Statistica Sinica **6**(1996), 79-95

ESTIMATION OF JUMP POINTS AND JUMP VALUES OF A DENSITY FUNCTION

C. K. Chu and P. E. Cheng

National Tsing Hua University and Academia Sinica

Abstract. The estimation of locations of jump points and corresponding jump sizes of a density function on a bounded interval of interest by the kernel method is considered. Strong convergence rates (SCR) and limiting distributions for the proposed estimators are obtained. The order of the SCR for estimators of locations for jump points is immune to the smoothness conditions imposed on the density function, but that for estimators of jump sizes is not. The limiting distributions are used to test the continuity of the density function and give asymptotic confidence intervals for locations of jump points and corresponding jump sizes. For applications of these estimators, the choices of bandwidths and kernel functions are considered. In the case that the number of jump points on a bounded interval of interest is known in advance, an approach is proposed to recover the density function on the interval such that the performance of the resulting density function estimate is not affected by these jump points. Simulations demonstrate that the asymptotic results hold for reasonable sample sizes.

Key words and phrases: Asymptotic normality, density estimation, jump point, jump size, kernel estimator, strong consistency.

1. Introduction

The problem of detecting and measuring discontinuity of an unknown density function arises in many statistical applications in science and technology, such as image processing, pattern recognition, tomography, etc. In X-ray transmission tomography, Johnstone and Silverman (1990) investigate discontinuities, in the form of sharp jumps, in the tissue density across the boundaries of various regions. It is noted in their work and the references cited therein that a difficulty with such practical examples is the assessment of the adequacy of a mathematical model for the physical process encountered. In essence, discontinuity detection together with density function estimation aims at an approximate image construction in practice.

So far in the statistics literature, most theoretical studies have considered analyzing discontinuity in the 1-dimensional marginal densities as a basic starting point. In this connection, the method of nonparametric kernel density estimation (cf. Silverman (1986) and Härdle (1990)) is a reasonably useful tool for exploiting the density structure, and is often simpler and more effective than using a seemingly correct parametric modeling. When discontinuities are involved, their effects on the asymptotic behavior of the kernel density estimator and, in particular, on an optimally chosen bandwidth have been studied by van Eeden (1985), Cline and Hart (1991), and van Es (1992). Rigorous theoretical analyses, by kernel density estimation, for locations and jump sizes of the discontinuities are yet unavailable. Lee (1990) has proposed algorithms for estimating an input signal density together with finding the discontinuity locations. In the related field of nonparametric regression, the problem of estimating the regression curve admitting jump discontinuities has been discussed by Yin (1988), Qiu et al. (1991), Müller (1992), and Wu and Chu (1993), but it is different from the problem of estimating the image intensity. The method considered here for the latter problem is similar to those applied for the former.

The limited goal of this study is to demonstrate that kernel density estimation can be skillfully employed in analyzing locations and jump sizes of the discontinuities of a 1-dimensional density. It is plausible that an extension of our method disigned in Section 2 would provide us with a rough idea of where the discontinuity edges or curves in a 2-dimensional image density are, although we feel that a carefully designed semiparametric approach could generate better insights and results in practice; the latter still demands further research (cf. Silverman et al. (1990)).

The organization of this paper is as follows. Section 2 describes the motivation and the precise formulation of the proposed estimators of locations of jump points and corresponding jump sizes of a density function. Section 3 gives strong convergence rates (SCR) and limiting distributions for these estimators. The order of the SCR for estimators of locations of jump points can be made close to n^{-1}. It is immune to the smoothness conditions imposed on the density function, but that for estimators of jump sizes is not. The limiting distributions are used to test the continuity of the density function and give asymptotic confidence intervals for locations of jump points and corresponding jump sizes of the density function on a bounded interval of interest. For applications of the proposed estimators, the choices of bandwidths and kernel functions are considered. In the case that the number of jump points on a bounded interval of interest is known in advance, an approach is proposed to recover the density function on the interval. The performance of the resulting density function estimate is not affected by these jump points. Section 4 contains simulation studies which support our theoretical findings. Finally, sketches of the proofs are given in Section 5.

2. The Proposed Estimators

Suppose without loss of generality that the density function f defined on the real line R has a few points of discontinuity on the bounded interval say, $[0,1]$. We are interested in estimating locations and jump sizes of the jump points in $[0,1]$ based on a random sample $X_1,\ldots,X_n$ from f. Note that f might have points of discontinuity outside of $[0,1]$, but we are not interested in detecting these jump points.

Let the density function $f(x)$, for $x\in[0,1]$, be expressed by

$$f(x)=\tau(x)+\psi(x). \tag{2.1}$$

Here the function $\tau(x)$ is defined and Lipschitz continuous on $[0,1]$, and $\psi(x)$ is defined by $\psi(x)=\sum_{j=1}^{p}d_jI_{[t_j,\infty)}(x)$, for $x\in[0,1]$. Note that p is a nonnegative integer representing the number of jump points of f on $[0,1]$, t_j are locations of jump points, $t_j\in[\xi,1-\xi]$, and d_j are nonzero real numbers representing jump sizes of f at t_j. Here ξ is an arbitrarily small positive constant. If $p=0$, then f is continuous. For simplicity of presentation, let $d_{p+1}=0$ and $|d_j|>|d_{j+1}|$, for $j=1,2,\ldots,p$. Also, assume that the distance between any two of these t_j is greater than ξ.

