EMPIRICAL LIKELIHOOD INFERENCE UNDER STRATIFIED RANDOM SAMPLING USING AUXILIARY INFORMATION

C.X.Bob Zhong and J.N.K. Rao, Carleton University
Dept. of Math. & Stats. Carleton University, Ottawa, ON. K1S 5B6

Key words: Distribution function; Jackknife; Likelihood ratio; Estimating equation; Confidence interval.

1 Introduction

Suppose we are required to make some inference about population parameters. In many practical problems, we have some information about the population. For example, some population values of the auxiliary variable x are known, such as the population mean or median. Another example is that we may know that the population variance is a function of the population mean. By using this kind of knowledge, we would like to provide improved inferences on the population parameters such as the population mean and the population distribution function. Empirical likelihood methods, recently introduced by Owen (1988, 90, 91) in the context of iid random variables, provide a systematic nonparametric approach to utilizing auxiliary information in making inference on the parameters of interest. Hartley and Rao (1968) gave the original idea of empirical likelihood in sample survey context, using their "scale-based" approach. Chen and Qin (1993) extended their results to cover distribution function and median. Qin and Lawless (1994, 95) used empirical likelihood and estimating equations in the iid case to deal with interval estimation or hypothesis testing. They obtained an empirical likelihood ratio test statistic (ELRS) and its asymptotic distribution under null hypothesis. The aim of this article is to study the case of stratified random sampling.

In section 2, we give the maximum empirical likelihood estimator (MELE) for the parameter of interest, where we know the mean of $\bar{X}$ of the auxiliary information x and samples are taken without replacement in each stratum and independently across strata. Some large sample properties of MELE are also discussed. The empirical likelihood ratio estimator (ELRE) is discussed. The results show that MELE is equal to ELRE. Empirical likelihood ratio tests (ELRT) are also discussed. In section 3, we use empirical likelihood and estimating equations to deal with interval estimation and hypothesis testing.

2 Empirical Likelihood Inference Using Stratified Random Sampling without Replacement

2.1 Introduction

Suppose that a target population is divided into H strata with known weight W_h for all stratum h, $\sum_h W_h = 1$. In stratum h there are N_h units with values $z_{hi}(i = 1,\cdots,N_h; h = 1,\cdots,H)$, where z_{hi} is a d-dimension variable, $z_{hi} = (x_{hi}^\tau, y_{hi}^\tau)^\tau$, x_{hi} and y_{hi} are $d-p$ and p dimensions respectively and τ denotes the transposition. Denote the h^{th} stratum population mean, median and distribution function by $(\bar{X}_h^\tau, \bar{Y}_h^\tau)^\tau, (m_{x_h}^\tau, m_{y_h}^\tau)^\tau$ and $F_{hN_h}(s) = N_h^{-1}\{\delta_{(z_{h1})} + \cdots + \delta_{(z_{hN_h})}\}$ respectively, where $\delta_{z_{hi}}$ is the point measure at z_{hi}. Also, let $(\bar{X}^\tau, \bar{Y}^\tau)^\tau, (m_X^\tau, m_Y^\tau)$ and $F_N(s) = \sum_{h=1}^H W_h F_{hN_h}(s)$ be the mean, median and distribution function of the target population respectively, where $N = \sum_{h=1}^H N_h$. Obviously, $\bar{X} = \sum_{h=1}^H W_h \bar{X}_h$, $\bar{Y} = \sum_{h=1}^H W_h \bar{Y}_h$. We want to make inference about target population parameters such as $\bar{Y}$ or m_Y.

2.2 MELE with $\bar{X}$ known

Suppose that $z_{h1},\cdots,z_{hn_h}$ is a simple random sample without replacement from stratum h with distribution function F_{hN_h} for all h, and that the samples are selected independently across the strata. The empirical likelihood for the above sampling scheme can be approximated by

$$L = \Pi_{h=1}^{H}\Pi_{i=1}^{n_h} p_{hi} \tag{2.1}$$

if $n_h << N_h$ and large N_h, where $p_{hi} = Pr(z_h = z_{hi})$ and $z_h = (x_h^\tau, y_h^\tau)^\tau$ has the distribution function F_{hN_h}.

