Consistent, Asymptotically Efficient Strategies

by

Ker-Chau Li*
Department of Statistics
Purdue University

Technical Report #83-1

January 1983

Department of Statistics
Purdue University
West Lafayette, IN

*Research supported by NSF Grant No. MCS-82-00631.

Consistent, Asymptotically Efficient Survey Strategies

Short title: Survey Strategies

by

Ker-Chau Li*
Department of Statistics
Purdue University

Summary

Survey strategies are studied under a linear regression superpopulation model. An asymptotic lower bound for the anticipated mean squared errors of all linear design-consistent strategies is derived. The achievability of this bound is then demonstrated for several commonly-used sampling methods including the rejective sampling, Samford-Durbin's method, successive sampling, and Rao-Hartley-Cochran's method, when the inclusion probabilities are proportional to the standard deviations of the random errors in the regression model and the regression estimates are used.

---

*Research supported by NSF Grant No. MCS-82-00631.

AMS 1980 Subject Classification: Primary 62D05; Secondary 62F12.

Keywords and Phrases: anticipated mean squared error, consistency, design-unbiasedness, efficiency, linear estimate, regression, rejective sampling, robustness, superpopulation model, survey strategy.

Consistent, Asymptotically Efficient Survey Strategies

1. Introduction.

Consider a finite population U with units labeled by 1,2,...,N. We shall write $U = \{1,\dots,N\}$ without ambiguity. Associated with unit i, $1 \le i \le N$, is a value y_i for the $\underset{\sim}{y}$ characteristic. To estimate the unknown population mean $\bar{Y} = N^{-1} \sum_{i=1}^{N} y_i$, we are allowed to take a sample $s \subset U$ either purposively or according to some random scheme. Let $\underset{\sim}{P}(s)$ denote the probability of selecting s. A probability measure $\underset{\sim}{P}$ is called a sampling design. An estimate of $\bar{Y}$ is a function of s and $\{y_i: i \in s\}$. By a strategy we mean a design-estimate pair. One problem in survey sampling is how to construct a strategy to estimate $\bar{Y}$ in a suitable way.

Quite often, associated with unit i, $1 \le i \le N$, there may exist a known p-dimensional vector $\underset{\sim}{x}_i = (x_{i1},\dots,x_{ip})'$ that is correlated with y_i in some sense. The information provided by these auxiliary vectors should be useful in obtaining a good strategy. To incorporate these $\underset{\sim}{x}$ vectors, a superpopulation regression model of the following form is often considered.

$$(1.1) \qquad y_i = \underset{\sim}{x}_i'\underset{\sim}{\beta} + \varepsilon_i, \; i = 1,2,\dots,N,$$

where $\underset{\sim}{\beta} = (\beta_1,\dots,\beta_p)'$ is an unknown p-vector and ε_i's are assumed to be uncorrelated <u>random</u> variables with means 0 and variances σ_i^2.

However, strictly depending on this superpopulation model, one is often led to the purposive sampling which may result a huge bias in estimation when (1.1) is inappropriate. One way to guard against this possible model violation is to impose the condition of the design-unbiasedness. But this rather severe principle is often more than necessary. In many cases allowing a small amount of bias, we may obtain a more useful estimate (see, e.g., Särndal (1980)). The idea of unbiasedness is useful only to the extent that greatly biased estimates are poor no matter what other properties they have (Hájek 1971). This concept naturally evolves into the conditions of the asymptotic design-unbiasedness and the asymptotic consistency with respect to the design-measure (e.g., Brewer 1979, Särndal 1980, Robinson and Tsui 1981, Isaki and Fuller 1982). As pointed out by Robinson and Tsui (1981), the asymptotic framework has to be carefully stated to avoid the obvious unreality implied by that in Brewer (1979) or Särndal (1980). The setup to be described below is in line with this viewpoint and that of Isaki and Fuller (1982).

Consider a sequence of populations $U^{(t)}$, $t = 1,2,\ldots$, such that the population size $N^{(t)}$ tends to ∞ as t tends to ∞. At t, the $\underset{\sim}{y}$ values associated with the units in $U^{(t)}$ will be denoted by $y_i^{(t)}$, $i = 1,\ldots,N^{(t)}$; similarly we define $\underset{\sim}{x}^{(t)}$, $\underset{\sim}{\beta}^{(t)}$, $\underset{\sim}{\sigma}_i^{(t)}$, $\underset{\sim}{p}^{(t)}$, $s^{(t)}$, etc.. However, these superscripts will be suppressed when the content is clear. Let $\mathcal{P}^{(t)}$ be the set of all survey designs $\underset{\sim}{p}^{(t)}$ with expected sample size $n^{(t)}$; i.e., $\mathcal{P}^{(t)} = \{\underset{\sim}{p}^{(t)}: \sum_s \#(s)\underset{\sim}{p}^{(t)}(s) = n^{(t)}\}$, where $\#(s)$ denotes the cardinality of s. We assume that $n^{(t)} \to \infty$ as $t \to \infty$. A linear estimate $\hat{Y}^{(t)}$ of $\bar{Y}^{(t)}$ is of the form $\sum_{i\in s} a_{i(s)}y_i + b_s$, where $a_{i(s)}$ and b_s are real numbers. In this paper we shall consider the linear estimates only. Now, given a sequence of sets $B^{(t)} \subset R^{N^{(t)}}$, $t = 1,2,\ldots$, a sequence of strategies $\{(\underset{\sim}{p}^{(t)},\hat{Y}^{(t)})\}_{t=1}^{\infty}$ is asymptotically consistent on $\{B^{(t)}\}_{t=1}^{\infty}$ if for any $\underset{\sim}{y}^{(t)} = (y_1,\ldots,y_{N^{(t)}})' \in B^{(t)}$, the mean squared error (MSE)

