CHAPTER 6

Inference Under Superpopulation Model

6.1 INTRODUCTION

In the previous chapters, we considered a fixed population setup, where the vector of variable under study $\mathbf{y} = (y_1,\ldots, y_N)$ is a fixed point in the N-dimensional Euclidean space R^N, and y_i is the unknown value of the study variable y associated with the unit $i(1,\ldots, N)$. Our problem was to find point estimates and confidence intervals of certain real-valued parametric functions of $\mathbf{y}$. Unlike the previous sections, in this section, we assume that the population vector $\mathbf{y}$ is a realization of a random vector $\mathbf{Y} = (Y_1,\ldots, Y_N)$ and its distribution is denoted by ξ. Here the population mean $\overline{Y} = \sum_{i=1}^{N} Y_i/N$ is a random variable. For example, if we consider a fixed set of households in a certain locality as a finite population, and the annual income of households as a study variable, then the ith household income Y_i is a random variable, which changes with time. The change of values of $\mathbf{Y}$ overtime may be characterized by a certain probability distribution ξ. The distribution ξ may be indexed by some parameter θ, which is generally unknown, but it may belong to a certain known parameter space Ω_θ. Such a distribution ξ is called a superpopulation model or simply a model. In most situations, the distribution ξ is related to an auxiliary variable $\mathbf{x} = (x_1,\ldots, x_N)$, whose values are fixed, assumed to be known and positive. We will denote E_ξ, V_ξ and C_ξ as the operators for expectation, variance, and covariance with respect to the model ξ. Our objective is to estimate some parametric function $\phi(\mathbf{y})$, where $\mathbf{y}$ is a particular realization of the random vector $\mathbf{Y} = (Y_1,\ldots, Y_i,\ldots, Y_N)$. For notational simplicity, we write $\mathbf{y}$ as a random vector $\mathbf{Y}$ and y_i for Y_i; $i = 1,\ldots, N$.

In this section we will discuss the concepts of design-based, model-based (dependent), and model design—based inference. Optimal sampling strategies (a combination of a sampling design and an estimator) in the presence of various superpopulation models will be established. The optimal strategy based on a model-based inference is highly dependent on the choice of the

Survey Sampling Theory and Applications
ISBN 978-0-12-811848-1
http://dx.doi.org/10.1016/B978-0-12-811848-1.00006-6

model and the strategy becomes inefficient if the chosen model is inappropriate. The concept of robustness of sampling designs and estimators has been discussed to guard against the use of inappropriate models. The concepts of Bayesian inferences will be introduced, and it is shown that the sampling design does not play any role in making inference about the parametric function $\phi(\mathbf{y})$. Finally, various sampling strategies, which are commonly used in practice, are compared under different superpopulation models.

6.2 DEFINITIONS

6.2.1 Sampling Strategy

A sampling strategy is a combination of sampling design p and an estimator t based on a sample selected using the design p. Let $\mathcal{P}$ be a class of sampling designs and C be a class of estimators. Then a sampling strategy is denoted by $h = (p, t)$ with $p \in \mathcal{P}, t \in C$.

6.2.2 Noninformative Sampling Design

A sampling design is said to be noninformative if the selection of a sample does not depend on the study variable y_i's, i.e., the sampling design is nonsequential. For a noninformative sampling design, the operators E_ξ and E_p are commutative in the sense $E_\xi E_p(\cdot) = E_p E_\xi(\cdot)$. In our present discussion, we will consider the noninformative sampling design only.

6.2.3 Design-Unbiased (or *p*-Unbiased) Estimator

An estimator t is said to be design-unbiased or p-unbiased or simply unbiased for the total Y if and only if $E_p(t) = Y$ for all $\mathbf{y} \in R^N$. The class of p-unbiased estimators (including linear and nonlinear) will be denoted by C_{pu}.

6.2.4 Model-Unbiased (or ξ-Unbiased) Estimator

An estimator t is said to be a model unbiased or ξ-unbiased estimator for Y if and only if $E_\xi(t) = E_\xi(Y) \ \ \forall \theta \in \Omega_\theta$. The class of ξ-unbiased estimators will be defined by $C_{\xi u}$.

6.2.5 Model Design—Unbiased (or $p\xi$-Unbiased) Estimator

An estimator is said to be a model design—unbiased or $p\xi$-unbiased estimator for the population total Y, if and only if $E_\xi E_p(t) = E_p E_\xi(t) = E_\xi(Y) \ \ \forall \theta \in \Omega_\theta$. The class of $p\xi$-unbiased estimators will be denoted by $C_{p\xi}$. If an estimator is p-unbiased or ξ-unbiased, then it is necessarily

$p\xi$-unbiased, but the converse is not true. So, the class $C_{p\xi}$ contains both the classes C_{pu} and $C_{\xi u}$. An estimator can be p-unbiased but may not be ξ-unbiased and vice versa.

Example 6.2.1

The Horvitz–Thompson estimator (HTE), $\widehat{Y}_{ht} = \sum_{i\in s} \frac{y_i}{\pi_i}$, is p-unbiased for Y but not ξ-unbiased for Y in general. But if π_i is proportional to $E_\xi(y_i)$ and p is a fixed effective sample of size n design (FESD(n)), then $\widehat{Y}_{ht}$ becomes ξ-unbiased for Y. The sample mean based on an simple random sampling without replacement (SRSWOR) method is p-unbiased but not ξ-unbiased for the population mean $\overline{Y}$ in general. But the HTE, based on any sampling design, is $p\xi$-unbiased for the total Y. The ratio estimator $\widehat{Y}_R = \frac{\overline{y}_s}{\overline{x}_s} X$ with $\overline{y}_s = \sum_{i\in s} y_i/n$, $\overline{x}_s = \sum_{i\in s} x_i/n$ and $X = \sum_{i\in U} x_i$ is ξ-unbiased for Y, when $E_\xi(y_i) \propto x_i$. The ratio estimator is, however, not p-unbiased in general.

6.2.6 Design-Based Inference

In a design-based inference, which is also called the probability sampling approach, we assume that the population vector $\mathbf{y}$ is fixed. From the population, a sample s of n units is selected by following a sampling design p, a man-made randomization procedure. The information concerning the study variable y is obtained from units selected in the sample s and a link between the observed values of y_i's, $i \in s$ and unobserved values y_i's, $i \notin s$ is established through the sampling design. Here, expectation of an estimator is the average value over all possible hypothetical samples.

6.2.7 Model-Based Inference

In a model-based inference, a sample is selected by using a sampling design p and the link between the observed y_i's, $i \in s$, and unobserved y_i's and $i \notin s$ is obtained through a superpopulation model ξ. The assumed model distribution ξ provides valid inferences with respect to the particular selected sample, irrespective of the chosen sampling design.

6.2.8 Model-Assisted Inference

The model-assisted or model-dependent inference is a hybrid of design-based and model-based inferences. Here, inference is based on the chosen sampling design p and a superpopulation model ξ. Model-assisted inference provides valid inferences under the assumed superpopulation model and also protects against model misspecification by providing a valid repeated sampling inference (Rao, 1994).

6.2.9 Optimal Estimator

An estimator $t^0(\in C)$ is said to be an optimal estimator (or simply optimal) in a certain class of estimators C for estimating Y under a given superpopulation model ξ and a sampling design p if

$$E_\xi E_p(t^0 - Y)^2 \leq E_\xi E_p(t - Y)^2 \quad \forall t(\neq t^0) \in C, \quad \theta \in \Omega_\theta \qquad (6.2.1)$$

and the inequality in Eq. (6.2.1) is strict at least for one $\theta \in \Omega_\theta$.

6.2.10 Optimal Strategy

Let $\mathscr{H}$ be a class of strategies. Then, the strategy $h_1 = (p_1, t_1)$ is said to be better than $h_2 = (p_2, t_2)$ for estimating Y if

$$E_\xi E_{p_1}(t_1 - Y)^2 \leq E_\xi E_{p_2}(t_2 - Y)^2, \quad \forall \theta \in \Omega_\theta$$

and

$$E_\xi E_{p_1}(t_1 - Y)^2 < E_\xi E_{p_2}(t_2 - Y)^2 \text{ for at least one value of } \theta \in \Omega_\theta.$$

The strategy $h^0 = (p^0, t^0)$ $(\in \mathscr{H})$ is said to be optimal in the class of strategies $\mathscr{H}$ if

$$E_\xi E_{p_0}(t^0 - Y)^2 \leq E_\xi E_p(t - Y)^2 \quad \forall h(\neq h^0) \in \mathscr{H}, \quad \theta \in \Omega_\theta \qquad (6.2.2)$$

and the inequality in Eq. (6.2.2) is strict for some $\theta \in \Omega_\theta$.

6.3 MODEL-ASSISTED INFERENCE

In this section we will establish optimal strategies in the class of design-unbiased and model design—unbiased estimators. Here we will call an estimator t a predictor when it is used to estimate a random variable. In the presence of a superpopulation model ξ the population total, Y, is a random variable and hence an estimator, t, of Y will be called a predictor of Y.

6.3.1 Optimal Design-Unbiased Predictors

6.3.1.1 Product Measure Model

Model $M1$: y_i's are independently distributed with mean, $E_\xi(y_i) = \mu_i(-\infty < \mu_i < \infty)$ and variance $V_\xi(y_i) = \sigma_i^2$ for $i = 1,\ldots, N$. Here μ_i's are assumed to be known and σ_i's are positive but unknown. Under the model $M1$, we have the following theorem (Godambe and Joshi, 1965).

Theorem 6.3.1

Under the model $M1$, for a given noninformative sampling design p with inclusion probability $\pi_i > 0 \ \forall i = 1,\ldots, N$, the expected variance of a predictor t belonging to the class of p-unbiased predictors C_{pu} of the population total of Y satisfies the following inequality.

$$E_\xi[V_p(t)] \geq \sum_{i=1}^{N} \sigma_i^2\left(\frac{1}{\pi_i} - 1\right) \quad \forall t \in C_{pu} \tag{6.3.1}$$

Proof

$$\begin{aligned}
E_\xi V_p(t) &= E_\xi[E_p(t^2) - Y^2] \\
&= E_p E_\xi(t^2) - V_\xi(Y) - [E_\xi(Y)]^2 \\
&\text{(since for a noninformative } p, E_\xi E_p(t) = E_p E_\xi(t)) \\
&= E_p[V_\xi(t)] + E_p[E_\xi(t)]^2 - V_\xi(Y) - [E_\xi(Y)]^2 \\
&= E_p[V_\xi(t)] + E_p[E_\xi(t) - E_\xi(Y)]^2 - V_\xi(Y) \qquad (6.3.2) \\
&\text{(noting } E_p E_\xi(t) = E_\xi E_p(t) = E_\xi(Y))
\end{aligned}$$

Let $t(s, y)$ be the value of the predictor t based on the sample s, selected with a probability $p(s)$.

Then we can write

$$t(s, y) = \widehat{Y}_{ht} + h(s, y) \tag{6.3.3}$$

where $\widehat{Y}_{ht} = \sum_{i=1}^{N} \frac{y_i}{\pi_i} I_{si}$, is the Horvitz–Thompson (HT) predictor, $I_{si} = 1$ if $i \in s$ and $I_{si} = 0$ if $i \notin s$, and $h(s, y)$ is a function of y_i's for $i \in s$ only.

Since $t(s, y)$ is p-unbiased of Y, we must have

$$\begin{aligned}
E_p[t(s, y)] &= \sum_s t(s, y)p(s) \\
&= E_p\left(\widehat{Y}_{ht}\right) + \sum_s h(s, y)p(s) \\
&= Y
\end{aligned}$$

$$\text{i.e.,} \quad \sum_s h(s, y)p(s) = 0$$

Let $\sum_{si}$ be the sum over those samples s that do not contain the unit i.

Then,

$$\sum_s h(s, y)p(s) = \sum_{s \supset i} h(s, y)p(s) + \sum_{si} h(s, y)p(s) = 0$$

implies

$$\sum_{s \supset i} h(s, y)p(s) = -\sum_{si} h(s, y)p(s) = 0 \tag{6.3.4}$$

Furthermore,

$$E_p V_\xi(t) = \sum_s \left[V_\xi\left(\widehat{Y}_{ht}\right) + V_\xi\{h(s, y)\} + 2C_\xi\left\{\widehat{Y}_{ht}(s), h(s, y)\right\}\right] p(s) \tag{6.3.5}$$

Now

$$\sum_s V_\xi\left(\widehat{Y}_{ht}\right)p(s) = \sum_s \left(\sum_{i=1}^{N} \frac{\sigma_i^2}{\pi_i^2} I_{si} p(s) \right) = \sum_{i=1}^{N} \frac{\sigma_i^2}{\pi_i} \tag{6.3.6}$$

and

$$\begin{aligned}
\sum_s C_\xi\left\{\widehat{Y}_{ht}, h(s, y)\right\}p(s) &= \sum_s C_\xi\left\{\widehat{Y}_{ht}, h(s, y)\right\}p(s) \\
&= \sum_s p(s) E_\xi \left\{ \sum_{i=1}^{N} I_{si} \frac{(y_i - \mu_i)}{\pi_i} h(s, y) \right\} \\
&= \sum_{i=1}^{N} E_\xi \left\{ \frac{(y_i - \mu_i)}{\pi_i} \sum_s I_{si} h(s, y) p(s) \right\} \\
&= \sum_{i=1}^{N} E_\xi \left\{ \frac{(y_i - \mu_i)}{\pi_i} \sum_{s \supset i} h(s, y) p(s) \right\} \\
&= -\sum_{i=1}^{N} E_\xi \left\{ \frac{(y_i - \mu_i)}{\pi_i} \sum_{si} h(s, y) p(s) \right\} \\
&= -\sum_{i=1}^{N} E_\xi \left\{ \frac{(y_i - \mu_i)}{\pi_i} E_\xi \sum_{si} h(s, y) p(s) \right\} \\
&= 0
\end{aligned} \tag{6.3.7}$$

$$\left(\text{since } \sum_{si} h(s, y)p(s) \text{ is independent of } y_i \text{ for } i \in s \right)$$

Eqs. (6.3.2) and (6.3.5)−(6.3.7) yield the required result

$$E_\xi[V_p(t)] = \sum_{i=1}^{N} \frac{\sigma_i^2}{\pi_i} + E_p[V_\xi\{h(s,y)\}] + E_p[E_\xi(t) - E_\xi(Y)]^2 - V_\xi(Y)$$
$$\geq \sum_{i=1}^{N} \sigma_i^2 \left(\frac{1}{\pi_i} - 1\right)$$

We note that $E_\xi V_p(t)$ attains the lower bound Eq. (6.3.1) when

$$\text{(i)}\ \ E_p[V_\xi\{h(s,y)\}] = 0 \ \text{ and } \ \text{(ii)}\ \ E_\xi(t) - E_\xi(Y) = 0 \tag{6.3.8}$$

The condition (ii) of Eq. (6.3.8) implies

$$E_\xi\{h(s,y)\} = \sum_{i=1}^{N} \mu_i - \sum_{i\in s} \frac{\mu_i}{\pi_i}$$

The condition (i) of Eq. (6.3.8) implies

$$h(s,y) = E_\xi\{h(s,y)\} = \sum_{i=1}^{N} \mu_i - \sum_{i\in s} \frac{\mu_i}{\pi_i}$$

Hence the optimum predictor attaining the lower bound Eq. (6.3.1) is given by

$$t = t_0 = \sum_{i=1}^{N} \frac{y_i - \mu_i}{\pi_i} I_{si} + \sum_{i=1}^{N} \mu_i \tag{6.3.9}$$

The predictor t_0 is known as the generalized difference predictor, which is usable when μ_i's are known for every $i = 1,\ldots, N$. Hence we have the following theorem (vide Chaudhuri and Stenger, 1992).

