Empirical Likelihood confidence regions for regression parameters under unequal probability sampling

Yves G.BERGER*
University of Southampton
Southampton Statistical Sciences Research Institute
Southampton, SO17 1BJ, United Kingdom
Y.G.Berger@soton.ac.uk

We propose a new empirical likelihood approach which can be used to construct design-based confidence regions of regression parameters under unequal probability sampling. The proposed approach gives confidence regions which may have better coverages than standard confidence regions and pseudo empirical likelihood confidence regions which rely on variance estimates and design-effects. The proposed approach does not rely on variance estimates, design-effects, re-sampling or linearisation, even when the regression parameter is not linear. It also gives suitable confidence regions when the point estimator is biased. The proposed approach can also be adjusted to account for large sampling fractions.

Key Words: Design-based approach, Estimating equations, Regression estimator, Unequal inclusion probabilities.

Introduction

Let U be a finite population of N units; where N is a fixed quantity which is not necessarily known. Let $\boldsymbol{\theta}_0$ be an r-dimensional population parameter of interest. Suppose that $\boldsymbol{\theta}_0$ is the unique solution of the following set of estimating equation (e.g. Qin and Lawless, 1994).

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

where $\boldsymbol{g}_i(\boldsymbol{\theta})$ is an r-dimensional function of $\boldsymbol{\theta}$ and of characteristics of the unit i. This function does not need to be differentiable.

Suppose we are interested in modelling the relationship between a response (or dependent) variable y_i and some r-dimensional covariates (or independent variables) $\boldsymbol{u}_i$. For example, with a linear model, $\boldsymbol{\theta}_0$ is a regression parameter and

$$\boldsymbol{g}_i(\boldsymbol{\theta}) = \boldsymbol{u}_i(y_i - \boldsymbol{u}_i'\boldsymbol{\theta}).$$

With a non-linear regression, we have

$$\boldsymbol{g}_i(\boldsymbol{\theta}) = \frac{\partial f(\boldsymbol{u}_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \left\{ y_i - f(\boldsymbol{u}_i, \boldsymbol{\theta}) \right\},$$

where $f(\boldsymbol{u}_i, \boldsymbol{\theta})$ is a non linear function. For a robust regression, we have

$$\boldsymbol{g}_i(\boldsymbol{\theta}) = \boldsymbol{u}_i \psi(y_i, \boldsymbol{u}_i'\boldsymbol{\theta}),$$

where $\psi(\cdot)$ is a robust function (Huber, 1981).

Suppose that we wish to estimate $\boldsymbol{\theta}_0$ from the data of a sample s of size n selected with a single stage unequal probabilities without replacement sampling design. We consider that the sample size n is a fixed (non random) quantity. We adopt a design-based (non-parametric) approach; where the sampling distribution is specified by the sampling design and where the values of the response variable and the covariates are fixed (non random) quantities.

Empirical likelihood approach

We propose to use the following empirical likelihood function (e.g. Owen, 2001).

$$L(m) = \prod_{i=1}^{n} \frac{m_i}{N}, \tag{2}$$

where m_i is the unit mass of unit i in the population (e.g. Deville, 1999).

Hartley and Rao (1969) showed that (2) is the an empirical likelihood function under unequal probability sampling with replacement, as m_i/N is the probability to observe the i-th unit. Owen (2001, Ch. 6) showed that (2) is a suitable empirical likelihood function when the units are selected independently with a Poisson sampling design. Although under fixed size sampling designs, the units are not selected independently, we propose to use the empirical likelihood function (2) under fixed size sampling designs. The aim is to show that this empirical likelihood function can be used for point estimation and to construct confidence regions (or to derive tests) under fixed size sampling designs.

