Handbook of Statistics, Volume 29

Sample Surveys: Theory, Methods and Inference

Empirical Likelihood Methods

J.N.K. Rao and Changbao Wu

(February 14, 2008, Final Version)

1 Likelihood-based Approaches

Let $U = \{1, 2, \cdots, N\}$ be the set of units in the finite population and y_i and $\boldsymbol{x}_i$ be respectively the values of the study variable y and the vector of the auxiliary variables $\boldsymbol{x}$ attached to the ith unit. In this chapter we restrict our discussion to the estimation of the population total $Y = \sum_{i=1}^N y_i$, the population mean $\overline{Y} = Y/N$ or the population distribution function $F_N(t) = N^{-1}\sum_{i=1}^N I(y_i \le t)$ using a survey sample $\{(y_i, \boldsymbol{x}_i), i \in s\}$, where s is the set of sample units selected by the probability sampling design, $p(s)$, and $I(y \le t)$ is the indicator function. The population totals $\boldsymbol{X} = \sum_{i=1}^N \boldsymbol{x}_i$ or means $\overline{\boldsymbol{X}} = \boldsymbol{X}/N$ may also be available and can be used at the estimation stage.

Likelihood-based estimation methods in survey sampling do not follow as special cases from classical parametric likelihood inferences. Under the conventional design-based framework, values of the study variable for the finite population, $\{y_1, y_2, \cdots, y_N\}$, are viewed as fixed. The only randomization is induced by the probability sampling selection of units. In the design-based setup, an unbiased minimum variance estimator or even an unbiased minimum variance linear estimator of Y does not exist (Godambe, 1955, Godambe and Joshi, 1965). If we consider a class of linear estimators of Y in the form of $\sum_{i\in s} c_i y_i$ where the weight c_i depends only on i, then the unique unbiased estimator in the class is the well-known Horvitz-Thompson (HT) estimator $\hat{Y}_{HT} = \sum_{i\in s} y_i/\pi_i$, where $\pi_i = P(i \in s)$ is the first order inclusion probability for unit i. The HT estimator therefore is often treated as a baseline estimator for inferences concerning Y.

One of the early attempts in formulating a likelihood-based approach was the flat likelihood function (Godambe, 1966). The population vector of parameters is specified as $\tilde{\boldsymbol{y}} = (\tilde{y}_1, \tilde{y}_2, \cdots, \tilde{y}_N)'$ where the tilde indicates that each y_i is treated as an unknown parameter. The likelihood function of $\tilde{\boldsymbol{y}}$ is the probability of observing the sample data $\{y_i, i \in s\}$ for the given $\tilde{\boldsymbol{y}}$. For a given sampling design we can write down the likelihood function as

$$L(\tilde{\boldsymbol{y}}) = P(y_i, i \in s | \tilde{\boldsymbol{y}}) = \begin{cases} p(s) & \text{if } \ y_i = \tilde{y}_i \ \text{ for } \ i \in s\,, \\ 0 & \text{otherwise}\,, \end{cases} \tag{1}$$

where $p(s)$ denotes the probability of selecting the sample s under the design. Although the likelihood function $L(\tilde{\boldsymbol{y}})$ is well defined, it is uninformative in the sense that all possible non-observed values $y_i, i \notin s$ lead to the same likelihood. This difficulty arises because of the distinct labels i associated with the units in the sample data that make the sample unique.

To circumvent the difficulty associated with Godambe's flat likelihood, one possible resolution is to take a Bayesian route (Ericson, 1969). Given a joint N-dimensional prior on $\tilde{\boldsymbol{y}}$ with probability density function $g(\tilde{\boldsymbol{y}})$ and assume that the sampling design is independent of $\tilde{\boldsymbol{y}}$, the posterior density is given by

$$h(\tilde{\boldsymbol{y}} | y_i, i \in s) = \begin{cases} g(\tilde{\boldsymbol{y}})/g(\tilde{\boldsymbol{y}}_s) & \text{if } \ y_i = \tilde{y}_i \ \text{ for } \ i \in s\,, \\ 0 & \text{otherwise}\,, \end{cases}$$

where $\tilde{\boldsymbol{y}}_s = \{\tilde{y}_i, i \in s\}$ and $g(\tilde{\boldsymbol{y}}_s)$ is the marginal prior density of $\tilde{\boldsymbol{y}}_s$. Any informative prior will lead to an informative posterior distribution of $\tilde{\boldsymbol{y}}$ given the sample data. A popular choice of $g(\cdot)$ is the so-called exchangeable prior which basically states that the labels i carry no information regarding the associated y_i and the finite population is effectively randomized. However, in addition to the difficulty of choosing a prior, inferences under the Bayesian formulation are independent of the sampling design, an undesirable feature under the design-based framework.

Hartley and Rao (1968) took a different route in searching for a likelihood-based approach. In their proposed *scale-load* approach, some aspects of the sample data are ignored to make the sample non-unique and in turn the likelihood informative. The basic feature of the Hartley-Rao approach is to assume that the variable y is measured on a scale with a finite set of known scale points y_t^*, $t = 1, 2, \cdots, T$. The number of scale points, T, is only conceptual and inferences do not require the specification of T.

Let N_t be the number of units in U having the value y_t^*. It follows that $N = \sum_{t=1}^T N_t$ and $Y = \sum_{t=1}^T N_t y_t^*$ which is completely specified by the population "scale-loads" $\boldsymbol{N} = (N_1, N_2, \cdots, N_T)'$. Let n be the total sample size and n_t be the number of units in the sample having the value y_t^*. The sample data is effectively reduced to the observed scale-loads $\boldsymbol{n} = (n_1, n_2, \cdots, n_T)'$, with $n_t \geq 0$ and $n = \sum_{t=1}^T n_t$. Under simple random sampling without replacement, the likelihood based on the reduced sample data is given by the multi-hypergeometric distribution that depends on the population parameter $\boldsymbol{N}$, unlike the flat likelihood of (1) based on the full sample data. If the sampling fraction is negligible, the likelihood may be approximated by using the multinomial distribution with the log likelihood given by $l(\boldsymbol{p}) = \sum_{t=1}^T n_t \log(p_t)$, where $\boldsymbol{p} = (p_1, \cdots, p_T)'$ and $p_t = N_t/N$. Without using any auxiliary information, the maximum likelihood estimator of $\overline{Y} = \sum_{t=1}^T p_t y_t^*$ is the sample mean $\bar{y} = \sum_{t=1}^T \hat{p}_t y_t^*$, where $\hat{p}_t = n_t/n$.

The scale-load approach also provides an effective method for using known population mean $\overline{X}$ of an auxiliary variable x in estimating $\overline{Y}$. Denoting the scale points of x as x_j^*, $j = 1, \cdots, J$, and the scale load of (y_t^*, x_j^*) as N_{tj}, we have $\overline{Y} = \sum_{t=1}^T \sum_{j=1}^J p_{tj} y_t^*$ and $\overline{X} = \sum_{t=1}^T \sum_{j=1}^J p_{tj} x_j^*$, where $p_{tj} = N_{tj}/N$ and $\sum_{t=1}^T \sum_{j=1}^J p_{tj} = 1$. The sample data reduces to the observed frequencies n_{tj} for the scale points (y_t^*, x_j^*) such that $\sum_{t=1}^T \sum_{j=1}^J n_{tj} = n$. The scale-load estimator of $\overline{Y}$ is computed as $\hat{\overline{Y}} = \sum_{t=1}^T \sum_{j=1}^J \hat{p}_{tj} y_{tj}^*$, where $\hat{p}_{tj}$ maximize the log likelihood $\sum_{t=1}^T \sum_{j=1}^J n_{tj} \log(p_{tj})$ subject to constraints $\sum_{t=1}^T \sum_{j=1}^J p_{tj} = 1$ and $\sum_{t=1}^T \sum_{j=1}^J p_{tj} x_{tj}^* = \overline{X}$. Hartley and Rao (1968) showed that $\hat{\overline{Y}}$ is asymptotically equivalent to the customary regression estimator of $\overline{Y}$. This result was later "rediscovered" when Chen and Qin (1993) applied Owen's 1988 formulation of the empirical likelihood method to the same settings. Section 2 provides more details.

Hartley and Rao (1969) generalized the *scale-load* approach to unequal probability sampling with replacement where selection probability is proportional to size (PPS). If y_i is approximately proportional to the size x_i, then it is reasonable to consider the scale points of $r_i = y_i/x_i$, say r_t^*, and the resulting scale-load estimator of Y is equal to the customary unbiased estimator in PPS sampling with replacement. Extending the scale-load approach to unequal probability sampling without replacement does not seem to be straightforward, and confidence intervals based on the likelihood ratio function were not studied under the scale-load approach.

2 Empirical Likelihood Method Under Simple Random Sampling

The *scale-load* approach of Hartley and Rao (1968, 1969) based on a multinomial distribution has the same spirit as *empirical likelihood* proposed later by Owen (1988). Let $y_1, y_2, \cdots, y_n$ be an independent and identically distributed (*iid*) random sample from y with cumulative distribution function $F(\cdot)$. Let $p_i = P(y = y_i) = F(y_i) - F(y_i-)$ be the probability mass assigned to y_i. The empirical likelihood (EL) function defined by Owen (1988) is $L(\boldsymbol{p}) = \prod_{i=1}^n p_i$. Maximizing $l(\boldsymbol{p}) = \log\{L(\boldsymbol{p})\} = \sum_{i\in s} \log(p_i)$ subject to $p_i > 0$ and $\sum_{i=1}^n p_i = 1$ leads to $\hat{p}_i = 1/n$, and the maximum empirical likelihood (MEL) estimator of $F(u)$ is given by $\hat{F}(u) = \sum_{i=1}^n \hat{p}_i I(y_i \leq u) = F_n(u)$, where $I(\cdot)$ is the indicator function and $F_n(u) = n^{-1}\sum_{i=1}^n I(y_i \leq u)$ is the empirical distribution function based on the *iid* sample.

There have been many important contributions to the development of the EL method in mainstream statistics since Owen's 1988 paper on the asymptotic χ^2 distribution of the empirical likelihood ratio statistic for the mean $\mu = E(y)$. This is evident from Owen's (2001) monograph on empirical likelihood. Among other results, the work by Qin and Lawless (1994), which showed that side information in the form of a set of estimating equations can be used to improve the maximum EL estimators and the EL ratio confidence intervals, is particularly appealing for inference from survey data in the presence of auxiliary information.