To construct the estimators $\hat{t}_j$ and $\hat{d}_j$ of t_j and d_j, respectively, for $j=1,2,\ldots,\rho$, the kernel density estimator is considered. Here ρ is a given positive integer since the true value of p is unknown. Given the bandwidth h and the kernel function K as a Lipschitz continuous probability density function supported on the interval $[-1,1]$, the kernel density estimator $\hat{f}(x)$ for $f(x)$ is given by

$$\hat{f}(x)=n^{-1}\sum_{i=1}^{n}K_h(x-X_i), \tag{2.2}$$

for $x\in R$, where $K_h(\cdot)=h^{-1}K(\cdot/h)$. The rest of this section is devoted to formulating the proposed estimators $\hat{t}_j$ and $\hat{d}_j$, for $j=1,2,\ldots,\rho$.

To estimate t_j, we first consider the magnitude of $\hat{f}(x)$, for $x\in[0,1]$. By (2.1) and (2.2), through a straightforward calculation, we have

$$\hat{f}(x)=E[\hat{f}(x)]+\hat{f}^*(x),$$

for $x\in[0,1]$, where

$$E[\hat{f}(x)]=\tau(x)+\sum_{j=1}^{p}d_j\int_{-1}^{(x-t_j)/h}K+O(h),$$

$$\hat{f}^*(x)=\hat{f}(x)-E[\hat{f}(x)].$$

Lemma 2.1.2 of Prakasa Rao (1983) shows that $\hat{f}^*(x)$ converges to 0 uniformly with probability one, in both cases of $p = 0$ and $p > 0$. This implies that $\hat{f}^*(x)$ is of small order in magnitude, for all $x \in [0,1]$, asymptotically. Thus, the effect of jump points on the magnitude of $\hat{f}(x)$ will only appear in the value of $E[\hat{f}(x)]$. Also, to consider the magnitude of $\hat{f}(x)$, it is enough to consider that of $E[\hat{f}(x)]$. Accordingly, to discover t_j by kernel density estimation, we construct a function $J(x)$ defined by

$$J(x) = \hat{f}_1(x) - \hat{f}_2(x),$$

for $x \in [0,1]$. Here $\hat{f}_1(x)$ and $\hat{f}_2(x)$ are kernel density estimators for $f(x)$ with different kernel functions K_1 and K_2, respectively, and the same bandwidth h.

From the above results, we find that the magnitude of $|J(x)|$ can be asymptotically expressed by

$$\begin{aligned} |J(x)| &= |E[\hat{f}_1(x) - \hat{f}_2(x)]| + O_p(n^{-1/2}h^{-1/2}) \\ &= \Big|\sum_{j=1}^{p} d_j \int_{-1}^{(x-t_j)/h} (K_1 - K_2)\Big| + O(h) + O_p(n^{-1/2}h^{-1/2}), \end{aligned} \tag{2.3}$$

for $x \in [0,1]$. To find locations of jump points t_j, some basic conditions imposed on K_1 and K_2 would be beneficial. Let K_1 and K_2 satisfy $K_2(z) = K_1(-z)$, $\int K_1 = \int K_2 = 1$, and $\int_{-1}^{0}(K_1 - K_2) \neq 0$. Given these K_1 and K_2, through a straightforward calculation, two results for the asymptotic value of $|J(x)|$ easily follow. It is symmetric about t_j and convex downward on some neighborhood of t_j, for each $j = 1, 2, \ldots, p$. The widths of these neighborhoods of t_j correspond to those of the intervals where K_{1h} and K_{2h} are supported. On the other hand, it is zero outside of the union of these neighborhoods of t_j.

Based on the above characteristics of the magnitude of $|J(x)|$, we propose to take local maximizers of $|J(x)|$ as estimators of locations of jump points. Since K_1 and K_2 are supported on $[-1,1]$, the widths of the above neighborhoods of t_j are not greater than $2h$. Combining this result with the fact that $|d_j| > |d_{j+1}|$, for $j = 1, 2, \ldots, p$, we take $\hat{t}_j$ as maximizers of $|J(x)|$ over the sets A_j, where

$$A_j = [0,1] - \bigcup_{k=1}^{j-1} [\hat{t}_k - 2h,\, \hat{t}_k + 2h],$$

for $j = 1, 2, \ldots, \rho$.

