A Minimax Approach to Sample Surveys

Ching-Shui Cheng*
Department of Statistics
Univ. of Calif., Berkeley

Ker-Chau Li
Department of Statistics
Purdue University

Mimeograph Series #81-55

Department of Statistics
Purdue University

December, 1981

*Research supported by NSF Grant No. MCS-79-09502

A Minimax Approach to Sample Surveys

by

Ching-Shui Cheng*
Department of Statistics
University of California, Berkeley

Ker-Chau Li
Department of Statistics
Purdue University

Summary

Suppose that there is a population $U = \{1,2,\ldots,N\}$ of N identifiable units and that two values x_i and y_i are associated with the ith unit. The values $x_1, x_2, \ldots, x_N$ are given but each y_i is determined only after the ith unit is selected and observed. The objective is to estimate the population total $\sum_{i=1}^{N} y_i$. It is assumed that $y_i = \theta x_i + \delta_i g(x_i)$, $1 \leq i \leq N$, where $\theta \in \Theta \subset R$, g is a real-valued function of x, and $(\delta_1,\ldots,\delta_N)$ is in some neighborhood L of $(0,\ldots 0)$. The Rao-Hartley-Cochran strategy and Hansen-Hurwitz strategy are shown to be approximately minimax under the model with $g(x)=x^{\frac{1}{2}}$ and $L = L_2(M) = \{(\delta_1,\ldots\delta_N) = \sum_{i=1}^{N} \delta_i^2 \leq M\}$, and that with $g(x) = x$ and $L = L_\infty(M) = \{(\delta_1,\ldots,\delta_N) = |\delta_i| \leq M$ for all $i\}$. The latter model applies to a problem considered by Scott and Smith (1975). These two strategies are then compared with some commonly-used strategies and are found to perform favorably when $g^2(x)/x$ is an increasing function of x. The problem of estimating θ is also considered and is solved for any g when L is $L_2(M)$. Finally, some exact minimax results are obtained for sample size one.

---

*Research supported by NSF Grant No. MCS-79-09502.

AMS 1980 Subject Classification: Primary 62D05

Secondary 62C20

Keywords and Phrases: Adjusted risk-generating matrix, Hansen-Hurwitz strategy, minimaxity, probability proportional to aggregate size sampling, Rao-Hartley-Cochran strategy, ratio estimator, risk-generating matrix, sample surveys, simple random sampling, strategy.

A MINIMAX APPROACH TO SAMPLE SURVEYS

1. Introduction

Suppose that there is a population $U = \{1,2,\ldots,N\}$ of N identifiable units and that two correlated values x_i and y_i, where $x_i \geq 0$, are associated with the ith unit. The values $x_1,x_2,\ldots,x_N$ are given but each y_i is determined only after the ith unit is selected and observed. The objective is to estimate the unknown population total $Y = \sum_{i=1}^{N} y_i$ based on a sample of size n. Without loss of generality, we shall assume that $x_1 \leq x_2 \leq \ldots \leq x_N$. In practice, x_i is often the value of y_i at some previous time when a complete census was taken. Another application is in cluster sampling where y_i is the ith cluster total and x_i is the size of the ith cluster. Many procedures have been suggested for using the auxiliary information provided by x to increase the precision of the estimate either at the design stage (e.g. using a probability proportional to size design) or at the estimation stage (e.g., using a ratio estimator) or both. The superpopulation approach of Brewer (1963) and Royall (1970a) incorporates the auxiliary information into a random superpopulation model which often leads to the selection of a purposive sample. Whether one should randomize or not had been one of the major controversies between the fixed population approach and the superpopulation approach. In this paper, we shall formulate a regression type model and use a minimax criterion to justify randomization for guarding against extreme population values.
We shall assume that

$$(1.1) \qquad y_i = \theta x_i + \varepsilon_i, \; 1 \leq i \leq N,$$

where $\theta \in \Theta \subset R$ (usually $\Theta = R$ or $\Theta = (0,\infty)$), and ε_i is the "error" or "deviation" from the strict linear relationship between y and x. We shall further assume that

(1.2) $$\varepsilon_i = \delta_i g(x_i), \quad 1 \le i \le N,$$

where g is a known function of x and $(\delta_1,\ldots,\delta_N)$ belongs to a fixed (usually small) neighborhood L of $(0,\ldots,0)$. Usually $g(x)$ is an increasing function of x (e.g., $g(x) = x^{\alpha}$, $\alpha \ge 0$), reflecting the idea that the error tends to increase with increasing x_i (which is usually the case in sample surveys). There are many possible choices for the neighborhood L depending on the measure of distance used. We may assume L to be the L_2-ball $L_2(M) = \{(\delta_1,\ldots,\delta_N) : \sum_{i=1}^{N} \delta_i^2 \le M\}$ or the L_∞-ball $L\{(M) = \{(\delta_1,\ldots,\delta_N) : |\delta_i| \le M$, for all $i\}$, where $M > 0$ is fixed. Our model is <u>not</u> a superpopulation model though it is somewhat similar to the commonly used superpopulation model in which $y_1,\ldots,y_N$ are assumed to be independent random variables with

(1.3) $$E(y_i) = \theta x_i \quad \text{and} \quad Var(y_i) = \sigma^2 \nu(x_i)$$

Note that $\nu(x_i)$ is the counterpart of $g^2(x_i)$.

Throughout this paper, we shall assume squared loss function and only consider sampling designs with fixed sample size n. Since a sampling design with replacement can always be improved by one without replacement, we <u>may consider sampling designs without replacement only</u>. Thus, without loss of generality, we define a <u>sampling design</u> to be a probability measure P on $\mathbb{S}$ $\mathbb{S} = \{S: S \subset U$ and $\#(S) = n\}$, where $\#(S)$ is the cardinality of S. We shall also restrict to linear homogeneous estimators of the population total Y, i.e., estimators of the form $\sum_{i \in S} a_{i(S)}\, y_i$, where S is the selected sample. <u>Such an estimator can also be written as</u> $\sum_{i=1}^{N} a_{i(S)}\, y_i$ <u>with</u> $a_{i(S)} = 0$ <u>for all</u> $i \notin S$.

Let $\underset{\sim}{a}_S = (a_{1(S)},\ldots,a_{N(S)})'$. Then a linear homogeneous estimator is specified by the vectors of coefficients $\{\underset{\sim}{a}_S\}_{S \in \mathcal{S}}$. A pair $d = (P,\{\underset{\sim}{a}_S\}_{S \in \mathcal{S}})$ of sampling design and estimator of Y is then called a strategy. The set of all such strategies will be denoted by $\mathcal{D}_n$.

For any $\theta \in \Theta$ and $\underset{\sim}{\delta} = (\delta_1,\ldots,\delta_N)' \in L$, the mean squared error (MSE) of a strategy $d = (p,\{\underset{\sim}{a}_S\}_{S \in \mathcal{S}})$, denoted by $R_n(d;\theta,\underset{\sim}{\delta})$ or $R_n(p,\{\underset{\sim}{a}_S\}_{S \in \mathcal{S}};\ \theta,\underset{\sim}{\delta})$, is defined to be $\sum_{S \in \mathcal{S}} P(S)\ (\sum_{i \in S} \underset{\sim}{a}_{(S)} y_i - \sum_{i=1}^{N} y_i)^2$, where $y_i = \theta x_i + \delta_i g(x_i)$.

Our goal then is to find a strategy to

$$(1.4) \qquad \text{minimize} \sup_{\theta \in \Theta,\ \underset{\sim}{\delta} \in L} R_n(d\ ;\ \theta,\ \underset{\sim}{\delta})$$

When exact minimax strategies are difficult to find, approximately minimax strategies will be desirable.

Note that all the strategies in $\mathcal{D}_n$ have nonrandomized estimators. One can also consider strategies with randomized linear homogeneous estimators. All such strategies will be denoted by $\mathcal{D}_n^R$. Clearly $\mathcal{D}_n \subset \mathcal{D}_n^R$. On the other hand, since the squared loss function is used, for each strategy d in $\mathcal{D}_n^R$, there is a Rao-Blackwellized strategy d' in $\mathcal{D}_n$ which is at least as good as d. However, sometimes Rao-Blackwellization can only provide a small amount of improvement with the cost of introducing a quite complicated estimator which makes the analysis difficult. Under such cases we may want to sacrifice a little bit of efficiency to gain some simplicity, and thus use a randomized strategy in $\mathcal{D}_n^R$. In this spirit, the well-known Rao-Hartley-Cochran strategy (Rao, Hartley and Cochran (1962)), a strategy with randomized estimator, is shown in Section 3 to be approximately minimax over $\mathcal{D}_n^R$ under model (1.1) - (1.2) with $g(x) = x$, $L = L_\infty(M)$, and that with $g(x) = x^{\frac{1}{2}}$ and $L = L_2(M)$, when some reasonable conditions are imposed on the configuration of $x_1,\ldots,x_N$. Also if the sampling fraction is small, then the Hansen-Hurwitz strategy (Hansen and Hurwitz(1943)), which is only slightly less efficient than Rao-Hartley-Cochran's strategy, but is much simpler to implement, is also shown to be approximately minimax.

