CHAPTER 22

Estimating Functions

22.1 INTRODUCTION

In the classical theory of estimation, we construct an estimator $t(\mathbf{y})$ of a population parameter θ based on a sample observation $\mathbf{y} = (y_1,\ldots, y_n)$. The estimator $t(\mathbf{y})$ is a function of $\mathbf{y}$ only (independent of θ) and expected to satisfy certain optimal criteria such as unbiasedness, sufficiency, and efficiency. The estimator $t(\mathbf{y})$ is constructed using some standard methods of estimation such as maximum likelihood (ML), least squares (LS), methods of moments, minimum chi-square, among others. These methods are ad hoc because no optimal criteria were used in developing them. However, in developing the minimum variance unbiased estimator by Rao–Blackwell and Lehmann–Scheffe approach, the criteria of sufficient statistic were used. The method of LS, ML, method of moments, and minimum chi-square have a similarity in which the estimator $t(\mathbf{y})$ is a solution of θ from the equation of the form $g(\mathbf{y},\theta) = 0$. The method of LS and ML are extensively used in statistical applications. Each method has some advantages and limitations. The method of estimating function (EF) was introduced by Godambe (1960a,b) and Durbin (1960). The EF method is more general since it includes the LS, ML, methods of moments and minimum chi-square methods of estimation. The EF has more strengths and fewer weaknesses than the other methods of estimation (Godambe and Kale, 1991). The EF method has wide applications in the areas of biostatistics, stochastic process, survey sampling, among others. In this chapter we will consider the application of EF only in the area of survey sampling. The confidence intervals in survey sampling will be determined using the framework of EFs.

22.2 ESTIMATING FUNCTION AND ESTIMATING EQUATIONS

Let $\mathbf{y} = (y_1, y_2,\ldots, y_n)$ be a random sample from a population with density function $f(\mathbf{y},\theta)$ with respect to the measure μ, where θ is an unknown parameter that belongs to a known parametric space Ω_θ. The methods of

Survey Sampling Theory and Applications
ISBN 978-0-12-811848-1
http://dx.doi.org/10.1016/B978-0-12-811848-1.00022-4

LS, ML, and minimum chi-square yield equation $g(\mathbf{y},\theta) = 0$, and by solving the equation we get an estimate of θ. If θ is a k-dimension vector, then we get k-independent equations $g_i(\mathbf{y},\theta) = 0$ for $i = 1,\ldots, k$. A function $g(\mathbf{y},\theta)$ of observation $\mathbf{y}$ and θ is called an estimating function if an estimate of θ can be derived from the equation $g(\mathbf{y},\theta) = 0$. The equation $g(\mathbf{y},\theta) = 0$ is called an estimating equation.

Example 22.2.1 (Godame and Kale, 1991)

Let $y_1,\ldots, y_n$ be independent random variables with

$$E(y_i) = \theta \text{ and } Var(y_i) = \sigma^2 \quad \text{for} \quad i = 1, \ldots, n \tag{22.2.1}$$

The LS estimator of θ is obtained by minimizing

$$S^2 = \sum_{i=1}^{n} (y_i - \theta)^2 \tag{22.2.2}$$

with respect to θ.

Now $\dfrac{\partial S^2}{\partial \theta} = 0$ gives $(\overline{y} - \theta) = 0$. If we choose the EF $g(\mathbf{y}, \theta) = g(y_1, \ldots, y_n; \theta) = (\overline{y} - \theta)$, then the estimating equation $g(\mathbf{y},\theta) = 0$ and the LS method provide the same solution of θ as $\overline{y}$. Furthermore, if we assume the distribution of y is normal, then the estimator of θ based on LS, ML, and EF method will be same as $\overline{y}$.

Now consider the class of estimation function G of the form

$$g(\mathbf{y}, \theta) = \sum_{i=1}^{n} (y_i - \theta) b_i \tag{22.2.3}$$

where b_i's are known constants such that $\sum_{i=1}^{n} b_i \neq 0$.

The estimating equation $g(\mathbf{y},\theta) = 0$ yields the solution of θ as

$$\widehat{\theta}_g = \sum_{i=1}^{n} b_i y_i \bigg/ \sum_{i=1}^{n} b_i \tag{22.2.4}$$

The variance of $\widehat{\theta}_g$ is

$$V\left(\widehat{\theta}_g\right) = \sigma^2 \frac{\sum b_i^2}{\left(\sum b_i\right)^2} \tag{22.2.5}$$

Furthermore, if we impose the restriction $\sum_{i=1}^{n} b_i = c$, a constant, then the variance of $g(\mathbf{y},\theta)$ is minimized when $b_i = c/n$ for $i = 1,\ldots, n$. Under this situation, both EF and Gauss–Markov approach lead the $\overline{y}$ as the solution of θ.

An estimation function $g \in G$ is unbiased if $E(g) = 0$.

The EF $g' = kg$ with k as a constant belongs to the class G. Although, the estimating equations $g = 0$ and $g' = 0$ yield the same unbiased estimate $\widehat{\theta}_g$, the variance of the EF g' is $Var(g') = k^2 Var(g)$, which can be made arbitrary small by appropriate choice of the constant k. Hence for the sake of comparability the concept of standardized estimation function has been introduced. A standardized estimation function is defined as

$$g_s = \frac{g}{E\left(\dfrac{\partial g}{\partial \theta}\right)} \tag{22.2.6}$$

Consider a larger class of EFs G_1 consisting of EFs of the form

$$g_1(y, \theta) = \sum b_i(\theta)(y_i - \theta) \tag{22.2.7}$$

where $b_i(\theta)$ is a differentiable function of θ.

The EF $g_1(\mathbf{y},\theta)$ is unbiased, but some solutions of the EF $g_1(\mathbf{y},\theta) = 0$ may not yield unbiased estimators of θ. The standardized EF derived from $g_1(\mathbf{y},\theta)$ is given by

$$g_{1s} = \frac{g_1}{E\left(\dfrac{\partial g_1}{\partial \theta}\right)} = \frac{g_1}{-\sum b_i(\theta)} \tag{22.2.8}$$

The variance of g_{1s} is

$$Var(g_{1s}) = \sigma^2 \frac{\sum b_i^2(\theta)}{\left(\sum b_i(\theta)\right)^2} \tag{22.2.9}$$

The variance $Var(g_{1s})$ is minimized when $b_i(\theta) = b(\theta) \neq 0$ for all $\theta \in \Omega_\theta$. Thus the optimum EF that minimizes the variance of $Var(g_{1s})$ in the class G_1 is

$$g_{10} = b(\theta) \sum (y_i - \theta) \tag{22.2.10}$$

The estimating Eq. (22.2.10) yields the optimal estimator of θ as $\bar{y} = \sum y_i / n$.

22.2.1 Optimal Properties of Estimating Functions

The following regularity conditions on $f(y,\theta)$ and $g(\mathbf{y},\theta)$ were imposed by Godambe (1960a,b) for derivation of the optimal properties of the estimation functions.

Regularity conditions on $f(y,\theta)$ (conditions A):

(i) Ω_θ is an open interval of the real line.

(ii) For almost all y, $\dfrac{\partial \log f(y,\theta)}{\partial\theta}$ and $\dfrac{\partial^2 \log f(y,\theta)}{\partial\theta^2}$ exist for $\forall\,\theta \in \Omega_\theta$.

(iii) $\int f(y,\theta)d\mu$ and $\displaystyle\int \frac{\partial \log f(y,\theta)}{\partial\theta} f(y,\theta)d\mu$ are differentiable under the sign of integration.

(iv) $E\left(\dfrac{\partial \log f(y,\theta)}{\partial\theta}\bigg|\,\theta\right)^2 > 0\,\forall\,\theta \in \Omega_\theta$.

Regularity conditions on $g(\mathbf{y},\theta)$ (conditions B):

(i) $E[g(\mathbf{y},\theta)] = 0\ \forall\,\theta \in \Omega_\theta$

(ii) For almost all $\mathbf{y}$, $\dfrac{\partial g(\mathbf{y},\theta)}{\partial\theta}$ exists $\forall\,\theta \in \Omega_\theta$.

