Empirical likelihood confidence intervals for complex sampling designs

Y.G. Berger and O. De La Riva Torres
University of Southampton, UK

Summary. We define an empirical likelihood approach which gives consistent design-based confidence intervals which can be calculated without the need of variance estimates, design-effects, re-sampling, joint-inclusion probabilities and linearisation, even when the point estimator is not linear. It can be used to construct confidence intervals for a large class of sampling designs and estimators which are solutions of estimating equations. It can be used for means, regressions coefficients, quantiles, totals or counts even when the population size is unknown. It can be used with large sampling fractions and naturally includes calibration constraints. It can be viewed as an extension of the empirical likelihood approach to complex survey data. This approach is computationally simpler than the pseudo empirical likelihood and the bootstrap approaches. The simulation study shows that the proposed confidence interval may give better coverages than the confidence intervals based on linearisation, bootstrap and pseudo empirical likelihood. Our simulation study shows that under complex sampling designs, standard confidence intervals based upon normality may have poor coverages, because point estimators may not follow a normal sampling distribution and their variance estimators may be biased.

Keywords: Calibration, Design-based approach, Estimating equations, Finite population corrections, Hájek estimator, Horvitz-Thompson estimator, Regression estimator, Stratification, Unequal inclusion probabilities.

1. Introduction

Survey data is often used to compute complex estimators, such as quantiles, poverty indicators, M-estimators or parameters of population models. The sampling distribution of these estimators may not be normal when the variable of interest is skewed. Furthermore, asymptotic linearised variance estimators may be biased. Therefore, standard confidence intervals based upon normality can have poor coverages and bounds out of the range of the parameter space. For example, standard lower bounds can be negative with positive parameters. The coverage and the tail error rates can be also different from the nominal levels (e.g. 95% and 2.5%). On the other hand, empirical likelihood confidence intervals may have better coverages in this situation, as empirical likelihood confidence intervals are determined by the distribution of the data (e.g. Owen, 2001) and as the range of the parameter space is preserved.

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 unique solution of the following estimating equation (Godambe, 1960).

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

where $g_i(\theta)$ is a function of θ and of the characteristics of the unit i, such as the variables of interest and the auxiliary variables. This function does not need to be differentiable. We assume that the $g_i(\theta_0)$ satisfy the conditions (21)-(24) which are discussed in § 5. Note that the $g_i(\theta)$ and θ_0 can be vectors, but for simplicity, we consider that they are scalars. For example, θ_0 is the population mean $N^{-1}\sum_{i\in U} y_i$, when $g_i(\theta) = y_i - \theta$; where the y_i denote the values of a variable of interest. Other examples are ratios, quantiles (see § 7.1), low income measures, regression coefficients, M-estimators (e.g. Qin and Lawless, 1994; Binder and Kovacević, 1995). The aim of this paper is to derive an empirical likelihood point estimator and an empirical likelihood confidence interval for θ_0. Suppose that θ_0 is estimated from the data of a sample s of size n selected randomly using a sampling design. The quantity nN^{-1} denotes the sampling fraction. We adopt a non-parametric design-based approach; where the sampling distribution is specified by the sampling design and where θ_0 and the values of the variables are fixed (non-random) quantities. First, we suppose that we do not have nonresponse. In § 7.3, we show how the proposed approach can be extended under a uniform response mechanism.

Under the design-based approach, the standard likelihood function is flat and cannot be used for inference (Godambe, 1966). Hartley and Rao (1968) introduced an empirical likelihood approach. Owen (1988) brought this approach into the mainstream statistics. The main purpose of this article is to extend empirical likelihood approaches for complex sampling designs.

Chen and Sitter (1999) pointed out that standard empirical likelihood approaches cannot be directly used without taking the sampling design into account. They proposed a pseudo empirical likelihood approach which can be used with complex sampling designs. This approach is not entirely satisfactory, because its empirical likelihood function is not a standard one and its empirical log-likelihood ratio function does not converge to a chi-squared distribution (Wu and Rao, 2006). For confidence intervals, the pseudo empirical log-likelihood ratio function needs to be adjusted by a ratio of variances (the design effect) which needs to be estimated. Wu and Rao (2006) proposed two pseudo empirical likelihood approaches denoted pseudo-EL1 and pseudo-EL2. The fact that the pseudo empirical likelihood approaches rely on variance estimates, limits the range of the parameters it can be applied to. The design effect needs to be estimated, incurring an additional variability which may affect the coverages of confidence intervals. The simulation study in § 7 shows that, for means, the proposed confidence interval may give better coverages and tail error rates than the pseudo empirical likelihood confidence intervals.

Empirical likelihood approaches under Poisson sampling are studied in detail by Kim (2009) and Chen and Kim (2014). Kim (2009) proposed an empirical likelihood point estimator under Poisson sampling with negligible sampling fraction. Chen and Kim (2014) stated that the empirical log-likelihood ratio function based on Kim (2009) empirical likelihood function follows a chi-squared distribution asymptotically under Poisson sampling with negligible sampling fractions. Chen and Kim (2014) proposed a population empirical log-likelihood ratio function which has a chi-squared distribution asymptotically with negligible sampling fractions. With large sampling fractions, these empirical log-likelihood ratio functions do not necessarily follow a chi-squared distribution asymptotically. In this paper, we do not assume that the sampling fraction is negligible. The sampling sampling designs considered are different and more complex than the Poisson sampling design.

We consider four sampling designs commonly used in practice: (i) (stratified) sampling with replacement (*pps sampling*), where n denotes the number of draws and the sample is the set of n observations (units can be selected more than once); (ii) (stratified) sampling without replacement (*πps sampling*); (iii) multi-stage sampling; (iv) two phase sampling with one phase being a uniform response mechanism. We show that the proposed confidence interval gives the correct nominal coverage asymptotically under these designs.

The calculation of the proposed confidence interval does not rely on variance estimates, or unknown population parameters, such as design effect or population size. The proposed approach is different from and not an adjustment of the pseudo empirical likelihood approaches or Kim (2009)'s approach. It is computationally simpler than the pseudo empirical likelihood and bootstrap approaches. It can be used with a wide class of parameters of interest and with large sampling fractions. Our simulation study shows that, for means and quantiles, the proposed confidence interval gives good coverages even when the variables of interest are skewed or contains outlying units which is a common situation with business surveys and social surveys on wealth or income.

The pseudo-EL2 approach can only be used when N is known. However, N may be unknown with social household surveys; for example, when the total number of households is unknown. The proposed approach can be used even when N is unknown.

In § 2, we define the proposed empirical likelihood function. In § 3, we define the maximum empirical likelihood estimator. In § 4, we define the asymptotic framework and give some asymptotic properties of this estimator. In § 5, we show that, under a series of regularity conditions, the empirical log-likelihood ratio function follows a chi-squared distribution asymptotically. We show how this property can be used to derive confidence intervals. In § 5.3, we show how the auxiliary variables can be taken into account. In § 6, we define a penalised empirical likelihood approach which takes the large sampling fractions into account (under πps sampling). In § 7, we compare, via a series of simulations, the proposed empirical likelihood approach with the pseudo empirical likelihood, the bootstrap and the linearisation approaches. We also show how the proposed approach can be implemented under multi-stage sampling and under a uniform response mechanism (two-phase sampling). In § 8, we have some concluding remarks and a summary of the proposed approaches in Table 5.

2. Empirical log-likelihood function

Consider the following *empirical log-likelihood function*

$$\ell(m) = \log\left(\prod_{i\in s} m_i\right) = \sum_{i\in s}\log(m_i), \qquad (2)$$

where $\prod_{i\in s}$ and $\sum_{i\in s}$ denote the product and the sum over the sampled units. The quantities m_i are unknown positive scale loads which will be estimated. Hartley and Rao (1969) showed that (2) is an log-empirical likelihood function under unequal probability sampling with replacement. Chen and Qin (1993) proposed to use (2) under simple random sampling. Zhong and Rao (2000) used (2) under stratified simple random sampling. The aim is to show that (2) can be used for inference: point estimation, confidence intervals and tests. Note that the pseudo empirical likelihood approach is not based on (2) and is based on the Kullback-Leibler distance.

The m_i can be estimated by the values $\widehat{m}_i$ which maximise $\ell(m)$ subject to the constraints $m_i \geq 0$ and

$$\sum_{i\in s} m_i \boldsymbol{c}_i \quad = \quad \boldsymbol{C}; \qquad (3)$$

where $\boldsymbol{c}_i$ is a $Q\times 1$ vector associated with the i-th sampled unit and $\boldsymbol{C}$ is a $Q\times 1$ vector. The vectors $\boldsymbol{c}_i$ are related to the design and the auxiliary variables. Possible choices for $\boldsymbol{c}_i$ and $\boldsymbol{C}$ are given in Table 5 and discussed throughout this paper. Note that the $\boldsymbol{c}_i$ and $\boldsymbol{C}$ cannot be any vectors, as they must obey the regularity conditions (9)-(13) given in § 3. The vector $\boldsymbol{C}$

is not necessarily a vector of fixed quantities (see Table 5). Hence $\boldsymbol{C}$ can be fixed or random. The constraint (3) resembles the constraint used in calibration (e.g. Huang and Fuller, 1978; Deville and Särndal, 1992). However, in this paper, we consider that the quantity $\boldsymbol{C}$ is not necessarily a vector of population totals of auxiliary variables. A comparison with calibration can be found in § 8.

Suppose that the sample size n is a fixed (non-random) quantity. Let π_i denote the first-order inclusion probability of unit i under sampling without replacement (πps sampling). Under sampling with replacement (pps sampling), $\pi_i = np_i$, where p_i is the probability to select unit i at the i-th draw (Hansen and Hurwitz, 1943) and n denotes the (fixed) number of draws. In this case, the sample s is a set of n labels of the units selected after n draws. This set may contain the same label several times, when some units are selected more than once. We consider that the $\boldsymbol{c}_i$ contain the π_i; that is, we assume that the $\boldsymbol{c}_i$ and $\boldsymbol{C}$ are such that there exists a $Q \times 1$ vector $\boldsymbol{t}$ such that $\boldsymbol{t}^\top \boldsymbol{c}_i = \pi_i$ and $\boldsymbol{t}^\top \boldsymbol{C} = \sum_{i \in U} \pi_i$. We have that (3) implies that $\sum_{i \in s} m_i \boldsymbol{t}^\top \boldsymbol{c}_i = \boldsymbol{t}^\top \boldsymbol{C}$ or equivalently

$$\sum_{i \in s} m_i \pi_i = \sum_{i \in U} \pi_i = n \cdot \tag{4}$$

In other words, the constraint (3) is such that that the constraint (4) always holds. For example, when we have a single stratum, we can use $Nn^{-1}\pi_i$ as the first component of $\boldsymbol{c}_i$ and N as the first component of $\boldsymbol{C}$. In this case, $\boldsymbol{t} = (nN^{-1}, 0, \ldots, 0)^\top$.

Note that the constraint (4) reduces to $\sum_{i \in s} m_i \pi_i = n$, because the sample size (or the number of draws) is fixed. Thus, the equation (4) can be interpreted as a design constraint. Under equal probability sampling, we have that $\pi_i = n/N$, and the constraint (4) reduces to $\sum_{i \in s} m_i = N$ which is the constraint adopted under equal probability sampling (e.g. Rao and Wu, 2009). Note that we do not impose that $\sum_{i \in s} m_i = N$ always holds (except when $\pi_i = n/N$), because N may be unknown. If N is known and we want to impose the constraint $\sum_{i \in s} m_i = N$, we need to consider an additional constraint $\sum_{i \in s} m_i x_i = N$ with $x_i = 1$, and treat x_i as an auxiliary variable (see § 5.3).

