The Canadian Journal of Statistics
Vol. 34, No. 3, 2006, Pages 359–375
La revue canadienne de statistique

Pseudo-empirical likelihood ratio confidence intervals for complex surveys

Changbao WU and J. N. K. RAO

Key words and phrases: Confidence interval; design effect; empirical likelihood; normal approximation; Rao–Sampford method; stratified sampling; unequal probability sampling.

MSC 2000: Primary 62D05; secondary 62G09.

Abstract: The authors show how an adjusted pseudo-empirical likelihood ratio statistic that is asymptotically distributed as a chi-square random variable can be used to construct confidence intervals for a finite population mean or a finite population distribution function from complex survey samples. They consider both non-stratified and stratified sampling designs, with or without auxiliary information. They examine the behaviour of estimates of the mean and the distribution function at specific points using simulations calling on the Rao–Sampford method of unequal probability sampling without replacement. They conclude that the pseudo-empirical likelihood ratio confidence intervals are superior to those based on the normal approximation, whether in terms of coverage probability, tail error rates or average length of the intervals.

Calcul d'intervalles de confiance fondés sur le rapport de pseudo-vraisemblances empiriques dans le cadre d'enquêtes complexes

Résumé : Les auteurs montrent comment une statistique du rapport de pseudo-vraisemblances empiriques ajustée dont la loi est asymptotiquement khi-carré peut servir à bâtir des intervalles de confiance pour la moyenne ou la fonction de répartition d'une population finie dans le cadre d'enquêtes complexes. Ils considèrent des plans d'échantillonnage stratifiés et non stratifiés pouvant éventuellement inclure de l'information auxiliaire. Ils examinent le comportement d'estimations de la moyenne et de la fonction de répartition en des points précis au moyen de simulations faisant appel à la méthode d'échantillonnage sans remise et à poids inégaux de Rao–Sampford. Ils concluent que les intervalles de confiance fondés sur le rapport de pseudo-vraisemblances empiriques sont supérieurs à ceux qui s'appuient sur l'approximation normale, tant en terme de probabilité de couverture que de taux d'erreur caudale et de longueur moyenne.

1. INTRODUCTION

Owen (1988) introduced a non-parametric likelihood, named empirical likelihood (EL), for the case of independent and identically distributed observations $y_1, \ldots, y_n$ from some distribution $F(\,\cdot\,)$ of Y. By putting probability masses $p_i = \Pr(Y = y_i)$ at the sample points y_i, the empirical likelihood function is defined as $L(F) = \prod_{i=1}^n p_i$ with $p_i > 0$ and $\sum_{i=1}^n p_i = 1$. Without further restrictions on $\boldsymbol{p} = (p_1, \ldots, p_n)^\top$, the empirical distribution function $F_n(t) = \sum_{i=1}^n \hat{p}_i I(y_i \leq t)$ with $\hat{p}_i = n^{-1}$ is the maximum empirical likelihood estimator (MELE) of $F(t)$, where $I(\,\cdot\,)$ is the indicator function, i.e., $I(y_i \leq t) = 1$ if $y_i \leq t$ and $I(y_i \leq t) = 0$ otherwise. The MELE of a scalar parameter $\theta_0 = \theta_0(F)$ is then given by $\hat{\theta}_0 = \theta_0(F_n)$; in particular, $\hat{\theta}_0 = \bar{y}$, the sample mean, if $\theta_0 = \mathrm{E}(Y)$. The MELE under further restrictions on $\boldsymbol{p}$ may also be obtained in a systematic manner, although often involving iterative solutions. A major advantage of the empirical likelihood approach is that it also provides non-parametric confidence intervals on parameters of interest θ_0, similar to parametric likelihood ratio confidence intervals, as shown by Owen (1988). The parameter θ_0 can also be defined as the unique solution of an estimating equation $\mathrm{E}\{g(Y, \theta_0)\} = 0$. For example, $g(Y, \theta_0) = Y - \theta_0$ and $g(Y, \theta_0) = I(Y \leq t) - \theta_0$ give $\theta_0 = \mathrm{E}(Y)$ and $\theta_0 = F(t)$, respectively. A profile likelihood ratio function is then defined as

$$R(\theta) = \max\left\{\prod_{i=1}^n (np_i) \;\Big|\; \sum_{i=1}^n p_i g(y_i, \theta) = 0,\ p_i > 0,\ \sum_{i=1}^n p_i = 1\right\}.$$

Under some mild moment conditions, Owen (1988) first proved for the case of $\theta_0 = \mathrm{E}(Y)$ that $r(\theta_0) = -2\log\{R(\theta_0)\}$ converges in distribution to χ_1^2, a chi-squared random variable with one degree of freedom, as $n \to \infty$. Hence, the $(1-\alpha)$-level empirical likelihood confidence interval on θ_0 is given by $\{\theta \,|\, r(\theta) < \chi_1^2(\alpha)\}$, where $\chi_1^2(\alpha)$ is the upper α-quantile of χ_1^2, similar to parametric likelihood ratio confidence intervals. Unlike intervals based on a normal approximation, empirical likelihood intervals do not require the evaluation of standard errors of estimators and provide more balanced tail error rates. Moreover, the shape and orientation of empirical likelihood intervals are determined entirely by the data, and the intervals are range preserving and transformation respecting. Owen's (2001) monograph provides an excellent account of the empirical likelihood approach, including extensions to regression models and dependent data.

Historically, the concept of empirical likelihood was first used in survey sampling by Hartley & Rao (1968, 1969) under the name "scale-load" approach. For simple random sampling with a negligible sampling fraction n/N, where N is the finite population size and n is the sample size, the scale-load likelihood is essentially the same as Owen's empirical likelihood. Hartley & Rao (1968) obtained the MELE of the finite population mean $\overline{Y}$ when the population mean $\overline{X}$ of an auxiliary variable, x, is known, and showed that it closely approximates the customary regression estimator of $\overline{Y}$. In the empirical likelihood setup, we use the additional constraint $\sum_{i\in s} p_i x_i = \overline{X}$, where s denotes the sample of fixed size n. Chen & Qin (1993) considered parameters of the form $\theta_0 = N^{-1}\sum_{i=1}^N g(y_i)$ for specified $g(\,\cdot\,)$ and constraints of the form $\sum_{i\in s} p_i a(x_i) = 0$ for known $a(\,\cdot\,)$ and obtained the MELE $\tilde{\theta}_0 = \sum_{i=1}^n \tilde{p}_i g(y_i)$ with positive weights $\tilde{p}_i$; the choice $g(y) = y$ and $a(x) = x - \overline{X}$ gives $\theta_0 = \overline{Y}$ and constraint $\sum_{i\in s} p_i x_i = \overline{X}$. By letting $g(y) = I(y \le t)$ for fixed t, we get $\widetilde{F}(t) = \sum_{i\in s} \tilde{p}_i I(y_i \le t)$ as the MELE of the finite population distribution function $\theta_0 = F(t)$. Note that $\widetilde{F}(t)$ is confined within the interval $[0\,,1]$ and non-decreasing in t, unlike the estimator of $F(t)$ based on the customary regression weights, and hence it can be used to obtain the MELE of population quantiles, in particular the population median. Zhong & Rao (1996, 2000) obtained the MELE of $\overline{Y}$ under stratified simple random sampling, assuming negligible sampling fractions n_h/N_h in each stratum h and known overall mean $\overline{X}$. They also studied empirical likelihood confidence intervals on $\overline{Y}$ by adjusting the empirical log-likelihood ratio statistic to account for within-strata sampling fractions and then showing that the adjusted statistic is asymptotically distributed as χ_1^2; the adjustment factor reduces to $1 - n/N$ in the special case of proportional allocation, $n_h/n = N_h/N$, where n is the total sample size.

It is not easy to obtain an empirical likelihood under general sampling designs. Because of this difficulty, Chen & Sitter (1999) proposed an alternative approach based on a pseudo-empirical log-likelihood (PELL) function. The finite population is regarded as a random sample from an infinite superpopulation, leading to the "census" log-likelihood $l_N(\boldsymbol{p}) = \sum_{i=1}^N \log(p_i)$. The Horvitz–Thompson (HT) estimator

$$\hat{l}_{\mathrm{HT}}(\boldsymbol{p}) = \sum_{i\in s} d_i \log(p_i) \tag{1}$$

of $l_N(\boldsymbol{p})$ is then used as a PELL, where $d_i = \pi_i^{-1}$ are the design weights and π_i are the inclusion probabilities. Maximizing $\hat{l}_{\mathrm{HT}}(\boldsymbol{p})$ subject to $p_i > 0$ and $\sum_{i\in s} p_i = 1$ leads to the pseudo-MELE of $\overline{Y}$ as $\widehat{\overline{Y}}_H = \sum_{i\in s} \hat{p}_i y_i = \sum_{i\in s} \tilde{d}_i(s) y_i$, the Hajek estimator, where $\tilde{d}_i(s) = d_i / \sum_{i\in s} d_i$ are the normalized design weights for the given sample, s. However, the Hajek estimator is significantly less efficient than the Horvitz–Thompson estimator $\widehat{\overline{Y}}_{\mathrm{HT}} = N^{-1}\sum_{i\in s} d_i y_i$ under unequal probability sampling without replacement with inclusion probabilities π_i proportional to known size measures z_i when y_i is approximately proportional to z_i. We propose an improved estimator of $\overline{Y}$ in the latter case. (See Section 2.1 for details.) Chen & Sitter (1999) also studied the case of known population mean $\overline{\boldsymbol{X}}$ of auxiliary variables $\boldsymbol{x}$, and obtained a pseudo-MELE of $\overline{Y}$ by imposing the additional constraint $\sum_{i\in s} p_i \boldsymbol{x}_i = \overline{\boldsymbol{X}}$. This estimator, $\sum_{i\in s} \tilde{p}_i y_i$, is closely approximated by a generalized regression (GREG) estimator of $\overline{Y}$, but unlike the latter it uses positive weights $\hat{p}_i$. As a result, the pseudo-MELE of $F(t)$ is non-decreasing and is itself a genuine distribution function. Chen, Sitter & Wu (2002) and Wu (2004) have given simple and efficient algorithms for computing the weights $\hat{p}_i$ used in pseudo-MELE.

