# Empirical likelihood approaches in survey sampling

Y.G. Berger
Southampton Statistical Sciences Research Institute,
University of Southampton, UK
y.g.berger@soton.ac.uk
http://yvesberger.co.uk

**The survey statistician** (No 78 July 2018)
http://isi-iass.org/home/services/the-survey-statistician/

**Abstract**

There are two different empirical likelihood approaches for complex sampling designs: the "*pseudo-empirical likelihood*" introduced by Chen and Sitter (1999) and the "*unequal probability empirical likelihood*" approach proposed by Berger and Torres (2014, 2016). Both approaches are described and reviewed critically. We do not pretend to give an exhaustive account of all the applications of empirical likelihood in survey sampling. This paper is an extended version of Berger (2018*b*).

*Keywords:* Design-based approach, estimating equations, inclusion probabilities, side information, stratification

# 1 Introduction

We consider a "*design-based approach*"; that is, we assume that survey variables are vectors of constants (Neyman, 1938). The design-based approach is often considered in survey sampling theory, because it gives a non-parametric distribution free inference, which does not rely on distributional assumptions about the variables of interest.

We use standard notation; that is, we have a finite population $U = \{1, \ldots, N\}$ of $N$ units. A vector of constant variables $y_i \in \mathbb{R}^{d_y}$ is measured for each unit $i$ in a sample $S \subset U$. The sampling design specifies the random selection of $S$ within $U$. Populations are often stratified into $H$ non-overlapping groups $U_1, \ldots, U_h, \ldots, U_H$ called strata; such that $\cup_{h=1}^{H} U_h = U$. Stratified sampling consists in selecting independent samples $S_h$ from each $U_h$. We assume that $S_h$ is a sample of $n_h$ units selected with unequal selection probabilities

$\pi_i$. We assume that $n_h$ are given constants. The overall sample and sample size are $S = \cup_{h=1}^{H} S_h$ and $n = \sum_{h=1}^{H} n_h$. First, we consider single-stage designs. Then, we shall show how empirical likelihood can be used with multi-stage designs. Unit non-response is another important feature of survey data. We will also show how empirical likelihood can accommodate unit non-response.

Godambe (1966) showed that under the design-based approach, the likelihood function cannot be used for inference, because this function is flat. As solution to this problem, Hartley and Rao (1968) showed that an empirical likelihood function can be used instead. Owen (1988), Qin and Lawless (1994) developed the properties of this approach under a semi-parametric framework (see also Owen, 2001). Most of the recent developments of empirical likelihood in survey sampling can be classified into two groups: Berger and Torres's (2016) unequal probability empirical likelihood approach and Chen and Sitter's (1999) pseudo-empirical likelihood approach (see also Wu and Rao, 2006). There is also Chen and Kim's (2014) population-empirical likelihood approach based on Poisson sampling, which will not be covered in this paper.

Customary approach in survey sampling focus on estimating totals, with well-defined variance estimators. However, for more complex parameters, linearisation is often needed for variance estimation. Linearisation consists in approximating a non-linear parameter by a total of a linearised variable, which depends on parameters, which need to be estimated. In fact, linearisation is essentially based on inference about totals. There is no unified linearisation approach, since different linearisation approaches have been developed. Bootstrap is also used for variance estimation. However, asymptotic theory of bootstrap is restricted to simple settings. Its properties are often limited to means, and solely based on simulations. Jackknife is another approach which is closely related to linearisation. The primary purpose of variance estimates is to construct confidence intervals. Empirical likelihood tackles the problem of measuring the precision of an estimator from a different angle. It focuses on confidence intervals which can be easier to obtain than variance estimates. Empirical likelihood does not rely on linearisation, even when the parameter of interest is not linear. The point estimator does not need to be normally distributed. It can handle nuisance parameters, which are often treated as constants with standard approaches.

We consider a large class of parameters defined by estimating equations. Let $\theta_U \in \mathbb{R}^{d_\theta}$ be an unknown population parameter which is defined as the solution to

$$\sum_{i \in U} e(y_i, \theta) = 0, \tag{1}$$

where $e(y_i, \theta) \in \mathbb{R}^{d_e}$ $(d_e \geqslant d_\theta)$ is a known function which defines $\theta_U$. For example, $\theta_U$ can be the coefficients of a generalised linear regression model, a mean, a total, a quantile, etc. Under a design-based approach, the parameter $\theta_U$ is a vector of unknown population values.