Under model (1.1) - (1.2), θ is not identifiable, i.e., for given $\{(y_i,x_i)\}_{i=1}^{N}$, there exist more than one θ such that (1.1) and (1.2) are satisfied. To make θ identifiable, we should restrict $\underset{\sim}{\delta}$ to a suitable subset $\tilde{L}$ of L. The choice of $\tilde{L}$ depends on what we think that θ means. For instance, if θ is interpreted as the population ratio $\sum_{i=1}^{N} y_i / \sum_{i=1}^{N} x_i$ and we have a situation where the individual ratios $\theta_i = y_i/x_i$ are approximately equal, then ε_i represents the error of approximating $y_i(= \theta_i x_i)$ by θx_i. Under this interpretation, one has $\theta \sum_{i=1}^{N} x_i = \sum_{i=1}^{N} y_i = \sum_{i=1}^{N} (\theta x_i + \delta_i g_{(x_i)}) = \theta \sum_{i=1}^{N} x_i + \sum_{i=1}^{N} \delta_i g(x_i)$, which implies that $\sum_{i=1}^{N} \delta_i g(x_i) = 0$. Then we are lead to $\tilde{L} = \{\underset{\sim}{\delta} \in L\colon \sum_{i=1}^{N} \delta_i g_{(x_i)} = 0\}$, i.e., it would be appropriate to

$$(1.5) \qquad \text{minimize} \sup_{\theta \in \Theta,\ \underset{\sim}{\delta} \in L,\ \sum_{i=1}^{N} \delta_i g(x_i) = 0} R_n(d;\theta,\underset{\sim}{\delta}).$$

Sometimes instead of having a definite meaning of θ, we merely know that the line $y = \theta x$ would be a good approximation to the data $\{(y_i,x_i)\}_{i=1}^{N}$ should we have the chance to observe all the y_i's. If θ is such that

$$\sum_{i=1}^{N} (y_i - \theta x_i)^2 / g(x_i)^2 = \min_{\theta'} \sum_{i=1}^{N} (y_i - \theta' x_i)^2/g(x_i)^2,$$

i.e., the value which give the weighted least squares fit, then

$$(1.6) \qquad \theta = \sum_{i=1} [g(x_1)]^{-2} x_i y_i / x_i^2[g(x_i)]^{-2} ,$$

and it is not hard to see that under model (1.1) $\sim$ (1.2), one has

(1.7) $$\sum_{i=1}^{N} \delta_i [g_{(x_i)}]^{-1} x_i = 0 .$$

Therefore one should

(1.8) $$\text{minimize} \quad \sup_{\theta \in \Theta,\ \underset{\sim}{\delta} \in L,\ \sum_{i=1}^{N} \delta_i [g(x_i)]^{-1} x_i = 0} R_n(d;\theta,\underset{\sim}{\delta}).$$

Clearly Problems (1.5) and (1.8) are the same when $g(x) = x$. It will be shown later that for unbounded Θ and $L = L_2(M)$, (1.4) and (1.8) are identical problems for any g. Thus the results on Rao-Hartley-Cochran strategy mentioned earlier also apply to (1.8).

In Section 2, the problem will be formulated in matrix notation. We shall introduce the important notion of <u>risk-generating matrix</u> and <u>adjusted risk-generating matrix</u>, which are very useful in assessing the performance of a sampling strategy when the estimator is linear. Section 3 is devoted to the approximate minimaxity of Rao-Harltey-Cochran strategy and Hansen-Hurwitz strategy. In Section 4, we shall compare the Rao-Hartley-Cochran strategy with two commonly used procedures: simple random sampling (SRS) together with ratio estimator and sampling with probabilities proportional to aggregate size (PPAS) together with ratio estimator. We find that the Rao-Hartley-Cochran strategy performs favorably when $g^2(x)/x$ is an increasing function of x, e.g. $g_{(x)} = x^{\alpha}$ with $\alpha \geq \frac{1}{2}$. This seems to be consistent with comparisons based on superpopulation models (see, e.g., Cassel, Särndal, and Wretman (1977) p.171). In section 5, we shall study the problem of estimating the paramter θ under the identifiability condition (1.7). For quite arbitrary g, approximately minimax strategies are derived. It turns out that if instead of selecting units with probability proportional to size x_i, we select units with probability proportional to $\frac{x_i^2}{g(x_i)^2}$, then the modified version of Rao-Hartley-Cochran strategy or Hansen-Hurwitz strategy are approximately minimax for estimating θ.

Before closing this section, we shall relate our problem to earlier works in the literature. Blackwell and Girshick (1954) gave the first minimax result for justifying the use of simple random sampling. Under the assumption that the space of all possible $\underset{\sim}{y} = (y_1,\ldots,y_N)'$ is permutation-invariant, they showed the minimaxity of simple random sampling for any permutation-invariant estimator and loss function. Works on minimax estimation of the population mean under <u>simple random sampling</u> include, e.g., Aggarawal (1959), Royall (1970b), Bickel and Lehmann (1981), Hodges and Lehmann (1981). Bickel and Lehmann (1981) essentially studied model (1.1) -(1.2) with $x_1 = x_2 = \ldots = x_N, \theta \in R$ and $L = L_2(M)$ while Hodges and Lehmann (1981) considered model (1.1) - (1.2) with $x_1 = \ldots = x_N$, θ known, and $L = L_\infty(M)$. They did not restrict to linear homogeneous estimators as we do. Bickel and Lehmann (1981) also extended their result to the minimaxity of (simple random sampling, sample mean). Stenger (1979) obtained the minimaxity of (simple random sampling, sample mean) under the restriction that the estimators are linear homogeneous and satisfy a condition similar to (2.2) in Section 2 of the present paper.

Scott and Smith (1975) probably is the first and only paper on the minimaxity of unequal probability sampling schemes. Their interesting results apparently stimulated our work. They considered the minimax estimation of a population total $Y = \sum_{i=1}^{N} y_i$, where $y_i = x_i z_i$, with the observable z_i taking values in a fixed interval, say $0 \leq z_i \leq B$. They fixed the estimator to be

$X \sum_{i=1} z_i/n$, where $z_1,\ldots z_n$ are the n observed values of z, and showed that under some condition, if a sampling scheme with replacement is to be used, then the probability proportional to size design is approximately minimax. However, there are several restrictions in this work. First, the estimator is fixed. As indicated by Scott and Smith (1975), their emphasis was on choosing a design to give a maximum protection to a given estimator, i.e., to find minimax designs for a given estimator, rather than minimax strategies. No doubt it is an important problem to vary the estimators and search for minimax strategies. Second, for sample size larger than 1, they restricted the competing sampling schemes to those with replacement (i.e., if you select the first observation by a certain scheme, then you should use the same scheme over and over again till the n-th observation is selected). Thus when combined with the restriction on estimators, many commonly used strategies including (SRS, ratio estimator) and (PPAS, ratio estimator) are ruled out. Third, they assumed that each z_i is in the interval [0,B]. This is only one of many possibilities that may arise. In fact, in Scott and Smith's model, one has $|y_i| \leq Bx_i$. Therefore y_i can be expressed as $y_i = \frac{B}{2} x_i + \varepsilon_i$ with $|\varepsilon_i| \leq \frac{B}{2} x_i$. Now is is clear that this is a special case of our model (1.1) - (1.2) with $\theta = B/2$, $g_{(x)} = x$, and $L = L_\infty$ (B/2). Thus our results also apply to Scott and Smith's problem; the above restrictions are then removed.

2. The risk-generating matrix and adjusted risk-generating matrix.

In this section we shall formulate our problem in matrix notation and derive an important matrix for assessing the performance of a sampling strategy. For convenience, the following notations will be adopted:

$\underset{\sim}{1}$ = the N x 1 vector of 1's

$\underset{\sim}{G}$ = the N x N diagonal matrix with $g(x_i)$ as the ith diagonal element,

$\underset{\sim}{x}$ = the N x 1 vector $(x_1,\ldots,x_N)'$,

$\underset{\sim}{y}$ = the N x 1 vector $(y_1,\ldots,y_N)'$,

and

$$X = \sum_{i=1}^{N} x_i .$$

Then for any $d = (P, \{\underset{\sim}{a}_S\}_{S\in\mathcal{S}}) \in \mathcal{D}_n$,

$$R_n(P, \{\underset{\sim}{a}_S\}_{S\in\mathcal{S}}; \theta, \underset{\sim}{\delta}) = \sum_{S\in\mathcal{S}} P(S) \left(\sum_{i=1}^{N} a_{i(S)}y_i - \sum_{i=1}^{N} y_i\right)^2$$

$$(2.1) \qquad = \sum_{S\in\mathcal{S}} P(S)\, \underset{\sim}{y}'(\underset{\sim}{a}_S - \underset{\sim}{1})(\underset{\sim}{a}_S - \underset{\sim}{1})'\underset{\sim}{y}$$

$$= \sum_{S\in\mathcal{S}} P(S)(\theta\underset{\sim}{x} + \underset{\sim}{G}\underset{\sim}{\delta})'(\underset{\sim}{a}_S-\underset{\sim}{1})(\underset{\sim}{a}_S-\underset{\sim}{1})'(\theta\underset{\sim}{x} + \underset{\sim}{G}\underset{\sim}{\delta}).$$

Clearly (2.1) is a quadratic function of θ if $(\underset{\sim}{a}_S - \underset{\sim}{1})'\,\underset{\sim}{x} \neq 0$ for some S with $P(S) > 0$. If Θ is unbounded, then the maximum MSE will be infinite unless we have $(\underset{\sim}{a}_S - \underset{\sim}{1})'\,\underset{\sim}{x} = 0$ for all S with $P(S) > 0$, i.e.,

$$(2.2) \qquad \sum_{i=1}^{N} \underset{\sim}{a}_{i(S)}x_i = X \quad \text{for all } P(S) > 0.$$

Therefore for unbounded Θ, e.g., $\Theta = R$ or $(0,\infty)$, we may restrict to estimators satisfying (2.2). We shall also make the same restriction when Θ is bounded. Condition (2.2) in fact is equivalent to the unbiasedness of an estimator in the usual superpopulation model (see, e.g., Cochran (1977), p. 159). Such a restriction is reasonable and is indispensable when Θ is unbounded. Clearly (2.2) holds for the ratio estimator and the estimator used by Scott and Smith (1975). A strategy satisfying (2.2) was called a <u>representative strategy</u> by

Hájek (1959). We shall denote the collection of all the representative strategies in $\mathcal{D}_n$ by $\bar{\mathcal{D}}_n$.

