Statistica Sinica **21** (2011), 1767-1783
doi:http://dx.doi.org/10.5705/ss.2009.117

CONSTRUCTING NONPARAMETRIC LIKELIHOOD CONFIDENCE REGIONS WITH HIGH ORDER PRECISIONS

Xiao Li[1], Jiahua Chen[2], Yaohua Wu[1] and Dongsheng Tu[3]

[1] *University of Science and Technology of China,*
[2] *University of British Columbia and* [3] *Queen's University*[3]

Abstract: Empirical likelihood is a natural tool for nonparametric statistical inference, and a member of nonparametric likelihoods. Inferences based on this class of likelihoods have the same first order asymptotic properties. One member of the class, exponential tilting likelihood, has been found to be stable to model mis-specification but is not as efficient as empirical likelihood. Exponentially tilted empirical likelihood, also called exponential empirical likelihood, was proposed to achieve both stability and efficiency. Unlike empirical likelihood, however, the hybrid likelihood is not Bartlett correctable, and the precision of its confidence regions is compromised when the sample size is not large. We introduce a novel adjustment procedure and show that it attains the high order precision that is not attained by the usual Bartlett correction. Simulation results confirm the improved precision in coverage probabilities.

Key words and phrases: Bartlett correction, edgeworth expansion, empirical likelihood, estimating equation, exponential empirical likelihood, exponential tilting, Wilks' theorem.

1. Introduction

Empirical likelihood (EL) as proposed by Owen (1990, 2001) is a popular tool for constructing nonparametric confidence regions and making other nonparametric inferences. Under mild conditions on the nonparametric distribution family, the profile empirical likelihood ratio statistic has a chi-square limiting distribution (Wilks' Theorem). The confidence regions based on this are transformation invariant, range respecting, and free of the burden of estimating scaling parameters, but often have lower than nominal coverage probabilities. However, DiCiccio, Hall and Romano (1991) found that the precision of the chi-square approximation can be substantially improved by a Bartlett correction.

Empirical likelihood belongs to a class of nonparametric likelihoods called empirical discrepancy (Baggerly (1998); Corcoran (1998); Schennach (2007)). Most nonparametric likelihoods have the same first-order asymptotic properties. One member of this class, exponential tilting likelihood (ET), has been found to be stable to model mis-specification. However, ET loses some optimality properties such as efficiency. A hybrid method, exponentially tilted empirical likelihood, also called exponential empirical likelihood (EEL), has been proposed as a procedure that is both stable and efficient (Schennach (2007)). The hybrid procedure, however, is not Bartlett correctable (Jing and Wood (1996); Corcoran (1998)) in the sense of DiCiccio, Hall and Romano (1991). Without Bartlett correction, the chi-square-approximation based confidence intervals/regions often have lower than nominal coverage probabilities.

Recently, Chen, Mulayath and Abraham (2008) proposed an adjustment technique for empirical likelihood. Empirical likelihood works well for inference problems in nonparametric distribution families characterized by estimating functions (Qin and Lawless (1994)). When the equation constraints defining the profile empirical likelihood have no solutions, however, the likelihood becomes undefined and a convention of assigning a 0-value is adopted. This convention does not affect the asymptotic properties, but can be a nuisance in applications. The adjustment proposed by Chen, Mulayath and Abraham (2008) eliminates this problem and has additional benefits. In particular, Liu and Chen (2010) show that by choosing an appropriate level of adjustment, the precision of the chi-square approximation can be improved to $O(n^{-2})$, the same as that attained by the Bartlett correction.

In this paper, we demonstrate further benefits. While exponential empirical likelihood is not Bartlett correctable, a specific adjustment is found that enables it to attain high order precision. Consequently, the adjusted exponential empirical likelihood is more stable and equally efficient, and has high order precision in constructing confidence regions.

2. Exponential Empirical Likelihood

Assume that we have a set of independent and identically distributed (i.i.d.) vector-valued observations $y_1, y_2, \ldots, y_n$ from an unknown distribution function $F(\cdot)$. Let $\theta = \theta(F)$ be a q-dimensional parameter, defined as the unique solution to some estimating equation $E\{g(Y;\theta)\} = 0$, where $g(\cdot)$ is a $p \geq q$ dimensional function. Empirical likelihood (Owen (1990)) is a natural tool for constructing nonparametric confidence regions. Let Y be a random vector with distribution F and write $p_i = P(Y = y_i)$, $i = 1, \ldots, n$. When there are no ties in y_i, the nonparametric likelihood of F is given by $L_n(F) = \prod_{i=1}^n p_i$. The likelihood is maximized when $p_i = n^{-1}$, or at $F(y) = F_n(y) = n^{-1}\sum_{i=1}^n I(y_i \leq y)$, where I is the usual indicator function. Thus the empirical distribution is the nonparametric maximum likelihood estimator of F. In general, because F is a distribution function, we must have $p_i \geq 0$ and $\sum_{i=1}^n p_i \leq 1$. Without loss of generality (Owen (2001)), we may simply require $\sum_{i=1}^n p_i = 1$.

The profile empirical likelihood is defined as

$$L_n(\theta) = \sup\Big\{\prod_{i=1}^{n} p_i : \sum_{i=1}^{n} p_i = 1; \sum_{i=1}^{n} p_i g(y_i;\theta) = 0; p_i > 0\Big\},$$

and the profile empirical log-likelihood ratio function as

$$W_{EL}(\theta) = -2\sup\Big\{\sum_{i=1}^{n}\log(np_i) : \sum_{i=1}^{n} p_i = 1; \sum_{i=1}^{n} p_i g(y_i;\theta) = 0\Big\}.$$

Let θ_0 be the parameter value that solves $E\{g(Y;\theta)\} = 0$, then $W_{EL}(\theta_0) \to \chi^2_p$ in distribution as $n \to \infty$ under some mild conditions. This is Wilks' Theorem and the basis for constructing confidence regions for θ.

