Statistica Sinica **13**(2003), 83-95

SMOOTHED EMPIRICAL LIKELIHOOD CONFIDENCE INTERVALS FOR THE DIFFERENCE OF QUANTILES

Wang Zhou and Bing-Yi Jing

Hong Kong University of Science and Technology

Abstract: This paper is concerned with inference for the population quantile difference (e.g., the interquartile range). Although normal approximations could be used for this purpose, they usually give poor performances unless the sample size is sufficiently large. In this paper, we investigate the use of the nonparametric likelihood method. Numerical results will be presented to compare its performance with other methods.

Key words and phrases: Bandwidth, confidence interval, quantile difference, smoothed empirical likelihood.

1. Introduction

Let $X_1, \ldots, X_n$ be a random sample from some unknown distribution function $F(x)$. For any distribution function G, we define its tth quantile by $G^{-1}(t) = \inf\{x : G(x) \geq t\}$, where $0 < t < 1$. In this paper, we are concerned with the inference for the quantile difference $\theta_0 = F^{-1}(q) - F^{-1}(p)$, where $0 < p < q < 1$. For instance, if we take $p = 1/4$ and $q = 3/4$, we obtain the interquartile range.

An obvious estimator for θ_0 is the sample quantile difference $\hat{\theta}_0 = F_n^{-1}(q) - F_n^{-1}(p)$, where F_n is the empirical distribution function given by $F_n(x) = n^{-1} \times \sum_{i=1}^n I\{X_i \leq x\}$, with $I\{\cdot\}$ being the indicator function. Therefore, one could use the sample quantile difference $\hat{\theta}_0$ (properly studentized) to construct a confidence interval for the population quantile difference θ_0. Under mild conditions, it can be shown that the (studentized) sample quantile difference $\hat{\theta}_0$ is approximately normally distributed. However, as our simulation results show (see Section 3), the performance of the normal approximation can be rather poor for small sample sizes.

We investigate some alternative nonparametric methods in the hope of improving the normal approximation method. The empirical likelihood method, introduced by Owen (1988, 1990), amounts to computing the profile likelihood of a general multinomial distribution supported on the data. The empirical likelihood enjoys some very nice properties. First, it avoids explicit studentization as this is done internally, and so is useful in cases where variance estimates are rather complicated or unstable. Second, the shape of the confidence region (in two or more dimensions) is determined automatically by the data configuration. Third, it is range preserving. For these reasons, the empirical likelihood has found application in smooth functions of means (DiCiccio, Hall and Romano (1991)), in nonparametric density and regression function estimation (Owen (1991), Chen (1996), Chen and Qin (2000)), in generalized linear models (Kolaczyk (1994)), in quantile estimation (Chen and Hall (1993)), in general estimating equation (Qin and Lawless (1994)), in dependent processes (Kitamura (1997)), and so on. For a more thorough review of the empirical likelihood method and its applications, the reader is referred to the recent monograph by Owen (2001).

Despite extensive studies of the empirical likelihood method, there has been little work on inference for population quantiles. Chen and Hall (1993) first applied the method of empirical likelihood to sample quantiles and obtained very accurate results. Smoothing has proved very useful in reducing the coverage errors for the empirical likelihood method in estimation of quantiles. In this paper, we investigate how to extend the empirical likelihood to the difference of quantiles of a certain distribution function, and compare it with the normal approximation method.

The paper is arranged as follows. In Section 2, we introduce the empirical likelihood method to the quantile difference problem. Some simulation results are given in Section 3 to compare the performances of the empirical likelihood and the normal approximation method. Proofs are deferred to Section 4.

2. Methodology and Main Results

2.1. Methodology

Let us first give some motivations for our definition of the empirical likelihood for $\theta(F)$. First notice that $\theta_0 = F^{-1}(q) - F^{-1}(p) = \theta(F)$. Let $(p_1, \ldots, p_n)$ be a probability vector, i.e., $\sum_{i=1}^n p_i = 1$ and $p_i \geq 0$ for $1 \leq i \leq n$. Let $\tilde{F}$ be the distribution function which assigns probability p_i at the ith observation X_i. Hence, $\theta(\widetilde{F}) = \widetilde{F}^{-1}(q) - \widetilde{F}^{-1}(p)$. Then, the empirical likelihood, evaluated at true parameter value θ_0, can be defined by

$$\tilde{L}(\theta_0) = \max_{\theta(\widetilde{F})=\theta_0, \sum p_i = 1} \prod_{i=1}^n p_i. \tag{2.1}$$

Similarly to Chen and Hall (1993), we replace $\tilde{F}$ by a smoothed version. Since $\tilde{F}$ assigns probability p_i at X_i, its formal density estimate can be given by

$$\tilde{f}_s(t) = \sum_{i=1}^n p_i \; h^{-1} K\left(\frac{t - X_i}{h}\right),$$

where $K(\cdot)$ is a kernel function and h is the bandwidth. (Note that when $p_i = 1/n$ for all i, this reduces to the usual kernel density estimate.) Therefore, the kernel distribution estimate is

$$\tilde{F}_s(x) = \int_{-\infty}^{x} \tilde{f}_s(t)dt = \sum_{i=1}^{n} p_i \int_{-\infty}^{x} \frac{1}{h} K\left(\frac{t - X_i}{h}\right) dt.$$

Replacing $\widetilde{F}$ in (2.1) by $\widetilde{F}_s$, our smoothed empirical likelihood, evaluated at true parameter value θ_0, can now be defined by

$$L(\theta_0) = \max_{\theta(\widetilde{F}_s)=\theta_0, \sum p_i = 1} \prod_{i=1}^{n} p_i.$$

