Empirical likelihood approaches under complex sampling designs

Y.G. Berger
Southampton Statistical Sciences Research Institute,
University of Southampton

This is the submitted version which will be published in final form in Wiley Statsref

http://wileyactual.com/wileystatsref/

Abstract

There are two different empirical likelihood approaches for complex sampling designs: the "*pseudoempirical likelihood*" introduced by Chen and Sitter (1999) and "*unequal probability empirical likelihood*" approach proposed by Berger and Torres (2016). Both approaches are described and reviewed critically. The key difference is the fact that the self-normalisation property of the pseudoempirical likelihood approach is limited to unidimensional parameters. This property holds for multidimensional parameters, with the unequal probability empirical likelihood approach. This is a brief description of the key empirical likelihood approaches for complex sampling. This is not an exhaustive account of all the applications of empirical likelihood in survey sampling.

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

1 Introduction

Consider a finite population $\mathcal{U} = \{1, \ldots, N\}$ of N units. For each unit $i \in \mathcal{U}$, we have a vector $\boldsymbol{y}_i$ of variables, where $\boldsymbol{y}_i \in \mathbb{R}^{d_{\boldsymbol{y}}}$. We consider a "*design-based approach*"; that is, we assume that $\boldsymbol{y}_i$ are vectors of constants (Neyman, 1938). A random sample $\boldsymbol{S}$ is selected from $\mathcal{U}$ according to a sampling design, which specifies the probability distribution of $\boldsymbol{S}$. The sampling distribution is therefore given by the sampling design. The design-based approach is the core of survey sampling theory. It offers a robust non-parametric approach for survey data, which does not rely on distributional assumption about $\boldsymbol{y}_i$.

We consider that the population $\mathcal{U}$ is split into H non-overlapping strata $\mathcal{U}_1, \ldots, \mathcal{U}_h, \ldots, \mathcal{U}_H$ such that $\cup_{h=1}^{H} \mathcal{U}_h = \mathcal{U}$. Within each stratum $\mathcal{U}_h$, a sample of n_h units is selected with unequal selection probabilities π_i. The overall sample size is the constant $n = \sum_{h=1}^{H} n_h$. With and without replacement sampling will be considered . We shall focus on single-stage designs.

Empirical likelihood has its origin in survey sampling theory. One of the first attempts to formulate a likelihood-based approach in survey sampling is due to Godambe (1966) who showed that under the design-based approach, the likelihood function is flat and cannot be used for inference. In order to solve this problem, Hartley and Rao (1968) developed the first application of an empirical likelihood approach under equal probability sampling. Owen (1988) popularised this approach into the mainstream statistics (see also Owen, 2001). There have been many recent developments of empirical likelihood based methods in survey sampling. Most of which can be classified into two categories: the unequal probability empirical likelihood approach (Berger and Torres, 2016) and the pseudoempirical likelihood approach (Chen and Sitter, 1999; Wu and Rao, 2006).

2 Parameter and side information

Let $\boldsymbol{\theta}_{\mathcal{U}} \in \mathbb{R}^{d_{\boldsymbol{\theta}}}$ be an unknown vector of $d_{\boldsymbol{\theta}}$ parameters. For example, $\boldsymbol{\theta}_{\mathcal{U}}$ can be the parameter of a generalised linear regression model. Let $\boldsymbol{\varphi}_{\mathcal{U}} \in \mathbb{R}^{d_{\boldsymbol{\varphi}}}$ denote another vector called "*side information*" which is assumed known without sampling errors. For example, $\boldsymbol{\varphi}_{\mathcal{U}}$ can be population-level means, counts or proportions, from large external censuses or surveys (see Example 2.1). Side information is often used in survey sampling (e.g. Kott, 2009), but not limited to this field. It can also be found in the empirical likelihood literature (Owen, 1991, 2001, §3.10) and in the econometric literature (Imbens and Lancaster, 1994). Taking into account of side information may increase the efficiency (Owen, 1991).

