Sigma_1 h_{ij} = h_{i.} + h_{j.}$, and $x_{ij} = D_{si} + D_{sj}$

Since

$$\sum_{i\varepsilon s} h_{i.} = \sum_{i\varepsilon s} D_{si} = 0,$$

and

$$\Sigma_1 w_{ij} = (n-2)w_{ij} \tag{3.38}$$

$$\Sigma_1 x_{ij} = (n-2)x_{ij} \text{ (by lemma 3.3)} \tag{3.39}$$

By (3.37), h_{ij} can be expressed as a linear combination of w_{ij} and x_{ij}. Hence, by (3.38) and (3.39),

$$\Sigma_1 h_{ij} = (n-2)h_{ij} \tag{3.40}$$

Applying the operator Σ_1 to equation (3.37), and using (3.38) and (3.39),

$$d_1\cdot(n-2)\Sigma_1 h_{ij} + d_2\Sigma_1 h_{ij} = (n-2)\ d_1\ (D_{si}+D_{sj}) \tag{3.41}$$

From (3.40) and (3.41),

$$h_{ij} = d_1\ (D_{si}+D_{sj})/t, \tag{3.42}$$

where

$$t_1 = (n-2)d_1 + d_2$$

From (3.33) and (3.36),

$$b_{sij} = C_{ij} + g_{ij} = C_{ij} + h_{ij} + \bar{g} \tag{3.43}$$

where h_{ij} is given by (3.42), and

$$\bar{g} = (a_o C_{s*} + 2d_1 C_{s\tilde{s}}/n)/t_o, \quad \text{[by (3.35)]}$$

$$t_o = \binom{n}{2} a_o + 2(n-1)d_1 + d_2$$

Theorem 3.6

For a model ξ satisfying (3.11) and the function $F_2 = \sum_{U(2)} C_{ij} f_{ij}$,

(1) The B_L - estimator is $e_2(s) = \sum_{s(2)} b_{sij} f_{ij}$, where

$$b_{sij} = C_{ij} + d_1(D_{si}+D_{si})/t_1 + [a_o C_{s*} + 2d_1 C_{s\tilde{s}}/n]/t_o \tag{3.44}$$

(2) The $B_{L\xi}$ estimator is $e_{2\xi}(s) = \sum_{s(2)} b^{\xi}_{sij} f_{ij}$, where

$$b^{\xi}_{sij} = C_{ij} + d_1(D_{si}+D_{si})/t_1 + C_{s*}/\binom{n}{2} \tag{3.45}$$

(3) The $B_{Lp\xi}$ estimator is $e_{2p\xi}(s) = \sum_{s(2)} b^{p\xi}_{sij} f_{ij}$, where

$$b^{p\xi}_{sij} = b_{sij} + \sum_{U(2)} C_{ij}[2(n-1)d_1(_1-\pi_i) + d_2(1-\pi_{ij})]/[\binom{n}{2}t_o] \tag{3.46}$$

Remarks Unlike in the case m = 1, knowledge of the ratios a_i/a_m is required for m = 2, to find the optimal estimator even under ξ - unbiasedness, except in certain special cases. One such case is when all C_{ij} are equal, and has been considered in theorem 3.3. When n = N, of course, all of the above reduce to F. When n = 2, $b^{\xi}_{sij} = C$, which obviously does not depend on a_o/a_2 and a_1/a_2. The same is true for the p-unbiased case where the solution is $b^p_{ij} = C_{ij}/\pi_{ij}$. For the unrestricted case,

$$b_{sij} = C_{ij} + (C - C_{ij}) \cdot a_1/a_2 \tag{3.47}$$

When ξ is exchangeable and $f_{ij} = (y_i - y_j)^2$, it is easy to show that $a_1/a_2 = 1/2$, and thus b_{sij} is known without any further assumptions.

Random Permutation Model

Definition The random permutation model is a model ξ_Y, under which the values assumed by the random vector $\underline{Y}$ are the permutations of a fixed vector $\underline{y}_o$, with probability 1/N! each.

Clearly, ξ_Y is exchangeable and hence satisfies the conditions (3.11). This model has been used by Kempthorne (1969), Godambe and Thompson (1971) and others to formalize the belief of "no association" between the unit labels and the y-values.

Population moments under ξ_Y

Let $\mu_1{}' = E_{\xi_Y}(Y_1)$

$$= \sum_{i=1}^{N} Y_i/N$$

or $$\mu_1{}' = \sum_{i=1}^{N} y_i/N = \bar{y}$$

since ℓ.h.s. is the same for all values $\underline{y}$ of $\underline{Y}$ under ξ_Y.

Similarly, all moments of Y_i under ξ_Y are the moments of the realized finite population.

In particular, $\mu_2{}' = \bar{y}^2 + (N-1)/N \cdot S^2$ (3.48)

where

$$S^2 = (N-1)^{-1} \sum_{i=1}^{N} (y_i - \bar{y})^2.$$

and

$$\mu_{11}{}' = E_{\xi_Y}(Y_1 Y_2) = \bar{y}^2 - S^2/N \qquad (3.49)$$

Similarly, we can define the central moments

$$\mu_r = E_{\xi_Y}(Y_1 - \bar{y})^r$$

$$\mu_{rs} = E_{\xi_Y}[(Y_i - \bar{y}_1)^r (Y_2 - \bar{y})^s]$$

etc..

We have the following relationships:

$$\mu_{rs} = [N(N-1)]^{-1} \left[\sum_{i \neq j}^{N} (y_i - \bar{y})^r (y_j - \bar{y})^s\right]$$

$$= [N(N-1)]^{-1} \left[\sum_i (y_i - \bar{y})^r \sum_j (y_j - \bar{y})^s - \sum_i (y_i \bar{y})^{r+s}\right]$$

$$= (N-1)^{-1} [N\mu_r \mu_s - \mu_{r+s}]$$

In particular, $\mu_{22} = (N-1)^{-1} (N\mu_2{}^2 - \mu_4)$ (3.50)

$$\mu_{31} = -\mu_4/(N-1) \tag{3.51}$$

since $\mu_1 = 0$.

Similarly, defining $N_{(k)} = N(N-1)\ldots(N-k+1)$,

$$\mu_{211} = [N_{(3)}]^{-1}\,[N_{(2)}\mu_{21}\cdot N\mu_1 - N_{(2)}\mu_{22} - N_{(2)}\mu_{31}]$$

$$= -\,(\mu_{22}+\mu_{31})/(N-2) \tag{3.52}$$

and

$$\mu_{1111} = -\,3\mu_{211}/(N-3) \tag{3.53}$$

Optimal linear estimators under ξ_Y,

For the model ξ_Y,

$$a_1/a_o = \mu_2'/\mu_{11}'$$

$$= 1 + S^2/(\bar{y}^2 - S^2/N)$$

$$= 1 + C_y^{\,2}/(1-C_y^{\,2}/N)$$

where $C_y = S/\bar{y}$, the coefficient of variation of y.

Substituting for a_1/a_o in (3.27), we get the following:

When C_y is known, the B_L - estimator of $F_1 = \sum_{i\varepsilon U} C_i y_i$ is

$$e_1(s) = \sum_{i\varepsilon s} C_i y_i + \frac{1-C_y^{\,2}/N}{1+C^2_{\bar{y}_s}} \cdot \Big(\sum_{i\not\varepsilon s} C_i\Big)\,\bar{y}_s \tag{3.54}$$

where

$$C^2_{\bar{y}_s} = (\frac{1}{n} - \frac{1}{N})\, C^2_y, \text{ the squared coefficient of variation of } \bar{y}_s.$$

The $B_{L\xi}$ - estimator does not depend on a_1/a_o and is given by (3.28). The $B_{Lp\xi}$ - estimator and the B_{Lp} - estimator in the case $\pi_i = n/N$ can be obtained from (3.29) and (3.30) by substituting for a_1/a_o.

For the population mean $\bar{y} = \sum_{i\varepsilon U} y_i/N$, the B_L - estimator is

$$e_{1*}(s) = \bar{y}_s/\ (1+C^2_{\bar{y}_s}) \qquad (3.55)$$

Remark The estimator e_{1*} can be obtained without using a super-population model as the estimator with minimum MSE among the multiples of the sample mean for simple random sampling. (A similar remark applies for the estimator given by theorem 3.4 for any U-statistic). It is a special case of the following more general result.

Let e be an unbiased estimator of a parameter θ for any population (not necessarily finite), with known coefficient of variation C_e. Then, the estimator $e/(1 + C_e^2)$ has the smallest MSE among the multiples of e.

A comparison of e with "shrinkage estimators" based on approximate knowledge of the coefficient of variation has been made by Searls (1964). He has further pointed out (Searls, 1967) that the mean of a sample of independent observations is improved upon by using any guess of C_y between zero and $\sqrt{2}\cdot C_y$, the upper limit being a large sample approximation. In particular, using any under-estimate of C_e gives an improvement. It follows that the sample mean is inadmissible as an estimator of the population mean when the parameter space is such that $C_y \geq k$, a positive constant.

