Biometrika (2006), **93,** 1, *pp.* 215–220

Printed in Great Britain

On Bartlett correction of empirical likelihood in the presence of nuisance parameters

By SONG XI CHEN

Department of Statistics, Iowa State University, Ames, Iowa 50011-1210, U.S.A.

songchen@iastate.edu

and HENGJIAN CUI

Department of Statistics and Financial Mathematics, Beijing Normal University, 100875, China

hjcui@bnu.eud.cn

Summary

Lazar & Mykland (1999) showed that an empirical likelihood defined by two estimating equations with a nuisance parameter need not be Bartlett-correctable. This paper shows that Bartlett correction of empirical likelihood in the presence of a nuisance parameter depends critically on the way the nuisance parameter is removed when formulating the likelihood for the parameter of interest. We establish in the broad framework of estimating functions that the empirical likelihood is still Bartlett-correctable if the nuisance parameter is profiled out given the value of the parameter of interest.

Some key words: Bartlett correction; Empirical likelihood; Estimation equation; Nuisance parameter.

1. Introduction

Since its introduction by Owen (1988, 1990), empirical likelihood has become a useful tool for conducting nonparametric or semiparametric inference. Empirical likelihood has been shown in a wide range of situations as outlined in Owen (2001) to admit limiting chi-squared distributions, which is a nonparametric version of the Wilks theorem in the context of parametric likelihood. Another key property of empirical likelihood which also resembles that of a parametric likelihood is Bartlett correction. Bartlett-correctability is a second-order property which implies that a simple mean adjustment to the likelihood ratio leads to its distributional approximation to the limiting chi-squared distribution being improved by one order of magnitude.

That the empirical likelihood is Bartlett-correctable has been established for a range of situations; see for example Hall & La Scala (1990) for the case of the mean parameter, DiCiccio et al. (1991) for smooth functions of means, Chen & Hall (1993) for quantiles, Chen (1993, 1994) for linear regression and Cui & Yuan (2001) for quantiles in the presence of auxiliary information. Jing & Wood (1996) showed that the exponentially tilted empirical likelihood for the mean is not Bartlett-correctable. Indeed, Corcoran (1998) showed that Kullback–Leibler divergence is the unique member of a large class of divergence measures that produces Bartlett-correctable empirical likelihood statistics. However, Lazar & Mykland (1999) showed that in some circumstances, where the empirical likelihood is defined by two estimating equations and when a nuisance parameter is present, even the use of Kullback–Leibler divergence can fail to guarantee Bartlett correctability.

In contrast to the result of Lazar & Mykland (1999), Chen (1994) had earlier proved that empirical likelihood is Bartlett-correctable in the context of simple linear regression when one coefficient is treated as a nuisance parameter. It appears that the result obtained by Lazar & Mykland (1999) is due to absence of a regular Edgeworth expansion for the signed square root of the empirical likelihood ratio.

In the present paper, we confirm that the result of Chen (1994) holds in general. We consider the Bartlett property in a broader situation where there are r estimating equations and the dimension of the nuisance parameter is p, with $p<r$, which is within the framework of the empirical likelihood for generalised estimating equations introduced in Qin & Lawless (1994). It is found that, if the nuisance parameter is profiled out given the parameter of interest, the empirical likelihood is still Bartlett-correctable. This indicates that the Bartlett-correctability of the empirical likelihood is dependent on the method of nuisance-parameter removal when formulating the likelihood for the parameter of interest, rather than on any fundamental differences between estimating equations and the smooth function of means. It is expected that a corresponding result holds for parametric likelihood as well, namely that the Bartlett correction property only holds in general when the nuisance parameter is 'profiled out'.