Empirical likelihood is one of many possible empirical dispersion measures between the unknown distribution F and the empirical distribution F_n. Most of them can be utilized to make nonparametric inferences in the same way. In this paper, we focus on exponential tilting likelihood, or ET (Efron (1981)). For brevity, we directly introduce its profile log-likelihood ratio function

$$W_{ET}(\theta) = 2\sup\Big\{\sum_{i=1}^{n}(np_i)\log(np_i) : \sum_{i=1}^{n} p_i = 1; \sum_{i=1}^{n} p_i g(y_i;\theta) = 0\Big\}.$$

Putting $g_i = g(y_i;\theta)$, the solution in p_i to ET is

$$p_i(\lambda) = \frac{\exp(\lambda^T g_i)}{\sum_{j=1}^{n}\exp(\lambda^T g_j)}$$

with the vector λ being the solution of $\sum_{i=1}^{n} p_i(\lambda) g_i = 0$.

The empirical likelihood procedure can be thought of as maximizing a likelihood and the exponential tilting likelihood as maximizing an entropy. EL inherits the efficiency of the usual parametric likelihood; ET, on the other hand, is stable to model mis-specification in the form of $E\{g(Y;\theta)\} \neq 0$.

To take advantage of both likelihoods, the (profile) log exponential empirical likelihood, or EEL, ratio is

$$W_{EEL}(\theta) = -2\sum_{i=1}^{n}\log\{np_i(\lambda)\},$$

with $p_i(\lambda)$ as defined previously. Under mild conditions, $W_{EEL}(\theta_0) \to \chi^2_p$ in distribution. Thus, a nonparametric confidence region for θ can be constructed accordingly. EEL is as efficient as, and more stable than, empirical likelihood (Schennach (2007)). However, unlike EL, EEL is not Bartlett correctable

(Jing and Wood (1996)). Searching for other methods to attain high order approximation precisions is an interesting research problem.

When the convex hull of $\{g_i; i = 1, \ldots, n\}$ does not contain 0, there is no solution to $\sum_{i=1}^{n} p_i g_i = 0$ in p_i such that $p_i > 0$ and $\sum p_i = 1$. Consequently, the EL, ET, and EEL values are all undefined, forcing researchers to assign a conventional 0-value. A more attractive alternative, proposed by Chen, Mulayath and Abraham (2008), is to create some pseudo-values g_{n+i}, $i = 1, \ldots, m$, such that the convex hull of $\{g_i; i = 1, \ldots, n+m\}$ contains 0. Applying the same idea to EEL, the adjusted (profile) log-likelihood ratio would be

$$W_A(\theta) = -2 \sum_{i=1}^{n+m} \log\{(n+m)p_i(\lambda)\}, \tag{2.1}$$

such that

$$p_i(\lambda) = \frac{\exp(\lambda^T g_i)}{\sum_{j=1}^{n+m} \exp(\lambda^T g_j)},$$

with the vector λ being the solution of $\sum_{i=1}^{n+m} p_i(\lambda) g_i = 0$. Under certain conditions, the first order asymptotic properties remain unaltered.

By careful choice of g_{n+1} and g_{n+2} with $m = 2$, the new $W_A(\theta)$ is always defined and we show that its chi-square approximation has a precision of order $O(n^{-2})$. We now introduce some notation before stating the main result.

Let Σ_0 be the variance-covariance matrix of $g(Y;\theta_0)$. Let g^r be the rth component of g, and similarly for other vectors. For any set of positive integers, $r, s, \ldots, t$, we write

$$\alpha^{rs\cdots t} = E[\{\Sigma_0^{-1/2} g\}^r \{\Sigma_0^{-1/2} g\}^s \cdots \{\Sigma_0^{-1/2} g\}^t]$$

for the standardized moments of g. Correspondingly, we let

$$A^{rs\cdots t} = n^{-1} \sum_{i=1}^{n} [\{\Sigma_0^{-1/2} g_i\}^r \{\Sigma_0^{-1/2} g_i\}^s \cdots \{\Sigma_0^{-1/2} g_i\}^t] - \alpha^{rs\cdots t}$$

be the centered sample moments. In addition, we adopt the tensor notation of DiCiccio, Hall and Romano (1991): when an index appears more than once in a term, the term represents the sum over the range of that index.

Theorem 1. *Let $y_1, \ldots, y_n$ be a set of independent and identically distributed vector observations from an unknown distribution F, and $\theta = \theta(F)$ be a q-dimensional parameter. Assume that*

(a) *F satisfies the generalized estimating equation $E\{g(Y, \theta)\} = 0$ for some p-dimensional estimating function g;*

(b) *the characteristic function of g satisfies Cramér's condition*

$$\lim \sup_{\|t\|\to\infty} E\|\exp\{\mathbf{i}t^T g(Y,\theta)\}\| < 1;$$

(c) $E\{\|g(Y,\theta)\|^{18}\} < \infty$.

At the true parameter θ_0, if $g^r_{n+1} = n(\alpha^{rst}\alpha^{tuv}A^sA^uA^v - \alpha^{rstu}A^sA^tA^u + A^rA^sA^s)/8$ and $g^r_{n+2} = -bA^r/2$, where $b = (\alpha^{rrss}/2 - \alpha^{rst}\alpha^{rst}/3)/q$, then, for W_A at (2.1), we have

$$P\{W_A(\theta_0) \le x\} = P(\chi^2_p \le x) + O(n^{-2}),$$

where χ^2_p is a chi-square distributed random variable with p degrees of freedom.

The proof is in the next section. Replacing the two pseudo-observations in the theorem by a single pseudo-observation $g_{n+1} + g_{n+2}$ does not change the precision. Because g_{n+2} is in the opposite direction of A, the current arrangement ensures the existence of a solution to the equation constraints. The moment requirement (c) is required because six cumulants of g^3 are needed for Edgeworth expansion.

In applications, we must evaluate W_A at all potential θ values. Yet at values far from θ_0, the size of $g_{n+1}(\theta)$ far exceeds what $O_p(n^{-1/2})$ might indicate. This results in bounded $W_A(\theta)$ and hence nonsensical confidence intervals, particularly when the data are from a severely skewed population. To prevent this problem, we scale down A^r. Let $\hat{\theta}$ be the corresponding maximum nonparametric likelihood estimate of θ, and

$$\nu^r = \left[\frac{\sum_{i=1}^n \{g^r_i(\hat{\theta})\}^2}{\sum_{i=1}^n \{g^r_i(\theta)\}^2}\right]^{1/2}.$$

We then replace A^r by $\nu^r A^r$ in g_{n+1} and g_{n+2}. Furthermore, we counter large g^r_{n+1} values by replacing them with

$$g^r_{n+1}(\theta)\exp[-\sqrt{n}\{A^r(\theta)\}^2].$$

Note that at $\theta = \theta_0$, we have $\nu = O_p(1 + n^{-1/2})$ and $\exp[-\sqrt{n}\{A^r(\theta_0)\}^2] = 1 + O_p(n^{-1/2})$. In addition, we also bound the estimated b by $\log n$. Because these changes are in high order terms, they do not invalidate Theorem 1 at $\theta = \theta_0$.

