Optimality Criteria in Survey Sampling

by

Ching-Shui Cheng and Ker-Chau Li
Department of Statistics, University of California, Berkeley
Department of Statistics, Purdue University

Mimeograph Series #83-40

Department of Statistics
Purdue University
August 1983

ABSTRACT

This paper attempts to fill the gap between some different and conflicting approaches in survey sampling. Based on a fixed population regression-type model, a class of optimality criteria similar to those well-accepted in the optimum experimental design theory is introduced. The minimax and the super-population approaches in survey sampling turn out to correspond to two extreme criteria of the proposed class. This helps understand the role of randomization. The strategy of simple random sampling with sample mean and the Rao-Hartley-Cochran strategy are shown to possess several desirable optimum properties.

Key words and phrases. Adjusted risk-generating matrix, optimum experimental design, Rao-Hartley-Cochran strategy, simple random sampling, superpopulation model.

1. Introduction

In this paper, some ideas and results from the theory of optimum experimental designs will be applied to survey sampling. This is an attempt to unify and fill the gap between some different and conflicting approaches in sampling theory (such as the minimax and the superpopulation approaches). A class of optimality criteria is used to interpret and understand the different viewpoints.

Consider a survey sampling problem where an unknown value y_i is associated with the ith unit in a finite population of size N. Our objective is to estimate the unknown population mean $\overline{Y} = \frac{1}{N}\sum_{i=1}^{N} y_i$ (or equivalently the population total) based on a sample of size n. The problem is how to select the sample and how to construct a good estimator. Quite often, correlated with each y_i there may exist a known q-vector $\underset{\sim}{x}_i = (x_{i1}, x_{i2}, \ldots, x_{iq})'$. The auxiliary information provided by these $\underset{\sim}{x}$ vectors, when suitably utilized, may greatly improve the estimation. For this purpose, a superpopulation model (Brewer (1963), Royall (1970a)) of the form

$$y_i = \underset{\sim}{x}_i'\underset{\sim}{\theta} + \epsilon_i \qquad , \qquad i = 1,2,\ldots, N. \tag{1.1}$$

is often assumed, where $\underset{\sim}{\theta} = (\theta_1, \theta_2, \ldots, \theta_q)'$ is an unknown q-vector and the ϵ_i's are uncorrelated random variables with mean zero and variances $\lambda^2 g_i$ (g_i's are known; whereas λ may or may not be known). Then an optimum strategy may be defined as one which minimizes the mean squared error of the best linear unbiased predictor of $\overline{Y}$ over all the possible samples (when randomization is involved) and all the possible populations generated by the superpopulation model (1.1). However, the superpopulation approach often leads to the selection of a purposive sample; randomization does not play a role here. To bring in randomization (mostly because it is thought to provide some robustness for estimation), one may impose the design-unbiasedness (or its weaker form "consistency" in asymptotic setups) and define an optimal strategy as one which minimizes the mean squared error under model (1.1). This approach, however, precludes all non-randomized strategies. Also, sometimes the payoff for strictly requiring design unbiasedness may be quite high (Godambe (1955, 1982), Godambe and Joshi (1965)).

Cheng and Li (1983) introduced a different approach which can easily be extended as follows. Instead of the superpopulation model (1.1) the following fixed population model is assumed:

$$y_i = \underset{\sim}{x}_i' \underset{\sim}{\theta} + \lambda \delta_i g_i^{\frac{1}{2}}, \qquad i = 1,2,\ldots,N \quad , \tag{1.2}$$

where $\underset{\sim}{\theta} \in \Theta \subset R^q$, g_i is known, λ may or may not be known, and $\underset{\sim}{\delta} = (\delta_1, \delta_2, \ldots, \delta_N)'$ is assumed to be in some neighborhood of $(0,0,\ldots,0)'$. For q = 1, Cheng and Li established some approximately minimax properties of the Rao-Hartley-Cochran (1962) and Hansen-Hurwitz (1943) strategies. An earlier related result justifying probability proportional to size sampling was obtained by Scott and Smith (1975). The minimax criterion has also been used to justify the use of simple random sampling and sample mean by many authors (Blackwell and Girshick (1954), Bickel and Lehmann (1981), Hodges and Lehmann (1981), Stenger (1979), Royall (1970b)). One crucial assumption made in these works is that $\underset{\sim}{y} = (y_1, y_2, \ldots, y_N)'$ be in a permutation-invariant subset of R^N. This amounts to assuming that, in (1.2), $\underset{\sim}{x}_1 = \underset{\sim}{x}_2 = \ldots = \underset{\sim}{x}_N$, $g_1 = g_2 = \ldots = g_N$, and $\underset{\sim}{\delta}$ lies in a permutation-invariant neighborhood of $\underset{\sim}{0}$.

