Journal of Nonparametric Statistics, 2014
Vol. 26, No. 3, 433–449, http://dx.doi.org/10.1080/10485252.2014.929676

Level-specific correction for nonparametric likelihoods

Yukun Liu[a]*, Jiahua Chen[b] and Ting Li[a]

[a]*School of Finance and Statistics, East China Normal University, 500 Dongchuan Road, Shanghai 200241, People's Republic of China;* [b]*Department of Statistics, University of British Columbia, Vancouver, BC, Canada V6T 1Z2*

(*Received 5 November 2013; accepted 24 May 2014*)

The popular empirical likelihood method not only has a convenient chi-square limiting distribution but is also Bartlett correctable, leading to a high-order coverage precision of the resulting confidence regions. Meanwhile, it is one of many nonparametric likelihoods in the Cressie–Read power divergence family. The other likelihoods share many attractive properties but are not Bartlett correctable. In this paper, we develop a new technique to achieve the effect of being Bartlett correctable. Our technique is generally applicable to pivotal quantities with chi-square limiting distributions. Numerical experiments and an example reveal that the method is successful for several important nonparametric likelihoods.

Keywords: Bartlett correction; empirical likelihood; exponential tilting likelihood; Euclidean likelihood; power divergence family

AMS Subject Classification: 62G15; 62G20

1. Introduction

Since the seminal work of Owen (1988), the empirical likelihood has found applications far and wide. It is particularly useful for estimating functions (Qin and Lawless 1994) and for over-identified models in econometrics (Chen and Cui 2007; Matsushita and Otsu 2013). It naturally accommodates auxiliary information for efficiency improvement in survey sampling (Chen and Qin 1993). A celebrated property of the empirical likelihood is that its likelihood ratio function has a chi-square limiting distribution under simple conditions. This result matches the famous Wilks theorem for parametric models. Moreover, the empirical likelihood is Bartlett correctable (DiCiccio et al. 1991; Chen and Cui 2006). That is, there exists a multiplication factor for the empirical likelihood ratio such that the distribution of the corrected statistic is approximated by a chi-square distribution with an $O(n^{-2})$ margin of error over the entire range, with n being the sample size. The Bartlett correction thus enables users to construct confidence regions with a second-order precision coverage probability. The empirical likelihood approach also has many other attractive properties (see Owen 2001).

The empirical likelihood is one of many nonparametric likelihoods in the power divergence family of Cressie and Read (1984). This family also includes the Euclidean likelihood (Owen 1991) and the exponential tilting likelihood (empirical entropy; Efron 1981), both of which are used in various applications (Imbens, Johnson, and Spady 1998; Antoine Bonnal, and Renault

*Corresponding author. Email: ykliu@sfs.ecnu.edu.cn

2007). These nonparametric likelihoods have the same first-order properties as the empirical likelihood, but they are not Bartlett correctable (Jing and Wood 1996; Baggerly 1998; Corcoran 1998). More precisely, there does not exist a multiplication factor such that the distribution of the corrected statistics is approximated by a chi-square distribution with an $O(n^{-2})$ margin of error over the entire range. It therefore seems that only the empirical likelihood can be used to construct high-precision confidence regions.

This paper introduces a new technique. In applications, the nominal level of a confidence region is often prespecified as 95% and the size of a hypothesis test as 5%. For any prespecified nominal level, a corresponding correction factor exists for each nonparametric likelihood in the power divergence family. We find the explicit form of this correction factor for each nonparametric likelihood; it is a simple function of the population moments and the target level of the confidence region. Our result is applicable to the well-known Hotelling's T^2 test statistic since it is simply the Euclidean likelihood ratio test statistic (Owen 1991) scaled by $(n-1)/n$. Our technique also applies to other inferential methods based on pivotal quantities with chi-square limiting distributions, such as the F-statistic in the analysis of variance.

The empirical likelihood and more generally all nonparametric likelihoods are also valued for inference on population parameters defined through estimating equations. Nonparametric-likelihood-based inference generally starts by obtaining a profile likelihood of the equation-defined parameters. The solution set of the sample-based estimating equations may be empty, leading to undefined or ad hoc nonparametric likelihoods. Many researchers have suggested ways to overcome this obstacle (Chen, Variyath, and Abraham 2008; Lahiri and Mukhopadhyay 2012; Tsao and Wu 2013). The adjusted empirical likelihood of Chen et al. (2008) is particularly simple and effective. The addition of a well-motivated pseudo-entry to the sample estimating equation ensures that the resulting adjusted empirical likelihood is always well defined. Liu and Chen (2010) further show that the adjusted empirical likelihood can be tuned to achieve the same second-order precision as the Bartlett-corrected empirical likelihood. Our technique can also develop level-specific adjustments of nonparametric likelihoods so that these likelihoods are always well defined and lead to confidence regions with high-order coverage precision.

The paper is organised as follows. Section 2 briefly reviews the empirical likelihood and nonparametric likelihoods. In Section 3, we present the technique for the nonparametric likelihood and its extensions. Monte Carlo simulations and an illustrative example are given in Section 4. Section 5 contains some discussion. All technical proofs are given in the appendix.

2. Cressie–Read nonparametric likelihoods

Suppose the d-dimensional observations $X_1, X_2, \ldots, X_n$ are independent and identically distributed copies of X with mean μ. Owen (1988) defines the empirical likelihood ratio function for the population mean as

$$\mathrm{EL}(\mu) = \inf\left\{-2\sum_{\mathrm{i}=1}^{\mathrm{n}} \log(\mathrm{np_i}) : \ \mathrm{p_i} \geq 0, \quad \sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{p_i} = 1, \quad \sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{p_i X_i} = \mu\right\}. \tag{1}$$

The empirical likelihood ratio function can be similarly defined for parameters defined by estimating equations (Qin and Lawless 1994). The results in this paper are generally applicable, but we will focus on the population mean.

Baggerly (1998) observes that the empirical likelihood is one of the more general Cressie–Read nonparametric likelihoods, which are defined, for a given user-specified $\gamma \in \mathbb{R}$, to be

$$L_\gamma(\mu) = \inf \left\{ \frac{2}{\gamma(\gamma+1)} \sum_{i=1}^{n} [(np_i)^{\gamma+1} - 1] : \; p_i \geq 0, \quad \sum_{i=1}^{n} p_i = 1, \quad \sum_{i=1}^{n} p_i X_i = \mu \right\}. \tag{2}$$

When $\gamma = -1$ and $\gamma = 0$, the nonparametric likelihood $L_\gamma(\mu)$ is defined through its continuous limits. When $\gamma \to -1$, the nonparametric likelihood reduces to the empirical likelihood $\mathrm{EL}(\mu)$, and when $\gamma \to 0$, it becomes

$$\mathrm{ET}(\mu) = \inf \left\{ 2 \sum_{i=1}^{n} (np_i) \log(np_i) : \; p_i \geq 0, \quad \sum_{i=1}^{n} p_i = 1, \quad \sum_{i=1}^{n} p_i X_i = \mu \right\},$$

which is called the exponential tilting or empirical entropy statistic (Efron 1981). Other popular nonparametric likelihoods include Neyman's modified χ^2 or the Euclidean likelihood (Owen 1991) with $\gamma = 1$, the Hellinger or Freeman–Tukey statistic with $\gamma = -1/2$, and Pearson's χ^2 statistic with $\gamma = -2$ (see Cressie and Read 1984).

Throughout this paper we take μ_0 as the true value of μ. When the covariance matrix of X_1 is nonsingular, the empirical likelihood ratio $\mathrm{EL}(\mu_0)$ has a χ^2_d limiting distribution (Owen 1988). Furthermore, if X_1 has finite third moments, then

$$P\{\mathrm{EL}(\mu_0) \leq x\} = P(\chi^2_d \leq x) + O(n^{-1}), \tag{3}$$

where χ^2_d denotes a χ^2-distributed random variable with d degrees of freedom. Moreover, there exists a nonrandom constant b, called the Bartlett correction factor, such that under some higher moment conditions,

$$P\left\{ \frac{\mathrm{EL}(\mu_0)}{(1+b/n)} \leq x \right\} = P(\chi^2_d \leq x) + O(n^{-2}). \tag{4}$$

That is, the empirical likelihood is Bartlett correctable (DiCiccio et al. 1991; Chen and Cui 2006).

Baggerly (1998) shows that the nonparametric Wilks theorem (3) holds for all nonparametric likelihoods defined by Equation (2). However, other than $\gamma = -1$, there exists no constant b such that Equation (4) holds, i.e. only the empirical likelihood is Bartlett correctable. When a likelihood is Bartlett correctable, users can construct confidence regions with highly accurate coverage probabilities through a single multiplication factor. It therefore seems that only the empirical likelihood can be used to construct confidence regions with a high-order coverage precision.