For a strategy in $\bar{\mathcal{D}}_n$, obviously one has

$$\begin{aligned} R(P,\{\underset{\sim}{a}_S\}_{S\in\mathcal{S}}, \theta, \underset{\sim}{\delta}) &= \sum_{S\in\mathcal{S}} P(S)\, \underset{\sim}{\delta}'\, \underset{\sim}{G}'\, (\underset{\sim}{a}_S - \underset{\sim}{1})\, (\underset{\sim}{a}_S - \underset{\sim}{1})'\, \underset{\sim}{G}\, \underset{\sim}{\delta} \\ &= \underset{\sim}{\delta}'\underset{\sim}{G}\, \{ \sum_{S\in\mathcal{S}} P(S)(\underset{\sim}{a}_S - \underset{\sim}{1})\, (\underset{\sim}{a}_S - \underset{\sim}{1})'\}\, \underset{\sim}{G}\, \underset{\sim}{\delta} \end{aligned} \tag{2.3}$$

which is independent of θ and is a quadratic form in the matrix $G\, \{ \sum_{S\in\mathcal{S}} P(S)\, (\underset{\sim}{a}_S - \underset{\sim}{1})\, (\underset{\sim}{a}_S - \underset{\sim}{1})'\}\, \underset{\sim}{G}$. This matrix plays an important role in assessing the performance of the strategy $d = (p,\{\underset{\sim}{a}_S\}_{S\in\mathcal{S}})$. We shall call $\sum_{S\in\mathcal{S}} P(S)\, (\underset{\sim}{a}_S - \underset{\sim}{1})\, (\underset{\sim}{a}_S - \underset{\sim}{1})'$ the <u>risk-generating matrix</u>, and $\underset{\sim}{G}\{ \sum_{S\in\mathcal{S}} P(S)(\underset{\sim}{a}_S-\underset{\sim}{1})'(\underset{\sim}{a}_S-\underset{\sim}{1})\}\underset{\sim}{G}$ the <u>adjusted risk-generating matrix</u> of the strategy $(P,\{\underset{\sim}{a}S\}_{S\in\mathcal{S}})$; and <u>we shall denote the adjusted risk-generating matrix of a strategy d by $\underset{\sim}{R}n(d)$ (or $\underset{\sim}{R}(d)$, when n is clear to us</u>).

We have the following basic property of $\underset{\sim}{R}(d)$.

<u>Proposition 2.1</u>. If d is representative, then $\underset{\sim}{R}(d)$ is singular and satisfies the condition $\underset{\sim}{R}(d)\underset{\sim}{G}^{-1}\underset{\sim}{x} = 0$.

This is a straightforward consequence of (2.2); the proof is thus omitted. The identifiability condition (1.7) can be written in vector form $\underset{\sim}{\delta}'\, \underset{\sim}{G}^{-1}\, \underset{\sim}{x} = \underset{\sim}{0}$. Now in view of (2.3) and Proposition 2.1, one has the following

<u>Proposition 2.2.</u> If (i) $L = L_2(M)$ and (ii) Θ is unbounded or Θ is bounded and the estimators satisfy (2.2), then (1.4) and (1.8) are identical problems. Moreover, a minimax strategy minimizes the maximum eigenvalue of $\underset{\sim}{R}(d)$.

This brings out an interesting connection to the theory of optimum designs. A minimax strategy in Proposition 2.2 is like an E-optimum design (see, Kiefer (1974) for some terminology of optimum design theory). In fact, when

$x_1 = x_2 = \ldots = x_N$ (i.e., there is no auxiliary information), one has $\underset{\sim}{R}(d)\ \underset{\sim}{1} = \underset{\sim}{0}$ under (2.2), i.e., $\underset{\sim}{R}(d)$ has zero row sums. In this case, if d* is the strategy (simple random sampling, sample mean), then R(d*) minimizes tr $\underset{\sim}{R}(d)$ and is completely symmetric in the sense that all the diagonal elements are the same and all the off-diagonals are the same, i.e., all the eigenvectors in the nondegenerate directions have the same eigenvalues. An argument similar to Proposition 1 of Kiefer (1975) then shows the minimaxity of d* when the space of all $\underset{\sim}{y} = (y_1,\ldots,y_N)$ is permutation-invariant. The so-called C-matrix of a balanced incomplete block design is also completely symmetric. Thus d* plays the same role as a balanced incomplete block design in block design setting.

For convenience, hereafter the maximum eigenvalue of a matrix $\underset{\sim}{A}$ will be denoted by $\lambda_{max}(\underset{\sim}{A})$. For a strategy with randomized estimator, a representativeness condition similar to (2.2) is also desirable in order that the MSE be finite when Θ is unbounded. The mean squared error of a strategy $d \in \mathcal{D}_n^R$ satisfying the representativeness condition is also a quadratic form $\underset{\sim}{\delta}'\underset{\sim}{R}(d)\underset{\sim}{\delta}$ with the adjusted risk-generating matrix $\underset{\sim}{R}(d)$ satisfying $\underset{\sim}{R}(d)\ \underset{\sim}{G}^{-1}\underset{\sim}{x} = \underset{\sim}{0}$, too. Later on, we shall show that when $L = L_2(M)$ and $g(x) = x^{\frac{1}{2}}$, all the nonzero eigenvalues of the adjusted risk-generating matrix of the Rao-Hartley-Cochran strategy are the same. This may explain why the Rao-Hartley-Cochran strategy performs very well under the minimax criterion; it is like a balanced design in experimental design settings.

3. Approximate minimaxity of Rao-Hartley-Cochran strategy.

For a sample size n, the Rao-Hartley-Cochran strategy first forms n random groups of units, one unit to be drawn from each group. The number of units $N_1, N_2, \ldots, N_n$ in the respective groups are made as equal as possible, i.e., $|N_i - N_j| \leq 1$ for all $i.j = 1,2,\ldots,n$. Let $X_j = \sum_{i \in \text{the jth group}} x_i$. Then the probability of selecting the ith unit in the jth group is x_i/X_j, and the estimate of the population total is

$$\hat{Y}_{RHC} = \sum_{j=1}^{n} X_j \frac{y_j}{x_j},$$

where y_j, x_j refer to the unit drawn from group j. This strategy will be denoted by d_{RHC}. Clearly d_{RHC} is not in $\mathcal{D}_n$ since the estimator $\hat{Y}_{RHC}$ is a randomized estimator. Therefore d_{RHC} can be improved by some strategy in $\mathcal{D}_n$. However, since d_{RHC} is well-known and is easy to implement, we shall ignore its improved version in $\mathcal{D}_n$ and state our main results in terms of d_{RHC}.

We shall calculate the adjusted risk-generating matrix of d_{RHC}, $\underset{\sim}{R}(d_{RHC})$ through the Hansen-Hurwitz strategy (denoted by d_{HH}) since d_{RHC} is closely related to d_{HH} and that the mean squared error of d_{HH} is easy to calculate. Recall that strategy d_{HH} is a with replacement scheme in which the probability of selecting the ith unit at each stage is x_i/X and the estimator of the population total is $\hat{Y}_{HH} = n^{-1} X \sum_{i=1}^{n} y_i/x_i$, where $(x_1, y_1), \ldots, (x_n, y_n)$ are the observed values of (x,y) with possible repetitions. When $n > 1$, this strategy is in neither $\mathcal{D}_n$ nor $\mathcal{D}_n^R$ because of the with replacement feature.