We assume that the $\boldsymbol{C}$ is an inner point of the convex conical hull formed by the sample observations $\{\boldsymbol{c}_i : i \in s\}$ so that a unique solution to (3) exists, as the objective function (2) is a strictly concave function. This solution can be found by using the Lagrangian function, $Q(m, \boldsymbol{\eta}) = \sum_{i \in s} \log(m_i) - (\boldsymbol{t}+\boldsymbol{\eta})^\top \left(\sum_{i \in s} m_i \boldsymbol{c}_i - \boldsymbol{C}\right)$. The values of m_i and $\boldsymbol{\eta}$ which maximise $Q(m, \boldsymbol{\eta})$ are the solutions of the following set of equations $\partial Q(m, \boldsymbol{\eta})/\partial m_i = 0$ and $\partial Q(m, \boldsymbol{\eta})/\partial(\boldsymbol{t}+\boldsymbol{\eta}) = 0$. The solution is

$$\widehat{m}_i = \left[(\boldsymbol{t}+\boldsymbol{\eta})^\top \boldsymbol{c}_i\right]^{-1} = \left(\pi_i + \boldsymbol{\eta}^\top \boldsymbol{c}_i\right)^{-1}, \tag{5}$$

as $\boldsymbol{t}^\top \boldsymbol{c}_i = \pi_i$. The quantity $\boldsymbol{\eta}$ is such that the constraint (3) holds. This quantity can be computed using a modified Newton-Raphson algorithm as in Chen *et al.* (2002). This algorithm ensures that $\widehat{m}_i > 0$. Note that it is not always necessary to know N in order to the compute $\boldsymbol{\eta}$ and $\widehat{m}_i$.

3. Maximum empirical likelihood estimator

In this section, we define the maximum empirical likelihood point estimate as the value which minimises the empirical log-likelihood ratio function defined by (6).

Let the $\widehat{m}_i$ be the values which maximise (2) subject to the constraints $m_i \geq 0$ and (3), for given $\boldsymbol{c}_i$ and $\boldsymbol{C}$. Let $\ell(\widehat{m}) = \sum_{i \in s} \log(\widehat{m}_i)$ be the maximum value of (2). Let the $\widehat{m}_i^\star(\theta)$ be the values which maximise (2) subject to the constraints $m_i \geq 0$ and $\sum_{i \in s} m_i \boldsymbol{c}_i^\star = \boldsymbol{C}^\star$ with

$\boldsymbol{c}_i^\star = (\boldsymbol{c}_i^\top, g_i(\theta))^\top$ and $\boldsymbol{C}^\star = (\boldsymbol{C}^\top, 0)^\top$, for a given θ. Let $\ell(\widehat{m}^\star, \theta) = \sum_{i\in s} \log(\widehat{m}_i^\star(\theta))$ be the maximum value of (2) subject to these constraints involving $\boldsymbol{c}_i^\star$. The *empirical log-likelihood ratio function* is defined by the following function of θ.

$$\widehat{r}(\theta) \quad = \quad 2\left\{\ell(\widehat{m}) - \ell(\widehat{m}^\star, \theta)\right\} \cdot \tag{6}$$

The *maximum empirical likelihood estimate* $\widehat{\theta}$ of θ_0 is defined by the value of θ which minimises the function $\widehat{r}(\theta)$. Note that $\widehat{r}(\theta) \geq 0$, for all θ. Thus $\widehat{\theta}$ is the solution of $\widehat{r}(\theta) = 0$. Assuming that the $g_i(\theta)$ are such that the estimating equation

$$\widehat{G}(\theta) = 0, \quad \text{with} \quad \widehat{G}(\theta) = \sum_{i\in s} \widehat{m}_i \; g_i(\theta), \tag{7}$$

has a unique solution, we have that $\widehat{\theta}$ is the solution of (7) as it implies $\widehat{m}_i^\star(\widehat{\theta}) = \widehat{m}_i$ (for all i) and $\widehat{r}(\widehat{\theta}) = 0$.

Note that when $\boldsymbol{c}_i = Nn^{-1}\pi_i$ and $\boldsymbol{C} = N$ (or equivalently $\boldsymbol{c}_i = \pi_i$ and $\boldsymbol{C} = n$), we have that $\boldsymbol{\eta} = 0$ and $\widehat{m}_i = \pi_i^{-1}$. Under πps sampling, $\widehat{G}(\theta)$ is given by (18) and $\widehat{\theta}$ is the Horvitz and Thompson (1952) estimator given by $\widehat{Y}_\pi = \sum_{i\in s} y_i \pi_i^{-1}$ when $g_i(\theta) = y_i - n^{-1}\theta\pi_i$. (Under pps sampling, we obtain the Hansen and Hurwitz (1943) estimator). When $g_i(\theta) = y_i - \theta N^{-1}$, $\widehat{\theta}$ is the Hájek (1971) ratio estimator $\widehat{Y}_H = N\widehat{N}_\pi^{-1}\widehat{Y}_\pi$, where $\widehat{N}_\pi = \sum_{i\in s} \pi_i^{-1}$. The estimator $\widehat{Y}_\pi$ is more efficient than $\widehat{Y}_H$ when the variable of interest is correlated with the inclusion probabilities (Rao, 1966). Note that we cannot obtain $\widehat{Y}_\pi$ with the pseudo-EL2 approach.

4. Asymptotic properties

In order to derive asymptotic properties of the proposed empirical likelihood approach, we need to define an asymptotic framework and assume a set of regularity conditions.

Consider that $n \to \infty$ and $N \to \infty$. The stochastic orders $O(\cdot)$, $o(\cdot)$, $O_p(\cdot)$ and $o_p(\cdot)$ are defined according to this asymptotic framework, where the convergence in probability is with respect to the sampling design (e.g. Isaki and Fuller, 1982). We do not assume $nN^{-1} = o(1)$. Other empirical likelihood approaches (Owen, 1988; Kim, 2009) assume that $nN^{-1} = o(1)$. This condition is restrictive because many surveys (e.g. business surveys) use non negligible sampling fractions.

Consider that the sampling design is such that the following regularity conditions hold.

$$\max\{nN^{-1}\pi_i^{-1} : \; i \in s\} \quad = \quad O_p(1), \tag{8}$$

$$N^{-1}\|\widehat{\boldsymbol{C}}_\pi - \boldsymbol{C}\| \quad = \quad O_p(n^{-\frac{1}{2}}), \tag{9}$$

$$\max\{\|\boldsymbol{c}_i\| : i \in s\} \quad = \quad o_p(n^{\frac{1}{2}}), \tag{10}$$

$$\|\widehat{\boldsymbol{S}}\| \quad = \quad O_p(1), \tag{11}$$

$$\|\widehat{\boldsymbol{S}}^{-1}\| \quad = \quad O_p(1), \tag{12}$$

$$\frac{1}{nN^\tau} \sum_{i\in s} \frac{\|\boldsymbol{c}_i\|^\tau}{\pi_i^\tau} \quad = \quad O_p(n^{-\tau}) \quad (\tau = 2,\ 3,\ 4), \tag{13}$$

with

$$\widehat{\boldsymbol{S}} \quad = \quad -\frac{n}{N^2} \sum_{i\in s} \frac{\boldsymbol{c}_i \boldsymbol{c}_i^\top}{\pi_i^2}, \quad \text{and} \quad \widehat{\boldsymbol{C}}_\pi = \sum_{i\in s} \frac{\boldsymbol{c}_i}{\pi_i} \cdot \tag{14}$$

The quantity $\|\boldsymbol{A}\| = \mathrm{trace}(\boldsymbol{A}^\top \boldsymbol{A})^{1/2}$ denotes the Euclidean (Frobenius) norm.

The condition (8) is the key condition. It ensures that the inclusion probabilities are not disproportionately small compared to the sampling fraction. This condition was proposed by Krewski and Rao (1981, p. 1014). The condition (9) holds when the law of large numbers holds. For unequal probability sampling, Isaki and Fuller (1982) gave conditions under which (9) holds (see also Krewski and Rao, 1981, p. 1014). Chen and Sitter (1999, Appendix 2) showed that the condition (10) holds for common unequal probability sampling designs. It can be shown that the conditions (11) and (12) hold when $-\widehat{\boldsymbol{S}}$ is positive definite and when there exists a positive definite matrix $-\boldsymbol{S}$ such that $\|\widehat{\boldsymbol{S}} - \boldsymbol{S}\| = o_p(1)$ and $\|\boldsymbol{S}\| = O(1)$. The condition (13) is a Lyapunov-type condition for the existence of moments (e.g. Krewski and Rao, 1981, p. 1014, Deville and Särndal, 1992, p. 381). In § 5, we will see that the conditions (9), (11), (12) and (13) are trivial when we do not have auxiliary variables. In § 5.3, we will see that these conditions are needed when we have auxiliary variables.

Using the Lemma 1 of the supplementary materials, we have that $Nn^{-1}\|\boldsymbol{\eta}\| = O_p(n^{-\frac{1}{2}})$, where $\boldsymbol{\eta}$ is given in (5). This implies the following approximation for $\boldsymbol{\eta}$ (see Lemma 2 in the Appendix A of the supplementary materials).

$$\boldsymbol{\eta} = nN^{-2}\widehat{\boldsymbol{S}}^{-1}(\boldsymbol{C} - \widehat{\boldsymbol{C}}_\pi) + nN^{-1}\widehat{\boldsymbol{e}}; \tag{15}$$

where $\widehat{\boldsymbol{C}}_\pi$ is defined in (14) and $\widehat{\boldsymbol{e}}$ is such that $\|\widehat{\boldsymbol{e}}\| = O_p(n^{-1})$. Furthermore, we have that $\widehat{m}_i = \pi_i^{-1} - \pi_i^{-2}\boldsymbol{c}_i^\top\boldsymbol{\eta}(1+v_i)^{-1}$ where $v_i = \pi_i^{-1}\boldsymbol{c}_i^\top\boldsymbol{\eta}$. Hence by substituting this equation into equation (7), we obtain $\widehat{G}(\theta) = \widehat{G}(\theta)_\pi - \sum_{i\in s} g_i(\theta)\boldsymbol{c}_i^\top\boldsymbol{\eta}\pi_i^{-2}(1+v_i)^{-1}$; where $\widehat{G}(\theta)_\pi$ is defined by (18). By replacing (15) into the previous expression of $\widehat{G}(\theta)$, we obtain (see Appendix A)

$$\widehat{G}(\theta) \quad = \quad \widehat{G}(\theta)_{reg} + o_p(Nn^{-\frac{1}{2}}); \tag{16}$$

where $\widehat{G}(\theta)_{reg}$ is the following regression estimator.

$$\widehat{G}(\theta)_{reg} \quad = \quad \widehat{G}(\theta)_\pi + \widehat{\boldsymbol{B}}(\theta)^\top(\boldsymbol{C} - \widehat{\boldsymbol{C}}_\pi), \tag{17}$$

where

$$\widehat{G}(\theta)_\pi \quad = \quad \sum_{i\in s} \breve{g}_i(\theta), \tag{18}$$

and $\breve{g}_i(\theta) = g_i(\theta)\pi_i^{-1}$. The quantity $\widehat{\boldsymbol{B}}(\theta)$ is a vector of regression coefficients defined by

$$\widehat{\boldsymbol{B}}(\theta) \quad = \quad \left(\sum_{i\in s} \frac{1}{\pi_i^2}\boldsymbol{c}_i\boldsymbol{c}_i^\top\right)^{-1} \sum_{i\in s} \frac{1}{\pi_i^2} g_i(\theta)\boldsymbol{c}_i\cdot \tag{19}$$

In the supplementary materials, we show rigorously that the equation (16) holds for sampling designs which are such that the conditions (8)-(13) hold. We also assume that θ is such that following condition holds.

$$\frac{1}{nN^2}\sum_{i\in s} \breve{g}_i(\theta)^2 \quad = \quad O_p(n^{-2})\cdot \tag{20}$$

In this paper, the vector $\boldsymbol{c}_i$ always contains the π_i or the stratification variables (see § 5.2). Thus, it can be shown that (17) converges to a design-optimal regression estimator under a single stage unequal probability pps sampling design (Berger *et al.*, 2003). Note that (16) implies that the maximum empirical likelihood is asymptotically design-consistent because (17) is a consistent regression estimator.

