ON ESTIMATION OF VARIANCE IN UNEQUAL PROBABILITY SAMPLING

by

Shriram H. Biyani

University of Minnesota

Technical Report No. 349

June, 1979

ABSTRACT

On Estimation of Variance in Unequal Probability Sampling

by

S.H. Biyani

Several estimators have been proposed for the variance of the Horvitz-Thompson estimator of the population total. Each of the well-known estimators can take values which are either negative, or positive but otherwise known to be impossible. In this paper, some alternative estimators are derived using a Random Permutation Model. It is shown that these estimators always take on possible values of the true variance. Relative efficiencies of these and some traditional estimators are numerically compared using a set of natural populations in which varying degrees of departure from the assumed model are observed. It is interesting to observe that the model-based estimators seem to have greater efficiencies relative to the "model-independent" ones in cases of severe model breakdown.

Key words and phrases: Finite population, Variance of Horwitz-Thompson estimator, Random Permutation model.

1. INTRODUCTION

Consider a finite population U consisting of N units labeled 1, 2, ..., N, and having values $y_1, y_2, \ldots, y_N$ respectively, for some characteristic y. Let p be a sampling design, which selects a sample s with probability p(s) from a collection S of possible samples. The data from a sample survey can be summarized as $d = \{(i,y_i), i\varepsilon s\}$, the collection of unit labels and the corresponding y-values. This includes any auxiliary information, which can be considered as a function of the unit lables. An estimator e is defined to be a function of the data d, and its value is denoted by $e(s,\underline{y})$, with the understanding that it depends on $\underline{y}$ only through $(y_i, i\varepsilon s)$.

<u>Definition 1.1.</u> With respect to a given sampling design p, an estimator e is said to be p-unbiased (design-unbiased) for a function $F(\underline{y})$, if

$$\sum_{s\varepsilon S} p(s)e(s,\underline{y}) = F(\underline{y}), \text{ for all } \underline{y}.$$

<u>Definition 1.2.</u> The mean square error (MSE) of an estimator e of a function $F(\underline{y})$, (with respect to a given design p), is

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

For a p-unbiased estimator e, the MSE is also the variance of the estimator, and is denoted by $V(e,\underline{y})$.

When there is no ambiguity, we will abbreviate $e(s,\underline{y})$ to e(s), $V(e,\underline{y})$ to V(e), etc.

For a given design p, let π_i be the inclusion probability of unit i, and let π_{ij} be the joint inclusion probability of units i and j, for i, j = 1,2,...,N. In particular, $\pi_{ii} = \pi_i$.

The sampling design will be assumed to be of fixed sample size throughout this paper.

The Horvitz-Thompson estimator of the population total is given by

$$e_{HT}(s) = \sum_{i \varepsilon s} z_i, \text{ where } z_i = y_i/\pi_i. \tag{1.1}$$

The variance of e_{HT} can be expressed in two equivalent forms as

$$V(e_{HT}) = \sum_{i,j=1}^{N} (\pi_{ij} - \pi_i \pi_j) z_i z_j, \tag{1.2}$$

and

$$V(e_{HT}) = \sum_{i<j}^{N} (\pi_i \pi_j - \pi_{ij})(z_i - z_j)^2. \tag{1.3}$$

Two p-unbiased estimators of $V(e_{HT})$, given by Horvitz and Thompson (1952), and Yates and Grundy (1953), respectively, are

$$v_{HT}(s) = \sum_{i,j \varepsilon s} (\pi_{ij} - \pi_i \pi_j) \pi_{ij}^{-1} z_i z_j, \tag{1.4}$$

and

$$v_{YG}(s) = \sum_{i<j \varepsilon s} \pi_i \pi_j - \pi_{ij}) \pi_{ij}^{-1} (z_i - z_j)^2. \tag{1.5}$$

It is well known that both estimators can take negative values, depending on the sampling design, and the population. A sufficient condition for the Yates-Grundy estimator v_{YG} to be nonnegative definite is, $\pi_{ij} \leq \pi_i \pi_j$ for all $i \neq j$. Therefore, v_{YG} is usually preferred to v_{HT}, which can take negative values for almost all designs. On the other hand, Godambe & Joshi (1965) proved that v_{HT} is admissible among p-unbiased estimators of $V(e_{HT})$, while the corresponding result for v_{YG} is only known to be true for sample size two, due to Joshi (1970). (Of course, the negative values of v_{HT} make it inadmissible, when we admit p-biased estimators.)

Biyani (1978) has shown that for any sample size greater than two, there are sampling designs, for which v_{YG} is inadmissible, even in the narrower class of nonnegative unbiased quadratic estimators. Further, even when it is nonnegative, it can take values known to be impossibly low.

It may be noted that the estimator e_{HT} for the population total is obtained by dividing the general term in $\sum_{i=1}^{N} y_i$ by the inclusion probability π_i, and restricting the sum to the sample instead of the population. Similarly the variance estimators v_{HT} and v_{YG} are obtained by dividing the general terms in the expressions (1.2) and (1.3) of $V(e_{HT})$, respectively, by the corresponding inclusion probabilities, and restricting the sum to the sample. The Horvitz-Thompson estimator of the total would intuitively seem reasonable, if y_i's are expected to be approximately proportional to π_i's. Formal optimality results for e_{HT} have been proved by Godambe and Joshi (1965), Godambe and Thompson(1971), and others, by invoking superpopulation models under which y_i/π_i, $i=1,2,\ldots, N$ have constant expectation and variance. However, no such justification can be given for v_{HT} and v_{YG}.

In this paper, we will consider alternative estimators of $V(e_{HT})$, based on a superpopulation model consistent with a mpdel under which e_{HT} is optimal. First, we consider a more general class of functions in Section 2. General solutions for the optimal "estimators" of these functions under a general symmetric model are derived. These solutions generally involve some unknown parameters, except in the special case of a symmetric function. In particular, some optimality results for the sample mean square as an estimator of the population mean square are obtained. In Section 3, a Random Permutation Model is considered as a special case of the general model mentioned above, and with some further assumptions, optimal estimators of $V(e_{HT})$ are obtained. In section 4, some "intuitive" and "optimal" model-based estimators are numerically compared with some of the proposed p-unbiased estimators of $V(e_{HT})$, using some natural populations.

2. MODEL-BASED APPROACH

In classical sample survey theory, the population vector $\underline{y}$ is treated as a fixed unknown parameter. A "Superpopulation model" regards $\underline{y}$ as the realized value of a random vector $\underline{Y} = (Y_1, Y_2, \ldots, Y_N)$, having some distribution ξ. The problem of estimating any function $F(\underline{y})$ can then be considered as a prediction problem, but we will make no distinction here between estimators and predictors. Making certain assumptions about ξ to represent our prior information about the population, we can obtain optimal estimators (predictors) for various problems. We will use the expected MSE as the optimality criterion.

Definition 2.1. An estimator e of a function $F(\underline{y})$ is said to be optimal (or best) in a class K of estimators, under a given model, if for every other estimator e' in K,

$$E_\xi \sum_{s\in S} p(s)[e(s,\underline{Y}) - F(\underline{Y})]^2 \leq E_\xi \sum_{s\in S} p(s)[e'(s,\underline{Y}) - F(\underline{Y})]^2,$$

for all possible distributions ξ of $\underline{Y}$ under the model.

In general, an estimator minimizing the expected mean square error among all estimators does not exist, and we restrict to some suitable class of estimators.

Two types of unbiasedness restrictions are often considered: p-unbiasedness defined in Section 1, and ξ-unbiasedness (model-unbiasedness) defined below.

Definition 2.2. An estimator e is said to be ξ-unbiased for a function F, if for every $s\in S$,

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

Another condition weaker than both p and ξ-unbiasedness is pξ-unbiasedness, introduced by Cassel, Särndal and Wretman (1977).

Definition 2.3. An estimator e is said to be pξ-unbiased for a function F, if

$$E_\xi \sum_{s\varepsilon S} p(s)[e(s,\underline{Y}) - F(\underline{Y})] = 0.$$

We will next consider a functional form which includes the population variance and $V(e_{HT})$ as particular cases.

2.1. A Class of Functions

Let $f(x,y)$ be a symmetric function, i.e., $f(x,y) = f(y,x)$. Define, $f_{ij}(\underline{y}) = f(y_i,y_j)$. For brevity, we will write f_{ij} for $f_{ij}(\underline{y})$. Consider a function of the form

$$F(\underline{y}) = \sum_{i<j}^{N} c_{ij} f_{ij}, \qquad (2.1)$$

where c_{ij}'s are known constants.

