 Vol 29: 89–99, 2002

Confidence Intervals Based on Local Linear Smoother

SONG XI CHEN

National University of Singapore

YONG SONG QIN

University of Science and Technology of China

ABSTRACT. Point-wise confidence intervals for a non-parametric regression function in conjunction with the popular local linear smoother are considered. The confidence intervals are based on the asymptotic normal distribution of the local linear smoother. Their coverage accuracy is evaluated by developing Edgeworth expansion for the coverage probability. It is found that the coverage error near the boundary of the support of the regression function is of a larger order than that in the interior, which implies that the local linear smoother is not adaptive to the boundary in terms of coverage. This is quite unexpected as the local linear smoother is adaptive to the boundary in terms of the mean squared error.

Key words: confidence interval, coverage probability, Edgeworth expansion, non-parametric regression, normal approximation

1. Introduction

Local polynomial smoothing (Stone, 1977; Cleveland, 1979; Cleveland & Devlin, 1988) has become a popular tool for smoothing curves. It has been shown by Fan (1993), Fan & Gijbels (1992), Hastie & Loader (1993), Ruppert & Wand (1994) and others that local polynomial smoothing is an effective smoothing method in non-parametric regression. It has the advantages of achieving full asymptotic minimax efficiency and automatically correcting for boundary bias. An excellent review of local polynomial smoothing is given in Fan & Gijbels (1996). While local polynomial smoothing is being used in a wide range of applications, the important issues of constructing confidence intervals and the coverage evaluation have not received much attention.

In this paper we consider confidence intervals for an unknown regression function with a bounded support at a fixed point based on the asymptotic normal distribution of the local linear smoother. The coverage accuracy is studied by developing Edgeworth expansions for the coverage probability. It is known in constructing confidence intervals associated with non-parametric curve estimation that the bias associated with the curve estimation has to be corrected in order to make the confidence intervals into ones for the curve function. There are two methods to correct for the bias. One is to estimate the bias and then carry out an explicit bias correction (Härdle, 1990; Fan & Gijbels, 1996; Rodríguez-Campos, 1999). The other is to reduce the bias via undersmoothing (Hall, 1992a; Neumann, 1995; Chen, 1996). Hall (1992a) in the context of the Nadaraya–Watson smoother and Neumann (1997) in the context of the Gasser–Müller smoother showed that better coverage accuracy is achieved by the undersmoothing method. We consider the latter approach in this paper.

A property of the local linear smoother is its adaptivity to the boundary in terms of the mean square error as both the bias and variance are of the same order throughout the support of the regression function. One would expect the same to be true for the coverage probability of confidence intervals based on the local linear estimators; that is, the coverage error is of the same order throughout the curve support. However, our study shows that the coverage error near the boundary is of a larger order of magnitude than that in the interior of the support, which implies that the local linear smoother is not automatically adaptive to the boundary in terms of coverage.

The paper is structured as follows. In section 2 we introduce some notation and the confidence intervals based on the local linear smoother. Their coverage accuracy in the interior and in the boundary are evaluated separately by developing Edgeworth expansions in section 3. The proofs are given in section 4.

2. Confidence intervals

Let $\{(X_i, Y_i)\}_{i=1}^n$ be n independent pairs of observations where Y_i are the responses of design points X_i for $i = 1, \ldots, n$ according to the following regression model:

$$Y_i = m(X_i) + \epsilon_i \tag{1}$$

where m is an unknown regression function with $\mathcal{S}$ being its support, ϵ_i are independent errors with zero mean and a variance function $V(X_i)$, and the design points $X_i \in R$ have a density f.

Let K be a kernel function which is a probability density itself and h be the smoothing bandwidth. Also define $V(x) = E[\{Y - m(x)\}^2 | X = x]$. Without loss of generality we assume $\mathcal{S} = [0, \infty)$ in our theoretical analysis as it can be easily generalized to the cases of $\mathcal{S} = (-\infty, \infty)$ and $\mathcal{S} = [0, 1]$.

The local linear regression estimator for $m(x)$ at any given $x \in \mathcal{S}$ is

$$\hat{m}(x) = \frac{\sum_1^n W_i Y_i}{\sum_1^n W_i} \tag{2}$$

where $W_i = K_h(x - X_i)\{s_{n,2} - (x - X_i)s_{n,1}\}$, $s_{n,l} = \sum_1^n K_h(x - X_i)(x - X_i)^l$ for $l = 0$, 1 and 2, and $K_h(\cdot) = h^{-1}K(\cdot/h)$.

We assume the following regularity conditions:

(i) K is a symmetric density function, bounded and strictly positive on, [−1, 1];
(ii) $h \to 0, nh \to \infty$ as $n \to \infty$;
(iii) f, m and V have continuous second derivatives in a neighbourhood of x; $f(x) > 0$ and $V(x) > 0$

For non-negative integers i, j, k, define

$$\bar{\omega}_{ijk} = n^{-1}h^{j-i-1}\sum_{l=1}^n (x - X_l)^i K_h^j(x - X_l)\{Y_l - m(x)\}^k \quad \text{and} \quad \mu_{ijk} = E\bar{\omega}_{ijk}. \tag{3}$$

It is noted that $\mu_{ij1} = O(h)$, and for $k > 1$

$$\mu_{ijk} = f(x)V_k(x)\int t^i K^j(t)\,dt + o(h).$$

It may be shown after some algebra that

$$\hat{m}(x) - m(x) = \frac{\bar{\omega}_{210}\bar{\omega}_{110} - \bar{\omega}_{110}\bar{\omega}_{111}}{\bar{\omega}_{210}\bar{\omega}_{010} - \bar{\omega}_{110}^2} =: \frac{\overline{U}_1}{\overline{U}_{21}}. \tag{4}$$

