Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.

This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

http://www.elsevier.com/copyright

Author's personal copy

Journal of Statistical Planning and Inference 139 (2009) 1462 – 1473

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Weighted empirical likelihood inference for multiple samples

Yuejiao Fu[a], Xiaogang Wang[a], Changbao Wu[b,*]

[a] *Department of Mathematics and Statistics, York University, Toronto, ON, Canada M3J 1P3*
[b] *Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, ON, Canada N2L 3G1*

ARTICLE INFO

Article history:
Received 29 October 2006
Received in revised form
4 October 2007
Accepted 22 July 2008
Available online 20 August 2008

Keywords:
Common mean models
Heteroscedasticity
Multiple samples
Related samples
Stratified sampling
Weighted likelihood

ABSTRACT

We propose a weighted empirical likelihood approach to inference with multiple samples, including stratified sampling, the estimation of a common mean using several independent and non-homogeneous samples and inference on a particular population using other related samples. The weighting scheme and the basic result are motivated and established under stratified sampling. We show that the proposed method can ideally be applied to the common mean problem and problems with related samples. The proposed weighted approach not only provides a unified framework for inference with multiple samples, including two-sample problems, but also facilitates asymptotic derivations and computational methods. A bootstrap procedure is also proposed in conjunction with the weighted approach to provide better coverage probabilities for the weighted empirical likelihood ratio confidence intervals. Simulation studies show that the weighted empirical likelihood confidence intervals perform better than existing ones.

© 2008 Elsevier B.V. All rights reserved.

1. Introduction

The empirical likelihood (EL) method was first introduced by Owen (1988) to construct confidence intervals for a population mean when there is one single sample of independent and identically distributed (iid) observations. This non-parametric and likelihood-based approach has since become one of the most popular statistical methods in the past 15 years. Owen's 2001 monograph provides an excellent account of the EL approach, including extensions to various non-iid scenarios such as regression models, biased and incomplete samples, and dependent data.

EL methods for multiple sample problems have not yet been fully explored. Chen and Sitter (1999) used a pseudo-EL approach for stratified sampling with unequal selection probabilities. Zhong and Rao (2000) used the EL method for stratified simple random sampling when the sampling fraction within each stratum is negligible. Tsao and Wu (2006) recently applied the EL method to the common mean problem in the presence of heteroscedasticity, using several independent samples. A naive EL approach which combines EL functions through the product of all EL components of involved samples could result in an EL confidence interval with extremely poor coverage properties for small samples. In some cases, the confidence interval could even be an empty set.

The use of the EL approach for multiple sample problems faces three major challenges, among others commonly encountered in the EL inference. First, the involved asymptotic development requires special technical treatment due to the special type of constraints imposed by the multiple samples; second, which is also closely related to the first one, the computational procedures do not follow directly from those developed for single sample problems; and third, the under-coverage problem of the EL confidence intervals with small samples is often more pronounced under multiple sample scenarios.

* Corresponding author. Tel.: +1 519 888 4567x35537.
E-mail address: cbwu@uwaterloo.ca (C. Wu).

0378-3758/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2008.07.015

Author's personal copy

One of the most crucial technical requirements for any EL-based approach is to show that the involved Lagrange multiplier is of the order of $O_p(n^{-1/2})$, where n is the sample size, so that the Taylor series expansion can be applied to the empirical log-likelihood function. The asymptotic distribution of the empirical log-likelihood ratio statistic is then determined by the asymptotic distribution of the leading term in the expansion. With multiple samples, the issue of equal or unequal sample sizes has to be dealt with along the often irregular type of constraints, which imposes difficulties for handling the involved Lagrange multipliers and the related asymptotic development.

The under-coverage problem of the EL confidence intervals under small samples is closely related to the non-existence of an EL solution for certain finite samples. For one single sample, the EL confidence interval for the population mean is confined by the convex hull of the sample data. The smaller the sample size the more restrictive the convex hull. With multiple samples, taking the common mean problem (Tsao and Wu, 2006) as an example, the EL confidence interval is bounded by the intersection of the convex hulls formed by each of the samples. Even if the asymptotic χ^2 distribution of the EL ratio function can be established for large samples, using such a distribution as an approximation when the sample sizes are small will unduly produce very undesirable results.

In this article, we propose a weighted EL (WEL) approach to inference involving multiple samples. Our method provides a unified framework for stratified sampling, the estimation of a common mean using several independent and non-homogeneous samples and inference on a particular population using other related samples. These three types of problems are quite different but they all involve multiple samples. Stratified sampling is one of the most frequently used procedures in sample surveys. Some practical examples on the common mean problem can be found in Tsao and Wu (2006) and references therein, and the use of related samples has been discussed, for instance, in Hu and Zidek (2002) and Wang (2006). Our method is also applicable to a variety of two-sample problems. Asymptotic derivations and computational procedures are effectively handled under the proposed weighting scheme. The under-coverage problem associated with the usual unweighted EL confidence intervals based on a χ^2 approximation can often be alleviated through a bootstrap calibration method (Owen, 2001). For the WEL approach, we develop bootstrap procedures for all three types of multiple sample problems considered in the paper. Finite sample performances of these methods are investigated through simulation studies. Bartlett correction to the EL intervals is potentially another alternative method, however, it involves non-trivial asymptotic development and has to be done one-at-a-time for different scenarios. The bootstrap method, on the other hand, can be applied in a unified manner and has major operational advantages as well.

The rest of the article is organized as follows. In Section 2, we motivate and present the proposed WEL function and establish a basic asymptotic result under stratified sampling. In Sections 3 and 4, we show that the proposed approach can ideally be applied to the other two types of multiple sample problems, namely, the estimation of a common mean using several independent and non-homogeneous samples, and inference on one particular population mean in the presence of related samples. Computational notes are included for each of the cases. In Section 5, results from a limited simulation study, which focuses on the comparison of the proposed method to existing ones, are reported. Some additional remarks and a discussion on applications to two-sample problems are provided in Section 6.

2. Inference for a population mean under stratified sampling

Let $\{Y_{ij},\ j=1,\ldots,n_i\}$, $i=1,\ldots,k$, be k independent samples such that Y_{ij}, $j=1,\ldots,n_i$ are iid observations with a common distribution function F_i. Let $E(Y_{ij})=\theta_i$ and $Var(Y_{ij})=\sigma_i^2$. Our inference is focused on an overall population mean $\theta_0=\sum_{i=1}^k w_i\theta_i$ for a fixed set of weights w_i satisfying $w_i\geqslant 0$ and $\sum_{i=1}^k w_i=1$. This setting arises from a typical stratified sampling design, where the overall population is stratified into k strata with θ_i being the stratum mean and w_i being the stratum weight. Samples are drawn independently from each of the k strata. We will ignore the finite population structure often attached to stratified sampling and assume the stratum samples themselves are iid.

A conventional formulation of the EL inference for stratified sampling is to use the k-sample approach to ANOVA as outlined in Owen (2001, p. 88) with the following empirical log-likelihood function:

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

where $p_{ij}=F_i(y_{ij})-F_i(y_{ij}-)$, $F_i(y_{ij})=P(Y_{ij}\leqslant y_{ij})$ and $F_i(y_{ij}-)=P(Y_{ij}<y_{ij})$. Unfortunately, for the inference on the overall population mean $\theta_0=\sum_{i=1}^k w_i\theta_i$, this formulation is typically asymptotically intractable as shown in subsequent discussions and explained in particular by the detailed remarks following the proof of Theorem 1.