Optimal Variance Estimation Under the Random Permutation Model

Consider $\phi_{ij}(\underline{y}) = (y_i - y_j)^2$. We have

$$a_2 = E_{\xi_Y}[\phi_{ij}(\underline{Y})]^2 = E_{\xi_Y}(Y_i - Y_j)^4$$

$$= 2\mu_4 - 8\mu_{31} + 6\mu_{22}$$

$$a_1 = E_{\xi_Y}[\phi_{ij}(\underline{Y})\phi_{ik}(\underline{Y})] \qquad \text{(i, j, k distinct)}$$

$$= E_{\xi_Y}[(Y_i - Y_j)^2(Y_i - Y_k)^2]$$

$$= \mu_4 - 4\mu_{31} + 3\mu_{22}$$

or
$$a_1 = a_2/2 \tag{3.56}$$

$$a_o = E_{\xi_Y}[(Y_i - Y_j)^2(Y_k - Y_\ell)^2] \qquad \text{(i, j, k,}\ell\text{ distinct)}$$

$$= 4\mu_{22} - 8\mu_{211} + 4\mu_{1111}$$

Using (3.50) - (3.53) and simplifying,

$$a_o = \frac{4N\mu_2^2}{(N-1)_{(3)}}[N^2 - (N-1)(\beta_2 + 3)] \tag{3.57}$$

and

$$a_1 = N\mu_2^2\,(\beta_2 + 3)/(N-1) \tag{3.58}$$

where

$$\beta_2 = \mu_4/\mu_2^2$$

Optimal estimator of S^2 when β_2 is known

Using theorem 3.4 with $m = 2$, $f_I = \phi_{ij}$ and the values of a_o, a_1, a_2 given by (3.56) - (3.58), we get the following B_L - estimator for the population mean square S^2.

$$v(s) = \frac{(N-2)(N-3)}{N(N-1)} \cdot \frac{n(n-1)s^2}{n(n+1)+(n-1)(\beta_2-3)+R} \tag{3.59}$$

where

$$R = (\beta_2+3)[n(n+1)/N^2 - (n^2+1)/N]$$

and

$$s^2 = \sum_{i\varepsilon s} (y_i-y_s)^2/(n-1) = \sum_{s(2)} \phi_{ij}/[n(n-1)]$$

For large N,

$$v(s) = n(n-1)s^2/[n(n+1) + (n-1)(\beta_2-3)] \tag{3.60}$$

For large N, and $\beta_2 \doteq 3$,

$$v(s) \doteq \sum_{i\varepsilon s} (y_i-\bar{y}_s)^2/(n+1) \tag{3.61}$$

Remark (3.60) gives the minimum MSE estimator of variance among the multiples of s^2, for a sample of independent observations from any population. The special case for a normal population ($\beta_2 = 3$) is fairly well-known.

When β_2 is not exactly known, substituting any under-estimate in (3.59) or (3.60) gives an estimator with smaller E_{ξ_Y} MSE than s^2, (smaller MSE for all $\underset{\sim}{y}$, in case of simple random sampling). In fact, we can show that for large N, the estimator obtained by substituting $\beta_2 = \tilde{\beta}_2$ in (3.60) is better than s^2, if

$$\frac{n-3}{n-1} < \tilde{\beta}_2 < \frac{2\beta_2-1+4/(n-1)^2}{(n-3)/(n-1) - (\beta_2-3)/n} \tag{3.62}$$

and in particular, if

$$1 \le \tilde{\beta}_2 \le 2\beta_2 - 1 \tag{3.63}$$

or, if $(\tilde{\beta}_2+1)/2 \le \beta_2 < \infty$

The lower limit in (3.62) is, of course, an impossible value of β_2. When one has "absolutely no idea" about β_2, one may still use the lowest possible value $\beta_2=1$, which gives the estimator

$$u_o(s) = \sum_{i\varepsilon s} (y_i-\bar{y}_s)^2/(n-1+\tfrac{2}{n}) \tag{3.64}$$

Of course, $v_o(s)$ is not substantially different from s^2, except for very small sample sizes and the improvement may be neglegible. However, this shows the inadmissibility of s^2 as an estimator of variance for any population. This is quite elementary, but does not seem to be very well known.

Optimal estimators of V_{HT}

Consider the random permutation model ξ_Z, under which $\underline{Z} = (Z_1, \ldots Z_N)'$, where $Z_i = Y_i/\pi_i$, assumes values which are permutations of a fixed unknown vector, with probability 1/N! each. This model may be interpreted as meaning that the "relevant information contained in the labels is exhausted" after transforming from $\underline{y}$ to $\underline{z}$, and in particular, there is no association between z_i's and π_i's.

From Godambe and Thompson (1973), we know that under ξ_Z, the optimal p-unbiased estimator of the population total is the Horwitz-Thomson estimator. Optimal estimators for its variance can be obtained by treating it as a special case of the function V_* considered below, with $C_{ij} = \pi_i\pi_j - \pi_{ij}$, $\phi_{ij}(\underline{y})$ replaced by $\phi_{ij}(\underline{z})$, and β_2 replaced by β_{2z}, defined analogously.

Let

$$V_*(\underline{y}) = \sum_{U(2)} C_{ij}\phi_{ij}(\underline{y}), \tag{3.65}$$

where

$$\phi_{ij}(\underline{y}) = (y_i - y_j)^2.$$

Coefficients giving the optimal estimators of V_* under the random permutation model are obtained by substituting for a_o, a_1 and a_2 from (3.56) - (3.58) into (3.44) - (3.46).

For large N and $\beta_2 \doteq 3$, we get the following approximations with a relative error of order N^{-1}. The expressions are exact for $\beta_2 \doteq 3(N-1)/(N+1)$.

$$b_{sij} = C_{ij} + (C_{\tilde{s}i}+C_{\tilde{s}j})/n + 2C_{\tilde{s}}/[n(n+1)] \tag{3.66}$$

$$b^{\xi}_{sij} = C_{ij} + (C_{\tilde{s}i}+C_{\tilde{s}j})/n + 2C_{\tilde{s}}/[n(n-1)]$$

$$+ 2C_{s\tilde{s}}/[n^2(n-1)] \tag{3.67}$$

$$b^{p\xi}_{sij} = b_{sij} + K,$$

where

$$K = 4 \sum_{U(2)} C_{ij} \cdot [(1-\pi_i) + (1-\pi_{ij})/(n-1)]/[n^2(n+1)] \tag{3.68}$$

Hence, the B_L - estimator is, $v(s) = \sum_{s(2)} b_{sij}\phi_{ij}$

$$= \sum_{s(2)} C_{ij}\phi_{ij} + \sum_{i\varepsilon s} C_{\tilde{s}i}\bar{\phi}_{is} + \frac{n}{n+1} C_{\tilde{s}}\bar{\phi}_s \tag{3.69}$$

where

$$\bar{\phi}_{is} = \sum_{j\varepsilon s} \phi_{ij}/n \text{ and } \bar{\phi}_s = \sum_{i\varepsilon s} \phi_{is}/n$$

The $B_{L\xi}$ - estimator is,

$$v_{\xi}(s) = \sum_{s(2)} C_{ij}\phi_{ij} + \sum_{i\varepsilon s} C_{\tilde{s}i}\bar{\phi}_{is} + (nC_{\tilde{s}}+C_{s\tilde{s}})\bar{\phi}_s/(n-1) \tag{3.70}$$

and the $B_{Lp\xi}$ - estimator is,

$$v_{p\xi}(s) = v(s) + Kn^2\bar{\phi}_s/2 \tag{3.71}$$

where K is defined in (3.68).

For $C_{ij} = \pi_i\pi_j - \pi_{ij}$, the following relationships facilitate the computations:

$$C_{\tilde{s}i} = \pi_i(1-\pi_i) - \sum_{\substack{j\varepsilon s \\ j\neq i}} C_{ij}$$

$$C_{\tilde{s}} = \frac{1}{2} \sum_{i\varepsilon U} \pi_i(1-\pi_i) - C_{s\tilde{s}} - C_s$$

$$= \frac{1}{2}\,(n - \sum_U \pi_i^2) - C_{s\tilde{s}} - C_s$$

<u>A desirable property of the optimal estimators</u>

We have seen in example 2.7 that the Yates-Grundy estimator may take values known to be impossible <u>given the data</u>. We will show that the estimators given by (3.69) - (3.71) overcome this drawback. We will use the "Intermediate Value Theorem" for continuous functions.

"If $h(\underline{x})$ is a continuous function defined on a convex set Q; $h(\underline{x}_1) = M_1$, $h(\underline{x}_2) = M_2$) for $\underline{x}_1, \underline{x}_2 \varepsilon Q$; and $M_1 \leq M \leq M_2$, then there exists $\underline{x}_3 \varepsilon Q$ such that $f(\underline{x}_3) = M$."

It follows that if $\underline{X}$ is a random variable (random vector) with a discrete distribution η on Q, then the function h attains the value $E_\eta h(\underline{X})$ at some point in Q.

<u>Theorem 3.7</u>

Let V, v, v_ξ, $v_{\,\xi}$ be as defined in (3.65) and (3.69) - (3.71), and let $= [a, b]^N$. Let s and y_i, iεs be given.

(1) For suitable choice of y_i, iεs, $v(s, \underline{y}) = V_*(\underline{y})$.

(2) A similar result holds for v_ξ, if the interval [a,b] is infinite at one or both ends.

(3) A similar result also holds for $v_{p\xi}$, if [a,b] is infinite at one or both ends, and V_* is non-negative definite and depends on at least one y_i, $i \notin s$.

Proof Consider a discrete distribution η of $\underline{Y} = (Y_1, \ldots Y_N)'$, such that

$$Y_i = y_i \text{ with probability one for } i \in s,$$

$$E_\eta(Y_k) = \bar{y}_s,$$

$$V_\eta(Y_k) = \sigma^2,$$

and

$$\mathrm{Cov}_\eta(Y_k, Y_\ell) = \rho\sigma^2 \quad \text{for } k, \ell \notin s$$

where ρ and σ^2 are to be suitably chosen.