Let $\alpha_l(x/h) = \int_{-1}^{x/h} u^l K(u)\,du$ for $l = 0$, 1, 2 and 3. Define

$$b(x/h) = \frac{\alpha_2^2(x/h) - \alpha_1(x/h)\alpha_3(x/h)}{\alpha_0(x/h)\alpha_2(x/h) - \alpha_1^2(x/h)}$$

and

$$v(x/h) = \frac{\int_{-1}^{x/h}\{\alpha_2(x/h) - u\alpha_1(x/h)\}^2 K^2(u)\,du}{\{\alpha_0(x/h)\alpha_2(x/h) - \alpha_1^2(x/h)\}^2}$$

which are respectively the coefficients of the asymptotic bias and variance of $\hat{m}(x)$. Put $\sigma_k^2 = \int u^2 K(u)\,du$ and $R(K) = \int K^2(u)\,du$. Then, clearly $b(x/h) = \sigma_k^2$ and $v(x/h) = R(K)$ for $x \geqslant h$ as $\alpha_0(x/h) = 1$, $\alpha_1(x/h) = \alpha_3(x/h) = 0$ and $\alpha_2(x/h) = \sigma_k^2$.

In may be shown, by noting the bias and the variance properties of $\hat{m}(x)$, that under the conditions in (3), if $m''(x) \neq 0$ then

$$T_n^{(0)} = \sqrt{\frac{nh}{v(x/h)}}\frac{\{\hat{m}(x) - m(x)\}}{\sqrt{V(x)/f(x)}} \to N(0, 1) \text{ in distribution} \tag{5}$$

if and only if $nh^5 \to 0$.

This implies that undersmoothing is required in order to make the effect of the bias negligible. Let $\hat{V}(x)$ and $\hat{f}(x)$ be some consistent estimators of $V(x)$ and $f(x)$, and z_α be the $(1+\alpha)/2$-quantile of $N(0, 1)$. Then, a normal approximation based confidence interval for $m(x)$, with nominal coverage α, is

$$I_\alpha = \hat{m}(x) \pm z_\alpha\sqrt{v(x/h)\hat{V}(x)/\{nh\hat{f}(x)\}}. \tag{6}$$

To ensure non-negativity under the square root, $\hat{V}(x)$ and $\hat{f}(x)$ should be non-negative. Let $\hat{f}_0(x) = n^{-1}\sum K_h(x - X_i)$ be the standard kernel estimator for $f(x)$. It is known that $\hat{f}_0(x)$ is not a consistent estimator of $f(x)$ if $x/h < 1$, but $\hat{f}(x) = \hat{f}_0(x)/\alpha_0(x/h)$, the re-normalizing density estimator, is. An estimator for $V(x)$ is

$$\hat{V}(x) = n^{-1}\sum_1^n K_h(x - X_i)\{Y_i - \hat{m}(x)\}^2/\hat{f}_0(x). \tag{7}$$

Then, (5) implies that the above confidence interval has correct asymptotic coverage as long as $h = o(n^{-1/5})$, which is of a smaller order than the usual bandwidth prescribed for standard curve estimation.

One may wonder why we do not use existing boundary kernels (Müller & Wang, 1994) or local linear kernels (Jones, 1993) instead of the re-normalizing kernel in estimating $f(x)$ near the boundary. The boundary/local linear kernel density estimators have a bias of $O(h^2)$ everywhere in $[0, \infty)$ and the re-normalizing kernel estimator has a bias of $O(h)$ near the boundary. The reasons are (a) the boundary/local linear kernels may produce negative estimates for $V(x)/f(x)$ which are not suitable for constructing the confidence interval; (b) using the boundary/local linear kernels does not lead to improvement of coverage error near the boundary area, as shown in the next section.

Let

$$T_n = \sqrt{\frac{nh}{v(x/h)}}\frac{\hat{m}(x) - m(x)}{\sqrt{\hat{V}(x)/\hat{f}(x)}}.$$

As the confidence interval I_α is based on T_n, we are going to develop Edgeworth expansions for the distribution of T_n to assess the coverage accuracy of I_α.

3. Coverage accuracy

In studying the coverage accuracy of above confidence interval, we assume three extra conditions

(iv) $h = o\{n^{1/5}\}$,
(v) $nh(\log n)^{-1} \to \infty$ and $E|Y|^{15} > \infty$, and (8)
(vi) $K(u)$ is continuous differentiable, and $\{1, K(u), uK(u), u^2K(u)\}$ are linearly independent in $C[-1, 1]$

where $C[-1, 1]$ is the space of all continuous functions in $[-1, 1]$. The condition (iv) implies undersmoothing as required in (5), and (v) and (vi) are needed in developing Edgeworth expansions for the coverage probability. Condition (vi) basically rules out the uniform kernel, in which case special treatment has to be carried out to establish the Edgeworth expansions as discussed in Hall (1991).

Let H_j be the Hermite polynomials of the jth order and ϕ be the density function of the standard normal distribution. We study the coverage probability of I_α separately for x in the interior and x in the boundary in the following theorems 1 and 2, whose proofs are deferred until the appendix.

Theorem 1
Under the conditions in (3) *and* (8), *if* $x/h \geqslant 1$ *then*

$$P\{m(x) \in I_\alpha\} = \alpha - \{b_1 nh^5 + b_2 h^2 + b_3 (nh)^{-1}\} + O\{(nh)^{1/2}h^4 + (nh)^{-1/2}h^2 + (nh)^{-3/2}\}$$

where $b_1 = \frac{1}{4}H_1(z_\alpha)\phi(z_\alpha)\{m''(x)\}^2 f(x)\sigma_k^4/\{R(K)V(x)\}$, *and* b_2 *and* b_3 *whose exact expressions are given in* (16) *and* (17), *are free of n and h.*

From the derivation in the appendix, we see that $b_1 nh^5$ is due to the bias of $\hat{m}(x)$ and $b_2 h^2$ is due to both the bias and the variance of $\hat{m}(x)$. While these two are unique terms only for non-parametric curve estimators, the $b_3(nh)^{-1}$ term is a variation of a standard term in Edgeworth expansions where n is replaced by nh, the effective sample size in the current curve estimation case.