(iii) $\int g(\mathbf{y},\theta) f(\mathbf{y},\theta)\partial\mu$ is differentiable under the sign of integration.

(iv) $E\left[\dfrac{\partial g(\mathbf{y},\theta)}{\partial\theta}\right]^2 \geq 0\ \forall\,\theta \in \Omega_\theta$

The condition (i) of **B** is known as an unbiasedness condition of an EF. If $L(\mathbf{y},\theta)$, the likelihood function of θ is differentiable in θ, then $g(\mathbf{y},\theta) = \dfrac{\partial L(\mathbf{y},\theta)}{\partial\theta}$ is an EF. An EF satisfying all the regularity conditions **B** is known as a regular EF.

Let $\widehat{\theta}_g$ be a solution of $g(\mathbf{y},\theta) = 0$, then using Taylor expansion, we can write

$$g(\mathbf{y},\theta) = g\left(\mathbf{y},\widehat{\theta}_g\right) + \left(\theta - \widehat{\theta}_g\right)\frac{\partial g(\mathbf{y},\theta^*)}{\partial\theta} \tag{22.2.11}$$

where $\theta^* \in \left(\widehat{\theta}_g,\theta\right)$.

Now noting $g\left(\mathbf{y},\widehat{\theta}_g\right) = 0$, we find

$$\theta - \widehat{\theta}_g = -g(\mathbf{y},\theta)/g'(\mathbf{y},\theta^*) \text{ assuming } g'(\mathbf{y},\theta^*) \neq 0 \tag{22.2.12}$$

A good EF should have the difference $\theta - \widehat{\theta}_g$ small. Hence it is desirable that the estimator $\widehat{\theta}_g$ should have

$$E\left(\theta - \widehat{\theta}_g\right)^2 = E(g^2)\Big/\left[E\left(\frac{\partial g}{\partial\theta}\right)\right]^2 \tag{22.2.13}$$

as small as possible. So the criterion

$$\lambda_g(\theta) = \frac{E(g^2)}{\left[E\left(\dfrac{\partial g}{\partial\theta}\right)\right]^2}$$

can be used as a measure of efficiency of an EF g.

Let G be the class of EF satisfying conditions (i) to (v) of B. Godambe (1960a,b) defined the EF $g_0 = g_0(\mathbf{y},\theta)$ as optimum EF in the class G if

$$\frac{E\left(g_0^2\right)}{\left[E\left(\frac{\partial g_0}{\partial \theta}\right)\right]^2} \le \frac{E\left(g^2\right)}{\left[E\left(\frac{\partial g}{\partial \theta}\right)\right]^2} \quad \forall\, \theta \in \Omega_\theta \tag{22.2.14}$$

The equation $g_0(\mathbf{y},\theta) = 0$ is called an optimal estimating equation and the corresponding solution of θ will be called the optimal estimate.

Let $g^s = g/E\left(\frac{\partial g}{\partial \theta}\right)$ be the standardized unbiased EF derived form $g(\in G)$. Then the Eq. (21.1.14) above can be expressed as

$$Var\left(g_0^s\right) \le Var\left(g^s\right) \; \forall g \in G \tag{22.2.15}$$

Under certain regularity conditions, Kale (1962) proved the following inequality analogous to the Rao−Cramer inequality

$$Var(g) \ge \frac{\left\{E\left(\frac{\partial g}{\partial \theta}\right)\right\}^2}{I(\theta)} \quad \forall g \in G \tag{22.2.16}$$

where $I(\theta) = E\left[\frac{\partial \log f(y,\theta)}{\partial \theta}\right]^2$ is the Fisher information.

The equality in Eq. (22.2.16) is attained when

$$g = g^* = \frac{\partial \log f}{\partial \theta}$$

$g^* = \frac{\partial \log f}{\partial \theta}$ is known as score function of the ML.

In particular if $g = t(\mathbf{y}) - \nu(\theta)$, where $t(\mathbf{y})$ is an unbiased estimator of $\nu(\theta)$, then the inequality (22.2.16) reduces to Rao−Cramer inequality

$$Var(t(\mathbf{y})) \ge \frac{\{\nu'(\theta)\}^2}{I(\theta)} \tag{22.2.17}$$

22.3 ESTIMATING FUNCTION FROM SUPERPOPULATION MODEL

Consider a finite population $U = (U_1, \ldots, U_i, \ldots, U_N)$ of N identifiable units, and let y_i be the value of the character under study y associated with the ith unit of the population U. Here we assume that the vector $\mathbf{y} = (y_1, \ldots, y_i, \ldots, y_N)$ is generated from a distribution ξ, which is known to be a member of a class C. The class C is called a superpopulation model. A superpopulation parameter θ is a real-valued function defined on C. Here we will define $g(\mathbf{y},\theta)$ as an unbiased EF for θ if

$$E_\xi\{g(y,\theta(\xi))\} = 0 \ \forall \xi \in C \tag{22.3.1}$$

where E_ξ denotes expectation under ξ.

The solution of the estimating equation $g(\mathbf{y},\theta) = 0$ has two interpretations:

(i) If the vector $\mathbf{y}$ is known, i.e., the population is completely enumerated, then $\theta_N(\mathbf{y})$, the solution of Eq. (22.3.1) is an estimate of the superpopulation parameter θ.

(ii) If the vector $\mathbf{y}$ is partially known, then $\theta_N(\mathbf{y})$ is the parameter of the survey population.

Consider a superpopulation model $C = \{\xi\}$ under which $y_1,\ldots, y_N$ are independent.

An unbiased EF $g^*(y,\theta) \in G$ is said to be optimal in class G and superpopulation $\xi \in C$ under the regularity conditions B stated in Section 22.2.1, if

$$\frac{E_\xi\left(g^{*2}\right)}{\left\{E_\xi\left(\dfrac{\partial g^*}{\partial\theta}\right)\Big|_{\theta=\theta(\xi)}\right\}^2} \le \frac{E_\xi\left(g^2\right)}{\left\{E_\xi\left(\dfrac{\partial g}{\partial\theta}\right)\Big|_{\theta=\theta(\xi)}\right\}^2} \tag{22.3.2}$$

for every $g \in G$ and every $\xi \in C$.

22.3.1 Optimal and Linearly Optimal Estimating Functions

Consider the superpopulation C under which $y_1,\ldots, y_N$ are independent. We define the EF $g_l(\mathbf{y},\theta)$ as linear in $\phi_1,\ldots, \phi_N$ if

$$g_l(\mathbf{y},\theta) = \sum_{i=1}^{N} \phi_i(y_i,\theta)a_i(\theta) \tag{22.3.3}$$

where $a_i(\theta)$ are real differentiable functions of θ and free from y_i's, and ϕ_i are functions satisfying $E_\xi\{\phi_i(y_i,\theta)\} = 0$ for $i = 1,\ldots, N$. An example, one may choose $\phi_i = y_i - \theta$, where $E_\xi(y_i) = \theta$ for $i = 1,\ldots, N$. We will denote the class of linear EF defined in Eq. (22.3.3) as G_L.

The EF $g_l^*(\mathbf{y},\theta)$ is said to be linearly optimal if it belongs to G_L and satisfies Eq. (22.3.2). The equation $g_l^*(y,\theta) = 0$ is called an optimal estimating equation, and the corresponding solution of θ will be called the optimal estimate of θ.

Theorem 22.3.1

For the superpopulation model $C = \{\xi\}$, where $y_1, \ldots, y_N$ are independently distributed, $E_\xi\{\phi_i(y_i, \theta)\} = 0$ for $i = 1, \ldots, N$ and $\xi \in C$, the EF $g_{l0}(\mathbf{y}, \theta) = \sum_{i=1}^{N} \phi_i(y_i, \theta)$ is linearly optimal if

$$E_\xi\left(\frac{\partial \phi_i(y_i, \theta)}{\partial \theta}\Big|_{\theta=\theta(\xi)}\right) = k\{(\theta(\xi)\} E_\xi\left\{\phi_i^2\right\} \quad \text{for } \forall i = 1, \ldots, N \qquad (22.3.4)$$

where $k\{(\theta(\xi)\}$ is a function of $\theta(\xi)$ only.

