A Unified Theory of Empirical Likelihood Confidence Intervals for Survey Data with Unequal Probabilities and Non Negligible Sampling Fractions

Y.G. Berger* O. De La Riva Torres †

Abstract
We propose a new empirical likelihood approach which can be used to construct design-based confidence intervals under unequal probability sampling without replacement. The proposed approach gives confidence intervals which may perform better than standard confidence intervals and pseudo empirical likelihood confidence intervals. They do not rely on variance estimates, re-sampling or linearisation, even when the parameter of interest is not linear. It can be applied to the Horvitz-Thompson estimator, the Hájek estimator or the regression estimator. It can be also used to construct confidence intervals of totals or counts even when the population size is unknown. We also show that the proposed maximum empirical likelihood estimator is asymptotically optimal. It also offers a likelihood-based justification for design-based approaches, such as calibration, used in sample surveys.

Key Words: Calibration, Design-based approach, Estimating equations, Finite population corrections, Hájek estimator, Horvitz-Thompson estimator, Length biased sampling, Regression estimator, Stratification, Unequal inclusion probabilities.

1. Introduction

Let \( U \) be a finite population of \( N \) units; where \( N \) is a fixed quantity which is not necessarily known. Suppose that the population parameter of interest \( \theta_0 \) is the solution of the following estimating equation (e.g. Binder & Kovačević, 1995).

\[
G(\theta) = 0, \quad \text{with} \quad G(\theta) = \sum_{i \in U} g_i(\theta); \tag{1}
\]

where \( g_i(\theta) \) is a function of \( \theta \) and of characteristics of the unit \( i \). This function does not need to be differentiable. Note that \( g_i(\theta) \) and \( \theta_0 \) can be vectors, but for simplicity, we consider that they are scalar. For example, \( \theta_0 \) is population mean \( \mu = N^{-1} \sum_{i \in U} y_i \), when \( g_i(\theta) = y_i - \theta \); where the \( y_i \) are the values of a variable of interest. Other examples are the low income measure and regression coefficients (Binder & Kovačević, 1995; Deville, 1999). In §5, we show how estimating equation can be used to estimate quantiles. The aim of this paper is to derive an empirical likelihood confidence intervals for \( \theta_0 \).

Suppose that we wish to estimate \( \theta_0 \) from the data of a sample \( s \) of size \( n \) selected with a single stage unequal probabilities without replacement sampling design. We consider that the sample size \( n \) is fixed quantity which is not random. We adopt a design-based approach; where the sampling distribution is specified by the sampling design. Let \( \pi_i \) denote the inclusion probability of unit \( i \). An unbiased estimator of the function (1) is given by the following Horvitz & Thompson (1952) estimator.

\[
\tilde{G}_n(\theta) = \sum_{i=1}^{n} \tilde{g}_i(\theta); \tag{2}
\]

*University of Southampton, Southampton Statistical Sciences Research Institute, Southampton, SO17 1BJ, UK. y.g.berger@soton.ac.uk
†University of Southampton, Southampton Social Statistics, Southampton, SO17 1BJ, UK. O.De-La-Riva@soton.ac.uk
where $\sum_{i=1}^{n}$ denotes the sum over the sampled units and $\hat{g}_i(\theta) = g_i(\theta)\pi_i^{-1}$. An estimator $\hat{\theta}$ of $\theta_0$ is the solution of $\hat{G}_\pi(\theta) = 0$. When $g_i(\theta) = y_i - n^{-1}\theta_1$, the solution of $\hat{G}_\pi(\theta) = 0$ is the Horvitz & Thompson (1952) estimator $\hat{Y}_{HT} = \sum_{i=1}^{n} y_i\pi_i^{-1}$ of the population total $Y = \sum_{i \in U} y_i$. When $g_i(\theta) = y_i - \theta N^{-1}$, the solution is the Hájek (1971) ratio estimator $\hat{Y}_H = N\hat{N}_\pi^{-1}\hat{Y}_{HT}$ of $Y$; where $\hat{N}_\pi = \sum_{i=1}^{n}\pi_i^{-1}$. The estimator $\hat{Y}_{HT}$ may not be as efficient as $\hat{Y}_H$ when $y_i$ and $\pi_i$ are correlated (Rao, 1966), which may be the case, for example, with business surveys.

Under the design-based approach, the standard likelihood function is flat and cannot be used for inference (Godambe, 1966). A possible solution is to assume a super-population models which can be used to derive likelihood function (e.g. Chambers et al., 2012). However these models are not always suitable for the production of survey estimates. Hartley & Rao (1968) introduced an empirical likelihood-based approach which does not rely on models. Owen (1988) brought this approach into the mainstream statistics (see also Owen, 2001). Since Chen & Qin (1993) suggested its first application in survey sampling, there have been many recent developments of empirical likelihood based methods in survey sampling (e.g. Rao & Wu, 2009) and adaptive sampling (Salehi et al., 2010).

Standard confidence intervals based upon the central limit theorem can perform poorly when the sampling distribution is not normal. For example, the lower bounds of a confidence interval can be negative even when the parameter of interest is positive. The coverage and the tail errors can be also lower than their intended levels. On the other hand, empirical likelihood confidence intervals may be better in this situation, as empirical likelihood confidence intervals are determined by the distribution of the data (Rao & Wu, 2009) and the range of the parameter space is preserved. Note empirical likelihood confidence intervals have better coverages when the variable of interest is skewed or contains many zeros (Chen et al., 2003) which is common with many surveys and with estimation of domains.

Chen & Sitter (1999) proposed a pseudo empirical likelihood approach which can be used to construct confidence intervals (Wu & Rao, 2006). The pseudo empirical likelihood approach is not entirely appealing from a theoretical point of view, as confidence intervals rely on variance estimates which can difficult to compute. The pseudo empirical log-likelihood ratio function depends on a population parameter (the design effect) which needs to be estimated, incurring an additional variability which may affect the coverage of the confidence intervals. The proposed approach does not rely on variance estimates, or population parameters.