Let us now consider the coverage accuracy of I_α for x near the boundary. Define for non-negative integers j, $\beta_j(x/h) = \int_{-\infty}^{x/h} u^j K^2(u)\,du$, and

$$\gamma(x/h) = \alpha_0(x/h)\alpha_2(x/h) - \alpha_1^2(x/h) \quad \text{and}$$
$$\rho(x/h) = f^3(x)V(x)\{\alpha_2^2(x/h)\beta_0(x/h) - 2\alpha_1(x/h)\alpha_2(x/h)\beta_1(x/h) + \alpha_1^2(x/h)\beta_2(x/h)\}. \quad (9)$$

To simplify the notation, we drop out the argument x/h in β_j, α_j and γ in the following theorem.

Theorem 2
Under the conditions in (3) *and* (8), *and if* $x/h < 1$, *then*

$$P\{m(x) \in I_\alpha\} = \alpha - \{d_1 nh^5 + d_2 h + d_3 (nh)^{-1}\} + O\{(nh)^{1/2}h^4 + (nh)^{-1/2}h^2 + (nh)^{-3/2}\},$$

where $d_1 = \frac{1}{4}\rho^{-1}(x/h) z_\alpha \phi(z_\alpha)\{\alpha_2^2(x/h) - \alpha_1(x/h)\alpha_3(x/h)\} f^2(x) m''(x)$ *and*

$$d_2 = \rho^{-1}(x/h)z_\alpha\phi(z_\alpha)[f^2(x)f'(x)V(x)(\alpha_2^2\beta_1 + 2\alpha_1\alpha_3\beta_1 - 2\alpha_2\alpha_3\beta_0 - \alpha_1^2\beta_3) + \{2\alpha_0\gamma f(x)V(x)\}^{-1}\{f'(x)V(x)(3\alpha_0\alpha_1\alpha_2 - \alpha_1^3 - 2\alpha_0^2\alpha_3) - f(x)V'(x)\alpha_1\gamma\} + f^3(x)V'(x)(-\alpha_2^2\beta_1 + 2\alpha_1\alpha_2\beta_2 - \alpha_1^2\beta_3)]. \quad (10)$$

The exact expression for d_3 is not given as it is tedious to write out. This is because many intermediate terms, which would be cancelled out for interior x, can not be cancelled out near the boundary. Notice that d_j depend on h via x/h. However, this does not have any effect on the order of magnitude of the coverage terms as we can treat x as a sequence $x_n = ch_n$ for a constant $0 < c < 1$. We deliberately leave x/h in its current form to preserve some original flavour.

Theorem 2 states that the coverage error at a boundary x is of a larger order than that at an interior x. This is surprising as the local linear smoother has the appealing feature of automatic correction to the boundary, shown by having the same order of bias and variance throughout the curve support.

People may suspect that this increased order near the boundary is due to the particular estimators $\hat{f}(x)$ and $\hat{V}(x)$ used. To show this is not the case, we replace $\hat{f}(x)$ and $\hat{V}(x)$ by $f(x)$ and $V(x)$ in the definition of I_α, and call the resulting interval $I_{\alpha 0}$. This interval is not practical and only serves a theoretical purpose here. Derivations similar to those in theorems 1 and 2 show that the Edgeworth expansions for the coverage probability of $I_{\alpha 0}$ admit the same forms as those of I_α except that the coefficients b_j and d_j are different. In particular, the d_2 that appeared in theorem 2, which is the coefficient of the h term given in (24), is replaced by

$$d_2' = \rho^{-1}(x/h)z_\alpha\phi(z_\alpha)[f^2(x)f'(x)V(x)(2\alpha_1\alpha_3\beta_1 - 2\alpha_2\alpha_3\beta_0 - \alpha_1^2\beta_3) - \{2\gamma f(x)\}^{-1}f'(x)(\alpha_1\alpha_2 - \alpha_0\alpha_3) + f^3(x)V'(x)(-\alpha_2^2\beta_1 + 2\alpha_1\alpha_2\beta_2 - \alpha_1^2\beta_3)]$$

Therefore, an increased coverage error near the boundary is still the case even when we know the values of $f(x)$ and $V(x)$.

A close look at the derivation given in the appendix reveals that the larger coverage error near the boundary is due to the fact that

$$\text{var}\{\hat{m}_l(x)\} = (nh)^{-1}v(x/h)V(x)/f(x) + O(n^{-1}); \quad (11)$$

that is, the second order term of the variance is of n^{-1} rather than $O(hn^{-1})$ as in the interior. In a related paper (Chen & Qin, 2000), the authors consider using the empirical likelihood (Owen, 1990) to construct confidence intervals for $m(x)$. The empirical likelihood, as shown previously in Chen (1996) in the context of non-parametric density estimation, has the ability to Studentize internally via the optimization algorithm it conducts, and hence is able to capture the second order variance structure better than the asymptotic variance of the local linear smoother. It is shown in that paper that the empirical likelihood based confidence intervals have the same order of coverage error throughout the support of the curve, and thus eliminate the boundary coverage problem associated with the confidence intervals based on the asymptotic normal approximation based intervals.

It may be shown from theorems 1 and 2 that the optimal order for h that minimizes the leading coverage error term is $O(n^{-1/3})$ in the interior and $O(n^{-1/2})$ in the boundary, and the optimal order for leading coverage error term is $O(n^{-2/3})$ in the interior and $O(n^{-1/2})$ in the boundary. It seems that finding a suitable h value for the purpose of constructing confidence intervals is a topic that requires research effort. The authors share a view expressed by a referee that "the bandwidth selection problem for confidence intervals is perhaps more complicated but not more difficult than for ordinary regression—the difficulties lies rather in proving the corresponding theorems for non-fixed bandwidths".