We consider the case of known vector of population means, $\bar{X}$, of the variable $x = (x_1, \cdots, x_{d-p})^\tau$:

$$\bar{X} = \sum_{h=1}^{H} W_h E(x_h) = \sum_{h=1}^{H} W_h N_h^{-1} \sum_{i=1}^{N_h} x_{hi}. \tag{2.2}$$

Clearly, the maximum likelihood estimator should be sought among distribution functions satisfying (2.2) . Using the same argument as in Owen (1990), we need consider only estimators of F_{hN_h} whose support is contained in the set of observations. The problem therefore reduces to maximizing

$$l = \sum_h \sum_i \log p_{hi} \tag{2.3}$$

subject to

$$\sum_i p_{hi} = 1 \ (h = 1, \cdots, H; p_{hi} \geq 0) \tag{2.4}$$

and

$$\sum_h W_h \sum_i p_{hi} x_{hi} = \bar{X}, \tag{2.5}$$

where $\sum_h = \sum_{h=1}^{H}$ and $\sum_i = \sum_{i=1}^{n_h}$. A unique solution for the above problem exists, provided that $\bar{X}$ is within the convex hull of the points $x_{11}, \cdots, x_{1n1}; \cdots; x_{H1}, \cdots, x_{Hn_H}$.

The estimators of p_{hi} are the solutions of the following system of equations

$$p_{hi} = 1/n_h[1 + m_h \psi^\tau (x_{hi} - \hat{\bar{X}}_h)], \tag{2.6}$$

$$\hat{\bar{X}}_h = \sum_i p_{hi} x_{hi}, \tag{2.7}$$

$$\sum_h W_h \sum_i p_{hi} x_{hi} = \bar{X}, \tag{2.8}$$

where $\hat{\bar{X}}_h = \sum_i p_{hi} x_{hi}$, $m_h = nW_h n_h^{-1}$ and p_{hi} subject to $0 \leq p_{hi} \leq 1$. We denote the solutions of (2.6), (2.7) and (2.8) as $\hat{p}_{hi}$. The estimator we get by using this method is called the Maximum Empirical Likelihood Estimator (MELE). An efficient method for getting $\hat{p}_{hi}$ is given in an unpublished report.

We will discuss the properties of $\hat{p}_{hi}$ and other related topics in the following sections. Proofs are omitted.

2.3 Asymptotic Results

We assume that both the sample size n_h and the stratum size N_h go to infinity as a certain index ν attached to n_ν and N_ν goes to infinity for all h, i.e., $n_{h\nu}$ and $N_{h\nu}$ go to infinity as $\nu \to \infty$. However, for convenience, we will suppress the index ν in the following whenever possible. And we will denote the solution $\hat{\psi}$ by $\hat{\psi}_n$ since we are going to deal with large sample problems.

Theorem 2.1. *Suppose that as $\nu \to \infty$, N_h, n_h, $N_h - n_h$ go to infinity, $n/n_h \to k_h$, $k_h > 0$, $n/N_h \to 0$, and both $\sum_h W_h^2 \sum_i \frac{N_{hi}}{N_h} \parallel x_{hi} \parallel^3$ and $\sum_h W_h^2 \sum_i \frac{N_{hi}}{N_h} \parallel y_{hi} \parallel^3$ have an upper bound independent of ν, and $\sigma_{h,xx} = Cov(x_h, x_h) > \sigma_1 > 0$ for all h and ν, where σ_1 is a fixed positive definite matrix. Then*

$$\sigma_\nu^{-\frac{1}{2}}(\hat{\bar{Y}} - \bar{Y}) \xrightarrow{L} N_p(0, I_p),$$

$$A^{-\frac{1}{2}} U \hat{\psi}_n \xrightarrow{L} N_{d-p}(0, I_{d-p}),$$

$$A^{-\frac{1}{2}}(\hat{F}_n(s) - F_N(s)) \xrightarrow{L} N_d(0, V),$$

as $\nu \to \infty$, where $F_n(s) = \sum_h \sum_i \hat{p}_{hi} 1(z_{hi} < s)$, $\sigma_\nu = \sum_h W_h^2(n_h^{-1} - N_h^{-1})[Var(y_h) - 2BCov(y_h, x_h) + BVar(x_h)B^\tau]$, $B = \sum_h k_h W_h^2 Cov(y_h, x_h) \cdot [\sum_h k_h W_h^2 Cov(x_h, x_h)]^{-1}$ and $U^2 = C\sigma_{xx}$, $\sigma_{xx} = \sum_h k_h W_h^2 Cov(x_h, x_h)$, $C = \sum_h k_h W_h^2$, $V = C^{-1}\sum_h W_h^2 k_h F_h(s)(1 - F_h(s)) - C^{-1} G \sigma_{xx}^{-1} G^\tau$, $G = \sum_h W_h^2 k_h Cov(1(z_h < s), x_h^\tau)$.

