The Canadian Journal of Statistics
Vol. 34, No. 1, 2006, Pages 45–59
La revue canadienne de statistique

Empirical likelihood inference for a common mean in the presence of heteroscedasticity

Min TSAO and Changbao WU

Key words and phrases: Combining information; confidence intervals; heteroscedasticity; independent samples; weighted empirical likelihood.

MSC 2000: Primary 62A10; secondary 62D05.

Abstract: The authors develop empirical likelihood (EL) based methods of inference for a common mean using data from several independent but nonhomogeneous populations. For point estimation, they propose a maximum empirical likelihood (MEL) estimator and show that it is $\sqrt{n}$-consistent and asymptotically optimal. For confidence intervals, they consider two EL based methods and show that both intervals have approximately correct coverage probabilities under large samples. Finite-sample performances of the MEL estimator and the EL based confidence intervals are evaluated through a simulation study. The results indicate that overall the MEL estimator and the weighted EL confidence interval are superior alternatives to the existing methods.

Inférence par la vraisemblance empirique pour une moyenne commune en présence d'hétéroscédasticité

Résumé : Les auteurs développent des méthodes d'inférence fondées sur la vraisemblance empirique (VE) pour la moyenne commune de populations indépendantes non-homogènes. Pour l'estimation ponctuelle, ils proposent un estimateur de vraisemblance empirique maximale (VEM) dont ils montrent qu'il est $\sqrt{n}$-convergent et asymptotiquement optimal. Pour l'estimation par intervalle, ils considèrent deux méthodes basées sur la VE et montrent que les intervalles correspondants ont à peu près la bonne couverture dans de grands échantillons. Les performances à taille finie de l'estimateur de VEM et des intervalles de confiance de VE sont évaluées par voie de simulation. Les résultats indiquent que globalement, l'estimateur de VEM et l'intervalle de confiance de VE pondéré sont supérieurs aux méthodes existantes.

1. INTRODUCTION

Estimation of an unknown quantity using information from several independent but nonhomogeneous samples is a classical problem in statistics. Applications can be found in many different contexts. For example, suppose two or more technicians perform assays on several sample materials. The technicians measure the same characteristic but their measurements differ in precision. To make inferences about the common characteristic, we wish to make use of the combined sample data. This example is a special case of the more general measurement error problems where several "instruments" are used to collect data on a common response variable. These instruments are believed to have no systematic biases but differ in precision. Heteroscedasticity clearly is the key feature of the combined sample data. Another example is in designed experiments involving a single factor. When one fails to reject the null hypothesis that the mean responses at different treatment levels are all the same, a point estimate and/or a confidence interval for the assumed common mean using the combined sample data will be the main focus of subsequent analyses.

Such common mean problems have attracted much attention over the years and have led to some very interesting solutions. Early work on such problems typically assumed that the underlying distributions are normal. Neyman and Scott (1948) gave an estimator which is more efficient than the maximum likelihood estimator when the number of samples approaches infinity. Cochran & Carroll (1953), Meier (1953) and Bement & Williams (1969) examined the relative efficiency of a weighted estimator to a simple unweighted estimator. Levy (1970) compared the weighted estimator with the maximum likelihood estimator. C. R. Rao (1970), J. N. K. Rao

(1973), and Hartley & Jayatillake (1973), among others, considered a more general problem of estimation for linear models with unequal variances and some of these have their emphasis on estimating the variances. For relative efficiencies of some commonly used estimators and further references, see J. N. K. Rao (1980) and the references therein.

In the present paper, we bring to bear a modern method of inference on this classical yet still frequently encountered problem. We develop empirical likelihood (EL) based methods for making inferences about the common mean; these methods are found to be very promising. The EL approach (Owen 1988) is nonparametric and requires only some mild finite-moment conditions. We define the maximum empirical likelihood (MEL) estimator and show that it is $\sqrt{n}$-consistent and is asymptotically optimal. For finite samples, simulation results show that the MEL estimator performs similarly to existing estimators when the parametric model is correctly specified but it performs much better otherwise. Past work on point estimators has sometimes avoided discussion of the construction of confidence intervals, partly because the methods involved do not readily lend themselves to dealing with interval estimation. With the EL approach, however, confidence intervals can be constructed in more than one way. We develop two different ways of constructing EL confidence intervals for the common mean. The weighted EL confidence interval emerges as the most reliable method when sample sizes are small or moderate, while the naive EL confidence interval performs poorly for cases where sample sizes are not large and/or the underlying population distributions are skewed. The MEL estimator and the weighted EL confidence interval are found to be superior to the existing methods. In particular, they are the most robust in that their performances are the least affected by the change of the underlying distributions.

The rest of the paper is organized as follows. In Section 2, we describe the two most commonly used estimators: the optimal estimator and the maximum likelihood estimator, and set out our notation. In Section 3, we present the MEL estimator and discuss its basic properties. In Section 4, we present two EL based confidence intervals. A major issue in the implementation of the EL methods is computation. In Section 5, we discuss computational algorithms. In Section 6, we examine finite-sample performances of the proposed methods, and compare them to the existing approaches through a limited simulation study. An application to a set of real data on assessing the quality of a newly formulated gasoline is presented in Section 7. We conclude with some additional remarks in Section 8.

2. OPTIMAL AND MAXIMUM LIKELIHOOD ESTIMATORS

The standard formulation of the common mean problem is as follows. Let Y_{ij}, $j = 1, \ldots, n_i$, $i = 1, \ldots, k$ be independent observations such that

$$\mathrm{E}\,(Y_{ij}) = \mu_0 \quad \text{and} \quad \mathrm{var}(Y_{ij}) = \sigma_i^2.$$

The variances σ_i^2 are unknown. The focus here is on estimating the common mean μ_0 and constructing confidence intervals using the combined sample data.

There are two asymptotic scenarios which are often of interest. One is that all n_i are bounded while the number of samples k goes to infinity; the other holds k fixed and allows each of the sample sizes n_i to approach infinity. In this paper, we restrict our discussion to the latter case. We also assume that $n_i/n \to f_i \neq 0$ as $n = n_1 + \cdots + n_k \to \infty$. Under this setting, it is not necessary to distinguish between $n_i \to \infty$ and $n \to \infty$, or between $O(n_i^{-1/2})$ and $O(n^{-1/2})$, etc.

The most straightforward and yet very attractive method is the optimal combination estimator. Let

$$\hat{\mu} = \sum_{i=1}^{k} c_i \overline{Y}_{i\cdot}, \quad \text{where } \overline{Y}_{i\cdot} = n_i^{-1} \sum_{j=1}^{n_i} Y_{ij},$$

with $c_i \geq 0$ and $c_1 + \cdots + c_k = 1$. The optimal choices of c_i which minimize $\mathrm{var}(\hat{\mu})$ are given by

$$c_i = \frac{n_i}{\sigma_i^2} \Big/ \sum_{i=1}^{k} \frac{n_i}{\sigma_i^2}.$$

Since the optimal choices of c_i depend on the unknown variances σ_i^2, the following *asymptotically optimal estimator* where σ_i^2 is replaced by

$$s_i^2 = (n_i - 1)^{-1} \sum_{j=1}^{n_i} (Y_{ij} - \overline{Y}_{i\cdot})^2$$

is used in practice instead:

$$\hat{\mu}_{\mathrm{op}} = \sum_{i=1}^{k} \left(\frac{n_i}{s_i^2}\right) \overline{Y}_{i\cdot} \Big/ \sum_{i=1}^{k} \left(\frac{n_i}{s_i^2}\right). \tag{1}$$

Since $\overline{Y}_{i\cdot}$ and s_i^2 are $\sqrt{n}$-consistent estimators for μ_0 and σ_i^2, respectively, it is clear that $\hat{\mu}_{\mathrm{op}}$ is also a $\sqrt{n}$-consistent estimator for μ_0.