The minimax criterion is sometimes criticized as being "too pessimistic" because it guards against the worst cases. But on the other hand, the superpopulation approach seems to be "too optimistic". It is the purpose of this paper to introduce other criteria to fill the gap between these two approaches.

Of particular interest is a family of criteria analogous to the Φ_p-criteria in the theory of optimum designs. It turns out that the superpopulation approach and the minimax criterion correspond to the two special cases $p = 1$ and $p = \infty$, respectively. This puts the two approaches in the same framework and may help to clarify the role of randomization in general.

The paper is organized as follows. Optimality criteria will be defined in Section 2. The relation of the minimax criterion and the superpopulation approach to the Φ_p-optimality criteria is discussed there. Kiefer's (1975) well-known result on universal optimality is applied to the strategy of simple random sampling and sample mean in Section 3. Section 4 contains another application of this powerful optimality tool to the Rao-Hartley-Cochran strategy.

The following notation will be used throughout the paper:

s : a subset of $\{1,2,\ldots,N\}$ with cardinality n.

S : the set of all samples of size n.

P : a probability distribution on S.

$\underset{\sim}{a}_s = (a_{s1}, a_{s2}, \ldots, a_{sN})'$: an N-vector such that $a_{si} = 0$ for all $i \notin s$.

$\underset{\sim}{a} = \{\underset{\sim}{a}_s : s \in S\}$.

Here s denotes a sample of size n, P denotes a sampling design, and $\sum_{i=1}^{N} a_{si}y_i = \sum_{i \in s} a_{si}y_i$ is a linear estimator of the population mean given that sample s is selected. Note that only linear estimators will be considered in this paper and we shall call $(P,\underset{\sim}{a})$ a <u>strategy</u>.

2. Optimality criteria

The mean squared error $R(\underset{\sim}{y};P,\underset{\sim}{a})$ of a strategy $(P,\underset{\sim}{a})$ is

$\sum_{s \in S} \{\sum_{i=1}^{N} (N^{-1} - a_{si})y_i\}^2 P(s)$ which, by (1.2), equals

$$\sum_{s \in S} P(s) \quad [\{\sum_{i=1}^{N} (N^{-1} - a_{si}) \underset{\sim}{x}_i'\}\underset{\sim}{\theta} + \lambda \sum_{i=1}^{N} (N^{-1} - a_{si}) \delta_i g_i]^2. \tag{2.1}$$

A strategy is said to be representative if

$$\sum_{i=1}^{N} (N^{-1} - a_{si}) \underset{\sim}{x}_i = \underset{\sim}{0} \text{ for all s with } P(s) > 0. \tag{2.2}$$

In view of (2.1), representative strategies are desirable when Θ is unbounded (since otherwise the risks would also be unbounded). In the rest of the paper, we will only consider representative strategies.

Now let $\underset{\sim}{G} = \text{diag}\,(g_1^{\frac{1}{2}}, g_2^{\frac{1}{2}}, \ldots, g_N^{\frac{1}{2}})$. Then the mean squared error of a representative strategy $(P, \underset{\sim}{a})$ can be simplified to

$$R(\underset{\sim}{y}; P, \underset{\sim}{a}) = \lambda^2 \underset{\sim}{\delta}' \underset{\sim}{G} \{ \sum_{s \in S} P(s) (\underset{\sim}{a}_s - N^{-1} \underset{\sim}{1}) (\underset{\sim}{a}_s - N^{-1} \underset{\sim}{1})' \} \underset{\sim}{G} \underset{\sim}{\delta},$$

where $\underset{\sim}{1}$ is the Nx1 vector of ones. This is clearly a quadratic form in $\underset{\sim}{\delta}$. We shall call the matrix $R(P, \underset{\sim}{a}) \equiv \underset{\sim}{G} \{ \sum_{s \in S} P(s) (\underset{\sim}{a}_s - N^{-1} \underset{\sim}{1}) (\underset{\sim}{a}_s - N^{-1} \underset{\sim}{1})' \} \underset{\sim}{G}$ the adjusted risk generating matrix of strategy $(P, \underset{\sim}{a})$. The comparison between two strategies now reduces to the comparison between their corresponding adjusted risk generating matrices; strategy $(P_1, \underset{\sim}{a}_1)$ is at least as good as strategy $(P_2, \underset{\sim}{a}_2)$ if and only if $\underset{\sim}{R}(P_2, \underset{\sim}{a}_2) - \underset{\sim}{R}(P_1, \underset{\sim}{a}_1)$ is non-negative definite. This is similar to the comparison of experimental designs in terms of their information matrices. It is well known that a design with information matrix $\underset{\sim}{M}_1$ is at least as good as another design with information matrix $\underset{\sim}{M}_2$ if and only if $\underset{\sim}{M}_1 - \underset{\sim}{M}_2$ is non-negative definite (see Ehrenfeld (1955)). Of course an adjusted risk generating matrix here should be thought of as a covariance matrix and therefore corresponds to the inverse (or generalized-inverse) of an information matrix.