Proof

$$\begin{aligned}
&E_\xi\left(\frac{g_l}{E_\xi\left(\dfrac{\partial g_l}{\partial \theta}\right)\Big|_{\theta=\theta(\xi)}} - \frac{g_{l0}}{E_\xi\left(\dfrac{\partial g_{l0}}{\partial \theta}\right)\Big|_{\theta=\theta(\xi)}}\right)^2 \\
&= \frac{E_\xi(g_l)^2}{\left\{E_\xi\left(\dfrac{\partial g_l}{\partial \theta}\right)\Big|_{\theta=\theta(\xi)}\right\}^2} + \frac{E_\xi(g_{l0})^2}{\left\{E_\xi\left(\dfrac{\partial g_{l0}}{\partial \theta}\right)\Big|_{\theta=\theta(\xi)}\right\}^2} \\
&\quad - 2\frac{E_\xi(g_l g_{l0})}{\left\{E_\xi\left(\dfrac{\partial g_l}{\partial \theta}\right)\Big|_{\theta=\theta(\xi)}\right\}\left\{E_\xi\left(\dfrac{\partial g_{l0}}{\partial \theta}\right)\Big|_{\theta=\theta(\xi)}\right\}} \\
&\geq 0
\end{aligned} \qquad (22.3.5)$$

Now using the condition $E_\xi\left(\dfrac{\partial \phi_i(y_i, \theta)}{\partial \theta}\Big|_{\theta=\theta(\xi)}\right) = k(\theta(\xi)) E_\xi\{\phi_i(y_i, \theta)\}^2$, we find

$$\begin{aligned}
&\frac{E_\xi(g_l g_{l0})}{\left\{E_\xi\left(\dfrac{\partial g_l}{\partial \theta}\right)\Big|_{\theta=\theta(\xi)}\right\}\left\{E_\xi\left(\dfrac{\partial g_{l0}}{\partial \theta}\right)\Big|_{\theta=\theta(\xi)}\right\}} \\
&= \frac{\sum_{i=1}^{N} a_i(\theta) E_\xi\left(\phi_i^2\right)}{\{k(\theta(\xi))\}^2 \left\{\sum_{i=1}^{N} a_i(\theta) E_\xi\left(\phi_i^2\right)\right\}\left\{\sum_{i=1}^{N} E_\xi\left(\phi_i^2\right)\right\}} \qquad (22.3.6) \\
&= \frac{1}{\{k(\theta(\xi))\}^2 \left\{\sum_{i=1}^{N} E_\xi\left(\phi_i^2\right)\right\}}
\end{aligned}$$

and

$$\frac{E_\xi(g_{l0})^2}{\left\{E_\xi\left(\frac{\partial g_{l0}}{\partial\theta}\right)\Big|_{\theta=\theta(\xi)}\right\}^2}=\frac{1}{\{k(\theta(\xi))\}^2\left(\sum_{i=1}^{N}E_\xi\left(\phi_i^2\right)\right)} \tag{22.3.7}$$

Substituting Eqs. (22.3.6) and (22.3.7) in Eq. (22.3.5), we get

$$E_\xi\left(\frac{g_l}{E_\xi\left(\frac{\partial g_l}{\partial\theta}\right)\Big|_{\theta=\theta(\xi)}}\right)^2\geq E_\xi\left(\frac{g_{l0}}{E_\xi\left(\frac{\partial g_{l0}}{\partial\theta}\right)\Big|_{\theta=\theta(\xi)}}\right)^2$$

Corollary 22.3.1

If $y_1,\ldots, y_N$ are independently and identically distributed and $\phi_i=\phi$ for $i=1,\ldots, N$, then $g_{l0}(\mathbf{y},\theta)=\sum_{i=1}^{N}\phi(y_i,\theta)$ is an optimal EF for θ(Godambe and Thompson, 1978, 1984) if

$$E_\xi\left(\frac{\partial\phi(y_i,\theta)}{\partial\theta}\Big|_{\theta=\theta(\xi)}\right)=k\{\theta(\xi)\}(E_\xi\{\phi(y_i,\theta)\}^2$$

Remark 22.3.1

The different choices of ϕ_i yield different survey parameters. For example, the choice $\phi_i(\mathbf{y}_i,\theta)=y_i-\theta$, the estimating equation $\sum_{i=1}^{N}\phi_i(y_i,\theta)=0$, yields $\theta_N=\overline{Y}=\sum_{i=1}^{N}y_i/N$, the mean of survey population. Similarly, for the choice $\phi_i(y_i,\theta)=(y_i-\theta x_i)$, θ_N becomes $\sum_{i=1}^{N}y_i\Big/\sum_{i=1}^{N}x_i$, the population ratio. Again, if we choose $\phi_i(y_i,\theta)=\varphi_i(y_i,\theta)-p$ with $\varphi_i(y_i,\theta)=1$ for $y_i\leq\theta$ and $\varphi_i(y_i,\theta)=0$ otherwise, then the $\sum_{i=1}^{N}\phi_i(y_i,\theta)=0$ yields θ_N as pth quantile of the survey population.

Example 22.3.1

Suppose that $y_1,\ldots, y_N$ are independent with $E_\xi(y_i)=\theta(\xi)x_i$ and $V_\xi(y_i)=\sigma_i^2$. Then condition (22.3.4) of Theorem 22.3.1 holds for $\phi_i=\frac{x_i(y_i-\theta x_i)}{\sigma_i^2}$. Hence, $g_{l0}(\mathbf{y},\theta)=\sum_{i=1}^{N}\frac{x_i(y_i-\theta x_i)}{\sigma_i^2}$ is linearly optimal and $\theta_N = \dfrac{\sum\limits_{i=1}^{N} y_i x_i / \sigma_i^2}{\sum\limits_{i=1}^{N} x_i^2 / \sigma_i^2}$ is the optimal estimate of θ as well as the parameter of the survey population (Chaudhuri and Stenger, 1992).

Example 22.3.2

In Example 22.3.1 above, if we assume $V_\xi(y_i) = \sigma^2$ (constant), then $\theta_N = \frac{\sum_{i=1}^{N} y_i x_i}{\sum_{i=1}^{N} x_i^2}$ is the optimal estimate of θ as well as the parameter of the survey population (Ghosh, 1991).

Example 22.3.3

In Example 22.3.2 if $x_i = 1$ for $i = 1,\ldots, N$, then $g_{l0}(\mathbf{y}, \theta) = \sum\limits_{i=1}^{N} (y_i - \theta)$ is an optimal EF of θ and $\theta_N = \sum\limits_{i=1}^{N} y_i / N$ becomes the optimal estimate of θ and also the survey parameter (Godambe and Thompson, 1986a,b).

22.4 ESTIMATING FUNCTION FOR A SURVEY POPULATION

Suppose a sample s of size n is selected from a finite population U of size N with probability $p(s)$ using a sampling design p. Let the inclusion probability for the ith unit be $\pi_i = \sum\limits_{s \supset i} p(s) > 0$ for $i = 1,\ldots, N$. Furthermore, let $y_1,\ldots, y_N$ be distributed independently as in Theorem 22.3.1 and θ_N be the solution of the linearly optimal EF $g_{l0}(\mathbf{y}, \theta) = \sum\limits_{i=1}^{N} \phi_i(y_i, \theta)$. Since y_i for $i \notin s$ are not known, we construct an EF $h(d,\theta)$ as a function of the collected data $d = (y_i, i \in s)$. Such a function $h(d,\theta)$ will be called a sample EF. The solution $\widehat{\theta}_s$ of the sample estimating equation $h(d,\theta) = 0$ provides an estimate of the survey parameter θ_N. The function $h(d,\theta)$ is said to be design unbiased of $\sum\limits_{i=1}^{N} \phi_i(y_i, \theta)$ if

$$E_p[h(d, \theta)] = \sum_s h(d, \theta) p(s) = \sum_{i=1}^{N} \phi_i(y_i, \theta) \tag{22.4.1}$$

for each population vector $\mathbf{y} = (y_1,\ldots, y_N)$ and θ, where E_p denotes expectation with respect to the sampling design p. The class of unbiased EFs $h(d,\theta)$ that satisfy Eq. (22.4.1) will be denoted by C_u. The function

$$h^*(d, \theta) = \sum_{i \in s} \frac{\phi_i(y_i, \theta)}{\pi_i} \tag{22.4.2}$$

is an unbiased EF of $\sum\limits_{i=1}^{N} \phi_i(y_i, \theta)$.