In the above mentioned work on the empirical likelihood method in survey sampling, the focus was on point estimation, excepting the work of Zhong & Rao (2000) where empirical likelihood confidence intervals under stratified simple random sampling were examined. In this article, our primary aim is to use the pseudo-empirical likelihood method for constructing confidence intervals on $\overline{Y}$ and $F(t)$ under unequal probability sampling without replacement. In Section 2, we consider both non-stratified sampling and stratified sampling, and formulate an alternative PELL that allows for a simple adjustment for the "design-effect". This formulation does not change the point estimators, but ensures that the resulting pseudo empirical log-likelihood ratio function, adjusted for the corresponding design-effect, is asymptotically distributed as χ^2_1, as shown in Section 3. These asymptotic results require special technical treatments of complex sampling designs and do not follow from standard empirical likelihood intervals or Chen & Sitter (1999). The involved design-effect depends not only on the sampling design but also on the set of additional constraints involving auxiliary information. Asymptotic results and the derivation of the design-effect under stratified sampling are greatly simplified by using a reformulation technique. Finite sample performance of the proposed pseudo-empirical likelihood confidence intervals on $\overline{Y}$ and $F(t)$ is examined through an extensive simulation study reported in Section 4. Our results show that the proposed intervals perform better than the intervals based on normal approximation in providing balanced tail error rates and improved coverage probabilities, particularly for the finite population distribution function $F(t)$. Section 5 contains some additional remarks. Proofs are relegated to the Appendix.

For asymptotic development, we assume that there is a sequence of finite populations indexed by ν such that the population size N_ν and the sample size n_ν both tend to infinity as $\nu \to \infty$. For stratified sampling designs, we assume that the total number of strata L is fixed and the stratum sample sizes all tend to infinity as $\nu \to \infty$. The index ν will be suppressed for notational simplicity. All limiting processes are under $\nu \to \infty$.

2. PSEUDO-EMPIRICAL LIKELIHOOD ESTIMATORS

2.1. Non-stratified sampling.

For non-stratified unistage sampling designs with fixed sample size n, the pseudo-empirical log-likelihood (PELL) function used in this article is defined as

$$l_{ns}(\boldsymbol{p}) = n \sum_{i \in s} \tilde{d}_i(s) \log(p_i), \tag{2}$$

where $\tilde{d}_i(s)$ are the normalized design weights as defined in Section 1. If the design weights d_i are all equal, then (2) reduces to the usual empirical log-likelihood $\sum_{i\in s} \log(p_i)$. The function (2) differs from (1) used by Chen & Sitter (1999) in that $\tilde{d}_i(s)$ is used instead of d_i, but maximizing (2) subject to a set of constraints on the p_is is equivalent to maximizing (1) subject to the same set of constraints.

In the absence of auxiliary population information, the pseudo-MELE of the parameter $\theta_0 = N^{-1}\sum_{i=1}^N g(y_i)$ is given by the Hajek estimator $\hat{\theta}_H = \sum_{i\in s} \hat{p}_i g(y_i) = \sum_{i\in s} \tilde{d}_i(s) g(y_i)$. Estimators $\widehat{\overline{Y}}_H$ and $\widehat{F}_H(t)$ are obtained from $\hat{\theta}_H$ by letting $g(y_i) = y_i$ and $g(y_i) = I(y_i \le t)$ respectively. As noted in Section 1, $\widehat{\overline{Y}}_H$ will be significantly less efficient than the Horvitz–Thompson estimator $\widehat{\overline{Y}}_{\rm HT}$ when $\pi_i \propto z_i$ and y_i and z_i are closely related. On the other hand, $\widehat{F}_H(t)$ will be efficient and more attractive than the Horvitz–Thompson estimator $\widehat{F}_{\rm HT}(t) = N^{-1}\sum_{i\in s} d_i I(y_i \le t)$ because $I(y_i \le t)$ and z_i are not closely related and, unlike $\widehat{F}_{\rm HT}(t)$,

$\widehat{F}_H(t)$ is itself a distribution function. Estimates of population quantiles can therefore be obtained through direct inversion of $\widehat{F}_H(t)$.

We now provide an improved pseudo-MELE of $\overline{Y}$ when $\pi_i \propto z_i$ by introducing the additional constraint

$$\sum_{i \in s} p_i z_i = \bar{Z}, \tag{3}$$

where $\overline{Z}$ is the known population mean of the size measures z_i. Note that (3) is equivalent to $\sum_{i \in s} p_i \pi_i = n/N$. Maximizing (2) subject to (3) and $\sum_{i \in s} p_i = 1$ gives the improved pseudo-MELE $\widehat{\overline{Y}}_E = \sum_{i \in s} \hat{p}_i y_i$. Note that the variance of $\widehat{\overline{Y}}_E$ is zero when $y_i \propto z_i$, unlike the Hajek estimator $\widehat{\overline{Y}}_H$, thus showing improved efficiency when $\pi_i \propto z_i$ and y_i and z_i are closely related. It follows from Chen & Sitter (1999) that $\widehat{\overline{Y}}_E$ is asymptotically equivalent to the GREG estimator

$$\widehat{\overline{Y}}_G = \widehat{\overline{Y}}_H + \widehat{B}(\bar{Z} - \widehat{\overline{Z}}_H), \tag{4}$$

where $\widehat{B} = \sum_{i \in s} \tilde{d}_i(s)(z_i - \widehat{\overline{Z}}_H) y_i / \sum_{i \in s} \tilde{d}_i(s)(z_i - \widehat{\overline{Z}}_H)^2$ is the weighted estimator of the population regression coefficient. This result holds under the conditions (i) $\max_{i \in s} |z_i - \overline{Z}\,| = o_p(n^{1/2})$ and (ii) $\sum_{i \in s} d_i(z_i - \overline{Z}\,)/ \sum_{i \in s} d_i(z_i - \overline{Z}\,)^2 = O_p(n^{-1/2})$, see Chen & Sitter (1999, p. 390). The asymptotic variance of $\widehat{\overline{Y}}_G$ is equivalent to the asymptotic variance of the Hajek estimator involving the residuals $r_i = y_i - \overline{Y} - B(z_i - \overline{Z}\,)$ which are weakly related to z_i even if y_i and z_i are closely related, where $B = \sum_{i=1}^N (z_i - \overline{Z}\,) y_i / \sum_{i=1}^N (z_i - \overline{Z}\,)^2$ is the population regression coefficient. Hence, the alternative estimator $\widehat{\overline{Y}}_E$ should perform better than $\widehat{\overline{Y}}_H$ in terms of efficiency. Our simulation results reported in Section 4 show that the inclusion of the constraint (3) when y_i and z_i are closely related and $\pi_i \propto z_i$ leads to a shorter pseudo-empirical likelihood confidence interval on $\overline{Y}$ relative to the interval not using (3).

We now turn to the case of known population mean $\overline{\boldsymbol{X}}$ of a vector of auxiliary variables $\boldsymbol{x}$ related to y. In this case, the pseudo-MELE $\tilde{\theta}_0$ that uses the auxiliary information at the estimation stage is computed as $\sum_{i \in s} \tilde{p}_i g(y_i)$, where the $\tilde{p}_i$ maximize $l_{ns}(\boldsymbol{p})$ subject to

$$\sum_{i \in s} p_i = 1 \quad \text{and} \quad \sum_{i \in s} p_i \boldsymbol{x}_i = \overline{\boldsymbol{X}}. \tag{5}$$

Chen & Sitter (1999) showed that $\tilde{\theta}_0$ is asymptotically equivalent to the GREG estimator of θ_0, but unlike the latter estimator, it uses positive weights $\hat{p}_i$. Note that $\tilde{\theta}_0$ is a calibration estimator in the sense of $\sum_{i \in s} \tilde{p}_i \boldsymbol{x}_i = \overline{\boldsymbol{X}}$, a set of equations often referred to as calibration equations or benchmark constraints. Maximizing $l_{ns}(\boldsymbol{p})$ subject to (5) gives $\tilde{p}_i = \tilde{d}_i(s)/\{1 + \boldsymbol{\lambda}^\top(\boldsymbol{x}_i - \overline{\boldsymbol{X}}\,)\}$ where the vector-valued Lagrange multiplier $\boldsymbol{\lambda}$ is the solution to

$$\sum_{i \in s} \frac{\tilde{d}_i(s)(\boldsymbol{x}_i - \overline{\boldsymbol{X}}\,)}{1 + \boldsymbol{\lambda}^\top(\boldsymbol{x}_i - \overline{\boldsymbol{X}}\,)} = \mathbf{0}. \tag{6}$$

Chen, Sitter & Wu (2002) showed that the solution exists and is unique if $\overline{\boldsymbol{X}}$ is an inner point of the convex hull formed by $\{\boldsymbol{x}_i,\ i \in s\}$. They proposed an efficient algorithm for solving (6). Wu (2005) has implemented the algorithm in R/S-PLUS.

