Acta Mathematicae Applicatae Sinica, English Series
Vol. 19, No. 3 (2003) 387–396

Acta Mathematicae Applicatae Sinica, English Series

Coverage Accuracy of Confidence Intervals in Nonparametric Regression

Song-xi Chen [1], Yong-song Qin[2]

[1]Department of Statistics and Applied Probability, National University of Singapore, Singapore 117543, Singapore (Email: stacsx@nus.edu.sg)

[2]Department of Mathematics, Guangxi Normal University, Guilin 541004, China (Email: ysqin@eyou.com)

Abstract Point-wise confidence intervals for a nonparametric regression function with random design points are considered. The confidence intervals are those based on the traditional normal approximation and the empirical likelihood. Their coverage accuracy is assessed by developing the Edgeworth expansions for the coverage probabilities. It is shown that the empirical likelihood confidence intervals are Bartlett correctable.

Keywords Confidence interval, empirical likelihood, Nadaraya-Watson estimator, normal approximation
2000 MR Subject Classification 62G05, 62E20

1 Introduction

Let $(X_1, Y_1), \cdots, (X_n, Y_n)$ be independent copies of R^{d+1} random vectors (X, Y) where $X \in R^d$ with density f. Let $m(x) = E(Y|X = x)$ be the conditional mean function and $\sigma^2(x) = \mathrm{Var}(Y|X = x)$ be the conditional variance function of Y given $X = x$. For any $x, u \in R^d$, let $V(u|x) = E[\{Y - m(x)\}^2|X = u]$. Clearly, $V(u|x) = \sigma^2(u) + \{m(u) - m(x)\}^2$.

The Nadaraya-Watson estimator for $m(x)$ at any given x is

$$\widehat{m}(x) = \frac{\sum K_h(x - X_i)Y_i}{\sum K_h(x - X_i)}, \tag{1.1}$$

where $K_h(t) = K(t/h)$, K is a kernel function and h is the smoothing bandwidth.

This paper is concerned with the construction of point-wise confidence intervals for $m(x)$ at any fixed x in conjunction with the Nadaraya-Watson estimator when the design points are random. Confidence intervals based on the asymptotic normality of the Nadaraya-Watson estimator and the percentile bootstrap are reviewed in [11]. Hall[8] considers coverage accuracy of the percentile-t bootstrap confidence intervals in the case of fixed design. A dominate issue in constructing confidence intervals in nonparametric curve estimation is the bias associated with the curve estimators. The bias has to be reduced to make the confidence intervals appropriate as confidence intervals for $m(x)$. One way of reducing the bias is to undersmooth by using a smaller bandwidth h; another way is to conduct explicit bias correction as used in [8].

The paper is aimed at studying coverage accuracy of the confidence intervals based on the normal approximation and the empirical likelihood by developing Edgeworth expansions for the coverage probabilities. It is shown that the empirical likelihood confidence interval is Bartlett correctable.

Manuscript received January 13, 2001.

The paper is structured as follows. Section 2 introduces confidence intervals originally for the mean of the Nadaraya-Watson estimator rather than $m(x)$. Section 3 reveals a sufficient and necessary condition in converting the confidence intervals to those for $m(x)$. The condition implies that a smaller than usual bandwidth should be used. The coverage accuracy of the confidence intervals is studied in Section 4 by developing Edgeworth expansions. Section 5 considers the Bartlett correction for the empirical likelihood confidence interval. Some proofs are given in Appendices.

2 Confidence Intervals for $E\{\widehat{m}(x)\}$

We outline two types of confidence intervals in nonparametric regression. The first type, based on the traditional normal approximation, exist already. The one based on the empirical likelihood is new. We first introduce some notations and assumptions.

The Nadaraya-Watson estimator $\widehat{m}(x)$ in (1.1) satisfies

$$\sum_{i=1}^{n} K_h(x-X_i)\{Y_i-\widehat{m}(x)\}=0.$$

Here a single bandwidth h is used, implying that the design points have the same scale in all directions. When this is not true, the method of Fukunaga[6] can be used to re-scale the design points.

We assume the following regularity conditions:

(i) K is a d-dimensional compactly supported symmetrical kernel which is a probability density itself satisfying $\int u_i u_j K(u)du=\sigma_i^2(K)\delta_{ij}$ where δ_{ij} is the Kronecker delta;

(ii) $h\to 0$ and $nh^d\to\infty$ as $n\to\infty$; and there is an $s\ge d+2$ such that $E|Y|^s<\infty$, $nh^{2s}\to 0$ and $n^{s-2}h^{sd}\to\infty$;

(iii) f and m have continuous partial derivatives up to the 2-nd order in a neighborhood of x, and $f(x)>0$;

(iv) $V(u)=V(u|x)$ has continuous first derivatives in a neighborhood of x and $V(x|x)>0$.