For n=1, the mean squared error of d_{HH} is $R_1(d_{HH};\ \theta\ \underset{\sim}{\delta}) = \underset{\sim}{\delta}' G\{\sum_{i=1}^{N} P(i)(\underset{\sim}{a}_i - \underset{\sim}{1})(\underset{\sim}{a}_i - \underset{\sim}{1})'\} \underset{\sim}{G}\underset{\sim}{\delta}$ where $\underset{\sim}{a}_i$ is the Nx1 vector with X/x_i as the ith coordinate and all the other coordinates are zero. Through some simple calculation, we obtain

$$(3.1) \qquad R_1(d_{HH};\ \theta, \underset{\sim}{\delta}) = \underset{\sim}{\delta}' \underset{\sim}{G}\{\mathrm{diag}(x_1^{-1} X, \ldots, x_N^{-1} X) - \underset{\sim}{J}_N\} \underset{\sim}{G}\ \underset{\sim}{\delta},$$

where $\underset{\sim}{J}_N$ is the N x N matrix of ones. For $n > 1$, $\hat{Y}_{HH}$ is an unbiased estimator of Y and is essentially the average of a random sample. Therefore

(3.2) $R_n(d_{HH};\ \theta,\underset{\sim}{\delta}) = n^{-1}\ R_1(d_{HH};\ \theta,\underset{\sim}{\delta}) = n^{-1}\ \underset{\sim}{\delta}'\underset{\sim}{G}\ \{\text{diag}\ (x_1^{-1}\ X,\ldots,x_N^{-1}X)-\underset{\sim}{J}_N\}\underset{\sim}{G}\ \underset{\sim}{\delta}$

Now, write N=nR+k, where R and k are integers with $0 \le k \le n$, and let

(3.3) $\mu = (N-1)^{-1}(N-n) + N^{-1}(N-1)^{-1}k(n-k).$

by (9A.66) and (9A.67) of Cochran (1977), we have

(3.4) $R_n(d_{RHC};\ \theta,\underset{\sim}{\delta}) = \mu \cdot R_n(d_{HH};\ \theta,\underset{\sim}{\delta}).$

Therefore, we have

(3.5) $\underset{\sim}{R}_n(d_{RHC}) = \mu \cdot n^{-1}\ \underset{\sim}{G}\ \{\text{diag}\ (x_1^{-1}X,\ldots,x_N^{-1}X) - \underset{\sim}{J}_N\}\ \underset{\sim}{G}.$

Furthermore, when $g(x) = x^{\frac{1}{2}}$, we obtain that

(3.6) $\underset{\sim}{R}_n(d_{RHC}) = \mu n^{-1}(X\underset{\sim}{I} - \underset{\sim}{G}\ J_N\ \underset{\sim}{G}),$

which has all the nonzero eigenvalues equal to $\mu n^{-1}X$. Thus we obtain the following proposition.

<u>Proposition 3.1</u>. Let $g(x) = x^{\frac{1}{2}}$. Then for any $M \ge 0$, we have

$$\sup_{\theta\in\Theta,\ \underset{\sim}{\delta}\in L_2(M)} R_n(d_{HH};\theta,\delta) = n^{-1}X\cdot M\ ,\ \text{and}$$

$$\sup_{\theta\in\Theta,\ \underset{\sim}{\delta}\in L_2(M)} R_n(d_{RHC};\ \theta,\delta) = \mu n^{-1}X\cdot M.$$

To establish the approximate minimaxity of d_{RHC}, we need to give a lower bound for the maximal MSE of an arbitrary strategy. The following theorem provide such a useful lower bound.

Theorem 3.1. Let $g(x) = x^{\frac{1}{2}}$ and Θ be unbounded. Then for $M \geq 0$, we have

$$(3.7) \qquad \min_{d \in \mathcal{D}_n} \sup_{\theta \in \Theta,\ \underset{\sim}{\delta} \in L_2(M)} R_n(d;\ \theta,\ \underset{\sim}{\delta}) \geq n^{-1} \times M \cdot \frac{\left(\sum_{i=1}^{N-n} x_i\right)^2 + n \cdot \sum_{i=1}^{N-n} x_i^2}{X^2 - \sum_{i=1}^{N} x_i^2}$$

If Θ is bounded, then (3.7) holds for $d \in \bar{\mathcal{D}}_n$.

Proof. For any $t > 0$, let

$$(3.8) \qquad \underset{\sim}{Z} = \underset{\sim}{R}_n(d) + t\,(\sqrt{x_1},\ldots,\sqrt{x_n})'\,(\sqrt{x_1},\ldots,\sqrt{x_N}).$$

By proposition 2.1, we have $\underset{\sim}{R}_n(d)(\sqrt{x_1},\ldots,\sqrt{x_N})' = 0$; i.e., each row of $\underset{\sim}{R}_n(d)$ is orthogonal to $(\sqrt{x_1},\ldots,\sqrt{x_N})'$. Therefore if $\lambda_1,\lambda_2,\ldots,\lambda_{N-1}$ and 0 are the eigenvalues of $\underset{\sim}{R}_n(d)$, then the eigenvalues of $\underset{\sim}{Z}$ will be

$\lambda_1,\ \lambda_2,\ldots,\lambda_{N-1}$, and $t \sum_{i=1}^{N} x_i = tX$. We shall first give a lower bound for the maximal eigenvalue of $\underset{\sim}{Z}$ and then by suitably choosing t we may obtain a good lower bound for the maximal eigenvalue of $\underset{\sim}{R}_n(d)$.

Now, the maximal eigenvalue of $\underset{\sim}{Z} \equiv \lambda_{max}(\underset{\sim}{Z})$

$$\geq \text{ the maximal diagonal element of } \underset{\sim}{Z}$$

$$= \max_{1 \leq i \leq N} \left\{ \sum_{S \in \mathcal{S}} x_i\, P(S)\, (a_{i(S)}-1)^2 + t\, x_i \right\}$$

$$= \max_{1 \leq i \leq N} \left\{ \left[\sum_{S \in \mathcal{S}} x_i^2\, P(S)\, (a_{i(S)}-1)^2 + t\, x_i^2 \right] / x_i \right\}$$

$$\geq \left\{ \sum_{i=1}^{N} \left[\sum_{S \in \mathcal{S}} x_i^2\, P(S)\, (a_{i(S)}-1)^2 + t\, x_i^2 \right] \right\} / \sum_{i=1}^{N} x_i$$

$$(3.9) \qquad = \left\{ \left[\sum_{S \in \mathcal{S}} P(S) \left[\sum_{i \in S} (a_{i(S)} x_i - x_i)^2 + \sum_{i \notin S} x_i^2 \right] \right] + t \sum_{i=1}^{N} x_i^2 \right\} / X.$$

Next, as was demonstrated in section 2, (2.2) should hold if we want the maximal MSE finite. An important consequence of (2.2) is the following inequality:

$$(3.10)\qquad \sum_{i\in S}(a_{i(s)}\,x_i - x_i)^2 \geq n(\frac{1}{n}\sum_{i\in S} a_{i(s)}\,x_i - \frac{1}{n}\sum_{i\in S} x_i)^2$$

$$= n(\frac{X}{n} - \frac{1}{n}\sum_{i\in S} x_i)^2 \text{ (by (2.2))}$$

$$= n^{-1}\,(\sum_{i\notin S} x_i)^2\ .$$

Now, due to the assumption that $x_1 \leq x_2 \leq \cdots \leq x_N$, we have $\sum_{i\notin S} x_i \geq \sum_{i=1}^{N-n} x_i$ and $\sum_{i\notin S} x_i^2 \geq \sum_{i=1}^{N-n} x_i^2$. Therefore backing to (3.9), we obtain that

$$(3.11)\qquad \lambda_{max}(\underset{\sim}{Z}) \geq \{[n^{-1}(\sum_{i=1}^{N-n} x_i)^2 + \sum_{i=1}^{N-n} x_i^2] + t\sum_{i=1}^{N} x_i^2\}/X\ .$$

Consider any t such that $t < [n^{-1}(\sum_{i=1}^{N-n} x_i)^2 + \sum_{i=1}^{N-n} x_i^2]/(X^2 - \sum_{i=1}^{N} x_i^2)$.

A simple computation leads to

$$t\,X < \{[n^{-1}(\sum_{i=1}^{N-n} x_i)^2 + \sum_{i=1}^{N-n} x_i^2] + t\sum_{i=1}^{N} x_i^2\}/X\ .$$

Therefore, t X can not be the maximal eigenvalue of $\underset{\sim}{Z}$. Thus by the discussions after (3.8) we obtain that $\lambda_{max}(\underset{\sim}{R}_n(d)) = \lambda_{max}(\underset{\sim}{Z})$. Finally, in (3.11), let t tend to $n^{-1}(\sum_{i=1}^{N-n} x_i)^2 + \sum_{i=1}^{N-n} x_i^2]\,/(X^2 - \sum_{i=1}^{N} x_i^2)$, Then we get the desired bound. □

How sharp is the bound on the right side of (3.7)? First, when $x_1=x_2=\ldots=x_N$, it can be verified that this bound equals $n^{-1}XM(N-1)^{-1}(N-n)$, which is exactly the maximal MSE of the (simple randomization, sample mean). Thus our bound is sharp in this degenerated case. Now, compared to maximal MSE of our candidate, d_{RHC} (or d_{HH} when μ is close to 1), the bound on the right side of (3.7) will be good provided that $\mu \cdot \dfrac{X^2 - \sum_{i=1}^{N} x_i^2}{(\sum_{i=1}^{N-n} x_i)^2 + n \sum_{i=1}^{N-n} x_i^2} \leq 1+\varepsilon$ for some small ε.