We now give the formulation of $\hat{d}_j$. To estimate d_j, based on the above $\hat{t}_j$, a direct method is to take the rescaled $J(\hat{t}_j)$ as $\hat{d}_j$, for $j = 1, 2, \ldots, \rho$. By (2.3), the scale factor c_J for $J(\hat{t}_j)$ is $c_J = [\int_{-1}^{0}(K_1 - K_2)]^{-1}$. However, there is a drawback to this simple approach. To address this drawback, consider the case $\hat{t}_j = t_j + \alpha h$, for some j and $\alpha \neq 0$. Then, based on the above arguments, the value of $\hat{d}_j$ is asymptotically equal to $c_s d_j \int_{-1}^{\alpha}(K_1 - K_2)$ which is not equal to d_j for $0 < |\alpha| < 1$ and is equal to 0 for $|\alpha| \geq 1$. By this, the drawback to the simple approach is that, even though $\hat{t}_j$ is close to t_j, d_j can not be estimated well.

To address the above drawback to the simple approach, we propose taking the rescaled $S(\hat{t}_j)$ as $\hat{d}_j$, for $j = 1, 2, \ldots, \rho$. The function $S(x)$ is defined by

$$S(x) = \hat{f}_3(x) - \hat{f}_4(x),$$

for $x \in [0, 1]$. Here $\hat{f}_3(x)$ and $\hat{f}_4(x)$ are kernel density estimators for $f(x)$ with different kernel functions K_3 and K_4, respectively, and the same bandwidth g. Note that K_3 and K_4 satisfy the above conditions given on K_1 and K_2, respectively, and the value of g is of larger order than that of h. Based on the above arguments, the scale factor c_s for $S(\hat{t}_j)$ is taken as $c_s = [\int_{-1}^{0}(K_3 - K_4)]^{-1}$.

We now give the effect of the order of magnitude of g on the performance of the proposed $\hat{d}_j$. Consider the above case $\hat{t}_j = t_j + \alpha h$, for some j and $\alpha \neq 0$. Following the same arguments, the value of $\hat{d}_j$ is roughly equal to $c_s d_j \int_{-1}^{\alpha h/g}(K_3 - K_4)$ which approaches d_j as h/g approaches 0, for each $\alpha \neq 0$. Combining this result with the fact that the value of g is of larger order than that of h, the proposed $\hat{d}_j$ does not have the above drawback.

Finally, the asymptotic behaviors of the proposed estimators $\hat{t}_j$ and $\hat{d}_j$ of t_j and d_j, respectively, will be studied in Section 3. For applications of the proposed estimators, the choices of bandwidths and kernel functions will be considered in Remark 2 of Section 3.

3. Results

In this section, we shall study the asymptotic behaviors of $\hat{t}_j$ and $\hat{d}_j$, for $j = 1, 2, \ldots, \rho$. For these, we impose the following assumptions:

(A.0) $X_1, \ldots, X_n$ are independence random variables with density function $f(x)$ as given in (2.1).

(A.1) The qth derivative $\tau^{(q)}$ of τ in (2.1) is Lipschitz continuous on the interval $[0, 1]$, where $q \geq 2$. Here and throughout this paper, the notation $m^{(j)}$ denotes the jth derivative of the given function m, for some integer $j \geq 0$.

(A.2) The kernel function K_1 is supported on the interval $[-1, \lambda]$, $\lambda \in [0, 1]$, and of order q. Recall that a kernel function G is said to be of order q if it satisfies $\int G = 1$, $\int z^\ell G = 0$, for $1 \leq \ell < q$, and $\int z^q G \neq 0$. Also, $K_1^{(1)}$ is Lipschitz continuous, $K_1^{(1)}(0) \neq 0$, and $K_1^{(\ell)}(-1) = K_1^{(\ell)}(\lambda) = 0$, for $\ell = 0, 1$. The kernel function K_2 is defined by $K_2(z) = K_1(-z)$, for all z. Finally, $\int_{-1}^{0}(K_1 - K_2) \neq 0$ and there is a constant $\eta > 0$ such that $|\int_0^{c_n}(K_1 - K_2)| > \eta c_n^\kappa$, for some $\kappa \geq 2$ and any sequence c_n of positive real numbers converging to 0 as $n \to \infty$.

(A.3) The kernel function K_3 is supported on the interval $[-1, \omega]$, $\omega \in [0,1]$, Lipschitz continuous, and of order q. The kernel function K_4 is defined by $K_4(z) = K_3(-z)$, for all z. The kernel functions K_3 and K_4 satisfy $\int_{-1}^{0}(K_3 - K_4) \neq 0$.

(A.4) The total number of observations in this density estimation setting is n, with $n \to \infty$. The bandwidths $h = h_n$ and $g = g_n$ satisfy $h \to 0$ with $nh \to \infty$, and $g \to 0$ with $ng \to \infty$, as $n \to \infty$.

