USE OF ESTIMATING FUNCTIONS FOR INTERVAL ESTIMATION FROM COMPLEX SURVEYS

David A. Binder, Statistics Canada
11-F R.H. Coats Building, Ottawa, Ontario, Canada K1A 0T6

KEY WORDS: Nuisance parameters, Poststratification, Quantiles, Ratios, Regression coefficients, Superpopulations.

1. Introduction

Point and interval estimation of population parameters is one of the cornerstones of modern statistical theory. For parametric infinite populations, these parameters completely describe the underlying distribution. In finite populations, the parameters are often of a descriptive nature, such as the population mean. Through the use of estimating functions, a unifying framework can be given for many of these problems. This unifying framework has been discussed in Godambe and Thompson (1989) and in Heyde (1989) for the case of sampling from infinite populations. We extend this to sampling from finite populations, with or without a superpopulation model. Because of the absence of a parametric likelihood in design-based inference, the issue of nuisance parameters poses special problems, for which we shall propose a solution.

Most population parameters of interest can be described through the population distribution function

$$F(\boldsymbol{y}) = \begin{cases} Pr\{\boldsymbol{Y} \le \boldsymbol{y}\} & \text{for infinite populations,} \\ \sum_{i=1}^{N} I\{\boldsymbol{Y}_i \le \boldsymbol{y}\}/N & \text{for finite populations,} \end{cases}$$

where $I\{\cdot\}$ is the indicator function taking the value one when the condition is satisfied and zero otherwise. At this point we do not distinguish between univariate and multivariate populations. We assume that the population parameter of interest, $\boldsymbol{\theta}_0$, can be formulated as the solution to the equation

$$\boldsymbol{U}(\boldsymbol{\theta}) = \int_{-\infty}^{\infty} \boldsymbol{u}(\boldsymbol{y},\boldsymbol{\theta})\, dF(\boldsymbol{y}) = \boldsymbol{0}. \quad (1.1)$$

Godambe (1984) has discussed such parameter defining estimating equations. For example, for many parametric infinite populations, we have

$$\boldsymbol{u}(\boldsymbol{y},\boldsymbol{\theta}) = \frac{\partial \log f(\boldsymbol{y};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}.$$

Godambe and Thompson (1986) discussed the use of this score function for superpopulation models.

Examples of estimating functions for finite populations are

$u(y,\theta) = y-\theta$ for the mean $\overline{Y}$,

$u(y,x,\theta) = y-\theta x$ for the ratio $\overline{Y}/\overline{X}$,

$u(y,\theta) = I\{y<\theta\}-p$ for the p-th percentile,

$\boldsymbol{u}(y,\boldsymbol{x},\boldsymbol{\theta}) = \boldsymbol{x}(y-\boldsymbol{x}'\boldsymbol{\theta})$ for ordinary least squares regression coefficients.

In Section 2 we consider point estimation using estimating equations. We discuss test inversion methods to derive confidence intervals for one-dimensional parameters in Section 3. The main new results are in Section 4, where we propose a method for eliminating nuisance parameters. The application of these methods to poststratification and regression is given in Section 5.

2. Parameter Estimation

To obtain point estimates of the parameters defined by (1.1), we first consider the estimation of $F(\boldsymbol{y})$. For independent sampling from an infinite population this may be estimated as

$$\hat{F}(\boldsymbol{y}) = \sum_{i=1}^{n} I\{\boldsymbol{Y}_i \le \boldsymbol{y}\}/n,$$

although other non-parametric density estimates are also available. For sampling from finite populations, we use the general framework given by Rao (1979). The well-known Horvitz-Thompson estimator is a special case of this framework when the sample size is fixed. For a given value of $\boldsymbol{y}$, we consider the finite population elements to be the $0-1$ variables given by $I\{\boldsymbol{Y_i} \le \boldsymbol{y}\}$. The linear estimator for $F(\boldsymbol{y})$ has the form

$$\hat{F}(\boldsymbol{y}) = \sum_{i=1}^{N} d_i(s)\, I\{\boldsymbol{Y_i} \le \boldsymbol{y}\}/N,$$

where s is the set of population units in the sample, selected through some probability mechanism $p(s)$, and $d_i(s)=0$ if $i \notin s$. We assume that N is known, although, as we shall see for many of our applications, this assumption is not necessary. It is often sufficient to have only an estimator of $N \cdot F(\boldsymbol{y})$. We denote the estimate of the population size by

$$\hat{N} = \sum_{i=1}^{N} d_i(s)\,.$$

Note that when $\hat{N} \ne N$ then $\hat{F}(\boldsymbol{y})$ is not a true distribution function, since $\hat{F}(\infty)$ would not be equal to 1.