Note that $\prod_{i=1}^{n} p_i$, subject to $\sum_{i=1}^{n} p_i = 1$, attains its maximum n^{-n} at $p_i = n^{-1}$. So we define the empirical likelihood ratio at θ_0 by

$$R(\theta_0) = L(\theta_0)/n^{-n} = \max_{\theta(\widetilde{F}_s)=\theta_0, \sum p_i = 1} \prod_{i=1}^{n} (np_i). \tag{2.2}$$

In the remainder of this section, we give an explicit expression for $R(\theta_0)$ in (2.2). First, let us introduce a new variable $\eta = \widetilde{F}_s^{-1}(p)$, i.e., $\widetilde{F}_s(\eta) = p$. Then it follows from $\theta(\widetilde{F}_s) = \theta_0$ that $\widetilde{F}_s(\eta + \theta_0) = q$. Therefore, we can rewrite $R(\theta_0)$ in (2.2) as

$$R(\theta_0) = \sup_{\eta, p_1, \ldots, p_n} \prod_{i=1}^{n} (np_i), \tag{2.3}$$

subject to

$$\sum_{i=1}^{n} p_i = 1, \qquad \widetilde{F}_s(\eta) = p, \qquad \widetilde{F}_s(\eta + \theta_0) = q. \tag{2.4}$$

Write

$$w_1(X_i, \eta) = \int_{-\infty}^{\eta} \frac{1}{h} K\left(\frac{t - X_i}{h}\right) dt - p, \tag{2.5}$$

$$w_2(X_i, \eta) = \int_{-\infty}^{\eta+\theta_0} \frac{1}{h} K\left(\frac{t - X_i}{h}\right) dt - q. \tag{2.6}$$

Using Lagrange multipliers, (2.3) is maximized subject to constraints (2.4) with

$$p_i = \frac{1}{n} \cdot \frac{1}{1 + \tilde{\lambda}_E w_1(X_i, \tilde{\eta}_E) + \tilde{t}_E w_2(X_i, \tilde{\eta}_E)}, \qquad i = 1, \ldots, n, \tag{2.7}$$

where $(\tilde{\lambda}_E, \tilde{t}_E, \tilde{\eta}_E)$ are solutions of (λ, t, η) to the equations

$$\frac{1}{n}\sum_{i=1}^{n}\frac{w_1(X_i,\eta)}{1+\lambda w_1(X_i,\eta)+tw_2(X_i,\eta)}=0, \tag{2.8}$$

$$\frac{1}{n}\sum_{i=1}^{n}\frac{w_2(X_i,\eta)}{1+\lambda w_1(X_i,\eta)+tw_2(X_i,\eta)}=0, \tag{2.9}$$

$$\frac{1}{nh}\sum_{i=1}^{n}\frac{\lambda K\left(\frac{\eta-X_i}{h}\right)+tK\left(\frac{\eta+\theta_0-X_i}{h}\right)}{1+\lambda w_1(X_i,\eta)+tw_2(X_i,\eta)}=0. \tag{2.10}$$

Equations (2.8)$-$(2.10) can be solved in two stages. First, fixing η, we can solve for λ and t from equations (2.8) and (2.9) and denote the solutions by $\lambda(\eta)$ and $t(\eta)$. Second, substituting these into equation (2.10), we can solve for η. Define

$$R(\theta_0,\eta)=\prod_{i=1}^{n}\frac{1}{1+\lambda(\eta)w_1(X_i,\eta)+t(\eta)w_2(X_i,\eta)}.$$

It is easy to see that $R(\theta_0)=\sup_\eta R(\theta_0,\eta)$. Hence

$$\log R(\theta_0)=\log R(\theta_0,\tilde{\eta}_E)=-\sum_{i=1}^{n}\log[1+\lambda(\tilde{\eta}_E)w_1(X_i,\tilde{\eta}_E)+t(\tilde{\eta}_E)w_2(X_i,\tilde{\eta}_E)], \tag{2.11}$$

where $\lambda(\tilde{\eta}_E), t(\tilde{\eta}_E)$ and $\tilde{\eta}_E$ are solutions to equations (2.8)$-$(2.10).

2.2. Main results

Before stating our main results, we first give some regularity conditions.

(i) Let $f(x)=F'(x)$. For some integer $r\geq 2$, $f^{(r-1)}(x)$ exists in a neighborhood of $\eta_p=F^{-1}(p)$ and $\eta_q=F^{-1}(q)$, and is continuous at η_p and η_q, respectively. Further assume that $f(\eta_p)f(\eta_q)>0$.

(ii) The kernel $K(\cdot)$ is bounded and compactly supported; $K^{(2)}$ exists and is bounded; assume that

$$\int u^jK(u)du=\begin{cases}1, & j=0,\\ 0, & 1\leq j\leq r-1,\\ C_0, & j=r,\end{cases}$$

where C_0 is some finite constant.

(iii) $nh^{4r}\to 0$, $n^{4s-1}h^4\to\infty$, as $n\to\infty$, where for some $1/3<s<1/2$.

Let us give some remarks about the conditions. Condition (i) requires that the distribution function F be sufficiently smooth in neighborhoods of η_p and η_q respectively. From a mathematical point of view, the restriction $f(\eta_p)f(\eta_q)>0$

ensures the asymptotic variance of the sample quantile difference is of order n^{-1}, as can be seen from (3.1) in Section 3. Without that assumption, the asymptotic theory is quite different. See Feldman and Tucker (1966). Condition (ii) is a typical requirement for the kernel function $K(\cdot)$ in nonparametric curve estimation. Finally, Condition (iii) implies that the convergence of the bandwidth h to zero is neither too fast nor too slow.

