# EMPIRICAL LIKELIHOOD INFERENCE IN THE PRESENCE OF MEASUREMENT ERROR

C.X.Bob Zhong, Carleton University
Dept. of Math. & Stats. Carleton University, Ottawa, ON. K1S 5B6

## 1 Introduction

Suppose that we want to measure a characteristic of interest of a target population. We have several different "instruments", one of them is "accurate", or it has smaller measurement error compared to the other ones. In practice we treat this instrument as "perfect", or no measurement error. All other instruments will have larger measurement error and we treat them as "imperfect". Here we use "instrument" to refer the method used to measure the characteristic. It may include factors such as the real instrument used, the personnel involved, working environment, etc. It may be due to the cost, or lack of highly trained personnel for using the perfect instrument. We use imperfect instruments to measure some samples drawn from the population, and use perfect instrument to some samples from the population. We combine all the data, both imperfect and perfect, to make inference on the parameters of the population. Of course, the inference should be more accurate compared to using only the perfect measurement data.

Survey researchers have long been cognizant of the ill effects of measurement error on estimators of means and totals of finite populations. In general, compared with perfect measurements, imperfect but unbiased measurements increase the variance, but do not affect the bias of estimators of means and totals. However, the sample cumulative distribution function(CDF) is no longer an unbiased or consistent estimator of the population CDF even if the measurement error has mean zero. Several authors discussed these problems (Overton 1989, Fuller 1995, Luo, Stokes and Sager and the references cited therein).

In this paper we propose to use the empirical likelihood method to make inference on parameters of interest by taking all the data into account, treating the imperfect measurements as auxiliary information. In section 2, we review Qin and Lawless's (1994) method of linking empirical likelihood and estimating functions or equations. Section 3 presents how to use the empirical likelihood to solve the problems above. Section 4 discusses the associated asymptotic results. Several examples will be given in section 5. All proofs will be omitted.

## 2 EL and estimating equations

The outline of the empirical likelihood method as discussed by Owen (1988, 1990) is as follows. Let $x_1, x_2, \cdots, x_n$ be *i.i.d.* observations from a population with a $d$-variate distribution function $F(a) = P(x \leq a)$ and nonsingular covariance matrix. The empirical likelihood function is

$$L(F) = \Pi_{i=1}^n dF(x_i) = \Pi_{i=1}^n p_i, \qquad (2.1)$$

where $p_i = dF(x_i) = Pr(x = x_i)$. Only distributions with an atom of probability on each $x_i$ have nonzero likelihood and (2.1) can be maximized by the empirical distribution function $F_n(x) = n^{-1}\sum_{i=1}^n 1(x_i < x)$, where $1(x_i < x) = (1(x_{i,1} < x_1), \cdots, 1(x_{i,d} < x_d))^\tau$, $x_{i,j}$ and $x_j$ are $j$-th component of vector $x_i$ and $x$ respectively, $1(x_{i,j} < x_j)$ is the indicator of set $(x_{i,j} < x_j)$. The empirical likelihood ratio is then defined as $R(F) = L(F)/L(F_n)$, which can be reduced to

$$R(F) = \Pi_{i=1}^n np_i. \qquad (2.2)$$

Suppose there is a $p$-dimensional parameter $\theta$ associated with $F$, and the information about $\theta$ and $F$ is available in the form of $r \geq p$ functionally independent unbiased estimating functions (as discussed by Qin and Lawless(1994)). That is functions $g_j(x,\theta)$, $j = 1, \cdots, r$, such that

$$E_F\{g(x,\theta)\} = 0, \tag{2.3}$$

where $g(x,\theta) = (g_1(x,\theta), \cdots, g_r(x,\theta))^\tau$. Linking (2.2) and (2.3) together to get an estimator of $\theta$, we define the profile empirical likelihood ratio function

$$\begin{aligned} R_E(\theta) &= sup\{\Pi_{i=1}^n np_i | p_i \geq 0, \sum_{i=1}^n p_i = 1, \\ &\sum_{i=1}^n p_i g(x_i,\theta) = 0\}. \end{aligned} \tag{2.4}$$