Rao (1979) placed the restriction on $\{d_i(s)\}$, such that if $Y_i \propto \omega_i$, for some $\{\omega_i\}$, then the mean square error of

$$\hat{Y} - Y = \sum_{i=1}^{N} d_i(s)\, Y_i - \sum_{i=1}^{N} Y_i$$

$$= N \int_{-\infty}^{\infty} y\, d[\hat{F}(y) - F(y)]$$

is zero. This implies that the mean square error is necessarily of the form

$$MSE(\hat{Y}) = -\sum_{i<j}\sum d_{ij}\, \omega_i \omega_j (Z_i - Z_j)^2,$$

where

$$Z_i = Y_i/\omega_i,$$

$$d_{ij} = \sum_s p(s)\,[d_i(s)-1]\,[d_j(s)-1]\,.$$

A non-negative unbiased quadratic estimator of $MSE(\hat{Y})$ has the form

$$mse(\hat{Y}) = -\sum_{\substack{i<j \\ i,j \in s}}\sum d_{ij}(s)\, \omega_i \omega_j (Z_i - Z_j)^2,$$

where

$$\sum_s p(s)\, d_{ij}(s) = d_{ij}, \quad i<j.$$

For example, for the case of Horvitz-Thompson estimation for fixed sample sizes, we have

$$\omega_i = \pi_i = \sum_{s \ni i} p(s)\,,$$

$$d_i(s) = \begin{cases} 1/\pi_i & \text{if } i \in s, \\ 0 & \text{if } i \notin s, \end{cases}$$

$$d_{ij} = \frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j},$$

where

$$\pi_{ij} = \sum_{s \ni i,j} p(s)\,.$$

If $d_{ij}(s) = d_{ij}/\pi_{ij}$, we have the well known Yates-Grundy estimator

$$mse(\hat{Y}) = \sum_{\substack{i<j \\ i,j \in s}}\sum \left(\frac{\pi_i \pi_j - \pi_{ij}}{\pi_{ij}}\right)\left(\frac{Y_i}{\pi_i} - \frac{Y_j}{\pi_j}\right)^2.$$

Now, our parameter of interest has been defined as the solution to the equation in (1.1). We consider a "model-assisted" approach to estimation analogous to Cassel, Särndal and Wretman (1976), Särndal, Swensson and Wretman (1989) and Godambe and Thompson (1986). For a given $\boldsymbol{\theta}$, we consider estimators for $\boldsymbol{U}(\boldsymbol{\theta})$ which can be expressed as

$$\hat{\boldsymbol{U}}_{\boldsymbol{\beta}}(\boldsymbol{\theta}) = \int_{-\infty}^{\infty} \boldsymbol{\alpha}(\boldsymbol{x}, \boldsymbol{\beta}, \boldsymbol{\theta})\, d[F(\boldsymbol{x}) - \hat{F}(\boldsymbol{x})]$$

$$+ \int_{-\infty}^{\infty} \boldsymbol{u}(\boldsymbol{y}, \boldsymbol{\theta})\, d\hat{F}(\boldsymbol{y})\,, \tag{2.1}$$

where

$$\boldsymbol{\alpha}(\boldsymbol{x}, \boldsymbol{\beta}, \boldsymbol{\theta}) = \mathscr{E}_\xi\{\boldsymbol{u}(\boldsymbol{y}, \boldsymbol{\theta})\,|\,\boldsymbol{x}\}, \tag{2.2}$$

under some model ξ, which may contain an unknown parameter $\boldsymbol{\beta}$. For example, if

$$u(y,\theta) = y-\theta$$
$$\alpha(\boldsymbol{x},\boldsymbol{\beta},\theta) = \boldsymbol{x}'\boldsymbol{\beta}-\theta$$

and we use Horvitz-Thompson estimation, we have

$$\hat{U}_{\beta}(\theta) = \bar{\boldsymbol{X}}'\boldsymbol{\beta}+\sum_{i=1}^{n}\left(\frac{Y_i-\boldsymbol{X}_i'\boldsymbol{\beta}}{N\pi_i}\right)-\theta,$$

which is the estimating function for a regression estimator of the mean.

If $\boldsymbol{\beta}$ is unknown, we include the functions

$$\boldsymbol{U}^*(\boldsymbol{\beta}) = \int_{-\infty}^{\infty}\boldsymbol{u}^*(\boldsymbol{x},\boldsymbol{y},\boldsymbol{\beta})\,dF(\boldsymbol{x},\boldsymbol{y}) \qquad (2.3)$$

and

$$\hat{\boldsymbol{U}}^*(\boldsymbol{\beta}) = \int_{-\infty}^{\infty}\boldsymbol{u}^*(\boldsymbol{x},\boldsymbol{y},\boldsymbol{\beta})\,d\hat{F}(\boldsymbol{x},\boldsymbol{y})$$

in our setup, where $\boldsymbol{\beta}_0$ is the solution to $\boldsymbol{U}^*(\boldsymbol{\beta})=\mathbf{0}$. In general $\hat{\boldsymbol{\theta}}$, our point estimator for $\boldsymbol{\theta}_0$, is given as the solution to

$$\hat{\boldsymbol{U}}^*(\hat{\boldsymbol{\beta}}) = \mathbf{0},$$
$$\hat{\boldsymbol{U}}_{\beta}(\hat{\boldsymbol{\theta}}) = \mathbf{0}. \qquad (2.4)$$

