Statistica Sinica **12**(2002), 1043-1060

REGRESSION MODELING FOR NONPARAMETRIC ESTIMATION OF DISTRIBUTION AND QUANTILE FUNCTIONS

Ming-Yen Cheng and Liang Peng

National Taiwan University and Georgia Institute of Technology

Abstract: We propose a local linear estimator of a smooth distribution function. This estimator applies local linear techniques to observations from a regression model in which the value of the empirical distribution function equals the value of true distribution plus an error term. We show that, for most commonly used kernel functions, our local linear estimator has a smaller asymptotic mean integrated squared error than the conventional kernel distribution estimator. Importantly, since this MISE reduction occurs through a constant factor of a second order term, any bandwidth selection procedures for kernel distribution estimator can be easily adapted for our estimator. For the estimation of a smooth quantile function, we establish a regression model of the empirical quantile function and obtain a local quadratic estimator. It has better asymptotic performance than the kernel quantile estimator in both interior and boundary cases.

Key words and phrases: Distribution function, empirical quantiles, kernel, local polynomial estimation, nonparametric estimation, quantile, smoothing.

1. Introduction

Suppose that $X_1, \ldots, X_n$ are independent and identically distributed random variables from a common univariate distribution function F. One obvious estimator of the distribution function F is the empirical distribution function $F_n(x) = \frac{1}{n}\sum_{i=1}^n I(X_i \le x)$, where $I(A)$ denotes the indicator function of set A. Though F_n has good properties, one may prefer a smooth estimator of F. An example is in the estimation of the receiver operating characteristic (ROC) curves for continuous diagnostic tests (see Zou, Hall and Shapiro (1997)). Kernel distribution estimators $\tilde{F}(x) = \frac{1}{n}\sum_{i=1}^n K(\frac{x-X_i}{h})$ are smooth and have been extensively studied in the literature. A potential problem is the boundary effect: one may have substantial bias near the boundaries of the data range.

We introduce smooth distribution estimators that avoid boundary effects. They are derived by writing

$$F_n(X_i) = F(X_i) + \epsilon_i, i = 1, \ldots, n, \tag{1.1}$$

where the $\epsilon_i's$ are error terms. Having the 'regression model' (1.1), one can employ nonparametric regression techniques, for example Nadaraya-Watson (1964b), and local polynomial (Fan and Gijbels (1996)) methods, to the data $(X_1, F_n(X_1)), \ldots, (X_n, F_n(X_n))$ to construct smooth estimators of F. In this paper we concentrate on local linear smoothing.

The mean integrated squared errors of $\tilde{F}(x)$ and our local linear distribution estimator both have the same leading term as that of the empirical distribution. However, for many commonly used kernel functions, the local linear distribution estimator has a smaller second order term, and hence asymptotic mean integrated squared error, than $\tilde{F}(x)$. The reduction in the second order term, which arises as effects of smoothing, is approximately 60% if the Epanechnikov, Biweight, or triangular kernels are employed. Furthermore, the asymptotic mean integrated squared error expressions suggest that any bandwidth selection procedure tailored for $\tilde{F}$ can be used, with a simple constant adjustment, for implementation of our estimator.

Quantile estimation plays an important role in a wide range of statistical applications: the Q-Q plot, Value-at-Risk in financial risk management, etc. A natural estimator for the quantile function $Q(p)$ is the empirical quantile function. Several smooth quantile estimators have appeared in the literature. Nadaraya (1964a) discussed a kernel estimator defined as the inverse of the kernel distribution estimator. Parzen (1979) proposed kernel quantile estimators, which were subsequently investigated by Yang (1985), Falk (1985), Zelterman (1990), and Sheather and Marron (1990). A unified kernel quantile estimator was given by Cheng and Parzen (1997). We consider local quadratic regression estimation for the quantile function through the relation

$$Q_n(s) = Q(s) + \text{error term}, \tag{1.2}$$

where Q_n is the empirical quantile function. We show that, under stronger smoothness assumptions, the local quadratic estimator of $Q(p)$ is better than the kernel quantile estimator $\hat{Q}_n^k(p)$, defined in (3.1), both when p is a fixed number in $(0,1)$ and when p tends to 0 or 1.

While F and Q are necessarily nondecreasing, it is possible that our estimators of these functions are decreasing at some places. However, since the response factors $F_n(x)$ in (1.1) and $Q_n(s)$ in (1.2) are nondecreasing, the locally fitted curves are not far from nondecreasing. When a nondecreasing estimated curve is desired, it can be achieved by applying isotonic regression techniques, see for example Mammen, Marron, Turlach and Wand (2001), to our estimator.

Smoothing regression techniques have been applied to estimation of distribution and density functions. Cheng, Fan and Marron (1997) obtained local linear density estimators by constructing regression data based on binning the original data. Wei and Chu (1994) regressed responses, obtain by differencing F_n on design points i/n, to estimate a density function. Lejeune and Sarda (1992) minimized a locally kernel-weighted L^2 norm between F_n and a polynomial to obtain a distribution estimator, and then used its derivative as an estimator of the density. This distribution estimator is essentially the same as the kernel quantile estimator in the interior. The regression functions in (1.1) and (1.2) are exactly the respective functions to be estimated. Thus our estimation methods are advantageous as they involve only denoising and not other operations, such as differentiation.

This paper is organized as follows. In Section 2, the local linear distribution estimator is derived, asymptotic results and a few remarks are given. Section 3 discusses local quadratic quantile estimation. Section 4 reports simulation studies on the finite sample performance of the distribution and quantile estimators. Proofs of the main results are given in the Appendix.

2. Distribution Estimation

The kernel distribution estimation was introduced by Nadaraya (1964a) and is defined by

$$\tilde{F}(x) = \frac{1}{n}\sum_{i=1}^{n} K\Big(\frac{x - X_i}{h}\Big), \tag{2.1}$$

where K is a given distribution function and $h = h(n) > 0$ ($h \to 0$ as $n \to \infty$) is the bandwidth. This estimator is the distribution function corresponding to the kernel density estimator based on the kernel $k(t) = K'(t)$ and bandwidth h. Note also that this estimator can be obtained by smoothing the empirical process $F_n(x)$, using kernel k, in the same way we smooth the quantile process, see Section 3. Theoretical properties of $\tilde{F}(x)$ as an estimator of the unknown true distribution function $F(x)$ have been investigated by several authors, see for example Yamato (1973), Reiss (1981) and Falk (1983). For the Edgeworth expansions, we refer to Garcia-Soidan, Gonzalez-Manteiga and Prada-Sanchez (1997). Altman and Léger (1995) and Bowman, Hall and Prvan (1998) investigated the optimal choice of bandwidth.

The optimal choice of the smoothing parameter h is obtained by minimizing the mean integrated squared error defined by

$$MISE(h, \tilde{F}) = E\int (\tilde{F}(x) - F(x))^2 W(x)\, dF(x) \tag{2.2}$$

in Altman and Lėger (1995), and

$$MISE^{\dagger}(h,\tilde{F}) = E\int(\tilde{F}(x)-F(x))^2dx \qquad (2.3)$$

in Bowman, Hall and Prvan (1998). Here, W is a bounded, nonnegative weight function supported on a compact set. As pointed out by Bowman, Hall and Prvan (1998), this kind of optimal choice of h is asymptotic to one that produces second order optimality. More specifically, the choice of bandwidth does not affect the first order expansion of $MISE^{\dagger}(h,\tilde{F})$ or $MISE(h,\tilde{F})$, i.e., the n^{-1} term, only if $\sqrt{n}h^2 \to 0$.

