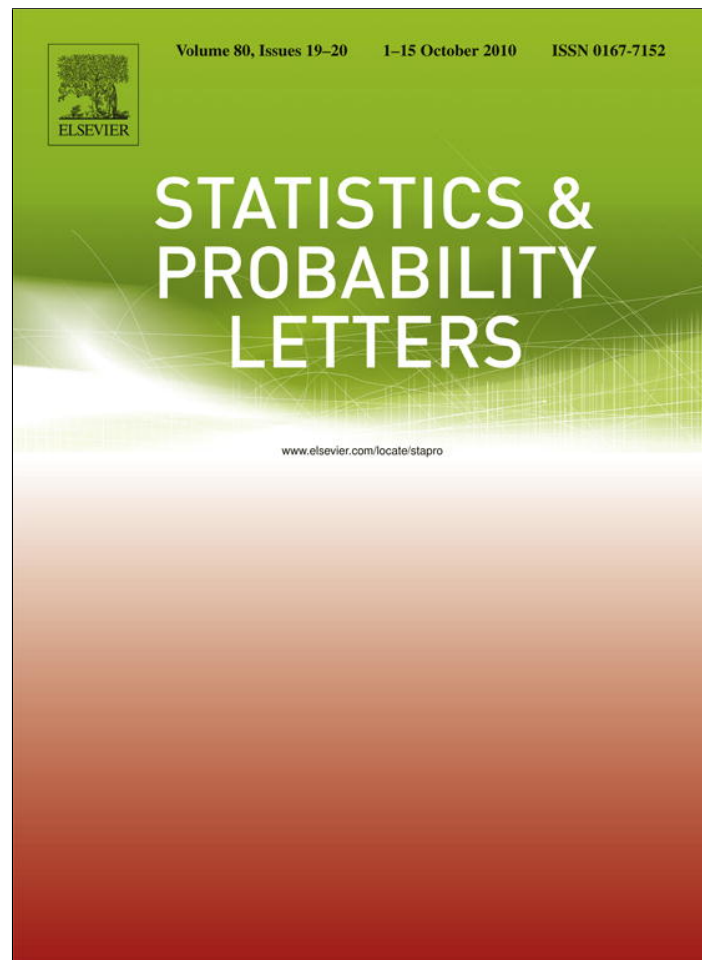


Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



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>



Contents lists available at ScienceDirect

Statistics and Probability Letters

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

Bootstrap procedures for the pseudo empirical likelihood method in sample surveys

Changbao Wu^{a,*}, J.N.K. Rao^b

^a Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ontario, N2L 3G1 Canada

^b School of Mathematics and Statistics, Carleton University, Ottawa, Ontario, K1S 5B6 Canada

ARTICLE INFO

Article history:

Received 28 May 2009

Received in revised form 15 May 2010

Accepted 23 May 2010

Available online 4 June 2010

Keywords:

Auxiliary information

Confidence interval

Design effect

Profile likelihood

Stratified sampling

Unequal probability sampling

ABSTRACT

Pseudo empirical likelihood ratio confidence intervals for finite population parameters are based on asymptotic χ^2 approximation to an adjusted pseudo empirical likelihood ratio statistic, with the adjustment factor related to the design effect. The calculation of the design effect involves variance estimation and hence requires second order inclusion probabilities. It also depends on how auxiliary information is used, and needs to be derived one-at-a-time for different scenarios. This paper presents bootstrap procedures for constructing pseudo empirical likelihood ratio confidence intervals. The proposed method bypasses the need for design effects and is valid under general single-stage unequal probability sampling designs with small sampling fractions. Different scenarios in using auxiliary information are handled by simply including the same type of benchmark constraints with the bootstrap procedures. Simulation results show that the bootstrap calibrated intervals perform very well and have much improved coverage probabilities over the χ^2 -based intervals when the sample sizes are small or moderate.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

The first major result in the development of the empirical likelihood (EL) method was the establishment of asymptotic χ^2 distribution of the EL ratio statistic for the population mean with independent and identically distributed (*iid*) sample data (Owen, 1988). The profile EL ratio confidence intervals have several advantages over the traditional normal theory intervals, such as range-respecting, transformation invariant and data-determined shapes. The EL intervals on the population mean based on the χ^2 approximation, however, tend to have coverage probabilities lower than nominal values when sample sizes are not large. Significant improvement can be made when the χ^2 approximation is replaced by a bootstrap calibration (Owen, 2001, Section 3.3). Bootstrap procedures for the EL method and related theoretical justifications are often straightforward for simple cases involving *iid* samples.