For $i, j \in s$ and $k, \ell \notin s$,

$$E_\eta[\phi_{ij}(\underline{Y})] = \phi_{ij}(\underline{y})$$

$$E_\eta[\phi_{ik}(\underline{Y})] = (y_i - \bar{y}_s)^2 + \sigma^2$$

$$= \bar{\phi}_{is} - \bar{\phi}_s/2+\sigma^2 \text{ (using lemma 2.1)}$$

and

$$E_\eta[\phi_{k\ell}(\underline{Y})] = 2\sigma^2(1-\rho)$$

Hence,

$$E_\eta[V_*(\underline{Y})] = \sum_{s(2)} C_{ij}\phi_{ij} + \sum_{\substack{i\varepsilon s\\ k\notin s}} C_{ik}(\bar{\phi}_{is}-\bar{\phi}_s/2 + \sigma^2)$$

$$+ \sum_{\tilde{s}(2)} C_{k\ell}.2\sigma^2(1-\rho)$$

$$= \sum_{s(2)} C_{ij}\phi_{ij} + \sum_{i\varepsilon s} C_{\tilde{s}i}\bar{\phi}_{is} + (\sigma^2-\bar{\phi}_s/2)C_{s\tilde{s}}$$

$$+ 2\sigma^2(1-\rho)C_{\tilde{s}} \tag{3.72}$$

For $\sigma^2 = \phi_s/2$ and $\rho = 1/(n+1)$, the r.h.s. reduces to $v(s,\underline{y})$. That such a distribution can be defined on $[a, b]^N$, can be seen from the following. Assume $s = \{1, 2,...n\}$, and consider the distributions η_1 and η_2, such that under η_1,

$$\Pr [(Y_{n+1},...Y_N) = (y_i,...y_i)] = \frac{1}{n},\ i\varepsilon s,$$

and under η_2, $Y_{n+1},...Y_N$ are independent, with

$$\Pr (Y_k = y_i) = \frac{1}{n},\ i\varepsilon s,\ k\notin s.$$

Under both η_1 and η_2, $\sigma^2 = \frac{1}{n} \sum_{i\varepsilon s} (y_i-\bar{y}_s)^2 = \phi_s/2$. Under η_1, $\rho = 1$ and under η_2, $\rho = 0$. Hence a suitable mixture of η_1 and η_2 gives the desired distribution. Under both η_1 and η_2, Y_k, $k\notin s$ take values in $\{y_i,\ i\varepsilon s\}$ and hence in $[a, b]$. It follows that the same holds for the mixture.

For $\sigma^2 = (n+1)\bar{\phi}_s/[2(n-1)]$ and $\rho = 1/(n+1)$, we get $E_\eta[V_*(\underline{Y}))] = v_\xi(s,\underline{y})$, which proves the second part.

To prove the last part, observe that when $V_*(\underline{y})$ is NND, and depends on at least one unobserved y_i, then it can be made to take arbitrarily large values by choosing y_i, i$\not\in$s sufficiently extreme. Since $V_*(\underline{y})$ attains the value $v(s, \underline{y})$, it can also take any larger value. It is thus sufficient to show that $v(s,\underline{y}) \le v_{p\xi}(s,\underline{y})$. Note that $w \cdot v(s,\underline{y})$ is an L-function.

By the optimality of v in C_L,

$$E_{\xi_Y} [w\cdot v(s,\underline{y}) - V_*(\underline{y})]^2 \ge E_{\xi_Y} [v(s,y) - V_*(\underline{Y})]^2$$

or

$$w^2 . V_{\xi_Y} [v(s,\underline{y})] \ge V_{\xi_Y} [v(s,\underline{y})] + (\text{bias})^2$$

Hence $w \ge 1$, which by (3.73) gives

$E_\xi [v(s,\underline{y}) - V_*(\underline{Y})] \le 0$. Since s is arbitrary,

$$\sum_{s\varepsilon S} p(s)\ E_\xi\ [v(s,\underline{Y}) - V_*(s,\underline{Y})] \le 0, \tag{3.74}$$

But

$$\sum_{s\varepsilon S} p(s)\ E_\xi\ [v_{p_\xi}(s,\underline{Y}) - V_*(s,\underline{Y})] = 0, \tag{3.75}$$

and

$v_{p_\xi}(s,\underline{Y}) = v(s,\underline{Y}) + K'\ \bar{\phi}_s$, where K' is a constant.

By (3.74) and (3.75), $K' \ge 0$ and the result follows.

CHAPTER IV. NUMERICAL COMPARISONS

Populations, Design and Estimators

Populations

An empirical study was made to compare several estimators of V_{HT}, the variance of the Horvitz-Thompson estimator (e_{HT}) of the population total. Ten natural populations listed in Table 4.1 were used. Most of these are random samples from real populations. The populations used were generally those for which e_{HT} was expected to perform reasonably well, i.e., where an auxiliary variable x, approximately proportional to y, was available with only weak or moderate trends in the departure from proportionality. However, for the purpose of illustration, population 3, containing at least one "wild" value of y/x was also included. To highlight the effect of the extreme values, population 4 was obtained by deleting two units from population 3, corresponding to the smallest and the largest values of y/x. Population 5 also contains a rather extreme value, producing interesting effects, which are discussed later.

Sampling design

The rejective sampling scheme of Sampford (1967) was used to draw the samples, with inclusion probabilities π_i proportional to x_i. Two hundred independent samples were drawn from each population, for each of the sample sizes 3, 5 and 10, the last one being restricted to the first 6 populations. Sample size 10 was not used for populations of less than 20 units, or where $\pi_i \propto x_i$ would have required some unit

Table 4.1. Populations used in the study

Pop. No.	Source	y	x	N	β_{2z}
1	Hanurav (1967, p.386)	1960 population	1950 population	20	9.0
2	Yates (1960, p.163)	number of absentees	total no. of persons	43	3.2
3	Yates (1960, p.159)	volume of timber	eye-estimate	25	19.9
4	Subset of 3 (see text)			23	2.6
5	Sukhatme and Sukhatme (1970, p.183)	1937 area under wheat	1936 area under wheat	34	3.4
6	(1970, p.166)	no. of banana bunches	no. of banana pits	20	3.1
7	(1970, p.51)	area under rice	total cultivated area	25	1.9
8	Rao (1963, p.207)	1960 area under corn	1958 area under corn	14	1.9
9	Cochran (1977, p.203)	weight of peaches	eye-estimate	10	1.4
10	Cochran (1977, p.325)	number of persons	number of rooms	10	2.1

to have inclusion probability greater than unity.

Estimators

The estimators compared and their abbreviations used in the tables are as follows:

1. YG, the Yates-Grundy estimator given by (1.4).
2. HT, The Horvitz-Thompson estimator of variance, given by (1.3).
3. A2, the estimator v_{A2} of Ajgaonkar, given in Chapter II.
4. F, the estimator of Fuller (1970), given by

$$F = [\sum_{s(2)} C_{ij}\pi_{ij}^{-1}\phi_{ij} / \sum_{s(2)} C_{ij}\pi_{ij}^{-1}] . \sum_{U(2)} C_{ij},$$

where

$$C_{ij} = \pi_i\pi_j - \pi_{ij},\ z_i = y_i/\pi_i \text{ and } \phi_{ij} = (z_i - z_j)^2$$

5. R, a "design-independent" ratio-type estimator, given by

$$R = [\sum_{s(2)} C_{ij}\phi_{ij} / \sum_{s(2)} C_{ij}] \cdot \sum_{U(2)} C_{ij}$$

6. B, the estimator v given by (3.69).
7. Bpx, the estimator $v_{p\xi}$, given by (3.71).

Also used in the study, but not tabulated, was the estimator v_ξ in (3.70), which performed nearly identically to $v_{p\xi}$.

The first four estimators are design-based, YG, HT and A2 being design-unbiased, and F approximately so. The last three estimators are model-based. F and R are ξ-unbiased and Bpx is $p\xi$-unbiased. R and B are design-independent, in the sense that they depend on the design only through C_{ij}'s. If we were to estimate the variance of e_{HT} under a design p_o, based on a sample drawn according to the design p, then these two estimators depend on p_o, but not on p. Bpx is slightly design-dependent, as it involves the design parameters π_i, π_{ij} in terms of order 1/n.

As shown in Chapter III, B and Bpx are "best" in the classes C_L, $C_{Lp\xi}$ respectively, under a random permutation model on $Z_i = Y_i/x_i$, i = 1,...N, with $\beta_{2z} \doteq 3$. The actual coefficients of kurtosis for the populations used ranged from 1.4 to 19.9.

Table 4.2 gives the efficiencies of other estimators relative to YG, measured by the inverse ratio of the empirical mean square errors. Geometric means of the efficiencies are computed over all populations except No. 3. The latter is excluded because e_{HT} is inappropriate as an estimator of the total for this population. For the same design, the ratio estimator, given in (1.6), was empirically found to be four to five times more efficient than e_{HT}.

The percentage contributions of bias to the mean square error are shown in Table 4.3. Based on a normal approximation for the 200-sample average of the estimators, a contribution of less than 2% is not significant at the 5% level, and may be attributed to sampling fluctuations, rather than any real bias.