2. Empirical likelihood with nuisance parameters

Consider a random vector X with unknown distribution function F which depends on an r-dimensional parameter $(\theta, \psi) \in R^{r-p} \times R^p$. Here interest is in the parameter θ with ψ as a p-dimensional nuisance parameter. We assume that the parameter (θ, ψ) is defined by r functionally unbiased estimating equations $g^j(x, \theta, \psi)$, for $j=1, 2, \ldots, r$, such that $E\{g^j(X, \theta_0, \psi_0)\}=0$, where (θ_0, ψ_0) is the true parameter value. In particularly, we define

$$g(X, \theta, \psi)=(g^1(X, \theta, \psi), g^2(X, \theta, \psi), \ldots, g^r(X, \theta, \psi))'.$$

Assume that $\{X_1, X_2, \ldots, X_n\}$ are independent and identically distributed observations drawn from F. We let $V=\text{cov}\{g(X_i, \theta_0, \psi_0)\}$ and we assume the following regularity conditions.

Condition 1. The $r \times r$ matrix V is positive definite and the rank of the $r \times p$ matrix $E\{\partial g(X, \theta_0, \psi_0)/\partial\psi\}$ is p.

Condition 2. For any j $(1 \leqslant j \leqslant p)$, all the fourth-order partial derivatives of $g^j(x, \theta_0, \psi)$ with respect to ψ are continuous in a neighbourhoood of θ_0 and are bounded by some integrable function $G(x)$ in that neighbourhood.

Condition 3. We require that $E\|g(X, \theta_0, \psi_0\}\|^{15}<\infty$ and the characteristic function of $g(X, \theta_0, \psi_0)$ satisfy the Cramér condition: $\limsup_{|t|\to\infty}|E[\exp\{it'g(X, \theta_0, \psi_0)\}]|<1$.

To simplify derivations, we first rotate the original estimating functions by defining

$$w_i(\theta, \psi) := TV^{-1/2} g(X_i, \theta, \psi),$$

where T is an $r \times r$ orthogonal matrix such that

$$TV^{-1/2} E\left\{\frac{\partial g(X, \theta_0, \psi_0)}{\partial\psi}\right\} U=(\Lambda \;\; 0)'_{r\times p},$$

$U=(u^{kl})_{p\times p}$ is an orthogonal matrix and $\Lambda=\text{diag}(\lambda_1, \ldots, \lambda_p)$ is a nonsingular $p \times p$ diagonal matrix. Furthermore, define $\Omega=(\omega^{kl})_{p\times p}=U\Lambda^{-1}$.

Let $p_1, \ldots, p_n$ be nonnegative weights allocated to the observations. Then the empirical likelihood for the parameter $\{\theta, \psi)$ is

$$L(\theta, \psi)=\prod_{i=1}^{n} p_i,$$

subject to $\sum p_i=1$ and the constraints $\sum p_i w_i(\theta, \psi)=0$. Let $\ell(\theta, \psi)=-2\log\{n^n L(\theta, \psi)\}$ be the log empirical likelihood ratio. Standard empirical likelihood derivations show that

$$\ell(\theta, \psi)=2\sum_{i=1}^{n}\log\{1+\lambda'(\theta, \psi) w_i(\theta, \psi)\},$$

where $\lambda(\theta, \psi)$ satisfies

$$n^{-1} \sum_{i=1}^{n} \frac{w_i(\theta, \psi)}{1 + \lambda'(\theta, \psi) w_i(\theta, \psi)} = 0. \tag{1}$$

To obtain the empirical likelihood ratio at θ_0, we profile out the nuisance parameter ψ. To simplify notation, write $w_i(\psi) = w_i(\theta_0, \psi)$, let $\tilde{\psi} := \tilde{\psi}(\theta_0)$ be the minimiser of $\ell(\theta_0, \psi)$ given $\theta = \theta_0$ and let $\tilde{\lambda} = \lambda(\theta_0, \tilde{\psi})$ be the solution of (1) at $(\theta_0, \tilde{\psi})$. Let $(\hat{\theta}, \hat{\psi})$ be the maximum empirical likelihood estimator of (θ, ψ). Since the number of estimating functions is equal to the dimension of (θ, ψ), then $\ell(\hat{\theta}, \hat{\psi}) = 0$. This means that the log empirical likelihood ratio for θ_0 is just

$$\ell(\theta_0) := \ell\{\theta_0, \tilde{\psi}(\theta_0)\} = 2 \sum_{i=1}^{n} \log\{1 + \tilde{\lambda}' w_i(\tilde{\psi})\}. \tag{2}$$