Theorem 6.3.2

Under the model $M1$, with $\pi_i > 0\ \ \forall i = 1,\ldots, N$

$$E_\xi[V_p(t)] \geq \sum_{i=1}^{N} \sigma_i^2 \left(\frac{1}{\pi_i} - 1\right) = E_\xi V_p(t_0) \ \ \forall t \in C_{pu}$$

Remark 6.3.1

Theorem 6.3.2 was derived by Godambe (1955), when t belongs to C_{lhp}, the class of linear homogeneous unbiased predictors of Y, which is a subclass of unbiased predictors C_{pu}.

Minimizing $\sum_{i=1}^{N}\sigma_i^2\left(\frac{1}{\pi_i}-1\right)$, the right hand side of Eq. (6.3.1) with respect to π_i keeping $\sum_{i=1}^{n}\pi_i = n$, the optimum value of π_i is obtained as $\pi_i = \pi_{i0} = np_i(\sigma)$, where $p_i(\sigma) = \sigma_i \big/ \sum_{i=1}^{N}\sigma_i$. The minimum value of $\sum_{i=1}^{N}\sigma_i^2\left(\frac{1}{\pi_i}-1\right)$ with $\pi_i = \pi_{i0} = np_i(\sigma)$ is obtained as

$$\frac{\left(\sum_{i=1}^{N}\sigma_i\right)^2}{n} - \sum_{i=1}^{N}\sigma_i^2 \tag{6.3.10}$$

Let $p(\pi,\ \sigma,\ \mu)$ be an FESD(n) with (i) $\pi_i = np_i(\sigma)$ and (ii) $\sum_{i=1}^{N}\frac{\mu_i}{np_i(\sigma)}I_{si} = \sum_{i=1}^{N}\mu_i$ for every s with $p(s) > 0$. Then the expected variance of the HT predictor $\widehat{Y}_{ht}$ based on the sampling design $p(\pi,\ \sigma,\ \mu)$ attains the lower bound Eq. (6.3.10). The findings of Godambe and Joshi (1965) aforementioned have been summarized in the following theorem:

Theorem 6.3.3

Let $\mathscr{P}_n$ be the class of FESD(n). Then under the model $M1$, the strategy $\left(p(\pi,\sigma,\mu),\ \widehat{Y}_{ht}\right)$ is the optimal in the class of strategies $\mathscr{H} = (p, t)$ with $p \in \mathscr{P}_n, t \in C_{pu}$, i.e.,

$$E_\xi[V_p(t)] \geq \frac{\left(\sum_{i=1}^{N}\sigma_i\right)^2}{n} - \sum_{i=1}^{N}\sigma_i^2 = E_\xi V_{p(\pi,\sigma,\mu)}\left(\widehat{Y}_{ht}\right)\ \ \forall p \in \mathscr{P}_n, t \in C_{pu}$$

Remark 6.3.2

Hájek (1959) derived Theorem 6.3.3, when t belongs to C_{lhp}, the class of linear homogeneous p-unbiased predictors.

Consider the model $M1$ with $\mu_i = \beta x_i$, where β is an unknown real number and x_i is the value of the auxiliary variable x for the ith unit, which is known and positive for every $i = 1,\ldots, N$. Let $p(\pi,\ x)$ be an FESD(n) with $\pi_i = np_i$, $p_i = x_i/X$, and $X = \sum_{i=1}^{N}x_i$. The design $p(\pi,\ x)$ is termed an inclusion probability proportional to size (IPPS) design in Chapter 5. For a $p(\pi,\ x)$ design, t_0 reduces to $\widehat{Y}_{ht} = \sum_{i=1}^{N}\frac{y_i}{np_i}I_{si}$ and we have the following theorem.

Theorem 6.3.4

Under the model $M1$ with $\mu_i = \beta x_i$

$$E_\xi\left[V_{p(\pi,x)}(t)\right] \geq \sum_{i=1}^{N} \sigma_i^2\left(\frac{1}{np_i} - 1\right) = E_\xi V_{p(\pi,x)}\left(\widehat{Y}_{ht}\right) \quad \forall\, t \in C_{pu}$$

Theorem 6.3.4 indicates that the strategy $h^* = \left(p(\pi, x), \widehat{Y}_{ht}\right)$ is optimal in the class of strategies $\mathcal{H}^* = (p(\pi, x), t), \;\; t \in C_{pu}$.

Furthermore, for the model $M1$ with $\mu_i = \beta x_i$ and $\sigma_i^2 = \sigma^2 x_i^2 (\sigma > 0)$, $p(\pi, \sigma, \mu)$ reduces to a $p(\pi, x)$ design and we get the following result.

Theorem 6.3.5

Under the model $M1$, with $\mu_i = \beta x_i$ and $\sigma_i^2 = \sigma^2 x_i^2$, $h^* = \left(p(\pi, x), \widehat{Y}_{ht}\right)$ is the optimal in the class of strategies $\mathcal{H}_1 = (p, t)$ with $p \in \mathcal{P}_n, t \in C_{pu}$, i.e.,

$$E_\xi[V_p(t)] \geq \sigma^2\left(\frac{X^2}{n} - \sum_{i=1}^{N} x_i^2\right) = E_\xi\left[V_{p(\pi,x)}\left(\widehat{Y}_{ht}\right)\right] \quad \forall\, p \in \mathcal{P}_n, t \in C_{pu}$$

Finally, when $x_i = 1$ for $i = 1, \ldots, N$, we get $\mu_i = \beta, \sigma_i^2 = \sigma^2$ and $p(\pi, x)$ reduces to a sampling design p_0, whose inclusion probability is $\pi_i = n/N$ for $\forall\, i = 1, \ldots, N$. The sampling design p_0 includes SRSWOR. In this case, $\widehat{Y}_{ht}$ reduces to the expansion predictor $N\overline{y}_s$. Hence, we have the following result as a corollary of the Theorem 6.3.5 above.

Theorem 6.3.6

Under the model $M1$, with $\mu_i = \beta$ and $\sigma_i^2 = \sigma^2$, $h_0 = \left(p_0, N\overline{y}_s\right)$ is the optimal in the class of strategies $\mathcal{H}_1 = (p, t)$ with $p \in \mathcal{P}_n, t \in C_{pu}$, i.e.,

$$E_\xi[V_p(t)] \geq \sigma^2 N\left(\frac{N}{n} - 1\right) = N^2 E_\xi\left[V_{p_0}\left(\overline{y}_s\right)\right] \quad \forall\, p \in \mathcal{P}_n, t \in C_p$$

Remark 6.3.3

Rao T.J. (1971) considered the model $M1g$, which is the model $M1$ with $\mu_i = \beta x_i$ and $\sigma_i^2 = \sigma^2 x_i^g$, where g is positive. He defined a $g - \pi ps$ sampling design, where (i) $\pi_i = \pi_{i0} = \nu_0\, x_i^{g/2} \Big/ \sum_{i=1}^{N} x_i^{g/2}$, (ii) $\dfrac{1}{\nu_0}\left(\sum_{i\in s} \dfrac{x_i}{x_i^{g/2}}\right)\left(\sum_{i\in U} x_i^{g/2}\right) = X$, and ν_0 is the expected effective sample size of the design. Rao T.J. (1971) showed that, under the model $M1g$, the HT predictor $\widehat{Y}_{ht}=\sum_{i=1}^{N}\frac{y_i}{\pi_{i0}}I_{si}$ with $g-\pi ps$ design is the optimal in the class of strategies $\mathscr{H}_g=(p,t)$ with $p\in\mathscr{P}_{\nu_0}, t\in C_{pu}$, where $\mathscr{P}_{\nu_0}$ is the class of sampling designs with ν_0 as the expected effective sample size.

6.3.1.2 Equicorrelation Model

The equicorrelation model is defined as follows:

$$\begin{aligned}&\text{Model } M2\text{: } E_\xi(y_i)=\mu_i(-\propto<\mu_i<\propto),\ V_\xi(y_i)=\sigma_i^2(\sigma_i>0)\ \text{ and}\\&\quad C_\xi(y_i,y_j)=\rho\sigma_i\sigma_j(-1/(N-1)<\rho<1)\\&\text{for } i\neq j; i,j=1,\ldots,N.\end{aligned} \tag{6.3.11}$$

Let C_{lp} be the class of linear p-unbiased predictors of Y consisting of the predictors of the form

$$t=a_s+\sum_{i\in s}b_{si}y_i$$

where a_s and b_{si}'s are constants free from y_i's and satisfy the following p-unbiasedness conditions:

$$\text{(i)}\sum_s a_sp(s)=0\quad\text{and}\quad\text{(ii)}\sum_{s\supset i}b_{si}p(s)=1\ \text{ for } i=1,\ldots,N \tag{6.3.12}$$

From Eq. (6.3.2), it follows

$$\begin{aligned}E_\xi[V_p(t)]&=E_p[V_\xi(t)]+E_p[E_\xi(t)-E_\xi(Y)]^2-V_\xi(Y)\\&\geq E_p[V_\xi(t)]-V_\xi(Y)\end{aligned} \tag{6.3.13}$$

Now,

$$\begin{aligned}E_p[V_\xi(t)]&=E_p\left[\sum_{i\in s}b_{si}^2\sigma_i^2+\rho\sum_{i\neq}\sum_{j\in s}b_{si}b_{sj}\sigma_i\sigma_j\right]\\&=\sum_s p(s)\sum_{i\in s}(b_{si}\sigma_i)^2-(1-\rho)A\end{aligned} \tag{6.3.14}$$

where $A=\sum_s p(s)\left\{\left(\sum_{i\in s}b_{si}\sigma_i\right)^2-\sum_{i\in s}b_{si}^2\sigma_i^2\right\}$.

Let us maximize A subject to the following condition:

$$\sum_{i=1}^{N} \sigma_i \sum_{s \supset i} b_{si} p(s) = \sum_{i=1}^{N} \sigma_i \tag{6.3.15}$$

Here we note that condition (ii) of Eq. (6.3.12) implies Eq. (6.3.15), but the converse is not true. To maximize A subject to Eq. (6.3.15), consider the following ϕ with λ as a Lagrange multiplier

$$\phi = \sum_{s} p(s) \left\{ \left(\sum_{i \in s} b_{si} \sigma_i \right)^2 - \sum_{i \in s} b_{si}^2 \sigma_i^2 \right\} - 2\lambda \left\{ \sum_{i=1}^{N} \sigma_i \sum_{s \supset i} b_{si} p(s) - \sum_{i=1}^{N} \sigma_i \right\}$$

$\dfrac{\partial \phi}{\partial b_{si}} = 0$ implies

$$\sum_{i \in s} \sigma_i b_{si} - \sigma_i b_{si} = \lambda \tag{6.3.16}$$

Summing Eq. (6.3.16) over $i \in s$ and assuming p is a fixed effective sample size n design, i.e., $p \in \mathcal{P}_n$, we get

$$\sum_{i \in s} \sigma_i b_{si} = n\lambda/(n-1) \tag{6.3.17}$$

Multiplying both sides of Eq. (6.3.17) by $p(s)$ and then summing over all possible samples s yields

$$\sum_{s} p(s) \left(\sum_{i \in s} \sigma_i b_{si} \right) = n\lambda/(n-1)$$

$$\text{i.e.,} \quad \lambda = \frac{(n-1)}{n} \sum_{i=1}^{N} \sigma_i \tag{6.3.18}$$

Eqs. (6.3.16)–(6.3.18) yield the optimum value of b_{si}'s, which maximize A subject to Eq. (6.3.15) as

$$b_{si} = b_{si0} = \frac{1}{np_i(\sigma)} \quad \text{with} \quad p_i(\sigma) = \sigma_i \Big/ \sum_{i=1}^{N} \sigma_i \tag{6.3.19}$$

The maximum value of A with $b_{si} = b_{si0} = \dfrac{1}{np_i(\sigma)}$ comes out as

$$A_0 = \frac{n-1}{n} \left(\sum_{i=1}^{N} \sigma_i \right)^2 \tag{6.3.20}$$

Eqs. (6.3.14) and (6.3.20) yield

$$\begin{aligned} E_p[V_\xi(t)] &\geq \sum_s p(s)\left(\sum_{i\in s}\sigma_i b_{si}\right)^2 - (1-\rho)A_0 \\ &\geq \left\{\sum_s p(s)\sum_{i\in s} b_{si}\sigma_i\right\}^2 \Big/ \sum_s p(s) - (1-\rho)A_0 \\ &= \left(\sum_{i=1}^N \sigma_i\right)^2 - (1-\rho)A_0 \end{aligned} \tag{6.3.21}$$

The aforementioned derivations can be stated in the following theorem as follows.

Theorem 6.3.7

Under model $M2$, $E_\xi[V_p(t)] \geq (1-\rho)\left[\dfrac{\left(\sum\limits_{i=1}^N \sigma_i\right)^2}{n} - \sum\limits_{i=1}^N \sigma_i^2\right]$ for $\forall p \in \mathscr{P}_n$, $t \in C_{lp}$

Equality in Theorem 6.3.7 holds if (i) $A = A_0$ and (ii) equality in Eq. (6.3.13) holds, i.e., $b_{si} = b_{si0} = \dfrac{1}{np_i(\sigma)}$ and $a_s = a_{s0} = \sum\limits_{i=1}^N \mu_i - \sum\limits_{i\in s}\dfrac{\mu_i}{np_i(\sigma)}$.

Hence the equality in Theorem 6.3.7 holds if

$$t = t^* = \sum_{i\in s}\frac{y_i - \mu_i}{np_i(\sigma)} + \sum_{i=1}^N \mu_i \tag{6.3.22}$$

Furthermore, the predictor t^* becomes design-unbiased if and only if the inclusion probability of the ith unit $\pi_i = np_i(\sigma)$ $\forall i = 1,\ldots, N$. We will call an FESD(n) with inclusion probability $\pi_i = np_i(\sigma)$ for $\forall i = 1,\ldots, N$ as $p(\pi, \sigma)$ design. Hence we have the following theorem.

Theorem 6.3.8

Under the model $M2$, the strategy $h_1^* = \left(t^*, p(\pi,\sigma)\right)$ is the optimum in the class of strategies $\mathscr{H}_2^* = (t, p)$, $t \in C_{lp}$, $p \in \mathscr{P}_n$ in the sense that

$$\begin{aligned} E_\xi[V_p(t)] &\geq (1-\rho)\left[\frac{\left(\sum\limits_{i=1}^N \sigma_i\right)^2}{n} - \sum_{i=1}^N \sigma_i^2\right] \\ &= E_\xi\left[V_{p(\pi,\sigma)}\left(t^*\right)\right] \quad \forall t \in C_{lp}, p \in \mathscr{P}_n \end{aligned}$$

where $t^* = \sum\limits_{i\in s}\dfrac{y_i - \mu_i}{np_i(\sigma)} + \sum\limits_{i=1}^N \mu_i$.

Remark 6.3.4

The optimal predictor t^* given in Theorems 6.3.7 and 6.3.8 cannot be used in practice unless μ_i and σ_i's are known. Arnab (1986) showed that for any sampling design p (not necessarily an FESD(n)), the following inequality holds when t belongs to C_{lhp}.

$$E_\xi[V_p(t)] \geq (1-\rho)\left[\frac{\left(\sum_{i=1}^{N}\sigma_i\right)^2}{n} - \sum_{i=1}^{N}\sigma_i^2\right] \quad \forall\, t \in C_{lhp}$$

6.3.1.3 Transformation Model

We define model $M3$ as the model $M2$ with $\mu_i = a_i + \beta x_i$ and $\sigma_i^2 = \sigma^2 x_i^2$, where a_i's are known constants, x_i's are known auxiliary variables, β, $\sigma(>0)$, and ρ are unknown. The model $M3$ is called the transformation model because the transformed variables $z_i = \dfrac{y_i - a_i}{x_i}$'s have the same means, variances, and covariances. Under the model $M3$, t^* reduces to $t_1 = \sum_{i\in s}\dfrac{y_i - a_i}{nx_i}X + \sum_{i=1}^{N} a_i$ and hence we get the optimum strategy under model $M3$ using Theorem 6.3.8 as follows.