To exemplify these concepts, we consider the poststratified estimator of the population mean. Here we have

$$u(y,\theta) = y-\theta,$$

and the $\boldsymbol{x}$'s are dummy variables for the first $H-1$ strata. We let $\boldsymbol{\beta}=(\beta_1,\ldots,\beta_{H-1})'$. For strata $h=1,\ldots,H-1$, we have

$$\mathscr{E}_{\xi}(y-\theta) = \beta_h-\theta,$$

and for stratum H, we have

$$\mathscr{E}_{\xi}(y-\theta) = \beta_H-\theta = \frac{1}{N_H}\sum_{h=1}^{H-1}N_h(\theta-\beta_h),$$

where $N_1,\ldots,N_H$ are the known strata sizes and

$$\theta = \frac{1}{N}\sum_{h=1}^{H}N_h\beta_h.$$

The estimating function for $\boldsymbol{\beta}$, given in (2.3) is

$$u_h^*(y,\boldsymbol{\beta}) = \delta_h(y-\beta_h),$$

where

$$\delta_h = \begin{cases}1 & \text{if observation is in stratum } h\\ 0 & \text{otherwise.}\end{cases} \qquad (2.6)$$

Solving the equations in (2.4), we obtain $\beta_h=\hat{Y}_h/\hat{N}_h$ where $\hat{Y}_h$ and $\hat{N}_h$ are the estimated total and size for stratum h and

$$\theta = \frac{1}{N}\sum_{h=1}^{H}N_h\frac{\hat{Y}_h}{\hat{N}_h},$$

the poststratified estimator of the population mean.

3. Confidence Intervals: One Dimensional Case

In the previous section we defined our parameter of interest as the solution to equation (1.1). We now consider the problem of constructing a confidence interval for this parameter. We first discuss the relatively simple case of only one unknown parameter θ. In Section 4, we extend this to the more complex case where the unknown parameters are multidimensional. The parameter of interest is defined as the solution to $U(\theta_0)=0$ and our estimator of $U(\theta)$ is $\hat{U}(\theta)$ with mean square error $MSE\{\hat{U}(\theta)\}$ estimated by $mse\{\hat{U}(\theta)\}$.

In Section 2 we discussed design-based estimators of the mean square error which could be used here when θ is given. However, we are not restricted to only design-based inferences as other randomization distributions are also permitted within the general framework. For example, the parameter of interest may be a superpopulation parameter, in which case the estimator of mean square error over the superpopulation distribution would be used.

To construct a confidence interval for θ_0 we propose the method of test inversion. In particular, we let the null and alternative hypotheses be

$$H_0: U(\theta) = 0$$
$$H_1: U(\theta) \neq 0$$

and we define the confidence interval for θ_0 to be the values of θ for which H_0 is accepted. Our acceptance region for H_0 is given as

$$\frac{\hat{U}(\theta)^2}{mse\{\hat{U}(\theta)\}} \leq z^2_{1-\frac{\alpha}{2}}, \qquad (3.1)$$

where z_α is the $\alpha-th$ percentile of a standard normal distribution. It is assumed here that $\hat{U}(\theta)$ is approximately normal with mean $U(\theta)$ and variance $MSE\{\hat{U}(\theta)\}$ and that $mse\{\hat{U}(\theta)\}$ is a consistent estimator of $MSE\{\hat{U}(\theta)\}$. This assumption will be valid for a large class of estimators and sample designs under various schemes for letting the population and sample sizes go to infinity. Sen (1988) has discussed the asymptotic theory for sampling from finite populations.

We see that confidence intervals which are constructed using (3.1) are preserved under one-to-one transformations of the parameter. This is to be contrasted with the commonly used intervals derived using Taylor expansions of the estimating equations. The Taylor expansion method (δ-method) can be derived by writing

$$\begin{aligned} 0 &= \hat{U}(\hat{\theta}) \\ &= U(\theta) + [\hat{U}(\theta) - U(\theta)] \\ &\quad + [U(\hat{\theta}) - U(\theta)] + R, \end{aligned} \qquad (3.2)$$

where

$$R = (\hat{U}-U)(\hat{\theta}) - (\hat{U}-U)(\theta).$$

At $\theta = \theta_0$ we have that $R = o_p(|\hat{\theta} - \theta_0|)$ and

$$\begin{aligned} U(\hat{\theta}) - U(\theta_0) &= (\hat{\theta} - \theta_0)\left[\frac{\partial U(\theta_0)}{\partial \theta}\right] \\ &\quad + o_p(|\hat{\theta} - \theta_0|). \end{aligned} \qquad (3.3)$$