Theorems 1 and 2 below will give the asymptotic behaviors of $\hat{t}_j$ and $\hat{d}_j$ in the cases $p \geq \rho \geq 1$ and $\rho > p \geq 0$, respectively. The proofs of these theorems are given in Section 5. To state these theorems, we need the following notation. Let $\beta = 2[q/2] + 1$, $\Lambda_j = ((n/h)^{1/2}(\hat{t}_j - t_j), (ng)^{1/2}(\hat{d}_j - d_j))^T$, for $j = 1, 2, \ldots, \rho$, θ a positive constant, where $\theta \in (0, \beta/\kappa)$, and δ an arbitrarily small positive constant. Here the notation $[x]$ denotes the largest integer which is smaller than x, and T the transpose of a vector. In these theorems, some conditions on the values of n, h, g, θ, and δ include:

$$n^{-1+\delta}h^{-1-2\kappa\theta} = o(1), \tag{B.1}$$

$$n^{1-\delta}g^{-1}h^{2+2\theta} = o(1), \tag{B.2}$$

$$n^{1-\delta}g^{1+2\beta} = o(1), \tag{B.3}$$

$$n\ h^{1+2\beta} = o(1), \tag{B.4}$$

$$n\ g^{-1}h^{2+2\theta} = o(1), \tag{B.5}$$

$$n\ g^{1+2\beta} = o(1), \tag{B.6}$$

$$h\ g^{-1} = o(1). \tag{B.7}$$

Theorem 1. *In the case $p \geq \rho \geq 1$, under the above assumptions* (A.0) *through* (A.4), *if* (B.1) *holds, then*

$$P(|\hat{t}_j - t_j| > h^{1+\theta} \quad i.o.) = 0, \tag{3.1}$$

for $j = 1, 2, \ldots, \rho$. Also, if (B.1) *through* (B.3) *hold, then*

$$n^{(1/2)-\delta}g^{1/2}|\hat{d}_j - d_j| \to 0 \quad a.s., \tag{3.2}$$

for $j = 1, 2, \ldots, \rho$. Furthermore, if (B.1) *and* (B.4) *through* (B.7) *hold, then*

$$\Lambda_j \Rightarrow N\Big((0,0)^T, \begin{bmatrix} d_j^{-2}F_jU & 0 \\ 0 & F_jV \end{bmatrix}\Big), \tag{3.3}$$

for $j = 1, 2, \ldots, \rho$, and these Λ_j are asymptotically independent, where

$$F_j = (1/2)[f(t_j^-) + f(t_j^+)],$$

$$U = [\int (K_1^{(1)} - K_2^{(1)})^2] \,/\, [2K_1^{(1)}(0)]^2,$$
$$V = [\int (K_3 - K_4)^2] \,/\, [\int_{-1}^{0} (K_3 - K_4)]^2.$$

Theorem 2. *In the case $\rho > p \geq 0$, under the above assumptions* (A.0) *through* (A.4)*, if* (B.1) *through* (B.3) *hold, then*

$$n^{(1/2)-\delta} g^{1/2} |\hat{d}_j - d_j| \to 0 \quad a.s. \tag{3.4}$$

for $j = 1, 2, \ldots, \rho$. Here $d_j = 0$, for $j > p$.

The following Theorem 3 will give a confidence band for $|J(x)|$. Using this confidence band, the test of the null hypothesis $H_0 : p = 0$ against the alternative hypothesis $H_1 : p > 0$ on the interval $[a, b]$ can be performed. Theorem 3 is obtained directly from (1.2) of Bickel and Rosenblatt (1973). Hence its proof is omitted.

Theorem 3. *Under the above assumptions* (A.0) *through* (A.4)*, if $f^{(q)}$ is Lipschitz continuous on the interval $[a, b]$ and $h = n^{-\gamma}$, where $\gamma \in (1/3, 1)$ for $q = 0$ and $\gamma \in ((2q+1)^{-1}, 1)$ for $q \geq 1$, then*

$$P(\sup_{z \in [a,b]} |J(z)| < a_n + b_n x) \to \exp(-2\exp(-x)),$$

where

$$a_n = w_n \Big[h_{ab} + h_{ab}^{-1} \log(\,(2\pi)^{-1} (\int (K_1^{(1)} - K_2^{(1)})^2 / \int (K_1 - K_2)^2)^{1/2})\Big],$$
$$b_n = w_n h_{ab}^{-1},$$

and

$$w_n = \Big[n^{-1} h^{-1} \hat{f}_1(x) \int (K_1 - K_2)^2\Big]^{1/2} \Big[\int_{-1}^{0} (K_1 - K_2)\Big]^{-1},$$
$$h_{ab} = [2 \log((b-a)/h)]^{1/2}.$$

We now close this section by the following remarks.

Remark 1. By (3.1) and (B.1), if $\kappa = 2$, $\theta = \delta$, and $h = \alpha_1 n^{-1+5\delta}$, $\alpha_1 > 0$, then the order of the SCR for $\hat{t}_j$ is $n^{-1+4\delta}$. This order of the SCR for $\hat{t}_j$ is independent of the value of q. Hence the smoothness condition imposed on the continuous part τ of f does not effect this order of the SCR for $\hat{t}_j$. On the other hand, by (3.2) and (B.1) through (B.3), if $g = \alpha_2 n^{(-1+\delta)/(1+2\beta)}$, $\alpha_2 > 0$, and the above values of κ, θ, and h are given, then the order of the SCR for $\hat{d}_j$ is $n^{-\beta/(1+2\beta)+\delta}$.