Discussion of the Results

General Observations

Among the three exactly design-unbiased estimators, the estimators

Table 4.2. Efficiencies relative to the Yates-Grundy estimator

Pop. no.	Sample size	HT	A2	F	R	B	Bpx
1	3	0.14	0.91	1.03	1.07	2.86	1.04
	5	.03	.37	1.05	1.12	1.92	1.10
	10	.004	.20	1.01	1.06	1.27	1.12
2	3	1.00	.98	1.03	1.35	1.81	1.07
	5	1.03	.24	1.06	1.63	1.73	1.20
	10	1.24	.01	1.20	1.94	1.92	1.58
3	3	.96	1.15	1.14	7.33	4.74	1.37
	5	.98	1.31	1.16	13.98	3.96	1.99
	10	1.15	.40	1.43	13.19	6.70	5.34
4	3	.82	.59	1.07	1.06	1.53	1.10
	5	.71	.32	1.11	1.27	1.77	1.22
	10	.24	.09	1.18	1.68	1.59	1.67
5	3	.23	.34	.88	.58	1.74	.78
	5	.07	.27	.88	.35	1.06	.58
	10	.03	.26	1.39	4.65	2.48	1.33
6	3	.61	.94	.99	.92	2.07	.97
	5	.23	.86	.96	.90	1.42	.93
	10	.03	.11	.94	.84	.99	.86
7	3	.16	.46	1.09	1.17	1.79	1.17
	5	.05	.06	1.24	1.35	1.68	1.43
8	3	.08	.68	1.19	1.37	1.67	1.33
	5	.02	.12	1.30	1.37	1.30	1.48
9	3	.002	.93	1.06	1.18	1.42	1.15
	5	.0003	.39	1.13	1.30	1.42	1.32
10	3	.16	.92	1.10	1.23	2.03	1.18
	5	.03	.50	1.18	1.37	1.47	1.38
Geo. mean exc. pop. 3	3	0.16	0.71	1.05	1.07	1.84	1.08
	5	0.05	0.28	1.09	1.10	1.51	1.14
	10	0.06	0.09	1.13	1.68	1.57	1.28

Table 4.3. Percentage contribution of bias to the MSE

Pop. no.	Sample size	F	R	B	Bpx
1	3	0	0	16	0
	5	0	0	6	0
	10	0	2	11	3
2	3	1	5	55	2
	5	0	2	30	0
	10	1	1	11	0
3	3	0	4	2	0
	5	0	22	2	0
	10	0	49	14	8
4	3	1	2	58	1
	5	1	0	22	0
	10	1	2	27	5
5	3	0	1	29	0
	5	0	10	3	5
	10	1	8	33	52
6	3	0	0	34	0
	5	0	0	17	0
	10	0	0	6	0
7	3	0	0	54	0
	5	0	0	37	0
8	3	0	2	58	0
	5	0	11	51	7
9	3	3	4	55	4
	5	0	0	29	1
10	3	0	0	36	0
	5	0	2	30	2

A2 and HT can be euphemistically described as uncompetitive. Ironically, the most miserable performer HT is the only one for which an admissibility result is available! The estimator F showed no significant bias, and was somewhat more efficient than YG.

All model-based estimators have performed generally better than the design-based ones, especially so in the presence of extreme values. In particular, B was never worse than any of the design-based estimators. For sample sizes 3 and 5, B was the most efficient estimator overall, while R did somewhat better than B for sample size 10.

The relative efficiencies of HT, A2, and B showed a decreasing trend with increasing sample size in a majority of the cases. Those of F, R and Bpx generally followed a reverse trend.

The biases of R and Bpx are small or moderate except for populations with extreme values, while B, not surprisingly, had the largest biases of all.

<u>The effect of extreme values</u>

Consider the problem of estimation of the ratio $\bar{y}/\bar{x}$ for two characteristics x and y of a finite population. Two ratio-type estimators are

$RM = \bar{y}_s/\bar{x}_s$, the ratio of means, and

$MR = \sum_s r_i/n$ the mean of ratios, where $r_i = y_i/x_i$

RM can be written $\sum_s x_i r_i/(n\bar{x}_s)$

If r_1 changes by an amount d, the changes in RM and MR are respectively, $d.x_1/(n\bar{x}_s)$ and d/n. It follows that an extreme value of

A2 and HT can be euphemistically described as uncompetitive. Ironically, the most miserable performer HT is the only one for which an admissibility result is available! The estimator F showed no significant bias, and was somewhat more efficient than YG.

All model-based estimators have performed generally better than the design-based ones, especially so in the presence of extreme values. In particular, B was never worse than any of the design-based estimators. For sample sizes 3 and 5, B was the most efficient estimator overall, while R did somewhat better than B for sample size 10.

The relative efficiencies of HT, A2, and B showed a decreasing trend with increasing sample size in a majority of the cases. Those of F, R and Bpx generally followed a reverse trend.

The biases of R and Bpx are small or moderate except for populations with extreme values, while B, not surprisingly, had the largest biases of all.

The effect of extreme values

Consider the problem of estimation of the ratio $\bar{y}/\bar{x}$ for two characteristics x and y of a finite population. Two ratio-type estimators are

$RM = \bar{y}_s/\bar{x}_s$, the ratio of means, and

$MR = \sum_s r_i/n$ the mean of ratios, where $r_i = y_i/x_i$

RM can be written $\sum_s x_i r_i/(n\bar{x}_s)$

If r_1 changes by an amount d, the changes in RM and MR are respectively, $d.x_1/(n\bar{x}_s)$ and d/n. It follows that an extreme value of

r_1 affects MR more than it affects RM, if $x_1 < \bar{x}_s$, and vice versa.

For the estimation of the population total, apart from a constant multiplier, the ratio estimator in (1.6) is of the type RM, while the Horvitz-Thompson estimator is of the type MR. In population 3, an extreme value of y/x is associated with a small value of x and therefore ratio estimator is more efficient than e_{HT}. For the estimation of V_{HT}, the estimator R is of the type RM, while B is essentially of the type MR, apart from some adjustment terms and some shrinkage, with the roles of x_i and r_i played by C_{ij} and ϕ_{ij} respectively.

An extreme value of y/x leads to an extreme value of $z_i = y_i/\pi_i$ and hence of ϕ_{ij} for all j ($\neq$i). The mean of the C_{ij}'s associated with a particular unit i is

$$\sum_{j \varepsilon U} C_{ij}/(N-1) = \pi_i \ (1-\pi_i)/(N-1).$$

Thus C_{ij}'s associated with the unit i are small on the average, if π_i is close to 0 or 1.

In population 3, z_i is extreme for a unit with very small π_i, and hence extreme values of ϕ_{ij}'s are associated with small values of C_{ij}. This explains the larger efficiencies of R than of B, although B is itself a considerable improvement over YG.

In population 5, an extreme value of z_i (the smallest) is associated with the unit with the largest x_i. For n = 3, 5 and 10, the inclusion probabilities of this unit are ·26, ·44 and ·88 respectively. Thus for the first two sample sizes, π_i $(1-\pi_i)$ and hence C_{ij}'s associated with this unit are relatively large, but for sample size 10, they are much smaller. This explains the low efficiency of R for n = 3 and 5, and the comeback for sample size 10. The effect is more dramatic because the unit involved has the largest inclusion probability.

In brief, we may conclude that R would be preferable, if units suspected to have extreme values of y/x have inclusion probabilities much below average, or close to 1. If the inclusion probabilities of the suspicious units are in the intermediate range, then B is likely to be more efficient. The choice is particularly critical when such units have large inclusion probabilities.

APPENDIX I: A GENERALIZATION OF THE YATES-GRUNDY FORMULA FOR VARIANCE

For a population U, let U(m) denote the collection of all subsets of U containing exactly m units. For a sample s, we can similarly define s(m).

For each $I = \{i_1, \ldots i_m\} \varepsilon U(m)$, let χ_I be a function of $y_{i_1}, \ldots, y_{i_m}$, and let

$$e(s) = \sum_{I\varepsilon s(m)} \chi_I \qquad \text{(A1.1)}$$

For a given design p and each $I\varepsilon U(m)$, let Π_I be the joint inclusion probability of all units in I, and let Π_{IJ} be the joint inclusion probability of all units in I and J. In symbols,

$$\Pi_I = \Pr (I \subset s) = \Pr (I\varepsilon s(m)), \qquad I\varepsilon U(m)$$

and

$$\Pi_{IJ} = \Pr (I,J \subset s) = \Pr (I, J\varepsilon s(m)), I, J\varepsilon U(m)$$

<u>Theorem A1.1</u> For a fixed sample size design p, variance of the estimator e in (A1.1) is given by

$$V (e) = 1/2 \sum_{I,J\varepsilon U(m)} (\Pi_I \Pi_J - \Pi_{IJ})(\chi_I - \chi_J)^2 \qquad \text{(A1.2)}$$

<u>Proof</u> Define random variables α_I $I\varepsilon U(m)$ as follows.

$$\alpha_I = 1 \text{ if } I \varepsilon s$$
$$= 0 \text{ otherwise.}$$

Clearly, $E(\alpha_I) = \Pi_I$

$E(\alpha_I \cdot \alpha_J) = \Pi_{IJ}$

$\mathrm{Cov}(\alpha_I, \alpha_J) = \Pi_{IJ} - \Pi_I \Pi_J$

For a design of fixed sample size n, there are exactly $\binom{n}{m}$ subsets of m units in the sample. Hence,

$$\sum_{I\varepsilon U(m)} \alpha_I = \binom{n}{m} = E\left(\sum_{I\varepsilon U(m)} \alpha_I \right) \tag{A1.3}$$

$$\text{Now } V(e) = V\left(\sum_{I\varepsilon s(m)} \chi_I \right)$$

$$= V\left(\sum_{I\varepsilon U(m)} \alpha_I \chi_I \right)$$

$$= \sum_{I,\ J\varepsilon U(m)} \chi_I \chi_J \,\mathrm{Cov}(\alpha_I, \alpha_J)$$

$$= \sum_{I,\ J\varepsilon U(m)} (\Pi_{IJ} - \Pi_I \Pi_J)\, \chi_I \chi_J \tag{A1.4}$$

$$= 1/2 \sum_{I,J} (\Pi_{IJ} - \Pi_I \Pi_J) \left[\chi_I^2 + \chi_J^2 - (\chi_I - \chi_J)^2 \right]$$

$$= 1/2 \sum_{I,J} (\Pi_I \Pi_J - \Pi_{IJ})(\chi_I - \chi_J)^2 + \sum_{I,J} (\Pi_{IJ} - \Pi_I \Pi_J)\chi_I^2$$

(using $\Pi_{IJ} = \Pi_{JI}$).

The second sum on r.h.s. is,

$$\sum_I \chi_I^2 \sum_J (\Pi_{IJ} - \Pi_I \Pi_J)$$

$$= \sum_I \chi_I^2 \sum_J [E(\alpha_I \alpha_J) - E(\alpha_I) E(\alpha_J)]$$

$$= \sum_I \chi_I^2 E \left[\alpha_I \left(\sum_J \alpha_J - E \sum_J \alpha_J\right)\right]$$

= 0, since the expression in the parentheses vanishes by (A1.3).