2.2. Stratified sampling.

For stratified unistage designs with samples of fixed sizes n_h drawn independently from each of the L strata, the pseudo-empirical log-likelihood (PELL) is defined as

$$l_{st}(\boldsymbol{p}_1, \ldots, \boldsymbol{p}_L) = n \sum_{h=1}^{L} W_h \sum_{i \in s_h} \tilde{d}_{hi}(s_h) \log(p_{hi}), \tag{7}$$

where $\tilde{d}_{hi}(s_h) = d_{hi}/\sum_{i\in s_h} d_{hi}$ are the normalized weights within strata with $d_{hi} = \pi_{hi}^{-1}$ denoting the design weights, s_h is the set of sample units in stratum h, $W_h = N_h/N$ are the known stratum weights with $\sum_{h=1}^{L} N_h = N$, $n = \sum_{h=1}^{L} n_h$ is the total sample size, and $\boldsymbol{p}_h = (p_{h1}, \ldots, p_{hn_h})^\top$ subject to $\sum_{i\in s_h} p_{hi} = 1$. Note that $l_{st}(\boldsymbol{p}_1, \ldots, \boldsymbol{p}_L)$ does not reduce to the empirical log-likelihood function $\sum_{h=1}^{L}\sum_{i\in s_h}\log(p_{hi})$ under stratified simple random sampling (Zhong & Rao 2000) unless $n_h = nW_h$ (proportional allocation).

In the absence of auxiliary information, maximizing (7) subject to

$$\sum_{i\in s_h} p_{hi} = 1, \quad h = 1, \ldots, L, \tag{8}$$

gives $\hat{p}_{hi} = \tilde{d}_{hi}(s_h)$. The pseudo-MELE of $\theta_0 = N^{-1}\sum_{h=1}^{L}\sum_{i=1}^{N_h} g(y_{hi}) = \sum_{h=1}^{L} W_h\theta_h$ is given by $\hat{\theta}_{H(st)} = \sum_{h=1}^{L} W_h\hat{\theta}_{Hh}$, where $\theta_h = N_h^{-1}\sum_{i=1}^{N_h} g(y_{hi})$ and $\hat{\theta}_{Hh} = \sum_{i\in s_h}\tilde{d}_{hi}(s_h)$ $g(y_{hi})$ is the Hajek estimator of θ_h. Again, we can get an improved pseudo-MELE of $\overline{Y}$, when the inclusion probabilities $\pi_{hi} \propto z_{hi}$ and y_{hi} and z_{hi} are closely related, by introducing the additional constraint

$$\sum_{i\in s_h} p_{hi} z_{hi} = \overline{Z}_h, \quad h = 1, \ldots, L, \tag{9}$$

where $\overline{Z}_h$ is the known stratum population mean of the size measures z_{hi}. Since the constraints (8) and (9) are separable for each h, maximizing (7) subject to (8) and (9) is equivalent to maximizing each component $\sum_{i\in s_h}\tilde{d}_{hi}(s_h)\log(p_{hi})$ of (7) separately. The resulting improved pseudo-MELE $\widehat{\widehat{\overline{Y}}}_{E(st)} = \sum_{h=1}^{L} W_h \sum_{i\in s_h}\hat{p}_{hi}y_i$ is asymptotically equivalent to a separate GREG estimator of the form $\widehat{\widehat{\overline{Y}}}_{G(st)} = \sum_{h=1}^{L} W_h\widehat{\widehat{\overline{Y}}}_{Gh}$, where $\widehat{\widehat{\overline{Y}}}_{Gh}$ is obtained from (4) for each stratum h.

We now turn to the case of known overall population mean $\overline{\boldsymbol{X}} = \sum_{h=1}^{L} W_h\overline{\boldsymbol{X}}_h$ of auxiliary variables $\boldsymbol{x}$ related to y, where the strata means $\overline{\boldsymbol{X}}_h$ are not known. In this case, the pseudo-MELE $\tilde{\theta}_0 = \sum_{h=1}^{L} W_h\sum_{i\in s_h}\tilde{p}_{hi}g(y_{hi})$ is obtained by maximizing (7) subject to

$$\sum_{i\in s_h} p_{hi} = 1,\ h = 1, \ldots, L \quad \text{and} \quad \sum_{h=1}^{L} W_h \sum_{i\in s_h} p_{hi}\boldsymbol{x}_{hi} = \overline{\boldsymbol{X}}\,. \tag{10}$$

To compute the $\tilde{p}_{hi}$, we follow Wu (2004) and reformulate the constraints (10) as

$$\sum_{h=1}^{L} W_h \sum_{i\in s_h} p_{hi} = 1 \quad \text{and} \quad \sum_{h=1}^{L} W_h \sum_{i\in s_h} p_{hi}\boldsymbol{x}_{hi}^* = \overline{\boldsymbol{X}}^*, \tag{11}$$

where $\overline{\boldsymbol{X}}^*$ is obtained by augmenting $\overline{\boldsymbol{X}}$ to include $W_1, \ldots, W_{L-1}$ as its first $L-1$ components and $\boldsymbol{x}_{hi}^*$ is obtained by augmenting $\boldsymbol{x}_{hi}$ to include the first $L-1$ stratum indicator variables. This reformulation makes all the steps for maximization under non-stratified sampling applicable to stratified sampling. The difference between the two cases is simply a matter of single or double summation. Maximizing (7) subject to (10), or equivalently (11), gives $\tilde{p}_{hi} = \tilde{d}_{hi}(s_h)/(1 + \boldsymbol{\lambda}^\top\boldsymbol{u}_{hi})$ where $\boldsymbol{u}_{hi} = \boldsymbol{x}_{hi}^* - \overline{\boldsymbol{X}}^*$ and the vector-valued $\boldsymbol{\lambda}$ is the solution to

$$\sum_{h=1}^{L} W_h \sum_{i\in s_h} \frac{\tilde{d}_{hi}(s_h)\boldsymbol{u}_{hi}}{1 + \boldsymbol{\lambda}^\top\boldsymbol{u}_{hi}} = \boldsymbol{0},$$

which can be solved in a similar way to (6) for non-stratified sampling designs. R/S-PLUS codes for doing this can be found in Wu (2005).

3. PSEUDO-EMPIRICAL LIKELIHOOD RATIO CONFIDENCE INTERVALS

In this section we show that the pseudo-empirical log-likelihood ratio functions associated with (2) and (7), adjusted to the respective design effects, are asymptotically distributed as χ_1^2 under certain regularity conditions. These results lead to pseudo-empirical likelihood ratio confidence intervals on θ_0, in particular on $\overline{Y}$ and $F(t)$.

3.1. Non-stratified sampling.

We first study the case of no auxiliary population information at the estimation stage. Maximizing $l_{ns}(\boldsymbol{p})$ given by (2), subject to $p_i > 0$ and $\sum_{i\in s} p_i = 1$ gives $\hat{p}_i = \tilde{d}_i(s)$. Let $\hat{p}_i(\theta)$ be the value of p_i obtained by maximizing (2) subject to

$$\sum_{i\in s} p_i = 1 \quad \text{and} \quad \sum_{i\in s} p_i g(y_i) = \theta \tag{12}$$

for a fixed θ. The solution to this constrained maximization problem exists and is unique when θ is an inner point of the convex hull formed by $\{g(y_i),\ i \in s\}$.

The pseudo-empirical log-likelihood ratio function is given by

$$r_{ns}(\theta) = -2\big\{l_{ns}(\hat{\boldsymbol{p}}(\theta)) - l_{ns}(\hat{\boldsymbol{p}})\big\}. \tag{13}$$

We use the following regularity conditions for studying the asymptotic distribution of $r_{ns}(\theta)$ as $n \to \infty$. For simplicity, we consider the case of $g(y_i) = y_i$, but similar results hold for general $g(y_i)$ with suitable conditions on $g(\cdot)$.

- C1 The sampling design $p(s)$ and the study variable y satisfy $\max_{i\in s}|y_i| = o_p(n^{1/2})$, where the stochastic order $o_p(\cdot)$ is with respect to the sampling design $p(s)$.
- C2 The sampling design $p(s)$ satisfies $N^{-1}\sum_{i\in s} d_i - 1 = O_p(n^{-1/2})$.
- C3 The Horvitz–Thompson estimator $\hat{\theta}_{\mathrm{HT}} = N^{-1}\sum_{i\in s} d_i y_i$ of $\theta_0 = \overline{Y}$ is asymptotically normally distributed.

Condition C1 imposes some restrictions on both the sampling design $p(s)$ and the finite population $\{y_1,\ldots,y_N\}$. If the finite population satisfies C1*: $\max\{|y_1|,\ldots,|y_N|\} = o(N^{1/2})$, then Condition C1 holds for any sampling design such that the sampling fraction $n/N \to f \neq 0$. If the population values $\{y_1,\ldots,y_N\}$ can be viewed as an independent and identically distributed sample from a random variable Y, then a sufficient condition for C1* is $\mathrm{E}(Y^2) < \infty$ (Owen 2001, Lemma 11.2). Under the condition that $N^{-1}\sum_{i=1}^{N} y_i^4 = O(1)$, as Chen & Sitter (1999) showed, Condition C1 holds for probability proportional to size (PPS) sampling with replacement, for the Rao–Hartley–Cochran method of PPS sampling without replacement, and for cluster sampling where the clusters are sampled with PPS and with replacement.