It is natural to ask whether the same technique could be used to enable the unadjusted ET likelihood to attain high order precision. Our investigation reveals that the answer is positive. The result for multidimensional θ, however, is messy; here we consider only a scale parameter θ.

Theorem 2. *Under the same conditions as in Theorem* 1*, let* $p = q = 1$*. Let* $g_{n+1} = n\alpha_3 A_1^2/6$*, and* $g_{n+2} = (-1/8 - \alpha_4/24 + \alpha_3^2/12)nA_1^3 - (\alpha_4/4 + \alpha_3^2/24)A_1$*. If*

$$W_{AET}(\theta) = 2\sum_{i=1}^{n+2}(n+2)p_i(\lambda)\log[(n+2)p_i(\lambda)] \tag{2.2}$$

is such that $p_i(\lambda) = \exp(\lambda g_i)/\sum_{j=1}^{n+2}\exp(\lambda g_j)$*, with the scalar* λ *being the solution of* $\sum_{i=1}^{n+2} p_i(\lambda)g_i = 0$*, then* $Pr\{W_{AET}(\theta_0) \le x)\} = Pr(\chi_1^2 \le x) + O(n^{-2})$*.*

The proof is in the next section.

3. Proofs

Proof of Theorem 1. Without loss of generality, assume $\Sigma_0 = I$, the identity matrix. Then $\alpha^{rs} = \delta^{rs}$, where δ is the Kronecker delta symbol. Before any pseudo-values are added to the data set, the Lagrange multiplier λ in EEL satisfies $\sum_{i=1}^{n} p_i(\lambda)g_i = 0$, and $\|\lambda\| = O_p(n^{-1/2})$. Under the model assumption, $\alpha^r = 0$, and we can easily verify that $\|g_{n+j}\| = O(n^{-1/2})$, $j = 1, 2$. Consequently, it can easily be seen that the first order asymptotic properties of EEL are preserved. In particular, denote the new Lagrange multiplier that solves $\sum_{i=1}^{n+2} p_i(\lambda)g_i = 0$ as $\tilde{\lambda}$. We still have $\|\tilde{\lambda}\| = O_p(n^{-1/2})$.

Applying routine Taylor expansion procedures, or taking advantage of the expansions in DiCiccio, Hall and Romano (1991) or Jing and Wood (1996), leads to the decomposition $\tilde{\lambda} = \tilde{\lambda}_1 + \tilde{\lambda}_2 + \tilde{\lambda}_3 + O_p(n^{-2})$, where

$$\begin{aligned}
\tilde{\lambda}_1^r &= -A^r;\\
\tilde{\lambda}_2^r &= A^s A^{rs} - \frac{1}{2}\alpha^{rst}A^s A^t;\\
\tilde{\lambda}_3^r &= \frac{1}{2}\alpha^{stu}A^t A^u A^{rs} + \alpha^{rst}A^s A^u A^{tu} + \frac{1}{6}\alpha^{rstu}A^s A^t A^u - A^t A^{rs} A^{st}\\
&\quad -\frac{1}{2}\alpha^{rst}\alpha^{suv}A^u A^v A^t - \frac{1}{2}A^s A^t A^{rst} - \frac{1}{n}(g_{n+1}^r + g_{n+2}^r)
\end{aligned}$$

and $\|\tilde{\lambda}_j\| = O_p(n^{-j/2})$ for $j = 1, 2, 3$.

Substituting the expansion of $\tilde{\lambda}$ into $W_A(\theta_0)$, we get

$$W_A(\theta_0) = n\widetilde{R}^T\widetilde{R} + O_p(n^{-3/2}) = n(\widetilde{R}_1 + \widetilde{R}_2 + \widetilde{R}_3)^T(\widetilde{R}_1 + \widetilde{R}_2 + \widetilde{R}_3) + O_p(n^{-3/2}),$$

with

$$
\begin{aligned}
\widetilde{R}_1^r &= A^r,\\
\widetilde{R}_2^r &= \frac{1}{3}\alpha^{rst}A^sA^t - \frac{1}{2}A^sA^{rs},\\
\widetilde{R}_3^r &= \frac{3}{8}A^tA^{rs}A^{st} - \frac{5}{12}\alpha^{rst}A^sA^uA^{tu} - \frac{5}{12}\alpha^{stu}A^tA^uA^{rs} + \frac{1}{3}A^sA^tA^{rst}\\
&\quad -\frac{1}{8}A^rA^sA^s - \frac{1}{8}\alpha^{rstu}A^sA^tA^u + \frac{23}{72}\alpha^{rst}\alpha^{tuv}A^sA^uA^v + \frac{1}{n}(g_{n+1}^r + g_{n+2}^r).
\end{aligned}
$$

To make sense of the above expansion, compare it to the expansion of empirical likelihood given by DiCiccio, Hall and Romano (1991):

$$W_{EL}(\theta_0) = n(R_1 + R_2 + R_3)^T(R_1 + R_2 + R_3) + O_p(n^{-3/2}),$$

where

$$
\begin{aligned}
R_1^r &= A^r,\\
R_2^r &= \frac{1}{3}\alpha^{rst}A^sA^t - \frac{1}{2}A^{rs}A^s,\\
R_3^r &= \frac{3}{8}A^tA^{rs}A^{st} - \frac{5}{12}\alpha^{rst}A^sA^uA^{tu} - \frac{5}{12}\alpha^{stu}A^tA^uA^{rs} + \frac{1}{3}A^sA^tA^{rst}\\
&\quad -\frac{1}{4}\alpha^{rstu}A^sA^tA^u + \frac{4}{9}\alpha^{rst}\alpha^{tuv}A^sA^uA^v.
\end{aligned}
$$

We note that $\widetilde{R}_1^r = R_1^r$, $\widetilde{R}_2^r = R_2^r$, and

$$\widetilde{R}_3^r = R_3^r - \frac{1}{8}\alpha^{rst}\alpha^{tuv}A^sA^uA^v + \frac{1}{8}\alpha^{rstu}A^sA^tA^u - \frac{1}{8}A^rA^sA^s + \frac{1}{n}(g_{n+1}^r + g_{n+2}^r).$$

With the choice of g_{n+1} and g_{n+2} given in the theorem, the third relationship further simplifies to $\widetilde{R}_3^r = R_3^r - bA^r/(2n)$.