Godambe and Thompson (1986a,b) defined an EF $h(d,\theta)$ satisfying Eq. (22.4.1) as optimum if it minimizes

$$E_\xi E_p\{h^2(d,\theta)\} \Big/ \left\{E_\xi E_p\left(\frac{\partial h(d,\theta)}{\partial\theta}\right)\right\}^2 \tag{22.4.3}$$

Here we note that

$$E_\xi E_p\left(\frac{\partial h(d,\theta)}{\partial\theta}\right) = E_\xi\left(\frac{\partial}{\partial\theta}\sum_{i=1}^{N}\phi_i\right) \tag{22.4.4}$$

as

$$E_\xi E_p\left(\frac{\partial h(d,\theta)}{\partial\theta}\right) = E_\xi\left\{\sum_s \frac{\partial h(d,\theta)}{\partial\theta}p(s)\right\} = E_\xi\left\{\frac{\partial}{\partial\theta}E_p\left(\sum_s h(d,\theta)\right)\right\} =$$
$$E_\xi\left(\frac{\partial}{\partial\theta}\sum_{i=1}^{N}\phi_i\right)$$

Since $E_\xi\left(\frac{\partial}{\partial\theta}\sum_{i=1}^{N}\phi_i\right)$ is independent of h, our problem is to find an $h(d,\theta)$, which minimizes $E_\xi E_p\{h^2(d,\theta)\}$ subject to $E_p[h(d,\theta)] = \sum_{i=1}^{N}\phi_i(y_i,\theta)$. The following theorem (Godambe and Thompson, 1986a,b) establishes optimality of the EF $h^*(d,\theta)$ given in Eq. (22.4.2).

Theorem 22.4.1

For the superpopulation model $C = \{\xi\}$ where $\phi_i(y_i,\theta)$ are independently distributed with $E_\xi\{\phi_i(y_i,\theta(\xi))\} = 0$ for $i = 1,\ldots, N$, the EF $h^*(d,\theta) = \sum_{i\in s}\frac{\phi_i(y_i,\theta)}{\pi_i}$ is the optimal (minimizes Eq. 22.4.3) in the class of unbiased EFs C_u (satisfying Eq. 22.4.1).

Proof

Let

$$Q(d,\theta) = h(d,\theta) + h^*(d,\theta) \tag{22.4.5}$$

The unbiasedness condition (22.4.1) yields

$$E_p[Q(d,\theta)] = \sum_s Q(d,\theta)p(s) = 0 \tag{22.4.6}$$

Furthermore,

$$E_p[h(d,\theta)]^2 = E_p[h^*(d,\theta)]^2 + E_p[Q(d,\theta)^2 + 2E_p[h^*(d,\theta)Q(d,\theta)] \tag{22.4.7}$$

and

$$E_\xi E_p[h^*(d,\theta)Q(d,\theta)] = E_\xi E_p\left[Q(d,\theta)\sum_{i\in s}\frac{\phi_i(y_i,\theta)}{\pi_i}\right]$$

$$= E_\xi \sum_s \left[Q(d,\theta)\sum_{i\in s}\frac{\phi_i(y_i,\theta)}{\pi_i}\right]p(s)$$

$$= E_\xi \sum_{i=1}^N \frac{\phi_i(y_i,\theta)}{\pi_i}\sum_{s\supset i} Q(d,\theta)p(s)$$

$$= -E_\xi \sum_{i=1}^N \frac{\phi_i(y_i,\theta)}{\pi_i}\sum_{\bar{s}_i} Q(d,\theta)p(s)$$

(noting $E_p[Q(d,\theta)] = \sum_{s\supset i} Q(d,\theta)p(s) + \sum_{\bar{s}_i} Q(d,\theta)p(s) = 0$, where $\sum_{\bar{s}_i}$ denotes sum over the samples that do not contain i)

Since $\phi_i(y_i,\theta)$ are independent and $E_\xi\{\phi_i(y_i,\theta)\} = 0$, we have

$$E_\xi E_p[h^*(d,\theta)Q(d,\theta)] = -\sum_{i=1}^N E_\xi \frac{\phi_i(y_i,\theta)}{\pi_i}\sum_{\bar{s}_i} E_\xi[Q(d,\theta)]p(s) = 0 \tag{22.4.8}$$

Finally, Eqs. (22.4.7) and (22.4.8) yield

$$E_p[h(d,\theta)]^2 \geq E_p[h^*(d,\theta)]^2$$

Hence the theorem.

For a fixed effective sample size n[FES(n)] design, $\sum_{i=1}^N \pi_i = n$ and the expression

$$E_\xi E_p[h^*(d,\theta)]^2 = \sum_s \sum_{i\in s} \frac{E_\xi[\phi_i(y_i,\theta)]^2}{\pi_i^2}p(s) = \sum_{i=1}^N \frac{E_\xi[\phi_i(y_i,\theta)]^2}{\pi_i}$$

is minimized when $\pi_i \propto \sqrt{E_\xi \phi_i^2(y_i,\theta)}$ and the minimum value of $E_\xi E_p[h^*(d,\theta)]^2$ is

$$\left(\sum_{i=1}^N \sqrt{E_\xi[\phi_i(y_i,\theta)]^2}\right)^2 \Big/ n \tag{22.4.9}$$

Thus we have the following as a corollary of Theorem 22.3.1.

Corollary 22.4.1

For the model ξ where $\phi_i(y_i,\theta)$ are independently distributed with $E_\xi\{\phi_i(y_i,\theta)\} = 0$, $i = 1,\ldots, N$, the sampling design p_0 with

$\pi_i = \pi_{i0} = n\sqrt{E_\xi\left(\phi_i^2(y_i,\theta)\right)}\Big/\sum_{i=1}^{N}\sqrt{E_\xi\left(\phi_i^2(y_i,\theta)\right)}$, is the optimal in the class $\mathcal{P}_n$ of FES(n) designs in the sense

$$E_\xi E_p(h(d,\theta))^2 \geq E_\xi E_{p_0}\left(\sum_{i\in s}\frac{\phi_i(y_i,\theta)}{\pi_{i0}}\right)^2$$
$$= \left(\sum_{i=1}^{N}\sqrt{E_\xi[\phi_i(y_i,\theta)]^2}\right)^2\Big/ n \ \forall p\in\mathcal{P}_n$$

Remark 22.4.1

The estimator $\widehat{\theta}_s$ of the survey parameter θ_N obtained from the optimal unbiased EF $h^*(d,\theta) = 0$ is called an optimal estimator. The estimator $\widehat{\theta}_s$ may not be design unbiased for θ_N. The optimality of $h^*(d,\theta)$ is independent of the variance structure $E_\xi\{\phi_i(y,\theta)\}^2$.

