Adjusted empirical likelihood for probability density functions under strong mixing samples

Jie Tang Yongsong Qin[1]

Department of Statistics, Guangxi Normal University
Guilin, Guangxi 541006, China

Abstract

To obtain a profile empirical likelihood ratio statistic is essentially to address a constrained maximization problem. However, when sample size is small, the optimal function may have no numerical solution, as the convex hull of the sample points may not contain the zero vector 0 as its interior point. The adjusted empirical likelihood (AEL) method introduced by Chen et al. (2008, adjusted empirical likelihood and its properties,Journal of Computational and Graphical Statistics, 17, 426-443) is very useful to solve this problem. The innovation of this paper is that we extend the AEL method to probability density functions (p.d.f.) under dependent samples, as there are few literatures using this method under dependent samples. It is shown that the AEL ratio statistic for a p.d.f. is asymptotically χ^2-type distributed under a strong mixing sample, which is used to obtain an AEL-based confidence interval for the p.d.f. Our simulations show that the AEL is faster to compute than unadjusted empirical likelihood and the coverage probability based on the AEL method is superior to the other two usual methods, namely the unadjusted empirical likelihood and the normal approximation methods, particularly under small sample sizes.

Keywords: probability density function; α-mixing; confidence interval; adjusted empirical likelihood

AMS 2020 subject classification: Primary 62G05 secondary 62E20

[1]Corresponding author.

Email address: ysqin@mailbox.gxnu.edu.cn; 1243711126@qq.com

1. Introduction

The probability density function (p.d.f.) is a fundamental concept in statistics as well as in other fields. An estimator of a p.d.f. can provide valuable indications such as skewness and multimodality in a given set of data. Let $f(x)$ be the p.d.f. of a population. There exist many estimation methods for $f(x)$ in a nonparametric setting, such as the histogram estimator, the kernel estimator and the nearest neighbor estimator. On the one hand, to construct a confidence interval for $f(x)$, one can use the normal approximation (NA) method based on the asymptotic normality of an estimator. On the other hand, the unadjusted / usual empirical likelihood (EL) method, proposed by Owen (1988, 1990), can be used to construct a confidence interval for $f(x)$. We also note that sometimes confidence intervals for $f(x)$ are required, for instance, to test a hypothesis about $f(x)$. Many researchers have contributed to the NA and EL methods for probability density functions under **independent samples**. See Silverman (1986) and Scott (1992) for comprehensive reviews of the estimation methods of a p.d.f., and Chen (1996) for the EL confidence intervals for $f(x)$, among others. Chen (1996) also compared the EL method and NA method and shown that the EL method outperforms the NA method.

In many application fields, data do not satisfy the condition of independence. Among these dependent data, α-mixing or strong-mixing structure is one of the most popular types. Bradley (2005) summarized various types of strong mixing to measure dependence and analyzed various properties of dependence conditions. Since α-mixing is implied by other types of mixing structures such as the widely used ϕ, ρ and β-mixings, the conditions of α-mixing is the weakest among well-known mixing structures. Therefore, α-mixing sequences are widely used in many applications. For clarity, the following definition of α-mixing is given (e.g., Rosenblatt, 1956).

Let $\{\eta_i, i \geq 1\}$ be a random variable sequence and $\mathcal{F}_s^t$ denote the σ-algerbra generated by $\{\eta_i, s \leq i \leq t\}$ for $s \leq t$. Random variables $\{\eta_i, i \geq 1\}$ are said to be strongly mixing or α-mixing, if

$$\alpha(n) = \sup_{k\geq 1} \sup_{A\in\mathcal{F}_1^k, B\in\mathcal{F}_{k+n}^\infty} |P(AB) - P(A)P(B)| \to 0$$

as $n \to \infty$, where $\alpha(n)$ is called α-mixing coefficient.

Under α-mixing samples, Robinson (1983) studied the asymptotic normality of the kernel density estimators and Qin and Lei (2016) studied the asymptotic normality of the nearest neighbor density estimators. Kitamura (1997) frist proposed a blockwise EL method to construct confidence intervals (CIs) for parameters under strong mixing samples. Xiong and Lin (2012) demonstrated that the usual EL method can be used to construct CIs for $f(x)$ under associated samples. Lei and Qin (2015) used usual EL method to construct CIs for $f(x)$ under strong mixing samples. The main advantage using usual EL method over blockwise EL is that we do not need to choose the block lengths in constructing the EL ratio statistics and the EL CIs. The shape and orientation of the EL confidence interval are determined by data and the CI is obtained without covariance estimation. These features of the EL method are becoming increasingly attractive in empirical economic research.

However, to obtain the EL ratio statistic, a prerequisite is that the convex hull of the sample points of the related estimation equations must have the zero vector 0 as its interior point. However, the usual EL method does not always satisfy this prerequisite. To address this issue, Chen et al. (2008) proposed an adjusted empirical likelihood (AEL) method by adding an artificial point to the data set, and shown that the confidence regions constructed via the AEL are found to have coverage probabilities closer to the nominal levels without employing complex procedures such as Bartlett correction or bootstrap calibration. Liu and Chen (2010) and Chen and Liu (2012) further investigated the high-order precision of the AEL. It is noticed that there are few publications studying the AEL method under dependent samples.

