Journal of Statistical Research *ISSN 0256 - 422 X*
2011, Vol. 45, No. 2, pp. 95-110
Bangladesh

EMPIRICAL LIKELIHOOD CONFIDENCE INTERVALS FOR DENSITY FUNCTION IN ERRORS-IN-VARIABLES MODEL

WEIXING SONG

Department of Statistics, Kansas State University, Manhattan, KS, 66503, USA

Email: weixing@ksu.edu

SUMMARY

The paper proposes empirical likelihood confidence intervals for the density function in the errors-in-variables model. We show that the empirical likelihood produces confidence intervals having theoretically accurate coverage rate for both ordinary and super smooth measurement errors. Some simulation studies are conducted to compare the finite sample performances of the empirical likelihood confidence intervals and the z-type confidence intervals based on Fan (1991)'s asymptotic normality theories.

Keywords and phrases: Confidence interval; Empirical likelihood; Deconvolution kernel density estimation.

AMS Classification: 62G05, 62G07

1 Introduction

Suppose X is a characteristic of interest from a random system and we want to estimate its probability density function $f_X(x)$. Many procedures exist in the literature for constructing either the point or the confidence interval estimations for $f_X(x)$, as long as a sample from X is available. From the simplest and intuitive graphical tools, such as the histogram, stem and leat plot, to the more technical and powerful kernel smoothing methods, nonparametric density estimation and its extension to various nonparametric curve estimations have attracted and still are drawing extensive attentions from both theoretical and applied statisticians. See Silverman (1986) and Scott (1992) for comprehensive introductions to the kernel density estimation. The existing literature mainly focuses on the point estimators and their asymptotic properties. The confidence interval, although not investigated thoroughly, can be constructed based on those well developed theories. Hall (1991) studied the coverage accuracy of the bootstrap confidence intervals by developing an Edgeworth expansion for the kernel density estimator; Hall and Owen (1993) proposed an empirical likelihood based simultaneous confidence interval for the density function, an analogue of Wilks' Theorem based on extreme value type asymptotic distribution was obtained. Inspired by a simulation study which shows that confidence intervals produced by the kernel based percentile-t bootstrap do not have the coverage claimed by the theory, Chen (1996) suggested using empirical likelihood in conjunction with the kernel method to construct confidence intervals for the density function. It is found that the coverage discrepancy is due to a conflict between the prescribed undersmoothing and the explicit variance estimate needed by the percentile-t method. Chen (1996) showed that the empirical likelihood avoids this conflict by studentising internally, and the resulting confidence intervals have theoretical coverage accuracy of the same order of magnitude! as the bootstrap. In the current research, we will try to apply Chen (1996)'s method to errors-in-variable model in which the random variable of interest cannot be observed directly, and the estimator of its density function has to be constructed from the samples of the contaminated observations.

To be specific, the errors-in-variable model has the form of

$$Y = X + u, \tag{1.1}$$

where X is the latent variable of interest but unobservable, instead, its surrogate Y, a contaminated version of X by an additive measurement error u, can be observed. X and u are independent. The distribution of u is usually assumed to be known. Although researchers tried to remove this assumption in some other scenarios, it seems that in the case of estimating the density function of X, assuming that the density function of u is known is necessary for the sake of identifiability.

It is well known that estimating the density function $f_X(x)$ of X is notorious difficult. The commonly used method of estimating $f_X(x)$ is the so called deconvolution method. Suppose $Y_1, Y_2, \ldots, Y_n$ is a sample of size n from model (1.1). Let K be a kernel function, $\phi_K(t)$ be the characteristic function of K, $\phi_u(t)$ be the characteristic function of u, and h be a positive number depending on n. Define

$$H_n(z) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp(-itz)\frac{\phi_K(t)}{\phi_u(t/h)}dt. \tag{1.2}$$

Then the deconvolution kernel density estimator of X is given by

$$\hat{f}_n(x) = \frac{1}{nh}\sum_{i=1}^{n} H_n\left(\frac{x - Y_i}{h}\right). \tag{1.3}$$

It is easy to show that, if u is symmetric around 0, $\hat{f}_n(x)$ is a real function. The limiting behavior of the estimator (1.3) heavily depends on the tail of $\phi_u(t)$. The literature separate the measurement error distributions into two cases:

(1). Ordinary smooth of order β: If the characteristic function $\phi_u(t)$ satisfies

$$d_0|t|^{-\beta} \le |\phi_u(t)| \le d_1|t|^{-\beta} \quad \text{as } t \to \infty, \tag{1.4}$$

for positive constants d_0, d_1 and β.

(2). Super smooth of order β: If the characteristic function $\phi_u(t)$ satisfies

$$d_0|t|^{\beta_0}\exp(-|t|^{\beta}/\gamma) \le |\phi_u(t)| \le d_1|t|^{\beta_1}\exp(-|t|^{\beta}/\gamma) \quad \text{as } t \to \infty, \tag{1.5}$$

for positive constants d_0, d_1, β and γ, real constants β_0, β_1.

For more discussion on this classification and the impact on the limiting properties of (1.3), see Fan (1991), Fan and Truong (1993) and the references therein. Typical examples of ordinary smooth error distribution include double exponential distribution, symmetric gamma distribution, and examples of super smooth error distribution include normal distribution and Cauchy distribution.

The literature seems scant in constructing the confidence intervals for $f_X(x)$ in errors-in-variables model. Although Fan (1991)'s asymptotic normality results might help us to construct a z-type confidence interval, the resulting confidence interval might not have the desired coverage, due to the confliction between the prescribed undersmoothing and the explicit variance estimate needed by the z-type procedure.