The EL method has been extended to complex survey data through a pseudo EL approach. The pseudo EL function, formulated by Chen and Sitter (1999), is the Narain–Horvitz–Thompson (NHT) estimator (Narain, 1951; Horvitz and Thompson, 1952) of the so-called “census” EL function and hence involves first order inclusion probabilities. It is shown by Wu and Rao (2006) that the pseudo EL ratio confidence intervals based on the χ^2 approximation require an adjustment factor which involves variance estimation and hence evaluation of second order inclusion probabilities. In addition, the adjustment factor also depends on how auxiliary information is used and needs to be derived one-at-a-time for different scenarios. Rao and Wu (2009) presented a detailed account of the EL methods for finite populations.

* Corresponding author. Tel.: +1 519 888 4567x35537; fax: +1 519 746 1875.
E-mail address: cbwu@uwaterloo.ca (C. Wu).

In this paper we present bootstrap procedures which bypass the adjustment factor previously required for the construction of pseudo EL ratio confidence intervals. The use of auxiliary information is handled in a unified manner by simply including the same type of benchmark constraints or calibration equations with the bootstrap procedure. The proposed methods are theoretically justified for single-stage with-replacement unequal probability sampling designs. They are also valid for without-replacement sampling designs when sampling fractions are small. Simulation results show that the bootstrap calibrated pseudo EL confidence intervals perform much better than those based on the χ^2 approximation using the adjustment factor when sample sizes are small. For simple random sampling without replacement with non-negligible sampling fractions, a simple correction can be made to the proposed bootstrap method and the resulting confidence intervals have correct asymptotic coverage probabilities. Simulation results show that the simple correction can also provide improved performance when it is applied to without-replacement unequal probability sampling designs with large sampling fractions.

A brief review of the pseudo EL ratio confidence intervals for complex surveys is given in Section 2. The proposed bootstrap procedures are presented in Section 3. Results from an extensive simulation study are reported in Section 4. We conclude with a few additional remarks in Section 5.

2. The pseudo empirical likelihood method

Consider a finite population \mathcal{U} consisting of N units. Let $\{(y_i, \mathbf{x}_i), i \in s\}$ be a non-stratified probability sample with fixed sample size n , where y_i and \mathbf{x}_i are the values of the response variable y and the vector of auxiliary variables \mathbf{x} associated with the i th unit, and s is the set of sample units selected using a probability sampling design. Let $\pi_i = P(i \in s)$ be the inclusion probabilities and $d_i = 1/\pi_i$ be the basic design weights. The pseudo EL function, first proposed by Chen and Sitter (1999), is given by $l(\mathbf{p}) = \sum_{i \in s} d_i \log(p_i)$ which is the NHT estimator of the so-called ‘‘census’’ empirical likelihood $\sum_{i=1}^N \log(p_i)$. This definition works fine for point estimation of population parameters but is not convenient for interval estimation or hypothesis testing.

The pseudo empirical likelihood (PEL) function defined in Wu and Rao (2006) is given by

$$l_{ns}(\mathbf{p}) = n \sum_{i \in s} \tilde{d}_i(s) \log(p_i), \tag{2.1}$$

where $\tilde{d}_i(s) = d_i / \sum_{i \in s} d_i$ are the normalized design weights and $\mathbf{p} = (p_1, \dots, p_n)'$ is the discrete probability measure imposed over the sampled units. Maximizing $l_{ns}(\mathbf{p})$ subject to $p_i > 0$ and $\sum_{i \in s} p_i = 1$ gives $\hat{p}_i = \tilde{d}_i(s)$. The maximum PEL estimator for the population mean $\bar{Y} = N^{-1} \sum_{i=1}^N y_i$ is given by $\hat{Y}_{PEL} = \sum_{i \in s} \hat{p}_i y_i = \sum_{i \in s} \tilde{d}_i(s) y_i$, and \hat{Y}_{PEL} is identical to the well-known Hájek estimator of \bar{Y} .

