CHAPTER 25

Empirical Likelihood Method in Survey Sampling

25.1 INTRODUCTION

Likelihood is the most important tool for parametric inference whereas empirical likelihood (EL) is a powerful nonparametric approach to statistical inference. EL was first introduced in survey sampling by Hartley and Rao (1968) in the name of scale load approach. The modern concept of EL was introduced by Owen (1988). The application of EL approach in survey sampling was introduced by Chen and Quin (1993), Chen and Sitter (1999), Sitter and Wu (2002), and Rao and Wu (2009), among others. In this chapter, we will present an overview of recent developments in EL methods in estimating finite population characteristics such as population mean, variance, and distribution function. We also introduce the concept of pseudo–empirical likelihood (PEL) for the probability sampling designs and their applications to superpopulation models and raking estimators. The uses of EL in determination of confidence intervals have also been discussed.

25.2 SCALE LOAD APPROACH

Consider a finite population U that consists of N units and y_i be the value of the study variable y for the ith unit of the population. In a fixed population approach the vector $\mathbf{y} = (y_1,\ldots, y_i,\ldots, y_N)$ is treated as a parameter and the corresponding parameter space is $\Omega_y = (-\infty < y_1 < \infty,\ldots, -\infty < y_i < \infty,\ldots, -\infty < y_N < \infty)$. Suppose a sample s is selected from the population U with probability $p(s)$ following a sampling design p. The likelihood function of $\mathbf{y}$ is the conditional probability of obtaining the data $d = (y_i, i \in s)$ for the given $\mathbf{y}$ and it can be written as

$$L(\mathbf{y}) = Prob(d|\mathbf{y}) = \begin{cases} p(s) & \text{for} \quad y_i, i \in s \\ 0 & \text{otherwise} \end{cases} \tag{25.2.1}$$

Survey Sampling Theory and Applications
ISBN 978-0-12-811848-1
http://dx.doi.org/10.1016/B978-0-12-811848-1.00025-X

Godambe (1966) concluded that though this aforementioned likelihood function (25.2.1) is well defined, it is noninformative in the sense that all possible nonobserved values $y_i, i \notin s$ have the same likelihood. To overcome this difficulty, Hartley and Rao (1968) introduced the scale load approach where the likelihood function becomes informative. In this approach, the variable y can take only a finite set of values $\tilde{y}_t$, $t = 1, \ldots, T$ and N_t is the frequency of $\tilde{y}_t$ so that $N = \sum_{t=1}^{T} N_t$, the population size and $\overline{Y} = \sum_{t=1}^{T} N_t \tilde{y}_t / N$, the population mean of $y \cdot \tilde{y}_t$ is called the scale point and N_t is the corresponding scale load. Let a sample s of size n be selected from the population U using simple random sampling without replacement (SRSWOR) sampling design and n_t be the number of units having y-values equal to $\tilde{y}_t$. Then the likelihood function of $\mathbf{N} = (N_1, \ldots, N_i, \ldots, N_T)$ is given by

$$L(\mathbf{N}) = \frac{\binom{N_1}{n_1} \cdots \binom{N_t}{n_t} \cdots \binom{N_T}{n_T}}{\binom{N}{n}} \tag{25.2.2}$$

For $N \to \infty$, $N_i/N \to p_i$, and $n/N \to 0$, the likelihood function of $\mathbf{p} = (p_1, \ldots, p_i, \ldots, p_T)$ can be approximated as

$$L(\mathbf{p}) \cong \frac{n!}{n_1! \cdots n_i! \cdots n_T!} \; p_1^{n_1} \cdots p_i^{n_i} \cdots p_T^{n_T} \tag{25.2.3}$$

The maximum likelihood estimate (MLE) of $\mathbf{p}$ is $\widehat{\mathbf{p}} = \left(\widehat{p}_1, \ldots, \widehat{p}_i, \ldots, \widehat{p}_T\right)$ where $\widehat{p}_i = n_i / n$ and the MLE of $\overline{Y} = \sum_{t=1}^{T} p_t \tilde{y}_t$ is $\overline{y} = \sum_{t=1}^{T} \widehat{p}_t \tilde{y}_t$, the sample mean.

Consider the situation where the information on the auxiliary variable x, closely related to the study variable y, is available and the scale points of x be $\tilde{x}_j$ with $j = 1, \ldots, J$. Let the scale load of $\left(\tilde{y}_t, \tilde{x}_j\right)$ in the population (sample of size n) be N_{tj} (n_{tj}). In this case we have $\overline{X} = \sum_{t=1}^{T} \sum_{j=1}^{J} \tilde{x}_j p_{tj}$ and $\overline{Y} = \sum_{t=1}^{T} \sum_{j=1}^{J} \tilde{y}_j p_{tj}$ where $p_{tj} = N_{tj}/N$ and $\sum_{t=1}^{T} \sum_{j=1}^{J} p_{tj} = 1$. The scale load estimator of $\overline{Y}$ proposed by Hartley and Rao (1968) is

$$\widehat{\overline{Y}}_{HR} = \sum_{t=1}^{T} \sum_{j=1}^{J} \widehat{p}_{tj} \tilde{y}_j \tag{25.2.4}$$

where $\widehat{p}_{tj} (= n_{tj}/n)$'s maximize the log likelihood function

$$\sum_{t=1}^{T} \sum_{j=1}^{J} n_{tj} \log(p_{tj})$$

subject to constraints $p_{tj} > 0, \sum_{t=1}^{T} \sum_{j=1}^{J} p_{tj} = 1$ and $\overline{X} = \sum_{t=1}^{T} \sum_{j=1}^{J} \widetilde{x}_j p_{tj}$.

Hartley and Rao (1968) showed that the estimator $\widehat{\overline{Y}}_{HR}$ is asymptotically equivalent to the customary regression estimator of $\overline{Y}$. Hartley and Rao (1968) also considered the probability proportional to size with replacement (PPSWR) sampling scheme with x as a measure of size variable and taking scale points of y_i/x_i as $\widetilde{r}_t$, $t = 1,\ldots, T$. The resulting scale load estimator was reduced to the well-known Hansen–Hurwitz estimator. Rao and Wu (2009) pointed out that scale load approach for unequal probability sampling scheme without replacement may not yield any useful solution.

25.3 EMPIRICAL LIKELIHOOD APPROACH

Let $y_1,\ldots, y_i,\ldots, y_n$ be independently and identically distributed random variables with the common cumulative distribution function (CDF) $F(y_i) = P(y \le y_i)$. The EL (nonparametric) function of the CDF (F) is

$$L(F) = \prod_{i=1}^{n} [F(y_i) - F(y_i-)] \tag{25.3.1}$$

where $F(y_i-) = P(y < y_i)$.

Thus $L(F)$ is the probability of getting exactly the observation $y_1,\ldots, y_i,\ldots, y_n$ from the distribution function F. Thus $L(F) = 0$ if F is continuous. Let $p_i = F(y_i) - F(y_i-)$ be the probability of $y = y_i$ and we denote the EL function as

$$L(F) = L(\mathbf{p}) = \prod_{i=1}^{n} p_i \tag{25.3.2}$$

The empirical CDF (ECDF) of $y_1,\ldots, y_i,\ldots, y_n$ is

$$S_n(y) = \frac{1}{n} \sum_{i=1}^{n} I(y_i \le y) \tag{25.3.3}$$

where

$$I(y_i \le y) = \begin{cases} 1 & \text{if } \; y_i \le y \\ 0 & \text{otherwise} \end{cases}$$

25.4 EMPIRICAL LIKELIHOOD FOR SIMPLE RANDOM SAMPLING

The application of EL method in survey sampling was introduced by Chen and Quin (1993). The method is described as follows. Let a sample s of size n be selected from a finite population U of N units by simple random sampling with or without replacement. For the case of without replacement the sampling fraction n/N is assumed to be negligible so that the Owen's EL results described in Section 25.3 can be applied directly. Chen and Quin (1993) defined EL function for the finite population as

$$L(\mathbf{p}) = \prod_{i \in s} p_i \tag{25.4.1}$$

where p_i is the probability mass associated with the ith unit.

Maximizing $L(\mathbf{p})$ subject to $p_i > 0$ and $\sum_{i \in s} p_i = 1$, the optimum value of p_i is obtained as $\widehat{p}_i = 1/n$. The maximum empirical likelihood (MEL) estimator of the parameter $\theta = \sum_{i \in U} g(y_i)/N$ is given by

$$\widehat{\theta}_{mel} = \sum_{i \in s} \widehat{p}_i g(y_i) = \sum_{i \in s} g(y_i)/n \tag{25.4.2}$$

where $g(y_i)$ is a known function of y_i.