We propose to use a weighted empirical (log) likelihood (WEL) function given by

$$l_w(F_1,\ldots,F_k)=\sum_{i=1}^k\frac{w_i}{n_i}\sum_{j=1}^{n_i}\log(p_{ij}). \tag{1}$$

The formulation of the WEL function (1) can be motivated using the argument of Chen and Sitter (1999). Suppose we have a stratified finite population, with N_i units in stratum i, $i=1,\ldots,k$. The stratum weight is $w_i=N_i/N$, where $N=N_1+\cdots+N_k$ is the

Author's personal copy

total population size. If one knows the entire finite population, the total empirical log-likelihood at the population level would be

$$l_N = \sum_{i=1}^{k}\sum_{j=1}^{N_i}\log(p_{ij}).$$

Under stratified random sampling, a sample-based estimate for the total likelihood l_N is given by

$$\hat{l}_N = N\sum_{i=1}^{k}\frac{w_i}{n_i}\sum_{j=1}^{n_i}\log(p_{ij}),$$

which differs from $l_w(F_1,\ldots,F_k)$ only by a multiplying constant, N. The WEL function represents the total likelihood at the population level. In view of the parameter $\theta_0=\sum_{i=1}^{k} w_i\theta_i$, the weight w_i reflects the contribution of the ith stratum to the overall population mean and n_i adjusts for the discrepancy among stratum sample sizes. Under proportional sample size allocations (i.e. $n_i \propto w_i$), the WEL function $l_w(F_1,\ldots,F_k)$ reduces to the unweighted one $l(F_1,\ldots,F_k)$.

The most important advantage of using this WEL formulation, however, is that the usual large sample properties of the EL approach can be rigorously established under the often irregular type of constraints induced by stratified or multiple samples. Computational procedures are also readily available under suitable reformulation of the constraints.

The maximum weighted empirical likelihood (MWEL) estimator of θ_0, denoted as $\hat{\theta}_w$, is defined as the maximizer of $l_w(\theta)=\sum_{i=1}^{k}(w_i/n_i)\sum_{j=1}^{n_i}\log\{\hat{p}_{ij}(\theta)\}$, where the $\hat{p}_{ij}(\theta)$ maximize the WEL function $l_w(F_1,\ldots,F_k)$ subject to $p_{ij}>0$ and

$$\sum_{j=1}^{n_i} p_{ij} = 1, \quad i=1,\ldots,k, \tag{2}$$

$$\sum_{i=1}^{k} w_i \sum_{j=1}^{n_i} p_{ij}Y_{ij} = \theta \tag{3}$$

for some fixed θ. It is apparent that without further restrictions the MWEL estimator of θ_0 is given by $\hat{\theta}_w=\sum_{i=1}^{k} w_i\bar{Y}_{i\cdot}$, the stratified sample mean, where $\bar{Y}_{i\cdot}=n_i^{-1}\sum_{j=1}^{n_i} Y_{ij}$.

We now turn to the asymptotic distribution of the WEL ratio statistic for constructing confidence intervals for θ_0. Let θ be fixed and used in the constraint (3). Let $r_w(\theta)=\sum_{i=1}^{k}(w_i/n_i)\sum_{j=1}^{n_i}\log\{n_i\hat{p}_{ij}(\theta)\}$ be the WEL ratio function. We assume that $n_i/n \to f_i \neq 0$, where $n=n_1+\cdots+n_k$ is the total sample size. With this assumption it is not necessary to distinguish between $O(n^{-1/2})$ and $O(n_i^{-1/2})$ and between $o(n^{-1/2})$ and $o(n_i^{-1/2})$. The following theorem establishes the asymptotic distribution of $r_w(\theta)$ at $\theta=\theta_0$.

Theorem 1. *Suppose $\{Y_{ij},\ j=1,\ldots,n_i\}$ is an iid sample from F_i with mean $\theta_i=E(Y_{ij})$ and finite variance $\sigma_i^2=Var(Y_{ij}),\ i=1,\ldots,k$, and the k samples are also independent of each other. Then $-2r_w(\theta_0)/c_1$ converges in distribution to a χ^2 random variable with one degree of freedom, where the scaling constant c_1 is given by* (12).

Proof. To ease presentation and without loss of generality, we consider $k=3$. Constraints (2) and (3) can be reformulated as

$$\sum_{i=1}^{3} w_i \sum_{j=1}^{n_i} p_{ij} = 1, \tag{4}$$

$$\sum_{i=1}^{3} w_i \sum_{j=1}^{n_i} p_{ij}\boldsymbol{Z}_{ij} = \boldsymbol{\eta}, \tag{5}$$

where the vector-valued variables $\boldsymbol{Z}_{ij}$ and $\boldsymbol{\eta}$ are given by

$$\boldsymbol{Z}_{1j}=\begin{bmatrix}1\\0\\Y_{1j}\end{bmatrix},\quad \boldsymbol{Z}_{2j}=\begin{bmatrix}0\\1\\Y_{2j}\end{bmatrix},\quad \boldsymbol{Z}_{3j}=\begin{bmatrix}0\\0\\Y_{3j}\end{bmatrix}\quad\text{and}\quad \boldsymbol{\eta}=\begin{bmatrix}w_1\\w_2\\\theta\end{bmatrix}.$$

We can also rewrite (5) as

$$\sum_{i=1}^{3} w_i \sum_{j=1}^{n_i} p_{ij}\boldsymbol{u}_{ij} = \boldsymbol{0}, \tag{6}$$

Author's personal copy

with $\boldsymbol{u}_{ij} = \boldsymbol{Z}_{ij} - \boldsymbol{\eta}$. Using the standard Lagrange multiplier method, we can show that the $\hat{p}_{ij}(\theta)$ which maximize $l_w(F_1, F_2, F_3)$ subject to (4) and (6) for a fixed θ are given by

$$\hat{p}_{ij}(\theta) = \frac{1}{n_i(1 + \boldsymbol{\lambda}'\boldsymbol{u}_{ij})}, \tag{7}$$

with the vector-valued Lagrange multiplier $\boldsymbol{\lambda}$ being the solution to

$$g(\boldsymbol{\lambda}) = \sum_{i=1}^{3} \frac{w_i}{n_i} \sum_{j=1}^{n_i} \frac{\boldsymbol{u}_{ij}}{1 + \boldsymbol{\lambda}'\boldsymbol{u}_{ij}} = \boldsymbol{0}. \tag{8}$$

The algorithm described in Wu (2004) can be used to solve (8). If we rewrite the numerator $\boldsymbol{u}_{ij}$ in (8) as $\boldsymbol{u}_{ij}\{(1 + \boldsymbol{\lambda}'\boldsymbol{u}_{ij}) - \boldsymbol{u}_{ij}'\boldsymbol{\lambda}\}$, we can re-express (8) as

$$\left\{ \sum_{i=1}^{3} \frac{w_i}{n_i} \sum_{j=1}^{n_i} \frac{\boldsymbol{u}_{ij}\boldsymbol{u}_{ij}'}{1 + \boldsymbol{\lambda}'\boldsymbol{u}_{ij}} \right\} \boldsymbol{\lambda} = \sum_{i=1}^{3} \frac{w_i}{n_i} \sum_{j=1}^{n_i} \boldsymbol{u}_{ij}. \tag{9}$$

Note that $\sum_{j=1}^{n_i} \{1/[n_i(1 + \boldsymbol{\lambda}'\boldsymbol{u}_{ij})]\} = 1$ for $i = 1, 2, 3$, the order of $\boldsymbol{\lambda}$ is related to the order of the right side of (9), which can be shown to be

$$\boldsymbol{U} = \sum_{i=1}^{3} \frac{w_i}{n_i} \sum_{j=1}^{n_i} \boldsymbol{u}_{ij} = \left(0, 0, \sum_{i=1}^{3} w_i \bar{Y}_{i\cdot} - \theta \right)'. \tag{10}$$

It follows that $\boldsymbol{U} = O_p(n^{-1/2})$ (component-wise) when $\theta = \theta_0$.