Condition C2 states that $\widehat{N} = \sum_{i\in s} d_i$ is a $\sqrt{n}$-consistent estimator of N. Under simple random sampling, stratified random sampling and single stage cluster sampling where clusters are sampled with probability proportional to size, we have $\widehat{N} = N$ which further implies C2.

Condition C3 is the central limit theorem for a Horvitz–Thompson estimator. Hajek (1960, 1964) established the asymptotic normality of $\widehat{\overline{Y}}_{\mathrm{HT}}$ under simple random sampling and rejective sampling with unequal selection probabilities. Visek (1979) established the asymptotic normality of $\widehat{\overline{Y}}_{\mathrm{HT}}$ for the well-known Rao–Sampford method of unequal probability sampling without replacement. Note that Condition C3 is also required for the conventional confidence intervals based on normal theory.

The design effect (abbreviated deff) associated with $\widehat{\overline{Y}}_H$ is defined as

$$\mathrm{deff}_H = V_p(\widehat{\overline{Y}}_H)/(S_y^2/n), \tag{14}$$

where S_y^2 is the population variance, S_y^2/n is the variance of $\widehat{\widehat{Y}}_H$ under simple random sampling (ignoring the finite population correction factor) and $V_p(\cdot)$ denotes the variance under the specified design $p(s)$. Using deff_H, we can also define the effective sample size as $n_e = n/\text{deff}_H$.

THEOREM 1. *Under the conditions C1–C3, the adjusted pseudo-empirical log-likelihood ratio statistic*

$$r_{ns}^{[a]}(\theta) = \{r_{ns}(\theta)\}/\text{deff}_H \tag{15}$$

is asymptotically distributed as χ_1^2 when $\theta = \overline{Y}$.

Note that deff_H is associated with $\widehat{\widehat{Y}}_H$, and it will change with the choice of $g(y_i)$ and/or with additional constraints on auxiliary variables. A result similar to Theorem 1 is obtained when the constraint $\sum_{i\in s} p_i z_i = \overline{Z}$ is used and the inclusion probabilities $\pi_i \propto z_i$. In this case the design effect is associated with the GREG estimator $\widehat{\widehat{Y}}_G$ given by (4). The result that the adjusted pseudo-EL ratio statistic is asymptotically χ_1^2 under this additional constraint follows as a special case of Theorem 2 below by changing $\boldsymbol{x}_i$ to z_i.

When $\overline{\boldsymbol{X}}$ is known and the calibration constraint $\sum_{i\in s} p_i \boldsymbol{x}_i = \overline{\boldsymbol{X}}$ is used, then the pseudo-empirical log-likelihood ratio function, adjusted by the design effect associated with the GREG estimator of $\overline{Y}$, is asymptotically χ_1^2 when $\theta = \overline{Y}$, as shown in Theorem 2 below. The GREG "estimator" of $\overline{Y}$, defined here for the purpose of theoretical development but not computable from the sample data, is given by $\widehat{\widehat{Y}}_{\text{GR}} = \widehat{\widehat{Y}}_H + \boldsymbol{B}^\top(\overline{\boldsymbol{X}} - \widehat{\widehat{\boldsymbol{X}}}_H)$, where $\boldsymbol{B}$ is the vector of population regression coefficients defined as

$$\boldsymbol{B} = \left\{\frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{x}_i - \overline{\boldsymbol{X}})(\boldsymbol{x}_i - \overline{\boldsymbol{X}})^\top\right\}^{-1}\left\{\frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{x}_i - \overline{\boldsymbol{X}})(y_i - \overline{Y})\right\}. \tag{16}$$

Note that we used the finite population quantity $\boldsymbol{B}$ instead of a sample based $\widehat{\boldsymbol{B}}$ in defining $\widehat{\widehat{Y}}_{\text{GR}}$. The design effect associated with $\widehat{\widehat{Y}}_{\text{GR}}$ is defined as

$$\text{deff}_{\text{GR}} = V_p(\widehat{\widehat{Y}}_{\text{GR}})/(S_r^2/n), \tag{17}$$

where $V_p(\widehat{\widehat{Y}}_{\text{GR}}) = V_p\{\sum_{i\in s}\tilde{d}_i(s)r_i\}$, $r_i = y_i - \overline{Y} - \boldsymbol{B}^\top(\boldsymbol{x}_i - \overline{\boldsymbol{X}})$, S_r^2/n is the variance of $\widehat{\widehat{Y}}_{\text{GR}}$ under simple random sampling (ignoring $1 - n/N$), and $S_r^2 = (N-1)^{-1}\sum_{i=1}^{N} r_i^2$. We require the following conditions (parallel to C1 and C3) on the auxiliary variables $\boldsymbol{x}_i$, where $\|\cdot\|$ denotes the L_1 norm.

C4 $\max_{i\in s}\|\boldsymbol{x}_i\| = o_p(n^{1/2})$.

C5 $\widehat{\widehat{\boldsymbol{X}}}_{\text{HT}} = N^{-1}\sum_{i\in s} d_i \boldsymbol{x}_i$ is asymptotically normally distributed.

THEOREM 2. *Let $\tilde{\boldsymbol{p}}$ be the maximizer of $l_{ns}(\boldsymbol{p})$ under the constraints $\sum_{i\in s} p_i = 1$ and $\sum_{i\in s} p_i \boldsymbol{x}_i = \overline{\boldsymbol{X}}$; let $\tilde{\boldsymbol{p}}(\theta)$ be obtained by maximizing $l_{ns}(\boldsymbol{p})$ subject to $\sum_{i\in s} p_i = 1$, $\sum_{i\in s} p_i \boldsymbol{x}_i = \overline{\boldsymbol{X}}$ and $\sum_{i\in s} p_i y_i = \theta$ for a fixed θ. Then under the conditions C1–C5, the adjusted pseudo-empirical log-likelihood ratio statistic*

$$r_{ns}^{(a)}(\theta) = \{\tilde{r}_{ns}(\theta)\}/\text{deff}_{\text{GR}} \tag{18}$$

is asymptotically distributed as χ_1^2 when $\theta = \overline{Y}$, where $\tilde{r}_{ns}(\theta) = -2\big\{l_{ns}(\tilde{\boldsymbol{p}}(\theta)) - l_{ns}(\tilde{\boldsymbol{p}})\big\}$.

In practice the design effect deff_H or $\mathrm{deff}_{\mathrm{GR}}$ will need to be estimated from the sample data. The asymptotic χ^2_1 distribution of the adjusted pseudo-empirical likelihood ratio statistics remains unchanged when the design effect is consistently estimated. The design effect $\mathrm{deff}_H = V_p(\widehat{\overline{Y}}_H)/(S^2_y/n)$ can be estimated by $v(\widehat{\overline{Y}}_H)/(\hat{S}^2_y/n)$, where $v(\widehat{\overline{Y}}_H)$ is the linearization estimator of $V_p(\widehat{\overline{Y}}_H)$ given by

$$v(\widehat{\overline{Y}}_H) = \frac{1}{\widehat{N}^2}\sum_{i\in s}\sum_{j>i}\frac{\pi_i\pi_j - \pi_{ij}}{\pi_{ij}}\left(\frac{e_i}{\pi_i} - \frac{e_j}{\pi_j}\right)^2$$

with $\widehat{N} = \sum_{i\in s} d_i$ and $e_i = y_i - \widehat{\overline{Y}}_H$ being the "residual variable" for $\widehat{\overline{Y}}_H$, and $\hat{s}^2_y$ is the unbiased estimator of S^2_y given by

$$\widehat{S}^2_y = \frac{1}{N(N-1)}\sum_{i\in s}\sum_{j>i}\frac{(y_i - y_j)^2}{\pi_{ij}},$$

following the well-known Laplace expression $S^2_y = \sum_{i=1}^N\sum_{j=i+1}^N (y_i - y_j)^2/(N(N-1))$. The design effect $\mathrm{deff}_{\mathrm{GR}} = V_p(\widehat{\overline{Y}}_{\mathrm{GR}})/(S^2_r/n)$ can similarly be estimated by $v(\widehat{\overline{Y}}_{\mathrm{GR}})/(\widehat{S}^2_r/n)$, where $v(\widehat{\overline{Y}}_{\mathrm{GR}})$ has the same format of $v(\widehat{\overline{Y}}_H)$ but uses the "residual variable" $r_i = y_i - \widehat{\overline{Y}}_H - \widehat{\boldsymbol{B}}^\top(\boldsymbol{x}_i - \overline{\boldsymbol{X}})$ which is associated with $\widehat{\overline{Y}}_{\mathrm{GR}}$, and $\widehat{S}^2_r$ is obtained as in $\widehat{S}^2_y$ but uses r_i in place of y_i. Section 4 contains further details on computing the design effects for scenarios considered in the simulation.