Kim (2009) proposed an empirical likelihood estimator which is equivalent to (17) in some particular cases. For example, if we use the auxiliary variables $x_i = \pi_i$ in Kim (2009)'s estimator, we obtain an estimator which is equivalent to (17), when $\boldsymbol{c}_i = (Nn^{-1}\pi_i, 1)^\top$ and $\boldsymbol{C} = (N, N)^\top$. However, (17) is different from Kim (2009)'s estimator when $\boldsymbol{c}_i$ does not contain the constant one or under stratified designs (see § 5.2). There are situations when we do not want to include the constant one within $\boldsymbol{c}_i$, for example, when N is unknown.

5. Empirical likelihood confidence intervals

In order to derive standard confidence intervals, we need unbiased point estimators following a normal distribution and unbiased variance estimators. However, the sampling distribution may not be normal for a given sample size, despite the fact that a point estimator may be asymptotically normal. Furthermore, linearised variance estimators may be biased. The main advantage of the empirical likelihood approach is its capability of deriving non-parametric confidence intervals which do not depend on variance estimates of $\widehat{\theta}$ and do not rely directly on the normality of $\widehat{\theta}$. However, this approach depends on the normality of $\widehat{G}(\theta_0)_\pi$ (see (21) below).

Empirical likelihood confidence intervals rely on the conditions (21)-(24) given below. Consider that the sampling design and the $g_i(\theta_0)$ are such that the following regularity conditions hold.

$$\widehat{G}(\theta_0)_\pi \; var[\widehat{G}(\theta_0)_\pi]^{-\frac{1}{2}} \;\to\; N(0,1), \tag{21}$$

$$N^{-1}\widehat{G}(\theta_0)_\pi \;=\; O_p(n^{-\frac{1}{2}}), \tag{22}$$

$$\max\{|g_i(\theta_0)| : i \in s\} \;=\; o_p(n^{\frac{1}{2}}), \tag{23}$$

$$\frac{1}{nN^\tau}\sum_{i\in s}|\breve{g}_i(\theta_0)|^\tau \;=\; O_p(n^{-\tau}) \quad (\tau = 2,\ 3,\ 4); \tag{24}$$

where $var[\widehat{G}(\theta_0)_\pi]$ denotes the design-based variance of $\widehat{G}(\theta_0)_\pi$ and $\breve{g}_i(\theta_0) = g_i(\theta_0)\pi_i^{-1}$.

The conditions (22), (23) and (24) ensure that the conditions (9)-(13) hold with $\boldsymbol{c}_i^\star$ when $\theta = \theta_0$. Chen and Sitter (1999) showed that (23) holds for common sampling designs. The condition (24) is a Lyapunov-type condition for the existence of moments. As θ_0 is a constant, $\widehat{G}(\theta_0)_\pi$ is an Horvitz and Thompson (1952) estimator. Isaki and Fuller (1982) gave regularity conditions under which (22) holds (law of large numbers). Hájek (1964), Víšek (1979) and Berger (1998) gave regularity conditions for the asymptotic normality of the Horvitz and Thompson (1952) estimator. Under pps sampling, the $\breve{g}_i(\theta_0)$ are independent and standard large sample theory can be used to show the normality (e.g. Prášková and Sen, 2009). Based on these evidences, it is reasonable to consider that (21) holds, as $E(\widehat{G}(\theta_0)_\pi) = G(\theta_0) = 0$. Note that the classical empirical likelihood approach and the pseudo empirical likelihood approaches also rely on (21) (e.g. Owen, 1988, p. 242, Owen, 2001, p. 219, Wu and Rao, 2006, p. 364).

The proposed empirical likelihood approach relies on the normality of $\widehat{G}(\theta_0)_\pi$, but not on the normality of $\widehat{\theta}$. That does not mean that $\widehat{\theta}$ is not asymptotically normal. It can be shown that the condition (21) implies that the point estimator is asymptotically normal under additional conditions such as $g_i(\theta)$ being differentiable and twice differentiable with respect to θ (Binder, 1983; Godambe and Thompson, 2009). Note that the differentiability is not necessary for empirical likelihood, although the differentiability holds for most parameters, except for quantiles which require additional conditions for normality (Francisco and Fuller, 1991). Thus, the condition (21) is weaker than the asymptotic normality of $\widehat{\theta}$, because the latter requires additional conditions. Even if $\widehat{\theta}$ is normal, we need an unbiased variance estimator, as a biased variance affects the coverage of confidence intervals. This can be an issue with quantiles (see Graf and Tillé (2014) and § 7.1). This is not a problem with empirical likelihood, because it does not rely on a variance estimate of $\widehat{\theta}$. This is the key advantage of empirical likelihood confidence intervals over standard confidence intervals. Note that we do not need to assume that $\widehat{\theta}$ is unbiased, although it will be asymptotically unbiased because of (17) and (22).

5.1. *Empirical likelihood confidence intervals for pps sampling with replacement*

In this §, we assume that the sample is selected according to a pps with replacement sampling design. The empirical log-likelihood ratio function (6) can be used to construct empirical likelihood confidence intervals. It requires that $\widehat{r}(\theta_0)$ follows asymptotically a chi-squared distribution with one degree of freedom under our regularity conditions. This property is the consequence of the Proposition 1 below.

The proposition 1 is valid for any sampling designs which satisfy the regularity conditions. However, in order for $\widehat{r}(\theta_0)$ to follow a chi-squared distribution asymptotically, (26) needs to be a consistent estimator for the variance. This is true under pps sampling with replacement.

Proposition 1. *Let $\boldsymbol{c}_i = Nn^{-1}\pi_i$, $\boldsymbol{C} = N$, $\boldsymbol{c}_i^\star = (\boldsymbol{c}_i^\top, g_i(\theta_0))^\top = (Nn^{-1}\pi_i, g_i(\theta_0))^\top$ and $\boldsymbol{C}^\star = (\boldsymbol{C}^\top, 0)^\top = (N, 0)^\top$, assuming that the sampling design is such that the conditions (8), (10), (22), (23) and (24) hold, we have that*

$$\widehat{r}(\theta_0) = \widehat{G}(\theta_0)_\pi^2 \ \widehat{var}_{pps}[\widehat{G}(\theta_0)_\pi]^{-1} + O_p(n^{-\frac{1}{2}}), \tag{25}$$

where

$$\widehat{var}_{pps}[\widehat{G}(\theta_0)_\pi] = \sum_{i\in s}\left(\breve{g}_i(\theta_0) - n^{-1}\widehat{G}(\theta_0)_\pi\right)^2 . \tag{26}$$

Find below a Sketch of the proof. The rigorous proof can be found in the supplementary materials (see Corollary 4). Chen and Kim (2014) obtained a result similar to (25) under Poisson sampling and $nN^{-1} = o(1)$.

The quantity Nn^{-1} in $\boldsymbol{c}_i$ guarantees that the conditions (11) and (12) hold. Note that the constraint $\sum_{i\in s} m_i\pi_i = n$ is equivalent to the constraint $\sum_{i\in s} m_i(Nn^{-1}\pi_i) = N$. Thus, we obtain the same $\widehat{m}_i$ and $\widehat{m}_i^\star(\theta)$, with $\boldsymbol{c}_i = \pi_i$ and $\boldsymbol{C} = n$, or with $\boldsymbol{c}_i = Nn^{-1}\pi_i$ and $\boldsymbol{C} = N$. Thus the equation (25) also holds with $\boldsymbol{c}_i = \pi_i$ and $\boldsymbol{C} = n$. This also means that $\widehat{m}_i$, $\widehat{m}_i^\star(\theta)$ and $\widehat{r}(\theta)$ can be calculated even when N is unknown.

The regularity conditions (9), (11), (12) and (13) do no appear in Proposition 1, because they are trivial with the $\boldsymbol{c}_i$ and $\boldsymbol{c}_i^\star$ given in the Proposition 1.

Sketch of the proof: Let $\ell(\pi) = \sum_{i\in s}\log(\pi_i)$. Using (5), we have that $-\log(\widehat{m}_i) - \log(\pi_i) = \log(1+v_i) = v_i - v_i^2/2 + O_p(v_i^3)$, where $v_i = \pi_i^{-1}\boldsymbol{c}_i^\top\boldsymbol{\eta}$. Hence, we have

$$\begin{aligned}-2\{\ell(\widehat{m}) + \ell(\pi)\} &= -2\sum_{i\in s}(\log(\widehat{m}_i) - \log(\pi_i)) = 2\sum_{i\in s} v_i - \sum_{i\in s} v_i^2 + O_p(n^{-\frac{1}{2}}) \\ &= 2\sum_{i\in s}(v_i + \widehat{m}_i\pi_i) - \sum_{i\in s} v_i^2 - 2n + O_p(n^{-\frac{1}{2}}) = \sum_{i\in s} v_i^2 + O_p(n^{-\frac{1}{2}}), \end{aligned} \tag{27}$$

because $\sum_{i\in s}\widehat{m}_i\pi_i = n$ (see (4)) and $\sum_{i\in s}(v_i+\widehat{m}_i\pi_i) = n+\sum_{i\in s} v_i^2 + O_p(n^{-\frac{1}{2}})$ (see (C.5) in the supplementary materials). Now using (15), we have

$$\sum_{i\in s} v_i^2 = \boldsymbol{\eta}^\top \sum_{i\in s}\frac{1}{\pi_i^2}\boldsymbol{c}_i\boldsymbol{c}_i^\top\boldsymbol{\eta} = (\widehat{\boldsymbol{C}}_\pi-\boldsymbol{C})^\top\widehat{\boldsymbol{\Sigma}}^{-1}(\widehat{\boldsymbol{C}}_\pi-\boldsymbol{C}) + O_p(n^{-\frac{1}{2}}); \tag{28}$$

where $\widehat{\boldsymbol{\Sigma}} = -N^2n^{-1}\widehat{\boldsymbol{S}}$. Thus, by using equations (27) and (28), we have

$$-2\{\ell(\widehat{m})+\ell(\pi)\} = (\widehat{\boldsymbol{C}}_\pi-\boldsymbol{C})^\top\widehat{\boldsymbol{\Sigma}}^{-1}(\widehat{\boldsymbol{C}}_\pi-\boldsymbol{C}) + O_p(n^{-\frac{1}{2}}), \tag{29}$$

and similarly

$$-2\{\ell(\widehat{m}^\star,\theta_0)+\ell(\pi)\} = (\widehat{\boldsymbol{C}}^\star_\pi-\boldsymbol{C}^\star)^\top\widehat{\boldsymbol{\Sigma}}^{\star-1}(\widehat{\boldsymbol{C}}^\star_\pi-\boldsymbol{C}^\star) + O_p(n^{-\frac{1}{2}}), \tag{30}$$