Since the value of β in (B.3) depends on q, the smoothness condition imposed on τ has different effects on the order of the SCR for $\hat{d}_j$.

Remark 2. In the case $p \geq \rho \geq 1$, to estimate t_j and d_j, for $j = 1, 2, \ldots, \rho$, a possible approach for practical choice of bandwidths and kernel functions is now given. Van Es (1992) has shown that the magnitude of the least squares cross-validated bandwidth produced in the case $p > 0$ is of order $n^{-1/2}$. By virtue of this, to estimate t_j, we suggest taking h as the least squares cross-validated bandwidth. Theoretically, given this value of h, $\kappa = 2$, and $\theta = (1/4) - \delta$, then the SCR of $\hat{t}_j$ to t_j is of order $n^{(-5/8)+\delta}$. To choose K_1 and K_2, by (5.8), (A.2), and (B.1), $\hat{t}_j$ have asymptotic mean square errors $n^{-1} h d_j^{-2} F_j U(1 + o(1))$. According to this, we suggest taking K_1 and K_2 to satisfy the conditions given in (A.2) and minimize the value of U over the class of $(q+4)$-th degree polynomials. For example, in the case $q = 0$,

$$K_1(x) = (0.4857 - 3.8560x + 2.8262x^2 + 19.1631x^3 + 11.9952x^4) I_{[-1, 0.2012]}(x) \quad (3.5)$$

and $K_2(x) = K_1(-x)$, for all x. The reason for choosing K_1 in the class of $(q+4)$-th degree polynomials is that, by (A.2), it must satisfy the following $q+4$ conditions $\int K_1 = 1$, $\int z^\ell K_1 = 0$, for $1 \leq \ell \leq q-1$, and $K_1(-1) = K_1^{(1)}(-1) = K_1(\lambda) = K_1^{(1)}(\lambda) = 0$. The same remark applies to K_2. To estimate d_j, we suggest choosing g as the least squares cross-validated bandwidth produced from the data in some subinterval $[a, b]$ of $[0, 1]$ on which the hypothesis test given in Theorem 3 has been performed and the null hypothesis accepted. The magnitude of the resulting g is of order $n^{-1/(1+2q)}$. For this, see, for example, Hall and Marron (1987). Theoretically, given this value of g and the above values of h, κ, and θ chosen for estimation of t_j, the SCR of $\hat{d}_j$ to d_j is of order $n^{(-q/(1+2q))+\delta}$. This order is close to the optimal rate of uniformly strong consistency of $\hat{f}$ to f in the case $p = 0$ given in Härdle (1991). Unfortunately, we do not know how to optimally choose K_3 and K_4. Finally, the performance of $\hat{t}_j$ and $\hat{d}_j$ derived by this approach needs further study.

Remark 3. We now consider the estimation of the density function $f(x)$, for $x \in [0, 1]$, when the value of p in (2.1) is known in advance. For this, using the results of Theorem 1 and Remark 2 in Section 3, the loctions of jump points can be estimated accurately, in the sense of SCR. In this case, we propose to estimate the density function by a kernel density estimator on subintervals separated by these estimates of locations of jump points. To avoid boundary effects on the kernel density estimator, the boundary modification method given in Gasser and Müller (1979) is applied. Through a straightforward calculation, the performance of the resulting kernel density estimate is the same as that given in Härdle (1991) for the case that the density function has q Lipschitz continuous derivatives, in the sense of the mean integrated square error over the interval $[0,1]$.

Remark 4. We now consider an intuitively simple estimation idea compared to the approach proposed above. If K_1 and K_2 are taken as the uniform kernel functions $K_1(x) = I_{[-1,0]}(x)$ and $K_2(x) = I_{[0,1]}(x)$, then $\hat{t}_j$ are simply constructed by comparing the number of data points within an interval of length h to the left of a location and that within a similar interval to the right. In this case, (3.1) still holds. Note that these rectangular kernel functions K_1 and K_2 exhibit jump points at endpoints of their support. In general, kernel functions with jump points will lead to bad finite sample behaviors of kernel estimators (see for example, Section 2.1 of Härdle (1991)). By this, the resulting $|J(x)|$ might have more local maximums (or sparks) than that using smooth kernel functions. The former will more often produce incorrect estimates of t_j than the latter. Simulation results given in Section 5 demonstrate that such particular $\hat{t}_j$ are inferior to the ones using other proper choices of K_1 and K_2, in the sense of having larger minimum sample mean square error over h.