Example 22.4.1

Consider the superpopulation model

$$y_i = \theta(\xi)x_i + \in_i \tag{22.4.10}$$

where $\in_i$'s are independent with $E_\xi(\in_i) = 0$, $V_\xi(\in_i) = \sigma_i^2$. The constants x_i's and σ_i^2's are positive and known. Since $\phi_i(y_i,\theta) = \dfrac{(y_i - \theta x_i)x_i}{\sigma_i^2}$ satisfies the condition (22.3.4), the linearly optimal EF is

$$g_{l0}(\mathbf{y},\theta) = \sum_{i=1}^{N}\frac{x_i(y_i-\theta x_i)}{\sigma_i^2}$$

as $\phi_i(y_i,\theta) = \dfrac{(y_i - \theta x_i)x_i}{\sigma_i^2}$ satisfies the condition (22.3.4). In this case the survey parameter is

$$\theta_N = \frac{\sum_{i=1}^{N} y_i x_i/\sigma_i^2}{\sum_{i=1}^{N} x_i^2/\sigma_i^2} \tag{22.4.11}$$

The survey estimator of θ_N is obtained from the optimum sample estimating equation $\sum_{i\in s}\dfrac{x_i(y_i - \theta x_i)}{\pi_i\sigma_i^2} = 0$ and it is given by

$$\widehat{\theta}_s = \frac{\sum_{i\in s} y_i x_i/\left(\pi_i\sigma_i^2\right)}{\sum_{i\in s} x_i^2/\left(\pi_i\sigma_i^2\right)} \tag{22.4.12}$$

In this case the inclusion probability for the optimum sampling design p_0 is given by

$$\pi_{i0} = \frac{nx_i/\sigma_i}{\sum_{i=1}^{N} x_i/\sigma_i} \tag{22.4.13}$$

For practical importance, consider the following special cases:

Case 1: $\sigma_i^2 \propto x_i$ yields $\theta_N = \sum_{i=1}^{N} y_i \Big/ \sum_{i=1}^{N} x_i$, $\widehat{\theta}_s = \dfrac{\sum_{i \in s} y_i/\pi_i}{\sum_{i \in s} x_i/\pi_i}$, and $\pi_{i0} = \dfrac{n\sqrt{x_i}}{\sum_{i=1}^{N} \sqrt{x_i}}$.

Case 2: $\sigma_i^2 \propto x_i^2$ yields $\theta_N = \dfrac{1}{N}\sum_{i=1}^{N} \dfrac{y_i}{x_i}$, $\widehat{\theta}_s = \dfrac{\sum_{i \in s} y_i/(\pi_i x_i)}{\sum_{i \in s} 1/\pi_i}$, and $\pi_{i0} = n/N$.

Case 3: $\sigma_i^2 = \sigma^2$ and $x_i = 1$ for $i = 1, \ldots, N$ yield $\theta_N = \overline{Y}$, $\widehat{\theta}_s = \dfrac{\sum_{i \in s} y_i/\pi_i}{\sum_{i \in s} 1/\pi_i}$, and $\pi_{i0} = n/N$. The estimator $\widehat{\theta}_s = \dfrac{\sum_{i \in s} y_i/\pi_i}{\sum_{i \in s} 1/\pi_i}$ was proposed by Brewer (1963a,b) and Hájek (1971).

Example 22.4.2

Consider the model

$$y_i = \theta + \in_i \tag{23.3.14}$$

where $\in_i$'s are independent with $E_\xi(\in_i) = 0$ and $V_\xi(\in_i) = \sigma_i^2$.

Substituting $x_1 = \cdots = x_N = 1$, in Example 23.4.1, we get

$$\theta_N = \frac{\sum_{i=1}^{N} y_i/\sigma_i^2}{\sum_{i=1}^{N} 1/\sigma_i^2},\ \widehat{\theta}_s = \frac{\sum_{i \in s} y_i/\left(\pi_i\sigma_i^2\right)}{\sum_{i \in s} 1/\left(\pi_i\sigma_i^2\right)},\ \text{and}\ \pi_{i0} = \frac{n/\sigma_i}{\sum_{i=1}^{N} 1/\sigma_i}.$$

Example 22.4.3

Consider the multiple regression model given in Mantel (1991), where

(i) y_i's are independent.

(ii) $E_\xi(y_i) = \mathbf{x}_i\boldsymbol{\beta}$, where $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})$, is a known vector of auxiliary information and $\boldsymbol{\beta}' = \left(\beta_1, \ldots, \beta_p\right)$ is a vector of unknown parameters.

(iii) $V_\xi(y_i) = \sigma^2 v_i$, where v_i is known but σ^2 is unknown.

Following Theorem 22.3.1, the linearly optimal estimating equation comes out as

$$g_{l0}(\mathbf{y},\theta)=\sum_{i=1}^{N}(y_i-\mathbf{x}_i\boldsymbol{\beta})\mathbf{x}_i'/v_i=0 \tag{22.4.15}$$

Eq. (22.4.15) yields the survey parameter as

$$\boldsymbol{\beta}_{\mathbf{N}}=\left(\mathbf{X}_N'\mathbf{V}_N^{-1}\mathbf{X}_N\right)^{-1}\mathbf{X}_N'\mathbf{V}_N^{-1}\mathbf{y}_N \tag{22.4.16}$$

where $\mathbf{y}_N'=(y_1,\ldots,y_N)$ and $\mathbf{V}_N=diag(v_1,\ldots,v_N)$; $\mathbf{X}_N$ is a matrix of order $N\times p$ with ith row, $\mathbf{x}_i$.

The survey estimator of $\boldsymbol{\beta}_N$ is found from the optimum sample estimating equation

$$h(d,\theta)=\sum_{i\in s}(y_i-\mathbf{x}_i\boldsymbol{\beta})\mathbf{x}_i'/(\pi_i v_i)=0 \tag{22.4.17}$$

to be

$$\widehat{\boldsymbol{\beta}}_s=\left(\mathbf{X}_s'\Pi_s^{-1}\mathbf{V}_s^{-1}\mathbf{X}_s\right)^{-1}\mathbf{X}_s'\Pi_s^{-1}\mathbf{V}_s^{-1}\mathbf{y}_s \tag{22.4.18}$$

where $\mathbf{y}_s$ is the vector of y_i for $i\in s$; Π_s and $\mathbf{V}_s$ are diagonal matrices with ith diagonal element π_i and v_i, respectively, for $i\in s$; and $\mathbf{X}_s$ is a matrix with ith row $\mathbf{x}_i$ for $i\in s$.

For estimating the finite population mean $\overline{Y}_N=\sum_{i=1}^{N}y_i/N$, we may consider the following generalized regression (*greg*) predictor proposed by Cassel et al. (1977) as

$$\overline{y}_{greg}=\overline{\mathbf{X}}_N'\widehat{\boldsymbol{\beta}}_s+\mathbf{1}_s'\Pi_s^{-1}\widehat{\mathbf{e}}_s/N \tag{22.4.19}$$

where $\widehat{\mathbf{e}}_s=\mathbf{y}_s-\mathbf{X}_s\widehat{\boldsymbol{\beta}}_s$ and I_s' is a vector of 1s whose length is the size of the sample s.

Mantel (1991) proposed an alternative estimator (*areg*) for $\overline{Y}_N$ as

$$\overline{y}_{areg}=\overline{\mathbf{X}}_N'\widehat{\boldsymbol{\beta}}_s+(c_1/c_2)\mathbf{1}_s'\Pi_s^{-1}\widehat{\mathbf{e}}_s/N \tag{22.4.20}$$

where $c_1=\mathbf{I}_N'\left(\mathbf{V}_N\mathbf{I}_N-\mathbf{X}_N\left(\mathbf{X}_s'\mathbf{V}_s^{-1}\boldsymbol{\Pi}_s^{-1}\mathbf{X}_s\right)^{-1}\mathbf{X}_s'\boldsymbol{\Pi}_s^{-1}\mathbf{I}_s\right)$,
$c_2=\mathbf{I}_s'\boldsymbol{\Pi}_s^{-1}\left(\mathbf{V}_s\mathbf{I}_s-\mathbf{X}_s\left(\mathbf{X}_s'\mathbf{V}_s^{-1}\boldsymbol{\Pi}_s^{-1}\mathbf{X}_s\right)^{-1}\mathbf{X}_s'\boldsymbol{\Pi}_s^{-1}\mathbf{I}_s\right)$, and $\mathbf{I}_N$ is a column vector of 1s.

The estimator $\overline{y}_{areg}$ is very similar to $\overline{y}_{greg}$ except for an adjusted weight c_1/c_2 in the second part. The estimator $\overline{y}_{areg}$ is more design based whereas $\overline{y}_{areg}$ is more model based. However, both the estimators are design consistent. Further details are given by Mantel (1991).