Define $\omega_i=K_h(x-X_i)\{Y_i-m(x)\}$ and, for positive integers j,

$$\overline{\omega}_j=(nh^d)^{-1}\sum_{i=1}^{n}\omega_i^j,\qquad \mu_j=E(\overline{\omega}_j)\qquad\text{and}\qquad R_j(K)=\int K^j(u)du.$$

The first type of confidence intervals is based on the fact that

$$\frac{(nh^d)^{1/2}\big[\widehat{m}(x)-E\{\widehat{m}(x)\}\big]}{\sqrt{V(x)R_2(K)/f(x)}}\xrightarrow{d}N(0,1),\tag{2.1}$$

which was derived by Schuster[17] by applying the Central Limit Theorem. Let $z_{\frac{1+\alpha}{2}}$ be the $\frac{1+\alpha}{2}$-quantile of $N(0,1)$. Then, a normal approximation based confidence interval for $E\{\widehat{m}(x)\}$ with nominal coverage α, as given in [11] (p.100), is

$$I_{\alpha,\text{nor}}=\widehat{m}(x)\pm z_{\frac{1+\alpha}{2}}\sqrt{R_2(K)\widehat{V}(x)/\{nh^d\widehat{f}(x)\}},\tag{2.2}$$

where $\widehat{f}(x)=(nh^d)^{-1}\sum\limits_{i=1}^{n}K_h(x-X_i)$ is the kernel estimator for the design density f and

$$\widehat{V}(x)=(nh^d)^{-1}\sum K_h(x-X_i)\{Y_i-\widehat{m}(x)\}^2/\widehat{f}(x).\tag{2.3}$$

The second type of confidence intervals considered is based on the empirical likelihood. The empirical likelihood, introduced by Owen[14,15], is a computer intensive statistical method like the bootstrap. However, instead of using an equal probability weight n^{-1} for all data values, the empirical likelihood chooses the weights, say p_i on the i-th data points (X_i, Y_i), by profiling a multinomial likelihood under a set of constraints. The constraints reflect the characteristics of the quantity of interests. A review of the empirical likelihood is given in [10].

Let $p_1, \dots, p_n$ be nonnegative numbers adding to unity. The empirical likelihood at θ, a candidate value of $E\{\widehat{m}(x)\}$, is defined as

$$L(\theta) = \sup_{\sum p_i K_h(x-X_i)(Y_i-\theta)=0} \prod_{i=1}^{n} p_i \,.$$

After using a Lagrange multiplier to find the optimal p_i, the log empirical likelihood ratio is

$$\ell(\theta) = -2\log\{L(\theta)n^n\} = 2\sum \log\big\{1+\lambda(\theta)K_h(x-X_i)(Y_i-\theta)\big\},$$

where $\lambda(\theta)$ satisfies

$$\sum_{1}^{n} K_h(x-X_i)(Y_i-\theta)\big\{1+\lambda(\theta)K_h(x-X_i)(Y_i-\theta)\big\}^{-1} = 0. \tag{2.4}$$

An empirical likelihood confidence interval with nominal coverage of α, denoted as I_α, is

$$I_{\alpha,\mathrm{el}} = \{\theta \,|\, \ell(\theta) \le c_\alpha\}, \tag{2.5}$$

where $c_\alpha = z^2_{\frac{1+\alpha}{2}}$ is the αth quantile of the χ^2_1 distribution. A special feature of the empirical likelihood confidence interval is that no explicit variance estimator, like the one in (2.3) for $V(x)$, is required in its construction as the studentizing is carried out internally via the optimization procedure.

The two confidence intervals introduced in this section are for $E\{\widehat{m}(x)\}$ which is $m(x)$ plus the bias of the regression estimator. To convert the confidence intervals into those of $m(x)$, the bias has to be corrected. We consider the approach of undersmoothing in this paper.

3 A Sufficient and Necessary Condition

In this section we show that, subject to Conditions (i) to (iv), a sufficient and necessary condition for the two confidence intervals to have correct asymptotic coverage is

$$nh^d\mu_1^2 = o(1). \tag{3.1}$$

The coverage probability of $I_{\alpha,\mathrm{nor}}$ for $m(x)$ is $P\{|T_n| \le z_{\frac{1+\alpha}{2}}\}$ where

$$T_n = \frac{\widehat{m}(x)-m(x)}{\sqrt{R_2(K)\widehat{V}(x)/\{nh^d\widehat{f}(x)\}}}.$$

An expansion of T_n is

$$\begin{aligned} T_n =& \big\{R_2(K)f(x)V(x)\big\}^{-1/2}\sqrt{nh^d}\,\overline{\omega}_1 + o_p\big\{(nh^d)^{1/2}(nh^d)^{-1/2}+h^2\big\} \\ =& Z + (nh^d)^{1/2}\mu_2^{-1/2}\mu_1 + o_p\big[(nh^d)^{1/2}\big\{(nh^d)^{-1/2}+h^2\big\}\big], \end{aligned} \tag{3.2}$$

where $Z = (nh^d)^{1/2}(\overline{\omega}_1 - \mu_1)\mu_2^{-1/2}$ is asymptotically distributed as $N(0,1)$. Thus, T_n is asymptotically $N(0,1)$ if and only if $(nh^d)^{1/2}\mu_2^{-1/2}\mu_1 = o(1)$. As $\mu_2 = E(\overline{\omega}_2) = f(x)V(x)R_2(K) + o(1) \neq 0$ almost surely under Condition (iii), the above condition is equivalent to (3.1).

