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 θ_0 is the solution of the following estimating equation (e.g. Binder & Kovačević, 1995).

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

where $g_i(\theta)$ is a function of θ and of characteristics of the unit i. This function does not need to be differentiable. Note that $g_i(\theta)$ and θ_0 can be vectors, but for simplicity, we consider that they are scalar. For example, θ_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 θ_0.

Suppose that we wish to estimate θ_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 π_i denote the inclusion probability of unit i. An unbiased estimator of the function (1) is given by the following Horvitz & Thompson (1952) estimator.

$$\widehat{G}_\pi(\theta) = \sum_{i=1}^{n} \breve{g}_i(\theta); \tag{2}$$

*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 $\breve{g}_i(\theta) = g_i(\theta)\pi_i^{-1}$. An estimator $\widehat{\theta}$ of θ_0 is the solution of $\widehat{G}_\pi(\theta) = 0$. When $g_i(\theta) = y_i - n^{-1}\theta\pi_i$, the solution of $\widehat{G}_\pi(\theta) = 0$ is the Horvitz & Thompson (1952) estimator $\widehat{Y}_{HT} = \sum_{i=1}^{n} y_i\pi_i^{-1}$ 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 $\widehat{Y}_H = N\widehat{N}_\pi^{-1}\widehat{Y}_{HT}$ of Y; where $\widehat{N}_\pi = \sum_{i=1}^{n} \pi_i^{-1}$. The estimator $\widehat{Y}_{HT}$ may not be as efficient as $\widehat{Y}_H$ when y_i and π_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, incuring 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 $\widehat{Y}_{HT}$ or $\widehat{Y}_H$ 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 situtation 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 parameters of the empirical likelihood function can be estimated. In §2.1, we define the empirical likelihood estimators. In §3 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)

$$F_s(y) = \frac{\sum_{i=1}^n \pi_i \, P_i \, \delta\{y_i \leq y\}}{\sum_{j=1}^n \pi_j \, P_j}; \tag{3}$$

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 $m_i = NP_i$ where m_i is the unit mass of unit i in the population (e.g. Deville, 1999). Thus (3) reduces to

$$F_s(y) = \frac{\sum_{i=1}^n \pi_i \, m_i \, \delta\{y_i \leq y\}}{\sum_{j=1}^n \pi_j \, m_j}. \tag{4}$$

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)

$$L(m) = \prod_{i=1}^n [F_s(y_i) - F_s(y_i^-)]; \tag{5}$$

where $F_s(y_i^-) = lim_{y \uparrow y_i} F_s(y)$. 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

$$L(m) = \prod_{i=1}^n \left(\frac{\pi_i \, m_i}{\sum_{j=1}^n \pi_j \, m_j} \right). \tag{6}$$

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 $\widehat{m}_i$ which maximise the *log-empirical likelihood function*

$$\ell(m) = \log(L(m)), \tag{7}$$

subject to the constraints $m_i \geq 0$ and

$$\sum_{i=1}^{n} m_i \boldsymbol{c}_i \;=\; \boldsymbol{C}; \tag{8}$$

where $\boldsymbol{c}_i$ is a known $Q \times 1$ vector associated with the i-th sampled unit and $\boldsymbol{C}$ 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 $\boldsymbol{C}$ is not necessarily a vector of fixed quantities. Hence $\boldsymbol{C}$ can be fixed or random. Possible choices for $\boldsymbol{c}_i$ and $\boldsymbol{C}$ are discussed in §3. Note that the $\boldsymbol{c}_i$ and $\boldsymbol{C}$ 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 $\boldsymbol{C}$ 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 find by using the Lagrangian function, $Q(m, \boldsymbol{\eta}) = \sum_{i=1}^{n} \log(\pi_i \, m_i) - n \log\left(\sum_{i=1}^{n} \pi_i \, m_i\right) - \boldsymbol{\eta}' \left(\sum_{i=1}^{n} m_i \boldsymbol{c}_i - \boldsymbol{C}\right)$. The values of m_i and $\boldsymbol{\eta}$ which minimise $Q(m, \boldsymbol{\eta})$ are the solutions of the following set of equations $\partial Q(m, \boldsymbol{\eta})/\partial m_i = 0$ and $\partial Q(m, \boldsymbol{\eta})/\partial \boldsymbol{\eta} = 0$. As (8) is such that $\sum_{i=1}^{n} m_i \pi_i = n$, the solution is

$$\widehat{m}_i = \left(\pi_i + \boldsymbol{\eta}' \boldsymbol{c}_i\right)^{-1}, \tag{9}$$