4. Proofs

Proof of theorem 1. Let $U_1 = \bar{\omega}_{210}\bar{\omega}_{011} - \bar{\omega}_{110}\bar{\omega}_{111}$ and $U_{21} = \bar{\omega}_{210}\bar{\omega}_{010} - \bar{\omega}_{110}^2$. From (4),

$$\hat{m}(x) - m(x) = \overline{U}_1/\overline{U}_{21}. \tag{12}$$

Define

$$\overline{U}_2 = \overline{U}_{21} - f^2(x)\int u^2K(u)\,du, \quad \overline{U}_3 = \bar{\omega}_{010} - f(x), \quad \overline{U}_4 = \bar{\omega}_{012} - f(x)V(x).$$

From the definitions of $\hat{f}$ and $\hat{V}$ given in section 2, $\hat{f}(x) = \bar{\omega}_{010}$, $\hat{f}(x)\hat{V}(x) = \bar{\omega}_{012} - 2\{\hat{m}(x) - m(x)\}\bar{\omega}_{011} + \{\hat{m}(x) - m(x)\}^2\bar{\omega}_{010}$. These together with (12) imply that

$$T_n = (nh)^{1/2}\{R(K)f^3(x)V(x)\sigma_k^4\}^{-1/2}\{1 + \overline{U}_3/f(x)\}\overline{U}_1/\sqrt{1+\Delta_1} \tag{13}$$

where

$$\Delta_1 = \{f^5(x)V(x)\sigma_k^4\}^{-1}[\overline{U}_2^2\overline{U}_4 - 2\overline{U}_1\overline{U}_2\bar{\omega}_{011} + \overline{U}_1^2\overline{U}_3 + f(x)\{\overline{U}_1^2 + V(x)\overline{U}_2^2\}$$
$$+ 2f^2(x)\sigma_k^2\{\overline{U}_2\overline{U}_4 - \overline{U}_1\bar{\omega}_{011}\} + 2f^3(x)V(x)\sigma_k^2\overline{U}_2 + f^4(x)\sigma_k^4\overline{U}_4].$$

Let

$$T_{n1} = \sqrt{nh}\{R(k)f^{11}(x)V^3(x)\sigma_k^{12}\}^{-1/2}[f^4(x)V(x)\sigma_k^4\overline{U}_1 - f^2(x)V(x)\sigma_k^2\overline{U}_1\overline{U}_2$$
$$- \tfrac{1}{2}f^3(x)\sigma_k^4\overline{U}_1\{\overline{U}_4 - 2V(x)\overline{U}_3\} + \tfrac{1}{2}f(x)\sigma_k^2\overline{U}_1\overline{U}_2\{\overline{U}_4 - 2V(x)\overline{U}_3\} + V(x)\overline{U}_1\overline{U}_2^2$$
$$+ f(x)\sigma_k^2\overline{U}_1^2\bar{\omega}_{011} - \tfrac{1}{2}\overline{U}_1^3 + f^2(x)\sigma_k^4\overline{U}_1\overline{U}_4\{\tfrac{3}{8}V^{-1}(x)\overline{U}_4 - \tfrac{1}{2}\overline{U}_3\}].$$

By Taylor expansion, it follows that $T_n = T_{n1} + R_1$, where $R_1 = O_p\,[\{(nh)^{-1/2} + h^2\}^3]$ by the fact that $\overline{U}_j = O_p\,[\{(nh)^{-1/2} + h^2\}]$ for 1, 2, 3 and 4.

It may be shown, using the delta method described in Hall (1992b, pp. 76–77), that

$$P(T_n \leqslant u) = P(T_{n1} \leqslant u) + O[\{(nh)^{-1/2} + h^2\}^3]. \tag{14}$$

Let $W = (\bar{\omega}_{010}, \bar{\omega}_{110}, \bar{\omega}_{210}, \bar{\omega}_{011}, \bar{\omega}_{111}, \bar{\omega}_{012})$. As $\overline{U}_j$s are polynomials of W, so T_{n1} is a polynomial of W as well. Let $\mu = \mu_{210}\mu_{011} - \mu_{110}\mu_{111}$, $\rho = R(K)f^3(x)V(x)\sigma_k^4, R_{ij}(K) = \int u^iK^j(u)\,du$ and $V_l(x) = E[\{Y - m(X)\}^l|X = x]$ for positive integers l. In particular, $R(K) = R_{02}(K)$ and $\sigma_k^2 = R_{21}(K)$.

Let $k_1, k_2, \ldots$ be the cumulants of T_{n1}. Using the formulae given in James & Mayne (1962) after deriving the multivariate cumulants of W we have

$$k_1 = \rho^{-1/2}\{\sqrt{nh}\mu + a_1(nh)^{-1/2}\} + O\{(nh)^{1/2}h^4 + (nh)^{-1/2}h^2 + (nh)^{-3/2}\},$$
$$k_2 = 1 + a_2h^2 + \rho^{-1}a_3(nh)^{-1} + O\{h^4 + (nh)^{-1}h^2 + (nh)^{-2}\},$$
$$k_3 = \rho^{-3/2}a_4(nh)^{-1/2} + O\{(nh)^{-1/2}h^2 + (nh)^{-3/2}\},$$
$$k_4 = \rho^{-2}a_5(nh)^{-1} + O\{(nh)^{-1}h^2 + (nh)^{-2}\},$$
$$k_l = O\{(nh^d)^{-(l-2)/2}\} \quad \text{for } l \geqslant 5,$$