as the regularity conditions hold with $\boldsymbol{c}_i^\star$ and $\boldsymbol{C}^\star$, when $\theta=\theta_0$, where $\widehat{\boldsymbol{C}}^\star_\pi = \sum_{i\in s}\pi_i^{-1}\boldsymbol{c}_i^\star$ and $\widehat{\boldsymbol{\Sigma}}^\star = \sum_{i\in s}\pi_i^{-2}\boldsymbol{c}_i^\star\boldsymbol{c}_i^{\star\top}$. Using equations (6), (29), (30) and $\widehat{\boldsymbol{C}}_\pi-\boldsymbol{C}=0$, we obtain

$$\begin{aligned}
\widehat{r}(\theta_0) &= (\widehat{\boldsymbol{C}}^\star_\pi-\boldsymbol{C}^\star)^\top\widehat{\boldsymbol{\Sigma}}^{\star-1}(\widehat{\boldsymbol{C}}^\star_\pi-\boldsymbol{C}^\star) + O_p(n^{-\frac{1}{2}}) && (31)\\
&= (0,\widehat{G}(\theta_0)_\pi)\begin{pmatrix} N^2n^{-1} & Nn^{-1}\widehat{G}(\theta_0)_\pi \\ Nn^{-1}\widehat{G}(\theta_0)_\pi & \sum_{i\in s}\breve{g}_i(\theta_0)^2\end{pmatrix}^{-1}(0,\widehat{G}(\theta_0)_\pi)^\top + O_p(n^{-\frac{1}{2}}) \\
&= \widehat{G}(\theta_0)^2_\pi\left(\sum_{i\in s}\breve{g}_i(\theta_0)^2 - n^{-1}\widehat{G}(\theta_0)^2_\pi\right)^{-1} + O_p(n^{-\frac{1}{2}}), && (32)
\end{aligned}$$

as $\widehat{\boldsymbol{C}}^\star_\pi-\boldsymbol{C}^\star = (0,\widehat{G}(\theta_0)_\pi)^\top$. The Proposition follows from (32). □

Under sampling with replacement with unequal probabilities (pps sampling), (26) is an unbiased consistent estimator for the variance (Durbin, 1953). Hence $\widehat{r}(\theta_0)$ follows asymptotically a chi-squared distribution with one degree of freedom, by Slutsky's theorem and (21). Thus, the consistent α level empirical likelihood type confidence interval (see Shao and Tu, 1996, Definition 4.1) for the population parameter θ_0 is given by (e.g. Wilks, 1938)

$$\left\{\theta\ :\ \widehat{r}(\theta)\le\chi^2_1(\alpha)\right\} = \left[\ \min\left\{\theta|\ \widehat{r}(\theta)\le\chi^2_1(\alpha)\right\};\ \max\left\{\theta|\ \widehat{r}(\theta)\le\chi^2_1(\alpha)\right\}\ \right]; \tag{33}$$

where $\chi^2_1(\alpha)$ is the upper α-quantile of the chi-squared distribution with one degree of freedom. The quantity α is called the nominal coverage level. Note that $\widehat{r}(\theta)$ is a convex non-symmetric function with a minimum at $\theta=\widehat{\theta}$. This interval can be found using any root search method. This involves calculating $\widehat{r}(\theta)$ for several values of θ. If $g_i(\theta)$ and θ are $R\times 1$ vectors, the random variable $\widehat{r}(\theta_0)$ will converge to a chi-squared distribution with R degrees of freedom (see Oguz-Alper and Berger, 2014).

The p-value to test $H_0:\theta_0=\theta_0^\bullet$ is given by $\int_{\widehat{r}(\theta_0^\bullet)}^\infty f(x)dx$, where $f(x)$ is the density of the chi-squared distribution with one degree of freedom.

5.2. Stratification

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 with replacement with unequal probabilities from U_h. We assume that the number of strata H is bounded ($H=O(1)$).

The empirical likelihood estimator is still the solution of (7) where $\widehat{m}_i$ are the values which maximise (2) under a set of constraints (3) with $\boldsymbol{c}_i = Nn^{-1}\boldsymbol{z}_i$ and $\boldsymbol{C} = Nn^{-1}\boldsymbol{n}$;

where $\boldsymbol{z}_i$ are the values of the design (or stratification) variables defined by

$$\boldsymbol{z}_i = (z_{i1}, \ldots, z_{iH})^\top \quad \text{and where} \quad \boldsymbol{n} = (n_1, \ldots, n_H)^\top \tag{34}$$

denotes the vector of the strata sample sizes, with $z_{ih} = \pi_i$ when $i \in U_h$ and $z_{ih} = 0$ otherwise. It can be shown that $\widehat{m}_i = \pi_i^{-1}$.

We assume that conditions (8), (10), (22), (23) and (24) hold with $\boldsymbol{c}_i = Nn^{-1}\boldsymbol{z}_i$, $\boldsymbol{C} = Nn^{-1}\boldsymbol{n}$, $\boldsymbol{c}_i^\star = (\boldsymbol{c}_i^\top, g_i(\theta))^\top$ and $\boldsymbol{C}^\star = (\boldsymbol{C}^\top, 0)^\top$, when $\theta = \theta_0$. The regularity conditions (9), (11), (12), (13) are trivial in this case. Note that for the computation of $\widehat{m}_i$ and $\widehat{m}_i^\star(\theta)$, we can use $\boldsymbol{c}_i = \boldsymbol{z}_i$ and $\boldsymbol{C} = \boldsymbol{n}$ instead.

The Corollary 3 in the supplementary materials shows that (25) holds where the variance estimator is now the following stratified pps variance estimator.

$$\widehat{var}_{st}[\widehat{G}(\theta_0)_\pi] = \sum_{h=1}^{H} \left[\sum_{i \in s_h} \left(\breve{g}_i(\theta_0) - n_h^{-1} \widehat{G}_{\pi;h}(\theta_0) \right)^2 \right]; \tag{35}$$

where $\widehat{G}_{\pi;h}(\theta_0) = \sum_{i \in s_h} \breve{g}_i(\theta_0)$. This can be proven using (31), where $\widehat{\boldsymbol{\Sigma}}^\star$ is now a $(H + 1) \times (H + 1)$ matrix. Details of the proof can be found in the Appendix B of the supplementary materials (see Corollary 3). The variance estimator (35) is design-consistent because $H = O(1)$. Hence, $\widehat{r}(\theta_0)$ follows a chi-squared distribution asymptotically and a consistent empirical likelihood confidence interval is given by (33).

Note that the same likelihood function (2) is used with or without stratification. The pseudo empirical likelihood function has to be modified to take the stratification into account (e.g. Rao and Wu, 2009, p. 195).

5.3. Auxiliary variables

In this §, we assume that the sample is selected according to a stratified pps with replacement sampling design described in § 5.2. Let $\boldsymbol{x}_i$ be a P vector of values of auxiliary variables attached to unit i. Suppose that these variables are such that their population totals $\boldsymbol{X} = \sum_{i \in U} \boldsymbol{x}_i$ are known. Let $\mathbf{f}_i(\boldsymbol{x}_i, \boldsymbol{X}) = \boldsymbol{x}_i - \boldsymbol{X}\pi_i n^{-1}$ be a P vector. Let $\widehat{m}_i(\boldsymbol{x})$ be the values which maximise (2) under the constraint (3) with $\boldsymbol{c}_i = (Nn^{-1}\boldsymbol{z}_i^\top, \mathbf{f}_i(\boldsymbol{x}_i, \boldsymbol{X})^\top)^\top$ and $\boldsymbol{C} = \sum_{i \in U} \boldsymbol{c}_i = (Nn^{-1}\boldsymbol{n}^\top, \boldsymbol{0}^\top)^\top$. Thus the maximum empirical likelihood estimator is the solution of $\sum_{i \in s} \widehat{m}_i(\boldsymbol{x}) g_i(\theta) = 0$ (see (6) and (7)). Note that $\widehat{m}_i(\boldsymbol{x})$ are calibrated weights because $\sum_{i \in s} \widehat{m}_i(\boldsymbol{x}) \mathbf{f}_i(\boldsymbol{x}_i, \boldsymbol{X}) = \boldsymbol{0}$ implies $\sum_{i \in s} \widehat{m}_i(\boldsymbol{x}) \boldsymbol{x}_i = \boldsymbol{X}$. When N is known, we recommend to include the variable $x_i = 1$ (intercept) within $\boldsymbol{x}_i$. This may improve the efficiency of the maximum empirical likelihood estimator. Note that for the calculation of $\widehat{m}_i(\boldsymbol{x})$, the quantity Nn^{-1} can be omitted within $\boldsymbol{c}_i$ and $\boldsymbol{C}$.

It is also possible to calibrate towards parameters more complex than totals. For example, we may want to calibrate with respect to population means, quantiles or variances (e.g. Owen, 1991; Chaudhuri *et al.*, 2008; Lesage, 2011). In this case, the calibration constraint is specified by the estimating equations $\sum_{i \in s} m_i \mathbf{f}_i(\boldsymbol{x}_i, \boldsymbol{\vartheta}_0) = \boldsymbol{0}$; where $\mathbf{f}_i(\boldsymbol{x}_i, \boldsymbol{\vartheta}_0)$ is a vector function of the auxiliary variables and of a known parameter $\boldsymbol{\vartheta}_0$ which is the solution of the estimating equation $\sum_{i \in U} \mathbf{f}_i(\boldsymbol{x}_i, \boldsymbol{\vartheta}_0) = \boldsymbol{0}$. In this case, we use $\boldsymbol{c}_i = (Nn^{-1}\boldsymbol{z}_i^\top, \mathbf{f}_i(\boldsymbol{x}_i, \boldsymbol{\vartheta}_0)^\top)^\top$ and $\boldsymbol{C} = (Nn^{-1}\boldsymbol{n}^\top, \boldsymbol{0}^\top)^\top$. For example, if we want to calibrate towards known population means, we need to use $\mathbf{f}_i(\boldsymbol{x}_i, \boldsymbol{\vartheta}_0) = \boldsymbol{x}_i - \boldsymbol{\vartheta}_0$, with $\boldsymbol{\vartheta}_0 = \boldsymbol{X}N^{-1}$. The most common situation is practice is to know a set of totals, means or proportions from large external censuses or surveys. Simultaneous calibration on totals, means or proportions is feasible with this § approach.

We assume that the conditions (8)-(13), (22)-(24) hold. Using Theorem 2 in the Appendix B of the supplementary materials, we have that $\widehat{r}(\theta_0)$ follows asymptotically a chi-squared distribution with one degree of freedom. This relies on the condition that the regression estimator (17) has a normal distribution asymptotically, when $\theta = \theta_0$. This condition can be supported by Scott and Wu (1981) regularity conditions for normality of the regression estimator.

6. πps sampling without replacement

In this §, we consider that we have a uni-stage sample selected without replacement (πps sampling). When n/N is negligible, the variance estimator (26) is approximately unbiased and the random variable (25) follows a chi-squared distribution asymptotically. When n/N is not negligible, (26) is biased under πps sampling, implying that (25) does not necessarily follow a chi-squared distribution. In this §, we assume that n/N is not negligible.

Under πps sampling, the point estimator is still given by the solution of (7) with $\boldsymbol{c}_i$ and $\boldsymbol{C}$ given in § 5. This gives a consistent estimator because (16) still holds. However, for confidence intervals, we use a penalised empirical log-likelihood ratio function (36) with constraints based on a different set of vectors $\widetilde{\boldsymbol{c}}_i$, $\widetilde{\boldsymbol{c}}_i^\star$, $\widetilde{\boldsymbol{C}}$, $\widetilde{\boldsymbol{C}}^\star$ given below. We will see that this ensures that the penalised empirical log-likelihood ratio function (40) follows a chi-squared distribution asymptotically, when $\theta = \theta_0$.