The maximum likelihood estimators of m_i are the values $\widehat{m}_i$ which maximise the *log-empirical likelihood function*

$$\ell(m) = \sum_{i=1}^{n} \log\left(m_i\right), \tag{3}$$

subject to the constraints $m_i \geq 0$ and

$$\sum_{i=1}^{n} m_i \boldsymbol{c}_i \;\; = \;\; \boldsymbol{C}; \tag{4}$$

where $\sum_{i=1}^{n}$ denotes the sum over the sampled units, $\boldsymbol{c}_i$ is a known $Q \times 1$ vector associated with the i-th sampled unit and $\boldsymbol{C}$ is a known $Q \times 1$ vector. We also assume that the constraint (4) is such that the fixed size constraint

$$\sum_{i=1}^{n} m_i \pi_i = n \tag{5}$$

always holds, where π_i denotes the inclusion probability of unit i. Under equal probability sampling, we have that $\pi_i = n/N$, and the constraint (5) reduces to $\sum_{i=1}^{n} m_i = N$ which is the constraint adopted under equal probability sampling (e.g. Rao and Wu, 2009). Berger and De La Riva Torres (2012) showed that the solution of this maximisation is given by

$$\widehat{m}_i = \left(\pi_i + \boldsymbol{\eta}' \boldsymbol{c}_i\right)^{-1}, \tag{6}$$

The quantity $\boldsymbol{\eta}$ is such that the constraint (4) holds. This quantity can be computed using an iterative Newton-Raphson procedure (e.g. Berger and De La Riva Torres, 2012; Rao and Wu, 2009).

The *maximum empirical likelihood estimator* $\widehat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}_0$ is defined by solution of the following estimating equation.

$$\widehat{\boldsymbol{G}}(\boldsymbol{\theta}) = \boldsymbol{0}, \quad \text{with} \quad \widehat{\boldsymbol{G}}(\boldsymbol{\theta}) = \sum_{i=1}^{n} \widehat{m}_i \; \boldsymbol{g}_i(\boldsymbol{\theta}); \tag{7}$$

where $\widehat{m}_i$ is defined by (6). We assume that the $\boldsymbol{g}_i(\boldsymbol{\theta})$ are such that $\widehat{\boldsymbol{G}}(\boldsymbol{\theta}) = \boldsymbol{0}$ has a unique solution. The estimator $\widehat{\boldsymbol{\theta}}$ is a maximum empirical likelihood estimator because it also minimises the empirical log-likelihood ratio function (or deviance) defined by (8).

For example, suppose that we would like to fit a linear model with only an intercept ($\boldsymbol{u}_i = 1$). In this case, $\boldsymbol{g}_i(\boldsymbol{\theta}) = y_i - \theta$. If $\boldsymbol{c}_i = \pi_i$ and $\boldsymbol{C} = n$, we have that $m_i = \pi_i^{-1}$ and

$$\widehat{\boldsymbol{G}}(\boldsymbol{\theta}) = \sum_{i=1}^{n} \frac{g_i(\theta)}{\pi_i}$$

is the Horvitz and Thompson (1952) estimator of the function (1). As $\boldsymbol{g}_i(\boldsymbol{\theta}) = y_i - \theta$, we have that $\widehat{\theta}$ is the Hájek (1971) estimator of the population mean; that is, $\widehat{\theta} = \widehat{N}_\pi^{-1} \sum_{i=1}^n y_i \pi_i^{-1}$, where $\widehat{N}_\pi = \sum_{i=1}^n \pi_i^{-1}$.

Empirical log-likelihood ratio function

The main advantage of the empirical likelihood approach is its capability of deriving non-parametric confidence regions which do not depend on variance estimates or on the normality of the point estimator. In this Section , we propose to use the empirical log-likelihood ratio function defined by (8) to derive empirical likelihood confidence regions.