where

$$a_1 = \tfrac{1}{2}\mu_{210}\mu_{023}/\{f(x)V(x)\},$$
$$\rho a_2 = \sigma_k^6[f(x)\{f'(x)\}^2V(x)\{2R(K) + 1\} - \tfrac{1}{2}f^3(x)m''(x)V_3(x)R(K))V^{-1}(x)]$$
$$+ \tfrac{1}{2}f^2(x)\sigma_k^4\{V(x)f(x)\}''\{R_{22}(K) - R(K)\}$$
$$- \sigma_k^4R_{22}(K)[2f(x)f'(x)\{V(x)f(x)\}' + f^3(x)\{m'(x)\}^2],$$
$$a_3 = \mu_{220}\mu_{222} + \mu_{210}[-2\mu_{232} + 2\mu_{220}\mu_{022}f^{-1}(x) - \mu_{022}\mu_{420}\{f(x)\sigma_k^2\}^{-1}$$
$$+ \mu_{210}^2\{f^2(x)V^2(x)\}^{-1}\{\mu_{022}\mu_{024} + \tfrac{7}{4}\mu_{023}^2 + 3V(x)\mu_{022}^2 - V^2(x)\mu_{020}\mu_{022}],$$

$$a_4 = \mu_{210}^3[\mu_{033} - 3\mu_{022}\mu_{023}\{f(x)V(x)\}^{-1}],$$
$$a_5 = \mu_{210}^4\{\mu_{044} - 6\{f(x)V(x)\}^{-1}[(\mu_{023}\mu_{033} + \mu_{022}\mu_{034}) + 2\mu_{022}^3 f^{-1}(x) + (3\mu_{022}\mu_{023}^2 + \tfrac{1}{2}\mu_{022}^2\mu_{024})\{f(x)V(x)\}^{-1}]\}.$$

Noted that $\mu_{ij1} = O(h)$, and that for $k > 1$

$$\mu_{ijk} = f(x)V_k(x)\int t^i K^j(t)\,dt + o(h)$$

and $V_2(x) = V(x)$. Thus, we have, by omitting the smaller order terms,

$$a_1 = -\tfrac{1}{2}\mu_{210}\mu_{023}\{f(x)V(x)\}^{-1} = -\tfrac{1}{2}f(x)V_3(x)R(K)\sigma_k^2/V(x),$$
$$\begin{aligned} a_2 = {} & \sigma_k^2 f^{-2}(x)\{f'(x)\}^2\{2 + 1/R(K)\} - \tfrac{1}{2}\sigma_k^2 m''(x)V_3(x)V^{-2}(x) \\ & - [2f^{-2}(x)f'(x)\{f(x)V(x)\}' - \{m'(x)\}^2]V^{-1}(x)R_{22}(K)/R(K) \\ & + \tfrac{1}{2}f^{-1}(x)V^{-1}(x)\{f(x)V(x)\}''\{R_{22}(K)/R(K) - 1\}, \end{aligned}$$
$$\begin{aligned} a_3 = {} & f^2(x)V(x)[\{R_{22}(K)\}^2 + 2R(K)\sigma_k^2R_{22}(K) - 2\sigma_k^2R_{23}(K) - R(K)R_{42}(K) \\ & + 2R^2(K)\sigma_k^4] + f^2(x)R^2(K)\sigma_k^4\{V_4(x)/V(x) + \tfrac{7}{4}V_3^2(x)/V^2(x)\}, \end{aligned}$$
$$a_4 = \rho^2 V_3(x)\{R_{03}(K) - 3R^2(K)\}\{f(x)V(x)R(K)\sigma_k\}^{-2},$$
$$\begin{aligned} a_5 = {} & f^5(x)\sigma_k^8[V_4(x)\{R_{04}(K) + 3R^3(K) - 6R(K)R_{03}(K)\} + 12R^3(K)V^2(x) \\ & + 6R(K)V_3^2(x)V^{-1}(x)\{3R^2(K) - R_{03}(K)\}]. \end{aligned}$$

By setting up the generating function of T_{n1} from the above cumulants and inverting it, we obtain the following Edgeworth expansion

$$\begin{aligned} P(|T_{n1}| \leqslant z_\alpha) = {} & \alpha - 2\phi(z_\alpha)[\{\rho^{-1}a_1H_1(z_\alpha) + \tfrac{1}{6}\rho^{-2}a_4H_3(z_\alpha)\}\mu \\ & + \tfrac{1}{2}H_1(z_\alpha)\rho^{-1}nh\mu^2 + \tfrac{1}{2}a_2H_1(z_\alpha)h^2 \\ & + \{\tfrac{1}{2}\rho^{-1}a_3H_1(z_\alpha) + \tfrac{1}{2}\rho^{-1}a_1^2H_1(z_\alpha) + \tfrac{1}{24}\rho^{-2}a_5H_3(z_\alpha) \\ & + \tfrac{1}{6}\rho^{-2}a_1a_4H_3(z_\alpha) + \tfrac{1}{72}\rho^{-3}a_4^2H_5(z_\alpha)\}(nh)^{-1}] \\ & + O\{(nh)^{1/2}h^4 + (nh)^{-1/2}h^2 + h^4 + (nh)^{-3/2}\} \\ = {} & \alpha - \{b_1nh^5 + b_2h^2 + b_3(nh)^{-1}\} \\ & + O\{(nh)^{1/2}h^4 + (nh)^{-1/2}h^2 + h^4 + (nh)^{-3/2}\}. \end{aligned} \tag{15}$$