In order to develop an expansion for $\ell(\theta_0)$, we need to derive expansions for $\tilde{\lambda}$ and $\tilde{\psi}$ first. We note from Qin & Lawless (1994) that $(\tilde{\lambda}, \tilde{\psi})$ are the solutions of

$$Q_{1n}(\lambda, \psi) := n^{-1} \sum_{i=1}^{n} \frac{w_i(\psi)}{1 + \lambda' w_i(\psi)} = 0, \tag{3}$$

$$Q_{2n}(\lambda, \psi) := n^{-1} \sum_{i=1}^{n} \frac{\{\partial w_i(\psi)/\partial \psi\}' \lambda}{1 + \lambda' w_i(\psi)} = 0. \tag{4}$$

Let $\eta = (\lambda', \psi')'$, $\eta_0 = (0, \psi_0)$,

$$Q(\eta) = \begin{pmatrix} Q_{1n}(\eta) \\ Q_{2n}(\eta) \end{pmatrix}, \quad S = E\frac{\partial Q(0, \psi_0)}{\partial \eta} = \begin{pmatrix} -I & S_{12} \\ S_{21} & 0 \end{pmatrix},$$

where $S_{21} = U(\Lambda, 0)$ and $S_{12} = S'_{21}$. To facilitate easy expressions, we standardise Q to $\Gamma(\eta) = S^{-1} Q(\eta)$. Let $w_i^j(\psi)$ and $\Gamma^j(\eta)$ denote respectively the jth components of $w_i(\psi)$ and $\Gamma(\eta)$. The following $\alpha - A$ system of notation was used by DiCiccio et al. (1991):

$$\alpha^{j_1 \ldots j_k} = E\{w^{j_1}(\psi_0) \ldots w^{j_k}(\psi_0)\}, \quad A^{j_1 \ldots j_k} = n^{-1} \sum_{i=1}^{n} w^{j_1}(\psi_0) \ldots w^{j_k}(\psi_0) - \alpha^{j_1 \ldots j_k}.$$

We also need to define

$$\beta^{j, j_1 \ldots j_k} = E\left\{\frac{\partial^k \Gamma^j(0, \psi_0)}{\partial \eta_{j_1} \ldots \partial \eta_{j_k}}\right\}, \quad B^{j, j_1 \ldots j_k} = \frac{\partial^k \Gamma^j(0, \psi_0)}{\partial \eta_{j_1} \ldots \partial \eta_{j_k}} - \beta^{j, j_1 \ldots j_k},$$

$$\gamma^{j, j_1 \ldots j_l; k, k_1 \ldots k_m; \ldots; p, p_1 \ldots p_n} = E\left\{\frac{\partial^l w_i^j(\psi_0)}{\partial \psi^{j_1} \ldots \partial \psi^{j_l}} \frac{\partial^m w_i^k(\psi_0)}{\partial \psi^{k_1} \ldots \partial \psi^{k_m}} \cdots \frac{\partial^n w_i^p(\psi_0)}{\partial \psi^{p_1} \ldots \partial \psi^{p_n}}\right\},$$

$$C^{j, j_1 \ldots j_l; k, k_1 \ldots k_m; \ldots; p, p_1 p_n} = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial^l w_i^j(\psi_0)}{\partial \psi^{j_1} \ldots \partial \psi^{j_l}} \frac{\partial^m w_i^k(\psi_0)}{\partial \psi^{k_1} \ldots \partial \psi^{k_m}} \cdots \frac{\partial^n w_i^p(\psi_0)}{\partial \psi^{p_1} \ldots \partial \psi^{p_n}} - \gamma^{j, j_1 \ldots j_l; k, k_1 \ldots k_m; \ldots; p, p_1 \ldots p_n}.$$

3. Main results

We only provide a brief sketch of the steps taken to obtain the Bartlett correction for this general case. A full account of the technical details is given in an Iowa State University technical report available from the authors.