The first formal use of the EL method in survey sampling was presented by Chen and Qin (1993) under simple random sampling with or without replacement. The sampling fraction is assumed to be negligible in the case of without replacement sampling, so that Owen's EL function for *iid* cases can be directly used. Let $\{(y_i, x_i), i \in s\}$ be the sample data, and $\theta_0 = N^{-1}\sum_{i=1}^N g(y_i)$ be the population parameter with $g(\cdot)$ being a known function. The known population auxiliary information is in the form of $E\{w(x)\} = 0$ for some known $w(\cdot)$. The log-likelihood function is given by $l(\boldsymbol{p}) = \log\{L(\boldsymbol{p})\} = \sum_{i\in s}\log(p_i)$. The maximum empirical likelihood (MEL) estimator of θ is defined as $\hat{\theta} = \sum_{i\in s}\hat{p}_i g(y_i)$, where $\hat{p}_i$ maximize $l(\boldsymbol{p})$ subject to $p_i > 0$, $\sum_{i\in s} p_i = 1$ and $\sum_{i\in s} p_i w(x_i) = 0$. The MEL estimator has no closed form expression. It can be shown that the solution is given by $\hat{p}_i = \{n(1+\lambda w(x_i)\}^{-1}$ where the Lagrange multiplier λ is the solution to $\sum_{i\in s} w(x_i)/\{1+\lambda w(x_i)\} = 0$. We give further computational details in Section 5. For now, we note that the choice of $g(y_i) = y_i$ gives $\theta_0 = \overline{Y}$ and incorporating the known population mean $\overline{X}$ translates into $w(x_i) = x_i - \overline{X}$ and $\sum_{i\in s} p_i x_i = \overline{X}$. The MEL estimator is uniquely defined when $\overline{X}$ is an inner point of the convex hull formed by $\{x_i, i \in s\}$. This happens with probability approaching one as $n \to \infty$. For cases where x is univariate, the convex hull becomes $(x_{(1)}, x_{(n)})$, where $x_{(1)} = \min_{i\in s} x_i$ and $x_{(n)} = \max_{i\in s} x_i$. Let $F_x(t)$ be the distribution function of x and assume simple random sampling with replacement, then $P\{x_{(1)} < \overline{X} < x_{(n)}\} = 1 - \{1 - F_x(\overline{X}-)\}^n - \{F_x(\overline{X})\}^n$ which goes to one at an exponential rate. The MEL estimator of $\overline{Y}$ is equivalent to the *scale-load* estimator of Hartley and Rao (1968).

By letting $g(y_i) = I(y_i \le t)$ for a fixed t, we get the MEL estimator of the population distribution function $F_N(t) = N^{-1}\sum_{i=1}^N I(y_i \le t)$ as $\hat{F}_N(t) = \sum_{i\in s}\hat{p}_i I(y_i \le t)$. The estimator $\hat{F}_N(t)$ is a genuine distribution function, that is, it is monotone non-decreasing and is confined within the range [0, 1]. Consequently, MEL estimators of population quantiles can be obtained through direct inversion of $\hat{F}_N(t)$.

The EL approach provides non-parametric confidence intervals through the profiling of empirical likelihood ratio statistics, similar to the parametric case. For $\theta_0 = \overline{Y}$ and in the absence of any auxiliary information, the log empirical likelihood ratio function is given by $r(\theta) = -2\sum_{i\in s}\log(n\hat{p}_i)$, where $\hat{p}_i$ maximize the EL function $l(\boldsymbol{p})$ subject to constraints $p_i > 0$, $\sum_{i\in s} p_i = 1$ and $\sum_{i\in s} p_i y_i = \theta$ for a fixed value of θ. It can be shown that under some moment conditions on y for the finite population and a suitable asymptotic framework which allows n and N simultaneously go to infinity while n/N goes to zero, the EL ratio function $r(\theta)$ converges in distribution to a χ^2 random variable with *one* degree of freedom when $\theta = \theta_0$. A $1 - \alpha$ level EL confidence interval for $\theta_0 = \overline{Y}$ is then given by $\mathcal{C}_{el} = \{ \theta \mid r(\theta) \leq \chi_1^2(\alpha) \}$, where $\chi_1^2(\alpha)$ is the upper α-quantile of the χ^2 distribution with one degree of freedom. Finding such an interval involves profiling. Section 5 again contains the computational detail. Unlike the symmetric interval based on the normal approximation to the Z-statistic $(\hat{\theta}_0 - \theta_0)/\{var(\hat{\theta}_0)\}^{1/2}$, the orientation of the EL interval $\mathcal{C}_{el}$ is determined by the data and the range of the parameter space is fully preserved.

Section 4.3 contains results from a limited simulation study on the EL interval for $\theta_0 = F_N(t)$. The results demonstrate several advantages of the EL intervals. One of them is that the upper and lower bounds of the EL intervals are always within the range of [0, 1], which is not the case for the conventional normal theory intervals.

Table 1: 95% Confidence Intervals for $\overline{Y}$ under PPS Sampling

ρ	Zeros (%)	CI	CP	L	U	LB
0.30	95%	NA	86.0	0.3	13.7	−0.01
		EL	91.0	2.4	6.6	0.09
	90%	NA	87.3	0.5	12.2	0.13
		EL	91.4	2.5	6.1	0.26
	80%	NA	92.0	1.0	7.0	0.58
		EL	93.6	2.5	3.9	0.72
	70%	NA	93.0	1.6	5.4	1.11
		EL	94.5	2.7	2.8	1.23
0.80	95%	NA	84.1	0.4	15.5	0.00
		EL	92.8	3.1	4.1	0.10
	90%	NA	89.6	1.3	9.1	0.17
		EL	92.1	3.3	4.6	0.27
	80%	NA	92.9	1.6	5.5	0.65
		EL	94.3	2.6	3.1	0.73
	70%	NA	93.8	2.1	4.1	1.18
		EL	94.9	2.8	2.3	1.24

Under simple random sampling, Chen, Chen and Rao (2003) compared EL intervals with several alternatives for the population mean of populations containing many zero values. Such populations are encountered, for instance, in audit sampling, where the response variable y denotes the amount of money owned to the government and the population mean $\overline{Y}$ is the average amount of excessive claims. Most of the claims are legitimate, with corresponding y being zeros, but a small portion of claims may be excessive. The lower bound of the 95% confidence interval on $\overline{Y}$ is often used to compute the total amount of money owned to the government. Since the total number of claims selected for auditing is usually not large, the normal approximation confidence intervals perform poorly in terms of the lower tail error rate and the average lower bound. Parametric likelihood ratio intervals based on parametric mixture distributions for the y variable have been used in auditing but the performance of such intervals depends heavily on the validity of the assumed parametric model. The EL intervals exhibit behaviour similar to intervals based on a correctly specified mixture model. More importantly, they perform better than the intervals based on incorrectly specified mixture models. That is to say, the EL intervals provide lower error rates at least as close to the nominal values while the intervals based on incorrectly specified mixture models lead to lower error rates much smaller that the nominal values. As a result, EL intervals have larger lower bounds, and methods that respect nominal error rates and at the same time provide larger lower bounds are regarded as desirable ones.

Under unequal probability sampling, the use of any parametric mixture model for such populations becomes difficult to justify. The EL intervals, however, are still available using the pseudo EL formulation described in Section 4.3. Table 1 reports the performance of the pseudo EL confidence intervals for the population mean from a simulation study. The finite population is first generated through Model I used by Wu and Rao (2006), with the correlation coefficient between the design variable z_i and the response variable y_i indicated by ρ, and a random portion of the y_is is then set to be zeros. Information on the design variable z is not further used in constructing the EL confidence intervals. The interval based on normal approximation (NA) is included for comparison. The reported results on coverage probability (CP), lower (L) and upper (U) tail error rates, and the average lower bound (LB) are based on 1000 simulated samples of size $n = 60$, selected by the PPS sampling method of Rao (1965) and Sampford (1967). While intervals based on normal approximations are clearly inappropriate, the EL interval maintains the same desirable performance observed under simple random sampling, with the lower tail error rates close to the nominal value and larger lower bound for all cases considered.

3 Stratified Simple Random Sampling

Stratified simple random sampling is commonly used when list frames within strata are available, as in many business surveys. Let $\{(y_{hi}, \boldsymbol{x}_{hi}), i \in s_h, h = 1, \cdots, L\}$ be a stratified simple random sample, where y_{hi} and $\boldsymbol{x}_{hi}$ are respectively the values of the study variable y and the vector of auxiliary variables $\boldsymbol{x}$ associated with the ith element in stratum h, L is the total number of strata in the population and s_h is the set of n_h sampled units from stratum h. Let N_h be the stratum size, $W_h = N_h/N$ be the stratum weight, and $N = \sum_{h=1}^{L} N_h$ is the overall population size. Zhong and Rao (1996, 2000) studied EL inferences on $\overline{Y}$ when the vector-valued population mean $\overline{\boldsymbol{X}}$ is known but the stratum means $\overline{\boldsymbol{X}}_h$ are unknown. Assuming negligible sampling fractions within strata and noting that samples from different strata are independent, the log empirical likelihood function under stratified simple random sampling is given by $l(\boldsymbol{p}_1, \cdots, \boldsymbol{p}_L) = \sum_{h=1}^{L} \sum_{i \in s_h} \log(p_{hi})$, where $\boldsymbol{p}_h = (p_{h1}, \cdots, p_{hn_h})'$ and p_{hi} is the probability mass assigned to y_{hi}, $i \in s_h$, $h = 1, \cdots, L$. The maximum empirical likelihood (MEL) estimator of $\overline{Y}$ is defined as $\hat{\overline{Y}} = \sum_{h=1}^{L} W_h \sum_{i \in s_h} \hat{p}_{hi} y_{hi}$ where the $\hat{p}_{hi}$ maximize $l(\boldsymbol{p}_1, \cdots, \boldsymbol{p}_L)$ subject to $p_{hi} > 0$, $\sum_{i \in s_h} p_{hi} = 1$, $h = 1, \cdots, L$ and $\sum_{h=1}^{L} W_h \sum_{i \in s_h} p_{hi} \boldsymbol{x}_{hi} = \overline{\boldsymbol{X}}$. For fixed L and large sample sizes n_h with negligible sampling fraction within strata, the MEL estimator is asymptotically equivalent to the randomization optimal linear regression estimator (Zhong and Rao, 2000; Kott, Chapter 25). While most design-based estimators can only be justified under unconditional repeated sampling, the optimal estimator leads to valid conditional inferences, with negligible conditional relative bias given the stratified mean $\bar{\boldsymbol{x}}_{st} = \sum_{h=1}^{L} W_h \bar{\boldsymbol{x}}_h$ where $\bar{\boldsymbol{x}}_h = n_h^{-1} \sum_{i \in s_h} \boldsymbol{x}_{hi}$ (Rao, 1994). Zhong and Rao (2000) also studied the EL ratio confidence intervals on $\overline{Y}$. An efficient computational algorithm was proposed by Wu (2004b) with simple R/SPLUS functions and codes available in Wu (2005). We will describe some of the details in Section 4.3 and Section 5. Note that the constrained maximization problem here is more involved than in the case of simple random sampling (Section 2), because it is not possible to impose separate constraints of the form $\sum_{i\in s_h} p_{hi}\boldsymbol{x}_{hi} = \overline{\boldsymbol{X}}_h$, $h = 1, \cdots, L$ when the strata means $\overline{\boldsymbol{X}}_h$ are unknown. The case of deep stratification (large L and small n_h within each stratum h) is not covered by Zhong and Rao (2000), and it had not been studied in the EL literature.