In particular if $g(y_i) = y_i$, we get MEL estimator of the population mean $\overline{Y} = \sum_{i \in U} y_i/N$ is the sample mean

$$\widehat{\overline{Y}}_{mel} = \sum_{i \in s} y_i/n = \overline{y} \tag{25.4.3}$$

For the choice of $g(y_i) = I(y_i \leq y)$, θ reduces to the population distribution function $F(y) = \sum_{i \in U} I(y_i \leq y)/N$ and the MEL of $F(y)$ is the sample distribution function

$$\widehat{F}_{mel}(y) = \sum_{i \in s} I(y_i \leq y)/n \tag{25.4.4}$$

Consider the situation where the auxiliary information of the finite population can be summarized as

$$E\{w(x)\} = \sum_{i=1}^{N} w(x_i)/N = \{w(x_1) + \cdots + w(x_i) + \cdots + w(x_N)\}/N = 0 \tag{25.4.5}$$

where $w(x)$ is a known function of the auxiliary variable x. In this case, MEL of θ is obtained by maximizing

$$l(\mathbf{p}) = \log\{L(\mathbf{p})\} = \sum_{i\in s} \log(p_i)$$

subject to

$$p_i > 0, \sum_{i\in s} p_i = 1 \text{ and } \sum_{i\in s} p_i w_i(x) = 0 \tag{25.4.6}$$

No explicit expression of the solution of Eq. (25.4.6) is obtained. However, numerical solution of p_i's is obtained from

$$\widehat{p}_i = \frac{1}{n\{1 + \lambda w(x_i)\}} \text{ for } i \in s \tag{25.4.7}$$

where λ is a Lagrange multiplier satisfying

$$\sum_{i\in s} \frac{w(x_i)}{1 + \lambda w(x_i)} = 0$$

As regards to the existence of solution of p_i, Chen and Quin (1993) pointed out that for large n, the solution is most likely to exist. However, Eq. (25.4.7) fails to have proper solution when the convex hull $\{w(x_i), i \in s\}$ does not contain zero. For $w(x_i) = x_i - \overline{X}$ with $\overline{X} = \sum_{i=1}^{N} x_i/N$, Rao and Wu (2009) pointed out that unique solution of p_i of Eq. (25.4.7) exists if $\overline{X}$ is an inner point of the convex hull $\{x_i, i \in s\}$ and this happen with probability 1 as $n \to \infty$.

Thus MEL estimator of the population mean $\overline{Y}$ and distribution function $F(y)$ are, respectively, given by

$$\widehat{\overline{Y}}_{mel} = \sum_{i\in s} \widehat{p}_i y_i \text{ and } \widehat{F}_{mel}(y) = \sum_{i\in s} \widehat{p}_i I(y_i \le y) \tag{25.4.8}$$

where $\widehat{p}_i$ satisfies Eq. (25.4.7).

It should be noted that $\widehat{F}_{mel}(y)$ is a proper distribution function in the sense that it satisfies all the properties of a distribution function.

25.5 PSEUDO—EMPIRICAL LIKELIHOOD METHOD

The expression of the EL function for the general unequal probability without replacement sampling design cannot be obtained because the expression for the joint probability function of the sample is not feasible.

Chen and Sitter (1999) proposed pseudo−empirical likelihood (PEL) for the unequal probability sampling schemes. Under this sampling scheme, it is assumed that population vector $\mathbf{y} = (y_1, \ldots, y_i, \ldots, y_N)$ is a sample from a superpopulation. For simplicity let us assume that y_i's are independently and identically distributed random variables with distribution function $F(y)$. Then the empirical likelihood of the entire population $\mathbf{y}$ is $L(\mathbf{p}) = \prod_{i=1}^{N} p_i$ and log empirical likelihood is

$$l(\mathbf{p}) = \sum_{i=1}^{N} \log\ p_i \tag{25.5.1}$$

Suppose a sample s is selected from the population with probability $p(s)$ so that the inclusion probability of the ith unit $\pi_i \left(= \sum_{s \supset i} p(s) \right)$ is positive for every $i = 1, \ldots, N$. An unbiased estimator of $l(\mathbf{p})$ based on the sample s is given by

$$\widehat{l}(\mathbf{p}) = \sum_{i \in s} w_i(s) \log p_i \tag{25.5.2}$$

where $w_i(s)$ are suitably chosen weights satisfying unbiasedness condition $\sum_{s \supset i} w_i(s) p(s) = 1$ for $i = 1, \ldots, N$. The estimator $\widehat{l}(\mathbf{p})$ is termed by Chen and Sitter (1999) as "pseudo−empirical log likelihood function." Maximizing $\widehat{l}(\mathbf{p})$ subject to constraints $p_i > 0$ and $\sum_{i \in s} p_i = 1$, yield the optimum value of p_i as

$$\widehat{p}_i = \frac{w_i(s)}{\sum_{i \in s} w_i(s)} \tag{25.5.3}$$

The maximum PEL (MPEL) for a parametric function $\overline{G} = \sum_{i \in U} g(y_i)/N$ is given by

$$\begin{aligned} \widehat{\overline{G}} &= \sum_{i \in s} \widehat{p}_i g(y_i) \\ &= \frac{\sum_{i \in s} w_i(s) g(y_i)}{\sum_{i \in s} w_i(s)} \end{aligned} \tag{25.5.4}$$

The estimator (25.5.4) is similar to the Hájek (1964) estimator.

25.5.1 MPEL Estimator for the Population Mean

Substituting $g(y_i) = y_i$ in Eq. (25.5.4), we get MPEL estimator for the population mean $\overline{Y}$ as

$$\widehat{\overline{Y}}_{mpel} = \frac{\sum_{i\in s} w_i(s) y_i}{\sum_{i\in s} w_i(s)} \tag{25.5.5}$$

In case $w_i(s) = 1/\pi_i$, $\widehat{\overline{Y}}_{mpel}$ reduces to $\widehat{\overline{Y}}_{mpel} = \dfrac{\sum_{i\in s} y_i/\pi_i}{\sum_{i\in s} 1/\pi_i}$. For SRSWOR sampling $\pi_i = n/N$ and $\widehat{\overline{Y}}_{pmel} = \overline{y}$ = sample mean.

25.5.2 MPEL Estimator for the Population Distribution Function

Substituting $g(y_i) = I(y_i \le t)$ in Eq. (25.5.4) MPEL estimator for the distribution function $F(t)$ is obtained as

$$\widehat{F}_{mpel}(t) = \frac{\sum_{i\in s} w_i(s) I(y_i \le t)}{\sum_{i\in s} w_i(s)} \tag{25.5.6}$$

In particular for $w_i(s) = 1/\pi_i$, the estimator $\widehat{F}_{pmel}(t)$ reduces to $\widehat{F}_{mpel}(t) = \dfrac{\sum_{i\in s} I(y_i \le t)/\pi_i}{\sum_{i\in s} 1/\pi_i}$. For SRSWOR $\widehat{F}_{mpel}(t)$ reduces to empirical distribution function $S(t) = \dfrac{1}{n}\sum_{i\in s} I(y_i \le t)$.

25.5.3 MPEL Estimator Under Linear Constraints

Suppose that the population mean $\overline{X}$ of the auxiliary variable x in is known. In this situation, p_i's for $i \in s$ are obtained by maximizing the pseudo−empirical log likelihood $\widehat{l}(\mathbf{p}) = \sum_{i\in s} w_i(s) \log p_i$ subject to the constraints

$$\text{(i)}\ \ p_i \ge 0, \text{(ii)} \sum_{i\in s} p_i = 1 \ \text{ and } \ \text{(iii)} \sum_{i\in s} p_i u(x_i) = 0 \tag{25.5.7}$$

where $u_i = u(x_i) = x_i - \overline{X}$.

The solution of the minimization problem above has no closed form. It is obtained numerically by solving the following equations.

$$\widehat{p}_i = \frac{\widetilde{w}_i(s)}{1+\lambda\, u_i} \tag{25.5.8}$$

where $\widetilde{w}_i(s) = w_i(s) \Big/ \sum_{i\in s} w_i(s)$ and λ is the Lagrange multiplier satisfying

$$\sum_{i\in s} \frac{\widetilde{w}_i(s) u_i}{1+\lambda\;\; u_i} = 0 \tag{25.5.9}$$

Rao and Wu (2009) presented a detailed algorithm for the solution of λ and $\widehat{p}_i$ given in Eq. (25.5.8).

Thus the MPEL estimator for the population distribution function $F(t)$ under constraint $\sum_{i\in U} u(x_i) = 0$ is given by

$$\widehat{F}_{mpel}(t) = \sum_{i\in s} \widehat{p}_i I(y_i \le t) \tag{25.5.10}$$

where $\widehat{p}_i$ is given in Eq. (25.5.8). The estimator $\widehat{F}_{mpel}(t)$ is a proper distribution function since $\widehat{p}_i > 0$ and $\sum_{i\in s} \widehat{p}_i = 1$ and estimators of quintiles can be obtained by direct inversion of $\widehat{F}_{mpel}(t)$.

25.6 ASYMPTOTIC BEHAVIOR OF MPEL ESTIMATOR

Chen and Sitter (1999) derived the asymptotic behavior of the MPEL estimator of the population mean. The theorem is given below without derivation.