Theorem 6.3.9

Under the model $M3$, the strategy $h_1 = (t_1, p(\pi, x))$ is the optimum in the class of strategies $\mathcal{H}_1^* = (t, p)$, $t \in C_{lp}$, $p \in \mathcal{P}_n$ in the sense

$$E_\xi[V_p(t)] \geq (1-\rho)\sigma^2\left(\frac{X^2}{n} - \sum_{i=1}^{N} x_i^2\right) = E_\xi\left[V_{p(\pi,x)}(t_1)\right] \quad \forall\, t \in C_{lp}, p \in \mathcal{P}_n$$

where $p(\pi, x)$ is an FESD(n) with inclusion probability $\pi_i = np_i$.

Corollary 6.3.1

If we put $a_i = 0$ in the model $M3$, t_1 reduces to the HT predictor $\widehat{Y}_{ht} = \sum_{i\in s}\dfrac{y_i}{np_i}$, and we find that $\widehat{Y}_{ht}$ based on a $p(\pi, x)$ sampling design provides the optimum strategy.

Corollary 6.3.2

If we put $a_i = 0$ and $x_i = 1$ in the model $M3$, t_1 reduces to $\dfrac{N}{n}\sum_{i\in s} y_i = N\overline{y}_s$, where $\overline{y}_s = \sum_{i\in s} y_i/n =$ the sample mean. In this case, $p(\pi, x)$ sampling design reduces to p_0, which includes SRSWOR. Hence the expansion estimator $N\overline{y}_s$ based on an SRSWOR becomes the optimal strategy.

Corollary 6.3.3

If we put $x_i = 1$ in the model $M3$, then t_1 reduces to the difference estimator $t_D = N\left\{\overline{y}_s - \left(\overline{a}_s - \overline{a}\right)\right\}$, where $\overline{a}_s = \sum\limits_{i\in s} a_i/n$ and $\overline{a} = \sum\limits_{i=1}^{N} a_i/N$. So in this situation, the difference estimator t_D based on SRSWOR sampling design becomes the optimum strategy.

Remark 6.3.5

Mukherjee and Sengupta (1989) considered the optimality of predictors under a more general model viz. $E_\xi(y_i) = \mu_i(-\propto < \mu_i < \propto)$, $V_\xi(y_i) = \sigma_{ii} = \sigma_i^2(\sigma_i > 0)$, and $C_\xi(y_i, y_j) = \sigma_{ij}\ i \neq j;\ i, j = 1,\ldots, N$. They derived the following result (proof is omitted here):

$$E_\xi\left[V_p(t)\right] \geq \mathbf{I}'\boldsymbol{\varphi}^{-1}\mathbf{I} - \mathbf{I}'\boldsymbol{\Sigma}\mathbf{I} \ \text{ for } \ p \in \mathscr{P}_n \ \text{ and } \ t \in C_{lp}$$

where $\boldsymbol{\Sigma} = (\sigma_{ij})$, $\mathbf{I}' = (1, \ldots, N)$, $\boldsymbol{\varphi} = (\phi_{ij})$, $\phi_{ij} = \sum\limits_{s\supset i,j} v^{ij}(s)p(s)$, $\left(v^{ij}(s)\right) = \boldsymbol{\Sigma}_{\boldsymbol{s}}^{-1}$, and $\underset{\boldsymbol{s}}{\boldsymbol{\Sigma}}$ is the $n \times n$ submatrix of $\boldsymbol{\Sigma}$ obtained by considering n units in s. They also derived the optimum estimator attaining the lower bound.

6.3.2 Optimal Model Design—Unbiased Prediction

The class of linear model design $(p\xi)$-unbiased predictors for the total Y will be denoted by $C_{lp\xi}$. The class $C_{lp\xi}$ consists of the estimators of the form $t = a_s + \sum\limits_{i\in s} b_{si}y_i$ satisfying the $p\xi$-unbiasedness condition

$$E_pE_\xi(t) = E_\xi E_p(t) = E_\xi(Y) \tag{6.3.23}$$

In this section, the optimal predictor for the total Y will be considered under the following superpopulation model:

$$\begin{aligned} \text{Model } M4\text{: } & E_\xi(y_i) = \beta\, x_i,\ V_\xi(y_i) = \sigma^2 x_i^2 \ \text{ and} \\ & C_\xi(y_i, y_j) = \rho\sigma^2 x_i x_j \ \text{ for } \ i \neq j \end{aligned} \tag{6.3.24}$$

Here we define $t_0^*(\in C_{lp\xi})$ as optimal in the class $C_{lp\xi}$ for estimating βX if

$$E_pE_\xi\left(t_0^* - \beta X\right)^2 \leq E_pE_\xi(t - \beta X)^2 \ \ \forall\, t \neq t_0^* \in C_{lp\xi}.$$

The $p\xi$-unbiasedness conditions of the estimator t under the model $M4$ yield

$$\text{(i) } \sum_s a_s p(s) = 0 \ \text{ and } \ \text{(ii) } \sum_s p(s)\sum_{i\in s} b_{si}x_i = X \tag{6.3.25}$$

Now,

$$E_pE_\xi(t-\beta X)^2 = E_pE_\xi(t^2) - \beta^2X^2 = A_1 + A_2 - \beta^2X^2 \tag{6.3.26}$$

where

$$\begin{aligned} A_1 = E_p[E_\xi(t)]^2 &= \sum_s p(s)\left(a_s + \beta\sum_{i\in s} b_{si}x_i\right)^2 \\ &\geq \left[\sum_s p(s)a_s\left(a_s + \beta\sum_{i\in s} b_{si}x_i\right)\right]^2 \Big/ \sum_s p(s) \\ &= \beta^2X^2 \quad \left(\text{noting } \sum_s p(s) = 1\right) \end{aligned} \tag{6.3.27}$$

and

$$\begin{aligned} A_2 &= E_p[V_\xi(t)] \\ &= \sigma^2\sum_s p(s)\left[\left(\sum_{i\in s} b_{si}x_i\right)^2 - (1-\rho)\left\{\left(\sum_{i\in s} b_{si}x_i\right)^2 - \sum_{i\in s} b_{si}^2x_i^2\right\}\right] \end{aligned} \tag{6.3.28}$$

The expression A_2 given in Eq. (6.3.28) can be obtained from the expression Eq. (6.3.14) by writing $\sigma_i^2 = \sigma^2x_i^2$. Hence the optimum value of b_{si} that minimizes A_2 subject to Eq. (6.3.22) is obtained from Eq. (6.3.19) by writing $\sigma_i^2 = \sigma^2x_i^2$. The optimum value of b_{si} therefore comes out as

$$b_{i0} = 1/(np_i) \tag{6.3.29}$$

It can be easily checked that the equality of Eq. (6.3.27) holds when $a_s = 0$ and $b_{si} = b_{i0} = 1/(np_i)$. From the aforementioned findings we get the following theorem derived by Cassel et al. (1976).

Theorem 6.3.10

Let p be any FESD(n) with $\pi_i > 0$ for $i = 1,\ldots, N$. Then, under model $M4$, $t_0^* = \frac{1}{n}\sum_{i\in s}\frac{y_i}{p_i}$ with $p_i = x_i/X$ is the optimum in the class of linear $p\xi$-unbiased estimators ($C_{lp\xi}$) for βX, i.e.,

$$E_pE_\xi(t-\beta X)^2 \geq \sigma^2X^2\{1+(n-1)\rho\}/n = E_pE_\xi\left(t_0^* - \beta X\right)^2 \quad \forall\, t \in C_{lp\xi}$$

Corollary 6.3.4

For a $p(\pi, x)$ sampling design, the HT estimator $\widehat{Y}_{ht} = \sum_{i\in s} y_i/\pi_i$ is the optimum in $C_{lp\xi}$, the class of linear $p\xi$-unbiased estimators for βX.

Corollary 6.3.5

Putting $a_i = 0$ and $x_i = 1$ $\forall i$ in the model $M4$, we get $E_\xi(y_i) = \beta$, $V_\xi(y_i) = \sigma^2$, and $C_\xi(y_i, y_j) = \rho\sigma^2$ for $i \neq j$. In this situation, the optimum $p\xi$-unbiased estimator t_0^* reduces to $N\overline{y}_s$ and hence $\overline{y}_s$, the sample mean is the optimal in $C_{lp\xi}(\beta)$, the class of linear $p\xi$-unbiased estimators of β, i.e.,

$$E_pE_\xi\left(\widehat{\beta} - \beta\right)^2 \geq \sigma^2\{1 + (n-1)\rho\}/n = E_pE_\xi\left(\overline{y}_s - \beta\right)^2 \quad \forall \widehat{\beta} \in C_{lp\xi}(\beta)$$

6.3.3 Exchangeable Model

The random variables $y_1,\ldots, y_i,\ldots, y_N$ follow exchangeable distribution (model), if the joint distribution of $y_{j_1}, \ldots, y_{j_i}, \ldots, y_{j_N}$ is the same for all $N!$ permutations of the labels $(j_1 \neq \cdots \neq j_i \neq \cdots \neq j_N) \subset (1,\ldots, j,\ldots, N)$. In other words the joint distribution of $\mathbf{y} = (y_1,\ldots, y_i,\ldots, y_N)$ is symmetric with respect to its coordinates. Hence if $\mathbf{y}$ follows exchangeable distribution, then y_i's must have the same means, variances, and covariances. The notion of exchangeability was introduced by de Finetti (1937) and used extensively in survey sampling by many, including Ericson (1969a,b). We will denote the exchangeable model by $MEX(y)$ when $\mathbf{y}$ follows an exchangeable distribution. Let us suppose that y_i's follow the transformation model $M3$ viz. $E_\xi(y_i) = \mu_i = a_i + \beta x_i$, $V_\xi(y_i) = \sigma_i^2 = \sigma^2 x_i^2$, and $C_\xi(y_i, y_j) = \rho\sigma^2 x_i x_j$, then the transformed variables $z_i = \dfrac{y_i - a_i}{x_i}\overline{X}$'s with $\overline{X} = X/N$, follow the exchangeable model with $E_\xi(z_i) = \beta\overline{X}$, $V_\xi(y_i) = \sigma^2\overline{X}^2$, and $C_\xi(y_i, y_j) = \rho\sigma^2\overline{X}^2$. The following theorem states that $\dfrac{1}{n}\sum_{i\in s} z_i + \overline{a} = \sum_{i\in s}\dfrac{y_i - a_i}{nx_i}\overline{X} + \overline{a} = \overline{t}_0$ with $\overline{a} = \sum_{i\in U} a_i/N$ is optimum for estimating $E_\xi\left(\dfrac{1}{N}\sum_{i\in U} z_i\right) + \overline{a} = \beta\overline{X} + \overline{a} = \overline{\Upsilon}$ (say) in $\overline{C}_{p\xi}$, the class of $p\xi$-unbiased estimator (linear or nonlinear) of $\overline{\Upsilon}$. The proof of the theorem is omitted but it is obtainable from Cassel et al. (1977).

Theorem 6.3.11

Let $\mathscr{P}_n$ be any FESD(n) design with $\pi_i > 0$ for $i = 1,\ldots, N$. Then, under the exchangeable model $MEX(z)$, any $p\xi$-unbiased estimator $\bar{t}$ satisfies

$$E_\xi E_p\left(\bar{t} - \overline{\Upsilon}\right)^2 \geq \frac{\{1 + (n-1)\rho\}\sigma^2}{n}\overline{X}^2$$
$$= E_\xi E_p\left(\bar{t}_0 - \overline{\Upsilon}\right)^2 \text{ for any } \bar{t} \in \overline{C}_{p\xi} \text{ and } p \in \mathscr{P}_n.$$

The class $\overline{C}_{p\xi}$ of $p\xi$-unbiased estimators of $\overline{\Upsilon}$ contains $\overline{C}_{pu}$, the class of p-unbiased predictors of $\overline{Y}$. Furthermore, the predictor $\bar{t}_0 = \overline{X}\sum_{i\in s}\frac{y_i - a_i}{nx_i} + \bar{a}$ is a p-unbiased predictor for $\overline{Y}$ when p is a $p(\pi, x)$ design. Hence, from Theorem 6.3.11, we get the following theorem.

Theorem 6.3.12

Let $\mathscr{P}_n$ be an FESD(n) with $\pi_i > 0$ for $i = 1,\ldots, N$. Then, under the exchangeable model $EX(z)$, the strategy $\overline{h}_0 = (\bar{t}_0, p(\pi, x))$ is optimal in the class of strategies $\overline{H}^* = (p, \bar{t})$, $p \in \mathscr{P}_n, \bar{t} \in \overline{C}_{pu}$ and we have

$$E_\xi E_p\left(\bar{t} - \overline{Y}\right)^2 \geq \frac{\{1 + (n-1)\rho\}\sigma^2}{n}\overline{X}^2$$
$$= E_\xi E_{p(\pi,x)}\left(\bar{t}_0 - \overline{Y}\right)^2 \text{ for any } \bar{t} \in \overline{C}_{pu} \text{ and } p \in \mathscr{P}_n$$

6.3.4 Random Permutation Model

Let us suppose that we know the marks of four students in Statistics as $50 = a_1$, $80 = a_2$, $30 = a_3$, and $90 = a_4$, but we do not know the students' identification. Let the marks of ith student be y_i. If the vector (y_1, y_2, y_3, y_4) can take any of the $4! = 24$ permutations of the number (a_1, a_2, a_3, a_4) with equal probabilities, then the resulting model is known as random permutation model (MRP). For this MRP, Prob $(y_1 = a_2, y_2 = a_3, y_3 = a_1, y_4 = a_4) = 1/24$ and $E_\xi(y_i) = \bar{a} = \sum_{i=1}^{4} a_i/4 = 62.5$, $V_\xi(y_i) = \frac{1}{4}\sum_{i=1}^{4}(a_i - \bar{a})^2 = 568.75$, and $C_\xi(y_i y_j) = \frac{1}{12}\sum_{i\neq}^{4}\sum_{j=1}^{4} a_i a_j - \bar{a}^2 = -V_\xi(y_i)/3 = -189.583$. For any pair $i, j (i \neq j)$ the correlation between y_i and y_j is $\rho = -1/3$. In general for an MRP, the population vector $\mathbf{y} = (y_1,\ldots, y_i,\ldots, y_N)$ of N units can take any one of the $N!$ permutations of the elements of the vector $\mathbf{a} = (a_1,\ldots, a_i,\ldots, a_N)$ with equal probability, i.e., Prob

$\left(y_1 = a_{i_1}, \ldots, y_j = a_{i_j}, \ldots, y_N = a_{i_N}\right) = 1/N!$ for $i_1 \neq \cdots i_j \neq \cdots \neq i_N \in$ (1, 2,…, N). So, for the MRP we have

$$E_\xi(y_i) = \bar{a} = \sum_{i=1}^{N} a_i/N,\ V_\xi(y_i) = \sum_{i=1}^{N} (a_i - \bar{a})^2/N = \sigma_a^2 \text{ and}$$
$$C_\xi(y_i, y_j) = -\sigma_a^2/(N-1) \tag{6.3.30}$$

Because MRP is a special case of the exchangeable model, by substituting $\beta = 0$, $x_i = 1$, $a_i = \bar{a}$, and $\overline{Y} = \bar{a}$ in Theorem 6.3.11, we get the following theorem.

Theorem 6.3.13

Let $\mathscr{P}_n$ be any FESD(n) with $\pi_i > 0$ for $i = 1,\ldots, N$. Then, under MRP

$$E_\xi E_p(\bar{t} - \bar{a})^2 \geq \frac{N-n}{(N-1)n}\sigma_a^2 = E_\xi E_{p_0}\left(\bar{y}_s - \bar{a}\right)^2 \text{ for any } \bar{t} \in \overline{C}_{p\xi} \text{ and } p \in \mathscr{P}_n$$

where $\bar{y}_s = \sum_{i \in s} y_i/n$ and p_0 is defined in Corollary 6.3.2.