$$(1.2) \qquad E(\bar{Y}^{(t)} - \hat{Y}^{(t)})^2 \to 0, \text{ as } t \to \infty,$$

where "E" denotes the expectation with respect to the design $\underset{\sim}{p}^{(t)}$. Let $C(\{B^{(t)}\}_{t=1}^{\infty})$ be the class of all such strategies. A reasonable choice of $\{B^{(t)}\}_{t=1}^{\infty}$ may be, for example, L_∞ balls, $B^{(t)} = \{\underset{\sim}{y}: |y_i| \le c,\ i=1,\ldots,N^{(t)}\}$, where c is a constant. Of course, other norms (like the L_2-norm) should also be studied. In addition, the centers of these balls need not be the origin. In view of the possible linear association between $\underset{\sim}{x}_i$ and y_i, one may want to take

$$(1.3) \qquad B^{(t)} = B_R^{(t)} \equiv \{\underset{\sim}{y}: y_i^{(t)} = \underset{\sim}{x}_i^{(t)\prime}\underset{\sim}{\beta}^{(t)} + e_i^{(t)},\ |e_i^{(t)}| \le c_0\sigma_i^{(t)},\ \underset{\sim}{\beta}^{(t)} \in R^p,$$

$$1 \le i \le N^{(t)}\},$$

with c_0 being a fixed positive number. In a non-asymptotic setting, this type of $B^{(t)}$ was considered in Cheng and Li (1983) where an approximate minimaxity was established for the well-known Rao-Hartley-Cochran and Hansen-Hurwitz strategies for $p = 1$.

Our objective in this paper is to find a strategy in $\mathcal{C}(\{B^{(t)}\}_{t=1}^{\infty})$ that asymptotically minimizes the anticipated mean squared error, $\mathcal{E}E(\bar{Y}^{(t)} - \hat{Y}^{(t)})^2$ where $\mathcal{E}$ denotes the expectation with respect to the probability measure due to the superpopulation model (1.1). Such a strategy possesses two desirable properties: (i) it is efficient when (1.1) is proper, and (ii) it is robust against the model failure, at least asymptotically. The robustness property is of course due to the consistency of (1.2). One step further, we may want to require the convergent rate of (1.2) to be, say n^{-1}. Let $\mathcal{C}'(\{B^{(t)}\}_{t=1}^{\infty})$ be the class all such strategies. The problem now is to find a strategy in $\mathcal{C}'(\{B^{(t)}\}_{t=1}^{\infty})$ that minimizes $\mathcal{E}E(\bar{Y}^{(t)} - \hat{Y}^{(t)})^2$ asymptotically. Since $\mathcal{C}'(\{B^{(t)}\}_{t=1}^{\infty})$ is typically a much smaller class than $\mathcal{C}(\{B^{(t)}\}_{t=1}^{\infty})$, we may anticipate different solutions for these two different problems. However, it turns out that this is not the case. For the reasonably-given $\{B^{(t)}\}_{t=1}^{\infty}$, we can find a strategy that solves both problems.

In Section 2, we shall find an asymptotic lower bound of $\mathcal{E}E(\bar{Y} - \hat{Y})^2$ for any strategy in $\mathcal{C}(\{B^{(t)}\}_{t=1}^{\infty})$. This bound turns out to be the same as that given in Godambe (1955) and Godambe and Joshi (1965) for the class of design-unbiased strategies. Under rather limited contents, the same bound was also obtained by many authors (see Remark 5 of Section 2). The achievability of this lower bound is rigorously demonstrated in Section 3. Write $\pi_i = \sum_{i \in s} \underset{\sim}{P}(s)$. Basically, the inclusion probabilities π_i, $i \leq i \leq N$, of the desired sampling designs should be proportional to σ_i's (or approximately so), and the estimates considered are the regression estimates of the form (3.2) of Section 3. This type of estimate was studied by Särndal (1980) and Hajek (1981, page 193), while the same condition on the inclusion probabilities was obtained by Brewer, Godambe, Hajek, Iasaki and Fuller, Robinson and Tsui and Särndal under various cases. It is clear that certain conditions on the $\underset{\sim}{x}_i$'s are necessary to obtain consistent and asymptotically efficient strategies. These conditions are carefully derived to cover rather general cases. In addition, several sampling methods including the rejective sampling, Sampford-Durbin's method, successive sampling, and Rao-Hartley-Cochran's method are studied. The common practice of rejecting a bad or highly-unbalanced sample turns out necessary to achieve both efficiency and consistency.

2. The lower bound.

Suppressing the superscript (t), for a linear strategy $(\underset{\sim}{P},\hat{Y})$ we write $d_i = \frac{1}{N} - \sum_{s:i \in s} a_{i(s)}\underset{\sim}{P}(s)$. Note that if $(\underset{\sim}{P},\hat{Y})$ is unbiased then $d_i = 0$, $i \leq i \leq N$. The following Lemma will be useful.