Therefore, evaluating (3.2) at $\theta = \theta_0$ and using (3.3) we have

$$\hat{\theta} - \theta_0 \approx -\left[\frac{\partial U(\theta_0)}{\partial \theta}\right]^{-1} \hat{U}(\theta_0)$$

from which the approximate variance of $\hat{\theta}$ may be derived and confidence intervals assuming normality may be obtained. Binder (1983) discussed this method for the more general case of multidimensional parameters.

The estimation of a ratio provides an example of confidence intervals obtained by solving expression (3.1) for θ. Here

$$\begin{aligned} &\frac{\hat{U}(\theta)}{mse\{\hat{U}(\theta)\}^{1/2}} \\ &= \frac{\hat{Y} - \theta\hat{X}}{(m_{yy} - 2\theta m_{yx} + \theta^2 m_{xx})^{1/2}}, \end{aligned} \qquad (3.4)$$

where m_{yy} and m_{xx} are the estimated mean square errors of $\hat{Y}$ and $\hat{X}$, respectively, and m_{yx} is the estimate for

$$\begin{aligned} &\mathscr{E}\{(\hat{Y}-Y)(\hat{X}-X)\} \\ &= -\sum_{i<j}\sum d_{ij}\,\omega_i\,\omega_j\left(\frac{Y_i}{\omega_i} - \frac{Y_j}{\omega_j}\right)\left(\frac{X_i}{\omega_i} - \frac{X_j}{\omega_j}\right). \end{aligned}$$

To obtain the confidence interval for the ratio, we see from (3.4) that it is necessary to solve a quadratic equation. This is similar to the interval proposed by Fieller (1932) under sampling from infinite normal populations. If, in (3.4), we use $\hat{\theta}$ instead of θ in the denominator, we obtain the usual Taylor expansion based confidence interval.

Another interesting example is the case of estimating the population percentile. In this case we have

$$u(y,\theta) = I\{y<\theta\}-p,$$

so that $\hat{U}(\theta) = \hat{F}(\theta) - (\hat{N}/N)p$. Confidence intervals defined by (3.1) yield those proposed by Francisco and Fuller (1991). If, in the denominator of (3.1), we evaluate $mse\{\hat{U}(\theta)\}$ at $\theta=\hat{\theta}$, we obtain the confidence intervals proposed by Woodruff (1952).

4. The Multidimensional Parameter Case

In general, our population parameters are estimated through a system of equations, such as those given in (2.4). Among the parameters $(\boldsymbol{\theta}, \boldsymbol{\beta})$, some of them may be nuisance parameters. We rewrite the system of equations given in (2.4) as $\hat{\boldsymbol{U}}(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\lambda}}) = \mathbf{0}$, where $\hat{\boldsymbol{U}}(\cdot)$ includes both $\hat{\boldsymbol{U}}^*(\cdot)$ and $\hat{\boldsymbol{U}}_{\boldsymbol{\beta}}(\cdot)$ in (2.4), and $\boldsymbol{\lambda}$ is the vector of nuisance parameters. We assume that $\hat{\boldsymbol{U}}(\boldsymbol{\theta}, \boldsymbol{\lambda})$ is approximately multivariate normal with mean $\boldsymbol{U}(\boldsymbol{\theta}, \boldsymbol{\lambda})$ and covariance $\boldsymbol{V}_U(\boldsymbol{\theta}, \boldsymbol{\lambda})$.

First, we consider the case where there are no nuisance parameters. Confidence intervals which are analogous to those defined by expression (3.1), are

$$\{\boldsymbol{\theta} \mid \hat{\boldsymbol{U}}(\boldsymbol{\theta})' \hat{\boldsymbol{V}}_U^{-1}(\boldsymbol{\theta})\, \hat{\boldsymbol{U}}(\boldsymbol{\theta}) \le \chi^2_{1-\alpha}(r)\}, \qquad (4.1)$$

where $\hat{\boldsymbol{V}}_U(\boldsymbol{\theta})$ is the estimator of the $r \times r$ matrix $\boldsymbol{V}_U(\boldsymbol{\theta})$, and $\chi^2_\alpha(r)$ is the $\alpha-th$ percentile of a χ^2 distribution with r degrees of freedom.

Note that, in general, confidence intervals defined by (4.1) will not be elliptical or even symmetric. If we were to linearize $\hat{\boldsymbol{U}}(\boldsymbol{\theta})$ around $\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}$, and evaluate $\hat{\boldsymbol{V}}_U(\boldsymbol{\theta})$ at $\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}$, we would obtain the confidence intervals derived in Binder (1983).