We first need to develop an expansion of the empirical likelihood ratio $\ell(\theta_0)$. Derivations given in our technical report show that the Lagrange multiplier and the nuisance parameter at θ_0 admit the expansions

$$\tilde{\lambda}^j = -B^j + B^{j,q}B^q - \tfrac{1}{2}\beta^{j,uq}B^uB^q - B^{j,u}B^{u,q}B^q + \tfrac{1}{2}\beta^{u,qs}B^{j,u}B^qB^s + \beta^{j,uq}B^{u,s}B^sB^q$$
$$-\tfrac{1}{2}\beta^{j,uq}\beta^{u,st}B^sB^tB^q - \tfrac{1}{2}B^{j,uq}B^uB^q + \tfrac{1}{6}\beta^{j,uqs}B^uB^qB^s + O_p(n^{-2}), \quad (5)$$

$$\tilde{\psi}^k = -B^{r+k} + B^{r+k,q}B^q - \tfrac{1}{2}\beta^{r+k,uq}B^uB^q - B^{r+k,u}B^{u,q}B^q + \tfrac{1}{2}\beta^{u,qs}B^{r+k,u}B^qB^s$$
$$+\beta^{r+k,uq}B^{u,s}B^sB^q - \tfrac{1}{2}\beta^{r+k,uq}\beta^{u,st}B^sB^tB^q - \tfrac{1}{2}B^{r+k,uq}B^uB^q + \tfrac{1}{6}\beta^{r+k,uqs}B^uB^qB^s + O_p(n^{-2}), \quad (6)$$

for $j \in \{1, \ldots, r\}$, $k \in \{1, \ldots, p\}$ and $q, s, t, u \in \{1, \ldots, r+p\}$. Here and in the rest of the paper, we use the tensor notation in which repetition of a superscript implies summation over that superscript; see McCullagh (1987) for a systematic description.

Substitution of the above expansions in (2) gives

$$n^{-1}\ell(\theta_0) = A^{p+a}A^{p+a} - A^{p+a\,p+b}A^{p+a}A^{p+b} - 2\omega^{kl}C^{p+a,k}A^{p+a}A^l$$
$$+2\gamma^{p+a;p+b,k}\omega^{kl}A^{p+a}A^{p+b}A^l + \tfrac{2}{3}\alpha^{p+a\,p+b\,p+c}A^{p+a}A^{p+b}A^{p+c}$$
$$+\gamma^{p+a,kl}\omega^{km}\omega^{ln}A^{p+a}A^mA^n + A^{ji}B^{i,q}B^qB^j[2, i, j] - B^{j,u}B^{j,q}B^uB^q$$
$$-2C^{j,k}B^{j,q}B^{r+k}B^q + \gamma^{j,kl}B^{r+k}B^{r+l}B^{j,q}B^q + 2\gamma^{j,kl}B^jB^{r+l}B^{r+k,q}B^q$$
$$-2\gamma^{j;i,l}(B^jB^iB^{r+l,q}B^q + B^{r+l}B^iB^{j,q}B^q[2, j, i]) + 2\alpha^{jih}B^jB^iB^{h,q}B^q$$
$$+(\tfrac{1}{2}\beta^{j,uq}\beta^{r+k,st}\gamma^{j,k} - \tfrac{1}{4}\beta^{j,uq}\beta^{j,st})B^uB^qB^sB^t - \tfrac{1}{2}\gamma^{j,kl}\beta^{j,uq}B^uB^qB^{r+l}B^{r+k}$$
$$+(\gamma^{i;j,k}\beta^{i,uq} + \gamma^{j;i,k}\beta^{i,uq} - \gamma^{j,kl}\beta^{r+l,pq})B^uB^qB^jB^{r+k} + 2\gamma^{j;i;h,k}B^jB^iB^hB^{r+k}$$
$$-(\gamma^{j;i,lk} + \gamma^{j,l;i,k})B^jB^iB^{r+a}B^{r+k} + \tfrac{1}{3}\gamma^{j,klm}B^jB^{r+k}B^{r+l}B^{r+m}$$
$$-\tfrac{1}{2}\alpha^{jihg}B^jB^iB^hB^g + (\gamma^{j;i,l}\beta^{r+l,uq} - \alpha^{jih}\beta^{h,uq})B^jB^iB^uB^q$$
$$-C^{j,kl}B^jB^{r+k}B^{r+l} + 2C^{j;i,l}B^jB^iB^{r+l} - \tfrac{2}{3}A^{jih}B^jB^iB^h + O_p(n^{-5/2}), \quad (7)$$