## 2 Unequal probability empirical likelihood approach

Berger and Torres's (2012; 2014; 2016) "*empirical log-likelihood function*" is defined by

$$\ell_{\max}(\theta) := \max_{p_i:\, i\in S}\Big\{\ell(p)\ :\ p_i > 0,\ \sum_{i\in S}\frac{p_i}{\pi_i}e(y_i,\theta) = 0,\ \sum_{i\in S}p_i z_i = \frac{\overrightarrow{n}}{n}\Big\}, \tag{2}$$

where

$$\begin{aligned}
\ell(p) &:= \sum_{i\in S}\log(p_i),\\
z_i &:= \big(z_{i1},\ldots,z_{ih},\ldots,z_{iH}\big)^\top,\\
z_{ih} &:= \begin{cases}1 & \text{if } i\in U_h,\\ 0 & \text{otherwise,}\end{cases}\\
\overrightarrow{n} &:= \sum_{i\in S} z_i = \big(n_1,\ldots,n_h,\ldots,n_H\big)^\top.
\end{aligned}$$

The $z_i$ are stratification variables and $\overrightarrow{n}$ is the strata allocation.

The key feature of (2) is the stratification constraint $\sum_{i\in S}p_i z_i = \overrightarrow{n}n^{-1}$, which is not motivated by moment conditions. This constraint is used to account for the sampling design. We also have that the constraint involving the parameter contains the standard sampling weights $\pi_i^{-1}$. Other approaches incorporate the information about the design within $\ell(p)$ (see Section 3). The function (2) reduces to Owen's (1988) empirical log-likelihood function when we have a single stratum and $\pi_i = n/N$, $\forall i\in U$. The advantage of (2) is that it can be used as a standard likelihood function for design-based inference.

The "*maximum empirical likelihood estimator*" $\widehat{\theta}$ is defined as the vector which maximises $\ell_{\max}(\theta)$. Berger and Torres (2016) showed that $\widehat{\theta}$ is also the solution to the sample estimating equation

$$\sum_{i\in S}\pi_i^{-1}\, e(y_i,\theta) = 0\cdot \tag{3}$$

For example, when $\theta_U$ is a population mean, we have $e(y_i,\theta) = y_i - \theta N n^{-1}\pi_i$ and $\widehat{\theta}$ is the Horvitz and Thompson's (1952) estimator. Hence, this estimator is a maximum empirical likelihood estimator. Traditional point estimators can be re-derived under the empirical likelihood framework. The key advantage is the self-normalising property which can be used for testing and model building.

The function (2) can be used for testing, by using the profile likelihood principle, or the self-normalisation property. Suppose we wish to test $H_0 : \theta_U^{(1)} = \theta_0^{(1)}$, against $H_a : \theta_U^{(1)} \neq \theta_0^{(1)}$, where $\theta_U^{(1)} \in \mathbb{R}^{d_{\theta^{(1)}}}$ is a sub-parameter of $\theta_U$; that is, $\theta_U = (\theta_U^{(1)\top}, \theta_U^{(2)\top})^\top$. Oğuz-Alper and Berger (2016) showed that under $H_0$,

$$\widehat{r}(\theta^{(1)}) := 2\Big\{\ell_{\max}(\widehat{\theta}) - \max_{\theta_U^{(2)}}\ell_{\max}(\theta)\Big\} \xrightarrow{d} \chi^2_{d_{\theta^{(1)}}}, \quad \text{if } \theta^{(1)} = \theta_0^{(1)}, \tag{4}$$

under with replacement stratified sampling, as $n \to \infty$, where $\chi^2_{d_{\theta^{(1)}}}$ denotes a $\chi^2$-distribution with $d_{\theta^{(1)}}$ degree of freedom and $\theta = (\theta^{(1)\top}, \theta^{(2)\top})^\top$. Extensions to without-replacement sampling are given in Section 2.3. The function $\widehat{r}(\theta^{(1)})$ is an empirical likelihood ancillary ratio statistics. Thus, the p-value of the test $H_0 : \theta_U^{(1)} = \theta_0^{(1)}$ is given by

$$\text{p-value} := \int_{\widehat{r}(\theta_0^{(1)})}^{\infty} \chi^2(x)\; dx, \tag{5}$$