Let $\boldsymbol{\psi}_{\mathcal{U}} = (\boldsymbol{\theta}_{\mathcal{U}}^{\top}, \boldsymbol{\varphi}_{\mathcal{U}}^{\top})^{\top}$. We assume that $\boldsymbol{\psi}_{\mathcal{U}}$ is the unique solution to the population multidimensional estimating equation (Godambe, 1960)

$$\sum_{i \in \mathcal{U}} \boldsymbol{g}(\boldsymbol{y}_i, \boldsymbol{\theta}, \boldsymbol{\varphi}) = \mathbf{0}_{d_{\boldsymbol{g}}}, \quad \text{if and only if } \boldsymbol{\theta} = \boldsymbol{\theta}_{\mathcal{U}} \text{ and } \boldsymbol{\varphi} = \boldsymbol{\varphi}_{\mathcal{U}}; \tag{1}$$

where

$$\boldsymbol{g}(\boldsymbol{y}_i, \boldsymbol{\theta}, \boldsymbol{\varphi}) := \left\{ \boldsymbol{e}(\boldsymbol{y}_i, \boldsymbol{\theta}, \boldsymbol{\varphi})^{\top}, \mathbf{f}(\boldsymbol{y}_i, \boldsymbol{\varphi})^{\top} \right\}^{\top} \in \mathbb{R}^{d_{\boldsymbol{g}}} \quad (d_{\boldsymbol{g}} \geqslant d_{\boldsymbol{\theta}}), \tag{2}$$

with $\mathbf{f}(\boldsymbol{y}_i, \boldsymbol{\varphi}_{\mathcal{U}}) \in \mathbb{R}^{d_{\mathbf{f}}}$ ($d_{\mathbf{f}} \geqslant d_{\boldsymbol{\varphi}}$) and $\boldsymbol{e}(\boldsymbol{y}_i, \boldsymbol{\theta}, \boldsymbol{\varphi}) \in \mathbb{R}^{d_{\boldsymbol{e}}}$ ($d_{\boldsymbol{e}} \geqslant d_{\boldsymbol{\theta}}$). Here, $\mathbf{0}_{d_{\boldsymbol{g}}}$ denotes a $d_{\boldsymbol{g}}$-vector of zeros. The estimating function $\boldsymbol{e}(\boldsymbol{y}_i, \boldsymbol{\theta}, \boldsymbol{\varphi})$ defines $\boldsymbol{\theta}_{\mathcal{U}}$. For example, $\boldsymbol{\theta}_{\mathcal{U}}$ can be the parameter of a generalised linear regression model (Chen and Van Keilegom, 2009). Under a design-based approach, the parameter $\boldsymbol{\theta}_{\mathcal{U}}$ is a vector of unknown constants. Note that the function $\boldsymbol{e}(\boldsymbol{y}_i, \boldsymbol{\theta}, \boldsymbol{\varphi})$ may or may not depends on $\boldsymbol{\varphi}$.

Note that (1) implies that $\boldsymbol{\varphi}_{\mathcal{U}}$ is such that

$$\sum_{i \in \mathcal{U}} \mathbf{f}(\boldsymbol{y}_i, \boldsymbol{\varphi}_{\mathcal{U}}) = \mathbf{0}, \tag{3}$$

by definition. The $\mathbf{f}(\boldsymbol{y}_i, \boldsymbol{\varphi}_{\mathcal{U}})$ are usually a function of a sub-vector $\widetilde{\boldsymbol{y}}_i$ of $\boldsymbol{y}_i$, with $\widetilde{\boldsymbol{y}}_i$ known $\forall i \in \mathcal{U}$. We assume that $\mathbf{f}(\boldsymbol{y}_i, \boldsymbol{\varphi}_{\mathcal{U}})$ are known $\forall i \in \boldsymbol{S}$. The $\mathbf{f}(\boldsymbol{y}_i, \boldsymbol{\varphi}_{\mathcal{U}})$ are called '*auxiliary variables*' in the survey sampling literature (e.g. Deville and Särndal, 1992). We shall treat $\boldsymbol{\varphi}_{\mathcal{U}}$ as a vector of constants, not as a parameter to estimate.