Unfortunately, the comparison in terms of matrix domination introduced above usually leaves a large class of admissible strategies. Following the theory of optimal designs, one could introduce some optimality criterion to define the "best" strategy. Let $\mathcal{R}$ be the set of all adjusted risk generating matrices and Ψ be a real-valued function defined on $\mathcal{R}$. Then a strategy $(P,\underset{\sim}{a})$ is called Ψ-optimal if $\Psi(\underset{\sim}{R}(P,\underset{\sim}{a})) \leq \Psi(\underset{\sim}{R}(P',\underset{\sim}{a}'))$ for any $(P',\underset{\sim}{a}')$.

One important class of criteria in the theory of optimal design is the Φ_p criteria defined in terms of the eigenvalues of the information matrices. They can easily be adapted to our problem. Note that because of (2.2), the dimension of the null space of any $\underset{\sim}{R}(P,\underset{\sim}{a})$ is at least q (here without loss of generality it is assumed that the rank of the qxN matrix $(\underset{\sim}{x}_1,\underset{\sim}{x}_2,\ldots,\underset{\sim}{x}_N)$ is q). Let $\lambda_{(1)} \geq \lambda_{(2)} \geq \cdots \geq \lambda_{(N-q)}$ be the N-q largest eigenvalues of $\underset{\sim}{R}(P,\underset{\sim}{a})$. For any $p > 0$, define a function Φ_p on $\mathcal{R}$ by

$$\Phi_p(\underset{\sim}{R}(P,\underset{\sim}{a})) = \left\{\sum_{i=1}^{N-q} \lambda_{(i)}^{\ p} \ / \ (N-q)\right\}^{1/p} .$$

For p=1, $\Phi_1(\underset{\sim}{R}(P,\underset{\sim}{a})) = \{\text{trace } \underset{\sim}{R}(P,\underset{\sim}{a})\} / (N-q)$ is simply the trace criterion (A-criterion). One can also show that $\lim_{p\to\infty} \Phi_p(\underset{\sim}{R}(P,\underset{\sim}{a})) = \lambda_{(1)}$, the maximum eigenvalue of $\underset{\sim}{R}(P,\underset{\sim}{a})$. Thus Φ_∞ is a kind of minimax criterion (E-criterion). In fact, if in (1.2), one assumes that $\underset{\sim}{\delta}$ belongs to an L_2-neighborhood $L_2(M) = \{\underset{\sim}{\delta} : \sum_{i=1}^{N} \delta_i^{\,2} \leq M\}$ for some $M > 0$, then minimizing $\Phi_\infty(\underset{\sim}{R}(P,\underset{\sim}{a}))$ is the same as minimizing the maximum mean squared error $\sup_{\underset{\sim}{\delta} \in L_2(M)} R(\underset{\sim}{y};P,\underset{\sim}{a})$. This is one approach taken by Cheng and Li (1983).

The following theorem establishes an interesting connection between the Φ_1-criterion and the superpopulation approach.

<u>THEOREM 2.1</u>. A strategy $(P,\underset{\sim}{a})$ is optimal under the superpopulation model (1.1) if and only if it is Φ_1-optimal (A-optimal) under model (1.2).