Theorem 2.1. *Assume that conditions* (i)$-$(iii) *hold. Then with probability* 1*, for sufficiently large n, there exists a solution $\tilde{\eta}_E$, $\lambda(\tilde{\eta}_E)$ and $t(\tilde{\eta}_E)$ to* (2.8)$-$(2.10)*, such that $R(\theta_0, \eta)$ attains its maximum value $R(\theta_0)$ at $\eta = \tilde{\eta}_E$. Furthermore, we have $-2 \log R(\theta_0, \tilde{\eta}_E) \rightarrow_L \chi_1^2$.*

The proof of Theorem (2.1) will be given in Section 4.

An approximate $(1-\alpha)$ level confidence interval for θ_0 can be taken as $I_{hc} = \{\theta : -2 \log R(\theta, \tilde{\eta}_E) \leq c\}$, where c is chosen to satisfy $P(\chi_1^2 \leq c) = 1 - \alpha$. From Theorem 2.1, we have $\lim_{n \to \infty} P\{\theta_0 \in I_{hc}\} = P(\chi_1^2 \leq c) = 1 - \alpha$.

3. Simulation Results

A Monte Carlo study was conducted to investigate the coverage accuracy of the empirical likelihood confidence interval. We generated 10,000 pseudorandom samples of various sizes from $F = \chi_1^2$. For each sample, we solved the equations (2.8)$-$(2.10) by routines in Numerical Recipes in C [Press, Flannery, Teukolsky and Vetterling (1989)]. After getting $\lambda(\tilde{\eta}_E)$, $t(\tilde{\eta}_E)$ and $\tilde{\eta}_E$, we calculated $\log R(\theta_0)$ using (2.11). The kernel function was $K(u) = \frac{15}{16}(1 - 2u^2 + u^4) I\{|u| \leq 1\}$, which satisfies Condition (ii) with $r = 2$. The bandwidths were $h = n^{-3/20}$, $n^{-1/10}$, and $n^{-1/2}$. Note that the first choice of bandwidths $h = n^{-3/20}$ satisfies Condition (iii), while the last two choices do not, being either too large or too small in comparison with Condition (iii).

Confidence intervals by the normal approximation method can be obtained as follows. From Reiss (1989), we know $n^{1/2}\left(\hat{\theta}_0 - \theta_0\right) \rightarrow_L N\left(0, \sigma^2\right)$, where

$$\sigma^2 = \frac{p(1-p)}{f^2(\eta_p)} - \frac{2p(1-q)}{f(\eta_p) f(\eta_q)} + \frac{q(1-q)}{f^2(\eta_q)}. \tag{3.1}$$

A consistent estimator of σ^2, $\hat{\sigma}^2$, can be obtained by replacing $f(\eta_p)$ and $f(\eta_q)$ in the above formula by their empirical versions (or smoothed ones when appropriate). Thus, a two-sided confidence interval based on the normal approximations can be taken to be $I_{1-\alpha}^{(N)} = (\hat{\theta}_0 - d_{1-\alpha/2}\hat{\sigma}/\sqrt{n},\ \hat{\theta}_0 + d_{1-\alpha/2}\hat{\sigma}/\sqrt{n})$, where $d_{1-\alpha}$ is the $1-\alpha$ quantile of a standard normal distribution. For simplicity, in the simulation studies conducted here, we employ the true value of σ^2 rather than its consistent estimator $\hat{\sigma}^2$.

Table 1. $p = 0.25$, $q = 0.75$.

	nominal level	0.80	0.90	0.95	0.99
$n = 20$	$h = n^{-3/20}$	0.813	0.912	0.957	0.990
	$h = n^{-1/2}$	0.811	0.896	0.958	0.992
	$h = n^{-1/10}$	0.813	0.911	0.956	0.988
	normal approximation	0.842	0.941	0.971	0.990
$n = 30$	$h = n^{-3/20}$	0.809	0.906	0.956	0.992
	$h = n^{-1/2}$	0.807	0.913	0.950	0.990
	$h = n^{-1/10}$	0.811	0.905	0.954	0.990
	normal approximation	0.814	0.910	0.951	0.983
$n = 100$	$h = n^{-3/20}$	0.800	0.901	0.951	0.990
	$h = n^{-1/2}$	0.806	0.903	0.952	0.989
	$h = n^{-1/10}$	0.800	0.900	0.951	0.990
	normal approximation	0.807	0.925	0.954	0.988

Table 2. $p = 0.1$, $q = 0.9$.

	nominal level	0.80	0.90	0.95	0.99
$n = 20$	$h = n^{-3/20}$	0.825	0.918	0.960	0.984
	$h = n^{-1/2}$	0.854	0.922	0.974	0.996
	$h = n^{-1/10}$	0.844	0.923	0.959	0.988
	normal approximation	0.845	0.950	0.980	0.993
$n = 30$	$h = n^{-3/20}$	0.781	0.927	0.961	0.991
	$h = n^{-1/2}$	0.801	0.965	0.975	0.994
	$h = n^{-1/10}$	0.816	0.924	0.958	0.989
	normal approximation	0.826	0.933	0.973	0.993
$n = 100$	$h = n^{-3/20}$	0.809	0.905	0.952	0.989
	$h = n^{-1/2}$	0.801	0.965	0.975	0.994
	$h = n^{-1/10}$	0.794	0.894	0.944	0.986
	normal approximation	0.806	0.929	0.956	0.990

Table 3. $p = 0.2$, $q = 0.8$.

	nominal level	0.80	0.90	0.95	0.99
$n = 20$	$h = n^{-3/20}$	0.827	0.916	0.967	0.992
	$h = n^{-1/2}$	0.832	0.894	0.973	0.995
	$h = n^{-1/10}$	0.826	0.916	0.965	0.991
	normal approximation	0.782	0.873	0.915	0.960
$n = 30$	$h = n^{-3/20}$	0.795	0.903	0.955	0.993
	$h = n^{-1/2}$	0.780	0.904	0.946	0.992
	$h = n^{-1/10}$	0.800	0.905	0.955	0.993
	normal approximation	0.791	0.881	0.927	0.970
$n = 100$	$h = n^{-3/20}$	0.803	0.904	0.952	0.991
	$h = n^{-1/2}$	0.803	0.904	0.951	0.991
	$h = n^{-1/10}$	0.800	0.900	0.951	0.990
	normal approximation	0.794	0.912	0.942	0.982

We make the following observations from the numerical studies.