In applications, the nominal level of a confidence region is often set to 95% and the size of a hypothesis test to 5%. We show that for any prespecified nominal level, a corresponding correction factor can be found for each nonparametric likelihood in the power divergence family. We propose a level-specific correction that improves the precision from $O(n^{-1})$ to $O(n^{-2})$ as given by Equation (4).

3. Level-specific correction

3.1. *Level-specific correction for the Cressie–Read family*

The main use of either Equation (3) or (4) is to construct confidence regions for μ. A confidence region for μ of size $(1-\alpha)$ is defined to be

$$\{\mu : \mathrm{EL}(\mu) \leq \left(1+\frac{\mathrm{b}}{\mathrm{n}}\right)\chi^2_{\mathrm{d}}(1-\alpha)\}, \tag{5}$$

where $\chi^2_d(1-\alpha)$ is the $(1-\alpha)$th quantile of the chi-square distribution with d degrees of freedom. Because of Equation (4), the coverage probability of this confidence region is $(1-\alpha)+O(n^{-2})$, differing from the target level with an error of the order of n^{-2}.

For the other nonparametric likelihoods, such a b does not exist unless it is made dependent on α. Suppose

$$P\{L_\gamma(\mu_0) \leq x\} = F_d(x) - n^{-1}K(x) + O(n^{-2}), \tag{6}$$

where $F_d(\cdot)$ is the cumulative distribution function of the χ^2_d distribution and $K(\cdot)$ is a smooth function to be specified. For any function $b(x)$, we have

$$\begin{aligned} P\left\{\frac{L_\gamma(\mu_0)}{(1+b(x)/n)} \leq x\right\} &= F_d\left(x+\frac{xb(x)}{n}\right) - n^{-1}K(x) + O(n^{-2}) \\ &= F_d(x) + n^{-1}\{xb(x)F'_d(x) - K(x)\} + O(n^{-2}), \end{aligned}$$

where $F'_d(x)$ is the derivative of $F_d(x)$. When $b(x) = K(x)/\{xF'_d(x)\}$, we get

$$P\left[\frac{L_\gamma(\mu_0)}{\{1+b(x)/n\}} \leq x\right] = F_d(x) + O(n^{-2}).$$

For any given nominal level $1-\alpha$, let $x = \chi^2_d(1-\alpha)$. Consequently, a level-specific corrected confidence region defined by

$$\left\{\mu : L_\gamma(\mu) \leq \left\{1+\frac{b(x)}{n}\right\}\chi^2_d(1-\alpha)\right\}$$

has coverage probability $(1-\alpha)+O(n^{-2})$. That is, a second-order precise nonparametric likelihood confidence region is possible once Equation (6) is established for some $K(x)$.

For ease of exposition we introduce some notation. Let X^r denote the rth component of a vector X and for $k = 1, 2, \ldots,$

$$\alpha^{r_1 r_2 \cdots r_k} = \mathbb{E}(X^{r_1}X^{r_2}\cdots X^{r_k}), \quad A^{r_1 r_2 \cdots r_k} = \frac{1}{n}\sum_{i=1}^n X_i^{r_1}X_i^{r_2}\cdots X_i^{r_k} - \alpha^{r_1 r_2 \cdots r_k} \tag{7}$$

with $r_1, \ldots, r_k \in \{1, 2, \ldots, d\}$. Without loss of generality we assume $\mu_0 = 0$ or $\alpha^r = 0$, and the covariance matrix of X is an identity matrix or $\alpha^{rs} = \delta^{rs}$ which is equal to 1 if $r = s$ and zero otherwise. Define

$$\xi_1 = \left\{\frac{(2-\gamma-\gamma^2)}{4}\right\}\alpha^{rrkk} - \left\{\frac{(\gamma^2+2\gamma+3)}{6}\right\}\alpha^{rkm}\alpha^{rkm} + \left\{\frac{(1+\gamma)^2}{6}\right\}\alpha^{rkk}\alpha^{rll}, \tag{8}$$

$$\xi_2 = \left\{\frac{(1-\gamma^2)}{4}\right\}\alpha^{rrss} + \left\{\frac{(2\gamma^2+3\gamma+1)}{3}\right\}\alpha^{rsk}\alpha^{rsk} - \left\{\frac{(1+\gamma)^2}{3}\right\}\alpha^{rkk}\alpha^{rss} + \left\{\frac{(1+\gamma)^2}{4}\right\}(d^2+2d), \tag{9}$$

$$\xi_3 = \left\{\frac{5(1+\gamma)^2}{12}\right\}\alpha^{rss}\alpha^{rtt}. \tag{10}$$

Here and in what follows, unless otherwise stated, we use the summation convention (or tensor notation) according to which if an index occurs more than once in an expression, summation over the index is understood.

THEOREM 3.1 *Suppose that $X_1, X_2, \ldots, X_n$ is a random sample from a d-variate population with mean zero. Assume that $\mathbb{V}\mathrm{ar}(X_1)$ is the d-dimensional identity matrix, $\mathbb{E}(\|X_1\|^{42}) < \infty$, and the characteristic function of $F(x)$ satisfies Cramér's condition*:

$$\limsup_{\|t\|\to\infty} |\mathbb{E}\{\exp(it^{\mathrm{T}}X_1)\}| < 1,$$

where t^{T} denotes the transpose of t and i is the imaginary unit. Then for each constant γ, the distribution of the nonparametric likelihood defined in Equation (2) has an Edgeworth expansion (6) with

$$K(x) = \left\{\left(\frac{\xi_1}{d}\right) + \left[\frac{\xi_2}{\{d(d+2)\}}\right]x + \left[\frac{\xi_3}{\{d(d+2)(d+4)\}}\right]x^2\right\} xF_d'(x), \tag{11}$$

where the ξ's are given in Equations (8)–(10).

Remark 1 The moment condition $\mathbb{E}(\|X_1\|^{42}) < \infty$ is inherited from condition (A2) of Theorem 3 of Bhattacharya and Ghosh (1978). Together with Cramér's condition, it ensures the validity of the formal Edgeworth expansion (A9).

Remark 2 Under the conditions of Theorem 3.1, the remainder term in Equation (6) is $O(n^{-2})$ uniformly in x which can be shown from Equation (A9).

It follows immediately from Theorem 3.1 that the level-specific correction factor should be chosen as

$$b(x) = \left(\frac{\xi_1}{d}\right) + \left[\frac{\xi_2}{\{d(d+2)\}}\right]x + \left[\frac{\xi_3}{\{d(d+2)(d+4)\}}\right]x^2. \tag{12}$$

When $\gamma = -1$, in which case the nonparametric likelihood becomes the empirical likelihood, we find

$$\xi_1 = \tfrac{1}{2}\alpha^{rrss} - \tfrac{1}{3}\alpha^{rst}\alpha^{rst}, \quad \xi_2 = \xi_3 = 0.$$

Therefore, the level-specific correction factor for the empirical likelihood is

$$b_{\mathrm{EL}}(x) = \frac{1}{d}\left\{\frac{1}{2}\alpha^{rrss} - \frac{1}{3}\alpha^{rst}\alpha^{rst}\right\},$$

which is no longer level dependent and equals the well-known Bartlett correction factor for the empirical likelihood. When $\gamma = 0$ and 1, we obtain the level-specific correction factors for the exponential tilting likelihood and the Euclidean likelihood, respectively. In particular, the level-specific correction factor for the Euclidean likelihood is given by

$$\begin{aligned} b_{\mathrm{EU}}(x) = {} & \frac{\{(2/3)\alpha^{rkk}\alpha^{rll} - \alpha^{rkm}\alpha^{rkm}\}}{d} \\ & + \left[\frac{\{2\alpha^{rsk}\alpha^{rsk} - (4/3)\alpha^{rkk}\alpha^{rss}\}}{\{d(d+2)\}} + 1\right]x \\ & + \left[\frac{\{(5/3)\alpha^{rss}\alpha^{rtt}\}}{\{d(d+2)(d+4)\}}\right]x^2. \end{aligned}$$

The Euclidean likelihood ratio is asymptotically equivalent to Hotelling's T^2 defined by

$$T^2(\mu) = n(\bar{X} - \mu)^{\mathrm{T}} S^{-1}(\bar{X} - \mu)$$

with $\bar{X}$ and S being, respectively, the sample mean and the sample variance–covariance matrix. Here is a brief explanation. The Euclidean likelihood ratio is obtained from the optimal solution to Equation (2) when $\gamma = 1$. It is straightforward to see that this is given by

$$p_i = \frac{1}{n}\{1 - (\bar{X} - \mu)^{\mathrm{T}} S_*^{-1}(X_i - \bar{X})\}$$

with $S_* = \{(n-1)/n\}S$. The Euclidean likelihood ratio function is then given by

$$\mathrm{EU}(\mu) = n(\bar{X} - \mu)^{\mathrm{T}} S_*^{-1}(\bar{X} - \mu) = \left\{\frac{n}{(n-1)}\right\} T^2(\mu).$$

Strictly speaking, we require $p_i \geq 0$ for all i in Equation (2), but this is satisfied asymptotically.