It is known that, under some regularity conditions, the mean integrated squared error for the kernel distribution estimation $\tilde{F}$ has the expansion

$$MISE(h,\tilde{F}) = v_1n^{-1} - 2c_2v_2n^{-1}h + \frac{c_1^2v_3}{4}h^4 + o\left(nh^{-1}\right) + o\left(h^4\right), \qquad (2.4)$$

where $v_1 = \int F(x)\,[1-F(x)]W(x)f(x)\,dx$, $v_2 = \int f^2(x)\,W(x)\,dx$, $v_3 = \int(f'(x))^2W(x)f(x)\,dx$, $c_1 = \int x^2k(x)\,dx$, $c_2 = \int xk(x)K(x)\,dx$ and $f(x) = F'(x)$. See, for example, Falk (1983). Obviously, kernel smoothing provides a second order correction (i.e., deficiency): mean integrated squared error of $F_n(x)$ has the same leading term v_1n^{-1} which is independent of the smoothing parameter h. The asymptotically optimal bandwidth that minimizes the second order correction induced by smoothing, $-2c_2v_2hn^{-1}+c_1^2v_3h^4/4$, is $h^*(\tilde{F}) = [2c_2v_2/(c_1^2v_3)]^{1/3}n^{-1/3}$. The optimal bandwidth $h^*(\tilde{F})$ gives rise to $MISE(h^*(\tilde{F}), \tilde{F}) = v_1n^{-1} - \frac{3}{4}(2c_2v_2)^{4/3}(c_1^2v_3)^{-1/3}n^{-4/3} + o(n^{-4/3})$. The second term in this asymptotic expression is negative. Therefore, kernel smoothing improves empirical distribution estimator by a second order effect.

Now we derive our estimator based on (1.1) and local linear smoothing techniques. Let k, called a kernel function, be a probability density and $h > 0$ be a bandwidth. For simplicity of notation, we take $k(t) = K'(t)$ with $K(t)$ defined in (2.1). Let $(\hat{a},\hat{b})$ be the value of (a,b) that minimizes the kernel weighted squared errors

$$\sum_{j=1}^{n}\{F_n(X_j)-a-b(x-X_j)\}^2k\Big(\frac{x-X_j}{h}\Big).$$

Then the local linear distribution estimator is defined as $\hat{a}$ and has the following explicit expression $\hat{F}(x) = \frac{\sum_{j=1}^n w_jF_n(X_j)}{\sum_{j=1}^n w_j}$, where $w_j = k(\frac{x-X_j}{h})[s_{n,2}-(x-X_j)s_{n,1}]$, $j = 1,\ldots,n$, with $s_{n,l} = \sum_{j=1}^n k(\frac{x-X_j}{h})(x-X_j)^l$ for $l = 1,2$. See Fan (1992) for the derivation of $\hat{F}(x)$.

Throughout this paper we assume that the kernel function k is symmetric about zero and has support $[-1,1]$. The mean integrated squared error for our local linear estimator, under some regularity conditions similar to those in Altman and Lé́ger (1995), is

$$MISE(h,\hat{F}) = v_1 n^{-1} - (4c_2 - c_3)v_2 n^{-1}h + \frac{c_1^2 v_3}{4}h^4 + o\left(nh^{-1}\right) + o\left(h^4\right), \quad (2.5)$$

for $C_0 n^{-1+\epsilon_0} \le h \le C_1 n^{-\epsilon_1}$, where $\epsilon_0 \in (0,2/3]$, $\epsilon_1 \in (0,1/3]$, C_0 and C_1 are positive constants, and $c_3 = \int x^2 k^2(x)\,dx$. Proof of (2.5) is given in the Appendix. The constant factor $4c_2 - c_3\ (\ge 2c_2 - c_3)$ in (2.5) is positive for most commonly used kernels, see Table 1. So the bandwidth which minimizes $-(4c_2 - c_3)v_2 n^{-1}h + \frac{c_1^2 v_3}{4}h^4$ is $h^*(\hat{F}) = \left[(4c_2 - c_3)v_2/(c_1^2 v_3)\right]^{1/3} n^{-1/3}$, and it results in $MISE(h^*(\hat{F}),\hat{F}) = v_1 n^{-1} - \frac{3}{4}[(4c_2 - c_3)v_2]^{4/3}(c_1^2 v_3)^{-1/3} n^{-4/3} + o(n^{-4/3})$.

Notice that $MISE(h^*(\hat{F}),\hat{F}) \le MISE(h^*(\tilde{F}),\tilde{F})$ is asymptotically equivalent to $2c_2 - c_3 \ge 0$. In addition, comparing the asymptotic expressions of $MISE(h^*(\hat{F}),\hat{F})$ and $MISE(h^*(\tilde{F}),\tilde{F})$, we see that $\{(2c_2 - c_3)/(2c_2)\}^{4/3}$ is the relative improvement of $\hat{F}$ over $\tilde{F}$ in terms of their second order performances. Values of $(2c_2 - c_3)$ and $\{(2c_2 - c_3)/(2c_2)\}^{4/3}$ are tabulated in Table 1 for some kernels that are commonly used in practice. In particular, the improvement is 58% for the Epanechnikov kernel, 62% for the Biweight kernel, and 64% for the triangular kernel. Such an improvement is particularly beneficial when the sample size is small or moderate, and is clearly seen in a simulation study given in Section 4.

Table 1. Comparison of the second order terms of $MISE\big(h^*(\tilde{F}),\tilde{F}\big)$ and $MISE\big(h(\hat{F}),\hat{F}\big)$ for some commonly used kernels.

Kernel	$2c_2 - c_3$	$\big\{(2c_2 - c_3)/(2c_2)\big\}^{4/3}$
Epanechnikov $k(x) = \dfrac{3}{4}(1-x^2)I(\lvert x\rvert \le 1)$	$\dfrac{6}{35}$	$(\frac{2}{3})^{4/3} \approx 0.5824$
Biweight $k(x) = \dfrac{16}{15}(1-x^2)^2 I(\lvert x\rvert \le 1)$	$\dfrac{5}{33}$	$(\dfrac{7}{10})^{4/3} \approx 0.6215$
Triangular $k(x) = (1-\lvert x\rvert)I(\lvert x\rvert \le 1)$	$\dfrac{1}{6}$	$(\dfrac{5}{7})^{4/3} \approx 0.6385$
Uniform $k(x) = \dfrac{1}{2}I(\lvert x\rvert \le 1)$	$\dfrac{1}{6}$	$(\dfrac{1}{2})^{4/3} \approx 0.3969$

Observe that the asymptotic expressions of $MISE(h,\hat{F})$ and $MISE(h,\tilde{F})$ differ only in the $n^{-1}h$ term, and that the optimal bandwidths $h^*(\hat{F})$ and $h^*(\tilde{F})$

differ by a constant multiplication factor. These facts suggest that any bandwidth rule for $\tilde{F}$ multiplied by a constant can be readily used for $\hat{F}$. Hence there is no need to invent new bandwidth selector in order to implement our local linear distribution estimator.