The population means square S^2, defined by

$$S^2 = (N-1)^{-1} \sum_{i=1}^{N} (y_i - \bar{y})^2, \text{ where } \bar{y} = \sum_{i=1}^{N} y_i/N$$

can be expressed in the form (2.1) with $c_{ij} = [N(N-1)]^{-1}$ and $f_{ij} = (y_i - y_j)^2$.

From (1.3), $V(e_{HT})$ is also seen to be of the form (2.1), with $\underline{y}$ replaced by $\underline{z} = (z_1,z_2,\ldots,z_N)$, $c_{ij} = \pi_i\pi_j - \pi_{ij}$, and $f_{ij} = (z_i - z_j)^2$.

Vijayan (1975) has shown that any nonnegative p-unbiased estimator v of $V(e_{HT})$ must be of the form

$$v(s) = \sum_{i<j\varepsilon s} b_{sij}(z_i - z_j)^2 \tag{2.2}$$

The same condition is obtained if the requirement of p-unbiasedness is replaced by the condition, that when $z_1 = z_2 = \ldots = z_N$, v(s) must vanish. This seems reasonable, because one would expect an estimator to be without error, when there is no variability in the population. Similarly, any nonnegative p-unbiased quadratic estimator of the population mean square must be of the form $\Sigma b_{sij}(y_i - y_j)^2$, if we require that the estimator vanish when $S^2 = 0$.

Henceforth, we will only consider estimators of the form

$$e(s) = \sum_{i<j\varepsilon s} b_{sij} f_{ij}. \tag{2.3}$$

2.2. A General Symmetric Model

Consider the following superpopulation model, to be referred to as Model I.

$$E_\xi[f_{ij}(\underline{x})] = m \tag{2.4}$$

$$E_\xi[f_{ij}(\underline{x}) f_{kh}(\underline{y})] = \begin{cases} a_2, & \text{if } i = k \neq j = h, \\ a_1, & \text{if } i = k, \text{ and } i,j,h \text{ are distinct}, \\ a_0, & \text{if } i,j,k,h \text{ are all distinct}, \end{cases} \tag{2.5}$$

where m, a_0, a_1, a_2 are constants.

Note that in (2.5), the product moment of two f's depends only on the number of common subscripts. In particular, all exchangeable distributions ξ of $\underline{Y}$ are included under Model I.

We will consider the following classes of estimators, defined for any given function.

$$K_0 = \{\text{all estimators of the form (2.3)}\}, \tag{2.6}$$

K_ξ = {all ξ-unbiased estimators of the form (2.3)}, and

K_p, $K_{p\xi}$ defined similarly.

It is easy to see, that within the classes K_0 and K_ξ, minimizing $E_\xi \sum_s p(s)[e(s,\underline{Y}) - F(\underline{Y})]^2$ is equivalent to minimizing $E_\xi[e(s,\underline{Y}) - F(\underline{Y})]^2$ for each s, and hence the best estimators in K_0 and K_ξ do not depend on the sampling design. The best estimators in K_p and $K_{p\xi}$, in general, depend on p.

2.3. Equations for Optimal Coefficients

Let $Q = E_\xi \sum_s p(s)[e(s,\underline{Y}) - F(\underline{Y})]^2$. The optimal estimator of $F(\underline{y}) = \sum_{i<j\varepsilon U} c_{ij} f_{ij}$ in the class K_0 is obtained by minimizing Q with respect to b_{sij}. The equations giving the optimal values of the b's are:

$$\frac{\partial Q}{\partial b_{sij}} = 0, \text{ or}$$

$$H_{sij} = 0, \quad i<j\varepsilon s, \ s\varepsilon S, \tag{2.7}$$

where $H_{sij} = E_\xi\ [\{e(s,\underline{Y}) - F(\underline{Y})\} f_{ij}(Y)]$.

To obtain the best estimator in K_ξ, we minimize Q, subject to the ξ-unbiasedness condition

$$E_\xi\ [\sum_{i<j\varepsilon s} b_{sij} f_{ij}(\underline{Y}) - \sum_{i<j\varepsilon U} c_{ij} f_{ij}(\underline{Y})] = 0, \ s\varepsilon S.$$

By (2.4), this reduces to

$$\sum_{i<j\varepsilon s} b_{sij} = C_U, \tag{2.8}$$

where $C_U = \sum_{i<j\varepsilon U} c_{ij}$.

The equations for b_{sij} are:

$$\frac{\partial}{\partial b_{sij}} [Q - \sum_{s\varepsilon S}\lambda_s(\sum_s b_{sij} - C_U)] = 0, \text{ or}$$

$$H_{sij} - \lambda_s = 0, \tag{2.9}$$

where λ_s, $s\varepsilon S$ are Lagrangian multipliers.

Similarly, the best estimator in K_p is given by

$$\sum_{s \ni i,j} p(s) b_{sij} = c_{ij}, \quad i<j\varepsilon U \tag{2.10}$$

$$\text{and } H_{sij} - \lambda_{ij} = 0, \ i<j\varepsilon s, \ s\varepsilon S, \tag{2.11}$$

where λ_{ij}, $i<j\varepsilon U$ are constants.

The best estimator in $K_{p\xi}$ is given by

$$\sum_{s\varepsilon S} p(s) \sum_{i<j\varepsilon s} b_{sij} = C_U, \tag{2.12}$$

$$\text{and } H_{sij} - \lambda = 0, \quad i<j\varepsilon s, \ s\varepsilon S. \tag{2.13}$$

For each sample s, (2.7) gives a system of $\binom{n}{2}$ equations in as many coefficients. Although this can be a very large number, the structure of the system enables us to solve it explicitly. Solutions of (2.8) - (2.9), and (2.12) - (2.13) can be easily derived from the solution of (2.7), but the solution of (2.10) - (2.11) seems to require the inversion of a large matrix, and will not be considered here. The solutions for the remaining systems are given below. The details of the derivation are given in the Appendix.

2.4. Solutions for the Optimal Coefficients

Let,

$$\tilde{s} = U - s$$

$$C_s = \sum_{i<j\varepsilon s} c_{ij}$$

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

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

$$C_{\tilde{s}} = \sum_{k,h\varepsilon\tilde{s}} c_{kh}$$

$$C_U = \sum_{i<j\varepsilon U} c_{ij} = C_s + C_{\tilde{s}} + C_{\tilde{s}s}$$

$$d_1 = a_1 - a_0$$

$$d_2 = a_2 - 2a_1 + a_0$$

$$t_0 = \binom{n}{2}a_0 + 2(n-1)d_1 + d_2 = \binom{n-2}{2}a_0 + 2(n-2)a_1 + a_2$$

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

The solution of (2.7) is,

$$b^o_{sij} = c_{ij} + d_1(C_{\tilde{s}i} + C_{\tilde{s}j} - 2C_{\tilde{s}s}/n)/t_1$$

$$+ [a_0(C_{\tilde{s}} + C_{\tilde{s}s}) + 2d_1C_{\tilde{s}s}/n)]/t_0, \qquad (2.14)$$

$$i<j\varepsilon s, \; s\varepsilon S.$$

The solution of (2.8) - (2.9) is

$$b^{\xi}_{sij} = c_{ij} + d_1(C_{\tilde{s}i} + C_{\tilde{s}j} - 2C_{\tilde{s}s}/n)/t_1 + (C_{\tilde{s}} + C_{\tilde{s}s})/\binom{n}{2}. \qquad (2.15)$$

The solution of (2.12) - (2.13) is

$$b^{p\xi}_{sij} = b^o_{sij} + K, \qquad (2.16)$$

where $K = \sum_{i<j\varepsilon U} c_{ij}[(n-1)d_1(2-\pi_i-\pi_j) + d_2(1-\pi_{ij})]/[\binom{n}{2}t_0]$. (2.17)

The solutions (2.14) - (2.16) depend on the ratios a_0/a_2 and a_1/a_2. When these ratios are not known no (uniformly) best estimator under Model I exists, in general. In order to obtain an optimal estimator, further assumptions about the distribution of $\underline{Y}$ are necessary. This is taken up in Section 3. A special case where the solutions (2.15) and (2.16) do not depend on a_0/a_2 and a_1/a_2, is considered below.