A level-specific correction factor of Hotelling's T^2 can hence be obtained from that of the Euclidean likelihood. For any given $x > 0$, we have

$$P\left(\mathrm{EU}(\mu_0) \leq \mathrm{x}\left\{1 + \frac{\mathrm{b_{EU}(x)}}{\mathrm{n}}\right\}\right) = F_d(x) + O(n^{-2}),$$

or equivalently

$$P\left(\mathrm{EU}(\mu_0)\left\{\frac{(\mathrm{n}-1)}{\mathrm{n}}\right\} \leq \mathrm{x}\left\{1 + \frac{\mathrm{b_{EU}(x)}}{\mathrm{n}}\right\}\left\{1 - \left(\frac{1}{\mathrm{n}}\right)\right\}\right) = F_d(x) + O(n^{-2}).$$

Since $\{1 + b_{\mathrm{EU}}(x)/n\}\{1 - (1/n)\} = 1 + \{b_{\mathrm{EU}}(x) - 1\}/n + O(n^{-2})$, it follows that for Hotelling's T^2, a level-specific corrected confidence region with level $(1-\alpha)$ can be constructed as

$$\left\{\mu : T^2(\mu) \leq \left[1 + \frac{\{b_{\mathrm{EU}}(\chi_d^2(1-\alpha)) - 1\}}{n}\right]\chi_d^2(1-\alpha)\right\}.$$

As for the empirical likelihood, the second-order precision is retained when $b(x)$ is replaced with one of its root-n consistent estimators. A natural root-n estimate is obtained using the method of moments.

3.2. *Estimation of the level-specific correction factor*

As suggested by Liu and Chen (2010), direct moment estimates often underestimate b. We adopt their finite-sample corrections as follows. Let $\mathbb{V}\mathrm{ar}(X) = U\Lambda U^{\mathrm{T}}$ be an eigenvalue decomposition of the variance–covariance matrix of X with Λ a $d \times d$ diagonal matrix of the eigenvalues, and U a $d \times d$ matrix of the eigenvectors. Let $Y_i = U^{\mathrm{T}}(X_i - \bar{X})$ and define

$$\hat{\alpha}^{r_1 r_2 \cdots r_k} = n^{-1} \sum_{i=1}^{n} Y_i^{r_1} Y_i^{r_2} \cdots Y_i^{r_k}$$

for $k = 1, 2, \ldots$ and $1 \le r_1, r_2, \ldots, r_k \le d$. We propose estimating the ξ_i's by

$$\tilde{\xi}_1 = \frac{2-\gamma-\gamma^2}{4}\frac{\tilde{\alpha}^{rrkk}}{\tilde{\alpha}^{rr,kk}} - \frac{\gamma^2+2\gamma+3}{6}\frac{\tilde{\alpha}^{rkm,rkm}}{\tilde{\alpha}^{rr,kk,mm}} + \frac{(1+\gamma)^2}{6}\frac{\tilde{\alpha}^{rkk,rll}}{\tilde{\alpha}^{rr,kk,ll}},$$

$$\tilde{\xi}_2 = \frac{1-\gamma^2}{4}\frac{\tilde{\alpha}^{rrkk}}{\tilde{\alpha}^{rr,kk}} + \frac{2\gamma^2+3\gamma+1}{3}\frac{\tilde{\alpha}^{rkm,rkm}}{\tilde{\alpha}^{rr,kk,mm}} - \frac{(1+\gamma)^2}{3}\frac{\tilde{\alpha}^{rkk,rll}}{\tilde{\alpha}^{rr,kk,ll}} + \frac{(1+\gamma)^2}{4}(d^2+2d),$$

$$\tilde{\xi}_3 = \frac{5}{12}(1+\gamma)^2\frac{\tilde{\alpha}^{rkk,rll}}{\tilde{\alpha}^{rr,kk,ll}},$$

where the $\tilde{\alpha}$'s are given as follows:

Parameter	Estimator	Expression
α^{rr}	$\tilde{\alpha}^{rr}$	$n\hat{\alpha}^{rr}/(n-1)$
α^{rrss}	$\tilde{\alpha}^{rrss}$	$\{n\hat{\alpha}^{rrss} - 2\tilde{\alpha}^{rr}\tilde{\alpha}^{ss} - 4\delta^{rs}\tilde{\alpha}^{rr}\tilde{\alpha}^{rr}\}/(n-4)$
α^{rst}	$\tilde{\alpha}^{rst}$	$n\hat{\alpha}^{rst}/(n-3)$
$\alpha^{rst}\alpha^{rst}$	$\tilde{\alpha}^{rst,rst}$	$\tilde{\alpha}^{rst}\tilde{\alpha}^{rst} - (\hat{\alpha}^{rrsstt} - \tilde{\alpha}^{rst}\tilde{\alpha}^{rst})/n$
$\alpha^{rss}\alpha^{rtt}$	$\tilde{\alpha}^{rss,rtt}$	$\tilde{\alpha}^{rss}\tilde{\alpha}^{rtt}$
$\alpha^{rr}\alpha^{ss}$	$\tilde{\alpha}^{rr,ss}$	$\tilde{\alpha}^{rr}\tilde{\alpha}^{ss} - \tilde{\alpha}^{rrss}/n$
$\alpha^{rr}\alpha^{ss}\alpha^{tt}$	$\tilde{\alpha}^{rr,ss,tt}$	$\tilde{\alpha}^{rr}\tilde{\alpha}^{ss}\tilde{\alpha}^{tt}$

We estimate $b(t)$ according to Equation (12) by

$$\tilde{b}(x) = \left(\frac{\tilde{\xi}_1}{d}\right) + \left\{\frac{\tilde{\xi}_2}{d(d+2)}\right\} x + \left[\frac{\tilde{\xi}_3}{\{d(d+2)(d+4)\}}\right] x^2.$$

3.3. *Extension*

Conceptually, the new technique is applicable to other pivotal quantities with chi-square limiting distributions. We state this result without a proof in Theorem 3.2.

THEOREM 3.2 *Suppose we have a univariate pivotal quantity $Q_n = nT_n^{\mathrm{T}}T_n + O_p(n^{-3/2})$ such that*

$$P(Q_n \le y) = P(nT_n^{\mathrm{T}}T_n \le y) + O(n^{-2}).$$

Suppose further that $n^{1/2}T_n$ admits an Edgeworth expansion up to order $o(n^{-2})$ and that for $1 \le r, s, t, u \le d$, its cumulants can be expanded as

$$\mathrm{cum}(n^{1/2}T_n^r) = n^{-1/2}\beta_1^r + n^{-3/2}c_1^r + o(n^{-2}),$$

$$\mathrm{cum}(n^{1/2}T_n^r, n^{1/2}T_n^s) = \delta^{rs} + n^{-1}\beta_2^{rs} + n^{-2}c_2^{rs} + o(n^{-2}),$$

$$\mathrm{cum}(n^{1/2}T_n^r, n^{1/2}T_n^s, n^{1/2}T_n^t) = n^{-1/2}\beta_3^{rst} + n^{-3/2}c_3^{rst} + o(n^{-2}),$$

$$\mathrm{cum}(n^{1/2}T_n^r, n^{1/2}T_n^s, n^{1/2}T_n^t, n^{1/2}T_n^u) = n^{-1}\beta_4^{rstu} + n^{-2}c_4^{rstu} + o(n^{-2}),$$

with $c_1^r, c_2^{rs}, c_3^{rst}, c_4^{rstu}$ being nonrandom constants. In addition, the cumulants of $n^{1/2}T_n$ of order 5 and 6 are $o(n^{-2})$.

Let $b^*(t) = (\zeta_1/d) + [\zeta_2/\{d(d+2)\}]t + [\zeta_3/\{d(d+2)(d+4)\}]t^2$ with

$$\zeta_1 = \sum_r (\beta_2^{rr} + \beta_1^r\beta_1^r) - \frac{1}{4}\sum_{r,s}(\beta_4^{rrss} + 4\beta_1^r\beta_3^{rss}) + \frac{5}{12}\sum_{r,s,t}\beta_3^{rrs}\beta_3^{stt},$$

$$\zeta_2 = \frac{1}{4}\sum_{r,s}(\beta_4^{rrss} + 4\beta_1^r\beta_3^{rss}) - \frac{5}{6}\sum_{r,s,t}\beta_3^{rrs}\beta_3^{stt},$$

$$\zeta_3 = \frac{5}{12}\sum_{r,s,t}\beta_3^{rrs}\beta_3^{stt}.$$

Then we have $P[Q_n/\{1 + b^*(t)/n\} \le t] = P(\chi_d^2 \le t) + O(n^{-2})$.

This condition-rich theorem primarily serves a conceptual purpose. The F-statistic constructed in the analysis of variance can be shown to satisfy these conditions. Thus, when the data are not normal, a level-specific correction is possible. We do not go as far as to recommend its use in the analysis of variance.