4 Pseudo Empirical Likelihood Method

One of the major difficulties for the empirical likelihood inferences under general unequal probability sampling designs is to obtain an informative empirical likelihood function for the given sample. The likelihood depends necessarily on the sampling design and a complete specification of the joint probability function of the sample is usually not feasible under any without replacement sampling. Because of this difficulty, Chen and Sitter (1999) proposed a *pseudo empirical likelihood* approach using a two-stage argument. Suppose the finite population $\{y_1, \cdots, y_N\}$ can be viewed as an *iid* sample from a superpopulation, then the population (or "census") log empirical likelihood would be $l_N(\boldsymbol{p}) = \sum_{i=1}^N \log(p_i)$, the population total of the $\log(p_i)$. For a given sample, the design-based Horvitz-Thompson estimator of $l_N(\boldsymbol{p})$ is given by $l_{HT}(\boldsymbol{p}) = \sum_{i\in s} d_i \log(p_i)$, where $d_i = 1/\pi_i$, and $\pi_i = P(i \in s)$ are the first order inclusion probabilities. Chen and Sitter termed $l_{HT}(\boldsymbol{p})$ the pseudo log empirical likelihood. Under simple random sampling and ignoring a multiplying constant $l_{HT}(\boldsymbol{p})$ reduces to $l(\boldsymbol{p})$ used by Chen and Qin (1993).

The pseudo EL function $l_{HT}(\boldsymbol{p})$ involves only the first order inclusion probabilities and does not catch the design effects under general unequal probability sampling without replacement. Wu and Rao (2006) defined the pseudo empirical log-likelihood (PELL) function under non-stratified (ns) sampling designs as

$$l_{ns}(\boldsymbol{p}) = n \sum_{i\in s} \tilde{d}_i(s) \log(p_i)\,, \tag{2}$$

where $\tilde{d}_i(s) = d_i / \sum_{i\in s} d_i$. The pseudo EL function (2) has likelihood-based motivation but is directly related to the "backward" Kullback-Leibler distance (DiCiccio and Romano, 1990) between $\boldsymbol{p} = (p_1, \cdots, p_n)'$ and $\boldsymbol{d}(s) = (\tilde{d}_1(s), \cdots, \tilde{d}_n(s))'$ in the form of

$D(\boldsymbol{d}(s), \boldsymbol{p}) = \sum_{i\in s} \tilde{d}_i(s)\log(\tilde{d}_i(s)/p_i)$. Since $D(\boldsymbol{d}(s), \boldsymbol{p}) = \sum_{i\in s} \tilde{d}_i(s)\log(\tilde{d}_i(s)) - l_{ns}(\boldsymbol{p})/n$, minimizing the Kullback-Leibler distance with respect to p_i subject to a set of constraints is equivalent to maximizing the pseudo EL function subject to the same set of constraints.

The PELL function given by (2) differs from $l_{HT}(\boldsymbol{p})$ used by Chen and Sitter (1999) in the sense that the normalized weights $\tilde{d}_i(s)$, also called the Hajek weights (Hajek, 1971), are used instead of d_i. But maximizing (2) subject to a set of constraints on the p_i is equivalent to maximizing $l_{HT}(\boldsymbol{p})$ subject to the same set of constraints, and the resulting MPEL estimators remain the same. However, it is shown in Section 4.3 that $l_{ns}(\boldsymbol{p})$ allows for simple adjustment for the design effect in constructing pseudo EL ratio confidence intervals.

For stratified (st) sampling with an arbitrary sampling design within each stratum, the PELL function of Wu and Rao (2006) is defined as

$$l_{st}(\boldsymbol{p}_1, \cdots, \boldsymbol{p}_L) = n\sum_{h=1}^{L} W_h \sum_{i\in s_h} \tilde{d}_{hi}(s_h)\log(p_{hi}), \tag{3}$$

where $\tilde{d}_{hi}(s_h) = d_{hi}/\sum_{i\in s_h} d_{hi}$ are the normalized weights within each stratum with $d_{hi} = \pi_{hi}^{-1}$ denoting the design weights, $\pi_{hi} = P(i \in s_h)$ the h-th stratum unit inclusion probabilities, $n = \sum_{h=1}^{L} n_h$, and n_h the stratum sample sizes. Note that $l_{st}(\boldsymbol{p}_1, \cdots, \boldsymbol{p}_L)$ does *not* reduce to the empirical log-likelihood function $\sum_{h=1}^{L}\sum_{i\in s_h}\log(p_{hi})$ under stratified simple random sampling (Zhong and Rao, 2000) unless $n_h = nW_h$, that is, the stratum sample sizes are proportionally allocated.

In the absence of auxiliary information, maximizing $l_{ns}(\boldsymbol{p})$ subject to $p_i > 0$ and $\sum_{i\in s} p_i = 1$ gives $\hat{p}_i = \tilde{d}_i(s)$. The resulting maximum pseudo empirical likelihood (MPEL) estimator of $\overline{Y}$, defined as $\hat{\overline{Y}}_{EL} = \sum_{i\in s}\hat{p}_i y_i$, is given by the Hajek estimator $\hat{\overline{Y}}_H = \sum_{i\in s}\tilde{d}_i(s) y_i$, and the MPEL estimator of the distribution function $F_N(t)$ is given by $\hat{F}_H(t) = \sum_{i\in s} d_i I(y_i \le t)/\sum_{i\in s} d_i$.

The Hajek estimator of $\overline{Y}$, however, can be less efficient than the Horvitz-Thompson estimator $\hat{\overline{Y}}_{HT} = N^{-1}\sum_{i\in s} d_i y_i$ under PPS sampling without replacement when the response variable y is highly correlated with the size variable. For designs with fixed sample size, Wu and Rao (2006) suggested a more efficient estimator by imposing the constraint

$$\sum_{i \in s} p_i \pi_i = \frac{n}{N}\,. \tag{4}$$

Equation (4) is the same as $\sum_{i\in s} p_i z_i = \overline{Z}$ where z_i is the size variable and $\overline{Z}$ is the population mean. The resulting MPEL estimator, which is a special case of the estimators discussed in Section 4.1, is equivalent to a regression type estimator with variance depending on the residuals. The Hajek estimator of $F_N(t)$ at a fixed t, on the other hand, is very efficient since the indicator variable $I(y_i \le t)$ is weakly correlated with the size variable, and $\hat{F}_H(t)$ itself is a genuine distribution function.

4.1 Pseudo empirical likelihood approach to calibration

Suppose the population mean $\overline{\boldsymbol{X}}$ of a vector of auxiliary variables $\boldsymbol{x}$ is known. In this case the MPEL estimator of $\overline{Y}$ is given by $\hat{\overline{Y}}_{EL} = \sum_{i\in s} \hat{p}_i y_i$, where $\hat{p}_i$ maximize $l_{ns}(\boldsymbol{p})$ subject to $p_i > 0$, $\sum_{i\in s} p_i = 1$ and

$$\sum_{i \in s} p_i \boldsymbol{x}_i = \overline{\boldsymbol{X}}\,. \tag{5}$$

Using the Lagrange multiplier method we can show that $\hat{p}_i = \tilde{d}_i(s)/(1 + \boldsymbol{\lambda}'\boldsymbol{u}_i)$, where $\boldsymbol{u}_i = \boldsymbol{x}_i - \overline{\boldsymbol{X}}$, and $\boldsymbol{\lambda}$ is the solution to

$$g(\boldsymbol{\lambda}) = \sum_{i \in s} \frac{\tilde{d}_i(s)\boldsymbol{u}_i}{1 + \boldsymbol{\lambda}'\boldsymbol{u}_i} = \boldsymbol{0}\,. \tag{6}$$

Constraints such as (5) are often referred to as benchmark constraints or calibration equations. A calibration estimator for $\overline{Y}$ can be defined as $\hat{\overline{Y}}_C = N^{-1}\sum_{i\in s} w_i y_i$, where the calibrated weights w_i minimize a distance measure $\Phi(\boldsymbol{w}, \boldsymbol{d})$ between $\boldsymbol{w} = (w_1, \cdots, w_n)'$ and the basic design weights $\boldsymbol{d} = (d_1, \cdots, d_n)'$ subject to $\sum_{i\in s} w_i \boldsymbol{x}_i = \boldsymbol{X}$. The simple chi-squared distance $\Phi(\boldsymbol{w}, \boldsymbol{d}) = \sum_{i\in s}(w_i - d_i)^2/(d_i q_i)$ with pre-specified q_i provides closed form solutions to w_i, leading to a generalized regression (GREG) estimator of Y (Särndal *et al.*, 1992), but the resulting w_i can take negative values under unbalanced sample configurations. Other distance measures that force the weights to be positive are available, but most of them suffer computational inefficiencies or other undesirable features.