This will be the case if the sampling fraction $\frac{n}{N}$ is small and (or) the x_i's do not vary too much. In fact, when N/n is integral and $x_1 = x_2 = \ldots = x_N$,

$$\mu \frac{X^2 - \sum_{i=1}^{N} x_i^2}{(\sum_{i=1}^{N-n} x_i)^2 + n \sum_{i=1}^{N-n} x_i^2} = 1.$$

We summarize this result as follows.

<u>Corrollary 3.1.</u> Let $g(x)=x^{\frac{1}{2}}$ and Θ is unbounded. Suppose $\mu\lambda \cdot \dfrac{X^2 - \sum_{i=1}^{N} x_i^2}{(\sum_{i=1}^{N-n} x_i)^2 + n(\sum_{i=1}^{N-n} x_i^2} \leq 1+\varepsilon$. Then the Rao-Hartley-Cochran strategy is $(1+\varepsilon)$-approximately minimax, in the sense that, for any sampling strategy $d \in \mathcal{D}_n$, we have

$$\sup_{\theta \in \Theta,\, \underset{\sim}{\delta} \in L_2(M)} R_n(d_{RHC};\, \theta, \underset{\sim}{\delta}) \leq (1+\varepsilon) \sup_{\theta \in \Theta,\, \underset{\sim}{\delta} \in L_2(M)} R_n(d;\, \theta, \underset{\sim}{\delta}) .$$

If Θ is bounded then the above inequality still holds if $d \in \bar{\mathcal{D}}_n$.

Note that because of Proposition 2.2, the Rao-Hartley-Cochran strategy is also $(1+\varepsilon)$-approximately minimax when the identifiability condition (1.7) is imposed. Furthermore, the Hansen-Hurwitz strategy will also be approximately minimax if μ is close to 1.

Now let us turn to the case $g(x) = x$ and $L = L_\infty(M)$. We need the following lemma:

Lemma 3.1. If $\underset{\sim}{M}$ is a NxN matrix such that $\underset{\sim}{1}'\,\underset{\sim}{M}\underset{\sim}{1} = 0$, then there is an Nx1 vector $\underset{\sim}{\delta}$ with +1,-1 entries such that $\underset{\sim}{\delta}'\underset{\sim}{M}\underset{\sim}{\delta} \geq N(N-1)^{-1}$ trace $\underset{\sim}{M}$ when N is even, and $\underset{\sim}{\delta}'\underset{\sim}{M}\underset{\sim}{\delta} \geq (N+1)N^{-1}$ trace $\underset{\sim}{M}$ when N is odd.

Proof. Let $\mathbb{P}$ be the set of all the NxN permutation matrices $\underset{\sim}{P}$. Clearly, the matrix $\frac{1}{N!} \sum_{\underset{\sim}{P} \in \mathbb{P}} \underset{\sim}{P}'\underset{\sim}{M}\underset{\sim}{P}$ is of the form $a\underset{\sim}{I}_N + b\underset{\sim}{J}_N$ for some real number a and b. Moreover, a and b can be determined by the fact that trace $a\underset{\sim}{I} + b\underset{\sim}{J}_N$ = trace $\underset{\sim}{M}$ and that $\underset{\sim}{1}'(a\underset{\sim}{I} + b\underset{\sim}{J}_N)\underset{\sim}{1} = \frac{1}{N!} \sum_{P \in \mathbb{P}} \underset{\sim}{1}'\underset{\sim}{P}'\underset{\sim}{M}\underset{\sim}{P}\underset{\sim}{1} = \frac{1}{N!} \sum_{P \in \mathbb{P}} \underset{\sim}{1}'\underset{\sim}{M}\underset{\sim}{1} = 0$. After some computation, we get

$$a = \frac{1}{N-1} \text{ trace } \underset{\sim}{M}, \text{ and } b = -\frac{1}{N(N-1)} \text{ trace } \underset{\sim}{M} .$$

Consider the case where N is even first. Let $\underset{\sim}{\delta}^{o}$ be the vector with the first $\frac{N}{2}$ coordinates equal to 1 and the last $\frac{N}{2}$ coordinates equal to -1. Then we have

$$\begin{aligned} &\text{Max } \{\underset{\sim}{\delta}'\underset{\sim}{M}\underset{\sim}{\delta} \mid \underset{\sim}{\delta} = \underset{\sim}{\delta}^{o}\underset{\sim}{P} \text{ for some } \underset{\sim}{P} \text{ in } \mathbb{P}\} \\ &\geq \frac{1}{N!} \sum_{\underset{\sim}{P} \in \mathbb{P}} \underset{\sim}{\delta}^{o\prime}\underset{\sim}{P}'\underset{\sim}{M}\underset{\sim}{P}\underset{\sim}{\delta}^{o} \\ &= \underset{\sim}{\delta}^{o\prime} (a\underset{\sim}{I} + b\underset{\sim}{J})\underset{\sim}{\delta}^{o} \\ &= N \cdot a = \frac{N}{N-1} \text{ trace } M . \end{aligned}$$

Thus the case where N is even is proved. Next, we consider the case where N is odd. Let $\underset{\sim}{\delta}^{o}$ be the vector with the first $\frac{N+1}{2}$ coordinates equal to 1 and the last $\frac{N-1}{2}$ coordinates equal to -1. Then similar computation leads to Max $\{\underset{\sim}{\delta}'\underset{\sim}{M}\underset{\sim}{\delta} \mid \underset{\sim}{\delta} = \underset{\sim}{\delta}^{o}\, \underset{\sim}{P}$ for some $\underset{\sim}{P}$ in $\underset{\sim}{\mathbb{P}}\} \geq Na + b = (N+1)N^{-1}$ trace M. The proof is now complete. □

We are ready to derive a useful lower bound for the maximal MSE of an arbitrary strategy.

Theorem 3.2. Let $g(x) = x$ and Θ be unbounded. Then we have

$$\min_{d\in\mathcal{D}_n}\ \sup_{\theta\in\Theta,\ \underset{\sim}{\delta}\in L_\infty(M)} R_n(d;\theta,\underset{\sim}{\delta}) \geq \left(n^{-1}\left(\sum_{i=1}^{N-n} x_i\right)^2 + \sum_{i=1}^{N-n} x_i^2\right) N(N-1)^{-1}M^2,$$

if N is even, and

$$\min_{d\in\mathcal{D}_n}\ \sup_{\theta\in\Theta,\ \underset{\sim}{\delta}\in L_\infty(M)} R_n(d;\theta\ \underset{\sim}{\delta}) \geq \left(n^{-1}\left(\sum_{i=1}^{N-n} x_i\right)^2 + \sum_{i=1}^{N-n} x_i^2\right)(N+1)N^{-1}M^2,$$

if N is odd.

If Θ is bound, then the above inequalities hold if $d\in\bar{\mathcal{D}}_n$.

Proof. Without loss of generality, we may assume $M = 1$. By Proposition 2.1, we have $\underset{\sim}{R}_n(d)\underset{\sim}{1} = 0$; i.e., all row sums of $\underset{\sim}{R}_n(d)$ are zero. Therefore by Lemma 3.1, there exists a $\underset{\sim}{\delta}^* \in L_\infty(1)$ such that $\underset{\sim}{\delta}^{*\prime}\underset{\sim}{R}_n(d)\ \underset{\sim}{\delta}^* \geq N(N-1)^{-1}$ tr $\underset{\sim}{R}_n(d)$ if N is even and $\underset{\sim}{\delta}^{*\prime}\underset{\sim}{R}_n(d)\underset{\sim}{\delta}^* \geq (N+1)N^{-1}$ tr $\underset{\sim}{R}_n(d)$ if N is odd. It remains to establish a lower bound for tr $\underset{\sim}{R}_n(d)$.

By a straightforward computation, we have

$$\begin{aligned}
\operatorname{tr} R_n(d) &= \sum_{i=1}^{N}\sum_{S\in\mathcal{S}} P(S)(a_{i(S)}-1)^2 x_i^2 \\
&= \sum_{S\in\mathcal{S}} P(S) \sum_{i=1}^{N}(a_{i(S)}-1)^2 x_i^2 \\
&= \sum_{S\in\mathcal{S}} P(S) \left(\sum_{i\in S}(a_{i(S)}\, x_i - x_i)^2 + \sum_{i\notin S} x_i^2\right) \\
&\geq \sum_{S\in\mathcal{S}} P(S) \cdot \left(n^{-1}\left(\sum_{i\notin S} x_i\right)^2 + \sum_{i\notin S} x_i^2\right) \qquad \text{(by(3.10)} \\
&\geq n^{-1}\left(\sum_{i=1}^{N-n} x_i\right)^2 + \sum_{i=1}^{N-n} x_i^2 \ .
\end{aligned}$$

Therefore, the proof is complete. □

Again, one can verify that this bound is sharp when $x_1 = \ldots = x_N$. On the other hand, we have the following bounds for the maximum MSE of d_{RHC} and d_{HH}.