The paper is organized as follows. After a brief introduction to empirical likelihood methodology, we shall extend Chen (1996)'s method in Section 2 to construct the empirical likelihood confidence intervals for the density function of X in the errors-in-variables model with both ordinary and super smooth measurement errors; z-type confidence intervals based on Fan (1991)'s asymptotic normality theories are developed in Section 3; some simulation studies are conducted in Section 4 to compare the finite performances of the empirical likelihood confidence intervals and the z-type confidence intervals, which are constructed based on Fan (1991)'s asymptotic normality theories; the proof of the main results is postponed to Section 5.

Throughout this paper, for any general random variable V, we use f_V and ϕ_V to denote its density function and characteristic function, "$\Longrightarrow_d$" to denote the convergence in distribution, and χ_1^2 to denote the χ^2 distribution with 1 degree of freedom.

2 Empirical Likelihood Confidence Interval

The empirical likelihood was introduced by Owen (1988). As an alternative to the bootstrap technique, empirical likelihood gained a striking success theoretically and practically in constructing the confidence intervals for the unknown parameters in various statistical models. See Owen (2001) for a comprehensive introduction to the empirical likelihood methodology. For the purpose of illustration, suppose $X_1, X_2, \ldots, X_n$ is a sample of size n from a population with mean μ. By assigning weights p_i to X_i, the empirical likelihood of μ is defined as $L(\mu) = \sup \prod_{i=1}^n p_i$, where the supremum is taken over the set $\{(p_1, \ldots, p_n) : \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i X_i = \mu\}$. For each μ value, the optimal p_i are found by some optimization procedures, the confidence interval for μ is then constructed by contouring the empirical likelihood function. Unlike the z-type intervals, no explicit variance estimation is required in the empirical likelihood procedure.

Let K be a smooth r-th order kernel function, that is, for some integer $r \geq 2$ and a constant κ,

$$\int v^j K(v)dv = \begin{cases} 1, & \text{if } j = 0, \\ 0, & \text{if } 1 \leq j \leq r-1, \\ \kappa, & \text{if } j = r. \end{cases}$$

Define the deconvolution kernel $H_n(z)$ and the corresponding deconvolution kernel density estimator $\hat{f}_n(x)$ as in (1.2) and (1.3). If we further assume that $f_X(x)$ has continuous derivatives up to the r-th order in a neighborhood of x, then one can show that for both ordinary and super smooth measurement errors, $E\hat{f}_n(x) = u(x)$, where

$$u(x) = f_X(x) + \frac{1}{r!}\kappa f^{(r)}(x)h^r + o_p(h^r). \tag{2.1}$$

Similar to Chen (1996), we first introduce the empirical likelihood for $u(x)$ as the mean of

$$H_{ni}(x) = h^{-1}H_n((x - Y_i)/h). \tag{2.2}$$

Then by properly removing the bias, we convert them into the confidence intervals for $f_X(x)$.

2.1 Ordinary Smooth Case

For this case, we will adopt the following technical assumptions:

(O1). The measurement error u is symmetric around 0.

(O2). $\phi_u(t)t^\beta \to c$ and $\phi'_u(t)t^{\beta+1} \to -\beta c$ as $t \to \infty$, with some constant $c \neq 0$ and $\beta \geq 0$. Moreover, $\phi_u(t) \neq 0$ for all t.

(O3). $\phi_K(t)$ is a symmetric function around 0, having $r + 2$ bounded integrable derivatives and $\phi_K(t) = 1 + O(|t|^r)$ as $t \to 0$, where $r \geq 2$ is fixed.

(O4). $\int_{-\infty}^{\infty}[|\phi_K(t)| + |\phi'_K(t)|]|t|^\beta dt < \infty$, $\int_{-\infty}^{\infty}|t|^{2\beta}|\phi_K(t)|^2 dt < \infty$.

(O5). $f_X(x)$ has continuous derivatives up to the r-th order.

(O6). The r-th order derivative of $f_X(x)$ is bounded.

(O7). $h \to 0$ and $nh^{2\beta+1} \to \infty$ as $n \to \infty$.

(O8). $h \to 0$ and $nh^{2\beta+2r+1} \to 0$ as $n \to \infty$.

Assumption (O2) is an alternative statement of (1.4) to indicate that the measurement error distribution is ordinary smooth. (O2)-(O5) are the same as in Fan (1991) to derive the asymptotic normality of $\hat{f}_n(x)$. Although Fan and Liu (1997) pointed out that a weaker condition than that used in Fan (1991) might lead to the same conclusion, we decide to keep Fan (1991)'s version in that these two set of assumptions are indeed equivalent because of (O1). (O3) is sufficient to ensure that the kernel K is of the order of r. (O7) is a typical condition used in deconvolution smoothing literature to trade off the asymptotic bias and variance of the deconvolution kernel density estimator. Assumptions (O1)-(O5) and (O7) suffice for the validity of empirical likelihood confidence interval for $u(x)$ as in Theorem 1, while (O1)-(O6) and (O8) are needed to guarantee the result stated in Theorem 2.