First we consider the case of a single stratum without auxiliary variables. In this case, we use $\widetilde{\boldsymbol{c}}_i = q_i N n^{-1} \pi_i$, $\widetilde{\boldsymbol{C}} = N n^{-1} \sum_{i \in s} q_i$. Let $\widetilde{\boldsymbol{c}}_i^\star = q_i (N n^{-1} \pi_i, g_i(\theta))^\top$ and $\widetilde{\boldsymbol{C}}^\star = (N n^{-1} \sum_{i \in s} q_i, \sum_{i \in s} (q_i - 1) \breve{g}_i(\theta))^\top$, with $q_i = (1 - \pi_i)^{\frac{1}{2}}$.

Consider the following *penalised empirical log-likelihood function.*

$$\widetilde{\ell}(m) = \log \left(\prod_{i \in s} m_i \exp(1 - \pi_i m_i) \right) = \sum_{i \in s} \log(m_i) + n - \sum_{i \in s} m_i \pi_i \cdot \qquad (36)$$

Let $\widetilde{\ell}(\widehat{m})$ be the maximum value of (36) under the constraints $m_i > 0$ and

$$\sum_{i \in s} m_i \widetilde{\boldsymbol{c}}_i = \widetilde{\boldsymbol{C}} \cdot \qquad (37)$$

The quantities $\widehat{m}_i$ which maximise (36) under the constraints (37) are given by $\widehat{m}_i = (\pi_i + \widetilde{\boldsymbol{\eta}}^\top \widetilde{\boldsymbol{c}}_i)^{-1}$; where $\widetilde{\boldsymbol{\eta}}$ is such that (37) holds. Note that for the computation of $\widehat{m}_i$, the quantity $N n^{-1}$ can be omitted within $\widetilde{\boldsymbol{c}}_i$ and $\widetilde{\boldsymbol{C}}$.

Let $\widetilde{\ell}(\widehat{m}^\star, \theta)$ be the maximum value of (36) under the constraint $m_i > 0$ and

$$\sum_{i \in s} m_i \widetilde{\boldsymbol{c}}_i^\star = \widetilde{\boldsymbol{C}}^\star, \qquad (38)$$

for a given θ. We obtain

$$\widehat{m}_i^\star(\theta) = \left(\pi_i + \widetilde{\boldsymbol{\eta}}^{\star\top} \widetilde{\boldsymbol{c}}_i^\star \right)^{-1}, \qquad (39)$$

where $\widetilde{\boldsymbol{\eta}}^\star$ is such that (38) holds. Note that for the computation of $\widehat{m}_i^\star(\theta)$, the quantity $N n^{-1}$ can be omitted within $\widetilde{\boldsymbol{c}}_i^\star$ and $\widetilde{\boldsymbol{C}}^\star$.

The *penalised empirical log-likelihood ratio function* is defined by

$$\widetilde{r}(\theta) \quad = \quad 2 \left\{ \widetilde{\ell}(\widehat{m}) - \widetilde{\ell}(\widehat{m}^\star, \theta) \right\} \cdot \qquad (40)$$

We may interpret the q_i as finite population corrections. The penalty $\exp(1-\pi_i m_i)$ in (36) ensures that the $\widehat{m}_i^{\star}(\theta)$ maximise (36) and that the random variable $\widehat{r}(\theta_0)$ follows a chi-squared distribution asymptotically (see Proposition 2 below). Note that (36) reduces to (2) when $\widetilde{\boldsymbol{c}}_i = \boldsymbol{c}_i$ and $\widetilde{\boldsymbol{c}}_i^{\star} = \boldsymbol{c}_i^{\star}$ given in § 2 because (4) holds in this case. However as $\widetilde{\boldsymbol{c}}_i \neq \boldsymbol{c}_i$ and $\widetilde{\boldsymbol{c}}_i^{\star} \neq \boldsymbol{c}_i^{\star}$, the equation (4) does not hold any longer. Thus, (2) is different from (36) and $\widetilde{r}(\theta) \neq \widehat{r}(\theta)$, for all θ.

PROPOSITION 2. *Assuming that the πps sampling design is such that the conditions (8), (10), (22), (23) and (24) hold with $\widetilde{\boldsymbol{c}}_i$, $\widetilde{\boldsymbol{c}}_i^{\star}$, $\widetilde{\boldsymbol{C}}$ and $\widetilde{\boldsymbol{C}}^{\star}$, when $\theta = \theta_0$, we have that*

$$\widetilde{r}(\theta_0) = \widehat{G}(\theta_0)_\pi^2 \ \widehat{var}[\widehat{G}(\theta_0)_\pi]^{-1} + O_p(n^{-\frac{1}{2}}); \tag{41}$$

where

$$\widehat{var}[\widehat{G}(\theta_0)_\pi] = \sum_{i \in s} q_i^2 \breve{g}_i(\theta_0)^2 - \widehat{d}^{-1} \mathring{G}(\theta_0)^2, \tag{42}$$

with $\mathring{G}(\theta_0) = \sum_{i \in s} q_i^2 \breve{g}_i(\theta_0)$ and $\widehat{d} = \sum_{i \in s} q_i^2$.

Note that the regularity conditions (9), (11), (12), (13) are trivial for the $\widetilde{\boldsymbol{c}}_i$, $\widetilde{\boldsymbol{c}}_i^{\star}$, $\widetilde{\boldsymbol{C}}$ and $\widetilde{\boldsymbol{C}}^{\star}$ considered.

The proof follows the same principle as the proof of Proposition 1. The rigorous proof can be found in the Appendix B of the supplementary materials (see Theorem 1 and Corollary 2).

The variance estimator (42) is the Hájek (1964) variance estimator which is consistent, for high entropy sampling designs when $d = \sum_{i \in U} \pi_i(1-\pi_i) \to \infty$ (e.g. Hájek, 1964, 1981; Berger, 1998; Prášková and Sen, 2009; Berger, 2011). Thus, the random variable $\widetilde{r}(\theta_0)$ follows a chi-squared distribution. Hence a consistent empirical likelihood confidence interval can be constructed with $\widetilde{r}(\theta)$ (see (33)). Berger (2011) gave regularity conditions under which (42) is consistent under a large class of high entropy sampling designs. For example the rejective (Hájek, 1964; Fuller, 2009), the Rao-Sampford (Rao, 1965; Sampford, 1967), the Pareto (Aires, 2000) and the Chao (1982) sampling designs are high entropy sampling designs. Although most sampling designs used in practice have large entropy, there are designs with low entropy, such as the non-randomised systematic sampling design and the Hartley-Rao-Cochran sampling design (Rao *et al.*, 1962). Berger (2014) proposed a set of $\widetilde{\boldsymbol{c}}_i$, $\widetilde{\boldsymbol{c}}_i^{\star}$, $\widetilde{\boldsymbol{C}}$ and $\widetilde{\boldsymbol{C}}^{\star}$ which can be used for the Hartley-Rao-Cochran sampling design (Rao *et al.*, 1962).

The q_i reduce the effect of the units with large π_i. For example, if $\pi_i = 1$, then $q_i = 0$ and $\widetilde{\boldsymbol{c}}_i^{\star} = \boldsymbol{0}$. Thus, $\widehat{m}_i = \widehat{m}_i^{\star}(\theta) = 1$ and this unit has no contribution towards the empirical likelihood functions and any confidence intervals. This is a natural property since this unit does not contribute towards the sampling distribution. Note that with small sampling fractions (π_i negligible), q_i is approximately equal to one. If we replace q_i by one, the approach proposed in this § reduces to the approach of § 5. This is in agreement with the fact that pps and πps sampling are equivalent when the π_i are negligible. Note we adjust the constraints by the quantities q_i which do not need to be estimated, unlike the pseudo empirical likelihood approach which adjusts the log-likelihood ratio function by the design effect which needs to be estimated. Note that the precision of (42) can be improved by substituting q_i by $(1-\lambda_i)^{\frac{1}{2}}$, where the λ_i are defined by the recursive formula (3.25) in Hájek (1981) . Using $q_i = (1-\lambda_i)^{\frac{1}{2}}$ instead of $q_i = (1-\pi_i)^{\frac{1}{2}}$ may improve the variance (42) for moderate sample sizes (see Hájek (1981) for more details).

For stratified designs, we use $\widetilde{\boldsymbol{c}}_i = q_i N n^{-1} \boldsymbol{z}_i$, $\widetilde{\boldsymbol{C}} = N n^{-1} \sum_{i \in s} q_i \breve{\boldsymbol{z}}_i$, $\widetilde{\boldsymbol{c}}_i^\star = q_i (N n^{-1} \boldsymbol{z}_i^\top, g_i(\theta))^\top$ and $\widetilde{\boldsymbol{C}}^\star = (N n^{-1} \sum_{i \in s} q_i \breve{\boldsymbol{z}}_i^\top, \sum_{i \in s} (q_i - 1) \breve{g}_i(\theta))^\top$, where $\breve{\boldsymbol{z}}_i = \boldsymbol{z}_i \pi_i^{-1}$ and $\boldsymbol{z}_i$ is defined by (34). We assume that conditions (8), (10), (22), (23) and (24) hold with $\widetilde{\boldsymbol{c}}_i$, $\widetilde{\boldsymbol{C}}$, $\widetilde{\boldsymbol{c}}_i^\star$ and $\widetilde{\boldsymbol{C}}^\star$, when $\theta = \theta_0$. The conditions (9), (11), (12) and (13) are trivial in this case. Using Corollary 1 in the Appendix B of the supplementary materials, we have that (41) holds; where $\widehat{var}[\widehat{G}(\theta_0)_\pi]$ is now the stratified Hájek (1964) variance estimator given by

$$\widehat{var}_{st}[\widehat{G}(\theta_0)_\pi] = \sum_{h=1}^{H} \left[\sum_{i \in s_h} q_i^2 \breve{g}_i(\theta_0)^2 - \widehat{d}_h^{-1} \mathring{G}_h(\theta_0)^2 \right]; \tag{43}$$

where $\widehat{d}_h = \sum_{i \in s_h} q_i^2$ and $\mathring{G}_h(\theta_0) = \sum_{i \in s_h} q_i^2 \breve{g}_i(\theta_0)$. This variance estimator is consistent when $d_h = \sum_{i \in U_h} \pi_i (1 - \pi_i) \to \infty$ and when $H = O(1)$. Hence $\widetilde{r}(\theta_0)$ follows a chi-squared distribution asymptotically.