Lemma 2.1. $\mathcal{E}E(\bar{Y} - \hat{Y})^2 \geq - N^{-2}\sum_{i=1}^{N} \sigma_i^2 + 2N^{-1}\sum_{i=1}^{N} \sigma_i^2 d_i + \sum_{i:\pi_i \neq 0} \sigma_i^2\pi_i^{-1}(\frac{1}{N} - d_i)^2.$

Proof. $\mathcal{E}E(\bar{Y} - \hat{Y})^2 = \sum_s \underset{\sim}{P}(s) \cdot \mathcal{E}(\bar{Y} - \sum_{i\in s} y_i - b_s)^2$

$$\geq \sum_s \underset{\sim}{P}(s) \cdot (N^{-2} \sum_{i=1}^{N} \sigma_i^2 + \sum_{i\in s} a_{i(s)}^2\sigma_i^2 - 2N^{-1} \sum_{i\in s} a_{i(s)}\sigma_i^2)$$

$$= N^{-2} \sum_{i=1}^{N} \sigma_i^2 - 2N^{-1} \sum_{i=1}^{N} \sigma_i^2 (\sum_{s:i\in s} a_{i(s)}\underset{\sim}{P}(s)) + \sum_{i=1}^{N} \sigma_i^2 (\sum_{s:i\in s} a_{i(s)}^2 \underset{\sim}{P}(s))$$

$$\geq -N^{-2} \sum_{i=1}^{N} \sigma_i^2 + 2N^{-1} \sum_{i=1}^{N} \sigma_i^2 d_i + \sum_{i:\pi_i\neq 0} \sigma_i^2\pi_i^{-1}(\frac{1}{N} - d_i)^2.$$

□

The equality in Lemma 2.1 is achieved when $a_i(s) = (\frac{1}{N} - d_i)/\pi_i$, which is then reduced to the Horvitz-Thompson estimate if $d_i = 0$. Now, fixing these d_i's, the right-hand term in this lemma will be minimized if $\pi_i = n|\frac{1}{N} - d_i|(\sum_{i=1}^{N} |\frac{1}{N} - d_i|)^{-1}$. To verify this, the standard method of Lagrangian multiplier can be used. A linear constraint $n = \sum_{i=1}^{N} \pi_i$ is involved here. Thus we obtain

Lemma 2.2. Given d_i, $1 \leq i \leq N$, we have

$$\mathcal{E}E(\bar{Y} - \hat{Y})^2 \geq -N^{-2} \sum_{i=1}^{N} \sigma_i^2 + 2N^{-2} \sum_{i=1}^{N} \sigma_i^2 d_i + n^{-1}(\sum_{i=1}^{N} \sigma_i|\frac{1}{N} - d_i|)^2.$$

Next, we shall show that the consistency of (1.2) implies that d_i's are negligible. The following type of $B^{(t)}$ will be considered first:

(2.1) $\qquad B^{(t)} = \{\underset{\sim}{y}: |y_i| \leq \delta_i^{(t)}, 1 \leq i \leq N^{(t)}\},$

where $\delta_i^{(t)}$'s are specified positive numbers.

Lemma 2.3. For any $(\underset{\sim}{P},\hat{Y}) \in \mathcal{C}(\{B^{(t)}\}_{t=1}^{\infty})$,

(2.2) $\qquad \lim_{t\to\infty} \sum_{i=1}^{N} |d_i^{(t)}|\delta_i^{(t)} \to 0.$

Proof. Clearly, (1.2) implies $\bar{Y} - E\hat{Y} \to 0$. First, letting $\underset{\sim}{y} = 0$, we get

$\sum_s b_{(s)} \underset{\sim}{P}(s) \to 0$. Now, taking $y_i = \mathrm{sgn}(d_i)\delta_i$, where $\mathrm{sgn}(d_i) = 1$ if $d_i \geq 0$ and $\mathrm{sgn}(d_i) = -1$ if $d_i < 0$, Lemma 2.3 follows easily. □

Now, we establish the following main theorem of this section.

Theorem 2.1. Assume that (2.1) holds and that

$$(2.3) \qquad \varliminf_{t\to\infty} N^{-1} \sum_{i=1}^{N} \sigma_i^{(t)} > 0, \text{ and } \varlimsup_{t\to\infty} N^{-1} \sum_{i=1}^{N} \sigma_i^{(t)} < \infty,$$

and

$$(2.4) \qquad \sup\{\sigma_i^{(t)}/\delta_i^{(t)}:\ 1 \leq i \leq N^{(t)},\ t = 1,2,\ldots\} < \infty.$$

Then for any $\{(\underset{\sim}{P}^{(t)}, \hat{Y}^{(t)}\}_{t=1}^{\infty} \in \mathcal{C}(\{B^{(t)}\}_{t=1}^{\infty})$, we have

$$\varliminf_{t\to\infty} \mathcal{E}E(\bar{Y} - \hat{Y})^2 / \{-N^{-2} \sum_{i=1}^{N} \sigma_i^2 + n^{-1}(N^{-1} \sum_{i=1}^{N} \sigma_i)^2\} \geq 1.$$

Proof. In view of (2.3) and Lemma 2.2, it suffices to show that

$$(2.5) \qquad \sum_{i=1}^{N} \sigma_i |d_i| \to 0$$

and

$$(2.6) \qquad N^{-1} \sum_{i=1}^{N} \sigma_i^2 d_i \to 0.$$

(2.5) follows from Lemma 2.3 and (2.4). To show (2.6), we simply observe that $|N^{-1} \sum_{i=1}^{N} \sigma_i^2 d_i| \leq (N^{-1} \sum_{i=1}^{N} \sigma_i)(\sum_{i=1}^{N} \sigma_i |d_i|)$. □

Remark 1. Theorem 2.1 holds for the $\{B_R^{(t)}\}_{t=1}^{\infty}$ defined by (1.3), because $B_R^{(t)} \supset B^{(t)}$. In addition, this theorem is translation invariant. More precisely, for any vector $\underset{\sim}{v}^{(t)} \subset R^{N^{(t)}}$, we have the same lower bound if $\{\underset{\sim}{P}^{(t)}, \hat{Y}^{(t)}\}_{t=1}^{\infty} \in \mathcal{C}(\{\underset{\sim}{v}^{(t)} + B^{(t)}\}_{t=1}^{\infty})$.