For PEL ratio confidence intervals on \bar{Y} , we consider the general case where the vector of population means, $\bar{\mathbf{X}} = N^{-1} \sum_{i=1}^N \mathbf{x}_i$, is known and needs to be incorporated into inferences. Let $\hat{p}_i, i \in s$ be the maximizer of $l_{ns}(\mathbf{p})$ subject to

$$\sum_{i \in s} p_i = 1, \tag{2.2}$$

$$\sum_{i \in s} p_i \mathbf{x}_i = \bar{\mathbf{X}}. \tag{2.3}$$

The maximum PEL estimator of \bar{Y} in this case is again defined as $\hat{Y}_{PEL} = \sum_{i \in s} \hat{p}_i y_i$. Let $\tilde{p}_i(\theta)$ be the maximizer of $l_{ns}(\mathbf{p})$ subject to (2.2), (2.3) and an additional constraint induced by the parameter of interest, \bar{Y} ,

$$\sum_{i \in s} p_i y_i = \theta \tag{2.4}$$

for a fixed θ . Let $r_{ns}(\theta) = -2\{l_{ns}(\tilde{\mathbf{p}}(\theta)) - l_{ns}(\hat{\mathbf{p}})\}$ be the PEL ratio function. It is shown by Wu and Rao (2006) that, under suitable regularity conditions, the adjusted PEL ratio function $r_{ns}^{[a]}(\theta) = r_{ns}(\theta)/\text{deff}_{GR}$ converges in distribution to a χ^2 random variable with one degree of freedom when $\theta = \bar{Y}$, where the design effect deff_{GR} is calculated based on a generalized regression (GR) estimator and requires second order inclusion probabilities $\pi_{ij} = P(i, j \in s)$.

A stratified probability sample is given by $\{(y_{hi}, \mathbf{x}_{hi}), i \in s_h, h = 1, \dots, H\}$ where $(y_{hi}, \mathbf{x}_{hi})$ is the value of (y, \mathbf{x}) associated with the i th unit in stratum h , s_h is the set of sample units selected from stratum h with fixed stratum sample size n_h , and H is the total number of strata in the population. Let $n = \sum_{h=1}^H n_h$ be the overall sample size. Let d_{hi} be the stratum design weights and $\tilde{d}_{hi}(s_h) = d_{hi} / \sum_{i \in s_h} d_{hi}$ be the normalized stratum design weights. The PEL function under stratified sampling is defined as

$$l_{st}(\mathbf{p}_1, \dots, \mathbf{p}_H) = n \sum_{h=1}^H W_h \sum_{i \in s_h} \tilde{d}_{hi}(s_h) \log(p_{hi}), \tag{2.5}$$

where $W_h = N_h/N$, N_h is the population size for stratum h , $N = \sum_{h=1}^H N_h$ is the overall population size, and $\mathbf{p}_h = (p_{h1}, \dots, p_{hn_h})'$ is the probability measure over the stratum sample s_h . The PEL ratio function $r_{st}(\theta)$ for the population mean $\bar{Y} = \sum_{h=1}^H W_h \bar{Y}_h$, where $\bar{Y}_h = N_h^{-1} \sum_{i=1}^{N_h} y_{hi}$ is the population mean for stratum h , is similarly defined as for the non-stratified case, with $l_{ns}(\mathbf{p})$ replaced by $l_{st}(\mathbf{p}_1, \dots, \mathbf{p}_H)$, and the constraints (2.2)–(2.4) respectively replaced by (2.6)–(2.8) below:

$$\sum_{i \in s_h} p_{hi} = 1, \quad h = 1, \dots, H, \tag{2.6}$$

$$\sum_{h=1}^H W_h \sum_{i \in s_h} p_{hi} \mathbf{x}_{hi} = \bar{\mathbf{X}}, \tag{2.7}$$

$$\sum_{h=1}^H W_h \sum_{i \in s_h} p_{hi} y_{hi} = \theta. \tag{2.8}$$

The adjusted PEL ratio function $r_{st}^{[a]}(\theta) = r_{st}(\theta)/\text{deff}_{GR(st)}$ is asymptotically distributed as χ_1^2 when $\theta = \bar{Y}$, where the design effect $\text{deff}_{GR(st)}$ is computed based on a combined generalized regression estimator using the stratified sample ($GR(st)$) as outlined in Wu and Rao (2006).

The $(1 - \alpha)$ -level profile PEL ratio confidence intervals on \bar{Y} can be constructed from

$$\mathcal{C}_a = \{\theta \mid r^{[a]}(\theta) < \chi_1^2(\alpha)\}, \tag{2.9}$$