The parameter $\boldsymbol{\eta}$ 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 $\boldsymbol{\eta}$, $\boldsymbol{f}(\boldsymbol{\eta}) = \sum_{i=1}^{n} \widehat{m}_i \boldsymbol{c}_i$. A Taylor approximation of $\boldsymbol{f}(\boldsymbol{\eta})$ in the neighbourhood of an initial guess $\boldsymbol{\eta}_0$ gives

$$\boldsymbol{\eta} \simeq \boldsymbol{\eta}_0 - \widehat{\boldsymbol{\Delta}}(\boldsymbol{\eta}_0)^{-1} (\boldsymbol{f}(\boldsymbol{\eta}_0) - \boldsymbol{C}), \tag{10}$$

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

$$\widehat{\boldsymbol{\Delta}}(\boldsymbol{\eta}) = \frac{\partial \boldsymbol{f}(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} = -\sum_{i=1}^{n} \boldsymbol{c}_i \boldsymbol{c}_i' \left(\pi_i + \boldsymbol{\eta}' \boldsymbol{c}_i\right)^{-2}. \tag{11}$$

The recursive formula (10) can be used to compute $\boldsymbol{\eta}$. For the first iteration, we used $\boldsymbol{\eta}_0 = \boldsymbol{0}$ which gives a new approximation of $\boldsymbol{\eta}$ using (10). This new approximation is used as a new value for $\boldsymbol{\eta}_0$ which is substituted into (10). We repeat this process until convergence. Note that it is not necessary to know N in order to the compute $\boldsymbol{\eta}$ and $\widehat{m}_i$.

Note that when $\boldsymbol{c}_i = \pi_i$ and $\boldsymbol{C} = n$, we have that $\boldsymbol{\eta} = 0$ and $\widehat{m}_i = \pi_i^{-1}$.

2.1 Maximum empirical likelihood estimator

The *maximum empirical likelihood estimator* $\widehat{\theta}$ of θ_0 is defined by solution of the following estimating equation.

$$\widehat{G}(\theta) = 0, \quad \text{with} \quad \widehat{G}(\theta) = \sum_{i=1}^{n} \widehat{m}_i \, g_i(\theta); \tag{12}$$

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

Note that when $\boldsymbol{c}_i = \pi_i$ and $\boldsymbol{C} = n$, we have that $\boldsymbol{\eta} = 0$ and $\widehat{m}_i = \pi_i^{-1}$. In this case, $\widehat{\theta}$ is the Horvitz & Thompson (1952) estimator $\widehat{Y}_{HT}$ when $g_i(\theta) = y_i - n^{-1}\theta\pi_i$ and $\widehat{\theta}$ is the Hájek (1971) ratio estimator $\widehat{Y}_H$ 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 $\widehat{Y}_{HT}$ and $\widehat{Y}_H$. 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 $d = \sum_{i \in U} \pi_i(1-\pi_i) \to \infty$. This assumption implies that $n \to \infty$ and $N \to \infty$, as $d < n < N$. The standard empirical likelihood approach (Owen, 1988) assumes that the sampling fraction is negligible ($n/N \to 0$). 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.

$$
\begin{aligned}
N^{-1}\|\widehat{\boldsymbol{C}}_\pi - \boldsymbol{C}\| &= O_p(n^{-\frac{1}{2}}), && (13)\\
N^{-1}\widehat{G}_\pi(\theta_0) &= O_p(n^{-\frac{1}{2}}), && (14)\\
nN^{-1}\pi_i^{-1} &= O(1), && (15)\\
\max\{\|\boldsymbol{c}_i\| : i \in s\} &= o_p(n^{\frac{1}{2}}), && (16)\\
\max\{|g_i(\theta)| : i \in s\} &= o_p(n^{\frac{1}{2}}), && (17)\\
\|\widehat{\boldsymbol{S}}\| &= O_p(1), && (18)\\
\|\widehat{\boldsymbol{S}}^{-1}\| &= O_p(1), && (19)\\
\frac{1}{nN^\tau}\sum_{i=1}^{n}\frac{\|\boldsymbol{c}_i\|^\tau}{\pi_i^\tau} &= O_p(n^{-\tau}), && (20)
\end{aligned}
$$

where $\tau \leq 3$,

$$
\widehat{\boldsymbol{S}} = \frac{n}{N^2}\widehat{\boldsymbol{\Delta}}(\boldsymbol{0}), \quad \text{and} \quad \widehat{\boldsymbol{C}}_\pi = \sum_{i=1}^{n}\frac{\boldsymbol{c}_i}{\pi_i}\cdot \qquad (21)
$$