Note that $\boldsymbol{D} = \sum_{i=1}^{3} (w_i/n_i) \sum_{j=1}^{n_i} \boldsymbol{u}_{ij}\boldsymbol{u}_{ij}' = O_p(1)$, it follows from (9) that we must have $\boldsymbol{\lambda} = O_p(n^{-1/2})$. Under the finite variance assumption, we also have $\max_{ij} |\boldsymbol{u}_{ij}| = o_p(n^{1/2})$ and $\boldsymbol{\lambda}'\boldsymbol{u}_{ij} = o_p(1)$ uniformly over all i and j (Owen, 2001, Lemma 11.2). We therefore obtain the following asymptotic expression for $\boldsymbol{\lambda}$

$$\boldsymbol{\lambda} = \boldsymbol{D}^{-1}\boldsymbol{U} + o_p(n^{-1/2}). \tag{11}$$

The WEL ratio function at θ_0 is given by

$$r_w(\theta_0) = -\sum_{i=1}^{3} \frac{w_i}{n_i} \sum_{j=1}^{n_i} \log(1 + \boldsymbol{\lambda}'\boldsymbol{u}_{ij}),$$

where both $\boldsymbol{\lambda}$ and $\boldsymbol{u}_{ij}$ depend on θ_0. Using the Taylor series expansion $\log(1 + x) = x - x^2/2 + o(x^2)$ at $x = \boldsymbol{\lambda}'\boldsymbol{u}_{ij}$, which is of the order $o_p(1)$ (uniform over all i and j) when $\theta = \theta_0$, we obtain

$$\begin{aligned} -2r_w(\theta_0) &= 2\sum_{i=1}^{3} \frac{w_i}{n_i} \sum_{j=1}^{n_i} \log(1 + \boldsymbol{\lambda}'\boldsymbol{u}_{ij}) \\ &= 2\sum_{i=1}^{3} \frac{w_i}{n_i} \sum_{j=1}^{n_i} \left(\boldsymbol{\lambda}'\boldsymbol{u}_{ij} - \frac{1}{2}\boldsymbol{\lambda}'\boldsymbol{u}_{ij}\boldsymbol{u}_{ij}'\boldsymbol{\lambda} \right) + o_p(n^{-1}) \\ &= \boldsymbol{U}'\boldsymbol{D}^{-1}\boldsymbol{U} + o_p(n^{-1}) \\ &= d^{(33)} \left(\sum_{i=1}^{3} w_i \bar{Y}_{i\cdot} - \theta_0 \right)^2 + o_p(n^{-1}), \end{aligned}$$

where the last step is due to (10) and $d^{(33)}$ is the last (the third for $k = 3$) diagonal element of $\boldsymbol{D}^{-1}$. Let

$$c_1 = d^{(33)} \sum_{i=1}^{3} w_i^2 \sigma_i^2 / n_i. \tag{12}$$

It immediately follows that $-2r_w(\theta_0)/c_1$ converges in distribution to a χ^2 random variable with one degree of freedom. □

One of the key asymptotic arguments in the EL-based inference is to show that the involved Lagrange multiplier $\boldsymbol{\lambda}$ is of the order $O_p(n^{-1/2})$, so that Taylor series expansions can be applied to $\log(1 + \boldsymbol{\lambda}'\boldsymbol{u}_{ij})$ when $\max_{ij} |\boldsymbol{u}_{ij}| = o_p(n^{1/2})$ and consequently

Author's personal copy

$\max_{ij}|\boldsymbol{\lambda}'\boldsymbol{u}_{ij}| = o_p(1)$. This is the case under our proposed formulation of the WEL function $l_w(F_1, \ldots, F_k)$, as shown in the proof of Theorem 1, but not so under the naive unweighted formulation $l(F_1, \ldots, F_k)$. It can be shown that for $k = 3$ and using $l(F_1, \ldots, F_k)$, the equation for determining the related Lagrange multiplier, similar to (9) under the weighted approach, involves

$$\boldsymbol{U} = \frac{1}{n}\sum_{i=1}^{3}\sum_{j=1}^{n_i}\boldsymbol{u}_{ij} = \left(\frac{n_1}{n} - w_1, \frac{n_2}{n} - w_2, \frac{n_1}{n}\bar{Y}_{1\cdot} + \frac{n_2}{n}\bar{Y}_{2\cdot} + \frac{n_3}{n}\bar{Y}_{3\cdot} - \theta\right)',$$

which is not of the order $O_p(n^{-1/2})$ unless $n_i/n = w_i + O(n^{-1/2})$ for all i. This essentially requires proportional sample size allocations.

The scaling constant c_1 involves the unknown parameters θ_0 and σ_i^2. It is easy to see (for $k = 3$ here) that replacing θ_0 by $\sum_{i=1}^{3} w_i\bar{Y}_{i\cdot}$ and σ_i^2 by $S_i^2 = (n_i - 1)^{-1}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_{i\cdot})^2$ in calculating c_1 will not change the asymptotic distribution of $-2r_w(\theta_0)/c_1$. When $k = 2$, the scaling constant is given by $c_1 = (w_1^2S_1^2/n_1 + w_2^2S_2^2/n_2)/(w_1S_1^2 + w_2S_2^2)$.

A $(1-\alpha)$-level WEL ratio confidence interval on θ_0 can therefore be constructed as $\{\theta \mid -2r_w(\theta)/c_1 < \chi_1^2(\alpha)\}$,where $\chi_1^2(\alpha)$ is the upper α-quantile from the χ^2 distribution with one degree of freedom. It can be seen that $r_w(\theta)$ is computable for any θ such that $\min_{ij}\{Y_{ij}\} < \theta < \max_{ij}\{Y_{ij}\}$, i.e. θ is an inner point of the convex hull formed by the combined sample data. This is not very restrictive as long as the total sample size n is not too small.

The quantile from the limiting χ^2 distribution may be replaced by the quantile from a bootstrap distribution. One advantage of the bootstrap procedure, to be described below, is that the scaling constant c_1 can be bypassed. It also improves the coverage probability of the WEL interval when the total sample size is not large.

Let $\{Y_{ij}^*,\ j = 1, \ldots, n_i\}$ be a bootstrap sample, randomly selected from the original sample $\{Y_{ij},\ j = 1, \ldots, n_i\}$ with replacement, $i = 1, \ldots, k$; let $r_w^*(\hat{\theta}_w)$ be computed in the same way as how $r_w(\theta)$ is computed but replacing the original sample by the bootstrap sample and use $\hat{\theta}_w$ for θ in the constraint (3). Let b_α^* be the upperα-quantile of the bootstrap distribution of $-2r_w^*(\hat{\theta}_w)$ obtained through the usual Monte Carlo approximation. The $(1-\alpha)$-level bootstrap calibrated confidence interval on θ_0 is then constructed as $\{\theta \mid -2r_w(\theta) < b_\alpha^*\}$. It can easily be argued that the interval has correct asymptotic coverage probability at the $(1-\alpha)$-level and should perform better than the χ^2 calibrated interval for samples of small or moderate sizes.

3. Inference for a common mean with multiple samples

Consider k independent samples $\{Y_{ij},\ j=1,\ldots,n_i\},\ i=1,\ldots,k$ with a common mean $\theta_0 = E(Y_{ij})$ but different variances $Var(Y_{ij}) = \sigma_i^2$. Such a scenario can arise from a variety of practical situations and inference about the common mean, θ_0, using information from all k samples have been addressed by several authors. Tsao and Wu (2006) recently explored the use of the EL method for both point estimation and confidence intervals on the common mean. Under their naive EL approach, the EL ratio function is given by $r(\theta) = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\log(n_i\hat{p}_{ij}(\theta))$, where the $\hat{p}_{ij}(\theta)$ maximize the unweighted EL function $l(F_1, \ldots, F_k) = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\log(p_{ij})$ subject to $\sum_{j=1}^{n_i} p_{ij} = 1$ and $\sum_{j=1}^{n_i} p_{ij}Y_{ij} = \theta,\ i = 1, \ldots, k$, for a fixed θ. Under some mild moment conditions and the common mean model, Tsao and Wu (2006) showed that $-2r(\theta_0)$converges in distribution to a χ^2 random variable with k degrees of freedom. Consequently, a $(1-\alpha)$-level confidence interval on θ_0 may be constructed in the form of $\mathscr{C} = \{\theta \mid -2r(\theta) < \chi_k^2(\alpha)\}$, where $\chi_k^2(\alpha)$ is the upper α-quantile from a χ^2 distribution with k degrees of freedom.