Hence the theorem.

<u>Corollary</u> An unbiased estimator of V(e) is,

$$1/2 \sum_{I,J \varepsilon s(m)} (\Pi_I \Pi_J - \Pi_{IJ})(\chi_I - \chi_J)^2 / \Pi_{IJ}$$

provided, $\Pi_{IJ} \neq 0$ whenever $\Pi_{IJ} - \Pi_I \Pi_J \neq 0$.

APPENDIX II: CHARACTERIZATION OF ADMISSIBLE LINEAR ESTIMATORS OF THE POPULATION TOTAL FOR SAMPLES OF SIZE ONE

Notation :

Let $p_i = p(\{i\})$, the probability of selecting unit i as the sample.

$$T = \sum_{i=1}^{n} y_i, \text{ the population total.}$$

e_a, e_b = estimators defined by

$$e_a(\{i\}) = a_i y_i$$

and $$e_b(\{i\}) = b_i y_i \qquad i = 1,\ldots..N. \tag{A2.1}$$

where a_i, b_i $i = 1,\ldots N$ are constants.

$$c_i = b_i - 1,$$

$$d_i = p_i(b_i - c_i) \text{ and } \underline{d} = (d_1,\ldots d_N). \tag{A2.2}$$

$$D = \underline{d}: \{d_i c_i > 0\}. \tag{A2.3}$$

$G_d = (g_{dij})_{NxN}$ and $H_d = (h_{dij})_{NxN}$, where

$$g_{dij} = d_i(2c_i - d_i/p_i) \qquad i = j$$

$$= -(d_i + d_j) \qquad i \neq j \tag{A2.4}$$

$$\text{and} \qquad h_{dij} = 2d_i c_i \qquad i = j$$

$$= g_{dij} \qquad i \neq j. \tag{A2.5}$$

The ranges of summations and products will be omitted or abbreviated after the first mention whenever there is no ambiguity.

Lemma A2.1

For a design of fixed sample size one, the estimator e_b defined in (A2.1) is admissible in the class of linear estimators of T, if and only if the matrix H_d in (A2.5) is positive definite (PD) for some $\underline{d} \varepsilon D$.

Proof We shall prove the lemma in three parts.

(i) MSE $(e_a) \leq$ MSE (e_b) for all $\underline{y}$, if and only if the matrix G_d in (A2.4) is non-negative definite (NND).

(ii) If G_d is NND and non-null, then $\underline{d} \varepsilon D$ and H_d is PD.

(iii) If H_d is PD for some $\underline{d}$, then $\underline{d} \varepsilon D$ and $G_{d'}$ is PD for some $\underline{d}'$.

$$\text{MSE } (e_a) - \text{MSE } (e_b) = \sum_{i=1}^{n} p_i \, (b_i y_i - T)^2 - \sum_{i=1}^{n} p_i \, (a_i y_i - T)^2$$

$$= \Sigma \, p_i \, (b_i - a_i) \, y_i \, (b_i y_i + a_i \, y_i - 2T)$$

$$= \sum_i d_i \, y_i \left[(2b_i - d_i/p_i - 2) \, y_i - 2 \sum_{\substack{j=1 \\ j \neq i}}^{N} y_j \right]$$

$$= \sum_i a_{ii} y_i^2 - 2 \sum_{\substack{i,j=1 \\ i \neq j}}^{N} d_i y_i y_j.$$

The last sum can be rearranged as

$$- \sum_{i \neq j} (d_i + d_j) y_i y_j = \sum_{i \neq j} g_{dij} y_i y_j,$$

giving MSE (e_a) - MSE $(e_b) = \underline{y}' G_d \underline{y}$, from which the first part follows.

To prove the second part, assume that G_d is NND and non-null.

If $d_i = 0$ for some i, then $g_{dii} = 0$, which implies $g_{dij} = 0$ j = 1, ...N since G_d is NND. This, along with $d_i = 0$ implies $d_j = 0$ j = 1 ...N, and hence $G_d = 0$, contrary to the hypothesis that it is non-null. Therefore, we must have,

$$d_i \neq 0 \qquad i = 1,\ldots N.$$

Further, $g_{dii} = d_i(2c_i - d_i/p_i) \geq 0$ for G_d to be NND. Hence, we must have $d_i c_i > 0$ i = 1...N, or $\underline{d} \varepsilon D$.

Positive definiteness of H_d follows easily since G_d is NND and $H_d - G_d = \text{diag.} (d_1^2/p_1,\ldots d_N^2/p_N)$.

Finally, to prove the third part, assume that H_d is PD for some $\underline{d}$. The necessity of $\underline{d} \varepsilon D$ is trivial.

We may assume without loss of generality,

$$d_i^2/p_i < 1 \qquad i = 1,\ldots N \tag{A2.6}$$

(If not, we can replace $\underline{d}$ by a suitable positive multiple of it without affecting the positive definiteness of H_d.)

Now, we use the fact that the eigenvalues of $H_d - \varepsilon I$ are obtained by subtracting ε from each eigenvalue of H_d. If λ is the smallest eigenvalue of H_d, then for $0 < \varepsilon < \lambda$, the eigenvalue of $H_d - \varepsilon I$ are all positive and hence it is PD.

Let $\varepsilon\underline{d} = \underline{d}' = (d_1',\ldots d_N')$.

$$G_{d'} = H_{d'} - \text{diag.}\ (d_1'^2/p_1,\ldots d_N'^2/p_N)$$

$$= \varepsilon H_d - \varepsilon^2.\ \text{diag.}\ (d_1^2/p_1,\ldots d_N^2/p_N)$$

$$= \varepsilon(H_d - \varepsilon I) + \varepsilon^2.\ \text{diag.}\ (1 - d_1^2/p_1,\ldots 1 - d_N^2/p_N),$$

which is PD because $H_d - \varepsilon I$ is PD and

$$1 - d_i^2/p_i > 0 \quad i = 1,\ldots N \ \text{ by (A2.6).}$$

This completes the proof of lemma A2.1.

Lemma A2.2.

If b_i, $d_i \neq 0$ $i = 1,2\ldots N$, then the determinant of H_d is given by

$$\left| H_d \right| = \prod_{i=1}^{N} (2\, b_i\, d_i) \cdot (1 - \sum_{i=1}^{N} b_i^{-1} - W/4) \qquad \text{(A2.7)}$$

where $\quad W = \Sigma\, b_i^{-1}\, d_i\ \Sigma\, b_i^{-1}\, d_i^{-1} - (\Sigma\, b_i^{-1})^2$

$$= \sum_{i,j=1}^{N} b_i^{-1}\, b_j^{-1}\, f_{ij}, \quad f_{ij} = d_i\, d_j^{-1} + d_i^{-1}\, d_j - 2 \qquad \text{(A2.8)}$$

Proof   It is easy to verify that

$$H_d = E - \underline{j}\underline{d}' - \underline{d}\underline{j}' = E - (\underline{j}\ \underline{d})(\underline{d}\ \underline{j})',$$

where $\quad E = \text{diag.}\ (2b_1\, d_1, \ldots\ 2b_N\, d_N)$ and $\underline{j} = (1, \ldots 1)'$.

Using the formula for the determinant of a partitioned matrix,

$$\begin{vmatrix} E & (\underline{j}\ \underline{d}) \\ (\underline{d}\ \underline{j})' & I_2 \end{vmatrix} = \left| I_2 \right| \cdot \left| E - (\underline{j}\ \underline{d})\, I_2^{-1}\, (\underline{d}\ \underline{j})' \right|$$

$$= \left| E \right| \cdot \left| I_2 - (\underline{d}\ \underline{j})'\, E^{-1}\, (\underline{j}\ \underline{d}) \right|$$

where I_2 is the 2 x 2 identity matrix.

From the last two expressions, we get

$$\left| H_d \right| = \left| E \right| \cdot \left| \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} \Sigma\, (d_i/2\, b_i d_i) & \Sigma\, (d_i^2/2\, b_i d_i) \\ \Sigma\, (1/2\, b_i d_i) & \Sigma\, (d_i/2\, b_i d_i) \end{pmatrix} \right|$$

$$= \prod_{i=1}^{N} (2\, b_i d_i) \cdot \begin{vmatrix} 1 - 1/2\ \Sigma\, b_i^{-1} & - 1/2\ \Sigma\, b_i^{-1} d_i \\ - 1/2\ \Sigma\, b_i^{-1} d_i^{-1} & 1 - 1/2\ \Sigma\, b_i^{-1} \end{vmatrix}$$

which simplifies to (A2.7).

Remark Since each principal minor of H_d has the same form as H_d itself, (A2.7) is applicable to it with obvious changes in the ranges of the product and the summations.

Now we are ready to prove theorem 1.1, restated below in a slightly modified form.

Theorem 1.1

For any design of fixed sample size one, with $p_i > 0$ for all units i, the estimator e_b is admissible in the class of linear estimators of T if and only if one of the following holds.

(i) $b_i = 1$ for some i.

(ii) $\sum_{i \varepsilon A} b_i^{-1} \geq 1$ where $A = \{i: b_i > 1\ (c_i > 0)\}$.

(iii) $0 \leq b_i < 1$ for at least two i.

(iv) $0 < b_i < 1$ for one i and $\sum_{i \varepsilon B} b_i^{-1} \leq 1$.

where

$B = \{i: b_i < 1\ (c_i < 0)\}$.

Remarks

(a) Parts (i) and (ii) above are combined in part (i) of the theorem as stated in Chapter I.

(b) If $p_i = 0$ for some i, then it is trivial to show that all linear estimators are admissible.