The matrix $\widehat{\boldsymbol{\Delta}}(\boldsymbol{0})$ is given by (11) with $\boldsymbol{\eta} = \boldsymbol{0}$. The quantity $\|\boldsymbol{A}\| = \text{trace}(\boldsymbol{A}'\boldsymbol{A})^{1/2}$ 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 $\widehat{\boldsymbol{S}}$ 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 $\|\widehat{\boldsymbol{S}}\|$ 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,

$$
\widehat{G}(\theta) = \widehat{G}_\pi(\theta) + \widehat{\boldsymbol{B}}'(\boldsymbol{C} - \widehat{\boldsymbol{C}}_\pi) + o_p(N), \qquad (22)
$$

where $\widehat{\boldsymbol{B}}$ is a vector of regression coefficients defined by

$$\widehat{\boldsymbol{B}} \;=\; \left(\sum_{i=1}^{n} \frac{1}{\pi_i^2}\boldsymbol{c}_i\boldsymbol{c}_i'\right)^{-1} \sum_{i=1}^{n} \frac{1}{\pi_i^2} g_i(\theta)\boldsymbol{c}_i. \tag{23}$$

where $\widehat{w}_i$ 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.

When $\boldsymbol{c}_i = (x_i,\ \pi_i)'$, we have that $\widehat{G}(\theta) = \widehat{G}_\pi(\theta) + \widehat{B}_x(X - \widehat{X}_{HT}) + o_p(N)$, where $\widehat{B}_x = \sum_{i=1}^{n}(\breve{x}_i - n^{-1}\widehat{X}_{HT})(\breve{g}_i(\theta) - \widehat{G}_\pi(\theta))\{\sum_{i=1}^{n}(\breve{x}_i - n^{-1}\widehat{X}_{HT})^2\}^{-1}$, $\breve{x}_i = x_i\pi_i^{-1}$ and $\breve{g}_i(\theta) = g_i(\theta)\pi_i^{-1}$. Note that $\widehat{B}_x$ is the estimator of the covariance between $\widehat{G}_\pi(\theta)$ and $\widehat{X}_{HT}$ divided by the estimator of the variance of $\widehat{X}_{HT}$ under a with replacement pps sampling design. Therefore $\widehat{B}_x$ is the optimal regression coefficient (e.g. Isaki & Fuller, 1982; Montanari, 1987; Särndal, 1996; Berger et al., 2003) when the sampling fraction is small. This results can be easily extended when we have more than one auxiliary variable. Hence, the empirical likelihood estimator is asymptotically optimal. When N is known, it is recommended to use $x_i = 1$ or to include a variable equal to one in $\boldsymbol{x}_i$. This will improve the efficiency of the maximum empirical likelihood estimator.

3. Empirical likelihood confidence intervals

The main advantage of empirical likelihood approach is its capability of deriving non-parametric confidence intervals which do not depend on variance estimates. In this §, we propose to use an empirical log-likelihood ratio function to derive empirical likelihood confidence intervals. The approaches proposed in this § are valid for small sampling fractions (i.e. when the π_i are negligible); that is, if we can treat the sample as if the units were selected with replacement. In §4, we will see how the proposed method can be adapted to account for large sampling fractions.

Empirical likelihood confidence intervals rely on the following property.

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

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

As θ_0 is a constant, $\widehat{G}_\pi(\theta_0)$ is a Horvitz & Thompson (1952) estimator. Hájek (1964), Víšek (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 $E(\widehat{G}_\pi(\theta_0)) = G(\theta_0) = 0$. 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 θ_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 $\widehat{m}_i$ be the values which maximise (7) subject to the constraints $m_i \geq 0$ and (8) when $\boldsymbol{c}_i = \pi_i$ and $\boldsymbol{C} = n$. Note that $m_i = \pi_i^{-1}$ in this situation. Hence the empirical likelihood point estimator is the solution of $\widehat{G}_\pi(\theta) = 0$. Let $\ell(\widehat{m})$ be the maximum value of the empirical log-likelihood function.

Let $\widehat{m}_i^*$ be the values which maximise (7) subject to the constraints $m_i \geq 0$ and (8) with $\boldsymbol{c}_i = \boldsymbol{c}_i^*$ and $\boldsymbol{C} = \boldsymbol{C}^*$, where $\boldsymbol{c}_i^* = (\pi_i, g_i(\theta))'$ and $\boldsymbol{C}^* = (n, 0)'$. Let $\ell(\widehat{m}^*, \theta)$ be the maximum value of of the empirical log-likelihood function.