Proposition 3.2. Assume $g(x) = x$. Then for any $M \geq 0$, we have

$$\sup_{\theta \in \Theta, \underset{\sim}{\delta} \in L_2(M)} R_n(d_{HH};\ \theta, \delta) \leq n^{-1} X^2 M^2$$

and

$$\sup_{\theta \in \Theta,\ \underset{\sim}{\delta} \in L_2(M)} R_n(d_{RHC};\ \theta, \delta) \leq \mu\, n^{-1} X^2 M^2.$$

Proof. Without loss of generality, we assume M=1. From (3.1), we obtain that $\underset{\sim}{R}_n(d_{HH}) = n^{-1}[\mathrm{diag}(x_1X, \ldots, x_NX) - \underset{\sim}{G}\underset{\sim}{J}_N\underset{\sim}{G}]$. Since $\underset{\sim}{G}\underset{\sim}{J}_N\underset{\sim}{G}$ is nonnegative definite, we conclude that for any $\underset{\sim}{\delta} \in L_2(1)$, $\underset{\sim}{\delta}' \underset{\sim}{R}_n(d_{HH})\underset{\sim}{\delta} \leq n^{-1} \sum_{i=1}^{N} x_i\ X = n^{-1}X^2$ as desired. Similarly, we obtain the bound for d_{RHC}. □

Note that equalities hold in Proposition 3.2 when there exists a number m such that $\sum_{i=1}^{m} x_i = \sum_{i=m+1}^{N} x_i$. Now we are ready to establish the approximate minimaxity of d_{RHC} when the x_i's do not vary too much, and that of d_{HH} when, in addition, the sampling fraction $\frac{n}{N}$ is small.

Corrollary 3.2. Let $g(x)=x$ and Θ be unbound. Suppose

$$\mu\lambda \cdot \frac{N-1}{N} \cdot \frac{X^2}{(\sum_{i=1}^{N-n} t_i)^2 + n \sum_{i=1}^{N-n} x_i^2} \leq 1 + \varepsilon \qquad \text{when N is even, and}$$

$$\mu\lambda \cdot \frac{N}{N+1} \quad \frac{X^2}{(\sum_{i=1}^{N-n} x_i)^2 + n \sum_{i=1}^{N-n} x_i^2} \leq 1 + \varepsilon$$ when N is odd. Then the Rao-Hartley-

Cochran strategy is $(1+\varepsilon)$-approximately minimax, in the sense that for any sampling strategy $d \in \mathcal{D}_n$, we have

$$\sup_{\theta \in \Theta,\ \underset{\sim}{\delta} \in L_\infty(M)} R_n(d_{RHC};\ \theta, \underset{\sim}{\delta}) \leq (1+\varepsilon) \sup_{\theta \in \Theta,\ \underset{\sim}{\delta} \in L_\infty(M)} R_n(d;\ \theta, \underset{\sim}{\delta}) \ .$$

If Θ is bounded, then the above inequality holds if $d \in \bar{\mathcal{D}}_n$.

Note that the Hansen-Hurwitz strategy is also approximately minimax if μ is close to 1.

Now take $\Theta = \{\frac{B}{2}\}$ and $M = \frac{B}{2}$. Then our results applies to the problem of Scott and Smith (1975). For their problem, the Rao-Hartley-Cochran strategy is approximately minimax over all the strategies in $\bar{\mathcal{D}}_n$ (note that $\Theta = \{ \frac{B}{2} \}$ is bounded), if $\mu X^2/[(\sum_{i=1}^{N-n} x_i)^2 + n \sum_{i=1}^{N-n} x_i^2]$ is close to 1. Note that we do not put any restriction on the sampling designs and the only restriction on the estimators is the representativeness condition (2.2) which is reasonable, while Scott and Smith (1975) fixed the estimator and restricted to sampling schemes with replacement. Furthermore, our result applies to much more general models.

So far the approximate minimaxity results were only established for $g(x) = x^{\frac{1}{2}}$, $L = L_2(M)$ and $g(x) = x$, $L = L_\infty(M)$. Results for other forms of $g(x)$ are rather difficult to derive. We do not have satisfactory results in this direction. Therefore in the next Section, we shall compare the Rao-Hartley-Cochran strategy with two commonly used strategies (SRS, ratio estimator) and (PPAS, ratio estimator) under a variety of functions g. It turns out that the Rao-Hartley - Cochran strategy performs when $g(x) = x^\alpha$ with $\alpha \geq \frac{1}{2}$.

4. Comparisons.

Comparisons of various sampling strategies had been done in the literature, see, e.g., Chapter 7 of Cassel, Särndal, and Whetman (1977) and the references cited there. They were mostly empirical studies or based on some superpopulations. The criteria used were often expected (with respect to superpopulation) mean squared error and hence were "average" type criteria. The comparison we shall make here is different and is based on a minimax criterion. Barrowing optimum design theory terminology, one can say that our criterion is like an E-criterion and the earlier comparisons were more or less based on something like the A-criterion.

We have seen that the Rao-Hartley-Cochran strategy is approximately minimax under model (1.1)-(1.2) with $L = L_2(M)$ and $g(x) = x^{\frac{1}{2}}$. A comparable superpopulation model is (1.3) with $\nu(x) = x$. Under this model Brewer (1963) and Royall (1970a) showed that the best strategy is to select a purposive sample S* which consists of the n units with the largest x values and then use the ratio estimator. Under our assumption that $x \le \ldots \le x_N$, we have $S^* = \{x_{N-n+1},\ldots,x_N\}$. Let this strategy be denoted by d_1. Now let us first compare d_1 with d_{RHC} and d_{HH}, when $g(x)=x^{\frac{1}{2}}$.

We compute the maximal eigenvalue of $\underset{\sim}{R}_n(d_1)$ below:

$\lambda\max\,(\underset{\sim}{R}_n(d_1)) \ge$ the largest diagonal element of $\underset{\sim}{R}_n(d_1)$

$$\ge x_N \left(\frac{X}{\sum\limits_{i=N-n+1}^{N} x_i} - 1\right)^2$$

$$\ge n^{-1}\frac{\left(\sum\limits_{i=1}^{N-n} x_i\right)^2}{\sum\limits_{i=N-n+1}^{N} x_i}$$

Comparing this bound with the result of Proposition 3.1, we get the following proposition.

Proposition 4.1. Assume $g(x) = x^{\frac{1}{2}}$. Suppose $\sum_{i=1}^{N-n} x_i \geq \frac{1+\sqrt{5}}{2} \cdot \sum_{i=N-n+1}^{N} x_i$. Then for any $M \geq 0$, we have

$$\sup_{\theta \in \Theta,\ \underset{\sim}{\delta} \in L_2(M)} R_n(d_1;\ \theta,\underset{\sim}{\delta}) \geq \sup_{\theta \in \Theta,\ \underset{\sim}{\delta} \in L_2(M)} R_n(d_{HH};\ \theta,\underset{\sim}{\delta})$$

$$\geq \sup_{\theta \in \Theta,\ \underset{\sim}{\delta} \in L_2(M)} R_n(d_{RHC};\ \theta,\underset{\sim}{\delta}) \quad .$$

Therefore, if the sampling fraction $\frac{n}{N}$ is small and the extreme values of x_i's are not too extreme, then d_1 is worse than d_{HH} and d_{RHC}.

Two sampling designs which are commonly used together with the ratio estimator are SRS and PPAS. Let (SRS, ratio estimator) and (PPS, ratio estimator) be denoted by d_2 and d_3, respectively. Recall that in a PPAS sampling scheme, each sample S is selected with probability proportional to $X_S \equiv \sum_{i \in S} x_i$. Such a sampling design was proposed to make the ratio estimator design-unbiased and is often associated with the names of Hájek, Lahiri, Midzuno, and Sen. We shall show that both d_2 and d_3 are inferior to d_{RHC} under model (1.1) - (1.2) with $g(x) = x^{\frac{1}{2}}$ and $L = L_2(M)$. Later we shall extend the result to an arbitrary function g such that $g^2(x)/x$ is increasing in x. Some conditions on the configuration of $x_1,\ldots,x_N$ are needed there.

Now assume $g(x) = x^{\frac{1}{2}}$ and write $X_S = \sum_{i \in S} x_i$. Then we have

$$(4.1)\quad \operatorname{tr} R_n(d_2) = \sum_{i=1}^{N} \{[\sum_{S: i \in S} P(S)\, x_i\, (X^2/X_S^2 - 2\, X/X_S)] + x_i\}$$

$$= \frac{1}{\binom{N}{n}} \sum_{S \in \mathcal{S}} \sum_{i \in S} x_i (X^2/X_S{}^2 - 2X/X_S) + X$$

$$= \frac{1}{\binom{N}{n}} \sum_{S \in \mathcal{S}} X^2/X_S - X$$

$$\geq \frac{X^2}{\binom{N}{n}^{-1} \sum_{S \in \mathcal{S}} X_S} - X$$

$$= n^{-1}\, (N-n)\, X\ .$$

The equality holds only if $x_1 = x_2 = \ldots = x_N$. Since $\underset{\sim}{R}_n(d_2)$ is singular, we conclude that $\lambda_{max}(\underset{\sim}{R}_n(d_2)) \geq \frac{1}{N-1} \operatorname{tr} \underset{\sim}{R}_n(d_2) \geq n^{-1}(N-1)^{-1}(N-n)\, X$, and the equality holds only if $x_1 = x_2 = \ldots = x_N$. Comparing with the result of Proposition 3.1, we establish the following proposition.