In this article, we employ the AEL to construct the CI for a p.d.f. under strong mixing samples, and show that the AEL method retains all the optimality properties as well as the EL method and guarantees a sensible value of the likelihood at any parameter value. Our simulation indicates that the CIs based on AEL method have coverage probabilities closer to the nominal level than the EL method again without employing complex procedures such as Bartlett correction or bootstrap calibration. In particular, the AEL method performs much better than the EL method when the sample size is small and provides a promising solution to the small-sample, under-coverage problem discussed by Tsao (2004). This is a similar finding as in Chen et al. (2008).

The article is organized as follows. Section 2 gives the main results. Section 3 presents the results from a simulation study and provides a website to download the R codes related to the simulations. All technical details are provided in Section 4.

2. Main results

Let $X_1, ..., X_n$ be a stationary α-mixing sample from a population $X \in R$ and $f(\cdot)$ be the p.d.f. of X. The kernel density estimator of $f(x)$ at a given $x \in R$ is

$$f_n(x) = (nh)^{-1} \sum_{i=1}^{n} K_h(x - X_i), \tag{1}$$

where K is a kernel function, $h = h_n$ are bandwidths, and $K_h(u) = K(u/h)$ for any $u \in R$. Denote $\theta = f(x)$ and $\omega_i(\theta) = h^{-1}K_h(x - X_i) - \theta, 1 \le i \le n$.

Following Chen (1996), one can obtain the following EL ratio

$$L(\theta) = \sup_{\hat{p}_1,\ldots,\hat{p}_n} \left\{ \prod_{i=1}^{n} n\hat{p}_i \Big| \sum_{i=1}^{n} \hat{p}_i = 1, \hat{p}_i \ge 0, \sum_{i=1}^{n} \hat{p}_i \omega_i(\theta) = 0 \right\}.$$

It is easy to obtain the (-2log) EL ratio statistic

$$\ell(\theta) = 2 \sum_{i=1}^{n} \log\{1 + \lambda(\theta)\omega_i(\theta)\}, \tag{2}$$

and

$$\hat{p}_i = \frac{1}{n} \cdot \frac{1}{1 + \lambda(\theta)\omega_i(\theta)}, 1 \le i \le n,$$

where $\lambda(\theta)$ is the solution of the following equation:

$$\frac{1}{n} \sum_{i=1}^{n} \frac{\omega_i(\theta)}{1 + \lambda(\theta)\omega_i(\theta)} = 0, \tag{3}$$

provided that 0 is an interior point of the convex hull of $\{\omega_i(\theta), i = 1, 2, \ldots, n\}$. Under some regularity conditions in Lei and Qin (2015), the convex hull $\{\omega_i(\theta_0), i = 1, 2, \ldots, n\}$ contains 0 as its interior point with probability 1 as $n \to \infty$, where θ_0 denotes the true value of θ. When θ is not close to θ_0, or when n is small, it would most likely cause that the solution $\lambda = \lambda(\theta)$ in (3) does not exist, which can be a serious limitation in some applications. Therefore, in this article, we propose to use the AEL method proposed by Chen et al. (2008) as follows.

Let $\{\omega_i(\theta), 1 \le i \le n\}$ be defined as above and define the AEL ratio

$$L^*(\theta) = \sup_{p_1,\ldots,p_{n+1}} \left\{ \prod_{i=1}^{n+1} np_i \Big| \sum_{i=1}^{n+1} p_i = 1, p_i \ge 0, \sum_{i=1}^{n+1} p_i \omega_i(\theta) = 0 \right\}.$$

where $\omega_{n+1}(\theta) = -a_n \bar{\omega}_n(\theta)$ for some positive constant a_n, with $\bar{\omega}_n(\theta) = n^{-1} \sum_{i=1}^{n} \omega_i(\theta)$. Obviously the (-2log) AEL ratio statistic

$$\ell^*(\theta) = 2 \sum_{i=1}^{n+1} \log\{1 + \lambda(\theta)\omega_i(\theta)\}, \tag{4}$$

and

$$p_i = \frac{1}{n+1} \cdot \frac{1}{1 + \lambda(\theta)\omega_i(\theta)}, 1 \le i \le n+1,$$

where $\lambda(\theta)$ is the solution of the following equation:

$$\frac{1}{n+1} \sum_{i=1}^{n+1} \frac{\omega_i(\theta)}{1 + \lambda(\theta)\omega_i(\theta)} = 0. \tag{5}$$

Since the convex hull of $\{\omega_i(\theta), i = 1, 2, \ldots, n, n+1\}$ for any given θ contains 0, $\ell^*(\theta)$ is well defined for all θ.

Chen et al. (2008) provided a guidance to the choice of the a_n in the AEL method. In general, a_n should be chosen according to the specific nature of the application. In theory, the first order asymptotic property of $\ell(\theta)$ is unchanged for $\ell^*(\theta)$ as long as $a_n = o(n^{2/3})$. In most applications, Chen et al. (2008) recommended to choose $a_n = \max\{1, \log n\}$. Investigation of the optimal choice of a_n in various practically important situations is still under way.

To obtain the asymptotical distribution of $\ell^*(\theta)$, we need following assumptions, which are presented and verified in Lei and Qin (2015).