We propose to use an empirical likelihood approach which is different from the pseudo empirical likelihood approach. It can be used to compute confidence intervals of totals or counts even when $N$ is unknown. Confidence intervals for $\hat{Y}_{HT}$ or $\hat{Y}_H$ can be computed, and it allows to take into account of auxiliary information. We show that the empirical likelihood estimator is asymptotically equivalent to an optimal regression estimator (Montanari, 1987). Note that pseudo empirical likelihood estimators are not asymptotically optimal. Wu & Rao (2006) proposed a more efficient pseudo empirical likelihood approach (EL2) when the variable of interest is correlated with the inclusion probabilities. However, this approach cannot be used to estimate totals and count when $N$ is unknown; which is a common situation with social surveys.

The main contribution of this paper is to show that under a series of regularity conditions, the distribution of the proposed empirical log-likelihood ratio function converges to a chi-squared distribution. This property depends on a set of constraints which takes account of the sampling design and the auxiliary variables. We show this property can be used to derive confidence intervals. We also show that the maximum empirical likelihood estimator is asymptotically optimal.

In §2, we define the proposed empirical likelihood function and we show how the pa-
Parameters of the empirical likelihood function can be estimated. In §2.1, we define the empirical likelihood estimators. In §3, we show how to compute non-parametric confidence intervals. In §3.2, we show how the auxiliary variables can be taken into account. In §4, we propose an adjusted empirical log-likelihood ratio function which takes into account of large sampling fractions. In §5, we show how the proposed approach can be used for quantiles. In §6, we show via a series of simulations that the proposed empirical likelihood approach gives better point estimators and confidence intervals, compared to the pseudo empirical likelihood approach.

2. Empirical likelihood approach under unequal probability sampling

Let \{y_1, \ldots, y_n\} denote a set of \(n\) independent and identically distributed values from the population distribution \(F(y) = N^{-1} \sum_{i \in U} \delta\{y_i \leq y]\); where \(y_i\) denotes the values of a variable of interest attached to unit \(i\). As the units are selected with unequal probabilities, we propose to use the length biased sampling approach proposed by Owen (2001, Ch. 6) who showed that under Poisson sampling, the sample distribution is given by (see also Kim, 2009)

\[
F_s(y) = \sum_{i=1}^{n} \pi_i m_i \frac{\delta\{y_i \leq y\}}{\sum_{j=1}^{n} \pi_j m_j},
\]

(3)

where the quantity \(P_i\) is the probability mass of unit \(i\) in the population and the function \(\delta\{A\}\) is the Dirac measure which is equal to one when \(A\) is true and zero otherwise. Let \(m_i = N P_i\) where \(m_i\) is the unit mass of unit \(i\) in the population (e.g. Deville, 1999). Thus (3) reduces to

\[
F_s(y) = \sum_{i=1}^{n} \pi_i m_i \frac{\delta\{y_i \leq y\}}{\sum_{j=1}^{n} \pi_j m_j}.
\]

(4)

Berger & De La Riva Torres (2012b) showed that under conditional Poisson sampling, the sample distribution is also given by (4).

The empirical likelihood function is defined by (see Owen, 2001, p. 7)

\[
L(m) = \prod_{i=1}^{n} [F_s(y_i) - F_s(y^{-})];
\]

(5)

where \(F_s(y^{-}) = \lim_{y \to y^{-}} F_s(y)\). The above definition is usually used in the context of independent and identically distributed observations. Despite the fact that under fixed size sampling designs, we do not have independent and identically distributed observations, we propose to use (5) as an approximation of the real empirical likelihood. Thus, the empirical likelihood function we propose to use is

\[
L(m) = \prod_{i=1}^{n} \left( \frac{\pi_i m_i}{\sum_{j=1}^{n} \pi_j m_j} \right).
\]

(6)

Note that Owen (2001, Ch. 6) and Kim (2009) proposed to use the same empirical likelihood function under Poisson sampling and with probability mass instead of the mass \(m_i\). The aim of this paper is to show that this empirical likelihood function can be used to construct confidence intervals under fixed size sampling designs.

The maximum likelihood estimators of \(m_i\) are the values \(\hat{m}_i\) which maximise the log-empirical likelihood function

\[
\ell(m) = \log(L(m)),
\]

(7)
subject to the constraints $m_i \geq 0$ and

$$\sum_{i=1}^{n} m_i c_i = C;$$  \hspace{1cm} (8)

where $c_i$ is a known $Q \times 1$ vector associated with the $i$-th sampled unit and $C$ is a known $Q \times 1$ vector. We consider that the constraint (8) is such that $\sum_{i=1}^{n} m_i \pi_i = n$ always holds. Note that the vector $C$ is not necessarily a vector of fixed quantities. Hence $C$ can be fixed or random. Possible choices for $c_i$ and $C$ are discussed in §3. Note that the $c_i$ and $C$ cannot be any vectors, as they must obey the regularity conditions given in §2.1.

The constraint (8) resembles the constraint used in calibration (e.g. Huang & Fuller, 1978; Deville & Särndal, 1992). However, we will see in §3 that $C$ is not necessarily a vector of population totals of auxiliary variables.

Deville & Särndal (1992) showed that such minimisation problem has a unique solution which can be calculated by using the Lagrangian function, $Q(m, \eta) = \sum_{i=1}^{n} \log(\pi_i m_i) - n \log(\sum_{i=1}^{n} \pi_i m_i) - \eta' (\sum_{i=1}^{n} m_i c_i - C).$ The values of $m_i$ and $\eta$ which minimise $Q(m, \eta)$ are the solutions of the following set of equations $\partial Q(m, \eta) / \partial m_i = 0$ and $\partial Q(m, \eta) / \partial \eta = 0.$ As (8) is such that $\sum_{i=1}^{n} m_i \pi_i = n,$ the solution is

$$\hat{m}_i = (\pi_i + \eta' c_i)^{-1},$$  \hspace{1cm} (9)