There are several attractive features with the pseudo empirical likelihood approach to calibration estimation. First, the weights $\hat{p}_i$ are intrinsically positive and normalized, i.e. $\hat{p}_i > 0$ and $\sum_{i\in s}\hat{p}_i = 1$. This is particularly appealing for the MPEL estimator of $F_N(t)$ computed as $\hat{F}_{EL}(t) = \sum_{i\in s}\hat{p}_i I(y_i \le t)$. Like the MEL estimator in the *iid* setting, it is a genuine distribution function, and quantile estimates can be obtained through a direct inversion of $\hat{F}_{EL}(t)$. Second, for the major computational task of finding the Lagrange multiplier $\boldsymbol{\lambda}$ as the solution to (6), a modified Newton-Raphson algorithm (Chen *et al.*, 2002), which guarantees fast convergence, is available as we show in Section 5. Third, the pseudo EL ratio confidence intervals, described in Section 4.3, have several advantages over the conventional normal theory intervals.

If confidence intervals are not of major interest and the focus is on reporting standard errors, the approximate design-based variance and a variance estimator on $\hat{\overline{Y}}_{EL}$ are also readily available. Under regularity conditions C1-C3 described in Section 4.3, we have $\boldsymbol{\lambda} = \left(\sum_{i\in s} d_i \boldsymbol{u}_i \boldsymbol{u}_i'\right)^{-1} \sum_{i\in s} d_i \boldsymbol{u}_i + o_p\left(n^{-1/2}\right)$ and $\hat{p}_i \doteq \tilde{d}_i(s)(1-\boldsymbol{\lambda}'\boldsymbol{u}_i)$, which lead to

$$\hat{\overline{Y}}_{EL} = \hat{\overline{Y}}_H + \hat{\boldsymbol{B}}'\left(\overline{\boldsymbol{X}} - \hat{\overline{\boldsymbol{X}}}_H\right) + o_p\left(n^{-1/2}\right) , \tag{7}$$

where $\boldsymbol{u}_i = \boldsymbol{x}_i - \hat{\overline{\boldsymbol{X}}}_H$, $\hat{\boldsymbol{B}} = (\sum_{i\in s} d_i \boldsymbol{u}_i \boldsymbol{u}_i')^{-1}\sum_{i\in s} d_i \boldsymbol{u}_i y_i$, $\hat{\overline{\boldsymbol{X}}}_H = \sum_{i\in s}\tilde{d}_i(s)\boldsymbol{x}_i$, and $\hat{\overline{Y}}_H + \hat{\boldsymbol{B}}'\left(\overline{\boldsymbol{X}} - \hat{\overline{\boldsymbol{X}}}_H\right)$ is a GREG estimator of $\overline{Y}$. It follows from(7) that linearization variance estimation techniques for GREG estimators can be applied to $\hat{\overline{Y}}_{EL}$. Similarly, for $\hat{F}_{EL}(t)$ by changing y_i in (7) to $I(y_i \le t)$.

In practice, one might wish to restrict the range of the calibrated weights so that $c_1 \le w_i/d_i \le c_2$ for some pre-specified $0 < c_1 < 1 < c_2$. Under the pseudo EL approach this amounts to imposing

$$c_1 \le p_i/\tilde{d}_i(s) \le c_2 \, , \ i \in s \, . \tag{8}$$

Chen *et al.* (2002) suggested a simple computational procedure to achieve (8) through a minimal relaxation of the benchmark constraints (5). If the MPEL solutions $\hat{p}_i$ using (5) do not satisfy (8), we replace (5) by

$$\sum_{i\in s} p_i \boldsymbol{x}_i = \overline{\boldsymbol{X}} + \delta(\hat{\overline{\boldsymbol{X}}}_H - \overline{\boldsymbol{X}})$$

for some $\delta \in [0,1]$ in finding the MPEL solutions $\hat{p}_i$. The choice of $\delta = 0$ (i.e., no relaxation from (5)) corresponds to the the initial MPEL solution. At the other end, the value of $\delta = 1$ gives $\hat{p}_i = \tilde{d}_i(s)$ which always satisfy (8). For any pre-chosen $c_1 < 1 < c_2$, the smallest value of δ with the corresponding MPEL solutions satisfying (8) can be found through a simple bi-section search method (Chen *et al.*, 2002).

There are two major motivations behind any type of calibration estimation method, including the MPEL method: (i) Internal consistency, achieved through the benchmark constraints (5); and (ii) Efficiency, due to the asymptotic equivalence of $\hat{\overline{Y}}_{EL}$ to the GREG estimator, as shown from (7). But using (5) for the estimation of $\overline{Y}$ may not be very efficient when the underlying relationship between y and $\boldsymbol{x}$ does not approximate a linear model. Wu and Sitter (2001) proposed a model-calibrated pseudo EL method when the underlying superpopulation model, linear or nonlinear, is specified as $E_\xi(y_i|\boldsymbol{x}_i) = \mu(\boldsymbol{x}_i, \boldsymbol{\theta})$ and $V_\xi(y_i|\boldsymbol{x}_i) = v(\boldsymbol{x}_i)\sigma^2$, where ξ denotes the superpopulation model, $\boldsymbol{\theta}$ and σ^2 are model parameters, $\mu(\cdot,\cdot)$ and $v(\cdot)$ are known functions. The traditional constraints (5) are replaced by

$$\sum_{i\in s} p_i\hat{\mu}_i = N^{-1}\sum_{i=1}^{N}\hat{\mu}_i\,, \tag{9}$$

where the calibration variable $\hat{\mu}_i = \mu(\boldsymbol{x}_i, \hat{\boldsymbol{\theta}})$ is the predicted value of y_i based on the model, and $\hat{\boldsymbol{\theta}}$ is a design-based estimator of $\boldsymbol{\theta}$. The PELL function (2) is maximized subject to the calibration constraint (9), leading to the model-calibrated MPEL estimator of the mean $\overline{Y}$.

Note that when $\mu(\boldsymbol{x}_i, \boldsymbol{\theta})$ has a nonlinear form, constraint (9) requires that complete information on $\boldsymbol{x}$, i.e. $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N$, be known. Under the linear model $\mu(\boldsymbol{x}_i, \boldsymbol{\theta}) = \boldsymbol{x}_i'\boldsymbol{\theta}$, the "population mean" $N^{-1}\sum_{i=1}^N \hat{\mu}_i$ reduces to $\overline{\boldsymbol{X}}'\hat{\boldsymbol{\theta}}$, so only $\overline{\boldsymbol{X}}$ is needed in (9) in this case. The resulting model-calibrated MPEL estimator of $\overline{Y}$ using (9) is asymptotically equivalent to the MPEL estimator using (5), under the linear model.

The model-calibrated MPEL estimator of $\overline{Y}$ is asymptotically *optimal* in the class of MPEL estimators $\hat{\overline{Y}}_u = \sum_{i\in s} p_i y_i$ satisfying the calibration constraint $\sum_{i\in s} p_i u(\boldsymbol{x}_i) = N^{-1}\sum_{i=1}^N u(\boldsymbol{x}_i)$, where $u(\cdot)$ is an arbitrary function satisfying finite moment conditions and $\mu(\cdot, \boldsymbol{\theta})$ is assumed to belong to this class (Wu, 2003). That is, the choice $u(\boldsymbol{x}_i) = \mu(\boldsymbol{x}_i, \boldsymbol{\theta})$ minimizes the anticipated asymptotic variance $E_\xi AV(\hat{\overline{Y}}_u)$, where AV denotes the asymptotic design variance. One immediate application of this result is the optimal calibration estimator of the distribution function $F_N(t)$ at a fixed t. The optimal calibration variable that should be used in (9) is given by $E_\xi\{I(y_i \le t)\} = P(y_i \le t)$. Chen and Wu (2002) compared the efficiency of the optimal model-calibrated MPEL estimator of $F_N(t)$ to several alternative estimators and also discussed the related quantile estimation problem.

The model-calibrated MPEL estimation method can be extended to cover quadratic population parameters in the form of $T = \sum_{i=1}^N \sum_{j=i+1}^N \phi(y_i, y_j)$, which includes the population variance, the covariance, and the variance of a linear estimator as special cases (Sitter and Wu, 2002). The basic idea is to view T as a total over a synthetic finite population, i.e. $T = \sum_{\alpha=1}^{N^*} t_\alpha$ where $\alpha = (ij)$ is relabelled from 1 to $N^* = N(N-1)/2$ and $t_\alpha = \phi(y_i, y_j)$. The synthetic sample consists of all the pairs from the original sample and the "first order" inclusion probabilities under this setting are $\pi_{ij} = P(i, j \in s)$. The "basic design weights" are $d_{ij} = 1/\pi_{ij}$. The extended pseudo log empirical likelihood function for quadratic parameters is defined as

$$l^*(\boldsymbol{p}) = \sum_{i\in s}\sum_{j>i} d_{ij}\log(p_{ij})\,,$$

where p_{ij} is the probability mass assigned to the pair (i, j). The model-calibrated MPEL estimator of T is defined as

$$\hat{T}_{EL} = N^* \sum_{i\in s}\sum_{j>i} \hat{p}_{ij}\phi(y_i, y_j)\,,$$

where the $\hat{p}_{ij}$ maximize $l^*(\boldsymbol{p})$ subject to

$$\sum_{i\in s}\sum_{j>i} p_{ij} = 1 \quad \text{and} \quad \sum_{i\in s}\sum_{j>i} p_{ij}u_{ij} = \frac{1}{N^*}\sum_{i=1}^{N}\sum_{j=i+1}^{N} u_{ij}\,.$$

The optimal calibration variable u_{ij} is given by $u_{ij} = E_\xi\{\phi(y_i, y_j)\}$. For the population variance $S^2 = (N-1)^{-1}\sum_{i=1}^N (y_i - \overline{Y})^2$, which can be expressed as $\{N(N-1)\}^{-1}\sum_{i=1}^N\sum_{j=i+1}^N (y_i - y_j)^2$, the optimal calibration variable is given by $u_{ij} = E_\xi\{(y_i - y_j)^2\} = \{\mu(\boldsymbol{x}_i, \boldsymbol{\theta}) - \mu(\boldsymbol{x}_j, \boldsymbol{\theta})\}^2 + \{v(\boldsymbol{x}_i) + v(\boldsymbol{x}_j)\}\sigma^2$ under the assumed model. In applications the unknown model parameters $\boldsymbol{\theta}$ and σ^2 will have to be replaced by suitable design-based estimators. The resulting model-calibrated MPEL estimator of T remains consistent under mild conditions.