4. Simulation and an example

In this section, we investigate the performance of the level-specific corrected nonparametric likelihoods. We examine eight representative nonparametric likelihoods corresponding to $\gamma = -2.5, -2.0, -1.5, -1.0, -0.5, 0.0, 0.5$, and 1.0. Confidence regions for the population mean are constructed at the nominal levels 90%, 95%, and 99%.

For the scalar observations, we generated data from two distributions: the standard normal $N(0,1)$ and the χ_3^2 with sample sizes $n = 20$ and 50. In the multivariate case, we chose the bivariate distributions (b) and (c) of Liu and Chen (2010) with the data generated as follows. A random observation D was first generated from the uniform distribution on the interval [1, 2]. Given D, we generated a datum $X = (X_{(1)}, X_{(2)})^{\mathrm{T}}$ from one of the following distributions: (b) $X_{(1)} \sim \text{Gamma}(D, 1)$, $X_{(2)} \sim \text{Gamma}(D^{-1}, 1)$, and (c) $X_{(1)} \sim 0.2N(5, D^2) + 0.8N(-1.25, D^{-2})$, $X_{(2)} \sim 0.2N(5, D^{-2}) + 0.8N(-1.25, D^2)$. The sample sizes are chosen to be $n = 30$ and 80.

We computed the coverage rates of the confidence regions for the population means based on two approaches, straight nonparametric likelihoods ($L_\gamma(\mu)$) and the nonparametric likelihoods with level-specific correction ($L_\gamma(\mu)/\{1 + \tilde{b}(t)/n\}$). The simulation size is 10,000 (Tables 1 and 2).

The straight nonparametric likelihood confidence regions with $\gamma = -1, -0.5, 0, 0.5$, and 1.0 tend to have lower than nominal coverage rates. For these likelihoods, the level-specific correction markedly improves the precision of the coverage probabilities. The improvement in the coverage precision is more apparent when the population is skewed. When the sample size increases, the correction still helps but is not really necessary.

When $\gamma = -2.0$ and -2.5 the level-specific correction is counterproductive. A superficial reason is that the level-specific correction factor b becomes negative for nonparametric likelihoods with γ values in this range. Because of this, the coverage rate decreases after the correction. Such a correction is warranted only if the corresponding nonparametric likelihoods have higher than nominal coverage probabilities when the chi-square distribution is used for calibration. However, over-coverage does not occur for these nonparametric likelihoods as might be suggested by the high-order expansion. Being confident in our mathematics, we regretfully conclude that the expansion does not kick in for the current small sample size. This is of course the case as pointed

Table 1. Simulated coverage probabilities (%) of nonparametric-likelihood-based confidence intervals and their level-specific counterparts for the population mean.

Population	n		γ	-2.5	-2.0	-1.5	-1.0	-0.5	0.0	0.5	1.0
N(0,1)	20	U	90	90.03	89.26	88.69	88.20	87.92	87.80	87.83	88.01
			95	94.71	94.26	93.91	93.50	93.26	93.04	93.02	93.24
			99	98.56	98.46	98.25	98.06	97.92	97.84	97.79	98.15
		C	90	86.11	87.92	88.04	89.49	89.86	89.95	89.97	89.93
			95	90.88	92.91	92.89	94.37	94.55	94.77	94.97	95.12
			99	94.83	97.23	97.98	98.48	98.74	98.88	99.02	99.32
	50	U	90	90.67	89.98	89.71	89.09	88.90	88.79	88.78	88.80
			95	96.19	95.50	95.24	94.72	94.43	94.30	94.31	94.37
			99	99.36	99.25	99.00	98.93	98.81	98.79	98.80	98.79
		C	90	89.24	89.45	90.30	89.71	89.70	89.75	89.76	89.71
			95	94.80	94.98	95.23	95.04	95.05	95.16	95.17	95.17
			99	99.01	99.06	98.93	99.06	99.03	99.02	99.08	99.16
χ^2_3	20	U	90	87.27	86.84	86.48	86.25	85.94	85.95	85.81	85.78
			95	92.70	92.50	92.19	91.84	91.42	91.26	91.14	90.93
			99	97.66	97.52	97.29	96.99	96.67	96.31	95.99	95.91
		C	90	80.57	84.22	86.36	88.07	88.74	88.85	88.79	88.50
			95	86.27	89.45	92.04	93.05	93.43	93.47	93.52	93.42
			99	92.91	95.05	96.97	97.51	97.74	97.82	97.91	98.04
	50	U	90	89.91	89.52	89.06	88.89	88.74	88.52	88.50	88.56
			95	95.10	94.74	94.49	94.26	94.09	93.92	93.67	93.55
			99	98.92	98.80	98.75	98.47	98.27	98.14	98.07	97.95
		C	90	86.91	88.38	89.19	89.62	89.94	90.01	90.07	89.95
			95	92.56	93.89	94.47	94.84	94.94	95.01	95.03	95.00
			99	97.62	98.18	98.73	98.69	98.74	98.75	98.74	98.77

Notes: U, uncorrected nonparametric likelihood ($L_\gamma(\mu)$); C, level-specific corrected nonparametric likelihood ($L_\gamma(\mu; n^{-2})/\{1 + \tilde{b}(x)/n\}$).

out by Tsao (2004). Yet when the sample size becomes large, high-order correction is no longer needed. Clearly, the high-order asymptotics fail to provide a useful guidance to a finite-sample problem here.

In conclusion, we have a rigorous proof of concept for the level-specific correction. The simulation shows that the technique works well only for nonparametric likelihoods with $\gamma \geq -1$. Fortunately, this range of γ includes most popular nonparametric likelihoods such as the empirical likelihood, the exponential tilting likelihood, and the Euclidean likelihood introduced earlier.

We next illustrate the proposed level-specific correction procedure using sweat data from Johnson and Wichern (2007, p. 215). The data set consists of perspiration from 20 healthy females in terms of sweat rate and sodium and potassium content; it is given in Table 3. We are interested in constructing confidence regions for the mean pair of sodium and potassium content in healthy females' sweat.

We constructed 90% and 99% confidence regions for the population mean of sodium and potassium content based on the empirical and Euclidean likelihoods ($\gamma = -1.0, 1.0$) and their level-specific corrected regions. The results are given in Figure 1. The top two plots reveal the differences in the sizes of the corrections. Because the Bartlett-correctable empirical likelihood ($\gamma = -1$) is not level dependent, the correction expands the two confidence regions at the same rate. In comparison, the correction for the Euclidean likelihood is heavily level dependent: its 99% region is expanded much more than its 90% region. The bottom two plots show another aspect of the comparison. At the 90% level, the two likelihoods give near-identical confidence regions, before or after the correction. At the 99% level and before the correction, the empirical likelihood confidence region is shifted to the north-west compared to the Euclidean one, but the two regions have roughly the same size. The correction has little effect on the empirical likelihood but a

Table 2. Simulated coverage probabilities (%) of nonparametric-likelihood-based confidence regions and their level-specific counterparts for bivariate population means.

Population	n		γ	−2.5	−2.0	−1.5	−1.0	−0.5	0.0	0.5	1.0
(b)	30	U	90	86.47	86.01	84.93	84.81	84.52	83.85	83.45	83.40
			95	91.78	91.64	91.15	90.43	89.87	89.17	88.72	88.59
			99	97.33	97.19	96.86	96.47	95.96	95.41	94.78	94.44
		C	90	75.51	81.65	84.95	87.30	88.04	88.33	88.26	87.81
			95	81.57	87.26	90.87	92.32	92.91	93.13	93.31	93.18
			99	89.95	93.84	96.34	97.25	97.61	97.85	97.96	98.11
	80	U	90	90.01	89.31	88.67	88.49	88.22	88.02	87.71	87.27
			95	95.23	94.81	94.49	94.00	93.31	92.98	92.70	92.52
			99	99.07	98.95	98.70	98.60	98.37	97.98	97.67	97.43
		C	90	86.50	87.96	88.68	89.49	89.88	89.91	89.83	89.80
			95	92.60	93.95	93.47	94.66	94.83	94.80	94.89	94.78
			99	98.15	98.55	98.31	98.84	98.94	98.96	98.91	98.97
(c)	30	U	90	89.16	88.63	87.38	87.03	86.43	85.37	84.72	84.24
			95	93.63	93.41	92.99	92.47	91.70	90.74	90.13	89.67
			99	97.98	97.92	97.70	97.52	97.05	96.31	95.65	95.16
		C	90	82.86	85.51	87.43	89.40	89.98	89.92	89.66	89.16
			95	88.59	90.83	92.62	94.00	94.73	94.80	94.77	94.31
			99	94.97	96.03	97.29	98.23	98.62	98.77	98.77	98.72
	80	U	90	90.60	89.90	89.75	89.05	88.51	87.98	87.74	87.57
			95	95.40	95.11	95.02	94.13	93.79	93.47	93.05	92.70
			99	99.11	99.06	98.96	98.77	98.46	98.21	97.84	97.64
		C	90	88.91	89.23	89.79	89.64	89.67	89.67	89.51	89.53
			95	94.38	94.48	94.89	94.63	94.67	94.71	94.80	94.80
			99	98.78	98.93	98.93	98.90	98.97	99.05	99.11	99.16

Notes: U, uncorrected nonparametric likelihood ($L_\gamma(\mu)$); C, level-specific corrected nonparametric likelihood ($L_\gamma(\mu; n^{-2})/\{1 + \tilde{b}(x)/n\}$).