(1). The empirical likelihood method with bandwidth $h = n^{-3/20}$, which satisfies Condition (iii), has almost uniformly more accurate coverage probabilities than those using the other two choices of bandwidths.

(2). Although we have chosen a wide range of bandwidths, the coverage probabilities seem to change rather slowly. This indicates that the smoothed empirical likelihood is robust with respect to the bandwidth selection.

(3). For small sample size n, the normal approximation method seems to perform worse than the smoothed empirical likelihood methods. However, the advantage of the latter method gradually disappears when the sample sizes get large, as one might expect.

4. Proof of Theorem 2.1.

Let us first introduce a few lemmas. Throughout the proofs, we use C to denote a generic constant, which may assume some different value at each occurrence. Furthermore, we denote $\delta := h^r + n^{-s}$, $1/3 < s < 1/2$.

Lemma 4.1. *Assume that conditions* (i) *and* (ii) *hold. For each η satisfying $|\eta - \eta_p| \le \delta$, we have*

$$Ew_1(X_i, \eta) = [F(\eta) - F(\eta_p)] + O(h^r), \tag{4.1}$$

$$Ew_2(X_i, \eta) = [F(\eta + \theta_0) - F(\eta_q)] + O(h^r), \tag{4.2}$$

$$\text{Var}\,[w_1(X_i, \eta)] = F(\eta)\,[1 - F(\eta)] + O(h), \tag{4.3}$$

$$\text{Var}\,[w_2(X_i, \eta)] = F(\eta + \theta_0)\,[1 - F(\eta + \theta_0)] + O(h), \tag{4.4}$$

$$E\,[w_1(X_i, \eta) w_2(X_i, \eta)] = F(\eta) - pq + O(\delta). \tag{4.5}$$

The proof uses Taylor expansions, and is omitted.

Denote $\overline{w}_{11}(\eta) = \frac{1}{n}\sum_{i=1}^n w_1(X_i, \eta)$, $\overline{w}_{12}(\eta) = \frac{1}{n}\sum_{i=1}^n w_1^2(X_i, \eta)$, $\overline{w}_{21}(\eta) = \frac{1}{n}\sum_{i=1}^n w_2(X_i, \eta)$, $\overline{w}_{22}(\eta) = \frac{1}{n}\sum_{i=1}^n w_2^2(X_i, \eta)$, and $\overline{w}_3(\eta) = \frac{1}{n}\sum_{i=1}^n w_1(X_i, \eta) w_2(X_i, \eta)$.

Lemma 4.2. *Under the conditions* (i)$-$(iii), *uniformly for $\eta \in \{\eta : |\eta - \eta_p| \le \delta\}$, we have $\overline{w}_{11}(\eta) = O_p(\delta)$, $\overline{w}_{21}(\eta) = O_p(\delta)$, $\overline{w}_{12}(\eta) = p(1 - p) + O_p(\delta + h)$, $\overline{w}_{22}(\eta) = q(1 - q) + O_p(\delta + h)$, and $\overline{w}_3(\eta) = p(1 - q) + O_p(\delta + h)$.*

Proof. We only prove $\overline{w}_{11}(\eta) = O_p(\delta)$, others are shown similarly. Since

$$w_1(X_i, \eta) - w_1(X_i, \eta_p) = \frac{(\eta - \eta_p)}{h} K\left(\frac{\eta_p - X_i}{h}\right) + \frac{(\eta - \eta_p)^2}{2h^2} K'\left(\frac{\eta_i - X_i}{h}\right),$$

where η_i is between η_p and η, we have

$$\begin{aligned}\overline{w}_{11}(\eta) &= \overline{w}_{11}(\eta_p) + \frac{(\eta - \eta_p)}{nh} \sum_{i=1}^n K\left(\frac{\eta_p - X_i}{h}\right) + \frac{(\eta - \eta_p)^2}{2nh^2} \sum_{i=1}^n K'\left(\frac{\eta_i - X_i}{h}\right) \\ &=: \overline{w}_{11}(\eta_p) + \Delta_{n1} + \Delta_{n2}, \quad \text{say.}\end{aligned} \tag{4.6}$$

From Lemma 4.1, $Ew_1(X_1,\eta_p)=O(h^r)$. From this and the CLT, it follows that

$$\overline{w}_{11}(\eta_p)=Ew_1(X_1,\eta_p)+[\overline{w}_{11}(\eta_p)-Ew_1(X_1,\eta_p)]=O(h^r)+O_p(n^{-1/2})=O_p(\delta). \tag{4.7}$$

Now let us look at Δ_{n1}. As at (4.24) later in the paper, we can show that

$$\frac{1}{nh}\sum_{i=1}^{n}\left[K\left(\frac{\eta_p-X_i}{h}\right)-EK\left(\frac{\eta_p-X_i}{h}\right)\right]\to 0 \quad \text{a.s.} \tag{4.8}$$

Also note that

$$h^{-1}EK\left(\frac{\eta_p-X_1}{h}\right)=\int_{-\infty}^{\infty}K(x)f(\eta_p-hx)dx=f(\eta_p)+O(h). \tag{4.9}$$

It then follows from (4.8) and (4.9) that $\Delta_{n1}=O_p(\delta)$.