Let $\widehat{m}_i$ be the values which maximise (3) subject to the constraints $m_i \geq 0$ and (4) when $\boldsymbol{c}_i = \pi_i$ and $\boldsymbol{C} = n$. Note that $m_i = \pi_i^{-1}$ in this situation. The maximum value of the empirical log-likelihood function is given by $\ell(\widehat{m}) = -\sum_{i=1}^n \log(\pi_i)$. Let $\widehat{m}_i^*$ be the values which maximise (3) subject to the constraints $m_i \geq 0$ and (4) with $\boldsymbol{c}_i = \boldsymbol{c}_i^*$ and $\boldsymbol{C} = \boldsymbol{C}^*$, where $\boldsymbol{c}_i^* = (\pi_i, \boldsymbol{g}_i(\boldsymbol{\theta}))'$ and $\boldsymbol{C}^* = (n, \boldsymbol{0})'$. Let $\ell(\widehat{m}^*, \boldsymbol{\theta})$ be the maximum value of the empirical log-likelihood function. The *empirical log-likelihood ratio function* (or deviance) is defined by the following function of $\boldsymbol{\theta}$.

$$\widehat{r}(\boldsymbol{\theta}) \quad = \quad 2\left\{\ell(\widehat{m}) - \ell(\widehat{m}^*, \boldsymbol{\theta})\right\}. \tag{8}$$

It can be easily shown that $\widehat{r}(\widehat{\boldsymbol{\theta}}) = \boldsymbol{0}$. Hence $\widehat{\boldsymbol{\theta}}$ is indeed the maximum empirical likelihood estimator of $\boldsymbol{\theta}_0$, because it minimises the empirical log-likelihood ratio function. Berger and De La Riva Torres (2012) showed how the stratification can be taken into account by including the stratification variables within the vectors $\boldsymbol{c}_i$ and $\boldsymbol{c}_i^*$.

We will show that under a set of regularity conditions, $\widehat{r}(\boldsymbol{\theta}_0)$ follows asymptotically a chi-squared distribution with r degree of freedom when the sampling fraction, n/N, is negligible. This property relies on the fact that

$$\widehat{\boldsymbol{G}}_\pi(\boldsymbol{\theta}_0) = \sum_{i=1}^n \frac{\boldsymbol{g}_i(\boldsymbol{\theta}_0)}{\pi_i}, \tag{9}$$

is a vector of Horvitz and Thompson (1952) estimator which follows a multivariate normal distribution asymptotically (Berger, 1998; Hájek, 1964; Vísek, 1979).

Empirical likelihood confidence regions

As $\widehat{r}(\boldsymbol{\theta}_0)$ follows asymptotically a chi-squared distribution, the $(1-\alpha)$ level empirical likelihood confidence region (Wilks, 1938) for the population parameter $\boldsymbol{\theta}_0$ is given by the following set

$$\left\{\boldsymbol{\theta} :\ \widehat{r}(\boldsymbol{\theta}) \leq \chi^2_{df=p}(\alpha)\right\}, \tag{10}$$

where $\chi^2_{df=p}(\alpha)$ is the upper α-quantile of the chi-squared distribution with r degree of freedom. Note that $\widehat{r}(\boldsymbol{\theta})$ is a convex non-symmetric function with a minimum when $\boldsymbol{\theta}$ is the maximum empirical likelihood estimator. This region can be found using a bijection search method. This involves calculating $\widehat{r}(\boldsymbol{\theta})$ for several values of $\boldsymbol{\theta}$.

Let $\boldsymbol{\theta}_0 = (\boldsymbol{\theta}_1', \boldsymbol{\theta}_2')'$; where $\boldsymbol{\theta}_1$ is an r_1-dimensional vector ($r_1 < r$). Suppose that we would like to derive a confidence region for the parameter $\boldsymbol{\theta}_1$. The parameter $\boldsymbol{\theta}_2$ is treated as a nuisance parameter. Consider

$$\widehat{r}(\boldsymbol{\theta}_1) \quad = \quad 2\left\{\ell(\widehat{m}) - \mathrm{Max}_{\boldsymbol{\theta}_2}\, \ell(\widehat{m}^*, \boldsymbol{\theta})\right\};$$

where $\mathrm{Max}_{\boldsymbol{\theta}_2}\, \ell(\widehat{m}^*, \boldsymbol{\theta})$ is the maximum value of $\ell(\widehat{m}^*, \boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}_2$, for a fixed value of $\boldsymbol{\theta}_1$. We will show that $\widehat{r}(\boldsymbol{\theta}_1)$ follows asymptotically a chi-squared distribution with r_1 degree of freedom when

the sampling fraction is negligible. Hence, the set $\{\boldsymbol{\theta}_1^\circ : \widehat{r}(\boldsymbol{\theta}_1^\circ) \leq \chi^2_{df=r_1}(\alpha)\}$ gives a confidence region for $\boldsymbol{\theta}_1$.