Using Theorems 1 and 2, we can construct $(1-\alpha)$-level pseudo-empirical likelihood ratio confidence intervals on $\theta_0 = \overline{Y}$ as $\{\theta \mid r^{[a]}_{ns}(\theta) \le \chi^2_1(\alpha)\}$ for the case of no auxiliary information and $\{\theta \mid r^{(a)}_{ns}(\theta) \le \chi^2_1(\alpha)\}$ for the case of known mean $\overline{\boldsymbol{X}}$ of the auxiliary variables $\boldsymbol{x}_i$, where $\chi^2_1(\alpha)$ is the $(1-\alpha)$-quantile of χ^2_1.

3.2. Stratified sampling.

We first study the case of no auxiliary information at the estimation stage. The $\hat{p}_{hi}$s which maximize $l_{st}(\boldsymbol{p}_1, \ldots, \boldsymbol{p}_L)$ of (7) subject to (8) are given by $\hat{p}_{hi} = \tilde{d}_{hi}(s_h)$. Let $\hat{p}_{hi}(\theta)$ be the value of p_{hi} obtained by maximizing (7) subject to

$$\sum_{i\in s_h} p_{hi} = 1,\ h = 1, \ldots, L \quad \text{and} \quad \sum_{h=1}^L W_h \sum_{i\in s_h} p_{hi} g(y_{hi}) = \theta \tag{19}$$

for a fixed value of θ. The related computational problem can be handled similarly, as in Section 2.2, through an augmented method. The set of constraints (19) is equivalent to

$$\sum_{h=1}^L W_h \sum_{i\in s_h} p_{hi} = 1 \quad \text{and} \quad \sum_{h=1}^L W_h \sum_{i\in s_h} p_{hi} g^*(y_{hi}) = \boldsymbol{\theta}^*, \tag{20}$$

where $\boldsymbol{\theta}^*$ is obtained by augmenting θ to include $W_1, \ldots, W_{L-1}$ as its first $L-1$ components, and $g^*(y_{hi})$ is obtained by augmenting $g(y_{hi})$ to include the first $L-1$ stratum indicator variables.

The pseudo-empirical log-likelihood ratio function is given by

$$r_{st}(\theta) = -2\big\{l_{st}(\hat{\boldsymbol{p}}_1(\theta), \ldots, \hat{\boldsymbol{p}}_L(\theta)) - l_{st}(\hat{\boldsymbol{p}}_1, \ldots, \hat{\boldsymbol{p}}_L)\big\}. \tag{21}$$

To study the asymptotic distribution of $r_{st}(\theta)$, we again consider the case of $g(y_{hi}) = y_{hi}$ and assume that $n_h/n \to f_h \neq 0$ as $n \to \infty$. We assume conditions C1–C3 to hold within each stratum h to avoid further notational changes.

THEOREM 3. *Under the conditions C1–C3 within each stratum h, the adjusted pseudo-empirical log- likelihood ratio statistic*

$$r_{st}^{[a]}(\theta) = \{r_{st}(\theta)\}/\text{deff}_{\text{GR}(st)} \tag{22}$$

is asymptotically distributed as χ_1^2 when $\theta = \overline{Y}$.

The proof of Theorem 3 and the definition of the design effect $\text{deff}_{\text{GR}(st)}$ follow as special cases of Theorem 4 below.

Now consider the case of known overall mean $\overline{\boldsymbol{X}}$ of auxiliary variables $\boldsymbol{x}$. Let $\boldsymbol{x}_{hi}^*$ and $\overline{\boldsymbol{X}}^*$ be the augmented variables as defined in Section 2.2. Let $\tilde{p}_{hi}$ be obtained by maximizing (7) subject to (10), or equivalently (11); let $\tilde{p}_{hi}(\theta)$ be obtained by maximizing (7) subject to (11) and the additional constraint

$$\sum_{h=1}^{L} W_h \sum_{i\in s_h} p_{hi} g(y_{hi}) = \theta$$

for a fixed θ. For simplicity, we again confine ourselves to $g(y_{hi}) = y_{hi}$ in which case the combined GREG estimator of $\overline{Y}$ is given by $\widehat{\overline{Y}}_{\text{GR}(st)} = \widehat{\overline{Y}}_{H(st)} + (\boldsymbol{B}^*)^\top(\overline{\boldsymbol{X}}^* - \widehat{\overline{\boldsymbol{X}}}^*_{H(st)})$, where $\widehat{\overline{Y}}_{H(st)} = \sum_{h=1}^{L} W_h \sum_{i\in s_h} \tilde{d}_{hi}(s_h) y_{hi}$ is the Hajek-type estimator under stratified sampling, $\widehat{\overline{\boldsymbol{X}}}^*_{H(st)}$ is similarly defined using the augmented $\boldsymbol{x}_{hi}^*$. The vector of population regression coefficients $\boldsymbol{B}^*$ is defined similarly to $\boldsymbol{B}$ of (16) but using $\boldsymbol{x}_i^*$ and $\overline{\boldsymbol{X}}^*$. The design effect associated with $\widehat{\overline{Y}}_{\text{GR}(st)}$ is given by

$$\text{deff}_{\text{GR}(st)} = \left\{\sum_{h=1}^{L} W_h^2 V_p\left(\sum_{i\in s} \tilde{d}_{hi}(s_h) r_{hi}\right)\right\} \Big/ \left(\frac{S_r^2}{n}\right),$$

where $r_{hi} = (y_{hi} - \overline{Y}) - (\boldsymbol{B}^*)^\top(\boldsymbol{x}_{hi}^* - \overline{\boldsymbol{X}}^*)$ and $S_r^2 = (N-1)^{-1}\sum_{h=1}^{L}\sum_{i=1}^{N_h} r_{hi}^2$.

THEOREM 4. *Under the conditions C1–C5 within each stratum h, the adjusted pseudo-empirical log-likelihood ratio statistic*

$$r_{st}^{(a)}(\theta) = \{\tilde{r}_{st}(\theta)\}/\text{deff}_{\text{GR}(st)} \tag{23}$$

is asymptotically distributed as χ_1^2 when $\theta = \overline{Y}$, where

$$\tilde{r}_{st}(\theta) = -2\{l_{st}(\tilde{\boldsymbol{p}}_1(\theta), \ldots, \tilde{\boldsymbol{p}}_L(\theta)) - l_{st}(\tilde{\boldsymbol{p}}_1, \ldots, \tilde{\boldsymbol{p}}_L)\}$$

is the pseudo-empirical log-likelihood ratio function under stratified sampling.

Because of the reformulated constraints (11), the proof of Theorem 4 under stratified sampling follows along the lines of the proof of Theorem 2 for non-stratified sampling and hence is omitted for brevity. The detailed proof can be found in an unpublished technical report (Wu & Rao 2004). It also follows that Theorem 3 is a special case of Theorem 4 since in the absence of auxiliary information the augmented variables $\boldsymbol{x}_{hi}^*$ reduce to the first $L-1$ stratum indicator variables and $\overline{\boldsymbol{X}}^*$ reduces to $(W_1, \ldots, W_{L-1})$.

Pseudo-empirical likelihood ratio confidence intervals on $F(t)$ for a given t can be obtained from Theorems 1–4 by simply changing y_i to $I(y_i \le t)$. However, the use of benchmark constraints such as $\sum_{i\in s} p_i \boldsymbol{x}_i = \overline{\boldsymbol{X}}$ in making inferences on $F(t)$ may not be very efficient due to the weak correlation between the indicator variable $I(y_i \le t)$ and $\boldsymbol{x}_i$. If the individual population values $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$ are known, then different benchmark constraints that lead to more efficient inference on $F(t)$ can be used (Chen & Wu 2002).

4. SIMULATION STUDY

In this section we report the results of a simulation study on the relative performance of pseudo-empirical likelihood ratio confidence intervals and intervals based on normal approximation to the usual Z-statistic, i.e., $(\hat{\theta}_0 - \theta_0)/\{v(\hat{\theta}_0)\}^{1/2}$, on the population mean $\overline{Y}$ and the distribution function $F(t)$. The latter also includes population proportions as special cases. In particular, we examine the performance of coverage probabilities, tail error rates and average lengths of confidence intervals under the Rao–Sampford method (Rao 1965, Sampford 1967) of sampling without replacement with inclusion probabilities π_i exactly proportional to size measures z_i. The major features of the Rao–Sampford method include: (a) easy implementation for any sample size n subject to $z_i < 1/n$ to ensure $\pi_i < 1$ for all $i = 1, \ldots, N$; (b) an exact recursive formula for computing the second order inclusion probabilities, π_{ij}, that ensure non-negativity of the Sen–Yates–Grundy-type variance estimators such as $v(\widehat{\overline{Y}}_H)$ or $v(\widehat{\overline{Y}}_{\rm GR})$ due to the property $\pi_i\pi_j - \pi_{ij} > 0$ for all (i, j); and (c) variance of the Horvitz–Thompson estimator is always smaller than the variance of the customary estimator under probability proportional to size sampling with replacement. We study three cases: (i) Non-stratified sampling with no auxiliary information at the estimation stage; (ii) Non-stratified sampling with auxiliary information at the estimation stage; and (iii) Stratified random sampling with disproportionate sample size allocation to ensure unequal π_i across strata. The Rao–Sampford sampling method was used for cases (i) and (ii).