To evaluate the coverage of $I_{\alpha,\mathrm{el}}$, we notice from (2.4) that

$$\sum \omega_i - \lambda(\theta) \sum \frac{\omega_i^2}{1+\lambda(\theta)\omega_i} = 0. \tag{3.3}$$

Put $Z_n = \max_i |\omega_i| = \circ(n^{1/s})$, then

$$\frac{|\lambda(\theta)|}{1+|\lambda(\theta)|Z_n}\overline{\omega}_2 = O_p\{(nh^d)^{-1/2} + h^2\}.$$

As $Z_n = \circ(n^{1/s})$ almost surely, and $nh^{2s} \to 0$ and $n^{s-2}h^{sd} \to \infty$ as assumed in Condition (ii), it may be shown, similarly to that given in [15], that

$$\lambda(\theta) = Op\{(nh^d)^{-1/2} + h^2\}.$$

From (3.3), we have $\lambda(\theta) = (\overline{\omega}_2)^{-1}\overline{\omega}_1 + O_p\big[\{(nh^d)^{-1/2} + h^2\}^2\big]$. Thus,

$$\begin{aligned}\ell(\theta) &= nh^d\mu_2^{-1}\overline{\omega}_1^2 + \circ_p\big[nh^d\big\{(nh^d)^{-1/2} + h^2\big\}^2\big] \\ &= \big\{Z + (nh^d)^{1/2}\mu_2^{-1/2}\mu_1\big\}^2 + \circ_p\big[nh^d\big\{(nh^d)^{-1/2} + h^2\big\}^2\big].\end{aligned}$$

Now the same argument as used after (3.2) implies that $\ell(\theta)$ is asymptotically χ_1^2 if and only if (3.1) is true.

Note that $\mu_1 = \frac{1}{2}h^2 B(f,m,K,x) + O(h^4)$ where

$$B(f,m,K,x) = \sum_{i=1}^{d} \sigma_i^2(K)\big\{f(x)m_i''(x) + 2f_i'(x)m_i'(x)\big\}. \tag{3.4}$$

Here $f_i'(x)$, $m_i'(x)$ and $m_i''(x)$, respectively, are the partial derivatives with respect to x_i, the i-th component of $x = (x_1, \cdots, x_n)$. When $d = 1$, $B(f,m,K,x) = \sigma_1^2(K)\{f(x)m''(x) + 2f'(x)m'(x)\}$. If $B(f,m,K,x) \neq 0$, then (3.1) is equivalent to

$$h = o\{n^{-1/(d+4)}\}. \tag{3.5}$$

Note that $n^{-1/(d+4)}$ is the standard order for the smoothing bandwidth in curve estimation when an nonnegative kernel is used. Thus, (3.5) implies undersmoothing.

Remark. If we take $h = O(n^{-l})$ with $1/(d+4) < l < (s-2)/(sd)$ for $s \geq d+2$, then condition (ii) and condition (3.5) are both satisfied.

4 Coverage Accuracy

In studying the coverage accuracy of the two types of confidence intervals, we assume two extra conditions:

(v) $h = o\{n^{-1/(d+4)}\}$, and

(vi) $nh^d(\log n)^{-1} \to \infty$ and $E|Y|^{15} < \infty$.

Condition (v) implies undersmoothing. Condition (vi) is needed to develop Edgeworth expansions for the coverage probabilities.

We need the following notations for describing the coverage errors:

$$V_j(u|x) = E\big[\{Y - m(x)\}^j \,|\, X = u\big] \quad \text{and} \quad V_j(x) = V_j(x|x) \qquad \text{for } j = 3, 4;$$
$$\mu_{ms} = h^{-d} E\big[K_h^m(x - X_i)\{Y_i - m(x)\}^s\big] \qquad \text{for positive integers } m > 1 \text{ and } s;$$

and H_j are the Hermite polynomials of the j-th order.

It may be shown that

$$\begin{aligned} &\mu_3 = V_3(x)f(x)R_3(K) + \circ(1), \qquad \mu_4 = V_4(x)f(x)R_4(K) + \circ(1),\\ &\mu_{23} = V_3(x)f(x)R_2(K) + \circ(1), \qquad \mu_{24} = V_4(x)f(x)R_2(K) + \circ(1),\\ &\mu_{34} = V_4(x)f(x)R_3(K) + \circ(1). \end{aligned} \tag{4.1}$$

The coverage accuracy of the confidence intervals is studied one by one in the following subsections.

4.1 Coverage Accuracy of $I_{\alpha,\text{nor}}$

Let $\phi(\cdot)$ be the density of the standard normal distribution. Derivations deferred until Appendix A.1 give the following Edgeworth expansion for the coverage probability of $I_{\alpha,\text{nor}}$:

$$\begin{aligned} P\{m(x) \in I_{\alpha,\text{nor}}\} =& \alpha - \big\{b_1(z_{\frac{1+\alpha}{2}})nh^{d+4} + b_2(z_{\frac{1+\alpha}{2}})h^2 + b_3(z_{\frac{1+\alpha}{2}})(nh^d)^{-1}\big\}\phi(z_{\frac{1+\alpha}{2}})\\ &+ O\big[\big\{(nh^d)^{1/2}h^2 + (nh^d)^{-1/2}\big\}^4 + \{h^2 + (nh^d)^{-1}\}^3\big], \end{aligned} \tag{4.2}$$

where $b_1(u) = \mu_1^2\mu_2^{-1}H_1(u)$ and the definitions for $b_2(u)$ and $b_3(u)$ are given in (A.3); they are all functions of μ_j,μ_{ij}, $V_j(x)$ and $H_j(z_{\frac{1+\alpha}{2}})$.