A1.(i) The r.v.s. $X_1, X_2, \ldots, X_n$ from a stationary sequence, and let f be the one-dimensional marginal p.d.f. (with respect to Lebesgue measure).

(ii) $\{X_i, 1 \le i \le n\}$ is a α-mixing sequence with mixing coefficient $\alpha(\cdot)$.

(iii) The joint p.d.f. $f_{1,j}$ of the r.v.s. X_1, X_{j+1} exists and $|f_{1,j}(u,v) - f(u)f(v)| \le C$ for all $u, v \in R$ and $j > 1$.

A2. The function K is bounded, has compact support and satisfies $\int K(u)\mathrm{d}u = 1$, $\int uK(u)\mathrm{d}u = 0$ and $0 < \int u^2 K(u)\mathrm{d}u < \infty$.

A3. Let $p = p(n)$ and $q = q(n)$ be positive integers satisfying $p + q \leq n$, and $k = [n/(p+q)]$, where $[t]$ denotes the integer part of t. p, q and h satisfy

(i) $p \to \infty$, $q/p \to 0$, $p/n \to 0$, $ph \to 0$, $p^2/(nh) \to 0$, $(1/h)\sum_{s=p}^{\infty} \alpha(s) \to 0$, $k\alpha(q) \to 0$.

(ii) $h \to 0, nh \to \infty$ and $nh^5 \to 0$ as $n \to \infty$.

Remark 1 *From (A3)(i), it can be shown that $pk/n \to 1$ and $qk/n \to 0$.*

We now state the main results.

Theorem 1 *Suppose that conditions A1-A3 hold and that $a_n = o(n^{2/3})$. Let θ_0 be the true value of θ. As $n \to \infty$,*

$$\ell^*(\theta_0) \xrightarrow{d} \chi_1^2,$$

where χ_1^2 is a chi-squared distributed random variable with one degree of freedom.

Let z_α satisfy $P(\chi_1^2 \leq z_\alpha) = 1 - \alpha$ for $0 < \alpha < 1$. It follows from Theorem 1 that an AEL-based CI for $\theta = f(x)$ with asymptotically correct coverage probability (CP) $1 - \alpha$ can be constructed as

$$\{\theta : \ell^*(\theta) \leq z_\alpha\}. \tag{6}$$

Theorem 2 *Assume that the conditions in Theorem 1 hold true and that $nh^{9/5} \to \infty$, $\theta \neq \theta_0$. Then, $n^{-1/3}\ell^*(\theta) \to \infty$ and $n^{-1/3}\ell(\theta) \to \infty$ in probability as $n \to \infty$.*

Based on Theorems 1, we can obtain a hypothesis testing approach based on the AEL for the hypothesis problem: $H_0 : \theta = \theta_0$ and $H_1 : \theta \neq \theta_0$, which implies asymptotically rejection region with the significance level α can be constructed as

$$W \hat{=} \{(X_1, \cdots, X_n) : \ell^*(\theta_0) > z_\alpha\}.$$

By Theorem 2, when $\theta \neq \theta_0$, we have $P(W|H_1) = P_{H_1}(\ell^*(\theta) > z_\alpha) = P_{H_1}(n^{-1/3}\ell^*(\theta) > n^{-1/3}z_\alpha) \to 1$. Thus, the power of this test tends to 1 as $n \to \infty$.

3. Simulations results

We denote the AEL-based CI for $\theta = f(x)$ given by equation (6) as AELCI. By Theorem 1 in Lei and Qin (2015), the EL-based CI for $\theta = f(x)$ is given by equation (7) and denoted as ELCI.

$$\{\theta : \ell(\theta) \leq z_\alpha\}. \tag{7}$$

The NA-based CI for $\theta = f(x)$ derived from Lemma 4 in Lei and Qin (2015) is shown in equation (8) and denoted as NACI.

$$\left[f_n(x) - \frac{\hat{\sigma}(x)u_\alpha}{(nh)^{1/2}}, f_n(x) + \frac{\hat{\sigma}(x)u_\alpha}{(nh)^{1/2}}\right], \tag{8}$$

where $\hat{\sigma}^2(x) = (n^{-1}h)\sum_{i=1}^n \omega_i^2(f_n(x))$ and $P(|N(0,1)| \leq u_\alpha) = 1 - \alpha$.

We conducted a small simulation study to compare the finite sample performances of the confidence intervals of AELCI, ELCI and NACI. In the simulations, we considered two AR models:

$$X_t = \left(\frac{1}{2}\right) X_{t-1} + \epsilon_t, \tag{9}$$

and

$$X_t = \left(\frac{3}{4}\right) X_{t-1} - \left(\frac{1}{8}\right) X_{t-2} + \epsilon_t, \tag{10}$$

where $\{\epsilon_t\}$ is an i.i.d. $N(0,1)$ random variable sequence. We can see that both models produce stationary and α-mixing sequences with $X_1 \sim N(0, \frac{4}{3})$ for model (9) and $X_1 \sim N(0, \frac{64}{35})$ for model (10).