Empirical Likelihood Approach with Auxiliary Variables

In practice, population control totals of auxiliary variables are often known and this information is often taken into account at the estimation stage. Let x_i be the value of an auxiliary variables attached to unit i. Suppose that the control total $X = \sum_{i \in U} x_i$ is known. Traditional approaches (Birch, 1963) consist in adding these auxiliary variables within the covariates $\boldsymbol{u}_i$. However, in this situation, a different model is fitted and in practice we may want not to include the variable x_i within $\boldsymbol{u}_i$. In the approach proposed, we do not include x_i within $\boldsymbol{u}_i$. The variable x_i is use to derive the weights $\widehat{m}_i$ of the estimating equation.

Let $\boldsymbol{c}_i = (x_i,\ \pi_i)'$ and $\boldsymbol{C} = (X,\ n)'$. Berger and De La Riva Torres (2012) showed that

$$\widehat{\boldsymbol{G}}(\boldsymbol{\theta}) = \widehat{\boldsymbol{G}}_\pi(\boldsymbol{\theta}) + \widehat{\boldsymbol{B}}(X - \widehat{X}_\pi) + \widehat{\boldsymbol{\epsilon}}, \tag{11}$$

where $\widehat{X}_\pi = \sum_{i=1}^n \breve{x}_i$, $\|\widehat{\boldsymbol{\epsilon}}\| = o_p(N)$,

$$\widehat{\boldsymbol{B}} = \frac{\sum_{i=1}^n (\breve{x}_i - n^{-1}\widehat{X}_\pi)(\breve{\boldsymbol{g}}_i(\boldsymbol{\theta}) - n^{-1}\widehat{G}_\pi(\theta)}{\sum_{i=1}^n (\breve{x}_i - n^{-1}\widehat{X}_\pi)^2},$$

$\breve{x}_i = x_i\pi_i^{-1}$ and $\breve{\boldsymbol{g}}_i(\boldsymbol{\theta}) = \boldsymbol{g}_i(\boldsymbol{\theta})\pi_i^{-1}$. Note that $\widehat{\boldsymbol{B}}$ is the estimator of the covariance between $\widehat{\boldsymbol{G}}_\pi(\boldsymbol{\theta})$ and $\widehat{X}_\pi$ divided by the estimator of the variance of $\widehat{X}_\pi$ under a with replacement pps sampling design (e.g. Särndal et al., 1992, p. 99). Therefore $\widehat{\boldsymbol{B}}$ is the optimal regression coefficient (e.g. Berger et al., 2003; Isaki and Fuller, 1982; Montanari, 1987; Rao, 1994; Särndal, 1996) when the sampling fraction is negligible. Hence, the empirical likelihood estimator is asymptotically optimal. The result (11) can be generalised when $\boldsymbol{x}_i$ is a vector of values of auxiliary variables. When N is known, we recommend to use $x_i = 1$ or to include a variable equal to one into $\boldsymbol{x}_i$. This may improve the efficiency of the maximum empirical likelihood estimator.