It may be shown that

$$b_1 = \tfrac{1}{4}H_1(z_\alpha)\phi(z_\alpha)\{m''(x)\}^2 f(x)\sigma_k^4/\{R(K)V(x)\},$$
$$\begin{aligned} b_2 = {} & H_1(z_\alpha)\phi(z_\alpha)\{\rho^{-1}a_1m''(x)f^2(x)\sigma_k^4 + a_2\} + \tfrac{1}{6}\rho^{-2}a_4H_3(z_\alpha)\phi(z_\alpha) \\ = {} & H_1(z_\alpha)\phi(z_\alpha)[\sigma_k^2f^{-2}(x)\{f'(x)\}^2\}2 + 1/R(K)\} - \sigma_k^2m''(x)V_3(x)V^{-2}(x) \\ & + \tfrac{1}{2}f^{-1}(x)V^{-1}(x)\{V(x)f(x)\}''\{R_{22}(K)/R(K) - 1\} \\ & - [2f^{-2}(x)f'(x)\{V(x)f(x)\}' - \{m'(x)\}^2]V^{-1}(x)R_{22}(K)/R(K)] \\ & + \tfrac{1}{6}H_3(z_\alpha)\phi(z_\alpha)f^{-2}(x)V^{-2}(x)\sigma_k^{-2}V_3(x)\{R_{03}(K)/R^2(K) - 3\}, \end{aligned} \tag{16}$$
$$\begin{aligned} b_3 = {} & H_1(z_\alpha)\phi(z_\alpha)\{f^{-1}(x)[\{R_{22}(K)\}^2/\{R(K)\sigma_k^4\} + 2R_{22}(K)/\sigma_k^2 \\ & - R_{42}(K)/\sigma_k^4 - 2R_{23}(K)/\{R(K)\sigma_k^2\} + 2R(K)] \\ & + f^{-1}(x)R(K)\{V_4(x)/V^2(x) + \tfrac{1}{2}V_3^2(x)/V^3(x)\}\} \\ & + \tfrac{1}{36}H_5(z_\alpha)\phi(z_\alpha)V_3^2(x)[R_{03}(K) - 3R^2(K)\}^2f^{-1}(x)V^{-3}(x)R^{-3}(K) \end{aligned} \tag{17}$$

$$+f^{-1}(x)R(K)+\tfrac{2}{3}f^{-1}(x)R^{-1}(K)V_3^2(x)V^{-3}(x)\{3R^2(K)-R_{03}(K)\}]$$
$$+H_3(z_\alpha)\phi(z_\alpha)\{\tfrac{1}{12}f^{-1}(x)R^{-2}(K)V^{-2}(x)[V_4(x)\{R_{04}(K)+3R^3(K)$$
$$-6R(K)R_{03}(K)\}]\}.$$

Combining with (14), theorem 1 is then verified.

It remains to check that the expansion (15) is valid. This can be done by first developing the Edgeworth expansion for the distribution of $\sqrt{nh}W$, with the form

$$P[\sqrt{nh}\{W-E(W)\}\in A]=\Phi_{0,\Sigma}(A)+\sum_{k=1}^{2}(nh)^{-k/2}\int_A p_k(x)\phi_{0,\Sigma}(x)\,dx+O\{(nh)^{-3/2}\} \tag{18}$$

uniformly for all A from $\mathscr{A}\in\mathscr{B}(R^6)$ of Borel sets satisfying

$$\sup_{A\in\mathscr{A}}\Phi_{0,\Sigma}\{(\partial A)^\epsilon\}=O(\epsilon)$$

as $\epsilon\to 0$. In this formulae, $\Phi_{0,\Sigma}$ and $\phi_{0,\Sigma}$ denote the distribution and density functions respectively of the normal $N(0,\Sigma)$ distribution; p_k is a polynomial of degree $k+2$ with uniformly bounded coefficients; and $(\partial A)^\epsilon$ is the set of all points distant at most ϵ from the boundary of A. After establishing the validity of (18), the validity of (15) can be obtained by applying a result of Skovgaard (1981) which proves that the validity of an Edgeworth expansion is preserved under a smooth transformation. As T_{n1} is a smooth function of W, the key at present is to show the expansion (18) is valid.

To establish the expansion in (18), it is sufficient to prove the following analogue of Cramér's condition; that is, for each $\epsilon>0$ there exists a positive $C(\epsilon)$ such that as $h\to 0$

$$\sup_{\sum_1^6|t_k|>\epsilon}\left|h\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\exp\left\{i\sum_{k=1}^{6}t_k\,g_k(u,v)\right\}f(x-hu)f(v|x-hu)\,du\,dv\right|\leqslant 1-C(\epsilon)h, \tag{19}$$

where $i=\sqrt{-1}$ and

$$g_1(u,v)=K(u),\quad g_2(u,v)=uK(u),\quad g_3(u,v)=u^2K(u),$$
$$g_4(u,v)=K(u)\{v-m(x)\},\quad g_5(u,v)=uK(u)\{v-m(x)\},\quad g_6(u,v)=K(u)\{v-m(x)\}^2,$$

sample.

We need the following lemma in proving (19).

Lemma 1

Assume that $K(u)$ has support $[-1,\ 1]$ on which $K(u)\neq 0$, $K(u)$ *is continuous differentiable and* $\{1,K(u),uK(u),u^2K(u)\}$ *are linearly independent* (*as elements of the vector space of continuous functions on* $[-1,\ 1]$). *Then*

$$\lim\sup_{\sum_1^6|t_k|\to\infty}\frac{1}{2}\left|\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\exp\left\{i\sum_{k=1}^{6}t_k\,g_k(u,v)\right\}f(v|x)\,du\,dv\right|<1. \tag{20}$$

Proof of Lemma 1. Notice that the linear independence of

$$\{1,g_1(u,v),g_2(u,v),\ldots,g_6(u,v)\}$$

is implied by $\{1,K(u),uK(u),u^2K(u)\}$ being linearly independent. Let $(U,\ V)$ be a bivariate random variable with a joint density $q(u,v)=\frac{1}{2}f(v|x)I(-1\leqslant u\leqslant 1)$. Clearly, the distribution of

(U, V) is absolutely continuous with respect to the Lebesgue measure in R^2 and the density q is positive in an open set in $[-1, 1] \times R$ where the $g_j s$ are continuously differentiable and $\{1, g_1(u, v), g_2(u, v), \ldots, g_6(u, v)\}$ are linearly independent. From lem. 2.2 of Bhattacharya & Ghosh (1978), the distribution of $(g_1(U, V), g_2(U, V), \ldots, g_6(U, V))$ has a non-zero absolutely continuous component. It may be shown from Hall (1992b, p. 65f) that this in turn implies the analogue of the Cramér's condition in (20).