Hiding the argument of b_j, the optimal h that minimizes the leading coverage error term is

$$h^\star_{nor} = \Big\{\frac{-b_2 + \sqrt{b_2^2 + d(d+4)b_1b_3}}{(d+4)b_1}\Big\}^{\frac{1}{d+2}} n^{-\frac{1}{d+2}},$$

when $b_2^2 + d(d+4)b_1b_4 > 0$. Choosing $h = O(n^{-\frac{1}{d+2}})$ as prescribed above,

$$P\{m(x) \in I_{\alpha,\text{nor}}\} = \alpha - O(n^{-\frac{2}{d+2}})$$

with a coverage error of $O(n^{-\frac{2}{d+2}})$.

4.2 Coverage Accuracy of $I_{\alpha,\text{el}}$

The derivation deferred until Appendix 2 shows that the coverage probability of $I_{\alpha,\text{el}}$ admits the following Edgeworth expansion:

$$\begin{aligned} P\{m(x) \in I_{\alpha,\text{el}}\} =& \alpha - \Big\{nh^d\mu_1^2\mu_2^{-1} + \Big(\frac{1}{2}\mu_2^{-2}\mu_4 - \frac{1}{3}\mu_2^{-3}\mu_3^2\Big)(nh^d)^{-1}\Big\}z_{\frac{1+\alpha}{2}}\phi(z_{\frac{1+\alpha}{2}})\\ &+ O\big\{nh^{d+6} + h^4 + (nh^d)^{-1}h^2 + (nh^d)^{-2}\big\}. \end{aligned} \tag{4.3}$$

Comparing (4.3) with (4.2) we see that both $I_{\alpha,\text{nor}}$ and $I_{\alpha,\text{el}}$ have the same term $nh^d\mu_1^2\mu_2^{-1}$ in their coverage errors. However, the leading coverage error term of $I_{\alpha,\text{nor}}$ has an extra term b_2h^2, and its coefficient of the $(nh^d)^{-1}$ term is more complicated than that of $I_{\alpha,\text{el}}$.

Using the expressions for μ_j given in (3.4) and (4.1), the dominant coverage error term becomes

$$\big\{\eta_1 nh^{d+4} + \eta_2(nh^d)^{-1}\big\}z_{\frac{1+\alpha}{2}}\phi(z_{\frac{1+\alpha}{2}}), \tag{4.4}$$

where

$$\eta_1 = \frac{1}{4}\frac{B^2(f,m,K,x)}{V(x)f(x)R_2(K)}, \qquad \eta_2 = \frac{3}{2}\frac{V_4(x)R_4(K)}{f(x)R_2^2(K)V^2(x)} - \frac{1}{3}\frac{V_3^2(x)R_3^2(K)}{f(x)V^3(x)R_2^3(K)}.$$

The optimal h that minimizes (4.4) is

$$h^\star_{el} = \big[d\eta_2(x)/\{(d+4)\eta_1(x)\}\big]^{1/(2d+4)} n^{-1/(d+2)}, \tag{4.5}$$

which is of the same order as $h^\star_{nor}$ given earlier. In practice, the plug-in estimates for $h^\star_{nor}$ and $h^\star_{\alpha,el}$ can be obtained by estimating the unknown quantities involved.

Choosing $h = O(n^{-\frac{1}{d+2}})$, $P\{m(x) \in I_{\alpha,\mathrm{el}}\} = \alpha - O(n^{-\frac{2}{d+2}})$. So, the coverage error is of the order $n^{-2/(d+2)}$ and is of the same order of magnitude as $I_{\alpha,\mathrm{nor}}$.

5 The Bartlett Correction

The results in the last section show that the optimal coverage errors of both $I_{\alpha,\mathrm{nor}}$ and $I_{\alpha,\mathrm{el}}$ are of the same order of $n^{-2/(d+2)}$. What we are going to show in this section is that the coverage error of the empirical likelihood confidence interval can be reduced by Bartlett correction.

The Bartlett correction is a novel and elegant property of classical parametric likelihood. A simple adjustment in the mean of the likelihood ratio statistic will improve the coverage accuracy of the likelihood ratio based confidence intervals by one order of magnitude. It has been shown by DiCiccio, Hall and Romano[5], Chen[1,2,3] and Chen, Hall[4] that the empirical likelihood possesses the Bartlett property for a wide range of situations. Thus far the only known case where the empirical likelihood does not admit the property is that found by Jing and Wood[13] by restricting the distributions within the exponential family.

We will show that in the current situation of random design regression the empirical likelihood admits the Bartlett property. It may be shown that

$$E\{\ell(\theta_0)\} = 1 + (nh^d)^{-1}\beta + o\{nh^{d+4} + (nh^d)^{-1}\},$$

where θ_0 is the true value of the parameter and

$$\beta = \mu_2^{-1}(nh^d\mu_1)^2 + \frac{1}{2}\mu_2^{-2}\mu_4 - \frac{1}{3}\mu_2^{-3}\mu_3^2. \tag{5.1}$$