We generated 2,000 random samples of data $\{X_i, i = 1, \ldots, n\}$ for $n = 50$, 100, 150, 200 and 250 from above models and set $a_n = \log n$ suggested by Chen et al. (2008) for the AEL method. In the simulations, we took $h = 4n^{-1/5}/\log n$ and K was chosen as

$$K(u) = \left(\frac{15}{16}\right)(1 - u^2)^2 I(|u| \leq 1).$$

The choice of the bandwidth is important and there is no satisfactory approach available now. Further study in choosing the bandwidth is definitely needed under mixing samples. Lei and Qin (2015) suggested to choose h as $Cn^{-1/5}/\log n$ with $3 \leq C \leq 5$ and we thus chosen $h = 4n^{-1/5}/\log n$ in the simulations.

Let f be the p.d.f. of X_1. For nominal confidence level $\alpha = 0.95$, using the simulated samples, we evaluated the coverage probabilities (CPs) and the average lengths (ALs) of AELCI, ELCI and NACI for $f(x)$ at $x = 0$ and 1. Table 1 reports the simulation results for model (9), while Table 2 reports the simulation results for model (10).

A Four Core Processor with 1.4 G and Intel Core i5, 8 G main memory and MacBook Pro 2020 was used to do the simulations. Seven and a half seconds, eight seconds, and one over three seconds were, respectively, used to obtain the AELCI, ELCI and NACI results in Table 1, which indicates that the AEL method is faster to compute than the EL method.

It can be seen from Tables 1-2 that the larger the sample size is, the closer the CP is to the nominal level, and the lengths of the estimated CIs decrease with increasing sample size for all CIs. It is clear that the CPs of the AELCI are closer to the target values than the other two CIs, but the ALs of the AELCI are larger than the others.

Table 1: CPs and ALs of CIs for $f(x)$ under model (9) at $x = 0$ and 1.

		CP			AL		
x	n	AELCI	ELCI	NACI	AELCI	ELCI	NACI
$x = 0$	50	0.9580	0.9365	0.9145	0.3475	0.3139	0.3159
	100	0.9470	0.9325	0.9200	0.3063	0.2906	0.2909
	150	0.9495	0.9415	0.9310	0.2799	0.2698	0.2697
	200	0.9390	0.9305	0.9210	0.2370	0.2304	0.2298
	250	0.9515	0.9440	0.9370	0.2201	0.2149	0.2145
$x = 1$	50	0.9420	0.9135	0.8785	0.3257	0.2940	0.2943
	100	0.9375	0.9230	0.9020	0.3201	0.3037	0.3039
	150	0.9355	0.9250	0.9075	0.2272	0.2191	0.2176
	200	0.9470	0.9385	0.9205	0.2302	0.2237	0.2232
	250	0.9465	0.9395	0.9280	0.1825	0.1783	0.1771

Further, we presented two figures to compare all three methods in terms of CPs and ALs as the bandwidth h varies. Figures 1 and 2 illustrated the performances under small and large sample size respectively. We can see from Figure 1 that there

Table 2: CPs and ALs of CIs for $f(x)$ under model (10) at $x = 0$ and 1.

x	n	CP			AL		
		AELCI	ELCI	NACI	AELCI	ELCI	NACI
$x = 0$	50	0.9440	0.9170	0.8850	0.3774	0.3410	0.3438
	100	0.9510	0.9395	0.9190	0.3154	0.2993	0.2995
	150	0.9335	0.9225	0.9125	0.2636	0.2541	0.2539
	200	0.9390	0.9325	0.9215	0.1865	0.1813	0.1803
	250	0.9475	0.9430	0.9315	0.2451	0.2440	0.2438
$x = 1$	50	0.9295	0.9020	0.8660	0.3597	0.3250	0.3253
	100	0.9295	0.9170	0.9015	0.2941	0.2790	0.2786
	150	0.9365	0.9260	0.9060	0.2561	0.2468	0.2461
	200	0.9315	0.9280	0.9175	0.1742	0.1693	0.1679
	250	0.9355	0.9300	0.9025	0.2062	0.2015	0.2008