The results of optimality in the MRPs were also provided by Kempthorne (1969), Rao C.R. (1971), and Rao and Bellhouse (1978) under various contexts.

6.4 MODEL-BASED INFERENCE

Let s be a sample selected using a sampling design p yielding the survey data $d = (i,\ y_i,\ i \in s)$. Our objective is to estimate the population total $Y = \sum_{i \in s} y_i + \sum_{i \notin s} y_i$ on the basis of the data d. Here $\sum_{i \in s} y_i = Y_s$, the sum of the distinct units in s, can be obtained from the data d. $\sum_{i \notin s} y_i = Y_{\bar{s}}$, the sum of the distinct units in $\bar{s} = U - s$, cannot be obtained from the data d. Hence we need to predict the unknown part $Y_{\bar{s}}$ through the observed data $d = (i,\ y_i,\ i \in s)$. In a model-based inference (which is also known as the prediction approach) we establish a link between y_i, $i \in s$ and y_i, $i \notin s$ by using an appropriate superpopulation model ξ, whereas in design-based inference, sampling design establishes such a link.

Let $t = t(s,\ y)$ be a predictor of the total Y. We can write

$$t = Y_s + U_s \tag{6.4.1}$$

where $U_s = U(s,\ y) = t - Y_s$ is the predictor of the unobserved total $Y_{\bar{s}}$.

The predictor U_s depends on the y_i's for $i \in s$ but is independent of y_i, $i \notin s$. Let t be a ξ-unbiased predictor of Y, i.e., $E_\xi(t) = E_\xi(Y)$, then U_s is a ξ-unbiased predictor of $Y_{\bar{s}}$ in the sense $E_\xi(U_s) = E_\xi(Y_{\bar{s}})$. The predictor U_s is called an implied predictor of t. Our objective is to find a predictor t belonging to $C_{\xi u}$, the class of ξ-unbiased predictors of Y, for which the expected mean-square error of t attains a minimum. Thus the predictor $t_1^0 (\in C_{\xi u})$ is optimal in $C_{\xi u}$ for predicting Y if

$$\begin{aligned} M_{\xi p}\left(t_1^0\right) &= E_\xi E_p\left(t_1^0 - Y\right)^2 \leq E_\xi E_p(t - Y)^2 \\ &= M_{\xi p}(t) \quad \forall\, t\left(\neq t_1^0\right) \in C_{\xi u} \ \text{ and } \ \theta \in \Omega_\theta \end{aligned} \tag{6.4.2}$$

and the inequality is strict in Eq. (6.4.2) for at least one $\theta \in \Omega_\theta$.

Here we will assume p is noninformative so that the commutativity of the operators E_ξ and E_p are valid. We will now prove the following theorem attributed to Royall (1970).

Theorem 6.4.1

Let U_s and U'_s be the implied ξ-unbiased predictors of t and t', respectively. If ξ is a product measure (y_i's are independent for $i = 1,\ldots, N$), then the inequality

$$V_\xi(U_s) \leq V_\xi(U'_s) \tag{6.4.3}$$

for every s with $p(s) > 0$ implies

$$M_{\xi,p}(t) \leq M_{\xi,p}(t') \tag{6.4.4}$$

If the inequality is strict for some s with $p(s) > 0$ in Eq. (6.4.3), then the latter inequality in Eq. (6.4.4) is also strict.

Proof

$$\begin{aligned} M_{\xi,p}(t) &= E_\xi E_p(t - Y)^2 \\ &= E_\xi E_p\left(\sum_{i \in s} y_i + U_s - Y\right)^2 \\ &= E_p E_\xi\left(U_s - \sum_{i \notin s} y_i\right)^2 \\ &= \sum_s p(s) E_\xi\left(U_s - \sum_{i \notin s} y_i\right)^2 \\ &= \sum_s p(s)\left\{V_\xi(U_s) + V_\xi\left(\sum_{i \notin s} y_i\right) - 2C_\xi\left(\sum_{i \notin s} y_i, U_s\right)\right\} \end{aligned}$$

Now noting $C_\xi\left(\sum_{i\notin s} y_i, U_s\right) = 0$ as U_s is a function of y_i's for $i \in s$ and y_i's are independent, we find

$$M_{\xi,p}(t) = \sum_s p(s)\left\{V_\xi(U_s) + V_\xi\left(\sum_{i\notin s} y_i\right)\right\}$$

Hence $V_\xi(U_s) \le V_\xi(U'_s)$ for every s with $p(s) > 0$ implies $M_{\xi,p}(t) \le M_{\xi,p}(t')$. Furthermore, if there exists one $s = s_0$ with $p(s_0) > 0$, for which $V_\xi(U_s) < V_\xi(U'_S)$ in addition to Eq. (6.4.3), we get $M_{\xi,p}(t) < M_{\xi,p}(t')$.

Corollary 6.4.1

Let t belong to $C_{l\xi}$, the class of linear ξ-unbiased predictors of Y. Then the implied estimator U_s is also linear and in that case $C_\xi\left(\sum_{i\notin s} y_i, U_s\right) = 0$, whenever y_i's are uncorrelated.

6.4.1 Optimal Model—Unbiased Prediction

6.4.1.1 Product Measure Model

Model M_1^*: $E_\xi(y_i) = \beta x_i$, $V_\xi(y_i) = \sigma^2 v(x_i)$ and $C_\xi(y_i, y_j) = 0$ for $i \ne j$ (6.4.5)

where $v(x_i)$ is a known function of x_i.

The class $C_{l\xi}$ consists of estimators of the form

$$t = a_s + \sum_{i\in s} b_{si} y_i = Y_s + U_s \tag{6.4.6}$$

where $U_s = a_s + \sum_{i\in s}(b_{si} - 1)y_i$.

Under the model M_1^*, $E_\xi(t) = E_\xi(Y) = \beta X$ implies $E_\xi(U_s) = \beta X_{\bar{s}}$, where $X_{\bar{s}} = \sum_{i\notin s} x_i$. Now using the generalized least square theory, we find that the best linear ξ-unbiased predictor (ξ-BLUP) for $\beta X_{\bar{s}}$ is $\widehat{\beta}_{blue} X_{\bar{s}}$, where

$$\widehat{\beta}_{blue} = \sum_{i\in s} \frac{y_i x_i}{v(x_i)}\left\{\sum_{i\in s} x_i^2/v(x_i)\right\}^{-1} \tag{6.4.7}$$

is the ξ-BLUP of β based on the sample s. Hence for a given s, $V_\xi(t)$ attains a minimum when

$$t = T_0 = Y_s + \widehat{\beta}_{blue} X_{\bar{s}} \tag{6.4.8}$$

Hence we have the following theorem.

Theorem 6.4.2

Under model M_1^*, for a given sampling design p

$$M_{p\xi}(t) = E_\xi E_p(t - Y)^2 \geq E_\xi E_p(T_0 - Y)^2 = M_{p\xi}(T_0) \text{ for } t \in C_{l\xi}$$

where $T_0 = \sum_{i \in s} y_i + \widehat{\beta}_{blue} X_{\bar{s}}$ with $\widehat{\beta}_{blue} = \sum_{i \in s} \frac{y_i x_i}{\nu(x_i)} \left\{ \sum_{i \in s} x_i^2 / \nu(x_i) \right\}^{-1}$.

It follows from Theorem 6.4.2 that the optimum estimators belonging to the class $C_{l\xi}$ for $\nu(x_i)$ equal to 1, x and x^2 become, respectively, as follows

$$T_{00} = Y_s + \frac{\sum_{i \in s} x_i y_i}{\sum_{i \in s} x_i^2} X_{\bar{s}}$$

$$T_{01} = Y_s + \frac{\sum_{i \in s} y_i}{\sum_{i \in s} x_i} X_{\bar{s}} = \frac{Y_s}{X_s} X = \widehat{Y}_R$$

$$T_{02} = Y_s + \left(\frac{1}{\nu_s} \sum_{i \in s} \frac{y_i}{x_i} \right) X_{\bar{s}}$$

where ν_s is the number of distinct units in s.

The estimator $T_{01} = \widehat{Y}_R$ is the well-known ratio estimator. It is interesting to note that when $\nu(x_i) = x_i^2$, the optimum p-unbiased predictor is the HT predictor $\widehat{Y}_{ht} = \frac{1}{n} \sum_{i \in s} \frac{y_i}{x_i}$ based on a $p(\pi, x)$ sampling design whereas the optimum ξ-unbiased predictor of Y is T_{02} as described earlier. In practice $\nu(x_i)$ is not known in advance and hence it is interesting to compare the performance of $\widehat{Y}_{ht}$ with T_{02} under various forms of the functions $\nu(x)$. The following theorem (Royall, 1970) compares performance of T_{02} and $\widehat{Y}_{ht}$.

Theorem 6.4.3

Let $\mathcal{P}_n$ be the class of FESD(n) and $\nu(x_i)/x_i^2$ be a nonincreasing function of x_i with $np_i \leq 1$ for $i = 1, \ldots, N$, then under the model M_1^*

$$E_p E_\xi (T_{02} - Y)^2 \leq E_p E_\xi \left(\widehat{Y}_{ht} - Y \right)^2 \quad \forall p \in \mathcal{P}_n$$

Proof

$$E_\xi\left(\widehat{Y}_{ht}-Y\right)^2-E_\xi\left(T_{02}-Y\right)^2=E_\xi\left\{\sum_{i\in s}\left(\frac{X}{nx_i}-1\right)y_i-Y_{\bar{s}}\right\}^2$$

$$-E_\xi\left\{\left(\frac{1}{n}\sum_{i\in s}\frac{y_i}{x_i}\right)X_{\bar{s}}-Y_{\bar{s}}\right\}^2$$

$$=\frac{\sigma^2}{n^2}X_{\bar{s}}^2\sum_{i\in s}\left\{\left(\frac{X-nx_i}{X_{\bar{s}}}\right)^2-1\right\}\frac{v(x_i)}{x_i^2}$$

$$=\frac{\sigma^2}{n^2}X_{\bar{s}}^2\sum_{i\in s}c_i d_i$$

where $c_i=\left(\dfrac{X-nx_i}{X_{\bar{s}}}\right)^2-1$ and $d_i=\dfrac{v(x_i)}{x_i^2}$

Let us consider the sample s contains labels 1, 2,…, n with $x_1\geq x_2\geq\ldots\geq x_n$, then $d_1\leq d_2\leq\ldots\leq d_n$ and $c_1\leq c_2\leq\ldots\leq c_n$ with $\sum_{i=1}^n c_i\geq 0$. Let k be the greatest integer i for which $c_i\leq 0$, then

$$\sum_{i=1}^n c_i d_i=\sum_{i=1}^k c_i d_i+\sum_{i=k+1}^n c_i d_i$$

$$\geq d_k\sum_{i=1}^k c_i+d_{k+1}\sum_{i=k+1}^n c_i$$

$$=d_k\sum_{i=1}^n c_i+\left(d_{k+1}-d_k\right)\sum_{i=k+1}^n c_i$$

$$\geq 0$$

Hence the proof.

6.4.1.1.1 Optimal Strategy and Purposive Sampling Design

Let $\boldsymbol{\Psi}_n$ be the collection of all possible $\binom{N}{n}$ samples of size n and $\mathcal{P}_n$ be the class of FESD(n). From Theorem 6.4.1, we note that under model M_1^*, $E_\xi(t-Y)^2$ attains a minimum value when $T_0=\sum_{i\in s}y_i+\widehat{\beta}_{blue}X_{\bar{s}}$ with $\widehat{\beta}_{blue}=\sum_{i\in s}\dfrac{y_i x_i}{v(x_i)}\left\{\sum_{i\in s}x_i^2/v(x_i)\right\}^{-1}$.

Hence

$$
\begin{aligned}
M_{p\xi}(t) &= \sum_{s\in\Psi_n} p(s)E_\xi(t-Y)^2 \\
&\geq \sum_{s\in\Psi_n} p(s)E_\xi(T_0-Y)^2 \\
&= \sigma^2 \sum_{s\in\Psi_n} p(s)\left\{\sum_{i\in\bar{s}} \nu(x_i) + X_{\bar{s}}^2 \Big/ \sum_{i\in s}\frac{x_i^2}{\nu(x_i)}\right\}
\end{aligned}
\tag{6.4.9}
$$

So, the optimum sampling strategy should be a combination of the estimator T_0 and a purposive sampling design, which consists of only one sample $s_{\max}(\in \Psi_n)$ with probability 1, for which $V_\xi(T_0)$ attains a minimum. Now if $\nu(x_i)$ is nondecreasing and $\nu(x_i)/x_i^2$ is nonincreasing in x_i, then the right hand side of Eq. (6.4.9) is minimized for the sample $s_{\max}$ consisting of those n units having the largest x-values.

Let $\max\limits_{s\in\Psi_n}\sum\limits_{i\in s} x_i = \sum\limits_{i\in s_{\max}} x_i = X_{s_{\max}}$ and $p_{n,\max}$ be a purposive sampling design that contains only the sample $s_{\max}$ with probability 1, i.e.,

$$
p_{n,\max}(s_{\max}) = 1 \tag{6.4.10}
$$

Then we have the following theorem (Royall, 1970).

Theorem 6.4.4

If $\nu(x_i)$ is nondecreasing and $\nu(x_i)/x_i^2$ is nonincreasing in x_i, then under the model M_1^*, the strategy $h_{\max} = (T_{\max}, p_{\max})$ with $T_{\max} = \sum\limits_{i\in s_{\max}} y_i + \widehat{\beta}_{blue}X_{\bar{s}_{\max}}$ is the optimal in the class of strategies $\mathcal{H} = (t,p)$, $t\in C_{l\xi}$, $p\in\mathcal{P}_n$ in the sense that

$$
M_{p\xi}(t) \geq M_{p_{\max}\xi}(T_{\max}) = \sigma^2\left\{\sum_{i\in\bar{s}_{\max}} \nu(x_i) + X_{\bar{s}_{\max}}^2 \Big/ \sum_{i\in s_{\max}}\frac{x_i^2}{\nu(x_i)}\right\}
$$

Corollary 6.4.2

Consider the model $\widetilde{M}1$, which is the model M_1^* with $\nu(x_i) = x_i$ and $p\in\mathcal{P}_n$. Under the model $\widetilde{M}1$ $\widehat{\beta}_{blue} = \bar{y}_s/\bar{x}_s$, where $\bar{x}_s = \sum\limits_{i\in s} x_i/n$, $\bar{y}_s = \sum\limits_{i\in s} y_i/n$, the optimal predictor T_0 reduces to the well-known ratio estimator $\widehat{Y}_R = X(Y_s/X_s)$. The mean-square error of $\widehat{Y}_R$ under $\widetilde{M}1$ is given by

$$M_{p\xi}\left(\widehat{Y}_R\right) = \sigma^2 X \sum_s \left(\frac{X}{X_s} - 1\right) p(s) \tag{6.4.11}$$

Here also Eq. (6.4.11) is minimized for the purposive sampling design $p_{\max}$, which selects the sample $s_{\max}$ consisting of n units having the largest x-values with $p(s_{\max}) = 1$. Then $M_{\xi p_{\max}}\left(\widehat{Y}_R\right)$ attains the minimum value $\sigma^2 X\left(\dfrac{X}{X_{s_{\max}}} - 1\right)$. Hence, $h_{0\max} = \left(\widehat{Y}_{R\max}, p_{\max}\right)$ becomes the optimal strategy in the class of strategies $\mathcal{H} = (t, p)$ with $t \in C_{l\xi}$, $p \in \mathcal{P}_n$, and

$$\widehat{Y}_{R\max} = \frac{Y_{s_{\max}}}{X_{s_{\max}}} X.$$

6.4.1.2 Transformation Model

For the transformation model

Model $M3$: $E_\xi(y_i) = a_i + \beta x_i$, $V_\xi(y_i) = \sigma^2 x_i^2$ and $C_m(y_i, y_j) = \rho\sigma^2 x_i x_j; i \neq j$

the linear predictor $t = a_s + \sum_{i\in s} b_{si} y_i = a_s + Y_s + \sum_{i\in s} d_{si} y_i$ with $d_{si} = b_{si} - 1$ will be ξ-unbiased for the total Y if

$$\text{(i)}\ \ a_s + \sum_{i\in s} d_{si} a_i = A_{\bar{s}} \ \text{ and } \ \text{(ii)} \ \sum_{i\in s} d_{si} x_i = X_{\bar{s}} \ \text{ where } \ A_{\bar{s}} = \sum_{i\in \bar{s}} a_i \tag{6.4.12}$$