Now we consider the case where the parameter of interest, θ, is one-dimensional and all the other population parameters are nuisance parameters. We partition the vector $\boldsymbol{U}(\boldsymbol{\theta}, \boldsymbol{\lambda})$ into

$$\boldsymbol{U}(\theta, \boldsymbol{\lambda}) = \begin{bmatrix} U_1(\theta, \boldsymbol{\lambda}) \\ \boldsymbol{U}_2(\theta, \boldsymbol{\lambda}) \end{bmatrix},$$

where $U_1(\theta_0, \boldsymbol{\lambda}) = 0$ and $\boldsymbol{U}_2(\theta, \boldsymbol{\lambda}_0) = \mathbf{0}$. Note that U_1 and $\boldsymbol{U}_2$ are the estimating functions for θ and $\boldsymbol{\lambda}$, respectively.

We denote the r-variate standard normal density function by $\phi_r(\cdot)$. We say that C is a $1-\alpha$ confidence region for $(\theta, \boldsymbol{\lambda})$ if

$$\int_{C^*} \phi_r(\hat{\boldsymbol{V}}_U^{-1/2}\hat{\boldsymbol{U}})\, d(\hat{\boldsymbol{V}}_U^{-1/2}\hat{\boldsymbol{U}}) = 1-\alpha, \qquad (4.2)$$

where

$$C = \{(\theta, \boldsymbol{\lambda}) \mid \hat{\boldsymbol{V}}_U^{-1/2}\hat{\boldsymbol{U}} \in C^*\}.$$

The integral in expression (4.2) may be rewritten as

$$\int_D \phi_r(\hat{\boldsymbol{V}}_U^{-1/2}\hat{\boldsymbol{U}})\, |\hat{\boldsymbol{V}}_U|^{-1/2}\, d\hat{\boldsymbol{U}}, \qquad (4.3)$$

where

$$D = \{\hat{\boldsymbol{U}} \mid \hat{\boldsymbol{V}}_U^{-1/2}\hat{\boldsymbol{U}} \in C^*\}.$$

A confidence interval, C_θ, for θ which does not depend on the nuisance parameters, satisfies $C = C_\theta \times \Lambda$, assuming the range for $\boldsymbol{\lambda}$ does not depend on the value of θ. If $\hat{\boldsymbol{U}}$ is differentiable, then the integral in (4.3) can be expressed as

$$\int_{D_\theta} \int_\Lambda \phi_r(\hat{\boldsymbol{V}}_U^{-1/2}\hat{\boldsymbol{U}}) \frac{|\boldsymbol{J}(\theta, \boldsymbol{\lambda})|}{|\boldsymbol{J}_{1\theta}(\theta, \boldsymbol{\lambda}_a)|} |\hat{\boldsymbol{V}}_U|^{-1/2}\, d\boldsymbol{\lambda}\, d\hat{U}_1(\theta, \boldsymbol{\lambda}_a), \qquad (4.4)$$

where $\boldsymbol{\lambda}_a$ is some arbitrary value of $\boldsymbol{\lambda}$,

$$C_\theta = \{\theta \mid \hat{U}_1(\theta, \boldsymbol{\lambda}_a) \in D_\theta\},$$

$$\boldsymbol{J} = \begin{bmatrix} J_{1\theta} & \boldsymbol{J}_{1\lambda} \\ \boldsymbol{J}_{2\theta} & \boldsymbol{J}_{2\lambda} \end{bmatrix} = \begin{bmatrix} \frac{\partial \hat{U}_1}{\partial \theta} & \frac{\partial \hat{U}_1}{\partial \boldsymbol{\lambda}} \\ \frac{\partial \hat{\boldsymbol{U}}_2}{\partial \theta} & \frac{\partial \hat{\boldsymbol{U}}_2}{\partial \boldsymbol{\lambda}} \end{bmatrix}.$$

The integration over the nuisance parameters should not be confused with Bayesian methods. The use of the integral here has been justified on purely repeated sampling principles.

Unless $\hat{\boldsymbol{U}}$ is linear in $\boldsymbol{\lambda}$ and $\hat{\boldsymbol{V}}_U$ is constant in $\boldsymbol{\lambda}$, expression (4.4) will often be difficult to calculate, even numerically. To overcome this difficulty, we propose taking a linear approximation for $\hat{\boldsymbol{U}}$ and we assume that $\hat{\boldsymbol{V}}_U$ is approximately constant in $\boldsymbol{\lambda}$. This is analogous to the Taylor series method given in Binder (1983), except that here we do not assume that $\hat{\boldsymbol{U}}$ is linear in θ or that $\hat{\boldsymbol{V}}_U$ is constant in θ. In general, this will result in asymmetric intervals similar to those given in (3.1) for the case of no nuisance parameters.