Remark 2. Suppose we take $B^{(t)} = \{\underset{\sim}{y}: \frac{1}{N^{(t)}} \sum_{i=1}^{N} |y_i/\delta_i^{(t)}|^k \leq 1\}$ with $k \geq 0$.

Then this theorem holds with (2.4) replaced by the following weaker condition:

$$(2.4') \qquad \sup\{\frac{1}{N^{(t)}} \sum_{i=1}^{N^{(t)}} (\sigma_i^{(t)}/\delta_i^{(t)})^k: \ t = 1,2,\ldots\} < \infty.$$

Remark 3. Suppose we consider $\mathcal{C}'(\{B^{(t)}\}_{t=1}^{\infty})$ (see Section 1 for the definition). Then the result of Theorem 2.1 holds with (2.4) being replaced by the weaker

$$(2.4'') \qquad \sup\{(\sqrt{n}\ \delta_i^{(t)})^{-1}\sigma_i^{(t)}: \ 1 \le i \le N^{(t)},\ t = 1,2,\ldots\} < \infty.$$

Remark 4. With all d_i's being 0 (i.e., the unbiasedness condition), Lemma 2.1 was obtained by Godambe (1955). Godambe and Joshi (1965) further obtained the same bound for any estimate (measurable with respect to the superpopulation model). Lemma 2.1 can be generalized similarly. For any estimate $e(s;y_i,i\in s)$, we decompose it into the linear and the non-linear parts by writing $e(s;y_i,i\in s) = \sum_{i\in s} a_{i(s)}y_i + f(s;y_i,i\in s)$ with $a_{i(s)} = \sigma_i^{-2}$. covariance of $e(s;y_i,i\in s)$ and y_i, where the covariance is taken with respect to the superpopulation model. Since $f(s;y_i,i\in s)$ is now uncorrelated with any y_i, $i \le j \le N$, our Lemma 2.1 follows easily, with d_i being defined as before. Godambe and Joshi's result now follows from the following interesting fact:

> "The design-unbiasedness of $e(s;y_i,i\in s)$ implies the design unbiasedness of the linear component $\sum_{i\in s} a_i(s)y_i$."

Proof. The definition of design-unbiasedness implies that for any $\underset{\sim}{y}$,

$$\frac{1}{N}\sum_{i=1}^{N} y_i - \sum_s(\sum_{i\in s} a_i(s)y_i)\underset{\sim}{P}(s) = \sum_s f(s;y_i,i\in s)\underset{\sim}{P}(s)\ .$$

Since each $f(s;y_i,i\in s)$ is uncorrelated with y_j, $1 \le j \le N$, a contradiction is now derived unless the coefficients of y_i's on the left hand side are all 0.

□

Remark 5. By restricting the consideration to be in certain subclasses of consistent strategies, the asymptotic lower bound in Theorem 2.1 has been obtained by Brewer (1979), Sarndal (1980), Robinson and Tsui (1981), and Isaki and Fuller (1982). Our Theorem strengthens and unifies these results by removing many unnecessary restrictions on the strategies; for example, there is no need to require $\{\pi_i;\ 1 \le i \le N\}$ to be bounded away from 0 (thus we can handle the realistic case that the sampling fraction $n^{(t)}/N^{(t)} \to 0$); the estimates need not be of any particular form. Also, it is interesting to see that the structure of the $\underset{\sim}{x}_i$'s is irrelevant in deriving the asymptotic lower bound. In contrast, all the previous literatures on this subject were tied with certain properties of the $\underset{\sim}{x}$ sequence.

Remark 6. We may have another interpretation about Theorem 2.1 by considering the following p-regressors version of the model studied in Godambe (1982):

$$(2.7)\qquad y_i^{(t)} = \underset{\sim}{x}_i^{(t)\prime}\underset{\sim}{\beta}^{(t)} + e_i^{(t)} + \varepsilon_i^{(t)},\ i = 1,\ldots,N^{(t)},$$

with random variables $\varepsilon_i^{(t)}$ defined as in (1.1) and $(e_1^{(t)},\ldots,e_N^{(t)})' \in B^{(t)}$ of (2.1). Let $\mathcal{C}''(\{B^{(t)}\}_{t=1}^{\infty})$ be the collection of all strategies with linear estimates such that the anticipated MSE (under (2.7)) tends to 0. Since $\mathcal{C}''(\{B^{(t)}\}_{t=1}^{\infty}) \subset \mathcal{C}'(\{B^{(t)}\}_{t=1}^{\infty})$ Theorem 2.1 holds for this class of consistent estimates.

3. Achieving the lower bound.

The strategies to be considered here (to some extent) possess the following two properties:

(3.1) The inclusion probabilities π_i's are proportion to σ_i's, i.e., $\pi_i = n\sigma_i(\sum_{i=1}^{N} \sigma_i)^{-1}$.

(3.2) The estimate $\hat{Y}$ is of the form

$$\hat{Y} = \bar{\underset{\sim}{x}}'\hat{\underset{\sim}{\beta}}_s + \frac{1}{N}\sum_{i\in s}(y_i - \underset{\sim}{x}_i'\hat{\underset{\sim}{\beta}}_s)/\pi_i ,$$

where $\bar{\underset{\sim}{x}} = \frac{1}{N}\sum_{i=1}^{N}\underset{\sim}{x}_i$ and $\hat{\underset{\sim}{\beta}}_s$ is a linear model-unbiased estimate of $\underset{\sim}{\beta}$ based on the sample s.

Hereafter, we assume that

$$(3.3) \qquad n \cdot \sup\{\sigma_i : 1 \le i \le N\} \le \sum_{i=1}^{N} \sigma_i ,$$

to assure π_i's to be no greater than 1.