<u>Proposition 4.2.</u> Assume $g(x) = x^{\frac{1}{2}}$ and $\frac{N}{n}$ is integral. Then, we have

$$\sup_{\theta \in \Theta, \underset{\sim}{\delta} \in L_2(M)} R_n(d_2;\ \theta, \underset{\sim}{\delta}) \geq \sup_{\theta \in \Theta, \underset{\sim}{\delta} \in L_2(M)} R_n(d_{RHC};\ \theta, \underset{\sim}{\delta})\ ,$$

and the equality holds only if $x_1 = x_2 = \ldots = x_N$.

After some computation we can show that $\operatorname{tr} \underset{\sim}{R}_n(d_3) = \operatorname{tr} \underset{\sim}{R}_n(d_2)$. Therefore a similar argument leads to the following proposition.

<u>Proposition 4.3.</u> Assume $g(x) = x^{\frac{1}{2}}$ and $\frac{N}{n}$ is integral. Then, we have

$$\sup_{\theta \in \Theta, \underset{\sim}{\delta} \in L_2(M)} R_n(d_3;\ \theta, \underset{\sim}{\delta}) \geq \sup_{\theta \in \Theta, \underset{\sim}{\delta} \in L_2(M)} R_n(d_{RHC};\ \theta, \underset{\sim}{\delta})\ ,$$

and the equality holds only if $x_1 = \ldots = x_N$.

Now consider a general function g such that $g^2(x)/x$ is increasing in x. Substituting $\underset{\sim}{G} = \mathrm{diag}\,(g(x_1),\ldots,g(x_N))$ into (3.5), we have

$$(4.2)\qquad \underset{\sim}{R}_n(d_{RHC}) = \mu \cdot n^{-1}\{\mathrm{diag}(X\, g^2(x_1)/x_1,\ldots,Xg^2(x_N)/x_N) - \underset{\sim}{G}\,\underset{\sim}{J}_N\underset{\sim}{G}\ .$$

Since $\underset{\sim}{G}\,\underset{\sim}{J}_N\underset{\sim}{G}$ is non-negative definite and $g^2(x)/x$ is an increasing function of x, it follows that

$$(4.3)\qquad \lambda_{max}(\underset{\sim}{R}_n(d_{RHC})) \leq \mu\, n^{-1}X\, g^2(x_N)/x_N.$$

On the other hand, we have

$$\begin{aligned}\lambda_{max}(\underset{\sim}{R}_n(d_2)) &\geq \text{ the N-th diagonal element of } \underset{\sim}{R}_n(d_2)\\ &\geq g^2(x_N)\cdot \binom{N}{n}^{-1}\cdot \sum_{S:N\in S}\left(\frac{X}{X_S}-1\right)^2\\ &\geq g^2(x_N)\binom{N}{n}^{-1}\binom{N-1}{n-1}\left\{\frac{X}{\binom{N-1}{n-1}^{-1}\sum_{S:N\in S}X_S}-1\right\}^2\\ &= g^2(x_N)\,\frac{n}{N}\left[\frac{(N-1)X}{(N-n)x_N+(n-1)\,X}-1\right]^2\end{aligned}$$

Therefore, to show that $\lambda_{max}(\underset{\sim}{R}_n(d_{RHC})) \leq \lambda_{max}(\underset{\sim}{R}_n(d_2))$, it suffices to demonstrate that

$$(4.4)\qquad \frac{n^2}{N}\cdot\frac{x_N}{X}\left[\frac{(N-1)X}{(N-n)x_N+(n-1)X}-1\right]^2 \geq 1\ .$$

Write $a = \dfrac{x_N}{N^{-1}X}$ and $f = \dfrac{n}{N}$. Then, we may rewrite (4.4) as

$$(4.5)\qquad N^{-1}(1-f)^2 f\, a^3 - (1-f)^2(2nf+1)a^2 + (1-f)(n^2(1-f)-2(n-1))a-(n-1)^2 \geq 0.$$

Discarding the first term of (4.5) and changing (n-1) and $(n-1)^2$ to n n and n^2 respectively, we obtain the following sufficient condition for (4.5) to hold as below:

(4.6) $$-(1-f)^2(2nf+1)a^2 + (1-f)n(n(1-f)-2)a-n^2 \geq 0.$$

Now, letting

(4.7) $$\ell = \tfrac{1}{2} + \tfrac{1}{2}\sqrt{1-\frac{4(1+f)}{n(1-f)^2}} - n^{-1}(1-f)^{-1} \quad ,$$

and solving (4.6), we get

(4.8) $$(1-f)^{-2}\,\ell^{-1} \leq a \leq \ell \cdot \frac{n^2N}{2n^2+N} \quad .$$

Therefore we obtain the following result.

Proposition 4.4. Assume that $g^2(x)/x$ is non-decreasing in x. Suppose (4.8) holds, where a is the ratio of x_N and the average of x_i's, i.e., $a = x_N/(N^{-1}X)$ and ℓ is defined by (4.7). Then, we have

$$\sup_{\theta \in \Theta, \underset{\sim}{\delta} \in L_2(M)} R_n(d_2;\, \theta, \underset{\sim}{\delta}) \geq \sup_{\theta \in \Theta, \underset{\sim}{\delta} \in L_2(M)} R_n(d_{RHC};\, \theta, \underset{\sim}{\delta}) \, .$$

Note that the possible value for a is between 1 and N; and ℓ is very close to 1 if the sampling fraction f is small. Therefore, (4.8) amounts to saying that $x_N \leq (N^{-1}X) \cdot \frac{n^2N}{2n^2+N}$, which is a reasonable condition if n or N is large enough, because it simply means that the largest x value is not too far away from the average x value. We further remark that because of the asymmetry of $\underset{\sim}{R}_n(d_2)$, we would expect that even if (4.8) does not hold, d_{RHC} may still be much better than d_2. However, we shall not elaborate here.

A similar argument also applies to the comparison between d_{RHC} and d_3. So d_{RHC} performs favorably when $g(x) = x^{\alpha}$ with $\alpha \geq \frac{1}{2}$. Some comparisons of

d_{RHC} with strategies using ratio estimator can be found in Chapter 7 of Cassel, Särndal, and Wretman. The comparison there was based on superpopulation model (1.3) with $\nu(x) = x^{\beta}$ (comparable to our model ith $g(x) = x^{2\beta}$) and the assumption that N is very large and that the frequency distribution of the auxiliary variable values $x_1,\ldots,x_N$ is approximately a gamma distribution. On p. 171, they wrote that d_{RHC} was a good choice for $1 \leq \beta \leq 2$. This seems to be consistent with our finding that d_{RHC} is good for α (comparable to $\frac{\beta}{2}$) $\geq \frac{1}{2}$.

5. Estimation of θ.

Sometimes one may be more interested in estimating θ than the population total Y especially when x_i's are the value of y_i at some previous time. In this section, we shall study the minimax estimation of θ under model (1.1) - (1.2) with identifiability condition (1.7). We shall focus the discussion on the case $L = L_2(M)$.

When $g(x) = x^{\frac{1}{2}}$, by (1.6), we have $\theta = Y/X$; the estimation of Y then is the same as that of θ. Therefore Theorem 3.1 and Corrollary 3.1 are applicable. In Section 3, for the estimation of Y and $L=L_2(M)$, we had only been able to derive satisfactory results for $g(x) = x^{\frac{1}{2}}$. In this section, however, we shall show that for estimating θ, parallel minimax results can be established for an arbitrary g. Thus theorem 3.1 and Corrollary 3.1 could be viewed as a special case of the results in this section.

Let us again restrict to linear homogeneous estimators and use the same notation as before. By (1.6),

$$\theta = \sum_{i=1}^{N} [g(x_i)]^{-2} x_i y_i \Big/ \sum_{i=1}^{N} [g(x_i)]^{-2} x_i^2 ,$$

and hence θ can be viewed as a population total $\sum_{i=1}^{N} \theta_i$ with

$$\theta_i = [g(x_i)]^{-2} x_i y_i \;/\; \sum_{i=1}^{N} [g(x_i)]^{-2} x_i^{\,2} \;. \tag{5.1}$$

Let

$$z_i = [g(x_i)]^{-2}\, x_i^2 \quad \text{and } Z = \sum_{i=1}^{N} z_i . \tag{5.2}$$

Then, we have

$$\theta_i = Z^{-1}\, \{[g(x_i)]^{-2}\, x_i(\theta x_i + \delta_i g(x_i))\}$$

$$= Z^{-1}(\theta z_i + \delta_i\, z_i^{\frac{1}{2}}).$$

Since Z is a known constant, the problem of estimating θ is now reduced to that of estimating a population total in model (1.1) - (1.2) with x and g(x) replaced by z and $g(z) = z^{\frac{1}{2}}$; respectively. Therefore by Theorem 3.1 and Corrollary 3.1 an approximately minimax strategy is to divide the N units into n random groups of sizes as equal as possible, choose one unit from each group with probability proportional to $z = x^2/[g(x)]^2$, and then estimate θ by $\hat{\theta}_{RHC} = Z^{-1}\, (\sum_{j=1}^{n} \frac{\theta_j Z}{z_j} Z_j)$, or, equivatently,

$$\hat{\theta}_{RHC} = Z^{-1}\, (\sum_{j=1}^{n} x_j^{-1}\, y_j\, Z_j),$$

where x_j, y_j refer to the unit drawn from group j, and $Z_j = \sum_{i \in \text{ the jth group}} z_i$. Let us again denote this strategy by d_{RHC}. Then we have

Theorem 5.1. Suppose $\mu \cdot \dfrac{Z^2 - \sum_{i=1}^{N} z_i^2}{(\sum_{i=1}^{N-n} z_i)^2 + N \sum_{i=1}^{N-n} z_i^2} \leq 1+\varepsilon$, and Θ is unbounded where Z and z_i are defined by (5.2) and μ is defined by (3.3). Then for any sampling strategy $d \in \mathcal{D}_n$,

$$\sup_{\theta\in\Theta,\ \underset{\sim}{\delta}\in\tilde{L}_2(M)} (\text{MSE for estimating } \theta \text{ under } d_{RHC} \leq (1+\varepsilon) \cdot \sup_{\theta\in\Theta,\underset{\sim}{\delta}\in\tilde{L}_2(M)} (\text{MSE for estimating } \theta \text{ under } d),$$

where $\tilde{L}_2(M) = \{\underset{\sim}{\delta} \in L_2(M) : \underset{\sim}{\delta}\, \underset{\sim}{G}^{-1} \underset{\sim}{x} = \underset{\sim}{0}\}$. If Θ is bounded, then the same inequality hold if d satisfies the representative condition that $\sum_{i\in S} a_{i(S)}\, x_i = 1$ for all S with $P(S) > 0$.