With auxiliary variables, the confidence regions have to be constructed using the following restricted empirical likelihood approach proposed by (Berger and De La Riva Torres, 2012), because the function (8) may not converge to a chi-square distribution. Let $\boldsymbol{c}_i = \dot{\boldsymbol{c}}_i$, $\boldsymbol{c}_i^* = (\dot{\boldsymbol{c}}_i', \boldsymbol{g}_i(\boldsymbol{\theta}_0))'$, $\boldsymbol{C} = (\boldsymbol{X}', \boldsymbol{n}')'$, and $\boldsymbol{C}^* = (\boldsymbol{X}', \boldsymbol{n}', \boldsymbol{0})'$, with $\dot{\boldsymbol{c}}_i = (\boldsymbol{x}_i', \widetilde{\pi}_i(\boldsymbol{x}))'$; where $\widetilde{\pi}_i(\boldsymbol{x}) = \widehat{m}_i(\boldsymbol{x})^{-1}$. Let $\ell(\widehat{m}, \boldsymbol{x})$ be the maximum value (3) under the constraint (4) with $\boldsymbol{c}_i$ and $\boldsymbol{C}$. Let $\ell(\widehat{m}^*, \boldsymbol{x}, \boldsymbol{\theta})$ be the maximum value (3) under the constraint (4) with $\boldsymbol{c}_i^*$ and $\boldsymbol{C}^*$. In both cases, we consider that the constraints (4) are such that $\sum_{i=1}^n m_i\widetilde{\pi}_i(\boldsymbol{x}) = n$. The *restricted empirical log-likelihood ratio function* is given by

$$\widehat{r}_x(\boldsymbol{\theta}) \quad = \quad 2\left\{\ell(\widehat{m}, \boldsymbol{x}) - \ell(\widehat{m}^*, \boldsymbol{x}, \boldsymbol{\theta})\right\}. \tag{12}$$

Berger and De La Riva Torres (2012) showed that $\widehat{r}_x(\boldsymbol{\theta}_0)$ follows asymptotically a chi-squared distribution with r degree of freedom. Note that $\widetilde{\pi}_i(\boldsymbol{x})$ can be replaced by the inverse of any calibration weights (e.g. Deville and Särndal, 1992) which are calibrated with respect to $(\boldsymbol{x}_i', \pi_i)'$, as the restricted empirical log-likelihood ratio function still follows a chi-squared distribution in this situation. Note that $\widetilde{\pi}_i(\boldsymbol{x})$ may be larger that one. However, this does not cause any problem. Berger and De La Riva Torres (2012) showed how the stratification can be taken into account by including the stratification variables within the vectors $\boldsymbol{c}_i$ and $\boldsymbol{c}_i^*$.

Non-negligible sampling fractions

With large sampling fractions, the empirical log-likelihood ratio function does not necessarily follow a chi-squared distribution. Berger and De La Riva Torres (2012) proposed to adjust the constraint in

order to obtain a chi-squared distribution asymptotically. Consider $\boldsymbol{c}_i = \pi_i$ and $\boldsymbol{C} = n$. We propose to use $\boldsymbol{c}_i^* = q_i(\pi_i, \boldsymbol{g}_i(\boldsymbol{\theta}))'$ and $\boldsymbol{C}^* = (\sum_{i=1}^n q_i, \sum_{i=1}^n (q_i - 1)\boldsymbol{g}_i(\boldsymbol{\theta})\pi_i^{-1})'$, with $q_i = (1-\pi_i)^{1/2}$. Let $\widehat{m}_i^*$ be defined by

$$\widehat{m}_i^* = \left(\pi_i + \boldsymbol{\eta}^{*'}\boldsymbol{c}_i^*\right)^{-1}, \tag{13}$$

where $\boldsymbol{\eta}^*$ is such that $\sum_{i=1}^n \widehat{m}_i^* \boldsymbol{c}_i^* = \boldsymbol{C}^*$ holds. We propose to use the same empirical log-likelihood ratio function (8). The empirical log-likelihood ratio function is still defined by (8) with $\ell(m)$ given by (3). We will show that under a set of regularity conditions, $\widehat{r}(\boldsymbol{\theta}_0)$ follows asymptotically a chi-squared distribution with r degree of freedom for any sampling fractions. Hence empirical likelihood confidence regions can be constructed using (10). Berger and De La Riva Torres (2012) showed how the stratification can be taken into account by including the stratification variables within the vectors $\boldsymbol{c}_i$ and $\boldsymbol{c}_i^*$.