where $\chi_1^2(\alpha)$ is the upper α quantile of the χ^2 distribution with one degree of freedom, and $r^{[a]}(\theta)$ is the adjusted PEL ratio function based on either the non-stratified or the stratified sample. While the interval \mathcal{C}_a shares all the attractive features of EL intervals for *iid* data, it also inherits the problem of low coverage probabilities when the sample size is not large. Moreover, the calculation of the design effect deff_{GR} or $\text{deff}_{GR(st)}$ depends on how auxiliary information is used and needs to be derived one-at-a-time for different scenarios. It also requires second order inclusion probabilities π_{ij} . From a theoretical point of view this is necessary since confidence intervals involve variances, either implicitly or explicitly; from a practical point of view this is often difficult, since computing the π_{ij} 's can be a daunting task, especially when the initial sampling design has to be modified or altered due to practical constraints. For example, Thompson and Wu (2008) encountered a practical case of multi-stage sampling where several upper level clusters, initially selected with probability proportional to cluster size using the randomized systematic pps sampling method (Hartley and Rao, 1962), needed to be replaced. They showed that under substitution of units it is possible in principle to calculate the first order inclusion probabilities but the second order inclusion probabilities become intractable.

3. Bootstrap procedures

The asymptotic χ^2 distribution of the adjusted PEL ratio function may be replaced by a bootstrap calibration. A major advantage of the bootstrap method in the current context is that it provides an approximation to the sampling distribution of the un-adjusted PEL ratio function and hence bypasses the adjustment factor required for the χ^2 approximation. It also improves the coverage probabilities of the resulting confidence intervals for samples of small or moderate sizes, as shown by the simulation results reported in Section 4.

The most crucial part of the proposed bootstrap procedures is that the basic design weights, d_i or d_{hi} , need to be treated as part of the sample data and a bootstrap version of the PEL function should be used. Consider a non-stratified single-stage unequal probability sample $\{(d_i, y_i, \mathbf{x}_i), i \in s\}$. Suppose that the population mean $\bar{\mathbf{X}}$ is known and the constraint (2.3) is used in defining the PEL ratio function $r_{ns}(\theta)$. Let b_α be the upper α quantile of the sampling distribution of $r_{ns}(\theta)$ at $\theta = \bar{Y}$. Then the profile $(1 - \alpha)$ -level confidence interval on \bar{Y} is constructed as

$$\mathcal{C}_u = \{\theta \mid r_{ns}(\theta) < b_\alpha\}. \tag{3.1}$$

The difference between \mathcal{C}_a given by (2.9) and \mathcal{C}_u specified by (3.1) is that \mathcal{C}_u uses the un-adjusted PEL ratio function and hence does not involve the adjustment factor and the design effect. The value of b_α in (3.1) can be approximated through the following bootstrap calibration procedure:

- (1) Select a bootstrap sample s^* of size n from the original sample s using simple random sampling with replacement and denote the bootstrap sample data by $\{(d_i^*, y_i^*, \mathbf{x}_i^*), i \in s^*\}$.
- (2) Let the bootstrap version of the PEL function be defined as

$$l_{ns}^*(\mathbf{p}) = n \sum_{i \in s^*} \tilde{d}_i^*(s^*) \log(p_i),$$

where $\tilde{d}_i^*(s^*) = d_i^* / \sum_{j \in s^*} d_j^*$.

- (3) Calculate the un-adjusted PEL ratio function $r_{ns}^*(\theta)$ at $\theta = \hat{Y}_{PEL}$ based on the bootstrap sample s^* and the bootstrap version of the PEL function $l_{ns}^*(\mathbf{p})$ in the same way as $r_{ns}(\theta)$ calculated from the original sample. The bootstrap version of the constraints (2.3) and (2.4) are given respectively by $\sum_{i \in s^*} p_i \mathbf{x}_i^* = \bar{\mathbf{X}}$ and $\sum_{i \in s^*} p_i y_i^* = \hat{Y}_{PEL}$.
- (4) Repeat Steps (1), (2) and (3) a large number of times, B , independently, to obtain the sequence $r_1^*(\theta), \dots, r_B^*(\theta)$, all at $\theta = \hat{Y}_{PEL}$. Let b_α^* be the upper 100α th sample quantile from this sequence.

Note that the maximum PEL estimator of $\bar{\mathbf{X}}$, computed as $\hat{\mathbf{X}}_{PEL} = \sum_{i \in s} \hat{p}_i \mathbf{x}_i$, is identical to $\bar{\mathbf{X}}$. The bootstrap version of the constraint (2.3) used in Step 3 is the same as $\sum_{i \in s^*} p_i \mathbf{x}_i^* = \hat{\mathbf{X}}_{PEL}$. The bootstrap calibrated PEL ratio confidence interval on \bar{Y} is constructed as $C_u^* = \{\theta \mid r_{ns}(\theta) < b_\alpha^*\}$, where the quantile b_α used in C_u of (3.1) is replaced by the bootstrap quantile b_α^* .