22.5 INTERVAL ESTIMATION

In survey sampling, the confidence interval of a parameter θ is obtained by inverting the pivotal $\left(\widehat{\theta}-\theta\right)\Big/\sqrt{\widehat{V}\left(\widehat{\theta}\right)}$, where $\widehat{\theta}$ is a suitable estimator of θ and $\widehat{V}\left(\widehat{\theta}\right)$ is an estimated variance of $\widehat{\theta}$. Godambe and Thompson (1999) provided a more direct method of constructing a pivotal quantity to find the confidence interval. This alternative method provides better confidence intervals than the conventional method. In this section we will mainly discuss the methods of determination of confidence intervals cited by Godambe (1991) and Godambe and Thompson (1999).

22.5.1 Confidence Interval for θ

Let $\phi_i(y_i,\theta)$ be independent elementary EF with

$$E_{\xi}[\phi_i(y_i,\theta)]=0 \text{ and } V_{\xi}[\phi_i(y_i,\theta)]=\sigma_i^2 \quad \text{for } i\in s \tag{22.5.1}$$

and $T(s,\theta)=\sum\limits_{i\in s} g_i(y,\theta)$ be an unbiased sample EF, which satisfies

$$E_p\{T(s,\theta)\}=\sum_{i=1}^{N}\phi_i(y_i,\theta)$$

The EF $T(s,\theta)$ may or may not be the optimum EF for estimating θ. An approximate $100(1-\alpha)\%$ confidence interval for θ under the super-population model (22.5.1) can be obtained by inverting

$$\left|\frac{\sum\limits_{i\in s} g_i(y_i,\theta)}{\sqrt{\sum\limits_{i\in s} g_i^2(y_i,\theta)}}\right|=z_{1-\alpha/2} \tag{22.5.2}$$

where $z_{1-\alpha/2}$ is the $(1-\alpha/2)$ quantile of the $N(0,1)$ distribution.

22.5.1.1 Confidence Interval for Survey Parameter θ_N

Let θ_N be the survey parameter obtained by solving the estimating equation $\sum\limits_{i=1}^{N}\phi_i(y_i,\theta)=0$. The confidence interval of θ_N can be obtained using any of the following methods.

Method 1: Confidence interval of θ_N is obtained by inverting

$$\left|\frac{\sum\limits_{i\in s} g_i(y_i,\theta_N)}{\sqrt{\widehat{V}_{\xi}\{T(s,\theta_N)\}}}\right|=z_{1-\alpha/2} \tag{22.5.3}$$

where $\widehat{V}_\xi\{T(s,\theta)\}$ is a model-unbiased estimator of the model variance $V_\xi\{T(s,\theta)\}$, i.e., $E_\xi\left[\widehat{V}_\xi\{T(s,\theta)\}\right] = V_\xi\{T(s,\theta)\}$.

Method 2: By inverting

$$\left|\frac{\sum_{i\in s} g_i(y_i,\theta_N)}{\sqrt{\widehat{V}_p\{T(s,\theta_N)\}}}\right| = z_{1-\alpha/2} \tag{22.5.4}$$

where $\widehat{V}_p\{T(s,\theta_N)\}$ is a design-unbiased estimator of the design variance $V_p\{T(s,\theta_N)\}$, i.e.,

$$E_p\left[\widehat{V}_p\{T(s,\theta)\}\right] = E_p\left\{\sum_{i\in s} g_i(y_i,\theta) - \sum_{i=1}^{N}\phi_i(y_i,\theta)\right\}^2.$$

The confidence intervals derived from the two methods above are expected to be close to one another if the chosen sampling design and superpopulation model are appropriate.

Example 22.5.1 (Godambe and Thompson, 2009)

Suppose that $y_1,\ldots, y_N$ are independent with $E_\xi(y_i) = \theta x_i$ and $V_\xi(y_i) = \sigma^2$. Let $\phi_i = y_i - \theta x_i$, then the estimating equation $\sum_{i=1}^{N}\phi_i(y_i,\theta) = 0$ gives $\theta = \theta_N = Y/X = R$, the population ratio. Let a sample of size n be selected by the simple random sampling without replacement (SRSWOR) method. Consider a sample estimating equation $T(s,R) = \sum_{i\in s}(y_i - Rx_i)$, which yields $\widehat{R}_s = \sum_{i\in s} y_i \Big/ \sum_{i\in s} x_i$ as an estimator of R. The design variance of $T(s,R)$ is $V_p[T(s,R)] = n^2\left(\frac{1}{n}-\frac{1}{N}\right)\frac{1}{N-1}\sum_{i=1}^{N}\left(z_i-\overline{Z}\right)^2$, where $z_i = y_i - Rx_i$. An approximate design-unbiased estimator for $V_p[T(s,R)]$ is $\widehat{V}_p\left[T\left(s,\widehat{R}_s\right)\right] = n^2\left(\frac{1}{n}-\frac{1}{N}\right)\frac{1}{n-1}\sum_{i\in s}\widehat{z}_i^2$, where $\widehat{z}_i = y_i - \widehat{R}_s x_i$. The $100(1-\alpha)\%$ confidence interval of R is obtained from the equation

$$\frac{\sum_{i\in s}(y_i - Rx_i)}{\sqrt{\left(1-\frac{n}{N}\right)\frac{n}{n-1}\sum_{i\in s}\left(y_j - \widehat{R}_s x_j\right)^2}} = \pm\, z_{1-\alpha/2} \tag{22.5.5}$$

Eq. (22.5.5) yields an approximate confidence interval for R as

$$\widehat{R}_s \pm z_{1-\alpha/2}\sqrt{\left(1-\frac{n}{N}\right)\frac{1}{n}\left(s_y^2 - 2\widehat{R}_s s_{xy} + \widehat{R}_s^2 s_x^2\right)}\Big/\overline{x}_s \tag{22.5.6}$$

where $(n-1)s_x^2 = \sum_{i\in s}(x_i - \overline{x}_s)^2$, $(n-1)s_y^2 = \sum_{i\in s}(y_i - \overline{y}_s)^2$ and $(n-1)s_{xy} = \sum_{i\in s}(x_i - \overline{x}_s)(y_i - \overline{y}_s)$.

22.5.1.2 Stratified Sampling

Consider a population U of N units classified into H strata and let N_h be the size of the hth stratum. From the hth stratum, a sample s_h of size n_h is selected by SRSWOR method. Let y_{hi} be the value of the study variable y of the ith unit of the hth stratum $i = 1,\ldots, N_h$; $h = 1,\ldots, H$. Suppose that $y_{hi}'s$ are distributed independently with

$$E_\xi(y_{hj}) = \theta \text{ and } V_\xi(y_{hj}) = \sigma_h^2 \quad \text{for } i = 1,\ldots,N_h; h = 1,\ldots,H \tag{22.5.7}$$

Consider the elementary EF $\phi_{hj}(y_{ij},\theta) = y_{hj} - \theta$, which yields $\theta_N = \overline{Y} = \sum_{h=1}^{H}\sum_{j=1}^{N_h} y_{hj}/N$ as the solution of the estimating equation $\frac{1}{N}\sum_{h=1}^{H}\sum_{j=1}^{N_h}\phi_{hj}(y_{hj},\theta) = 0$. Noting that the inclusion probability for the jth unit of the hth stratum under SRSWOR sampling is n_h/N_h, we find the optimal sample EF is

$$T = \sum_{h=1}^{H}\frac{W_h}{n_h}\sum_{j\in s_h}\phi_{hj}\left(y_{hj},\overline{Y}\right) = \sum_{h=1}^{H}\frac{W_h}{n_h}\sum_{j\in s_h}\left(y_{hj} - \overline{Y}\right) \tag{22.5.8}$$

where $W_h = N_h/N$.

Thus the optimal estimate of $\overline{Y}$ is

$$\widehat{\theta}_s = \sum_{h=1}^{H} W_h\overline{y}_h = \overline{y}_{st}$$

where $\overline{y}_h = \sum_{j\in s_h} y_{hj}/n_h$.