Theorem 25.6.1

Under the regularity conditions stated below, $\widehat{\overline{Y}}_{mpel}$, the MPEL estimator of the population mean $\overline{Y}$, when $\overline{X}$, the population mean of the auxiliary variable is known, is asymptotically equivalent to the generalized regression estimator (GREG), i.e.,

$$\widehat{\overline{Y}}_{mpel} = \widehat{\overline{Y}}_{greg} + o_p\left(n^{-1/2}\right)$$

where $\widehat{\overline{Y}}_{greg} = \overline{y}_w - \widehat{B}_{greg}\left(\overline{x}_w - \overline{X}\right)$, $\overline{x}_w = \sum_{i\in s} \widetilde{w}_i(s)\, x_i$, $\overline{y}_w = \sum_{i\in s} \widetilde{w}_i(s)\, y_i$,

$$\widetilde{w}_i(s) = w_i(s) \Big/ \sum_{i\in s} w_i(s), \text{ and } \widehat{B}_{greg} = \frac{\sum_{i\in s} \widetilde{w}_i(s)(x_i - \overline{x}_w) y_i}{\sum_{i\in s} \widetilde{w}_i(s)(x_i - \overline{x}_w)^2}.$$

In this case the Lagrange multiplier λ is given by

$$\lambda = \frac{\overline{x}_w - \overline{X}}{\sum_{i \in s} \widetilde{w}_i(s)\left(x_i - \overline{x}_w\right)^2} + o_p\left(n^{-1/2}\right)$$

The regularity conditions:

$$\text{(i)} \quad u^* = \max_{i \in s}|u_i| = o_p\left(n^{1/2}\right) \quad \text{and} \quad \text{(ii)} \quad \frac{\sum_{i \in s} w_i(s)u_i}{\sum_{i \in s} w_i(s)u_i^2} = O_p\left(n^{-1/2}\right),$$

where $u_i = x_i - \overline{X}$.

Remark 25.6.1

Chen and Sitter (1999) showed that many commonly used sampling designs satisfy the regularity conditions above such as PPSWR and Rao-Hartley-Cochran (1962) sampling designs.

25.6.1 GREG Estimator Versus MPEL Estimator

Deville and Särndal (1992) proposed calibrated Horvitz–Thompson estimator $\widehat{\overline{Y}}_{HT} = \frac{1}{N}\sum_{i \in s}\frac{y_i}{\pi_i}$ of the population mean $\overline{Y}$ as

$$\widehat{\overline{Y}}_c = \sum_{i \in s} w_i y_i \tag{25.6.1}$$

Here w_i's are the calibrated design weights that are obtained by minimizing distance between w_i's and $d_i\left(= \frac{1}{N\pi_i}\right)$'s subject to the constraints

$$\sum_{i \in s} w_i x_i = \overline{X} \tag{25.6.2}$$

The commonly used distance measure is the chi-squared distance

$$\phi = \sum_{i \in s}\frac{\left(w_i - d_i\right)^2}{d_i q_i} \tag{25.6.3}$$

with q_i's as prespecified weights. Several alternative distance measures were proposed by Deville and Särndal (1992).

The minimization of Eq. (25.6.3) subject to Eq. (25.6.2) yields

$$w_i = d_i - \frac{d_i x_i q_i}{\sum_{i \in s} d_i x_i^2 q_i}\left(\widehat{\overline{X}}_{HT} - \overline{X}\right) \tag{25.6.4}$$

where $\widehat{\overline{X}}_{HT} = \sum_{i \in s} d_i x_i$.

Substituting the values of w_i obtained from Eq. (25.6.4) in Eq. (25.6.1) yields calibrated estimator as

$$\widehat{Y}_c = \widehat{\overline{Y}}_{ht} - \widehat{B}_{ht}\left(\widehat{\overline{X}}_{HT} - \overline{X}\right) \tag{25.6.5}$$

where

$$\widehat{B}_{ht} = \frac{\sum_{i\in s} y_i d_i q_i x_i}{\sum_{i\in s} y_i d_i q_i x_i^2} \tag{25.6.6}$$

The estimator Eq. (25.6.6) is the well-known generalized regression estimator. It is worth to note that the weights w_i's given in Eq. (25.6.4) may take negative values whereas the weights $\widehat{p}_i$'s of the MPEL estimator are always nonnegative. Thus the GREG estimator for a distribution function may not be a genuine distribution function whereas an MPEL estimator for a distribution function is always a genuine distribution function.

25.7 EMPIRICAL LIKELIHOOD FOR STRATIFIED SAMPLING

Suppose that a population of N units is stratified into H strata of sizes $N_1,\ldots,$ N_H and $W_h = N_h/N$. Let y_{hi} and x_{hi} be the values of the study and auxiliary variables for the ith unit of the hth stratum, $i = 1,\ldots,\ N_h;\ h = 1,\ldots,\ H$. Suppose a sample s_h of size n_h is selected from the hth stratum by SRSWOR. Let the samples be selected independently from each of the stratum and sampling fractions n_h/N_h are negligible for each $h = 1,\ldots,\ H$. Assuming $z_{hi} = (x_{hi}, y_{hi})$'s are independently distributed with distribution function $F_h, i = 1,\ldots,\ n_h, h = 1,\ldots,\ H$. Following Chen and Quin (1993), the log empirical likelihood for the stratified random sampling is obtained as

$$l(\mathbf{p}_1,\ldots,\mathbf{p}_\mathbf{H}) = \sum_{h=1}^{H}\sum_{i\in s_h}\log(p_{hi}) \tag{25.7.1}$$

where p_{hi} is the probability mass assigned to z_{hi}, $i \in\ s_h$, and $\mathbf{p_h} = (p_{h1},\ldots,p_{hn_h})'$. In case the population mean $\overline{X}$ of the auxiliary variable x is known, the MEL estimator of the population mean $\overline{Y}$ and distribution function F are, respectively, given by

$$\widehat{\overline{Y}}_{EL}^{st} = \sum_{h=1}^{H} W_h \sum_{i\in s_h}\widehat{p}_{hi}y_{hi} \tag{25.7.2}$$

and

$$\widehat{F}_{EL}^{st}(t) = \sum_{h=1}^{H} W_h \sum_{i\in s_h}\widehat{p}_{hi}I(y_{hi}\le t) \tag{25.7.3}$$

where $\widehat{p}_{hi}$'s maximize $l(\mathbf{p_1},\ldots,\mathbf{p_H})$ subject to,

(i) $p_{hi} \geq 0$, (ii) $\sum_{i\in s_h} p_{hi} = 1$, and (iii) $\sum_{h=1}^{H} W_h \sum_{i\in s_h} p_{hi}x_{hi} = \overline{X}$.

For maximization of $l(\mathbf{p_1},\ldots,\mathbf{p_H})$, we consider the following function ϕ where λ_h and λ are Lagrange multipliers and $n = \sum_h n_h$.

$$\begin{aligned}\phi = \sum_h \sum_{i\in s_h} \log p_{hi} - \sum_h \lambda_h\left(\sum_{i\in s_h} p_{hi} - 1\right)\\ - n\,\lambda\left(\sum_h W_h \sum_{i\in s_h} p_{hi}x_{hi} - \overline{X}\right)\end{aligned} \tag{25.7.4}$$

Maximization of $l(\mathbf{p_1},\ldots,\mathbf{p_H})$ yields

$$p_{hi} = \frac{1}{n_h\left[1 + m_h\,\lambda\left(x_{hi} - \overline{X}_h^*\right)\right]} \tag{25.7.5}$$

where $\overline{X}_h^* = \sum_{i\in s_h} p_{hi}x_{hi}$ and $m_h = nW_h/n_h$.

Zhong and Rao (2000) showed that the solution of the system of Eq. (25.7.3) exists with probability tending to 1 as the sample sizes tend to infinity for each stratum. However, for deep stratification where the number of strata H is large and n_h's are small, the nature of solution is unknown.

25.7.1 Asymptotic Properties

Zhong and Rao (2000) derived the following result relating to asymptotic properties of empirical likelihood estimators by considering a sequence of stratified populations indexed by a label ν. The result is stated below without derivation.