Finally, we turn to Δ_{n2}. Conditions (ii) and (iii) imply that

$$\left|K'\left(\frac{\eta-X_i}{h}\right)\right|\le C, \quad \text{and} \quad h^{-2}\delta\le h^{r-2}+h^{-2}n^{-s}=h^{r-2}+\frac{n^{s-1/2}}{(h^4n^{4s-1})^{1/2}}\le C.$$

It then follows easily that $\Delta_{n2}=O(\delta)$. The proof is complete.

Lemma 4.3. *Under the conditions* (i)$-$(iii), *for each η satisfying $|\eta-\eta_p|\le\delta$, as $n\to\infty$, we have*

$$\overline{w}_{11}(\eta)=O\left(\delta+h^r+n^{-1/2}(\log n)^{1/2}\right)=O(\delta) \quad a.s., \tag{4.10}$$

$$\overline{w}_{21}(\eta)=O\left(\delta+h^r+n^{-1/2}(\log n)^{1/2}\right)=O(\delta) \quad a.s., \tag{4.11}$$

$$\overline{w}_{12}(\eta)=p(1-p)+O\left(\delta+h+n^{-1/2}(\log n)^{1/2}\right)=p(1-p)+O(h) \quad a.s., \tag{4.12}$$

$$\overline{w}_{22}(\eta)=q(1-q)+O\left(\delta+h+n^{-1/2}(\log n)^{1/2}\right)=q(1-q)+O(h) \quad a.s., \tag{4.13}$$

$$\overline{w}_{3}(\eta)=p(1-q)+O\left(\delta+h+n^{-1/2}(\log n)^{1/2}\right)=p(1-q)+O(h) \quad a.s. \tag{4.14}$$

Proof. Recall the Bernstein inequality

$$P(|\bar{Y}-\bar{\mu}|\ge t)\le 2\exp\left(-\frac{nt^2}{2\mathrm{Var}\,(Y_1)+\frac{2}{3}mt}\right), \tag{4.15}$$

where $Y_1,\ldots,Y_n$ are i.i.d. r.v.'s satisfying $P(|Y_i-EY_i|\le m)=1$ and $t>0$. (see Serfling (1980), p.95) Choosing $Y_i=w_1^k(X_i,\eta)$ for $k=1,\ 2,$ and $t=dn^{-1/2}(\log n)^{1/2}$ for some $d>0$ to be determined later, we get

$$\begin{aligned}&\sum_{n=1}^{\infty}P\left(|\overline{w}_{1k}(\eta)-E\overline{w}_{1k}(\eta)|\ge dn^{-1/2}(\log n)^{1/2}\right)\\ &\quad\le 2\sum_{n=1}^{\infty}\exp\left(\frac{-d^2\log n}{2C+\frac{2}{3}dn^{-1/2}(\log n)^{1/2}}\right)\le 2\sum_{n=1}^{\infty}\exp\left(-2\log n\right)\end{aligned}$$

for d sufficiently large, and this is finite. By the Borel-Cantelli Lemma, $|\overline{w}_{1k}(\eta) - E\overline{w}_{1k}(\eta)| < dn^{-1/2}(\log n)^{1/2}$ a.s., $k = 1,\ 2$. Thus, (4.10) and (4.12) follows from these inequalities and Lemma 4.1. Similarly, we can prove (4.11), (4.13) and (4.14).

Lemma 4.4. *Under the conditions* (i)$-$(iii), *uniformly for* $\eta \in \{\eta : |\eta - \eta_p| \le \delta\}$, *we have*

$$\lambda(\eta) = O_p(\delta), \qquad t(\eta) = O_p(\delta). \tag{4.16}$$

Furthermore, on the boundary points, we have

$$\lambda(\eta_p \pm \delta) = O\left(\delta + h^r + n^{-1/2}(\log n)^{1/2}\right) \quad a.s., \tag{4.17}$$

$$t(\eta_p \pm \delta) = O\left(\delta + h^r + n^{-1/2}(\log n)^{1/2}\right) \quad a.s. \tag{4.18}$$

Proof. Let $w_i(\eta) = (w_1(X_i, \eta),\ w_2(X_i, \eta))^T$, $\rho = [\lambda^2(\eta) + t^2(\eta)]^{1/2}$, $\rho(\xi_1, \xi_2)^T = (\lambda(\eta),\ t(\eta))^T$, $\xi^T = (\xi_1, \xi_2)^T$, and $Z_n = \max_{1 \le i \le n} |w_i(\eta)|$. From (2.8)$-$(2.9), we have

$$\frac{\rho \xi^T S \xi}{1 + \rho Z_n} \le |\overline{w}_{11}(\eta)| + |\overline{w}_{21}(\eta)|, \quad \text{where} \quad S =: \begin{pmatrix} \overline{w}_{12}(\eta) & \overline{w}_3(\eta) \\ \overline{w}_3(\eta) & \overline{w}_{22}(\eta) \end{pmatrix}. \tag{4.19}$$

By Lemma 4.2, we have that, uniformly for $\eta \in \{\eta : |\eta - \eta_p| \le \delta\}$,

$$S = \begin{pmatrix} p(1-p) & p(1-q) \\ p(1-q) & q(1-q) \end{pmatrix} + O_p(\delta + h) =: S_0 + O_p(\delta + h). \tag{4.20}$$

Let σ_p be the minimal eigenroot of S_0. Then

$$\xi^T S_0 \xi \ge \sigma_p / 2. \tag{4.21}$$

We also have $Z_n \le C$. Combining (4.19)$-$(4.21), we obtain $\rho = O_p(\delta)$ uniformly for $\eta \in \{\eta : |\eta - \eta_p| \le \delta\}$. This proves (4.16).