2.5. The Symmetric Case

If c_{ij} = constant for all (i,j) i<j, then $F(\underline{Y}) = \sum_{i<j\varepsilon U} c_{ij} f_{ij}$ becomes a "U-statistic", and the corresponding sample U-statistic is its optimal estimator in the classes $K_{p\xi}$ and K_ξ , irrespective of the sampling design. However, no optimal estimator in K_0 exists, unless the ratios a_0/a_2 and a_1/a_2 are specified.

Theorem 2.1: Under Model I, the best estimator of $F(\underline{Y}) = \binom{N}{2}^{-1} \sum_{i<j\varepsilon U} f_{ij}$ in the classes K_ξ and $K_{p\xi}$ (defined by 2.6) is given by $e^*(s) = \binom{n}{2}^{-1} \sum_{i<j\varepsilon s} f_{ij}$.

Proof: Putting $c_{ij} = \binom{N}{2}^{-1}$ in (2.15) and simplifying, we get $b^\xi_{sij} = \binom{n}{2}^{-1}$, which shows the optimality of e* in the class K_ξ. The result for $K_{p\xi}$ similarly follows from (2.16).

Consider the special case $f_{ij} = (y_i - y_j)^2$, and $c_{ij} = [n(n-1)]^{-1}$, giving $F(\underline{y}) = \sum_U c_{ij} f_{ij} = S^2$. The corresponding optimal estimator in K_ξ and $K_{p\xi}$ is the sample mean square $s^2 = \sum_{i\varepsilon s} (y_i - \bar{y}_s)^2$, where $\bar{y}_s = n^{-1} \sum_{i\varepsilon s} y_i$. Since K_ξ and $K_{p\xi}$ include, respectively, the nonnegative quadratic ξ- and pξ-unbiased estimators of S^2 vanishing when $S^2 = 0$, we obtain the following corollary.

Corollary 2.1. Under Model I, the sample mean square S^2 is the best ξ- and the best pξ-unbiased estimator of the population mean square S^2 among all nonnegative quadratic estimators vanishing when $S^2 = 0$.

3. RANDOM PERMUTATION MODELS

A superpopulation model requiring minimal input of prior information is a Random Permutation Model (RPM). Under the simplest RPM, the random vector $\underline{Y}$ is a random permutation of a fixed vector, i.e.,

$$P(\underline{Y} = \underline{w}^*) = \frac{1}{N!}, \qquad (3.1)$$

for every permutation $\underline{w}^*$ of a fixed, unknown vector $\underline{w} = (w_1,\ldots,w_N)$. We will denote this as Model II. This may be interpreted as the assumption of a lack of association between the unit labels and the y-values. This model has been considered, in particular, by Madow and Madow (1944) and Kempthorne (1969), for the problem of estimation of the population mean.

The distribution of $\underline{Y}$ under Model II is exchangeable, and hence it is a special case of Model I. Other RPMs can be obtained by considering some function $g(i,y_i)$ of the unit label and the y value, and regarding $[g(1,y_1),g(2,y_2),\ldots,g(N,y_N)]$ as a random permutation of a fixed, unknown vector. In particular, we will consider the case $g(i,y_i) = y_i/\pi_i$ in Section 3.3.

3.1. Moments under Model II

An important feature of a RPM is that the moments and product moments of the components of the random vector under the model are the corresponding moments of the realized finite population. For example, let ξ denote the distribution under which $\underline{Y}$ is a random permutation of a fixed vector $\underline{w} = (w_1,\ldots,w_N)$. Then

$$E_\xi(Y_1) = \frac{1}{N}\sum_{i=1} w_i = \frac{1}{N}\sum_{i=1} y_i = \bar{y},$$

since the realized value $\underline{y}$ is a permutation of $\underline{w}$.

$$\text{Let } \mu = E_\xi(Y_1)$$
$$\mu_r = E_\xi(Y_1-\mu)^r$$
$$\mu_{rs} = E_\xi[(Y_1-\mu)^r(Y_2-\mu)^s], \text{ etc.}$$

The following relations can be easily proved.

$$\mu_{22} = (N-1)^{-1}(N\mu_2^2-\mu_4) \tag{3.2}$$
$$\mu_{31} = -\mu_4/(N-1) \tag{3.3}$$
$$\mu_{211} = -(\mu_{22}+\mu_{31})/(N-2) \tag{3.4}$$
$$\mu_{1111} = -3\mu_{211}/(N-3) \tag{3.5}$$

For $f_{ij} = (y_i - y_j)^2$, the values of a_0, a_1, a_2 [defined by (2.5)] can be expressed as follows. With i, j, k, ℓ distinct,

$$a_0 = E_\xi[(Y_i-Y_j)^2(Y_k-Y_\ell)^2] = 4(\mu_{22}-2\mu_{211}+\mu_{1111}) \tag{3.6}$$
$$a_1 = E_\xi[(Y_i-Y_j)^2(Y_i-Y_k)^2] = \mu_4-4\mu_{31}+3\mu_{22} \tag{3.7}$$
$$a_2 = E_\xi(Y_i-Y_j)^4 = 2(\mu_4-4\mu_{31}+3\mu_{22}) \tag{3.8}$$

From (3.7) and (3.8),

$$a_1/a_2 = 1/2 \tag{3.9}$$

From (3.6) and (3.8), using (3.1)-(3.5) and simplifying, we obtain

$$a_0/a_2 = 2[N^2 - (N-1)(\beta_2+3)] \;/\; [(N-2)(N-3)(\beta_2+3)], \tag{3.10}$$

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

3.2. Optimal Estimator of S^2 under Model II

Since Model II is a special case of Model I, s^2 remains the optimal estimator of S^2 under model II, in the classes K_ξ and $K_{p\xi}$ [defined in (2.6)]. To obtain the optimal estimator in K_0 under Model II, we put $c_{ij} = 1/[N(N-1)]$, and a_0/a_2, a_1/a_2 from (3.9)-(3.10) in (2.14). After simplification, we get

$$b^o_{sij} = \frac{(N-2)(N-3)}{N(N-1)} \cdot \frac{1}{n(n+1) + (n-1)(\beta_2-3) + R}, \qquad (3.11)$$

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

Since, the solution depends on β_2, no optimal estimator of S^2 exists in K_0, under Model. II, unless the model is further restricted by specifying β_2. For large N, and $\beta_2 = 3$ (corresponding to a normal shape), we get $b^o_{sij} = 1/n(n+1)$, giving the well-known estimator

$$[n(n+1)]^{-1} \sum_{i<j\epsilon s} (y_i-y_j)^2 = (n-1)/(n+1) \cdot s^2.$$

<u>Remark 3.1.</u> When β_2 is not known, it can be easily shown that substituting any underestimate of β_2 would give an estimator better than s^2. In particular, substituting the smallest value $\beta_2 = 1$ gives, for large N,

$b^o_{sij} = (n^2-n+2)^{-1}$, giving the estimate

$$e^*(s) = \sum_{i<j\epsilon s} (n^2-n+2)^{-1} (y_i-y_j)^2 = [1 + \frac{2}{n(n-1)}]^{-1} s^2. \qquad (3.12)$$

<u>Remark 3.2.</u> If the sampling design is Simple Random Sampling (SRS), then the MSE of any multiple of s^2 is constant over all permutations of the population vector, and hence the expected MSE under Model II is the same as the actual MSE for the realized population. It follows that s^2 is inadmissible as an estimator of S^2, the estimator e^* given by (3.12) being uniformly better than s^2.

<u>Remark 3.3.</u> If we have a sample of n independent observations from any population (not necessarily finite) then e* is uniformly better than s^2 as an estimator of the population variance, if the population has a finite fourth moment. This result, easily obtainable by considering multiples of s^2, does not seem to be well known.