The estimating function $\mathbf{f}(\boldsymbol{y}_i, \boldsymbol{\varphi}_{\mathcal{U}})$ cannot be a function of $\boldsymbol{\theta}_{\mathcal{U}}$, because $\boldsymbol{\varphi}_{\mathcal{U}}$ is assumed known and defined by (3). Indeed, if $\mathbf{f}(\boldsymbol{y}_i, \boldsymbol{\varphi})$ depends on $\boldsymbol{\theta}_{\mathcal{U}}$, the vector $\boldsymbol{\varphi}_{\mathcal{U}}$ cannot be known, unless $\boldsymbol{\theta}_{\mathcal{U}}$ is known, which is a trivial situation where inference is not necessary.

Example 2.1. *Let $\boldsymbol{y}_i = (\boldsymbol{x}_i^\top, \delta_i)^\top$, where $\boldsymbol{x}_i$ is some covariates and δ_i is a binary variable: $\delta_i = 1$ for a success and $\delta_i = 0$ for a failure. Suppose that we know the population success rate $\boldsymbol{\varphi}_{\mathcal{U}} = N^{-1} \sum_{i \in \mathcal{U}} \delta_i$, and we wish to fit a binary logistic model with δ_i as dependent variable and $\boldsymbol{x}_i$ as covariates. The estimating functions are*

$$\begin{aligned} \boldsymbol{e}(\boldsymbol{y}_i, \boldsymbol{\theta}, \boldsymbol{\varphi}) &= \boldsymbol{x}_i^\top \delta_i - \boldsymbol{x}_i^\top \exp(\boldsymbol{x}_i^\top \boldsymbol{\theta}) \big\{1 + \exp(\boldsymbol{x}_i^\top \boldsymbol{\theta})\big\}^{-1}, \\ \mathbf{f}(\boldsymbol{y}_i, \boldsymbol{\varphi}) &= \delta_i - \boldsymbol{\varphi}. \end{aligned}$$

More examples can be found in Imbens and Lancaster (1994), Berger and Torres (2016) and Oğuz-Alper and Berger (2016).

3 Unequal probability empirical likelihood

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

$$\ell_{\max}(\boldsymbol{\theta}, \boldsymbol{\varphi}_{\mathcal{U}}) := \max_{\boldsymbol{p}} \Big\{ \ell(\boldsymbol{p}) \ : \ p_i > 0, \ \sum_{i \in \boldsymbol{S}} \frac{p_i}{\pi_i} \boldsymbol{g}(\boldsymbol{y}_i, \boldsymbol{\theta}, \boldsymbol{\varphi}_{\mathcal{U}}) = \mathbf{0}_{d_{\boldsymbol{g}}}, \ \sum_{i \in \boldsymbol{S}} p_i \boldsymbol{z}_i = \frac{\boldsymbol{n}_H}{n} \Big\}, \tag{4}$$

where $\boldsymbol{p} = (p_i : i \in \boldsymbol{S})^\top$ is the n-vector of p_i,

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

The $\boldsymbol{z}_i$ are stratification variables and $\boldsymbol{n}_H$ is the strata allocation, and $\ell(\boldsymbol{p})$ is Owen's (1988) empirical log-likelihood function . Berger and Torres's (2016) empirical log-likelihood function is expressed in term of $m_i := np_i\pi_i^{-1}$. If we substitute p_i within (4) by $n^{-1}m_i\pi_i$, straightforward algebra shows (4) reduces to Berger and Torres's (2016) empirical log-likelihood function plus a quantity that does not depend on $\boldsymbol{\theta}$. We prefer expressing (4) in term of p_i in order to ease the comparison with other empirical likelihood approaches.

The constraints within (4) differs from Owen's (1988) constraints. We have π_i within the first constraint involving the parameter. We also have a second stratification constraint not motivated by moment conditions. This constraint implies $\sum_{i\in\boldsymbol{S}} p_i = 1$, since $\mathbf{1}_H^\top \boldsymbol{z}_i = 1$ and $\mathbf{1}_H^\top \boldsymbol{n}_H = n$, where $\mathbf{1}_H$ is the H-vector of 1. Hence, the stratification constraint can be viewed as a generalisation of Owen's (1988) leading constraint $\sum_{i\in\boldsymbol{S}} p_i = 1$.