4.1. Case (i).

We generated three finite populations, each of size $N = 800$, from the model (Model I)

$$y_i = \beta_0 + \beta_1 z_i + \sigma\varepsilon_i \tag{24}$$

with $\beta_0 = \beta_1 = 1$, where the z_is follow the standard exponential distribution and $\varepsilon_i \sim \chi_1^2 - 1$ which ensures that $\mathrm{E}(\varepsilon_i) = 0$. A constant number was added to all z_i to eliminate extremely small values of z_i. Three different values of σ were used to reflect a weak, moderate and strong correlation between y and z: $\varrho(y, z) = 0.3$, 0.5 and 0.8. The finite populations so generated remain fixed under repeated simulation runs.

The task of computing the π_{ij}s needed for normal approximation and pseudo-empirical likelihood ratio confidence intervals is very heavy for repeated simulation runs. We used $n = 40$ and 80, corresponding to sampling fractions 5% and 10%, in the simulation. Our simulations were programmed in R/S-PLUS using the algorithms outlined in Wu (2005). The R source codes are available from the authors upon request. We denote the pseudo-empirical likelihood intervals without using the additional constraint (3) on the size measures z_i as EL1 and those with the constraint as EL2. The design effect associated with EL1 for $\theta_0 = \overline{Y}$ is deff$_H$ given by (14) and is estimated by $v(\widehat{\overline{Y}}_H)/(\widehat{S}_y^2/n)$ as outlined in Section 3.1; for EL2, the design effect is computed as $v(\widehat{\overline{Y}}_{\rm GR})/(\widehat{S}_r^2/n)$, where the residual variable is defined as $r_i = y_i - \widehat{\overline{Y}}_H - \widehat{B}(z_i - \overline{Z})$. The estimated regression coefficient is computed as $\widehat{B} = \left\{N^{-1}\sum_{i\in s} d_i(z_i - \widehat{\overline{Z}}_{\rm HT})^2\right\}^{-1}\left\{N^{-1}\sum_{i\in s} d_i(z_i - \widehat{\overline{Z}}_{\rm HT})y_i\right\}$. For the distribution function $F(t)$, y_i is replaced by $I(y_i \le t)$ for estimating the associated design effect.

We used 1000 simulation runs for each sample size n and correlation $\varrho(y, z)$. Table 1 reports the simulated values of coverage probability (CP), lower (L) and upper (U) tail error rates, average length of the interval (AL) and average lower bound (LB) for the 95% confidence intervals on $\overline{Y}$ based on $\widehat{\overline{Y}}_{\rm HT}$ and normal approximation (NA), EL1 and EL2 for $\varrho(y, z) = 0.3$ and 0.8. Wu & Rao (2004) contains simulation results for additional parameter combinations. The results can be summarized as follows: (1) In terms of balanced tail error rates, EL1 and EL2 clearly outperform NA. The latter leads to much smaller L and larger U than the nominal 2.5% rate at each tail. For example, with $n = 80$ and $\varrho(y, z) = 0.3$, L=0.7 and U=6.3 for NA compared to L=2.5 and U=3.8 for EL2; (2) EL2 has better coverage probability than NA with similar average length; EL1 has better coverage probability but the average length is bigger, and indeed much bigger when $\varrho(y,z) = 0.80$: AL=0.51 for EL1 compared to AL=0.33 for EL2 when $n = 80$; (3) EL2 has the largest lower bound (LB) and close to the 2.5% lower tail nominal error rate in all cases, a feature which is desirable in some applications such as audit sampling. The smaller lower tail error rate (L) for NA is associated with the smaller lower bound LB.

TABLE 1: 95% confidence intervals for the population mean (Model I).

ϱ	n	CI	CP	L	U	AL	LB
0.30	40	NA	90.2	0.5	9.3	1.85	5.26
		EL1	92.7	2.3	5.0	1.97	5.37
		EL2	91.4	2.8	5.8	1.86	5.44
	80	NA	93.0	0.7	6.3	1.30	5.55
		EL1	93.4	2.5	4.1	1.38	5.61
		EL2	93.7	2.5	3.8	1.32	5.64
0.80	40	NA	91.7	0.6	7.7	0.48	5.84
		EL1	94.7	2.3	3.0	0.75	5.76
		EL2	92.3	2.7	5.0	0.47	5.89
	80	NA	94.2	1.5	4.3	0.34	5.91
		EL1	94.6	1.8	3.6	0.51	5.85
		EL2	93.8	2.5	3.7	0.33	5.94

It appears from Table 1 that EL2 is the most reliable method. When the correlation between y_i and the size measures z_i is weak (e.g. $\varrho(y,z) = 0.30$), EL1 may be used. It has better coverage probability with small to moderate inflation in length. The EL1 interval is closely related to the Hajek estimator $\widehat{\bar{Y}}_H$. The much increased length of EL1 under strong correlations between y and z is due to the inefficiency of $\widehat{\bar{Y}}_H$ under such situations, as discussed in Section 2.1.

Table 2 reports the simulated values of CP, L, U and AL for the distribution function $F(t)$ with $t = t_p$ and $p = 0.10$, 0.50 and 0.90, where t_p satisfies $F(t_p) = p$ and $\varrho(y,z) = 0.50$. Additional simulation results for $\varrho(y,z) = 0.30$ and 0.80 and for $p = 0.20$ and 0.80 can be found in Wu & Rao (2004). The 95% NA interval is computed as $(\widehat{F}_H(t) - 1.96\{v(\widehat{F}_H(t)\}^{1/2}$, $\widehat{F}_H(t) + 1.96\{v(\widehat{F}_H(t)\}^{1/2})$, with the lower and upper bounds respectively truncated at 0 and 1. It is evident from Table 2 that both EL1 and EL2 perform uniformly better than NA in terms of coverage probability, balanced tail error rates and average length. NA performs well only for the case of $p = 0.50$ where the underlying distribution of $\widehat{F}_H(t)$ is nearly symmetric. The only unsatisfactory case for the pseudo-empirical likelihood method is $n = 40$ and $p = 0.10$ where the sample size is not large enough to handle the extreme quantile t_p and the resulting tail error rates are not balanced. Our simulation results show that EL1 has a good and stable performance for all cases considered. The correlation between $I(y_i \le t)$ and z_i is generally weak and the Hajek estimator $\widehat{F}_H(t)$ performs well regardless of $\varrho(y,z)$. The EL2 interval is shorter when $\varrho(y,z) = 0.80$ but its coverage probabilities are also deteriorated as compared to EL1.

TABLE 2: 95% confidence intervals for the distribution function (Model I, $\varrho(y, z) = 0.5$).

n	p	CI	CP	L	U	AL
40	0.10	NA	86.0	0.5	13.5	0.183
		EL1	96.9	3.1	0.0	0.194
		EL2	97.0	2.8	0.2	0.183
	0.50	NA	94.4	2.6	3.0	0.307
		EL1	95.0	2.7	2.3	0.296
		EL2	94.7	2.9	2.4	0.265
	0.90	NA	88.3	10.8	0.9	0.160
		EL1	92.4	5.3	2.3	0.162
		EL2	91.1	6.6	2.3	0.156
80	0.10	NA	90.7	0.2	9.1	0.134
		EL1	94.1	1.7	4.2	0.134
		EL2	94.5	1.9	3.6	0.127
	0.50	NA	95.3	2.4	2.3	0.212
		EL1	95.5	2.4	2.1	0.208
		EL2	95.4	2.8	1.8	0.187
	0.90	NA	93.9	5.0	1.1	0.116
		EL1	95.2	2.7	2.1	0.115
		EL2	93.5	4.0	2.5	0.110

4.2. Case (ii).

In this case, we generated finite populations, each of size $N = 800$, from the model (Model II)

$$y_i = \beta_0 + \beta_1 z_i + \beta_2 x_i + \sigma\varepsilon_i \tag{25}$$

with $\beta_0 = 1$, and $\beta_1 = 1$, $\beta_2 = 1$ (Model II-1), $\beta_1 = 1$, $\beta_2 = 2$ (Model II-2) and $\beta_1 = 2$ and $\beta_2 = 1$ (Model II-3), where z_i and x_i follow the standard exponential distribution and $\varepsilon_i \sim \chi_1^2 - 1$. The z-variable is used as the design variable for the Rao–Sampford sampling method and the x-variable is used at the estimation stage. Three values of σ were used for each model to reflect weak, moderate and strong multiple correlation between y and $\beta_0 + \beta_1 z + \beta_2 x$, denoted by ϱ. We studied the performance of confidence intervals for $\overline{Y}$ based on the following methods: (a) Normal approximation to the GREG estimator calibrated over $(1, x)$ (GR1); (b) Pseudo-EL interval with benchmark constraint over x (EL1); (c) Normal approximation to the GREG estimator calibrated over $(1, z, x)$ (GR2); and (d) Pseudo-EL interval with benchmark constraints over both z and x (EL2). The design effect for EL1 or EL2 is estimated as $\{v(\widehat{\overline{Y}}_{\rm GR})\}/(\widehat{S}_r^2/n)$, where $\widehat{\overline{Y}}_{\rm GR}$ involves only the x-variable for EL1 and both z and x for EL2.

Table 3 reports the simulated values of CP, L, U and AL for the mean $\overline{Y}$ under Model II-1, based on 1,000 simulation runs for $n = 40$ and 80. When we compare EL1 to GR1, or EL2 to GR2, the pseudo-EL interval is clearly better than the GR interval in terms of coverage probability and balanced tail error rates and is comparable to GR in terms of average length. For example, CP$=94.2$, L$=2.3$, U$=3.5$ and AL$=1.88$ for EL2 compared to CP$=92.5$, L$=1.2$, U$=6.3$ and AL$=1.85$ for GR2 when $\varrho = 0.30$ and $n = 80$. To choose between EL1 and EL2, it appears reasonable to use EL1 for most cases unless the multiple correlation is strong ($\varrho = 0.80$) and the sample size is large ($n = 80$), where EL2 has good coverage probability but is considerably shorter. Results under Models II-2 and II-3 (not reported here) demonstrate similar trends except that the superiority of the pseudo-EL interval over the GR interval is more pronounced in terms of CP under Model II-3.