From the results of Theorem 2.1, we know $\hat{F}_n(s)$ is asymptotically more efficient than the empirical c.d.f. $F_n(s)$. Also, $\hat{\bar{Y}}$ has the same asymptotically variance as the optional regression estimator $\bar{y}_{st} + B(\bar{X} - \bar{x}_{st})$.

In above discussion we consider auxiliary information of the form of (2.2), but we can easily generalize it to other form of auxiliary information such as

$$\sum_h W_h N_h^{-1} \sum_{i=1}^{N_h} w(x_{hi}) = 0.$$

The choice $w(x_{hi}) = x_{hi} - \bar{X}$ gives (2.2). When the population median m_x is known and x is a scalar, we let $w(x_{hi}) = I_{[x_{hi} \leq m_X]} - 0.5$.

2.4 Variance Estimation

In this section, we will consider how to estimate σ, the variance of $\hat{\bar{Y}}$ and $A^{\frac{1}{2}}V$, the variance of $\hat{F}_n(s)$.

Under the conditions of Theorem 2.1, we can see

$$\begin{aligned}\hat{\sigma} &= \sum_h W_h^2 \frac{n_h^{-1} - N_h^{-1}}{n_h - 1} \\ & \sum_{i=1}^{n_h} [y_{hi} - \hat{\bar{Y}}_h - \hat{B}(x_{hi} - \hat{\bar{X}}_h)] \\ & \cdot [y_{hi} - \hat{\bar{Y}}_h - \hat{B}(x_{hi} - \hat{\bar{X}}_h)]^\tau\end{aligned}$$

and

$$\begin{aligned}A^{\frac{1}{2}}\hat{V} &= A^{-\frac{1}{2}}[\sum_h W_h^2 \frac{1}{n_h} F_{h,n_h}(s)[1 - F_{h,n_h}(s)] \\ & -\frac{1}{n} G_n S_{xx}^{-1} G_n^\tau]\end{aligned}$$

are consistent estimators of σ and $A^{\frac{1}{2}}V$ respectively, where

$$\hat{B} = [\sum_h n W_h^2 n_h^{-2} \sum_i (y_{hi} - \hat{\bar{Y}}_h)(x_{hi} - \hat{\bar{X}}_h)^\tau] S_{xx}^{-1}.$$

We can use this result to get confidence intervals for $\bar{Y}$ and $F(s)$, which are asymptotically correct.

Another consistent estimator of σ is the jackknife variance estimator. We omit the theorem here.

2.5 ELRE estimator

Now we turn to the empirical likelihood ratio estimator (ELRE) of $\bar{Y}$. Here we denote $E(z) = \sum_h W_h E(z_h) = \theta$, where $\theta = (\theta_k^\tau, \theta_u^\tau)^\tau$, $\theta_k = \bar{X}, \theta_u = \bar{Y}$, and k, u mean "known", "unknown" respectively. By (2.1) we know the empirical likelihood function is

$$L(F) = \Pi_{h=1}^H \Pi_{i=1}^{n_h} F_{h,N_h}(z_{hi}) = \Pi_{h=1}^H \Pi_{i=1}^{n_h} p_{hi}, \tag{2.9}$$

where $p_{hi} = dF_{h,N_h}(z_{hi}) = Pr(z_h = z_{hi})$ and F denote the population distribution. Noting that $p_{hi} \geq 0, \sum_i p_{hi} = 1$ and $F(s) = \sum_h W_h F_{h,N_h}(s)$, we know (2.9) is maximized by $F_n(s) = \sum_h W_h F_{h,n_h}(s)$, where $F_{h,n_h} = \frac{1}{n_h}\sum_i I(z_{hi} < s)$. The empirical likelihood ratio is then defined as $R(F) = L(F)/L(F_n)$ which reduces to

$$R(F) = \Pi_{h=1}^H \Pi_{i=1}^{n_h} n_h p_{hi}. \tag{2.10}$$

Since we are interested in the parameter $\theta_u = \bar{Y}$, and we have auxiliary information (2.2), we define the profile empirical likelihood ratio function

$$\begin{aligned}R_E(\theta) &= sup\{R(F) \mid p_{hi} \geq 0, \sum_i p_{hi} = 1, \\ & \sum_h W_h \sum_i p_{hi} z_{hi} = \theta\}. \end{aligned} \tag{2.11}$$