Theorem 25.7.1

Suppose that, as $\nu \to \infty$, N_h, n_h and $N_h - n_h$ tends to infinity, $n_h/n \to k_h(>0)$, both $\sum_h W_h^2 \frac{1}{N_h}\sum_{i=1}^{N_h} |x_{hi}|^3$ and $\sum_h W_h^2 \frac{1}{N_h}\sum_{i=1}^{N_h} |y_{hi}|^3$ have an upper bound independent of ν, and $S_{hxx} = \frac{1}{N_h - 1}\sum_{i=1}^{N_h}\left(x_{hi} - \overline{X}_h\right)^2 \geq S_0 > 0$ $\forall\, h$ and ν. Then

$$\frac{\widehat{\overline{Y}}{}_{EL}^{st} - \overline{Y}}{\sigma_E}$$

asymptotically normally distributed with mean zero and variance unity as $\nu \to \infty$ where

$$\sigma_E^2 = \sum_h W_h^2\left(\frac{1}{n_h} - \frac{1}{N_h}\right)\left(S_{hyy} - 2BS_{hxy} + B^2 S_{hxx}\right),$$

$$B = \left(\sum_h \frac{W_h^2 S_{hxy}}{n_h}\right) \Bigg/ \left(\sum_h \frac{W_h^2 S_{hxx}}{n_h}\right),$$

$$S_{hyy} = \frac{1}{N_h - 1}\sum_{i=1}^{N_h}\left(y_{hi} - \overline{Y}_h\right)^2 \text{ and } S_{hxy} = \frac{\sum_{i=1}^{N_h}\left(x_{hi} - \overline{X}_h\right) y_{hi}}{N_h - 1}.$$

From the Theorem 25.7.1 above, it follows that $\widehat{\overline{Y}}_{EL}^{st}$ approximately possesses the same asymptotic variance of the optimum regression estimator

$$\widehat{\overline{Y}}_{lr} = \overline{y}_{st} - \widehat{B}_0\left(\overline{x}_{st} - \overline{X}\right) \qquad (25.7.6)$$

where $\overline{x}_{st} = \sum_{h=1}^{H} W_h \frac{1}{n_h}\sum_{i\in s_h} x_{hi}$, $\overline{y}_{st} = \sum_{h=1}^{H} W_h \frac{1}{n_h}\sum_{i\in s_h} y_{hi}$,

$\widehat{B}_0 = \sum_h W_h^2\left(\frac{1}{n_h} - \frac{1}{N_h}\right) s_{hxy} \Big/ \sum_h W_h^2\left(\frac{1}{n_h} - \frac{1}{N_h}\right) s_{hxx}$ with

$$s_{hxy} = \frac{\sum_{i\in s_h}\left(x_{hi} - \overline{x}_h\right) y_{hi}}{n_h - 1}, \text{ and } s_{hxx} = \frac{\sum_{i\in s_h}\left(x_{hi} - \overline{x}_h\right)^2}{n_h - 1}.$$

25.7.1.1 Variance Estimation

Under the conditions of Theorem 25.7.1, it can be shown that a consistent estimator of σ_E^2 (the asymptotic variance of $\widehat{\overline{Y}}_{EL}^{st}$) is

$$s_E^2 = \sum_h W_h^2\left(\frac{1}{n_h} - \frac{1}{N_h}\right)\frac{1}{n_h - 1}\sum_{i\in s_h}\left\{y_{hi} - \overline{y}_h^* - \widehat{B}\left(x_{hi} - \overline{x}_h^*\right)\right\}^2$$

where

$$\widehat{B} = \sum_h \frac{W_h^2}{n_h}\sum_{i\in s_h}\left(y_{hi} - \overline{y}_h^*\right)\left(x_{hi} - \overline{x}_h^*\right) \Big/ \sum_h \frac{W_h^2}{n_h}\sum_{i\in s_h}\left(x_{hi} - \overline{x}_h^*\right)^2,$$

$\overline{y}_h^* = \sum_{i\in s_h}\widehat{p}_{hi} y_{hi}$, and $\overline{x}_h^* = \sum_{i\in s_h}\widehat{p}_{hi} x_{hi}$.

25.7.1.2 Jackknife Variance Estimation

Let $\widehat{\overline{Y}}_{EL}^{st}(h_j)$ be the estimator of $\overline{Y}$ when jth sample observation of the h stratum is deleted, then the Jackknife variance estimator of σ_E^2 is given by

$$s_{JE}^2 = \sum_h (n_h - 1)\left(\frac{1}{n_h} - \frac{1}{N_h}\right) \sum_{j=1}^{n_h} \left(\widehat{\overline{Y}}_{EL}^{st}(h_j) - \widehat{\overline{Y}}_{EL}\right)^2$$

Zhong and Rao (2000) showed that $n\left(s_{JE}^2 - \sigma_E^2\right)$ converges in probability to 0.

25.7.2 Pseudo–empirical Likelihood Estimator

Suppose that samples s_h of size n_h are selected independently from each of the strata by using unequal probability sampling schemes so that the inclusion probability of the ith unit of the hth stratum π_i^h is positive for every $i = 1,\ldots, N_h$, $h = 1,\ldots, H$. Let us further assume that z_{hi}'s independently distributed with distribution function F_h, then the log likelihood function of the entire stratified population becomes

$$l(\mathbf{p}_1, \ldots, \mathbf{p}_\mathbf{H}) = \sum_{h=1}^{H} \sum_{i=1}^{N_h} \log(p_{hi})$$

A design-unbiased estimate of the population empirical log likelihood $l(\mathbf{p_1},\ldots, \mathbf{p_H})$ is given by

$$\widehat{l}(\mathbf{p}_1, \ldots, \mathbf{p}_\mathbf{H}) = \sum_{h=1}^{H} \sum_{i\in s_h} d_{hi} \log(p_{hi}) \tag{25.7.7}$$

where $d_{hi} = 1/\pi_i^h$.

In case $\overline{X}$ is known, maximizing $\widehat{l}(\mathbf{p}_1, \ldots, \mathbf{p}_\mathbf{H})$ subject to (i) $p_{hi} \geq 0$, (ii) $\sum_{i\in s_h} p_{hi} = 1$, and (iii) $\sum_{h=1}^{H} \sum_{i\in s_h} p_{hi}\left(x_{hi} - \overline{X}\right) = 0$, Chen and Sitter (1999) derived the expression of the MPEL estimator of the population mean $\overline{Y}$ as

$$\widehat{\overline{Y}}_{MPEL}^{st} = \overline{y}_w - \widehat{B}_{greg}^{st}\left(\overline{x}_w - \overline{X}\right) + o_p\left(n^{-1/2}\right) \tag{25.7.8}$$

where $\widehat{B}_{greg}^{st} = \dfrac{\sum_h \sum_{i\in s_h} w_{hi}(s)(x_{hi} - \overline{x}_w)y_{hi}}{\sum_h \sum_{i\in s_h} w_{hi}(s)(x_{hi} - \overline{x}_w)^2}$, $n = \sum_h n_h$, $\overline{y}_w = \sum_h \sum_{i\in s_h} w_{hi}(s)\ y_{hi}$, $\overline{x}_w = \sum_h \sum_{i\in s_h} w_{hi}(s)\ x_{hi}$, and $w_{hi}(s) = d_{hi}\Big/\sum_h \sum_{i\in s_h} d_{hi}$

For SRSWOR, $d_{hi} = N_h/n_h$ and the above expression (25.7.8) reduces to

$$\widehat{\overline{Y}}_{MPEL} = \overline{y}^{st}_{greg} + o_p\left(n^{-1/2}\right) \tag{25.7.9}$$

where $\overline{y}^{st}_{greg} = \overline{y}_{st} - \dfrac{\sum\limits_h \sum\limits_{i \in s_h} W_h(x_{hi} - \overline{x}_{st})y_{hi}\Big/ n_h}{\sum\limits_h \sum\limits_{i \in s_h} W_h(x_{hi} - \overline{x}_{st})^2\Big/ n_h}\left(\overline{x}_{st} - \overline{X}\right), \quad \overline{y}_{st} = \sum W_h \overline{y}_h$

and $\overline{x}_{st} = \sum W_h \overline{x}_h$.

25.7.2.1 Multistage Sampling

Let us consider a finite population U, which is stratified into H strata. The hth strata consists of N_h first-stage units (fsu's) and the jth fsu of the hth stratum consists of N_{hj} second-stage units (ssu's). Let y_{hij} and x_{hij} be the values of the study and auxiliary variables for the jth ssu of the ith fsu of the hth stratum. Then the population totals of y and x are $Y = \sum\limits_{h=1}^{H} \sum\limits_{i=1}^{N_h} \sum\limits_{j=1}^{N_{hi}} y_{hij}$ and $X = \sum\limits_{h=1}^{H} \sum\limits_{i=1}^{N_h} \sum\limits_{j=1}^{N_{hi}} x_{hij}$, respectively. From the h (=1,…, H)th stratum, a subsample s_h of n_h fsu's is selected by same suitable sampling design with inclusion probability $\pi_i^h(>0)$ for the ith fsu of the hth stratum. If the ith fsu is selected in s_h, a subsample s_{hi} of n_{hi} units is selected from it by some suitable sampling design with inclusion probability $\pi_j^{ih}(>0)$ for the jth ssu of the ith fsu of the hth stratum. Here the sample sizes n_{hj} are predetermined numbers. Let $y_{hi1},\ldots, y_{hiNi}$ be distributed independently with distribution function F_{hi} and p_{hij} probability mass attached to y_{hij}. The log likelihood function of the entire population is given by

$$L = \sum_{h=1}^{H} \sum_{i=1}^{N_h} \sum_{j=1}^{N_{hi}} \log(p_{hij})$$

An unbiased estimate of the log likelihood L is given by

$$\widehat{L} = \sum_{h=1}^{H} \frac{1}{\pi_i^h} \sum_{i \in s_h} \frac{1}{\pi_j^{hi}} \sum_{j \in s_{hi}} \log(p_{hij}) = \sum_{h=1}^{H} \sum_{i \in s_h} \sum_{j \in s_{hi}} d_{hij} \log(p_{hij})$$

where $d_{hij} = \dfrac{1}{\pi_i^h \pi_j^{hi}}$.