Proof. It is clear that the unbiasedness of $\sum_{i=1}^{N} a_{si} y_i$ is equivalent to the representativeness condition (2.2). Write $\underset{\sim}{\epsilon} = (\epsilon_1, \epsilon_2, \ldots, \epsilon_N)'$. For a representative strategy, the mean squared error under model (1.1) is

$$E \sum_{s \in S} (\bar{Y} - \sum_{i=1}^{N} a_{si} y_i)^2 P(s) = E \sum_{s \in S} \left\{ \sum_{i=1}^{N} (N^{-1} - a_{si}) \epsilon_i \right\}^2 P(s)$$

$$= E \sum_{s \in S} \underset{\sim}{\epsilon}' (N^{-1} \underset{\sim}{1} - \underset{\sim}{a}_s)(N^{-1} \underset{\sim}{1} - \underset{\sim}{a}_s)' \underset{\sim}{\epsilon} P(s)$$

$$= E \sum_{s \in S} P(s) \text{ trace}\{(N^{-1} \underset{\sim}{1} - \underset{\sim}{a}_s)(N^{-1} \underset{\sim}{1} - \underset{\sim}{a}_s)' \underset{\sim}{\epsilon}\underset{\sim}{\epsilon}'\}$$

$$= \lambda^2 \text{ trace}\left\{ \sum_{s \in S} P(s) (N^{-1} \underset{\sim}{1} - \underset{\sim}{a}_s)(N^{-1} \underset{\sim}{1} - \underset{\sim}{a}_s)' \underset{\sim}{G}^2 \right\}$$

$$= \lambda^2 \text{ trace } \underset{\sim}{R}(P, \underset{\sim}{a}).$$

Therefore to minimize the mean squared error is the same as to minimize $\Phi_1(\underset{\sim}{R}(P,\underset{\sim}{a}))$. This completes the proof.

Thus we see that the superpopulation approach of (1.1) and the minimax criterion are two special cases of the Φ_p-criteria with P=1 and ∞, respectively. Let $\mathfrak{M}$ be the confex hull of $\mathcal{R}$. Then it is clear that the extreme points of $\mathfrak{M}$ are adjusted risk generating matrices of non-randomized strategies. For $p \leq 1$, Φ_p is a concave function and hence its minimum usually is attained at some extreme points of $\mathfrak{M}$. This explains why a superpopulation model (p=1) often leads to a non-randomized optimum strategy. On the other hand, for $p > 1$, Φ_p is strictly convex and its minimum is often (but not always) attained at some mixture of extreme points, i.e., a randomized strategy. In particular, for $p = \infty$, one sees why the minimax approach often yields a randomized design. This provided a unified view of the two different approaches. The difference lies in the criteria used. Wynn (1977 a,b) also applied some theory on the D-optimum experimental designs to survey sampling, but his approach is very different from ours. He assumed superpopulation model (1.1) and was interested in minimizing the maximum mean squared error of "prediction" over the unsampled units. This different notion of minimaxity leads to a purposive optimal sampling design. The techniques he used were mainly the kind usually found in the approximate theory of optimal design, while our methods are borrowed from the exact theory.

3. Universal optimality of the strategy of simple random sampling and sample mean

The minimaxity of the strategy (simple random sampling, sample mean) is well-known. In this section, we shall use an important result in optimal design theory to show that it is also optimal with respect to a large class of criteria including all the Φ_p-criteria, $p \geq 1$.

Consider the case where $\underset{\sim}{x}_1=\underset{\sim}{x}_2=\ldots=\underset{\sim}{x}_N$ and $g_1=g_2=\ldots=g_N$. The representativeness condition (2.2) is reduced to

$$\sum_{i=1}^{N} a_{si} = 1 \text{ for all } s \text{ with } P(s) > 0.$$

An important consequence of the above condition is that $\underset{\sim}{R}(P,\underset{\sim}{a})$ has zero row and column sums. Kiefer's (1975) Proposition 1 becomes relevant here.

Suppose $\mathcal{B}_N$ is the set of all the NxN nonnegative definite symmetric matrices with zero row and column sums, and $\mathcal{C} \subset \mathcal{B}_N$. Kiefer (1975) showed that if there exists a matrix $\underset{\sim}{C}^*$ in $\mathcal{C}$ such that

(a) $\underset{\sim}{C}^*$ is completely symmetric in the sense that $\underset{\sim}{C}^*$ is of the form $a\underset{\sim}{I}_N + b\underset{\sim}{J}_N$, where $\underset{\sim}{I}_N$ is the identity matrix of order N and $\underset{\sim}{J}_N$ is the NxN matrix of ones,

and

(b) $\underset{\sim}{C}^*$ maximizes tr $\underset{\sim}{C}$ over $\mathcal{C}$,

then $\underset{\sim}{C}^*$ minimizes $\Phi(\underset{\sim}{C})$ for all the real-valued functions Φ defined on $\mathcal{B}_N$ satisfying the following three conditions.

(i) Φ is convex,

(ii) $\Phi(b\underset{\sim}{C})$ is nonincreasing in the scalar $b \geq 0$,

(iii) Φ is invariant under any simultaneous permutation of rows and columns (permutation of coordinates).