As discussed in section 2.2, for a given θ_u, a unique value for the right-hand side of (2.11) exists, provided that θ is inside the convex hull of points $z_{11}, \cdots, z_{1n_1}; \cdots; z_{H1}, \cdots, z_{Hn_H}$. Using the Lagrange multiplier method, let

$$\begin{aligned}l &= \sum_h \sum_i \log p_{hi} - \sum_h \lambda_h (\sum_i p_{hi} - 1) \\ & -nt^\tau (\sum_h W_h \sum_i p_{hi} z_{hi} - \theta)\end{aligned}$$

where $t = (t_1, t_2, \cdots, t_d)^\tau$ are Lagrange multipliers. Taking derivatives with respect to p_{hi}, we have

$$\begin{aligned}\frac{\partial l}{\partial p_{hi}} &= p_{hi}{}^{-1} - \lambda_h - n w_h t^\tau z_{hi} = 0, \ i = 1, \cdots, n_h; \\ & h = 1, \cdots, H,\end{aligned}$$

$$\sum_i p_{hi} \frac{\partial l}{\partial p_{hi}} = n_h - \lambda_h - n W_h t^\tau \sum_i p_{hi} z_{hi} = 0$$

$$\Rightarrow \lambda_h = n_h - n W_h t^\tau \sum_i p_{hi} z_{hi}.$$

Hence,

$$p_{hi} = 1/n_h[1 + m_h t^\tau (z_{hi} - \hat{\bar{Z}}_h)], \tag{2.12}$$

where $\hat{\bar{Z}}_h = \sum_i p_{hi} z_{hi}$, $m_h = n W_h n_h^{-1}$ and p_{hi} with the restriction

$$\begin{aligned}\theta &= \sum_h W_h \sum_i p_{hi} z_{hi} \\ &= \sum_h \frac{W_h}{n_h} \sum_i \frac{1}{1 + m_h t^\tau (z_{hi} - \hat{\bar{Z}}_h)} z_{hi}.\end{aligned}$$

We know that t can be determined in terms of θ, and $t = t(\theta)$ is actually a continuous differentiable function of θ.

The empirical likelihood function for θ is now defined as

$$L_E(\theta) = \Pi_{h=1}^H \Pi_{i=1}^{n_h} (\frac{1}{n_h}) \frac{1}{1 + m_h t^\tau (z_{hi} - \hat{\bar{Z}}_h)},$$

which leads to the empirical log-likelihood ratio statistic:

$$l_E(\theta) = \sum_h \sum_i \log[1 + m_h t^\tau (z_{hi} - \hat{\bar{Z}}_h)]. \tag{2.13}$$

We minimize $l_E(\theta)$ to obtain an estimator $\tilde{\theta}_u$ of the parameter θ_u, $\tilde{\theta} = (\bar{X}^\tau, \tilde{\theta}_u^\tau)^\tau$, called empirical likelihood ratio estimator. In addition, this yields estimators $\tilde{p}_{hi}$, and an estimator for the distribution function F:

$$\tilde{F}_n(s) = \sum_i W_h \sum_i \tilde{p}_{hi} \, 1(z_{hi} < s). \tag{2.14}$$