3.3. Optimal Estimators for $V(e_{HT})$

To obtain optimal estimators for $V(e_{HT})$, we consider the model under which $\underline{Z} = (Z_1,\ldots,Z_N)$, with $Z_i = Y_i/\pi_i$, is a random permutation of a fixed vector, i.e.,

$$P(\underline{Z} = u^*) = \frac{1}{N!}, \tag{3.13}$$

for every permutation u* of a fixed vector u. We will denote this as Model III. Godambe and Thompson (1971) showed the optimality of e_{HT} under this model, in the class of p-unbiased estimators of the population total. Define $f_{ij}(\underline{z}) = (z_i - z_j)^2$, and a_o, a_1, a_2 as in (2.5), with $\underline{Z}$ in place of $\underline{Y}$. The values of a_1/a_2 and a_0/a_2 under Model III are given by (3.9) and (3.10), with β_2 replaced by $\beta_{2z} = [\ \sum^N (z_i - \bar{z})^4/N\ ] \ /\ [\ \sum^N (z_i - \bar{z})^2/N\ ]^2$.

The optimal coefficients can be obtained from (2.14)-(2.16), substituting for a_0/a_2 and a_1/a_2. Since a_0/a_2 depends on β_{2z}, in general no best estimator of $V(e_{HT})$ exists in K_0, K_ξ or $K_{p\xi}$, unless β_{2z} is specified.

For $\beta_{2z} = 3$, and large N, we get the following approximations for the optimal coefficients in the classes K_0, K_ξ and $K_{p\xi}$ respectively. [These are the exact solutions for $\beta_{2z} = 3(N-1)/(N+1)$.]

$$b^{o}_{sij} = c_{ij} + (C_{\tilde{s}i}+C_{\tilde{s}j})/n + 2C_{\tilde{s}}/[n(n+1)] \tag{3.14}$$

$$b^{\xi}_{sij} = c_{ij} + (C_{\tilde{s}i}+C_{\tilde{s}j})/n + 2C_{\tilde{s}}/[n(n-1)] + 2C_{\tilde{s}s}/[n^2(n-1)], \tag{3.15}$$

$$b^{p\xi}_{sij} = b^{o}_{sij} + K, \tag{3.16}$$

where $K = 2\sum_{i<j}^{N} c_{ij}[2 - \pi_i - \pi_j + 2(1-\pi_{ij})/(n-1)]/[n^2(n+1)]$, (3.17)

$c_{ij} = \pi_i\pi_j - \pi_{ij}$, $C_{\tilde{s}i} = \sum_{k\notin s} c_{ik} = \pi_i(1-\pi_i) - \sum_{k\in s} c_{ik}$, $C_{\tilde{s}s} = \sum_{i\in s} C_{\tilde{s}i}$, and

$C_{\tilde{s}} = \sum_{k<\ell\in\tilde{s}} c_{k\ell} = \sum_{i=1}^{N}\pi_i(1-\pi_i)/2 - C_{s\tilde{s}} - \sum_{i<j\in s} c_{ij}$.

The optimal estimators are obtained by substituting the above values for the coefficients in (2.3).

<u>Theorem 3.1.</u> Under Model III (defined by (3.13), with $\beta_{2z} = 3$ and N large, the optimal estimators of $V(e_{HT})$ in the classes K_0, K_ξ and $K_{p\xi}$ [defined by (2.6)] are, respectively,

$$v_0(s) = \sum_{i<j\in s} c_{ij}f_{ij} + (n-1)/n \sum_{i\in s} C_{\tilde{s}i}\bar{f}_{is} + (n-1)/(n+1)\cdot C_{\tilde{s}}\bar{f}_s, \tag{3.18}$$

$$v_\xi(s) = \sum_{i<j\in s} c_{ij}f_{ij} + \sum_{i\in s} C_{\tilde{s}i}[(n-1)\bar{f}_{is}+\bar{f}_s]/n + C_{\tilde{s}}\bar{f}_s, \tag{3.19}$$

and

$$v_{p\xi}(s) = v_0(s) + K\binom{n}{2}\bar{f}_s, \tag{3.20}$$

where K is defined by (3.17), $f_{ij} = (y_i/\pi_i - y_j/\pi_j)^2$, $\bar{f}_{is} = \sum_{j\in s} f_{ij}/(n-1)$ and $\bar{f}_s = \sum_{i<j\in s} f_{ij}/\binom{n}{2} = \sum_{i\in s}\bar{f}_{is}/n$, and c_{ij}, $C_{\tilde{s}i}$, $C_{\tilde{s}s}$ and $C_{\tilde{s}}$ are as defined above. (3.21)

Remark 3.4. In (3.18) and (3.19), the first sum represents the terms in expression (1.3) of $V(e_{HT})$, corresponding to the sampled pairs of units. The second term predicts the contribution from pairs with one unit observed and one unobserved, and the last term predicts the contribution from entirely unobserved pairs. Note that the last two terms of v_0 are "shrunken" by factors of $(n-1)/n$ and $(n-1)/(n+1)$ respectively. The term $K\binom{n}{2}\bar{f}_s$ in (3.20) makes an adjustment to v_0 for achieving $p\xi$-unbiasedness.

Remark 3.5. The computation of the estimators v_0 and v_ξ does not require much additional work compared to the Yates-Grundy estimator, since most effort is required for the computation of the π_{ij}'s. However, the computation of K, and hence of $v_{p\xi}$, may become impractical for very large N.

We next show that the estimators v_0, v_ξ, and $v_{p\xi}$ always take possible values of $V(e_{HT})$ given the sample.

3.4. A desirable property of the optimal estimators.

In the following theorem, the notation $v_0(s, z_i, i\varepsilon s)$ is used to denote the value of v_0 for given s and $z_i, i\varepsilon s$. A similar notation is used for other estimators. The value of $V(e_{HT})$ for given $(z_1, z_2, \ldots, z_N)$ is denoted by $V(e_{HT}, z_1, z_2, \ldots, z_N)$.

Theorem 3.2: Let p be any fixed sample size design with $\pi_i > 0$ for $i = 1,2,\ldots,N$, and let v_0, v_ξ, $v_{p\xi}$ be as defined in (3.18)-(3.20).

(i) For a given sample s, and z_i $i\varepsilon s$, there exists a choice of values of z_i, $i\notin s$, such that

$$V(e_{HT}, z_1, z_2, \ldots, z_N) = v_0(s, z_i i\varepsilon s)$$

(ii) A similar result holds for v_ξ and $v_{p\xi}$

Proof: (i) Assume without loss of generality, that $s = \{1,2,\ldots n\}$, and let $z_1, z_2, \ldots z_n$ be given. We need to show that for suitable choice of $z_{n+1}, \ldots z_N$,

$$V(e_{HT}, z_1, \ldots z_n, z_{n+1}, \ldots, z_N) = v_0(s, z_1, \ldots, z_n) \quad . \tag{3.22}$$

To prove the result, we will define a posterior distribution η for $(Z_{n+1}, \ldots, Z_N)$, such that

$$E_\eta V(e_{HT}, z_1, \ldots, z_n, Z_{n+1}, \ldots, Z_N) = v_0(s, z_1, \ldots, z_n) \ , \tag{3.23}$$

and the possible values of $(Z_{n+1}, \ldots, Z_N)$ under η are limited to a finite set. This would imply that the right hand side lies between two possible values of $V(e_{HT})$ given $z_1, \ldots, z_n$. Since $V(e_{HT}, z_1, \ldots, z_N)$ is a continuous function of $z_{n+1}, \ldots, z_N$, by the "intermediate-value property", there exists a choice of $z_{n+1}, \ldots, z_N$ for which (3.22) holds. It remains to find a distribution η for which (3.23) holds. Consider two distributions η_1 and η_2 defined as follows.