Proof of the "only if" part Suppose, any of the conditions (i) - (iv) do not hold. We must show that for some $\underline{d}$, H_d is PD. The desired result will then follow from lemma A2.1.

Assume, without loss of generality,

$$C_1, \ldots C_M > 0 \text{ and } C_{M+1}, \ldots C_N < 0 \quad (0 \le M \le N), \text{ i.e.,}$$

$$A = \{1, \ldots M\} \text{ and } B = \{M+1, \ldots N\}.$$

The possibility $C_i = 0$ is ruled out because (i) does not hold.

Let $d_* = (d_{*1}, \ldots . d_{*N})'$, where

$$\begin{aligned} d_{*i} &= 1 \quad i\varepsilon A \\ &= -1 \quad i\varepsilon B \end{aligned}$$

Let $H = (h_{ij}) = H_{d*}$, the value of H_d for $\underline{d} = \underline{d}_*$.

Since $h_{ij} = -(d_{*i} + d_{*j})$ vanishes for $i\varepsilon A$ and $j\varepsilon B$, or vice versa, H can be partitioned as

$$H = \begin{pmatrix} H_A & 0 \\ 0 & H_B \end{pmatrix}$$

To show H is PD, we show that H_A and H_B are PD.

From (A2.7), using $d_{*i} = 1$ for all $i\varepsilon A$, the leading r x r minor of H_A has determinant

$$\prod_{i=1}^{r} (2b_i) \cdot (1 - \sum_{i=1}^{r} b_i^{-1}), \quad r = 1,\ldots M$$

which is positive because $b_i > 1$ for $i\varepsilon A$, and the last factor is not smaller than $1 - \sum_{i\varepsilon A} b_i^{-1}$ which is positive because (ii) does not hold. Thus H_A is PD.

To show H_B is PD, consider the following three cases, which exhaust the possibilities because (iii) and (iv) do not hold.

(a) $b_i < 0$ for all $i\varepsilon B$

(b) $0 < b_i < 1$ for some i, say i = N,

$$b_i < 0 \text{ for } i\varepsilon B - \{N\}, \quad \text{and } 1 - \sum_{i\varepsilon B} b_i^{-1} < 0$$

(c) $b_i = 0$ for i = N (say), and $b_i < 0$ for $i\varepsilon B - \{N\}$.

The result for case (a) follows along the same lines as that for H_A, noting that, for $i\varepsilon B$

$$b_i < 0, \; d_{i*} = -1 \text{ and } 1 - \sum_{i=M+1}^{M+r} b_i^{-1} > 0$$

$$r = 1, 2\ldots N - M.$$

Case (b) is also treated similarly except for r = N - M, for which the determinant is,

$$\left| H_B \right| = \prod_{i=M+1}^{N} (-2b_i) \cdot (1 - \sum_{i\varepsilon B} b_i^{-1})$$

This contains two negative factors ($-2\ b_N$) and $1 - \sum_{i\varepsilon B} b_i^{-1}$, all other factors being positive. Thus, $|H_B| > 0$ and the proof of case (b) is complete.

For case (c), replace b_N in H_B by $b_{N*} > 0$, such that

$$1 - \sum_{i\varepsilon B-\{N\}} b_i^{-1} - b_{N*}^{-1} < 0,$$

and call the resulting matrix H_{B*}.

H_{B*} is PD by case (b).

$$H_B - H_{B*} = \text{diag. } (0,0\ldots 2d_{*N}\ (b_N - b_{N*}))$$

$$= \text{diag. } (0,0,\ldots 2b_{N*}), \text{ since}$$

$$d_{*N} = -1 \text{ and } b_N = 0.$$

Hence $H_B - H_{B*}$ is NND and therefore H_B is PD.

Proof of the "only if" part is now complete.

<u>Proof of "if" part</u> Suppose one of the conditions (i) - (iv) holds. We will show that H_d cannot be PD for any $\underline{d}\varepsilon D$ and the theorem will follow from lemma A2.1.

If (i) holds, i.e. $c_i = 0$ for some i, then $h_{dii} = 0$, whatever $\underline{d}$ may be and hence H_d cannot be PD.

Let H_{dA} be the principal minor of H_d containing rows and columns

$i \varepsilon A$, with a similar definition for H_{dB}. To show H_d is not PD, it is sufficient to show that either H_{dA} or H_{dB} is not PD.

If (ii) holds, i.e $\sum_{i \varepsilon A} b_i^{-1} \geq 1$, then

$$\left|H_{dA}\right| = \prod_{i \varepsilon A} (2d_i b_i)(1 - \sum_{i \varepsilon A} b_i^{-1}) \leq 0$$

because $b_i > c_i > 0$ for $i \varepsilon A$ and

$d_i > 0$ for $i \varepsilon A$, $\underline{d} \varepsilon D$.

If (iii) holds, i.e. $0 \leq b_i, b_j < 1$ for two units i and j, then d_i, $d_j < 0$ for $\underline{d} \varepsilon D$.

The determinant of the 2 x 2 minor of H_{dB} containing rows and columns i and j is

$$\begin{vmatrix} 2d_i \, b_i & -2\,(d_i + d_j) \\ -(d_i + d_j) & 2d_j \, b_j \end{vmatrix} = 4\, d_i\, d_j\, b_i\, b_j - (d_i + d_j)^2$$

$$= 4\, d_i\, d_j\, (b_i\, b_j - 1) - (d_i - d_j)^2$$

$$\leq 0.$$

Hence H_{dB} cannot be PD.

Finally, consider the case when (iv) holds, i.e., $0 < b_i < 1$ for some $i \varepsilon B$, say $i = N$, $b_i < 0$ for $i \varepsilon B - \{N\} = B1$ (say), and

$$1 - \sum_{i \varepsilon B} b_i^{-1} \geq 0.$$

The last condition can be written

$$1 - q - b_N^{-1} \geq 0 \tag{A2.9}$$

where

$$q = \sum_{i \varepsilon B1} b_i^{-1}$$

Now consider the principal minor H_{B1} containing all rows of H_B except the last one.

$$\left|H_{B1}\right| = \prod_{i \varepsilon B1} (2d_i\ b_i) \cdot (1 - q - W/4)$$

where

$$W = \sum_{i,j \varepsilon B1} b_i^{-1}\ b_j^{-1}\ f_{ij} \quad \text{and}$$

$$f_{ij} = d_i\ d_j^{-1} + d_j\ d_i^{-1} - 2$$

Since b_i, $b_j < 0$, d_i, $d_j < 0$ for $i, j \varepsilon B1$ and $\underline{d} \varepsilon D$.

$$W \geq 0 \tag{A2.10}$$

If $\left|H_{B1}\right| \leq 0$, the proof is complete. Otherwise, we have $\left|H_{B1}\right| > 0$, implying $1 - q - W/4 > 0$ or, $W < 4\,(1 - q)$ (A2.11)

$$\left| H_B \right| = \prod_{i\varepsilon B} (2d_i\, b_i) \cdot K \tag{A2.12}$$

where

$$K = 1 - q - b_N^{-1} - 1/4 \left(W + \sum_{j\varepsilon B1} b_N^{-1}\, b_j^{-1}\, f_{Nj}\right)$$

$$= 1 - q - W/4 - b_N^{-1} + q\, b_N^{-1}/2 - b_N^{-1}\, d_N^{-1}/4$$

$$\times \left(\Sigma\, b_j^{-1}\, d_j + d_N^2\, \Sigma\, b_j^{-1}\, d_j^{-1}\right) \tag{A2.13}$$

Let $\alpha = (\Sigma\, b_j^{-1}\, d_j)^{1/2}$ and $\beta = -\, d_N\, (\Sigma\, b_j^{-1}\, d_j^{-1})^{1/2}$

The expression in the parentheses in (A2.13) is

$$\alpha^2 + \beta^2 \geq 2\alpha\beta$$

$$= -\, 2d_N\, (\Sigma\, b_j^{-1}\, d_j \,.\, \Sigma\, b_j^{-1}\, d_j^{-1})^{1/2}$$

$$= -\, 2d_N\, (q^2 + W)^{1/2}$$

or

$$\alpha^2 + \beta^2 \geq -\, 2d_N U \tag{A2.14}$$

where

$$U = (q^2 + W)^{1/2} \tag{A2.15}$$

From (A2.11),

$$U \le [q^2 + 4\ (1 - q)]^{1/2}$$

or

$$U \le 2 - q \tag{A2.16}$$

From (A2.10),

$$U \ge |q| = -\ q \tag{A2.17}$$

From (A2.13 - A2.15),

$$K \ge 1 - q - (U^2 - q^2)/4 - b_N^{-1} + q\ b_N^{-1}/2 + Ub_N^{-1}/2$$

$$= (1 - q/2 - U/2)\ (1 - q/2 + U/2 - b_N^{-1}).$$

The first factor on r.h.s. ≥ 0 by (A2.16), and the second factor ≥ 0 by (A2.17) and (A2.9). Hence $K \ge 0$, which by (A2.12) gives

$$\left| H_B \right| \le 0 \quad \text{as } b_N > 0 \text{ and } d_N < 0.$$

Hence the theorem.

APPENDIX III. DECOMPOSABILITY OF LINEARLY DEPENDENT SYMMETRIC KERNELS

Notation :

$$U = \{1, 2 \ldots N\}$$

For any set A, define

$$A(m) = \{\text{subsets of } A \text{ containing exactly } m \text{ elements}\}$$

e.g., $U(2) = \{ \{i,j\}: \; i,j \in U, \; i \neq j\}$

$$I = \{i_1, \ldots i_m\}$$

$$I_k = \{i_k, \ldots i_m\} \qquad 1 \leq k \leq m$$

For a symmetric function f,

$$f_I(y) = f_{i_1 \ldots i_{k-1} I_k}(y) = f_{i_1 \ldots i_m}(y)$$

$$= f(y_{i_1}, \ldots y_{i_m})$$

For brevity '$(\underline{y})$' may be omitted when there is no ambiguity, giving f_I, $f_{i_1 I_2}$ etc.