A unique value for the right side of (2.4) exists for a given $\theta$, provided that 0 is inside the convex hull of the points $g(x,\theta) = (g_1(x,\theta), \cdots, g_r(x,\theta))^\tau$. An explicit expression for $R_E(\theta)$ can be derived by a Lagrange multiplier argument. The maximum of $\Pi_{i=1}^n np_i$ subject to the constraints $p_i \geq 0$, $\sum_{i=1}^n p_i = 1$ and $\sum_{i=1}^n p_i g(x_i,\theta) = 0$ is attained when

$$p_i = p_i(\theta) = n^{-1}\{1 + t^\tau g(x_i,\theta)\}^{-1}, \tag{2.5}$$

where $t = t(\theta)$ is a $d$-dimension vector given as the solution to

$$\frac{1}{n}\sum_{i=1}^n \{1 + t^\tau g(x_i,\theta)\}^{-1} g(x_i,\theta) = 0. \tag{2.6}$$

The (profile) empirical (negative) log-likelihood ratio function for $\theta$ is then defined as

$$l_E(\theta) = \sum_{i=1}^n \log[1 + t^\tau(\theta) g(x_i,\theta)]. \tag{2.7}$$

Minimizing $l_E(\theta)$, we can get an estimator $\tilde{\theta}$ of the parameter $\theta$, called the maximum empirical likelihood estimator (MELE). In addition, this yields estimator $\tilde{p}_i$, from (2.5) and an estimator for the distribution function $F$, as

$$\tilde{F}_n\,(x) = \sum_{i=1}^n \tilde{p}_i\; 1(x_i < x). \tag{2.8}$$

Qin and Lawless have established some asymptotic results on $\tilde{\theta}$ and $\tilde{F}_n\,(x)$ under mild conditions. Large sample tests and confidence limits for parameters are also obtained.

In the next section we are going to show how to apply this method to solve the problems mentioned in section 1.

# 3 EL method in the presence of measurement error

Suppose that there are $H$ different instruments used in measuring the characteristic of interest. The distribution function associated with instrument $h$ ($h = 1, \cdots, H$) is $F_h(a) = P(x_h < a)$, with unknown parameter $\theta$, where $\theta$ is $p$-dimension vector, and the $H$-th measuring instrument is taken as perfect. These $H$ different populations are related by the common parameter $\theta$. We also assume that information about $\theta$ and $F_h$ is available in the form of $r_h \geq p$ functionally independent unbiased estimation functions, that is functions $g_h(x,\theta)$ such that

$$E_{F_h}(g_h(x,\theta)) = 0, \tag{3.1}$$

where $g_h(x,\theta)$ is a $r_h$-dimension vector function. We further suppose that $x_{h1}, \cdots, x_{hn_h}$ is an $i.i.d.$ sample from $F_h(x)$, and samples measured by different instruments are independent. The sampling scheme considered here is different from the method of Luo and others. They used two-phase sampling with the imperfect sample as the first phase sample and the second phase sample taken from the first phase sample. That means that the second phase sample is dependent on the first phase sample, and usually the second phase sample is a very small portion of the first phase sample. We consider independent samples here for two main reasons. First, in the case different instruments are involved in measuring, the makers may have done a lot of testing and will give the accuracies of different instruments. Or a great number of testing can be made to know the difference between different instruments before practical measuring. The same thing can be said to the different-degree of trained personnel. Our method can take this kind of information into account, hence we can make more accurate inference. Second, as pointed by Luo and others, when the second phase sample is quite smaller compared to the first phase sample, two samples can be treated as independent samples. Actually their recommended estimator for the CDF is based on independent samples.