where $\chi^2(x)$ is the density of the $\chi^2$-distribution with $d_{\theta^{(1)}}$ degree of freedom. Inverse testing can be used to construct confidence intervals, when $\theta_U^{(1)}$ is unidimensional ($d_{\theta^{(1)}} = 1$); that is, the $\alpha$-level confidence interval of a scalar $\theta_U^{(1)}$ is

$$\text{CI}(\theta_U^{(1)}) := \Big\{\theta^{(1)} : \widehat{r}(\theta^{(1)}) \leqslant \chi^2_1(\alpha)\Big\},$$

where $\chi^2_1(\alpha)$ is the upper $\alpha$-quantile of the $\chi^2$-distribution with one degree of freedom. Note that $\widehat{r}(\theta^{(1)})$ is a convex non-symmetric function with a minimum at $\theta^{(1)} = \widehat{\theta}^{(1)}$. This interval can be found by using any root search method. This involves calculating $\widehat{r}(\theta^{(1)})$ for several values of $\theta^{(1)}$. An algorithm to compute (4) can be found in Oğuz-Alper and Berger (2016, Appendix).

The ancillary statistics $\widehat{r}(\theta^{(1)})$ can be also used for model building, when comparing two nested models. In this case, $\theta_U = (\theta_U^{(1)\top}, \theta_U^{(2)\top})^\top$ is the parameter of the full model and $\theta_U^{(2)}$ is the parameter of the reduced model. The p-value (5) gives the significance of the relative fit.

**Remark 2.1.** *Berger and Torres (2016) and Oğuz-Alper and Berger (2016) used a different parametrisation based upon $m_i := np_i\pi_i^{-1}$; that is,*

$$L_{\max}(\theta) := \max_{m_i :\, i \in S} \Big\{ \sum_{i\in S} \log(m_i) \;:\; m_i > 0,\; \sum_{i\in S} m_i e(y_i, \theta) = 0,\; \sum_{i\in S} m_i \pi_i z_i = \overrightarrow{n} \Big\}. \tag{6}$$

*By substituting $m_i$ by $np_i\pi_i^{-1}$ within* (6)*, we obtain $L_{\max}(\theta) = \ell_{\max}(\theta) + \sum_{i\in S}\log(n\pi_i^{-1})$. Note that the quantity $\sum_{i\in S}\log(n\pi_i^{-1})$ does not depend on $\theta$ and $m_i$. Therefore* (2) *and* (6) *give the same maximum empirical likelihood estimate. Straightforward algebra shows that the same function* (4) *is obtained by using* (2) *or* (6)*. The empirical log-likelihood function* (6) *may be more suitable in a survey sampling context, because the $m_i$ are scale-loads which are estimated by the weights $\pi_i^{-1}$ within* (3)*. However, we prefer using* (2) *in order to simplify the comparison with other empirical likelihood approaches.*

## 2.1 Side information

Consider a different parameter $\varphi_U \in \mathbb{R}^{d_\varphi}$, which denotes some "*side information*" assumed known without sampling error, and such that

$$\sum_{i\in U} f(y_i, \varphi_U) = 0, \tag{7}$$

where $f(y_i, \varphi_U) \in \mathbb{R}^{d_f}$ ($d_f \geqslant d_\varphi$) is a known vector-function, which is often called "*auxiliary information*" in the survey sampling literature (e.g. Deville and Särndal, 1992; Lesage, 2011). For example, the most common situation in practice is to know a set of totals, means or proportions from large external censuses or surveys. Examples can be found in Imbens and Lancaster (1994), Berger and Torres (2016) and Oğuz-Alper and Berger (2016). Side information is the core of survey sampling theory (e.g. Kott, 2009). It can also be found in the mainstream empirical likelihood literature (Owen, 1991, 2001, §3.10) and in econometrics (Imbens and Lancaster, 1994).

Obviously, it will not be necessary to estimate $\varphi_U$ because it is known. We shall treat $\varphi_U$ as a vector of constants, not as a parameter to estimate. The idea is to combine $\theta_U$ and $\varphi_U$ to improve the precision of $\theta_U$. Let $\psi_U := (\theta_U^\top, \varphi_U^\top)^\top$ be the unique solution to

$$\sum_{i \in U} g(y_i, \theta, \varphi) = 0, \tag{8}$$

where

$$g(y_i, \theta, \varphi) := \left\{ e(y_i, \theta, \varphi)^\top, f(y_i, \varphi)^\top \right\}^\top \in \mathbb{R}^{d_g} \quad (d_g \geqslant d_\theta), \tag{9}$$

with $e(y_i, \theta, \varphi)$ defined as in (1). Now, we write $e(y_i, \theta, \varphi)$ as a function of $\varphi$, because it may indeed depend on $\varphi$, as in Example 2.1 below. Note that (8) implies (7).