The parameter $\eta$ is such that the constraint (8) holds. This parameter can be computed using an iterative Newton-Raphson procedure. Consider the following $Q \times 1$ vector function of $\eta$, $f(\eta) = \sum_{i=1}^{n} \hat{m}_i c_i.$ A Taylor approximation of $f(\eta)$ in the neighbourhood of an initial guess $\eta_0$ gives

$$\eta \approx \eta_0 - \hat{\Delta}(\eta_0)^{-1} (f(\eta_0) - C),$$  \hspace{1cm} (10)

as the constraint (8) can be re-written as $f(\eta) = C.$ The $Q \times Q$ matrix $\hat{\Delta}(\eta)$ is the following gradient.

$$\hat{\Delta}(\eta) = \partial f(\eta) / \partial \eta = - \sum_{i=1}^{n} c_i c_i' (\pi_i + \eta' c_i)^{-2}.$$  \hspace{1cm} (11)

The recursive formula (10) can be used to compute $\eta.$ For the first iteration, we used $\eta_0 = 0$ which gives a new approximation of $\eta$ using (10). This new approximation is used as a new value for $\eta_0$, which is substituted into (10). We repeat this process until convergence. Note that it is not necessary to know $N$ in order to compute $\eta$ and $\hat{m}_i.$

Note that when $c_i = \pi_i$ and $C = n,$ we have that $\eta_0 = 0$ and $\hat{m}_i = \pi_i^{-1}.$

2.1 Maximum empirical likelihood estimator

The maximum empirical likelihood estimator $\hat{\theta}$ of $\theta_0$ is defined by solution of the following estimating equation.

$$\hat{G}(\theta) = 0, \text{ with } \hat{G}(\theta) = \sum_{i=1}^{n} \hat{m}_i g_i(\theta);$$  \hspace{1cm} (12)

where $\hat{m}_i$ is defined by (9). We assume that the $g_i(\theta)$ are such that $\hat{G}(\theta) = 0$ has a solution.

Note that when $c_i = \pi_i$ and $C = n,$ we have that $\eta_0 = 0$ and $\hat{m}_i = \pi_i^{-1}.$ In this case, $\hat{\theta}$ is the Horvitz & Thompson (1952) estimator $\hat{Y}_{HT}$ when $g_i(\theta) = y_i - n^{-1} \theta \pi_i,$ and $\hat{\theta}$ is the Hájek (1971) ratio estimator $\hat{Y}_{H}$ when $g_i(\theta) = y_i - \theta N^{-1}.$ Wu & Rao (2006, p. 362)
proposed to use the pseudo empirical likelihood estimator with a similar constraint. This gives a pseudo empirical likelihood estimator which is different from $\hat{Y}_{HT}$ and $\hat{Y}_{H}$. In §6, we will compare the proposed approach with the Wu & Rao (2006) empirical likelihood approach via simulation.

In order to derive asymptotic properties of the proposed empirical likelihood approach, it is necessary to define the asymptotic framework and a set of regularity conditions. We use the Hájek (1964) asymptotic framework, which consists in assuming that $d = \sum_{i \in U} \pi_i (1 - \pi_i) \rightarrow \infty$. This assumption implies that $n \rightarrow \infty$ and $N \rightarrow \infty$, as $d < n < N$. The standard empirical likelihood approach (Owen, 1988) assumes that the sampling fraction is negligible ($n/N \rightarrow 0$). However, many surveys (e.g. business surveys) use sampling fractions which are not necessarily negligible. The proposed empirical likelihood approach does not rely on this assumption. The stochastic order $O(\cdot)$, $o(\cdot)$, $O_p(\cdot)$ and $o_p(\cdot)$ are defined according to this asymptotic framework, where the convergence in probability is with respect to the sampling design.

Consider the following regularity conditions.

\begin{align}
N^{-1} \| \hat{C}_\pi - C \| &= O_p(n^{-\frac{1}{2}}), \\
N^{-1} \hat{G}_\pi(\theta_0) &= O_p(n^{-\frac{1}{2}}), \\
nN^{-1} \pi_i^{-1} &= O(1), \\
\max \{ \| e_i \| : i \in s \} &= o_p(n^{\frac{1}{2}}), \\
\max \{ |g_i(\theta)| : i \in s \} &= o_p(n^{\frac{1}{2}}), \\
\| \hat{S} \| &= O_p(1), \\
\| \hat{S}^{-1} \| &= O_p(1), \\
\frac{1}{nN^\tau} \sum_{i=1}^{n} \frac{\| e_i \|^r}{\pi_i^r} &= O_p(n^{-\tau}),
\end{align}

where $\tau \leq 3$.

\begin{equation}
\hat{S} = \frac{n}{N^2} \hat{\Delta}(0), \quad \text{and} \quad \hat{C}_\pi = \sum_{i=1}^{n} \frac{c_i}{\pi_i}. \tag{21}
\end{equation}

The matrix $\hat{\Delta}(0)$ is given by (11) with $\eta = 0$. The quantity $\| A \| = \text{trace}(A'A)^{1/2}$ denotes the Euclidean norm.

The conditions (13) and (14) hold when the central limit theorem holds. For unequal probability sampling, Isaki & Fuller (1982) gave conditions under which (13) holds (see also Krewski & Rao, 1981, p. 1014). The condition (15) was proposed by Krewski & Rao (1981, p. 1014). Chen & Sitter (1999, Appendix 2) showed that the conditions (16) and (17) hold for common unequal probability sampling designs. The matrix $\hat{S}$ is equal to a covariance matrix between totals multiplied by $-n/N^2$. Thus the condition (18) holds when the norm of this covariance matrix variance decreases with rate $n^{-1}$. The condition (19) means that $\| \hat{S} \|$ is larger than a positive lower bound which is similar to the Cramér-Rao lower bound (see also Zhong & Rao, 2000, p. 932). The condition (20) is a Lyapunov-type condition for the existence of moments (e.g. Krewski & Rao, 1981, p. 1014).