Remark 1. One can derive a Nadaraya-Watson type estimator of the distribution function based on the model (1.1). It can be shown that this estimator (with an optimal bandwidth) has a greater MISE than $\tilde{F}$ for large n and any kernel function.

Remark 2. The assumption that the kernel function k has support $[-1, 1]$ may be dropped through a more careful analysis.

3. Quantile Estimation

In this section we discuss estimation of quantile functions. The quantile function corresponding to the distribution function F is defined as $Q(p) = \inf\{x : F(x) \geq p\}$ for $p \in (0, 1]$. The empirical quantile estimator is

$$Q_n(p) = \begin{cases} X_{n,s}, & \text{if} \quad (s-1)/n < p \leq s/n, \quad s = 1, \ldots, n, \\ X_{n,1}, & \text{if} \quad p = 0, \end{cases}$$

where $X_{n,1} \leq \cdots \leq X_{n,n}$ denote the order statistics of $X_1, \ldots, X_n$. The kernel quantile estimator, proposed by Parzen (1979), is given by

$$\hat{Q}_n^k(p) = \int_0^1 h^{-1} k\Big(\frac{s-p}{h}\Big) Q_n(s) ds, \tag{3.1}$$

where k is a probability density function and $h > 0$ is the bandwidth. As we see later, this type of kernel quantile estimators has a slower rate of convergence when p is a boundary point than when p is a fixed interior point.

To estimate the quantile $Q(p)$ we utilize (1.2), which establishes a regression relation between the empirical quantile function and the true quantile function, and apply a local quadratic technique as follows. Find the values of a, b and c that minimize the weighted integral of squared error of a quadratic approximation

$$\int_0^1 \Big[Q_n(s) - a - b(p-s) - c(p-s)^2\Big]^2 k\Big(\frac{p-s}{h}\Big) ds,$$

where k is a density function and $h > 0$ is the bandwidth. Then the local quadratic quantile estimator is defined to be the value of a in the above solution. It has the form

$$\hat{Q}(p) = \frac{(a_2a_4 - a_3^2)A_0(p) - (a_1a_4 - a_2a_3)A_1(p) + (a_1a_3 - a_2^2)A_2(p)}{a_0(a_2a_4 - a_3^2) - a_1(a_1a_4 - a_2a_3) + a_2(a_1a_3 - a_2^2)},$$

where $a_i = \int_0^1 (p-s)^i k(\frac{p-s}{h})ds$, $i = 0, 1, \ldots, 6$, and $A_i(p) = \int_0^1 (p-s)^i k(\frac{p-s}{h}) Q_n(s)ds$, $i = 0, 1, 2$. Note that the $a_i's$ are functions of p and h. For simplicity of notation we suppress this dependence. Throughout this section we assume that k is a symmetric density about zero and has support $[-1, 1]$. In the case that p is a fixed interior point in $(0, 1)$, $\hat{Q}_n^k(p)$ is the same as a local linear quantile estimator (see Section 3.1), which is defined as the solution in a that minimizes $\int_0^1 [Q_n(s) - a - b(p-s)]^2 k(\frac{p-s}{h})ds$. This is the reason why we consider local quadratic estimator instead of local linear estimator. Another local quadratic estimator is obtained by minimizing $\sum_{i=1}^n [Q_n(X_i) - a - b(p - X_i) - c(p - X_i)^2]^2 k(\frac{p-X_i}{h})$. This estimator is asymptotically equivalent to $\hat{Q}_n^k(p)$. Asymptotic properties of the quantile estimators $\hat{Q}_n^k(p)$ and $\hat{Q}(p)$ are considered in the following two subsections.

3.1. Interior quantiles

In this subsection p is a fixed interior point in $(0, 1)$. Under this circumstance, as n is large enough and h is small enough, $a_0 = h$, $a_1 = a_3 = a_5 = 0$, $a_2 = h^3 \int_{-1}^1 s^2 k(s)\, ds$, $a_4 = h^5 \int_{-1}^1 s^4 k(s)\, ds$, $a_6 = h^7 \int_{-1}^1 s^6 k(s)\, ds$. Hence the local quadratic quantile estimator becomes $\hat{Q}(p) = \frac{a_4 A_0(p) - a_2 A_2(p)}{a_0 a_4 - a_2^2}$. Also, if we define a kernel function k_2 by $k_2(u) = \frac{h(a_4 - a_2 u^2 h^2)}{a_0 a_4 - a_2^2} k(u)$ then we can write $\hat{Q}(p) = \int_0^1 \frac{1}{h} k_2(\frac{p-s}{h}) Q_n(s)ds$, which is a kernel quantile estimator, see (3.1), with the kernel k_2.

If the second derivative of Q is continuous in a neighborhood of p, then the asymptotic mean squared error of $\hat{Q}_n^k(p)$ is

$$MSE(\hat{Q}_n^k(p)) = n^{-1}p(1-p)[Q'(p)]^2 - 2n^{-1}h[Q'(p)]^2 \int_{-1}^1 sk(s)K(s)ds + \frac{1}{4}h^4[Q''(p)]^2 \left[\int_{-1}^1 s^2 k(s)\, ds\right]^2 + o(n^{-1}h) + o(h^4), \quad (3.2)$$

where $K(u) = \int_{-1}^u k(s)\, ds$ (see Sheather and Marron (1990)). If the fourth derivative of Q is continuous in a neighborhood of p and $EX_1^2 < \infty$, then the mean squared error of our local quadratic estimator $\hat{Q}(p)$ has the asymptotic expression

$$\begin{aligned} MSE(\hat{Q}(p)) = {} & n^{-1}p(1-p)[Q'(p)]^2 - 2n^{-1}h[Q'(p)]^2\sigma_1^2 \\ & + \frac{1}{24^2}h^8[Q^{(4)}(p)]^2 \left[\int_{-1}^1 s^4 k_2(s)ds\right]^2 \\ & + o(n^{-1}h) + O(n^{-3/2}\log n) + o(h^8), \end{aligned} \quad (3.3)$$

where $\sigma_1^2 = \int_{-1}^1 sk_2(s)K_2(s)\, ds$ with $K_2(s) = \int_{-1}^s k_2(t)\, dt$. Proof of (3.3) is given in the Appendix. We remark tha.t the condition $EX_1^2 < \infty$ may be removed by a more careful analysis, similar to that in Falk (1984), for example.

From (3.2) and (3.3) we have the following conclusions. First, $\hat{Q}(p)$ and $\hat{Q}_n^k(p)$ have the same leading term in their mean squared errors and the leading term is independent of the smoothing. Second, the minimal asymptotic mean squared error of $\hat{Q}_n^k(p)$, with respect to the smoothing parameter h, is $n^{-1}p(1-p)[Q'(p)]^2 - C_1 n^{-4/3}$ and that for $\hat{Q}(p)$ is $n^{-1}p(1-p)[Q'(p)]^2 - C_2 n^{-8/7}$. Here, C_1 and C_2 are some positive constants. Therefore, $\hat{Q}(p)$ has a better mean squared error performance than $\hat{Q}_n^k(p)$.