Parametric approaches can be easily explored. Under normality assumptions where $Y_{ij} \sim \mathsf{N}(\mu_0, \sigma_i^2)$, the maximum likelihood (ML) estimator of μ_0, denoted by $\hat{\mu}_{\mathrm{ml}}$, can be shown to be the solution to

$$\sum_{i=1}^{k} \frac{n_i(\overline{Y}_{i\cdot} - \mu)}{n_i^{-1} \sum_{j=1}^{n_i} (Y_{ij} - \mu)^2} = 0.$$

The $\hat{\mu}_{\mathrm{ml}}$ under the assumed normal distributions is also approximately the same as the optimal combination estimator since we can rewrite $\hat{\mu}_{\mathrm{ml}}$ as

$$\hat{\mu}_{\mathrm{ml}} = \sum_{i=1}^{k} \left(\frac{n_i}{\hat{\sigma}_i^2}\right) \overline{Y}_{i\cdot} \Big/ \sum_{i=1}^{k} \left(\frac{n_i}{\hat{\sigma}_i^2}\right),$$

where

$$\hat{\sigma}_i^2 = n_i^{-1} \sum_{j=1}^{n_i} (Y_{ij} - \hat{\mu}_{\mathrm{ml}})^2.$$

Under the normality condition, the maximum likelihood estimator $\hat{\mu}_{\mathrm{ml}}$ is $\sqrt{n}$-consistent and is asymptotically optimal in the sense that $\hat{\mu}_{\mathrm{ml}} = \hat{\mu}_{\mathrm{op}} + o_p(n^{-1/2})$.

Confidence intervals based on these point estimators are not easy to construct. For the optimal estimator $\hat{\mu}_{\mathrm{op}}$, the major difficulty is the associated variance estimation problem if one wishes to construct a confidence interval through normal approximation to the distribution of $\hat{\mu}_{\mathrm{op}}$. Meier (1953) derived an approximate variance estimator for $\hat{\mu}_{\mathrm{op}}$ under the normality assumption, but in that case it may be preferable to use a confidence interval based on the maximum likelihood estimator $\hat{\mu}_{\mathrm{ml}}$. Also, for parametric approaches their vulnerability to model misspecifications makes these intervals undesirable in many practical situations.

3. THE MAXIMUM EMPIRICAL LIKELIHOOD ESTIMATOR

The empirical likelihood approach requires only some finite-moment conditions on the underlying population distributions. To define the maximum EL estimator, we first formulate the empirical likelihood function using a k-sample approach similar to the ANOVA setting discussed in Owen (2001, p. 87). Let $F_1, \ldots, F_k$ be the underlying distribution functions for the k samples, respectively. The joint empirical log-likelihood function is given by

$$\ell(F_1, \ldots, F_k) = \sum_{i=1}^{k} \sum_{j=1}^{n_i} \log(p_{ij}),$$

where $p_{ij} = F_i(Y_{ij}) - F_i(Y_{ij}-)$ is the probability of getting the value Y_{ij} in a sample from F_i. The maximum empirical likelihood estimator of the common mean μ_0 is defined as

$$\hat{\mu}_{\mathrm{el}} = \sum_{j=1}^{n_1} \hat{p}_{1j} Y_{1j} = \cdots = \sum_{j=1}^{n_k} \hat{p}_{kj} Y_{kj},$$

where the $\hat{p}_{ij}$ maximize $\ell(F_1, \ldots, F_k)$ subject to

$$\sum_{j=1}^{n_i} p_{ij} = 1 \ (p_{ij} > 0),\ i = 1, \ldots, k \quad \text{and} \quad \sum_{j=1}^{n_1} p_{1j} Y_{1j} = \cdots = \sum_{j=1}^{n_k} p_{kj} Y_{kj}. \tag{2}$$

Let $\mathcal{H}_i = \mathcal{H}(Y_{i1}, \ldots, Y_{in_i})$ be the convex hull formed by the ith sample data, $i = 1, \ldots, k$. Since we only consider a univariate response variable, it follows that $\mathcal{H}_i = (Y_{(i1)}, Y_{(in_i)})$ and the intersection

$$\mathcal{H}^* = \mathcal{H}_1 \cap \cdots \cap \mathcal{H}_k = \big(\max\{Y_{(11)}, \ldots, Y_{(k1)}\}, \min\{Y_{(1n_1)}, \ldots, Y_{(kn_k)}\}\big),$$

where $Y_{(i1)} = \min(Y_{i1}, \ldots, Y_{in_i})$ and $Y_{(in_i)} = \max(Y_{i1}, \ldots, Y_{in_i})$ are the smallest and the largest order statistics for the ith sample. The empirical likelihood estimator $\hat{\mu}_{\mathrm{el}}$ is defined if and only if $\mathcal{H}^* \neq \emptyset$. It is not difficult to show that as $n_i \to \infty$, with probability tending to one μ_0 is an interior point of $\mathcal{H}_i$. Since k is fixed, it follows that when $\min_i(n_i) \to \infty$, with probability tending to one μ_0 is an interior point of all $\mathcal{H}_i$. Thus $\mathcal{H}^* = \mathcal{H}_1 \cap \cdots \cap \mathcal{H}_k \neq \emptyset$ as all $n_i \to \infty$. In the following, we shall assume that $\mathcal{H}^* \neq \emptyset$ and in particular $\mu_0 \in \mathcal{H}^*$. The following theorem gives some basic properties of $\hat{\mu}_{\mathrm{el}}$.

THEOREM 1. *Let $Y_{i1}, \ldots, Y_{in_i}$ be independent random variables with common distribution F_i, $i = 1, \ldots, k$, and let the k samples be independent. Suppose* $\mathrm{E}\,(Y_{ij}) = \mu_0$, $\mathrm{var}(Y_{ij}) = \sigma_i^2 < \infty$ *for all i. Then*

(i) *The maximum empirical likelihood estimator $\hat{\mu}_{\mathrm{el}}$ exists and is unique.*

(ii) *$\hat{\mu}_{\mathrm{el}}$ is a $\sqrt{n}$-consistent estimator of μ_0.*

(iii) *$\hat{\mu}_{\mathrm{el}}$ is asymptotically optimal in the sense that $\hat{\mu}_{\mathrm{el}} = \hat{\mu}_{\mathrm{op}} + o_p(n^{-1/2})$.*

(iv) *$\hat{\mu}_{\mathrm{el}}$ is asymptotically normally distributed.*

A proof of Theorem 1 is given in the Appendix. It should be noted that $\sqrt{n}$-consistency follows from the asymptotic normality. A direct proof of the asymptotic normality of the maximum empirical likelihood estimator, however, can be quite involved. See Zhong, Chen & Rao (2000) for such a proof. The argument presented in the Appendix takes advantage of the specific structure of the k-sample problem and is straightforward. The asymptotic normality is typically used for constructing pivotal statistics for confidence intervals. It is not, however, of great interest here since the empirical likelihood ratio confidence intervals to be presented in Section 4 are more appealing than the usual symmetric intervals based on normal approximations.