In particular, we take the first order Taylor expansion of $\hat{\boldsymbol{U}}$ around $\hat{\boldsymbol{\lambda}}_\theta$, the solution to $\hat{\boldsymbol{U}}_2(\theta, \hat{\boldsymbol{\lambda}}_\theta) = \boldsymbol{0}$, so that

$$\hat{U}_1(\theta, \boldsymbol{\lambda}) \approx \hat{U}_1(\theta, \hat{\boldsymbol{\lambda}}_\theta) + \boldsymbol{J}_{1\lambda}(\theta, \hat{\boldsymbol{\lambda}}_\theta)(\boldsymbol{\lambda} - \hat{\boldsymbol{\lambda}}_\theta)$$

$$\hat{\boldsymbol{U}}_2(\theta, \boldsymbol{\lambda}) \approx \boldsymbol{J}_{2\lambda}(\theta, \hat{\boldsymbol{\lambda}}_\theta)(\boldsymbol{\lambda} - \hat{\boldsymbol{\lambda}}_\theta).$$

In order to integrate out $\boldsymbol{\lambda}$ in (4.3) we must assume that $J_{1\theta}$ and $\boldsymbol{J}_{2\theta}$ are independent of $\boldsymbol{\lambda}$. Fortunately, many common applications satisfy this condition. For example, if, in (2.2), $\boldsymbol{\beta}$ is the nuisance parameter and

$$\frac{\partial^2 \boldsymbol{\alpha}(\boldsymbol{x}, \boldsymbol{\beta}, \theta)}{\partial \theta \partial \boldsymbol{\beta}} = \boldsymbol{0},$$

then the estimating equations in (2.4) satisfy the requirement. This is true , for example, when $\boldsymbol{\alpha}$ is a linear model.

Under our assumptions, we integrate (4.4) with respect to $\boldsymbol{\lambda}$ and obtain

$$\int_{D_\theta} \phi_1 [W^{-1/2} \hat{U}_1(\theta, \hat{\boldsymbol{\lambda}}_\theta)] \, |W|^{-1/2} d\hat{U}_1(\theta, \hat{\boldsymbol{\lambda}}_\theta), \tag{4.5}$$

where

$$W^{-1} = V^{11} - (V^{11}\boldsymbol{J}_{1\lambda} + \boldsymbol{V}^{12}\boldsymbol{J}_{2\lambda})$$
$$(\boldsymbol{J}_\lambda' \boldsymbol{V}^{-1} \boldsymbol{J}_\lambda)^{-1} (\boldsymbol{J}_{1\lambda}' V^{11} + \boldsymbol{J}_{2\lambda}' \boldsymbol{V}^{21}),$$

$$\hat{\boldsymbol{V}}_U^{-1} = \begin{bmatrix} V^{11} & \boldsymbol{V}^{12} \\ \boldsymbol{V}^{21} & \boldsymbol{V}^{22} \end{bmatrix}, \quad \boldsymbol{J}_\lambda = \begin{bmatrix} \boldsymbol{J}_{1\lambda} \\ \boldsymbol{J}_{2\lambda} \end{bmatrix}.$$

Note that $\boldsymbol{J}_\lambda$ is evaluated at $\boldsymbol{\lambda} = \hat{\boldsymbol{\lambda}}_\theta$. Applying the Binomial Inverse Theorem (see, for example, Press 1972, p.23) and defining

$$\boldsymbol{K} = -\begin{bmatrix} c & \boldsymbol{J}_{1\lambda} \\ \boldsymbol{J}_{1\lambda}' & \boldsymbol{J}_{2\lambda} \end{bmatrix}^{-1} = \begin{bmatrix} K_{11} & \boldsymbol{K}_{12} \\ \boldsymbol{K}_{21} & \boldsymbol{K}_{22} \end{bmatrix},$$

for some positive constant c, we obtain

$$W = K_{11}^{-1} [\boldsymbol{K} \hat{\boldsymbol{V}}_U \boldsymbol{K}]_{11} K_{11}^{-1}. \tag{4.6}$$

It is interesting to compare this result with the usual classical parametric model case. Here, $\hat{\boldsymbol{U}}$ is usually the score function, and when $c = \partial \hat{U}_1 / \partial \theta$, we have that $\hat{\boldsymbol{V}}_U$ and $\boldsymbol{K}$ are the estimated Fisher information matrix and its inverse, respectively, evaluated at $\boldsymbol{\lambda} = \hat{\boldsymbol{\lambda}}_\theta$. In this case $W = K_{11}^{-1}$, which is the usual classical result.