Now maximizing $\widehat{L}$ subject to (i) $p_{hij} > 0$ and (ii) $\sum\limits_{i \in s_h} \sum\limits_{j \in s_{hi}} p_{hij} = 1$, the MPEL estimator for the population mean becomes a ratio estimator

$$\widehat{\overline{Y}}_{mst} = \sum_{h=1}^{H} \sum_{i \in s_h} \sum_{j \in s_{hi}} \widehat{p}_{hij} y_{hij} \tag{25.7.10}$$

where $\widehat{p}_{hij} = d_{hij} \Big/ \sum\limits_{h=1}^{H} \sum\limits_{i \in s_h} \sum\limits_{j \in s_{hi}} d_{hij}$.

25.8 MODEL-CALIBRATED PSEUDOEMPIRICAL LIKELIHOOD

Wu and Sitter (2001) introduced the model-calibrated PEL method where the study variable y is related to a vector of auxiliary variable $\mathbf{x}_i' = (1, x_{i1}, \ldots, x_{ip})$ through the following superpopulation model ξ

$$E_\xi(y_i|\mathbf{x}_i) = \mu(\mathbf{x}_i, \boldsymbol{\theta}) = \mu_i \text{ and } V_\xi(y_i|\mathbf{x}_i) = \sigma^2 v_i^2; i = 1, \ldots, N \qquad (25.8.1)$$

where $\boldsymbol{\theta} = (\theta_0, \theta_1, \ldots, \theta_p)'$ and σ^2 are unknown model parameters, $\mu(\mathbf{x}_i, \boldsymbol{\theta})$ is a known function of $\mathbf{x}_i$ and $\boldsymbol{\theta}$, v_i is a known function of $\mathbf{x}_i$, and E_ξ and V_ξ denote the expectation and variance, respectively, with respect to the model ξ. We also assume that $(y_1, \mathbf{x}_1), \ldots, (y_N, \mathbf{x}_N)$ are mutually independent.

Let a sample s of size n be selected from the population with inclusion probability for the ith, and i and jth unit $(i \neq j)$ be π_i and π_{ij}, respectively. In the model-based approach, $(y_i, \mathbf{x}_i), i \in s$ are regarded as a random sample from the superpopulation ξ. Here we consider the design-based approach where $(y_i, \mathbf{x}_i), i \in s$ are not viewed as a random sample from the superpopulation. The model parameter $\boldsymbol{\theta}$ is regarded an estimate of $\boldsymbol{\theta}$ based on the entire population of N units and will be denoted by $\boldsymbol{\theta}_N$. For a linear regression model $\mu_i = \mathbf{x}_i'\boldsymbol{\theta}$, $\boldsymbol{\theta}_N$ is defined as

$$\boldsymbol{\theta}_N = (\mathbf{X}_N'\mathbf{X}_N)^{-1}\mathbf{X}_N'\mathbf{y}_N \qquad (25.8.2)$$

where $\mathbf{X}_N = \begin{pmatrix} \mathbf{x}_1' \\ \cdots \\ \mathbf{x}_N' \end{pmatrix}$, $\mathbf{y}_N = \begin{pmatrix} y_1 \\ \cdots \\ y_N \end{pmatrix}$.

A design-based estimate of $\boldsymbol{\theta}_N$ is

$$\widehat{\boldsymbol{\theta}} = \left(\mathbf{X}_n'\Pi^{-1}\mathbf{X}_n\right)^{-1}\mathbf{X}_n'\Pi^{-1}\mathbf{y}_n$$

where $\Pi = diag(\pi_1, \ldots, \pi_n)$, $\mathbf{X}_n = \begin{pmatrix} \mathbf{x}_1' \\ \cdots \\ \mathbf{x}_n' \end{pmatrix}$, $\mathbf{y}_n = \begin{pmatrix} y_1 \\ \cdots \\ y_n \end{pmatrix}$ assuming $s = (1, \ldots, n)$.

25.8.1 Estimation of the Population Mean

Let $\widehat{\mu}_i = \mu\left(\mathbf{x}_i, \widehat{\boldsymbol{\theta}}\right)$ is the predicted value of y_i based on the model ξ and $\widehat{\boldsymbol{\theta}}$ is design-based estimator of $\boldsymbol{\theta}$. Here we maximize the PEL function

$$\widehat{l}(\mathbf{p}) = \sum_{i \in s} d_i \log p_i \text{ with } d_i = 1/\pi_i \qquad (25.8.3)$$

$$\text{under constraints (i)} \sum_{i \in s} p_i = 1 \text{ and (ii)} \sum_{i \in s} p_i \widehat{\mu}_i = \frac{1}{N} \sum_{i \in U} \widehat{\mu}_i \quad (25.8.4)$$

It is to be noted that for the linear model $\mu_i = \mathbf{x}_i'\boldsymbol{\theta}$, $\frac{1}{N}\sum_{i \in U} \widehat{\mu}_i$ reduces to $\overline{\mathbf{X}}'\widehat{\boldsymbol{\theta}}$, so only the vector of the population mean $\overline{\mathbf{X}}$ of the auxiliary variables are in need for the constraint (25.8.4) but nonlinear model $\mu_i = \mathbf{x}_i'\boldsymbol{\theta}$, the constraint (25.8.4) requires complete information of $\mathbf{x}_1, \ldots, \mathbf{x}_N$.

The model-calibrated MPEL estimator of the population mean $\overline{Y}$ is given by

$$\widehat{\overline{Y}}_{CMPEL} = \sum_{i \in s} \widehat{p}_i y_i$$

where $\widehat{p}_i$'s maximize $\widehat{l}(\mathbf{p})$ subject to Eq. (25.8.4).

It follows from Eq. (25.5.8) that $\widehat{p}_i = \dfrac{w_i}{1 + \lambda u_i}$ where the Lagrange multiplier λ satisfies

$$\sum_{i \in s} \frac{w_i u_i}{1 + \lambda u_i} = 0$$

with $w_i = d_i \Big/ \sum_{i \in s} d_i$ and $u_i = \widehat{\mu}_i - \dfrac{1}{N}\sum_{i \in U} \widehat{\mu}_i$.

25.8.2 Estimation of the Population Distribution Function

The population distribution function is $F_N(t) = \dfrac{1}{N}\sum_{i \in U} I(y_i \leq t)$ and the Horvitz–Thompson estimator of $F_N(t)$ is $\widehat{F}_{ht}(t) = \dfrac{1}{N}\sum_{i \in s} d_i I(y_i \leq t)$.

Under superpopulation model ξ,

$$E_\xi\{I(y_i \leq t)\} = P(y_i \leq t) = G\left(\frac{t - \mu_i}{\nu(x_i)}\right)$$

where $G(\cdot)$ is the CDF of the error term $\in_i = \dfrac{y_i - \mu_i}{\nu(x_i)}$.

For a given value of t, let us define $\widehat{G}_i = \dfrac{1}{n}\sum_{i \in s} I\left(\widehat{\in}_i \leq \dfrac{t - \hat{\mu}_i}{\nu(\mathbf{x}_i)}\right)$ with $\widehat{\in}_i = \dfrac{y_i - \hat{\mu}_i}{\nu(x_i)}$.

The MPEL estimator of $F_N(t)$ is given by

$$\widehat{F}_{MPEL}(t) = \sum_{i \in s} \widehat{p}_i I(y_i \leq t) \quad (25.8.5)$$

where $\widehat{p}_i$ maximizes the PEL $\widehat{l}(\mathbf{p})$ given in Eq. (25.8.3) subject to

$$\text{(i)}\ \ p_i \geq 0, \text{(ii)}\ \ \sum_{i\in s} p_i = 1 \ \text{ and } \ \sum_{i\in s} p_i \widehat{G}_i = \frac{1}{N}\sum_{i\in U}\widehat{G}_i$$

The MPEL estimator $\widehat{F}_{MPEL}(t)$ derived in Eq. (25.8.5) always satisfies the properties of distribution function.

25.8.3 Model-Calibrated MPEL Estimation for Population Quadratic Parameters

The population variance and covariances can be expressed as a quadratic function of the form

$$T = \frac{1}{2}\sum_{i\neq}\sum_{j\in s}\phi(y_i, y_j) \tag{25.8.6}$$

where $\phi(y_i,y_j)$ is a symmetric function of y_i and y_j.

The pseudo log empirical likelihood function for the quadratic parameters is defined by

$$\widehat{l}^*(\mathbf{p}) = \frac{1}{2}\sum_{i\neq}\sum_{j\in s} d_{ij}\log(p_{ij})$$

where $d_{ij} = 1/\pi_{ij}$, π_{ij} inclusion probability for the ith and jth$(i \neq j)$ unit (assuming positive) and p_{ij} is the probability mass assigned to the pair of units (i,j).