Theorem 1 implies that the three point estimators $\hat{\mu}_{\mathrm{el}}$, $\hat{\mu}_{\mathrm{op}}$ and $\hat{\mu}_{\mathrm{ml}}$ are asymptotically equivalent. For finite samples, however, results from a limited simulation study reported in Section 6 show that $\hat{\mu}_{\mathrm{el}}$ has comparable performance to $\hat{\mu}_{\mathrm{ml}}$ or $\hat{\mu}_{\mathrm{op}}$ under normal distributions but performs much better when the underlying distributions are skewed. An algorithm for computing $\hat{\mu}_{\mathrm{el}}$ is briefly discussed in Section 5.

4. EMPIRICAL LIKELIHOOD RATIO CONFIDENCE INTERVALS

One of the major advantages of using the EL approach is the easy construction of the empirical likelihood ratio confidence intervals. The EL confidence interval has several distinctive features including (i) range respecting for the parameter space; (ii) not necessarily symmetric but data-driven shape; and (iii) invariant coverage probabilities and length of the interval for the mean under scale and location transformations. We now present two methods of constructing EL confidence intervals for μ_0.

4.1. The k-sample empirical likelihood approach.

A naive empirical likelihood approach for constructing a confidence interval for μ_0 proceeds as follows. For an arbitrarily fixed μ, let

$$r(\mu) = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\log(n_i p_{ij})$$

be the empirical log-likelihood ratio statistic, where the p_{ij} maximize $\ell(F_1,\ldots,F_k)$ subject to

$$\sum_{j=1}^{n_i} p_{ij} = 1\ (p_{ij} > 0),\ i = 1,\ldots,k \quad \text{and} \quad \sum_{j=1}^{n_1} p_{1j}Y_{1j} = \cdots = \sum_{j=1}^{n_k} p_{kj}Y_{kj} = \mu. \tag{3}$$

This formulation is equivalent to the one discussed by Owen (2001, p. 89). If we rewrite $r(\mu)$ as

$$r(\mu) = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\log(p_{ij}/\hat{p}_{ij}),$$

where $\hat{p}_{ij} = 1/n_i$ is the maximizer of $\ell(F_1,\ldots,F_k)$ subject to $p_{i1}+\cdots+p_{in_i} = 1$, $i = 1,\ldots,k$, then the difference between the number of constraints used for computing p_{ij} and the one for $\hat{p}_{ij}$ is k. We have the following result based on which a confidence interval for μ_0 may be constructed.

THEOREM 2. *Suppose that the conditions of Theorem 1 hold. Then $-2r(\mu_0)$ converges in distribution to a χ^2 random variable with k degrees of freedom as $\min_i(n_i) \to \infty$.*

A proof of Theorem 2 can be found in the Appendix. It follows that a $100 \times (1-\alpha)$% level confidence interval for μ_0 can be constructed as

$$C_{elk} = \big\{\mu \mid -2r(\mu) < \chi^2_{[k]}(\alpha)\big\}, \tag{4}$$

where $\chi^2_{[k]}(\alpha)$ is the $1-\alpha$ quantile from a χ^2 distribution with k degrees of freedom. The confidence interval (4) has asymptotically correct coverage probability and performs reasonably well when sample sizes n_i are all large but it also has several restrictions under small samples. This is further discussed in Sections 6–8.

4.2. The weighted empirical likelihood approach.

When the variances σ_i^2 are equal, one could simply treat $\{Y_{ij},\ j = 1,\ldots,n_i,\ i = 1,\ldots,k\}$ as a single pooled sample from a population with mean μ_0 and variance, say, σ^2. For such cases, even though each of the n_i might be small or moderate, the combined sample size $n = n_1 + \cdots + n_k$ is often large enough that the usual one sample EL confidence interval for μ_0 can provide more reliable and accurate results. The negative impact from the rather restrictive condition of nonempty joint convex hull of the k samples (see further discussion in Section 8) is not usually a problem.

When the σ_i^2 are different but known, the weighted empirical likelihood approach proposed by Wu (2004a) can be used where the pooled data are viewed as a single sample and the unequal variance structure of the data is accommodated through an explicit weighting of the empirical log-likelihood function using the σ_i^2. Wu's original formulation of the weighted empirical likelihood deals with the situation where $Y_1, \ldots, Y_n$ are independent with common mean $\mu_0 = \mathrm{E}\,(Y_i)$ and nonhomogeneous variance $\mathrm{var}(Y_i) = v_i \sigma^2$. The v_i are assumed to be known. The weighted log-likelihood function is defined as

$$\ell_w(F) = C_n \sum_{i=1}^{n} v_i \{\log(p_i) - np_i\},$$

where $C_n = n/(v_1 + \cdots + v_n)$ is a scaling constant. To establish the asymptotic χ^2 distribution of the weighted empirical likelihood ratio statistic $r_w(\mu_0) = \ell_w(\mu_0) - \ell_w(\hat{\mu})$, several finite moment conditions on the Y_i as well as on the constant sequence v_i are required. See Wu (2004a) for further details.

Under the current setting, we view $\{Y_{ij},\ j = 1, \ldots, n_i,\ i = 1, \ldots, k\}$ as a single independent sample such that $\mathrm{E}\,(Y_{ij}) = \mu_0$ and $\mathrm{var}(Y_{ij}) = \sigma_i^2$. Since the σ_i^2 are typically unknown, we use

$$s_i^2 = (n_i - 1)^{-1} \sum_{j=1}^{n_i} (Y_{ij} - \overline{Y}_{i\cdot})^2$$

in the formulation of a weighted empirical log-likelihood function. Let

$$C_n = n \Bigg/ \sum_{i=1}^{k} n_i s_i^2 \quad \text{and} \quad \ell_w(\mu) = C_n \sum_{i=1}^{k} s_i^2 \sum_{j=1}^{n_i} \{\log(p_{ij}) - np_{ij}\},$$

where the p_{ij} maximize

$$\ell_w = \sum_{i=1}^{k} s_i^2 \sum_{j=1}^{n_i} \{\log(p_{ij}) - np_{ij}\}$$

subject to

$$\sum_{i=1}^{k} \sum_{j=1}^{n_i} p_{ij} = 1 \ (p_{ij} > 0) \quad \text{and} \quad \sum_{i=1}^{k} \sum_{j=1}^{n_i} p_{ij} Y_{ij} = \mu.$$

Let

$$\ell_w(\hat{\mu}) = C_n \sum_{i=1}^{k} s_i^2 \sum_{j=1}^{n_i} \{\log(n^{-1}) - 1\}.$$

The following theorem establishes the asymptotic χ^2 distribution of the weighted empirical log-likelihood ratio statistic $r_w(\mu) = \ell_w(\mu) - \ell_w(\hat{\mu})$.