We also note a strong analogy with the proposal in Godambe (1991) where it is suggested that the appropriate estimating function for θ is

$$\hat{U}_1(\theta, \hat{\boldsymbol{\lambda}}_\theta) - \mathscr{E}_{\hat{\lambda}_\theta}(\hat{U}_1 | \hat{\boldsymbol{U}}_2).$$

This is because the variance matrix W in (4.6) may be expressed as

$$W = var_{\boldsymbol{\lambda} = \hat{\boldsymbol{\lambda}}_\theta} \{\hat{U}_1 + K_{11}^{-1} \boldsymbol{K}_{12} \hat{\boldsymbol{U}}_2\}$$
$$= var_{\boldsymbol{\lambda} = \hat{\boldsymbol{\lambda}}_\theta} \{\hat{U}_1 - \boldsymbol{J}_{1\lambda} \boldsymbol{J}_{2\lambda}^{-1} \hat{\boldsymbol{U}}_2\},$$

where $\boldsymbol{J}_\lambda$ is assumed fixed.

To compute this variance it is often convenient to use the variable

$$u^*(\boldsymbol{y},\theta) = \tag{4.7}$$

$u_1(y,\theta,\hat{\lambda}_\theta) - \boldsymbol{J}_{1\lambda}\boldsymbol{J}_{2\lambda}^{-1}\boldsymbol{u}_2(y,\theta,\hat{\lambda}_\theta)$ and estimate the mean square error of $\hat{U}^*$.

Finally, to obtain confidence intervals for θ, we see that the distribution in (4.5) may be treated just as in the case of no nuisance parameters. The analogous expression for (3.1) becomes

$$\frac{\hat{U}_1(\theta,\hat{\lambda}_\theta)^2}{W(\theta,\hat{\lambda}_\theta)} \le z^2_{1-\frac{\alpha}{2}}, \tag{4.8}$$

where W is given in (4.6). If we had taken a Taylor expansion of $\hat{U}_1$ around $\theta=\hat{\theta}$ and evaluated W at $\theta=\hat{\theta}$, we would obtain the same intervals as those in Binder (1983). However, the use of (4.8) to derive confidence intervals requires fewer assumptions than the other methods.

5. Examples

In this section we consider the application of these results to poststratification and to estimating a regression coefficient.

As described in Section 2, for poststratification, we have from (2.1)

$$\hat{U}_1 = \frac{\hat{Y}}{N} - \frac{1}{N}\sum_{h=1}^{H-1}\left(\hat{N}_h - N_h\frac{\hat{N}_H}{N_H}\right)\beta_h - \left(\frac{\hat{N}_H}{N_H}\right)\theta. \tag{5.1}$$

Evaluating this at

$$\beta_h = \frac{\hat{Y}_h}{\hat{N}_h}, \quad h=1,\ldots,H-1,$$

we have

$$\hat{U}_1 = \frac{\hat{N}_H}{N_H}(\hat{\theta}-\theta),$$

where $\hat{\theta}$ is the poststratified estimator for the mean. Our nuisance parameters in (5.1) are $\{\beta_h,\ h=1,\ldots,H-1\}$. We also let $\hat{\lambda}=\hat{N}_H$ since $\hat{N}_H$ appears in the expression for $\hat{U}_1$. To ensure that $\boldsymbol{J}_{1\theta}$ is independent of the nuisance parameters we multiply the right hand side of (5.1) by $N_H/\hat{\lambda}$ to obtain

$$\hat{U}_1 = \frac{N_H}{\hat{\lambda}}\frac{\hat{Y}}{N} - \frac{1}{N}\sum_{h=1}^{H-1}\left(\hat{N}_h\frac{N_H}{\hat{\lambda}} - N_h\right)\beta_h - \theta.$$

Our estimating functions for the nuisance parameters are

$$\hat{U}_{2h} = \frac{1}{N}(\hat{Y}_h - \hat{N}_h\beta_h), \qquad h=1,\ldots,H-1$$

$$\hat{U}_{2H} = \frac{1}{N}(\hat{N}_H - \lambda).$$

Applying our results, we find that u^* in (4.7) is

$$\sum_{h=1}^{H}\frac{\delta_h}{N}\frac{N_h}{\hat{N}_h}\left(y - \frac{\hat{Y}_h}{\hat{N}_h}\right) - (\theta-\hat{\theta}),$$

where δ_h is defined in (2.6). Therefore, the test statistic corresponding to (4.8) is

$$\frac{(\hat{\theta}-\theta)^2}{mse\left\{\sum_{i=1}^{N} d_i(s)\,u^*(Y_i,\theta)\right\}}.$$

When the sample size is constant, so that $\hat{N}=N$ for all samples, the term in $\theta-\hat{\theta}$ can be ignored and the mean square error corresponds to that given by Rao (1985). He argued, on the grounds of conditional inference, that it is better to leave the term $N_h/\hat{N}_h$ in the variance expression rather than setting it to its asymptotic value of one. We see that in our formulation this happens naturally. Also the calculation of the MSE can be over the conditional distribution, or even a superpopulation distribution, if so desired.