Next we prove (4.17) (the proof of (4.18) is similar). It follows from Lemma 4.3 that $S = S_0 + O(\delta + h + n^{-1/2}(\log n)^{1/2})$ a.s. Hence if n is sufficiently large,

$$\xi^T S \xi \ge \sigma_p / 2 \quad \text{a.s.} \tag{4.22}$$

Also from (4.3),

$$\overline{w}_{j1}(\eta_p + \delta) = O(\delta + h^r + n^{-1/2}(\log n)^{1/2}) \quad \text{a.s.}, \quad j = 1,\ 2. \tag{4.23}$$

Combining (4.19), (4.22) and (4.23), we have $\rho = O(\delta)$ a.s. This completes the proof.

Lemma 4.5. *Under the conditions* (i)$-$(iii), *with probability one, for sufficiently large n, there exists a solution $\tilde{\eta}_E$, $\lambda(\tilde{\eta}_E)$ and $t(\tilde{\eta}_E)$ to* (2.8)$-$(2.10) *such that $R(\theta_0, \eta)$ attains its maximum value $R(\theta_0)$ at $\eta = \tilde{\eta}_E$.*

Proof. Let $\eta_1 = \eta_p + \delta$. From (2.8)$-$(2.9) and (4.17)$-$(4.18), we have $S(\eta_1)(\lambda(\eta_1), t(\eta_1))^T = (\overline{w}_{11}(\eta_1), \overline{w}_{21}(\eta_1))^T + O(\delta^2)$ a.s., where

$$S(\eta_1) =: \begin{pmatrix} \overline{w}_{12}(\eta_1) & \overline{w}_3(\eta_1) \\ \overline{w}_3(\eta_1) & \overline{w}_{22}(\eta_1) \end{pmatrix} = S_0 + O(\delta + h^r + n^{-1/2}(\log n)^{1/2}) \text{ a.s. (by Lemma 4.3).}$$

Thus, $(\lambda(\eta_1), t(\eta_1))^T = S^{-1}(\eta_1)(\overline{w}_{11}(\eta_1), \overline{w}_{21}(\eta_1))^T + O(\delta^2)$ a.s. Taylor expansion of $-\log R(\theta_0, \eta_1)$ gives

$$\begin{aligned}
&-\log R(\theta_0, \eta_1) \\
&= n\lambda(\eta_1)\overline{w}_{11}(\eta_1) + nt(\eta_1)\overline{w}_{21}(\eta_1) - \frac{1}{2}\sum_{i=1}^{n}[\lambda(\eta_1)w_1(X_i, \eta_1) + t(\eta_1)w_2(X_i, \eta_1)]^2 \\
&\quad + O(n\delta^3) \quad \text{a.s.} \\
&= \frac{n}{2}(\overline{w}_{11}(\eta_1), \overline{w}_{21}(\eta_1))S^{-1}(\eta_1)(\overline{w}_{11}(\eta_1), \overline{w}_{21}(\eta_1))^T + O(n\delta^3) \quad \text{a.s.} \\
&= \frac{n}{2}\left(\overline{w}_{11}(\eta_p) + (nh)^{-1}\sum_{i=1}^{n} K\left(\frac{\eta_i' - X_i}{h}\right)\delta, \overline{w}_{21}(\eta_p) + (nh)^{-1}\sum_{i=1}^{n} K\left(\frac{\eta_i'' - X_i}{h}\right)\delta\right) S^{-1}(\eta_1) \\
&\quad \times \left(\overline{w}_{11}(\eta_p) + (nh)^{-1}\sum_{i=1}^{n} K\left(\frac{\eta_i' - X_i}{h}\right)\delta, \overline{w}_{21}(\eta_p) + (nh)^{-1}\sum_{i=1}^{n} K\left(\frac{\eta_i'' - X_i}{h}\right)\delta\right)^T \\
&\quad + O(n\delta^3) \quad \text{a.s.,}
\end{aligned}$$

where η_i' is between η_p and $\eta_p + \delta$, η_i'' is between $\eta_p + \theta_0$ and $\eta_p + \theta_0 + \delta$. Note that the η_i''s are independent. Hence the Bernstein inequality implies

$$\begin{aligned}
&\sum_{n=1}^{\infty} P\left(\left|\frac{1}{nh}\sum_{i=1}^{n}\left[K\left(\frac{\eta_i' - X_i}{h}\right) - EK\left(\frac{\eta_i' - X_i}{h}\right)\right]\right| \geq dh\right) \\
&\qquad \leq 2\sum_{n=1}^{\infty}\exp\left(\frac{-nd^2h^4}{2C + \frac{2}{3}dh^2}\right) \leq 2\sum_{n=1}^{\infty}\exp(-2\log n) < \infty,
\end{aligned}$$

for d sufficiently large, and this is finite. We have used $nh^4n^{4s-2} \to \infty$ in the second to last inequality. By the Borel-Cantelli lemma, we get

$$\frac{1}{nh}\sum_{i=1}^{n}\left[K\left(\frac{\eta_i' - X_i}{h}\right) - EK\left(\frac{\eta_i' - X_i}{h}\right)\right] \to 0 \qquad \text{a.s.}$$

We also have

$$h^{-1}EK\left(\frac{\eta_i' - X_1}{h}\right) = h^{-1}\int_{-\infty}^{\infty} K\left(\frac{\eta_i' - x}{h}\right) f(x)dx$$

$$
\begin{aligned}
&= h^{-1}\int_{-\infty}^{\infty} K\left(\frac{\eta_i' - \eta_p}{h} + \frac{\eta_p - x}{h}\right) f(x)dx– \\
&= \int_{-\infty}^{\infty} K\left(\frac{\eta_i' - \eta_p}{h} + x\right) f(\eta_p - hx)dx– \\
&= \int_{-\infty}^{\infty} K\left(\frac{\eta_i' - \eta_p}{h} + x\right) f(\eta_p)dx + O(h)– \\
&= f(\eta_p) + O(h),
\end{aligned}
$$

where in the last equality, we have used $|\eta_i' - \eta_p/h| \leq \delta h^{-1} \leq Ch$ for each i. Hence

$$
\frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{\eta_i' - X_i}{h}\right) \to f(\eta_p) \quad \text{a.s.} \tag{4.24}
$$