**Example 2.1.** *Suppose we wish to fit a logistic regression model with a known success rate in the population. In this case, $y_i = (x_i^\top, \delta_i)^\top$, where $x_i$ is some covariates and $\delta_i$ is the (dependent) binary variable specifying the successes and failures. Suppose that $\varphi_U = N^{-1} \sum_{i \in U} \delta_i$ is known. The estimating functions are*

$$\begin{aligned} e(y_i, \theta, \varphi) &= x_i^\top \delta_i - x_i^\top \exp(x_i^\top \theta) \left\{ 1 + \exp(x_i^\top \theta) \right\}^{-1}, \\ f(y_i, \varphi) &= \delta_i - \varphi. \end{aligned}$$

With side information, Berger and Torres's (2016) "*empirical log-likelihood function*" is defined by

$$\ell_{\max}(\theta, \varphi_U) := \max_{p_i :\, i \in S} \left\{ \ell(p) \;:\; p_i > 0,\; \sum_{i \in S} \frac{p_i}{\pi_i} g(y_i, \theta, \varphi_U) = 0,\; \sum_{i \in S} p_i z_i = \frac{\overrightarrow{n}}{n} \right\}. \tag{10}$$

It can be shown that the "*maximum empirical likelihood estimator*" $\widehat{\theta}$ which maximises (10), is also the solution to

$$\sum_{i \in S} \widehat{m}_i(\varphi_U)\, e(y_i, \theta, \varphi_U) = 0,$$

where $\widehat{m}_i(\varphi_U)$ are the empirical likelihood weights defined by

$$\begin{aligned} \widehat{m}_i(\varphi_U) &:= n\, \widehat{p}_i(\varphi_U) \pi_i^{-1}, \\ \widehat{p}_i(\varphi_U) &:= n^{-1} \left\{ 1 + \eta(\varphi_U)^\top c_i(\varphi_U) \pi_i^{-1} \right\}^{-1}, \end{aligned} \tag{11}$$

$$c_i(\varphi_U) \;\; := \;\; \left\{ f(y_i, \varphi_U)^\top, \pi_i z_i^\top \right\}^\top \cdot$$

Here, $\eta(\varphi_U)$ is a Lagrangian parameter which is such that

$$\sum_{i \in S} \widehat{m}_i(\varphi_U)\, c_i(\varphi_U) = \left(0^\top, \overrightarrow{n}^\top\right)^\top \cdot$$

A modified Newton-Raphson algorithm (e.g. Polyak, 1987) can be used to compute $\eta(\varphi_U)$. Oğuz-Alper and Berger (2016) showed that the self-normalisation property holds; that is, (4) holds after replacing $\ell_{\max}(\theta)$ replaced by $\ell_{\max}(\theta, \varphi_U)$. Hence, (10) can be used for testing, confidence intervals and model building.

## 2.2 Empirical likelihood versus calibration

Empirical likelihood should not be viewed as a particular case of calibration (Deville and Särndal, 1992). Calibration relies on auxiliary information. On the other hand, empirical likelihood can be used without auxiliary information. However, the calibration property indeed holds with the empirical likelihood weights $\widehat{m}_i(\varphi_U)$ because the constrain within (10) and (9) imply

$$\sum_{i \in S} \widehat{m}_i(\varphi_U)\, f(y_i, \varphi_U) = 0 \cdot$$

This property is the consequence of the maximisation of (2) and the fact that $\varphi_U$ is constant. Here, calibration is a property which is the results of a maximum likelihood principle. In survey sampling literature, calibration is viewed from a different angle. It is mainly a weighting procedure, rather than the consequence of the maximisation of likelihood function.

Calibration relies on a distance function between the sampling weights $\pi_i^{-1}$ and the calibrated weights (Deville and Särndal, 1992). This function is only used for weighting and does not serve any other purpose, other than obtaining cosmetically acceptable weights. The distance function is also disconnected from the mainstream statistical theory. With empirical likelihood, we have an objective function $\ell(p)$, rather than a distance function, because $\ell(p)$ does not depend on $\pi_i$. This function is related to the likelihood principle in mainstream statistics. This function is used for point estimation, for tests and confidence intervals.