Theorem 1. *The bootstrap calibrated PEL ratio confidence interval C_u^* on \bar{Y} has asymptotically correct coverage probability at $(1 - \alpha)$ -level if the original sample is selected using unequal probability sampling with replacement.*

Proof. Following the same lines of the proof of Theorem 1 in Wu and Rao (2006) and assuming the same set of regularity conditions C1–C3, we can show that both $r_{ns}(\theta)$ at $\theta = \bar{Y}$ and $r_{ns}^*(\theta)$ at $\theta = \hat{Y}_{PEL}$ converge in distribution to the same scaled χ^2 random variable with one degree of freedom. \square

In practice, survey samples are usually selected by without-replacement sampling procedures. The interval C_u^* can be used for without-replacement sampling designs if the sampling fraction $f = n/N$ is small. When f is not small, the interval C_u^* under the proposed with-replacement bootstrap procedure tends to have an over-coverage problem, as shown by the simulation results reported in Section 4. For simple random sampling without replacement, the over-coverage problem can be corrected easily through the finite population correction factor, $1 - f$.

Theorem 2. *Let b_α^* be obtained by the with-replacement bootstrap steps (1)–(4) and assume $n/N \rightarrow \gamma \in (0, 1)$ as $n \rightarrow \infty$ and $N \rightarrow \infty$. Then the confidence interval $C_u^*(f) = \{\theta \mid r_{ns}(\theta) < (1 - f)b_\alpha^*\}$ on \bar{Y} has correct asymptotic coverage probability at the $(1 - \alpha)$ -level if the original sample is selected by simple random sampling without replacement.*

Proof of Theorem 2 follows from a straightforward asymptotic expansion of $r_{ns}(\theta)$ at $\theta = \bar{Y}$ and $r_{ns}^*(\theta)$ at $\theta = \bar{y} = n^{-1} \sum_{i \in s} y_i$. Note that, under simple random sampling without replacement, the normalized design weights $\tilde{d}_i(s) = n^{-1}$, where n is the sample size. The asymptotic sampling distributions of $r_{ns}(\theta)$ and $r_{ns}^*(\theta)$ differ by a scaling factor $1 - f$, due to the fact that $Var(\bar{y}) = (1 - f)S^2/n$ under the sampling design but $Var(\bar{y}^*) = S^2/n$ under the with-replacement bootstrap procedure, where S^2 is the population variance, s^2 is the sample variance, and $\bar{y}^* = n^{-1} \sum_{i \in s^*} y_i^*$ is the mean of the bootstrap sample.

For a general without-replacement unequal probability sampling design with non-negligible sampling fraction f , the interval $C_u^*(f)$ does not necessarily have coverage probabilities at the $(1 - \alpha)$ -level. In Section 4 we explore several scenarios involving different sampling method and sample size combinations through simulation studies. The results show that the $(1 - f)$ -adjusted interval $C_u^*(f)$ performs reasonably well for most of the cases considered.

The bootstrap procedure described for non-stratified samples can be readily extended to stratified sampling. Let $\{(d_{hi}^*, y_{hi}^*, \mathbf{x}_{hi}^*), i \in s_h^*\}$ be a bootstrap sample of size n_h selected from $\{(d_{hi}, y_{hi}, \mathbf{x}_{hi}), i \in s_h\}$ using simple random sampling with replacement, $h = 1, \dots, H$. The bootstrap version of the un-adjusted PEL ratio function $r_{st}(\theta)$ is computed by using $l_{st}^*(\mathbf{p}_1, \dots, \mathbf{p}_H) = n \sum_{h=1}^H W_h \sum_{i \in s_h^*} \tilde{d}_{hi}^*(s_h^*) \log(p_{hi})$, where $\tilde{d}_{hi}^*(s_h^*) = d_{hi}^* / \sum_{i \in s_h^*} d_{hi}^*$. Bootstrap versions of constraints (2.7) and (2.8) are given respectively by $\sum_{h=1}^H W_h \sum_{i \in s_h^*} p_{hi} \mathbf{x}_{hi}^* = \bar{\mathbf{X}}$ and $\sum_{h=1}^H W_h \sum_{i \in s_h^*} p_{hi} y_{hi}^* = \theta$, with θ replaced by \hat{Y}_{PEL} . Similar to non-stratified cases, the bootstrap version of constraint (2.7) is the same as $\sum_{h=1}^H W_h \sum_{i \in s_h^*} p_{hi} \mathbf{x}_{hi}^* = \hat{\mathbf{X}}_{PEL}$. The bootstrap calibrated PEL ratio confidence intervals are valid if sampling fractions within each of the strata are small.