The empirical likelihood for $u(x)$ is defined by $L(u(x)) = \sup \prod_{i=1}^n p_i$, where the supremum is over the set of $\{(p_1, p_2, \ldots, p_i) : \sum p_i = 1, \sum p_i H_{ni}(x) = u(x)\}$. The maximizer of $p_i, i =$

$1, 2, \ldots, n$ is obtained by using Lagrangian multiplier, the log empirical likelihood ratio for $u(x)$ has the form of

$$l(u(x)) = -2\log[n^n L(u(x))] = 2\sum_{i=1}^{n} \log[1 + \lambda(H_{ni}(x) - u(x))],$$

where λ satisfies the following equation

$$\frac{1}{n}\sum_{i=1}^{n} \frac{H_{ni}(x) - u(x)}{1 + \lambda[H_{ni}(x) - u(x)]} = 0. \tag{2.3}$$

The following theorem states the asymptotic distribution of $l(u(x))$:

Theorem 1. *Suppose the conditions (O1)-(O5) and (O7) hold. Then for any fixed x, $l(u(x)) \Longrightarrow_d \chi_1^2$.*

Thus an empirical likelihood confidence interval for $u(x)$ with nominal confidence level $1 - \alpha$ is $\{u(x) : l(u(x)) \le \chi_1^2(1 - \alpha)\}$, where $\chi_1^2(1 - \alpha)$ denotes the $(1 - \alpha)$100-th percentile of χ_1^2.

The above confidence interval is for $u(x)$, but what we want is a confidence interval for $f_X(x)$. As pointed out by Hall (1991) and Chen (1996), there are two ways to obtain the confidence interval for $f_X(x)$. The first method is to shift the confidence interval for $u(x)$ by the estimated dominant bias term $\kappa \hat{f}^{(r)}(x)h^r/r!$ as in (2.1), where $\hat{f}^{(r)}(x)$ is defined by

$$\hat{f}^{(r)}(x) = \frac{1}{nh^{r+1}}\sum_{i=1}^{n} H_n^{(r)}\left(\frac{x - Y_i}{h}\right);$$

The second approach to implicitly correct the bias by under-smoothing, which can be implemented by choosing h properly to reduce the bias of $u(x)$. For this purpose, similar to derive $l(u(x))$, define

$$l(f_X(x)) = 2\sum_{i=1}^{n} \log[1 + \lambda(H_{ni}(x) - f_X(x))],$$

and λ satisfies

$$\sum_{i=1}^{n} \frac{H_{ni}(x) - f_X(x)}{1 + \lambda[H_{ni}(x) - f_X(x)]} = 0. \tag{2.4}$$

The following theorem states the asymptotic distribution of $l(f_X(x))$:

Theorem 2. *Suppose the conditions (O1)-(O6) and (O8) hold. Then for any fixed x, $l(f_X(x)) \Longrightarrow_d \chi_1^2$.*

Thus an empirical likelihood confidence interval for $f_X(x)$ with nominal confidence level $1 - \alpha$ is $\{f_X(x) : l(f_X(x)) \le \chi_1^2(1 - \alpha)\}$.

2.2 Super Smooth Case

In the super smooth case, the second moment of $H_{ni}(x)$ does not have an explicit order as in the ordinary smooth case, which makes it much harder to obtain similar results as Theorem 1 and 2. Denote $\delta_n = \text{Var}(H_{ni})$. To begin with, we state the following assumptions needed for developing the empirical likelihood confidence intervals in the super smooth case:

(S1). The measurement error u is symmetric around 0.

(S2). $c_1 \leq |\phi_u(t)||t|^{-\beta_0} \exp(|t|^\beta/\gamma) \leq c_2$ as $t \to \infty$ with $\beta, \gamma, c_1, c_2 > 0$ and some real number β_0. $\phi_u(t) \neq 0$ for all t.

(S3). $\phi_K(t)$ is symmetric and supported with $[-1, 1]$, having the first $r+2$ continuous derivatives. Moreover, $\phi_K(t) \geq c_3(1-t)^{r+3}$ for $t \in [1-\eta, 1]$ for some $c_3, \eta > 0$.

(S4). $\phi_K(t) = 1 + O(|t|^r)$ as $t \to \infty$.

(S5). $f_X(x)$ has continuous derivatives up to the r-th order.

(S6). $\delta_n \to \infty$ and $n/\delta_n \to \infty$ as $n \to \infty$.

(S7). $\delta_n \to \infty$ and $h^{2r}n/\delta_n \to 0$ as $n \to \infty$.

Assumption (S2) is an alternative statement of (1.5) to indicate that the measurement error distribution is super smooth. It is slightly different from the one in Fan (1991) because of (S1). (S2)-(S5) are similar to the conditions adopted in Fan (1991) to derive the asymptotic normality of $\hat{f}_n(x)$. As an example of a compacted support characteristic function of a kernel function, see Example 1 in Fan and Truong (1993) or Simulation 2 in Section 4. Condition (S6) is needed for constructing the empirical likelihood confidence interval of $u(x)$, and (S7) is for $f_X(x)$, they depend on the bandwidth h and play the same role in the super smooth case as (O6) and (O7) in ordinary smooth case. Based on the upper and lower bounds of δ_n provided in Fan (1991), one may state (S6) and (S7) in terms of n and h. We are not going to do so here for the sake of brevity.

The following theorem states the asymptotic distribution of $l(u(x))$:

Theorem 3. *Suppose the conditions (S1)-(S6) hold. Then for any fixed x, $l(u(x)) \Longrightarrow_d \chi_1^2$.*

Like the ordinary smooth case, one can obtain the confidence interval for $f_X(x)$ by shifting an estimator of the dominant bias term in the expansion of $u(x)$, which is the same as in the ordinary smooth case, or implicitly correcting the bias by undersmoothing using the bandwidth suggested in (S7). In fact, we have the following result:

Theorem 4. *Suppose the conditions (S1)-(S5), and (S7) hold. Then for any fixed x, $l(f_X(x)) \Longrightarrow_d \chi_1^2$.*

As a consequence, the empirical likelihood confidence intervals for $u(x)$ and $f_X(x)$ can be constructed in a similar fashion as in the ordinary smooth case.