Notice that β appears in the leading coverage error term in (4.3). On the basis of (4.3) and choosing $h = O(n^{-\frac{1}{d+2}})$, we have

$$\begin{aligned} & P\big[\ell\{m(x)\} \le c_\alpha\{1+\beta(nh^d)^{-1}\}\big] \\ =& P\big[\chi_1^2 \le c_\alpha\{1+\beta(nh^d)^{-1}\}\big] - (nh^d)^{-1}\beta c_\alpha^{1/2}\{1+\beta(nh^d)^{-1}\}^{1/2} \\ & \cdot \phi\big[c_\alpha^{-1/2}\{1+\beta(nh^d)^{-1}\}^{1/2}\big] + O\{(nh^d)^{-2}\} \\ =& P(\chi_1^2 \le c_\alpha) + (nh^d)^{-1}\beta z_{\frac{1+\alpha}{2}}\phi(z_{\frac{1+\alpha}{2}}) - (nh^d)^{-1}\beta z_{\frac{1+\alpha}{2}}\phi(z_{\frac{1+\alpha}{2}}) + O\{(nh^d)^{-2}\} \\ =& \alpha + O(n^{-\frac{4}{d+2}}). \end{aligned} \tag{5.2}$$

Therefore, the empirical likelihood is Bartlett correctable in the current case of nonparametric regression.

Let $I_{\alpha,\mathrm{bcel}} = \big[\theta \,|\, \ell(\theta) \le c_\alpha\{1+\beta(nh^d)^{-1}\}\big]$ be the Bartlett corrected empirical likelihood confidence interval. From (5.2), we see that $I_{\alpha,\mathrm{bcel}}$ has coverage errors of $n^{-4/(d+2)}$ if h is chosen to be $O(n^{-\frac{1}{d+2}})$. In practice, the Bartlett factor β has to be estimated which in turn requires the estimation of μ_j for $j=1,2,3$ and 4. Estimators for μ_j, $j\geq 2$, can be defined as

$$\widehat{\mu}_j=(nh^d)^{-1}\sum K_h(x-X_i)^j\{Y_i-\widehat{m}(x)\}^j.$$

An estimator for μ_1 has to involve the estimation of derivatives of f and m as required in the definition of $B(f,m,K,x)$ given in (3.4). To avoid estimating the derivatives, we define $\beta_0=\frac{1}{2}\mu_2^{-2}\mu_4-\frac{1}{3}\mu_2^{-3}\mu_3^2$ which is only part of β given in (5.1). A partial Bartlett correction confidence interval is

$$I_{\alpha,\mathrm{pbcel}}=\big[\theta\,|\,\ell(\theta)\leq c_\alpha\{1+\beta_0(nh^d)^{-1}\}\big].$$

It may be shown that

$$P\big[\ell\{m(x)\}\leq c_\alpha\{1+\beta_0(nh^d)^{-1}\}\big]=\alpha-nh^d\mu_1^2\mu_2^{-1}z_{\frac{1+\alpha}{2}}\phi(z_{\frac{1+\alpha}{2}})+O\{(nh^d)^{-2}\}.$$

That is, the partial Bartlett correction only removes part of the leading coverage error term, and still has the term involving the bias μ_1. However, the term can be of a smaller order by further reducing h to be of a smaller order than $n^{-\frac{1}{d+2}}$, see [3] for an implementation in the density estimation.

Appendix. Derivations

A.1. Derivation of (4.2)

Let $Q=(nh^d)^{-1}\sum K_h(x-X_i)\{Y_i-m(x)\}^2-f(x)V(x)$. A Taylor expansion for T_n is

$$\begin{aligned}T_n=&\frac{\widehat{m}(x)-m(x)}{\sqrt{R_2(K)\widehat{V}(x)/\{nh^d\widehat{f}(x)\}}}\\=&\mu_2^{-1/2}\sqrt{nh^d}\Big\{\overline{\omega}_1-\frac{Q\overline{\omega}_1}{2f(x)V(x)}+\frac{\overline{\omega}_1^3}{2f^2(x)V(x)}+\frac{3}{8}\frac{Q^2\overline{\omega}_1}{f^2(x)V^2(x)}\Big\}+O_p\big[\big\{(nh^d)^{-1/2}+h^2\big\}^3\big]\\\widehat{=}&\mu_2^{-1/2}\sqrt{nh^d}g(\overline{\omega}_1,Q)+O_p\big[\big\{(nh^d)^{-1/2}+h^2\big\}^3\big].\end{aligned}$$

Let $\mu=(\mu_1,\mu_Q)$ and $k_1,k_2,\cdots$ be the cumulates of $\mu_2^{-1/2}\sqrt{nh^d}g(\overline{\omega}_1,Q)$. Using the formulae for the cumulates given in [12] after deriving the multivariate cumulates of $(\overline{\omega}_1,Q)$ we obtain that

$$\begin{aligned}k_1=&\mu_2^{-1/2}\big\{\sqrt{nh^d}\mu_1-a_1(nh^d)^{-1/2}\big\}\\&+O\big\{(nh^d)^{1/2}h^4+(nh^d)^{1/2}n^{-1}+(nh^d)^{-1/2}h^2+(nh^d)^{-3/2}\big\},\\k_2=&1-a_2h^2-a_3\mu_1-\mu_2^{-1}a_4(nh^d)^{-1}+O\big\{h^4+n^{-1}+(nh^d)^{-2}\big\},\\k_3=&\mu_2^{-3/2}a_5(nh^d)^{-1/2}+O\big\{(nh^d)^{-1/2}h^2+(nh^d)^{-3/2}\big\},\\k_4=&\mu_2^{-2}a_6(nh^d)^{-1}+O\big\{(nh^d)^{-1}h^2+n^{-1}+(nh^d)^{-2}\big\},\\k_l=&O\big\{(nh^d)^{-(l-2)/2}\big\}\qquad\text{for }\ l\geq 5,\end{aligned}$$