3.2. Boundary quantiles

Throughout this subsection we assume that the distribution function F has a finite left end point, i.e., $Q(0-) \in (-\infty, \infty)$. In order to investigate the boundary effect of the kernel quantile estimator and the local quadratic estimator, we assume $p = lh$ where $l \in (0,1)$. We can investigate the right boundary case similarly by taking $p = 1 - lh$ where $l \in (0,1)$.

In the boundary case, $\hat{Q}_n^k(p) - Q(p)/\int_{-1}^{l} k(s)ds \xrightarrow{p} 0$ and $\hat{Q}_n^k(p)$ is no longer a consistent estimator of $Q(p)$ unless $Q(p) = 0$. Consider the modified kernel estimator

$$\bar{Q}_n^k(p) = \frac{\int_0^1 k(\frac{s-p}{h}) Q_n(s)\, ds}{\int_0^1 k(\frac{s-p}{h})\, ds}.$$

Then, if Q' is continuous in a neighborhood of zero and $n^{1/2}h/\log n \to \infty$ as $n \to \infty$, we may show, in a similar way as Falk (1984), that

$$h^{-1}\{\bar{Q}_n^k(p) - Q(p)\} \xrightarrow{p} \int_{-1}^{l} sk(s)\, ds \Big/ \int_{-1}^{l} k(s)\, ds. \tag{3.4}$$

Note that the optimal choice of $h = O(n^{-1/3})$ for kernel quantile estimation, in the sense of minimizing the second order error term in (3.2), satisfies the condition $n^{1/2}h/\log n \to \infty$ (see Sheather and Marron (1990)). Hence we may conclude from (3.4) that the kernel quantile estimator does not perform as well at near boundary points as at interior points.

Next we study the boundary effect of the local quadratic estimator. In this case, for $p = lh$, $l \in (0,1)$, with n tending to infinity and h tending to zero, we have $a_i = h^{i+1}\int_{-1}^{l} s^i k(s)\, ds$, $i = 0,1,2,3,4,5$, and $k_2(u) = \frac{h}{d}[(a_2a_4 - a_3^2) - (a_1a_4 - a_2a_3)hu + (a_1a_3 - a_2^2)h^2u^2]k(u)$, where $d = a_0(a_2a_4 - a_3^2) - a_1(a_1a_4 - a_2a_3) + a_2(a_1a_3 - a_2^2)$. If the third derivative of Q is continuous in a neighborhood of zero and $EX_1^2 < \infty$, then the mean squared error of $\hat{Q}(p)$ has the asymptotic expression

$$\begin{aligned} MSE(\hat{Q}(p)) = {} & n^{-1}h[Q'(p)]^2 l - 2n^{-1}h[Q'(p)]^2\sigma_2^2 + \frac{1}{36}h^6[Q^{(3)}(p)]^2\left[\int_{-1}^{l} s^3 k_2(s) ds\right]^2 \\ & + o(n^{-1}h) + O(n^{-3/2}\log n) + o(h^6), \end{aligned} \tag{3.5}$$

where $\sigma_2^2 = \int_{-1}^{l} sk_2(s)K_2(s)ds$ with $K_2(s) = \int_{-1}^{s} k_2(t)dt$. Proof of (3.5) is similar to proof of (3.3). We remark again that the condition $EX_1^2 < \infty$ may be removed by a more careful analysis similar to that in Falk (1984). Note that (3.4) implies $MSE(\bar{Q}_n^k(p)) = O(n^{-2/3})$, which is much larger than $MSE(\hat{Q}(p)) = O(n^{-1})$ in this boundary case.

4. Simulation Studies

4.1. Distribution estimation

A Monte Carlo study was conducted to compare the mean integrated squared error performances of local linear and kernel distribution estimators. A discrete approximation to $MISE(h, \tilde{F})$ is $ASE(h, \tilde{F}) = n^{-1} \sum_{i=1}^{n} [\tilde{F}(X_i) - F(X_i)]^2 W(X_i)$. The Epanechnikov kernel $k(x) = \frac{3}{4}(1 - x^2)I(|x| \leq 1)$ was used to construct the two estimators. To compute the average squared errors of $\tilde{F}(x)$ and $\hat{F}(x)$, i.e., $ASE(h, \tilde{F})$ and $ASE(h, \hat{F})$, we generated 500 samples from Weibull(θ, τ) distributions $F_{\theta,\tau}(x) = 1 - \exp(-\theta x^{\tau})$ $(x > 0)$, where $\theta, \tau > 0$. The weight function W was chosen as $W(x) \equiv 1$. The sample size was n = 10, 30, 50 or 70. Noting that $h^*(\hat{F})$ differs from $h^*(\tilde{F})$ only by a constant factor, the plug-in approach proposed by Altman and Léger (1995) was employed to choose the bandwidths for both estimators. In this simulation study we replaced $\hat{F}(x)$ by $F_n(x)$ whenever the value of the denominator $\sum_{j=1}^{n} w_j$ defined in $\hat{F}(x)$ is zero. Moreover $\hat{F}(x)$ can be easily modified to be a distribution, for example, by defining $\hat{F}(x) = 0$ if $x \leq \inf\{y : \hat{F}(y) > 0\}$, $\hat{F}(x) = 1$ if $x \geq \sup\{y : \hat{F}(y) < 1\}$ and $\hat{F}(x) = \hat{F}(x)$ otherwise.

The ratio of the empirical mean of $ASE(h^*(\tilde{F}), \tilde{F})$ to that of $ASE(h^*(\hat{F}), \hat{F})$ is reported in Table 2. In Table 3 we report the average of the ratios of $ASE(h^*(\tilde{F}),\ \tilde{F})$ to $ASE(h^*(\hat{F}),\ \hat{F})$ with the corresponding standard error in parentheses. The figures of Tables 2 and 3 show that our local linear estimator performs better than the kernel distribution estimator in all cases considered. In particular, Table 3 demonstrates clear gains of $\hat{F}$; the average ratios are all significantly greater than 1.

In Figure 1 we plot the empirical distribution $F_n(x)$, kernel distribution estimate $\tilde{F}(x)$, and local linear estimate $\hat{F}(x)$ based on one random sample from the Weibull(6,2) or Weibull(3,2) distribution. Observe that $\hat{F}(x)$ is better than $F_n(x)$ and $\tilde{F}(x)$ as $F(x)$ is away from zero and one. In another simulation, which is not reported here, of 500 samples of size 50 from the same two distributions, we also observed that $\hat{F}(x)$ has the smallest mean squared error among the three estimators for x that has $F(x)$ away from zero and one.

Table 2. Ratio of the empirical mean of $ASE(h^*(\tilde{F}),\ \tilde{F})$ to the empirical mean of $ASE(h^*(\hat{F}),\ \hat{F})$.