4.2 Pseudo empirical likelihood alternative to raking

Raking ratio estimation can be viewed as a special application of the calibration method where the auxiliary information is in the form of known marginal totals of a contingency table of two or more dimensions. Unfortunately, the number of benchmark constraints involved is often very large and the related computational procedures can be problematic. The pseudo EL alternative to raking offers a major advantage in computational efficiency and stability. Like raking, and unlike standard linear calibration approaches, the successful completion of pseudo EL will always produce positive weights. Moreover, convergence is guaranteed and fast. Confidence intervals can be constructed using either the pseudo EL ratio function to be described in Section 4.3 or the linearization variance estimator and normal approximation. Auxiliary population information other than the marginal totals and features of complex sampling designs can also be incorporated into the estimation procedure.

We first describe the EL alternative to raking under the classical setting of Deming and Stephan (1940). Suppose the finite population is cross-classified into $r \times c$ cells with a total number of N_{ij} units in the (i,j)-th cell, $i = 1, \cdots, r$, $j = 1, \cdots, c$. Let $N = \sum_{i=1}^{r} \sum_{j=1}^{c} N_{ij}$ be the total population size. The marginal totals $N_{i\cdot} = \sum_{j=1}^{c} N_{ij}$, $i = 1, \cdots, r$ and $N_{\cdot j} = \sum_{i=1}^{r} N_{ij}$, $j = 1, \cdots, c$ are known. The cell totals N_{ij} are unknown and need to be estimated. Let n be the size of a simple random sample drawn from the population and n_{ij} be the sample frequency for the (i,j)-th cell. Noting that $N_{i\cdot}$, $N_{\cdot j}$ and N are all known, we could estimate N_{ij} by $n_{ij}(N/n)$ or $n_{ij}(N_{i\cdot}/n_{i\cdot})$ or $n_{ij}(N_{\cdot j}/n_{\cdot j})$, where $n_{i\cdot} = \sum_{j=1}^{c} n_{ij}$ and $n_{\cdot j} = \sum_{i=1}^{r} n_{ij}$. But none of these estimators will necessarily match the known marginal totals in both dimensions (i.e., equal $N_{i\cdot}$ when summed across the rows and $N_{\cdot j}$ when summed across the columns).

The classical raking ratio estimator of N_{ij} in the form of $m_{ij}(N/n)$, obtained through the so-called iterative proportional fitting procedure (IPFP) (Deming and Stephan, 1940), was initially conceived to minimize the least square distance $\Phi = \sum_{i=1}^{r} \sum_{j=1}^{c} (m_{ij} -$

$n_{ij})^2/n_{ij}$ subject to the set of constraints

$$\sum_{j=1}^{c} m_{ij} = N_{i\cdot}n/N\,, \quad i = 1, \cdots, r\,, \tag{10}$$

$$\sum_{i=1}^{r} m_{ij} = N_{\cdot j}n/N\,, \quad j = 1, \cdots, c-1\,. \tag{11}$$

Although the m_{ij} obtained through the IPFP satisfy (10) and (11), they do not minimize the least square distance Φ (Stephan, 1942). Ireland and Kullback (1968) showed that the estimates $\hat{p}_{ij} = m_{ij}/n$ in fact minimize the "forward" discrimination information (also called the "forward" Kullback-Leibler distance by DiCiccio and Romano, 1990)

$$I(\boldsymbol{p}, \boldsymbol{q}) = \sum_{i=1}^{r}\sum_{j=1}^{c} p_{ij} \log\left(\frac{p_{ij}}{q_{ij}}\right)$$

with respect to p_{ij} subject to (10) and (11), where $q_{ij} = n_{ij}/n$ are the observed cell proportions.

Our proposed EL alternative to raking, under simple random sampling, is to estimate the cell proportions N_{ij}/N by the $\hat{p}_{ij}$ that maximize the EL function

$$l_0(\boldsymbol{p}) = \sum_{i=1}^{r}\sum_{j=1}^{c} n_{ij} \log(p_{ij})$$

subject to

$$\sum_{i=1}^{r}\sum_{j=1}^{c} p_{ij} = 1\,, \tag{12}$$

$$\sum_{j=1}^{c} p_{ij} = N_{i\cdot}/N\,, \quad i = 1, \cdots, r-1\,, \tag{13}$$

$$\sum_{i=1}^{r} p_{ij} = N_{\cdot j}/N\,, \quad j = 1, \cdots, c-1\,. \tag{14}$$

The EL function $l_0(\boldsymbol{p})$ is related to the "backward" discrimination information

$$I(\boldsymbol{q}, \boldsymbol{p}) = \sum_{i=1}^{r}\sum_{j=1}^{c} q_{ij} \log\left(\frac{q_{ij}}{p_{ij}}\right) = \sum_{i=1}^{r}\sum_{j=1}^{c} q_{ij}\log(q_{ij}) - \frac{1}{n}\, l_0(\boldsymbol{p})\,.$$

It is apparent that minimizing $I(\boldsymbol{q}, \boldsymbol{p})$ with respect to p_{ij} is equivalent to maximizing $l_0(\boldsymbol{p})$ with respect to p_{ij}. The EL function $l_0(\boldsymbol{p})$ is indeed the true multinomial likelihood function under simple random sampling with replacement.

The use of "forward" discrimination information $I(\boldsymbol{p}, \boldsymbol{q})$ for classical raking ratio estimation is equivalent to the *multiplicative* method described in Deville *et al.* (1993) and Deville and Särndal (1992). The resulting $\hat{p}_{ij}$ are guaranteed to be positive but often contain some extremely large values compared to q_{ij}. Deville *et al.* (1993) also discussed other alternative distance measures that force the ratio p_{ij}/q_{ij} to be confined within certain range. The major challenge in using these alternative methods, as well as the multiplicative method, is the computational implementation. Efficient algorithms are not available and the convergence of the involved iterative procedures is not guaranteed.

The most important feature of the EL approach, however, is the availability of a simple and efficient algorithm for the constrained maximization problem. Let $x_{(1)ij}, \cdots,$ $x_{(r-1)ij}$ be the first $r-1$ row indicator variables and $x_{ij(1)}, \cdots, x_{ij(c-1)}$ be the first $c-1$ column indicator variables. For instance, $x_{(1)ij} = 1$ if $i = 1$ and zero otherwise. Let

$$\boldsymbol{x}_{ij} = (x_{(1)ij}, \cdots, x_{(r-1)ij}, x_{ij(1)}, \cdots, x_{ij(c-1)})'$$

and

$$\overline{\boldsymbol{X}} = \left(\frac{N_{1\cdot}}{N}, \cdots, \frac{N_{(r-1)\cdot}}{N}, \frac{N_{\cdot 1}}{N}, \cdots, \frac{N_{\cdot (c-1)}}{N} \right)' .$$

The two sets of constraints (13) and (14) can be re-written as

$$\sum_{i=1}^{r} \sum_{j=1}^{c} p_{ij} \boldsymbol{x}_{ij} = \overline{\boldsymbol{X}} . \tag{15}$$

Using the standard Lagrange multiplier method it can be shown that the $\hat{p}_{ij}$ which maximize the pseudo EL function $l_0(\boldsymbol{p})$ subject to the normalization constraint (12) and the benchmark constraint (15) are given by $\hat{p}_{ij} = n_{ij}/\{n(1+\boldsymbol{\lambda}'\boldsymbol{u}_{ij})\}$, where $\boldsymbol{u}_{ij} = \boldsymbol{x}_{ij} - \overline{\boldsymbol{X}}$, and the vector-valued Lagrange multiplier $\boldsymbol{\lambda}$ is the solution to

$$g_0(\boldsymbol{\lambda}) = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{n_{ij} \boldsymbol{u}_{ij}}{n(1 + \boldsymbol{\lambda}' \boldsymbol{u}_{ij})} = \boldsymbol{0} . \tag{16}$$

A unique solution to (16) exists if none of the observed sample marginal totals $n_{i\cdot}$ or $n_{\cdot j}$ is zero. If a particular sample cell frequency $n_{ij} = 0$, we set $\hat{p}_{ij} = 0$ for that cell, and this does not change the existence and uniqueness of the solution. On the other hand, the classical raking ratio algorithm may not converge if some $n_{ij} = 0$. The solution to (16) can be found using the same algorithm of Chen *et al.* (2002) for solving (6).

Under a general probability sampling design and with known population mean $\overline{\boldsymbol{Z}}$ on a vector of auxiliary variables $\boldsymbol{z}$ in addition to known marginal totals in, for instance, a two dimensional contingency table, a pseudo EL alternative to raking is as follows. Let π_{ijk} be the inclusion probability and $d_{ijk} = 1/\pi_{ijk}$ be the basic design weight associated with the k-th unit in the (i, j)-th cell, $k = 1, \cdots, N_{ij}$. Let $\boldsymbol{z}_{ijk}$ be the additional vector-valued auxiliary variable, observed only for units in the sample but with known population mean $\overline{\boldsymbol{Z}}$. The population size N is also known from the contingency table. The goal is to estimate the population total $Y = \sum_{i=1}^{r} \sum_{j=1}^{c} \sum_{k=1}^{N_{ij}} y_{ijk}$, or equivalently the mean $\overline{Y} = Y/N$, of a study variable y. Estimation of the cell totals N_{ij} is a special case of Y where y is the (i, j)-th cell indicator variable.