Table 3. Sweat data set.

Female No.	1	2	3	4	5	6	7	8	9	10
Sodium content	48.5	65.1	47.2	53.2	55.5	36.1	24.8	33.1	47.4	54.1
Potassium content	9.3	8.0	10.9	12.0	9.7	7.9	14.0	7.6	8.5	11.3
Female No.	11	12	13	14	15	16	17	18	19	20
Sodium content	36.9	58.8	27.8	40.2	13.5	56.4	71.6	52.8	44.1	40.9
Potassium content	12.7	12.3	9.8	8.4	10.1	7.1	8.2	10.9	11.2	9.4

large effect on the Euclidean likelihood. The Euclidean confidence region is markedly expanded after the correction. We note that after the correction, the Euclidean confidence region covers the empirical likelihood region. Is this justified? Recall the asymptotic equivalence between the Euclidean likelihood and Hotelling's T^2. The latter gives a completely symmetric confidence region regardless of the data. Thus, when the data are not symmetric, the correction has to expand the original confidence region more to attain the nominal level of coverage. The last plot may thus indicate the superiority of the empirical likelihood in constructing confidence intervals for the population mean.

The properties of the level-specific corrected nonparametric confidence regions for $-1 < \gamma < 1$ or at the 95% level can be interpolated from what we have presented. They are not included to save space and to make the plots easy to read.

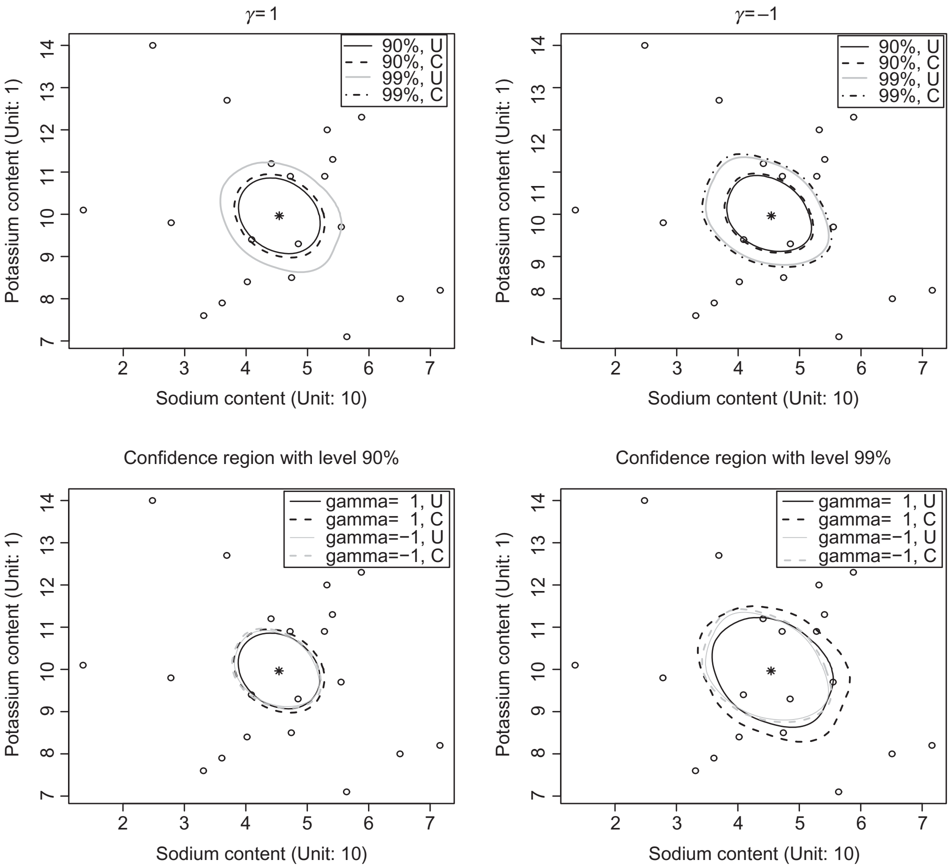

Figure 1. Nonparametric likelihood confidence regions and their level-specific counterparts for the population mean of the sweat data. The circles and the star stand for the observations and the sample mean, respectively.

5. Discussion

It is well known that of all the nonparametric likelihoods in the Cressie–Read family, only the empirical likelihood is Bartlett correctable. In this paper, we have established a proof of concept for the level-specific correctability of all the Cressie–Read nonparametric likelihoods. Simulations based on constructing confidence regions for the population mean show that the level-specific correction helps to improve the coverage precision for many important nonparametric likelihoods, but it is counterproductive for nonparametric likelihoods with $\gamma < -1$.

Another way to improve the precision of the empirical likelihood confidence regions is via the adjusted empirical likelihood proposed by Chen et al. (2008). The adjusted empirical likelihood was invented to overcome the empty-solution problem which may occur when the parameter is defined by over-identified estimating equations (Grendar and Judge 2009; Bergsma, Croon, and van der Ark 2012; Tsao 2013). By tuning its level of adjustment according to the Bartlett correction factor, Liu and Chen (2010) prove that the adjusted empirical likelihood confidence regions also have high-order coverage precision. Can nonparametric likelihoods be adjusted to achieve high-order coverage precision? The answer is positive and the theory parallels the Bartlett correctability of the nonparametric likelihoods. We do not bother readers with the repetitive details and tedious algebra.

Acknowledgements

The authors thank the editor and two anonymous referees for helpful comments, which have significantly improved the quality of this paper.

Funding

The research was supported by the National Natural Science Foundation of China [grant numbers 11371142, 11171112, and 11101156] and Chinese Ministry of Education, the 111 Project [B14019]. The research of Jiahua Chen was partially supported by the Natural Science and Engineering Research Council of Canada and by a UBC Killam Faculty Research Fellowship.

References

Antoine, B., Bonnal, H., and Renault, E. (2007), 'On the Efficient Use of the Informational Content of Estimating Equations: Implied Probabilities and Euclidean Empirical Likelihood', *Journal of Econometrics*, 138, 461–487.

Baggerly, K.A. (1998), 'Empirical Likelihood as a Goodness-of-Fit Measure', *Biometrika*, 85, 535–547.

Bahr, B.V. (1967), 'On the Central Limit Theorem in R_k', *Arkiv fór Matematik*, 7, 61–69.

Bergsma, W., Croon, M., and van der Ark, L.A. (2012), 'The Empty Set and Zero Likelihood Problems in Maximum Empirical Likelihood Estimation', *Electronic Journal of Statistics*, 6, 2356–2361.

Bhattacharya, R.N., and Ghosh, J.K. (1978), 'On the Validity of the Formal Edgeworth Expansion', *Annals of Statistics*, 6, 434–451.

Chen, S.X., and Cui, H.J. (2006), 'On Bartlett Correction of Empirical Likelihood in Presence of Nuisance Parameters', *Biometrika*, 93, 215–220.

Chen, S.X., and Cui, H.J. (2007), 'On the Second-Order Properties of Empirical Likelihood with Moment Restrictions', *Journal of Econometrics*, 141, 492–516.

Chen, J., and Qin, J. (1993), 'Empirical Likelihood Estimation for Finite Populations and the Effective Usage of Auxiliary Information', *Biometrika*, 80, 107–116.

Chen, J., Variyath, A.M., and Abraham, B. (2008), 'Adjusted Empirical Likelihood and Its Properties', *Journal of Computational and Graphical Statistics*, 17, 426–443.

Corcoran, S.A. (1998), 'Bartlett Adjustment of Empirical Discrepancy Statistics', *Biometrika*, 85, 967–972.

Cressie, N., and Read, T.R.C. (1984), 'Multinomial Goodness-of-Fit Tests', *Journal of the Royal Statistical Society: Series B*, 46, 440–464.

DiCiccio, T.J., Hall, P., and Romano, J.P. (1991), 'Empirical Likelihood is Bartlett-Correctable', *Annals of Statistics*, 19, 1053–1061.

Efron, B. (1981), 'Nonparametric Standard Errors and Confidence Intervals (with Discussion)', *Canadian Journal of Statistics*, 9, 139–172.

Grendar, M., and Judge, G. (2009), 'Empty Set Problem of Maximum Empirical Likelihood Methods', *Electronic Journal of Statistics*, 3, 1542–1555.