3 Normal Theory Based Confidence Interval

Different confidence intervals for $f_X(x)$ can be constructed based on the asymptotic normality theories developed in Fan (1991). For the sake of completeness, below we reproduce the results in Fan (1991). For the ordinary smooth case, we have

Lemma 3.1. *Under assumptions (O2)-(O6), if* $h = o(n^{-1/(2\beta+2r+1)})$*, then*

$$\frac{\sqrt{n}(\hat{f}_n(x) - f_X(x))}{s_n} \Longrightarrow_d N(0,1), \tag{3.1}$$

where $s_n^2 = n^{-1}\sum_{i=1}^n H_{ni}^2(x)$ *or the sample variance defined by*

$$s_n^2 = n^{-1}\sum_{i=1}^n \left[H_{ni}(x) - \bar{H}_n(x)\right]^2,$$

and $\bar{H}_n(x) = n^{-1}\sum_{i=1}^n H_{ni}(x)$*,* $H_{ni}(x)$ *is defined by (2.2).*

For the super smooth case, we have

Lemma 3.2. *Under assumptions (S2)-(S5), if* $h \sim (a\gamma \log n/2)^{-1/\beta}$ *for some* $a > 1$*, then*

$$\frac{\sqrt{n}(\hat{f}_n(x) - f_X(x))}{s_n} \Longrightarrow_d N(0,1), \tag{3.2}$$

where $s_n^2 = n^{-1}\sum_{i=1}^n H_{ni}^2(x)$*,* $H_{ni}(x)$ *is defined by (2.2), and* β, γ *are defined in (S2).*

Therefore, by (3.1) and (3.2), a confidence interval of $f_X(x)$ with confidence level $1-\alpha$ for both ordinary and super smooth measurement errors can be constructed as follows:

$$[\hat{f}_n(x) - z_{1-\alpha/2}s_n, \hat{f}_n(x) + z_{1-\alpha/2}s_n], \tag{3.3}$$

where $z_{1-\alpha/2}$ is the $(1-\alpha/2)100$-th percentile of standard normal distribution.

It is easy to see that explicit estimators of the variance of $\hat{f}_n(x)$ are required in order to construct the confidence interval (3.3), while these estimators are not needed in constructing the empirical likelihood confidence interval.

4 Simulation Studies

In this section, we conduct two simulation studies to examine the finite sample performance of the empirical likelihood confidence intervals proposed in Section 2, and compare them with the z-type confidence intervals proposed in Section 3. In both simulation studies, 100 samples with sample sizes $n = 50, 100, 200$ are generated from $Y = X + u$ with ordinary and super smooth measurement error distributions. The distribution $f_X(x)$ of X is chosen to be the standard normal, and the confidence intervals are constructed for $f_X(0)$. The nominal confidence level is chosen to be 0.95, and for each simulation, we report the empirical coverage rate and the average length of confidence interval. The computation for empirical likelihood method is made possible by the R package `emplik`.

Simulation 1 (Ordinary Smooth Case): In this simulation, u follows the double exponential distribution with variance 1. The characteristic function $\phi_u(t)$ of u has the form of $2/(2+t^2)$ which is an ordinary smooth distribution of order $\beta = 2$. The kernel function K is chosen to be standard normal which is a second order kernel ($r = 2$). Based on the assumptions (O7) and (O8), the band width h is chosen to be $n^{-1/7}$, and from Fan and Truong (1993), the deconvolution kernel function $H_n(z)$ has the form of

$$H_n(z) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{z^2}{2}\right) \left[1 - \frac{1}{2h^2}(z^2 - 1)\right]. \tag{4.1}$$

The simulation results are reported in Table 1.

Table 1: Ordinary Smooth Case: Double Exponential Distribution

Sample Size	$n = 50$		$n = 100$		$n = 200$	
Types of CI	EL-CI	z-CI	EL-CI	z-CI	EL-CI	z-CI
Cov. Rate	0.92	0.90	0.94	0.93	0.95	0.94
Ave. Length	0.3771	0.4267	0.3296	0.3609	0.2918	0.3101

In Table 1, EL is the abbreviation for empirical likelihood, CI for confidence interval, Cov. for coverage and Ave. for average. Similar abbreviations are used in Table 2.

From Table 1, we can see that the empirical coverage rates of empirical likelihood confidence intervals are slightly closer to the nominal confidence level 0.95 than that of z-type confidence intervals, and the average lengths of empirical coverage rates of empirical likelihood confidence intervals are less than than that of z-type confidence intervals. This indicates that the empirical likelihood confidence intervals outperforms the z-type confidence intervals.

Simulation 2 (Super Smooth Case): In this simulation, u has the standard normal distribution. The characteristic function $\phi_u(t)$ of u has the form of $\exp(-t^2/2)$ which is a super smooth distribution of order $\beta = 2$. The kernel function K is chosen so that it has the characteristic function $\phi_K(t) = (1-t^2)^3$, $0 \le t \le 1$, and and from Fan and Truong (1993), the deconvolution kernel function $H_n(z)$ has the form of

$$H_n(z) = \frac{1}{\pi} \int_0^1 \cos(tz)(1-t^2)^3 \exp\left(\frac{t^2}{2h^2}\right) dt.$$

Based on the assumptions (S6) and (S7), the band width h is chosen to be $(2\log n)^{-1/2}$. The simulation results are reported in Table 2.