Now we turn to the calculation of $\tilde{\theta}$, i.e., minimization of $l_E(\theta)$. From the context above we know that it is equivalent to maximizing

$$L = \Pi_h \Pi_i p_{hi} \quad (2.15)$$

subject to:

$$\sum_i p_{hi} = 1, p_{hi} \geq 0, \quad (2.16)$$

$$\sum_h W_h \sum_i p_{hi} x_{hi} = \theta_k, \quad (2.17)$$

$$\sum_h W_h \sum_i p_{hi} y_{hi} = \theta_u \quad (2.18)$$

where θ_u is also unknown. Since θ_u is unknown, (2.18) does not add any information into the estimation problem. Hence, L will be maximized dropping (2.18). We can show this as follows. Let

$$\begin{aligned} l_1 &= \sum_h \sum_i \log p_{hi} - \sum_h \lambda_h (\sum_i p_{hi} - 1) \\ & -nt_k^\tau (\sum_h w_h \sum_i p_{hi} x_{hi} - \theta_k) \\ & -nt_u^\tau (\sum_h W_h \sum_i p_{hi} y_{hi} - \theta_u), \end{aligned}$$

where t_k,t_u are multipliers with the same dimension as θ_k, θ_u respectively. Then:

$$\frac{\partial l_1}{\partial \theta_u} = nt_u = 0 \Rightarrow t_u = 0.$$

This shows that we can drop (2.18). Therefore, $\tilde{p}_{hi}$'s are the same as $\hat{p}_{hi}$'s in section 2.2, and $\tilde{\theta}_u = \hat{\hat{Y}} = \sum_h W_h \sum_i \tilde{p}_{hi}\, y_{hi}$.

We note here that under the hypothesis $H_0 : \theta_u = \theta_{u_0}$, the estimator for p_{hi} will be changed to

$$\hat{p}_{hi} = \frac{1}{n_h[1 + m_h \hat{t}^\tau(\theta_0)(z_{hi} - \hat{\hat{Z}}_h)]},$$

where $\hat{\hat{Z}}_h = \sum_i \hat{p}_{hi} z_{hi}$, $\theta_0 = (\bar{X}^\tau, \theta_{u_0}^\tau)^\tau$.

The empirical likelihood statistic for testing H_0 is given by

$$\begin{aligned} W_E(\theta_0) &= 2l_E(\theta_0) - 2l_E(\tilde{\theta}) \\ &= 2\sum_h \sum_i \log[1 + m_h \hat{t}^\tau(\theta_0)(z_{hi} - \hat{\hat{Z}}_h)] - \\ & -2\sum_h \sum_i \log[1 + m_h\, \tilde{t}^\tau\, (\tilde{\theta})(z_{hi} - \tilde{\tilde{Z}}_h)], \end{aligned}$$

where $\tilde{\tilde{Z}}_h = \sum_i \tilde{p}_{hi}\, z_{hi}$.

Theorem 2.3. *Under the hypothesis H_0 and the conditions of Theorem 2.1 and $Var(z_h) \geq \sigma_1 > 0$ for any h and ν.*

$$W_E(\theta_0) \xrightarrow{L} \chi^2(p) \quad as \quad \nu \to \infty,$$

where $\theta_0 = (\theta_k^\tau, \theta_{u_0}^\tau)^\tau$.

3 Empirical Likelihood and General Estimating Equations

3.1 Introduction

Likelihood and estimating equations provide the most common approaches to parametric inference. Our purpose is to combine empirical likelihood, estimating equations and stratified sampling technology together. In order to simplify our discussion, we assume i.i.d. sampling in each stratum, but the discussion can be carried over to simple random sampling without replacement.

Suppose that a target population with d-dimensional characteristic x has the unknown distribution function F and a p-dimensional parameter θ associated with F. We are interested in making inference on θ. The sampling scheme is as follows. Suppose that the target population is divided into H strata with known weight W_h's for all strata. We get an i.i.d. sample $x_{h1}, \cdots, x_{hn_h}$ from x_h, the stratum h which has distribution F_h, where x_{hi}'s are d-dimensional and $h = 1, \cdots, H$, $i = 1, \cdots, n_h$ and F_h's are unknown, and the samples across strata are also independent. We have

$$F = \sum_h W_h F_h. \quad (3.1)$$

We also assume that information about θ and F is available in the form of $r \geq p$ functionally independent unbiased estimation functions, that is functions $g_j(x, \theta), j = 1, \cdots, r$ such that $E_F\{g_j(x,\theta)\} = 0$. In vector form,

$$g(x, \theta) = (g_1(x, \theta), \cdots, g_r(x, \theta))^\tau,$$

where $g(x, \theta)$ satisfies

$$E_F\{g(x,\theta)\} = \sum_h W_h E_{F_h}\{g(x_h, \theta)\} = 0. \quad (3.2)$$

In the following we will show how to use such information to estimate θ and F, in conjunction with empirical likelihood.