A formal Edgeworth expansion of $\widetilde{R}$, according to Bhattacharya and Ghosh (1978), can be obtained through its first four cumulants. Let $Q_n = \sqrt{n}\widetilde{R}$, and let the cumulants of Q_n be denoted by $\kappa_{r,s,\ldots,t}$. Some tedious but straightforward computation gives

$$
\begin{aligned}
\kappa_r &= n^{-1/2}k_{1,1}^r + n^{-3/2}k_{1,2}^r + o(n^{-2}),\\
\kappa_{r,s} &= \delta^{rs} + n^{-1}k_{2,2}^{rs} + n^{-2}k_{2,3}^{rs} + o(n^{-2}),\\
\kappa_{r,s,t} &= n^{-3/2}k_{3,1}^{rst} + o(n^{-2}),\\
\kappa_{r,s,t,u} &= n^{-2}k_{4,1}^{rstu} + o(n^{-2}),
\end{aligned}
$$

where $k_{1,1}^r = -\alpha^{rss}/6$, $k_{2,2}^{rs} = \alpha^{rstt}/2 - \alpha^{ruv}\alpha^{suv}/3 - \alpha^{ruu}\alpha^{svv}/36 - b\delta^{rs}$ and the exact values of the remaining constants $k_{1,2}^r$, $k_{3,1}^{rst}$, and $k_{4,1}^{rstu}$ are not needed.

Let $\chi_n(\cdot)$, $f_{Q_n}(\cdot)$ be the characteristic and density functions of Q_n, respectively. From the relationship between the characteristic function and the cumulants of a random vector, we have

$$\begin{aligned}\chi_n(\tau) &= E\{\exp(\mathbf{i}\tau^T Q_n)\}\\ &= \exp\{\mathbf{i}\kappa_r\tau^r - \frac{1}{2}\kappa_{r,s}\tau^r\tau^s - \frac{1}{6}\mathbf{i}\kappa_{r,s,t}\tau^r\tau^s\tau^t + \frac{1}{24}\kappa_{r,s,t,u}\tau^r\tau^s\tau^t\tau^u + o(n^{-2})\}\\ &= \exp\{-\frac{1}{2}\tau^T\tau + \mathbf{i}n^{-1/2}k_{1,1}^r\tau^r - \frac{1}{2}n^{-1}k_{2,2}^{rs}\tau^r\tau^s + O(n^{-3/2})\}\\ &= \exp\{-\frac{1}{2}\tau^T\tau\}\{1 + \mathbf{i}n^{-1/2}k_{1,1}^r\tau^r - \frac{1}{2}n^{-1}(k_{1,1}^r k_{1,1}^s + k_{2,2}^{rs})\tau^r\tau^s\} + O(n^{-3/2}).\end{aligned}$$

To justify that the above expansion has retained all leading terms, the cumulants of Q_n, of orders five and six must be $o(n^{-2})$. The order assessments of the cumulants of Q_n can be done easily through the relationship between R and $\widetilde{R}$, and the existing results on R.

Let $\phi(x)$ be the density function of the standard multivariate normal distribution, $r_1(\tau) = k_{1,1}^r\tau^r$, and $r_2(\tau) = (k_{1,1}^r k_{1,1}^s + k_{2,2}^{rs})\tau^r\tau^s/2$. The density function with characteristic function $\chi_n(\tau)$ is then

$$f_{Q_n}(x) = \phi(x) + n^{-1/2}r_1(-\frac{d}{dx})\phi(x) + n^{-1}r_2(-\frac{d}{dx})\phi(x) + O(n^{-3/2}).$$

According to Bhattacharya and Ghosh (1978), this formal Edgeworth expansion has $O(n^{-3/2})$ precision. Consequently, the cumulative distribution function of $W_A = Q_n^T Q_n$ is obtained through a simple integration as

$$\begin{aligned}&Pr\{W_A(\theta_0) \le y\}\\ &\quad = Pr\{Q_n^T Q_n \le y\} + O(n^{-3/2})\\ &\quad = \int_{x^T x \le y}\left\{1 + n^{-1/2}r_1(-\frac{d}{dx}) + n^{-1}r_2(-\frac{d}{dx})\right\}\phi(x)dx + O(n^{-3/2}).\end{aligned}$$

Notice that $r_1(-d/dx)\phi(x) = k_{1,1}^r x^r\phi(x)$ is an odd function. Hence, when integrated over a symmetric region,

$$\int_{x^T x \le y} r_1(-\frac{d}{dx})\phi(x)dx = 0.$$

For the second polynomial, we find that

$$r_2(-\frac{d}{dx})\phi(x) = \frac{1}{2}(k_{1,1}^r k_{1,1}^s + k_{2,2}^{rs})(x^r x^s - \delta^{rs})\phi(x).$$

Recall that when an index repeats, the corresponding term sums over its range. The above expression is a sum of many ordinary terms. Among them, the terms corresponding to $r \neq s$ are odd functions. Hence, their integrations over symmetric regions are zero. We need only keep the terms with $r = s$ in the expansion of $Pr(W_A(\theta_0) \leq y)$. After these 0 terms are removed, we arrive at

$$\begin{aligned}
&Pr(W_A(\theta_0) \leq y) \\
&= \int_{x^T x \leq y} \phi(x)dx + \frac{1}{2n}(k^r_{1,1}k^r_{1,1} + k^{rr}_{2,2}) \int_{x^T x \leq y} (x^r x^r - \delta^{rr})\phi(x)dx + O(n^{-3/2}).
\end{aligned}$$