TABLE 3: 95% confidence intervals for the population mean (Model II-1).

ϱ	n	CI	CP	L	U	AL
0.30	40	GR1	90.5	1.4	8.1	2.68
		EL1	91.2	2.9	5.9	2.72
		GR2	88.5	1.3	10.2	2.59
		EL2	89.9	2.7	7.4	2.60
	80	GR1	94.5	1.2	4.3	1.89
		EL1	94.8	2.3	2.9	1.92
		GR2	92.5	1.2	6.3	1.85
		EL2	94.2	2.3	3.5	1.88
0.80	40	GR1	92.3	2.8	4.9	0.85
		EL1	93.6	2.7	3.7	0.87
		GR2	89.3	1.2	9.5	0.63
		EL2	90.8	2.1	7.0	0.63
	80	GR1	94.1	1.5	4.4	0.59
		EL1	95.0	1.5	3.5	0.60
		GR2	92.8	0.7	6.5	0.45
		EL2	94.2	1.6	4.2	0.45

TABLE 4: 95% confidence intervals for the population mean (Model III: Stratified random sampling.)

ϱ	n_h	CI	CP	L	U	AL
0.30	20	HT	93.4	1.4	5.2	12.22
		EL1	94.4	2.7	2.9	12.47
		GR	93.7	1.2	5.1	11.98
		EL2	94.5	2.7	2.8	12.13
	40	HT	92.7	1.7	5.6	8.51
		EL1	92.8	3.2	4.0	8.65
		GR	92.1	1.5	6.4	8.38
		EL2	93.5	2.5	4.0	8.49
0.80	20	HT	94.6	1.3	4.1	3.10
		EL1	94.7	2.6	2.7	3.16
		GR	94.0	2.3	3.7	2.52
		EL2	94.5	3.1	2.4	2.27
	40	HT	93.6	2.4	4.0	2.13
		EL1	95.1	2.7	2.2	2.16
		GR	93.2	1.9	4.9	1.77
		EL2	93.3	2.3	4.4	1.59

4.3. Case (iii).

For stratified random sampling, we generated finite populations, each consisting of $L = 4$ strata with strata sizes $N_1 = 800$, $N_2 = 600$, $N_3 = 400$ and $N_4 = 200$, from the model (Model III)

$$y_{hi} = \alpha_h + \beta_h x_{hi} + \sigma \varepsilon_{hi}. \tag{26}$$

The x_{hi}s were generated from $\exp(\lambda_h)$ with the parameter $\lambda_h = 1/h$ and $\varepsilon_{hi} \sim \chi_1^2 - 1$. The regression coefficients in (26) were fixed as $\alpha_h = 2h$ and $\beta_h = h$. Note that $\mathrm{E}(X) = 1/\lambda$ and $\mathrm{var}(X) = 1/\lambda^2$ if $X \sim \exp(\lambda)$, under the current setting, smaller strata ($h = 3, 4$) have larger population means and variances, a scenario often seen in survey practice. We used three values of σ such that the overall population correlations between y and x, $\varrho(y, x)$, are 0.3, 0.5 and 0.8. Equal sample size allocations were used so that the sampling fractions, n_h/N_h, are larger for smaller strata, resulting in unequal selection probabilities across strata.

We studied the performance of confidence intervals on $\overline{Y}$ based on (a) $\widehat{\overline{Y}}_{\mathrm{HT}}$ and normal approximation (HT); (b) pseudo-EL without using the x-variable (EL1); (c) GREG estimator calibrated over x at the population level and normal approximation (GR); and (d) Pseudo-EL with benchmark constraint on x at the population level (EL2). Table 4 reports the simulation results for $n_h = 20$ and 40. When we compare EL1 to HT, or EL2 to GR, the pseudo-EL interval has more balanced tail error rates and improved coverage probability with similar average length. For example, CP$=94.4$, L$=2.7$, U$=2.9$ and AL$=12.47$ for EL1 compared to CP$=93.4$, L$=1.4$, U$=5.2$ and AL$=12.22$ for HT when $\varrho = 0.30$ and $n_h = 20$. EL2 is better than EL1 in terms of AL since EL2 calibrates over x at the population level and is also comparable to EL1 in terms of CP.

5. SOME ADDITIONAL REMARKS

In survey sampling, confidence intervals are customarily constructed based on normal approximations. The performance of such intervals is often unsatisfactory when the underlying distribution is skewed and/or the parameter is confined within a restricted range. The pseudo-empirical likelihood ratio confidence interval proposed in this paper can be an attractive alternative approach. The orientation of the pseudo-EL interval is automatically determined by the data and the range of the parameter space is fully respected. It has similar performance to the normal interval when the latter is satisfactory and performs better otherwise in terms of balanced tail error rates and coverage probabilities. Unlike the EL method in other areas of statistics, the pseudo-EL ratio function for complex surveys requires adjustment to reflect features of the sampling design and the use of auxiliary information at the estimation stage. Finding the pseudo-EL interval in the form of $\{\theta \,|\, r^{(a)}(\theta) \le \chi_1^2(\alpha)\}$ involves profile analysis. For high dimensional problems this is a daunting task. When θ_0 is a scalar, such as the population mean $\overline{Y}$ or the distribution function $F(t)$ for a specified t, the lower and upper bounds of this interval can be found through a simple bisection search method as outlined in Wu (2005).

Our main results can also be modified, using general design weights, to cover unequal probability sampling *with replacement* and the Rao–Hartley–Cochran probability proportional to size sampling method. The asymptotic χ^2 distributions established in Theorems 1–4, however, do not cover cases where the response variable is not a scalar and therefore cannot be used directly to construct confidence regions for vector-valued population means.

APPENDIX: PROOFS

Proof of Theorem 1. Let $\theta = \overline{Y}$ (i.e., θ_0). Using the standard Lagrange multiplier argument, the $\hat{p}_i(\theta)$ which maximize $l_{ns}(\boldsymbol{p})$ subject to (12) are given by $\hat{p}_i(\theta) = \tilde{d}_i(s)/\{1 + \lambda(y_i - \overline{Y}\,)\}$, with the λ being the solution to

$$\sum_{i \in s} \frac{\tilde{d}_i(s)(y_i - \overline{Y}\,)}{1 + \lambda(y_i - \overline{Y}\,)} = 0. \tag{27}$$

By rewriting $\tilde{d}_i(s)(y_i - \overline{Y}\,)$ as $\tilde{d}_i(s)(y_i - \overline{Y}\,)\big[1 + \lambda(y_i - \overline{Y}\,) - \lambda(y_i - \overline{Y}\,)\big]$, we can rearrange (27) to obtain

$$\lambda \sum_{i \in s} \frac{\tilde{d}_i(s)(y_i - \overline{Y}\,)^2}{1 + \lambda(y_i - \overline{Y}\,)} = \sum_{i \in s} \tilde{d}_i(s) y_i - \overline{Y}\,. \tag{28}$$

It follows from (28) that

$$\frac{|\lambda|}{1+|\lambda|u^*}\sum_{i\in s}\tilde{d}_i(s)(y_i-\overline{Y})^2 \le \left|\sum_{i\in s}\tilde{d}_i(s)y_i-\overline{Y}\right|, \tag{29}$$

where $u^*=\max_{i\in s}|y_i-\overline{Y}|$ which is of order $o_p(n^{1/2})$ by condition C1. Under conditions C2 and C3, we have $\widehat{\overline{Y}}_{\rm HT}=\overline{Y}+O_p(n^{-1/2})$ and $\widehat{N}/N=1+O_p(n^{-1/2})$, where $\widehat{N}=\sum_{i\in s}d_i$, which imply $\sum_{i\in s}\tilde{d}_i(s)y_i=\widehat{\overline{Y}}_{\rm HT}/(\widehat{N}/N)=\overline{Y}+O_p(n^{-1/2})$. Noting that $\sum_{i\in s}\tilde{d}_i(s)(y_i-\overline{Y})^2$ is the Hajek-type estimator of S_y^2 which is of order $O(1)$, it follows from (29) that we must have $\lambda=O_p(n^{-1/2})$ and, consequently, $\max_{i\in s}|\lambda(y_i-\overline{Y})|=o_p(1)$. This together with (28) leads to

$$\lambda=\left\{\sum_{i\in s}\tilde{d}_i(s)(y_i-\overline{Y})^2\right\}^{-1}\left(\sum_{i\in s}\tilde{d}_i(s)y_i-\overline{Y}\right)+o_p(n^{-1/2}).$$

Using a Taylor series expansion of $\log(1+x)$ at $x=\lambda(y_i-\overline{Y})$ up to the second order, we obtain

$$\begin{aligned} r_{ns}(\overline{Y}) &= 2n\sum_{i\in s}\tilde{d}_i(s)\log\{1+\lambda(y_i-\overline{Y})\} \\ &= n\left(\sum_{i\in s}\tilde{d}_i(s)y_i-\overline{Y}\right)^2 \Big/ \left(\sum_{i\in s}\tilde{d}_i(s)(y_i-\overline{Y})^2\right)+o_p(1). \end{aligned}$$

Since $\sum_{i\in s}\tilde{d}_i(s)(y_i-\overline{Y})^2=S_y^2+o_p(1)$, and $\sum_{i\in s}\tilde{d}_i(s)y_i$ is asymptotically normal with mean $\overline{Y}$ and variance $V_p(\widehat{\overline{Y}}_H)$ under Conditions C2 and C3, the conclusion that the adjusted pseudo-empirical likelihood ratio statistic converges in distribution to χ_1^2 follows immediately since $r_{ns}^{[a]}(\overline{Y})=\left\{\sum_{i\in s}\tilde{d}_i(s)y_i-\overline{Y}\right\}^2/V_p\left\{\sum_{i\in s}\tilde{d}_i(s)y_i\right\}+o_p(1)$.