The "*maximum empirical likelihood estimator*" $\widehat{\boldsymbol{\theta}}$ maximises $\ell_{\max}(\boldsymbol{\theta}, \boldsymbol{\varphi}_{\mathcal{U}})$ by definition . Berger and Torres (2016) show that $\widehat{\boldsymbol{\theta}}$ is also the solution of the sample estimating equation

$$\sum_{i\in\boldsymbol{S}} \widehat{m}_i(\boldsymbol{\varphi}_{\mathcal{U}})\, \boldsymbol{e}(\boldsymbol{y}_i, \boldsymbol{\theta}, \boldsymbol{\varphi}_{\mathcal{U}}) = \mathbf{0}_{d_{\boldsymbol{g}}}, \tag{5}$$

where

$$\begin{aligned} \widehat{m}_i(\boldsymbol{\varphi}_{\mathcal{U}}) &:= n\,\widehat{p}_i(\boldsymbol{\varphi}_{\mathcal{U}})\pi_i^{-1}, \\ \widehat{p}_i(\boldsymbol{\varphi}_{\mathcal{U}}) &:= n^{-1}\big\{1 + \boldsymbol{\eta}(\boldsymbol{\varphi}_{\mathcal{U}})^\top \boldsymbol{c}_i(\boldsymbol{\varphi}_{\mathcal{U}})\pi_i^{-1}\big\}^{-1}, \\ \boldsymbol{c}_i(\boldsymbol{\varphi}_{\mathcal{U}}) &:= \big\{\mathbf{f}(\boldsymbol{y}_i, \boldsymbol{\varphi}_{\mathcal{U}})^\top, \pi_i \boldsymbol{z}_i^\top\big\}^\top. \end{aligned} \tag{6}$$

Here, $\boldsymbol{\eta}(\boldsymbol{\varphi}_{\mathcal{U}})$ is a Lagrangian parameter which is such that

$$\sum_{i\in\boldsymbol{S}} \widehat{m}_i(\boldsymbol{\varphi}_{\mathcal{U}})\, \boldsymbol{c}_i(\boldsymbol{\varphi}_{\mathcal{U}}) = \big(\mathbf{0}_{d_{\mathbf{f}}}^\top, \boldsymbol{n}_H^\top\big)^\top.$$

A modified Newton-Raphson algorithm (e.g. Polyak, 1987) can be used to compute $\boldsymbol{\eta}(\boldsymbol{\varphi}_{\mathcal{U}})$. We assume that $(\mathbf{0}_{d_{\mathbf{f}}}^\top, \boldsymbol{n}_H^\top)^\top$ is an inner point of the convex conical hull of $\{\boldsymbol{c}_i(\boldsymbol{\varphi}_{\mathcal{U}}) : i \in \boldsymbol{S}\}$, so that a unique solution $\boldsymbol{\eta}(\boldsymbol{\varphi}_{\mathcal{U}})$ exists.

The quantities $\widehat{m}_i(\boldsymbol{\varphi}_{\mathcal{U}})$ are the empirical likelihood weights. If we do not have side information, $\mathbf{f}(\boldsymbol{y}_i, \boldsymbol{\varphi})$ is removed from (2) and $\widehat{m}_i(\boldsymbol{\varphi}_{\mathcal{U}})$ reduces to the standard sampling weights π_i^{-1}. If $\boldsymbol{\theta}_{\mathcal{U}}$ is a population mean, we obtain the Horvitz and Thompson's (1952) estimator with $\boldsymbol{e}(\boldsymbol{y}_i, \boldsymbol{\theta}, \boldsymbol{\varphi}_{\mathcal{U}}) = \boldsymbol{y}_i - \boldsymbol{\theta} N n^{-1}\pi_i$. Note that the calibration property (Deville and Särndal, 1992) holds because of $\sum_{i\in\boldsymbol{S}} \widehat{m}_i(\boldsymbol{\varphi}_{\mathcal{U}})\, \mathbf{f}(\boldsymbol{y}_i, \boldsymbol{\varphi}_{\mathcal{U}}) = \mathbf{0}_{d_{\mathbf{f}}}$ and (3) . We have this property because we maximise (4) and $\boldsymbol{\varphi}_{\mathcal{U}}$ is constant. In survey sampling literature, calibration is viewed as a weighting procedure, rather than the consequence of the maximisation of likelihood function.