4. Simulations

To investigate the practical implications of the asymptotic results of the proposed estimators of locations of jump points presented in Section 3, an empirical study was carried out. The simulation settings were as follows. The sample size was $n = 100$. Three density functions with the same location of jump point $x = 0$ were considered. The density functions were $f_\sigma(x) = \phi_\sigma(x)I_{[x\leq 0]} + \phi_{0.5}(x)I_{[x>0]}$, where $\sigma = 2, 1$, and $2/3$. The corresponding jump sizes of $f_\sigma(x)$ at $x = 0$ are $d(2\pi)^{-1/2}$, where $d = 3/2, 1$, and $1/2$. Here $\phi_z(x)$ denotes the probability density function of Normal$(0, z^2)$. For each density function, 100 independent sets of observations X_i were generated. Two sets of the kernel functions K_1 and K_2 were used. The first were those given in (3.5). The second were the uniform kernel functions given in Remark 4. The resulting estimates of the location of the jump point are denoted by $\hat{t}_1$ and $\hat{t}_1^*$, respectively.

We now describe the calculation of $\hat{t}$. Here $\hat{t}$ stands for $\hat{t}_1$ and $\hat{t}_1^*$. For each data set, the location of the jump point was searched for on the interval $[-6, 2]$. This interval was chosen arbitrarily. For this, 6 values of h, $h = 0.03, 0.1, 0.2, \ldots, 0.5$, were chosen. For each data set and each value of h, the values of $|J(x)|$ were calculated on an equally spaced grid of 801 values on $[-6, 2]$. The maximizer $\hat{t}$ of $|J(x)|$ over $[-6, 2]$ was calculated. After evaluation on the grid, a one-step interpolation was done, with the result taken as $\hat{t}$.

Figure 1 shows $|J(x)|$ with the kernel functions K_1 and K_2 given in (3.5) and $h = 0.03$ (solid curve) and $h = 0.3$ (dashed curve) derived from the simulated data from f_σ with $\sigma = 2$ (stars at the bottom). Given a small value of $h = 0.03$, the maximizer of $|J(x)|$ over the interval $[-6, 2]$ shows the location of jump point incorrectly. On the other hand, increasing the value of h as $h = 0.3$, the maximizer of $|J(x)|$ over $[-6, 2]$ shows the location of the jump point $x = 0$ accurately in this example. Based on this simulated data, we might consider that the underlying density has a peak to the right of 0, a bump in $[-2, -1]$, a long-left tail, and a short-right tail. It is difficult to distinguish visually from the simulated data alone that the underlying density function has a jump point at $x = 0$.

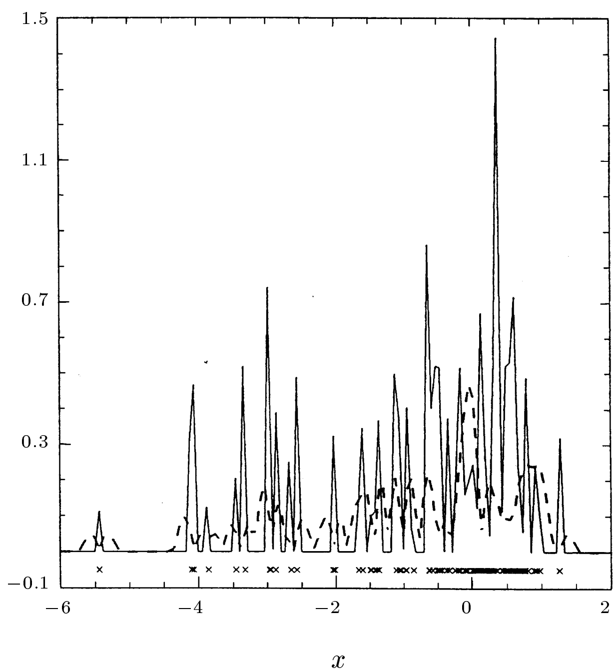

Figure 1. Plot of $|J(x)|$ with $h = 0.03$ (solid curve) and $h = 0.3$ (dashed curve) derived form the simulated data set (stars at the bottom).

For each density function $f_\sigma(x)$, Table 1 gives the sample bias, standard deviation, and mean square error of $\hat{t}_1$ and those of $\hat{t}_1^*$ in parentheses. For each $f_\sigma(x)$, when h increased from 0.03 to 0.4, there is a tendency that $\hat{t}_1$ moved from the right to the left. But $\hat{t}_1^*$ does not show this tendency. For each $f_\sigma(x)$, Table 1 also shows the minimum absolute sample bias, minimum sample standard deviation, and minimum sample mean square error of $\hat{t}_1$ and those of $\hat{t}_1^*$ in the last row. These values show the performance of the estimators with the ideal choice of the optimal bandwidth. It is clear that $\hat{t}_1$ has smaller minimum sample mean square error over h than $\hat{t}_1^*$, for each $f_\sigma(x)$.

Table 1. The sample bias, standard deviation (SD), and mean square error (MSE) of $\hat{t}_1$ and those (given in parentheses) of $\hat{t}_1^*$.