Theorem 3.1 is restated below and proved in several stages.

Theorem 3.1

Let $f: A^m \to V$ be a symmetric function, where A is arbitrary and V is a vector space. For $\underline{y} \varepsilon A^N$ and $I = \{i_1, \ldots i_m\} \; \varepsilon \; U(m)$, let $f_I(y) = f(y_{i_1}, \ldots y_{i_m})$.

If there exist constants C_I, not all zero, such that

$$\sum_{U(m)} C_I f_I(\underline{y}) = 0 \qquad \text{for all } \underline{y}\varepsilon A^N, \tag{A3.1}$$

and m>1, then there exists a symmetric function g: $A^{m-1} \to V$, such that

$$f_I(\underline{y}) = \sum_{J\varepsilon I(m-1)} g_J(\underline{y}) \tag{A3.2}$$

where I(m-1) = {subsets of I containing m-1 elements}, and for $J = \{j_1,\ldots j_{m-1}\}$, $g_J(\underline{y}) = g(y_{j_1},\ldots y_{j_{m-1}})$.

We will first prove the theorem for the special case when C_I are all non-negative (or all non-positive). The device used is similar to that of theorem 1 of Royall (1968) or theorem 4.1 of Godambe and Joshi (1965).

<u>Lemma A3.1</u>

If (A3.1) holds for m≥1, $C_I \geq o$ for all I and $C_I > o$ for some I, then f identically vanishes, and hence (A3.2) holds trivially.

<u>Proof</u> (by induction) Let 'a' be an arbitrary element of A, and let f (a,...a) = α.

Putting $\underline{y}$ = (a,.....a) in (A3.1), $\alpha \sum_{U(m)} C_I = o$.

The second factor is positive by hypothesis, hence α = o. That is, f vanishes when all m of its arguments are equal. Now assume the result to be true when at least m-k+1 arguments of f are equal, i.e.,

$f(a_1,\ldots a_k,\ldots a_k) = o$ for all $a_1,\ldots a_k \varepsilon A$.

We will show f vanishes when m-k of its arguments are equal, i.e., $f(a_1,\ldots a_k,a_{k+1}\ldots a_{k+1}) = o$ for $a_1,\ldots a_{k+1}\varepsilon A$.

Since $C_I > o$ for some I, assume this holds for $I = \{1, 2,\ldots m\}$.

Let $B = \{I\varepsilon U(m)\colon\ 1, 2,\ldots k\varepsilon I\}$.

Put $\underline{y} = (a_1,\ldots a_k,a_{k+1},\ldots a_{k+1})$ in (A3.1). If $I \not\varepsilon B$, then $f_I(\underline{y}) = o$ by the induction hypothesis, as at least m-k+1 of its arguments are equal to a_{k+1}. If $I\varepsilon B$, $f_I(\underline{y}) = f(a_1,\ldots a_k,a_{k+1},\ldots a_{k+1}) = \alpha_k$ (say). Thus (A3.1) reduces to

$$\alpha_k \sum_{I\varepsilon B} C_I = o$$

The second factor is positive because $C_I > o$ for $I = \{1, 2,\ldots m\}\ \varepsilon B$. Hence $\alpha_k = o$. By induction, we get $f\ (a_1,\ldots a_m) = o$ for all $a_1,\ldots a_m \varepsilon A$.

<u>Lemma A3.2</u>

Let $f\colon A \to V$ and $\sum_{i\varepsilon U} C_i f(y_i) = o$ for all $\underline{y}\varepsilon A^N$, where C_i are constants, not all zero. Then f is a constant function.

<u>Proof</u> If C_i are all equal, the result follows from lemma A3.1. Otherwise, suppose $C_1 \neq C_2$.

Let $q(y_1,\ldots y_N) = \sum_{i\varepsilon U} C_i f(y_i)$.

Since q is identically zero,

$$0 = q(y_1,y_2,\ldots y_N) - q(y_2,y_1,y_3,\ldots y_N)$$

$$= (C_1 - C_2)\ [f(y_1) - f(y_2)]$$

Since $C_1 \neq C_2$, $f(y_1) = f(y_2)$. But y_1, y_2 are arbitrary, hence f is a constant function.

Lemma A3.3

For $\ell < m$, let $I(\ell)$ = {subsets of I containing ℓ elements}, and let q^J, $J = \{j_1,\dots j_\ell\} \in I(\ell)$ be functions, not necessarily symmetric, defined on A^ℓ, such that

$$f_I(\underline{y}) = \sum_{J\in I(\ell)} q^J(y_{j_1},\dots y_{j_\ell}) \text{ for all } \underline{y}\ A^N \tag{A3.2}$$

Then there exists a symmetric function $g: A^\ell \to V$ such that

$$f_I(\underline{y}) = \sum_{J\in I(\ell)} g(y_{j_1},\dots y_{j_\ell}) \tag{A3.3}$$

Remark In (A3.2) f_I is expressed as a sum of $\binom{m}{\ell}$ possibly different and asymmetric functions, each evaluated at one of the ℓ-tuples $(y_{j_1},\dots y_{j_\ell})$. In (A3.3), it is expressed as the sum of a single symmetric function evaluated at all $(y_{j_1},\dots y_{j_\ell})$. The result is proved by averaging (A3.2) over all permutations of I and using the symmetry of f. A formal general proof is not difficult, but a particular case may be more illuminating.

Let $m = 3$, $\ell = 2$, and

$$f_{123}(\underline{y}) = q^{12}(y_1,y_2) + q^{13}(y_1,y_3) + q^{23}(y_2,y_3).$$

Averaging over all permutations of {1, 2, 3} gives

$$f_{123}(\underline{y}) = g(y_1,y_2) + g(y_1,y_3) + g(y_2,y_3),$$

where $$g(u,v) = [q^{12}(u,v) + q^{12}(v,u) + q^{13}(u,v) + q^{13}(v,u) + q^{23}(u,v) + q^{23}(v,u)]/6.$$

<u>Proof of theorem 3.1 for m = 2</u>

Let $f\colon A^2 \to V$ be symmetric and let $f_{ij} = f(y_i, y_j)$, and C_{ij}, $\{i, j\} \varepsilon U(2)$ be constants, not all zero, such that

$$\sum_{U(2)} C_{ij} f_{ij}(\underline{y}) = 0 \quad \text{for all } \underline{y} \varepsilon A^N.$$

Then there exists a function $g\colon A \to V$, such that $f(y_i, y_j) = g(y_i) + g(y_j)$.

<u>Proof</u> If all C_{ij}'s are equal, then the result follows from lemma A3.1. Otherwise, suppose $C_{ij} \neq C_{k\ell}$. We have, either $C_{ij} \neq C_{kj}$ or $C_{kj} \neq C_{k\ell}$, i.e., there exist two unequal coefficients with a common subscript. Assume without loss of generality, $C_{13} \neq C_{23}$. Let $U_3 = U - \{1,2\} = \{3, \ldots N\}$.

$$\text{Let } q(y_1, \ldots y_N) = \sum_{U(2)} C_{ij} f_{ij}$$

$$= C_{12} f_{12} + \sum_{k \varepsilon U_3} (C_{1k} f_{1k} + C_{2k} f_{2k}) + \sum_{\substack{k<\ell \\ \varepsilon U_3}} C_{k\ell} f_{k\ell}$$

$$0 = q(y_1, y_2, \ldots y_N) - q(y_2, y_1, y_3 \ldots y_N)$$

$$= \sum_{k=3}^{N} (C_{1k} - C_{2k})(f_{1k} - f_{2k})$$

For fixed y_1, y_2, define $f_*(x) = f(y_1, x) - f(y_2, x)$.

We have $\sum_{k=3}^{N} (C_{1k} - C_{2k})\, f_*(y_k) = 0$, with $C_{1k} - C_{2k} \neq 0$ for some k.

By lemma A3.2, $f_*(x)$ is constant, and hence, in particular,

$f_*(y_1) = f_*(y_2)$. That is,

$$f(y_1,y_1) - f(y_2,y_1) = f(y_1,y_2) - f(y_2,y_2),$$

giving $\quad f(y_1,y_2) = g(y_1) + g(y_2)$, where $g(x) = f(x,x)/2$.

Proof of theorem 3.1 for m>2 (by induction)

We have proved the theorem for $m = 2$. Now, assume it to be true for $m = k$. We will show it is true for $m = k+1$.

Suppose $f: A^{k+1} \to V$ is symmetric and (A3.1) holds for some constants C_I, not all zero. Along the lines of the proof for $m = 2$, we can show there are two unequal coefficients differing in only one subscript. Else, all C_I's must be equal and the result follows by lemma A3.1. We can therefore assume $C_{1J} \neq C_{2J}$ for some $J = \{j_1,..j_k\}$ $\varepsilon U_3(k)$ = {subsets of U_3 containing k elements}.