THEOREM 3. *Suppose* $\{Y_{ij},\ j = 1, \ldots, n_i,\ i = 1, \ldots, k\}$ *are independent random variables and for each of the* $i = 1, \ldots, k$, *the* i*th subgroup* $\{Y_{i1}, \ldots, Y_{in_i}\}$ *follows a common distribution with mean* μ_0. *If* $\mathrm{E}\,(|Y_{ij}|^{2+\delta}) < \infty$ *for some fixed* $\delta > 0$, *then* $-2r_w(\mu_0) = -2\{\ell_w(\mu_0) - \ell_w(\hat{\mu})\}$ *converges in distribution to a* χ^2 *random variable with one degree of freedom as* $n \to \infty$.

A proof of Theorem 3 is given in the Appendix. By Theorem 3, the weighted empirical likelihood ratio confidence interval for μ_0 is constructed as

$$C_{wel} = \big\{\mu \mid -2r_w(\mu) < \chi^2_{[1]}(\alpha)\big\}. \tag{5}$$

This interval has an asymptotically correct coverage probability at $1 - \alpha$ level and compares very favorably to the k-sample interval (4), as we will see from the simulation study reported in Section 6. It is also less restrictive due to the combined sample size.

5. NOTES ON COMPUTATIONAL ALGORITHMS

The major computational task is to maximize $\ell(F_1,\ldots,F_k)$ subject to the set of constraints (2) or (3). For the latter case, the problem reduces to

$$\text{maximizing} \sum_{j=1}^{n_i}\log(p_{ij}) \text{ subject to } \sum_{j=1}^{n_i} p_{ij}=1 \text{ and } \sum_{j=1}^{n_i} p_{ij}Y_{ij}=\mu$$

for $i=1,\ldots,k$, a standard situation in empirical likelihood inference.

To maximize $\ell(F_1,\ldots,F_k)$ subject to (2), a strategy similar to the one used in Wu (2004b, p. 24) can be adapted. Without loss of generality, we use $k=3$ to illustrate the procedure. Let $q_{ij}=p_{ij}/3$ and define the variable $\boldsymbol{u}_{ij}$ as follows:

$$\boldsymbol{u}_{1j}=\begin{pmatrix}1\\1\\Y_{1j}\\Y_{1j}\end{pmatrix},\quad \boldsymbol{u}_{2j}=\begin{pmatrix}-1\\0\\-Y_{2j}\\0\end{pmatrix},\quad \boldsymbol{u}_{3j}=\begin{pmatrix}0\\-1\\0\\-Y_{3j}\end{pmatrix}.$$

It is straightforward to show that maximizing $\ell(F_1,\ldots,F_k)$ subject to (2) for $k=3$ is equivalent to

$$\text{maximizing} \sum_{i=1}^{3}\sum_{j=1}^{n_i}\log(q_{ij}) \text{ subject to } \sum_{i=1}^{3}\sum_{j=1}^{n_i} q_{ij}=1 \text{ and } \sum_{i=1}^{3}\sum_{j=1}^{n_i} q_{ij}\boldsymbol{u}_{ij}=\mathbf{0}.$$

The final maximizer of $\ell(F_1,\ldots,F_k)$ is computed as $\hat{p}_{ij}=3\hat{q}_{ij}$, where $\hat{q}_{ij}=1/\{n(1+\boldsymbol{\lambda}'\boldsymbol{u}_{ij})\}$ and the vector-valued Lagrange multiplier $\boldsymbol{\lambda}$ is the solution to

$$g(\boldsymbol{\lambda})=\frac{1}{n}\sum_{i=1}^{3}\sum_{j=1}^{n_i}\frac{\boldsymbol{u}_{ij}}{1+\boldsymbol{\lambda}'\boldsymbol{u}_{ij}}=\mathbf{0}.$$

The modified Newton–Raphson procedure proposed by Chen, Sitter & Wu (2002) and an R function developed by Wu (2005) can be used to solve $g(\boldsymbol{\lambda})=\mathbf{0}$.

Except for some notational modifications, the algorithm described in Wu (2004a) for handling the weighted empirical likelihood method remains valid in the current context.

The final EL confidence interval (4) or (5) can be found through profiling. Note that for the EL interval (4), the minimum value of $-2r(\mu)$ is achieved at $\mu=\hat{\mu}_{\text{el}}$. The interval (4) is also bounded by (L,R) where $L=\max\{Y_{(11)},\ldots,Y_{(k1)}\}$ and $R=\min\{Y_{(1n_1)},\ldots,Y_{(kn_k)}\}$. The empirical likelihood ratio function $r(\mu)$ is convex for $\mu\in(L,R)$, which implies that $r(\mu)$ is monotone increasing for $\mu\in(L,\hat{\mu}_{\text{el}})$ and monotone decreasing for $\mu\in(\hat{\mu}_{\text{el}},R)$. The upper and lower bound of the interval can be determined by using a bisection search method within $(L,\hat{\mu}_{\text{el}})$ and $(\hat{\mu}_{\text{el}},R)$, respectively. Wu (2005) contains sample R codes for finding such intervals.

6. SIMULATION STUDY

In this section, we present results from a limited simulation study on finite-sample performances of proposed methods. We considered $k=3$ and generated sample data from the model $Y_{ij}=\mu_0+\sigma_i\varepsilon_{ij}$, where the ε_{ij} are independent and identically distributed random variables with zero mean and unit variance. The simulation study was conducted for $\varepsilon_{ij}\sim \mathsf{N}(0,1)$ and for $\varepsilon_{ij}\sim(\chi^2_{[1]}-1)/\sqrt{2}$, respectively. The value of μ_0 was chosen to be 1.

Three scenarios of variance structure were examined: (i) the case of equal variances $(\sigma_1^2,\sigma_2^2,\sigma_3^2)=(1,1,1)$; (ii) the case of moderate nonhomogeneity $(\sigma_1^2,\sigma_2^2,\sigma_3^2)=\left(1/\sqrt{2},1,\sqrt{2}\right)$; and (iii) the case of severe heteroscedasticity $(\sigma_1^2,\sigma_2^2,\sigma_3^2)=(1/2,1,2)$ where $\max(\sigma_i^2)/\min(\sigma_i^2)=4$. Various combinations of the sample sizes (n_1,n_2,n_3) were considered in the simulation, with values of n_i ranging from 10 to 60.

To evaluate the relative performance of the proposed MEL estimator $\hat{\mu}_{\mathrm{el}}$, we compared it to the two commonly used estimators described in Section 2: the parametric maximum likelihood estimator $\hat{\mu}_{\mathrm{ml}}$ and the optimal estimator $\hat{\mu}_{\mathrm{op}}$. The maximum likelihood estimator is based on the normal model, and thus the model is misspecified when the data are generated from the transformed chi-square distribution.