Let, under η_1, $Z_{n+1}, \ldots, Z_N$ be independently distributed, each with the empirical distribution of the sample, i.e.,

$$P_{\eta_1}(Z_k = z_i) = 1/n, \ i=1,\ldots,n \ ; \ k=n+1,\ldots,N \ .$$

Let, under η_2, $Z_{n+1}, \ldots, Z_N$ be perfectly correlated, each with the empirical distribution of the sample, i.e.,

$$P_{\eta_2}[(Z_{n+1}, \ldots, Z_N) = (z_i, \ldots, z_i)] = 1/n, \ i=1,\ldots,n \ .$$

For $i=1,\ldots,n$ and $n+1\leq k\neq h\leq N$,

$$E_{\eta_1}(z_i-Z_k)^2 = E_{\eta_2}(z_i-Z_k)^2 = \sum_{j=1}^{n}(z_i-z_j)^2/n = (n-1)n^{-1}\bar{f}_{is} , \tag{3.24}$$

$$E_{\eta_1}(Z_k-Z_h)^2 = n^{-2}\sum_{i,j=1}^{n}(z_i-z_j)^2 = (n-1)n^{-1}\bar{f}_s , \tag{3.25}$$

and $$E_{\eta_2}(Z_k-Z_h)^2 = 0 , \tag{3.26}$$

where $\bar{f}_{is}$ and $\bar{f}_s$ are defined by (3.21)

Define $\eta = (n\eta_1 + \eta_2)/(n+1)$. From (3.24),

$$E_\eta(z_i-Z_k)^2 = (n-1)n^{-1}\bar{f}_{is} , \qquad 1\leq i\leq n, \quad n+1\leq k\leq N . \tag{3.27}$$

From (3.25) and 3.26) ,

$$E_\eta(Z_k-Z_h)^2 = (n-1)(n+1)^{-1}\bar{f}_s , \quad n+1\leq k\neq h\leq N . \tag{3.28}$$

We have

$$V(e_{HT},z_1,\ldots z_n,Z_{n+1},\ldots,Z_N) = \sum_{i<j\varepsilon s} c_{ij}f_{ij} + \sum_{\substack{i\varepsilon s\\ k\notin s}} c_{ik}(z_i-Z_k)^2 + \sum_{k<h\varepsilon\tilde{s}} c_{kh}(Z_k-Z_h)^2 .$$

Taking expectation with respect to η, it is easy to see that $E_\eta V(e_{HT})$ reduces to $v_0(s)$ given by (3.18). This proves the first part of the theorem.

(ii) To prove the second part, let s_0 be any given sample. If s_0 is the entire population, the result is trivial. Otherwise, let $k\notin s_0$. We must have $0<\pi_k<1$. Coefficient of z_k^2 in $V(e_{HT})$ is $\pi_k(1-\pi_k)>0$. Thus, we can make $V(e_{HT})$ arbitrarily large by choosing z_k sufficiently large. It follows that if M is any possible value of $V(e_{HT})$ given $z_i i\varepsilon s$, then any real number greater than M is also a possible value. To prove that $v_\xi(s_0)$ and $v_{p\xi}(s_0)$ are possible values of $V(e_{HT})$ given $z_i i\varepsilon s_0$, it is sufficient to prove that

$v_\xi(s_0) \geq v_0(s_0)$ and $v_{p\xi}(s_0) \geq v_0(s_0)$.

Consider first, $v_\xi(s_0)$. From (3.18) and (3.19),

$$v_\xi(s_0) - v_0(s_0) = h_{s_0} \bar{f}_{s_0} , \tag{3.29}$$

where h_{s_0} is a constant. We need to show $h_{s_0} \geq 0$.

Let ξ be any distribution of $\underline{Z}$ under Model III. By Lemma 3.1 (given following the theorem),

$$E_\xi v_0(s_0) \leq E_\xi V(e_{HT})$$

$$= E_\xi[v_\xi(s_0)] \qquad \text{by } \xi\text{-unbiasedness of } v_\xi.$$

By (3.29), $h_{s_0} E_\xi(\bar{f}_{s_0}) \leq 0$. Since $\bar{f}_{s_0}$ is nonnegative, and not identically zero, $h_{s_0} \leq 0$, which proves the result for v_ξ.

Finally, to prove the result for $v_{p\xi}$, observe that

$$v_{p\xi}(s_0) - v_0(s_0) = K\bar{f}_{s_0} , \tag{3.31}$$

and it is sufficient to show that $K \geq 0$.

Since v_ξ and $v_{p\xi}$ are both $p\xi$ unbiased under Model III,

$$E_\xi \sum_{s \in S} p(s)[v_\xi(s) - v_{p\xi}(s)] = 0 .$$

By (3.29) and (3.31) , $\quad \sum_{s \in S} p(s)(h_s - K) E_\xi \bar{f}_s = 0$.

But, under Model III, $E_\xi \bar{f}_s$ is independent of s. Thus, $\sum_{s \in S} p(s)(h_s - K) = 0$. But $h_s \geq 0$, as we have shown above. Therefore, $K \geq 0$ and the proof is complete.

Remark 3.6: Theorem 3.2 implies, in particular, that the estimators v_0, v_ξ and $v_{p\xi}$ are nonnegative.

Lemma 3.1: Let $(\underline{X},\underline{\theta})$ be a random vector with distribution ξ. Let $e_0(\underline{X})$ be a nonnegative predictor of a nonnegative function $F(\underline{X},\underline{\theta})$, with the smallest mean square error (with respect to ξ) among all constant multiples of e_0. Then, $E_\xi e_0(\underline{X}) \leq E_\xi F(\underline{X},\underline{\theta})$.

Proof: If $E_\xi e_0(\underline{X}) = 0$, the result is trivial. Otherwise, let

$$t = E_\xi F(\underline{X},\underline{\theta}) / E_\xi e_0(\underline{X}), \text{ or } E_\xi[te_0(\underline{X}) - F(\underline{X},\underline{\theta})] = 0. \quad (3.32)$$

We must show $t \geq 1$.

Proof: Let $M_\xi(e)$ and $V_\xi(e)$ denote, respectively, the mean square error and the variance of a predictor e of $F(\underline{X},\underline{\theta})$. Since $e = e_0$, minimizes $M_\xi(e)$ among all multiples of e_0, we have, in particular,

$$\begin{aligned} M_\xi(e_0) &\leq M_\xi(te_0) \\ &= V_\xi(te_0) \qquad \text{(since } te_0 \text{ is } \xi\text{-unbiased by 3.32)} \\ &= t^2 V_\xi(e_0) \\ &\leq t^2 M_\xi(e_0) \quad . \end{aligned}$$

Hence, $1 \leq t^2$. By (3.32), t must be nonnegative, since e_0 and F are nonnegative. Hence $t \geq 1$.

To apply the Lemma to (3.30), let $\underline{X} = (Z_1,\ldots,Z_n)$, $\theta = (Z_{n+1},\ldots,Z_N)$, $F = V(e_{HT})$, $e_0 = v_0(s_0)$, and observe that for each sample s, $v(s) = v_0(s)$ minimizes $E_\xi[v(s,Z_i \ i\varepsilon s) - V(e_{HT},Z_1,\ldots,Z_N)]^2$ in the class K_0, which includes all multiples of v_0.

4. NUMERICAL COMPARISONS

An empirical study was made to compare the relative efficiencies of several estimators of $V(e_{HT})$, the variance of the Horvitz-Thompson estimator. The estimators compared include v_{HT}, v_{YG}, v_0, $v_{p\xi}$ defined by (1.4), (1.5), (3.18) and (3.20) respectively, and three other estimators defined below.

$$v_A(s) = \sum_{i<j\varepsilon s} c_{ij} f_{ij} / [\binom{N-2}{n-2} p(s)] \quad \text{(due to Ajgaonkar 1967)},$$

$$v_F(s) = (\sum_s c_{ij} f_{ij}/\pi_{ij})(\sum_s c_{ij}/\pi_{ij})^{-1} \sum_{i<j}^{N} c_{ij} \quad \text{(due to Fuller 1970)},$$

and

$$v_R(s) = (\sum_s c_{ij} f_{ij})(\sum_s c_{ij})^{-1} \sum_{i<j}^{N} c_{ij},$$

where $c_{ij} = \pi_i \pi_j - \pi_{ij}$, $z_i = y_i/\pi_i$, and $f_{ij} = (z_i - z_j)^2$.

The estimators v_{HT}, v_{YG}, and v_A are design-unbiased, while v_F is approximately so. Under Model I, the ratio-type estimators v_F and v_R are both ξ-unbiased and $v_{p\xi}$ is $p\xi$-unbiased. It should be noted that v_0 and v_R are design-independent in the sense that they depend on the design only through the coefficients c_{ij}. The estimator $v_{p\xi}$ is basically model-based, but it is slightly design-dependent, as it includes a term (of order 1/n) involving the design parameters π_i, π_{ij}.