where

$$\begin{aligned}
a_1 &= \frac{1}{2}\mu_2^{-1}\mu_{23}R_2(K), \qquad a_2 = \frac{\gamma(f,V,m)}{2f(x)V(x)}\Big\{1-\frac{1}{R_2(K)}\Big\},\\
a_3 &= \mu_2^{-2}\mu_{23}R_2(K),\\
a_4 &= \mu_2^{-1}\mu_{34}R_2(K) - 3V(x)R_2^2(K) - \frac{7}{4}\mu_{23}^2\mu_2^{-2}R_2^2(K) - \mu_{24}\mu_2^{-1}R_2^2(K),\\
a_5 &= \mu_3 - 3\mu_{23}R_2(K),\\
a_6 &= \mu_4 - 6\mu_{34}R_2(K) - 6\mu_3\mu_{23}\mu_2^{-1}R_2(K) + 12\mu_2V(x)R_2^2(K)\\
&\quad + 18\mu_{23}^2\mu_2^{-1}R_2^2(K) + 3\mu_{24}R_2^2(K).
\end{aligned}$$

In the above definition for a_2,

$$\gamma(f,V,m) = \sum_{i=1}^{d}\sigma_i^2(K)\big[V_i''(x)f(x) + 2V_i'(x)f_i'(x) + V(x)f_i''(x)\big].$$

The generating function for T_{n1} is

$$\begin{aligned}
M(T_{n1},t) = {}&\exp(t^2/2)\Big[1 + k_1t + \frac{1}{2}(k_2-1)t^2 + \frac{1}{6}k_3t^3 + \frac{1}{24}k_4t^4\\
&+ \frac{1}{2}\Big\{k_1^2t^2 + \frac{1}{4}(k_2-1)^2t^4 + \frac{1}{36}k_3^2t^6 + \frac{1}{3}k_1k_3t^4\Big\}\Big]\\
&+ O\big[\{(nh^d)^{1/2}h^2 + (nh^d)^{-1/2}\}^3 + \{h^2 + (nh^d)^{-1}\}^3\big].
\end{aligned}$$

Converting the above generating functions, we have the following Edgeworth expansion:

$$\begin{aligned}
P(T_{n1}\le u) = \Phi(u) - \phi(u)\Big[&\mu_2^{-1/2}(nh^d)^{1/2}\mu_1 + \frac{1}{2}\mu_2^{-1}nh^d\mu_1^2H_1(u)\\
&- \frac{1}{2}a_2h^2H_1(u) + \Big\{\frac{1}{6}\mu_2^{-2}a_5H_3(u) - \Big(\mu_2^{-1}a_1 + \frac{1}{2}a_3\Big)H_1(u)\Big\}\mu_1\\
&+ \Big\{\frac{1}{6}\mu_2^{-3/2}a_5H_2(u) - \mu_2^{-1/2}a_1\Big\}(nh^d)^{-1/2} + \Big\{\frac{1}{2}\mu_2^{-1}(a_1^2-a_4)H_1(u)\\
&+ \mu_2^{-2}\Big(\frac{1}{24}a_6 - \frac{1}{6}a_1a_5\Big)H_3(u) + \frac{1}{72}\mu_2^{-3}a_5^2H_5(x)\Big\}(nh^d)^{-1}\Big]\\
&+ O\big[\{(nh^d)^{1/2}h^2 + (nh^d)^{-1/2}\}^3 + \{h^2 + (nh^d)^{-1}\}^3\big]. \qquad (A.1)
\end{aligned}$$

From (A.1), we have the following Edgeworth expansion for $|T_{n1}|$ for $u>0$:

$$\begin{aligned}
P(|T_{n1}|\le u) = 2\Phi(u) - 1 - \phi(u)\big[&nh^d\mu_1^2\mu_2^{-1}H_1(u) - \xi_1H_1(u)h^2\\
&+ \xi_2H_3(u)h^2 + \{\xi_3H_1(u) + \xi_4H_3(u) + \xi_5H_5(u)\}(nh^d)^{-1}\big]\\
&+ O\big[\{(nh^d)^{1/2}h^2 + (nh^d)^{-1/2}\}^4 + \{h^2 + (nh^d)^{-1}\}^3\big], \qquad (A.2)
\end{aligned}$$

where

$$\begin{aligned}
\xi_1 &= \gamma(f,V,m)\{2f(x)V(x)\}^{-1}\{1 - 1/R_2(K)\} + \mu_2^{-2}\mu_{23}B(f,m,K,x),\\
\xi_2 &= \frac{1}{6}B(f,m,K,x)\{\mu_3 - 3\mu_{23}R_2(K)\}\mu_2^{-2},\\
\xi_3 &= 2\mu_{23}^2R_2(K)\mu_2^{-3} - \mu_{34}R_2(K)\mu_2^{-2} + 3R_2(K)/f(x) + \mu_{24}R_2^2(K)\mu_2^{-2},\\
\xi_4 &= \frac{1}{12}\mu_2^{-2}\big\{\mu_4 - 6\mu_{34}R_2(K) - 8\mu_3\mu_{23}R_2(K)\mu_2^{-1} + 12R_2(K)\mu_2^2/f(x)\\
&\quad + 24\mu_{23}^2R_2^2(K) + 3\mu_{24}^2R_2^2(K)\big\},\\
\xi_5 &= \frac{1}{36}\mu_2^{-3}\{\mu_3 - 3\mu_{23}R_2(K)\}^2.
\end{aligned}$$