3.2 MELR estimators

We define empirical likelihood as

$$L(F) = \Pi_{h=1}^{H}\Pi_{i=1}^{n_h} p_{hi}, \quad (3.3)$$

where $p_{hi} = Pr(x_h = x_{hi})$. Only those F distributions with F_h's which have an atom of probability on each x_{hi} have nonzero likelihood. Noting (3.2), we know (3.3) is maximized by the empirical distribution function $F_n(s) = \sum_h F_{h,n_h}(s)$, where $n = \sum_h n_h, F_{h,n_h}(s) = \frac{1}{n_h}\sum_i 1(x_{hi} < s)$. The empirical likelihood ratio is then defined as $R(F) = L(F)/L(F_n)$, and after a little calculation we get

$$R(F) = \Pi_{h=1}^{H}\Pi_{i=1}^{n_h} n_h p_{hi}. \quad (3.4)$$

Since we are interested in estimating the parameter θ, and we know the estimating equation (3.2), we define the empirical likelihood ratio function

$$\begin{aligned} R_E(\theta) \quad = \quad & sup\{R(F) \mid p_{hi} \geq 0, \sum_i p_{hi} = 1, \\ & \sum_h W_h \sum_i p_{hi} g_{hi}(\theta) = 0\}, \end{aligned} \quad (3.5)$$

where $g_{hi}(\theta) = g(x_{hi}, \theta)$ for all h and i. For any given θ, a unique value for the right side of (3.5) exists, provided 0 is inside the convex hull of the points $g_{11}(\theta), \cdots, g_{1n_1}(\theta); \cdots; g_{H1}(\theta), \cdots, g_{Hn_H}(\theta)$. The maximum may be found via Lagrange multiplier method. The estimators for $\hat{p}_{hi}$ are solutions of following equations

$$p_{hi} = \frac{1}{n_h[1 + m_h t^\tau (g_{hi}(\theta) - \hat{g}_h(\theta))]}, \quad (3.6)$$

$$\hat{g}_h(\theta) = \sum_i \frac{g_{hi}(\theta)}{n_h[1 + m_h t^\tau (g_{hi}(\theta) - \hat{g}_h(\theta))]}, \quad (3.7)$$

$$\sum_h W_h \hat{g}_h(\theta) = 0, \quad (3.8)$$

from which t can be determined in terms of θ, and $t = t(\theta)$ is actually a continuous differentiable function of θ. Therefore, the empirical negative log-likelihood ratio statistic

$$l_E(\theta) = \sum_h \sum_i \log[1 + m_h t^\tau(\theta)(g_{hi}(\theta) - \hat{g}_h(\theta))]. \quad (3.9)$$

We may minimize $l_E(\theta)$ to obtain an estimate $\tilde{\theta}$ of the parameter θ (called MELR). In addition, this yields estimates $\tilde{p}_{hi}$'s from (3.6), and an estimate for distribution F as

$$\tilde{F}_n(s) = \sum_h W_h \sum_i \tilde{p}_{hi}\, 1(x_{hi} < s)) \quad (3.10)$$

Here are a few conditions for the following theorems to be true.

(1) As $n \to \infty$, $n/n_h \to k_h > 0$ for all h. And suppose that in a neighborhood of the true value θ_0, $E_h[(g(x_h, \theta_0) - E_h g(x_h, \theta_0))(g(x_h, \theta_0) - E_h g(x_h, \theta_0))^\tau] = \sigma_h(\theta_0) > \sigma_1 > 0$ for all h, where σ_1 is positive definite, $\| g(x, \theta) \|^3$ is bounded by some integrable function $G(x)$ in this neighbourhood, then for given θ.

(2) $\partial g(x, \theta)/\partial\theta$ is continuous in a neighborhood of the true value θ_0; that $\| \partial g(x, \theta)/\partial\theta \|$ is bounded by some integrable function $G(x)$ in this neighbourhood; and that the rank of $\sum_h W_h E_h[\partial g(x_h, \theta_0)/\partial\theta]$ is p.

(3) $\frac{\partial^2 g(x,\theta)}{\partial\theta\partial\theta^\tau}$ is continuous in θ in a neighbourhood of the true value θ_0, then if $\| \frac{\partial^2 g(x,\theta)}{\partial\theta\partial\theta^\tau} \|$ is bounded by some integrable function $G(x)$ in this neighbourhood.