One of the major drawbacks of the naive EL approach for the common mean is that the EL ratio conference intervals have severe under-coverage problems and may even not be constructible. There are two major causes behind this. First, the interval $\mathscr{C}$ is confined by the intersection of the k convex hulls formed respectively by the k samples, which is very restrictive and particularly so when any one of the sample sizes is small. The asymptotic χ^2 distribution provides a poor approximation to the actual finite sample distribution of the EL ratio statistic. Second, there is another unusual restriction arising from this particular application: for any given samples the minimum value of the profile function $-2r(\theta)$ is not zero unless $\bar{Y}_{1\cdot} = \cdots = \bar{Y}_{k\cdot}$. Tsao and Wu (2006) presented a real example involving two samples where $-2r(\theta) > \chi_2^2(0.05)$ for all θ, and consequently the desired 95% level EL confidence interval $\{\theta \mid -2r(\theta) < \chi_2^2(0.05)\}$ is an empty set!

The critical model assumption, i.e. the k samples have a common mean, is only used in the derivation of the asymptotic distribution ofthe EL ratio statistic. It is not explicitly used in forming a suitable constraint such as (13) below, which will be used in the current article. A detailed examination of the asymptotic derivations, not presented here to save space, reveals that including (13) as part of the constraints under the naive EL approach makes the derivation intractable.

We now present a WEL approach to constructing confidence intervals on the common mean θ_0 with improved coverage properties. We use the WEL formulation of Section 2 and the resulting WEL ratio function eliminates the non-zero minimum value problem; we further improve the interval by replacing the poor asymptotic χ^2 approximation by a conditional bootstrap calibration method, which provides consistent results under large samples but works dramatically better for samples of small or moderate sizes.

Author's personal copy

With k independent samples $\{Y_{ij},\ j=1,\ldots,n_i\},\ i=1,\ldots,k$, the WEL function l_w is given by (1) with equal "weights" $w_1=\cdots=w_k=1/k$ (see further comments in the proof of Theorem 2 on the choice of equal weights). Let $\hat{p}_{ij}$ be the maximizer of l_w under the normalization constraints (2) and

$$\sum_{j=1}^{n_1} p_{1j}Y_{1j}=\cdots=\sum_{j=1}^{n_k} p_{kj}Y_{kj}, \tag{13}$$

which is the constraint induced by the common mean structure. Let $\tilde{p}_{ij}(\theta)$ be the maximizer of l_w subject to (2), (13) and an additional constraint induced by the parameter of interest, i.e. the common mean:

$$\sum_{j=1}^{n_i} p_{ij}Y_{ij}=\theta,\quad i=1,\ldots,k. \tag{14}$$

Note that (14) only adds one constraint on top of (13). The WEL ratio statistic for θ is given by

$$r_w(\theta)=-2\{l_w(\tilde{\boldsymbol{p}}(\theta))-l_w(\hat{\boldsymbol{p}})\},$$

where $l_w(\hat{\boldsymbol{p}})=\sum_{i=1}^{k}(w_i/n_i)\sum_{j=1}^{n_i}\log(\hat{p}_{ij})$ and $l_w(\tilde{\boldsymbol{p}}(\theta))$ is similarly defined using $\tilde{p}_{ij}(\theta)$.

We have the following asymptotic result regarding $r_w(\theta)$.

Theorem 2. *Suppose $\{Y_{ij},\ j=1,\ldots,n_i\}$ is an iid sample from F_i with common mean $\theta_0=E(Y_{ij})$ and finite variance $\sigma_i^2=Var(Y_{ij}),\ i=1,\ldots,k$, and the k samples are also independent of each other. Then $-2r_w(\theta_0)/c_2$ converges in distribution to a χ^2 random variable with one degree of freedom, where the scaling constant c_2 is given by* (18).

Proof. Once again, we consider $k=3$. To derive an asymptotic expansion for $l_w(\hat{\boldsymbol{p}})$, we used the same technique from the proof of Theorem 1 with reformulated variables

$$\boldsymbol{Z}_{1j}=\begin{bmatrix}1\\0\\Y_{1j}\\Y_{1j}\end{bmatrix},\quad \boldsymbol{Z}_{2j}=\begin{bmatrix}0\\1\\-Y_{2j}\\0\end{bmatrix},\quad \boldsymbol{Z}_{3j}=\begin{bmatrix}0\\0\\0\\-Y_{3j}\end{bmatrix}\quad\text{and}\quad \boldsymbol{\eta}=\begin{bmatrix}w_1\\w_2\\0\\0\end{bmatrix}.$$

It is straightforward to show that the constraints (2) and (13) can equivalently be written as (4) and (6) with $\boldsymbol{u}_{ij}=\boldsymbol{Z}_{ij}-\boldsymbol{\eta}$. A critical point to observe is that

$$\boldsymbol{U}=\sum_{i=1}^{3}\frac{w_i}{n_i}\sum_{j=1}^{n_i}\boldsymbol{u}_{ij}=(0,0,w_1\bar{Y}_{1\cdot}-w_2\bar{Y}_{2\cdot},w_1\bar{Y}_{1\cdot}-w_3\bar{Y}_{3\cdot})', \tag{15}$$

which is of the order $O_p(n^{-1/2})$ under the choice of equal weights $w_i=1/k$ and the common mean structure. For instance, $w_1\bar{Y}_{1\cdot}-w_2\bar{Y}_{2\cdot}=\{(\bar{Y}_{1\cdot}-\theta_0)-(\bar{Y}_{2\cdot}-\theta_0)\}/k=O_p(n^{-1/2})$. It follows from the steps in the proof of Theorem 1 that

$$-2l_w(\hat{\boldsymbol{p}})=2\sum_{i=1}^{3}w_i\log(n_i)+\boldsymbol{U}'\boldsymbol{D}^{-1}\boldsymbol{U}+o_p(n^{-1}), \tag{16}$$

where $\boldsymbol{U}$ is given by (15) and $\boldsymbol{D}=\sum_{i=1}^{3}(w_i/n_i)\sum_{j=1}^{n_i}\boldsymbol{u}_{ij}\boldsymbol{u}_{ij}'$. To derive a similar expansion for $l_w(\tilde{\boldsymbol{p}}(\theta))$, we reformulate the additional constraint (14) as

$$\sum_{i=1}^{3}w_i\sum_{j=1}^{n_i}p_{ij}(Y_{ij}-\theta)=0. \tag{17}$$

The two equations (14) and (17) are equivalent if (4) and (13) also hold.

Let $B=[\sum_{i=1}^{3}w_iE(\boldsymbol{u}_{ij}\boldsymbol{u}_{ij}')]^{-1}\sum_{i=1}^{3}w_iE[(Y_{ij}-\theta)\boldsymbol{u}_{ij}]$, which can be estimated by

$$\hat{B}=\left[\sum_{i=1}^{3}\frac{w_i}{n_i}\sum_{j=1}^{n_i}\boldsymbol{u}_{ij}\boldsymbol{u}_{ij}'\right]^{-1}\sum_{i=1}^{3}\frac{w_i}{n_i}\sum_{j=1}^{n_i}(Y_{ij}-\theta)\boldsymbol{u}_{ij},$$

with the usual $\sqrt{n}$ consistency, i.e. $\hat{B}=B+O_p(n^{-1/2})$. If we let $X_{ij}=Y_{ij}-\theta-B'\boldsymbol{u}_{ij}$, then

(i) The set of constraints (6) and (17) is equivalent to the set of constraints (6) and $\sum_{i=1}^{3}w_i\sum_{j=1}^{n_i}p_{ij}X_{ij}=0$.