To prove (19), we notice that

$$\begin{aligned} I(t_1, \ldots, t_6) &= h\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\left\{i\sum_{k=1}^{6} t_k g_k(u, v)\right\} f(x-hu) f(v|x-hu)\, du\, dv \\ &= 1 - h\int_{-1}^{1} f(x-hu)\, du \\ &\quad + h\int_{-1}^{1}\int_{-\infty}^{\infty} \exp\left\{i\sum_{k=1}^{6} t_k g_k(u, v)\right\} f(x-hu) f(v|x-hu)\, du\, dv \\ &= 1 - h\int_{-1}^{1} f(x-hu)\, du \\ &\quad + h\left[\int_{-1}^{1}\int_{-\infty}^{\infty} \exp\left\{i\sum_{k=1}^{6} t_k g_k(u, v)\right\} f(x) f(v|x)\, du\, dv + a_n\right] \end{aligned} \tag{21}$$

where $a_n = \circ(1)$ as $h \to 0$.

From lemma 1, it can be shown that there exists $\epsilon' > 0$ such that

$$\sup_{\sum_1^6 |t_k| \to \epsilon} \frac{1}{2}\left|\int_{-1}^{1}\int_{-\infty}^{\infty} \exp\left\{i\sum_{k=1}^{6} t_k g_k(u, v)\right\} f(x) f(v|x)\, du\, dv\right| < f(x)(1-3\epsilon'). \tag{22}$$

Let h be small enough such that $|a_n| < 2f(x)\epsilon'$ and

$$\int_{-1}^{1} f(x-hu)\, du \geqslant 2f(x)(1-\epsilon').$$

Thus, from (21) we have

$$\begin{aligned} |I(t_1, \ldots, t_6)| &\leqslant 1 - 2hf(x)(1-\epsilon') + 2hf(x)(1-3\epsilon') + 2hf(x)\epsilon' \\ &= 1 - 2hf(x)\epsilon', \end{aligned}$$

which gives (19) and thus (18).

Proof of Theorem 2. Put

$$\overline{U}_2 = \overline{U}_{21} - \gamma(x/h) f^2(x), \quad \overline{U}_3 = \bar{\omega}_{010} - \alpha_0(x/h) f(x), \quad \overline{U}_4 = \bar{\omega}_{012} - \alpha_0(x/h) f(x) V(x).$$

To simplify the notation, we will hide the argument x/h of α_j and β_j in the following. The definitions of β_j and γ are in (9). Notice that

$$T_n = \sqrt{\frac{nh}{\rho(x/h)}}\left\{1 + \frac{\overline{U}_3}{\alpha_0 f(x)}\right\} \frac{\overline{U}_1}{\sqrt{1+\Delta_2}}$$

where

$$\begin{aligned} \Delta_2 = \{\alpha_0\gamma^2 f^5(x) V(x)\}^{-1} \{&\overline{U}_4 \overline{U}_2^2 + 2\gamma f^2(x) \overline{U}_2 \overline{U}_4 + \gamma^2 f^4(x) \overline{U}_4 - 2\overline{U}_1 \overline{U}_2 \bar{\omega}_{011} \\ &- 2\gamma f^2(x) \overline{U}_1 \bar{\omega}_{011} + \overline{U}_1^2 \overline{U}_3 + \alpha_0 f(x) \overline{U}_1^2\} + \{\gamma^2 f^4(x)\}^{-1} \{\overline{U}_2^2 + 2\gamma f^2(x) \overline{U}_2\}. \end{aligned}$$

Let

$$T_{n2} = \sqrt{nh/\rho(x/h)}\,\overline{U}_1[1 - \overline{U}_2/\{\gamma f^2(x)\} - \overline{U}_4/\{2\alpha_0 f(x)V(x)\} + \overline{U}_3/\{\alpha_0 f(x)\}$$
$$+ \{\gamma^2 f^4(x)V(x)\}^{-1}\{V(x)\overline{U}_2^2 - \tfrac{1}{2}\overline{U}_1^2\} + \tfrac{3}{8}\{\alpha_0^2 f^2(x)V^2(x)\}^{-1}\{\overline{U}_4^2 - \tfrac{4}{3}V(x)\overline{U}_3\overline{U}_4\}$$
$$+ \{\alpha_0\gamma f^3(x)V(x)\}^{-1}\{\overline{U}_1\bar{\omega}_{011} + \tfrac{1}{2}\overline{U}_2\overline{U}_4 - V(x)\overline{U}_2\overline{U}_3\}].$$

By Taylor expansion, we have $T_n = T_{n2} + R_2$ where $R_2 = O_p[\{(nh)^{-1/2} + h\}^3]$. It may be shown, again using the method described in Hall (1992b, pp. 76–77), that

$$P(T_n \leqslant u) = P(T_{n2} \leqslant u) + O[\{(nh)^{-1/2} + h\}^3].$$

Computation shows that the cumulants of T_{n2} are

$$k_1 = \rho(x/h)^{-1/2}\{\sqrt{nh}\mu + a_1'(nh)^{-1/2}\} + O\{(nh)^{1/2}h^4 + (nh)^{-1/2}h^2 + (nh)^{-3/2}\},$$
$$k_2 = 1 + a_2'h + \rho(x/h)^{-1}a_3'(nh)^{-1} + O\{h^2 + (nh)^{-1}h + (nh)^{-2}\},$$
$$k_3 = \rho(x/h)^{-3/2}a_4'(nh)^{-1/2} + O\{(nh)^{-1/2}h + (nh)^{-3/2}\},$$
$$k_4 = \rho(x/h)^{-2}a_5'(nh)^{-1} + O\{(nh)^{-1}h + (nh)^{-2}\},$$
$$k_l = O\{(nh^d)^{-(l-2)/2}\} \quad \text{for } l \geqslant 5,$$