Distribution	$n = 10$	$n = 30$	$n = 50$	$n = 70$
Weibull(6,2)	1.371	1.267	1.199	1.156
Weibull(3,2)	1.032	1.092	1.090	1.088
Weibull(6,1)	1.659	1.812	1.933	1.838
Weibull(3,1)	1.133	1.390	1.434	1.392

Table 3. Average of the ratios of $ASE(h^*(\tilde{F}),\ \tilde{F})$ to $ASE(h^*(\hat{F}),\ \hat{F})$ with the corresponding standarderror in parentheses.

Distribution	$n = 10$	$n = 30$	$n = 50$	$n = 70$
Weibull(6,2)	2.438 (0.097)	1.635 (0.037)	1.382 (0.024)	1.213 (0.015)
Weibull(3,2)	1.690 (0.064)	1.300 (0.029)	1.185 (0.020)	1.110 (0.013)
Weibull(6,1)	2.100 (0.051)	2.240 (0.058)	2.605 (0.064)	2.792 (0.088)
Weibull(3,1)	1.448 (0.036)	1.877 (0.047)	2.195(0.079)	2.172(0.080)

4.2. Quantile estimation

Next we report results of a Monte Carlo study which was conducted to compare the performance of the local quadratic estimator $\hat{Q}(p)$ with the modified kernel quantile estimator $\bar{Q}_n^k(p)$ for a range of values of p. We generated 300 pseudo-random samples of size 100 from the exponential distribution with mean 1 and the Weibull(3,2) distribution. The Epanechnikov kernel $k(x) = \frac{3}{4}(1 - x^2)I(|x| \leq 1)$ was used. The MSE of the two estimators were calculated for $p = 0.05, 0.10$ and with values of the bandwidth ranging from 0.002 to 0.4.

Figures 2 and 3 show that local quadratic estimator $\hat{Q}(p)$ with its optimal bandwidth behaves better than the modified kernel estimator $\bar{Q}_n^k(p)$ with its own optimal bandwidth for smaller quantiles. Notice that the optimal bandwidths that achieved minimal mean squared errors of $\bar{Q}_n^k(p)$ are much smaller than those for $\hat{Q}(p)$. Thus, these $\bar{Q}_n^k(p)$ estimates would have very similar appearance to the empirical quantile estimate which is not smooth. By contrast, our estimator $\hat{Q}(p)$ does not have this problem. More importantly, $\hat{Q}(p)$ is much less sensitive to the bandwidth choice than $\bar{Q}_n^k(p)$. This is important if we use global bandwidth, for example, choosing $h^* = \text{argmin} \int_\alpha^{1-\alpha} |\hat{Q}(s) - Q(s)|^2\, ds$, where $\alpha \in (0, 1/2)$.

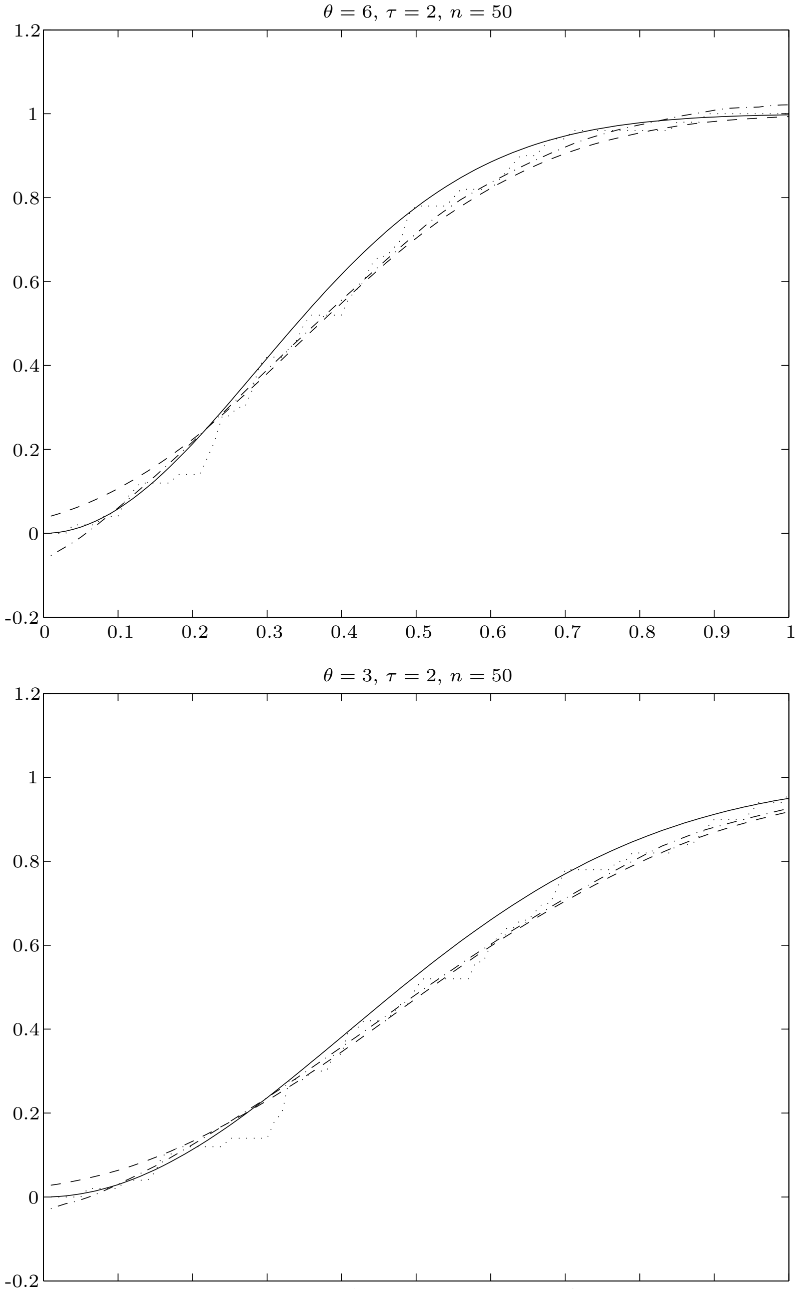
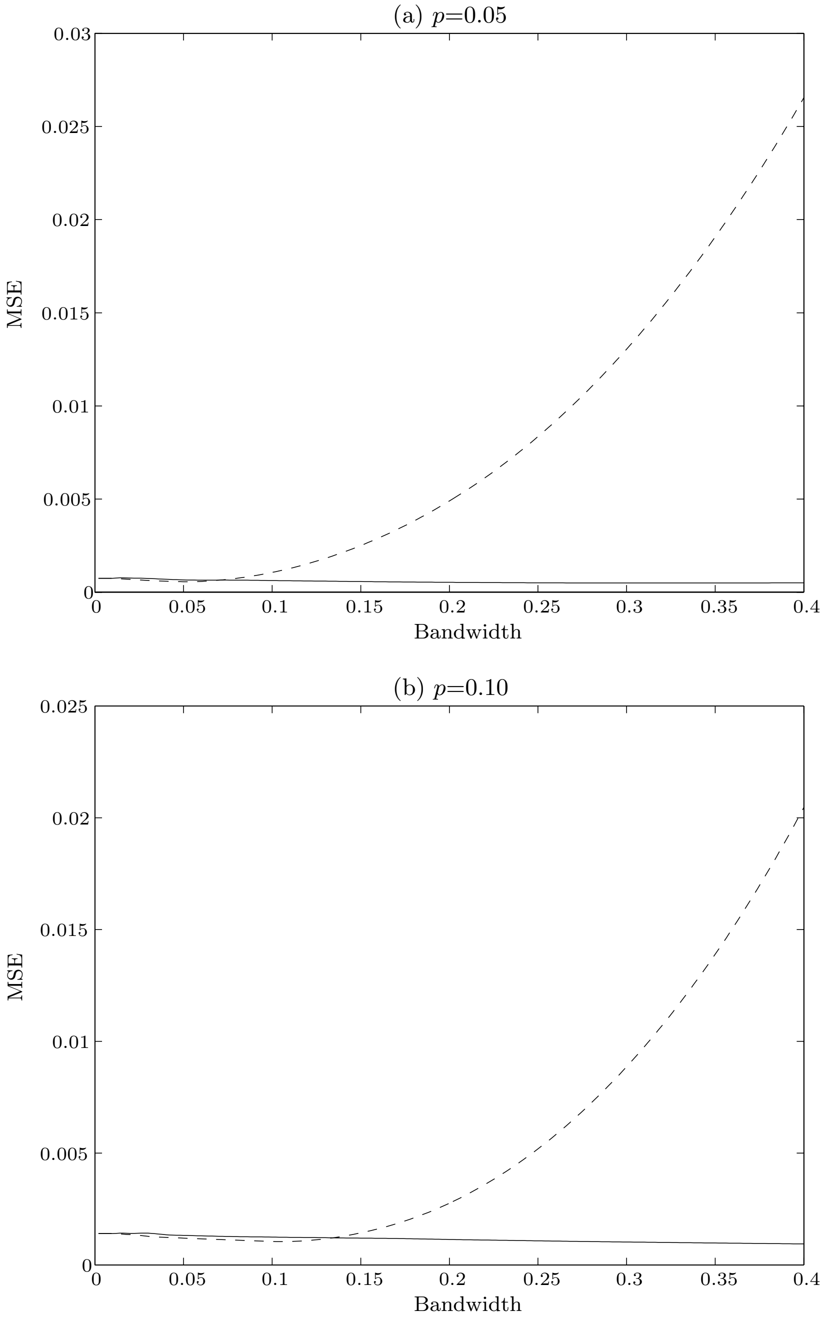