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

$$\widehat{r}(\theta) \quad = \quad 2\left\{\ell(\widehat{m}) - \ell(\widehat{m}^*, \theta)\right\} \cdot \tag{25}$$

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

$$\widehat{r}(\theta_0) = \widehat{G}_\pi(\theta_0)^2\ \widehat{V}_{pps}[\widehat{G}_\pi(\theta_0)]^{-1} + o_p(1) \cdot \tag{26}$$

where θ_0 denotes the population parameter to estimate and where $\widehat{V}_{pps}[\widehat{G}_\pi(\theta_0)]$ is the following *pps* variance estimator (e.g. Durbin, 1953; Särndal et al., 1992, p. 99).

$$\widehat{V}_{pps}[\widehat{G}_\pi(\theta_0)] = \sum_{i=1}^{n} \left(\breve{g}_i(\theta_0) - n^{-1}\widehat{G}_\pi(\theta_0)\right)^2 ; \tag{27}$$

where $\breve{g}_i(\theta) = g_i(\theta)/\pi_i$. When the sampling fractions are negligible, $\widehat{V}_{pps}[\widehat{G}_\pi(\theta)]$ is a consistent estimator for the variance (Durbin, 1953). Hence the property (24) implies that $\widehat{r}(\theta_0)$ follows asymptotically a chi-squared distribution with one degree of freedom, by Slutsky's theorem. Thus, the $(1-\alpha)$ level empirical likelihood confidence interval (e.g. Wilks, 1938; Hudson, 1971) for the population parameter θ_0 is given by

$$\left[\ \min\left\{\theta |\ \widehat{r}(\theta) \leq \chi_1^2(\alpha)\right\} ;\ \max\left\{\theta |\ \widehat{r}(\theta) \leq \chi_1^2(\alpha)\right\}\ \right] ; \tag{28}$$

where $\chi_1^2(\alpha)$ is the upper α-quantile of the chi-squared distribution with one degree of freedom. Note that $\widehat{r}(\theta)$ 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 $\widehat{r}(\theta)$ 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 $U_1, \dots, U_h, \dots, U_H$; where $\cup_{h=1}^{H} U_h = U$. 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 $d_h = \sum_{i \in U_h} \pi_i(1-\pi_i) \to \infty$ for all h and that the number of strata H is bounded.

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

$$\boldsymbol{z}_i = (z_{i1}, \dots, z_{iH})' \quad \text{and} \quad \boldsymbol{n} = (n_1, \dots, n_H)' \tag{29}$$

denotes the vector of the strata sample sizes, with $z_{ih} = \pi_i\ \delta\{i \in U_h\}$. It can be shown that $\widehat{m}_i = \pi_i^{-1}$.

For confidence intervals, we propose to use $\boldsymbol{c}_i = \boldsymbol{z}_i$, $\boldsymbol{c}_i^* = (\boldsymbol{z}_i', g_i(\theta))'$, $\boldsymbol{C} = \boldsymbol{n}$, and $\boldsymbol{C}^* = (\boldsymbol{n}, 0)'$. Berger & De La Riva Torres (2012b) showed that (26) holds where $\widehat{V}_{pps}[\widehat{G}_\pi(\theta_0)]$ is now the stratified variance pps estimator which is consistent because the number of strata is bounded. Hence $\widehat{r}(\theta_0)$ 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 $\boldsymbol{x}_i$ be a P vector of values of auxiliary variables attached to unit i. These variables are such that their population control totals $\boldsymbol{X} = \sum_{i \in U} \boldsymbol{x}_i$ are known. Let $\widehat{m}_i(\boldsymbol{x})$ be the values which maximise (6) under the constraint (8) with $\boldsymbol{c}_i = (\boldsymbol{x}_i', \boldsymbol{z}_i')'$ and $\boldsymbol{C} = \sum_{i \in U} \boldsymbol{c}_i$. In §2.1, we showed that $\widehat{G}(\theta)$ 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).

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

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

$$\widehat{r}_x(\theta) \quad = \quad 2 \left\{ \ell(\widehat{m}(\boldsymbol{x})) - \ell(\widehat{m}^*(\boldsymbol{x})) \right\} \cdot$$

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

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

where $\widehat{G}_x(\theta_0) = \sum_{i=1}^{n} g_i(\theta_0) \widehat{m}_i(\boldsymbol{x})$ and $\widehat{V}_{st}[\widehat{G}_x(\theta_0)]$ is an estimator of the variance of $\widehat{G}_x(\theta_0)$. 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 $\widehat{r}_x(\theta_0)$ 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 §, 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 $\widehat{r}(\theta)$ by the ratio of variances $\widehat{\phi}(\theta) = \widehat{V}_{pps}[\widehat{G}_\pi(\theta)] \widehat{V}[\widehat{G}_\pi(\theta)]^{-1}$, where $\widehat{V}[\widehat{G}_\pi(\theta)]$ is an unbiased estimator of $\widehat{G}_\pi(\theta)$. This makes the computation of confidence intervals more intensive, as $\widehat{\phi}(\theta)$ needs to be computed for several values of θ. 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 $\boldsymbol{c}_i = q_i \pi_i$, $\boldsymbol{c}_i^* = q_i(\pi_i, g_i(\theta))'$, $\boldsymbol{C} = \sum_{i=1}^{n} q_i$ and $\boldsymbol{C}^* = (\sum_{i=1}^{n} q_i, \sum_{i=1}^{n} (q_i - 1) g_i(\theta) \pi_i^{-1})'$, 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 π_i. For example, if $\pi_i = 1$, then