Using the delta method described in [9], we know that T_n has the same Edgeworth expansion as T_{n1}. Replacing u with $z_{\frac{1+\alpha}{2}}$ in (A.2), we have the coverage accuracy of $I_{\alpha,\text{nor}}$ is

$$P\{m(x)\in I_{\alpha,\text{nor}}\}=\alpha-\big\{b_1(z_{\frac{1+\alpha}{2}})nh^{d+4}+b_2(z_{\frac{1+\alpha}{2}})h^2+b_3(z_{\frac{1+\alpha}{2}})(nh^d)^{-1}\big\}\phi(z_{\frac{1+\alpha}{2}})$$
$$+O\big[\big\{(nh^d)^{1/2}h^2+(nh^d)^{-1/2}\big\}^4+\{h^2+(nh^d)^{-1}\}^3\big],$$

where b_j are functions of $\mu_j,\ \mu_{ij},\ V_j(x)$ and $H_j(z_{\frac{1+\alpha}{2}})$ and are defined as

$$\begin{aligned}&b_1(u)=\mu_1^2\mu_2^{-1}H_1(u),\qquad b_2(u)=-\xi_1H_1(u)+\xi_2H_3(u),\\&b_3(u)=\xi_3H_1(u)+\xi_4H_3(u)+\xi_5H_5(u).\end{aligned}\tag{A.3}$$

Thus, we have obtained (4.2).

A.2. Derivation of (4.3)

Note that $\lambda(\theta)=O_p\{(nh^d)^{-1/2}+h^2\}$. By (2.4), for each integer $j\geq 1$,

$$\begin{aligned}0&=(nh^d)^{-1}\sum\omega_i\big\{1-\lambda(\theta)\omega_i+\big(\lambda(\theta)\omega_i\big)^2-\big(\lambda(\theta)\omega_i\big)^3+\cdots\big\}\\&=\sum_{k=1}^{j}\big(-\lambda(\theta)\big)^{k-1}\overline{\omega}_k+\big\{\big((nh^d)^{-1/2}+h^2\big)^j\big\}.\end{aligned}$$

Similarly to (A1.3) in [4], we obtain the following Taylor expansion of $\ell(\theta)$:

$$\begin{aligned}\ell(\theta)=&(nh^d)\Big\{\overline{\omega}_2^{-1}\overline{\omega}_1^2+\frac{2}{3}\overline{\omega}_2^{-3}\overline{\omega}_3\overline{\omega}_1^3+\Big(\overline{\omega}_2^{-5}\overline{\omega}_3^2-\frac{1}{2}\overline{\omega}_2^{-4}\overline{\omega}_4\Big)\overline{\omega}_1^4\\&+\Big(8\overline{\omega}_2^{-6}\overline{\omega}_3\overline{\omega}_4-8\overline{\omega}_2^{-7}\overline{\omega}_3^3-\frac{8}{5}\overline{\omega}_2^{-5}\overline{\omega}_5\Big)\overline{\omega}_1^5\Big\}\\&+nh^d\sum_{k=5}^{j}R_{1k}\overline{\omega}_1^{k+1}+O_p\big[nh^d\{(nh^d)^{-1/2}+h^2\}^{j+2}\big],\end{aligned}$$

where R_{1k} denotes $\overline{\omega}_2^{-(2k-1)}$ multiplied by a polynomial in $\overline{\omega}_2,\cdots,\overline{\omega}_{k+1}$, with constant coefficients.

As in [4], we may write $\ell(\theta)=\{(nh^d)^{1/2}S_j'\}^2$, where

$$\begin{aligned}S_j'&=\overline{\omega}_2^{-1/2}\Big\{\overline{\omega}_1+\frac{1}{3}\overline{\omega}_2^{-2}\overline{\omega}_3\overline{\omega}_1^2+\Big(\frac{4}{9}\overline{\omega}_2^{-4}\overline{\omega}_3^2-\frac{1}{4}\overline{\omega}_2^{-3}\overline{\omega}_4\Big)\overline{\omega}_1^3\\&\quad+\Big(-\frac{112}{27}\overline{\omega}_2^{-6}\overline{\omega}_3^3+\frac{49}{12}\overline{\omega}_2^{-5}\overline{\omega}_3\overline{\omega}_4-\frac{4}{5}\overline{\omega}_2^{-4}\overline{\omega}_5\Big)\overline{\omega}_1^4+\sum_{k=5}^{j}T_k\overline{\omega}_1^k\Big\}+U_j\\&=S_j+U_j,\end{aligned}$$

where $U_j=O_p[\{(nh^d)^{-1/2}+h^2\}^{j+1}]$, and T_k denotes $\overline{\omega}_2^{-2(k-1)}$ multiplied by a polynomial in $\overline{\omega}_2,\cdots,\overline{\omega}_k$ with constant coefficients.