The q_i are finite population corrections factors proposed by Berger (2005). The q_i reduce the effect on the confidence region of units with large π_i. For example, if $\pi_i = 1$, then $\widehat{m}_i\pi_i = \widehat{m}_i^*\pi_i = 1$. This implies that this unit will have no contribution towards the empirical likelihood functions and any confidence regions. This is a natural property as this unit does not contribute towards the sampling distribution. Note that we propose to adjust the constraints by quantities which do not need to be estimated, unlike the pseudo empirical likelihood approach (Wu and Rao, 2006) which adjusts the empirical log-likelihood ratio function by a quantity that needs to be estimated (the design effect).

Conclusions

Standard confidence regions based upon the central limit theorem can perform poorly when the sampling distribution is not normal. For example, the lower bounds of a confidence region can be negative even when the parameter of interest is positive. The coverage and the tail errors can be also different from their intended levels. On the other hand, empirical likelihood confidence regions may be better in this situation, as empirical likelihood confidence regions are determined by the distribution of the data (Rao and Wu, 2009) and as the range of the parameter space is preserved. Note that the distribution of a point estimator of is not necessarily normal, and the proposed empirical likelihood approach does not rely on the normality of the point estimator.

Standard confidence regions based on the central limit theorem require normality and variance estimates which often involve linearisation or re-sampling. The proposed method does not rely on normality, variance estimates, linearisation or re-sampling, even if the parameter of interest is not linear. Empirical likelihood confidence regions can be easier to compute than standard confidence regions based on variance estimates. It provides an alternative to bootstrap, when linearisation cannot be used. The proposed approach has some advantages over the bootstrap approach. It is less computationally intensive than the bootstrap. The proposed approach naturally includes auxiliary information and the information about the sampling design. This is particularly useful under informative sampling (Pfeffermann, 2009).

REFERENCES

Berger, Y. G. (1998), "Rate of Convergence to Normal Distribution for the Horvitz-Thompson Estimator," *Journal of Statistical Planning and Inference*, 67, 209–226.

Berger, Y. G. (2005), "Variance Estimation with Highly Stratified Sampling Designs with Unequal Probabilities," *Australian and New Zealand Journal of Statistics*, 47(3), 365–373.

Berger, Y. G., and De La Riva Torres, O. (2012), "A unified theory of empirical likelihood ratio confidence intervals for survey data with unequal probabilities and non negligible sampling fractions," *Southampton Statistical Sciences Research Institute*, `http://eprints.soton.ac.uk/337688`.

Berger, Y. G., Tirari, M. E. H., and Tillé, Y. (2003), "Towards Optimal Regression Estimation in Sample Surveys," *Australian and New Zealand Journal of Statistics*, 45, 319–329.

Birch, M. W. (1963), "Maximum likelihood in three-way contingency tables," *Journal of the Royal Statistical Society*, B(25), 220–233.

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

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

Hájek, J. (1964), "Asymptotic Theory of Rejective Sampling with Varying Probabilities from a Finite Population," *The Annals of Mathematical Statistics*, 35(4), 1491–1523.

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

Hartley, H. O., and Rao, J. N. K. (1969), *A new estimation theory for sample surveys, II*, A Symposium on the Foundations of Survey Sampling held at the University of North Carolina, Chapel Hill, North Carolina: Wiley-Interscience, New York.

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.

Huber, P. J. (1981), *Robust Statistics* Wiley, New York.

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

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

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

Pfeffermann, D. (2009), "Inference under informative sampling," *Handbook of Statistics 29: Sample Surveys: Inference and Analysis. Elsevier. Danny Pfeffermann and C.R. Rao eds*, pp. 455–487.

Qin, J., and Lawless, J. (1994), "Empirical Likelihood and General Estimating Equations," *The Annals of Statistics*, 22(1), pp. 300–325.

Rao, J. N. K. (1994), "Estimating total and distribution function using auxiliary information at the estimation stage," *Journal of Official Statistics*, 10(2), 153–165.

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

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

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

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

Wilks, S. S. (1938), "Shortest Average Confidence Intervals from Large Samples," *The Annals of Mathematical Statistics*, 9(3), 166–175.

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