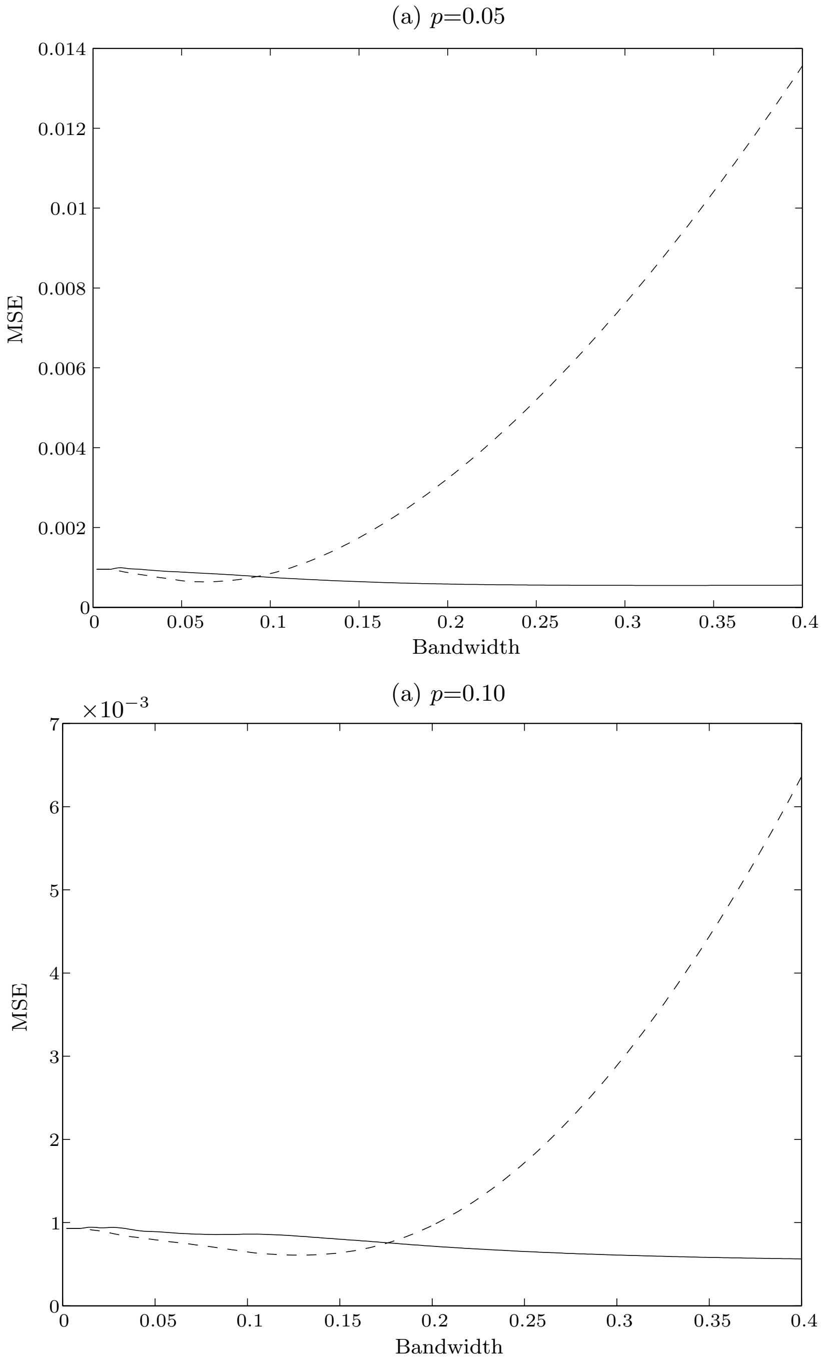
Figure 1. Distribution Function Estimation Based on One Sample. Solid line represents the true distribution, Weibull (6,2) (top panel) or Weibull (3,2) (bottom panel). Dashed line, dot-and-dash line, and dotted line represent the kernel distribution estimate $\tilde{F}(x)$, the local linear estimate $\hat{F}(x)$ and the empirical estimate $F_n(x)$, respectively, based on one sample of size 50.



(a) p=0.05

(b) p=0.10

Figure 2. Quantile Estimation for Exponential (1) distribution. Solid line and broken line plot the mean squared errors, against bandwidth h, for the local quadratic estimator $\hat{Q}(p)$ and the modified kernel quantile estimator $\bar{Q}_n^k(p)$, respectively. Sample size n is 100.



(a) p=0.05

(a) p=0.10

Figure 3. Quantile Estimation for Weibull (3,2) distribution. Solid line and broken line represent the mean squared errors for the local linear estimator $\hat{Q}(p)$ and the modified kernel quantile estimator $\bar{Q}_n^k(p)$, respectively. Sample size n is 100.