4. Simulation studies

We examined the performance of the proposed bootstrap methods through simulation studies. Our study focused on comparisons among three versions of the PEL ratio confidence intervals: (i) the interval based on the adjusted PEL ratio function and asymptotic χ^2 approximation, denoted by EL(χ^2); (ii) the interval C_u^* with bootstrap calibration to the un-adjusted PEL ratio function, denoted by EL(B); and (iii) the interval $C_u^*(f)$ which replaces the bootstrap quantile b_α^* used in C_u^* by $(1 - f)b_\alpha^*$, denoted by EL(B_f). We consider different scenarios involving (a) non-stratified or stratified sampling designs; (b) different sampling procedures for selecting non-stratified samples; and (c) the use of auxiliary information through the inclusion or exclusion of calibration constraints. Our simulation is programmed in R using algorithms outlined in Wu (2004, 2005).

For non-stratified sampling designs, finite populations of size $N = 800$ were generated from the model $y_i = \beta_0 + \beta_1 x_i + \sigma \varepsilon_i$. This is the Model I used in Wu and Rao (2006); see Section 4 of Wu and Rao (2006) for a detailed description of the model. The finite population correlation coefficient between y and x was set to be 0.80 for the current study. The finite population,

Table 1

95% Confidence intervals on \bar{Y} : non-stratified samples by Rao–Sampford PPS sampling method.

| n | | CI | CP | L | U | AL |
|-----|---|----------------|------|-----|-----|------|
| 40 | – | NA | 91.7 | 0.6 | 7.7 | 0.48 |
| | | EL(χ^2) | 94.7 | 2.3 | 3.0 | 0.75 |
| | | EL(B) | 94.6 | 3.7 | 1.7 | 0.79 |
| | x | EL(B_f) | 93.9 | 4.1 | 2.0 | 0.77 |
| | | EL(χ^2) | 92.3 | 2.7 | 5.0 | 0.47 |
| | | EL(B) | 96.0 | 1.3 | 2.7 | 0.56 |
| 80 | – | EL(B_f) | 95.4 | 1.7 | 2.9 | 0.54 |
| | | NA | 94.2 | 1.5 | 4.3 | 0.34 |
| | | EL(χ^2) | 94.6 | 1.8 | 3.6 | 0.51 |
| | x | EL(B) | 95.4 | 2.8 | 1.8 | 0.55 |
| | | EL(B_f) | 94.3 | 3.2 | 2.5 | 0.52 |
| | | EL(χ^2) | 93.8 | 2.5 | 3.7 | 0.33 |
| | x | EL(B) | 96.3 | 1.3 | 2.4 | 0.37 |
| | | EL(B_f) | 94.6 | 1.7 | 3.7 | 0.35 |

Table 2

95% Confidence intervals on \bar{Y} : non-stratified samples by randomized systematic PPS sampling method.

| n | | CI | CP | L | U | AL |
|-----|---|-------------|------|-----|-----|------|
| 40 | – | EL(B) | 94.9 | 3.2 | 1.9 | 0.79 |
| | | EL(B_f) | 94.7 | 3.3 | 2.0 | 0.77 |
| | x | EL(B) | 96.8 | 1.5 | 1.7 | 0.57 |
| | | EL(B_f) | 96.6 | 1.7 | 1.7 | 0.56 |
| 80 | – | EL(B) | 96.4 | 2.0 | 1.6 | 0.55 |
| | | EL(B_f) | 95.6 | 2.3 | 2.1 | 0.52 |
| | x | EL(B) | 96.0 | 1.4 | 2.6 | 0.38 |
| | | EL(B_f) | 94.1 | 2.3 | 3.6 | 0.36 |

once generated, was held fixed and repeated simulation samples were drawn from the fixed population. We considered three different sample selection methods: the Rao–Sampford unequal probability sampling method (Rao, 1965; Sampford, 1967); the randomized systematic pps sampling method (Hartley and Rao, 1962); and simple random sampling without replacement. For the Rao–Sampford method and the randomized systematic pps method, the inclusion probabilities were set to be proportional to the values of the size measure, x . For each simulated sample, $B = 1000$ bootstrap samples were used in finding b_{α}^* . The total number of simulation runs was also 1000.