The design variance of T is

$$V_p(T) = \sum_{h=1}^{H} W_h^2(1-f_h)S_{hi}^2/n_h \tag{22.5.9}$$

where $f_h = n_h/N_h$, $S_{hi}^2 = \sum_{j=1}^{N_h}\left(y_{hj} - \overline{Y}_h\right)^2/(N_h - 1)$ and $\overline{Y}_h = \sum_{j=1}^{N_h} y_{hj}/N_h$.

The model $E_\xi(y_{ij}) = \theta$ suggests that the strata means $\overline{Y}_h$ should be approximately equal to $\overline{Y}$ when the strata sizes N_h are large. Hence, we find a design-unbiased estimate of $V_p(T)$ as

$$\begin{aligned}\widehat{V}_p(T) &= \sum_{h=1}^{H} W_h^2 \frac{(1-f_h)N_h}{n_h(N_h-1)} \sum_{j\in s_h}\left(y_{hj}-\overline{Y}\right)^2 \Big/ n_h \\ &= \widehat{V}_0 + Q\end{aligned} \tag{22.5.10}$$

where $\widehat{V}_0 = \sum_{h=1}^{H} W_h^2(1-f_h)s_h^2/n_h$, $s_h^2 = \sum_{j\in s_h}\left(y_{hj}-\overline{y}_h\right)^2 \Big/ (n_h-1)$ and Q is of the order $O\left(1/n_h^2\right)$. When n_h's are large for every h, we can neglect Q and assume $T\Big/\sqrt{\widehat{V}_0} \sim N(0,1)$ and find the confidence interval of $\overline{Y}$ by inverting $T\Big/\sqrt{\widehat{V}_0}$. On the other hand if all n_h's are not large, we can find the confidence interval of $\overline{Y}$ by inverting the distribution of $T\Big/\sqrt{\widehat{V}_p(T)}$, which is asymptotically $N(0,1)$.

The model variance $V_\xi(T)$ can by estimated by

$$\widehat{V}_\xi(T) = \sum_{h=1}^{H} \frac{W_h^2}{n_h^2} \sum_{j\in s_h}\left(y_{hj}-\overline{Y}\right)^2 = \widehat{V}_\xi \tag{22.5.11}$$

Hence the confidence interval of $\overline{Y}$ can be obtained by inverting $T\Big/\sqrt{\widehat{V}_\xi}$, which is asymptotically $N(0,1)$. Consider the superpopulation model

$$E_\xi(y_{hj}) = \theta x_{hj} \text{ and } V_\xi(y_{hj}) = \sigma_h^2 \text{ for } j=1,\ldots,N_h; h=1,\ldots,H \tag{22.5.12}$$

Here the EF

$$\frac{1}{N}\sum_{h=1}^{H}\sum_{j=1}^{N_h}(y_{hj}-\theta x_{hj}) \tag{22.5.13}$$

yields the population ratio as the survey parameter, which is given by

$$\theta_N = \frac{\sum_{h=1}^{H}\sum_{j=1}^{N_h} y_{hj}}{\sum_{h=1}^{H}\sum_{j=1}^{N_h} x_{hj}} = \frac{\sum_{h=1}^{H} Y_h}{\sum_{h=1}^{H} X_h} = \frac{Y}{Y} = R \tag{22.5.14}$$

where $Y = \sum_{h=1}^{H} Y_h$, $Y_h = \sum_{j=1}^{N_h} y_{hj}$, $X = \sum_{h=1}^{H} X_h$, and $X_h = \sum_{j=1}^{N_h} x_{hj}$.

The optimal sample EF for estimating R is

$$T = \sum_{h=1}^{H} \frac{W_h}{n_h} \sum_{j=1}^{n_h} (y_{hj} - R\, x_{hj}) \tag{22.5.15}$$

The EF (22.5.15) yields the estimate of R as. $\widehat{R}_s = \sum_{h=1}^{H} W_h \overline{y}_h \Big/ \sum_{h=1}^{H} W_h \overline{x}_h$

For the model (22.5.12) with large strata sizes N_h, we may assume that $R = Y_h/X_h$ and neglect the differences $Y_h - RX_h$. In this case an approximate estimator of the design variance of T is given by

$$\widehat{V}_p(T) = \sum_{h=1}^{H} \frac{W_h^2(1 - f_h)}{n_h} \frac{N_h}{N_h - 1} \frac{1}{n_h} \sum_{j \in s_h} (y_{hj} - Rx_{hj})^2 \tag{22.5.16}$$

Finally the confidence interval of R is obtained by inversion of the distribution of $T/\sqrt{\widehat{V}_p(T)}$, which is asymptotically $N(0,1)$.

22.5.1.3 Confidence Intervals for Quantiles

Consider the stratified population as described in Section 22.5.1.2 above. Here we define

$$\delta(y, \theta) = \begin{cases} 1 & \text{if} \quad y \leq \theta \\ 0 & \text{otherwise} \end{cases}$$

The pth quantile of the survey population θ_N is a solution of the equation

$$F = \sum_{h=1}^{H} \sum_{j=1}^{N_h} \left\{ \delta(y_{hj}, \theta) - p \right\} = 0 \tag{22.5.17}$$

The survey estimate $\widehat{\theta}_s$ of θ_N based on the stratified sample is obtained by equating the sample optimum EF

$$\widehat{F} = \sum_{h=1}^{H} \frac{N_h}{n_h} \sum_{j \in s_h} \left\{ \delta(y_{hj}, \theta) - p \right\} \tag{22.5.18}$$

to zero.

The expectation of $\widehat{F}$ is zero for $\theta = \theta_N$. The estimated variance of $\widehat{F}$ is

$$\widehat{V}\left(\widehat{F}\right) = \sum_{h=1}^{H} N_h^2 (1 - f_h) \frac{\alpha_h (1 - \alpha_h)}{n_h - 1} \tag{22.5.19}$$

where $\alpha_h = \frac{1}{n_h}\sum_{j\in s_h}\delta\left(y_{hj},\theta_N\right)$ is the proportion of y_{hj}'s that is less than equal to the pth quantile θ_N in the selected sample s_h of the hth stratum. Since θ_N is unknown, we replace θ_N by $\widehat{\theta}_s$ in Eq. (22.5.19) for computation $\widehat{V}\left(\widehat{F}\right)$. Hence the confidence interval of θ_N is obtain by inversion of the distribution

$$\widehat{F}_{\theta=\theta_N}\Big/\widehat{V}\left(\widehat{F}\right)_{\theta=\hat{\theta}_s} \tag{22.5.20}$$

which is asymptotically $N(0,1)$.

The confidence interval for the $p-th$ quantile was derived by Woodruff (1952) while Godambe and Thompson (1999) considered Eq. (22.5.20) from the angle of the EF. A lucid presentation was provided by Godambe (1991). The main feature of the derivation of this confidence interval is that it is not obtained from the traditional method of inverting the distribution of the estimator $\widehat{\theta}_s$ but it was derived from the distribution of $\widehat{F}$.

22.6 NONRESPONSE

Let a sample s of size n be selected from a finite population U with probability $p(s)$ using a sampling design p and let $\pi_i(>0)$ be the inclusion probability of the ith unit. Suppose $r(\geq 0)$ of the selected n units responded and the remaining $n-r$ did not response. Let the set of response sample be denoted by $s'(\subset s)$. Here the data in hand are $d(s,s') = \{(s,s') : (i,y_i), i\in s'\}$ with y_i as the value of the study variable y for the ith unit. Let us assume that the response probability of the unit i is $q_i(>0)$ and it is known for $i = 1,\ldots, N$.

Consider the superpopulation model where $y_1,\ldots, y_N$ are independent with

$$E_\xi(y_i) = \theta x_i, \theta \in \Omega_\theta \text{ for } i = 1,\ldots,N. \tag{22.6.1}$$

Suppose we are interested in estimating the survey parameter $\theta_N = \sum_{i=1}^{N}\alpha_i y_i\Big/\sum_{i=1}^{N}\alpha_i x_i$ from the estimating equation

$$g = \sum_{i=1}^{N}(y_i - \theta x_i)\alpha_i = 0 \tag{22.6.2}$$

where x_i and α_i are known constants.