This § 's approach can be extended for calibration constraints. For $\widetilde{\ell}(m)$, we use $\widetilde{\boldsymbol{c}}_i = q_i (N n^{-1} \boldsymbol{z}_i^\top, \mathbf{f}_i(\boldsymbol{x}_i, \boldsymbol{\vartheta}_0)^\top)^\top$ and $\widetilde{\boldsymbol{C}} = (N n^{-1} \sum_{i \in s} q_i \breve{\boldsymbol{z}}_i^\top, \sum_{i \in s} (q_i - 1) \breve{\mathbf{f}}_i(\boldsymbol{x}_i, \boldsymbol{\vartheta}_0)^\top)^\top$, where $\breve{\mathbf{f}}_i(\boldsymbol{x}_i, \boldsymbol{\vartheta}_0) = \mathbf{f}_i(\boldsymbol{x}_i, \boldsymbol{\vartheta}_0) \pi_i^{-1}$. For $\widetilde{\ell}(\widehat{m}^\star, \theta)$, we use $\widetilde{\boldsymbol{c}}_i^\star = (\widetilde{\boldsymbol{c}}_i^\top, q_i g_i(\theta))^\top$ and $\widetilde{\boldsymbol{C}}^\star = (\widetilde{\boldsymbol{C}}^\top, \sum_{i \in s} (q_i - 1) \breve{g}_i(\theta))^\top$. We assume that the conditions (8)-(13), (22)-(24) hold. Using the Theorem 2 in the Appendix B of the supplementary materials, we have that $\widetilde{r}(\theta_0)$ follows asymptotically a chi-squared distribution with one degree of freedom. Note that for the computation of $\widehat{m}_i$ and $\widehat{m}_i^\star(\theta)$, the quantity $N n^{-1}$ can be omitted within $\widetilde{\boldsymbol{c}}_i$, $\widetilde{\boldsymbol{C}}$, $\widetilde{\boldsymbol{c}}_i^\star$ and $\widetilde{\boldsymbol{C}}^\star$.

This § 's results are based on the Theorem 1 (Appendix B of the supplementary materials) which holds under any stratified sampling designs which satisfy the regularity conditions (9)-(13) with $\widetilde{\boldsymbol{c}}_i$ and $\widetilde{\boldsymbol{c}}_i^\star$. This is true for the high entropy πps and pps sampling designs with the $\widetilde{\boldsymbol{c}}_i$ and $\widetilde{\boldsymbol{c}}_i^\star$ we considered. The Theorem 1 shows that $\widetilde{r}(\theta_0)$ converges to a quadratic form which follows a chi-squared distribution asymptotically only for specific choices of $\widetilde{\boldsymbol{c}}_i$ and $\widetilde{\boldsymbol{c}}_i^\star$ which depends on the design and on the auxiliary variables (see Table 5), because these choices induce a design consistent variance within the quadratic form.

7. Simulation study

In this §, we compare the performance of the proposed empirical likelihood confidence interval with alternative approaches. The parameters of interest are population quantiles (in §§ 7.1 and 7.3) and population means (in § 7.2). In §§ 7.1 and 7.2, we consider a single stage πps randomised systematic sampling design (e.g. Tillé, 2006, §7.2). This is a high entropy design (Brewer and Donadio, 2003) which is implemented by arranging the population randomly before selecting each πps systematic sample. In the Appendix C of the supplementary materials, we consider a two-stage design. We consider that we have a single stratum. For the bootstrap approaches, we used 1000 bootstrap samples. The Hartley and Rao (1962) variance estimator is used for standard confidence intervals and for the pseudo empirical likelihood approaches. In § 7.3, we show how empirical likelihood confidence intervals can be constructed when we have unit nonresponse. The population data are either generated from a model or based upon the 1998-1999 British Family Expenditure Survey (FES). The variables of interest considered are skewed. With non-skewed variables of interest, we did not observe major differences between the competitive approaches. These simulation results are not presented.

We investigate the Monte-Carlo performance of the 95% confidence intervals. All the simulation studies are based on 10,000 samples randomly selected. The sample size is $n = 500$

in §§ 7.1, 7.2 and 7.3. In § A, the average sample size is 1000. We used the statistical software R (R Development Core Team, 2014). The algorithms were coded in C.

7.1. Quantiles

Let θ_0 be the q quantile Y_q of the population distribution of a variable of interest y_i; where $0 < q < 1$. We use $g_i(\theta) = \varrho(y_{(i)}, \theta) - q$, with

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

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 function $\delta\{y \leq \theta\} = 1$ when $y \leq \theta$ and $\delta\{y \leq \theta\} = 0$ otherwise. The empirical likelihood estimator of Y_q is the solution of the equation $\widehat{G}(\theta) = 0$ which reduces to $\widetilde{F}(\theta) = q$; where $\widetilde{F}(\theta) = \left(\sum_{i\in s} \widehat{m}_{(i)}\right)^{-1} \sum_{i\in s} \widehat{m}_{(i)} \varrho(y_{(i)}, \theta)$ is a distribution function. 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

$$\widehat{F}(\theta) = \frac{\sum_{i\in s} \widehat{m}_{(i)} \delta\{y_{(i)} \leq \theta\}}{\sum_{i\in s} \widehat{m}_{(i)}}. \tag{44}$$

This interpolation consists in joining the steps of $\widehat{F}(\theta)$ by straight lines segments (Harms and Duchesne, 2006). It can be easily shown that $N^{-1}\widehat{G}(\theta_0)_\pi = N^{-1} \sum_{i\in s}[\varrho(y_{(i)}, \theta_0) - q]\pi_i^{-1}$ is approximately equal to $N^{-1} \sum_{i\in s}[\delta\{y_i \leq \theta_0\} - q]\pi_i^{-1}$ which is an Horvitz and Thompson (1952) estimator. Thus, (21) and (22) hold, and the log-likelihood ratio function has a chi-squared distribution asymptotically and it can be used to derive confidence intervals for Y_q. This approach can be generalised to take the stratification and auxiliary variables into account (see § 7.3). Note that the empirical likelihood estimator is different from the classical estimator of a quantile $\widehat{Y}_q = \inf\{\theta : \widehat{F}(\theta) \geq q\}$. We did not observe significant differences between the mean squared errors of both estimators.

We generated several skewed population data according to the following model $y_i = 3 + a_i + \varphi\ e_i$ (Wu and Rao, 2006); where a_i follows an exponential distribution with rate parameters equal to 1 and $e_i \sim \chi^2_1 - 1$. The π_i are proportional to $a_i + 2$. Populations of size $N = 2000$ and $N = 25000$ are generated. The values y_i, x_i and a_i generated are treated as fixed. The parameter φ is used to specify the correlation $\rho(y, \pi)$ between the values y_i and π_i ($\rho(y, \pi) = 0.8$ with $\varphi = 0.5$; $\rho(y, \pi) = 0.3$ with $\varphi = 2.3$). We used the πps randomised systematic sampling design to select 10,000 samples of size $n = 500$. We use the approach described in § 6. We consider the 5% and 25% quantiles.

The results are given in Table 1. The values not within brackets are the values for the populations of size $N = 2000$ (large sampling fractions). The values within brackets are the values for the populations of size $N = 25{,}000$ (small sampling fractions). The ratio of average length is the average length of the confidence intervals divided by the average length of the confidence intervals based on linearisation. We measure the stability of the confidence intervals using the standard deviation of the lengths. The standard deviations are divided by the standard deviation of lengths of the linearisation confidence intervals. These values are given in the last column of Table 1.

Based on the Shapiro-Wilks test for normality, the point estimators do not follow a normal distribution. This lack of normality may affect the coverages. For the standard confidence intervals based on the linearised variance (Deville, 1999), we observed that the coverages and

Table 1. Quantiles Y_q ($q = 5\%$ and 25%). $n = 500$. The values not within brackets for $N = 2000$ (large sampling fractions). The values within brackets for $N = 25,000$ (small sampling fractions).

q	$\rho(y,\pi)$	Approaches	Overall Cov. %	Lower tail err. rates%	Upper tail err. rates%	Ratio Av. Length	Ratio SD Length
5%	0.8	Linear. (Std.)	99.3* (98.0*)	0.7* (1.8*)	0.0* (0.2*)	1.00 (1.00)	1.00 (1.00)
		Rescaled boot.	97.2* (95.4)	1.4* (2.3)	1.4* (2.4)	0.74 (0.82)	2.95 (2.21)
		Direct boot.	95.0 (93.6*)	1.9* (3.6*)	3.2* (2.8)	0.67 (0.83)	2.83 (2.36)
		Woodruff	94.9 (95.2)	2.3 (2.0*)	2.8 (2.8*)	0.67 (0.81)	2.82 (2.19)
		Emp. Lik.	94.8 (95.0)	1.9* (2.1*)	3.5* (3.0*)	0.65 (0.80)	2.78 (2.18)
	0.3	Linear. (Std.)	98.8* (98.8*)	1.2* (1.2*)	0.0* (0.0*)	1.00 (1.00)	1.00 (1.00)
		Rescaled boot.	97.2* (95.2)	1.4* (2.1*)	1.4* (2.8)	0.67 (0.73)	2.59 (2.22)
		Direct boot.	92.9* (93.8*)	4.9* (3.1*)	2.2 (3.1*)	0.65 (0.74)	2.85 (2.35)
		Woodruff	95.0 (95.2)	2.5 (1.7*)	2.4 (3.1*)	0.62 (0.73)	2.54 (2.19)
		Empirical Lik.	94.9 (94.9)	2.2* (1.9*)	3.0* (3.2*)	0.60 (0.72)	2.48 (2.16)
25%	0.8	Linear. (Std.)	94.2* (95.4)	2.5 (2.0*)	3.4* (2.6)	1.00 (1.00)	1.00 (1.00)
		Rescaled boot.	97.0* (95.1)	1.5* (2.3)	1.5* (2.6)	1.10 (0.98)	3.78 (2.29)
		Direct boot.	92.0* (94.2*)	4.2* (2.9*)	3.8* (2.9*)	1.03 (0.99)	4.16 (2.41)
		Woodruff	94.8 (95.1)	2.8* (2.4)	2.4 (2.5)	1.00 (0.98)	3.58 (2.25)
		Empirical Lik.	94.9 (95.0)	2.4 (2.2)	2.7 (2.8)	0.99 (0.97)	3.52 (2.23)
	0.3	Linear. (Std.)	97.6* (97.1*)	1.7* (1.6*)	0.7* (1.3*)	1.00 (1.00)	1.00 (1.00)
		Rescaled boot.	97.0* (94.9)	1.4* (2.3)	1.6* (2.7)	0.97 (0.91)	3.36 (2.45)
		Direct boot.	94.3* (95.0)	3.1* (2.7)	2.5 (2.3)	0.87 (0.91)	3.29 (2.55)
		Woodruff	95.0 (94.9)	2.6 (2.6)	2.4 (2.5)	0.86 (0.91)	3.12 (2.41)
		Empirical Lik.	94.9 (94.9)	2.1* (2.3)	2.9* (2.8)	0.86 (0.91)	3.10 (2.39)

$*$ Coverages (or tail error rates) significantly different from 95% (or 2.5%). p-value ≤ 0.05.

tail error rates are significantly different from their nominal levels 95% and 2.5% respectively, except with $Y_{0.25}$, $N = 25{,}000$ and a correlation of 0.8. This can be explained by the bias of the linearised variance estimator. This was also observed by Wu (1999).

The rescaled bootstrap confidence interval (the histogram approach) is based upon the observed 2.5% and 97.5% quantiles of the set of bootstrap values (Rao *et al.*, 1992). The rescaled bootstrap approach gives acceptable coverages for small sampling fractions. However, for large sampling fraction, it gives coverages and tail error rates significantly different from 95% and 2.5% respectively. This is not surprising, as rescaled bootstrap is design for small sampling fractions. The direct bootstrap approach (Antal and Tillé, 2011) is designed for πps sampling with large sampling fractions. The coverages and tail error rates of the confidence interval proposed by Antal and Tillé (2011) are significantly different, except in two situations: (i) with $Y_{0.25}$ with a small sampling fraction and a small correlation between y_i and π_i, (ii) and $Y_{0.05}$ with a large sampling fraction and a large correlation.

Chen and Wu (2002) proposed a Woodruff (1952) approach for confidence intervals of pseudo empirical likelihood estimators of quantiles. The Woodruff (1952) confidence intervals are obtained by inverting the distribution function (44). This approach gives the correct coverage in all situations, except that the tail error rates of $Y_{0.05}$ are significantly different from 2.5% with small sampling fractions. The empirical likelihood confidence intervals have also good coverages comparable to the Woodruff (1952) approach. The tail error rates of $Y_{0.05}$ can also be significantly different from 2.5%.