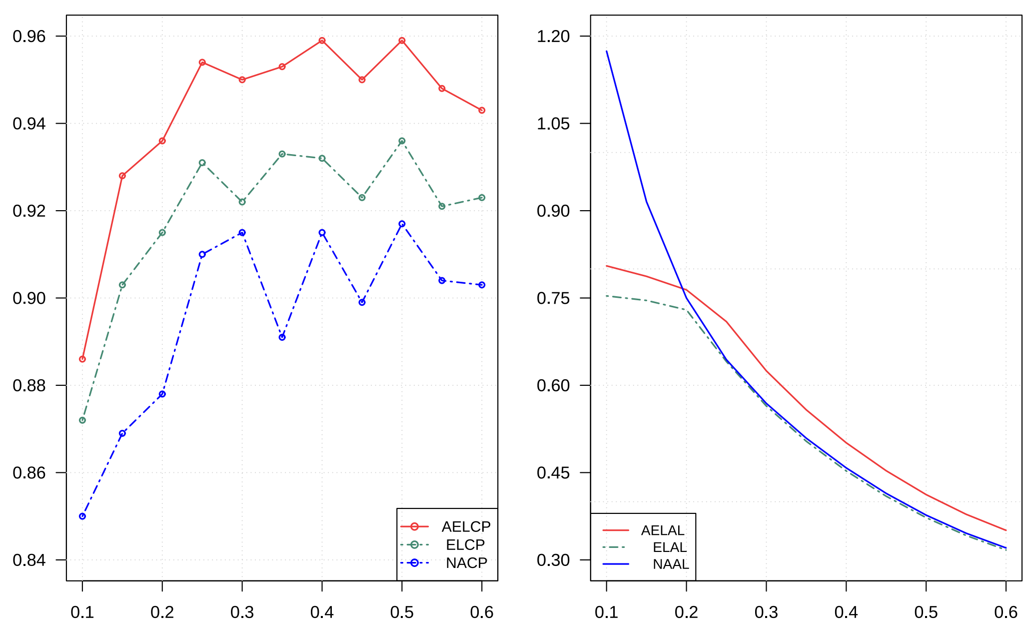

Figure 1: CPs and ALs of CIs for $f(0)$ under model (9) via the three methods (AEL, EL & NA) with $n = 50$ as h varies: the left is for CP and the right is for AL.

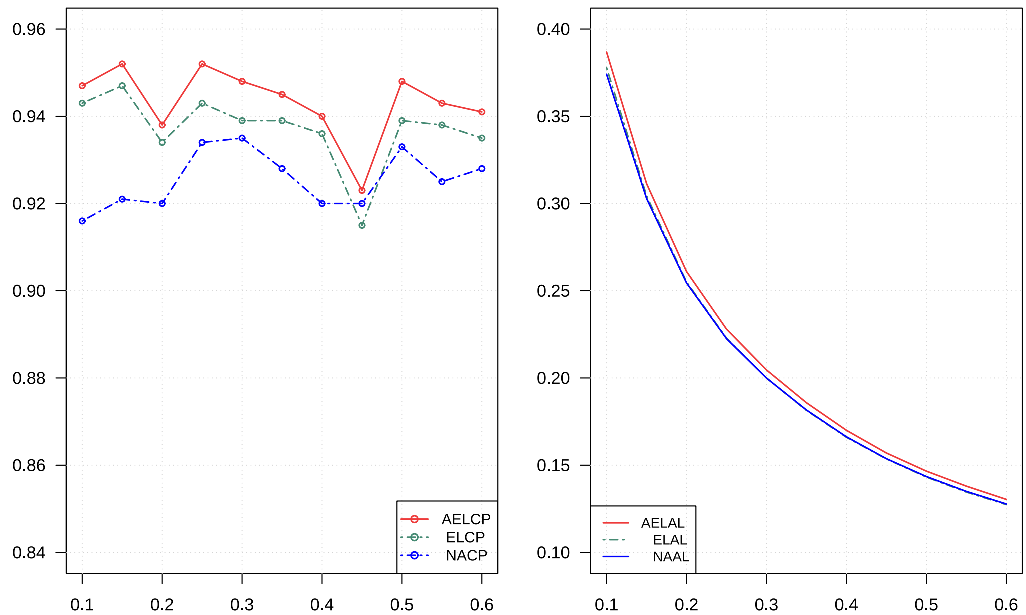

Figure 2: CPs and ALs of CIs for $f(0)$ under model (9) via the three methods (AEL, EL & NA) with $n = 250$ as h varies: the left is for CP and the right is for AL.

are significant differences among the three methods and the AEL is the best one in terms of the CPs for almost all h in the range. However, the AEL is a littler poorer than other two methods in terms of ALs. On the other hand, Figure 2 shows that there are some differences among the three methods and the AEL is still the best one in terms of the CPs for almost all h in the range. Moreover, for the large sample size, all three methods almost agree in terms of ALs. We may conclude from the simulations that the AEL is effective in improving the CPs of the CIs. Therefore, the AEL method may be suitable to solve the under-coverage problem indicated by Tsao (2004).

For the computation issue in the simulations, we used modified Newton-Raphson algorithm to solve Equation (5) to obtain $\lambda(\theta)$, combining with the computation method shown in Chen et al. (2008). One may also refer to Chen et al. (2002) for the computation of the EL method. One can download R codes related to this article at https://github.com/Tang-Jay/Mixing-Sample.

4. Proofs

To prove the main results, we need following lemma.

LEMMA 1 *Let $\theta = \theta_0$. Under the conditions of Theorem 1, as $n \to \infty$,*

$$\max_{1 \le i \le n} |\omega_i(\theta)| \le Ch^{-1}, \ a.s.,$$

$$\sqrt{n^{-1}h} \sum_{i=1}^{n} \omega_i(\theta) \xrightarrow{d} N(0, \sigma^2(x)),$$

$$(n^{-1}h) \sum_{i=1}^{n} \omega_i^2(\theta) = \sigma^2(x) + o_p(1),$$

$$(n^{-1}h)^{3/2} \sum_{i=1}^{n} |\omega_i(\theta)|^3 = o_p(1),$$

where $\sigma^2(x) = f(x) \int K^2(u)\mathrm{d}u$.

Proof See Lemma 4 in Lei and Qin (2015).