Under the current setting, the non-stratified pseudo EL function is defined as

$$l_1(\boldsymbol{p}) = \sum_{i=1}^{r} \sum_{j=1}^{c} \sum_{k=1}^{n_{ij}} d_{ijk} \log(p_{ijk}) \,,$$

where p_{ijk} is the probability mass assigned to the k-th unit in the (i, j)-th cell and the p_{ijk} are subject to

$$\sum_{i=1}^{r} \sum_{j=1}^{c} \sum_{k=1}^{n_{ij}} p_{ijk} = 1 \,. \tag{17}$$

The MPEL estimator of $\overline{Y}$ is computed as

$$\hat{\overline{Y}} = \sum_{i=1}^{r} \sum_{j=1}^{c} \sum_{k=1}^{n_{ij}} \hat{p}_{ijk} y_{ijk} \,,$$

where the $\hat{p}_{ijk}$ maximize the pseudo EL function $l_1(\boldsymbol{p})$ subject to (17) and the set of benchmark constraints

$$\sum_{i=1}^{r} \sum_{j=1}^{c} \sum_{k=1}^{n_{ij}} p_{ijk} \boldsymbol{x}_{ijk} = \overline{\boldsymbol{X}} \,. \tag{18}$$

The vector-valued $\boldsymbol{x}_{ijk}$ consists of the first $r-1$ row indicator variables and the first $c-1$ column indicator variables as well as $\boldsymbol{z}_{ijk}$. It is very important to note that the row and column indicator variables used here are defined at the unit level, while those used in (15) are defined at the (i, j) cell level. For instance, the first row indicator variable used in (18) is defined as $x_{(1)ijk} = 1$ if $i = 1$ and zero otherwise. The population mean $\overline{\boldsymbol{X}}$ used in (18) consists of the first $r-1$ marginal row proportions $N_{i\cdot}/N$ and the first $c-1$ marginal column proportions $N_{\cdot j}/N$ as well as the mean $\overline{\boldsymbol{Z}}$.

It can be shown that the $\hat{p}_{ijk}$ are given by $\hat{p}_{ijk} = \tilde{d}_{ijk}(s)/(1+\boldsymbol{\lambda}'\boldsymbol{u}_{ijk})$, where $\tilde{d}_{ijk}(s) = d_{ijk}/\sum_{i=1}^{r}\sum_{j=1}^{c}\sum_{k=1}^{n_{ij}} d_{ijk}$ are the normalized design weights over the sample s, $\boldsymbol{u}_{ijk} = \boldsymbol{x}_{ijk} - \overline{\boldsymbol{X}}$, and $\boldsymbol{\lambda}$ is the solution to

$$g_1(\boldsymbol{\lambda}) = \sum_{i=1}^{r}\sum_{j=1}^{c}\sum_{k=1}^{n_{ij}} \frac{\tilde{d}_{ijk}(s)\boldsymbol{u}_{ijk}}{1+\boldsymbol{\lambda}'\boldsymbol{u}_{ijk}} = \boldsymbol{0}\,. \tag{19}$$

A unique solution to (19) exists if none of the observed sample marginal totals $n_{i\cdot}$ or $n_{\cdot j}$ is zero and $\overline{\boldsymbol{Z}}$ is an inner point of the convex hull formed by the $\boldsymbol{z}_{ijk}$ observed in the sample. One of the major features of the pseudo EL approach under the above formulation is that $g_1(\boldsymbol{\lambda}) = \boldsymbol{0}$ is very well structured and can be solved using the same algorithm of Chen *et al.* (2002).

The cell totals N_{ij} are estimated by $\hat{N}_{ij} = N\sum_{k=1}^{n_{ij}} \hat{p}_{ijk}$. It is apparent that

$$\sum_{j=1}^{c} \hat{N}_{ij} = N_{i\cdot}\,, i = 1, \cdots, r \quad \text{and} \quad \sum_{i=1}^{r} \hat{N}_{ij} = N_{\cdot j}\,, j = 1, \cdots, c\,.$$

Under simple random sampling and without the additional auxiliary variables $\boldsymbol{z}_{ijk}$, we have $d_{ijk} = N/n$ and the probability mass p_{ijk} take the same value for all the units k in the same (i, j)-th cell. Apart from a multiplying constant, the pseudo EL function $l_1(\boldsymbol{p})$ reduces to the EL function $l_0(\boldsymbol{p})$ used for the classical simple random sampling set-up.

4.3 Pseudo empirical likelihood ratio confidence intervals

One of the major attractive features of the EL approach is the nonparametric confidence intervals constructed through profiling the EL ratio function. With designs other than simple random sampling, the pseudo EL ratio statistic needs to be adjusted for the design effect. The exact definition of the design effect depends not only on the probability sampling design but also on the auxiliary information used. We consider pseudo EL intervals for the population mean $\overline{Y}$ and the distribution function $F_N(t)$ and discuss three scenarios.

For a non-stratified sampling design, the pseudo empirical log-likelihood ratio function for $\overline{Y}$ without using any auxiliary information at the estimation stage is given by

$$r_{ns}(\theta) = -2\{l_{ns}(\hat{\boldsymbol{p}}(\theta)) - l_{ns}(\hat{\boldsymbol{p}})\}\,, \tag{20}$$

where the $\hat{p}_i = \tilde{d}_i(s)$ maximize $l_{ns}(\boldsymbol{p})$ given by (2) subject to $p_i > 0$ and $\sum_{i\in s} p_i = 1$, and the $\hat{p}_i(\theta)$ are the values of p_i obtained by maximizing $l_{ns}(\boldsymbol{p})$ subject to

$$\sum_{i\in s} p_i = 1 \;\text{ and }\; \sum_{i\in s} p_i y_i = \theta \tag{21}$$

for a fixed θ. The design effect (abbreviated deff) in this case is associated with the estimator $\hat{\overline{Y}}_H$ and is defined as

$$\text{deff}_H = V_p(\hat{\overline{Y}}_H)/(S_y^2/n)\,, \tag{22}$$

where S_y^2 is the population variance and $V_p(\cdot)$ denotes the variance under the specified design $p(s)$. Under regularity conditions C1-C3 specified below, the pseudo EL ratio function $r_{ns}(\theta)$ converges in distribution to a scaled χ^2 random variable with one degree of freedom when $\theta = \overline{Y}$, where the scale factor is equal to the design effect deff_H (Wu and Rao, 2006). Hence, the adjusted pseudo EL ratio function

$$r_{ns}^{[a]}(\theta) = \{r_{ns}(\theta)\}/\text{deff}_H \tag{23}$$

converges in distribution to a χ^2 random variable with one degree of freedom when $\theta = \bar{Y}$. The regularity conditions are as follows:

C1 The sampling design $p(s)$ and the study variable y satisfy $\max_{i\in s}|y_i| = o_p(n^{1/2})$, where the stochastic order $o_p(\cdot)$ is with respect to the sampling design $p(s)$.

C2 The sampling design $p(s)$ satisfies $N^{-1}\sum_{i\in s} d_i - 1 = O_p(n^{-1/2})$.

C3 The HT estimator $\hat{\theta}_{HT} = N^{-1}\sum_{i\in s} d_i y_i$ of $\theta_0 = \bar{Y}$ is asymptotically normally distributed.

A $(1-\alpha)$-level confidence interval on $\overline{Y}$ can be constructed as $\{\theta \,|\, r_{ns}^{[a]}(\theta) \le \chi_1^2(\alpha)\}$, where $\chi_1^2(\alpha)$ is the upper α quantile of the χ_1^2 distribution. Finding such an interval, however, involves profile analysis described in Section 5.

For non-stratified sampling designs and a vector of auxiliary variables with known population means $\overline{\boldsymbol{X}}$, the pseudo empirical log-likelihood ratio function for $\overline{Y}$ is similarly defined as $r_{ns}(\theta)$ given in (20) but with the benchmark constraints (5) included in finding both $\hat{p}_i$ and $\hat{p}_i(\theta)$. The design effect under this scenario is associated with the estimator $\hat{\bar{Y}}_{GR}$ and is defined as

$$\text{deff}_{GR} = V_p(\hat{\bar{Y}}_{GR})/(S_r^2/n)\,, \tag{24}$$

where $V_p(\hat{\bar{Y}}_{GR}) = V_p\{\sum_{i\in s}\tilde{d}_i(s)r_i\}$, $r_i = y_i - \bar{Y} - \boldsymbol{B}'\boldsymbol{u}_i$, $\boldsymbol{B} = (\sum_{i=1}^N \boldsymbol{u}_i\boldsymbol{u}_i')^{-1}\sum_{i=1}^N \boldsymbol{u}_i y_i$, $\boldsymbol{u}_i = \boldsymbol{x}_i - \overline{\boldsymbol{X}}$, and $S_r^2 = (N-1)^{-1}\sum_{i=1}^N r_i^2$. Under conditions C1-C3 and assuming that C1 and C3 also apply to the components of $\boldsymbol{x}$, the adjusted pseudo empirical log-likelihood ratio statistic

$$r_{ns}^{(a)}(\theta) = \{r_{ns}(\theta)\}/\text{deff}_{GR} \tag{25}$$

is asymptotically distributed as χ_1^2 when $\theta = \bar{Y}$ (Wu and Rao, 2006).

Under stratified sampling and with known overall population mean $\overline{\boldsymbol{X}} = \sum_{h=1}^L W_h \overline{\boldsymbol{X}}_h$ but unknown strata means $\overline{\boldsymbol{X}}_h$, the pseudo EL ratio function of $\overline{Y}$ is defined as

$$r_{st}(\theta) = -2\{l_{st}(\hat{\boldsymbol{p}}_1(\theta),\cdots,\hat{\boldsymbol{p}}_L(\theta)) - l_{st}(\hat{\boldsymbol{p}}_1,\cdots,\hat{\boldsymbol{p}}_L)\}\,, \tag{26}$$

where $\hat{p}_{hi}$ maximize $l_{st}(\boldsymbol{p}_1,\cdots,\boldsymbol{p}_L)$ defined by (3) subject to the set of constraints

$$\sum_{i\in s_h} p_{hi} = 1\,,\ h = 1,\cdots,L \quad \text{and} \quad \sum_{h=1}^L W_h \sum_{i\in s_h} p_{hi}\boldsymbol{x}_{hi} = \overline{\boldsymbol{X}}\,, \tag{27}$$

and $\hat{p}_{hi}(\theta)$ maximize $l_{st}(\boldsymbol{p}_1,\cdots,\boldsymbol{p}_L)$ subject to (27) plus an additional constraint

$$\sum_{h=1}^L W_h \sum_{i\in s_h} p_{hi} y_{hi} = \theta$$

for a fixed θ. Under suitable regularity conditions on the sampling design and variables involved within each stratum, the adjusted pseudo EL ratio statistic

$$r_{st}^{[a]}(\theta) = \{r_{st}(\theta)\}/\text{deff}_{GR(st)}$$

is asymptotically distributed as χ_1^2 when $\theta = \overline{Y}$ (Wu and Rao, 2006). The design effect $\text{deff}_{GR(st)}$ is defined through an augmented vector of variables and is given in Section 5.