$\widehat{m}_i\pi_i = \widehat{m}_i^*\pi_i = 1$. 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.

$$\widehat{r}(\theta)^{(a)} = \widehat{r}(\theta) + \widehat{\xi}(\theta), \tag{32}$$

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

$$\widehat{r}(\theta_0)^{(a)} = \widehat{G}_\pi(\theta_0)^2\ \widehat{V}[\widehat{G}_\pi(\theta_0)]^{-1} + o_p(1); \tag{33}$$

where

$$\widehat{V}[\widehat{G}_\pi(\theta_0)] = \sum_{i=1}^{n} q_i^2\breve{g}_i(\theta_0)^2 - \widehat{d}^{-1}\mathring{G}(\theta_0)^2 \tag{34}$$

is the Hájek (1964) variance estimator, with $\mathring{G}(\theta_0) = \sum_{i=1}^n q_i^2\breve{g}_i(\theta_0)$ and $\widehat{d} = \sum_{i=1}^n q_i^2$. If this variance estimator is consistent, we have that $\widehat{r}(\theta_0)^{(a)}$ follows a chi-squared distribution, by Slutsky's theorem. Hence Empirical likelihood confidence intervals can be constructed with $\widehat{r}(\theta)^{(a)}$. 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 $\widehat{V}[\widehat{G}_\pi(\theta)]$ equals (27) and $\widehat{\xi}(\theta) = 0$, implying that $\widehat{r}(\theta)^{(a)} = \widehat{r}(\theta)$.

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 $\widehat{\xi}(\theta)$ is a correction which takes into account of the dispersion between the q_i. Indeed, Berger & De La Riva Torres (2012b) showed that $\widehat{\xi}(\theta) = 0$ under simple random sampling. Because of (33), we see that this correction ensures that that $\widehat{r}(\theta_0)^{(a)}$ follows a chi-squared distribution asymptotically. Berger & De La Riva Torres (2012b) showed that by making Isaki & Fuller (1982) assumptions about the asymptotic behaviour of the joint-inclusion probabilities, we have that $\widehat{\xi}(\theta) = O_p(n^{1-2\psi})$. Thus $\widehat{\xi}(\theta) \to 0$, if these asymptions hold when $\psi > 1/2$. Although $\widehat{\xi}(\theta)$ 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 $\boldsymbol{c}_i = q_i\boldsymbol{z}_i$, $\boldsymbol{c}_i^* = q_i(\boldsymbol{z}_i', g_i(\theta))'$, $\boldsymbol{C} = \sum_{i=1}^n q_i\boldsymbol{z}_i'\pi_i^{-1}$, and $\boldsymbol{C}^* = (\sum_{i=1}^n q_i\boldsymbol{z}_i'\pi_i^{-1}, \sum_{i=1}^n(q_i - 1)g_i(\theta)\pi_i^{-1})'$. Berger & De La Riva Torres (2012b) showed that (33) holds; where $\widehat{V}[\widehat{G}_\pi(\theta_0)]$ is now the stratified Hájek (1964) variance estimator which is consistent because the number of strata is bounded. Hence $\widehat{r}(\theta_0)^{(a)}$ follows a chi-squared distribution asymptotically.

With calibration constraints, we propose to use $\boldsymbol{c}_i = \dot{\boldsymbol{c}}_i$, $\boldsymbol{c}_i^* = \dot{\boldsymbol{c}}_i^*$, $\boldsymbol{C} = \sum_{i=1}^n \widehat{m}_i(\boldsymbol{x})\dot{\boldsymbol{c}}_i$, and $\boldsymbol{C}^* = (\sum_{i=1}^n \widehat{m}_i(\boldsymbol{x})\dot{\boldsymbol{c}}_i, \sum_{i=1}^n(q_i - 1)g_i(\theta)\widehat{m}_i(\boldsymbol{x}))'$, with $\dot{\boldsymbol{c}}_i = q_i(\boldsymbol{x}_i', \dot{\boldsymbol{z}}_i')'$, $\dot{\boldsymbol{c}}_i^* = q_i(\boldsymbol{x}_i', \dot{\boldsymbol{z}}_i', g_i(\theta))'$. In this case, we need to use an *adjusted restricted empirical log-likelihood ratio function* given by $\widehat{r}_x^{(a)}(\theta) = \widehat{r}_x(\theta) + \widehat{\xi}_x(\theta)$; with $\widehat{\xi}_x(\theta) = -2\boldsymbol{\eta}^{*'}\boldsymbol{C}^*$; where $\boldsymbol{\eta}^*$ is the Lagrangian multiplier obtain with $\boldsymbol{c}_i^*$ and $\boldsymbol{C}^*$. Berger & De La Riva Torres (2012b) showed that (31) holds. Thus $\widehat{r}_x^{(a)}(\theta)$ 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 appproach 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