***Proof of Theorem* 1 :**

Let $\theta = \theta_0$ and $\lambda = \lambda(\theta)$, where λ is the solution to (5). We first show that $\lambda = O_p((h/n)^{1/2})$. By Lemma 1, we have

$$\bar{\omega}_n(\theta) = O_p\left((nh)^{-1/2}\right) \text{ and } \omega^* = O_p(h^{-1}). \tag{11}$$

where $\bar{\omega}_n(\theta) = n^{-1}\sum_{i=1}^{n} \omega_i(\theta)$ and $\omega^* = \max_{1 \le i \le n} |\omega_i(\theta)|$. Let $\rho = \lambda/|\lambda|$. By (5), we get

$$\begin{aligned}
0 &= \frac{\rho}{n} \sum_{i=1}^{n+1} \frac{\omega_i(\theta)}{1 + \lambda\omega_i(\theta)} \\
&= \frac{\rho}{n} \sum_{i=1}^{n+1} \omega_i(\theta) - \frac{\rho}{n} \sum_{i=1}^{n+1} \frac{\lambda\omega_i^2(\theta)}{1 + \lambda\omega_i(\theta)} \\
&= \frac{\rho}{n} \sum_{i=1}^{n+1} \omega_i(\theta) - \frac{|\lambda|}{n} \sum_{i=1}^{n+1} \frac{\rho^2\omega_i^2(\theta)}{1 + \lambda\omega_i(\theta)} \\
&= \rho\bar{\omega}_n(\theta)(1 - a_n/n) - \frac{|\lambda|}{n} \sum_{i=1}^{n} \frac{\rho^2\omega_i^2(\theta)}{1 + \lambda\omega_i(\theta)} - \frac{|\lambda|}{n} \frac{\rho^2\omega_{n+1}^2(\theta)}{1 + \lambda\omega_{n+1}(\theta)} \\
&\le \rho\bar{\omega}_n(\theta)(1 - a_n/n) - \frac{|\lambda|}{n} \sum_{i=1}^{n} \frac{\rho^2\omega_i^2(\theta)}{1 + |\lambda|\omega^*} \\
&= \rho\bar{\omega}_n(\theta) - \frac{|\lambda|}{n(1 + |\lambda|\omega^*)} \sum_{i=1}^{n} \omega_i^2(\theta) + O_p(a_n(nh^{1/3})^{-3/2}).
\end{aligned} \tag{12}$$

The inequality above is valid because the $(n+1)$th term of the second summation is nonnegative. By Lemma 1, we have

$$\sum_{i=1}^{n} \omega_i^2(\theta) = O_p\left(n/h\right). \tag{13}$$

Therefore, as long as $a_n = o(n)$, equations (11) and (12) imply that

$$\frac{|\lambda|}{n(1+|\lambda|\omega^*)} \sum_{i=1}^{n} \omega_i^2(\theta) = O_p\left((nh)^{-1/2}\right),$$

and hence

$$\lambda = O_p\left(\left(h/n\right)^{1/2}\right). \tag{14}$$

Next, we will express λ as the function of the linear and quadratic sums of $\{\omega_i(\theta)\}$. Let $\gamma_i = \lambda\omega_i(\theta)$, we have $\max_{1\le i\le n} |\gamma_i| \le |\lambda| \cdot \omega^*$, $|\gamma_{n+1}| = |\lambda\omega_{n+1}(\theta)| = |a_n\lambda\bar{\omega}_n(\theta)| \le a_n|\lambda| \cdot |\bar{\omega}_n(\theta)|$. By equations (11) and (14), we know that

$$\max_{1\le i\le n+1} |\gamma_i| = o_p(1) \tag{15}$$

Let $S = (n^{-1}h)\sum_{i=1}^{n} \omega_i^2(\theta)$. We find that

$$\begin{aligned}
0 &= \frac{1}{n}\sum_{i=1}^{n+1} \frac{\omega_i(\theta)}{1+\lambda\omega_i(\theta)} \\
&= \frac{1}{n}\sum_{i=1}^{n+1} \omega_i(\theta) - \frac{1}{n}\sum_{i=1}^{n+1} \lambda\omega_i^2(\theta) + \frac{1}{n}\sum_{i=1}^{n+1} \frac{\lambda^2\omega_i^3(\theta)}{1+\lambda\omega_i(\theta)} \\
&= \bar{\omega}_n(\theta) - \lambda h^{-1}S + \frac{1}{n}\omega_{n+1}(\theta) - \frac{1}{n}\omega_{n+1}(\theta)\gamma_{n+1} \\
&\quad + \frac{1}{n}\sum_{i=1}^{n} \frac{\lambda^2\omega_i^3(\theta)}{1+\gamma_i} + \frac{1}{n}\cdot\frac{\lambda^2\omega_{n+1}^3(\theta)}{1+\gamma_{n+1}}
\end{aligned} \tag{16}$$

From Lemma 1, equations (11), (14) and (15), the last four term have the norms bounded by

$$n^{-1}|\omega_{n+1}(\theta)| = n^{-1}|a_n\bar{\omega}_n(\theta)| = o_p\left((nh)^{-1/2}\right),$$

$$n^{-1}|\omega_{n+1}(\theta)\gamma_{n+1}| = o_p\left((nh)^{-1/2}\right),$$

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\lambda^2|\omega_i(\theta)|^3}{|1+\gamma_i|}=o_p\left((nh)^{-1/2}\right),$$

and

$$\frac{1}{n}\cdot\frac{\lambda^2|\omega_{n+1}(\theta)|^3}{|1+\gamma_{n+1}|}=o_p\left((nh)^{-1/2}\right).$$