On the main advantage of this §'s approach is that the function (4) can be used for testing. Suppose we wish to test $H_0 : \boldsymbol{\theta}_{\mathcal{U}}^{(1)} = \boldsymbol{\theta}_0^{(1)}$, where $\boldsymbol{\theta}_{\mathcal{U}}^{(1)} \in \mathbb{R}^{d_{\boldsymbol{\theta}^{(1)}}}$ is a sub-parameter of $\boldsymbol{\theta}_{\mathcal{U}}$; that is, $\boldsymbol{\theta}_{\mathcal{U}} = (\boldsymbol{\theta}_{\mathcal{U}}^{(1)\top}, \boldsymbol{\theta}_{\mathcal{U}}^{(2)\top})^\top$. Oğuz-Alper and Berger (2016) showed that under H_0,

$$\widehat{r}(\boldsymbol{\theta}^{(1)}, \boldsymbol{\varphi}_{\mathcal{U}}) := 2\Big\{\ell_{\max}(\widehat{\boldsymbol{\theta}}, \boldsymbol{\varphi}_{\mathcal{U}}) - \max_{\boldsymbol{\theta}_{\mathcal{U}}^{(2)}\in\boldsymbol{\Theta}^{(2)}} \ell_{\max}(\boldsymbol{\theta}, \boldsymbol{\varphi}_{\mathcal{U}})\Big\} \xrightarrow{d} \chi^2_{d_{\boldsymbol{\theta}^{(1)}}}, \quad \text{if } \boldsymbol{\theta}^{(1)} = \boldsymbol{\theta}_0^{(1)}, \tag{7}$$

under with replacement stratified sampling, as $n \to \infty$. Inverse testing can be used to construct confidence intervals, when $\boldsymbol{\theta}_{\mathcal{U}}^{(1)}$ is unidimensional. The property (7) is also known as the self-normalisation property. Oğuz-Alper and Berger (2016) also showed how this approach can be extended to multi-stage designs.

3.1 Extension for large sampling fractions

It is common practice for business surveys to use non-negligible sampling fractions n/N. In this case, sampling without-replacement is preferred over sampling with-replacement, in order to avoid selecting the same units several times. With large sampling fractions and without-replacement sampling, the maximum empirical likelihood estimator is still the solution of (5), but the property (7) does not hold. Berger and Torres (2016) proposed a "*penalised empirical likelihood function*" as a solution to this problem. This function is defined by

$$\begin{aligned}\widetilde{\ell}_{\max}(\boldsymbol{\theta}, \boldsymbol{\varphi}_{\mathcal{U}}) \;\; := \;\; \max_{\boldsymbol{p}} \Big\{ \widetilde{\ell}(\boldsymbol{p}) \;:\; p_i > 0,\; \sum_{i \in \boldsymbol{S}} (p_i q_i - \psi_i) \frac{1}{\pi_i} \boldsymbol{g}(\boldsymbol{y}_i, \boldsymbol{\theta}, \boldsymbol{\varphi}_{\mathcal{U}}) = \mathbf{0}_{d_{\boldsymbol{g}}}, \\ \sum_{i \in \boldsymbol{S}} (p_i q_i - \psi_i) \boldsymbol{z}_i = \frac{\boldsymbol{n}_H}{n} \Big\}, \end{aligned} \tag{8}$$

where

$$\begin{aligned}\widetilde{\ell}(\boldsymbol{p}) \;\; &:= \;\; \sum_{i \in \boldsymbol{S}} \log(p_i) - n \sum_{i \in \boldsymbol{S}} p_i + n, \\ q_i \;\; &:= \;\; (1 - \pi_i)^{1/2}, \\ \psi_i \;\; &:= \;\; (q_i - 1) n^{-1}, \cdot \end{aligned}$$

The penalty $n \sum_{i \in \boldsymbol{S}} p_i + n$ is necessary for the pivotal property (9) to hold . The q_i are Hájek's (1964) finite population corrections .