Proof of Theorem 2. The arguments on the order of magnitude and the asymptotic expansion of the involved Lagrange multiplier are similar to those given in the proof of Theorem 1. There are two crucial arguments, however, which are unique to this proof. The $\tilde{p}_i$ which maximize $l_{ns}(\boldsymbol{p})$ subject to $\sum_{i\in s}p_i=1$ and $\sum_{i\in s}p_i\boldsymbol{x}_i=\overline{\boldsymbol{X}}$ are given by $\tilde{p}_i=\tilde{d}_i(s)/\left\{1+\boldsymbol{\lambda}^\top(\boldsymbol{x}_i-\overline{\boldsymbol{X}})\right\}$, where the $\boldsymbol{\lambda}$ is the solution to (6). Under Conditions C2, C4 and C5, we can show that $\|\boldsymbol{\lambda}\|=O_p(n^{-1/2})$ and

$$\boldsymbol{\lambda}=\left\{\sum_{i\in s}\tilde{d}_i(s)(\boldsymbol{x}_i-\overline{\boldsymbol{X}})(\boldsymbol{x}_i-\overline{\boldsymbol{X}})^\top\right\}^{-1}\left(\sum_{i\in s}\tilde{d}_i(s)\boldsymbol{x}_i-\overline{\boldsymbol{X}}\right)+o_p(n^{-1/2}).$$

With the term $n\sum_{i\in s}\tilde{d}_i(s)\log(\tilde{d}_i(s))$ omitted, we obtain the following asymptotic expansion for $l_{ns}(\tilde{\boldsymbol{p}})$:

$$-\frac{n}{2}\left(\sum_{i\in s}\tilde{d}_i(s)\boldsymbol{x}_i-\overline{\boldsymbol{X}}\right)^\top\left\{\sum_{i\in s}\tilde{d}_i(s)(\boldsymbol{x}_i-\overline{\boldsymbol{X}})(\boldsymbol{x}_i-\overline{\boldsymbol{X}})^\top\right\}^{-1}\left(\sum_{i\in s}\tilde{d}_i(s)\boldsymbol{x}_i-\overline{\boldsymbol{X}}\right)+o_p(1). \tag{30}$$

To obtain a similar expansion for $l_{ns}(\tilde{\boldsymbol{p}}(\overline{Y}))$ where $\tilde{\boldsymbol{p}}(\overline{Y})$ maximize $l_{ns}(\boldsymbol{p})$ subject to

$$\sum_{i\in s}p_i=1,\quad \sum_{i\in s}p_i\boldsymbol{x}_i=\overline{\boldsymbol{X}}\quad\text{and}\quad\sum_{i\in s}p_iy_i=\overline{Y}, \tag{31}$$

our first crucial argument is to reformulate the constrained maximization problem as follows: let $r_i=y_i-\overline{Y}-\boldsymbol{B}^\top(\boldsymbol{x}_i-\overline{\boldsymbol{X}})$ where $\boldsymbol{B}$ is defined by (16). Then the set of constraints (31) is equivalent to

$$\sum_{i\in s}p_i=1,\quad \sum_{i\in s}p_i\boldsymbol{x}_i=\overline{\boldsymbol{X}}\quad\text{and}\quad\sum_{i\in s}p_ir_i=0. \tag{32}$$

With complete parallel development that leads to $l_{ns}(\tilde{\boldsymbol{p}})$ given by (30), maximizing $l_{ns}(\boldsymbol{p})$ subject to (32) leads to the following expansion for $l_{ns}(\tilde{\boldsymbol{p}}(\overline{Y}))$ (with the same term $n\sum_{i\in s}\tilde{d}_i(s)\log(\tilde{d}_i(s))$ omitted):

$$-\frac{n}{2}\left(\sum_{i\in s}\tilde{d}_i(s)\boldsymbol{u}_i-\overline{\boldsymbol{U}}\right)^{\top}\left\{\sum_{i\in s}\tilde{d}_i(s)(\boldsymbol{u}_i-\overline{\boldsymbol{U}})(\boldsymbol{u}_i-\overline{\boldsymbol{U}})^{\top}\right\}^{-1}\left(\sum_{i\in s}\tilde{d}_i(s)\boldsymbol{u}_i-\overline{\boldsymbol{U}}\right)+o_p(1), \quad (33)$$

where $\boldsymbol{u}_i = (\boldsymbol{x}_i^{\top}, r_i)^{\top}$ and $\overline{\boldsymbol{U}} = (\overline{\boldsymbol{X}}^{\top}, 0)^{\top}$. Our second crucial argument is the observation that $\sum_{i=1}^{N}(\boldsymbol{x}_i - \overline{\boldsymbol{X}})r_i = \mathbf{0}$, i.e., the matrix involved in the middle of (33) is an estimate for its population counterpart which is block diagonal. It is straightforward to show that

$$\tilde{r}_{ns}(\overline{Y}) = -2\{l_{ns}(\tilde{\boldsymbol{p}}(\overline{Y})) - l_{ns}(\tilde{\boldsymbol{p}})\} = n\left(\sum_{i\in s}\tilde{d}_i(s)r_i\right)^2 \Big/ \left(\frac{1}{N}\sum_{i=1}^{N}r_i^2\right) + o_p(1). \quad (34)$$

The conclusion of the theorem follows since $\sum_{i\in s}\tilde{d}_i(s)r_i$ is asymptotically normal with mean 0 and variance $V_p\{\sum_{i\in s}\tilde{d}_i(s)r_i\}$ under Conditions C3 and C5.

ACKNOWLEDGEMENTS

The authors would like to thank the Editor, an Associate Editor and two referees for valuable comments which have led to improvements in the presentation of this paper. This research was supported by grants from the Natural Sciences and Engineering Research Council of Canada.

REFERENCES

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

J. Chen & R. R. Sitter (1999). A pseudo empirical likelihood approach to the effective use of auxiliary information in complex surveys. *Statistica Sinica*, 9, 385–406.

J. Chen & C. Wu (2002). Estimation of distribution function and quantiles using the model-calibrated pseudo empirical likelihood method. *Statistica Sinica*, 12, 1223–1239.

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.

J. Hajek (1960). Limiting distributions in simple random sampling from a finite population. *A Magyar Tudományos Akadémia Matematikai Kutató Intézetének Közlényei*, 5, 361–374.

J. Hajek (1964). Asymptotic theory of rejective sampling with varying probabilities from a finite population. *The Annals of Mathematical Statistics*, 35, 1491–1523.

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

H. O. Hartley & J. N. K. Rao (1969). A new estimation theory for sample surveys, II. In *New Developments in Survey Sampling: A Symposium on the Foundations of Survey Sampling held at the University of North Carolina, Chapel Hill, North Carolina* (N. L. Johnson & H. Smith, Jr., eds.), Wiley-Interscience, New York, pp. 147–169.

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/CRC.

J. N. K. Rao (1965). On two simple schemes of unequal probability sampling without replacement. *Journal of the Indian Statistical Association*, 3, 173–180.

M. R. Sampford (1967). On sampling without replacement with unequal probabilities of selection. *Biometrika*, 54, 499–513.

J. A. Víšek (1979). Asymptotic distribution of simple estimate for rejective, Sampford and successive sampling. In *Contributions to Statistics: Jaroslav Hájek Memorial Volume* (J. Jurecková, ed.), Academia, Prague & D. Reidel, Dordrecht, pp. 263–275.

C. Wu (2004). Some algorithmic aspects of the empirical likelihood method in survey sampling. *Statistica Sinica*, 14, 1057–1067.

C. Wu (2005). Algorithms and R codes for the pseudo empirical likelihood method in survey sampling. *Survey Methodology*, 31, 239–243.

C. Wu & J. N. K. Rao (2004). Pseudo empirical likelihood ratio confidence intervals for complex surveys. Working paper 2004-06, Department of Statistics and Actuarial Science, University of Waterloo.

B. Zhong & J. N. K. Rao (1996). Empirical likelihood inference under stratified random sampling using auxiliary information. In *ASA Proceedings of the Section on Survey Research Methods*, American Statistical Association, pp. 798–803.

B. Zhong & J. N. K. Rao (2000). Empirical likelihood inference under stratified random sampling using auxiliary population information. *Biometrika*, 87, 929–938.

Received 1 December 2005
Accepted 22 April 2006

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

J. N. K. RAO: jrao@math.carleton.ca
School of Mathematics and Statistics
Carleton University
Ottawa, ON, Canada K1S 5B6

SSC Statistical Society of Canada
Société Statistique du Canada