Empirical likelihood is based on a likelihood principle based on a maximisation of an objective function. This gives point estimates and an ancillary statistics (4). Calibration is just a procedure to obtain weights satisfying a given constraint. It is worth noticing that the first works related to calibration (Hartley and Rao, 1969; Owen, 1991; Imbens and Lancaster, 1994) are linked with likelihood principles.

## 2.3 Large sampling fractions and sampling without-replacement

The self-normalising property (4) is based on sampling with-replacement or sampling without-replacement with negligible sampling fraction $n/N$, since with and without-replacement sampling are equivalent when $n/N$ is negligible. In this section, we show how empirical likelihood can be extended to accommodate non-negligible sampling fractions and sampling without-replacement. This approach is limited to single stage sampling. An extension to large sampling fraction with multi-stage sampling can be found in Berger (2018*a*).

The approach of Section 2.1 can still be used for point estimation, but the property (4) does not hold. A solution is to use Berger and Torres's (2016) "*penalised empirical likelihood function*", which is defined by

$$\widetilde{\ell}_{\max}(\theta, \varphi_U) := \max_{p_i:\, i\in S}\Big\{\widetilde{\ell}(p) \ : \ p_i > 0,\ \sum_{i\in S}\frac{\nu_i}{\pi_i}g(y_i, \theta, \varphi_U) = 0, \sum_{i\in S}\nu_i z_i = \frac{\overrightarrow{n}}{n}\Big\},$$

where

$$\widetilde{\ell}(p) \ := \ \sum_{i\in S}\log(p_i) - n\sum_{i\in S}p_i + n,$$

$\nu_i := p_i q_i - \psi_i$ are penalties, with $q_i := (1-\pi_i)^{1/2}$, $\psi_i := (q_i - 1)n^{-1}$. Note that $\nu_i$ is a function of $p_i$. The $q_i$ are Hájek's (1964) finite population corrections. Under without-replacement stratified sampling and Hájek's (1964) asymptotic framework, Berger and Torres (2016) showed that under $H_0 : \theta_U = \theta_0$, we have that

$$\widetilde{r}(\theta, \varphi_U) := 2\Big\{\widetilde{\ell}_{\max}(\varphi_U) - \widetilde{\ell}_{\max}(\theta, \varphi_U)\Big\} \xrightarrow{d} \chi^2_{d_\theta}, \quad \text{if } \theta = \theta_0, \tag{12}$$

where

$$\widetilde{\ell}_{\max}(\varphi_U) \ := \ \max_{p_i:\, i\in S}\Big\{\widetilde{\ell}(p) \ : \ p_i > 0,\ \sum_{i\in S}\frac{\nu_i}{\pi_i}f(y_i, \varphi_U) = 0, \sum_{i\in S}\nu_i z_i = \frac{\overrightarrow{n}}{n}\Big\}.$$

Berger (2016) extended this Section's approach to Rao et al.'s (1962) sampling design with large sampling fraction. Tests and confidence regions can be derived from (12).

## 2.4 Multi-stage sampling

Berger (2018*a*) showed how empirical likelihood can be modified to accommodate multi-stage sampling and non-response, when the primary sampling units (PSUs) are sampled with unequal probabilities. The sample of PSUs can be stratified. Side information at PSU-level or at a lower level can be taken into account. The key assumption is a negligible sampling fraction at PSU-level.

The idea is to use a "*PSU-level empirical likelihood function*", which can be found in Berger (2018*a*). We not give the actual expression of this function, because the notations are heavy due to the multi-stage structure of the design. Berger (2018*a*) gives the regularity conditions under which the PSU-level empirical likelihood ratio statistics is ancillary as in (4). In Berger (2018*a*), this empirical likelihood approach is applied to a logistic model based on the 2006 PISA survey data (OECD, 2006) for the United Kingdom. The empirical likelihood p-value (as in (5)) can be significant when naïve p-values are not significant.

This assumption of negligible sampling fraction can be relaxed, but there is a price to pay. Berger (2018*a*) showed how empirical likelihood can be modified to obtain an ancillary empirical log-likelihood ratio statistic likelihood under multi-stage sampling with large sampling fraction. Unfortunately, this involves calculating an adjustment factor based on variance estimates. In Berger (2018*a*), a simulation study suggests that the effect of this factor is small. Non-adjusted confidence intervals based on the PSU-level version of (4) gives more conservative intervals, and even better coverages than the approach involving an adjustment factor. Thus, this suggests that the more conservative confidence intervals based on (4) may be preferable.