Note that the empirical likelihood confidence intervals have the shortest average length among the approaches which give the correct coverage. The bootstrap confidence intervals are more unstable (see last column of Table 1) because of re-sampling. Linearisation gives the most stable confidence intervals, but with usually poor coverage and tail error rates.

Table 2. Quantiles Y_q with $q = 5\%$ and $q = 25\%$ (values within brackets). FES Data. $n = 500$. $N = 19,890$.

Approaches	Overall Cov. %	Lower tail error rates %	Upper tail error rates %	Ratio Av. Length	Ratio SD Length
Linearisation (Std.)	95.6* (95.8*)	3.1* (1.8*)	1.3* (2.4)	1.00 (1.00)	1.00 (1.00)
Rescaled bootstrap	95.5* (95.0)	2.6 (2.6)	2.1* (2.4)	0.95 (0.95)	1.99 (2.03)
Direct bootstrap	93.2* (94.3*)	3.1* (2.9*)	3.6* (2.8*)	0.97 (0.95)	2.17 (2.11)
Woodruff	95.8* (94.9)	1.7* (2.6)	2.6 (2.6)	0.96 (0.94)	1.97 (2.00)
Empirical likelihood	95.5* (94.8)	2.1* (2.4)	2.5 (2.8*)	0.94 (0.94)	1.94 (1.98)

$*$ Coverages (or tail error rates) significantly different from 95% (or 2.5%). p-value ≤ 0.05.

With $Y_{0.25}$, $N = 25{,}000$ and a correlation of 0.8, the linearisation approach gives acceptable coverages with slightly more stable confidence intervals, but only in this case. In the other cases, the proposed confidence interval is the most stable one among the confidence intervals with acceptable coverages.

We duplicated the FES data three times to create an artificial population of $N = 19{,}890$ households. Samples of size $n = 500$ are selected with π_i proportional to first-order inclusion probabilities given in the FES dataset. The parameters of interest are quantiles of the equivalent total weekly household expenditure (Department of Social Security, 2001). The results are given in Table 2. For $Y_{0.05}$, the coverages and tail error rates are significantly different from 95% and 2.5%. For $Y_{0.25}$ (the values within brackets), the rescaled bootstrap, the Woodruff and the empirical likelihood confidence intervals have similar coverages. However, the upper tail error rate of the empirical likelihood confidence interval is significantly larger than 2.5%. The linearisation and the direct bootstrap confidence intervals have significantly different coverage and tail error rates. The empirical likelihood approach gives slightly more stable confidence intervals than the rescaled bootstrap and the Woodruff approaches.

7.2. Means

Suppose that the parameter of interest θ_0 is the population mean, and that we have a vector $\boldsymbol{x}_i = (1, x_i)^\top$ of auxiliary variables with known population totals $\boldsymbol{X} = (N, X)^\top$; that is, $\mathbf{f}_i(\boldsymbol{x}_i, \boldsymbol{X}) = \boldsymbol{x}_i - \boldsymbol{X}\pi_i n^{-1}$. We use $g_i(\theta) = y_i - n^{-1}N\theta\pi_i$. The standard confidence interval is based on the standard regression estimator defined by (6.4.2) in Särndal *et al.* (1992). Note that the regression estimator, the pseudo empirical likelihood point estimators (pseudo-EL1 & pseudo-EL2) and the empirical likelihood point estimator are different.

Let a_i and x_i be generated from independent exponential distributions with rate parameters equal to 0.5. The π_i are proportional to $a_i + 2$. We generate 80% of the values of y_i from a normal distribution with mean 8 and variance 1. The remaining 20% are outlying values generated from $y_i = 3 + a_i + \beta x_i + \varphi\, e_i$, where $\varphi = 1.5$. We select 10,000 samples of size $n = 500$ from populations of size $N = 2000$ and $N = 25{,}000$. Based on the Shapiro-Wilks test for normality, the point estimators do not follow a normal distribution, except for the empirical likelihood and pseudo empirical likelihood estimators when $N = 25{,}000$.

The simulation results are given in Table 3. The column "Ratio MSE" gives the relative efficiency given by the ratio between the mean squared error (MSE) of the point estimator and the regression estimator. The proposed empirical likelihood approach gives the correct coverage in all cases. The regression estimator has also good coverages. However, the proposed empirical likelihood approach gives shorter and more stable confidence intervals. The MSE of the empirical likelihood point estimator is about 50% lower than the MSE of the regression estimator. The pseudo empirical likelihood estimators have similar MSE. With small sam-

Table 3. Mean. $n = 500$. The values not within brackets for $N = 2000$ (large sampling fractions). The values within brackets for $N = 25,000$ (small sampling fractions).

Approaches	Overall Cov. %	Lower tail err. rates %	Upper tail err. rates %	Ratio Av. Length	Ratio SD Length	Ratio MSE (Rel. Eff.)
Reg. Est. (Std.)	95.2 (94.8)	2.3 (2.7)	2.5 (2.5)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)
Rescaled boot.	97.4* (94.3*)	0.2* (1.0*)	2.4 (4.6*)	1.19 (1.01)	1.29 (1.03)	1.00 (1.00)
Direct boot.	94.3* (91.3*)	0.9* (1.4*)	4.7* (7.2*)	0.99 (0.90)	0.50 (0.37)	1.00 (1.00)
Pseudo-EL1	94.5* (94.8)	3.0* (2.5)	2.5 (2.6)	0.54 (0.52)	0.54 (0.41)	0.53 (0.49)
Pseudo-EL2	90.3* (92.3*)	5.5* (4.0*)	4.2* (3.6*)	0.47 (0.47)	0.45 (0.36)	0.52 (0.48)
Empirical Lik.	95.1 (94.9)	2.6 (2.4)	2.4 (2.7)	0.53 (0.50)	0.45 (0.37)	0.52 (0.48)

$*$ Coverages (or tail error rates) significantly different from 95% (or 2.5%). p-value ≤ 0.05.

pling fraction ($N = 25{,}000$), the proposed empirical likelihood approach and the pseudo-EL1 approach give similar coverages, but the proposed confidence intervals are slightly shorter and more stable. The bootstrap and the pseudo-EL2 approaches give coverages and tail error rates significantly different from 95% and 2.5%. The coverages observed for the bootstrap approaches are due to a lack of normality because both bootstrap variance estimators are unbiased. The small coverages for the pseudo-EL2 approach is due to the instability of the design effect. The pseudo-EL1 approach gives the correct nominal coverage when the point estimator is normal; that is, for small sampling fraction ($N = 25{,}000$).

In the Appendix C of the supplementary materials, we give the result of a simulation study based on a two-stage sampling design. The coverages and tail error rates of the proposed approach are not significantly different 95% and 2.5% (see Table C1). The coverages and tail errors rates of the bootstrap and the pseudo-EL approaches are significantly different from 95% and 2.5%

7.3. Unit non-response

Suppose that we have unit non-response according to a uniform response mechanism; that is, we assume that all the units respond independently with the same response probability p_r. Let r_i be the response indicator: $r_i = 1$ if the unit i is a respondent and $r_i = 0$ otherwise. Consider a reverse approach (Fay, 1991); that is, we have a two-phase design with the response mechanism being the first phase and the second phase being a stratified pps sampling design or a stratified πps sampling design with negligible sampling fraction ($n/N = o_p(1)$). The uniform response assumption is often unrealistic in practice. It is common practice to form a finite number of adjustment cells and to assume uniform response within cells. The proposed approach can be extended in this case.

The approaches described in § 5 can be used after replacing $g_i(\theta)$ by $r_i g_i(\theta)$. The estimator of θ_0 is the solution of $\sum_{i \in s} \widehat{m}_i r_i g_i(\theta) = 0$. Assuming that the two-phase design is such that the regularity conditions (8)-(13), (22)-(24) hold and that $n/N = o(1)$, the empirical log-likelihood ratio function $\widehat{r}(\theta_0)$ is asymptotically equal to a quadratic form with a variance (35) which incorporates the r_i (see §§ 5.1, 5.2, 5.3 and 6). When $n/N = o(1)$, Shao and Steel (1999) showed that this variance is a consistent variance estimator. Hence, $\widehat{r}(\theta_0)$ follows a chi-squared distribution asymptotically.

We used the FES population data described in § 7.1. The parameters of interest are quantiles of the equivalent total weekly household expenditure. The auxiliary information is the numbers of individuals within age-sex groups (0-19, 20-39, 40-59, 60+). We consider two situations: with and without the auxiliary variables. Nonrespondents are generated according to a uniform response mechanism with the average response rates 60%, 70% and

Table 4. Empirical likelihood coverages (%) for quantiles Y_q. Unit nonresponse. FES data.

		Av. resp. rate = 60%			Av. resp. rate = 70%			Av. resp. rate = 80%		
	q	Overall	Lower	Upper	Overall	Lower	Upper	Overall	Lower	Upper
Without	10%	95.2	2.1*	2.7	95.0	2.6	2.4	94.8	2.4	2.8*
auxiliary	25%	94.9	2.2	2.9*	94.8	2.4	2.8	94.8	2.5	2.6
variables	50%	95.1	2.3	2.7	95.0	2.3	2.7	95.3	2.1*	2.5
	75%	94.5*	2.4	3.1*	94.8	2.2	2.9*	95.4	1.9*	2.7
With	10%	94.4*	2.8	2.8*	94.6	2.6	2.7	94.8	2.4	2.8*
auxiliary	25%	94.5*	2.7	2.8	94.9	2.5	2.6	95.0	2.2	2.7
variables	50%	94.9	2.2	2.9*	94.8	2.4	2.8*	94.5*	2.3	3.2*
	75%	94.4*	2.2	3.4*	94.6*	2.3	3.1*	94.6*	2.2	3.2*

$*$ Coverages (or tail error rates) significantly different from 95% (or 2.5%). p-value ≤ 0.05.

80%. The aim of this simulation study is to show the performance of the proposed confidence interval. Comparing with alternative methods is beyond the scope of this §. The results are given in Table 4. Most of the coverages and tail error rates are not significantly different from 95% and 2.5%. With auxiliary variables, the upper tail error rates can be slightly larger than 2.5%. Note that these confidence intervals take the effects of the nonresponse, the calibration and the sampling design into account. A variance estimator which incorporate these features is more complex to derive and not necessarily unbiased.

8. Conclusion and discussion

The vectors $\boldsymbol{c}_i$ and $\boldsymbol{C}$ used for point estimation and the vectors $\widetilde{\boldsymbol{c}}_i$, $\widetilde{\boldsymbol{C}}$, $\widetilde{\boldsymbol{c}}_i^\star$ and $\widetilde{\boldsymbol{C}}^\star$ used for confidence intervals are summarised in Table 5. In this table, the quantity Nn^{-1} has been omitted because this quantity does not change the constraints. We notice that the $\boldsymbol{c}_i \neq \widetilde{\boldsymbol{c}}_i$ and $\boldsymbol{C} \neq \widetilde{\boldsymbol{C}}$ unless $\psi_i = 1$. Note that by setting $\psi_i = 1$, the approaches described in §§ 5 and 6 are the same, because in this case, $\widetilde{\boldsymbol{c}}_i = \boldsymbol{c}_i$, $\widetilde{\boldsymbol{C}} = \boldsymbol{C}$, $\widetilde{\boldsymbol{c}}_i^\star = \boldsymbol{c}_i^\star$ and $\widetilde{\boldsymbol{C}}^\star = \boldsymbol{C}^\star$. This implies that (2) is equal to (36) and $\widetilde{r}(\theta) = \widehat{r}(\theta)$. Note that the finite population correction q_i plays no role for point estimation. However, it has an effect on the confidence intervals.