Author's personal copy

(ii) $\sum_{i=1}^{3}(w_i/n_i)\sum_{j=1}^{n_i} X_{ij}\boldsymbol{u}_{ij} = O_p(n^{-1/2})$ when $\theta = \theta_0$.
(iii) The $\tilde{p}_{ij}$'s which maximize l_w subject to (4), (6) and the additional constraint (14) will lead to the same type of expansion (16) for $l_w(\tilde{\boldsymbol{p}}(\theta))$, if we replace $\boldsymbol{u}_{ij}$ by $\boldsymbol{u}_{ij}^* = (\boldsymbol{u}_{ij}', X_{ij})'$.

To finish the proof, we further notice that

$$\boldsymbol{U}^* = \sum_{i=1}^{3} \frac{w_i}{n_i} \sum_{j=1}^{n_i} \boldsymbol{u}_{ij}^* = (\boldsymbol{U}', \bar{X})',$$

where $\bar{X} = \sum_{i=1}^{3}(w_i/n_i)\sum_{j=1}^{n_i} X_{ij}$, which is of the order $O_p(n^{-1/2})$ when $\theta = \theta_0$, and

$$\boldsymbol{D}^* = \sum_{i=1}^{3} \frac{w_i}{n_i} \sum_{j=1}^{n_i} \boldsymbol{u}_{ij}^* \boldsymbol{u}_{ij}^{*'} = \begin{pmatrix} \boldsymbol{D} & 0 \\ 0 & A \end{pmatrix} + O_p(n^{-1/2}),$$

where $A = \sum_{i=1}^{3}(w_i/n_i)\sum_{j=1}^{n_i} X_{ij}^2$. The fact that $\boldsymbol{U}^* = O_p(n^{-1/2})$ further leads to

$$\boldsymbol{U}^{*'}(\boldsymbol{D}^*)^{-1}\boldsymbol{U}^* = \boldsymbol{U}'\boldsymbol{D}^{-1}\boldsymbol{U} + (\bar{X})^2/A + o_p(n^{-1}).$$

The final asymptotic expansion for the WEL ratio statistic is given by $-2r_w(\theta) = (\bar{X})^2/A + o_p(n^{-1})$. If we let

$$c_2 = Var(\bar{X})/A, \tag{18}$$

then the conclusion of Theorem 2 follows from the fact that $\bar{X}$ is asymptotically normally distributed with zero mean when $\theta = \theta_0$. □

The weighted formulation of the EL ratio statistic presented in Theorem 2 eliminates the non-zero minimum value problem associated with the naive EL approach of Tsao and Wu (2006), since $-2r_w(\theta) = 0$ when $\theta = \sum_{j=1}^{n_1} \hat{p}_{1j} Y_{1j}$ $(= \sum_{j=1}^{n_i} \hat{p}_{ij} Y_{ij}$ for $i = 2, \ldots, k$ due to constraint (13)). But the problem of poor asymptotic χ^2 approximation remains since the interval is confined by the intersection of the k convex hulls formed respectively by the k samples. In addition, the scaling constant c_2 involves unknown parameters that need to be estimated from the sample. The bootstrap calibration method described below not only bypasses the estimation of c_2 but also provides better approximation to the sampling distribution of $-2r_w(\theta_0)$ when sample sizes are not large. The procedure provides an approximation to the conditional distribution of $-2r_w(\theta_0)$ given that θ_0 is an inner point of the intersection of the convex hulls, a condition required for $-2r_w(\theta_0)$ to be computable.

Let $\hat{\theta}_0 = \sum_{j=1}^{n_1} \hat{p}_{1j} Y_{1j}$ be the MWEL estimator of the common mean θ_0. It follows from the same argument as in Tsao and Wu (2006) that $\hat{\theta}_0 = \theta_0 + O_p(n^{-1/2})$. Let $\{Y_{ij}^*,\ j = 1, \ldots, n_i,\ i = 1, \ldots, k\}$ be a bootstrap sample, i.e. $Y_{ij}^*,\ j = 1, \ldots, n_i$ are randomly selected from $\{Y_{ij},\ j = 1, \ldots, n_i\}$ with replacement, $i = 1, \ldots, k$. Let $-2r_w^*(\hat{\theta}_0)$ be computed in the same way as for $-2r_w(\theta)$ but using the bootstrap sample with $\theta = \hat{\theta}_0$ being used in (14). To obtain a Monte Carlo approximation to the conditional distribution of $-2r_w^*(\theta)$ given that it is computable, a sequence of independent bootstrap samples are used but samples for which the intersection of the convex hulls formed by the bootstrap samples does not include $\hat{\theta}_0$as an inner point are disregarded. An asymptotic expansion to $-2r_w^*(\hat{\theta}_0)$ can easily be established by following the same argument used in the proof of Theorem 2. The conclusion that the sampling distribution of $-2r_w^*(\hat{\theta}_0)$ is a consistent estimator of the sampling distribution of $-2r_w(\theta)$ follows from the fact that both have the same scaled asymptotic χ^2 distribution.

Let b_α^* be the αth upper quantile obtained from the Monte Carlo approximation to the distribution of $-2r_w^*(\hat{\theta}_0)$. The bootstrap calibrated $(1-\alpha)$-level WEL ratio confidence interval on θ_0 can be constructed as

$$\mathscr{C}^* = \{\theta \mid -2r_w(\theta) < b_\alpha^*\}.$$

This interval has correct $1-\alpha$ coverage probability under large samples but performs dramatically better than the naive EL ratio confidence intervals for samples of small or moderate sizes.

4. Inference on a population mean using related samples

One unique scenario of multiple sample problems is when we have multiple samples available but our inference is focused on one particular sample. This is the case, for instance, when we have a combined stratified sample from the overall population but we are only interested in estimating the mean of one particular stratum. In a more general setting, we could have data collected from other occasions which are related to the current study. These related samples might contain very useful information that

Author's personal copy

could be used to improve the inference on the population of interest. Wang et al. (2004) studied the point estimation problem under such settings, using a parametric weighted likelihood approach. In this section, we show that the WEL method proposed in Section 2 can be used to construct confidence intervals on the population mean using information from related samples.

We first briefly introduce the parametric weighted likelihood proposed by Wang et al. (2004). Suppose $Y_{ij},\ j=1,\ldots,n_i$ are iid random variables with probability density $f_i(y_{ij};\theta_i)$. The parameter of interest is θ_1 of the first density f_1. Wang et al. (2004) proposed a parametric weighted likelihood to integrate information from related but different samples to yield a more reliable estimator of θ_1 than the classical maximum likelihood estimator based on the single sample. The parametric weighted likelihood for θ_1 using related samples is defined as

$$\mathrm{WL}(\theta_1,\boldsymbol{y})=\prod_{i=1}^{k}\prod_{j=1}^{n_i}f_1(y_{ij};\theta_1)^{\lambda_i}, \tag{19}$$

where $\boldsymbol{y}=(\boldsymbol{y}_1,\ldots,\boldsymbol{y}_k),\ \boldsymbol{y}_i=(y_{i1},\ldots,y_{in_i})',\ i=1,\ldots,k$, and the λ_i's are likelihood weights that must be specified. The weight assigned to each related sample must accurately reflect the importance or relevance of the information contained in that sample. The parametric weighted likelihood given by (19) is closely related to the relevant weighted likelihood proposed by Hu (1997) and the local likelihood method proposed by Tibshirani and Hastie (1987) and Eguchi and Copas (1998). There exist some other types of parametric weighted likelihood methods proposed in the literature for different purposes. Hu and Zidek (2002) provided an excellent review of various weighted parametric likelihood approaches to date. These weighted likelihoods are primarily designed for point estimation. Confidence intervals have not been considered in previous studies of parametric weighted likelihood approaches.