Observe that S_j is a function of $\overline{\omega}_1,\cdots,\overline{\omega}_j$. Denote that function by s_j. Put $\mu_k=E(\overline{\omega}_k),\ \mu=(\mu_1,\cdots,\mu_j)^T,\ u=(u_1,\cdots,u_j)^T,\ V_k=\overline{\omega}_k-\mu_k,\ V=(V_1,\cdots,V_j)^T$,

$$d_{k_1,\cdots,k_m}=\Big(\prod_{l=1}^{m}\frac{\partial}{\partial u_{k_l}}\Big)s_j(u_1,\cdots,u_j)|_{u=\mu},$$

$$p(u)=s_j(\mu)+\sum_{m=1}^{6}(m!)^{-1}\sum_{k_1,\cdots,k_m\in\{1,\cdots,j\}}d_{k_1,\cdots,k_m}u_{k_1}\cdots u_{k_m}.$$

Let $k_1, k_2, \cdots$ be the cumulants of $(nh^d)^{1/2}p(V)$. After calculating the the multivariate cumulants of $V = (V_1, \cdots, V_j)$, we have

$$k_1 = n^{1/2}s_j(\mu)h^{d/2} - \frac{1}{6}\mu_2^{-3/2}\mu_3 h^{-d/2}n^{-1/2} + O\big\{(nh^d)^{-1/2}h^2 + (nh^d)^{-3/2}\big\},$$
$$k_2 = \sigma^2 + \Big(\frac{1}{2}\mu_2^{-2}\mu_4 - \frac{13}{36}\mu_2^{-3}\mu_3^2\Big)(nh^d)^{-1} + O\big\{(nh^d)^{-1}h^2 + (nh^d)^{-2}\big\},$$
$$k_3 = O\big\{(nh^d)^{-1/2}h^2 + (nh^d)^{-5/2}\big\}, \qquad k_4 = O\big\{(nh^d)^{-1}h^2 + (nh^d)^{-2}\big\},$$
$$k_l = O\big\{(nh^d)^{-(l-2)/2}\big\} \qquad \text{for} \quad l \geq 5,$$

where

$$\sigma^2 = 1 + \frac{1}{3}\mu_2^{-2}\mu_3\mu_1 + O(h^4).$$

Thus we could develop a formal Edgeworth expansion for the distribution of $(nh^d)^{1/2}p(V)$:

$$\begin{aligned} &P\{n^{1/2}h^{d/2}p(V) \leq t\} \\ =&\Phi(t) - \frac{1}{12}(nh^d)^{-1}\big\{6\mu_2^{-1}(nh^d\mu_1)^2 + 3\mu_2^{-2}\mu_4 - 2\mu_2^{-3}\mu_3^2\big\}t\phi(t) \\ &+ (\text{even polynomial in } t)\phi(t) + O\big\{nh^{d+6} + h^4 + (nh^d)^{-1}h^2 + (nh^d)^{-2}\big\}, \end{aligned}$$

which, in turn as in [4], gives the Edgeworth expansion for $\ell(\theta)$ in (4.3).

References

[1] Chen, S.X. On the coverage accuracy of empirical likelihood confidence regions for linear regression model. *Ann. Inst. Statist. Math.*, 45: 621–637 (1993)

[2] Chen, S.X. Empirical likelihood confidence intervals for linear regression coefficients. *J. Mult. Anal.*, 49: 24–40 (1994)

[3] Chen, S.X. Empirical likelihood for nonparametric density estimation. *Biometrika*, 83: 329–341 (1996)

[4] Chen, S.X., Hall, P. Smoothed empirical likelihood confidence intervals for quantiles. *Ann. Statist.*, 21: 1166–1181 (1993)

[5] DiCiccio, T.J., Hall, P., Romano, J.P. Bartlett adjustment for empirical likelihood. *Ann. Statist.*, 19: 1053–1061 (1991)

[6] Fukunaga, K. Introduction to Statistical Pattern Recognition. New York: Academic Press, 1972

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

[8] Hall, P. On bootstrap confidence intervals in nonparametric regression. *Ann. Statist.*, 20: 695–711 (1992)

[9] Hall, P. The bootstrap and Edgeworth Expansions. New York: Springer-Verlag, 1992

[10] Hall, P., La Scala, B. Methodology and algorithms of empirical likelihood. *Internat. Statist. Rev.*, 58: 109–127 (1990)

[11] Härdle, W. Applied nonparametric regression. Cambridge: Cambridge University Press, 1990

[12] James, G.S., Mayne, A.J. Cumulants of functions of random variables. *Sankhya* (Ser. A) 24: 47–54 (1962)

[13] Jing, B.Y., Wood, A.T.A. Exponential empirical likelihood is not Bartlett correctable. *Ann. Statist.*, 24: 365–369 (1996)

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

[15] Owen, A. Empirical likelihood ratio confidence regions. *Ann. Statist.*, 18: 90–120 (1990)

[16] Press, W.H., Flannery, B.F., Teukolsky, S.A., Vetterling, W.T. Numerical recipes in C: the art of scientific computing. Cambridge: Cambridge University Press, 1992

[17] Schuster, E.F. Joint asymptotic distribution of the estimated regression function at a finite number of distinct points. *Ann. Math. Statist.*, 43: 84–88 (1972)