By explicitly spelling out the summations over the repetitive indices, it becomes simple to show that

$$\sum_r (k^r_{1,1}k^r_{1,1} + k^{rr}_{2,2}) = \frac{1}{2}\sum_{r,s} \alpha^{rrss} - \frac{1}{3}\sum_{r,s,t} \alpha^{rst}\alpha^{rst} - b\sum_r \delta^{rr} = 0.$$

Without summing over r, $\int_{x^T x \leq y}(x^r x^r - \delta^{rr})\phi(x)dx$ is a constant, say C, which does not depend on r. Therefore

$$(k^r_{1,1}k^r_{1,1} + k^{rr}_{2,2}) \int_{x^T x \leq y} (x^r x^r - \delta^{rr})\phi(x)dx = C\sum_r (k^r_{1,1}k^r_{1,1} + k^{rr}_{2,2}) = 0.$$

That is, the second term in the expansion disappears, and

$$Pr(W_A(\theta_0) \leq y) = \int_{x^T x \leq y} \phi(x)dx + O(n^{-3/2}) = P(\chi^2_p \leq y) + O(n^{-3/2}).$$

Because we are working on a symmetric region in Q_n, the expansion does not contain terms of order $n^{-j/2}$ if j is odd. Thus, the remainder term must have the next possible order, $O(n^{-2})$. This completes our proof.

Proof of Theorem 2. Without loss of generality, $E\{g(X, \theta_0)\}^2 = 1$. It can be seen that $g_{n+1} = O_p(1)$ and $g_{n+2} = O_p(n^{-1/2})$. Because we consider only the case where $p = q = 1$, we do not need tensor notations and, for $r = 1, 2, 3$,

$$\alpha_r = E[g(Y; \theta_0)]^r, \quad A_r = n^{-1}\sum_{i=1}^{n}[g(y_i; \theta_0)]^r - \alpha_r.$$

Similar to the proof of Theorem 1, the Lagrange multiplier λ can be expanded as the sum of

$$\begin{aligned}
\lambda_1 &= -A_1, \\
\lambda_2 &= A_1A_2 - \frac{1}{2}\alpha_3 A_1^2 - n^{-1}g_{n+1}, \\
\lambda_3 &= \frac{3}{2}\alpha_3 A_1^2 A_2 + \frac{1}{6}\alpha_4 A_1^3 - A_1A_2^2 - \frac{1}{2}\alpha_3^2 A_1^3 - \frac{1}{2}A_1^2 A_3 \\
&\quad + n^{-1}(A_2 - \alpha_3 A_1 + g_{n+1}A_1)g_{n+1} - n^{-1}g_{n+2},
\end{aligned}$$

after ignoring higher order terms. Clearly, $\lambda = O_p(n^{-1/2})$. Recall that

$$W_{AET}(\theta_0) = 2\sum_{i=1}^{n+2}(n+2)p_i(\lambda)\log\{(n+2)p_i(\lambda)\}$$
$$= 2(n+2)\log(n+2) - 2(n+2)\log\Big[\sum_{j=1}^{n+2}\exp(\lambda g_j)\Big]$$
$$+2(n+2)\lambda\frac{\sum_{i=1}^{n+2} g_i\exp(\lambda g_i)}{\sum_{j=1}^{n+2}\exp(\lambda g_j)}.$$

Expanding in powers of λ, we get

$$W_{AET}(\theta_0) = (1 + A_2 - A_1^2 + n^{-1}g_{n+1}^2)n\lambda^2 + \frac{2n}{3}(\alpha_3 + A_3 - 3A_1)n\lambda^3$$
$$+(\frac{1}{4}\alpha_4 - \frac{3}{4})n\lambda^4 + O_p(n^{-3/2}).$$

Substituting the expansion of λ into this expression, we get

$$W_{AET}(\theta_0) = nV^2 + O_p(n^{-3/2}) = n(V_1 + V_2 + V_3)^2 + O_p(n^{-3/2})$$

with

$$V_1 = A_1,$$
$$V_2 = -\frac{1}{2}A_1A_2 + \frac{1}{6}\alpha_3A_1^2 + n^{-1}g_{n+1},$$
$$V_3 = \frac{3}{8}A_1A_2^2 + \frac{1}{6}A_1^2A_3 + \frac{1}{9}\alpha_3^2A_1^3 - \frac{5}{12}\alpha_3A_1^2A_2 - \frac{1}{24}\alpha_4A_1^3 + \frac{1}{8}A_1^3$$
$$+n^{-1}g_{n+1}\left(-\frac{1}{2}A_2 + \frac{1}{3}\alpha_3A_1 - \frac{1}{2}g_{n+1}A_1\right) + n^{-1}g_{n+2}.$$

Comparing this to the expansion of $W_{EL}(\theta_0)$ for the case of $q = 1$, and with the choice of g_{n+1} and g_{n+2} given in the theorem, we find $V_1 = R_1$, $V_2 = R_2$, and

$$V_3 = R_3 - \frac{1}{6}A_1^2A_3 - \frac{7}{36}\alpha_3^2A_1^3 + \frac{1}{3}\alpha_3A_1^2A_2 + \frac{1}{6}\alpha_4A_1^3$$
$$-n\frac{1}{72}\alpha_3^2A_1^5 - n^{-1}A_1(\frac{1}{4}\alpha_4 + \frac{1}{24}\alpha_3^2).$$

Because the cumulants of R are given in DiCiccio, Hall and Romano (1991), we can compute the cumulants of $T_n = \sqrt{n}V$, $\kappa_j, j = 1,2,3,4$, relatively easily. They are

$$\kappa_1 = n^{-1/2}k_{1,1} + n^{-3/2}k_{1,2} + o(n^{-2}),$$
$$\kappa_2 = 1 + n^{-1}k_{2,2} + n^{-2}k_{2,3} + o(n^{-2}),$$
$$\kappa_3 = n^{-3/2}k_{3,1} + o(n^{-2}),$$
$$\kappa_4 = n^{-2}k_{4,1} + o(n^{-2}),$$

Table 1. Less Biased Moment Estimators when $p = 1$.