TABLE 1: Simulated mean squared errors $(\times 100)$ for estimating the common mean.

		$\mathsf{N}(0,1)$			$\chi^2_{[1]}$		
$(\sigma_1^2, \sigma_2^2, \sigma_3^2)$	(n_1, n_2, n_3)	$\hat{\mu}_{\mathrm{ml}}$	$\hat{\mu}_{\mathrm{op}}$	$\hat{\mu}_{\mathrm{el}}$	$\hat{\mu}_{\mathrm{ml}}$	$\hat{\mu}_{\mathrm{op}}$	$\hat{\mu}_{\mathrm{el}}$
(1, 1, 1)	(15,15,15)	2.42	2.42	2.44	3.57	3.65	2.72
	(10,15,20)	2.44	2.45	2.86	3.78	3.91	3.14
	(20,15,10)	2.46	2.44	2.79	3.63	3.77	3.04
$(1/\sqrt{2}, 1, \sqrt{2})$	(15,15,15)	2.33	2.33	2.36	3.32	3.48	2.59
	(10,15,20)	2.57	2.55	2.88	3.69	3.87	3.06
	(20,15,10)	2.19	2.19	2.55	3.29	3.50	2.80
$(1/2,\ 1,\ 2)$	(15,15,15)	2.07	2.08	2.10	2.97	3.09	2.25
	(10,15,20)	2.48	2.46	2.67	3.36	3.56	2.71
	(20,15,10)	1.83	1.83	2.15	2.61	3.03	2.35
(1, 1, 1)	(30,30,30)	1.16	1.16	1.17	1.61	1.60	1.20
	(20,30,40)	1.15	1.15	1.27	1.67	1.68	1.41
	(40,30,20)	1.15	1.16	1.30	1.67	1.67	1.35
$(1/\sqrt{2}, 1, \sqrt{2})$	(30,30,30)	1.10	1.10	1.11	1.54	1.53	1.15
	(20,30,40)	1.20	1.19	1.30	1.70	1.72	1.43
	(40,30,20)	1.01	1.02	1.15	1.47	1.50	1.22
$(1/2,\ 1,\ 2)$	(30,30,30)	0.97	0.97	0.98	1.32	1.34	1.02
	(20,30,40)	1.14	1.14	1.22	1.57	1.60	1.32
	(40,30,20)	0.84	0.84	0.94	1.18	1.23	1.01

At each simulation run, sample data were first generated using the chosen model and a particular setting of parameters. The three estimators were then computed using the sample data. The process was repeated independently for $B = 2000$ times. The performances of these estimators are evaluated using the simulated mean squared error (MSE)

$$\mathrm{MSE} = B^{-1} \sum_{b=1}^{B} (\hat{\mu}_b - \mu_0)^2$$

and the simulated relative bias (RB)

$$\mathrm{RB} = B^{-1} \sum_{b=1}^{B} (\hat{\mu}_b - \mu_0)/\mu_0,$$

where $\hat{\mu}_b$ is the estimate given by one of the three estimators for the bth simulated sample. Our simulation was carried out with an R/S-PLUS program and the program is available from the authors upon request.

The simulated mean squared errors reported in Table 1 indicate that the MEL estimator $\hat{\mu}_{\mathrm{el}}$ is slightly less efficient than the maximum likelihood estimator $\hat{\mu}_{\mathrm{ml}}$ under the correctly specified normal model. The gain of efficiency from using $\hat{\mu}_{\mathrm{el}}$ under the chi-square model, however, can be substantial with reductions in mean squared errors of up to 37%. Further, it is interesting to note that for the chi-squared case, the optimal estimator $\hat{\mu}_{\rm op}$ is also outperformed by the MEL estimator. In terms of their relative biases, simulation results (not reported here) suggest that all three estimators are essentially unbiased under the normal model. When the chi-square model is used, however, both $\hat{\mu}_{\rm ml}$ and $\hat{\mu}_{\rm op}$ have nonnegligible biases but the bias of the MEL estimator is considerably smaller. A referee has correctly pointed out that $\hat{\mu}_{\rm op}$ is, strictly speaking, not optimal for finite-sample sizes as the weights in (1) are estimated. Also, the maximum likelihood estimator $\hat{\mu}_{\rm ml}$ may perform poorly in finite-sample situations. The scenarios considered here for comparing the three estimators are clearly not exhaustive. However, given the lack of a more commonly used estimator and the difficulty in identifying a correct parametric model for practical situations, the robust behaviour and relative accuracy of $\hat{\mu}_{\rm el}$ to $\hat{\mu}_{\rm ml}$ or $\hat{\mu}_{\rm op}$ suggest that the MEL estimator is a very attractive and favourable alternative approach.

TABLE 2: Simulated results of 90% confidence intervals for the common mean.

			$\mathsf{N}(0,1)$				$\chi^2_{[1]}$			
$(\sigma_1^2,\sigma_2^2,\sigma_3^2)$	(n_1,n_2,n_3)	CI	L	CP	R	AL	L	CP	R	AL
(1, 1, 1)	(15,15,15)	PLR	6.2	87.2	6.6	0.49	1.3	66.3	32.4	0.40
		ELk	7.5	84.3	8.2	0.53	4.5	70.2	25.3	0.42
		WEL	5.2	89.4	5.4	0.49	4.4	86.1	9.5	0.48
	(30,30,30)	PLR	5.1	89.3	5.6	0.35	1.1	74.7	24.2	0.31
		ELk	6.0	87.8	6.2	0.40	4.6	79.2	16.2	0.35
		WEL	4.5	90.5	5.0	0.35	3.9	89.4	6.7	0.34
	(45,45,45)	PLR	6.0	88.7	5.3	0.28	1.6	79.3	19.1	0.26
		ELk	4.8	89.5	5.7	0.34	5.6	82.9	11.5	0.30
		WEL	5.3	89.9	4.8	0.28	4.4	89.4	6.2	0.28
(1/2, 1, 2)	(15,15,15)	PLR	6.0	87.0	7.0	0.45	1.3	68.7	30.0	0.38
		ELk	8.1	84.3	7.6	0.49	4.5	70.2	25.3	0.39
		WEL	5.2	89.5	5.3	0.53	4.1	86.5	9.4	0.51
	(30,30,30)	PLR	4.6	90.0	5.4	0.32	1.1	75.6	23.3	0.29
		ELk	6.2	87.8	6.0	0.37	5.0	79.1	15.9	0.32
		WEL	5.4	89.3	5.3	0.38	4.1	88.8	7.1	0.37
	(45,45,45)	PLR	5.2	90.1	4.7	0.26	1.3	80.2	18.5	0.24
		ELk	5.0	89.5	5.5	0.31	6.0	82.9	11.1	0.28
		WEL	4.7	90.1	5.2	0.31	4.5	88.9	6.6	0.30

To evaluate the confidence intervals, our simulation study included the parametric likelihood ratio (PLR) interval based on the normality assumption, the naive EL interval based on Theorem 2 (denoted by ELk where the limiting chi-square distribution has k degrees of freedom) and the weighted EL interval (WEL) from (5). Since a general variance formula for $\hat{\mu}_{\rm op}$ is not available, it is not clear how to construct a confidence interval based on $\hat{\mu}_{\rm op}$.