The populations used in the study are listed in Table 1. In each case, we have an auxiliary variable x approximately proportional to the variable y of interest. For most of the populations, the Horvitz-Thompson estimator can be expected to perform reasonably well as an estimator of the population total. However, for the purpose of illustration, population 10, containing at least one "wild" observation is also included. Population 9 is a trimmed version of 10, excluding two units with the smallest and the

Table 1. Population 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	Sukhatme and Sukhatme (1970, p.166)	no. of banana bunches	no. of banana pits	20	3.1
4	Sukhatme and Sukhatme (1970, p.51)	area under rice	total cultivated area	25	1.9
5	Rao (1963, p.207)	1960 area under corn	1958 area under corn	14	1.9
6	Cochran (1977, p.203)	weight of peaches	eye-estimate	10	1.4
7	Cochran (1977, p.325)	number of persons	number of rooms	10	2.1
8	Sukhatme and Sukhatme (1970, p.183)	1937 area under wheat	1936 area under wheat	34	3.4
9	Subset of 10 (see text)			23	2.6
10	Yates (1960, p.159)	volume of timber	eye-estimate	25	19.9

largest values of y/x. The kurtosis coefficients β_{2z} range from 1.4 to 19.9, whereas the optimality of v_0 and $v_{p\xi}$ under the Random Permutation Model (Model III) holds for $\beta_{2z} \doteq 3$.

The sampling scheme of Sampford (1967) was used to draw samples with inclusion probabilities proportioned to x. For this design, all estimators except v_{HT} are nonnegative, as the relation $\pi_i\pi_j - \pi_{ij} \geq 0$ holds. The design was implemented using Sampford's rejective algorithm, with a computer program written in FORTRAN. Sample sizes 3, 5 and 10 were considered. For four populations, sample size 10 was not used because the condition $\pi_i \propto x_i$ would have forced the largest π_i to exceed unity for n = 10. From each population, 200 samples were drawn of each sample size used.

The true variance $V(e_{HT})$ and the values of all estimators for each sample were calculated using exact values of π_{ij}. The empirical efficiencies (ratios of mean squared errors over 200 samples) relative to the Yates-Grundy estimator are given in Table 2.

4.1. Discussion of the Results

Among the three design-unbiased estimators, v_{HT} and v_A have clearly performed worse than the Yates-Grundy estimator, although in a few cases they are somewhat more efficient. Ironically, the worst performer v_{HT} is the only one for which an admissibility result is available! The approximately design-unbiased estimator v_F is somewhat more efficient than v_{YG} in most cases.

The "intuitive" model-based estimator v_R as well as the "optimal" estimators v_0 and $v_{p\xi}$ have performed generally (but not uniformly) better than the rest. For n = 3, v_0 is the best of the seven estimators for all populations except no. 10. For n = 5, v_0 is again the best with two exceptions. For n = 10, the picture is mixed, with v_R performing the best for 4 out of 6 populations.

Table 2. Efficiencies Relative to the Yates-Grundy Estimator

Pop. No.	Sample Size	v_{HT}	v_A	v_F	v_R	v_0	$v_{p\xi}$
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	.61	.94	.99	.92	2.07	.97
	5	.23	.86	.96	.90	1.42	.93
	10	.03	.11	.94	.84	.99	.86
4	3	.16	.46	1.09	1.17	1.79	1.17
	5	.05	.06	1.24	1.35	1.68	1.43
5	3	.08	.68	1.19	1.37	1.67	1.33
	5	.02	.12	1.30	1.37	1.30	1.48
6	3	.002	.93	1.06	1.18	1.42	1.15
	5	.0003	.39	1.13	1.30	1.42	1.32
7	3	.16	.92	1.10	1.23	2.03	1.18
	5	.03	.50	1.18	1.37	1.47	1.38
8	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
9	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
10	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

A simple intuitive explanation for the high relative efficiency of v_R for population 10 and the unusual trend for population 8 is given below.

4.2. Effect of Extreme Values

Consider the problem of estimating the ratio $\bar{y}/\bar{x}$ of the population means of two characteristic x and y. Two basic estimators of the ratio are:

$$R = \bar{y}_s/\bar{x}_s \qquad \text{(the ratio of sample means)}$$
$$= \sum_{i\epsilon s} r_i x_i / n\bar{x}_s$$

and $$M = \sum_{i\epsilon s} r_i/n, \qquad \text{(the mean of ratios)}$$

where $r_i = y_i/x_i$.

If r_i changes by an amount d, then R and M would change by $dx_i/n\bar{x}_s$ and d/n respectively. It follows that an extreme value of r_i will have a greater influence on the ratio of means R, if $x_i > \bar{x}_s$, and on M, if $x_i < \bar{x}_s$. Thus the ratio of means can be expected to be more efficient, if an extreme value of r_i is associated with a unit with small x_i, and vice-versa.

For the problem of estimating $V(e_{HT})$, the estimators v_R and v_0 are essentially of the types R and M respectively, apart from a constant multiplier $\sum_{i<j}^{N} c_{ij}$ (and some shrinkage in case of v_0), the roles of x_i and r_i being played by c_{ij} and f_{ij} respectively. Thus, v_R can be expected to be more efficient if extreme values of f_{ij} are associated with small values of c_{ij}.

Now observe that for a fixed i, the average of c_{ij}'s is $(N-1)^{-1}\sum_{j(\neq i)} c_{ij} = (N-1)^{-1} \pi_i(1-\pi_i)$. Hence, if π_i is either very small or close to 1, then c_{ij}'s would be relatively small. Also, if z_i is extreme, then f_{ij} is extreme for every j ($\neq$i). It is found from the data of population 10, that an extreme value of z_i is associated with a small value of x_i, and hence a small values of π_i. It follows that extreme values of f_{ij} are associated with small values of c_{ij}, making v_R relatively more efficient.

In population 8, an extreme value of z is associated with a unit having a large x value. For n=3, 5 and 10, the inclusion probability π_i of this unit takes the values .26, .44 and .88, respectively. Thus $\pi_i(1-\pi_i)$ takes the values .19, .25 and .11 respectively. As we might expect, the relative efficiency of V_R is small for n=3, gets even smaller for n=5, but becomes large for n=10. The changes are quite dramatic because the "wild" unit is the one with the largest inclusion probability.

To summarize the effect of extreme values, we might expect v_R to be more efficient when extreme ratios y/x are associated with units having inclusion probabilities either substantially below average, or close to unity. Otherwise, v_0 can be expected to be more efficient. The choice of estimator is particularly critical when the extreme observation has a large inclusion probability.

4.3. Concluding Remarks

Using the mean square error as the criterion, the model-based estimators v_R, v_0 and $v_{p\xi}$ appear to be generally more efficient than the design-based estimators. The relative efficiency of v_0 does not appear to be as much affected by deviation of β_{2z} from the value required for optimality, as by other deviations from the RPM, in particular, the association of extreme values of z with extreme values of x. In the presence of the extreme values, v_0 is, in some cases, less efficient than v_R. However, in these cases, its efficiency relative to the design-based estimators is higher than usual. In other words, although the optimality of v_0 is affected by model breakdown, the design-based estimators fail to solve this problem. In order to obtain more robust estimators, it seems necessary to consider a wider range of models, rather than no model.

REFERENCES

1. Ajgaonkar, S.G.P. (1967), "Unbiased estimator of variance of Narain, Horvitz and Thompson estimator," Sankhya Ser. A., 29, 55-60.

2. Biyani, S.H. (1978), "On some estimation problems in finite population sampling," Ph.D. thesis, Iowa State University, Ames, Iowa.

3. Cassel, C.M., Särndal, C.E. and Wretman, J.H. (1977), Foundations of inference on survey sampling, John Wiley and Sons, New York.

4. Cochran, W.G. (1977), Sampling Techniques. John Wiley and Sons, New York.

5. Fuller, W.A. (1970), "Sampling with random stratum boundaries," Journal of the Royal Statistical Society, Ser. B., 32, 209-226.

6. Godambe, V.P. and Joshi, V.M. (1965). "Admissibility and Bayes estimation in sampling finite populations -- I." Annals of Mathematical Statistics, 36, 1707-1722.

7. Godambe, V.P. and Thompson, M.E. (1973), "Estimation in sampling theory with exchangeable priors." Annals of Statistics, 1, 1212-1221.

8. Hanurav, T.V. (1967), "Optimal utilization of auxiliary information in πps sampling of two units from a stratum." Journal of Royal Statistical Society, Ser. B., 29, 374-391.

9. Horvitz, D.G. and Thompson, D.J. (1952), "A generalization of sampling without replacement from a finite universe." Journal of American Statistical Association, 47, 663-685.

10. Joshi, V.M. (1970), "Note on the admissibility of Sen-Yates-Grundy estimator." Sankhya Ser. A., 32, 431-438.

11. Kempthorne, O. (1979), "Some remarks on statistical inference in finite sampling," in New Developments in Survey Sampling, N.L. Johnson and H. Smith, editors, John Wiley and Sons, New York.