Parameter	Estimator	Expression
β_2	$\tilde{\beta}_2$	$n\hat{\beta}_2/(n-1)$
β_4	$\tilde{\beta}_4$	$(n\hat{\beta}_4 - 6\tilde{\beta}_2^2)/(n-3)$
β_2^2	$\tilde{\beta}_{22}$	$\tilde{\beta}_2^2 - \tilde{\beta}_4/n$
β_3	$\tilde{\beta}_3$	$n\hat{\beta}_3/(n-3)$
β_3^2	$\tilde{\beta}_{33}$	$\tilde{\beta}_3^2 - (\hat{\beta}_6 - \tilde{\beta}_3^2)/n$
β_2^3	$\tilde{\beta}_{222}$	$\tilde{\beta}_2^3$

with $k_{1,1} = -\alpha_3/6$, $k_{2,2} = -\alpha_3^2/36$, and $k_{1,2}$, $k_{3,1}$, $k_{4,1}$ being non-random constants whose exact values are not needed. The higher order cumulants of T_n in the orders five and six are $o(n^{-2})$.

Because the leading terms in these cumulants are the same as these of $\tilde{R}$ in the proof of Theorem 1 (when $q = 1$), the Edgeworth expansion of $\sqrt{n}V$ must be the same as that of $Q_n = \sqrt{n}\tilde{R}$. Consequently, the same expansion applies and $Pr(W_{AET}(\theta_0) \le y) = P(\chi_1^2 \le y) + O(n^{-2})$. This completes the proof.

4. Estimation of the Coefficients in Adjusted Exponential Empirical Likelihood

Our adjustment method contains unknown parameters $\alpha^{rs\cdots t}$ that must be estimated. Moment estimators underestimate and do not allow the method to achieve its full potential (Liu and Chen (2010)). Through some simple bias analysis, Liu and Chen (2010) suggested a set of less biased estimators. We follow their example, with details omitted.

Consider the case p=1. Let $\beta_r = E\{g(Y, \theta_0)\}^r$, with moment estimator $\hat{\beta}_r = n^{-1}\sum_{i=1}^n (g_i - \bar{g})^r$. If needed, we replace the unknown θ_0 by the maximum adjusted exponential tilt empirical likelihood estimator $\hat{\theta} = \arg\min_\theta\{W_{EEL}(\theta)\}$. The pseudo-observations g_{n+1} and g_{n+2} are functions of $\alpha^r = \beta_2^{-r/2}\beta_r$ for $r = 2, 3, 4$. Closer investigation reveals that they depend only on $\beta_2, \beta_2^2, \beta_2^3$, β_3, β_3^2, and β_4. In the simulation, we used the less-biased estimators of these parameters given in Table 1. Note that we do not estimate, for example, β_2^2 by $(\tilde{\beta}_2)^2$, but by $\tilde{\beta}_{22}$ to reduce potential bias.

When $p > 1$, we estimate Σ_0 by the sample variance $\hat{\Sigma}_0$ of $g(y, \theta)$ at $\theta = \hat{\theta}$. We then compute $x_i = \hat{\Sigma}_0^{-1/2} g(y_i, \hat{\theta})$ and the corresponding moment estimator $\hat{\alpha}^{rs\cdots t} = n^{-1}\sum_{i=1}^n {x_i}^r {x_i}^s \cdots {x_i}^t$. Similarly to the case where $p = 1$, we use the estimators given in Table 2.

5. Simulations

A classical problem is the construction of confidence regions or tests of a hypothesis about a specific value of the population mean based on a set of n

Table 2. Less Biased Moment Estimators when $p > 1$.

Parameter	Estimator	Expression
α^{rr}	$\tilde{\alpha}^{rr}$	$n\hat{\alpha}^{rr}/(n-1)$
α^{rrss}	$\tilde{\alpha}^{rrss}$	$[n\hat{\alpha}^{rrss} - 2\hat{\alpha}^{rr}\hat{\alpha}^{ss} - 4I(r=s)\hat{\alpha}^{rr}\hat{\alpha}^{rr}]/(n-4)$
α^{rrrs}	$\tilde{\alpha}^{rrrs}$	$n\hat{\alpha}^{rrrs}/(n-4)$
α^{rst}	$\tilde{\alpha}^{rst}$	$n\hat{\alpha}^{rst}/(n-3)$
$\alpha^{rst}\alpha^{rst}$	$\tilde{\alpha}^{rst,rst}$	$\tilde{\alpha}^{rst}\tilde{\alpha}^{rst} - (\hat{\alpha}^{rrsstt} - \tilde{\alpha}^{rst}\tilde{\alpha}^{rst})/n$
$\alpha^{rr}\alpha^{ss}$	$\tilde{\alpha}^{rr,ss}$	$\hat{\alpha}^{rr}\hat{\alpha}^{ss} - \tilde{\alpha}^{rrss}/n$
$\alpha^{rr}\alpha^{ss}\alpha^{tt}$	$\tilde{\alpha}^{rr,ss,tt}$	$\tilde{\alpha}^{rr}\tilde{\alpha}^{ss}\tilde{\alpha}^{tt}$

independent and identically distributed observations $x_1, \ldots, x_n$. When we consider the population mean, the estimating equation is $E\{g(X,\theta)\} = 0$, where $g(x,\theta) = x - \theta$. We investigate the coverage probabilities of level 0.90, 0.95, and 0.99 confidence intervals and corresponding interval lengths based on six methods.

1. Hotellling's T^2 with $T_n^2 = n(\bar{X}_n - \theta)^T S_n^{-1}(\bar{X}_n - \theta)$, where $\bar{X}_n$ is the vector sample mean and S_n is the sample covariance matrix.
2. Empirical likelihood, EL.
3. Bartlett corrected empirical likelihood, BEL.
4. Adjusted empirical likelihood, AEL.
5. Exponential empirical likelihood, EEL.
6. Adjusted exponential empirical likelihood, AEEL.

We generated 10,000 samples from four distributions: the standard normal; the exponential distribution with mean 1; a mixture of normal $0.2N(5,1) + 0.8N(-1.25,1)$; and the χ_1^2 distribution. The coverage probability results are presented in Table 3 and the corresponding interval lengths in Table 4. As expected, AEEL improves EEL, and is comparable to other high order methods in coverage probabilities without inflating average lengths. We conclude that AEEL lives up to nice properties acclaimed by Theorem 1.