## 2.5 Unit non-response

Non-response is another important aspect which is covered by Berger (2018*a*). In order to simplify the notation, we now consider a single-stage sampling. The general approach involving multi-stage sampling and non-response can be found in Berger (2018*a*). Response propensities can be used to adjust for missing data. This involves adding an additional non-response constraint to the empirical likelihood function. Adding the non-response constraint is equivalent of modifying the function (9); that is, (9) needs to be replaced by

$$g(y_i, \theta, \lambda, \varphi) := \left\{ r_i P_i(\lambda)^{-1} e(y_i, \theta, \varphi)^\top, \xi_i^\top \{r_i - P_i(\lambda)\}, f(y_i, \varphi)^\top \right\}^\top,$$

where $r_i = 0$ if the unit $i$ is missing and $r_i = 1$ otherwise. Here,

$$P_i(\lambda) := \Psi^{-1}(\xi_i^\top \lambda),$$

where $\Psi^{-1} : \mathbb{R} \rightarrow (0, 1]$ is the inverse of a link function $\Psi$ (e.g. logit, probit, complementary log-log). The vector $\lambda$ is a non-response parameter and $\xi_i$ denotes (non-missing) variables, which defines the non-response mechanism. The function $\Psi$ may describe re-weighting classes, when the $\xi_i$ are dichotomous variables describing categories. In this case, $P_i(\widehat{\lambda})$ reduces to response rates. For point estimation, the sample estimating equation (3) reduces to

$$\sum_{i \in S} \widehat{m}_i(\varphi_U) \, \frac{r_i}{P_i(\lambda)} e(y_i, \theta, \varphi_U) \;=\; 0, \tag{13}$$

$$\sum_{i \in S} \widehat{m}_i(\varphi_U) \, \xi_i^\top \{r_i - P_i(\lambda)\} \;=\; 0\cdot \tag{14}$$

The quantities $r_i P_i(\lambda)^{-1}$ are "*propensity-score adjustments*". The non-response parameter $\lambda$ is estimated from (14), which is the weighted estimating equation of a generalised linear model. The parameter $\theta_U$ is estimated from the equation (13), which includes the propensity-scores $r_i P_i(\lambda)^{-1}$.

Berger (2018*a*) showed that the independence between the response mechanism and the sampling design implies that the empirical log-likelihood ratio statistic likelihood (4) is ancillary and does not need to be adjusted for missing data. It is common practice to treat the estimated response propensities as deterministic within variance estimators. This may lead to shorter confidence intervals. Since the empirical log-likelihood ratio statistic likelihood possesses the self-normalising property, the confidence intervals reflect the estimation of these propensities.

## 3 Pseudo-empirical likelihood

Chen and Sitter's (1999) developed a different empirical likelihood approach called "*pseudo-empirical likelihood*" (see also Wu and Rao, 2006; Rao and Wu, 2009). For simplicity, non-response is not considered in this Section The "*pseudo-empirical log-likelihood function*" is defined by

$$\mathcal{L}_{\max}(\theta, \varphi_U) := \max_{p_i:\, i \in S} \Big\{ \mathcal{L}(p) \;:\; p_i > 0,\; \sum_{i \in S} p_i\, g(y_i, \theta, \varphi_U) = 0,\; \sum_{i \in S} p_i z_i = 1_H \Big\}, \tag{15}$$

where $1_H$ is the $H \times 1$ unit vector,

$$\mathcal{L}(p) \;:=\; n \sum_{i \in S} \frac{\phi_i}{\pi_i} \log(p_i), \quad \phi_i := \frac{1}{N} \sum_{h=1}^{H} \frac{N_h}{\widehat{N}_h} z_{ih},$$

$N_h := \sum_{i \in U} z_{ih}$, $\widehat{N}_h := \sum_{i \in S} z_{ih} \pi_i^{-1}$. The function (15) is not an empirical likelihood function, because $\mathcal{L}(p)$ is different from $\ell(p)$. This is the reason why the approach is called "*pseudo-empirical likelihood*". The function $\mathcal{L}(p)$ is adjusted to take the design into account. The $\pi_i$ are incorporated within $\mathcal{L}(p)$ and $\phi_i$ takes into account of the stratification. The stratification constraint $\sum_{i \in S} p_i z_i = 1_H$ is also different from the stratification constraint $\sum_{i \in S} p_i z_i = \overrightarrow{n}\, n^{-1}$ used within (2).