Regression estimate of (3.2), studied in Särndal (1980) and Hajek (1981), are generalizations of difference estimates (Cassel, Särndal and Wretman 1976; 1977). Of course, these estimates originate from the Horvitz-Thompson estimate $\frac{1}{N}\sum_{i\in s} y_i/\pi_i$. However, estimates of (3.2) are no longer design-unbiased. Some conditions are needed to ensure that the bias is small and consistency of (1.2) holds.

Write the mean squared error as

$$E(\bar{Y} - \hat{Y})^2 = E\{(\bar{y} - \frac{1}{N}\sum_{i\in s} y_i/\pi_i) + (\bar{\underset{\sim}{x}} - \frac{1}{N}\sum_{i\in s}\underset{\sim}{x}_i/\pi_i)'\hat{\underset{\sim}{\beta}}_s\}^2 .$$

Our strategies will be in $\mathcal{C}(\{B^{(t)}\}_{t=1}^{\infty})$, if for any $\underset{\sim}{y}\in B^{(t)}$,

$$E(\bar{Y} - \frac{1}{N}\sum_{i\in s} y_i/\pi_i)^2 \to 0$$

and

$$E((\bar{\underset{\sim}{x}} - \frac{1}{N}\sum_{i\in s}\underset{\sim}{x}_i/\pi_i)'\hat{\underset{\sim}{\beta}}_s)^2 \to 0.$$

In fact, for many strategies, under mild conditions we can show that

$$(3.4) \qquad E(\bar{Y} - \frac{1}{N}\sum_{i\in s} y_i/\pi_i)^2 = O(n^{-1}), \text{ for } \underset{\sim}{y}\in B^{(t)},$$

$$(3.5) \qquad E(\bar{x}_j - \frac{1}{N}\sum_{i\in s} x_{ij}/\pi_i)^2 = O(n^{-1}), \ 1 \le j \le p,$$

and

(3.6) $\quad E||\hat{\beta}_s||^2 = o(n)$, for $\underset{\sim}{y} \in B^{(t)}$,

where $\bar{x}_j = \frac{1}{N}\sum_{i=1}^{N} x_{ij}$ and $||\cdot||$ is the Euclidean norm. In addition, our strategies will be root-n consistent (i.e., in $\mathcal{C}'(\{B^{(t)}\}_{t=1}^{\infty})$) if instead of (3.6) we have

(3.7) $\quad E||\hat{\beta}_s||^2 = O(1)$, for $\underset{\sim}{y} \in B^{(t)}$.

Next, consider the anticipated mean squared error

$$\mathcal{E}E(\bar{Y} - \hat{Y})^2 = \mathcal{E}E\{(\bar{\varepsilon} - \frac{1}{N}\sum_{i\in s}\varepsilon_i/\pi_i) + (\bar{\underset{\sim}{x}} - \frac{1}{N}\sum_{i\in s}\underset{\sim}{x}_i/\pi_i)'(\underset{\sim}{\beta} - \hat{\underset{\sim}{\beta}}_s)\}^2,$$

where $\bar{\varepsilon} = \frac{1}{N}\sum_{i=1}^{N}\varepsilon_i$. By straightforward computation, we have

$$\mathcal{E}E(\bar{\varepsilon} - \frac{1}{N}\sum_{i\in s}\varepsilon_i/\pi_i)^2 = -\frac{1}{N^2}\sum_{i=1}^{N}\sigma_i^2 + \frac{1}{n}(\frac{1}{N}\sum_{i=1}^{N}\sigma_i)^2.$$

Thus the anticipated mean squared error will be as desired in Theorem 2.1 if $\mathcal{E}E\{(\bar{\underset{\sim}{x}} - \frac{1}{N}\sum_{i\in s}\underset{\sim}{x}_i/\pi_i)'(\underset{\sim}{\beta} - \hat{\underset{\sim}{\beta}}_s)\}^2 = o(n^{-1})$. To establish this, it suffices to show that (3.5) holds and

(3.8) $\quad E\mathcal{E}||\underset{\sim}{\beta} - \hat{\underset{\sim}{\beta}}_s||^2 \to 0$.

Now, it remains to derive some conditions to assure (3.4) $\sim$ (3.8). It is clear that here the second order inclusion probabilities π_{ij}'s will play an important role. By the well-known Yates and Grundy (1953) formula, the desired conditions that involve π_{ij}'s (and, of course, $\underset{\sim}{x}_i$'s, σ_i's etc.) may be obtained. However, they are bound to be either complicate (to gain generality) or restrictive (to gain simplicity). Thus this approach will not be attempted here. Instead, we shall consider some well-known sampling methods (with some mild modification to be specified later) and set up the desired conditions accordingly. These conditions turn out relatively simple and without much loss of generality. The first one to be considered is the rejective sampling. Then some other methods, Sampford-Durbin modification of rejective sampling, successive sampling, and the Rao-Hartley-Cochran's method, will also be discussed. The description of these methods and their properties can be found in Hajek (1981) and its references; the details will be omitted.

Consider the rejective sampling design $\underset{\sim}{P}$ with probabilities of inclusion specified by (3.1). Recall the following formula for the mean squared error of the Horvitz-Thompson estimate from (6.3) of Hajek (1964, page 1512), or from (17.21) of Hajek (1981, page 167) (the notations are changed):

$$(3.9) \qquad E(\bar{Y} - \frac{1}{N}\sum_{i\in s} y_i/\pi_i)^2 = \frac{1}{N^2}(1 + o(1))\sum_{i=1}^{N} (y_i - \theta\pi_i)^2[\frac{1}{\pi_i} - 1]$$

where $\theta = \sum_{i=1}^{N} y_i(1 - \pi_i)/\sum_{i=1}^{N}\pi_i(1 - \pi_i)$, and $o(1) \to 0$ if $\sum_{i=1}^{N} \pi_i(1 - \pi_i) \to \infty$.