where $h, i, j \in \{1, \ldots, r\}$, $k, l, m \in \{1, \ldots, p\}$, $a, b, c \in \{1, \ldots, r-p\}$ and $q, s, t, u \in \{1, \ldots, r+p\}$. When $p = 0$, in which case there is no nuisance parameter, the above expansion takes the form given in DiCiccio et al. (1991) with $S = -I$, $B^j = -A^j$, $B^{i,j} = A^{ij}$, $\beta^{j,j_1j_2} = -2\alpha^{jj_1j_2}$ and all the γ's and C's equal to zero.

Let $R = R_1 + R_2 + R_3$ be a signed square root decomposition of $n^{-1}\ell(\theta_0)$ such that

$$n^{-1}\ell(\theta_0) = R^qR^q + O(n^{-5/2}),$$

where $R_j = O_p(n^{-j/2})$ for $j = 1$, 2 and 3. Clearly, R_1 and R_2 can be determined from the terms of $O_p(n^{-1})$ and $O_p(n^{-3/2})$ respectively in (7). To be specific, for $a, b, c, d, e \in \{1, \ldots, r-p\}$ and $l, k, m, n, o, v, m', n' \in \{1, \ldots, p\}$,

$$R_1^q = R_2^q = 0 \quad (q \leqslant p), \quad R_1^{p+a} = A^{p+a},$$

$$R_2^{p+a} = -\tfrac{1}{2}A^{p+a\,p+b}A^{p+a} - \omega^{kl}C^{p+a,k}A^l + \gamma^{p+a;p+b,k}\omega^{kl}A^{p+b}A^l$$
$$+\tfrac{1}{3}\alpha^{p+a\,p+b\,p+c}A^{p+b}A^{p+c} + \tfrac{1}{2}\gamma^{p+a,kl}\omega^{km}\omega^{ln}A^mA^n.$$

The expression for R_3 is obtained after removing terms induced by $R_1^{p+a}R_1^{p+a}$ and $R_2^{p+a}R_2^{p+a}$ from (7); see our technical report for details.

The key to checking the Bartlett correctability of the empirical likelihood is to examine if the third- and fourth-order joint cumulants of the signed square root R are of the orders of n^{-3} and n^{-4} respectively. This is the path taken by DiCiccio et al. (1991), Chen (1994), Jing & Wood (1996) and Lazar & Mykland (1999). A formal establishment of the Bartlett correction can be then made by developing Edgeworth expansions for the empirical likelihood ratio under Conditions 1–3. After a rather lengthy calculation documented in our technical report, it is found that the joint third- and fourth-order cumulants of R are indeed of the orders of n^{-3} and n^{-4} respectively. Hence, the empirical likelihood is still Bartlett-correctable in this general case of estimating equations with nuisance parameters.

4. Discussion

The established Bartlett correction provides the theoretical justification for an empirical Bartlett correction to the confidence regions. Let c_α be the upper α quantile of the χ^2_{r-p} distribution. Based on Wilks' theorem, a $1-\alpha$ level confidence region for θ is $I_\alpha = \{\theta | \ell(\theta) \leqslant c_\alpha\}$. It may be shown by developing Edgeworth expansions using the derived cumulants given in our technical report that the coverage error of I_α is of $O(n^{-1})$. As $\ell(\theta_0)$ is Bartlett-correctable, it may be shown under Conditions 1–3 that

$$\text{pr}\{\ell(\theta_0) < c_\alpha(1 + n^{-1}B_c)\} = 1 - \alpha + O(n^{-2}),$$

where B_c is the Bartlett factor whose expression is given in our report. The rather complicated form of B_c means that the direct plug-in method may not work efficiently for its estimation. We propose a method based on the following bootstrap procedure.