Therefore, (16) implies that

$$\lambda h^{-1}S=\bar{\omega}_n(\theta)+o_p\left((nh)^{-1/2}\right),$$

which further implies that

$$\lambda=hS^{-1}\bar{\omega}_n(\theta)+\tau, \tag{17}$$

where

$$\tau=o_p\left(\left(h/n\right)^{1/2}\right). \tag{18}$$

Finally, we expand $\ell^*(\theta)$ as follows. By equation (15) we may expand $\log(1+\gamma_i)=\gamma_i-\gamma_i^2/2+u_i$ where, for some finite $B>0$,

$$P(|u_i|\le B|\gamma_i|^3, 1\le i\le n+1)\to 1, \text{as } n\to\infty.$$

Therefore, from equations (4), (17) and Taylor expansion, we have

$$\begin{aligned}
\ell_n^*(\theta) &= 2\sum_{j=1}^{n+1}\log(1+\gamma_j)=2\sum_{j=1}^{n+1}\gamma_j-\sum_{j=1}^{n+1}\gamma_j^2+2\sum_{j=1}^{n+1}u_j\\
&= 2\sum_{j=1}^{n}\gamma_j-\sum_{j=1}^{n}\gamma_j^2+2\gamma_{n+1}-\gamma_{n+1}^2+2\sum_{j=1}^{n+1}u_j\\
&= 2\lambda n\bar{\omega}_n(\theta)-\lambda^2nh^{-1}S+2\gamma_{n+1}-\gamma_{n+1}^2+2\sum_{j=1}^{n+1}u_j\\
&= nhS^{-1}\bar{\omega}_n^2-nh^{-1}\tau^2S+2\gamma_{n+1}-\gamma_{n+1}^2+2\sum_{j=1}^{n+1}u_j.
\end{aligned}$$

From equations (15), (18) and Lemma 1, last four terms above are all $o_p(1)$ shown in the following:

$$nh^{-1}\tau^2S=nh^{-1}o_p(n^{-1}h)O_p(1)=o_p(1),$$

$$2\gamma_{n+1} - \gamma_{n+1}^2 = o_p(1),$$

and

$$\left|\sum_{j=1}^{n+1} u_j\right| \le \sum_{j=1}^{n+1} |u_j| \le \sum_{j=1}^{n+1} B|\gamma_j|^3 = \sum_{j=1}^{n} B|\gamma_j|^3 + B|\gamma_{n+1}|^3 = o_p(1).$$

By Lemma 1 again, we have

$$nhS^{-1}\bar{\omega}_n^2 = \left(\frac{\sqrt{n^{-1}h}\sum_{i=1}^n \omega_i(\theta)}{\sigma(x)}\right)^2 + o_p(1).$$

Therefore, by the Cramer-Wold's theorem, we have $\ell_n^*(\theta) \xrightarrow{d} \chi_1^2$, which completes the proof of Theorem 1.

***Proof of Theorem* 2 :**

Based on the notations in Section 2, it is obvious that

$$\omega^* \hat{=} \max_{1\le i\le n} |\omega_i(\theta)| = O_p(1/h).$$

By Lemma 1 and note that

$$\bar{\omega}_n(\theta) = n^{-1}\sum_{i=1}^n \omega_i(\theta_0) - (\theta - \theta_0),$$

we have

$$\bar{\omega}_n^2(\theta) = \delta^2 + o_p(1),$$

where $\delta = \theta - \theta_0$. Let $\tilde{\lambda} = n^{-2/3}\bar{\omega}_n(\theta)M, \tilde{\gamma}_i = \tilde{\lambda}\omega_i(\theta), 1 \le i \le n+1$, for some positive constant M. Hence, using the condition $nh^{9/5} \to \infty$ as $n \to \infty$, we have

$$\max_{1\le i\le n} |\tilde{\gamma}_i| \le |\tilde{\lambda}| \cdot \omega^* = O_p(n^{-2/3}h^{-1}) = o_p(1),$$

$$|\tilde{\gamma}_{n+1}| = |\tilde{\lambda}\omega_{n+1}(\theta)| = o_p(1),$$

i.e.,

$$\max_{1\le i\le (n+1)} |\tilde{\gamma}_i| = o_p(1). \tag{19}$$

By (19) and Taylor expansion $\log(1+\tilde{\gamma}_i) = \tilde{\gamma}_i - \tilde{\gamma}_i^2/2 + \tilde{u}_i$, where for some $B > 0$, as $n \to \infty$,

$$P(|\tilde{u}_i| \le B|\tilde{\gamma}_i|^3, 1 \le i \le n+1) \to 1.$$

Thus, by (19), we expand

$$\begin{aligned}\sum_{j=1}^{n+1}\log(1+\tilde{\gamma}_j) &= \sum_{j=1}^{n+1}\tilde{\gamma}_j - \frac{1}{2}\sum_{j=1}^{n+1}\tilde{\gamma}_j^2 + \sum_{j=1}^{n+1}\tilde{u}_j \\ &= \sum_{j=1}^{n}\tilde{\gamma}_j - \frac{1}{2}\sum_{j=1}^{n}\tilde{\gamma}_j^2 + \sum_{j=1}^{n+1}\tilde{u}_j + \tilde{\gamma}_{n+1} - \frac{1}{2}\tilde{\gamma}_{n+1}^2 \\ &= \sum_{j=1}^{n}\tilde{\gamma}_j - \frac{1}{2}\sum_{j=1}^{n}\tilde{\gamma}_j^2 + \sum_{j=1}^{n+1}\tilde{u}_j + o_p(1). \end{aligned} \tag{20}$$