Under without-replacement stratified sampling design and Hájek's (1964) asymptotic framework, Berger and Torres (2016) showed that under $H_0 : \boldsymbol{\theta}_{\mathcal{U}} = \boldsymbol{\theta}_0$, we have

$$\widetilde{r}(\boldsymbol{\theta}, \boldsymbol{\varphi}_{\mathcal{U}}) := 2 \Big\{ \widetilde{\ell}_{\max}(\boldsymbol{\varphi}_{\mathcal{U}}) - \widetilde{\ell}_{\max}(\boldsymbol{\theta}, \boldsymbol{\varphi}_{\mathcal{U}}) \Big\} \xrightarrow{d} \chi^2_{d_{\boldsymbol{\theta}}}, \quad \text{if } \boldsymbol{\theta} = \boldsymbol{\theta}_0, \tag{9}$$

where

$$\begin{aligned}\widetilde{\ell}_{\max}(\boldsymbol{\varphi}_{\mathcal{U}}) \;\; := \;\; \max_{\boldsymbol{p}} \Big\{ \widetilde{\ell}(\boldsymbol{p}) \;:\; p_i > 0,\; \sum_{i \in \boldsymbol{S}} (p_i q_i - \psi_i) \frac{1}{\pi_i} \mathbf{f}(\boldsymbol{y}_i, \boldsymbol{\varphi}_{\mathcal{U}}) = \mathbf{0}_{d_{\mathbf{f}}}, \\ \sum_{i \in \boldsymbol{S}} (p_i q_i - \psi_i) \boldsymbol{z}_i = \frac{\boldsymbol{n}_H}{n} \Big\}. \end{aligned}$$

In Berger and Torres (2016), the penalised empirical likelihood function uses the parametrization $m_i := n p_i \pi_i^{-1}$. Straightforward algebra shows that it indeed reduces to (8). Note that the constraints within (8) do not imply $\sum_{i \in \boldsymbol{S}} p_i = 1$. We have $\widetilde{\ell}_{\max}(\boldsymbol{\theta}, \boldsymbol{\varphi}_{\mathcal{U}}) = \ell_{\max}(\boldsymbol{\theta}, \boldsymbol{\varphi}_{\mathcal{U}})$ if we replace q_i by 1, because in this case, $\psi_i = 0$ and $\sum_{i \in \boldsymbol{S}} p_i = 1$. In fact, $q_i \to 1$, as $n/N \to 0$. Thus, with n/N negligible, $\widetilde{\ell}_{\max}(\boldsymbol{\theta}, \boldsymbol{\varphi}_{\mathcal{U}}) \approx \ell_{\max}(\boldsymbol{\theta}, \boldsymbol{\varphi}_{\mathcal{U}})$ and $\widetilde{r}(\boldsymbol{\theta}, \boldsymbol{\varphi}_{\mathcal{U}}) \approx \widehat{r}(\boldsymbol{\theta}^{(1)}, \boldsymbol{\varphi}_{\mathcal{U}})$. Berger (2016) extended this § 's approach to Rao et al.'s (1962) sampling design with large sampling fraction.

4 Pseudoempirical likelihood

Wu and Rao's (2006) "*pseudoempirical log-likelihood function*" (see also Chen and Sitter, 1999) is defined by

$$L_{\max}(\boldsymbol{\theta}, \boldsymbol{\varphi}_{\mathcal{U}}) := \max_{\boldsymbol{p}} \Big\{ L(\boldsymbol{p}) \ : \ p_i > 0, \ \sum_{i \in \boldsymbol{S}} p_i \, \boldsymbol{g}(\boldsymbol{y}_i, \boldsymbol{\theta}, \boldsymbol{\varphi}_{\mathcal{U}}) = \mathbf{0}_{d_{\boldsymbol{g}}}, \ \sum_{i \in \boldsymbol{S}} p_i \boldsymbol{z}_i = \mathbf{1}_H \Big\}, \quad (10)$$

where

$$\begin{aligned} L(\boldsymbol{p}) &:= n \sum_{i \in \boldsymbol{S}} \frac{\phi_i}{\pi_i} \log(p_i), \\ \phi_i &:= \frac{1}{N} \sum_{h=1}^{H} \frac{N_h}{\widehat{N}_h} z_{ih}, \\ N_h &:= \sum_{i \in \mathcal{U}} z_{ih}, \\ \widehat{N}_h &:= \sum_{i \in \boldsymbol{S}} \frac{z_{jh}}{\pi_j}. \end{aligned}$$