The empirical likelihood from the independent samples

$$\{x_{h1}, \cdots, x_{hn_h}; h = 1, \cdots, H\}$$

is given by

$$L(F_1, \cdots, F_H) = \Pi_{i=1}^{H}\Pi_{i=1}^{n_h} p_{hi}, \qquad (3.2)$$

where $p_{hi} = Pr(x_h = x_{hi})$ and $\sum_{i=1}^{n_h} p_{hi} = 1$ for each $h$. Only those $F_h$ distributions which have an atom of probability on each $x_{hi}$ have nonzero likelihood. (3.2) can be maximized by the empirical distribution function $F_{h,n_h}(x) = n_h^{-1}\sum_i 1(x_{hi} < x)$, the empirical likelihood ratio is then defined as

$$\begin{aligned} R(F_1, &\cdots, F_H) \\ &= L(F_1, \cdots, F_H)/L(F_{1,n_1}, \cdots, F_{H,n_H}), \end{aligned}$$

which reduces to

$$R(F_1, \cdots, F_H) = \Pi_{h=1}^{H}\Pi_{i=1}^{n_h} n_h p_{hi}. \qquad (3.3)$$

We remark that formulas here and elsewhere in this paper do not require that the $x_{hi}$'s be distinct. Since we are interested in estimating the parameter $\theta$, and we know the estimating equation (3.1), we define the empirical likelihood ratio function

$$\begin{aligned} R_E(\theta) \quad = \quad & sup\{\Pi_{h=1}^{H}\Pi_{i=1}^{n_h} n_h p_{hi} \mid p_{hi} \geq 0, \\ & \sum_i p_{hi} = 1, \ \sum_i p_{hi} g_{hi}(\theta) = 0\}, \end{aligned}$$

where $g_{hi}(\theta) = g_h(x_{hi}, \theta)$ for all $h$ and $i$. Here function $g_{hi}(\theta)$ depends on $h$ also, which is different from the one discussed by Zhong and Rao(1996).

As discussed by Qin and Lawless(1994), for any given $\theta$ and $h$, $\Pi_{i=1}^{n_h} p_{hi}$ can be maximized, provided 0 is inside the convex hull of the point $g_{h1}(\theta), \cdots, g_{hn_h}(\theta)$. The maximum of $\Pi_{i=1}^{n_h} p_{hi}$ subject to the constraints $p_{hi} \geq 0$, $\sum_i p_{hi} = 1$, and $\sum_i p_{hi} g_{hi}(\theta) = 0$ is attained when

$$p_{hi} = p_{hi}(\theta) = n_h^{-1}\{1 + t_h^{\tau} g_{hi}(\theta)\}^{-1}, \qquad (3.4)$$

where $t_h = t_h(\theta)$ is a $r_h \times 1$ vector given as the solution to

$$\frac{1}{n_h}\sum_{i=1}^{n_h}\{1 + t_h^{\tau} g_{hi}(\theta)\}^{-1} g_{hi}(\theta) = 0. \qquad (3.5)$$

Hence the left side of (3.4) is

$$R_E(\theta) = \Pi_h \Pi_i \{1 + t_h^{\tau} g_{hi}(\theta)\}^{-1}, \qquad (3.6)$$

and the empirical (negative) log-likelihood ratio of $\theta$ is

$$l_E(\theta) = \sum_h \sum_i \log[1 + t_h^{\tau} g_{hi}(\theta)]. \qquad (3.7)$$

We can minimize $l_E(\theta)$ to obtain an estimator $\tilde{\theta}$ of the parameter $\theta$, called the maximum empirical likelihood ratio estimator (MELRE). In addition, this yields estimators $\tilde{p}_{hi}$, from (3.5), and an estimator for the true distribution function $F_H$ as

$$\tilde{F}_{H,n_H}(x) = \sum_{i=1}^{n_H} \tilde{p}_{Hi}\, 1(x_{Hi} < x). \qquad (3.8)$$

$\tilde{\theta}$ may be obtained by solving

$$\frac{\partial l_E(\theta)}{\partial \theta} = \sum_h \sum_i \frac{1}{1 + t_h^{\tau} g_{hi}(\theta)} \left(\frac{\partial g_{hi}(\theta)}{\partial \theta^{\tau}}\right) t_h = 0 \qquad (3.9)$$

and (3.6) together.

In the following section we will discuss the existence of $\tilde{\theta}$ and study the asymptotic properties of $\tilde{\theta}$, $\tilde{F}_{H,n_H}(x)$.

# 4 Asymptotic properties of MELRE

First, we give the conditions for the existence of $\tilde{\theta}$ which will minimize $l_E(\theta)$ defined by (3.8).

In the following, we will use $\|\cdot\|$ to denote Euclidean norm and $n = \sum_h n_h$.