We can estimate θ_N using the following two approaches:

Approach 1: Let $\mathscr{H}^{(1)}$ be the class of unbiased estimating functions (UEFs), which comprises of EFs h_1 based on the data $d(s,s')$ satisfying

$$E_pE_R(h_1) = \sum_{i=1}^{N}(y_i - \theta x_i)\alpha_i \quad \forall \mathbf{y} = (y_1,\ldots,y_N) \text{ and } \theta\in\Omega_\theta \tag{22.6.3}$$

where E_R denotes expectation over the response mechanism R.

The optimum EF in the class $\mathscr{H}^{(1)}$ is one which $E_\xi E_p E_R\left(h_1^2\right)$ minimizes for $h_1 \in H^1$. Since the inclusion probability for the ith unit in s' is $\pi_i q_i$, the optimal EF is obtained from Theorem 22.3.1 as

$$\widehat{h}_1^* = \sum_{i \in s'} \frac{(y_i - \theta x_i)\alpha_i}{\pi_i q_i} \tag{22.6.4}$$

Approach 2: In case 100% response was available, i.e., $r = n$, then the optimum UEF based on the data $d(s) = \{s:(i,y_i), i \in s\}$ for estimating θ_N would be

$$h^*(s) = \sum_{i \in s} \frac{(y_i - \theta x_i)\alpha_i}{\pi_i} \tag{22.6.5}$$

Since $h^*(s)$ is unknown, we consider the class $\mathscr{H}^{(2)}$ UEFs $h_2(s, s')$ satisfying

$$E_p[h_2(s, s') - h^*(s)|s] = 0 \quad \forall \mathbf{y} \text{ and } \theta \in \Omega_\theta \tag{22.6.6}$$

The optimum UEFs is defined by $\widehat{h}_2^*$ that satisfies

$$E_\xi E_p\left(\widehat{h}_2^*\right)^2 \leq E_\xi E_p\left(h_2^2\right) \quad \forall \mathrm{h}_2 \in \mathscr{H}^{(2)} \text{ and } \theta \in \Omega_\theta$$

The optimal EF in the class $\mathscr{H}^{(2)}$ is obtained from Theorem 22.4.1 as

$$\widehat{h}_2^* = \sum_{i \in s'} \frac{(y_i - \theta x_i)\alpha_i}{\pi_i q_i} \tag{22.6.7}$$

Thus the two approaches lead to the same optimal EF $\widehat{h}_1^* = \widehat{h}_2^* = \sum_{i \in s'} \frac{(y_i - \theta x_i)\alpha_i}{\pi_i q_i} = \widehat{h}^*$.

The following theorems of Godambe and Thompson (1986a,b, 1987) summarize the discussions above.

Theorem 22.6.1

For the superpopulation model (Eq. (22.6.1)) the EF $\widehat{h}^* = \sum_{i \in s'} \frac{(y_i - \theta x_i)\alpha_i}{\pi_i q_i}$ is optimal in the class of unbiased EFs $\mathscr{H}^{(1)}$.

Theorem 22.6.2

Let $\mathscr{H}^{(2)}$ be a subclass of $\mathscr{H}^{(1)}$ for which $h_2(s, s')$ depends only on s'. Then the EF $\widehat{h}^*$ given in Eq. (22.6.7) is also optimum in the subclass $\mathscr{H}^{(2)}$.

Example 22.6.1

Let us assume $E_\xi(y_i - \theta x_i)^2 = \sigma^2 v(x_i)$, where $v(x_i)$ is a known function of x_i and σ^2 is positive but unknown. In this case

$$E_\xi E_p\left(\widehat{h}^*\right)^2 = \sigma^2 \sum_{i=1}^{N} \frac{\alpha_i^2 v(x_i)}{\pi_i q_i} \tag{22.6.8}$$

The optimum values of the π_i's, which minimize (25.5.8) are obtained under the following two constraints.

In case $\sum_{i=1}^{N} \pi_i = n =$ expected sample size $= E_p|s|$ is kept fixed, the optimum value of π_i becomes

$$\pi_{i0} = n \frac{\alpha_i \sqrt{v(x_i)/q_i}}{\sum_{i=1}^{N} \alpha_i \sqrt{v(x_i)/q_i}} \tag{22.6.9}$$

On the other hand in case $\sum_{i=1}^{N} \pi_i q_i = n^* =$ expected sample size of s' is kept fixed, the optimum value of π_i is obtained as

$$\pi_{i0}^* = n^* \frac{\left(\alpha_i \sqrt{v(x_i)}\right)/q_i}{\sum_{i=1}^{N} \alpha_i \sqrt{v(x_i)}} \tag{22.6.10}$$

22.7 CONCLUDING REMARKS

The concept of the EF was developed by Godambe (1960a,b) and Durbin (1960). It unifies the principal methods in the theory of estimation, namely, the method of LS, ML, minimum chi-square, and the method of minimum variance unbiased estimation. The EFs provide suitable solutions when the least square theory and ML methods fail to give reasonable solutions. Godambe and Thompson (1986a,b) provided a unified approach of finding optimal EFs in presence of a superpopulation model. They also provided methods of determination of confidence intervals through EFs. The theory of estimation functions has been applied successfully in various fields of estimation problems such as survey sampling, biostatistics, econometrics, time series, and stochastic processes. Interested readers are referred to Kalbfleisch and Lawless (1988), Godambe (1985, 1991), Godambe and Thompson (1986a,b), Heyde and Lin (1991), Vijayan (1991) and Baswa (2000), among others.

22.8 EXERCISES

22.8.1 Find the optimal EF for θ in the class of linear EFs when $y_1, \ldots, y_N$ are independent and follow the following superpopulation models. In each case, obtain the survey parameter θ_N and its sample estimate $\widehat{\theta}_s$. (i) $y_i = \theta \log x_i + \in_i$, (ii) $y_i = \theta / x_i + \in_i$, and (iii) $y_i = \theta x_i + \in_i$.

$$E_\xi(\in_i) = 0,\ V_\xi(\in_i) = \sigma_i^2,$$

22.8.2 Let $y_1, \ldots, y_N$ be independent with $E_\xi(y_i) = \theta x_i$, $V_\xi(y_i) = \sigma^2 x_i^g, g \geq 0$. Find θ_N, the optimal estimator of the superpopulation parameter θ and also optimal sample estimator of θ_N when $g = 0$, 1 and 2.

22.8.3 Consider the model $y_i = \theta x_i + \in_i$ with $\in_i > 0$, $\in_i$'s are independent with $E_m(\in_i) = 0$ and $V_m(\in_i) = \sigma_i^2$. Find the survey parameter θ_N and sample estimate $\widehat{\theta}_s$ of θ_N. Find also the optimum sampling design when (i) $\sigma_i^2 \propto x_i$ and (ii) $\sigma_i^2 \propto x_i^2$.

22.8.4 Consider the superpopulation model $y_i = \alpha_i(\theta) + \in_i$, $i = 1, \ldots, N$, $\in_i$'s are independent with mean zero and variance $V(\in_i) = \sigma^2 x_i$

and x_i's are known positive constants. Show that the EF $g^*(\mathbf{y}, \theta) =$

$a(\theta) \sum_{i=1}^{N} \left(\frac{\partial \alpha_i(\theta)}{\partial \theta} \right) \left(\frac{y_i - \alpha_i(\theta)}{x_i} \right)$ is optimal in the class of unbiased

EFs of the form $g(\mathbf{y}, \theta) = \sum_{i=1}^{N} a_i(\theta)(y_i - \alpha_i(\theta))$ (Vijayan, 1991).

22.8.5 Let $y_1, \ldots, y_N$ be independent with $E_\xi(y_i) = \theta x_i$, $V_\xi(y_i) = \sigma^2 x_i^g, g \geq 0$. Find θ_N, the optimal estimator of the superpopulation parameter θ and also optimal sample estimator of θ_N when $g = 0$, 1, and 2.