Imbens, G., Johnson, P., and Spady, R. (1998), 'Information Theoretic Approaches to Inference in Moment Condition Models', *Econometrica*, 66, 333–357.

Jing, B.Y., and Wood, A.T.A. (1998), 'Exponential Empirical Likelihood Is Not Bartlett Correctable', *Annals of Statistics*, 24, 365–369.

Johnson, R.A. and Wichern, D.W. (2007), *Applied Multivariate Statistical Analysis*, Upper Saddle River, NJ: Prentice Hall Press.

Lahiri, S.N., and Mukhopadhyay, S. (2012), 'A Penalized Empirical Likelihood Method in High Dimensions', *Annals of Statistics*, 40, 2511–2540.

Liu, Y., and Chen, J. (2010), 'Adjusted Empirical Likelihood with High-Order Precision', *Annals of Statistics*, 38, 1341–1362.

Matsushita, Y., and Otsu, T. (2013), 'Second-Order Refinement of Empirical Likelihood for Testing Overidentifying Restrictions', *Econometric Theory*, 29, 324–353.

Owen, A.B. (1988), 'Empirical Likelihood Ratio Confidence Intervals for a Single Functional', *Biometrika*, 75, 237–249.

Owen, A.B. (1991), 'Empirical Likelihood for Linear Models', *Annals of Statistics*, 19, 1725–1747.

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

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

Tsao, M. (2004), 'Extending the Empirical Likelihood by Domain Expansion', *Annals of Statistics*, 32, 1215–1221.

Tsao, M. (2013), 'Extending the Empirical Likelihood by Domain Expansion', *The Canadian Journal of Statistics*, 40, 1–8.

Tsao, M., and Wu, F. (2013), 'Empirical Likelihood on the Full Parameter Space', *Annals of Statistics*, 41, 2176–2196.

Appendix. Proof of Theorem 3.1

There are two main steps in the proof of Theorem 1. In the first step, we seek a smooth function, say $T(Z)$, of the sum Z of independent and identically distributed random vectors, such that the nonparametric likelihood ratio function

$$L_\gamma(\mu_0) = n\{T(Z)\}^{\mathrm{T}}\{T(Z)\} + \epsilon_n$$

and the remainder

$$P(|\epsilon_n| \geq c_1 n^{-5/2}\{\log(n)\}^{7/2}) \leq c_2 n^{-2}\{\log(n)\}^{-3}$$

for some constants c_1 and c_2. We will write the above result as $\epsilon_n = O_p^*(n^{-5/2}\{\log(n)\}^{7/2})$ with the $*$ emphasising that the probability of exception is of the order $n^{-2}\{\log n\}^{-3}$. We will use o^* and O^* in this spirit and liberally.

In the second step, we obtain a formal Edgeworth expansion of $T(Z)$ to order $o(n^{-2})$, and infer that it leads to an expansion of the distribution function of $L_\gamma(\mu_0)$ with the same precision, as assured by Theorem 2 of Bhattacharya and Ghosh (1978).

Since the index γ is not relevant in the proof, it will sometimes be omitted in the presentation. Define

$$H(\lambda) = \frac{2}{\gamma(\gamma+1)} \sum_{i=1}^{n} \{(1 + \gamma\lambda^{\mathrm{T}} Y_i)^{1/\gamma+1} - 1\},$$

where $Y_i^{\mathrm{T}} = (1, (X_i - \mu_0)^{\mathrm{T}})$. Let $\hat{\lambda}$ be the solution to the following equation:

$$G(\lambda) \equiv \frac{1}{n} \sum_{i=1}^{n} (1 + \gamma\lambda^{\mathrm{T}} Y_i)^{1/\gamma} Y_i = e_1, \tag{A1}$$

where e_r is a vector of length $d+1$ with the rth component being one and the remaining components being zero. Solving the optimisation problem in Equation (2) by the Lagrange multiplier method, we obtain a simplified form of the nonparametric likelihood ratio $L_\gamma(\mu_0) = H(\hat{\lambda})$. See Baggerly (1998) for a detailed derivation.

A.1 The magnitude of $\hat{\lambda}$ and its polynomial expansion

The sample means of a set of independent and identically distributed random variables are known to differ from their expectation by $O_p(n^{-1/2})$ in general. The following lemma of Bahr (1967) gives a more precise order assessment, and it will be used in our proof.

Lemma A1 *Suppose $X_1, X_2, \ldots, X_n$ are independent and identically distributed random vectors with mean zero. Let $\bar{X}$ denote the sample mean. If $\mathbb{E}(\|X_1\|^k) < \infty$ for some integer $k \geq 3$, then there exist positive constants c_1 and c_2 such that*

$$P(\|\bar{X}\| > c_1 n^{-1/2}\{\log(n)\}^{1/2}) < c_2 n^{-(k-2)/2}\{\log(n)\}^{-k/2}.$$

As a first step in approximating $L_\gamma(\mu_0)$, we quantify the order of $\hat{\lambda}$ as follows:

$$\hat{\lambda} = O_p^*(n^{-1/2}\{\log(n)\}^{1/2}).$$

To begin with, we investigate the properties of the function $G(\lambda)$ defined in Equation (A1) when $\|\lambda\| \leq c_1 n^{-1/2}\{\log(n)\}^{1/2}$. Expanding $G(\lambda)$ at 0 to order k, we get

$$G(\lambda) = G(0) + n^{-1} \sum_{v=1}^{k-1} a_v \sum_{i=1}^{n} (\lambda^{\mathrm{T}} Y_i)^v Y_i + a_k n^{-1} \sum_{i=1}^{n} (\tilde{\lambda}^{\mathrm{T}} Y_i)^k Y_i, \tag{A2}$$

where the a_v's are generalised binomial coefficients and $\tilde{\lambda}$ is between 0 and λ.

In view of the size restriction on λ and for $k = 2$, we have

$$\left\| n^{-1} \sum_{i=1}^{n} (\tilde{\lambda}^{\mathrm{T}} Y_i)^2 Y_i \right\| \leq \|\tilde{\lambda}\|^2 n^{-1} \sum_{i=1}^{n} \|Y_i\|^3 = \|\tilde{\lambda}\|^2 O_p^*(1).$$

Note that $G(0) = \bar{Y}_n = O_p^*(n^{-1/2}\{\log(n)\}^{1/2})$ and the second term in the expansion has $a_1 = 1$. Hence,

$$n^{-1} a_1 \sum_{i=1}^{n} (\lambda^{\mathrm{T}} Y_i) Y_i = I_{d+1} + O_p^*(n^{-1/2}\{\log(n)\}^{1/2}),$$

where I_{d+1} denotes the $(d+1)$-dimensional identity matrix. The equation $G(\lambda) = e_1$ is therefore equivalent to

$$G(0) - e_1 + I_{d+1} + O_p^*(n^{-1}\{\log n\}) = \lambda, \tag{A3}$$

which leads to the claim $\|\hat{\lambda}\| = O_p^*(n^{-1/2}\{\log(n)\}^{1/2})$.

We now approximate $\hat{\lambda}$ by a polynomial of a sum of independent and identically distributed random vectors. Let $\tilde{\lambda}$ be the solution to $\Gamma(\lambda) = e_1$ with

$$\Gamma(\lambda) = G(0) + n^{-1} \sum_{\nu=1}^{5} a_\nu \sum_{i=1}^{n} (\lambda^{\mathrm{T}} Y_i)^\nu Y_i.$$

The difference between $\Gamma(\lambda)$ and $G(\lambda)$ is $O^*(\|\hat{\lambda}\|^6)$. This leads to $\|\hat{\lambda} - \tilde{\lambda}\| = O_p^*(n^{-3}\{\log(n)\}^3)$.

Clearly, $\Gamma(\lambda)$ is a polynomial in Y_i. Let Z_i be a stacked vector with components Y_i^r, $Y_i^r Y_i^s$, and so on up to degree 5, and let Z represent $n^{-1} \sum Z_i$. The solution to $\Gamma(\lambda) = e_1$ may be written as $\tilde{\lambda} = \tilde{W}(Z)$ for a sufficiently smooth function $\tilde{W}(\cdot)$.

Under the moment condition $\mathbb{E}\{\|X_1\|^{42}\} < \infty$, it follows from Lemma 1 that

$$\left\| n^{-1} \sum \{Z_i - \mathbb{E}(Z_i)\} \right\|^k = O^*(n^{-7/2}\{\log(n)\}^{7/2})$$

when $k \geq 7$. Expanding $\tilde{W}(Z)$ at $Z = E(Z)$ to order 5, we get a polynomial $W(Z)$ such that

$$\hat{\lambda} = W(Z) + O_p^*(n^{-3}\{\log(n)\}^3). \tag{A4}$$

A.2 High-order approximation of $L_\gamma(\mu_0)$

The goal of this subsection is to approximate the nonparametric likelihood ratio with a polynomial of Z, a sum of independent and identically distributed random vectors.