Proceeding further along the lines of the proof for $m = 2$, we get

$$\sum_{J\varepsilon U_3(k)} (C_{1J} - C_{2J})(f_{1J} - f_{2J}) = 0 \qquad \text{(A3.4)}$$

where

$$f_{iJ} = f(y_i,\ y_{j_1},...y_{j_k}),\ i = 1,2.$$

For fixed y_1, y_2, define $f_*(y_{j_1},...y_{j_k}) = f(y_1,y_{j_1},...y_{j_k})$ -

$$f(y_2,y_{j_2},...y_{j_k}),\ \text{or}\ f_{*J} = f_{1J} - f_{2J}.$$

Since f_{*J} is symmetric in $y_{j_1},...y_{j_k}$ and $\Sigma\,(C_{1J} - C_{2J})\ f_{*J} = 0$ with at least one $C_{1J} - C_{2J} \neq 0$, by the induction hypothesis, there exists a symmetric function h_*: $A^{k-1} \to V$, such that

$$f_{*J} = f_{1J} - f_{2J} = \sum_{L_2 \varepsilon J(k-1)} h_{*L_2} \tag{A3.5}$$

where J(k-1) = {subsets of J containing (k-1) elements},

$$L_r = (\ell_r, \ldots \ell_k),\ r = 2,\ 3,\ \text{and}\ h_{*L_2} = h_*(y_{\ell_2}, \ldots y_{\ell_k}).$$

Note that f_* and h_* above are defined for particular y_1 and y_2. If we let y_1 and y_2 vary, h_* in (A3.5) must be replaced by a function h which also depends on y_1 and y_2, (but not necessarily symmetric in all arguments). Substituting h $(y_1 \cdot y_2; y_{\ell_2}, \ldots y_{\ell_k})$ for $h_*(y_{\ell_2}, \ldots y_{\ell_k})$, (A3.5) can be written

$$f_{*J} = f_{1J} - f_{2J} = \sum_{L_2} h(y_1, y_2; y_{\ell_2}, \ldots y_{\ell_k})$$

Putting $(y_{j_1}, \ldots y_{j_k}) = (y_2, \ldots y_{k+1})$,

$$f(y_1, y_2, \ldots y_{k+1}) - f(y_2, y_2, \ldots y_{k+1}) = \sum_{L_2} h(y_1, y_2; y_{\ell_2}, \ldots y_{\ell_k})$$

$$= h(y_1, y_2; y_3, \ldots y_{k+1}) + \sum_{L_3 \varepsilon M_3} h(y_1, y_2; y_2, y_{\ell_3}, \ldots y_{\ell_k}) \tag{A3.6}$$

where

$$M_3 = \{\text{subsets of } \{3, \ldots. k+1\} \text{ containing } k-2 \text{ elements}\}$$

Similarly, putting $(y_{j_1}, \ldots y_{j_k}) = (y_1, y_3, \ldots y_{k+1})$, we get

$$f(y_1, y_1, y_3, \ldots y_{k+1}) - f(y_1, y_2, y_3, \ldots y_{k+1}) = h(y_1, y_2; y_3 \ldots y_{k+1})$$

$$+ \sum_{L_3 \varepsilon M_3} h(y_1,y_2;\ y_1,y_{\ell_3},\ldots y_{\ell_k}) \tag{A3.7}$$

Subtracting (A3.6) from (A3.7),

$$2f(y_1,y_2,\ldots k+1) = f(y_1,y_1,y_3,\ldots y_{k+1}) + f(y_2,y_2,y_3,\ldots y_{k+1})$$

$$+ \sum_{L_3 \varepsilon M_3} [h(y_1,y_2;y_2,y_{\ell_3},\ldots y_{\ell_k}) - h(y_1,y_2;y_1,y_{\ell_3},\ldots y_{\ell_k})]$$

We have thus expressed $f(y_1,\ldots y_{k+1})$ in the form

$$\sum_{J \varepsilon M} q^J(y_{j_1},\ldots y_{j_k}).$$

where M = {subsets of {1,...,k+1} containing k elements}.

The proof is now completed by lemma A3.3.

REFERENCES

Ajgaonkar, S. G. P. 1967. Unbiased estimator of variance of Narain, Horvitz and Thompson estimator. Sankhya, Ser. A., 29: 55-60.

Basu, D. 1971. An essay on the logical foundations of survey sampling. Pages 203-232 in V. P. Godambe and D. A. Sprott, eds. Foundations of statistical inference. Holt, Rinehart and Winston of Canada, Ltd., Toronto.

Cassel, C. M., C. E. Särndal and J. H. Wretman. 1977. Foundations of inference in survey sampling. John Wiley and Sons, New York.

Cochran, W. G. 1946. Relative accuracy of systematic and stratified random sampling for a certain class of populations. Ann. Math. Statist. 17: 164-177.

Cochran, W. G. 1977. Sampling techniques. John Wiley and Sons, New York.

Ericson, W. A. 1965. Optimal stratified sampling using prior information. J. Amer. Statist. Assoc. 60: 750-771.

Fellegi, I. P. 1963. Sampling with varying probabilities without replacement, rotating and non-rotating samples. J. Amer. Statist. Assoc. 58: 183-201.

Fuller, W. A. 1970. Sampling with random stratum boundaries. J. Roy. Statist. Soc., Ser. B, 32: 209-226.

Godambe, V. P. 1955. A unified theory of sampling from finite populations. J. Roy. Statist. Soc., Ser. B, 17: 269-278.

Godambe, V. P. and V. M. Joshi. 1965. Admissibility and Bayes estimation in sampling finite populations - I. Ann. Math. Statist. 36: 1707-1722.

Godambe, V. P. and M. E. Thompson. 1971. The specification of prior knowledge by classes of prior distributions in survey sampling estimation. Pages 243-258 in V. P. Godambe and D. A. Sprott, eds. Foundations of statistical inference. Holt, Rinehart and Winston of Canada, Ltd., Toronto.

Godambe, V. P. and M. E. Thompson. 1973. Estimation in sampling theory with exchangeable priors. Ann. Statist. 1: 1212-1221.

Hajek, J. 1959. Optimal strategy and other problems in probability sampling. Casopis Pest. Mat. 84: 382-423.

Hajek, J. 1964. Asymptotic theory of rejective sampling with varying probabilities for a finite population. Ann. Math. Statist. 35: 1491-1523.

Hanurav, T. V. 1965. Optimal sampling strategies and some related problems. Ph.D. Thesis. Indian Statistical Institute, Calcutta.

Hanurav, T. V. 1966. Some aspects of unified sampling theory. Sankhya, Ser. A, 28: 175-204.

Hanurav, T. V. 1967. Optimal Utilization of auxiliary information in πps sampling of two units from a stratum. J. Roy. Statist. Soc., Ser. B, 29: 374-91.

Hartley, H. O. and J. N. K. Rao. 1969. A new estimation theory for sample surveys - II. Pages 147-169 in N. L. Johnson and H. Smith, eds. New Developments in survey sampling. John Wiley and Sons, New York.

Hoeffding, W. 1948. A class of statistics with asymptotically normal distribution. Ann. Math. Statist. 19: 293-325.

Horvitz, D. G. and D. J. Thompson. 1952. A generalization of sampling without replacement from a finite universe. J. Amer. Statist. Assoc. 47: 663-685.

Joshi, V. M. 1965a. Admissibility and Bayes estimation in sampling finite populations - II. Ann. Math. Statist. 36: 1723-1729.

Joshi, V. M. 1965b. Admissibility and Bayes estimation in sampling finite populations - III. Ann. Math. Statist. 36: 1730-1742.

Joshi, V. M. 1966. Admissibility and Bayes estimation in sampling finite populations - IV. Ann. Math. Statist. 37: 1658-1670.

Joshi, V. M. 1969. Admissibility of estimators of mean of a finite population. Pages 188-212 in N. L. Johnson and H. Smith, eds. New developments in survey sampling. John Wiley and Sons, New York.

Joshi, V. M. 1970. Note on the admissibility of Sen-Yates-Grundy estimator. Sankhya, Ser. A, 32: 431-438.

Kempthorme, O. 1969. Some remarks on statistical inference in finite sampling. Pages 671-695 in N. L. Johnson and H. Smith, eds: New developments in survey sampling. John Wiley and Sons, New York.

Lanke, J. 1974. On non-negative variance estimation in sample surveys. Sankhya, Ser. C., 1936: 33-42.

Ogawa, J. 1974. Statistical theory of the analysis of experimental design. Marcel Dekker and Company, New York.

Rao, J. N. K. 1963. On three procedures of unequal probability sampling without replacement. J. Amer. Statist. Assoc. 58: 202-215.

Royall, R. M. 1968. An old approach to finite population sampling theory. J. Amer. Statist. Assoc. 63: 1269-1279.

Royall, R. M. 1970. On finite population sampling theory under certain linear regression models. Biometrika. 57: 377-387.

Royall, R. M. and J. Herson. 1973. Robust estimation in finite populations - I, II. J. Amer. Statist. Assoc. 68: 880-893.

Sampford, M. R. 1967. On sampling without replacement with unequal probabilities of selection. Biometrika. 54: 499-513.

Särndal, C. E. 1976. On uniformly minimum variance estimation in finite populations. Ann. Statist. 4: 994-7.

Searls, D. T. 1964. The utilization of a known coefficient of variation in the estimation procedure. J. Amer. Statist. Assoc. 59: 1225-1226.

Searls, D. T. 1967. Note on use of approximately known coefficient of variation. Amer. Statist. 21(3): 20-21.

Sekkappan, Rm. and M. E. Thompson. 1975. On a class of uniformly admissible estimators for finite populations. Ann. Statist. 3: 492-499.

Sukhatme, P. V. and B. V. Sukhatme. 1970. Sampling theory of surveys with applications. Iowa State University Press. Ames, Iowa.

Vijayan, K. 1975. On estimating the variance in unequal probability sampling. J. Amer. Statist. Assoc. 70: 713-716.

Yates, F. 1960. Sampling methods for censuses and surveys. Hafner Publishing Company, New York.

Yates, F. and P. M. Grundy. 1953. Selection without replacement from within strata with probability proportional to size. J. Roy Statist. Soc., Ser. B, 15: 235-261.

ACKNOWLEDGMENTS

I wish to sincerely thank Dr. B. V. Sukhatme for suggesting the area of research and for valuable guidance during various stages. I am also thankful to the faculty of the Department of Statistics, and in particular, Dr. H. A. David for constant encouragement and help during my stay at Iowa State. For financial support, I am grateful to the Statistical Laboratory and Agricultural Experiment Station of Iowa State University. Thanks are also due to Mrs. Kathy Crouch for her skill and patience in typing the manuscript and making numerous corrections.