Step 1. Generate bootstrap resamples of size n by sampling with replacement from the original sample $\{X_i\}_{i=1}^n$; compute the empirical likelihood ratio $\ell^*(\hat{\theta})$ based on the resamples, where $\hat{\theta}$ is the global maximum empirical likelihood estimator of θ based on the original sample.

Step 2. Repeat Step 1 B times to obtain $\ell^{*1}(\hat{\theta}), \ldots, \ell^{*B}(\hat{\theta})$ and $B^{-1}\sum_{b=1}^{B} \ell^{*b}(\hat{\theta})$, which is the bootstrap estimator of $E\{\ell(\hat{\theta})\}$.

The bootstrap estimator of $1 + n^{-1}B_c$ is $\tau = (r-p)^{-1}B^{-1}\sum_{b=1}^{B} \ell^{*b}(\hat{\beta})$, which can be used to construct $I_{\alpha,bc} = \{\theta | \ell(\theta) \leqslant \tau c_\alpha\}$, the Bartlett-corrected confidence region. It can be shown, based on the Edgeworth expansion mentioned earlier, that the coverage error of this Bartlett-corrected region is $O(n^{-3/2})$, which is one order of magnitude smaller than that of I_α.

Acknowledgement

The authors would like to thank Professor Nicole Lazar for informative discussions, and two anonymous reviewers, the associate editor and the editor for comments and suggestions which improved the presentation of the paper. The project was supported by grants from the National University of Singapore and Iowa State University, and by grants from National Natural Science Foundation of China and the Chinese Research Fund for Doctoral Programme of Higher Education.

References

Chen, S. X. (1993). On the coverage accuracy of empirical likelihood regions for linear regression model. *Ann. Inst. Statist. Math.* **45**, 621–37.

Chen, S. X. (1994). Empirical likelihood confidence intervals for linear regression coefficients. *J. Mult. Anal.* **49**, 24–40.

Chen, S. X. & Hall, P. (1993). Smoothed empirical likelihood confidence intervals for quantiles. *Ann. Statist.* **21**, 1166–81.

Corcoran, S. A. (1998). Bartlett adjustment of empirical discrepancy statistics. *Biometrika* **85**, 967–72.

Cui, H. J. & Yuan, X. J. (2001). Smoothed empirical likelihood confidence interval for quantile in the partial symmetric auxiliary information. *J. Syst. Sci. Math. Sci.* **21**, 172–81.

DiCiccio, T. J., Hall, P. & Romano, J. P. (1991). Empirical likelihood is Bartlett-correctable. *Ann. Statist.* **19**, 1053–61.

HALL, P. & LA SCALA, B. (1990). Methodology and algorithms of empirical likelihood. *Int. Statist. Rev.* **58**, 109–27.

JING, B. Y. & WOOD, A. T. A. (1996). Exponential empirical likelihood is not Bartlett-correctable. *Ann. Statist.* **24**, 365–9.

LAZAR, N. A. & MYKLAND, P. A. (1999). Empirical likelihood in the presence of nuisance parameters. *Biometrika* **86**, 203–11.

MCCULLAGH, P. (1987). *Tensor Methods in Statistics.* London: Chapman and Hall.

OWEN, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* **75**, 237–49.

OWEN, A. B. (1990). Empirical likelihood ratio confidence regions. *Ann. Statist.* **18**, 90–120.

OWEN, A. B. (2001). *Empirical Likelihood.* Boca Raton, FL: Chapman and Hall/CRC.

QIN, J. & LAWLESS, J. (1994). Empirical likelihood and general estimation equations. *Ann. Statist.* **22**, 300–25.

[*Received August* 2004. *Revised March* 2005]