The expected mean-square error of t is given by

$$\begin{aligned} M_{p\xi}(t) &= E_p E_\xi (t - Y)^2 = E_p E_\xi \left(a_s + \sum_{i\in s} d_{si} y_i - Y_{\bar{s}}\right)^2 \\ &= E_p V_\xi \left(a_s + \sum_{i\in s} d_{si} y_i - Y_{\bar{s}}\right) \end{aligned} \tag{6.4.13}$$

Now,

$$V_\xi\left(a_s+\sum_{i\in s}d_{si}y_i-Y_{\bar{s}}\right)=\sigma^2\left[\left(\sum_{i\in s}d_{si}^2x_i^2+\rho\sum_{i\neq}\sum_{j\in s}d_{si}d_{sj}x_ix_j\right)\right.$$

$$\left.+\sum_{i\in\bar{s}}x_i^2+\rho\sum_{i\neq}\sum_{j\in\bar{s}}x_ix_j-2\rho X_{\bar{s}}\sum_{j\in\bar{s}}d_{sj}x_j\right]$$

$$=\sigma^2\left[\left\{(1-\rho)\sum_{i\in s}d_{si}^2x_i^2+\rho\left(\sum_{j\in s}d_{si}x_i\right)^2\right\}\right.$$

$$\left.+(1-\rho)\sum_{i\in\bar{s}}x_i^2+\rho X_{\bar{s}}^2-2\rho X_{\bar{s}}\sum_{j\in s}d_{si}x_i\right]$$

$$\geq\sigma^2(1-\rho)\left\{\frac{1}{\nu_s}X_{\bar{s}}^2+\sum_{i\in\bar{s}}x_i^2\right\}\tag{6.4.14}$$

(using the unbiasedness condition (ii) of Eq. (6.4.12) and noting $\sum_{i\in s}d_{si}^2x_i^2 \geq \left(\sum_{j\in s}d_{si}x_i\right)^2\Big/\nu_s=X_{\bar{s}}^2/\nu_s$)

The lower bound Eq. (6.4.14) is attained by a predictor t with $d_{si}x_i=X_{\bar{s}}/\nu_s$ and $a_s=A_{\bar{s}}-\dfrac{X_{\bar{s}}}{\nu_s}\sum_{i\in s}\dfrac{a_i}{x_i}$. So, the optimum predictor attaining the lower bound is given by

$$T_0^*=\sum_{i\in s}y_i+\frac{X_{\bar{s}}}{\nu_s}\sum_{i\in s}\frac{y_i-a_i}{x_i}+A_{\bar{s}}\tag{6.4.15}$$

Hence we have the following theorem.

Theorem 6.4.5

For a given sampling design p and under the transformation model $M3$, the best linear ξ-unbiased predictor Y is

$$T_0^*=\sum_{i\in s}y_i+\frac{X_{\bar{s}}}{\nu_s}\sum_{i\in s}\frac{y_i-a_i}{x_i}+A_{\bar{s}}\ \text{ with }\ A_{\bar{s}}=\sum_{i\in\bar{s}}a_i.$$

The predictor T_0^* satisfies

$$E_pE_\xi(t-Y)^2 \geq \sigma^2(1-\rho)\sum_s\left(\frac{1}{\nu_s}X_{\bar{s}}^2+\sum_{i\in\bar{s}}x_i^2\right)p(s)=E_pE_\xi\left(T_0^*-Y\right)^2 \text{ for } t\in C_{l\xi}$$

Remark 6.4.1

For an FESD(n), the optimum linear ξ-unbiased predictor $T_0^*=\sum_{i\in s}y_i+\frac{X_{\bar{s}}}{n}\sum_{i\in s}\frac{y_i - a_i}{x_i}+A_{\bar{s}}$ is quite different from the optimum p-unbiased predictor $t_1=\frac{X}{n}\sum_{i\in s}\frac{y_i - a_i}{x_i}+A$, where $A=\sum_{i\in U}a_i$. In particular, when $a_i=a=$ constant for $i=1,\ldots,N$, the predictor $T_0^*=\sum_{i\in s}y_i+\frac{X_{\bar{s}}}{n}\sum_{i\in s}\frac{y_i}{x_i}$ is different from the optimum p-unbiased predictor $\widehat{Y}_{ht}=\frac{X}{n}\sum_{i\in s}\frac{y_i}{x_i}$.

For $x_i=1$ and $i=1,\ldots,N$, the predictor T_0^* reduces to the well-known difference estimator

$$t_D=N\left[\overline{y}_s-\left(\overline{a}_s-\overline{a}\right)\right].$$

Furthermore, if $x_i=1$ and $a_i=a$ for $i=1,\ldots,N$, the predictor T_0^* reduces to $N\overline{y}_s$.

6.4.1.3 Multiple Regression Model

Suppose that the study variable y is related to the k-auxiliary variables $x_1,\ldots, x_i,\ldots, x_k$ through the following model,

$$\begin{aligned}\text{Model } M_{reg}: E_\xi(y_i)&=\beta_0x_i(0)+\beta_1x_i(1)+\cdots+\beta_jx_i(j)+\cdots+\beta_kx_i(k);\\ V_\xi(y_i)&=\sigma^2\nu_i \text{ and } C_\xi(y_i,y_j)=0 \text{ for } i\neq j\end{aligned} \tag{6.4.16}$$

where $x_i(0)=1$ and $x_i(j)$ is the value of the jth auxiliary variable for the ith unit, which is assumed to be known for $j=1,\ldots,k$; $i=1,\ldots,N$; $\beta_0,\ldots,\beta_j,\ldots,\beta_k$ and σ^2 are unknown model parameters, and ν_i is a known function of $x_i(0),\ldots,x_i(j),\ldots,x_i(k)$.

Without loss of generality, let us assume that the sample s contains ν_s distinct units with labels $1,\ldots,\nu_s$ and $\bar{s}$ contains units with labels $\nu_s+1,\ldots,N$.

Let $t=Y_s+U_s$ be a linear ξ-unbiased predictor of Y, where U_s is the linear ξ-unbiased predictor of $Y_{\bar{s}}$. Since t is ξ-unbiased,

$$E_\xi(U_s)=\beta_0X_{\bar{s}}(0)+\cdots+\beta_jX_{\bar{s}}(j)+\cdots+\beta_kX_{\bar{s}}(k)$$

where $X_{\bar{s}}(0) = N - \nu_s$ and $X_{\bar{s}}(j) = \sum_{i \in \bar{s}} x_i(j)$ for $j = 2, \ldots, k$. Using the generalized Gauss–Markov theorem, we note that ξ-BLUP of $Y_{\tilde{s}}$ is

$$U_s(blue) = \widehat{\beta}_0 X_{\bar{s}}(0) + \cdots + \widehat{\beta}_j X_{\bar{s}}(j) + \cdots + \widehat{\beta}_k X_{\bar{s}}(k)$$

where $\widehat{\beta}_0, \ldots, \widehat{\beta}_j, \ldots, \widehat{\beta}_k$ are the least square estimators of $\beta_0, \ldots, \beta_j$ and β_k, respectively. The estimates $\widehat{\beta}_j$'s are given by

$$\widehat{\boldsymbol{\beta}}_{\mathbf{s}} = \left(\mathbf{X}_{\mathbf{s}}^{\mathbf{T}} \mathbf{V}_{\mathbf{s}}^{-1} \mathbf{X}_{\mathbf{s}}\right)^{-1} \mathbf{X}_{\mathbf{s}}^{\mathbf{T}} \mathbf{V}_{\mathbf{s}}^{-1} \mathbf{y}_{\mathbf{s}} \tag{6.4.17}$$

where

$$\mathbf{y}_{\mathbf{s}} = \begin{bmatrix} y_0 \\ \cdot \\ y_i \\ \cdot \\ y_{\nu_s} \end{bmatrix}, \mathbf{X}_{\mathbf{s}} = \begin{bmatrix} x_1(0) & .. & x_1(j) & .. & x_1(k) \\ .. & .. & .. & .. & .. \\ x_i(0) & .. & x_i(j) & .. & x_i(k) \\ .. & .. & .. & .. & .. \\ x_{\nu_s}(0) & .. & x_{\nu_s}(j) & .. & x_{\nu_s}(k) \end{bmatrix},$$

$$\widehat{\boldsymbol{\beta}}_{\mathbf{s}} = \begin{bmatrix} \widehat{\boldsymbol{\beta}}_{\mathbf{0}} \\ \cdot \\ \widehat{\boldsymbol{\beta}}_{\mathbf{j}} \\ \cdot \\ \widehat{\boldsymbol{\beta}}_{\mathbf{k}} \end{bmatrix} \text{ and } \mathbf{V}_{\mathbf{s}} = \begin{bmatrix} \mathrm{v}_1 & .. & 0 & .. & 0 \\ .. & .. & .. & .. & .. \\ 0 & .. & \mathrm{v}_i & .. & 0 \\ .. & .. & .. & .. & .. \\ 0 & .. & 0 & .. & \mathrm{v}_{\nu_s} \end{bmatrix}$$

So for a given s, the best linear ξ-unbiased predictor of Y is

$$T_{BLUE} = Y_s + \widehat{\boldsymbol{\beta}}_{\mathbf{s}}^{\mathbf{T}} \mathbf{X}_{\bar{\mathbf{s}}}$$

where $\mathbf{X}_{\bar{\mathbf{s}}} = [X_{\bar{s}}(0), \ldots, X_{\bar{s}}(j), \ldots, X_{\bar{s}}(k)]^{\mathbf{T}}$ and $\widehat{\boldsymbol{\beta}}_{\mathbf{s}}$ is given in Eq. (6.4.17).

The expected mean-square error of T_{BLUE} is given by

$$\begin{aligned} E_p E_\xi (T_{BLUE} - Y)^2 &= E_p\left[V_\xi(Y_{\bar{s}}) + V_\xi\left(\widehat{\boldsymbol{\beta}}_{\mathbf{s}}^{\mathbf{T}} \mathbf{X}_{\bar{\mathbf{s}}}\right)\right] \\ &= \sigma^2 \sum_s \left\{ \sum_{i \in \bar{s}} \nu_i + \mathbf{X}_{\bar{\mathbf{s}}}^{\mathbf{T}} \left(\mathbf{X}_{\mathbf{s}}^{\mathbf{T}} \mathbf{V}_{\mathbf{s}}^{-1} \mathbf{X}_{\mathbf{s}}\right)^{-1} \mathbf{X}_{\bar{\mathbf{s}}} \right\} p(s) \end{aligned}$$

The aforementioned findings have been summarized in the following theorem.

Theorem 6.4.6

Under model M_{reg} and a given sampling design p

$$E_\xi E_p(t-Y)^2 \geq E_\xi E_p(T_{BLUE}-Y)^2$$
$$= \sigma^2 \sum_s \left\{ \sum_{i \in \bar{s}} \nu_i + \mathbf{X}_{\bar{\mathbf{s}}}^{\mathbf{T}} \left(\mathbf{X}_{\mathbf{s}}^{\mathbf{T}} \mathbf{V}_{\mathbf{s}}^{-1} X_s \right)^{-1} \mathbf{X}_{\bar{\mathbf{s}}} \right\} p(s) \quad \text{for any } t \in C_{\xi l}$$

where $T_{BLUE} = Y_s + \widehat{\boldsymbol{\beta}}_{\mathbf{s}}^{\mathbf{T}} \mathbf{X}_{\bar{\mathbf{s}}}$, $\widehat{\boldsymbol{\beta}}_{\mathbf{s}} = \left(\mathbf{X}_{\mathbf{s}}^{\mathbf{T}} \mathbf{V}_{\mathbf{s}}^{-1} \mathbf{X}_{\mathbf{s}} \right)^{-1} \mathbf{X}_{\mathbf{s}}^{\mathbf{T}} \mathbf{V}_{\mathbf{s}}^{-1} \mathbf{y}_{\mathbf{s}}$,
$\mathbf{X}_{\bar{\mathbf{s}}} = [X_{\bar{s}}(0), \ldots, X_{\bar{s}}(j), \ldots, X_{\bar{s}}(k)]^{\mathbf{T}}$ and $X_{\bar{s}}(j) = \sum_{i \in \bar{s}} x_i(j)$.

Corollary 6.4.3

For $k = 1$ and $\nu_i = 1$, M_{reg} reduces to $E_\xi(y_i) = \beta_0 + \beta_1 x_i(1)$; $V_\xi(y_i) = \sigma^2$ and $C_\xi(y_i, y_j) = 0$ for $i \neq j$. Now writing x_i for $x_i(1)$ and $\bar{x}_s = \sum_{i \in s} x_i/\nu_s$, the least square estimates of β_0 and β_1 become $\widehat{\beta}_0 = \bar{y}_s - \widehat{\beta}_1 \bar{x}_s$ and $\widehat{\beta}_1 = \sum_{i \in s} \left(y_i - \bar{y}_s\right)(x_i - \bar{x}_s) \big/ \sum_{i \in s} (x_i - \bar{x}_s)^2$, respectively. In this case, the optimal ξ-unbiased estimator T_{BLUE} reduces to the well-known regression estimator

$$t_{reg} = N\left\{ \bar{y}_s - \widehat{\beta}_1 \left(\bar{x}_s - \overline{X} \right) \right\} \text{ with } \overline{X} = \sum_{i \in U} x_i/N \tag{6.4.18}$$

and

$$E_\xi\left(t_{reg} - Y\right)^2 = \sigma^2 \left\{ \frac{N(N-\nu_s)}{\nu_s} + \frac{N^2\left(\overline{X} - \bar{x}_s\right)^2}{\sum_{i \in s} (x_i - \bar{x}_s)^2} \right\} \tag{6.4.19}$$

For an FESD(n) design $E_\xi(t_{reg} - Y)^2$ reduces to $\sigma^2 \left\{ \frac{N(N-n)}{n} + \frac{N^2\left(\overline{X} - \bar{x}_s\right)^2}{\sum_{i \in s} (x_i - \bar{x}_s)^2} \right\}$. The optimal sampling strategy in this case is again a purposive sampling, which selects a sample with probability one consisting of those units for which either (i) $\bar{x}_s = \overline{X}$ or (ii) $\frac{\left(\overline{X} - \bar{x}_s\right)^2}{\sum_{i \in s} (x_i - \bar{x}_s)^2}$ is minimum.

6.5 ROBUSTNESS OF DESIGNS AND PREDICTORS

In the previous sections, we have derived optimal sampling strategies in presence of various superpopulation models. For example, $\widehat{Y}_{ht}$, based on a $p(\pi, x)$ sampling design, was found to be optimum for estimating a finite population total under the superpopulation model $M1$ with $\mu_i = \beta x_i$ and $\sigma_i^2 = \sigma^2 x_i^2$. The optimum $p(\pi, x)$ design is seldom used in practice because it is difficult to construct when the sample size n is greater than 2. Apart from this, unbiased estimation of the variance of $\widehat{Y}_{ht}$ becomes tedious because the second-order inclusion probabilities are difficult to compute. It should be noted that although $\widehat{Y}_{hh}$, the Hansen–Hurwitz (HH) estimator based on probability proportional to size with replacement (PPSWR) is inadmissible, it is extensively used in practice because of its elegant and simple expression of its unbiased variance estimator. In most situations, one is not sure about the validity of a particular superpopulation model. Hence the estimator $\widehat{Y}_{ht}$ based on $p(\pi, x)$ will no longer be optimal if the assumed model $M1$ fails to hold in practice. It is very difficult to choose an appropriate model in a particular situation. Furthermore, in multicharacter surveys, we collect information about more than one character at a time. The $p(\pi, x)$ sampling design provides efficient estimators for estimation of totals of those characters, which are approximately proportional to the auxiliary variable x, but the $p(\pi, x)$ sampling design yields inefficient estimators of the totals for those characters, which are not well related to the auxiliary variable x, as reported by Rao (1966a,b). We call an estimator "robust" if it is still optimum in wider classes of designs and/or models. In this section, we will consider the robustness of predictors for misspecification of the model only.