Performances of confidence intervals are evaluated and compared in terms of the simulated coverage probability (CP), lower (L) and upper (U) tail error rates and average length (AL). Table 1 contains results for the 95% confidence intervals on \bar{Y} where samples are selected by the Rao–Sampford method. The second column indicates whether the design variable x was calibrated to the known population mean \bar{X} in constructing the pseudo EL intervals. The sample size $n = 40$ corresponds to $f = 40/800 = 5\%$ and $n = 80$ corresponds to $f = 10\%$. The conventional confidence intervals based on normal approximations (NA) are also included for comparison, since variance estimators are available for the Rao–Sampford sampling method. Both EL(B) and EL(B_f) performed well, with coverage probabilities closer to the nominal value than EL(χ^2) when calibration over x is involved. The EL(B) intervals are a bit wider, resulting in coverage probabilities greater than the nominal value for a few cases. The EL(B_f) intervals, however, do not seem to have such problem and perform well for all cases considered. All PEL-based intervals perform better than the NA intervals in terms of coverage probabilities and balanced tail error rates.

We repeated the simulation study described above with the same parameter setting but using two different sampling methods. First, we replaced the Rao–Sampford method by the randomized systematic pps sampling method (Hartley and Rao, 1962). The results are reported in Table 2. The NA and EL(χ^2) intervals are not included since the required second order inclusion probabilities are not readily computable. The results are very similar to those under the Rao–Sampford method. Second, we considered simple random sampling without replacement with sample sizes $n = 40, 80, 120$ and 160 . The corresponding sampling fractions are $f = 5\%, 10\%, 15\%$ and 20% . This is the scenario for which the result of Theorem 2 applies. The results are shown in Table 3. The EL(B) intervals are generally wider than the EL(B_f) intervals, and are “overshooting” (in terms of coverage probability) for some cases. The EL(B_f) intervals, on the other hand, perform well regardless of sample sizes.

For stratified sampling designs, we generated finite populations from the model $y_{hi} = \alpha_h + \beta_h x_{hi} + \sigma \varepsilon_{hi}, i = 1, \dots, N_h, h = 1, \dots, H$, with $H = 4, N_1 = 800, N_2 = 600, N_3 = 400$ and $N_4 = 200$. This is the Model III used in Wu and Rao (2006); see Section 4 of Wu and Rao (2006) for a detailed description of the model. The finite population correlation coefficient between y and x was again set to be 0.80. Stratified samples were selected, with simple random sampling without replacement used within each of the four strata. Equal sample size allocations were used so that the sampling fractions, n_h/N_h , are larger for

Table 3

95% Confidence intervals on \bar{Y} : non-stratified samples by simple random sampling without replacement.

| n | CI | CP | L | U | AL |
|-----|-------------|------|-----|-----|------|
| 40 | EL(B) | 95.1 | 2.0 | 2.9 | 0.86 |
| | EL(B_f) | 94.6 | 2.3 | 3.1 | 0.83 |
| 80 | EL(B) | 96.4 | 1.7 | 1.9 | 0.59 |
| | EL(B_f) | 95.5 | 2.1 | 2.4 | 0.56 |
| 120 | EL(B) | 94.8 | 2.3 | 2.9 | 0.47 |
| | EL(B_f) | 93.2 | 3.3 | 3.5 | 0.44 |
| 160 | EL(B) | 96.3 | 2.0 | 1.7 | 0.41 |
| | EL(B_f) | 94.0 | 2.9 | 3.1 | 0.37 |

Table 4

95% Confidence intervals based on stratified random samples.

| n_h | | CI | CP | L | U | AL |
|-------|---|----------------|------|-----|-----|------|
| 20 | – | EL(χ^2) | 94.7 | 2.6 | 2.7 | 3.16 |
| | | EL(B) | 95.3 | 1.8 | 2.9 | 3.23 |
| | | EL(B_f) | 95.1 | 1.9 | 3.0 | 3.16 |
| | x | EL(χ^2) | 94.5 | 3.1 | 2.4 | 2.27 |
| | | EL(B) | 95.4 | 1.7 | 2.9 | 2.41 |
| | | EL(B_f) | 95.1 | 1.8 | 3.1 | 2.36 |
| 40 | – | EL(χ^2) | 95.1 | 2.7 | 2.2 | 2.16 |
| | | EL(B) | 95.4 | 2.1 | 2.5 | 2.27 |
| | | EL(B_f) | 94.7 | 2.4 | 2.9 | 2.18 |
| | x | EL(χ^2) | 93.3 | 2.3 | 4.4 | 1.59 |
| | | EL(B) | 95.6 | 1.7 | 2.7 | 1.67 |
| | | EL(B_f) | 94.8 | 2.0 | 3.2 | 1.60 |

smaller strata, resulting in unequal selection probabilities across strata. The overall sampling fractions are $f = 80/2000 = 4\%$ for $n_h = 20$ and $f = 160/2000 = 8\%$ for $n_h = 40$. These are the values of f used for the EL(B_f) intervals.