$f_\sigma(x)$	h value	Bias	SD	MSE
$\sigma = 2$				
	0.03	0.129(0.064)	0.266(0.307)	0.087(0.098)
	0.1	0.090(0.070)	0.224(0.231)	0.058(0.058)
	0.2	0.068(0.037)	0.158(0.128)	0.029(0.018)
	0.3	0.011(0.008)	0.159(0.127)	0.025(0.016)
	0.4	−0.004(0.014)	0.107(0.147)	0.011(0.021)
	0.5	−0.008(0.052)	0.128(0.191)	0.016(0.039)
	min abs	0.004(0.008)	0.107(0.127)	0.011(0.016)
$\sigma = 1$				
	0.03	0.028(−0.060)	0.330(0.386)	0.109(0.151)
	0.1	0.055(0.006)	0.293(0.334)	0.088(0.110)
	0.2	0.007(−0.043)	0.365(0.344)	0.132(0.119)
	0.3	−0.076(−0.112)	0.384(0.342)	0.152(0.128)
	0.4	−0.083(−0.003)	0.384(0.372)	0.153(0.137)
	0.5	0.034(0.124)	0.416(0.397)	0.173(0.171)
	min abs	0.007(0.003)	0.293(0.334)	0.088(0.110)
$\sigma = 2/3$				
	0.03	−0.004(−0.127)	0.276(0.316)	0.075(0.115)
	0.1	−0.026(−0.040)	0.343(0.330)	0.117(0.109)
	0.2	−0.077(−0.098)	0.395(0.363)	0.160(0.140)
	0.3	−0.123(−0.156)	0.391(0.377)	0.167(0.165)
	0.4	−0.165(−0.170)	0.429(0.457)	0.210(0.236)
	0.5	−0.064(−0.062)	0.498(0.505)	0.250(0.256)
	min abs	0.004(0.040)	0.276(0.316)	0.075(0.109)

5. Sketches of the Proofs

The following notation will be used throughout this section. Set $\eta_h = n^{(1/2)-\delta}h^{1/2}$ and $\eta_g = n^{(1/2)-\delta}g^{1/2}$. Let $E[S(\hat{t}_j)] = E[S(x)]|_{x=\hat{t}_j}$, and $E[J^{(1)}(\hat{t}_j)] = E[J^{(1)}(x)]|_{x=\hat{t}_j}$, for $j = 1, 2, \ldots, \rho$. Let z_i, $i \in Z$, denote partition points of

$[0,1]$ satisfying $z_i - z_{i-1} = n^{-(1+q)}$, Ψ the interval $[0,1]$, $\Omega = \{i : z_i \in \Psi\}$, $\Psi_j = \{x : x \in A_j,\ I_{[|x-t_j|>h^{1+\theta}]} = 1\}$, and u_j the partition point satisfying $|u_j - t_j| = \min\{|z_i - t_j| : i \in \Omega\}$, for $j = 1, 2, \ldots, p$.

To prove Theorem 1 and 2, we require the following lemma.

Lemma 1. *In the case $p \geq 0$, under (2.1), (A.0), and (A.1), if K is compactly supported on $[-1,1]$, Lipschitz continuous, and of order q, then the kernel density estimator $\hat{f}(x)$ of (2.2) has the following properties*:

$$E[\hat{f}(x)] = \tau(x) + \sum_{j=1}^{p} d_j \int_{-1}^{(x-t_j)/h} K + h^q (-1)^q \tau^{(q)}(x) \int z^q K/(q!) + O(h^{q+1}), \quad (5.1)$$

uniformly on Ψ,

$$\eta_h \sup_{x \in \Psi} |\hat{f}(x) - E[\hat{f}(x)]| \to 0 \quad a.s. \quad (5.2)$$

Proof. The proof of (5.1) follows through straightforward calculation. Hence it is omitted. We now give the proof of (5.2). For this, let $\hat{f}^*(x) = \hat{f}(x) - E[\hat{f}(x)]$. Consider the inequality

$$\eta_h \sup_{x \in \Psi} |\hat{f}^*(x)| \leq \varphi_1 + \varphi_2,$$

where

$$\varphi_1 = \eta_h \sup_{i \in \Omega} |\hat{f}^*(z_i)|, \qquad \varphi_2 = \eta_h \sup_{i \in \Omega} \sup_{|x - z_i| \leq n^{-(q+1)}} |\hat{f}^*(x) - \hat{f}^*(z_i)|.$$

The proof of (5.2) is complete by showing that $\varphi_i \to 0$ a.s. for $i = 1, 2$. To check the strong consistency of φ_1, using the result that $E[\hat{f}^*(x)^{2k}] = O(n^{-k}h^{-k})$ uniformly on Ψ, for any integer $k \geq 1$, and taking $\eta > 0$, then, for any $k = 1, 2, \ldots$, there is a constant a_k such that

$$\sum_{n=1}^{\infty} P(\eta_h \sup_{i \in \Omega} |\hat{f}^*(z_i)| > \eta) \leq \sum_{n=1}^{\infty} a_k \eta^{-2k} n^{(1+q)-2k\delta}.$$

According to this result, the strong consistency of φ_1 follows by using the Borel-Cantelli lemma and the fact that there is a sufficiently large k such that $(1 + q) - 2k\delta < -1$. The strong consistency of φ_2 is a consequence of the Lipschitz continuity of K. Hence, the proof of (5.2) is complete, i.e. the proof of Lemma 1 is complete.