To insure this, we simply assume that the sampling fraction is small; precisely,

$$(3.10) \qquad \frac{n}{N}(\frac{1}{N}\sum_{i=1}^{N}\sigma_i^2)(\frac{1}{N}\sum_{i=1}^{N}\sigma_i)^{-2} < 1$$

will be assumed. (3.9) is however complicated for further development. In order to obtain simpler conditions, we shall use the following cruder formula to replace (3.9)

$$(3.11) \qquad E(\bar{Y} - \frac{1}{N}\sum_{i\in s} y_i/\pi_i)^2 \leq (1 + o(1))n^{-1}(\frac{1}{N}\Sigma\sigma_i)(\frac{1}{N}\sum_{i=1}^{N} y_i^2/\sigma_i).$$

The above inequality is obtained from (3.9) by deleting $\theta\pi_i$ and changing $[\frac{1}{\pi_i} - 1]$ to $\frac{1}{\pi_i}$.

Recall the definition of $\{B^{(t)}\}$ from (2.1). In view of (3.11), (3.4) will hold if (2.3) and

(3.12) $\quad \sup\{\delta_i^{(t)}/\sigma_i^{(t)}:\ 1 \leq i \leq N^{(t)},\ t = 1,2,\ldots\} < \infty.$

Similarly, (3.5) holds under (2.3) and

(3.13) $\quad \frac{1}{N}\sum_{i=1}^{N} x_{ij}^2/\sigma_i = O(1),\ 1 \leq j \leq P.$

We now turn to (3.6) $\sim$ (3.8). We shall only consider the case that $\hat{\beta}_s$ is the best linear estimate under the superpopulation model; i.e.,

(3.14) $\quad \hat{\beta}_s = (X_s'\sigma_s^{-2}X_s)^{-1}X_s'\sigma_s^{-2}y_s,$

where X_s is the nxp design matrix, σ_s^{-2} is the diagonal matrix with diagonal elements being σ_i^{-2}, $i\in s$, and y_s is the vector of observations of y_i, $i\in s$.

We shall establish (3.8) first. Observe that

$$\mathcal{E}||\beta - \hat{\beta}_s||^2 = \text{trace } (X_s'\sigma_s^{-2}X_s)^{-1}.$$

Denote the minimum eigenvalue of a symmetric matrix A by $\lambda(A)$. $\hat{\beta}_s$ will be inconsistent unless the (adjusted) information matrix $X_s'\sigma_s^{-2}X_s$ tends to ∞ in the sense that $\lambda(X_s'\sigma_s^{-2}X_s) \to \infty$. Thus it is clear that (3.8) follows from

(3.15) $\quad \lambda(X_s'\sigma_s^{-2}X_s) \to \infty$ with (design-) probability 1,

and

(3.16) $\quad$ there exists a positive constant c such that $\lambda(X_s'\sigma_s^{-2}X_s) \geq c$ for any s such that $P(s) > 0$.

When $p = 1$, (3.15) will hold if the expected value of the random variable $X_s'\sigma_s^{-2}X_s$ tends to ∞ and its coefficient of variation tends to 0. Generalizing this idea to $p \geq 1$, it is clear that (3.15) follows from

(3.17) $\lambda\{E(X_s'\underset{\sim}{\sigma}_s^{-2}X_s)\} \to \infty$

and

(3.18) $[\mathrm{Var}\{[X_s'\underset{\sim}{\sigma}_s^{-2}X_s]_{jj'}\}]^{\frac{1}{2}}/\lambda\{E(X_s'\underset{\sim}{\sigma}_s^{-2}X_s)\} \to 0$ for $1 \leq j,j' \leq p$,

where $[A]_{jj'}$ denoted the j, j'th element of the matrix A and Var denotes the variance with respect to the design measure.

Write $\lambda_N = \lambda\{\sum_{i=1}^{N}\sigma_i^{-1}\underset{\sim}{x}_i\underset{\sim}{x}_i'\}$. Under (2.3), (3.17) follows from

(3.19) $n\lambda_N/N \to \infty$.

Since $\mathrm{Var}\{[X_s'\underset{\sim}{\sigma}_s^{-2}X_s]_{jj'}\} = \mathrm{Var}(\sum_{i\in s} x_{ij}x_{ij'}\sigma_i^{-2})$, we may use (3.11) with $y_i = x_{ij}x_{ij'}\sigma_i^{-2}\pi_i$ to obtain the following sufficient condition for (3.18):

(3.20) $n^{-1}\lambda_N^{-2}N(\sum_{i=1}^{N} x_{ij}^2x_{ij'}^2\sigma_i^{-3}) \to 0$, for $1 \leq j$, $j' \leq p$.

By the Cauchy-Schwartz inequality, (3.20) holds if

(3.21) $n^{-1}\lambda_N^{-2}N(\sum_{i=1}^{N} x_{ij}^4\sigma_i^{-3}) \to 0$, for $1 \leq j \leq p$.