Table 4 contains results from stratified sampling. The second column indicates whether the calibration constraint (2.7) is used or not. The most interesting observation here is that the EL(B) intervals are not only good for cases with $f = 4\%$ but also acceptable for cases with $f = 8\%$. Over-coverage does not seem to be a problem for these later cases. The EL(B_f) intervals, adjusted by the overall finite population correction $1 - f$, perform very close to the EL(χ^2) intervals for all cases considered.

5. Concluding remarks

The bootstrap method proposed in this paper is used to obtain an approximation to the asymptotic sampling distribution of the PEL ratio statistic for constructing confidence intervals or conducting statistical hypothesis tests. This is different from the conventional normal approximation-based approach used in survey sampling where point estimation of variance of the mean estimator is the focus of bootstrap methods. Most existing bootstrap procedures for complex surveys are developed for that purpose. For without-replacement complex sampling designs, consistent bootstrap variance estimators often require special adjustment to the usual with-replacement bootstrap procedures. The Rao–Wu re-scaling bootstrap method (Rao and Wu, 1988) and the mirror-match method (Sitter, 1992) are two well-known examples on variance estimation. Bootstrap approximations to the sampling distribution of a statistic based on complex survey samples are much more demanding than point estimation of variance of the mean estimator. The proposed bootstrap methods perform well for single-stage without-replacement sampling designs when sampling fractions are small. For simple random sampling without replacement, the interval $C_u^*(f)$, which replaces the bootstrap quantile b_{α}^* by $(1-f)b_{\alpha}^*$, is asymptotically valid even if f is large. When sampling fractions are appreciable and the sampling designs are complex, a simple adjustment by the finite population correction factor $1 - f$ may not always work. Bootstrap methods which provide consistent estimation of the sampling distribution of the PEL ratio statistics in these later cases would be of great interest.

Our discussion in this paper is restricted to cases where the parameter θ is a scalar and the PEL ratio confidence intervals (2.9) or (3.1) can be computed using a simple algorithm described in Wu (2005). Extensions to cases where θ is a vector parameter are currently under investigation.

Acknowledgement

This research was supported by grants from the Natural Sciences and Engineering Research Council of Canada.

References

Chen, J., Sitter, R.R., 1999. A pseudo empirical likelihood approach to the effective use of auxiliary information in complex surveys. *Statistica Sinica* 9, 385–406.
 Hartley, H.O., Rao, J.N.K., 1962. Sampling with unequal probabilities and without replacement. *Annals of Mathematical Statistics* 33, 350–374.

- Horvitz, D.G., Thompson, D.J., 1952. A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association* 47, 663–685.
- Narain, R.D., 1951. On sampling with varying probabilities. *Journal of the Indian Society of Agricultural Statistics* 3, 169–174.
- 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.
- Rao, J.N.K., 1965. On two simple schemes of unequal probability sampling without replacement. *Journal of Indian Statistical Association* 3, 173–180.
- Rao, J.N.K., Wu, C., 2009. Empirical likelihood methods. In: Pfeiffermann, D., Rao, C.R. (Eds.), *Sample Surveys: Inference and Analysis*. In: *Handbook of Statistics*, vol. 29B. pp. 189–207.
- Rao, J.N.K., Wu, C.F.J., 1988. Resampling inference with complex survey data. *Journal of the American Statistical Association* 83, 231–241.
- Sampford, M.R., 1967. On sampling without replacement with unequal probabilities of selection. *Biometrika* 54, 499–513.
- Sitter, R.R., 1992. A resampling procedure for complex survey data. *Journal of the American Statistical Association* 87, 755–765.
- Thompson, M.E., Wu, C., 2008. Simulation-based randomized systematic PPS sampling under substitution of units. *Survey Methodology* 34, 3–10.
- Wu, C., 2004. Some algorithmic aspects of the empirical likelihood method in survey sampling. *Statistica Sinica* 14, 1057–1067.
- Wu, C., 2005. Algorithms and R codes for the pseudo empirical likelihood methods in survey sampling. *Survey Methodology* 31, 239–243.
- Wu, C., Rao, J.N.K., 2006. Pseudo-empirical likelihood ratio confidence intervals for complex surveys. *The Canadian Journal of Statistics* 34, 359–375.