Chen and Sitter (1999) showed that the "*maximum pseudo-empirical likelihood estimator*", which maximises $\mathcal{L}_{\max}(\theta, \varphi_U)$, is the solution to

$$\widehat{E}(\theta, \varphi_U) := \sum_{i \in S} \widehat{w}_i(\varphi_U)\, e(y_i, \theta, \varphi_U) = 0, \tag{16}$$

where $\widehat{w}_i(\varphi_U)$ are the pseudo-empirical likelihood weights which are different from (11) (see Berger, 2018*b*, for an expression for $\widehat{w}_i(\varphi_U)$ using this paper's notation). Hence, maximum pseudo-empirical likelihood and empirical likelihood estimates are different. However, simulation studies in Berger and Torres (2016) shows that the differences are usually negligible, as long as the same side information is used.

The main issue with the pseudo-empirical likelihood approach is that the pseudo-empirical log-likelihood ratio statistic likelihood is not ancillary. To solve this problem, Wu and Rao (2006) proposed to multiply this statistics by a "*design effect*". However, we will see that this also bring other issues. Suppose that $d_\theta = d_e = 1$; that is, we have a scalar parameter $\theta_U$. Let $\widehat{\theta}$ be the maximum pseudo-empirical likelihood estimator. Wu and Rao (2006) showed that under $H_0 : \theta_U = \theta_0$, we have that

$$\widehat{r}(\theta, \varphi_U)_{\mathrm{PEL}} := \frac{2\big\{\mathcal{L}_{\max}(\widehat{\theta}, \varphi_U) - \mathcal{L}_{\max}(\theta, \varphi_U)\big\}}{\mathrm{Deff}(\theta_0, \varphi_U)} \xrightarrow{d} \chi_1^2, \quad \text{if } \theta = \theta_0, \tag{17}$$

where $\mathrm{Deff}(\theta_0, \varphi_U)$ is called the "*design effect*" and is given by

$$\mathrm{Deff}(\theta_0, \varphi_U) := \mathrm{Var}\{\widehat{E}(\theta_0, \varphi_U)\}\, \mathrm{Var}_{\mathrm{SRS}}\{\widehat{E}(\theta_0, \varphi_U)\}^{-1}. \tag{18}$$

Here, $\widehat{E}(\theta_0, \varphi_U)$ is defined by (16), when $d_e = 1$. The quantity $\mathrm{Var}\{\widehat{E}(\theta, \varphi_U)\}$ is the variance under the sampling design and $\mathrm{Var}_{\mathrm{SRS}}\{\widehat{E}(\theta, \varphi_U)\}$ is the variance under simple random sampling. This design effect is a population value that would need to be estimated. We refer to Wu and Rao (2006), for more details about the estimation of (18).

Pseudo-empirical likelihood can be applied in principle to any complex sampling designs, because the design effect takes the complexity of the design into account. Note that the approach of Section 2 covers most of the designs used in practice. The property (4) is limited to multi-stage design with small sampling fractions. The property (12) holds for single stage design with large and small sampling fractions, under Hájek (1964) asymptotic framework.

The function (17) has the disadvantage of relying on variance estimates, which can be tedious to compute under complex sampling. The estimation of the design effect adds some additional variability that may affect the convergence of $\widehat{r}(\theta_U, \varphi_U)_{\mathrm{PEL}}$ towards the $\chi^2$ distribution. Berger and Torres (2016) showed via a series of simulation that coverages of confidence intervals obtained from (4) and (12) are closer to the nominal value, than coverages obtained from (17).

The main disadvantage of pseudo-empirical likelihood is that (17) is based on a scalar parameter ($d_\theta = d_e = 1$). It cannot be used with multidimensional parameters, because the design effect has to be scalar; that is, $\widehat{E}(\cdot)$ and the variances have to be unidimensional. Thus, the pseudo-empirical log-likelihood ratio statistic likelihood cannot be used for multi-dimensional regression parameters. It is recommended to use a traditional approach based on linearised variances estimates computed from (16) and use pseudo-empirical likelihood as a method to derive calibrated weights. This has modest advantages over traditional approaches. The key advantage of empirical likelihood is the self-normalising property which does not hold with pseudo-empirical likelihood for multidimensional parameters. On the other hand, the self-normalising property holds with multidimensional parameters under the approach of Section 2.