The model-calibrated MPEL estimator of T is given by

$$\widehat{T}_{CMPEL} = \frac{1}{2N^*}\sum_{i\neq}\sum_{j\in s}\widehat{p}_{ij}\phi(y_i, y_j)$$

where $N^* = N(N-1)$ and $\widehat{p}_{ij}$ maximizes $\widehat{l}^*(\mathbf{p})$ subject to

$$\text{(i)}\ \ p_{ij} \geq 0, \text{(ii)}\ \ \frac{1}{2}\sum_{i\neq}\sum_{j\in s} p_{ij} = 1 \ \text{ and }\ \text{(iii)}\ \ \frac{1}{2}\sum_{i\neq}\sum_{j\in s} p_{ij}\widehat{u}_{ij} = \frac{1}{2N^*}\sum_{i\neq}\sum_{j\in U}\widehat{u}_{ij} \tag{25.8.7}$$

In the constraint (25.8.7) above, $\widehat{u}_{ij}$ is the estimate of $u_{ij} = E_\xi\{\phi(y_i,y_j)\}$, which is obtained by replacing the model parameters by their suitable design-based estimators. For example, consider the finite population variance

$$S_y^2 = \frac{1}{N-1}\sum_{i\in U}\left(y_i - \overline{Y}\right)^2 = \frac{1}{2N(N-1)}\sum_{i\neq}\sum_{j\in U}\left(y_i - y_j\right)^2$$

Here $u_{ij} = E_\xi\left(y_i - y_j\right)^2 = \{\mu(\mathbf{x}_i, \boldsymbol{\theta}) - \mu(\mathbf{x}_j, \boldsymbol{\theta})\}^2 + \sigma^2\{v_i + v_j\}$ and $\widehat{u}_{ij} = \left\{\mu\left(\mathbf{x}_i, \widehat{\boldsymbol{\theta}}\right) - \mu\left(\mathbf{x}_j, \widehat{\boldsymbol{\theta}}\right)\right\}^2 + \widehat{\sigma}^2(v_i + v_j)$, where $\widehat{\boldsymbol{\theta}}$ and $\widehat{\sigma}^2$ are design-based estimators of $\boldsymbol{\theta}$ and σ^2, respectively.

25.9 PSEUDO–EMPIRICAL LIKELIHOOD TO RAKING

Let U be a population of N units, which is cross-classified into $r \times c$ table and let U_{ij} be the set of N_{ij} units falling in the cell (i,j) such that $\sum_{i=1}^{r}\sum_{j=1}^{c} N_{ij} = N$. Let the marginal totals $N_i\cdot = \sum_{j=1}^{r} N_{ij}$ and $N._j = \sum_{i=1}^{c} N_{ij}$ of the cells are known but N_{ij}'s are unknown. To estimate the cell frequencies N_{ij}'s, a sample s of size n is selected from the population U and let $s_{ij} = s\cap U_{ij}$ be the sample of size n_{ij} for the (i,j) the cell. We can estimate N_{ij} by using estimates $\widehat{N}_{ij}(1) = n_{ij}(N/n)$, $\widehat{N}_{ij}(2) = n_{ij}(N_i\cdot/n_i\cdot)$, or $\widehat{N}_{ij}(3) = n_{ij}(N\cdot_j/n\cdot_j)$ where $n_i\cdot = \sum_{j=1}^{c} n_{ij}$ and $n\cdot_j = \sum_{i=1}^{r} n_{ij}$. But none of the estimates $\widehat{N}_{ij}(k)$ satisfies $\sum_{j=1}^{c}\widehat{N}_{ij}(k) = N_i\cdot$ and $\sum_{i=1}^{r}\widehat{N}_{ij}(k) = N\cdot_j$ for $k = 1,2,3$.

The purpose of the raking procedure is to find estimates $\widehat{N}_{ij}$ satisfying constraints

$$\sum_{j=1}^{c}\widehat{N}_{ij} = N_i\cdot \quad\text{and}\quad \sum_{i=1}^{r}\widehat{N}_{ij} = N._j. \tag{25.9.1}$$

Deming and Steaphan (1940) proposed a raking estimator of N_{ij} as

$$\widehat{N}_{ij}^{ds} = m_{ij}(N/n)$$

where m_{ij}'s are obtained by minimizing

$$\phi = \sum_{i=1}^{r}\sum_{j=1}^{c}\left(m_{ij} - n_{ij}\right)^2/n_{ij} \tag{25.9.2}$$

subject to constraints

$$\sum_{j=1}^{c} m_{ij} = N_i\cdot n/N, i = 1, \ldots, r-1 \quad\text{and}\quad \sum_{i=1}^{r} m_{ij} = N._j n/N, \qquad j = 1, \ldots, c-1 \tag{25.9.3}$$

Deming and Steaphan (1940) used the iterative proportional fitting procedure (IPFP) to solve the minimization problem. Rao and Wu (2009) pointed out that although IPFP satisfies Eq. (25.9.1), the solution does not minimize the least square distance (25.9.2). Rao and Wu (2009) gave the following solutions:

Let p_{ij} be the probability that an observation will fall in (i,j)th cell, then the log likelihood function is given by

$$l_0(\mathbf{p}) = \sum_{i=1}^{r}\sum_{j=1}^{c} n_{ij}\log(p_{ij}) \tag{25.9.4}$$

An empirical likelihood estimate of N_{ij} is given by

$$\widehat{N}_{ij}^{EL} = \widehat{p}_{ij}N \tag{25.9.5}$$

where $\widehat{p}_{ij}$'s are obtained by maximizing $l_0(\mathbf{p})$ subject to (i) $p_{ij} \geq 0$, (ii) $\sum_{i=1}^{r}\sum_{j=1}^{c} p_{ij} = 1$, and (iii) $\sum_{j=1}^{c} p_{ij} = N_{i\cdot}/N$, $i = 1,\ldots, r-1$ and $\sum_{i=1}^{r} p_{ij} = N_{\cdot j}/N$, $j = 1,\ldots,c-1$.

The solution (25.9.5) satisfies the raking conditions in (25.9.1).

25.10 EMPIRICAL LIKELIHOOD RATIO CONFIDENCE INTERVALS

Owen (1988) showed that the empirical likelihood approach provides nonparametric confidence intervals for the parameter of interest θ similar to parametric likelihood ratio confidence intervals. Consider the situation where the parameter of interest θ can be defined as a unique solution of the estimating equation $E[g(y,\theta)] = 0$. For example, $g(y,\theta) = y - \theta$ and $g(y,\theta) = I(y \leq t) - \theta$ yield $\theta = E(y)$, the population mean, and $\theta = E[I(y \leq t)] = F(t)$, the population distribution function, respectively.

25.10.1 Simple Random Sampling

Let a sample of size n is selected from a finite population of size N by simple random sampling with or without replacement. For SRSWOR, we assume that the sampling fraction is negligible so that we can write the likelihood function as $\prod_{i=1}^{n} p_i$. A profile likelihood function is then defined as

$$R(\theta) = \text{Max}\left\{\prod_{i=1}^{n}(np_i)\,\middle|\,\sum_{i=1}^{n} p_i g(y_i,\theta) = 0, p_i > 0, \sum_{i=1}^{n} p_i = 1\right\} \tag{25.10.1}$$

Owen (1988) proved that under mild moment conditions, for the case $\theta = E(y)$ or $F(t)$ that

$$r(\theta) = -2\log\{R(\theta)\} \tag{25.10.2}$$

converges to χ_1^2 (chi-square distribution with one degree of freedom) as $n \to \infty$. Hence $(1-\alpha)$-level confidence interval for θ is obtained as $\left\{\theta \middle| r(\theta) \le \chi_1^2(\alpha)\right\}$ where $\chi_1^2(\alpha)$ is the upper α-quantile of χ_1^2. It should be noted that unlike confidence interval based on normal approximation, empirical likelihood intervals do not require estimation of standard error of estimators and yield more balanced tail error rates.

25.10.2 Complex Sampling Designs

Let a sample of size n be selected from the population using a sampling design p and $\pi_i(>0)$ be the inclusion probability of the ith unit. In this case the pseudo−empirical log likelihood is given by

$$l_{ns}(\mathbf{p}) = \sum_{i\in s} d_i \log\ p_i \ \text{ with } \ d_i = 1/\pi_i \tag{25.10.3}$$

Maximizing $l_{ns}(\mathbf{p})$ with respect to $\mathbf{p}$ subject to (i) $p_i > 0$ and (ii) $\sum_{i\in s} p_i = 1$ yield

$$\widehat{p}_i = \widetilde{d}_i(s) = d_i \Big/ \sum_{i\in s} d_i$$

Similarly, let $\widehat{p}_i(\theta)$ be the value of p_i that maximizes the log likelihood $l_{ns}(\mathbf{p})$ for a fixed θ subject to

$$\text{(i)}\ \ p_i > 0, \text{(ii)}\ \ \sum_{i\in s} p_i = 1 \ \text{ and } \ \text{(iii)}\ \ \sum_{i\in s} p_i g(y_i) = \theta$$

where $g(y_i)$ is a function of y_i.

The pseudo−empirical log likelihood function is given by

$$r_{ns}(\theta) = -2[l_{ns}\{\widehat{\mathbf{p}}(\theta)\} - l_{ns}(\widehat{\mathbf{p}})] \tag{25.10.4}$$

where $\widehat{\mathbf{p}}(\theta)$ and $\widehat{\mathbf{p}}$ are the vector $\mathbf{p}$ with $p_i = \widehat{p}_i(\theta)$ and $p_i = \widehat{p}_i$, respectively.