Clearly, $H(\hat{\lambda})$ also admits a binomial expansion:

$$L_\gamma(\mu_0) = H(\hat{\lambda}) = \sum_{\nu=1}^{6} a_\nu (\hat{\lambda}^{\mathrm{T}} Y_i)^\nu + a_7 (\hat{\lambda}^{\mathrm{T}} Y_i)^7, \tag{A5}$$

where the a_ν's are a combination of binomial coefficients and γ, and $\tilde{\lambda}$ lies between 0 and λ.

For the same reason as before, we have

$$a_7 (\hat{\lambda}^{\mathrm{T}} Y_i)^7 = O_p^*(n^{-5/2}\{\log(n)\}^{7/2}).$$

Let $\Omega(\hat{\lambda}) = \sum_{\nu=1}^{6} a_\nu (\hat{\lambda}^{\mathrm{T}} Y_i)^\nu$. It follows from Equation (A4) that

$$\Omega(\hat{\lambda}) = \Omega(W(Z)) + O_p^*(n^{-5/2}\{\log(n)\}^3).$$

Note that $\Omega(\lambda)$ is a polynomial in λ with coefficients that are linear functions of Z. Hence, $\Omega(W(Z))$ is a polynomial in Z, and we denote it $Q(Z)$:

$$L_\gamma(\mu_0) = Q(Z) + O_p^*(n^{-5/2}\{\log(n)\}^{7/2}). \tag{A6}$$

Because the lowest degree in the polynomial $Q(Z)$ is two, it admits a decomposition of the form

$$Q(z) = nT^{\mathrm{T}}(z)T(Z) + nR_3(Z), \tag{A7}$$

where $T(Z)$ is a polynomial in Z of degree 5, and the remainder $R_3(Z)$ contains only seventh-order or higher order monomials of Z.

Combining expansions (A6) and (A7), we have

$$L_\gamma(\mu_0) = n\{T(Z)\}^{\mathrm{T}} T(Z) + O_p^*(n^{-5/2}\{\log(n)\}^{7/2}). \tag{A8}$$

The algebra in this subsection has been used to demonstrate that $L_\gamma(\mu_0)$ has a decomposition as above. The exact form of $T(Z)$ has been ignored but will be worked out with patience and without much mathematical complexity. We will postpone the tedious details to the last moment to maintain readability.

A.3 Formal Edgeworth expansion

Let the formal expansion of the cumulative distribution function of $\sqrt{n}T(Z)$ be, symbolically,

$$\psi_{6,n}(x) = \left\{1 + \sum_{r=1}^{4} n^{-r/2}\pi_r(x)\right\}\phi(x)$$

with x a d-variate input variable, $\phi(x)$ the d-variate standard normal density function, and $\pi_r(x)$ some polynomials. Since $T(Z)$ is a smooth function of the sample means, Theorem 2 of Bhattacharya and Ghosh (1978) states that

$$\sup_{B\in\mathcal{B}} |P(\sqrt{n}T(Z) \in B) - \int_{x\in B} \psi_{6,n}(x)\,\mathrm{d}x| = o(n^{-2}) \tag{A9}$$

for any class $\mathcal{B}$ of Borel sets having a certain boundary property. This class contains all d-dimensional rectangles, spheres, and so on; we will not give a precise definition. This leads to

$$P(L_\gamma(\mu_0) \le y) = \int_{x^{\mathrm{T}}x\le y} \psi_{6,n}(x)\,\mathrm{d}x + o(n^{-2}) \tag{A10}$$

with $o(n^{-2})$ uniformly in y. That is, the distribution function of $L_\gamma(\mu_0)$ has an accurate and workable Edgeworth expansion.

It can be verified that $\pi_1(x)$ and $\pi_3(x)$ are odd polynomials so that $\int_{x^{\mathrm{T}}x\le y} \pi_r(x)\phi(x)\,\mathrm{d}x = 0$ for $r = 1$ and 3. Hence, the Edgeworth expansion (A10) is reduced to

$$\begin{aligned} P(L_\gamma(\mu_0) \le y) &= \int_{x^{\mathrm{T}}x\le y} \{1 + n^{-1}\pi_2(x) + n^{-2}\pi_4(x)\}\phi(x)\,\mathrm{d}x + o(n^{-2}) \\ &= \int_{x^{\mathrm{T}}x\le y} \{1 + n^{-1}\pi_2(x)\}\phi(x)\,\mathrm{d}x + O(n^{-2}). \end{aligned} \tag{A11}$$

The integration with respect to $\pi_4(x)$ has been absorbed into the $O(n^{-2})$ term.

We cannot postpone the tedious algebra much longer, so we now work out $T(Z)$. Its expression will be used to verify that $\pi_1(x)$ and $\pi_3(x)$ are odd polynomials and to work out the exact form of $\pi_2(x)$. Our task is simplified by quoting Baggerly (1998) that

$$\begin{aligned} n^{-1}L_\gamma(\mu_0) &= A^jA^j - A^{jk}A^jA^k + \frac{1-\gamma}{3}\alpha^{jkl}A^jA^kA^l + A^{jl}A^{kl}A^jA^k + \frac{1-\gamma}{3}A^{jkl}A^jA^kA^l \\ &\quad + (\gamma-1)\alpha^{jkm}A^{lm}A^jA^kA^l + \frac{(\gamma-1)^2}{4}\alpha^{jkn}\alpha^{lmn}A^jA^kA^lA^m \\ &\quad - \frac{(\gamma-1)(2\gamma-1)}{12}\alpha^{jklm}A^jA^kA^lA^m + \frac{(\gamma+1)^2}{4}A^jA^jA^kA^k + O_p(n^{-5/2}), \end{aligned}$$

where $\alpha^{j_1\cdots j_k}$ and $A^{j_1\cdots j_k}$ are defined in Equation (7). Note that the leading terms are polynomials of degree 4 in centred sample moments $A^{j_1\cdots j_k}$. The expansion matches the algebra sketched so far.

The following decomposition of $L_\gamma(\mu_0)$ can be confirmed by matching terms on the two sides:

$$L_\gamma(\mu_0) = nR_n^jR_n^j + O_p(n^{-3/2}),$$

where $R_n = R_{n1} + R_{n2} + R_{n3}$ and

$$\begin{aligned} R_{n1}^j &= A^j, \\ R_{n2}^j &= -\frac{1}{2}A^{jk}A^k + \frac{1-\gamma}{6}\alpha^{jkl}A^kA^l, \\ R_{n3}^j &= \frac{3}{8}A^{jl}A^{kl}A^k + \frac{1-\gamma}{6}A^{jkl}A^kA^l - \frac{5(1-\gamma)}{36}\alpha^{lkm}A^{jm}A^lA^k \\ &\quad - \frac{5(1-\gamma)}{18}\alpha^{jkm}A^{lm}A^kA^l + \frac{(1-\gamma)^2}{9}\alpha^{jkn}\alpha^{lmn}A^kA^lA^m \\ &\quad - \frac{(1-\gamma)(1-2\gamma)}{24}\alpha^{jklm}A^kA^lA^m + \frac{(1+\gamma)^2}{8}A^jA^kA^k. \end{aligned}$$

The R_n in the above decomposition is not $T(Z)$ stated in Equation (A8), which is a polynomial of degree 5. However, we must have $T(Z) = R_n + R_{n4} + R_{n5}$ with R_{n4} and R_{n5} being degree 4 and 5 entries of the A's. Because of the high degrees of R_{n4} and R_{n5}, the first and third cumulants of $T(Z)$ and R_n differ by $O(n^{-3/2})$, and the second and fourth cumulants differ by $O(n^{-2})$. At the same time, $\pi_2(x)$ in the formal Edgeworth expansion of $\sqrt{n}T(Z)$ is completely determined by its first four cumulants. That is, the exact expression for $\pi_2(x)$ for $\sqrt{n}T(Z)$ is found through that of the simpler R_n.