6.5.1 Robustness of Predictors

The ratio predictor $\widehat{Y}_R = X \sum_{i\in s} y_i \Big/ \sum_{i\in s} x_i$ is the ξ-BLUP, when the following model $\widetilde{M}1$ is true.

$$\begin{aligned}&\widetilde{M}1: y_i = \beta x_i + \epsilon_i; E_\xi(\epsilon_i) = 0, V_\xi(\epsilon_i) = \sigma^2 x_i \text{ and}\\ &C_\xi(\epsilon_i, \epsilon_j) = 0 \text{ for } i \neq j\end{aligned} \tag{6.5.1}$$

Suppose that the model $\widetilde{M}1$ was not true but an alternative model $\widetilde{M}2$, which is similar to $\widetilde{M}1$ in all respects except $V_\xi(\epsilon_i) = \sigma^2 x_i^2$, is true. Then the predictor $\widehat{Y}_R$ is no longer the ξ-BLUP for Y under the model $\widetilde{M}2$.

But $\widehat{Y}_R$ retains its ξ-unbiasedness property with respect to $\widetilde{M}2$ as $E_\xi\left(\widehat{Y}_R\right) = Y$. So, we say that the predictor $\widehat{Y}_R$ is robust with respect to ξ-bias. But as far as expected mean-square error is concerned, the optimum estimator $Y_s + \dfrac{X_{\bar{s}}}{\nu_s}\sum_{i\in s}\dfrac{y_i}{x_i}$ has a much lower expected mean-square error under the model $\widetilde{M}2$ than that of $\widehat{Y}_R$. Hence we say that $\widehat{Y}_R$ is not robust with respect to the expected mean-square error.

6.5.2 Balanced Sampling Design

Let us consider the performance of $\widehat{Y}_R$ when the following linear model $\widetilde{M}3$ is true.

$$\widetilde{M}3: y_i = \alpha + \beta x_i + \epsilon_i \text{ with } E_\xi(\epsilon_i) = 0, V_\xi(\epsilon_i) = \sigma^2 \text{ and}$$
$$C_\xi(\epsilon_i, \epsilon_j) = 0 \text{ for } i \neq j$$

Under the model $\widetilde{M}3$ the ξ-bias of $\widehat{Y}_R$ is

$$B_\xi\left(\widehat{Y}_R\right) = N\alpha\left(\overline{X}/\overline{x}_s - 1\right) \text{ where } \overline{x}_s = \sum_{i\in s} x_i/\nu_s \tag{6.5.2}$$

The ξ-bias $B_\xi\left(\widehat{Y}_R\right)$ is nonzero if $\alpha \neq 0$. However, if the selected sample s is such that $\overline{x}_s = \overline{X}$, then for the given sample s, the estimator $\widehat{Y}_R$ is ξ-unbiased.

A sample s for which $\overline{x}_s = \overline{X}$ is known as a balanced sample. A sampling design that selects a balanced sample with probability 1 is called a balanced sampling design or simply a balanced design, and it will be denoted by p_B.

Under a balanced sampling design, the ratio estimator $\widehat{Y}_R$ reduces to $T_{BL} = N\overline{y}_s$, where $\overline{y}_s = \sum_{i\in s} y_i/\nu_s$. Hence the estimator T_{BL} based on a balanced sample is robust with respect to ξ-bias when the model $\widetilde{M}3$ is true. For an FESD(n), the expected mean square of $T_{BL} = N\overline{y}_s$ based on a balanced design under the model $\widetilde{M}1$ is

$$M_\xi(T_{BL}) = N^2\sigma^2(1-f)\overline{X}/n$$

The expected mean square for the optimum estimator $\widehat{Y}_{R\max}$ under the model $\widetilde{M}1$, based on the optimum design $p_{\max}$ is obtained from Eq. (6.4.11) as

$$M_\xi\left(\widehat{Y}_{R\max}\right) = N^2\sigma^2\frac{(1-f)\overline{X}}{n}\frac{\overline{x}_{\bar{s}_{\max}}}{\overline{x}_{s_{\max}}} \le M_\xi(T_{BL}) \tag{6.5.3}$$

where $\overline{x}_{s_{\max}} = \sum_{i\in s_{\max}} x_i/n$ and $\overline{x}_{\bar{s}_{\max}} = (X - n\overline{x}_{s_{\max}})/(N-n)$.

Although a balanced design provides a ξ-unbiased estimator, one should accept lower efficiency if a balance strategy $h_b = \left(p_B, N\overline{y}_s\right)$ instead of the optimum strategy $h_{\max} = \left(p_{\max}, \widehat{Y}_{\max}\right)$ is used. Obviously, the balanced design cannot be constructed unless a sample s exists with $\overline{x}_s = \overline{X}$. In this case, one may choose an approximate balance sample for which $\overline{x}_s$ is approximately equal to $\overline{X}$. The simple random sampling with replacement (SRSWR) and SRSWOR provide approximately balanced samples when the sample size n is large.

6.5.3 Polynomial Regression Model

Following Royall and Herson (1973), we denote the polynomial regression model of order J as follows:

$$\xi[\delta_0, \ldots, \delta_j, \ldots, \delta_J : v(x)] : E_\xi(y_i) = \sum_{j=0}^{J} \delta_j \beta_j x_i^j, \; V_\xi(y_i) = \sigma^2 v(x_i) \;\text{ and }\; C_\xi(y_i, y_j) = 0 \;\text{ for }\; i \neq j \tag{6.5.4}$$

where $\delta_j = 0$ if the term x_i^j is absent in the model and $\delta_j = 1$ if x_i^j is present in the model Eq. (6.5.4).

For example, $\xi[1,\ 0,\ 1 : x]$ represents the model $E_\xi(y_i) = \beta_0 + \beta_2 x_i^2$, $V_\xi(y_i) = \sigma^2 x_i$ and $C_\xi(y_i,\ y_j) = 0$ for $i \neq j$.

Let $T = Y_s + U_s$ be a linear ξ-unbiased estimator of Y. Then $E_\xi(U_s) = E_\xi(Y_{\bar{s}}) = \sum_{j=0}^{J} \delta_j \beta_j X_{\bar{s}}^{(j)}$, where $X_{\bar{s}}^{(j)} = \sum_{i \in \bar{s}} x_i^j$. Using the generalized Gauss–Markov theorem, we note that ξ-BLUP of $E_\xi(Y_{\bar{s}})$ is $\sum_{j=0}^{J} \delta_j \widehat{\beta}_j X_{\bar{s}}^{(j)}$, where $\widehat{\beta}_j$ is the weighted least square predictor of β_j. Hence, we have the following theorem by Royall and Herson (1973).

Theorem 6.5.1

Under the model $\xi[\delta_0, \ldots, \delta_j, \ldots, \delta_J : v(x)]$, the ξ-BLUP for the total Y is

$$\widehat{T} = Y_s + \sum_{j=0}^{J} \delta_j \widehat{\beta}_j X_{\bar{s}}^{(j)}$$

where $\widehat{\beta}_j$ is the weighted least square estimate of β_j and $X_{\bar{s}}^{(j)} = \sum_{i \in \bar{s}} x_i^j$ for $j = 0, \ldots, J$.

The ξ-bias of the ratio estimator $\widehat{Y}_R = \dfrac{\sum_{i\in s} y_i}{\sum_{i\in s} x_i} X$ under the model $\xi[\delta_0,\ldots, \delta_j,\ldots, \delta_J : \nu(x)]$ is

$$B_\xi\left(\widehat{Y}_R\right) = E_\xi\left(\widehat{Y}_R - Y\right) = N\overline{X}\sum_{j=0}^{J} \delta_j\beta_j\left(\frac{\overline{x}_s^{(j)}}{\overline{x}_s} - \frac{\overline{X}^{(j)}}{\overline{X}}\right) \tag{6.5.5}$$

where $\overline{x}_s^{(j)} = \sum_{i\in s} x_i^j \Big/ \nu_s$ and $\overline{X}^{(j)} = \sum_{i\in U} x_i^j \Big/ N$.

From Eq. (6.5.5), we note that the bias of $\widehat{Y}_R$ is independent of the regression coefficient β_1 and the bias is of order N, the population size. Furthermore, the ξ-bias of $\widehat{Y}_R$ is zero under the model $\xi[\delta_0,\ldots, \delta_j,\ldots, \delta_J : \nu(x)]$ if and only if

$$\frac{\overline{x}_s^{(j)}}{\overline{x}_s} = \frac{\overline{X}^{(j)}}{\overline{X}} \tag{6.5.6}$$

for all $j = 1,\ldots, J$ with $\delta_j = 1$.

For example, $\widehat{Y}_R$ is ξ-unbiased for the model $\xi[1, 1 : \nu(x)]$

$$\text{if} \quad \frac{\overline{x}_s^{(0)}}{\overline{x}_s} = \frac{\overline{X}^{(0)}}{\overline{X}} \quad \text{i.e.,} \quad \overline{x}_s = \overline{X}.$$

Similarly, $\widehat{Y}_R$ is ξ-unbiased under the model $\xi[1, 1, 1 : \nu(x)]$ if and only if

$$\overline{x}_s = \overline{X} \text{ and } \overline{x}_s^{(2)} = \overline{X}^{(2)}.$$

6.5.4 Balanced Sample of Order *k*

A sample s is said to be a balanced sample of order k if

$$\overline{x}_s^{(j)} = \overline{X}^{(j)} \text{ for } j = 1, \ldots, k(\leq J) \tag{6.5.7}$$

A sampling design that selects a balanced sample of order k with probability 1 is called a balanced sampling design of order k or simply a balanced design of order k. Construction of a balanced sample of order k is certainly difficult when k is large (say 5 or more). We will denote a balanced sampling design of order k by p_{Bk}. However, SRSWOR and SRSWR provide an approximate balanced sample when the sample size n is large.

6.5.5 Optimality of Balanced Sampling

Consider an FESD(n) with model $\xi[1,\ 1:x]$. Under this model the ξ-BLUP is given by

$$T = Y_s + (N-n)\widehat{\beta}_0 + \widehat{\beta}_1 X_{\bar{s}} \tag{6.5.8}$$

where

$$\widehat{\beta}_0 = \left(n\overline{x}_s \sum_{i\in s}\frac{y_i}{x_i} - n^2\,\overline{y}_s\right)\Bigg/ D,\ \widehat{\beta}_1 = \left(Y_s\sum_{i\in s}\frac{1}{x_i} - n\sum_{i\in s}\frac{y_i}{x_i}\right)\Bigg/ D \text{ and}$$
$$D = n^2\left\{\overline{x}_s\left(\frac{1}{n}\sum_{i\in s}\frac{1}{x_i}\right) - 1\right\}$$

For a balanced sampling design $\overline{x}_s = \overline{X}$ and hence the ξ-BLUP reduces to the expansion estimator $N\overline{y}_s$. Royall and Herson (1973) proved that ξ-BLUP for the polynomial regression model reduces to the expansion estimator when a balanced sampling design is used.

Theorem 6.5.2

Let p_{BJ} be a balanced sampling design of order J with fixed effective sample size n. Then,

$$\widehat{T}[1,\delta_1,\delta_2,\ldots,\delta_J:1] = \widehat{T}[\delta_0,1,\delta_2,\ldots,\delta_J:x] = \cdots =$$
$$\widehat{T}[\delta_0,\delta_1,\delta_2,\ldots,1:x^J] = N\overline{y}_s$$

where $\delta_j = 0$ or 1 for $j = 1,\ldots,J$ and $\widehat{T}[\delta_0,\delta_1,\ldots,\delta_{j-1},1,\delta_{j+1},\ldots,\delta_J : x^j]$ is the ξ-BLUP for Y under the model $[\delta_0,\ \delta_1,\ldots,\ \delta_{j-1},\ 1,\ \delta_{j+1},\ldots,\ \delta_J : x^j]$.

Proof

Using Theorem 6.5.1, we find

$$\widehat{T}\left[\delta_0,\delta_1,\ldots,\delta_{j-1},1,\delta_{j+1},\ldots,\delta_J : x^j\right] = Y_s + \sum_{j=0}^{J}\delta_j\widehat{\beta}_j X_{\bar{s}}^{(j)} \tag{6.5.9}$$

The coefficients $\widehat{\beta}_j$ of Eq. (6.5.9) are obtained by minimizing

$$\sum_{i\in s}\frac{1}{x_i^j}\left(y_i - \sum_{r=0}^{J}\delta_r\beta_r x_i^r\right)^2$$

with respect to $\beta_0,\ldots,\ \beta_r,\ldots,\ \beta_J$.

In particular, $\dfrac{\partial}{\partial \beta_j} \sum_{i\in s} \dfrac{1}{x_i^j}\left(y_i - \sum_{r=0}^{J} \delta_j \beta_j x_i^r\right)^2 = 0$ yields

$$\sum_{i\in s} y_i = \sum_{r=0}^{J} \delta_r \widehat{\beta}_r \sum_{i\in s} x_i^r \tag{6.5.10}$$

For a balanced sampling design $\sum_{i\in s} x_i^r = \dfrac{n}{N-n} \sum_{i\in \bar{s}} x_i^r = \dfrac{n}{N-n} X_{\bar{s}}^{(r)}$, Eq (6.5.10) reduces to

$$\sum_{i\in s} y_i = \frac{n}{N-n} \sum_{r=0}^{J} \delta_r \widehat{\beta}_r X_{\bar{s}}^{(r)} \tag{6.5.11}$$

Now substituting Eq. (6.5.11) in Eq. (6.5.9), we get

$$\widehat{T}\left[\delta_0, \delta_1, \ldots, \delta_{j-1}, 1, \delta_{j+1}, \ldots, \delta_J : x^j\right] = N\overline{y}_s.$$

6.6 BAYESIAN INFERENCE

Let s be an unordered sample selected from a finite population U using a noninformative sampling design p. Let the unordered sample s contains ν_s distinct units with labels $i_1, \ldots, i_k, \ldots, i_{\nu_s}$, where $i_1 < \ldots < i_k < \ldots < i_{\nu_s}$ and $\bar{s}$, the complementary of the sample s, contains $N - \nu_s$ distinct units with labels $i_{\nu_s+1}, \ldots, i_N$, where $i_{\nu_s+1} < \ldots < i_N$. Let the unordered data be denoted by $d = \{(i, y_i);\ i \in s\}$. In Bayesian inference, we assume that the population vector $\mathbf{y} = (y_1, \ldots, y_i, \ldots, y_N)$ has a prior distribution $h(\mathbf{y})$, which may be regarded as the statistician's belief or ignorance about parameter $\mathbf{y}$. Here we assume that the parametric space of $\mathbf{y}$ is the N-dimensional Euclidean space $R^N = \Omega_N = \{-\propto < y_1 < \propto, \ldots, -\propto < y_N < \propto\}$. After observing the data d, the parametric space will be reduced to Ω_d, where $i_1, \ldots, i_k$ and i_{ν_s}th coordinates are respectively fixed as $y_{i_1}, \ldots, y_{i_k}$ and $y_{i_{\nu_s}}$ and the rest $N - \nu_s$ coordinates can take any value between $-\propto$ and $\propto$. We technically say Ω_d is consistent with data d (details given in Section 2.7). The Bayesian approach in finite population sampling was mainly popularized by Godambe (1968), Basu (1969), Ericson (1969a,b), and Cassel et al. (1977). The likelihood of $\mathbf{y}$ given the data d is

$$L(\mathbf{y}|d) = \begin{cases} p(s) & \text{for} \quad \mathbf{y} \in \Omega_d \\ 0 & \text{for} \quad \mathbf{y} \notin \Omega_d \end{cases} \tag{6.6.1}$$

The posterior distribution of $\mathbf{y}$ taking the prior of $\mathbf{y}$ as $h(\mathbf{y})$ is given by

$$h(\mathbf{y}|d) \propto \begin{cases} p(s)h(\mathbf{y}) & \text{for} \quad \mathbf{y} \in \Omega_d \\ 0 & \text{for} \quad \mathbf{y} \notin \Omega_d \end{cases} \tag{6.6.2}$$

Writing the marginal distribution of $y^{(s)} = (y_i, i \in s) = h\left(\mathbf{y}^{(s)}\right) = \int_{-\propto}^{\propto} \ldots \int_{-\propto}^{\propto} h(\mathbf{y}) dy_{\nu_{s+1}} \ldots dy_N$, the constant of proportionality in Eq. (6.6.2) is obtained as $\{p(s) \cdot h(\mathbf{y}^{(s)})\}^{-1}$. Hence the posterior distribution of $\mathbf{y}$ taking the prior of $\mathbf{y}$ as $h(\mathbf{y})$ is given by

$$h(\mathbf{y}|d) = \begin{cases} h(\mathbf{y})/h\left(\mathbf{y}^{(s)}\right) & \text{for} \quad \mathbf{y} \in \Omega_d \\ 0 & \text{for} \quad \mathbf{y} \notin \Omega_d \end{cases} \tag{6.6.3}$$

From the expression (6.6.3) above, we note that the posterior distribution of $\mathbf{y}$, given the data d, is independent of the sampling design. Thus Bayesian inference about any function of $\mathbf{y}$ should depend only on the data but not the sampling design, through which the data were collected. We summarize the aforementioned discussions in the following theorem.