Using Lemma 1 again, we get that $\sum_{i=1}^{n}\omega_i^2(\theta) = \sum_{i=1}^{n}\{\omega_i(\theta_0)-\delta\}^2 = O_p(nh^{-1})$ and $\sum_{i=1}^{n}|\omega_i(\theta)|^3 = o_p((nh^{-1})^{3/2})$ in the same way, which respectively means that

$$\sum_{j=1}^{n}\tilde{\gamma}_j^2 = \tilde{\lambda}^2\sum_{j=1}^{n}\omega_j(\theta)^2 = O_p(n^{-2/3})^2 O_p(nh^{-1}) = O_p((nh^3)^{-1/3})$$

and

$$|\sum_{j=1}^{n+1}\tilde{u}_j| \leq B\sum_{j=1}^{n}|\tilde{\gamma}_j|^3 + B|\tilde{\gamma}_{n+1}|^3 = o_p((nh^3)^{-1/2}) + o_p(1).$$

Thus, by $nh^{9/5}\to\infty$, (20) can be written as

$$\begin{aligned}\sum_{j=1}^{n+1}\log(1+\tilde{\gamma}_j) &= \sum_{j=1}^{n}\tilde{\gamma}_j + O_p((nh^3)^{-1/3}) + o_p((nh^3)^{-1/2}) + o_p(1) \\ &= n^{1/3}M\bar{\omega}_n^2(\theta) + O_p((nh^3)^{-1/3}) + o_p((nh^3)^{-1/2}) + o_p(1) \\ &= n^{1/3}M\delta^2 + o_p(n^{1/3}).\end{aligned}$$

Using the duality of the maximization problem, we find that

$$\begin{aligned} W^*(\theta) &\hat{=} \log L^*(\theta) = \sup_{p_i, 1\leq i\leq n+1}\left\{\sum_{i=1}^{n+1}\log(n+1)p_i\right\} = -\sup_{\lambda}\left\{\sum_{i=1}^{n+1}\log(1+\lambda\omega_i(\theta))\right\} \\ &\leq -\sum_{j=1}^{n+1}\log(1+\tilde{\lambda}\omega_j(\theta)) = -n^{1/3}M\delta^2 + o_p(n^{1/3}).\end{aligned}$$

Since M can be arbitrarily large, we have $-2n^{-1/3}W^*(\theta)\to\infty$ in probability for any $\theta\neq\theta_0$, which means $n^{-1/3}\ell_n^*(\theta)\to\infty$ in probability. The proof of $n^{-1/3}\ell_n(\theta)\to\infty$ is similar. Therefore, the proof of Theorem 2 is complete.

Acknowledgements This work was partially supported by the National Natural Science Foundation of China (12061017) and the Research Fund of Guangxi Key Lab of Multi-source Information Mining & Security (22-A-01-01).

References

Bradley, R. C. (2005). Basic properties of strong mixing conditions: a survey and some open questions. *Probability Survey, 2*, 107-144.

Chen, J. H. and Liu, Y. K. (2012). Adjusted empirical likelihood with high-order one-sided coverage precision. *Statistics and its Interface, 5*, 281-292.

Chen, J., Sitter, R. R., and Wu, C. (2002). Using empirical likelihood methods to obtain range restricted weights in regression estimators for surveys. *Biometrika, 89*, 230-237.

Chen, J., Variyath, A. M., and Abraham, B. (2008). Adjusted empirical likelihood and its properties. *Journal of Computational and Graphical Statistics, 17*, 426-443.

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

Kitamura, Y. (1997). Empirical likelihood methods with weakly dependent processes. *The Annals of Statistics, 25*, 2084-2102.

Lei, Q. Z. and Qin, Y. S. (2015). Confidence intervals for probability density functions under strong mixing samples. *Journal of Nonparametric Statistics, 27*, 181-193.

Liu, Y. K. and Chen, J. H. (2010). Adjusted empirical likelihood with high-order precision. *The Annals of Statistics, 38*, 1341-1362.

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. *The Annals of Statistics, 18*, 90-120.

Qin, Y. S. and Lei, Q. Z. (2016). Asymptotic normality of the nearest neighbor density estimators under strong mixing samples. *Chinese Journal of Contemporary Mathematics, 37*, 463-472.

Robinson, P. M. (1983). Nonparametric estimators for time series. *Journal of Time Series Analysis, 4,* 185-207.

Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition, *Pro-*

ceedings of the National Academy of Sciences of the United States of America, 42, 43-47.

Scott, D. W. (1992). *Multivariate Density Estimation.* New York: Wiley.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis.* London: Chapman and Hall.

Tsao, M. (2004). Bounds on coverage probabilities of the empirical likelihood ratio confidence regions. *The Annals of Statistics, 32*, 1215-1221.

Xiong, X., and Lin, Z. (2012). Empirical likelihood inference for probability density functions under association. *Journal of Statistical Planning and Inference, 42*, 986-992.