In the multivariate case, we conducted simulation experiments for $p = q = 2$, and generated data from the bivariate standard normal distribution and three other distributions as follows. We first generated a random observation D from the uniform distribution on the interval [1, 2]. Given D, we generated X_1, X_2 as

(a) $X_1 \sim N(0, D^2)$, $X_2 \sim Gamma(D^{-1}, 1)$,
(b) $X_1 \sim Gamma(D, 1)$, $X_2 \sim Gamma(D^{-1}, 1)$,
(c) $X_1 \sim Poisson(D)$, $X_2 \sim Poisson(D^{-1})$.

For each population, 10,000 data sets were generated with sample sizes $n = 20$ and 30. Table 5 presents the simulation results of coverage probability. From

Table 3. Coverage probabilities for one-sample population mean.

	n	nominal	T^2	EL	BEL	AEL	EEL	AEEL
$N(0,1)$	20	90	89.6	88.2	89.5	89.4	87.4	89.6
		95	95.0	93.6	94.5	94.5	92.6	94.6
		99	98.8	98.2	98.6	98.6	97.7	98.6
	30	90	89.8	89.5	90.4	90.4	89.0	90.5
		95	94.9	94.5	95.1	95.1	93.9	95.2
		99	99.0	98.7	98.9	98.9	98.3	98.9
$Exp(1)$	20	90	87.0	86.2	87.8	87.9	85.5	88.4
		95	91.6	91.5	92.6	92.6	90.7	93.0
		99	96.6	96.6	97.2	97.2	96.0	97.2
	30	90	88.2	87.6	88.8	88.9	87.1	89.3
		95	92.9	93.0	93.9	94.0	92.3	94.3
		99	97.3	97.7	98.0	98.0	97.4	98.1
$Mixture$	20	90	89.1	88.6	90.4	90.0	88.1	90.6
		95	93.3	93.7	95.0	94.8	93.3	94.9
		99	97.3	98.0	98.2	98.2	97.8	98.1
	30	90	89.2	89.6	90.5	90.4	89.3	90.9
		95	93.7	94.3	95.1	95.0	94.0	95.4
		99	97.9	98.7	99.0	98.9	98.4	98.9
χ_1^2	20	90	84.1	83.6	85.6	85.6	82.8	86.6
		95	88.8	89.2	90.6	90.7	88.5	91.3
		99	94.4	95.2	96.0	95.9	94.6	96.3
	30	90	85.9	86.0	87.5	87.5	85.4	88.3
		95	90.4	91.7	92.8	92.8	91.9	93.3
		99	95.5	96.8	97.3	97.3	96.3	97.5

this table, we can see that AEEL improves EEL substantially, but we also notice that T^2 has the best performance in terms of the coverage probability, followed by BEL, AEL and AEEL. This was also observed by Baggerly (1998) who noticed that, for construction confidence interval of a univariate mean based on empirical likelihood, Bartlett correction was better than the method using the critical values from a scaled F-distribution only when the nominal level was at 50%.

We further computed the average areas of the 95% confidence regions of T^2 and AEEL based on 1,000 data sets. For population (a) with sample sizes $n = 20$ and 30, the ratios of these two areas were respectively $2.237/1.935 = 1.156$ and $1.432/1.373 = 1.036$. The gains for AEEL remained even after taking the observed coverage probabilities into account (based on the current 1,000 data sets). We attribute this gain to the data driven shape of the AEEL confidence regions. The comparisons were similar for populations (b) and (c) and and other nominal levels. We do not report these details here.

Table 4. Average length of confidence interval for one-sample population mean.

	n	nominal	T^2	EL	BEL	AEL	EEL	AEEL
$N(0,1)$	20	90	0.75	0.73	0.76	0.76	0.71	0.77
		95	0.91	0.87	0.91	0.91	0.85	0.92
		99	1.25	1.16	1.21	1.20	1.10	1.21
	30	90	0.61	0.60	0.62	0.62	0.59	0.62
		95	0.73	0.72	0.74	0.74	0.70	0.74
		99	0.99	0.96	0.99	0.98	0.92	0.99
$Exp(1)$	20	90	0.73	0.71	0.75	0.75	0.70	0.77
		95	0.89	0.85	0.90	0.90	0.83	0.92
		99	1.21	1.12	1.19	1.20	1.08	1.22
	30	90	0.59	0.60	0.62	0.62	0.58	0.63
		95	0.71	0.71	0.75	0.74	0.70	0.76
		99	0.97	0.95	0.99	0.99	0.92	1.00
$Mixture$	20	90	2.03	1.90	1.98	1.98	1.88	2.00
		95	2.46	2.27	2.36	2.36	2.22	2.39
		99	3.35	2.96	3.09	3.09	2.88	3.12
	30	90	1.64	1.58	1.62	1.62	1.57	1.63
		95	1.98	1.88	1.93	1.93	1.86	1.94
		99	2.67	2.47	2.53	2.53	2.41	2.55
χ_1^2	20	90	1.00	0.98	1.04	1.04	0.96	1.08
		95	1.21	1.17	1.25	1.25	1.14	1.29
		99	1.66	1.55	1.66	1.67	1.50	1.70
	30	90	0.82	0.83	0.87	0.87	0.81	0.89
		95	0.99	0.99	1.04	1.04	0.97	1.06
		99	1.34	1.32	1.39	1.39	1.28	1.41

6. An Application Example

We applied the AEEL and other methods to a data set from Efron and Tibshirani (1993, Table 2.1). In a small experiment, 7 out of 16 mice were randomly selected to receive a new medical treatment, while the remaining 9 were assigned to the non-treatment (control) group. The survival times following surgery, in days, for all 16 mice are shown in Table 6. The objective of this experiment was to test whether the new treatment would prolong survival.

The 90% confidence intervals of the average survival time for treatment and control groups based on methods introduced in the simulation section, and their lengths, are given in Table 7. We also calculated 90% confidence interval based on the bootstrap-t method. From the table, we can see that all nonparametric empirical likelihood confidence intervals shifted to the right compared to that based on T^2, which is a desirable result, and the AEEL interval was a lot shorter than the Bootstrap-t interval. The AEEL intervals were slightly longer than other nonparametric likelihood confidence intervals, but the differences were negligible.