The function (10) clearly differs from (4). In (10), $L(\boldsymbol{p})$ is different from Owen's (1988) empirical log-likelihood function $\ell(\boldsymbol{p})$ and $\sum_{i \in \boldsymbol{S}} p_i = H$. In (4), $\ell(\boldsymbol{p})$ is used and $\sum_{i \in \boldsymbol{S}} p_i = 1$. The main difference between (4) and (10) is the way in which the π_i are used. In (4), the π_i appears within the constraint. On the other hand, in (10), the π_i only appear within $L(\boldsymbol{p})$. The function $L(\boldsymbol{p})$ is adjusted to take the π_i and the stratification into account by using ϕ_i. The constraints involving the stratification variable $\boldsymbol{z}_i$ also differ.

Wu and Rao (2006) showed that the "*maximum pseudoempirical likelihood estimator*", which maximises $L_{\max}(\boldsymbol{\theta}, \boldsymbol{\varphi}_{\mathcal{U}})$, is also the solution to

$$\widehat{\boldsymbol{E}}(\boldsymbol{\theta}, \boldsymbol{\varphi}_{\mathcal{U}}) := \sum_{i \in \boldsymbol{S}} \widehat{w}_i(\boldsymbol{\varphi}_{\mathcal{U}}) \, \boldsymbol{e}(\boldsymbol{y}_i, \boldsymbol{\theta}, \boldsymbol{\varphi}_{\mathcal{U}}) = \mathbf{0}_{d_{\boldsymbol{g}}}, \quad (11)$$

where

$$\begin{aligned} \widehat{w}_i(\boldsymbol{\varphi}_{\mathcal{U}}) &:= n \widehat{\rho}_i(\boldsymbol{\varphi}_{\mathcal{U}}) \pi_i^{-1}, \\ \widehat{\rho}_i(\boldsymbol{\varphi}_{\mathcal{U}}) &:= n^{-1} \big\{ 1 + \boldsymbol{\lambda}(\boldsymbol{\varphi}_{\mathcal{U}})^{\top} \boldsymbol{u}_i(\boldsymbol{\varphi}_{\mathcal{U}}) \big\}^{-1} \phi_i, \\ \boldsymbol{u}_i(\boldsymbol{\varphi}_{\mathcal{U}}) &:= \big\{ \mathbf{f}(\boldsymbol{y}_i, \boldsymbol{\varphi}_{\mathcal{U}})^{\top}, \boldsymbol{z}_i^{(H-1)\top} - \boldsymbol{N}_{H-1} \, N^{-1} \big\}^{\top}, \\ \boldsymbol{z}_i^{(H-1)} &:= \big\{ z_{i1}, \ldots, z_{ih}, \ldots, z_{i(H-1)} \big\}^{\top}, \\ \boldsymbol{N}_{H-1} &:= \big(N_1, \ldots, N_h, \ldots, N_{H-1} \big)^{\top}. \end{aligned}$$

Here, $\boldsymbol{\lambda}(\boldsymbol{\varphi}_{\mathcal{U}})$ is a Lagrangian parameter which is such that

$$\sum_{i \in \boldsymbol{S}} \widehat{w}_i(\boldsymbol{\varphi}_{\mathcal{U}}) \, \boldsymbol{u}_i(\boldsymbol{\varphi}_{\mathcal{U}}) = \mathbf{0}_{d_{\mathbf{f}}+H-1}.$$

We see that the weights $\widehat{w}_i(\boldsymbol{\varphi}_{\mathcal{U}})$ differ from $\widehat{m}_i(\boldsymbol{\varphi}_{\mathcal{U}})$ given by (6), because $\widehat{\rho}_i(\boldsymbol{\varphi}_{\mathcal{U}}) \neq \widehat{p}_i(\boldsymbol{\varphi}_{\mathcal{U}})$ and $\boldsymbol{u}_i(\boldsymbol{\varphi}_{\mathcal{U}}) \neq \boldsymbol{c}_i(\boldsymbol{\varphi}_{\mathcal{U}})$. Hence maximum pseudoempirical likelihood estimates are different from maximum empirical likelihood estimates. However, we expect minor numerical differences between them. The main difference between the approaches of §§ 4 and 3 is in the pivotal property of the empirical likelihood ratio statistics.