Theorem 6.6.1

Let an unordered data d be selected following a sampling design p and let the parameter vector $\mathbf{y}$ has a prior distribution $h(\mathbf{y})$. Then the posterior distribution of $\mathbf{y}$, given the data d, is independent of the sampling design.

Corollary 6.6.1

Let $y_1, \ldots, y_N$'s be mutually independent and the density of y_i be $h_i(y_i)$. Then the joint density of $\mathbf{y}$ is $\prod_{i=1}^{N} h_i(y_i)$ and the posterior distribution of $\mathbf{y}$ given d is

$$h(\mathbf{y}|d) = \begin{cases} \prod_{i \notin s}^{N} h_i(y_i) & \text{for} \quad \mathbf{y} \in \Omega_d \\ 0 & \text{for} \quad \mathbf{y} \notin \Omega_d \end{cases}$$

Corollary 6.6.2

In case the prior distribution of $\mathbf{y}$ depends on an unknown parameter vector θ, we write the density of $\mathbf{y}$ as $h(\mathbf{y}|\theta)$. Let the prior distribution of θ be given by $F(\theta)$, then the posterior density of $\mathbf{y}$, given the data d, is

$$h(\mathbf{y}|d) = \begin{cases} \dfrac{\displaystyle\int_{\Theta} h(\mathbf{y}|\theta)dF(\theta)}{\displaystyle\int_{\Theta} h\left(y^{(s)}|\theta\right)dF(\theta)} & \text{for} \quad \mathbf{y} \in \Omega_d \\ 0 & \text{for} \quad \mathbf{y} \notin \Omega_d \end{cases}$$

where $\mathbf{\Theta}$ is the parameter space of θ.

6.6.1 Bayes Estimator

An estimator T^* in the class C of estimators is said to be the Bayes estimator for a population total Y under a given sampling design p and prior distribution $h(\mathbf{y})$ if for every other estimator T in C satisfies

$$\int_{\Omega_N}\left[\sum_s p(s)\left(T^* - Y\right)^2\right]h(\mathbf{y})d\mathbf{y} \le \int_{\Omega_N}\left[\sum_s p(s)(T - Y)^2\right]h(\mathbf{y})d\mathbf{y}$$

Following Godambe (1969), the Bayes estimator is obtained as

$$\begin{aligned} T^* &= \sum_{i\in s} y_i + E_{\mathbf{y}}\left(\sum_{i\notin s} y_i|d\right) \\ &= N\left[f_s\overline{y}_s + (1 - f_s)\psi(d)\right] \end{aligned} \tag{6.6.4}$$

where $f_s = \nu_s/N$, $\psi(d) = E_{\mathbf{y}}\left(\overline{y}_{\bar{s}}\middle|d\right)$ and $\overline{y}_{\bar{s}} = \sum_{i\in\bar{s}} y/(N - n_s)$.

Remark 6.6.1

In case $h(\mathbf{y}|\theta)$, the distribution of $\mathbf{y}$, depends on some known parameter vector θ, then $\psi(d)$ involves θ and we write $\psi(d, \theta)$ in place of $\psi(d)$. In particular, if y_i's are independently and identically distributed with common mean $\mu(\theta)$, then $\psi(d, \theta) = \mu(\theta)$ and the Bayes estimator T^* reduces to

$$T^* = N\left[f_s\overline{y}_s + (1 - f_s)\mu(\theta)\right] \tag{6.6.5}$$

Remark 6.6.2

In case $h(\mathbf{y}|\theta)$ involves an unknown parameter θ and θ has a prior distribution $F(\theta|\gamma)$, where γ is a known parameter vector, then the Bayes estimator T^* may be written as

$$T^* = N\left[f_s\overline{y}_s + (1-f_s)\psi(d,\gamma)\right] \tag{6.6.6}$$

In particular, if y_i's are independently and identically distributed with common mean $\mu(\theta)$, then the Bayes estimator T^* becomes

$$T^* = N\left[f_s\overline{y}_s + (1-f_s)E_\theta\{\mu(\theta|d,\gamma)\}\right] \tag{6.6.7}$$

where $E_\theta\{\mu(\theta|d,\gamma)\}$ is the conditional expectation $\mu(\theta)$, given the data d.

Example 6.6.1 (Ericson, 1969a,b)

Let y_i's $(i = 1,\ldots, N)$ be independently normally distributed random variables with unknown mean θ and known variance σ^2. Suppose further that θ is normally distributed with the known mean θ_0 and known variance σ_0^2. Here the parameter θ is normally distributed with the known parameter vector $\psi = (\mu_0, \sigma_0)$. It is well known that the posterior distribution of θ given the data $d = \{(i, y_i);\ i \in s\}$ is normally distributed with mean $E_\theta(\theta|d,\gamma) = w_0\overline{y}_s + (1-w_0)\mu_0$ and variance $\left(\nu_s/\sigma^2 + 1/\sigma_0^2\right)^{-1}$, where $w_0 = (\nu_s/\sigma^2)\left(\nu_s/\sigma^2 + 1/\sigma_0^2\right)^{-1}$. Hence using Eq. (6.6.7), the Bayes estimator for the population total is obtained as

$$T^* = N\left[f_s\overline{y}_s + (1-f_s)\left\{w_0\overline{y}_s + (1-w_0)\mu_0\right\}\right].$$

Example 6.6.2

Suppose that a sample s of size n_s is selected from a finite population of size N. Let n_A and N_A be the number of units that belong to the certain category A (HIV+, say) in the sample and population, respectively. Here $y_i = 1$ if $i \in A$ and $y_i = 0$ if $i \notin A$. Let us assume that y_i's are independent Bernoulli variables with parameter π, and the prior distribution of π has the form $p(\pi) \propto \pi^{\alpha-1}(1-\pi)^{\beta-1}$ with $\alpha, \beta > 0$. Here, n_A follows the binomial distribution $B(n_s, \pi)$ and the posterior distribution of π, given n_A, has the form $p(\pi|n_A) \propto \pi^{\alpha+n_A-1}(1-\pi)^{\beta+n_s-n_A-1}$, a beta distribution with parameter $(\alpha + n_A,\ \beta + n_s - n_A)$. The Bayes estimator of the population proportion $\pi_A = N_A/N$ is given by

$$\begin{aligned}\widehat{\pi}_A &= \left[f_sP_A + (1-f_s)E(\pi|n_A)\right]\\ &= \left[f_sP_A + (1-f_s)\frac{\alpha + n_sP_A}{\alpha+\beta+n_s}\right]\end{aligned}$$

where $P_A = n_A/n_s =$ proportion of HIV+ persons in the sample.

6.7 COMPARISON OF STRATEGIES UNDER SUPERPOPULATION MODELS

In this section, we will compare the relative efficiencies of a few well-known sampling strategies under the superpopulation model

$$M_1^*: E_\xi(y_i) = \beta x_i,\ V_\xi(y_i) = \sigma^2 \nu(x_i) \text{ and } C_\xi(y_i, y_j) = 0 \text{ for } i \neq j = 1, \ldots, N \tag{6.7.1}$$

(i) HH strategy (H_1): HH estimator $\widehat{Y}_{hh} = \dfrac{1}{n}\sum_{r=1}^{n}\dfrac{y(r)}{p(r)}$ (defined in Section 5.2.3) based on PPSWR sampling design of size.

(ii) HT strategy (H_2): HT estimator $\widehat{Y}_{ht} = \sum_{i\in s}\dfrac{y_i}{\pi_i}$ based on an IPPS sampling design of fixed effective size n.

(iii) Rao−Hartley−Cochran (RHC) strategy (H_3): RHC estimator $\widehat{Y}_{rhc} = \sum_{i\in s}\dfrac{y_i}{p_i}P_i$ based on RHC sampling scheme (defined in Section 5.6).

(iv) Lahiri−Midzuno−Sen (LMS) strategy (H_4): Unbiased ratio estimator $\widehat{Y}_{lms} = \dfrac{y_s}{x_s}X$ based on probability proportional to aggregate size (PPAS) sampling scheme (defined in Section 5.5).

(v) Ratio estimator strategy (H_5): Ratio estimator $\widehat{Y}_R = \dfrac{y_s}{x_s}X$ based on SRSWOR sampling scheme.

(vi) Ratio estimator strategy (H_6): Ratio estimator $\widehat{Y}_R^* = \dfrac{y_s}{x_s}X$ based on SRSWR sampling scheme.

The expected variances of the design-unbiased estimators $\widehat{Y}_{hh}$, $\widehat{Y}_{ht}$, $\widehat{Y}_{rhc}$, $\widehat{Y}_{lms}$ and approximate mean-square errors of the approximate design unbiased estimators $\widehat{Y}_R$ and $\widehat{Y}_R^*$ are given under the model M_1^* as follows:

$$\begin{aligned}
E_\xi\left[V\left(\widehat{Y}_{hh}\right)\right] &= \frac{1}{n}E_\xi\left(\sum_{i\in U}\frac{y_i^2}{p_i} - Y^2\right)\\
&= \frac{\sigma^2}{n}\sum_{i\in U}\nu(x_i)\left(\frac{1}{p_i} - 1\right)\\
&= \frac{\sigma^2}{n}\sum_{i\in U}\frac{\nu(x_i)}{x_i}(X - x_i)\\
&= Q_1(\text{say})
\end{aligned}$$

$$E_\xi\left[V\left(\widehat{Y}_{ht}\right)\right] = E_\xi\left[\sum_{i\in U} y_i^2\left(\frac{1}{\pi_i}-1\right) + \sum_{i\neq}\sum_{j\in U} y_i y_j\left(\frac{\pi_{ij}}{\pi_i\pi_j}-1\right)\right]$$

$$= \sum_{i\in U}\left\{\sigma^2\nu(x_i)+\beta^2x_i^2\right\}\left(\frac{1}{\pi_i}-1\right) + \beta^2\sum_{i\neq}\sum_{j\in U} x_ix_j\left(\frac{\pi_{ij}}{\pi_i\pi_j}-1\right)$$

For a $p(\pi, x)$ sampling $\pi_i = np_i = nx_i/X$ and we get

$$E_\xi\left[V\left(\widehat{Y}_{ht}\right)\right] = \sigma^2\sum_{i\in U}\nu(x_i)\left(\frac{1}{np_i}-1\right)$$

$$= \sigma^2\sum_{i\in U}\frac{\nu(x_i)}{x_i}\left(\frac{X}{n}-x_i\right)$$

$$= Q_2(\text{say})$$

$$E_\xi\left[V\left(\widehat{Y}_{rhc}\right)\right] = \frac{N-n}{n(N-1)}E_\xi\left(\sum_{i\in U}\frac{y_i^2}{p_i}-Y^2\right)$$

$$= \sigma^2\frac{N-n}{n(N-1)}\sum_{i\in U}\nu(x_i)\left(\frac{1}{p_i}-1\right)$$

$$= \sigma^2\frac{N-n}{n(N-1)}\sum_{i\in U}\frac{\nu(x_i)}{x_i}(X-x_i)$$

$$= Q_3\ (\text{say})$$

$$E_\xi\left[V\left(\widehat{Y}_{lms}\right)\right] = E_\xi\left(\sum_{i\in U}T_iy_i^2 + \sum_{i\neq}\sum_{j\in U}T_{ij}y_iy_j - Y^2\right)\ (\text{using Theorem 5.5.1})$$

where $T_i = X\sum_{s\supset i}\frac{1}{x_s}\Big/\binom{N-1}{n-1}$ and $T_{ij} = X\sum_{s\supset i,j}\frac{1}{x_s}\Big/\binom{N-1}{n-1}$.

Hence,

$$E_\xi\left[V\left(\widehat{Y}_{lms}\right)\right] = \sigma^2\sum_{i\in U}(T_i-1)\nu(x_i) + \beta^2\left(\sum_{i\in U}T_ix_i^2 + \sum_{i\neq}\sum_{j\in U}T_{ij}x_ix_j - X^2\right)$$

Finally, noting $\sum_{i\in U} T_i x_i^2 + \sum_{i\neq}\sum_{j\in U} T_{ij} x_i x_j = X^2$ (Rao, T.J., 1967a,b), we get

$$E_\xi\left\{V\left(\widehat{Y}_{lms}\right)\right\} = \sigma^2 \sum_{i\in U} \frac{\nu(x_i)}{x_i}\left(T_i x_i - x_i\right)$$

$$= Q_4 \text{ (say)}$$

An approximate mean-square error of $\widehat{Y}_R$ is given by

$$MSE\left(\widehat{Y}_R\right) = \frac{N-n}{Nn}\frac{1}{N-1}\sum_{i\in U}\left(y_i - Rx_i\right)^2 \text{ where}$$

$$R = Y/X \text{ (vide Eq. 8.4.7)}$$

and

$$E_\xi\left[MSE\left(\widehat{Y}_R\right)\right] = \frac{N-n}{Nn}\frac{1}{N-1}E_\xi\sum_{i\in U}\left(y_i - Rx_i\right)^2$$

$$= Q_5 \text{ (say)}$$

The expected approximate mean-square error of $\widehat{Y}_R^*$ is

$$E_\xi\left[MSE\left(\widehat{Y}_R^*\right)\right] = \frac{1}{n}\frac{1}{N}E_\xi\sum_{i\in U}\left(y_i - Rx_i\right)^2$$

$$= Q_6 \text{ (say)}$$

6.7.1 Hansen—Hurwitz Strategy With Others

The strategy H_i is said to be better than that of H_j under the model M_1^* if $Q_i < Q_j$ for $i \neq j$. It is well known that H_1, the HH strategy, is inferior to RHC strategy H_3 since $V\left(\widehat{Y}_{rhc}\right) < V\left(\widehat{Y}_{hh}\right)$. The strategy H_1 is also inferior to the HT strategy H_2 under the superpopulation model M_1^* since $Q_1 > Q_2$. However, no exact comparisons between the strategies H_1 and H_6 or H_3 and H_5 are available under the model M_1^*. However, if $\nu(x_i) = x_i^g$, following Cassel et al. (1977), we note that for a large N and n, the strategy H_1 is approximately superior, equal, or inferior to the alternative H_6 if $g > 1$, $=$ or <1, respectively.