It is well known that estimating the density function $f_X(x)$ in super smooth measurement error models is rather difficult, the simulation results in Table 2 also confirm this point. From Table 2, we can see that the empirical coverage rates from both methods are all less than the nominal confidence

Table 2: Super Smooth Case: Standard Normal Distribution

Sample Size	$n = 50$		$n = 100$		$n = 200$	
Types of CI	EL-CI	z-CI	EL-CI	z-CI	EL-CI	z-CI
Cov. Rate	0.85	0.86	0.88	0.87	0.93	0.93
Ave. Length	0.3734	0.4492	0.4058	0.4916	0.4545	0.5780

level 0.95 when the sample size is small, but they approach to 0.95 as the sample size gets bigger. Like Simulation 1, the average lengths of empirical coverage rates of empirical likelihood confidence intervals are all less than than that of z-type confidence intervals. Once again, this indicates that the empirical likelihood confidence intervals behave better than the z-type confidence intervals.

One interesting phenomenon found in the simulation studies above is that the average length of the both types of confidence intervals are decreasing in Table 1 and increasing in Table 2 when sample size gets bigger.

5 Proofs of Main Results

This section includes the proofs of main results in Section 2.

Proof of Theorem 1: To derive the asymptotic distribution of $l(u(x))$, we have to find the magnitude of λ first. For this purpose, let $Z_{ni} = H_{ni}(x) - u(x)$, $M_n = \max_{1\le i\le n} |Z_{ni}|$. Then equation (2.3) can be written as

$$0 = \left| \frac{1}{n}\sum_{i=1}^{n} Z_{ni} - \frac{\lambda}{n}\sum_{i=1}^{n} \frac{Z_{ni}^2}{1+\lambda Z_{ni}} \right|.$$

Applying triangular inequality on the right hand side, we have

$$\frac{|\lambda|}{n}\sum_{i=1}^{n} \frac{Z_{ni}^2}{1+\lambda|Z_{ni}|} \le \left| \frac{1}{n}\sum_{i=1}^{n} Z_{ni} \right|. \tag{5.1}$$

From Fan (1991), one can show that

$$EZ_{ni}^2 = E|H_{ni}(x) - u(x)|^2 = \frac{v(x)}{h^{2\beta+1}}[1+o(1)],$$

where

$$v(x) = \frac{f_X(x)}{2\pi|c|}\int_{-\infty}^{\infty} |t|^{2\beta}|\phi_K(t)|^2 dt, \quad 0 \ne c = \lim_{t\to\infty} t^\beta \phi_u(t).$$

Therefore, $E\left(h^{(2\beta+1)/2} Z_{ni}\right)^2 = O(1)$, which implies $M_n = \sup_{1\le i\le n} |Z_{ni}| = o_p(\sqrt{n}h^{-\beta-1/2})$. Also, one can easily show that

$$\frac{1}{n}\sum_{i=1}^{n} Z_{ni} = O_p\left(\frac{1}{\sqrt{nh^{2\beta+1}}}\right), \quad \frac{1}{n}\sum_{i=1}^{n} Z_{ni}^2 = O_p\left(\frac{1}{h^{2\beta+1}}\right). \tag{5.2}$$

By (5.1), one has

$$\frac{|\lambda|}{1+\lambda M_n} \cdot \frac{1}{n}\sum_{i=1}^{n} Z_{ni}^2 \le \left|\frac{1}{n}\sum_{i=1}^{n} Z_{ni}\right|$$

and from (5.2), we obtain

$$|\lambda| \cdot O_p(h^{-2\beta-1}) \le [1+|\lambda| o_p(\sqrt{n}h^{-\beta-1/2})] \cdot O_p(n^{-1/2}h^{-\beta-1/2})$$

which implies

$$|\lambda| = O_p\left(\frac{h^{\beta+1/2}}{\sqrt{n}}\right). \tag{5.3}$$

Denote $r_{ni} = \lambda Z_{ni}$, then from (5.3)

$$\max_{1\le i\le n} |r_{ni}| = \lambda M_n = O_p(n^{-1/2}h^{\beta+1/2}) \cdot o_p(n^{1/2}h^{-\beta-1/2}) = o_p(1). \tag{5.4}$$

Using the notations defined above, (2.3) can be written as

$$\frac{1}{n}\sum_{i=1}^{n} Z_{ni}\left(1-r_{ni}+\frac{r_{ni}^2}{1+r_{ni}}\right) = \frac{1}{n}\sum_{i=1}^{n} Z_{ni} - \frac{\lambda}{n}\sum_{i=1}^{n} Z_{ni}^2 + \frac{\lambda^2}{n}\sum_{i=1}^{n}\frac{Z_{ni}^3}{1+r_{ni}} = 0.$$

Note that

$$\left|\frac{\lambda^2}{n}\sum_{i=1}^{n}\frac{Z_{ni}^3}{1+r_{ni}}\right| \le \frac{\lambda^2 M_n}{n}\sum_{i=1}^{n} Z_{ni}^2(1+r_{ni})^{-1} = n^{-1}h^{2\beta+1}\cdot O_p(h^{-2\beta-1})\cdot o_p(n^{1/2}h^{-\beta-1/2})$$
$$= o_p(1/\sqrt{nh^{2\beta+1}}).$$

Therefore, we may write

$$\lambda = \left(\frac{1}{n}\sum_{i=1}^{n} Z_{ni}^2\right)^{-1} \cdot \frac{1}{n}\sum_{i=1}^{n} Z_{ni} + \beta_n, \quad \beta_n = o_p\left(\sqrt{\frac{h^{2\beta+1}}{n}}\right). \tag{5.5}$$

Also, from (5.4) we can obtain that

$$\log(1+r_{ni}) = r_{ni} - \frac{r_{ni}^2}{2} + \eta_{ni},$$

where

$$P(|\eta_{ni}| \le B|r_{ni}|^3, 1 \le i \le n) \to 1, \text{ for some finite } B > 0 \tag{5.6}$$

as $n \to \infty$.