There are analogies between the proposed empirical likelihood approach and calibration (Huang and Fuller, 1978; Owen, 1991; Deville and Särndal, 1992). The empirical likelihood estimator is asymptotically equivalent to a calibrated regression estimator (16). The objective function (2) is related to the concept of empirical likelihood and can be used with or without auxiliary information. The advantage of the proposed empirical likelihood approach over standard calibration is the fact that it gives positive weights, and the empirical log-likelihood ratio function (6) can be used to construct confidence intervals and to test hypotheses. The proposed approach can be naturally extended for balanced samples (e.g. Deville and Tillé, 2004), by including the balancing constraints within (3).

Non-parametric bootstrap is an alternative approach which can be used to derive non-parametric confidence intervals. The consistency of the bootstrap confidence intervals is limited to smooth function of means and for quantiles with small sampling fraction (e.g. Shao and Tu, 1996, Ch.6). The direct bootstrap (Antal and Tillé, 2011) is limited to variance estimation of totals, because it provides a second-moment matching in this case. For complex parameters (such as quantiles), only simulation evidence are provided. Results on the consistency of the direct bootstrap confidence interval is not available. The proposed empirical likelihood confidence interval is consistent for a wider class of parameters (which are solution of estimating equations) with large and small sampling fractions. The proposed approach is simpler to implement and less computationally intensive than the bootstrap,

Table 5. The vectors $\boldsymbol{c}_i$ and $\boldsymbol{C}$ (for point estimation) and the vectors $\widetilde{\boldsymbol{c}}_i$, $\widetilde{\boldsymbol{C}}$, $\widetilde{\boldsymbol{c}}_i^\star$ and $\widetilde{\boldsymbol{C}}^\star$ (for confidence intervals) For without replacement sampling: $\psi_i = q_i = (1-\pi_i)^{\frac{1}{2}}$ (or $\psi_i = (1-\lambda_i)^{\frac{1}{2}}$, see § 6). For with replacement sampling designs: $\psi_i = 1$. The $\boldsymbol{z}_i$ and $\boldsymbol{n}$ are defined by (34). With a single stratum, $\boldsymbol{z}_i = \pi_i$ and $\boldsymbol{n} = n$. The parameter $\boldsymbol{\vartheta}_0$ is a known parameter specifying the auxiliary information (see § 5.3).

	$\boldsymbol{c}_i$ and $\boldsymbol{C}$ for point estimation	$\widetilde{\boldsymbol{c}}_i$, $\widetilde{\boldsymbol{C}}$, $\widetilde{\boldsymbol{c}}_i^\star$ and $\widetilde{\boldsymbol{C}}^\star$ for confidence intervals
Without auxiliary information	$\boldsymbol{c}_i = \boldsymbol{z}_i$ $\boldsymbol{C} = \sum_{i\in U} \boldsymbol{c}_i = \boldsymbol{n}$	$\widetilde{\boldsymbol{c}}_i = \psi_i \boldsymbol{z}_i^\top$ $\widetilde{\boldsymbol{C}} = \sum_{i\in s} \psi_i \breve{\boldsymbol{z}}_i^\top$ $\widetilde{\boldsymbol{c}}_i^\star = (\widetilde{\boldsymbol{c}}_i^\top, \psi_i g_i(\theta))^\top$ $\widetilde{\boldsymbol{C}}^\star = (\widetilde{\boldsymbol{C}}^\top, \sum_{i\in s}(\psi_i - 1)\breve{g}_i(\theta))^\top$
With auxiliary information	$\boldsymbol{c}_i = (\boldsymbol{z}_i^\top, \mathbf{f}_i(\boldsymbol{x}_i, \boldsymbol{\vartheta}_0)^\top)^\top$ $\boldsymbol{C} = \sum_{i\in U} \boldsymbol{c}_i = (\boldsymbol{n}^\top, \boldsymbol{0}^\top)^\top$	$\widetilde{\boldsymbol{c}}_i = \psi_i (\boldsymbol{z}_i^\top, \mathbf{f}_i(\boldsymbol{x}_i, \boldsymbol{\vartheta}_0)^\top)^\top$ $\widetilde{\boldsymbol{C}} = (\sum_{i\in s} q_i \breve{\boldsymbol{z}}_i^\top, \sum_{i\in s}(\psi_i - 1)\breve{\mathbf{f}}_i(\boldsymbol{x}_i, \boldsymbol{\vartheta}_0)^\top)^\top$ $\widetilde{\boldsymbol{c}}_i^\star = (\widetilde{\boldsymbol{c}}_i^\top, \psi_i g_i(\theta))^\top$ $\widetilde{\boldsymbol{C}}^\star = (\widetilde{\boldsymbol{C}}^\top, \sum_{i\in s}(\psi_i - 1)\breve{g}_i(\theta))^\top$

especially with calibration weights. Like bootstrap, the proposed approach does not rely on analytic derivation. Our simulation studies show that, for means and quantiles, bootstrap confidence intervals may have coverages and tail error rates significantly different from their nominal levels. The empirical likelihood approach may give better coverages.

Unlike the pseudo empirical likelihood approach, the computation of the proposed confidence interval does not rely on variance estimates and design effects. This means that it can be applied to a wide class of parameters. The proposed approach is also simpler to implement than the pseudo empirical likelihood. The simulation studies show that, for means, the empirical likelihood confidence interval may give better coverages than the pseudo empirical likelihood confidence intervals.

There are other issues, such as imputation, two-stage designs with large sampling fraction, heavily stratified designs, non-randomised systematic designs and weight trimming adjustment, which are not tackled in this paper and are beyond the scope of this paper.

Acknowledgement

We wishes to thank the anonymous reviewers, Prof. V. Patilea (ENSAI-CREST, France), Prof. Li-Chun Zhang (University of Southampton, UK), M. Oguz-Alper (University of Southampton, UK) and E. Kabzinska (University of Southampton, UK) for helpful comments. We are also grateful to Prof. J.N.K. Rao (Carleton University, Canada) for introducing us to the challenges of empirical likelihood approaches and to E. Antal (FORS, Geneva, Switzerland) for providing the R codes for implementing the direct bootstrap approach.

References

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

Antal, E. and Tillé, Y. (2011) A direct bootstrap method for complex sampling designs from a finite population. *Journal of the American Statistical Association*, **106**, 534–543.

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

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**, 407–426.

Berger, Y. G. (2014) Empirical likelihood confidence intervals under the Rao-Hartley-Cochran sampling design. *Proceedings of the Survey Methods Section, American Statistical Association*.

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

Binder, D. A. (1983) On the variance of asymptotically normal estimators from complex surveys. *International Statistical Review*, **51**, 279–292.

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

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

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

Chaudhuri, S., Handcock, M. S. and Rendall, M. S. (2008) Generalized linear models incorporating population level information: An empirical-likelihood-based approach. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, **70**, pp. 311–328.

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

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

Chen, J., Sitter, R. R. and Wu, C. (2002) Using empirical likelihood methods to obtain range restricted weights in regression estimators for surveys. *Biometrika*, **89**, 230–237.

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

Chen, S. and Kim, J. K. (2014) Population empirical likelihood for nonparametric inference in survey sampling. *Statistica Sinica*, **24**, 335–355.

Department of Social Security (2001) Households below average income 1999/00, appendix 2. *London: Department of Social Security.*

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

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

Deville, J. C. and Tillé, Y. (2004) Efficient balanced sampling: The cube method. *Biometrika*, **91**, 893–912.

Deville, J. C. and 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**, 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*, 429–440.

Francisco, C. A. and Fuller, W. A. (1991) Quantile estimation with a complex survey design. *The Annals of Statistics*, **19**, pp. 454–469.

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.

Godambe, V. P. (1960) An optimum property of regular maximum likelihood estimation. *The Annals of Mathematical Statistics*, **31**, pp. 1208–1211.

Godambe, V. P. and Thompson, M. . (2009) Estimating functions and survey sampling. *Handbook of Statistics: Design, Method and Applications: D. Pfeffermann and C.R. Rao.(editors). Elsevier*, **29B**, 83–101.

Graf, E. and Tillé, Y. (2014) Variance estimation using linearization for poverty and social exclusion indicators. *Survey Methodology*, 61–79.

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

Hansen, M. H. and Hurwitz, W. N. (1943) On the theory of sampling from finite populations. *The Annals of Mathematical Statistics*, **14**, pp. 333–362.

Harms, T. and Duchesne, P. (2006) On calibration estimation for quantiles. *Survey Methodology*, **32**, 37–52.

Hartley, H. O. and Rao, J. N. K. (1962) Sampling with unequal probabilities without replacement. *Ann. math. Statist. Assoc.*, **33**, 350–374.

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

Hartley, H. O. and Rao, J. N. K. (1969) A new estimation theory for sample surveys, ii. *New Developments in survey Sampling (Johnson, N.L., and Smith, H.Jr., Eds.) Wiley, New York*, 147–169.

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

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

Isaki, C. T. and Fuller, W. A. (1982) Survey design under the regression super-population model. *Journal of the American Statistical Association*, **77**, 89–96.

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

Krewski, D. and 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.

Lesage, E. (2011) The use of estimating equations to perform a calibration on complex parameters. *Survey Methodology*, **37**, 103–108.

Oguz-Alper, M. and Berger, Y. G. (2014) Empirical likelihood confidence intervals and significance test for regression parameters under complex sampling designs. *Proceedings of the Survey Research Method Section of the American Statistical Association, Joint Statistical Meeting, Boston*, 10pp.

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

Owen, A. B. (1991) Empirical likelihood for linear models. *The Annals of Statistics*, **19**, 1725–1747.

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

Prášková, Z. and Sen, P. K. (2009) Asymptotic in finite population sampling. *Handbook of*

Statistics 29: Sample Surveys: Inference and Analysis. Elsevier. Danny Pfeffermann and C.R. Rao eds, 489–522.

Qin, J. and Lawless, J. (1994) Empirical likelihood and general estimating equations. *The Annals of Statistics*, **22**, pp. 300–325.

R Development Core Team (2014) *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing. http://www.R-project.org, Vienna, Austria.

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. and Cochran, W. G. (1962) On a simple procedure of unequal probability sampling without replacement. *Journal of the Royal Statistical Society. Series B (Methodological)*, **24**, pp. 482–491.

Rao, J. N. K., Wu, C. F. J. and Yue, K. (1992) Some recent work on resampling methods for complex surveys. *Survey Methodology*, **18**, 209–217.

Rao, J. N. K. and Wu, W. (2009) Empirical likelihood methods. *Handbook of statistics: Sample Surveys: Inference and Analysis, D. Pfeffermann and C. R. Rao eds. The Netherlands (North-Holland)*, **29B**, 189–207.

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

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

Scott, A. and Wu, C. F. (1981) On the asymptotic distribution of ratio and regression estimators. *Journal of the American Statistical Association*, **76**, pp. 98–102.

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

Shao, J. and Tu, D. (1996) *The Jackknife and Boostrap.* New York: Springer.

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*, 71–78.

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

Woodruff, R. S. (1952) Confidence intervals for medians and other position measures. *Journal of the American Statistical Association*, **47**, 635–646.

Wu, C. (1999) The effective use of complete auxiliary information from survey data. *Unpublished doctoral dissertation, Simon Fraser University, Canada.*

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

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