Berger & De La Riva Torres (2012b) showed that under these regularity conditions,

\begin{equation}
\hat{G}(\theta) = \hat{G}_\pi(\theta) + \hat{B}'(C - \hat{C}_\pi) + o_p(N), \tag{22}
\end{equation}
where $\hat{B}$ is a vector of regression coefficients defined by

$$\hat{B} = \left( \sum_{i=1}^{n} \frac{1}{\pi_i} \pi_i c_i \right)^{-1} \sum_{i=1}^{n} \frac{1}{\pi_i} g_i(\theta)c_i,$$  \((23)\)

where $\hat{w}_i$ are regression weights. There is a clear analogy between the proposed empirical likelihood approach and calibration (e.g. Huang & Fuller, 1978; Deville & Särndal, 1992), as the function (7) can be viewed as a calibration distance function, and the empirical likelihood estimator is asymptotically equivalent to a calibrated regression estimator (22). The distance functions used in calibration are disconnected from mainstream statistical theory. However, the proposed distance function (7) is clearly related to the concept of likelihood. The advantage of the proposed empirical likelihood approach over standard calibration is the fact that the empirical likelihood function can be used to construct likelihood ratio confidence intervals. The approaches proposed in this section on Survey Research Methods – JSM 2012

Empirical likelihood confidence intervals rely on the following property.

$$\hat{G}_\pi(\theta_0) V[\hat{G}_\pi(\theta_0)]^{-\frac{1}{2}} \rightarrow N(0, 1);$$  \((24)\)

where $V[\hat{G}_\pi(\theta_0)]$ denotes the design-based variance of $\hat{G}_\pi(\theta_0)$.

As $\theta_0$ is a constant, $\hat{G}_\pi(\theta_0)$ is a Horvitz & Thompson (1952) estimator. Hájek (1964), Vísek (1979), Ohlsson (1986), Zhong & Rao (1996) and Berger (1998b) gave regularity conditions for the asymptotic normality of the Horvitz & Thompson (1952) estimator. Based on these evidences, it is reasonable to assume (24), as $E(\hat{G}_\pi(\theta_0)) = G(\theta_0) = 0$. Note that the classical empirical likelihood approach and the pseudo empirical likelihood approach also rely on (24) (e.g. Owen, 1988, p. 242, Owen, 2001, p. 219, Wu & Rao, 2006, p. 364). Note that the distribution of a point estimator of $\theta_0$ is not necessarily normal, and we will not need the normality of the point estimator which is necessary to derive standard confidence intervals.

Let $\hat{m}_i$ be the values which maximise (7) subject to the constraints $m_i \geq 0$ and (8) when $c_i = \pi_i$ and $C = n$. Note that $m_i = \pi_i^{-1}$ in this situation. Hence the empirical
likelihood point estimator is the solution of $\hat{G}_π(θ) = 0$. Let $ℓ(\hat{m})$ be the maximum value of the empirical log-likelihood function.

Let $\hat{m}^*_τ$ be the values which maximise (7) subject to the constraints $m_τ ≥ 0$ and (8) with $c_τ = c_τ^*$ and $C = C^*$, where $c_τ^* = (π_τ, g_τ(θ))^T$ and $C^* = (n, 0)^T$. Let $ℓ(\hat{m}^*, θ)$ be the maximum value of $ℓ$ of the empirical log-likelihood function.

The empirical log-likelihood ratio function is defined by the following function of $θ$.

$$\hat{r}(θ) = 2 \{ℓ(\hat{m}) - ℓ(\hat{m}^*, θ)\}.$$  \hfill (25)

Berger & De La Riva Torres (2012b) showed that

$$\hat{r}(θ_0) = \hat{G}_π(θ_0)^2 \hat{V}_{pps}[\hat{G}_π(θ_0)]^{-1} + α_p(1),$$  \hfill (26)

where $θ_0$ denotes the population parameter to estimate and where $\hat{V}_{pps}[\hat{G}_π(θ_0)]$ is the following $pps$ variance estimator (e.g. Durbin, 1953; Särndal et al., 1992, p. 99).

$$\hat{V}_{pps}[\hat{G}_π(θ_0)] = \sum_{i=1}^n \left(\hat{g}_i(θ_0) - n^{-1}\hat{G}_π(θ_0)\right)^2;$$  \hfill (27)

where $\hat{g}_i(θ) = g_i(θ)/π_i$. When the sampling fractions are negligible, $\hat{V}_{pps}[\hat{G}_π(θ)]$ is a consistent estimator for the variance (Durbin, 1953). Hence the property (24) implies that $\hat{r}(θ)$ follows asymptotically a chi-squared distribution with one degree of freedom, by Slutsky’s theorem. Thus, the $(1 - α)$ level empirical likelihood confidence interval (e.g. Wilks, 1938; Hudson, 1971) for the population parameter $θ_0$ is given by

$$[\min \{θ | \hat{r}(θ) ≤ χ^2_1(α)\}; \max \{θ | \hat{r}(θ) ≤ χ^2_1(α)\}];$$  \hfill (28)

where $χ^2_1(α)$ is the upper $α$-quantile of the chi-squared distribution with one degree of freedom.

Note that $\hat{r}(θ)$ is a convex non-symmetric function with a minimum when $θ$ is the maximum empirical likelihood estimator. This interval can be found using a bijection search method (e.g. Wu, 2005). This involves calculating $\hat{r}(θ)$ for several values of $θ$.

3.1 Empirical likelihood approach for stratified sampling designs

Assume that the sample $s$ is randomly selected by a uni-stage stratified probability sampling design $p(s)$. Suppose that the finite population $U$ is stratified into $H$ strata denoted by $U_1, \ldots, U_h, \ldots, U_H$, where $\bigcup_{h=1}^H U_h = U$. Suppose that a sample $s_h$ of fixed size $n_h$ is selected without replacement with unequal probabilities $π_i$ from $U_h$. We assume that $d_h = \sum_{i∈U_h} π_i(1 - π_i) → ∞$ for all $h$ and that the number of strata $H$ is bounded.

The empirical likelihood estimator is still the solution of (12) where $\hat{m}_i$ are still the values which maximise (7) under a set of constraints with $c_i = z_i$ and $C = n$; where

$$z_i = (z_{i1}, \ldots, z_{iH})' \quad \text{and} \quad n = (n_1, \ldots, n_H)'$$  \hfill (29)

denotes the vector of the strata sample sizes, with $z_{ih} = π_i δ\{i ∈ U_h\}$. It can be shown that $\hat{m}_i = π_i^{-1}$.