The WEL approach described in Section 2 can be applied here to make inference on the parameter of interest, θ_1. Let us again consider k related samples $\boldsymbol{Y}_i=(Y_{i1},\ldots,Y_{in_i})',\ i=1,\ldots,k$ with $\theta_i=E(Y_{ij})$ being the population means and the goal is to construct WEL confidence intervals on θ_1 of the first population, using all related samples. Our proposed strategy is to put the first sample along with other related samples under a synthetic stratified population. We then consider the "overall population mean" $\theta=\sum_{i=1}^k w_i\theta_i$ with judiciously chosen weights w_i to achieve the following two goals: (i) the WEL confidence interval on θ is asymptotically equivalent to the single sample EL confidence interval on θ_1 when the first sample size n_1 is large; (ii) the WEL interval on θ, treated as if it is for θ_1, performs better than the single sample EL interval on θ_1 when n_1 is small or even moderate. For a given set of weights $\boldsymbol{w}=(w_1,\ldots,w_k)'$, the MWEL estimator of $\theta=\sum_{i=1}^k w_i\theta_i$ is given by $\hat\theta=\sum_{i=1}^k w_i\bar Y_{i\cdot}$.

The central issue for the WEL inference when relevant samples are available is the choice of the weights w_i, which is parallel to the selection of λ_i in the parametric weighted likelihood case. How to choose a set of optimal likelihood weights is indeed a very challenging research problem. Various choices of likelihood weights currently available in the literature all face different difficulties. For example, the kernel weights often used in the local likelihood would require the determination of the bandwidth parameter h. Likelihood weights without parametric forms such as the cross-validated (cv) weights proposed by Wang and Zidek (2005a) also face difficulties on numerical instabilities. A thorough study of optimal weights is beyond the scope of this article. In this section, we propose to choose the weights for the WEL approach using a supervised cross-validation procedure based on the likelihood weights of Wang and Zidek (2005a).

For the purpose of point estimation, Wang and Zidek (2005a) propose to choose the likelihood weights using a cross-validation procedure for the parametric weighted likelihood. More specifically, consider the following natural measure of the total prediction error when we use the MWEL estimator $\hat\theta$ to estimate θ_1,

$$D(\boldsymbol{w},\boldsymbol{Y}_1,\ldots,\boldsymbol{Y}_k)=\sum_{j=1}^{n_1}(Y_{1j}-\hat\theta^{[-j]})^2,$$

where $\hat\theta^{[-j]}$ is the MWEL estimator of θ without using the jth observation Y_{1j} from the first sample. The optimum cv weights can then be chosen to minimize the total prediction error, i.e.

$$\boldsymbol{w}^{\mathrm{cv}}=arg\min_{\boldsymbol{w}}D(\boldsymbol{w},\boldsymbol{Y}_1,\ldots,\boldsymbol{Y}_k).$$

The cv weights are intended to obtain likelihood weights without the knowledge of the functional form of the underlying distribution function. Thus, it could be used for choosing weights for WEL. However, the direct use of the cv likelihood weights for constructing confidence intervals could be difficult for the following two reasons. Firstly, the weights chosen by the cross-validation are designed to achieve better point estimation for θ_1 as $D(\boldsymbol{w},\boldsymbol{Y}_1,\ldots,\boldsymbol{Y}_k)$ is closely related to the mean squared error. However, our experiences suggest that the optimal set of weights for point estimation does not necessarily lead to optimal results for confidence intervals. Secondly, the cross-validation procedure might not work well for very small sample sizes as stated in Wang and Zidek (2005a). To see this, we consider the case $k=2$ with equal sample sizes $n_1=n_2$, denoted by m, and $\hat\theta^{[-j]}$ is obtained by the alternative deletion scheme where the jth observations from both samples are removed. Wang and Zidek (2005a) showed that the optimum cv weights take the following form:

$$w_1^{\mathrm{cv}}=1-w_2^{\mathrm{cv}}\quad\text{and}\quad w_2^{\mathrm{cv}}=S_2/S_1,$$

Author's personal copy

where

$$S_1 = \frac{m(m-2)}{(m-1)^2}(\bar{Y}_{1\cdot} - \bar{Y}_{2\cdot})^2 + \frac{1}{m(m-1)^2}\sum_{j=1}^{m}(Y_{1j} - Y_{2j})^2,$$

$$S_2 = \frac{m}{(m-1)^2}(\hat{\sigma}_1^2 - \hat{cov}),$$

with $\hat{\sigma}_1^2$ being the first sample variance and $\hat{cov}$ being the sample covariance between the two samples. It can be seen that the first term in S_1 is dominant when m is large. For small m, however, the term $(\bar{Y}_{1\cdot} - \bar{Y}_{2\cdot})^2$ has a chance of being very small. If this happens, then the S_1, which serves as the denominator in w_2^{cv}, could be very small and result in a very unreasonably large value for w_2^{cv} that could be bigger than 1 which is clearly not a reasonable weight for the relevant sample.

Wang and Zidek (2005b) showed that the optimal weights should be non-negative if the loss function is chosen to be the Kullback–Leibler function also known as the relative entropy. One simple solution is to truncate the cv weights so that they all lie between 0 and 1. For example, if w_2^{cv} is too large or negative due to numerical instability, we could set it to be zero. This approach, however, could result in losing all relevant information from related samples when such information is desperately needed for small sample inference.

Instead of using the simple truncation, we propose the following supervised cross-validation procedure. We again consider the simple case of $k=2$ with equal sample sizes $n_1 = n_2 = m$. We utilize the classical t statistic when the cv weights are unstable. To be more specific, we consider the following truncation for the second weight:

$$w_2^s = \frac{\min(t/m, 0.5)}{0.5 + \min(t/m, 0.5)}, \tag{20}$$

where $t = |\bar{Y}_{1\cdot} - \bar{Y}_{2\cdot}|/\sqrt{\hat{\sigma}_1^2/m + \hat{\sigma}_2^2/m}$ with $\hat{\sigma}_2^2$ being the second sample variance. It is apparent that the weight w_2^s will not exceed 0.5, and consequently, the weight assigned to the first sample will always assume a bigger value than that for the second sample. The second weight can be approximated by

$$w_2^s = \frac{2\min(t/m, 0.5)}{1 + 2\min(t/m, 0.5)} \approx 2t/m \tag{21}$$

when t/m is small. Let

$$w_1^{scv} = 1 - w_2^{scv}, \quad w_2^{scv} = \min(w_2^{cv}, 2t/m, 0.5). \tag{22}$$

We term w_i^{scv}, $i=1,2$ the supervised cross-validated (scv) weights. They are designed to correct the numerical instability of the unsupervised cv weights. When the unsupervised weight w_2^{cv} is too big, which is often related to cases where $|\bar{Y}_{1\cdot} - \bar{Y}_{2\cdot}|$ is small, we use the classical t statistic for the truncation. The weight for the second sample is also controlled so that the inference will be dominated by the first sample.

The supervised weights also have the same order of asymptotic convergence as the cv weights. This is due to the fact that $2t/m$ converges to 0 in probability and is of the order $O_p(m^{-1})$. It is then straightforward to show that

$$(w_1^{scv}, w_2^{scv}) \xrightarrow{p} (1, 0).$$

Therefore, the WEL confidence interval on θ obtained using the result of Theorem 1 and the weights w_i^{scv} reduce to the usual single sample EL confidence interval on θ_1 as $m \to \infty$.