Such a matrix $\underset{\sim}{C}^*$ is called "universally optimal". The above result is an important tool for showing the optimality of balanced incomplete block designs and many other symmetric designs such as Latin squares and Youden squares, wherein each matrix in $\mathcal{C}$ is an information matrix comparable to a generalized inverse of an adjusted risk generating matrix in the present setting. In order to apply Kiefer's result to our problem, one has to compute the generalized inverses of adjusted risk generating matrices, which is a formidable task. One way to get around this inversion is to use a weaker form of universal optimality, as employed in Kiefer and Wynn (1981). Let $\overline{\mathcal{C}}$ be a *subset of* $\mathcal{B}_N$. *The matrices in* $\overline{\mathcal{C}}$ can be thought of as the adjusted risk generating matrices or the generalized inverses of information matrices. It can be shown that if there exists a matrix $\underset{\sim}{C}^*$ in $\mathcal{C}$ such that

subset of $\mathcal{B}_N$. It can be shown that if there exists a matric $\overline{\underset{\sim}{C}}^*$ in $\overline{\mathcal{C}}$ such that

(a') $\overline{\underset{\approx}{C}}^*$ is completely symmetric,

and

(b') $\overline{\underset{\sim}{C}}^*$ minimizes tr $\overline{\underset{\sim}{C}}$ over $\overline{\mathcal{C}}$,

then $\overline{\underset{\sim}{C}}^*$ minimizes $\Psi(\overline{\underset{\sim}{C}})$ over $\overline{\mathcal{C}}$ for any real-valued function Ψ defined on $\mathcal{B}_N$ satisfying conditions (i), (iii) and

(ii') $\Psi(b\overline{\underset{\sim}{C}})$ is nondecreasing in the scalar $b \geq 0$.

Such a matrix $\overline{\underset{\sim}{C}}^*$ is called <u>weakly universally optimal</u> by Kiefer and Wynn (1981). This is a weaker form of optimality because the class of convex decreasing functionals Φ of matrices in $\mathcal{B}_N$ is more general than the class of convex increasing functionals Ψ of the generalized inverses of matrices in $\mathcal{B}_N$. For example, when $p < 1$, the Φ_p-criterion is covered by Proposition 1 in Kiefer (1975) but is a strictly concave function on the convex hull of $\mathcal{B}$. Nevertheless, the class of criteria satisfying (i), (ii'), and (iii) is still quite large; e.g., it contains all the Φ_p-criteria with $p \geq 1$.

Note that condition (b') is simply the Φ_1-optimality. Therefore for a strategy with completely symmetric adjusted risk-generating matrix, the optimality under superpopulation model (1.1) implies the optimality under model (1.2) with respect to a large class of criteria. It is straightforward to see that the adjusted risk-generating matrix of the strategy d*≡ (simple random sampling, sample mean) is $a\underset{\sim}{I}_N + b\underset{\sim}{J}_N$ where $a = (N-n)n^{-1}N^{-1}(N-1)^{-1}$ and $b = -a/N$. Furthermore, it is fairly easy to show that d* is Φ_1-optimal (In fact, any strategy using the sample mean as an estimator of $\overline{Y}$ is Φ_1-optimal). This proves the "weakly universal optimality" of the strategy of simple random sampling and sample mean; in particular, it is Φ_p-optimal for all $p \geq 1$.

The above argument shows how a well-known theorem in optimal design theory finds an application in survey sampling. Our result is different from the various minimax properties proven in the literature. Stenger (1979) considered general convex loss functions, but he still used a minimax criterion. Furthermore, our proof clearly reveals an interesting connection to design theory: the strategy of simple random sampling and sample mean plays the same role as a balanced incomplete block design.

4. More results on universal optimality

The discussion in the last section shows that the strategy of simple random sampling and sample mean is a balanced strategy; all the nonzero eigenvalues of the adjusted risk generating matrix are the same. This property is crucial to its weakly universal optimality. Another example of balanced strategy is the

Rao-Hartley-Cochran strategy which was shown by Cheng and Li (1983) to enjoy some approximately minimax properties.