For confidence intervals, we propose to use $c_i = z_i$, $c_i^* = (z_i', g_i(θ))^T$, $C = n$, and $C^* = (n, 0)^T$. Berger & De La Riva Torres (2012b) showed that (26) holds where $\hat{V}_{pps}[\hat{G}_π(θ_0)]$ is now the stratified variance $pps$ estimator which is consistent because the number of strata is bounded. Hence $\hat{r}(θ_0)$ follows a chi-squared distribution asymptotically and the empirical likelihood confidence intervals can be computed with (25).

Note that we propose to use the same likelihood function (12) with or without stratification. With the pseudo empirical likelihood approach, the pseudo empirical likelihood function without stratification is different from the pseudo empirical likelihood function with stratification (e.g. Rao & Wu, 2009, p. 195).
3.2 Empirical likelihood approach with auxiliary variables

Let $x_i$ be a $P$ vector of values of auxiliary variables attached to unit $i$. These variables are such that their population control totals $X = \sum_{i \in U} x_i$ are known. Let $\hat{m}_i(x)$ be the values which maximise (6) under the constraint (8) with $c_i = (x_i', z_i')'$ and $C = \sum_{i \in U} c_i$.

In §2.1, we showed that $\hat{G}(\theta)$ is asymptotically equal to the generalised optimal regression estimator of $G(\theta)$.

For confidence intervals, we propose to use the following restricted empirical log-likelihood function instead of the function (7).

$$\ell(\hat{m}(x)) = \sum_{i=1}^{n} \log \left( \frac{m_i \hat{m}_i(x)^{-1}}{\sum_{j=1}^{n} m_j \hat{m}_j(x)^{-1}} \right), \quad (30)$$

which will be used for the calculation of confidence intervals and not for point estimation. Note that the function (30) reduces to the function (7) when we do not have auxiliary variables.

Let $c_i = \hat{c}_i$, $c_i^* = \hat{c}_i^*$, $C = (X', n')'$, and $C^* = (X', n', 0)'$, with $\hat{c}_i = (x_i', \hat{z}_i')'$, $\hat{c}_i^* = (x_i', \hat{z}_i, g_i(\theta))'$ and $\hat{z}_i = z_i/\{\pi_i \hat{m}_i(x)\}$. Let $\ell(\hat{m}^*(x))$ be the maximum value of of the empirical log-likelihood function. The restricted empirical log-likelihood ratio function is given by

$$\hat{r}_x(\theta) = 2 \{ \ell(\hat{m}(x)) - \ell(\hat{m}^*(x)) \}.$$  

Berger & De La Riva Torres (2012b) showed that

$$\hat{r}_x(\theta_0) = \hat{G}_x(\theta_0)^2 \hat{V}_{st}[\hat{G}_x(\theta_0)]^{-1} + o_p(1); \quad (31)$$

where $\hat{G}_x(\theta_0) = \sum_{i=1}^{n} g_i(\theta_0) \hat{m}_i(x)$ and $\hat{V}_{st}[\hat{G}_x(\theta_0)]$ is an estimator of the variance of $\hat{G}_x(\theta_0)$. This variance takes into account of the calibration constraint and of the fixed sizes constraints. Deville & Tillé (2005) showed that this estimator is consistent under fixed size sampling designs. Thus $\hat{r}_x(\theta_0)$ follows a chi-squared distribution asymptotically.

4. Empirical likelihood approach for non negligible sampling fractions

With large sampling fractions, the pps variance estimator (27) is biased, implying that the empirical log-likelihood ratio function does not necessarily follow a chi-squared distribution. Hence the empirical log-likelihood ratio function described in §3 cannot be used for confidence intervals, and needs to be adjusted to allow for large sampling fractions. Note that for point estimation, the approaches describes in the previous section are still valid even if we have large sampling fractions. In this §4, we propose to adjust the empirical log-likelihood ratio function in order to obtain a chi-squared distribution asymptotically.

A simple solution consists in using the approaches described in the previous § and multiplying $\hat{r}(\theta)$ by the ratio of variances $\hat{\phi}(\theta) = \hat{V}_{pps}[\hat{G}_x(\theta)]\hat{V}[\hat{G}_x(\theta)]^{-1}$, where $\hat{V}[\hat{G}_x(\theta)]$ is an unbiased estimator of $\hat{G}_x(\theta)$. This makes the computation of confidence intervals more intensive, as $\hat{\phi}(\theta)$ needs to be computed for several values of $\theta$. This approach is not entirely satisfactory, as it relies on variance estimates. We propose an alternative approach which does not rely on variance estimates.

When we have a single stratum, we propose to use $c_i = q_i \pi_i$, $c_i^* = q_i(\pi_i, g_i(\theta))'$, $C = \sum_{i=1}^{n} q_i$ and $C^* = \left\{ \sum_{i=1}^{n} q_i, \sum_{i=1}^{n} (q_i - 1) g_i(\theta) \pi_i^{-1} \right\}'$, with $q_i = (1-\pi_i)^{1/2}$. The $q_i$ are finite population corrections factors proposed by Berger (2005b). The $q_i$ reduce the effect on the confidence interval of units with large $\pi_i$. For example, if $\pi_i = 1$, then
\( \hat{m}_i \pi_i = \hat{m}^*_i \pi_i = 1 \). This implies that this unit will have no contribution towards the empirical likelihood functions and any confidence intervals. This is a natural property as this unit does not contribute towards the sampling variation.