We may also generalize the Hurwiz-Hansen strategy for the estimation of θ for arbitrary g as follows. The sampling scheme is to select n units with replacement such that at each stage, the probability of selecting the i-th unit is proportional to z_i $(= x_i^2/g(x_i)^2)$, and the estimator is $\frac{1}{n}\sum_{j=1}^{n} y_j/x_j$ where y_j, x_j refer to the unit being selected at the j-th stage. If μ is close to 1, then this strategy will also be approximately minimax.

It is interesting to note that although Theorem 5.1 is proved by transforming to a model covered by Theorem 3.1 and Corrollary 3.1, Corrollary 3.1 itself is indeed a special case of Theorem 5.1.

6. Some exact minimax results for n=1.

Scott and Smith (1975) derived some exact minimaxity results for n=1. We shall show that similar results also hold for our problem.

When the sample size n=1, there is only one estimator satisfying the representativeness condition (2.2), i.e., $\hat{Y} = x_i^{-1} y_i X$. Therefore the problem is purely the choice of designs. Let P* be the design such that the ith unit is selected with probability proportional to x_i, then we have the following

Theorem 6.1. Let n=1, $g(x) = x^{\frac{1}{2}}$, and $\hat{Y}$ be the estimator $x_i^{-1} y_i X$. If $\sum_{i \in S_0} x_i = \frac{1}{2} \sum_{i=1}^{N} x_i$ for some $S_0 \subset \{1,2,\ldots,N\}$, the P^* minimizes $\sup_{\theta \in \Theta, \underset{\sim}{\delta} \in L_2(M)} R_1(P,\hat{Y};\theta,\underset{\sim}{\delta})$ over all sampling designs P for any Θ and M > 0.

Proof. By proposition 2.2, without loss of generality, we may impose the indentifiability condition (1.7) on θ, i.e.,

$$\theta = Y/X \quad \text{and} \quad \sum_{i=1}^{N} \delta_i \, x_i^{\frac{1}{2}} = 0. \tag{6.1}$$

Then the mean squared error of $(p^*, \hat{Y})$ is

$$\begin{aligned} R_1(P^*,\hat{Y};\ \theta,\underset{\sim}{\delta}) &= X^{-1} \sum_{i=1}^{N} x_i \ (x_i^{-1} y_i X - Y)^2 \\ &= X \sum_{i=1}^{N} \delta_i^2 \ . \end{aligned}$$

The last equality follows from (6.1) and the assumption that $y_i = \theta x_i + \delta_i x_i^{\frac{1}{2}}$. Let $\tilde{L}_2(M) = \{ \underset{\sim}{\delta} \in L_2(M): \ \sum \delta_i \ x_i^{\frac{1}{2}} = 0\}$. Then

$$\sup_{\theta \in \Theta, \underset{\sim}{\delta} \in \tilde{L}_2(M)} R_1(P^*, \hat{Y};\ \theta,\underset{\sim}{\delta}) \ \leq \ MX. \tag{6.2}$$

On the other hand, let

$$\underset{\sim}{\delta}_i^* = \begin{cases} (Mx_i/X)^{\frac{1}{2}} , & \text{if } i \in S_0, \\ -(Mx_i/X)^{\frac{1}{2}} , & \text{if } i \notin S_0. \end{cases}$$

Then obviously $\sum \delta_i^{*2} = M$ and $\sum \delta_i^* \ x_i^{\frac{1}{2}} = 0$. Therefore $\underset{\sim}{\delta}^* = (\delta_1^*,\ldots,\delta_N^*)' \in \tilde{L}_2(M)$ and for any P,

$$R_1(P,\hat{Y};\ \theta,\underset{\sim}{\delta}^*) = \sum_{i=1}^{N} p_i(x_i^{-1}y_iX - Y)^2$$

$$= \sum_{i=1}^{N} p_i\ x_i^{-1}\ \delta_i^{*2}\ X^2$$

$$= MX\ (\sum_{i=1}^{N} p_i)$$

$$= MX\ .$$

It follows that $\sup_{\theta \in \Theta, \underset{\sim}{\delta} \in \tilde{L}_2(M)} R_1(P,\hat{Y};\ \theta,\underset{\sim}{\delta}) \geq MX \geq \sup_{\theta \in \Theta, \underset{\sim}{\delta} \in \tilde{L}_2(M)} R_1(P^*,\hat{Y};\ \theta,\underset{\sim}{\delta})$. □

Readers familiar with Scott and Smith (1975) could easily recognize the similarity between Theorem 6.1 and their Theorem 1. In fact Theorem 1 of Scott and Smith (1975) can be extenced to the following

Theorem 6.2. Let $n = 1$, $g(x) = x$, and $\hat{Y}$ be the estimator $x_i^{-1}y_iX$. If $\sum_{i \in S_0} x_i = \frac{1}{2}\sum_{i=1}^{N} x_i$ for some $S_0 \subset \{1,2,\ldots,N\}$, then P^* minimizes

$\sup_{\theta \in \Theta, \underset{\sim}{\delta} \in L_\infty(M)} R_1(p,\hat{Y};\ \theta,\underset{\sim}{\delta})$ over all sampling designs for any Θ and $M > 0$.

Theorem 1 of Scott and Smith (1975) becomes a special case of Theorem 6.2 with $\Theta = \{B/2\}$ and $M = B/2$. One can also write down a similar result for the problem of estimating θ. The condition needed is then $\sum_{i \in S_0} z_i = \frac{1}{2}\sum_{i=1}^{N} z_i$ for some S_0, where $z_i = x_i^2/[g(x_i)]^2$.

If $|\sum_{i \in S_0} x_i - \frac{1}{2}| < \varepsilon$ for some S_0, then results similar to Theorem 2 of Scott and Smith (1975) (i.e., approximate minimaxity) can also be established.

References

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

[2] Bickel, P.J. and Lehmann, E.L. (1981). A minimax property of the sample mean in finite populations. Ann. Statist. **9** 1119-1122.

[3] Blackwell, David and Girshick, M.A. (1954). Theory of Games and Statistical Decisions. Wiley, New York.

[4] Brewer, K.W.R. (1963). Ratio estimation in finite populations: Some results deducible from the assumption of an underlying stochastic process. Australian Jour. Stat., **5**, 93-105.

[5] Cassel, C.M., Särndal, C.E., and Wretman, J.H. (1977). Foundations of Inference in Survey Sampling. Wiley, New York.

[6] Hajék, J. (1959). Optimum strategy and other problems in probability sampling. Casopis. Pest. Math. **84**, 387-423.

[7] Hansen, M.H., and Hurwitz, W.N. (1943). On the theory of sampling from finite populations. Ann. Math. Stat., **14**, 333-362.

[8] Hodges, J.L., Jr. and Lehman, E.L. (1981). Minimax estimation in simple random sampling (To appear in Essays in Statistics and Probability in honor of C.R. Rao (P.R. Krishnaiah, ed.). North-Holland.

[9] Kiefer, J. (1974). General equivalence theory for optimum designs (approximate theory). Ann. Stat., **2**, 849-879.

[10] Kiefer, J. (1975). Constructions and optimality of generalized Youden designs, in A Survey of Statistical Designs and Linear Models (J.N. Srivastava, ed.), 333-353. North Holland, Amsterdam.

[11] Rao, J.N.K., Hartley, H.O., and Cochran, W.G. (1962). A simple procedure of unequal probability sampling without replacement. Jour. Roy. Stat. Soc. **B** **24**, 482-491.

[12] Royall, R.M. (1970a). On finite population sampling theory under certain linear regression models. Biometrika, **57**, 377-387.

[13] Royall, R.M. (1970b). Finite population sampling - on labels in estimation., Ann. Math. Stat., **41**, 1774-1779

[14] Scott, A.J., and Smith, T.M.F. (1975). Minimax designs for sample surveys. Biometrika **62**, 353-357.

[15] Stenger, H. (1979). A minimax approach to randomization and estimation in survey sampling. Ann Stat. **7**, 395-399.