Denote

$$S_{n1} = \frac{1}{n}\sum_{i=1}^{n} Z_{ni}, \quad S_{n2} = \frac{1}{n}\sum_{i=1}^{n} Z_{ni}^2.$$

Then from (5.5), we obtain

$$\begin{aligned} l(u(x)) &= 2\sum_{i=1}^{n}\log(1+r_{ni}) = 2\sum_{i=1}^{n} r_{ni} - \sum_{i=1}^{n} r_{ni}^2 + 2\sum_{i=1}^{n}\eta_{ni} \\ &= 2nS_{n1}[S_{n2}^{-1}S_{n1}+\beta_n] - n[S_{n2}^{-1}S_{n1}+\beta_n]^2 S_{n2} + 2\sum_{i=1}^{n}\eta_{ni} \\ &= nS_{n1}^2 S_{n2}^{-1} - n\beta_n^2 S_{n2} + 2\sum_{i=1}^{n}\eta_{ni} \\ &= nS_{n1}^2 S_{n2}^{-1} + 2\sum_{i=1}^{n}\eta_{ni} + o_p(1) \end{aligned}$$

from (5.2) and (5.5). By (5.6),

$$|2\sum_{i=1}^{n}\eta_{ni}| \le 2B\sum_{i=1}^{n}|r_{ni}|^3 = 2B|\lambda|^3\sum_{i=1}^{n}|Z_{ni}|^3 \le 2nB|\lambda|^3 M_n S_{n2} = o_p(1)$$

from (5.2), (5.5) and the fact $M_n = \sup_{1\le i\le n}|Z_{ni}| = o_p(\sqrt{n}h^{-\beta-1/2})$.

From Theorem 2.1 in Fan (1991), we have $nS_{n1}^2 S_{n2}^{-1} \Longrightarrow_d N(0,1)$. Therefore, we finally obtain

$$l(u(x)) \Longrightarrow_d \chi_1^2.$$

This finishes the proof. □

Proof of Theorem 2: Similar to Chen (1996), let

$$w_i(x) = H_{ni}(x) - f_X(x), \quad \bar{w}_j = \frac{1}{n}\sum_{i=1}^{n} w_i^j(x), \quad j = 1, 2.$$

First we need to find the order of $\bar{w}_1$ and $\bar{w}_2$. Consider the expectation of $\bar{w}_2$,

$$E\bar{w}_2 = E[H_{ni}(x) - EH_{ni}(x)]^2 + [EH_{ni}(x) - f_X(x)]^2 = O\left(\frac{1}{h^{2\beta+1}}\right) + O(h^{2r}) = O\left(\frac{1}{h^{2\beta+1}}\right),$$

therefore, $\bar{w}_2 = O_p(1/h^{2\beta+1})$. For $\bar{w}_1$, we have

$$\begin{aligned} E\bar{w}_1^2 &= E\left[\frac{1}{n}\sum_{i=1}^{n} H_{ni}(x) - EH_{ni}(x) + EH_{ni}(x) - f_X(x)\right]^2 \\ &\le 2E\left[\frac{1}{n}\sum_{i=1}^{n} H_{ni}(x) - EH_{ni}(x)\right]^2 + 2\left[EH_{ni}(x) - f_X(x)\right]^2 \\ &= O\left(\frac{1}{nh^{2\beta+1}}\right) + O(h^{2r}) \end{aligned}$$

which implies $\bar{w}_1 = O_p(1/\sqrt{nh^{2\beta+1}} + h^r)$. Based on the result for $\bar{w}_2$, we can also verify that

$$\sup_{1\le i\le n} |H_{ni}(x) - f_X(x)| = o_p(n^{1/2}h^{-\beta-1/2}). \tag{5.7}$$

By similar argument as in $u(x)$ case, we get

$$\frac{|\lambda|\bar{w}_2}{1 + |\lambda| \sup_{1\le i\le n} |H_{ni}(x) - f_X(x)|} \le |\bar{w}_1|.$$

Therefore, based on the asymptotic order of $\bar{w}_1$, $\bar{w}_2$ and (5.7), by (O7), we obtain that

$$\lambda = O_p\left(\sqrt{\frac{h^{2\beta+1}}{n}} + h^{2\beta+r+1}\right). \tag{5.8}$$

Similar to the case of $u(x)$, we can write (2.4) as

$$\bar{w}_1 - \lambda\bar{w}_2 + \lambda^2 \frac{1}{n}\sum_{i=1}^n \frac{w_i^3}{1+\lambda w_i} = 0.$$