Consider a simple situation where $g(y_i) = y_i$ and $\theta = \overline{Y}$. The design effect of estimating $\overline{Y}$ using the Hájek (1964) estimator $\widehat{\overline{Y}}_H = \sum_{i\in s} d_i y_i \Big/ \sum_{i\in s} d_i$ is given by

$$deff_H = V_p\left(\widehat{\overline{Y}}_H\right) \Big/ \left(S_y^2 \big/ n\right) \tag{25.10.5}$$

where $V_p\left(\widehat{\overline{Y}}_H\right)$ is the design variance of $\widehat{\overline{Y}}_H$ with respect to the sampling design p, S_y^2 is the population variance, and S_y^2/n is the variance of $\widehat{\overline{Y}}_H$ under SRSWOR sampling design (ignoring the sampling fraction).

Wu and Rao (2006) derived the following theorem (given without derivation) relating to the asymptotic distribution of $r_{ns}(\theta)$.

Theorem 25.10.1

Under the regularity conditions stated below, the adjusted pseudo–empirical log likelihood ratio statistic

$$r_{ns}^{[H]}(\theta) = \{r_{ns}(\theta)\}/deff_H$$

is asymptotically distributed as χ_1^2 when $\theta = \overline{Y}$.

The regularity conditions comprise the following:

C1: The sampling design p and the study variable y satisfy $\max_{i\in s}|y_i| = o_p(n^{1/2})$, where the stochastic order $o_p(\cdot)$ is with respect to the sampling design p.

C2: The sampling design p satisfies $\dfrac{1}{N}\sum_{i\in s} d_i - 1 = O_p\left(n^{-1/2}\right)$.

C3: The Horvitz–Thompson estimator $\widehat{\overline{Y}}_{HT} = \dfrac{1}{N}\sum_{i\in s} d_i y_i$ of the population mean $\overline{Y}$ is asymptotically normally distributed.

Using Theorem 25.10.1, we can set a $(1-\alpha)$ level confidence interval for $\theta = \overline{Y}$ as

$$\left\{\overline{Y}\middle| r_{ns}^{[H]}\left(\overline{Y}\right) \le \chi_1^2(\alpha)\right\}.$$

Now let us suppose that the we have a vector of auxiliary variable $\mathbf{x}$ with known population mean $\overline{\mathbf{X}}$. In this case the pseudoempirical log likelihood function for $\overline{Y}$ given in Eq. (25.10.3) should be obtained incorporating the additional constraint $\sum_{i\in s} p_i\mathbf{x}_i = \overline{\mathbf{X}}$ in finding both $\widehat{p}_i$ and $\widehat{p}_i(\theta)$. In this case the pseudo–empirical log likelihood ratio function is adjusted by the design effect associated with the GREG estimator

$$\widehat{\overline{Y}}_G = \widehat{\overline{Y}}_H - \mathbf{B}'\left(\widehat{\overline{\mathbf{X}}}_H - \overline{\mathbf{X}}\right)$$

where $\widehat{\overline{\mathbf{X}}}_H = \sum_{i\in s}\tilde{d}_i(s)\mathbf{x}_i$, $\widehat{\overline{\mathbf{Y}}}_H = \sum_{i\in s}\tilde{d}_i(s)y_i$ and $\mathbf{B} = \dfrac{\sum_{i\in U}(\mathbf{x}_i - \overline{\mathbf{x}})\left(y_i - \overline{\mathbf{Y}}\right)}{\sum_{i\in U}(\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x}_i - \overline{\mathbf{x}})'}$.

The design effect associated with $\widehat{\overline{Y}}_G$ is given by

$$deff_G = V_p\left(\widehat{\overline{Y}}_G\right)\Big/\left(S_r^2/n\right) \tag{25.10.7}$$

where $V_p\left(\widehat{\overline{Y}}_G\right) = V_p\left(\sum_{i\in s} \widetilde{d}_i(s)r_i\right)$, $r_i = y_i - \overline{Y} - \mathbf{B}'\mathbf{u}_i$, $\mathbf{u}_i = \mathbf{x}_i - \overline{\mathbf{X}}$, S_r^2/n is the variance of $\widehat{\overline{Y}}_G$ under SRSWOR ignoring sampling fraction n/N, and $S_r^2 = \sum_{i=1}^{N} r_i^2/(N-1)$.

Wu and Rao (2006) derived the asymptotic distribution of distribution of the adjusted pseudoempirical log likelihood ratio statistic $r_{ns}^{(G)}(\theta) = \{\widetilde{r}_{ns}(\theta)\}/deff_G$ under the following additional constraints.

C4: $\max_{i\in s}||\mathbf{x}_i|| = o_p(n^{1/2})$

C5: $\widehat{\overline{\mathbf{X}}}_{HT} = \dfrac{1}{N}\sum_{i\in s} d_i\mathbf{x}_i$ is asymptotically normally distributed.

The result has been given in the following theorem without derivation.

Theorem 25.10.2

Let $\widetilde{\mathbf{p}}$ maximizer of $l_{ns}(\widetilde{\mathbf{p}})$ subject to (i) $p_i > 0$, (ii) $\sum_{i\in s} p_i = 1$, and (iii) $\sum_{i\in s} p_i\mathbf{x}_i = \overline{\mathbf{X}}$; and let $\widetilde{\mathbf{p}}(\theta)$ be obtained by maximizing $l_{ns}(\widetilde{\mathbf{p}})$ under the constraints (i) $p_i > 0$, (ii) $\sum_{i\in s} p_i = 1$, (iii) $\sum_{i\in s} p_i\mathbf{x}_i = \overline{\mathbf{X}}$, and (iv) $\sum_{i\in s} p_i y_i = \theta$ for a fixed θ. Then under the regularity conditions stated C1 to C5, the adjusted pseudoempirical log likelihood ratio statistic

$$r_{ns}^{(G)}(\theta) = \{\widetilde{r}_{ns}(\theta)\}/deff_G$$

is asymptotically distributed as χ_1^2 when $\theta = \overline{Y}$ where
$\widetilde{r}_{ns}(\theta) = -2\ [l_{ns}\{\widetilde{\mathbf{p}}(\theta)\} - l_{ns}(\widetilde{\mathbf{p}})]$.

In practice the design effects $deff_H$ and $deff_G$ are unknown because of involvement of unknown parameters. The common practice is to replace the unknown parameters by their consistent estimators. Wu and Rao (2006) reported that asymptotic distributions of $r_{ns}^{(H)}(\theta)$ and $r_{ns}^{(G)}(\theta)$ remain unchanged if designs effects $deff_H$ and $deff_G$ are replaced by their consistent estimators. The estimated design effects are

$$\widehat{d}eff_H = \widehat{V}_p\left(\widehat{\overline{Y}}_H\right)\Big/\left(\widehat{S}_y^2\big/n\right) \text{ and } \widehat{d}eff_G = \widehat{V}_p\left(\widehat{\overline{Y}}_G\right)\Big/\left(\widehat{S}_y^2\big/n\right)$$

where $\widehat{V}_p\left(\widehat{\overline{Y}}_H\right) = \dfrac{1}{\widehat{N}^2}\dfrac{1}{2}\sum_{i\neq}\sum_{j\in s}\dfrac{\pi_i\pi_j - \pi_{ij}}{\pi_{ij}}\left(\dfrac{e_i}{\pi_i} - \dfrac{e_j}{\pi_j}\right)^2$ with $\widehat{N} = \sum_{i\in s} d_i$

and $\quad e_i = y_i - \widehat{\overline{Y}}_H; \quad \widehat{S}_y^2 = \dfrac{1}{N(N-1)}\dfrac{1}{2}\sum_{i\neq}\sum_{j\in s}\dfrac{(y_i - y_j)^2}{\pi_{ij}};$

$$\widehat{V}_p\left(\widehat{\overline{Y}}_G\right) = \frac{1}{\widehat{N}^2}\frac{1}{2}\sum_{i\neq}\sum_{j\in s}\frac{\pi_i\pi_j - \pi_{ij}}{\pi_{ij}}\left(\frac{\hat{r}_i}{\pi_i} - \frac{\hat{r}_j}{\pi_j}\right)^2$$

with $\widehat{r}_i = y_i - \widehat{\overline{Y}}_H - \widehat{\mathbf{B}}'\left(\mathbf{x}_i - \overline{\mathbf{X}}\right)$, $\widehat{B} = \dfrac{\sum_{i\in s}\left(\mathbf{x}_i - \hat{\overline{\mathbf{X}}}_h\right)y_i/\pi_i}{\sum_{i\in s}\left(\mathbf{x}_i - \hat{\overline{\mathbf{X}}}_h\right)\left(\mathbf{x}_i - \hat{\overline{\mathbf{X}}}_h\right)'\Big/\pi_i}$,

and $\widehat{S}_r^2 = \dfrac{1}{N(N-1)}\dfrac{1}{2}\sum_{i\neq}\sum_{j\in s}\dfrac{(r_i - r_j)^2}{\pi_{ij}}$.