Let $(\hat{L}_b, \widehat{U}_b)$ be a confidence interval for μ_0 computed from the bth simulated sample. The simulated average length (AL) of the interval, the coverage probability (CP), the lower tail error rate (L) and the upper tail error rate (U) are respectively computed as

$$\text{AL} = B^{-1}\sum_{b=1}^{B}(\widehat{U}_b - \hat{L}_b), \quad \text{CP} = B^{-1}\sum_{b=1}^{B} I(\hat{L}_b < \mu_0 < \widehat{U}_b) \times 100,$$

$$L = B^{-1} \sum_{b=1}^{B} I(\mu_0 \le \hat{L}_b) \times 100, \quad U = B^{-1} \sum_{b=1}^{B} I(\mu_0 \ge \widehat{U}_b) \times 100,$$

where I is the usual indicator function. Simulation results for a 90% confidence interval based on $B = 2000$ independent runs are presented in Table 2 for the case of equal variance $(\sigma_1^2, \sigma_2^2, \sigma_3^2) = (1, 1, 1)$ and the case of severe heteroscedasticity $(\sigma_1^2, \sigma_2^2, \sigma_3^2) = (1/2, 1, 2)$. Results for the case of moderate nonhomogeneity, i.e., $(\sigma_1^2, \sigma_2^2, \sigma_3^2) = \big(1/\sqrt{2}\,, 1, \sqrt{2}\,\big)$, are somewhat in between and are not reported. Only cases where $n_1 = n_2 = n_3$ are reported but many other cases with unequal sample sizes have also been examined and the results do not seem to differ dramatically. Our major findings based on all cases are summarized as follows.

1. Under the normal model, all three intervals perform well when the n_i are around 30 or more. When the n_i are around 15, the weighted EL interval (WEL) performs the best in terms of coverage probabilities but is slightly wider than the PLR interval under severe heteroscedasticity.

2. Under the chi-square model, the PLR interval based on normality is totally off target; so is the interval ELk. The weighted interval WEL, on the other hand, has coverage probabilities very close to the nominal value at almost all cases. It also has the most balanced tail errors among all methods considered.

3. The naive interval ELk does not show any robustness against change of underlying distributions, a property often expected from nonparametric approaches. Under the chi-square model the coverage probabilities are too low and the method shouldn't be recommended for practical uses (see Section 8 for more discussion on this method).

Finally, we have also considered a few other models in our simulation study and results (not included here) are consistent with what have been reported above.

7. APPLICATION TO ENVIRONMENTAL PROTECTION AGENCY DATA SET

Yu, Sun & Sinha (2002) presented a data set on evaluating gasoline quality based on what is known as Reid vapor pressure (RVP), collected by the Environmental Protection Agency of the United States. Two types of Reid vapor pressure measurements X and Y are included in the data set. Values of X are obtained by an Agency inspector who visits gas pumps in a city, takes samples of gasoline of a particular brand, and measures the Reid vapor pressure right on the spot; values of Y, on the other hand, are produced by shipping gasoline samples to the laboratory for measurements of presumably higher precision at a high cost. The original data set has a double sampling structure, with a subset of the sample units having measurements on both X and Y. Table 3 contains two independent samples of a new reformulated gasoline, one related to X with sample size 30 and the other, to Y with sample size 15.

TABLE 3: Field and lab data on Reid vapor pressure for newly reformulated gasoline.

X (Field)	8.09	8.46	7.37	8.80	7.59	8.62	7.88	7.98	7.47	8.90
	8.51	8.69	7.93	7.96	7.45	8.02	7.32	7.45	7.86	7.88
	7.39	8.03	7.31	7.44	7.95	7.92	7.53	8.01	7.16	7.31
Y (Lab)	8.28	8.63	9.28	7.85	8.62	9.14	7.86	7.90	8.52	7.92
	7.89	8.48	7.95	8.32	7.60					

One of the assumptions of Yu, Sun & Sinha (2002) is that the field measurement X and the lab measurement Y have common mean μ. The two types of measurements differ, however, in terms of precision. Yu, Sun & Sinha (2002) also assume that (X, Y) is bivariate normal, which is not required under our proposed empirical likelihood approach. We compute three versions of the point estimate and three types of 90% confidence interval for μ, as described in Section 6. The results are given in Table 4.

TABLE 4: Point estimates and 90% confidence intervals for Reid vapor pressure.

$\hat{\mu}_{\rm ml}$	$\hat{\mu}_{\rm op}$	$\hat{\mu}_{\rm el}$	PLR	ELk	WEL
7.989	8.008	8.124	(7.848 , 8.137)	(NA)	(7.888 , 8.145)

The two sample means are $\overline{X} = 7.876$ and $\overline{Y} = 8.283$. The maximum EL estimator $\hat{\mu}_{\rm el}$ relies more on the second sample Y and is larger than the other two. For 90% confidence intervals, it turns out that the naive EL interval (4) based on two degrees of freedom is not available. The related EL ratio statistic $-2r(\mu)$ has minimum value at $\mu = \hat{\mu}_{\rm el}$. One of the constraints for (4) to be computable is that $-2r(\hat{\mu}_{\rm el}) < \chi^2_{[k]}(\alpha)$, which is not the case for this particular example. There are also other restrictions associated with this naive EL interval (see Section 8 for more discussion). On the other hand, the WEL interval is shorter and has a larger lower bound than the parametric likelihood ratio confidence interval based on the normal assumption. We do not intend to draw any general conclusion here, but our theoretical and simulated results suggest that the maximum EL estimator $\hat{\mu}_{\rm el} = 8.124$ and the WEL confidence interval $(7.888\,, 8.145)$ should be used for this particular application.

8. CONCLUDING REMARKS

In the past decade, the empirical likelihood method has attracted increased attention from many statisticians. Applications of the method have been found in many areas of statistics, as evidenced by the wide range of topics included in the recent book by Owen (2001). For the majority of these developments, the main focus is the construction of confidence intervals or significance tests using the empirical likelihood ratio statistics. It is shown in this article that the EL method provides both efficient point estimates and reliable confidence intervals for a common mean when the data are combined from several independent but nonhomogeneous samples. Even when the underlying distribution is correctly specified, the parametric approach does not seem to have any major advantages over the EL method. The proposed maximum empirical likelihood estimator and the weighted empirical likelihood confidence interval, on the other hand, are robust against model misspecifications and have good to satisfactory performance in almost all cases investigated in the current study.

The standard error of the MEL estimator is not easily available because the estimator does not have an explicit analytic expression. At a price of considerably more computing effort, we may estimate the standard error by bootstrap methods. Nevertheless, we note that the lack of an easy means to compute the standard error should have little impact on the usefulness and competitiveness of the MEL estimator. This is so because the standard error is often computed for the sole purpose of constructing confidence intervals and, in our case, there is already a complementary weighted EL confidence interval. Further, the standard errors of its main competitors, the optimal and maximum likelihood estimators, are equally difficult to compute.