Acknowledgements

We thank the co-editor, an associate editor and a referee for their helpful comments.

Appendix

Proof of (2.5). In order to simplify the proof, we work on

$$\hat{F}(x)-F(x)=\frac{\sum_{j=1}^n w_jF_n(X_j)}{\sum_{j=1}^n w_j+n^{-q}}-F(x)=\frac{n^{-2}h^{-4}\sum_{j=1}^n w_j\left[F_n(X_j)-F(x)\right]}{n^{-2}h^{-4}\sum_{j=1}^n w_j+n^{-q}},$$

where $q>0$ is some large constant. This trick was used by Fan (1993) for analysis of the local linear regression estimator. First we have

$$(nh^3)^{-1}s_{n,1}\xrightarrow{a.s.}-f'(x)c_1,(nh^3)^{-1}s_{n,2}\xrightarrow{a.s.}f(x)c_1,(n^2h^4)^{-1}\sum_{i=1}^n w_i\xrightarrow{a.s.}f^2(x)c_1. \quad (4.1)$$

Let $U_j=F(X_j), j=1,\ldots,n$. Put $G_n(u)=\frac{1}{n}\sum_{i=1}^n I(U_i\le u)$ and $\alpha_n(u)=\sqrt{n}(G_n(u)-u)$. From Csörgő, Csörgő, Horváth and Mason (1986), there exists a sequence of Brownian bridges $B_n(u)$, $0\le u\le 1$, $n=1,\ldots$, such that

$$\limsup_{n\to\infty}\sup_{0\le u\le 1} n^{1/4}|\alpha_n(u)-B_n(u)|/((\log n)^{1/2}(\log\log n)^{1/4})=2^{-1/4}\quad \text{a.s.} \quad (4.2)$$

Because of the new version of $\hat{F}(x)$, we can treat equations (4.1) and (4.2) as true in the whole space instead of almost surely. Note that

$$\begin{aligned}
&(n^2h^4)^{-1}\sum_{j=1}^n w_j\left[F_n(X_j)-F(x)\right]\\
&= n^{-2}h^{-4}\sum_{j=1}^n k\left(\frac{x-X_j}{h}\right)\left[s_{n,2}-(x-X_j)s_{n,1}\right]\left[F_n(X_j)-F(x)\right]\\
&= (nh)^{-1}\sum_{j=1}^n k\Big(\frac{x-X_j}{h}\Big)\Big[n^{-1}h^{-3}s_{n,2}-f(x)c_1\Big]\left[F_n(X_j)-F(x)\right]\\
&\quad -(nh)^{-1}\sum_{j=1}^n (x-X_j)k\Big(\frac{x-X_j}{h}\Big)\Big[n^{-1}h^{-3}s_{n,1}+f'(x)c_1\Big]\left[F_n(X_j)-F(x)\right]\\
&\quad +(nh)^{-1}c_1\sum_{j=1}^n k\Big(\frac{x-X_j}{h}\Big)[f(x)+(x-X_j)f'(x)][F_n(X_j)-F(x)]\\
&:= I_1+I_2+I_3.
\end{aligned}$$

Write

$$
\begin{aligned}
I_3 &= h^{-1}c_1\int_{-\infty}^{\infty}[F_n(s)-F(x)]k\Big(\frac{x-s}{h}\Big)[f(x)+(x-s)f'(x)]\,dF_n(s)\\
&=(2h)^{-1}c_1\int_{-1}^{1}k(y)[f(x)+hyf'(x)]\,d[F_n(x-yh)-F(x)]^2\\
&=(2h)^{-1}c_1\int_{-1}^{1}[F_n(x-yh)-F(x)]^2[k'(y)f(x)+yk'(y)hf'(x)+k(y)hf'(x)]\,dy\\
&=(2h)^{-1}c_1\int_{-1}^{1}[G_n(F(x-yh))-F(x)]^2g(y)\,dy\\
&=(2nh)^{-1}c_1\int_{-1}^{1}[\alpha_n(F(x-yh))-B_n(F(x-yh))]^2g(y)\,dy\\
&\quad+(nh)^{-1}c_1\int_{-1}^{1}[\alpha_n(F(x-yh))-B_n(F(x-yh))]\\
&\qquad\qquad\times[B_n(F(x-yh))+n^{1/2}(F(x-yh)-F(x))]g(y)\,dy\\
&\quad+(2nh)^{-1}c_1\int_{-1}^{1}B_n^2(F(x-yh))g(y)\,dy\\
&\quad+(n^{1/2}h)^{-1}c_1\int_{-1}^{1}B_n(F(x-yh))[F(x-yh)-F(x)]g(y)\,dy\\
&\quad+(2h)^{-1}c_1\int_{-1}^{1}[F(x-yh)-F(x)]^2g(y)\,dy\\
&=II_1+\cdots+II_5,
\end{aligned}
$$

where $g(y)=k'(y)f(x)+yk'(y)hf'(x)+k(y)hf'(x)$. By (4.2) we have

$$
II_1+II_2=O(n^{-3/2}h^{-1}(\log n)^2+n^{-5/4}h^{-1}\log n+n^{-3/4}\log n),
$$

$$
E(II_3^2)+E(II_3II_4)=O(n^{-2}h^{-1}+n^{-3/2}) \quad\text{and}\quad II_5=2^{-1}h^2c_1^2f^2(x)f'(x)+O(h^3).
$$

Furthermore,

$$
\begin{aligned}
E(II_4^2)=c_1^2n^{-1}h^{-2}&\int_{-1}^{1}\int_{-1}^{1}E\,[B_n(F(x-y_1h))B_n(F(x-y_2h))]\,[F(x-y_1h)-F(x)]\\
&\times[F(x-y_2h)-F(x)]g(y_1)g(y_2)\,dy_1dy_2\\
&=c_1^2n^{-1}f^4(x)[F(x)-F^2(x)]+c_1^2c_3n^{-1}hf^5(x)-4c_1^2c_2n^{-1}hf^5(x)+O(n^{-1}h^2),
\end{aligned}
$$

where the last equality follows from moments of Brownian bridges, Taylor expansions and integration by parts. Thus we may show that

$$
\begin{aligned}
E\Big(\frac{I_3}{n^{-2}h^{-4}\sum_{j=1}^{n}w_j}\Big)^2 &= b_n^2(x)+\sigma_n^2(x)+O\left(n^{-2}h^{-2}n^{-1/2}(\log n)^2\right)\\
&\quad+O\left(n^{-2}h^{-1})+O(n^{-3/2}(\log n)^2\right)+o\left(h^4\right)+o\left(n^{-1}h\right),
\end{aligned}
$$

where $b_n^2(x) = 4^{-1}(f'(x))^2 c_1^2 h^4$ and $\sigma_n^2(x) = (F(x) - F^2(x))n^{-1} - 4f(x)c_2 n^{-1}h + f(x)c_3 n^{-1}h$. The above result and the fact that $E\,(I_1+I_2+I_3)^2 = E(I_3^2)\,(1+o(1))$ yield $E(\hat{F}(x) - F(x))^2 = \left(b_n^2(x) + \sigma_n^2(x)\right)(1+o(1))$. Further calculations lead to (2.5).