$$L(m) = \prod_{i=1}^{n} \frac{p_i \, m_i}{(\sum_{j=1}^{n} p_j \, m_j)}.$$

By maximising this function under the constraint (8) with $\boldsymbol{c}_i = p_i$ and $\boldsymbol{C} = n$, we obtain $\widehat{m}_i = p_i^{-1}$. When $g_i(\theta) = y_i - n^{-1} p_i \theta$, the maximum empirical likelihood estimator $\widehat{\theta}$, 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 $\boldsymbol{c}_i = q_i^{\circ} p_i$, $\boldsymbol{c}_i^{*} = (q_i^{\circ} p_i, q_i^{\bullet} g_i(\theta))'$, $\boldsymbol{C} = \sum_{i=1}^{n} q_i^{\circ}$ and $\boldsymbol{C}^{*} = (\sum_{i=1}^{n} q_i^{\circ}, \sum_{i=1}^{n} (q_i^{\bullet} - 1) g_i(\theta) p_i^{-1})'$, with $q_i^{\circ} = a_i^{1/2}$ and $q_i^{\bullet} = (\widehat{\varsigma}\, n \, a_i^{-1})^{1/2}$; where $\widehat{\varsigma} = (\sum_{i=1}^{n} N_i^2 - N)[(N^2 - \sum_{i=1}^{n} N_i^2)$ is the finite population correction proposed by Rao et al. (1962). Berger & De La Riva Torres (2012b) showed that $\widehat{r}(\theta_0)^{(a)} = \widehat{G}_R(\theta_0)^2 \widehat{V}[\widehat{G}_R(\theta_0)]^{-1} + o_p(1)$ where $\widehat{G}_R(\theta_0)$ is the Rao et al. (1962) estimator of a total and $\widehat{V}[\widehat{G}_R(\theta_0)]$ is its variance estimator. Hence $\widehat{r}(\theta_0)^{(a)}$ follows a chi-squared distribution asymptotically, as $\widehat{G}_R(\theta_0)$ is asymptotically normal under regularity conditions proposed by Ohlsson (1986).

5. Estimation of Quantiles

Suppose that the parameter θ_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 $\sum_{i=1}^{n} \widehat{m}_i (\delta\{y_i \leq \theta\} - q) = 0$ 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 $g_i(\theta) = \varrho(y_{(i)}, \theta) - q$; 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 $\widehat{G}(\theta) = 0$ which becomes $\widetilde{F}(\theta) = q$; where

$$\widetilde{F}(\theta) = \left(\sum_{i=1}^{n} \widehat{m}_i \right)^{-1} \sum_{i=1}^{n} \widehat{m}_{(i)} \varrho(y_{(i)}, \theta).$$

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

$$\widehat{F}(\theta) = \left(\sum_{i=1}^{n} \widehat{m}_i \right)^{-1} \sum_{i=1}^{n} \widehat{m}_i \delta\{y_{(i)} \leq \theta\}.$$

This interpolation consists in joining the steps of $\widehat{F}(\theta)$ by straight lines segments. It can be easily shown that

$$\frac{1}{N}\widehat{G}_{\pi}(\theta_0) = \frac{1}{N}\sum_{i=1}^{n}\pi_i^{-1}[\varrho(y_{(i)}, \theta_0) - q] \simeq \frac{1}{N}\sum_{i=1}^{n}\pi_i^{-1}[\delta\{y_i \leq \theta_0\} - q]$$

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 $g_i(\theta) = y_i - Nn^{-1}\theta\pi_i$. The approach described in §3 is used to compute confidence intervals. For the pseudo empirical likelihood (EL2) approach, the point estimator is the Hájek (1971) estimator.

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

6. Simulation study

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

$$y_i = 3 + a_i + \varphi\, e_i, \tag{35}$$

where a_i follows an exponential distributions with rate parameters equal to one and $e_i \sim \chi_1^2 - 1$. The π_i are proportional to $a_i + 2$. The constant 2 is added to a_i to avoid having very small π_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 φ 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 θ_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.

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

(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 nonresponse 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 linearisation 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

Aires, N. (2000). Comparisons between conditional Poisson sampling and Pareto πps sampling designs. *Journal of Statistical Planning and Inference*, *82*, 1–15.

Berger, Y. G. (1998a). Rate of convergence to asymptotic variance for the Horvitz-Thompson estimator. *Journal of Statistical Planning and Inference*, *74*, 149–168.

Berger, Y. G. (1998b). Rate of convergence to normal distribution for the Horvitz-Thompson estimator. *Journal of Statistical Planning and Inference*, *67*, 209–226.