**Lemma 1.** *Suppose as $n \to \infty$, $n/n_h \to k_h > 0$ for all $h$. And suppose that in a neighborhood of the true value $\theta_0$, $Var[g_h(x_h, \theta_0)] = \sigma_h(\theta_0) > 0$ for all $h$, $\| \partial g_h(x_h, \theta)/\partial\theta \|$ and $\| g_h(x_h, \theta) \|^4$ are bounded by some integrable function $G(x)$ in this neighbourhood, and the rank of $E[\partial g_h(x_h, \theta)/\partial\theta]$ is $p$. Then, as $n \to \infty$, with probability 1 $l_E(\theta)$ attains its minimum value at some point $\tilde{\theta}$ in the interior of the ball $\| \theta - \theta_0 \| \leq n^{-\frac{1}{3}}$, and $\tilde{\theta}$ and $\tilde{t}_h = \tilde{t}_h(\tilde{\theta})$ satisfy*

$$Q_h(\tilde{\theta}, \tilde{t}_h) = 0, \ h = 1, \cdots, H,$$

$$Q_{H+1}(\tilde{\theta}, \tilde{t}_1, \cdots, \tilde{t}_H) = 0, \qquad (4.1)$$

*where*

$$Q_h(\theta, t_h) = \frac{1}{n_h}\sum_{i=1}^{n_h} \frac{g_{hi}(\theta)}{1 + t_h^{\tau} g_{hi}(\theta)}, h = 1, \cdots, H, \qquad (4.2)$$

$$Q_{H+1}(\theta, t_1, \cdots, t_H) = \sum_h \sum_i \frac{1}{1+t_h^\tau g_{hi}(\theta)} \frac{\partial g_{hi}(\theta)}{\partial \theta} t_h = 0. \quad (4.3)$$

**Theorem 1.** *In addition to the conditions of Lemma 1 above, we further assume that* $\frac{\partial^2 g_h(x_h,\theta)}{\partial\theta\partial\theta^\tau}$ *is continuous in* $\theta$ *in a neighbourhood of the true value* $\theta_0$*, then if* $\| \frac{\partial^2 g_h(x_h,\theta)}{\partial\theta\partial\theta^\tau} \|$ *can be bounded by some integrable function* $G(x)$ *in the neighbourhood for all* $h$*, then*

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

$$\sqrt{n}(\tilde{F}_{H,n_H}(x) - F_H(x)) \rightarrow N(0, W),$$

*where*

$$\tilde{F}_{H,n_H}(x) = \sum_i \tilde{p}_{Hi}\, 1(x_{Hi} < x),$$

$$\tilde{p}_{Hi} = \frac{1}{n_H} \frac{1}{1+ \tilde{t}_H^\tau\, g_{Hi}(\tilde{\theta})},$$

$$\begin{aligned} M_{11} &= diag\{\sigma_1(\theta_0), \cdots, \sigma_H(\theta_0)\} \\ &= diag\{\sigma_1, \cdots, \sigma_H\}, \end{aligned}$$

$$M_{12} = \left(-E\frac{\partial g_1(\theta_0)}{\partial\theta}, \cdots, -E\frac{\partial g_H(\theta_0)}{\partial\theta}\right)^\tau,$$

$$M_{21} = M_{12}^\tau, \quad M_{22.1} = -M_{21}M_{11}^{-1}M_{12},$$

$$V = M_{22.1}^{-1}\left[\sum_h k_h E\frac{\partial g_h(\theta_0)}{\partial\theta}\sigma_h^{-1}E\frac{\partial g_h(\theta_0)}{\partial\theta^\tau}\right] \cdot M_{22.1}^{-1}, \quad (4.4)$$