Proof of (3.3). Let $U_i = F(X_i), i = 1, \ldots$. Csörgő et al. (1986) have constructed a probability space carrying $U_1, U_2, \ldots$ and a sequence of Brownian bridges $B_n(s), 0 \le s \le 1$, $n = 1, \ldots$, such that, for the quantile process $\beta_n(s) = n^{1/2}\{s - U_n(s)\}$, $0 \le s \le 1$, and

$$U_n(s) = \begin{cases} U_{n,k}, & \text{if} \quad (k-1)/n < s \le k/n, k = 1, \ldots, n \\ U_{n,1}, & \text{if} \quad s = 0, \end{cases}$$

with $U_{n,1} \le \cdots \le U_{n,n}$ denoting the order statistics of $U_1, \ldots, U_n$, we have

$$\begin{cases} \sup\limits_{0 \le s \le 1} n^{1/2}|\beta_n(s) - B_n(s)| = O(\log n), \quad \text{a.s.} \\ \sup\limits_{0 \le s \le 1} |\beta_n(s)| = O(\log n) \quad \text{a.s.} \end{cases} \tag{4.3}$$

Since $EX_1^2 < \infty$ we can treat (4.3) as true in the whole space instead of a.s. Write

$$\begin{aligned} &(a_0a_4 - a_2^2)\left[\hat{Q}(p) - Q(p)\right] \\ &= n^{-1/2}\int_0^1 [-a_4 + a_2(p-s)^2]k(\frac{p-s}{h})Q'(s)[\beta_n(s) - B_n(s)][1 + O(n^{-1/2}\log n)]\,ds \\ &\quad + n^{-1/2}\int_0^1 [-a_4 + a_2(p-s)^2]k(\frac{p-s}{h})Q'(s)B_n(s)[1 + O(n^{-1/2}\log n)]ds \\ &\quad + \int_0^1 [a_4 - a_2(p-s)^2]k(\frac{p-s}{h})[Q(s) - Q(p)]ds \\ &= I_1 + I_2 + I_3. \end{aligned}$$

One can check that

$$|I_1| = O\left(h^6 n^{-1}\log n\right), |I_3| = \frac{1}{24}Q^{(4)}(p)(a_0a_4 - a_2^2)h^4\int_{-1}^1 s^4 k_2(s)ds + o(h^{10}),$$

$$\begin{aligned} E\left\{I_2^2\right\} &= n^{-1}[1 + O(n^{-1/2}\log n)][Q'(p)]^2\Big\{(a_0a_4 - a_2^2)^2(p - p^2) + o(h^3) \\ &\quad -2h^3\int_{-1}^1\int_{-1}^{y_1} [a_4^2 y_1 + a_2^2 h^4 y_1^3 y_2^2 - a_2a_4h^2(y_1y_2^2 + y_1^3)]k(y_1)k(y_2)dy_2dy_1\Big\}. \end{aligned}$$

Hence (3.3) follows from the above results and the identity

$$2(a_0a_4 - a_2^2)^2\sigma_1^2 = 2h^2\int_{-1}^1\int_{-1}^{y_1}(a_4y_1 - a_2h^2y_1^3)k(y_1)(a_4 - a_2h^2y_2^2)k(y_2)dy_2dy_1.$$

References

Altman, N. and Léger, C. (1995). Bandwidth selection for kernel distribution function estimation. *J. Statist. Plann. Inference* **46**, 195-214.

Bowman, A., Hall, P. and Prvan, T. (1998). Bandwidth selection for the smoothing of distribution functions. *Biometrika* **85**, 799-808.

Cheng, M.-Y., Fan, J. and Marron, J. S. (1997). On automatic boundary corrections. *Ann. Statist.* **25**, 1691-1708.

Cheng, C. and Parzen, E. (1997). Unified estimators of smooth quantile and quantile density functions. *J. Statist. Plann. Inference* **59**, 291-307.

Csörgő, M., Csörgő, S., Horváth, L. and Mason, D. M. (1986). Weighted empirical and quantile processes. *Ann. Probab.* **14**, 31-85.

Falk, M. (1983). Relative efficiency and deficiency of kernel type estimators of smooth distribution functions. *Statist. Neerlandica* **37**, 73-83.

Falk, M. (1984). Relative deficiency of kernel type estimators of quantiles. *Ann. Statist.* **12**, 261-268.

Falk, M. (1985). Asymptotic normality of kernel type estimators of quantiles. *Ann. Statist.* **13**, 428-433.

Fan, J. (1992). Design-adaptive nonparametric regression. *J. Amer. Statist. Assoc.* **87**, 998-1004.

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

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications.* Chapman and Hall, London.

Lejeune, M. and Sarda, P. (1992). Smooth estimators of distribution and density functions. *Comput. Statist. Data Anal.* **14**, 457-471.

Garcia-Soidan, P. H., Gonzalez-Manteiga, W. and Prada-Sanchez, J. M. (1997). Edgeworth expansions for nonparametric distribution estimation with applications. *J. Statist. Plann. Inference* **65**, 213-231.

Mammen, E., Marron, J. S., Turlach, B. A. and Wand, M. P. (1999) A general projection framework for constrained smoothing. *Statist. Sci.* **16**, 232-248.

Nadaraya, E. A. (1964a). Some new estimators for distribution functions. *Theory Probab. Appl.* **9**, 497-500.

Nadaraya, E. A. (1964b). On estimating regression. *Theory Probab. Appl.* **9**, 141-142.

Parzen, E. (1979). Nonparametric statistical data modeling. *J. Amer. Statist. Assoc.* **74**, 105-131.

Reiss, R.-D. (1981). Nonparametric estimation of smooth distribution functions. *Scand. J. Statist.* **8**, 116-119.

Sheather, S. J. and Marron, J. S. (1990). Kernel quantile estimators. *J. Amer. Statist. Assoc.* **85**, 410-416.

Wei, C. Z. and Chu, C. K. (1994). A regression point of view toward density estimation. *J. Nonparametr Statist.* **4**, 191-201.

Yamato, H. (1973). Uniform convergence of an estimator of a distribution function. *Bull. Math. Statist.* **15**, 69-78.

Yang, S. S. (1985). A smooth nonparametric estimator of a quantile function. *J. Amer. Statist. Assoc.* **80**, 1004-1011.

Zelterman, D. (1990). Smooth nonparametric estimation of the quantile function. *J. Statist. Plann. Inference* **26**, 339-352.

Zou, K. H., Hall, W. J. and Shapiro, D. E. (1997). Smooth non-parametric receiver operating characteristic (ROC) curves for continuous diagnostic tests. *Statist. Medicine* **16**, 2143-2156.

Department of Mathematics, National Taiwan University, Taipei 106, Taiwan.

E-mail: cheng@math.ntu.edu.tw

School of Mathematics, Georgia Institute of Technology, Atlanta GA 30332-0160, U.S.A

E-mail: peng@math.gatech.edu

(Received March 2001; accepted March 2002)