Now assume that $q = 1$ and $g_i = x_{i1}$ which we shall write as x_i. Then $\underset{\sim}{G} = \text{diag}\,(x_1^{\frac{1}{2}}, x_2^{\frac{1}{2}}, \ldots, x_N^{\frac{1}{2}})$. In this case, the representativeness condition (2.2) implies that for any $\underset{\sim}{R}$ in $\mathcal{R}$, $\underset{\sim}{R}(x_1^{\frac{1}{2}}, x_2^{\frac{1}{2}}, \ldots, x_N^{\frac{1}{2}})' = \underset{\sim}{0}$ (see Proposition 2.1 of Cheng and Li (1983)). Let $\underset{\sim}{R}$ (RHC) be the adjusted risk generating matrix of the Rao-Hartley-Cochran strategy. Then it was calculated in Cheng and Li (1983) that

$$\underset{\sim}{R}(\text{RHC}) = a\underset{\sim}{I}_N + b\underset{\sim}{G}\underset{\sim}{J}_N\underset{\sim}{G}, \tag{4.1}$$

where $a = N^{-2}n^{-1}\mu \sum_{i=1}^{N} x_i$ and $b = -N^{-2}n^{-1}\mu$, with μ being the finite population correction $(N-n)/(N-1) + k(n-k)/N(N-1)$ where $k = N-n[N/n]$. Thus $\underset{\sim}{R}(\text{RHC})$ acts as a multiple of the identity matrix in all the directions orthogonal to the vector $(x_1^{\frac{1}{2}}, x_2^{\frac{1}{2}}, \ldots, x_N^{\frac{1}{2}})'$.

Unlike in Section 3, now the matrices in $\mathcal{R}$ no longer have zero row and column sums. The key step in the proof of Proposition 1 in Kiefer (1975) is to take the average of an arbitrary matrix in $\mathcal{C}$ with respect to all the simultaneous permutations of rows and columns. Such an argument fails here since the set of matrices orthogonal to $(x_1^{\frac{1}{2}}, \ldots, x_N^{\frac{1}{2}})'$ is not permutation invariant. Instead one can take the average with respect to all the orthogonal transformations that leave the vector $(x_1^{\frac{1}{2}}, x_2^{\frac{1}{2}}, \ldots, x_N^{\frac{1}{2}})'$ invariant, and prove the following analogue of Proposition 1' in Kiefer (1975).

Proposition 4.1 Let $\mathcal{B}_N'$ be the set of all the NxN nonnegative definite symmetric matrices $\underset{\sim}{C}$ such that $\underset{\sim}{C}\underset{\sim}{a} = \underset{\sim}{0}$ for some nonzero vector $\underset{\sim}{a}$, and $\overline{\mathcal{C}}$ is a subset of $\mathcal{B}_N$. If there exists a matrix $\underset{\sim}{\overline{C}}^*$ such that all the N-1 largest eigenvalues of $\underset{\sim}{\overline{C}}^*$ are equal and $\underset{\sim}{\overline{C}}^*$ minimizes tr $\underset{\sim}{\overline{C}}$ over $\overline{\mathcal{C}}$, then $\underset{\sim}{\overline{C}}^*$ minimizes $\Psi(\overline{C})$ over $\overline{\mathcal{C}}$ for all real-valued functions Ψ defined on $\mathcal{B}_N'$ satisfying conditions (i), (ii') in section 3 and

(iii') $\Psi(\underset{\sim}{O}\overline{\underset{\sim}{C}}\underset{\sim}{O}') = \Psi(\overline{\underset{\sim}{C}})$ for any orthogonal matrix $\underset{\sim}{O}$ such that $\underset{\sim}{O}\underset{\sim}{a} = \underset{\sim}{a}$.

Such a matrix $\overline{\underset{\sim}{C}}^*$ is again said to be "weakly universally optimal". (Note that it is well-known that Proposition 1' in Kiefer (1975) contains an error; the condition of "permutation invariance" should be replaced by "orthogonal invariance". A correct version in fact appeared in an earlier paper by Kiefer (1971)). Condition (iii') implies that Ψ is a function of the eigenvalues. One may rephrase conditions (i), (ii') and (iii') in terms of the eigenvalues. Another variation is to define weakly universal optimality in terms of Schur-convex functions of the N-1 largest eigenvalues of the adjusted risk generating matrices. Again, all the Φ_p-criteria with $p \geq 1$ satisfy the three conditions (i), (ii'), and (iii').