Theorem 3.1. *Under conditions (1), (2) and (3) above, we have*

$$\sqrt{n}(\tilde{\theta} - \theta_0) \to N(0, V), \ \sqrt{n}(\tilde{t} - 0) \to N(0, U),$$

$$\sqrt{n}(\tilde{F}_n(s) - F(s)) \to N(0, W),$$

where $\tilde{F}_n(x) = \sum_h W_h \sum_i \tilde{p}_{hi}\, 1(x_{hi} < s)$, $\tilde{p}_{hi} = \frac{1}{n_h}[1 + m_h \tilde{t}^\tau (g_{hi}(\theta) - \hat{g}_h(\theta))]^{-1}$, *and* U, V, W *are defined below:* $U = M_{11}^{-1}(I + M_{12}M_{22.1}^{-1}M_{21}M_{11}^{-1})$, *where* $-\sum_h \frac{W_h}{n_h}\sum_i \frac{\partial g_{hi}(\theta_0)}{\partial\theta^\tau} \xrightarrow{p} M_{12}$, $M_{21} = M_{12}^\tau$, $-\sum_h \frac{nW_h^2}{n_h^2}\sum_i(g_{hi}(\theta_0) - \bar{g}_h(\theta_0))(g_{hi}(\theta_0) - \bar{g}_h(\theta_0))^\tau \xrightarrow{p} M_{11}$, $M_{22.1} = -M_{21}M_{11}^{-1}M_{12}$, $V = (M_{21}M_{11}^{-1}M_{12})^{-1}$, *and* $W = \sum_h W_h^2 k_h F_h(s)[1 - F_h(s)] - BUB^\tau$, *where* $B = \sum_h W_h^2 k_h E_h[(g(x_h, \theta_0) - E_h g(x_h, \theta_0))^\tau 1(x_h < s)]$.

Theorem 3.2 *In the semiparametric model (3.2), for testing* $H_0 : \theta = \theta_0$ *the empirical likelihood ratio test statistic*

$$\begin{aligned} R \quad = \quad & 2\{\sum_h \sum_i \log[1 + m_h t^\tau(\theta_0)(g_{hi}(\theta_0) - M_h(\theta_0)] \\ & - \sum_h \sum_i \log[1 + m_h \tilde{t}^\tau (\tilde{\theta})(g_{hi}(\theta)t - \hat{g}_h(\theta)t]\} \end{aligned}$$

is asymptotically χ_p^2 *under* H_0, *assuming (1), (2) and (3).*

3.3 MELR estimators and testing with constraints

In this section we extend the empirical likelihood methods to deal with the case in which there are constraints on parameters. Suppose there are q-dimensional constraints on θ:

$$r(\theta) = 0, \tag{3.11}$$

where $r(\theta)$ is a $q \times 1$ ($q \leq p$) and the $q \times p$ matrix $R(\theta) = \frac{\partial r}{\partial \theta^\tau}$ is of full rank q. To minimize $l_E(\theta)$ defined by (3.9) subject to $r(\theta) = 0$, we consider

$$G_2 = \frac{1}{n} l_E(\theta) + \nu^\tau r(\theta),$$

where ν is a $q \times 1$ vector of Lagrange multipliers. Differentiating G_2 with respect to θ and ν, we have

$$\frac{1}{n}\frac{\partial l_E(\theta)}{\partial \theta^\tau} + \nu^\tau R(\theta) = 0, \ r(\theta) = 0. \tag{3.12}$$

Consequently, to minimize $l_E(\theta)$ subject to $r(\theta) = 0$, we consider the solution of

$$Q_{1n}(\theta, t, \nu) = 0, \ Q_{2n}(\theta, t, \nu) = 0, \ Q_{3n}(\theta, t, \nu) = 0, \tag{3.13}$$

where

$$Q_{1n}(\theta, t, \nu) = \sum_h \frac{W_h}{n_h} \sum_i \frac{g_{hi}(\theta) - \hat{g}_h(\theta)}{1 + m_h t^\tau [g_{hi}(\theta) - \hat{g}_h(\theta)]},$$

$$Q_{2n}(\theta, t, \nu) = [\sum_h \frac{W_h}{n_h} \sum_i \frac{1}{1 + m_h t^\tau [g_{hi}(\theta) - \hat{g}_h(\theta)]} \frac{\partial g^\tau_{hi}(\theta)}{\partial \theta}] t + R^\tau(\theta)\nu, \tag{3.14}$$

$$Q_{3n}(\theta, t, \nu) = r(\theta),$$

where $\hat{g}_h(\theta)$ defined by (3.7). We denote the solution to (3.13) as $(\tilde{\theta}_r, \tilde{t}_r, \tilde{\nu}_r)$.