In summary, (3.19) and (3.21) imply (3.15). To obtain (3.16), we shall slightly modify the rejective sampling design $\underset{\sim}{P}$. Our new sampling design $\underset{\sim}{P}'$ is derived from the rejective sampling $\underset{\sim}{P}$ by conditioning on the set $\Lambda_1 = \{\underset{\sim}{s}: \lambda(X_s'\underset{\sim}{\sigma}_s^{-2}X_s) \geq c\}$ for a fixed positive constant c. Now, (3.15) implies that $\underset{\sim}{P}\{s\in\Lambda_1\}\to 1$. Thus asymptotically all formulae involving $\underset{\sim}{P}$ still hold when $\underset{\sim}{P}$ is replaced by $\underset{\sim}{P}'$. For example, (3.11) holds because $\underset{\sim}{P}(\Lambda_1) \to 1$ and

$$E_{\underset{\sim}{P}'}(\bar{Y} - \frac{1}{N}\sum_{i\in s} y_i/\pi_i)^2 = \underset{\sim}{P}(\Lambda_1)^{-1}\sum_{s\in\Lambda_1}(\bar{Y} - \frac{1}{N}\sum_{i\in s} y_i/\pi_i)^2\underset{\sim}{P}(s)$$

$$\leq \underset{\sim}{P}(\Lambda_1)^{-1}E_{\underset{\sim}{P}}(\bar{Y} - \frac{1}{N}\sum_{i\in s} y_i/\pi_i)^2.$$

Of course, (3.1) is only asymptotically valid; but we still have

$$\mathcal{E}\, E_{\underset{\sim}{P}'}(\bar{\varepsilon} - \frac{1}{N}\sum_{i\in s}\varepsilon_i/\pi_i)^2 \le (1 + o(1))\{-\frac{1}{N^2}\sum_{i=1}^{N}\sigma_i^2 + \frac{1}{n}(\frac{1}{N}\sum_{i=1}^{N}\sigma_i)^2\}.$$

Therefore, we have established (3.4) , (3.5) and (3.8) for the sampling design $\underset{\sim}{P}'$; it remains to derive conditions to verify (3.7).

For this purpose, we consider $\Lambda_2 = \{s\colon \lambda(X_s'\underset{\sim}{\sigma}_s^{-2}X_s)n^{-1}\lambda_N^{-1}N \ge c\}$ with $0 < c < 1$ being fixed. Let $\underset{\sim}{P}''$ be the conditional probability measure of $\underset{\sim}{P}$ on Λ_2. Then inview of (3.17) $\sim$ (3.19), $\underset{\sim}{P}(\Lambda_2) \to 1$ and all the results obtained for $\underset{\sim}{P}$ and $\underset{\sim}{P}'$ also hold for $\underset{\sim}{P}''$. In addition, we have

Lemma 3.1. Assume that

(3.22) $\qquad \underline{\lim}\ \lambda_N N^{-1} > 0.$

Then under (2.3), (3.12), (3.13) and (3.21), (3.7) holds for $\underset{\sim}{P}''$.

The proof of this lemma will be given in the Appendix. We now summarize our results by the following theorems.

Theorem 3.1. Under (2.3), (3.3), (3.10), (3.13), (3.19), and (3.21), the anticipated MSE of the modified version of the rejective sampling design, $\underset{\sim}{P}'$ (or $\underset{\sim}{P}''$), together with the estimate specified by (3.2) and (3.14), achieves the asymptotic lower bound given by Theorem 2.1.

Theorem 3.2. Under (2.3), (3.3), (3.10), (3.12), (3.13), (3.22), and

(3.23) $\qquad n^{-1}N^{-1}\sum_{i=1}^{N} x_{ij}^4\sigma_i^{-3} \to 0,$

the sampling design $\underset{\sim}{P}''$, together with the estimate specified by (3.2) and (3.14) belongs to the class $\mathcal{C}'\{B^{(t)}\}_{t=1}^{\infty}$ with $B^{(t)}$ being defined by (2.1).

Note that (3.21) is implied by (3.22) and (3.23).

Remark 7. Theorem 3.2 also holds for the case that $B^{(t)} = B_R^{(t)}$ (see (1.3)) without the condition of (3.12). This is because our estimate is representative (Hajek 1981, page 157) in the sense that when $\underset{\sim}{y} = (x_{1j}, x_{2j}, \ldots, x_{Nj})'$, for any $1 \le j \le p$, we have $\bar{Y} = \hat{Y}$. In addition, the anticipated MSE of the proposed strategy under (2.7) converges to 0 at rate n^{-1}.

Remark 8. When applying any probability sampling method, if the sample drawn is highly-unbalanced in the sense that the information matrix $X_s'\underset{\sim}{\sigma}_s^{-2}X_s$ is small or nearly degenerated in certain directions, a common practice is to reject this sample and take another one by the same sampling method. Our $\underset{\sim}{P}'$ and $\underset{\sim}{P}''$ are obtained exactly in this way. Without doing so, it seems unlikely that (3.7) and (3.8) will hold.

Remark 9. In view of Hajek (1981, page 167 ∿ 168), (3.11) also holds for the successive sampling and the Samford-Durbin Sampling. Therefore Theorems 3.1 and 3.2 are also valid for these two methods.

Remark 10. Consider the Rao-Hartley-Cochran's sampling (Cochran 1977; Rao, Hartley and Cochran 1962) with the selection probabilities proportional to σ_i. We may replace (3.2) by using

$$\hat{Y} = \bar{\underset{\sim}{X}}'\hat{\underset{\sim}{\beta}}_s + \frac{1}{N}\hat{Y}_{RHC}$$

where $\hat{Y}_{RHC}$ is the Rao-Hartley-Cochran's estimate of the population total when the population values are $y_i - \underset{\sim}{x}_i'\hat{\underset{\sim}{\beta}}_s$; i.e., in terms of Cochran 1977, page 266,

$$\hat{Y}_{RHC} = \sum_{j=1}^{n} \Big(\sum_{i \in \text{group } j} \sigma_i\Big)(y_j - \underset{\sim}{x}_j'\hat{\underset{\sim}{\beta}}_s)/\sigma_j .$$

Then in view of (9A.66) and (9A.67) of Cochran (1977), (3.11) is also valid when the Horvitz-Thompson estimate is replaced by the Rao-Hartley-Cochran's estimate. Therefore, Theorem 3.2 holds. In addition, when $nN^{-1} \to 0$, Theorem 3.1 is also valid.