Consider the following adjusted empirical log-likelihood ratio function.

\[
\hat{r}(\theta)^{(a)} = \hat{r}(\theta) + \hat{\xi}(\theta),
\]

where \( \hat{\xi}(\theta) \) is a correction factor for large sampling fraction. This factor is defined by \( \hat{\xi}(\theta) = -2\eta^* C^* \); where \( \eta^* \) is the Lagrangian multiplier obtained with \( c_i^* \) and \( C^* \). In Appendix B, we show that

\[
\hat{r}(\theta_0)^{(a)} = \hat{G}_\pi(\theta_0)^2 \hat{V}[\hat{G}_\pi(\theta_0)]^{-1} + \alpha_p(1);
\]

where

\[
\hat{V}[\hat{G}_\pi(\theta_0)] = \sum_{i=1}^{n} q_i^2 \tilde{g}_i(\theta_0)^2 - \tilde{d}^{-1} \hat{G}(\theta_0)^2
\]

is the Hájek (1964) variance estimator, with \( \hat{G}(\theta_0) = \sum_{i=1}^{n} q_i^2 \tilde{g}_i(\theta_0) \) and \( \tilde{d} = \sum_{i=1}^{n} q_i^2 \).

If this variance estimator is consistent, we have that \( \hat{r}(\theta_0)^{(a)} \) follows a chi-squared distribution, by Slutsky’s theorem. Hence Empirical likelihood confidence intervals can be constructed with \( \hat{r}(\theta)^{(a)} \). The result (33) is consistent with (26), as when all the \( q_i \) are equal to one, Berger & De La Riva Torres (2012b) showed that \( \hat{V}[\hat{G}_\pi(\theta)] \) equals (27) and \( \hat{\xi}(\theta) = 0 \), implying that \( \hat{r}(\theta)^{(a)} = \hat{r}(\theta) \).

The variance estimator (34) is a consistent estimator for the variance, for high entropy sampling designs (e.g. Hájek, 1964, 1981; Berger, 1998a; Deville, 1999; Brewer, 2002; Brewer & Donadio, 2003; Haziza et al., 2004; Henderson, 2006; Tillé, 2006; Prášková & Sen, 2009; Fuller, 2009; Berger, 2007, 2011). For example the rejective (Hájek, 1964; Fuller, 2009), the Rao-Sampford (Rao, 1965; Sampford, 1967), the Chao (1982) and the Pareto sampling designs (Aires, 2000) are high entropy sampling designs (Berger, 2005a, 2011). Note that most sampling designs used in practice have large entropy, except the non-randomized systematic sampling design and the Rao et al. (1962) sampling design (see §4.1).

The adjustment term \( \hat{\xi}(\theta) \) is a correction which takes into account of the dispersion between the \( q_i \). Indeed, Berger & De La Riva Torres (2012b) showed that \( \hat{\xi}(\theta) = 0 \) under simple random sampling. Because of (33), we see that this correction ensures that that \( \hat{r}(\theta_0)^{(a)} \) follows a chi-squared distribution asymptotically. Berger & De La Riva Torres (2012b) showed that by making using Isaki & Fuller (1982) assumptions about the asymptotic behaviour of the joint-inclusion probabilities, we have that \( \hat{\xi}(\theta) = O_p(n^{1-2\psi}) \). Thus \( \hat{\xi}(\theta) \to 0 \), if these asymptions hold when \( \psi > 1/2 \). Although \( \hat{\xi}(\theta) \) may be negligible, we prefer to keep it in (32), as these conditions on the joint-inclusion probabilities can be hard to verify.

For stratified designs, we propose to use \( c_i = q_i z_i \), \( c_i^* = q_i \{ z_i', g_i(\theta) \}' \), \( C = \sum_{i=1}^{n} q_i z_i' \pi_i^{-1} \), and \( C^* = \{ \sum_{i=1}^{n} q_i z_i' \pi_i^{-1} \} \{ \sum_{i=1}^{n} (q_i - 1) g_i(\theta) \pi_i^{-1} \} \). Berger & De La Riva Torres (2012b) showed that (33) holds; where \( \hat{V}[\hat{G}_\pi(\theta_0)] \) is now the stratified Hájek (1964) variance estimator which is consistent because the number of strata is bounded. Hence \( \hat{r}(\theta_0)^{(a)} \) follows a chi-squared distribution asymptotically.

With calibration constraints, we propose to use \( c_i = \bar{c}_i \), \( c_i^* = \bar{c}_i^* \), \( C = \sum_{i=1}^{n} \hat{m}_i(x) \bar{c}_i \), and \( C^* = \{ \sum_{i=1}^{n} \hat{m}_i(x) \bar{c}_i \} \{ \sum_{i=1}^{n} (q_i - 1) g_i(\theta) \hat{m}_i(x) \} \). In this case, we need to use an adjusted restricted empirical log-likelihood ratio function given by \( \hat{r}_x(\theta) = \hat{r}_x(\theta) + \hat{\xi}_x(\theta) \); with \( \hat{\xi}_x(\theta) = -2\eta^* C^* \); where \( \eta^* \) is the Lagrangian multiplier obtain with \( c_i^* \) and \( C^* \). Berger & De La Riva Torres (2012b) showed that (31) holds. Thus \( \hat{r}_x(\theta) \) follows a chi-squared distribution asymptotically.
4.1 Empirical likelihood approach for the Rao-Hartley-Cochran strategy

The Hartley-Rao-Cochran sampling design (Rao et al., 1962) is a popular unequal probability sampling design which does not belong to the class of high entropy sampling designs. In this §, we show how the proposed approach can be used in this situation.

Suppose that the population is divided randomly into \( n \) groups \( A_1, \ldots, A_i, \ldots, A_n \) of sizes \( N_1, \ldots, N_i, \ldots, N_n \), where \( \sum_{i=1}^{n} N_i = N \). One unit is selected independently from each group with probability \( p_i = \pi_i/a_i \); where \( a_i = \sum_{j \in A_i} \pi_j \). As the units are selected independently, the empirical likelihood function is given by

\[
L(m) = \prod_{i=1}^{n} \frac{p_i m_i}{\left( \sum_{j=1}^{n} p_j m_j \right)}.
\]

By maximising this function under the constraint (8) with \( c_i = p_i \) and \( C = n \), we obtain \( \hat{m}_i = p_i^{-1} \). When \( g_i(\theta) = y_i - n^{-1} p_i \theta \), the maximum empirical likelihood estimator \( \hat{\theta} \), defined by (12), is the Hartley-Rao-Cochran estimator (Rao et al., 1962) of a total.