The proposed scv weights can be generalized for multiple samples as well as samples of unequal sizes. However, when there are more than two samples and the sample sizes are unequal, the unsupervised cross-validation procedure is computationally intensive and the cv weights are difficult to compute since the derivation of the optimal weights involves handling matrices which are not invertible (Wang and Zidek, 2005a). Therefore, we propose the following pair-wise approach to determine the likelihood weights when there are k samples with unequal sizes. Let

$$w_1^{scv} = 1 - \sum_{i=2}^{k} w_i^{scv}, \quad w_i^{scv} = \frac{\min(w_i^{cv}, 2t_i/n_1, 0.5)}{(k-1)/2 + \sum_{i=2}^{k}\min(w_i^{cv}, 2t_i/n_1, 0.5)}, \quad i = 2, \ldots, k, \tag{23}$$

where w_i^{cv} is the unsupervised cv likelihood weight and t_i is the t statistic using samples 1 and i only. Therefore, the relevance of each related sample is evaluated against the first sample on a pair-wise fashion. The term $(k-1)/2$ is introduced as part of the normalizing constant to ensure that the sum of the weights of all related samples will not exceed the weight assigned to the first sample.

The WEL confidence interval on θ_1 can be constructed using the methods described in Section 2, with the weights w_i's pre-calculated from the aforementioned supervised cross-validation procedure and treating the "overall population mean" θ as if it is θ_1. For the bootstrap calibration method, the weights w_i remain the same and are not re-calculated for each of the bootstrap samples. Since $w_1 \xrightarrow{p} 1$ and $w_i \xrightarrow{p} 0$ for $i > 1$, the WEL interval reduces to the usual single sample EL interval under large samples.

Author's personal copy

We demonstrate through a simulation study reported in the next section that the WEL confidence interval on θ_1 using related samples performs much better than the single sample EL confidence interval when sample sizes are small or moderate.

5. Simulation studies

In this section, we report results from an extensive simulation study on the finite sample performance of the WEL confidence intervals for the three types of problems discussed in the previous sections, and compare them with existing methods. The total number of simulation runs for each of the parameter settings is 2000. For the bootstrap method, 1000 bootstrap samples are used in the Monte Carlo approximation.

5.1. Stratified sampling

We consider the case in which $k=3$. Stratified samples are generated from the following model:

$$Y_{ij} = \theta_i + \sigma_i \varepsilon_{ij}, \quad j=1,\ldots,n_i, \;\; i=1,2,3,$$

where ε_{ij} are iid $(\chi_1^2-1)/\sqrt{2}$ with zero mean and unit variance. We choose $\theta_i=i$ and $(\sigma_1^2,\sigma_2^2,\sigma_3^2)=(\frac{1}{2},1,2)$ such that $E(Y_{ij})=i$ and $Var(Y_{ij})=\sigma_i^2$. The stratum weights are chosen as $(w_1,w_2,w_3)=(0.5,0.3,0.2)$. Under this setting, the smaller stratum has bigger mean and variance, a scenario often seen in practice. The choice of a non-symmetric distribution for ε_{ij} and hence for Y_{ij} is also common in real world situations. The true population mean is $\theta_0=0.5\times 1+0.3\times 2+0.2\times 3=1.7$.

For each simulated sample of size (n_1,n_2,n_3), three confidence intervals on θ_0 are computed: (i) the conventional normal theory interval (NTI) based on the stratified mean estimator and its estimated variance; (ii) the WEL interval using the χ^2 calibration method (WEL$_1$) based on Theorem 1; and (iii) the WEL interval using the bootstrap calibration method (WEL$_2$).

Table 1 reports the simulated coverage probability (CP), lower (L) and upper (U) tail error rates, the average length (AL) of the 90% confidence intervals on θ_0 for different sample size combinations. We note that NTI and WEL$_1$ have similar coverage probabilities and they are both lower than the nominal level when the sample sizes are (10, 10, 10). The bootstrap calibrated WEL$_2$ interval has much improved coverage probabilities but its average length is a bit enlarged for the case of $(n_1,n_2,n_3)=(10,10,10)$. For the three sample size combinations where the total sample size n is 60, both NTI and WEL$_1$ intervals are improved and WEL$_2$ remains the best. The two WEL intervals become virtually identical when $n_i \geqslant 60$. As for tail error rates, the WEL-based intervals (WEL$_1$ and WEL$_2$) are more balanced than NTI under all sample size combinations.

5.2. The common mean problem

Once again, we consider $k=3$. The three samples are generated from the common mean model:

$$Y_{ij} = \theta_0 + \sigma_i \varepsilon_{ij}, \quad j=1,\ldots,n_i, \;\; i=1,2,3,$$

where the ε_{ij}'s follow the same location and scale transformed χ_1^2 distribution as in the model of Section 5.1. The true value of the common mean θ_0 is set to be 1. Several variance structures $(\sigma_1^2,\sigma_2^2,\sigma_3^2)$ are considered to reflect the severity of heteroscedasticity.

Table 1
Simulated results of 90% confidence intervals under stratified sampling.

(n_1,n_2,n_3)	CI	L	CP	U	AL
(10,10,10)	NTI	1.9	84.7	13.4	0.31
	WEL$_1$	4.5	85.4	10.1	0.31
	WEL$_2$	3.2	88.6	8.2	0.36
(20,20,20)	NTI	2.3	87.2	10.5	0.23
	WEL$_1$	5.0	87.1	7.9	0.23
	WEL$_2$	4.0	88.9	7.1	0.25
(10,20,30)	NTI	2.5	85.7	11.8	0.25
	WEL$_1$	4.8	86.7	8.5	0.26
	WEL$_2$	4.4	87.4	8.2	0.27
(30,20,10)	NTI	1.7	85.7	12.6	0.24
	WEL$_1$	5.1	85.2	9.7	0.24
	WEL$_2$	3.9	87.5	8.6	0.26
(60,60,60)	NTI	3.0	89.9	7.1	0.13
	WEL$_1$	4.6	90.5	4.9	0.14
	WEL$_2$	4.4	90.8	4.8	0.14
(90,90,90)	NTI	2.7	89.6	7.7	0.11
	WEL$_1$	3.8	90.3	5.9	0.11
	WEL$_2$	3.8	90.2	6.0	0.11

Author's personal copy

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

(n_1,n_2,n_3)	CI	L	CP	U	AL
(20,20,20)	EL-3	4.3	75.9	19.8	0.37
	WEL_1	4.1	78.3	17.6	0.33
	WEL_2	1.0	87.5	11.5	0.50
(10,20,30)	EL-3	4.8	74.4	20.8	0.40
	WEL_1	4.1	77.3	18.6	0.38
	WEL_2	2.1	81.3	16.6	0.55
(30,20,10)	EL-3	4.1	73.3	22.6	0.33
	WEL_1	3.5	75.0	21.5	0.33
	WEL_2	1.0	83.6	15.4	0.51
(30,30,30)	EL-3	4.9	80.1	15.0	0.33
	WEL_1	4.0	83.5	12.5	0.29
	WEL_2	1.5	89.2	9.3	0.36
(20,30,40)	EL-3	5.0	79.4	15.6	0.35
	WEL_1	4.3	82.2	13.5	0.32
	WEL_2	1.9	88.2	9.9	0.42
(40,30,20)	EL-3	4.8	79.6	15.6	0.31
	WEL_1	4.0	83.0	13.0	0.28
	WEL_2	1.5	89.9	8.6	0.38
(60,60,60)	EL-3	5.9	84.1	10.0	0.26
	WEL_1	5.5	86.1	8.4	0.22
	WEL_2	3.0	89.6	7.4	0.24
(90,90,90)	EL-3	5.6	86.1	8.3	0.22
	WEL_1	5.1	87.5	7.4	0.18
	WEL_2	3.0	90.1	6.9	0.19

For each simulated sample of sizes (n_1,n_2,n_3), three confidence intervals of the common mean θ_0 are computed: the unweighted EL interval based on three degrees of freedom (EL-3) and the two WEL intervals, WEL_1 and WEL_2, similarly defined as in Section 5.1 but based on Theorem 2.