Berger, Y. G. (2005a). Variance estimation with Chao's sampling scheme. *Journal of Statistical Planning and Inference*, *127*, 253–77.

Berger, Y. G. (2005b). Variance estimation with highly stratified sampling designs with unequal probabilities. *Australian and New Zealand Journal of Statistics*, *47*(3), 365–373.

Berger, Y. G. (2007). A jackknife variance estimator for unistage stratified samples with unequal probabilities. *Biometrika*, *94*(4), 953–964.

Berger, Y. G. (2011). Asymptotic consistency under large entropy sampling designs with unequal probabilities. *Pakistan Journal of Statistics, Festschrift to honour Ken Brewer's 80th birthday*, *27*(4), 407–426.

Berger, Y. G., & De La Riva Torres, O. (2012a). Estimation of confidence intervals using a new empirical likelihood approach. *Proceedings of European Conference on quality in Official Statistics - Q2012*, (p. 10pp).

Berger, Y. G., & De La Riva Torres, O. (2012b). A unified theory of empirical likelihood ratio confidence intervals for survey data with unequal probabilities and non negligible sampling fractions. *S3RI Methodology Working Papers* `http://eprints.soton.ac.uk/id/eprint/337688`, (p. 24pp).

Berger, Y. G., Tirari, M. E. H., & Tillé, Y. (2003). Towards optimal regression estimation in sample surveys. *Australian and New Zealand Journal of Statistics*, *45*, 319–329.

Binder, D. A., & Kovačević, M. S. (1995). Estimating some measure of income inequality from survey data: an application of the estimating equation approach. *Survey Methodology*, *21*(2), 137–145.

Brewer, K. R. W. (2002). *Combined Survey Sampling Inference*. London: Arnold.

Brewer, K. R. W., & Donadio, M. E. (2003). The high entropy variance of the Horvitz-Thompson estimator. *Survey Methodology*, *29*, 189–196.

Chambers, R. L., Steel, D. G., Wang, S., & Welsh, A. (2012). *Maximum Likelihood Estimation for Sample Surveys*. New York: Chapman and Hall/CRC.

Chao, M. T. (1982). A general purpose unequal probability sampling plan. *Biometrika*, *69*, 653–656.

Chen, J., Chen, S. R., & Rao, J. N. K. (2003). Empirical likelihood confidence intervals for the mean of a population containing many zero values. *The Canadian Journal of Statistics*, *31*(1), 53–68.

Chen, J., & Qin, J. (1993). Empirical likelihood estimation for finite populations and the effective usage of auxiliary information. *Biometrika*, *80*(1), 107–116.

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

Deville, J. C. (1999). Variance estimation for complex statistics and estimators: linearization and residual techniques. *Survey Methodology*, *25*, 193–203.

Deville, J. C., & Särndal, C. E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association*, *87*(418), 376–382.

Deville, J. C., & Tillé, Y. (2005). Variance approximation under balanced sampling. *Journal of Statistical Planning and Inference*, *128*, 569–591.

Durbin, J. (1953). Some results in sampling theory when the units are selected with unequal probabilities. *Journal of the Royal Statistical Society Series B*, *15*(2), 262–269.

Fay, B. E. (1991). A design-based perspective on missing data variance. *Proceeding of the 1191 Annual Research Conference. U.S. Bureau of the Census*, (pp. 429–440).

Fuller, W. A. (2009). Some design properties of a rejective sampling procedure. *Biometrika*, *96*, 933–944.

Godambe, V. (1966). A new approach to sampling from finite population I, II. *Journal of the Royal Statistical Society, Series B*, *28*, 310–328.

Hájek, J. (1964). Asymptotic theory of rejective sampling with varying probabilities from a finite population. *The Annals of Mathematical Statistics*, *35*(4), 1491–1523.

Hájek, J. (1971). Comment on a paper by D. Basu. in Foundations of Statistical Inference. Toronto: Holt, Rinehart and Winston.

Hájek, J. (1981). *Sampling from a Finite Population*. New York: Marcel Dekker.

Hartley, H. O., & Rao, J. N. K. (1968). A new estimation theory for sample surveys. *Biometrika*, *55*(3), 547–557.

Haziza, D., Mecatti, F., & Rao, J. N. K. (2004). Comparison of variance estimators under Rao-Sampford method: a simulation study. *Proceedings of the Survey Methods Section, American Statistical Association*.

Henderson, T. (2006). Estimating the variance of the Horvitz-Thompson estimator. *M.Phil. thesis, School of Finance and Applied Statistics, The Australian National University.*.

Horvitz, D. G., & Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, *47*(260), 663–685.

Huang, E. T., & Fuller, W. A. (1978). Nonnegative regression estimation for survey data. *Proceedings Social Statistics Section American Statistical Association*, (pp. 300–303).