Suppose that $d_{\boldsymbol{\theta}} = d_{\boldsymbol{e}} = 1$; that is, we have a scalar parameter $\theta_{\mathcal{U}}$. Let $\widehat{\theta}$ be the maximum pseudoempirical likelihood estimator. Wu and Rao (2006) showed that under $H_0 : \theta_{\mathcal{U}} = \theta_0$,

$$\widehat{r}(\theta, \boldsymbol{\varphi}_{\mathcal{U}})_{\text{PEL}} := 2\Big\{ L_{\max}(\widehat{\theta}, \boldsymbol{\varphi}_{\mathcal{U}}) - L_{\max}(\theta, \boldsymbol{\varphi}_{\mathcal{U}}) \Big\} \text{Deff}(\theta, \boldsymbol{\varphi}_{\mathcal{U}})^{-1} \xrightarrow{d} \chi^2_1, \quad \text{if } \theta = \theta_0,$$

where $\text{Deff}(\theta, \boldsymbol{\varphi}_{\mathcal{U}})$ is called the "*design effect*" given by

$$\text{Deff}(\theta, \boldsymbol{\varphi}_{\mathcal{U}}) := \frac{\text{Var}\{\widehat{E}(\theta, \boldsymbol{\varphi}_{\mathcal{U}})\}}{\text{Var}_{\text{SRS}}\{\widehat{E}(\theta, \boldsymbol{\varphi}_{\mathcal{U}})\}}. \tag{12}$$

Here, $\widehat{E}(\theta, \boldsymbol{\varphi}_{\mathcal{U}})$ is defined by (11), when $d_{\boldsymbol{e}} = 1$. The quantity $\text{Var}\{\widehat{E}(\theta, \boldsymbol{\varphi}_{\mathcal{U}})\}$ denotes the variance under the sampling design and $\text{Var}_{\text{SRS}}\{\widehat{E}(\theta, \boldsymbol{\varphi}_{\mathcal{U}})\}$ is the variance under simple random sampling ; that is, a sampling design with equal probability. The design effect would need to be estimated. We refer to Wu and Rao (2006) for more details about the estimation of (12).

The advantage of pseudoempirical likelihood is that it can be applied in principle to any complex sampling designs, because the design effect capture the complexity of any designs. The properties (7) and (9) are limited to single stage design or multi-stage design with small sampling fractions (Berger and Torres, 2016; Oğuz-Alper and Berger, 2016). The property (9) is valid under Hájek (1964) framework involving maximum entropy. Note that most designs used in practice have large entropy (Berger, 2011).

The disadvantage of pseudoempirical likelihood is that (12) only holds when the parameter is scalar ($d_{\boldsymbol{\theta}} = d_{\boldsymbol{e}} = 1$). For example, it cannot be used with multidimensional parameters, because the design effect implicitly assumes that (11) is scalar. Thus, for example, we do not a pivotal statistics for regression parameters. The properties (7) and (9) holds for multidimensional parameters. Thus, the approach of §3 can be used for generalised linear models. In (12), the design effect needs to be estimated, which adds some additional variability that may compromise to the convergence of $\widehat{r}(\theta, \boldsymbol{\varphi}_{\mathcal{U}})_{\text{PEL}}$ towards the χ^2 distribution . Berger and Torres (2016) showed via a series of simulation that coverages of confidence intervals obtained from (7) and (9) are closer to the nominal value, than coverages obtained from (12). From a computational point of view (12) has the disadvantage of relying on variance estimates, which can be tedious to compute under complex sampling. Furthermore, (12) shows that the pseudoempirical likelihood ratio statistics is not pivotal, because of presence of a design effect.

References

Berger, Y. G. (2011), "Asymptotic consistency under large entropy sampling designs with unequal probabilities," *Pakistan Journal of Statistics*, 27(4), 407–426.

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

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 Van Keilegom, I. (2009), "A review on empirical likelihood methods for regression," *Test*, 18, 415–447.

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.

Godambe, V. P. (1960), "An Optimum Property of Regular Maximum Likelihood Estimation," *The Annals of Mathematical Statistics*, 31(4), pp. 1208–1211.

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.

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.

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.

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.

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.

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