Table 2 summarizes the simulated results of CP, L, U and AL of the 90% confidence intervals for the case of $(\sigma_1^2,\sigma_2^2,\sigma_3^2)=(\frac{1}{2},1,2)$ with selected sample size combinations. When $\min\{n_1,n_2,n_3\}=10$, none of the three EL type intervals is acceptable, since their coverage probabilities are too low; when $\min\{n_1,n_2,n_3\}=20$, the bootstrap calibrated WEL_2 interval has coverage probability close to the nominal value while the other two alternatives still have serious under-coverage problems. The naive EL method (EL-3) has low coverage probabilities even for $n_i \geqslant 60$. The bootstrap calibrated WEL_2 interval has improved coverage probabilities in all cases considered but it comes with a price of inflated length when the sample sizes are small. This is also the case for the EL method in many other scenarios as well. For the common meanproblem discussed here, the inflation in length disappears when $n_i \geqslant 60$. In general, the bootstrap calibrated WEL method proposed in this article offers a significant improvement in terms of coverage probabilities over the naive EL approach discussed in Tsao and Wu (2006) .

5.3. *Inference using related samples*

We also consider $k=3$. The parameter of interest is the mean of the first population, θ_1. We consider several combinations of normal and exponential distributions as the initial population setting from which the three samples are generated. For each simulated sample of sizes (n_1,n_2,n_3), four confidence intervals on θ_1 are computed: (i) the χ^2 calibrated unweighted EL interval based on the first sample only (EL_1); (ii) the bootstrap calibrated EL interval based on the first sample only (EL_2); (iii) the χ^2 calibrated WEL interval using information from related samples (WEL_1); and (iv) the bootstrap calibrated WEL interval using information from related samples (WEL_2).

Table 3 presents the simulated results of CP, L, U and AL of the 90% confidence intervals for selected sample size combinations under several population settings. There are two striking patterns shown from Table 3. First, compared with the single sample based EL_1, WEL_1 not only improves the coverage probability but also dramatically reduces the average length of the interval. Second, when bootstrap calibration is used, the WEL method (WEL_2) provides similar coverage probabilities as EL_2 but with much shorter interval length. The reduction of average length ranges from 18% to 24%.

6. Some additional remarks

Multiple sample problems are quite common in many statistical applications. A naive formulation of the EL-based approach is either not very efficient or asymptotically intractable due to the irregular types of constraints induced by the multiple samples.

Author's personal copy

Table 3
Simulated results of 90% confidence intervals on the first population mean.

1st S	2nd S	3rd S	(n_1, n_2, n_3)	CI	L	CP	U	AL
N(0, 1)	N(0.1, 1)	N(0.3, 1)	(15, 15, 15)	EL_1	7.0	85.8	7.2	0.81
				WEL_1	7.3	87.6	5.1	0.72
				EL_2	5.4	88.7	6.0	0.90
				WEL_2	6.8	88.5	4.8	0.74
N(0, 1)	N(−0.1, 1)	N(0.3, 1)	(15, 15, 15)	EL_1	7.2	86.3	6.6	0.82
				WEL_1	6.9	88.5	4.6	0.72
				EL_2	6.1	89.1	4.9	0.91
				WEL_2	6.4	89.4	4.3	0.75
Exp(1)	Exp(10/9)	Exp(10/8)	(15,20,20)	EL_1	4.9	83.9	11.3	0.79
				WEL_1	2.7	86.1	11.3	0.68
				EL_2	3.6	87.9	8.6	0.97
				WEL_2	2.3	87.1	10.7	0.75
Exp(1/3)	Exp(1/2)	Exp(1/4)	(15,20,20)	EL_1	5.8	83.8	10.5	2.35
				WEL_1	4.0	86.6	9.5	2.02
				EL_2	3.5	88.7	7.9	2.91
				WEL_2	3.3	88.0	8.7	2.20

The proposed WEL approach provides a unified framework for the EL-based inference. We have demonstrated this through three scenarios, namely, inference under stratified sampling, the estimation of a common mean using multiple samples, and inference of one particular population mean in the presence of related samples. Our weighted approach facilitates the involved asymptotic derivations as well as computational procedures. The WEL confidence intervals under the three scenarios considered in this article are shown to be more efficient than existing ones.

Our proposed method can also be applied to a variety of two-sample problems. Jing (1995) investigated the EL inference on the difference of two population means, $\theta = \mu_1 - \mu_2$, using the simple unweighted approach. He used three constraints corresponding to the three parameters: $\sum_{i=1}^{n_1} p_{1i} Y_{1i} = \mu_1$, $\sum_{i=1}^{n_2} p_{2i} Y_{2i} = \mu_2$ and $\sum_{i=1}^{n_1} p_{1i} Y_{1i} - \sum_{i=1}^{n_2} p_{2i} Y_{2i} = \theta$. But clearly the last constraint is redundant and his profile likelihood function for θ is essentially for both μ_1 and μ_2. It is not clear how inference about θ is directly handled under such an approach. Under our proposed WEL approach, the WEL function is given by (1) with $k=2$ and $w_1 = w_2 = \frac{1}{2}$. In addition to the normalization constraints $\sum_{i=1}^{n_1} p_{1i} = 1$ and $\sum_{i=1}^{n_2} p_{2i} = 1$, we only need to impose one single constraint corresponding to the true parameter of interest, the difference of the two population means: $\sum_{i=1}^{n_1} p_{1i} Y_{1i} - \sum_{i=1}^{n_2} p_{2i} Y_{2i} = \theta$. It can be shown that a result similar to Theorem 2 regarding θ can be established. Some further examination along this line, with applications to case–control studies, is currently under investigation.

Acknowledgements

This research was supported by grants from the Natural Sciences and Engineering Research Council of Canada. We thank the editor, the associate editor and two anonymous referees for constructive comments and suggestions which lead to improved presentation of the paper.

References

Chen, J., Sitter, R.R., 1999. A pseudo empirical likelihood approach to the effective use of auxiliary information in complex surveys. Statist. Sinica 9, 385–406.
Eguchi, S., Copas, J., 1998. A class of local likelihood methods and near-parametric asymptotics. J. Roy. Statist. Soc. B 60, 709–724.
Hu, F., 1997. The asymptotic properties of the maximum-relevance weighted likelihood estimators. Canad. J. Statist. 25, 45–59.
Hu, F., Zidek, J.V., 2002. The weighted likelihood. Canad. J. Statist. 30, 347–371.
Jing, B.Y., 1995. Two-sample empirical likelihood method. Statist. Probab. Lett. 24, 315–319.
Owen, A.B., 1988. Empirical likelihood ratio confidence intervals for a single functional. Biometrika 75, 237–249.
Owen, A.B., 2001. Empirical Likelihood. Chapman & Hall, CRC, London, Boca Raton, FL.
Tibshirani, R., Hastie, T., 1987. Local likelihood estimation. J. Amer. Statist. Assoc. 82, 559–567.
Tsao, M., Wu, C., 2006. Empirical likelihood inference for a common mean in the presence of heteroscedasticity. Canad. J. Statist. 34, 45–59.
Wang, X., 2006. Approximating Bayesian inference by weighted likelihood. Canad. J. Statist. 34, 279–298.
Wang, X., Zidek, J.V., 2005a. Selecting likelihood weights by cross-validation. Ann. Statist. 33, 463–500.
Wang, X., Zidek, J.V., 2005b. Derivation of mixture distribution and weighted likelihood as minimizers of the KL divergence subject to constraints. Ann. Inst. Statist. Math. 51, 687–701.
Wang, X., van Eeden, C., Zidek, J.V., 2004. Asymptotic properties of weighted likelihood estimators. J. Statist. Plann. Inference 119, 37–54.
Wu, C., 2004. Some algorithmic aspects of the empirical likelihood method in survey sampling. Statist. Sinica 14, 1057–1067.
Zhong, C.X.B., Rao, J.N.K., 2000. Empirical likelihood inference under stratified random sampling using auxiliary information. Biometrika 87, 929–938.