Theorem 3.3. *Under conditions (1) (2) (3) and*

(4) $g(x, \theta)$ is continuous and differentiable in the neighbourhood of θ_0 and $E \parallel g(x, \theta) \parallel^3 < \infty$. And $\sigma_h(\theta_0) > 0$ for all h and that $\sum_h E_h\{\frac{\partial g(x_h, \theta_0)}{\partial \theta}\}$ is full rank.

(5) in a neighbourhood of θ_0, $h(\theta)$ is a continuous differentiable function; that the $q \times p$ matrix $R(\theta)$ is of rank q; and that

$$\frac{\partial^2 h(\theta)}{\partial \theta \partial \theta^\tau}, \ \frac{\partial^2 g(x, \theta)}{\partial \theta \partial \theta^\tau}$$

exist and are bounded by some constant and some integrable function respectively, we have

$$\sqrt{n}\,\tilde{\xi}_r = \sqrt{n}\begin{pmatrix} \tilde{\theta}_r - \theta_0 \\ \tilde{\nu}_r \end{pmatrix} \xrightarrow{L} N\left(0, \begin{pmatrix} P & 0 \\ 0 & H \end{pmatrix}\right) \tag{3.15}$$

Now we turn to the problem of testing $H_0 : r(\theta) = 0$. There are three popular methods based on likelihood: likelihood-ratio test, Lagrange-multiplier test and Wald test. Of course here they should be based on empirical likelihood. These statistics are defined respectively as:

$$ELR = 2l_E(\tilde{\theta}_r), \ LA = n\, \tilde{\nu}_r^\tau\, H^{-1}_{\tilde{\theta}_r}\, \tilde{\nu}_r, \tag{3.16}$$

$$WA = nr^\tau(\hat{\theta}) H_{\hat{\theta}} r(\hat{\theta}), \tag{3.17}$$

where $\hat{\theta}$ is a solution of the estimating equation $\sum_h W_h n_h^{-1} \sum_i g_{hi}(\theta) = 0$, and H_θ defined in (3.15), is the asymptotic covariance matrix of $\sqrt{n}\ \tilde{\nu}_r$. The following theorem gives the asymptotic behavior and relationship between the test statistics defined in (3.16) and (3.17).

Theorem 3.4. *Under the assumptions of Theorem 3.3, the three test statistics in (3.16) and (3.17) are asymptotically equivalent, and each of them is asymptotically distributed as χ^2_q under H_0.*

References

Aitchison, J., & Silvey, S.D. (1958). Maximum-likelihood estimation of parameters subject to restraints. Ann. Math. Statist., 29, 813-828.

Bickel, P.J. & Freedman, D.A. (1984). Asymptotic normality and the bootstrap in stratified sampling. *Ann. Statist.* 12, 470-82.

Chen, J. & Qin, J. (1993). Empirical likelihood estimation for finite population and the effective usage of auxiliary information . *Biometrika* 80 107-116.

Cochran, W.G. (1977). *Sampling Techniques*, 3rd ed. New York: Wily.

Hartley, H.O. & Rao, J.N.K. (1968). A new estimation theory for sample surveys. *Biometrika* 55, 547-557.

Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* 75 237-249.

Owen, A. B. (1990). Empirical likelihood confidence regions. *Ann. Statist.* 18 90-120.

Owen, A. B. (1991). Empirical likelihood for linear models. *Ann. Statist.* 19 1725-1747.

Qin, J. & Lawless, J.F. (1994). Empirical likelihood and general estimating equations. *Ann. Statist.* 22 300-325

Qin, J. & Lawless, J.F. (1995). Estimating equations, empirical likelihood and constraints on parameters. *The Canadian Journal of Statistics.* 23 145-159