Hudson, D. J. (1971). Interval estimation from the likelihood function. *Journal of the Royal Statistical Society*, (pp. 256–262).

Isaki, C. T., & Fuller, W. A. (1982). Survey design under the regression super-population model. *Journal of the American Statistical Association*, (pp. 89–96).

Kim, J. K. (2009). Calibration estimation using empirical likelihood in survey sampling. *Statistica Sinica*, *19*, 145–157.

Krewski, D., & Rao, J. N. K. (1981). Inference from stratified sample: properties of linearization jackknife, and balanced repeated replication methods. *The Annals of Statistics*, *9*, 1010–1019.

Montanari, G. (1987). Post sampling efficient qr-prediction in large sample survey. *International Statistical Review*, *55*, 191–202.

Ohlsson, E. (1986). Normality of the Rao, Hartley, Cochran estimator: An application of the martingale CLT. *Scandinavian Journal of Statistics*, *13*(1), 17–28.

Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, *75*(2), 237–249.

Owen, A. B. (2001). *Empirical Likelihood*. New York: Chapman & Hall.

Prášková, Z., & Sen, P. K. (2009). *Sample Surveys: Inference and Analysis*, vol. 29B of *Handbook of statistics*, chap. Asymptotic in Finite Population Sampling, (pp. 489–522). The Netherlands: North-Holland.

R Development Core Team (2006). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
URL http://www.R-project.org

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

Rao, J. N. K. (1966). Alternative estimators in pps sampling for multiple characteristics. *Sankhyā*, *A28*, 47–60.

Rao, J. N. K., Hartley, H. O., & Cochran, W. G. (1962). On a simple procedure of unequal probability sampling without replacement. *Journal of the Royal Statistical Society. Series B (Methodological)*, *24*(2), pp. 482–491.

Rao, J. N. K., & Wu, W. (2009). *Sample Surveys: Inference and Analysis*, vol. 29B of *Handbook of statistics*, chap. Empirical Likelihood Methods, (pp. 189–208). The Netherlands: North-Holland.

Salehi, M., Mohammadi, M., Rao, J. N. K., & Berger, Y. G. (2010). Empirical likelihood confidence intervals for adaptive cluster sampling. *Environmental and Ecological Statistics*, *17*, 111–123.

Sampford, M. R. (1967). On sampling without replacement with unequal probabilities of selection. *Biometrika*, *54*(3/4), 499–513.

Särndal, C. E. (1996). Efficient estimators with simple variance in unequal probability sampling. *Journal of the American Statistical Association*, *91*, 1289–1300.

Särndal, C. E., & Lundström (2005). *Estimation in Surveys with Nonresponse*. Chichester: Wiley.

Särndal, C.-E., Swensson, B., & Wretman, J. (1992). *Model Assisted Survey Sampling*. New York: Springer-Verlag.

Sen, P. K. (1953). On the estimate of the variance in sampling with varying probabilities. *Journal of the Indian Society of Agricultural Statistics*, (5), 119–127.

Shao, J., & Steel, P. (1999). Variance estimation for survey data with composite imputation and nonnegligible sampling fractions. *Journal of the American Statistical Association*, *94*, 254–265.

Tillé, Y. (2006). *Sampling Algorithms*. Springer Series in Statistics. New York: Springer.

Víšek, J. (1979). Asymptotic distribution of simple estimate for rejectif, sampford and successive sampling. *Contribution to Statistics, Jaroslav Hajek Memorial Volume. Academia of Prague, Czech Republic*, (pp. 71–78).

Wilks, S. S. (1938). Shortest average confidence intervals from large samples. *The Annals of Mathematical Statistics*, *9*(3), 166–175.

Wu, C. (2005). Algorithms and R codes for the pseudo empirical likelihood method in survey sampling. *Survey Methodology*, *31*(2), 239–243.

Wu, C., & Rao, J. N. K. (2006). Pseudo-empirical likelihood ratio confidence intervals for complex surveys. *The Canadian Journal of Statistics*, *34*(3), 359–375.

Wu, C., & Rao, J. N. K. (2010). Bootstrap procedures for the pseudo empirical likelihood method in sample surveys. *Statistics and Probability Letters*, *80*, 1472–1478.

Yates, F., & Grundy, P. M. (1953). Selection without replacement from within strata with probability proportional to size. *Journal of the Royal Statistical Society, series B*, *1*, 253–261.

Zhong, B., & Rao, J. N. K. (1996). Empirical likelihood inference under stratified random sampling using auxiliary information. *ASA Proceedings of the Section on Survey Research Methods*, *87*, 798–803.

Zhong, B., & Rao, J. N. K. (2000). Empirical likelihood inference under stratified random sampling using auxiliary population. *Biometrika*, *87*(4), 929–938.