After tedious algebra, we obtain the cumulants of R_n as follows:

$$\begin{aligned}\mathrm{cum}(n^{1/2}R_n^r) &= n^{-1/2}\beta_1^r + O(n^{-3/2}),\\ \mathrm{cum}(n^{1/2}R_n^r, n^{1/2}R_n^s) &= \delta^{rs} + n^{-1}\beta_2^{rs} + O(n^{-2}),\\ \mathrm{cum}(n^{1/2}R_n^r, n^{1/2}R_n^s, n^{1/2}R_n^t) &= n^{-1/2}\beta_3^{rst} + O(n^{-3/2}),\\ \mathrm{cum}(n^{1/2}R_n^r, n^{1/2}R_n^s, n^{1/2}R_n^u, n^{1/2}R_n^v) &= n^{-1}\beta_4^{rsuv} + O(n^{-2}),\end{aligned}$$

where

$$\begin{aligned}\beta_1^r =& -\left\{\frac{(2+\gamma)}{6}\right\}\alpha^{rkk},\\ \beta_2^{rs} =& -\left\{\frac{(2\gamma^2+\gamma-3)}{4}\right\}\alpha^{rskk} + \left\{\frac{(3\gamma^2+4\gamma-1)}{6}\right\}\alpha^{skm}\alpha^{rkm}\\ &+ \left\{\frac{(1+\gamma)^2(2+d)}{4}\right\}\delta^{rs} + \left\{\frac{(8\gamma^2+14\gamma+5)}{36}\right\}\alpha^{rkk}\alpha^{sll},\\ \beta_3^{rst} =& -(\gamma+1)\alpha^{rsu},\\ \beta_4^{rsuv} =& (1-\gamma^2)\alpha^{rsuv} + (1+\gamma)^2(\delta^{rs}\delta^{uv} + \delta^{ru}\delta^{sv} + \delta^{rv}\delta^{us})\\ &+ \left\{\frac{2(\gamma+1)(2\gamma+1)}{3}\right\}(\alpha^{rsk}\alpha^{uvk} + \alpha^{ruk}\alpha^{svk} + \alpha^{rvk}\alpha^{suk}).\end{aligned} \tag{A12}$$

With these cumulants and the standard Edgeworth expansion technique, we obtain the functions $\pi_1(x)$ and $\pi_2(x)$ for $\sqrt{n}T(Z)$ as follows:

$$\begin{aligned}\pi_1(x) =& \sum_r \beta_1^r x^r + \frac{1}{6}\sum_{r,s,t}\beta_3^{rst}(x^r x^s x^t - [\delta^{rs}x^t]_3),\\ \pi_2(x) =& \frac{1}{2}\sum_{r,s}(\beta_2^{rs} + \beta_1^r\beta_1^s)(x^r x^s - \delta^{rs})\\ &+ \frac{1}{24}\sum_{r,s,t,u}(\beta_4^{rstu} + 4\beta_1^r\beta_3^{stu})(x^r x^s x^t x^u - [\delta^{rs}x^t x^u]_6 + [\delta^{rs}\delta^{tu}]_3)\\ &+ \frac{1}{72}\sum_{r,s,t,u,v,w}\beta_3^{rst}\beta_3^{uvw}(x^r x^s x^t x^u x^v x^w - [\delta^{rs}x^t x^u x^v x^w]_{15}\\ &+ [\delta^{rs}\delta^{tu}x^v x^w]_{45} - [\delta^{rs}\delta^{tu}\delta^{vw}]_{15}),\end{aligned}$$

where the superscripts run from 1 to d. Here $[\delta^{rs}x^t]_3$ and $[\delta^{rs}\delta^{tu}]_3$ denote, respectively, $\delta^{rs}x^t + \delta^{rt}x^s + \delta^{st}x^r$ and $\delta^{rs}\delta^{tu} + \delta^{rt}\delta^{su} + \delta^{ru}\delta^{su}$, similar to the definite of U-statistics. Expressions such as $[\cdot]_6$ are defined in the same way.

The expressions for $\pi_1(x)$ and $\pi_2(x)$ confirm our previous claim that they are, respectively, odd and even polynomials of x. We omit the expression for $\pi_3(x)$; it is odd but its exact expression is not needed.

Using the explicit expression of $\pi_2(x)$, we now compute the exact expression (A11) through integration. Let $\pi_{2k}(x)$ with $k = 1, 2, 3$ denote the three terms in $\pi_2(x)$, respectively. We calculate their integrals one by one.

Clearly, $\int_{x^{\mathrm{T}}x\le y} x^r x^s \phi(x)\,\mathrm{d}x = 0$ for $r \neq s$. Therefore,

$$\int_{x^{\mathrm{T}}x\le y}\pi_{21}(x)\phi(x)\mathrm{d}x = \frac{1}{2}\sum_{r=1}^{d}(\beta_2^{rr} + \beta_1^r\beta_1^r)\int_{x^{\mathrm{T}}x\le t}(x^r x^r - 1)\phi(x)\,r\mathrm{d}x.$$

For fixed $1 \le r \le d$, we find that

$$\int_{x^{\mathrm{T}}x\le y}\phi(x)\,\mathrm{d}x = F(y|d), \quad \int_{x^{\mathrm{T}}x\le y}x^r x^r\phi(x)\,\mathrm{d}x = F(y|d+2),$$

where $F(t|d)$ is the cumulative distribution function of the χ_d^2 distribution. Thus, we obtain

$$\int_{x^{\mathrm{T}}x\le y}\pi_{21}(x)\phi(x)\,\mathrm{d}x = \frac{1}{2}\sum_r(\beta_2^{rr} + \beta_1^r\beta_1^r)\{F(y|d+2) - F(y|d)\}.$$

Next we consider $\int_{x^{\mathrm{T}}x\leq y} \pi_{22}(x)\phi(x)\,\mathrm{d}x$. We observe that the integral $\int_{x^{\mathrm{T}}x\leq y} \phi(x)x^r x^s x^t x^u\,\mathrm{d}x = 0$ except for $r = s \neq t = u$ or its permutations, or when $r = s = t = u$. In the first case, the integral equals

$$\int_{x^{\mathrm{T}}x\leq y} \phi(x)x^1x^1x^2x^2\,\mathrm{d}x = F(y|d+4).$$

In the second case, the integral is equal to

$$\int_{x^{\mathrm{T}}x\leq y} \phi(x)x^1x^1x^1x^1\,\mathrm{d}x = 3F(y|d+4).$$

The other integrals in $\int_{x^{\mathrm{T}}x\leq y} \pi_{22}(x)\phi(x)\,\mathrm{d}x$ can be calculated in a similar way. This leads to

$$\int_{x^{\mathrm{T}}x\leq y} \pi_{22}(x)\phi(x)\,\mathrm{d}x = \frac{1}{8}\sum_{r,s}(\beta_4^{rrss} + 4\beta_1^r\beta_3^{rss})\{F(y|d+4) - 2F(y|d+2) + F(y|d)\}.$$

Similarly,

$$\int_{x^{\mathrm{T}}x\leq y} \pi_{23}(x)\phi(x)\,\mathrm{d}x = \frac{5}{24}\sum_{r,s,t}\beta_3^{rrs}\beta_3^{stt}\{F(y|d+6) - 3F(y|d+4) + 3F(y|d+2) - F(y|d)\}.$$

The above integrals can be further simplified. According to the properties of the χ^2 distribution, we have $F(y|d+2) = F(y|d) - (2y/d)F'(y|d)$ and $F'(y|d+2) = (y/d)F'(y|d)$, where $F'(y|d)$ is the density function of the χ^2_d distribution. Applying these properties, we obtain

$$\int_{x^{\mathrm{T}}x\leq y} \pi_{21}(x)\phi(x)\,\mathrm{d}x = -yF'(y|d)\frac{1}{d}\sum_{r}(\beta_2^{rr} + \beta_1^r\beta_1^r),$$

$$\int_{x^{\mathrm{T}}x\leq y} \pi_{22}(x)\phi(x)\,\mathrm{d}x = -yF'(y|d)\left\{\frac{y}{d(d+2)} - \frac{1}{d}\right\}\frac{1}{4}\sum_{r,s}(\beta_4^{rrss} + 4\beta_1^r\beta_3^{rss}),$$

$$\int_{x^{\mathrm{T}}x\leq y} \pi_{23}(x)\phi(x)\,\mathrm{d}x = -yF'(y|d)\left\{\frac{y^2}{d(d+2)(d+4)} - \frac{2y}{d(d+2)} + \frac{1}{d}\right\}\frac{5}{12}\sum_{r,s,t}\beta_3^{rrs}\beta_3^{stt}.$$

Summing the above three terms gives

$$\int_{x^{\mathrm{T}}x\leq y} \pi_2(x)\phi(x)\,\mathrm{d}x = -yF'(y|d)\left\{\frac{\xi_1}{d} + \frac{\xi_2}{d(d+2)}y + \frac{\xi_3}{d(d+2)(d+4)}y^2\right\},$$

where

$$\xi_1 = \sum_r(\beta_2^{rr} + \beta_1^r\beta_1^r) - \frac{1}{4}\sum_{r,s}(\beta_4^{rrss} + 4\beta_1^r\beta_3^{rss}) + \frac{5}{12}\sum_{r,s,t}\beta_3^{rrs}\beta_3^{stt},$$

$$\xi_2 = \frac{1}{4}\sum_{r,s}(\beta_4^{rrss} + 4\beta_1^r\beta_3^{rss}) - \frac{5}{6}\sum_{r,s,t}\beta_3^{rrs}\beta_3^{stt},$$

$$\xi_3 = \frac{5}{12}\sum_{r,s,t}\beta_3^{rrs}\beta_3^{stt}.$$

The function $K(t)$ is then obtained from the above by replacing the β's with Equation (A12). This completes the proof of Theorem 1.