For the computation of confidence intervals, we propose to use \( c_i = q_i^* p_i \), \( c_i^* = (q_i^* p_i, q_i^* g_i(\theta))^\prime \), \( C = \sum_{i=1}^{n} q_i^* \) and \( C^* = (\sum_{i=1}^{n} q_i^*)^2 \), with \( q_i^* = a_i^{1/2} \) and \( q_i^* = (\bar{\zeta} a_i^{-1})^{1/2} \), where \( \bar{\zeta} = (\sum_{i=1}^{n} N_i^2 - N)/(N^2 - \sum_{i=1}^{n} N_i^2) \) is the finite population correction proposed by Rao et al. (1962). Berger & De La Riva Torres (2012b) showed that \( \hat{\tau}(\theta_0)(a) = \hat{G}_R(\theta_0)^2 \hat{V}[\hat{G}_R(\theta_0)]^{-1} + o_p(1) \) where \( \hat{G}_R(\theta_0) \) is the Rao et al. (1962) estimator of a total and \( \hat{V}[\hat{G}_R(\theta_0)] \) is its variance estimator. Hence \( \hat{\tau}(\theta_0)(a) \) follows a chi-squared distribution asymptotically, as \( \hat{G}_R(\theta_0) \) is asymptotically normal under regularity conditions proposed by Ohlsson (1986).

5. Estimation of Quantiles

Suppose that the parameter \( \theta_0 \) of interest is the \( q \) quantile \( Y_q \) of the population distribution of a variable of interest \( y_i \); where \( 0 < q < 1 \). As the estimating equation \( \sum_{i=1}^{n} \hat{m}_i(\delta(y_i \leq \theta) - q) = 0 \) does not always have a solution, it cannot be used directly to derive a empirical log-likelihood ratio function (e.g. Owen, 2001, p. 45). In order to avoid this problem, we propose to use the following function \( g_i(\theta) = \theta - y_i \); where

\[
\varrho(y_{(i)}, \theta) = \delta(y_i \leq \theta) + \frac{\theta - y_{(i-1)}}{y_{(i)} - y_{(i-1)}} \delta(y_{(i-1)} \leq \theta)(1 - \delta(y_i \leq \theta));
\]

where the \( y_{(i)} \) is the values of the \( i \)-th sampled units arranged in increasing order, with \( y_{(0)} = y_{(1)} - (y_{(2)} - y_{(1)}) \). The empirical likelihood estimator of \( Y_q \) is the solution of the equation \( \hat{G}(\theta) = 0 \) which becomes \( \hat{G}(\theta) = q \); where

\[
\hat{G}(\theta) = \left( \sum_{i=1}^{n} \hat{m}_i \right)^{-1} \sum_{i=1}^{n} \hat{m}_i \varrho(y_{(i)}, \theta).
\]

Note that \( \hat{G}(\theta) = q \) has always a unique solution because \( \hat{F}(y) \) is a bijective function given by a piecewise linear interpolation of the step distribution function

\[
\hat{F}(\theta) = \left( \sum_{i=1}^{n} \hat{m}_i \right)^{-1} \sum_{i=1}^{n} \hat{m}_i \delta(y_{(i)} \leq \theta).
\]
This interpolation consists in joining the steps of \( \hat{F}(\theta) \) by straight lines segments. It can be easily shown that

\[
\frac{1}{N} \sum_{i=1}^{n} \pi_i^{-1} [\varrho(y(i), \theta_0) - q] \leq \frac{1}{N} \sum_{i=1}^{n} \pi_i^{-1} [\delta \{ y_i \leq \theta_0 \} - q]
\]

which is an Horvitz & Thompson (1952) estimator. Thus, (24) holds, and the empirical log-likelihood ratio function has a chi-squared distribution asymptotically. Therefore, the empirical log-likelihood ratio function can be used to derive confidence intervals for \( Y_q \).

**Table 1**: Coverages of the 95% confidence intervals. \( N = 800 \). \( \theta_0 = \mu \) and \( g_i(\theta) = y_i - N n^{-1} \theta \pi_i \). The approach described in \S 3 is used to compute confidence intervals. For the pseudo empirical likelihood (EL2) approach, the point estimator is the Hájek (1971) estimator.

<table>
<thead>
<tr>
<th>( cor(y_i, \hat{y}_i) )</th>
<th>( n )</th>
<th>Type of confidence intervals</th>
<th>Coverage Probabilities</th>
<th>Lower tail error rates</th>
<th>Upper tail error rates</th>
<th>Average Lengths</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.3</td>
<td>40</td>
<td>Proposed</td>
<td>91.5%</td>
<td>2.2%</td>
<td>6.3%</td>
<td>1.96</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pseudo EL2</td>
<td>91.2%</td>
<td>2.5%</td>
<td>6.3%</td>
<td>1.85</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Normal</td>
<td>90.7%</td>
<td>0.7%</td>
<td>8.6%</td>
<td>1.87</td>
</tr>
<tr>
<td></td>
<td>80</td>
<td>Proposed</td>
<td>95.0%</td>
<td>2.5%</td>
<td>2.5%</td>
<td>1.42</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pseudo EL2</td>
<td>93.1%</td>
<td>3.1%</td>
<td>3.8%</td>
<td>1.32</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Normal</td>
<td>92.3%</td>
<td>1.5%</td>
<td>6.2%</td>
<td>1.33</td>
</tr>
<tr>
<td>0.8</td>
<td>40</td>
<td>Proposed</td>
<td>94.2%</td>
<td>2.1%</td>
<td>3.7%</td>
<td>0.62</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pseudo EL2</td>
<td>91.9%</td>
<td>2.6%</td>
<td>5.5%</td>
<td>0.46</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Normal</td>
<td>93.3%</td>
<td>1.3%</td>
<td>5.4%</td>
<td>0.59</td>
</tr>
<tr>
<td></td>
<td>80</td>
<td>Proposed</td>
<td>95.6%</td>
<td>1.4%</td>
<td>3.0%</td>
<td>0.45</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pseudo EL2</td>
<td>93.5%</td>
<td>2.5%</td>
<td>4.0%</td>
<td>0.33</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Normal</td>
<td>93.9%</td>
<td>1.1%</td>
<td>5.0%</td>
<td>0.41</td>
</tr>
</tbody>
</table>