For each of the three scenarios discussed above, the design effect is defined as a population quantity and needs to be replaced by sample-based estimates for the construction of the pseudo EL ratio confidence intervals. The asymptotic coverage level remains valid

Table 2: 95% Confidence Intervals for the Distribution Function

q	CI	CP	L	U	AL
0.10	NA	90.7	0.2	9.1	0.134
	EL1	94.1	1.7	4.2	0.134
	EL2	94.5	1.9	3.6	0.127
0.50	NA	95.3	2.4	2.3	0.212
	EL1	95.5	2.4	2.1	0.208
	EL2	95.4	2.8	1.8	0.187
0.90	NA	93.9	5.0	1.1	0.116
	EL1	95.2	2.7	2.1	0.115
	EL2	93.5	4.0	2.5	0.110

if the design effect is consistently estimated; see Wu and Rao (2006) for details on the estimation of design effects.

Pseudo empirical likelihood ratio confidence intervals on $F(t)$ for a given t can be obtained by simply changing y_i to $I(y_i \leq t)$. Table 2 contains results on 95% confidence intervals on $\theta_0 = F_N(t)$ at the qth population quantile $t = t_q$ for selected values of q. This was part of an extensive simulation study originally reported in Wu and Rao (2006). The well-known Rao-Sampford unequal probability sampling method is used and the sample size is $n = 80$. Three types of confidence intervals were examined: normal approximation (NA) interval in the form of $(\hat{\theta}_0 - 1.96\{var(\hat{\theta}_0)\}^{1/2},\ \hat{\theta}_0 + 1.96\{var(\hat{\theta}_0)\}^{1/2})$ with truncation of the lower bound at 0 or upper bound at 1 if necessary; pseudo EL ratio interval ($EL1$) without using any additional auxiliary information; and pseudo EL ratio interval ($EL2$) using the constraint (4). The performance of these intervals is measured in terms of simulated coverage probability (CP), lower (L) and upper (U) tail error rates and average length (AL).

The message conveyed by Table 2 highlights the advantages of the pseudo EL ratio confidence intervals. For $q = 0.50$ and $t = t_{0.5}$, the population median, the sampling distribution of $\hat{\theta}_0 = \hat{F}_{EL}(t_{0.5})$ is nearly symmetric. In this case the NA interval usually performs well in terms of coverage probabilities and balanced tail error rates. But for small or large population quantiles ($q = 0.10$ or 0.90) where the underlying sampling distribution of $\hat{\theta}_0$ is skewed, NA intervals perform poorly: coverage probabilities are lower than the nominal level and tail error rates are not balanced. The EL-based intervals $EL1$ and $EL2$, on the other hand, perform well in all cases in terms of coverage probabilities and tail error rates, and $EL2$ gives shorter average length than $EL1$. In addition, both the lower bound and the upper bound of the EL-based intervals automatically locate within the range of the parameter space, $(0,1)$, which is not always the case for the conventional NA intervals.

5 Computational Algorithms

There are three major computational tasks for implementing the EL-based methods: (i) to find the Lagrange multiplier $\boldsymbol{\lambda}$ as the solution to (6) with a single non-stratified sample; (ii) to obtain the MPEL solutions for stratified sampling, raking ratio estimation and other "irregular" cases; and (iii) to construct the pseudo EL ratio confidence intervals through profiling.

We assume that the population mean $\overline{\boldsymbol{X}}$ is an inner point of the convex hull formed by the sample observations $\{\boldsymbol{x}_i, i \in s\}$ so that a unique solution to (6) exists. Chen *et al.* (2002) proposed a simple algorithm for solving (6) with guaranteed convergence if the solution exists. The uniqueness of the solution and the convergence of the algorithm are proved based on a duality argument: maximizing $l_{ns}(\boldsymbol{p})$ *with respect to* $\boldsymbol{p}$ subject to $p_i > 0$, $\sum_{i\in s} p_i = 1$ and the benchmark constraints (5) is a dual problem of maximizing $H(\boldsymbol{\lambda}) = \sum_{i\in s} \tilde{d}_i(s)\log(1+\boldsymbol{\lambda}'\boldsymbol{u}_i)$ *with respect to* $\boldsymbol{\lambda}$ with no restrictions on $\boldsymbol{\lambda}$. In both cases the solution $\boldsymbol{\lambda}$ solves the equation system (6). Since $H(\boldsymbol{\lambda})$ is a concave function of $\boldsymbol{\lambda}$ with the matrix of second order derivatives negative definite, a unique maximum point to $H(\boldsymbol{\lambda})$ exists and can be found using the Newton-Raphson search algorithm. Denoting $\boldsymbol{x}_i - \overline{\boldsymbol{X}}$ by $\boldsymbol{u}_i$, the algorithm of Chen *et al.* (2002) for solving (6) is as follows.

Step 0: Let $\boldsymbol{\lambda}_0 = \boldsymbol{0}$. Set $k=0$, $\gamma_0 = 1$ and $\epsilon = 10^{-8}$.

Step 1: Calculate $\Delta_1(\boldsymbol{\lambda}_k)$ and $\Delta_2(\boldsymbol{\lambda}_k)$ where

$$\Delta_1(\boldsymbol{\lambda}) = \sum_{i\in s} \tilde{d}_i(s)\frac{\boldsymbol{u}_i}{1+\boldsymbol{\lambda}'\boldsymbol{u}_i} \quad \text{and} \quad \Delta_2(\boldsymbol{\lambda}) = \left\{ -\sum_{i\in s} \tilde{d}_i(s)\frac{\boldsymbol{u}_i\boldsymbol{u}_i'}{(1+\boldsymbol{\lambda}'\boldsymbol{u}_i)^2} \right\}^{-1} \Delta_1(\boldsymbol{\lambda})\,.$$

If $\|\Delta_2(\boldsymbol{\lambda}_k)\| < \epsilon$, stop the algorithm and report $\boldsymbol{\lambda}_k$; otherwise go to Step 2.

Step 2: Calculate $\boldsymbol{\delta}_k = \gamma_k \Delta_2(\boldsymbol{\lambda}_k)$. If $1 + (\boldsymbol{\lambda}_k - \boldsymbol{\delta}_k)'\boldsymbol{u}_i \le 0$ for some i, let $\gamma_k = \gamma_k/2$ and repeat Step 2.

Step 3: Set $\boldsymbol{\lambda}_{k+1} = \boldsymbol{\lambda}_k - \boldsymbol{\delta}_k$, $k = k+1$ and $\gamma_{k+1} = (k+1)^{-1/2}$. Go to Step 1.

It turns out that this modified Newton-Raphson algorithm for a single non-stratified sample is also applicable to stratified samples and a variety of other "irregular" cases after suitable reformulation. Under stratified sampling with known $\overline{\boldsymbol{X}}$, the basic problem is to maximize $l_{st}(\boldsymbol{p}_1, \cdots, \boldsymbol{p}_L)$ subject to (27). If we use $\overline{\boldsymbol{X}}^*$ to denote the augmented $\overline{\boldsymbol{X}}$ to include $W_1, \cdots, W_{L-1}$ as its first $L-1$ components and $\boldsymbol{x}_{hi}^*$ to denote the augmented $\boldsymbol{x}_{hi}$ to include the first $L-1$ stratum indicator variables, then the set of constraints (27) can equivalently be re-written as (Wu, 2004b)

$$\sum_{h=1}^{L} W_h \sum_{i \in s_h} p_{hi} = 1 \quad \text{and} \quad \sum_{h=1}^{L} W_h \sum_{i \in s_h} p_{hi} \boldsymbol{x}_{hi}^* = \overline{\boldsymbol{X}}^* . \tag{28}$$

This re-formulation makes all the steps for maximization under non-stratified sampling applicable to stratified sampling. The difference between non-stratified and stratified sampling is simply a matter of single or double summation. Maximizing (3) subject to (27), or equivalently (28), gives $\tilde{p}_{hi} = \tilde{d}_{hi}(s_h)/(1 + \boldsymbol{\lambda}'\boldsymbol{u}_{hi}^*)$ where $\boldsymbol{u}_{hi}^* = \boldsymbol{x}_{hi}^* - \overline{\boldsymbol{X}}^*$ and the vector-valued $\boldsymbol{\lambda}$ is the solution to

$$\sum_{h=1}^{L} W_h \sum_{i \in s_h} \frac{\tilde{d}_{hi}(s_h)\boldsymbol{u}_{hi}^*}{1 + \boldsymbol{\lambda}'\boldsymbol{u}_{hi}^*} = \boldsymbol{0} ,$$

which can be solved using the same algorithm of Chen *et al.* (2002). Using the augmented variables $\boldsymbol{x}_i^*$ and $\overline{\boldsymbol{X}}^*$, the design effect $\text{deff}_{GR(st)}$ required for the pseudo EL ratio confidence intervals discussed in Section 4.3 is defined as

$$\text{deff}_{GR(st)} = \left\{ \sum_{h=1}^{L} W_h^2 \, V_p \left(\sum_{i \in s_h} \tilde{d}_{hi}(s_h) r_{hi} \right) \right\} / \left(\frac{S_r^2}{n} \right) ,$$

where $r_{hi} = (y_{hi} - \bar{Y}) - (\boldsymbol{B}^*)'(\boldsymbol{x}_{hi}^* - \bar{\boldsymbol{X}}^*)$, $\boldsymbol{B}^*$ is the population vector of regression coefficients similarly defined as $\boldsymbol{B}$ for deff_{GR} given by (24) but using $\boldsymbol{x}_i^*$ and $\bar{\boldsymbol{X}}^*$, and $S_r^2 = (N-1)^{-1} \sum_{h=1}^{L} \sum_{i=1}^{N_h} r_{hi}^2$.

The pseudo EL alternative to raking discussed in Section 4.2 involves finding solutions to (16) or (19), which has the same structure as (6) and can be solved using the same algorithm. Other "irregular" EL-based methods include combining information from multiple surveys (Wu, 2004) where the algorithm of Chen *et al.* (2002) once again is the fundamental piece for computational procedures.