Note that

$$|\lambda| \cdot \sup_{1\le i\le n} |w_i| = O_p\left(\sqrt{\frac{h^{2\beta+1}}{n}} + h^{2\beta+r+1}\right) \cdot o_p(n^{1/2}h^{-\beta-1/2}) = o_p(1),$$

we have

$$\begin{aligned}\left|\frac{\lambda^2}{n}\sum_{i=1}^n \frac{w_i^3}{1+\lambda w_i}\right| &\le \frac{\lambda^2 \sup_{1\le i\le n}|w_i|}{n}\sum_{i=1}^n w_i^2(1+\lambda w_i)^{-1} \\ &= O_p\left(\left(\sqrt{\frac{h^{2\beta+1}}{n}} + h^{2\beta+r+1}\right)^2\right) \cdot O_p(h^{-2\beta-1}) \cdot o_p(n^{1/2}h^{-\beta-1/2}) \\ &= o_p\left(\frac{1}{\sqrt{nh^{2\beta+1}}}\right) + o_p(\sqrt{n}h^{\beta+2r+1/2}) + o_p(h^r) = o_p(1).\end{aligned}$$

Therefore, we may write

$$\lambda = \bar{w}_2^{-1}\bar{w}_1 + \beta_n, \quad \beta_n = o_p\left(\sqrt{\frac{h^{2\beta+1}}{n}}\right) + o_p(\sqrt{n}h^{3\beta+2r+3/2}) + o_p(h^{2\beta+r+1}). \tag{5.9}$$

Also, from the fact that $\max_{1\le i\le n}|\lambda w_i| = o_p(1)$, we can obtain that

$$\log(1+\lambda w_i) = \lambda w_i - \frac{\lambda^2 w_i^2}{2} + \eta_{ni},$$

where

$$P(|\eta_{ni}| \le B|\lambda w_i|^3, 1 \le i \le n) \to 1, \text{ for some finite } B > 0 \tag{5.10}$$

as $n \to \infty$.

Then from (5.9), we obtain

$$\begin{aligned} l(f_X(x)) &= 2\sum_{i=1}^{n}\log(1+\lambda w_i) = 2\lambda\sum_{i=1}^{n} w_i - \lambda^2\sum_{i=1}^{n} w_i^2 + 2\sum_{i=1}^{n}\eta_{ni} \\ &= 2n\bar{w}_2[\bar{w}_2^{-1}\bar{w}_1 + \beta_n] - n[\bar{w}_2^{-1}\bar{w}_1+\beta_n]^2\bar{w}_2 + 2\sum_{i=1}^{n}\eta_{ni} \\ &= n\bar{w}_1^2\bar{w}_2^{-1} - n\beta_n^2\bar{w}_2 + 2\sum_{i=1}^{n}\eta_{ni} \\ &= n\bar{w}_1^2\bar{w}_2^{-1} + 2\sum_{i=1}^{n}\eta_{ni} + o_p(1) \end{aligned}$$

from the order of $\bar{w}_2$ and (5.9). By (5.10),

$$\left|2\sum_{i=1}^{n}\eta_{ni}\right| \le 2B\lambda^3\sum_{i=1}^{n}|w_i|^3 = 2B|\lambda|^3\sum_{i=1}^{n}|w_i|^3 \le 2nB|\lambda|^3\bar{w}_2\cdot\sup_{1\le i\le n}|w_i| = o_p(1) \tag{5.11}$$

from (5.8), the order of $\bar{w}_2$ and the fact $\sup_{1\le i\le n}|w_i| = o_p(\sqrt{n}h^{-\beta-1/2})$.

From Corollary 2.1 in Fan (1991), we have $n\bar{w}_1^2\bar{w}_2^{-1} \Longrightarrow_d N(0,1)$. The result of Theorem 2 follows. □

Proof of Theorem 3: As in the ordinary smooth case, denote $Z_{ni} = H_{ni}(x) - u(x)$, where $u(x) = EH_{ni}(x)$. We have

$$M_n = \sup_{1\le i\le n}|Z_{ni}| = o_p(\sqrt{n}\delta_n),\ S_{n1} = \frac{1}{n}\sum_{i=1}^{n} Z_{ni} = O_p(\sqrt{\delta_n/n}),\ S_{n2} = \frac{1}{n}\sum_{i=1}^{n} Z_{ni}^2 = O_p(\delta_n).$$

By assumption (S6), one can obtain that

$$\lambda = O_p(1/\sqrt{n\delta_n}). \tag{5.12}$$

Let $r_i = \lambda[H_{ni}(x) - u(x)]$. From (5.12), we have

$$\max_{1\le i\le n}|r_i| = \lambda\max_{1\le i\le n}|H_{ni}(x) - u(x)| = O_p(1/\sqrt{n\delta_n})\cdot O_p(\sqrt{n\delta_n}) = o_p(1). \tag{5.13}$$

Therefore,

$$\frac{\lambda^2}{n}\sum_{i=1}^{n}\frac{[H_{ni}(x)-u(x)]^3}{1+r_i} = o_p(\sqrt{\delta_n/n}).$$

Similar to the ordinary smooth case, by expanding (2.3), we have

$$\lambda = S_{n2}^{-1}S_{n1} + \beta_n, \quad \beta_n = o_p(1/\sqrt{n\delta_n}).$$

From (5.13), one also has

$$\log(1+r_i) = r_i - \frac{r_i^2}{2} + \eta_i,$$

where

$$P(|\eta_i| \leq B|r_i|^3, 1 \leq i \leq n) \to 1$$

for some finite B as $n \to \infty$. As in the ordinary smooth case, we have

$$l(u(x)) = nS_{n2}^{-1}S_{n1}^2 - n\beta_n^2 S_{n2} + 2\sum_{i=1}^n \eta_i.$$

But

$$n\beta_n^2 S_{n2} = n \cdot o_p(1/n\delta_n) \cdot \delta_n = o_p(1), \quad |2\sum_{i=1}^n \eta_i| \leq 2B|\lambda|^3 M_n = o_p(1).$$