As a second example, we consider the estimation of the slope of a regression line in simple linear regression. The extension to one of the coefficients in a multiple regression model is straightforward. We let θ be the slope and λ be the nuisance parameter for the intercept. In this case we have

$$u_1(y,x,\theta,\lambda) = x(y-\lambda-\theta x)$$

$$u_2(y,x,\theta,\lambda) = (y-\lambda-\theta x),$$

so we have

$$u_1\big|_{\lambda=\hat{\lambda}} = x\left[y-\frac{\hat{Y}}{\hat{N}}-\theta\left(x-\frac{\hat{X}}{\hat{N}}\right)\right]$$

and from (4.7)

$$u^* = \left(x-\frac{\hat{X}}{\hat{N}}\right)\left(y-\frac{\hat{Y}}{\hat{N}}\right)-\theta\left(x-\frac{\hat{X}}{\hat{N}}\right)^2. \qquad (5.2)$$

The test statistic obtained from u^* given in (5.2) has the same form as that obtained for the ratio in (3.4). Therefore, the intervals can be obtained, in general, by solving a quadratic equation.

6. Summary

The application of the theory of estimating functions has been well established in the classical parametric framework. Godambe and Thompson (1986) have shown how some of this theory can be applied in sample surveys. However, the theory of optimal estimating equations depends on model considerations. Some of these model assumptions could be weakened as was shown by Godambe and Thompson (1989). We have shown that the formulation can also be applied for descriptive parameters of a finite population, using only design-based inferences.

The issue of removing nuisance parameters has been addressed in the literature by conditioning and orthogonality considerations. We have shown that this issue can be resolved within a general framework which includes both design-based and model-based inference. The examples give appealing results which are consistent with the theory of orthogonal and conditional methods.

Finally, we note that when the estimating function is not differentiable, such as is the case for population quantiles, it may be more difficult to eliminate nuisance parameters. The integral in (4.2) still applies, but the approximation of the integral may be difficult.

Acknowledgments

I would like to thank V.P. Godambe, H.J. Mantel, D. Pfeffermann, and J.N.K. Rao for their useful comments to an earlier draft of this paper.

References

Binder, D.A. (1983), "On the Variances of Asymptotically Normal Estimators from Complex Surveys", *International Statistical Review, 51*, 279-292.

Cassel, C.M., C.E. Särndal, and J.H. Wretman (1976), "Some Results on Generalized Difference Estimation and Generalized Regression Estimation for Finite Populations", *Biometrika, 63*, 615-620.

Fieller, E.C. (1932), "The Distribution of the Index in a Normal Bivariate Population", *Biometrika, 24*, 428-440.

Francisco, C.A. and W.A. Fuller (1991), "Quantile Estimation with a Complex Survey Design", *Annals of Statistics, 19*, 454-469.

Godambe, V.P. (1984), "Robust Estimation through Estimating Equations", *Biometrika, 71*, 115-125.

Godambe, V.P. (1991), "Orthogonality of Estimating Functions and Nuisance Parameters", *Biometrika 78*, 143-151.

Godambe, V.P. and M.E. Thompson (1986), "Parameters of Superpopulation and Survey Population: Their Relationships and Estimation", *International Statistical Review, 54*, 127-138.

Godambe, V.P. and M.E. Thompson (1989), "An Extension of Quasi-Likelihood Estimation", *Journal of Statistical Planning and Inference, 22*, 137-152.

Heyde C.C. (1989), "Quasi-Likelihood and Optimality for Estimating Functions: Some Current Unifying Themes", *Bulletin of the International Statistical Institute, Proceedings of the 47th Session, Book 1*, 19-28.

Press S.J. (1972), *Applied Multivariate Analysis*, Holt, Rinehart and Winston: New York.

Rao, J.N.K. (1979), "On Deriving Mean Square Errors and their Non-Negative Unbiased Estimators in Finite Population Sampling", *Journal of the Indian Statistical Association, 17*, 125-136.

Rao, J.N.K. (1985), "Conditional Inference in Survey Sampling", *Survey Methodology, 11*, 15-31.

Särndal, C.E., B. Swensson, and J.H. Wretman (1989), "The Weighted Residual Technique for Estimating the Variance of the General Regression Estimator of the Finite Population Total", *Biometrika, 76*, 527-537.

Sen, P.K. (1988), "Asymptotics in Finite Population Sampling", *Handbook of Statistics, Vol.6* (P.R. Krishnaiah and C.R. Rao, eds.), 291-331. Elsevier Science: Amsterdam.

Woodruff, R.S. (1952), "Confidence intervals for medians and other position measures", *Journal of the American Statistical Association, 47*, 635-646.