# References


Berger, Y. G. (2016), "Empirical Likelihood Inference for the Rao-Hartley-Cochran Sampling Design," *Scandinavian Journal of Statistics*, 43, 721–735.

Berger, Y. G. (2018*a*), "An empirical likelihood approach under cluster sampling with missing observations," *Annals of the Institute of Statistical Mathematics (accepted for publication)*, `https://eprints.soton.ac.uk/421738/`.

Berger, Y. G. (2018*b*), "Empirical likelihood approaches under complex sampling designs," *Wiley StatsRef: Statistics Reference Online*, `doi:10.1002/9781118445112.stat08066`, 20pp.

Berger, Y. G., and Torres, O. D. L. R. (2012), "A unified theory of empirical likelihood ratio confidence intervals for survey data with unequal probabilities," *Proceedings of the Survey Research Method Section of the American Statistical Association, Joint Statistical Meeting, San Diego*, p. 15pp.

Berger, Y. G., and Torres, O. D. L. R. (2014), "Empirical likelihood confidence intervals:

an application to the EU-SILC household surveys," *Contribution to Sampling Statistics, Contribution to Statistics: F. Mecatti, P. L. Conti, M. G. Ranalli (editors). Springer*, p. 20pp.

Berger, Y. G., and Torres, O. D. L. R. (2016), "An empirical likelihood approach for inference under complex sampling design," *Journal of the Royal Statistical Society Series B*, 78(2), 319–341.

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

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

Deville, J. C., and Särndal, C.-E. (1992), "Calibration Estimators in Survey Sampling," *Journal of the American Statistical Association*, 87(418), 376–382.

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.

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

Hartley, H. O., and Rao, J. N. K. (1969), "A new estimation theory for sample surveys, II," in *New Developments in survey Sampling*, eds. N. L. Johnson, and H. J. Smith, New York: John Wiley and Sons, pp. 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(260), 663–685.

Imbens, G. W., and Lancaster, T. (1994), "Combining Micro and Macro Data in Microeconometric Models," *The Review of Economic Studies*, 61(4), 655–680.

Kott, P. S. (2009), "Calibration weighting: combining probability samples and linear prediction models," in *Sample Surveys: Design, Methods and Applications*, eds. D. Pfeffermann, and C. Rao, Handbook of Statistics, Amsterdam: Elsevier, pp. 55–82.

Lesage, E. (2011), "The use of estimating equations to perform a calibration on complex parameters," *Survey Methodology*, 37(1), 103–108.

Neyman, J. (1938), "On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection," *Journal of the Royal Statistical Society*, 97(4), 558–625.

OECD (2006), *PISA 2006 Technical Report*, `https://www.oecd.org/pisa/data/42025182.pdf`.

Oğuz-Alper, M., and Berger, Y. G. (2016), "Empirical likelihood approach for modelling survey data," *Biometrika*, 103(2), 447–459.

Owen, A. B. (1988), "Empirical Likelihood Ratio Confidence Intervals for a Single Functional," *Biometrika*, 75(2), 237–249.

Owen, A. B. (1991), "Empirical Likelihood for Linear Models," *Ann. Statist.*, 19(4), 1725–1747.

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

Polyak, B. T. (1987), *Introduction to Optimization*, New York: Optimization Software, Inc., Publications Division.

Qin, J., and Lawless, J. (1994), "Empirical Likelihood and General Estimating Equations," *Ann. Statist.*, 22(1), pp. 300–325.

Rao, J. N. K., Hartley, H. O., and Cochran, W. G. (1962), "On a Simple Procedure of Unequal Probability Sampling without Replacement," *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 24(2), pp. 482–491.

Rao, J. N. K., and Wu, C. (2009), "Empirical Likelihood Methods," in *Sample Surveys: Inference and Analysis*, eds. D. Pfeffermann, and C. Rao, Handbook of Statistics, Amsterdam: Elsevier, pp. 189–207.

Wu, C., and Rao, J. N. K. (2006), "Pseudo-empirical likelihood ratio confidence intervals for complex surveys," *Canad. J. Statist.*, 34(3), 359–375.