Table 5. Coverage probabilities for one-sample bivariate population mean.

	n	nominal	T^2	EL	BEL	AEL	EEL	AEEL
$N(0, I)$	20	90	90.3	84.7	87.3	87.3	82.8	86.6
		95	95.0	91.0	92.6	92.8	89.0	92.1
		99	99.0	96.6	97.4	97.5	95.4	97.1
	30	90	90.8	87.0	88.6	88.5	85.8	88.0
		95	95.7	92.4	93.6	93.6	91.2	93.0
		99	99.3	97.6	98.1	98.1	96.9	97.7
(a)	20	90	86.1	81.5	84.4	84.7	79.9	83.8
		95	91.2	87.9	90.0	90.2	86.0	89.4
		99	96.4	94.5	95.7	95.8	93.3	95.4
	30	90	88.1	84.9	86.9	87.0	83.6	86.3
		95	93.2	90.6	92.2	92.2	89.3	91.5
		99	97.8	96.6	97.3	97.4	95.8	97.0
(b)	20	90	84.2	81.0	84.0	84.3	79.4	83.7
		95	89.5	87.1	89.4	89.7	85.6	89.1
		99	95.4	93.7	95.1	95.3	92.5	94.8
	30	90	86.6	84.5	86.9	87.0	83.4	86.5
		95	91.7	90.4	91.9	92.1	89.5	91.6
		99	96.7	96.5	97.2	97.2	95.6	96.8
(c)	20	90	88.1	83.6	86.4	86.4	82.2	85.7
		95	93.2	89.9	91.6	91.7	88.3	91.0
		99	97.7	95.9	96.7	96.8	94.9	96.4
	30	90	89.2	86.2	87.8	87.8	85.1	87.1
		95	94.5	92.1	93.2	93.2	91.0	92.6
		99	98.6	97.4	97.9	97.9	96.6	97.5

Table 6. Survival times of the mice in the application example.

Group	Survival time
Treatment:	94, 197, 16, 38, 99, 141, 23
Control:	52, 104, 146, 10, 50, 31, 40, 27, 46

In conclusion, this example shows that, as anticipated, the AEEL is a suitable method for the construction of confidence intervals.

7. Conclusion and Discussion

This paper shows that the technique developed in Chen, Mulayath and Abraham (2008) is useful to other nonparametric likelihoods in addition to the empirical likelihood. Unlike the Bartlett correction, this technique is more widely applicable and works when the accompanying estimating equations do not have a solution.

The straightforward adjustment in Chen, Mulayath and Abraham (2008)

Table 7. Confidence intervals (90%) and their lengths for the mean survival time of the treatment and control groups.

	Treatment		Control	
Method	Confidence interval	Length	Confidence interval	Length
T^2	(37.82, 135.89)	98.07	(29.93, 82.51)	52.58
EL	(52.95, 127.86)	74.91	(38.39, 82.34)	43.95
BEL	(51.62, 129.85)	78.23	(36.77, 85.57)	48.80
AEL	(51.35, 129.67)	78.32	(36.14, 85.03)	48.89
EEL	(53.83, 126.57)	72.74	(39.25, 81.83)	42.58
AEEL	(52.11, 128.32)	76.21	(35.95, 86.46)	50.51
Bootstrap-t	(42.99, 146.42)	103.43	(35.82, 116.74)	80.92

results in a bounded likelihood ratio function. When data are from a severely biased population and the sample size is small, some AEL and AEEL confidence regions can be unbounded. We employed some corrections to overcome this shortcoming. The requirement of this remedy points to imperfectness of the method and the need for more research. We plan to investigate these issues in the future.

Acknowledgement

This research was partially supported by the Natural Science and Engineering Research Council of Canada, Mathematics of Information Technology and Complex Systems, and the National Nature Science Foundation of China. We thank referees and an associate editor for their insightful comments.

References

Baggerly, K. A. (1998). Empirical likelihood as a goodness-of-fit measure. *Biometrika* **85**, 535-547.

Bhattacharya, R. N. and Ghosh, J. K. (1978). On the validity of the formal Edgeworth expansion. *Ann. Statist.* **6**, 434-451.

Chen, J., Mulayath, V. A. and Abraham, B. (2008). Adjusted empirical likelihood and its properties. *J. Comput. Graph. Statist.* **17**, 1-18.

Corcoran, S. A. (1998). Bartlett adjustment of empirical discrepancy statistics. *Biometrika* **85**, 967-972.

DiCiccio, T., Hall, P. and Romano, J. (1991). Empirical likelihood is Bartlett correctable. *Ann. Statist.* **19**, 1053-1061.

Efron, B. (1981). Nonparametric standard errors and confidence intervals (with discussion). *Canad. J. Statist.* **9**, 139-172.

Efron, B. and Tibshirani, R. J. (1993). An Introduction to the Bootstrap. Chapman and Hall/CRC New York. Monographs on Statistics and Applied Probability.

Jing, B. and Wood, T. A. (1996). Exponential empirical likelihood is not Bartlett correctable. *Ann. Statist.* **24**, 365-369.

Liu, Y. and Chen, J. (2010). Adjusted empirical likelihood with high-order precision. *Ann. Statist.* **38**, 1341-1362

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

Owen, A. B. (2001). *Empirical Likelihood*. Chapman and Hall/CRC, New York.

Qin, J. and Lawless, J. (1994). Empirical likelihood and general equations. *Ann. Statist.* **22**, 300-325.

Schennach, S. M. (2007). Point estimation with exponentially tilted empirical likelihood. *Ann. Statist.* **35**, 634-672.

Department of Statistics and Finance, University of Science and Technology of China, Hefei, 230026, China.

E-mail: lix@mail.ustc.edu.cn

Department of Statistics, University of British Columbia, Vancouver, BC, V6T 1Z2, Canada.

E-mail: jhchen@stat.ubc.ca

Department of Statistics and Finance, University of Science and Technology of China, Hefei, 230026, China.

E-mail: wuyh@ustc.edu.cn

NCIC Clinical Trials Group, Queen's University, Kingston, ON, K7L 3N6, Canada.

E-mail: DTu@ctg.queensu.ca

(Received May 2009; accepted June 2010)