12. Madow, W.G. and Madow, L.H. (1944), "On the theory of systematic sampling," Annals of Mathematical Statistics, 15, 1-24.

13. Rao, J.N.K. (1963), "On three procedures of unequal probability sampling without replacement." Journal of American Statistical Association, 58, 202-215.

14. Sampford, M.R. (1967), "On sampling without replacement with unequal probabilities of selection." Biometrika, 54, 499-513.

15. Sukhatme, P.V. and Sukhatme, B.V. (1970), Sampling theory of surveys with applications, Iowa State University Press, Ames, Iowa.

16. Vijayan, K. (1975) "On estimating the variance in unequal probability sampling," Journal of the American Statistical Association, 70, 713-716.

17. Yates, F. (1960) Sampling methods for censuses and surveys. Hafner Publishing Company, New York.

18. Yates, F. and Grundy, P.M. (1953), "Selection without replacement from within strata with probability proportional to size." Journal of the Royal Statistical Society, Ser. B., 15, 235-261.

APPENDIX

SOLUTIONS FOR OPTIMAL COEFFICIENTS

The notation of Section 2.4 will be used in this appendix.

The following linear operator for two-way arrays is useful in the derivation.

Given a two-way array of elements x_{ij}, define

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

$$= \sum_{k(\neq i,j)} (x_{ik} + x_{kj}) + 2x_{ij} .$$

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

<u>Lemma A</u>: Let $u_1, \ldots, u_n$ be numbers satisfying $\sum_{i=1}^{n} u_i = 0$, and let $x_{ij} = u_i + u_j$. Then

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

<u>Proof</u>: $\Sigma_1 x_{ij} = \sum_{k=1}^{n} (u_i+u_k+u_k+u_j) - 2u_i - 2u_j$

$$= n(u_i+u_j) + 2\sum_{k=1}^{n} u_k - 2(u_i+u_j)$$

$$= (n-2)\ (u_i+u_j) \text{ by hypothesis.}$$

Now consider the problem of solving (2.7) to obtain the coefficients of the optimal estimator in K_o. The equations are:

$$E_\xi[\sum_{k<h\varepsilon s} b_{skh}\ f_{kh}(\underline{Y})\ f_{ij}(\underline{Y})] = E_\xi[\sum_{k<h}^{N} c_{kh} f_{kh}(\underline{Y}) f_{ij}(\underline{Y})] , \qquad (A.1)$$

$$i\neq j\varepsilon s,\ s\varepsilon S .$$

Defining $g_{skh} = b_{skh} - c_{kh}$, (A.2)

$$\sum_{k<h\epsilon s} g_{skh} E_\xi\{f_{kh}(\underline{Y})f_{ij}(\underline{Y})\} = \left(\sum_{\substack{k\epsilon s\\ h\epsilon\tilde{s}}} + \sum_{k<h\epsilon\tilde{s}}\right)\{c_{kh} E_\xi f_{kh}(\underline{Y})f_{ij}(\underline{Y})\} .$$

Separating terms with 0,1,2 subscripts common with (i,j) and using (2.5),

$$a_0 \sum_{\substack{k<h\epsilon s\\ (\neq i,j)}} g_{skh} + a_1 \sum_{\substack{k\epsilon s\\ (\neq i,j)}} (g_{sik} + g_{skj}) + a_2 g_{sij}$$

$$= \sum_{h\epsilon\tilde{s}} [(c_{ih} + c_{jh}) a_1 + \sum_{\substack{k\epsilon s\\ (\neq i,j)}} c_{kh} a_0] + \sum_{k<h\epsilon\tilde{s}} c_{kh} a_0 .$$

Putting $C_{\tilde{s}k} = \sum_{h\epsilon\tilde{s}} c_{kh}$, $C_{\tilde{s}s} = \sum_{k\epsilon s} c_{\tilde{s}k}$,

$C_{\tilde{s}} = \sum_{k<h\epsilon\tilde{s}} c_{kh}$, $g_s = \sum_{k<h\epsilon s} g_{shk}$, and

$$\Sigma_1 g_{sij} = \sum_{\substack{k\epsilon s\\ (\neq i,j)}} (g_{sik} + g_{skj}) + 2g_{sij} ,$$

$$a_0 (g_s - \Sigma_1 g_{sij} + g_{sij}) + a_1 (\Sigma_1 g_{sij} - 2g_{sij})$$

$$+ a_2 g_{sij} = (C_{\tilde{s}i} + C_{\tilde{s}j})a_1 + (C_{\tilde{s}s} - C_{\tilde{s}i} - C_{\tilde{s}j}) a_0 + C_{\tilde{s}} a_0 .$$

Putting $d_1 = a_1 - a_0$, $d_2 = a_2 - 2a_1 + a_0$,

$$a_0 g_s + d_1 \Sigma_1 g_{sij} + d_2 g_{sij} = d_1(C_{\tilde{s}i}+C_{\tilde{s}j}) + a_0 (C_{\tilde{s}s} + C_{\tilde{s}}),$$

$$i \neq j\epsilon s. \qquad \text{(A.3)}$$

Next, we average (A.3) over all pairs (i,j), i<jεs . Note that, $\Sigma_1 g_{sij}$ contains 2(n-1) terms, and by symmetry, each has average

$\sum_{k<h\varepsilon s} g_{skh} / \binom{n}{2} = g_s / \binom{n}{2}$. Average of (A.3) is

$$a_0 g_s + d_1 \cdot 2(n-1)\, g_s / \binom{n}{2} + d_2 g_s / \binom{n}{2}$$

$$= d_1 \cdot 2C_{\tilde{s}s}/n + a_0(C_{\tilde{s}s} + C_{\tilde{s}}), \text{ or} \qquad \text{(A.4)}$$

$$g_s / \binom{n}{2} = [d_1 \cdot 2C_{\tilde{s}s}/n + a_o(C_{\tilde{s}s} + C_{\tilde{s}})]/t_0 , \qquad \text{(A.5)}$$

where $t_0 = \binom{n}{2} a_0 + 2(n-1)\, d_1 + d_2$.

Subtracting (A.4) from (A.3) and putting

$$h_{sij} = g_{sij} - g_s / \binom{n}{2} , \text{ and} \qquad \text{(A.6)}$$

$$D_{\tilde{s}i} = C_{\tilde{s}i} - C_{\tilde{s}s}/n , \qquad \text{(A.7)}$$

$$d_1 \Sigma_1 h_{sij} + d_2 h_{sij} = d_1(D_{\tilde{s}i} + D_{\tilde{s}j}) , \qquad \text{(A.8)}$$

$$i \neq j \varepsilon s .$$

Let $h_{si\cdot} = \sum_{k\varepsilon s\,(\neq i)} h_{sik}$. Then, by definition of Σ_1 , (and using $h_{sij} = h_{sji}$) ,

$$\Sigma_1 h_{sij} = h_{si\cdot} + h_{sj\cdot} . \qquad \text{(A.9)}$$

Thus, (A.8) becomes $d_1(h_{si\cdot} + h_{sj\cdot}) + d_2 h_{sij} = d_1(D_{\tilde{s}i} + D_{\tilde{s}j})$. (A.10)

By (A.6) $\sum_{i\varepsilon s} h_{si\cdot} = \sum_{i\neq j\varepsilon s} h_{sij} = 0$.