Remark 11. Isaki and Fuller (1982) obtained an interesting connection between (3.2) and the best linear unbiased estimate. By assuming conditions similar to (3.5) and (3.8) and others, (e.g., the inclusion probabilities are bounded below), they obtained strategies achieving the asymptotic lower bound of Theorem 2.1. In contrast our conditions involve only the structure of the problem instead of the properties of the strategies studied. Also our results do not require σ_i and σ_i^2 to be included in the coordinates of the $\underset{\sim}{x}_i$. Finally, it is not difficult to extend our results to cover the case that the independence of ε_i's is replaced by the assumption that $\mathcal{E}\, \varepsilon_i \varepsilon_j = \rho\sigma_i\sigma_j$, $1 \leq i \neq j \leq N$, with $-(N-1)^{-1} \leq \rho \leq 1$.

Appendix

Proof of Lemma 3.1. By (3.14), $||\hat{\underset{\sim}{\beta}}_s||^2 \leq [\lambda(X_s' \underset{\sim}{\sigma}_s^{-2} X_s)]^{-2} ||X_s' \underset{\sim}{\sigma}_s^{-2} \underset{\sim}{y}_s||^2$. Computing $E||X_s' \underset{\sim}{\sigma}_s^{-2} \underset{\sim}{y}_s||^2$ by using (3.11), we have

$$E||X_s'\sigma_s^{-2}y_s||^2 \leq \sum_{j=1}^{P} \{n^2(\sum_{i=1}^{N} x_{ij}y_i\sigma_i^{-1})^2(\sum_{i=1}^{N}\sigma_i)^{-2} + (1+o(1))n(\sum_{i=1}^{N}\sigma_i)^{-1}$$

$$(\sum_{i=1}^{N} x_{ij}^2 y_i^2 \sigma_i^{-3})\} \leq \sum_{j=1}^{P} \{n^2(\sum_{i=1}^{N} x_{ij}^2\sigma_i^{-1})(\sum_{i=1}^{N}\delta_i^2\sigma_i^{-1})(\sum_{i=1}^{N}\sigma_i)^{-2} + (1+o(1))n$$

$$(\sum_{i=1}^{N}\sigma_i)^{-1}(\sum_{i=1}^{N} x_{ij}^2\delta_i^2\sigma_i^{-3})\}.$$

It follows that

$$[\lambda(EX_s'\underset{\sim}{\sigma}_s^{-2}X_s)]^{-2}E||X_s'\underset{\sim}{\sigma}_s^{-2}\underset{\sim}{y}_s||^2$$

$$= n^{-2}(\sum_{i=1}^{N}\sigma_i)^2\lambda_N^{-2}E||X_s'\underset{\sim}{\sigma}_s^{-2}\underset{\sim}{y}_s||^2$$

$$\leq \sum_{j=1}^{P} \{(\sum_{i=1}^{N} x_{ij}^2\sigma_i^{-1})(\sum_{i=1}^{N}\delta_i^2\sigma_i^{-1})\lambda_N^{-2} + (1 + o(1))n^{-1}(\sum_{i=1}^{N}\sigma_i)(\sum_{i=1}^{N} x_{ij}^2\delta_i^2\sigma_i^3)\lambda_N^{-2}\}$$

$$= O(1) \quad \text{(by (3.22), (3.13), (3.12), and (2.3))}.$$

Now by the dominating convergence Theorem, the proof will be complete if $\lambda(X_s'\underset{\sim}{\sigma}_s^{-2}X_s)/\lambda(EX_s'\underset{\sim}{\sigma}_s^{-2}X_s) \to 1$ in probability with respect to $\underset{\sim}{P}''$. This condition is guaranteed by (3.18) which follows from (3.21).

References

Brewer, K.R.W. (1979). A class of robust sampling designs for large scale surveys. J. Amer. Statist. Assoc. 74, 911-915.

Cassel, C., Särndal, C., and Wretman, J.H. (1976). Some results on generalized difference estimation and generalized regression estimation for finite populations. Biometrika, 63, 615-620.

Cassel, C., Särndal, C., and Wretman, J.H. (1977). Foundations of Inference in Survey Sampling, John Wiley, New York.

Cheng, C.S., and Li, K.C. (1983). A minimax approach to sample surveys. To appear in Ann. Statist.

Cochran, W.G. (1977). Sampling Techniques, third edition. John Wiley & Sons, New York.

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.

Hajek, J. (1964). Asymptotic theory of rejective sampling with varying probabilities from a finite population. Ann. Math. Statist., 35, 1491-1523.

Hajek, J. (1971). Contribution to discussion of paper by D. Basu. In Foundations of Statistical Inference. Eds. V.P. Godambe and D.A. Sprott, p. 236. Toronto: Holt, Rinehart and Winston.

Hajek, J. (1981). Sampling from a finite population. Marcel Dekker; New York and Basel.

Isaki, C.T. and Fuller, W.A. (1982). Survey design under the regression superpopulation model. J. Amer. Statist. Assoc. 77, 89-96.

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. B, 24, 482-491.

Robinson, P.M., and Tsui, K.W. (1981). Optimal asymptotically design-unbiased estimation of a population total. Technical report No. 634. Department of Statistics, University of Wisconsin, Madison.

Särndal, C.E. (1981). On π-inverse weighting versus BLU weighting in probability sampling. Biometrika 67, 639-650.

Yates, F., and Grundy, P.M. (1953). Selection without replacement from within strata with probability proportionate to size. J. Roy. Stat. Soc., B, 15, 236-261.