$W$

$$\begin{aligned} & F_H(x)[1 - F_H(x)] \\ & -2B\sigma_H^{-1}U_{HH}B^\tau \\ & +B\sigma_H^{-1}U_{HH}\sigma_H U_{HH}^\tau \sigma_H^{-1}B^\tau \\ & +B\sigma_H^{-1}\sum_{h=1}^{H-1}\left(\frac{k_h}{k_H}\right)U_{Hh}\sigma_h U_{Hh}^\tau\sigma_H^{-1}B^\tau \quad (4.5) \end{aligned}$$

$$U_{HH} = \left(I + E\frac{\partial g_H(\theta_0)}{\partial\theta^\tau}M_{22.1}^{-1}E\frac{\partial g_H(\theta_0)}{\partial\theta}\sigma_H^{-1}\right),$$

$$U_{Hh} = \left(I + E\frac{\partial g_H(\theta_0)}{\partial\theta^\tau}M_{22.1}^{-1}E\frac{\partial g_h(\theta_0)}{\partial\theta}\sigma_h^{-1}\right),$$

$$B = Eg_H^\tau(x_H, \theta_0)1(x_H < x).$$

Since usually $n_h >> n_H$ for $h = 1, \cdots, H-1$, we can see the main part of the variance of $\tilde{F}_{H,n_H}(x)$ will come from the first term of (4.5), which is smaller than the variance of the estimator $F_{H,n_H}(x)$, the cumulative distribution function from the precise data $x_{H1}, \cdots, x_{Hn_H}$ only. Some further discussion will be given in Section 6.

The estimators for variance of $\tilde{\theta}$ and $\tilde{F}_{H,n_H}(x)$ can be obtained by replacing each component of $V$ and $W$ by their obvious estimators. These estimators are asymptotically correct.

# 5 Examples

We present several illustrations of the estimation procedures. Procedures about how to solve equations (4.1) through (4.3) will be given. Large-sample aspects of these estimators will be discussed.

We suppose only two different instruments are used, i.e., $H = 2$, for the following examples.

**Example 1. Common Mean Model**

Suppose we only know that they are unbiased, i.e.,

$$Ex_1 = \theta, Ex_2 = \theta,$$

and $x_2$ refers to the perfect measurement. Suppose that $x_{11}, \cdots, x_{1n_1}$ and $x_{21}, \cdots, x_{2n_2}$ are independent samples from $x_1$ and $x_2$ respectively. Let

$$g_1(x_1, \theta) = x_1 - \theta, g_2(x_2, \theta) = x_2 - \theta,$$

then

$$\frac{\partial g_1}{\partial\theta} = -1 = \frac{\partial g_2}{\partial\theta} = E\frac{\partial g_1}{\partial\theta} = E\frac{\partial g_2}{\partial\theta},$$

and equation (4.3) becomes

$$n_1t_1 + n_2t_2 = 0, i.e., t_2 = -\frac{n_1}{n_2}t_1. \quad (5.1)$$

Substituting (5.1) into (4.2), we get

$$\frac{1}{n_1}\sum_{i=1}^{n_1}\frac{x_{1i} - \theta}{1 + t(x_{1i} - \theta)} = 0, \quad (5.2)$$

$$\frac{1}{n_2}\sum_{i=1}^{n_2}\frac{x_{2i} - \theta}{1 - \frac{n_1}{n_2}t(x_{2i} - \theta)} = 0. \quad (5.3)$$

Solving (5.2) and (5.3) by Newton method or using NAG Fortran library function with initial value $(\theta, t) = (\bar{x}, 0)$ often works, where $\bar{x} = \frac{n_1}{n}\bar{x}_1 + \frac{n_2}{n}\bar{x}_2$.

Let $Var(x_1) = \sigma_1, \quad Var(x_2) = \sigma_2$. Since

$$\sum_h k_h E \frac{\partial g_h}{\partial \theta} \sigma_h^{-1} \frac{\partial g_h}{\partial \theta^\tau} = \sum_h k_h \sigma_h^{-1},$$

replacing $k_h$ by $n/n_h$ we can get the asymptotic variance of $\tilde{\theta}$ as

$$\begin{aligned} \frac{V}{n} &= \frac{1}{n}\left(\sum_h \sigma_h^{-1}\right)^{-1} \\ &\cdot \left(\sum_h \frac{n}{n_h}\sigma_h^{-1}\right)^{-1}\left(\sum_h \sigma_h^{-1}\right)^{-1} \\ &= \left(\frac{\sigma_2}{n_1} + \frac{\sigma_1}{n_2}\right)\frac{\sigma_1\sigma_2}{(\sigma_1+\sigma_2)^2}, \end{aligned}$$

which is the same as the variance of the optimal linear combination "estimator"

$$\hat{\theta} = \frac{\sigma_2}{\sigma_1+\sigma_2}\bar{x}_1 + \frac{\sigma_1}{\sigma_1+\sigma_2}\bar{x}_2.$$

Note $\hat{\theta}$ is actually not an estimator since we do not know the variances $\sigma_1$ and $\sigma_2$, but we use it for comparison here and in the following.