Similarly,

$$
\frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{\eta_i'' - X_i}{h}\right) \to f(\eta_q) \quad \text{a.s.} \tag{4.25}
$$

From (4.24)$-$(4.25), and also noting that for $j = 1, 2$, $\overline{w}_{j1}(\eta_p) - E\overline{w}_{j1}(\eta_p) = O(n^{-1/2}(\log n)^{1/2})$ a.s., and $E\overline{w}_{j1}(\eta_p) = O(h^r)$, we have $-2\log R(\theta_0, \eta_1) \geq (C - \epsilon_n)n\delta^2$ a.s., where $\epsilon_n \to 0$. But $-2\log R(\theta_0, \eta_p) = n(\overline{w}_{11}(\eta_p), \overline{w}_{21}(\eta_p))S^{-1}(\eta_p)$ $(\overline{w}_{11}(\eta_p), \overline{w}_{21}(\eta_p))^T + O(n\delta^3) = o(n\delta^2)$ a.s. Hence $-2\log R(\theta_0, \eta_p+\delta) > -2\log R(\theta_0, \eta_p)$ a.s. Similarly $-2\log R(\theta_0, \eta_p - \delta) > -2\log R(\theta_0, \eta_p)$ a.s.

Since $-2\log R(\theta_0, \eta)$ is a differentiable function of η for $\eta \in [\eta_p - \delta, \eta_p + \delta]$, $-2\log R(\theta_0, \eta)$ attains its minimum in the region $(\eta_p - \delta, \eta_p + \delta)$, say at $\tilde{\eta}_E$, such that $\tilde{\eta}_E, \lambda(\tilde{\eta}_E), t(\tilde{\eta}_E)$ satisfy equations (2.8)$-$(2.10).

Lemma 4.6. *Under the conditions* (i)$-$(iii)*, for $\tilde{\eta}_E$ given in Lemma* 4.5*, we have* $\sqrt{n}\,(\tilde{\eta}_E - \eta_p) \to_L N\left(0, c_1^2 p(1-p) + c_2^2 q(1-q) + 2c_1c_2 p(1-q)\right)$, $\lambda(\tilde{\eta}_E) = -\{f(\eta_q)/f(\eta_p)\}t(\tilde{\eta}_E) + o_p(n^{-1/2})$, $\sqrt{n}t(\tilde{\eta}_E) \to_L N(0, d_1^2 p(1-p) + d_2^2 q(1-q) -2d_1d_2p(1-q))$, *where* $c_0 = f^2(\eta_p)q(1-q) - 2p(1-q)f(\eta_p)f(\eta_q) + f^2(\eta_q)p(1-p)$, $c_1 = (1-q)[qf(\eta_q) - pf(\eta_p)]/c_0$, $c_2 = p[(1-p)f(\eta_q) - (1-q)f(\eta_p)]/c_0$, $d_1 = f(\eta_p)f(\eta_q)/c_0$, *and* $d_2 = f^2(\eta_p)/c_0$.

Proof. Define $\lambda = \lambda(\eta)$, $\tilde{\lambda}_E = \lambda(\tilde{\eta}_E)$, $t = t(\eta)$, and $\tilde{t}_E = t(\tilde{\eta}_E)$. Further denote the left hand sides of (2.8)$-$(2.10) by $Q_{jn}(\eta, \lambda, t)$, $j = 1,\ 2,\ 3$, respectively. It follows from Lemma 4.5 that $Q_{in}\left(\tilde{\eta}_E, \tilde{\lambda}_E, \tilde{t}_E\right) = 0$ for $i = 1,\ 2,\ 3$. Therefore, by Lemma 4.4, Lemma 4.5 and a Taylor expansion, we get

$$
\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} Q_{1n}(\eta_p, 0, 0) \\ Q_{2n}(\eta_p, 0, 0) \\ Q_{2n}(\eta_p, 0, 0) \end{pmatrix} + \hat{S}_1 \begin{pmatrix} \tilde{\eta}_E - \eta_p \\ \tilde{\lambda}_E \\ \tilde{t}_E \end{pmatrix} + O_p\left(\delta^2\right),
$$

where, by some simple calculations,

$$\hat{S}_1 =: \begin{pmatrix} \frac{\partial Q_{1n}(\eta_p,0,0)}{\partial \eta} & \frac{\partial Q_{1n}(\eta_p,0,0)}{\partial \lambda} & \frac{\partial Q_{1n}(\eta_p,0,0)}{\partial t} \\ \frac{\partial Q_{2n}(\eta_p,0,0)}{\partial \eta} & \frac{\partial Q_{2n}(\eta_p,0,0)}{\partial \lambda} & \frac{\partial Q_{2n}(\eta_p,0,0)}{\partial t} \\ \frac{\partial Q_{3n}(\eta_p,0,0)}{\partial \eta} & \frac{\partial Q_{3n}(\eta_p,0,0)}{\partial \lambda} & \frac{\partial Q_{3n}(\eta_p,0,0)}{\partial t} \end{pmatrix}$$
$$\longrightarrow_{a.s.} \begin{pmatrix} f(\eta_p) & -p(1-p) & -p(1-q) \\ f(\eta_q) & -p(1-q) & -q(1-q) \\ 0 & f(\eta_p) & f(\eta_q) \end{pmatrix} =: S_1.$$

Condition (iii) implies $\delta^2 = (h^{2r} + 2h^r n^{-s} + n^{-2s}) = o(n^{-1/2})$. Hence

$$\begin{pmatrix} \tilde{\eta}_E - \eta_p \\ \tilde{\lambda}_E \\ \tilde{t}_E \end{pmatrix} = -S_1^{-1} \begin{pmatrix} Q_{1n}(\eta_p, 0, 0) \\ Q_{2n}(\eta_p, 0, 0) \\ 0 \end{pmatrix} + o_p(n^{-1/2}).$$