Computational issues and the lack of explicit analytic forms may discourage some potential users from using the proposed methods which we believe are superior alternatives to existing methods. We intend to develop a reliable R/S-PLUS program for handling the computation and make it available in a public R/S-PLUS software library. Our experience with the simulation studies shows that the computation can be handled as routine practice once a few crucial codings are available. Some of these codings in the context of survey sampling have already been implemented in R by Wu (2005).

Empirical likelihood ratio confidence intervals such as (4) are often associated with an under-coverage problem when sample sizes are small or even moderate. Among possible causes for this is the fact that the finite-sample distribution of the empirical log-likelihood ratio statistic is actually a mixture distribution with an atom at infinity. As such, this distribution may be poorly approximated by the limiting, continuous, chi-squared distribution (Tsao 2004). In general, the larger the atom, the poorer the chi-square approximation and the more serious the under-coverage problem. In the current setting, the atom of the k-sample EL ratio statistic is the probability that the joint convex hull $\mathcal{H}^*$ does not cover μ_0. Since

$$\mathcal{H}^* = \big(\max\{Y_{(11)}, \ldots, Y_{(k1)}\}, \min\{Y_{(1n_1)}, \ldots, Y_{(kn_k)}\}\big),$$

the bigger the k and the smaller the n_i, the bigger the atom will be and consequently the more serious the under-coverage problem. The joint convex hull $\mathcal{H}^*$ is more restrictive under skewed distributions than that under symmetric ones (Tsao 2004). This can be seen from the results of our simulation studies reported in Section 6 and the unavailability of the interval (4) using the real Environmental Protection Agency data. Naively applying the empirical likelihood approach without knowing the severity of this atom could lead to a very inefficient method. Presently, work is continuing to find ways to alleviate this problem.

The asymptotic framework used in this article keeps k fixed, while letting all the n_i go to infinity. When some of the n_i are very small, cautions must be exercised in using the proposed methods. Some preliminary investigation through simulation might be helpful when one faces such situations. Our results do not cover cases where all n_i are bounded while the number of independent samples k becomes large, a scenario discussed by J. N. K.Rao (1980) and others. As a future research problem, it may be of interest to determine the asymptotic behaviour of the empirical log-likelihood ratio statistic when both k and the n_i are allowed to go to infinity at some fixed relative speed. Finally, it may be possible to extend the empirical likelihood approach discussed here to regression models with nonhomogeneous variances. We are presently studying this possibility. There are some technical difficulties in extending the weighted EL method to such situations. We hope to report our findings on these issues when they become available.

APPENDIX: PROOFS FOR THEOREMS

Proof of Theorem 1. For a given combined sample and under the assumed conditions, the MEL estimator $\hat{\mu}_{\mathrm{el}}$ exists and is unique since the empirical log-likelihood function is convex in a small neighbourhood of μ_0, and $\hat{\mu}_{\mathrm{el}}$ must be located within such a neighbourhood, as shown below.

For the ith sample, let

$$\ell_i(\mu) = \sum_{j=1}^{n_i} \log(p_{ij}),$$

where p_{ij} maximize $\sum_{j=1}^{n_i} \log(p_{ij})$ subject to

$$\sum_{j=1}^{n_i} p_{ij} = 1 \quad \text{and} \quad \sum_{j=1}^{n_i} p_{ij} Y_{ij} = \mu$$

for some fixed μ. We observe that (a) $\ell_i(\mu)$ is maximized at $\hat{\mu}_i = \overline{Y}_{i\cdot}$ and (b) $\ell_i(\mu)$ is monotone increasing when $\mu < \overline{Y}_{i\cdot}$ and is monotone decreasing when $\mu > \overline{Y}_{i\cdot}$.

Let

$$\ell(\mu) = \sum_{i=1}^{k} \sum_{j=1}^{n_i} \log(p_{ij})$$

be the joint empirical log-likelihood function where p_{ij} maximize $\ell(F_1, \ldots, F_k)$ subject to

$$\sum_{j=1}^{n_i} p_{ij} = 1, \;\; i = 1, \ldots, k \quad \text{and} \quad \sum_{j=1}^{n_1} p_{1j} Y_{1j} = \cdots = \sum_{j=1}^{n_k} p_{kj} Y_{kj} = \mu$$

for the fixed μ. It is clear that $\ell(\mu) = \ell_1(\mu) + \cdots + \ell_k(\mu)$ and $\ell(\mu)$ is monotone increasing when $\mu < \min_i(\overline{Y}_{i\cdot})$ and monotone decreasing when $\mu > \max_i(\overline{Y}_{i\cdot})$. It follows that the MEL estimator $\hat{\mu}_{\rm el}$ must satisfy

$$\min_{1\le i\le k}(\overline{Y}_{i\cdot}) \le \hat{\mu}_{\rm el} \le \max_{1\le i\le k}(\overline{Y}_{i\cdot}).$$

Noting that $\overline{Y}_{i\cdot} = \mu_0 + O_p(n^{-1/2})$ for $i = 1, \ldots, k$ when k is fixed, we conclude that $\hat{\mu}_{\rm el} = \mu_0 + O_p(n^{-1/2})$. This completes the proof of (1) and (2). □

By using the standard Lagrange multiplier method it can be shown that the p_{ij} which maximize $\ell(F_1, \ldots, F_k)$ subject to $p_{i1} + \cdots + p_{in_i} = 1$ and $p_{i1}Y_{i1} + \cdots + p_{in_i}Y_{in_i} = \mu$, $i = 1, \ldots, k$, are given by

$$p_{ij} = \frac{1}{n_i\{1 + \lambda_i(Y_{ij} - \mu)\}}, \tag{6}$$

where the Lagrange multiplier λ_i is the solution to

$$\sum_{j=1}^{n_i}(Y_{ij} - \mu)/\{1 + \lambda_i(Y_{ij} - \mu)\} = 0. \tag{7}$$

The resulting profile empirical log-likelihood function (omitting a constant term) is given by

$$\ell(\mu) = -\sum_{i=1}^{k}\sum_{j=1}^{n_i}\log\{1 + \lambda_i(Y_{ij} - \mu)\}.$$

The MEL estimator $\hat{\mu}_{\rm el}$ is the solution to $\partial\ell(\mu)/\partial\mu = 0$. We have

$$\frac{\partial\ell(\mu)}{\partial\mu} = -\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left\{\frac{\partial\lambda_i}{\partial\mu}(Y_{ij} - \mu) - \lambda_i\right\}\{1 + \lambda_i(Y_{ij} - \mu)\}^{-1} = \sum_{i=1}^{k} n_i\lambda_i,$$

where the last step used the identities (6), $p_{i1} + \cdots + p_{in_i} = 1$ and (7). For any fixed μ such that $\mu = \mu_0 + O(n^{-1/2})$, we have $\overline{Y}_{i\cdot} - \mu = O_p(n^{-1/2})$. Again by using standard arguments (Owen, 2001, pp. 219–222), we can show that

$$\lambda_i = (\overline{Y}_{i\cdot} - \mu)\Big/\left\{n_i^{-1}\sum_{j=1}^{n_i}(Y_{ij} - \mu)^2\right\} + o_p(n_i^{-1/2}).$$