Thus Lemma A applies to $x_{ij} = h_{si\cdot} + h_{sj\cdot}$, giving

$$\Sigma_1 (h_{si\cdot} + h_{sj\cdot}) = (n-2)\,(h_{si\cdot} + h_{sj\cdot}) . \qquad \text{(A.11)}$$

Similarly, $\sum_{i\varepsilon s} D_{\tilde{s}i} = 0$, and by Lemma A,

$$\Sigma_1 (D_{\tilde{s}i} + D_{\tilde{s}j}) = (n-2)\,(D_{\tilde{s}i} + D_{\tilde{s}j}) . \qquad \text{(A.12)}$$

Applying the operator Σ_1 to (A.10),

$$d_1\Sigma_1(h_{si\cdot} + h_{sj\cdot}) + d_2\Sigma_1 h_{sij} = d_1\Sigma_1(D_{\tilde{s}i\cdot} + D_{\tilde{s}j\cdot}) .$$

Using (A.9), (A.11) and (A.12),

$$d_1(n-2)(h_{si\cdot} + h_{sj\cdot}) + d_2(h_{si\cdot} + h_{sj\cdot}) = d_1(n-2)(D_{\tilde{s}i} + D_{\tilde{s}j}) ,$$

or

$$h_{si\cdot} + h_{sj\cdot} = (n-2)d_1(D_{\tilde{s}i\cdot} + D_{\tilde{s}j\cdot})/t_1 , \qquad (A.13)$$

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

Substituting from (A.13) into (A.10),

$$h_{sij} = d_1(D_{\tilde{s}i\cdot} + D_{\tilde{s}j\cdot})/t_1$$

$$= d_1(C_{\tilde{s}i\cdot} + C_{\tilde{s}j\cdot} - 2C_{\tilde{s}s}/n)/t_1 . \qquad (A.14)$$

By (A.2), the solution $b_{sij} = b^o_{sij}$ of (2.7) is given by

$$b^o_{sij} = c_{ij} + g_{sij}$$

(by A.6)

$$= c_{ij} + h_{sij} + g_s/\binom{n}{2} .$$

Using (A.5) and (A.14),

$$b^o_{sij} = c_{ij} + [a_0(C_{\tilde{s}s} + C_{\tilde{s}}) + 2d_1C_{\tilde{s}s}/n]/t_0$$

$$+ d_1(C_{\tilde{s}i\cdot} + C_{\tilde{s}j\cdot} - 2C_{\tilde{s}s}/n)/t_1 . \qquad (A.15)$$

Next, consider the problem of finding the optimal coefficients under ξ-unbiasedness, given by (2.8) and (2.9). This can be derived from the solution of (2.7), (given by A.15) as follows.

Let the solution of (2.7) $(H_{sij} = 0)$ be $b_{sij} = b^o_{sij}$, and let the solution of (2.9) $(H_{sij} = \lambda_s)$ be $b_{sij} = b^{\xi}_{sij}$, where λ_s is to be chosen to satisfy (2.8).

For fixed i, j, we can write $H_{sij} = \sum_{k<h\varepsilon s} \alpha_{kh} b_{skh} - \alpha_0$,

where $\alpha_{kh} = E_\xi [f_{kh}(\underline{Y}) f_{ij}(\underline{Y})]$ and

$$\alpha_0 = E_\xi [F(\underline{Y}) f_{ij}(\underline{Y})] .$$

$\alpha_{kh} = a_0$, a_1 or a_2 according as {k,h} has 0, 1 or 2 elements common with {i,j}. The number of times α_{kh} takes the values a_0, a_1, and a_2 [as (k,h) varies over all pairs of units in s] is $\binom{n-2}{2}$, 2(n-2) and 1 respectively.

$$\text{Thus,} \quad \sum_{k<h\varepsilon s} \alpha_{kh} = \binom{n-2}{2} a_0 + 2(n-2)a_1 + a_2 = t_0 .$$

Therefore, $H_{sij} - \lambda_s = \sum_{k<h\varepsilon s} \alpha_{kh} (b_{skh} - \lambda_s/t_0) - \alpha_0$.

Thus, if $b_{sij} = b^o_{sij}$ is a solution of $H_{sij} = 0$, then $b_{sij} - \lambda_s/t_0$ is a solution of $H_{sij} = \lambda_s$. To get the required solution b^ξ_{sij} eliminating λ_s, put $b_{sij} = b^o_{sij} - \lambda_s/t_0$

$$= c_{ij} + d_1(C_{\tilde{s}i} + C_{\tilde{s}j} - 2C_{\tilde{s}s}/n)/t_1 + \mu_s , \quad \text{where} \tag{A.16}$$

μ_s is a constant.

From (2.8),

$$\sum_{i<j\varepsilon s} \{c_{ij} + d_1 (C_{\tilde{s}i} + C_{\tilde{s}j} - 2C_{\tilde{s}s}/n)/t_1 + \mu_s\} = C_U ,$$

or $C_s + 0 + \binom{n}{2}\mu_s = C_U$.

$$\text{Thus} \quad \mu_s = (C_U - C_s)/\binom{n}{2} = (C_{\tilde{s}} + C_{\tilde{s}s})/\binom{n}{2} ,$$

and the solution of (2.8) - (2.9) is

$$b^\xi_{sij} = c_{ij} + d_1(C_{\tilde{s}i} + C_{\tilde{s}j} - 2C_{\tilde{s}s}/n)/t_1 + (C_{\tilde{s}s} + C_{\tilde{s}})/\binom{n}{2} . \tag{A.17}$$

Similarly, the solution of (2.12) - (2.13) must be of the form $b^{b\xi}_{sij} = b^{o}_{sij} - \lambda/t_0 = b^{o}_{sij} + K$ (say), where K is chosen to satisfy (2.12). That is,

$$\sum_{s\varepsilon S} p(s) \sum_{i<j\varepsilon s} [c_{ij} + d_1(C_{\tilde{s}i} + C_{\tilde{s}j} - 2C_{\tilde{s}s}/n)/t_1$$

$$+ \{a_0(C_{\tilde{s}s} + C_{\tilde{s}}) + 2d_1 C_{\tilde{s}s}/n\}/t_0 + K] = C_U . \tag{A.18}$$

We have $\sum_{s\varepsilon S} p(s) \sum_{i<j\varepsilon s} c_{ij} = \sum_{i<j}^{N} c_{ij} \sum_{s\ni i,j} p(s)$

$$= \sum_{i<j}^{N} c_{ij}\pi_{ij} , \tag{A.19}$$

$$\sum_{s\varepsilon S} p(s) (C_{\tilde{s}s} + C_{\tilde{s}}) = \sum_{s\varepsilon S} p(s) (C_U - C_s)$$

$$= C_U - \sum_{s\varepsilon S} p(s) C_s = C_U - \sum_{i<j}^{N} c_{ij}\pi_{ij} , \tag{A.20}$$

$$\sum_{i<j\varepsilon s} (C_{\tilde{s}i} + C_{\tilde{s}j} - 2C_{\tilde{s}s}/n) = 0 , \tag{A.21}$$

$$\sum_{s\varepsilon S} p(s) C_{\tilde{s}s} = \sum_{s\varepsilon S} p(s) \sum_{\substack{i\varepsilon s \\ j\varepsilon \tilde{s}}} c_{ij}$$

$$= \sum_{i<j}^{N} c_{ij} [\sum_{\substack{s\ni i \\ s\not\ni j}} p(s) + \sum_{\substack{s\not\ni i \\ s\ni j}} p(s)]$$

$$= \sum_{i<j}^{N} c_{ij} (\pi_i - \pi_{ij} + \pi_j - \pi_{ij}) = \sum_{i<j}^{N} c_{ij}(\pi_i + \pi_j - 2\pi_{ij}) . \tag{A.22}$$

From (A.18) - (A.22) ,

$$\binom{n}{2} K + \sum_{i<j}^{N} c_{ij}\pi_{ij} + \binom{n}{2} t_o^{-1} \Big[a_0(C_U - \sum_{i<j}^{N} c_{ij}\pi_{ij})$$

$$+ 2d_1 n^{-1} \sum_{i<j}^{N} c_{ij}(\pi_i + \pi_j - 2\pi_{ij}) \Big] = C_U ,$$

$$\binom{n}{2} K = \sum_{i<j}^{N} c_{ij}(1 - \pi_{ij}) - \binom{n}{2} t_0^{-1} a_0 \sum_{i<j}^{N} c_{ij}(1 - \pi_{ij})$$

$$- d_1(n-1) t_0^{-1} \sum_{i<j}^{N} c_{ij} (\pi_i + \pi_j - 2\pi_{ij}) ,$$

$$K = [\binom{n}{2} t_0]^{-1} \sum_{i<j}^{N} c_{ij}[t_2(1-\pi_{ij}) - \binom{n}{2} a_0(1-\pi_{ij}) - d_1(n-1)(\pi_1 + \pi_j - 2\pi_{ij})]$$

$$= [\binom{n}{2} t_0]^{-1} \sum_{i<j}^{N} c_{ij} \ [d_2(1 - \pi_{ij}) + (n-1)d_1 \ (2 - \pi_i - \pi_j)] \quad . \qquad \text{(A.23)}$$