Therefore, $l(u(x)) = nS_{n2}^{-1}S_{n1}^2 + o_p(1)$. Hence Theorem 3 follows from $\sqrt{n}S_{n2}^{-1/2}S_{n1} \Longrightarrow_d N(0,1)$, see Fan (1991). □

Proof of Theorem 4: The second approach is made possible by the facts

$$E\bar{w}_1^2 = O(\delta_n/n + h^{2r})$$

which implies $\bar{w}_1 = O_p(\sqrt{\delta_n/n} + h^r)$, and $E\bar{w}_2 = \delta_n$. Therefore,

$$\frac{|\lambda| O_p(\delta_n)}{1 + |\lambda| o_p(\sqrt{n\delta_n})} = O_p\left(\sqrt{\delta_n/n} + h^r\right).$$

Then from (S7), we have

$$\lambda = O_p\left(\frac{1}{\sqrt{n\delta_n}} + \frac{h^r}{\delta_n^2}\right). \tag{5.14}$$

Note that $\sup_{1\leq i\leq n}|w_i| = o_p(\sqrt{n\delta_n} + \sqrt{nh^{2r}}) = o_p(\sqrt{n\delta_n})$, hence

$$|\lambda| \cdot \sup_{1\leq i\leq n}|w_i| = O_p\left(\sqrt{\frac{1}{n\delta_n}} + \frac{h^r}{\delta_n}\right) \cdot o_p(\sqrt{n\delta_n}) = o_p(1),$$

we have

$$\begin{aligned}\left|\frac{\lambda^2}{n}\sum_{i=1}^n \frac{w_i^3}{1+\lambda w_i}\right| &\leq \frac{\lambda^2 \sup_{1\leq i\leq n}|w_i|}{n}\sum_{i=1}^n w_i^2(1+\lambda w_i)^{-1} \\ &= O_p\left(\left(\frac{1}{\sqrt{n\delta_n}} + \frac{h^r}{\delta_n}\right)^2\right) \cdot O_p(\delta_n) \cdot o_p(\sqrt{n\delta_n}).\end{aligned}$$

Therefore, from

$$\bar{w}_1 - \lambda\bar{w}_2 + \lambda^2\frac{1}{n}\sum_{i=1}^n \frac{w_i^3}{1+\lambda w_i} = 0,$$

we have

$$\lambda = \bar{w}_2^{-1}\bar{w}_1 + \beta_n, \quad \beta_n = o_p\left(\frac{1}{\sqrt{n\delta_n}}\right) + o_p(\sqrt{n}h^{2r}/\sqrt{\delta_n^3}) + o_p(h^r/\delta_n). \tag{5.15}$$

Also, from the fact that $\max_{1\le i\le n}|\lambda w_i| = o_p(1)$, we can obtain that

$$\log(1+\lambda w_i) = \lambda w_i - \frac{\lambda^2 w_i^2}{2} + \eta_{ni},$$

where

$$P(|\eta_{ni}| \le B|\lambda w_i|^3, 1 \le i \le n) \to 1, \text{ for some finite } B > 0 \tag{5.16}$$

as $n \to \infty$.

Then from (5.15) and $w_2 = O_p(\delta_n)$, we obtain

$$l(f_X(x)) = n\bar{w}_1^2\bar{w}_2^{-1} + 2\sum_{i=1}^n \eta_{ni} + o_p(1).$$

By (5.16),

$$\left|2\sum_{i=1}^n \eta_{ni}\right| \le 2B\lambda^3\sum_{i=1}^n |w_i|^3 = 2B|\lambda|^3\sum_{i=1}^n |w_i|^3 \le 2nB|\lambda|^3\bar{w}_2 \cdot \sup_{1\le i\le n}|w_i| = o_p(1)$$

from (5.14), the order of $\bar{w}_2$ and the fact $\sup_{1\le i\le n}|w_i| = o_p(\sqrt{n\delta_n})$.

From Corollary 2.2 in Fan (1991), we have $n\bar{w}_1^2\bar{w}_2^{-1} \Longrightarrow_d N(0,1)$. Therefore, we finally obtain $l(f_X(x)) \Longrightarrow_d \chi_1^2$. This finishes the proof. □

References

[1] Chen, S. X. (1996). Empirical likelihood confidence intervals for nonparametric density estimation. *Biometrika*, **83**(2), 329-341.

[2] Fan, J. Q. (1991). Asymptotic normality for deconvolution kernel density estimators. *Sankhyā*, **53**, Series A, 97-110.

[3] Fan, J. Q. and Truong, Y. K. (1993). Nonparametric regression with errors in variables. *Ann. Statist.* **21**, 1900-1925.

[4] Fan, Y. Q. and Liu, Y. J. (1997). A note on asymptotic normality for deconvolution kernel density estimators. *Sankhyā*, **59**, Series A, 138-141.

[5] Hall, P. (1991). Edgeworth expansions for nonparametric density estimators, with applications. *Statistics*, **22**, 215-232.

[6] Hall, P. and Owen, A. (1993). Empirical likelihood confidence bands in density estimation. *J. Computat. Graph. Statist.* **2**, 273-289.

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

[8] Owen, A. (2001). *Empirical likelihood*. Chapman and Hall/CRC.

[9] Scott, D. W. (1992). *Multivariate density estimation*. New York: Wiley.

[10] Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. London: Chapman and Hall.