It follows that $\hat{\mu}_{\rm el}$ satisfies

$$\sum_{i=1}^{k}\frac{n_i(\overline{Y}_{i\cdot} - \mu)}{n_i^{-1}\sum_{j=1}^{n_i}(Y_{ij} - \mu)^2} = o_p(\sqrt{n}),$$

which leads to

$$\hat{\mu}_{\rm el} = \sum_{i=1}^{k}\left(\frac{n_i}{\tilde{\sigma}_i^2}\right)\overline{Y}_{i\cdot}\Big/\sum_{i=1}^{k}\left(\frac{n_i}{\tilde{\sigma}_i^2}\right) + o_p\left(\frac{1}{\sqrt{n}}\right), \tag{8}$$

where

$$\tilde{\sigma}_i^2 = n_i^{-1}\sum_{j=1}^{n_i}(Y_{ij} - \hat{\mu}_{\rm el})^2.$$

Since $a_i = n_i/n = O(1)$, $\overline{Y}_{i\cdot} - \mu_0 = O_p(n^{-1/2})$ and $\tilde{\sigma}_i^2 = \sigma_i^2 + O_p(n^{-1/2})$, it follows from (8) that

$$\hat{\mu}_{\rm el} - \mu_0 = \sum_{i=1}^{k}\left(\frac{a_i}{\sigma_i^2}\right)(\overline{Y}_{i\cdot} - \mu_0)\Big/\sum_{i=1}^{k}\left(\frac{a_i}{\sigma_i^2}\right) + o_p\left(\frac{1}{\sqrt{n}}\right). \tag{9}$$

The asymptotic optimality of $\hat{\mu}_{\rm el}$ follows from (8) and the asymptotic normality of $\hat{\mu}_{\rm el}$ follows from (9) when we apply the central limit theorem to $\overline{Y}_{i\cdot}$.

Proof of Theorem 2. Following the standard argument of Owen (2001, p. 220), one can show that

$$-2r(\mu_0) = \sum_{i=1}^{k} \{ n_i(\overline{Y}_{i\cdot} - \mu_0)^2 / \widetilde{S}_i^2 \} + o_p(1),$$

where

$$\widetilde{S}_i^2 = n_i^{-1} \sum_{j=1}^{n_i} (Y_{ij} - \mu_0)^2.$$

Since the k samples are independent, the conclusion of the theorem follows from the fact that $n_i(\overline{Y}_{i\cdot} - \mu_0)^2/\widetilde{S}_i^2$ converges in distribution to a χ^2 random variable with one degree of freedom as $n_i \to \infty$.

Proof of Theorem 3. Note that $Y_{i1}, \ldots, Y_{in_i}$ have a common distribution with mean μ_0. Let $\sigma_i^2 = \mathrm{var}(Y_{ij})$ and $\tau_i = \mathrm{E}\,|Y_{ij} - \mu_0|^{2+\delta}$. Since

$$\sum_{i=1}^{k} \sum_{j=1}^{n_i} \mathrm{E}\,|Y_{ij} - \mu_0|^{2+\delta} = \sum_{i=1}^{k} n_i \tau_i = o(B_n^{2+\delta}),$$

where

$$B_n^2 = \sum_{i=1}^{k} \sum_{j=1}^{n_i} \mathrm{var}(Y_{ij}) = \sum_{i=1}^{k} n_i \sigma_i^2,$$

the Liapunov condition holds for the sequence of random variables Y_{ij}. The asymptotic χ^2 distribution of $-2r_w(\mu_0)$ follows directly from the proof of Theorem 1 in Wu (2004a) if we replace s_i^2 by σ_i^2 in the formulation of $r_w(\mu_0)$. To complete the proof, we note that $n_i/n = O(1)$ and $s_i^2 = \sigma_i^2 + O_p(n^{-1/2})$, and so it is straightforward to show by following the lines of the proof in Wu (2004a) that replacing σ_i^2 by s_i^2 does not change the leading term in the expansion of $-2r_w(\mu_0)$ and therefore results in the same asymptotic distribution for $-2r_w(\mu_0)$. $\square$

ACKNOWLEDGEMENTS

The authors would like to thank the Editor, an Associate Editor and a referee for valuable comments which led to many improvements. This research was supported by grants from the Natural Sciences and Engineering Research Council of Canada.

REFERENCES

T. R. Bement & J. S. Williams (1969). Variance of weighted regression estimators when sampling errors are independent and heteroscedastic. *Journal of the American Statistical Association*, 64, 1369–1382.

J. Chen, R. R. Sitter & C. Wu (2002). Using empirical likelihood methods to obtain range restricted weights in regression estimators for surveys. *Biometrika*, 89, 230–237.

W. G. Cochran & S. P. Carroll (1953). A sampling investigation of the efficiency of weighting inversely as the estimated variance. *Biometrics*, 9, 447–459.

H. O. Hartley & K. S. E. Jayatillake (1973). Estimation for linear models with unequal variances. *Journal of the American Statistical Association*, 68, 189–192.

P. S. Levy (1970). Combining independent estimators-an empirical study. *Technometrics*, 12, 162–165.

P. Meier (1953). Variance of a weighted mean. *Biometrics*, 9, 59–73.

J. Neyman & E. Scott (1948). Consistent estimators based on partially consistent observations. *Econometrica*, 16, 1–32.

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

A. B. Owen (2001). *Empirical Likelihood*. Chapman & Hall, London.

C. R. Rao (1970). Estimation of heteroscedastic variances in linear models. *Journal of the American Statistical Association*, 65, 161–172.

J. N. K. Rao (1973). On the estimation of heteroscedastic variances. *Biometrics*, 29, 11–24.

J. N. K. Rao (1980). Estimating the common mean of possibly different normal populations: a simulation study. *Journal of the American Statistical Association*, 75, 447–453.

M. Tsao (2004). Bounds on coverage probabilities of the empirical likelihood ratio confidence regions. *The Annals of Statistics*, 32, 1215–1221.

C. Wu (2004a). Weighted empirical likelihood inference. *Statistics & Probability Letters*, 66, 67–79.

C. Wu (2004b). Combining information from multiple surveys through the empirical likelihood method. *The Canadian Journal of Statistics*, 32, 15–26.

C. Wu (2005). Algorithms and R codes for the pseudo empirical likelihood method in survey sampling. *Survey Methodology*, 31, in press.

P. L. H. Yu, Y. Sun & B. Sinha (2002). Estimation of the common mean of a bivariate normal population. *Annals of the Institute of Statistical Mathematics*, 54, 861–878.

B. Zhong, J. Chen & J. N. K. Rao (2000). Empirical likelihood inference in the presence of measurement error. *The Canadian Journal of Statistics*, 28, 841–852.

Received 14 July 2004
Accepted 13 July 2005

Min TSAO: tsao@math.uvic.ca
Department of Mathematics and Statistics
University of Victoria, Victoria, BC, Canada V8W 3P4

Changbao WU: cbwu@uwaterloo.ca
Department of Statistics and Actuarial Science
University of Waterloo, Waterloo, ON, Canada N2L 3G1

SSC Statistical Society of Canada
Société Statistique du Canada