Using Theorems 25.10.1 and 25.10.2 we can set a $(1-\alpha)$ level PEL confidence interval for $\theta = \overline{Y}$ as $\left\{\overline{Y}\middle|\widehat{r}_{ns}^{(H)}\left(\overline{Y}\right) \leq \chi_1^2(\alpha)\right\}$ for the case of no auxiliary information is available. In case the vector of the population mean $\overline{\mathbf{X}}$ is available, the confidence interval of $\theta = \overline{Y}$ is obtained as $\left\{\overline{Y}\middle|r_{ns}^{(\hat{G})}\left(\overline{Y}\right) \leq \chi_1^2(\alpha)\right\}$

where $\widehat{r}_{ns}^{(H)}\left(\overline{Y}\right) = \{r_{ns}(\theta)\}\Big/\widehat{deff}_H$ and $r_{ns}^{(\hat{G})}\left(\overline{Y}\right) = \left\{\widetilde{r}_{ns}\left(\overline{Y}\right)\right\}\Big/\widehat{deff}_G$.

25.10.3 Stratified Sampling

Consider the stratified sampling described in Section 25.7, where independent samples s_h's of sizes n_h's are selected from the hth stratum with inclusion probability $\pi_i^h = 1/d_{hi}$ (>0) for the ith unit of the hth stratum and $\sum_{h=1}^{H} n_h = n$ be the overall sample size. The PEL function under stratified sampling was given by Wu and Rao (2006) as

$$l_{st}(\widehat{\mathbf{p}}_1, \ldots, \widehat{\mathbf{p}}_H) = n\sum_{h=1}^{H} W_h \sum_{i\in s_h}\widetilde{d}_{hi}\log(p_{hi}) \tag{25.10.9}$$

where $\widetilde{d}_{hi} = d_{hi}\big/\sum_{i\in s_h} d_{hi}$.

Suppose that the mean vectors $\overline{\mathbf{X}}_h$' s of auxiliary variable of the strata are unknown but the overall population mean $\overline{\mathbf{X}} = \sum_{h=1}^{H} W_h\overline{\mathbf{X}}_h$ is known. In this case the pseudo—empirical log likelihood ratio statistic $r_{st}(\theta)$ is given by Rao and Wu (2009) as

$$r_{st}(\theta) = -2[l_{st}\{\widehat{\mathbf{p}}_1(\theta), \ldots, \widehat{\mathbf{p}}_H(\theta)\} - l_{st}\{\widehat{\mathbf{p}}_1, \ldots, \widehat{\mathbf{p}}_H\}] \tag{25.10.10}$$

where

(a) $\widehat{p}_{hi}$ maximize $l_{st}(\widehat{\mathbf{p}}_1, \ldots, \widehat{\mathbf{p}}_H) = n\sum_{h=1}^{H} W_h \sum_{i\in s_h}\widetilde{d}_{hi}\log(p_{hi})$ subject to the set of constraints

$$\text{(i)}\ \ p_{hi} > 0,\ \ \text{(ii)}\ \ \sum_{i\in s_h} p_{hi} = 1, h = 1, \ldots, H\ \ \text{and}\ \ \text{(iii)}\ \ \sum_{h=1}^{H} W_h \sum_{i\in s_h} p_{hi}\mathbf{x}_{hi} = \overline{\mathbf{X}} \tag{25.10.11}$$

and

(b) $\widehat{p}_{hi}(\theta)$ maximize $l_{st}(\widehat{\mathbf{p}}_1, \ldots, \widehat{\mathbf{p}}_H)$ subject to (25.10.11) with an additional constraint $\sum_{h=1}^{H} W_h \sum_{i\in s_h} p_{hi} y_{hi} = \theta$.

Theorem 25.10.3

Under the regularity condition C1 to C5 within each stratum h, the adjusted PEL ratio statistic

$$r_{st}^{[a]}(\theta) = \{r_{st}(\theta)\} / deff_{G(st)}$$

is asymptotically distributed as χ_1^2 when $\theta = \overline{Y}$. The design effect $deff_{G(st)}$ is given by Wu and Rao (2006).

The $(1-\alpha)$ level confidence interval for the population mean can be constructed from

$$\left\{\overline{Y} \middle| r_{st}^{[a]}\left(\overline{Y}\right) \le \chi_1^2(\alpha)\right\} \tag{25.10.12}$$

25.10.4 Confidence Interval for Distribution Function

Pseudo−empirical likely ratio confidence intervals of the distribution function $F(t)$ for a given t can be obtained from Theorems 25.10.1−25.10.3 by writing $I(y_i \le t)$ in place of y_i. The bench mark condition $\sum_{i\in s} p_i \mathbf{x}_i = \overline{\mathbf{X}}$ for making inference on $F(t)$ may not produce efficient result as the correlation between the indicator variable $I(y_i \le t)$ and $\mathbf{x}_i$ is seldom high. However, if the entire auxiliary variables of the population are known, different constraints may lead more efficient inference on $F(t)$ (Chen and Wu, 2002).

25.11 CONCLUDING REMARKS

Hartley and Rao (1968) introduced the concept of scale load approach, which produces useful results for simple random sampling in absence and presence of auxiliary variables. However, the generalization of scale load approach in unequal probability sampling without replacement does not provide any useful result. The modern concept of EL was introduced by Owen (1988). Chen and Quin (1993) provided the asymptotic variance of EL estimator of the finite population mean and a consistent estimator of the variance of the proposed estimator. Their method was limited to simple random sampling only. Chen and Sitter (1999) brought the concept of PEL approach, which can be used for any complex survey design. For simple random sampling, the PEL reduces to EL approach proposed by Chen and Quin (1993). For estimating the finite population mean or distribution function under complex survey design, the PEL approach yields asymptotically generalized regression estimator when the population means of the vector of auxiliary variables are known. Wu (2004) developed a PEL approach that combines information from two or more independent surveys from the same population with some common variable of interest. The method can be used to handle data from independent samples taken from two or more incomplete frames covering the entire population of interest and to produce efficient estimators. Sitter and Wu (2002) proposed the model-calibrated PEL method for estimating quadratic population parameters, which includes population variance and covariance. Chen and Quin (1993) used EL approach for nonparametric confidence intervals of the population mean and distribution function. Quin and Lawless (1994) used EL and estimating equations for constructing confidence intervals and hypothesis testing problems. Wu and Rao (2006) used PEL method for estimating confidence interval for finite population parameter with or without auxiliary information. They pointed out that PEL ratio confidence intervals are better than those based on normal approximation in terms of coverage probability and length of the confidence intervals. PEL confidence intervals are based on the asymptotic distribution of PEL ratio statistic with the adjustment factor related to design effect. Estimation of design effect involves variance estimation, which may not be easy for complex surveys. Wu and Rao (2009) introduced bootstrap procedures for PEL ratio confidence intervals, which do not require variance estimation and at the same time provide superior intervals in respect of coverage probability for moderate or small sample sizes. Wu (2005) provided computer software for computation of confidence intervals. Wang and Rao (2002) and Quin et al. (2006) used empirical and PEL methods for kernel regression imputation for nonresponse. Rao (2010) introduced Baysian PEL intervals for complex survey designs. Some good reviews have been given by Rao (2006) and Rao and Wu (2009).

25.12 EXERCISES

25.12.1 Let $y_1,\ldots, y_i,\ldots, y_n$ be independent random variables with common CDF F_0 and S_n be their ECDF. Let F be any CDF, then $L(F) < L(S_n)$ for $F \neq S_n$ (Owen, 2001).

25.12.2 Let $y_1,\ldots, y_i,\ldots, y_n$ be a random sample from a discrete distribution with $P(y = y_i) = p_i$, $p_i > 0, \sum_{i=1}^{N} p_i = 1$. Find the MEL estimator for $\mu_k = E(y^k)$.

25.12.3 A sample of n units using PPSWR method with normed size measure q_i attached to the ith unit. Show that the MPEL estimator for the population mean $\overline{Y}_N$ without using any auxiliary information is $\widehat{\overline{Y}}_N = \left(\sum_{i=1}^{N} y_i/p_i\right) \Big/ \left(\sum_{i=1}^{N} 1/p_i\right)$ (Chen and Sitter, 1999).

25.12.4 Suppose a sample of n is selected from a population by Rao–Hartley–Cochran sampling scheme using p_i as normed size measure for the ith unit. Find the MPEL estimator for the population mean when no auxiliary variable is available (Chen and Sitter, 1999).

25.12.5 Let a sample of n is selected by Poisson sampling scheme with inclusion probability π_i for the ith unit.

(i) Show that the MPEL estimator for the population mean is $\left(\sum_{j=1}^{n} y_i/\pi_i\right) \Big/ \left(\sum_{j=1}^{n} 1/\pi_i\right)$ when no auxiliary information is available.

(ii) If the auxiliary information is used as $\left(\sum_{j=1}^{n} p_i x_i\right) = \mu_x$,

show that p_i is obtainable as a solution of the equations

$$p_i = \frac{1}{\lambda_1\pi_i + \lambda_2(x_i - \mu_x)}, \sum_{i=1}^{n} \frac{1}{\lambda_1\pi_i + \lambda_2(x_i - \mu_x)} = 1$$

$$\text{and} \quad \sum_{i=1}^{n} \frac{x_i - \mu_x}{\lambda_1\pi_i + \lambda_2(x_i - \mu_x)} = 0 \quad \text{(Kim, 2009).}$$