As for the asymptotic variance of $\tilde{F}_{H,n_H}(x)$, i.e., $W$ given by (4.5), after a little calculation we get

$$\begin{aligned} W &= F_2(x)[1-F_2(x)] + B^2\sigma_2^{-1} \\ &\cdot \left[\frac{\sigma_1^{-2} + k_1k_2^{-1}\sigma_1\sigma_2^{-3}}{(\sigma_1^{-1}+\sigma_2^{-1})^2} - \frac{2\sigma_1^{-1}}{\sigma_1^{-1}+\sigma_2^{-1}}\right] \\ &= F_2(x)[1-F_2(x)] + B^2\frac{\sigma_2^{-3}}{(\sigma_1^{-1}+\sigma_2^{-1})^2} \\ &\cdot \left[-\left(\frac{\sigma_2}{\sigma_1}\right)^2 - 2\left(\frac{\sigma_2}{\sigma_1}\right) + \frac{k_1}{k_2}\left(\frac{\sigma_2}{\sigma_1}\right)\right]. \end{aligned}$$

From above we can see that if $\sigma_2 << \sigma_1$, then combining the two samples will not improve the efficiency of CDF estimator compared to using only the perfect data. On the other hand, if $\sigma_2/\sigma_1$ is reasonable large, such as greater than 0.5, then for example, $n_2/n_1 < 5/8$, we will have a more efficient estimator for CDF.

### Example 2 Additive Model

We consider the following model

$$Ex_1 = \theta_1, \quad Var(x_1) = \theta_2 + \sigma_0,$$

$$Ex_2 = \theta_2, \quad Var(x_2) = \theta_2,$$

i.e., the variance of imperfect measurement differs from that of the perfect measurement by $\sigma_0$, where $\sigma_0(> 0)$ is known. We note here that if $Ex_1 = \theta_1 + \nu_0$, where $\nu_0$ is known, we can change the data $x_1$ to $x_1 - \nu_0$, and all the following discussion still applies.

Suppose two independent samples from $x_1$ and $x_2$ are $x_{11}, \cdots, x_{1n_1}$ and $x_{21}, \cdots, x_{2n_2}$ respectively. Denote $z_1 = (x_1, x_1^2)^\tau$, $z_2 = (x_2, x_2^2)^\tau$, $\theta = (\theta_1, \theta_2)^\tau$, $g_1(z_1, \theta) = (x_1 - \theta_1, x_1^2 - \theta_1^2 - \theta_2 - \sigma_0^2)^\tau$, $g_2(z_2, \theta) = (x_2 - \theta_1, x_2^2 - \theta_1^2 - \theta_2)^\tau$, then $Eg_1 = 0, \quad Eg_2 = 0$,

$$\frac{\partial g_1}{\partial \theta} = \begin{pmatrix} -1 & -2\theta_1 \\ 0 & -1 \end{pmatrix} = \frac{\partial g_2}{\partial \theta} = E\frac{\partial g_1}{\partial \theta} = E\frac{\partial g_2}{\partial \theta},$$

and equation (4.3) becomes

$$n_1t_{11} + n_2t_{21} = 0, \quad n_1t_{12} + n_2t_{22} = 0,$$

or

$$t_{21} = -\frac{n_2}{n_1}t_{11}, \quad t_{22} = -\frac{n_2}{n_1}t_{12}.$$

Hence, (4.2) becomes

$$\sum_{i=1}^{n_1} \frac{(1/n_1)(x_{1i} - \theta_1)}{dd} = 0,$$

$$\sum_{i=1}^{n_1} \frac{(1/n_1)(x_{1i}^2 - \theta_1^2 - \theta_2 - \sigma_0^2)}{dd} = 0,$$

$$\frac{1}{n_2}\sum_{i=1}^{n_1} \frac{x_{2i} - \theta_1}{dd} = 0,$$

$$\frac{1}{n_2}\sum_{i=1}^{n_1} \frac{x_{2i}^2 - \theta_1^2 - \theta_2}{dd} = 0.$$

where

$$dd = 1 + t_{11}(x_{1i} - \theta_1) + t_{12}(x_{1i}^2 - \theta_1^2 - \theta_2 - \sigma_0^2).$$