This leads to

$$\begin{aligned} \tilde{\eta}_E - \eta_p &= c_0^{-1}[(1-q)(qf(\eta_q) - pf(\eta_p))Q_{1n}(\eta_p, 0, 0) \\ &\quad + p((1-p)f(\eta_q) - (1-q)f(\eta_p))Q_{2n}(\eta_p, 0, 0)] + o_p(n^{-1/2}), \\ \tilde{\lambda}_E &= c_0^{-1}[-f^2(\eta_q)Q_{1n}(\eta_p, 0, 0) + f(\eta_p)f(\eta_q)Q_{2n}(\eta_p, 0, 0)] + o_p(n^{-1/2}), \\ \tilde{t}_E &= c_0^{-1}[f(\eta_p)f(\eta_q)Q_{1n}(\eta_p, 0, 0) - f^2(\eta_p)Q_{2n}(\eta_p, 0, 0)] + o_p(n^{-1/2}). \end{aligned}$$

Finally, the lemma follows directly from the above and the fact that

$$\sqrt{n} \begin{pmatrix} Q_{1n}(\eta_p, 0, 0) \\ Q_{2n}(\eta_p, 0, 0) \end{pmatrix} \to_L N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} p(1-p) & p(1-q) \\ p(1-q) & q(1-q) \end{pmatrix} \right).$$

Proof of Theorem 2.1. By (2.8), (2.9) and Lemma 4.4, $\overline{w}_{11}(\tilde{\eta}_E) = \lambda(\tilde{\eta}_E)\overline{w}_{12}(\tilde{\eta}_E) + t(\tilde{\eta}_E)\overline{w}_3(\tilde{\eta}_E) + o_p(n^{-1/2})$, and $\overline{w}_{21}(\tilde{\eta}_E) = \lambda(\tilde{\eta}_E)\overline{w}_3(\tilde{\eta}_E) + t(\tilde{\eta}_E)4\overline{w}_{22}(\tilde{\eta}_E) + o_p(n^{-1/2})$. From these, Lemma 4.6, and a Taylor expansion, we get

$$\begin{aligned} &-2\log R(\theta_0, \tilde{\eta}_E) \\ &= 2\Big(n\lambda(\tilde{\eta}_E)\overline{w}_{11}(\tilde{\eta}_E) + nt(\tilde{\eta}_E)\overline{w}_{21}(\tilde{\eta}_E)\Big) \\ &\quad - \sum_{i=1}^{n}[\lambda(\tilde{\eta}_E)w_1(X_i, \tilde{\eta}_E) + t(\tilde{\eta}_E)w_2(X_i, \tilde{\eta}_E)]^2 + o_p(1) \\ &= n\lambda^2(\tilde{\eta}_E)\overline{w}_{12}(\tilde{\eta}_E) + 2n\lambda(\tilde{\eta}_E)t(\tilde{\eta}_E)\overline{w}_3(\tilde{\eta}_E) + nt^2(\tilde{\eta}_E)\overline{w}_{22}(\tilde{\eta}_E) + o_p(1) \\ &= nt^2(\tilde{\eta}_E)\left[\frac{f^2(\eta_q)}{f^2(\eta_p)}\overline{w}_{12}(\tilde{\eta}_E) - 2\frac{f(\eta_q)}{f(\eta_p)}\overline{w}_3(\tilde{\eta}_E) + \overline{w}_{22}(\tilde{\eta}_E)\right] + o_p(1). \end{aligned}$$

Finally, the proof follows from this, Lemma (4.6) and the fact that $\overline{w}_{12}(\tilde{\eta}_E) \to p(1-p)$, $\overline{w}_{22}(\tilde{\eta}_E) \to q(1-q)$, $\overline{w}_3(\tilde{\eta}_E) \to p(1-q)$ a.s.

Acknowledgements

The authors would like to thank the Co-Editor, an associate editor and the referees for their valuable suggestions and remarks.

References

Chen, S. X. (1996). Empirical likelihood confidence intervals for nonparametric density estimation. *Biometrika* **83**, 329-341.

Chen, S. and Hall, P. (1993). Smoothed empirical likelihood confidence intervals for quantiles. *Ann. Statist.* **21**, 1166-1181.

Chen, S. X. and Qin, Y. S. (2000). Empirical likelihood confidence intervals for local linear smoothers. *Biometrika* **87**, 946-953.

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

Feldman, D. and Tucker, H. G. (1966). Estimation of non-unique quantiles. *Ann. Math. Statist.* **37**, 451-457.

Kitamura, Y. (1997) Empirical likelihood methods with weakly dependent processes. *Ann. Statist.* **25**, 2084-2102.

Kolaczyk, E. D. (1994). Empirical likelihood for generalized linear models. *Statist. Sinica* **4**, 199-218.

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

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

Owen, A. B. (1991). Empirical likelihood for linear models. *Ann. Statist.* **19**, 1725-1747.

Owen, A. B. (2001). *Empirical Likelihood.* Chapman and Hall, London.

Press, W. H., Flannery, B. F., Teukolsky, S. A. and Vetterling, W. T. (1989). *Numerical Recipes in C: The Art of Scientific Computing.* Cambridge Univ. Press.

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

Reiss, R. D. (1989). *Approximate Distributions of Order Statistics.* Springer-Verlag, New York.

Serfling, R. J. (1980). *Approximation Theorem of Mathematical Statistics.* Wiley, New York.

Department of Mathematics, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong.

E-mail: wangzhou@ust.hk

Department of Mathematics, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong.

E-mail: majing@ust.hk

(Received April 2001; accepted June 2002)