Although the Rao-Hartley-Cochran strategy is a balanced strategy in the present setting, it is not "weakly universally optimal" since it is not Φ_1-optimal (A-optimal). As shown earlier, a Φ_1-optimal strategy is the optimal strategy under the superpopulation model which, according to the results of Brewer (1963) and Royall (1970a), comprises a purposive sampling design selecting the n largest x-values and the ratio estimator. Let $\underset{\sim}{R}(BR)$ be the Φ_1-optimal strategy and $x_{(i)}$ be the ith smallest x-value. Then $\underset{\sim}{R}(BR)$ has only one nonzero eigenvalue $N^{-2}X\sum_{i=1}^{N-n} x_{(i)} \Big/ \sum_{i=N-n+1}^{N} x_{(i)}$, where $X = \sum_{i=1}^{N} x_i$. By (4.1), all the N-1 nonzero eigenvalues of $\underset{\sim}{R}(RHC)$ are equal to $N^{-2}n^{-1}\mu X$. Obviously $\text{tr}\,\underset{\sim}{R}(BR) = N^{-2}X\sum_{i=1}^{N-n} x_{(i)} \Big/ \sum_{i=N-n+1}^{N} x_{(i)} \leq N^{-2}(N-1)n^{-1}\mu X = \text{tr}\,\underset{\sim}{R}(RHC)$. For $p > 1$,

$$\Phi_p(\underset{\sim}{R}(BR)) = (N-1)^{-1/p} N^{-2}X \sum_{i=1}^{N-n} x_{(i)} \Big/ \sum_{i=N-n+1}^{N} x_{(i)} \text{ and } \Phi_p(\underset{\sim}{R}(RHC)) = N^{-2}n^{-1}\mu X.$$

By direct comparison we see that the Rao-Hartley-Cochran strategy is Φ_p - better than the Φ_1-optimal strategy if $(N-1)^{-1/p} \sum_{i=1}^{N-n} x_{(i)} \geq n^{-1} \sum_{i=N-n+1}^{N} x_{(i)}$.

In particular, if $\sum_{i=1}^{N-n} x_{(i)} \geq n^{-1} \sum_{i=N-n+1}^{N} x_{(i)}$, then the Rao-Hartley-Cochran strategy is Φ_∞-better than the Φ_1-optimal strategy.

Although the Rao-Hartley-Cochran strategy is not weakly universally optimal, by Proposition 4.1 it is weakly universally optimal over the subclass $\{\underset{\sim}{R} : \underset{\sim}{R} \in \mathcal{R}$ and trace $\underset{\sim}{R} \geq$ trace $\underset{\sim}{R}(RHC)\}$. We shall show that this is a quite large class that contains some commonly used competitors to the Rao-Hartley-Cochran strategy. More specifically, we have the following

Theorem 4.2 Assume that $q = 1$ and $q_i = x_i$, $i=1,2,\ldots,N$. If $n|N$, then $\Psi(\underset{\sim}{R}(RHC)) \leq \Psi(\underset{\sim}{R}(P,\underset{\sim}{a}))$ for any Ψ satisfying (i), (ii'), and (iii'), and any representative strategy with $\sum_{s \in S} p(s)/X_s \geq (n\overline{X})^{-1}$, where $\overline{X} = N^{-1} \sum_{i=1}^{N} x_i$ and $X_s = \sum_{i \in s} x_i$.

Proof When $n|N$, $\mu = (N-n)/(N-1)$ and therefore tr $\underset{\sim}{R}(RHC) = n^{-1}N^{-2}(N-n)X$. By proposition 4.1, it is enough to show that if $\sum_{s \in S} p(s)/X_s \geq (n\overline{X})^{-1}$, then tr $\underset{\sim}{R}(P,\underset{\sim}{a}) \geq n^{-1}N^{-2}(N-n)X$. Now for any strategy $(P,\underset{\sim}{a})$, we have

$$\text{tr } \underset{\sim}{R}(P,\underset{\sim}{a}) = \sum_{i=1}^{N} \sum_{s \in S} P(s)\,(a_{si}-N^{-1})^2 x_i$$

$$= \sum_{s \in S} P(s)\{\sum_{i \notin s} N^{-2}x_i + \sum_{i \in s}(a_{si}-N^{-1})^2 x_i\}. \qquad (4.2)$$

Recall the representativeness condition $\sum_{i \in s} a_{si}x_i = \overline{X}$. Minimizing $\sum_{i \in s} (a_{si}-N^{-1})^2 x_i$

in (4.2) subject to the constraint $\sum_{i\in s} a_{si}x_i = \bar{X}$ yields $a_{si} = \bar{X}/X_s$. Thus if $\sum_{s\in S} P(s)/X_s \geq (n\bar{X})^{-1}$, then

$$\begin{aligned} \operatorname{tr} \underset{\sim}{R}(P,\underset{\sim}{a}) &\geq \sum_{s\in S} P(s)\{\sum_{i\notin s} N^{-2}x_i + \sum_{i\in s} (\bar{X}/X_s - N^{-1})^2 x_i\} \\ &= \sum_{s\in S} P(s)\{\sum_{i=1}^{N} N^{-2}x_i + \bar{X}^2/X_s - 2\bar{X}/N\} \\ &= \sum_{s\in S} P(s)\bar{X}^2/X_s - \bar{X}/N \\ &\geq n^{-1}N^{-2}(N-n)X, \end{aligned}$$

where the last inequality follows from the assumption that $\sum_{s\in S} p(s)/X_s \geq (n\bar{X})^{-1}$. This completes the proof. □