6.7.2 Horvitz–Thompson and Rao–Hartley–Cochran Strategy

Following Hanurav (1967), we get

$$Q_2 - Q_3 = \sigma^2 \sum_{i\in U} \frac{\nu(x_i)}{x_i}\left(\frac{X}{n} - x_i - \frac{N-n}{n(N-1)}X + \frac{N-n}{n(N-1)}x_i\right)$$

$$= -\frac{N(n-1)}{n(N-1)}\sigma^2 \sum_{i\in U} \frac{\nu(x_i)}{x_i}\left(x_i - \overline{X}\right)$$

$$= -\frac{N^2(n-1)}{n(N-1)}\sigma^2 Cov\left(\frac{\nu(x_i)}{x_i}, x_i\right)$$

Thus $Q_2 - Q_3$ is $<$, $=$ or >0 if $\nu(x_i)/x_i$ is increasing, constant, or decreasing function of x_i, respectively.

6.7.3 Horvitz–Thompson and Lahiri–Midzuno–Sen Strategy

Following Rao T.J. (1967a,b), we find

$$Q_2 - Q_4 = \sigma^2 \sum_{i\in U} \frac{\nu(x_i)}{x_i}\left(\frac{X}{n} - x_i - T_i x_i + x_i\right)$$

$$= -\sigma^2 N Cov\left(\frac{\nu(x_i)}{x_i}, T_i x_i\right)$$

$$\left(\text{since } \sum_{i\in U} T_i x_i = NX/n\right)$$

Thus $Q_2 - Q_4$ is $<$, $=$ or >0 if $\nu(x_i)/x_i$ is increasing, constant, or decreasing function of x_i respectively, because $T_i x_i$ is an increasing function of x_i as observed by Rao T.J. (1967a,b).

6.7.4 Rao–Hartley–Cochran and Lahiri–Midzuno–Sen Strategy

Following Chaudhuri and Arnab (1979a,b), we note that

$$Q_3 - Q_4 = \sigma^2 \sum_{i\in U} \frac{\nu(x_i)}{x_i}\left(\frac{N-n}{n(N-1)}X - \frac{N-n}{n(N-1)}x_i - T_i x_i + x_i\right)$$

$$= \sigma^2 \sum_{i\in U} \frac{\nu(x_i)}{x_i} z_i$$

where $z_i = \dfrac{N - n}{n(N - 1)}X - T_i x_i + \dfrac{N(n - 1)}{n(N - 1)}x_i.$

Noting $\sum_{i\in U} z_i = \frac{(N - n)N}{n(N - 1)} X - \frac{N}{n} X + \frac{N(n - 1)}{n(N - 1)} X = 0$ as $\sum_{i\in U} T_i x_i = NX/n$, we can write

$$Q_3 - Q_4 = \sigma^2 NCov\left(\frac{v(x_i)}{x_i}, z_i\right) \tag{6.7.2}$$

Furthermore, noting $T_i = X\sum_i \frac{1}{x_i + x_{i_2} + \ldots + x_{i_n}} \Big/ \binom{N-1}{n-1}$, where $\sum_i$ is the sum over distinct $n-1$ integers i_2, $i_3,\ldots,$ i_n $(i_2 \neq i_3 \neq \ldots \neq i_n)$ from 1, 2,..., $N(\neq i)$

and

$$T_i + x_i\frac{\partial T_i}{\partial x_i} = \left[X\sum_i \frac{1}{x_i + x_{i_2} + \cdots + x_{i_n}} + x_i\left\{\sum_i \frac{1}{x_i + x_{i_2} + \cdots + x_{i_n}} - X\sum_i \frac{1}{(x_i + x_{i_2} + \cdots + x_{i_n})^2}\right\}\right] \Big/ \binom{N-1}{n-1}$$

$$= \sum_i \frac{x_i(x_i + x_{i_2} + \cdots + x_{i_n}) + X(x_{i_2} + \cdots + x_{i_n})}{(x_i + x_{i_2} + \cdots + x_{i_n})^2} \Big/ \binom{N-1}{n-1}$$

$$\geq \sum_i \frac{x_i(x_i + x_{i_2} + \cdots + x_{i_n}) + (x_i + x_{i_2} + \cdots + x_{i_n})(x_{i_2} + \cdots + x_{i_n})}{(x_i + x_{i_2} + \cdots + x_{i_n})^2} \Big/ \binom{N-1}{n-1}$$

$$= \sum_i 1 \Big/ \binom{N-1}{n-1} = 1$$

We conclude that z_i is a decreasing function of x_i as $\frac{\partial z_i}{\partial x_i} = 1 - \left(T_i + x_i\frac{\partial T_i}{\partial x_i}\right) \leq 0.$

Hence, $Q_3 - Q_4$ is $<, =$ or >0 if $\nu(x_i)/x_i$ is increasing, constant, or decreasing function of x_i, respectively. These findings lead to the following theorem (Chaudhuri and Arnab, 1979a,b).

Theorem 6.7.1

Under the superpopulation model M_1^*,

$$E_\xi\left[V\left(\widehat{Y}_{ht}\right)\right] < E_\xi\left[V\left(\widehat{Y}_{rhc}\right)\right] < E_\xi\left[V\left(\widehat{Y}_{lms}\right)\right]$$
$$\text{if } \nu(x_i)/x_i \text{ is an increasing function of } x_i$$

$$E_\xi\left[V\left(\widehat{Y}_{ht}\right)\right] = E_\xi\left[V\left(\widehat{Y}_{rhc}\right)\right] = E_\xi\left[V\left(\widehat{Y}_{lms}\right)\right]$$
$$\text{if } \nu(x_i)/x_i \text{ is a constant}$$

$$E_\xi\left[V\left(\widehat{Y}_{ht}\right)\right] > E_\xi\left[V\left(\widehat{Y}_{rhc}\right)\right] > E_\xi\left[V\left(\widehat{Y}_{lms}\right)\right]$$
$$\text{if } \nu(x_i)/x_i \text{ is a decreasing function of } x_i$$

Corollary 6.7.1

In particular if $\nu(x_i) = \sigma^2 x_i^g$, the theorem above reduces to

$$\begin{aligned}
E_\xi\left[V\left(\widehat{Y}_{ht}\right)\right] &< E_\xi\left[V\left(\widehat{Y}_{rhc}\right)\right] < E_\xi\left[V\left(\widehat{Y}_{lms}\right)\right] \text{ if } g > 1\\
E_\xi\left[V\left(\widehat{Y}_{ht}\right)\right] &= E_\xi\left[V\left(\widehat{Y}_{rhc}\right)\right] = E_\xi\left[V\left(\widehat{Y}_{lms}\right)\right] \text{ if } g = 1\\
E_\xi\left[V\left(\widehat{Y}_{ht}\right)\right] &> E_\xi\left[V\left(\widehat{Y}_{rhc}\right)\right] > E_\xi\left[V\left(\widehat{Y}_{lms}\right)\right] \text{ if } g < 1
\end{aligned}$$

Remark 6.7.1

Rao (1966a) proved that the inequality $E_\xi\left\{V\left(\widehat{Y}_{rhc}\right)\right\} - E_\xi\left\{V\left(\widehat{Y}_{lms}\right)\right\} <, = \text{ or } >1$ above, for values of $g >, =$ or <1, respectively, when both the population and sample sizes N and n are large. But the results by Chaudhuri and Arnab (1979a,b) above are exact.

Remark 6.7.2

Under the models M_1^* with $\nu(x_i) = x_i^g$, Rao (1966a) showed that for a large n $E_\xi MSE\left(\widehat{Y}_R\right) \approx E_\xi MSE\left(\widehat{Y}_{lms}\right)$. Hence using the Corollary 6.7.1, we note that for $g > 1$, both the HT and RHC strategies are more efficient than the ratio strategies based on SRSWOR; whereas for $g < 1$, the ratio strategy fares better than the HT and RHC Strategies. For $g = 1$, all three strategies are equally efficient. Cassel et al. (1977) cited that H_3 is approximately superior, equal, and inferior to H_6 according to $g \geq$ or <1. It is well known that H_5 and H_6 are approximately equal for large N.

6.8 DISCUSSIONS

In a design-based inference or a fixed population setup, the vector of the variable under study **y** is a fixed point in the N-dimensional parametric space R^N. Here, a sample is selected by using a sampling design and expectation of an estimator is the average value of all possible samples that can be selected through the sampling design. It was found that the minimum variance unbiased estimator for estimating a finite population total or mean under a given sampling design p does not exist, and hence one cannot find the best strategy.

In the superpopulation approach, which is also known as model-based or prediction approach, we assume that the vector of the study variable **y** is a realization of a random vector **Y**. The distribution of the random vector **Y** is denoted here by ξ and it is called the superpopulation model or simply model. The superpopulation model may depend on a set of known auxiliary variables involving unknown model parameters. The model-based approach yields the minimum variance unbiased (optimal) estimator in presence of a given sampling design, and it also yields the best sampling strategy in certain classes of strategies. The best strategy very often involves a purposive sampling design. The main demerit of the model-based approach is that its performance is highly dependent on the assumed model. The performance of the optimum strategy becomes poor if the assumed model fails to hold in reality. We call an estimator "robust" if it remains optimum, not only when the assumed model is true, but also in a wider class of models. A balanced sampling design retains optimality of the estimators in wider class of models. It is very difficult to find a balanced design that can control both the bias and the mean-square error together.

In a Bayesian model, the random vector **Y** is assumed to be distributed with one or more unknown parameters and the parameters have some known prior distribution. In Bayesian inference, the posterior distribution of the vector **Y,** given the data, is independent of the sampling design, and hence here the sampling design p has no role in making the inference of any parametric function of **Y**.

The model-design based inference or model-assisted inference is a hybrid of the design-based and model-based inference. Here also the optimum strategies for estimating certain parametric functions such as total or mean are available. The model-assisted strategies are found to be robust in the sense that it provides valid inferences in the event of failure of the assumed model.

Attempts are made to compare performances of the existing popular strategies under various superpopulation models. A good review is given by Cassel et al. (1977). Details of the comparisons of the HH, HT, LMS, and RHC strategies are given under model M_1^*. It is found that each of the strategies HT, LMS, and RHC perform better that than the HH strategy. The HT strategy is the best if $\nu(x)$ is an increasing function of x whereas LMS is the best if $\nu(x)$ is the decreasing function of x. Strategies HT, LMS, and RHC are equally efficient if $\nu(x)$ is proportional to x.

6.9 EXERCISES

6.9.1 Explain the concept of design-based, model-based, and model-assisted inference with reference to a finite population sampling. Show that the design and model unbiasedness imply model design unbiasedness but the converse may not be true.

6.9.2 Find optimum ξ-unbiased linear predictors for the population mean under a given sampling design and the following models, where α, β, γ and σ are unknown but ρ is known.

(a) $E_\xi(y_i) = \alpha + \beta i$, $V_\xi(y_i) = \sigma^2 i$ and $C_\xi(y_i, y_j) = 0$ for $i \neq j$; $i, j = 1, \ldots, N$

(b) $E_\xi(y_i) = \alpha + \beta i + \gamma i^2$, $V_\xi(y_i) = \sigma^2$ and $C_\xi(y_i, y_j) = 0$ for $i \neq j$; $i, j = 1, \ldots, N$

(c) $E_\xi(y_i) = \alpha + \beta x_i$, $V_\xi(y_i) = \sigma^2 x_i^2$ and $C_\xi(y_i, y_j) = \rho\sigma^2 x_i x_j$ for $i \neq j; i, j = 1, \ldots, N$

6.9.3 State the optimality of the HT estimators under different superpopulation models. Derive optimality of the expansion estimator based on an SRSWOR sampling design as a special case.

6.9.4 Consider the superpopulation model $E_\xi(y_i) = \mu$, $V_\xi(y_i) = \sigma^2$ and $C_\xi(y_i, y_j) = 0$.

(i) Derive the optimal estimators for the population mean for a given sampling design p in the class of (a) p-unbiased, (b) ξ-unbiased, and (c) $p\xi$-unbiased predictors.

(ii) Derive optimal sampling strategies under p-unbiased, ξ-unbiased, and $p\xi$-unbiased class predictors, respectively.

6.9.5 Let $y_1, \ldots, y_n$ be independently distributed random variables with $E(y_i) = \mu(\theta)$ and $V(y_i) < \propto$ for $i = 1, \ldots, n$ and let the posterior expectation of $\mu(\theta)$ be given $\mathbf{y} = (y_1, \ldots, y_n)$ is $E\{\mu(\theta)|\mathbf{y}\} = \alpha\bar{y} + \beta$, where α and β are independent of $\mathbf{y}$. Then show that $E\{\mu(\theta)|\mathbf{y}\} = \dfrac{\bar{y}V\{\mu(\theta)\} + mE_\theta V(\bar{y}|\theta)}{V\{\mu(\theta)\} + E_\theta V(\bar{y}|\theta)}$, where $\bar{y} = \sum_{i=1}^{n} y_i/n$ (Ericson, 1969a,b).

6.9.6 Let a finite population $y_1,\ldots, y_n$ be viewed as a sample from a normal superpopulation with unknown mean θ and unknown variance $1/h$. Suppose further that the prior distribution of θ and h have joint density $f(\theta, h) \propto h^{-1}$, $h > 0, -\propto < \theta < \propto$. Obtain the Bayes estimator of the population mean $\overline{Y}$ based on a sample s of n distinct units (Ericson, 1969a,b).

6.9.7 Consider the superpopulation model: $E_\xi(y_i) = \alpha_i + \beta x_i$, $V_\xi(y_i) = \sigma^2 f_i$, $C_\xi(y_i, y_j) = \rho\sigma^2(f_i f_j)^{1/2}$, $i \neq j$; $i, j = 1,\ldots, N$ where α_i, β, x_i, and f_i are known and ρ and σ^2 are unknown. Let C_{lu} be a linear unbiased estimator of the population total Y and $\mathscr{P}_n$ be FESD(n). Then show that

(i) $E_p E_\xi(T - Y)^2 \geq (1-\rho)\sigma^2 \left\{ \left(\sum_{i\in U} f_i^{1/2} \right)^2 \Big/ n - \sum_{i\in U} f_i \right\}$ for any $T \in C_{lu}$ and $p \in \mathscr{P}_n$

(ii) $E_{p_0} E_\xi(T_0 - Y)^2 = (1-\rho)\sigma^2 \left\{ \left(\sum_{i\in U} f_i^{1/2} \right)^2 \Big/ n - \sum_{i\in U} f_i \right\}$

where $T_0 = \sum_{i\in s}(y_i - \alpha_i - \beta x_i)/\pi_{i0} + \sum_{i\in U}(\alpha_i + \beta x_i)$

and p_0 is an FESD(n) with inclusion probability $\pi_i = \pi_{i0} = n f_i^{1/2} \big/ \sum_{i\in U} f_i^{1/2}$ (Tam, 1984).

6.9.8 Let $\widehat{T}_g$ be the ξ-BLUE for the total Y for a given sampling design under the model

$$E_\xi(y_i) = \beta x_i,\ V_\xi(y_i) = \sigma^2 \nu(x_i) \text{ and } C_\xi(y_i, y_j) = 0$$
$$\text{for } i \neq j; i, j = 1, \ldots, N.$$

Then, for $0 \leq g \leq h$, show that (a) $MSE\left(\widehat{T}_g\right) \leq MSE\left(\widehat{T}_h\right)$ if $\nu(x_i)/x_i^g$ is nonincreasing and (b) $MSE\left(\widehat{T}_g\right) \leq MSE\left(\widehat{T}_h\right)$, if $\nu(x_i)/x_i^g$ is nondecreasing (Royall, 1970).