### 6. Simulation study

We generated several population data according to the following model proposed by Wu & Rao (2006).

\[
y_i = 3 + a_i + \varphi e_i,
\]

where \( a_i \) follows an exponential distributions with rate parameters equal to one and \( e_i \sim \chi_1^2 \). The \( \pi_i \) are proportional to \( a_i + 2 \). The constant 2 is added to \( a_i \) to avoid having very small \( \pi_i \). Populations of size \( N = 800 \) and \( N = 150 \) will be generated using (35). The values \( y_i \) and \( a_i \) generated will be treated as fixed. The parameter \( \varphi \) is used to obtain a weak and a strong correlation between the values \( y_i \) and \( \hat{y}_i = 3 + a_i \). Let \( \rho(y, \hat{y}) \) denote this correlation. The parameter of interest \( \theta_0 \) is the population mean.

We use the Chao (1982) sampling design to select 1000 samples with unequal probabilities in order to compare the Monte-Carlo performance of the 95% empirical likelihood confidence interval with the standard confidence interval based on the central limit theorem and the pseudo empirical likelihood (EL2) confidence interval proposed by Wu & Rao.
Table 2: Coverages of the 95% confidence intervals. \( N = 150 \). \( \theta_0 = \mu \) and \( g_i(\theta) = y_i - Nn^{-1} \theta \pi_i \). The approach described in §4 is used to compute confidence intervals. For the pseudo empirical likelihood (EL2) approach, the point estimator is the Hájek (1971) estimator.

<table>
<thead>
<tr>
<th>( cor(y_i, \hat{y}_i) )</th>
<th>( n )</th>
<th>Type of confidence intervals</th>
<th>Coverage Probabilities</th>
<th>Lower tail error rates</th>
<th>Upper tail error rates</th>
<th>Average Lengths</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.3</td>
<td>40</td>
<td>Proposed</td>
<td>91.6%</td>
<td>2.6%</td>
<td>5.8%</td>
<td>2.03</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pseudo EL2</td>
<td>90.6%</td>
<td>2.1%</td>
<td>7.3%</td>
<td>1.88</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Normal</td>
<td>89.6%</td>
<td>0.4%</td>
<td>10.0%</td>
<td>1.90</td>
</tr>
<tr>
<td></td>
<td>80</td>
<td>Proposed</td>
<td>92.5%</td>
<td>4.3%</td>
<td>3.2%</td>
<td>1.29</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pseudo EL2</td>
<td>93.0%</td>
<td>2.0%</td>
<td>5.0%</td>
<td>1.15</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Normal</td>
<td>93.5%</td>
<td>0.8%</td>
<td>5.7%</td>
<td>1.15</td>
</tr>
<tr>
<td>0.8</td>
<td>40</td>
<td>Proposed</td>
<td>94.6%</td>
<td>3.4%</td>
<td>2.0%</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pseudo EL2</td>
<td>93.9%</td>
<td>1.8%</td>
<td>4.3%</td>
<td>0.38</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Normal</td>
<td>94.7%</td>
<td>1.1%</td>
<td>4.2%</td>
<td>0.48</td>
</tr>
<tr>
<td></td>
<td>80</td>
<td>Proposed</td>
<td>89.9%</td>
<td>9.6%</td>
<td>0.5%</td>
<td>0.29</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pseudo EL2</td>
<td>93.7%</td>
<td>2.4%</td>
<td>3.9%</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Normal</td>
<td>93.5%</td>
<td>1.4%</td>
<td>5.1%</td>
<td>0.25</td>
</tr>
</tbody>
</table>

(2006, p. 362). We consider that we have a single stratum. The Sen-Yates-Grundy variance estimator (Sen, 1953; Yates & Grundy, 1953) is used for standard confidence intervals and for pseudo empirical likelihood approach. We used the statistical software R (R Development Core Team, 2006). The observed coverage probability, the lower and the upper tail error rates and the average length of the 95% confidence intervals are reported in Tables 1 and 2.

In Table 1, we used the approach described in §3, as the sampling fraction is negligible. The confidence intervals computed with the proposed empirical likelihood approach perform better than the confidence intervals computed with the other approaches. The coverages of the proposed approach are closer to 95% and the lower and upper tail error rates are closer to 2.5%. In Table 2, we used the approach described in §4, as the sampling fraction is not negligible. We see that the proposed approach gives better coverages which are better than the pseudo empirical likelihood approach, except when the sample size \( n = 80 \).

7. Discussion

The proposed empirical likelihood approach can be easily generalised for multi-stage designs (e.g. Särndal et al., 1992, §4.3.2), by using an ultimate cluster approach; where the primary sampling units’ totals play the role of the units. This approach gives consistent confidence intervals when the sampling fractions are small.

The proposed empirical likelihood approach can be generalised in the presence of non-response by using Fay (1991) reverse approach (Shao & Steel, 1999) which can accommodate imputation and weighting adjustment. Another approach consists in using auxiliary variables to compensate for nonresponse (e.g. Särndal & Lundström, 2005).

Standard confidence intervals based on the central limit theorem and pseudo empirical likelihood confidence intervals require variance estimates which often involve linearisa-
tion or re-sampling. Even if the parameter of interest is no linear, the proposed method does not rely on variance estimates, linearisation or re-sampling, and empirical likelihood confidence intervals can be easier to compute than standard confidence intervals based on variance estimates. It provides an alternative to more computationally intensive methods such as bootstrap or jackknife, when linearisation cannot be used.

Bootstrap is an alternative approach which can be used to derive non-parametric confidence intervals. The proposed approach is less computationally intensive than the bootstrap. It is also possible to combine the empirical likelihood and the bootstrap approaches to improve the coverage of the empirical likelihood confidence intervals, by replacing the threshold $\chi^2_1(\alpha)$ in (28) by a quantity obtained by bootstrapping the empirical likelihood ratio function (e.g. Owen, 2001, §3.3, Wu & Rao, 2010).

REFERENCES