Note that the condition in Theorem 4.2 is mainly on sampling design; <u>no other assumption is made on the estimator except the condition of representativeness</u>. This indeed covers a very large class of strategies which, by the following proposition, includes the two important sampling designs commonly used with the ratio estimator: <u>simple random sampling</u> and the scheme of Lahiri and Midzuno (also known as <u>probability proportional to aggregate size sampling)</u> in which each sample s is selected with probability proportional to X_s.

<u>Proposition 4.3</u> Let P be a simple random sampling or a probability proportional to aggregate size sampling. Then $\sum_{s\in S} P(s)/X_s \geq (n\bar{X})^{-1}$.

<u>Proof</u> For probability proportional to aggregate size sampling, one has the equality $\sum_{s\in S} P(s)/X_s = (n\bar{X})^{-1}$. For simple random sampling, the result follows from a simple application of the Jensen inequality. □

Thus we see that the Rao-Hartley-Cochran strategy is better than <u>any</u> representative strategy using simple random sampling or probability proportional to aggregate size sampling with respect to <u>all</u> the criteria satisfying (i), (ii'), and (iii') including all the Φ_p-criteria with $p \geq 1$. In particular it beats the strategies of (simple random sampling, ratio estimator), (probability proportional to aggregate size sampling, ratio estimator) and many other unbiased ratio-type estimates with simple random sampling (Cochran 1977, page 174). This extends the comparison made in Section 4 of Cheng and Li (1983).

This kind of "universal optimality" holds because we have a "balanced" strategy. The striking result of Kiefer (1975) again finds another powerful application.

Acknowledgement

This research was prepared with the support of National Science Foundation Grants No. MCS-82-00909 and MCS-82-00631.

References

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

Blackwell, D. and Girshick, M. A. (1954). Theory of Games and Statistical Decisions. Wiley, New York.

Brewer, K. W. R. (1963). Ratio estimation in finite populations: Some results deducible from the assumption of an underlying stochastic process. Australian J. Statist. 5, 93-105.

Cheng, C. S. and Li, K. C. (1983). A minimax approach to sample surveys. Ann. Statist. 11, 552-563.

Cochran, W. G. (1977) Sampling Techniques. 3rd ed. Wiley, New York.

Ehrenfeld, S. (1955). Complete class theorem in experimental design. Proceedings of the Third Berkeley Symp. Math. Statist. Prob. 1, 69-75.

Godambe, V. P. (1955). A unified theory of sampling from finite populations. J. Roy. Statist. Soc. Ser. B 17, 269-278.

Godambe, V. P. (1982). Estimation in survey sampling: robustness and optimality (with discussion). J. Amer. Statist. Assoc. 77, 393-406.

Godambe, V. P., and Joshi, V. M. (1965). Admissibility and Bayes estimation in sampling finite populations I. Ann. Math. Statist., 36, 1707-1722.

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

Hodges, J. L., Jr. and Lehmann, 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, Amsterdam.

Kiefer, J. (1971). The role of symmetry and approximation in exact design optimality, in Statistical Decision Theory and Related Topics I, (Gupta, S. and Yackel, J. ed.), 109-118, Academic Press, New York.

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

Kiefer, J. and Wynn, H. (1981). Optimum balanced block and Latin square designs for correlated observations. Ann. Statist. 9, 737-757.

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

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

Royall, R. M. (1970b). Finite population sampling - on labels in estimation. Ann. Math. Statist. 41, 1774-1779.

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

Stenger, H. (1979). A minimax approach to randomization and estimation in survey sampling. Ann. Statist. 7, 395-399.

Wynn, H. (1977a). Minimax purposive survey sampling design. J. Amer. Statist. Assoc. 72, 655-657.

Wynn, H. (1977b). Optimum designs for finite populations sampling. In Statistical Decision Theory and Related Topics II (Gupta, S. and Moore, D. ed.), 471-478, Academic Press, New York.