Construction of the pseudo EL ratio confidence intervals for $\theta_0 = \overline{Y}$ involves two steps: (i) calculate $r^{[a]}(\theta)$ for a given θ; and (ii) find the lower and upper bound for $\{\theta \mid r^{[a]}(\theta) \leq \chi_1^2(\alpha)\}$. The first step invokes no additional complications since, for instance, the constraint $\sum_{i \in s} p_i y_i = \theta$ for a given θ can be treated as an additional component for the benchmark constraints $\sum_{i \in s} p_i \boldsymbol{x}_i = \overline{\boldsymbol{X}}$. For the second step, the lower and upper bounds of the interval can be found through a simple bi-section search method since the interval is confined in between $y_{(1)} = \min_{i \in s} y_i$ and $y_{(n)} = \max_{i \in s} y_i$, and the adjusted pseudo EL ratio function $r^{[a]}(\theta)$ is monotone decreasing for $\theta \in (y_{(1)}, \hat{\overline{Y}}_{EL})$ and monotone increasing for $\theta \in (\hat{\overline{Y}}_{EL}, y_{(n)})$. Wu (2005) contains the detailed argument and also provides R/SPLUS functions and codes for several key computational procedures.

6 Discussion

The EL and PEL approaches are flexible enough to handle a variety of other problems. In particular, data from two or more independent surveys from the same target population can be combined naturally through the PEL approach and efficient point estimators and pseudo EL ratio confidence intervals for the population mean can be obtained. For example, suppose $\{(y_i, \boldsymbol{x}_{1i}, \boldsymbol{z}_i), i \in s_1\}$ are the sample data from the first survey and $\{(\boldsymbol{x}_{2j}, \boldsymbol{z}_j), j \in s_2\}$ are the data from the second survey, where the auxiliary variables $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ have known population means $\overline{\boldsymbol{X}}_1$ and $\overline{\boldsymbol{X}}_2$ but the population mean of the "common" auxiliary variable $\boldsymbol{z}$ is unknown. The goal here is to make inference on the population mean $\overline{Y}$ associated with the first survey, taking advantage of the supplementary data from the two sources. The maximum PEL estimator of $\overline{Y}$ is obtained by maximizing a combined pseudo empirical log likelihood function $l(\boldsymbol{p}_1, \boldsymbol{p}_2)$ based on the two independent samples subject to the following normalization, benchmarking and internal consistency constraints (Wu, 2004):

$$\sum_{i\in s_t} p_{ti} = 1\,, \quad \sum_{i\in s_t} p_{ti}\boldsymbol{x}_{ti} = \overline{\boldsymbol{X}}_t\,, \quad t=1,2 \quad \text{and} \quad \sum_{i\in s_1} p_{1i}\boldsymbol{z}_i = \sum_{i\in s_2} p_{2i}\boldsymbol{z}_i\,.$$

The estimator is given by $\hat{\overline{Y}} = \sum_{i\in s_1}\hat{p}_{1i}y_i$, where the $\hat{p}_{ti}$ $(t=1,2)$ maximize $l(\boldsymbol{p}_1,\boldsymbol{p}_2)$. Similarly, the PEL ratio confidence intervals for $\overline{Y}$ can be obtained. The above approach can be extended to handle data from independent samples taken from two or more incomplete frames together covering the population of interest. The PEL approach is flexible in combining data from different sources as demonstrated above. Depending on what is available, new constraints can be added to and existing ones can be removed from the system of constraints.

Another problem of interest is to make inference on the population parameters of interest in the presence of imputation for item non-response. Again, the EL and pseudo EL approaches can be applied in a systematic manner to handle imputed data and any auxiliary population information. Recent work has focused on EL inference on the mean, distribution function and quantiles of a variable of interest y, assuming that an *iid* sample $\{(y_i,x_i), i=1,\cdots,n\}$ subject to missing y_i is available. The missing y-values are imputed using regression imputation, assuming a missing at random (MAR) response mechanism and a linear regression model (Wang and Rao, 2002a; Qin, Rao and Ren, 2006). EL inference using kernel regression imputation, assuming only that the conditional expectation of y given x is a smooth function of x, has likewise been studied (Wang and Rao, 2002b; Wang and Chen, 2006). Various extensions have also been analyzed. The pseudo EL approach can be applied to extend the above work to survey data.

In summary, the EL and pseudo EL approaches have several advantages over the traditional approaches to inference from survey data. The advantages include (i) likelihood based motivations; (ii) intrinsic positive weights and efficient point and interval estimation taking account of benchmark and other constraints naturally; (iii) orientation of the confidence intervals is determined by the data and the range of the parameter space is fully preserved, unlike the customary normal theory intervals; and (iv) flexibility in handling a variety of problems in a systematic manner. However, the true empirical likelihood is not available for unequal probability sampling without replacement and other complex designs, and the rationale in Section 4 for using a pseudo empirical likelihood to overcome this problem is not entirely appealing from a theoretical point of view.

Acknowledgement

This work was supported by grants from the Natural Sciences and Engineering Research Council of Canada. We are grateful to Phil Kott for many useful comments and suggestions.

References

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

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

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

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

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

Deming, W.E. and F.F. Stephan (1940). On a least squares adjustment of a sample frequency when the expected Marginal totals are known. *Ann. Math. Statist.*, 11, 427-444.

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

Deville, J.C., C.E. Särndal and O. Sautory (1993). Generalized raking procedures in survey sampling. *Journal of the American Statistical Association*, 88, 1013-1020.

DiCiccio, T. and J. Romano (1990). Nonparametric confidence limits by resampling methods and least favorable families. *International Statistical Review*, 58, 59-76.

Ericson, W.A. (1969). Subjective Bayesian models in sampling finite populations. *Journal of the Royal Statistical Society*, B, 31, 195-224.

Godambe, V.P. (1955). A unified theory of sampling from finite populations. *Journal of the Royal Statistical Society*, B, 17, 267-278.

Godambe, V.P. (1966). A new approach to sampling from finite populations, I and II. *Journal of the Royal Statistical Society*, B, 28, 310-328.

Godambe, V.P. and V.M. Joshi (1965). Admissibility and Bayes estimation in sampling from finite populations, I. *Annals of Mathematical Statistics*, 36, 1707-1722.

Hajek, J. (1971). Comments on a paper by Basu. In *Foundations of Statistical Inference* (V.P. Godambe and D.A. Sprott eds.), Holt, Rinehart and Winston, 203-242.

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

Hartley, H.O. and J.N.K. Rao (1969). A new estimation theory for sample surveys, II. In *New Developments in Survey Sampling*, Wiley Inter-science, New York, 147-169.

Ireland, C.T. and S. Kullback (1968). Contingency tables with given marginals. *Biometrika*, 55, 179-188.

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

Owen, A.B. (2001). *Empirical likelihood.* Chapman and Hall, New York.

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

Qin, Y., J.N.K. Rao and Q. Ren (2006). Confidence intervals for parameters of the response variable in a linear model with missing data. Technical Report, Laboratory for Research in Statistics and Probability, Carleton University, Ottawa, Canada.

Rao, J.N.K. (1994). Estimating totals and distribution functions using auxiliary information at the estimation stage. *Journal of Official Statistics*, 10, 153-165.

Rao, J.N.K. and C. Wu (2005). Empirical likelihood approach to calibration using survey data. Proceedings of 55th Session of International Statistics Institute, Sydney.

Särndal, C.E., B. Swensson and J. Wretman (1992). *Model Assisted Survey Sampling.* Springer.

Sitter, R.R. and C. Wu (2002). Efficient estimation of quadratic finite population functions in the presence of auxiliary information. *Journal of the American Statistical Association*, 97, 535-543.

Stephan, F.F. (1942). An iterative method of adjusting sample frequency tables when expected marginal totals are known. *Ann. Math. Statist.*, 13, 166-178.

Wang, D. and S.X. Chen (2006). Nonparametric imputation of missing values for estimating equation based inference - a full report. Technical Report, Department of Statistics, Iowa State University, Ames, Iowa.

Wang, Q. and J.N.K. Rao (2002a). Empirical likelihood-based inference in linear models with missing data. *Scandinavian Journal of Statistics*, 29, 563-576.

Wang, Q. and J.N.K. Rao (2002b). Empirical likelihood-based inference under imputation for missing response data. *Annals of Statistics*, 30, 896-924.

Wu, C. (2003). Optimal calibration estimators in survey sampling. *Biometrika*, 90, 937-951.

Wu, C. (2004a). Combining information from multiple surveys through the empirical likelihood method. *The Canadian Journal of Statistics*, 32, 15-26.

Wu, C. (2004b). Some algorithmic aspects of the empirical likelihood method in survey sampling. *Statistica Sinica*, 14, 1057-1067.

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

Wu, C. and R.R. Sitter (2001). A model-calibration approach to using complete auxiliary information from survey data. *Journal of the American Statistical Association*, 96, 185-193.

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

Zhong, C.X.B. and J.N.K. Rao (1996). Empirical likelihood inference under stratified random sampling using auxiliary information. *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 798-803.

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

Empirical Likelihood Methods

1. **Likelihood-based Approaches**
2. **Empirical Likelihood Method Under Simple Random Sampling**
3. **Stratified Simple Random Sampling**
4. **Pseudo Empirical Likelihood Method**
   - **4.1 Pseudo empirical likelihood approach to calibration**
   - **4.2 Pseudo empirical likelihood alternative to raking**
   - **4.3 Pseudo empirical likelihood ratio confidence intervals**
5. **Computational Algorithms**
6. **Discussion**

Key Words:

Auxiliary information; Benchmark constraint; Calibration equation; Calibration estimator; Confidence interval; Constrained maximization; Convex hull; Coverage probability; Design effect; Design weight; Distribution function; Effective sample size; Empirical likelihood; Empirical likelihood ratio; Horvitz-Thompson estimator; Hajek estimator; Imputation; Inclusion probability; Indicator variable; Lagrange multiplier; Marginal totals; Non-response; Normalized weights; Optimal estimator; Population mean; Pseudo empirical likelihood; Scale-load; Raking; Range-restricted weights; Regression estimator; Simple random sampling; Stratified sampling; Tail error rates; Unequal probability sampling;