where

$$\mu = \tfrac{1}{2}\{\alpha_2^2(x/h) - \alpha_1(x/h)\alpha_3(x/h)\}f^2(x)m''(x)h^2,$$
$$a_2' = \rho^{-1}(x/h)\{f^2(x)f'(x)V(x)(\alpha_2^2\beta_1 + 2\alpha_1\alpha_3\beta_1 - 2\alpha_2\alpha_3\beta_0 - \alpha_1^2\beta_3)$$
$$+ f^3(x)V'(x)(-\alpha_2^2\beta_1 + 2\alpha_1\alpha_2\beta_2 - \alpha_1^2\beta_3)\} \tag{23}$$
$$+ \{2\alpha_0\gamma f(x)V(x)\}^{-1}\{f'(x)V(x)(3\alpha_0\alpha_1\alpha_2 - \alpha_1^3 - 2\alpha_0^2\alpha_3) - f(x)V'(x)\alpha_1\gamma\}$$

and $a_j', j = 1, 3, 4, 5$, depends only on f, V and x/h whose detailed forms are not given here as they are quite tedious.

By deriving the generating function for T_{n2} and converting it we have the following Edgeworth expansion:

$$P(|T_{n2}| \leqslant z_\alpha) = \alpha - 2\phi(z_\alpha)[\tfrac{1}{2}H_1(z_\alpha)\rho(x/h)^{-1}nh\mu^2 + \tfrac{1}{2}a_2'H_1(z_\alpha)h$$
$$+ \{\tfrac{1}{2}\rho(x/h)^{-1}a_3'H_1(z_\alpha) + \tfrac{1}{2}\rho(x/h)^{-1}(a_1')^2H_1(z_\alpha) + \tfrac{1}{24}\rho(x/h)^{-2}a_5'H_3(z_\alpha)$$
$$+ \tfrac{1}{6}\rho(x/h)^{-2}a_1'a_4'H_3(z_\alpha) + \tfrac{1}{72}\rho(x/h)^{-3}(a_4')^2H_5(z_\alpha)\}(nh)^{-1}]$$
$$+ O\{(nh)^{1/2}h^4 + (nh)^{-1/2}h + h^2 + (nh)^{-3/2}\}$$
$$=: \alpha - \{d_1nh^5 + d_2h + d_3(nh)^{-1}\}$$
$$+ O\{(nh)^{1/2}h^4 + (nh)^{-1/2}h + h^2 + (nh)^{-3/2}\}$$

where

$$d_1 = \tfrac{1}{4}\rho(x/h)^{-1}z_\alpha\phi(z_\alpha)\{\alpha_2^2(x/h) - \alpha_1(x/h)\alpha_3(x/h)\}f^2(x)m''(x)$$

and $d_2 = z_\alpha\phi(z_\alpha)a_2'$ where a_2' is defined in (23). The exact expression for d_3 is quite tedious to write out. It can be derived by working out a_1', a_3', a_4' and a_5' as we have done in the proof of theorem 1.

Acknowledgement

The authors thank two referees for comments and suggestions that improved the presentation of the paper.

References

Bhattacharya, R. N. & Ghosh, J. K. (1978). On the validity of the formal Edgeworth expansion. *Ann. Statist*. **6**, 434–451.

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

Chen, S. X. & Qin, Y. S. (2000). Empirical likelihood confidence interval for a local linear smoother. *Biometrika* **87**, 946–955.

Cleveland, W. S. (1979) Robust locally weighted regression and smoothing scatterplots. *J. Amer. Statist. Assoc*. **74**, 829–836.

Cleveland, W. S. & Devlin, S. (1988). Locally weighted regression: an approach to regression analysis by local fitting. *J. Amer. Statist. Assoc*. **83**, 596–610.

Fan, J. (1993). Local linear regression smoothers and their minimax efficiencies. *Ann. Statist*. **21**, 196–216.

Fan, J. & Gijbels, I. (1992). Variable bandwidth and local linear regression smoothers. *Ann. Statist*. **20**, 2008–2036.

Fan, J. & Gijbels, I. (1996). *Local polynomial modelling and its applications*. Chapman & Hall, London.

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

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

Hall, P. (1992b). *The bootstrap and Edgeworth expansions*. Springer-Verlag, New York.

Härdle, W. (1990). *Applied non-parametric regression*. Cambridge University Press, Cambridge.

Hastie, T. & Loader, C. (1993). Local regression: automatic kernel carpentry. *Statist. Sci*. **8**, 120–143.

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

Jones, M. C. (1993). Simple boundary correction for kernel density estimation. *Statist. Comput*. **3**, 135–146.

Müller, H.-G. & Wang, J.-L. (1994). Hazard rate estimation under random censoring with varying kernel and bandwidth. *Biometrics* **50**, 61–67.

Neumann, M. H. (1995). Automatic bandwidth choice and confidence intervals in nonparametric regression. *Ann. Statist*. **23**, 1937–1959.

Neumann, M. H. (1997). Pointwise confidence intervals in non-parametric regression with heteroscedastic error structure. *Statistics* **29**, 1–36.

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

Rodríguez-Campos, M. C. (1999). On confidence intervals in nonparametric binary regression via Edgeworth expansions. *J. Multivariate Anal*. **69**, 218–214.

Ruppert, D. & Wand, M. P. (1994). Multivariate locally weighted least squares regression. *Ann. Statist*. **22**, 1346–1370.

Skovgaard, I. B. M. (1981). Transformation of an Edgeworth expansion by a sequence of smooth functions. *Scand. J. Statist*. **8**, 207–217.

Stone, C. J. (1977). Consistent nonparametric regression. *Ann. Statist*. **5**, 595–645.

Received March 2000, in final form October 2000

Song X. Chen, Department of Statistics and Applied Probability, National University of Singapore, Singapore 117543.
E-mail: stacsx@nus.edu.sg