We can solve the above four equations by letting the initial values

$$\theta_{10} = \frac{n_1}{n_1+n_2}\bar{x}_1 + \frac{n_2}{n_1+n_2}\bar{x}_2,$$

$$\theta_{20} = \frac{n_1}{n_1+n_2}S_{1n_1}^2 + \frac{n_2}{n_1+n_2}S_{2n_2}^2$$

$$t_{11,0} = 0, \quad t_{12,0} = 0.$$

Using Newton's method will usually give us a solution to $\theta_1$, $\theta_2$, and hence $\tilde{p}_{hi}$, $\tilde{F}_{2n_2}(x)$.

In order to get an idea of the asymptotic properties of these estimators, we suppose $x_1$ and $x_2$ have normal distribution. We can get

$$V = (k_1\theta_2 + k_2\theta_*)\theta_2\theta_*/(\theta_2 + \theta_*)^2.$$

If we replace $k_h$ by $n/n_h$, then $Var(\tilde{\theta}) \approx (n_1^{-1}\theta_2 + n_2^{-1}\theta_*)\theta_2\theta_*/(\theta_2 + \theta_*)^2$, which is the same as the variance of the optimal linear combination estimator

$$\hat{\theta} = \frac{\theta_2}{\theta_2 + \theta_*}\bar{x}_1 + \frac{\theta_*}{\theta_2 + \theta_*}\bar{x}_2.$$

**Example 3. Product Model**

Suppose

$$Ex_1 = \theta_1, \; Var(x_1) = c\theta_2,$$

$$Ex_2 = \theta_2, \; Var(x_2) = \theta_2,$$

i.e., the ratio of variances of imperfect measurement to perfect measurement is a known constant($c > 0$). The two independent samples from $x_1$ and $x_2$ are $x_{11}, \cdots, x_{1n_1}$ and $x_{21}, \cdots, x_{2n_2}$ respectively. The method to solve (4.1) to (4.3) is similar to example 2 and is omitted.

In the following discussion, we suppose $x_1$ and $x_2$ are normally distributed. Then we can get

$$Var(\tilde{\theta}) \approx \left(\frac{c}{1+c}\right)^2 \left(\frac{1}{n_1 c} + \frac{1}{n_2}\right)\theta_2,$$

which is the same as the variance of the optimal linear combination estimator

$$\hat{\theta} = \frac{\theta_2}{c\theta_2 + \theta_2}\bar{x}_1 + \frac{c\theta_2}{c\theta_2 + \theta_2}\bar{x}_2.$$

## 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.

DiCiccio, T.J., P. Hall & J.P. Romano (1991). Empirical likelihood is Bartlett-correctable. *Ann. Statist. 19, 1053-1061*

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

Holl, P. (1990). Pseudo-likelihood theory for empirical likelihood. *Ann. Statist.* 18, 121-140

Luo, M. & Stokes, L. & Sager, T. Estimation of the CDF of a Finite Population in the Presence of a Calibration Sample

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

Overton, W. S. (1989). Effects of measurement and other extraneous errors on estimated distribution functions in the national surface water surveys. *Technical report 129, Dept. of Statistics, Oregon State University.*

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

Rao, J.N.K. & Liu, J. (1992) On estimating distribution function from sample survey data using supplementary information at the estimation stage. *Nonparametric Statistics and Related Topics (A.K.Md.E. Saleh, Ed.), New York: Elswior, 399-4.7*

Zhong, C. & Rao, J.N.K. (1996). Empirical likelihood inference under stratified sampling using auxiliary information. *The proceeding of JSM 96*