Proof of Theorem 1.

We first give the proof of (3.1). The proof for $\hat{t}_1$ is complete by showing

$$P(\sup_{x \in \Psi_1} |J(x)| \geq |J(u_1)| \quad \text{i.o.}) = 0. \quad (5.3)$$

To check (5.3), by (5.1), (A.1), (A.2), $|d_j| > |d_{j+1}|$, for $j = 1, 2, \ldots, p$, and the fact that $\theta \in (0, \beta/\kappa)$, through straightforward calculation, then

$$|E[J(u_1)]| - \sup_{x\in\Psi_1} |E[J(x)]| = \left| d_1 \int_0^{h^\theta} (K_1 - K_2) \right| + O(h^\beta) \geq 2C + O(h^\beta),$$

where

$$C = (1/2)|d_1|\eta h^{\kappa\theta}.$$

Using this result, then, through straightforward calculation,

$$\sup_{x\in\Psi_1} |J(x)| - |J(u_1)| \leq 2 \sup_{x\in\Psi} |J(x) - E[J(x)]| - 2C + O(h^\beta).$$

Combining this inequality with (5.2), (B.1), and the fact that $\theta \in (0, \beta/\kappa)$, we then have

$$P(\sup_{x\in\Psi} |J(x) - E[J(x)]| + O(h^\beta) \geq C \quad \text{i.o.}) = 0.$$

Hence, the proof for $\hat{t}_1$ is complete.

We now give the proof for $\hat{t}_2$. The proofs for the rest of $\hat{t}_j$ follow similarly. Since the distance between any two of t_j, $j = 1, 2, \ldots, p$, is greater than ξ and $h = o(1)$, then, for sufficiently large n, we have $|u_2 - t_1| > 3h$. Using this result and the property of $\hat{t}_1$ in (3.1), we have

$$P(u_2 \in [\hat{t}_1 - 2h,\, \hat{t}_1 + 2h] \quad \text{i.o.}) = 0.$$

Following essentially the same proof of (5.3), through a straightforward calculation, it follows,

$$P(\sup_{x\in\Psi_2} |J(x)| \geq |J(u_2)| \quad \text{i.o.}) = 0.$$

According to the property of $\hat{t}_1$ in (3.1) and the fact that

$$|\hat{t}_2 - t_1| \geq |\hat{t}_2 - \hat{t}_1| - |\hat{t}_1 - t_1| \geq 2h - |\hat{t}_1 - t_1|,$$

then

$$P(|\hat{t}_2 - t_1| < h \quad \text{i.o.}) = 0.$$

Combining these results with the definition of Ψ_2, we have

$$P(|\hat{t}_2 - t_2| > h^{1+\theta} \quad \text{i.o.}) = 0.$$

Hence the proof of (3.1) is complete.

We now give the proof of (3.2). Here we shall only give the proof for $\hat{d}_1 - d_1$. The proofs for the rest of $\hat{d}_j - d_j$ follow similarly. For this, subtracting and adding the terms $E[S(\hat{t}_1)]$ and $E[S(t_1)]$, then

$$\hat{d}_1 - d_1 = \tau_1 + \tau_2 + \tau_3,$$

where $\tau_1 = c_s(S(\hat{t}_1) - E[S(\hat{t}_1)])$, $\tau_2 = c_s(E[S(\hat{t}_1)] - E[S(t_1)])$, $\tau_3 = c_s E[S(t_1)] - d_1$. By this, the proof that $\eta_g|\hat{d}_1 - d_1| \to 0$ a.s. is complete by showing $\eta_g \tau_1 \to 0$ a.s. and $\tau_2 + \tau_3 = o(\eta_g^{-1})$. Note that the strong consistency of $\eta_g \tau_1$ follows by the result of (5.2). To check $\tau_2 + \tau_3 = o(\eta_g^{-1})$, multiplying τ_2 by $I_{[|\hat{t}_1 - t_1| \geq h^{1+\theta}]} + I_{[|\hat{t}_1 - t_1| < h^{1+\theta}]}$, and combining the result with (A.3) and (5.1), through a straightforward calculation, it follows that

$$\tau_2 + \tau_3 = c_s(E[S(\hat{t}_1)] - E[S(t_1)]) I_{[|\hat{t}_1 - t_1| \geq h^{1+\theta}]} + O(h^{1+\theta} g^{-1} + g^{\beta}).$$

Combining this result with (B.2), (B.3), and the property of $\hat{t}_1$ in (3.1), then $\tau_2 + \tau_3 = o(\eta_g^{-1})$. Hence the proof of (3.2) is complete.

We now give the proof of (3.3). Here we shall only give the proof of the asymptotic normality for Λ_1. The proofs for the rest of Λ_j follow similarly. The proof of the asymptotic normality for $\hat{t}_1 - t_1$ is based on the expansion

$$0 = J^{(1)}(\hat{t}_1) = E[J^{(1)}(\hat{t}_1)] + (J^{(1)}(\hat{t}_1) - E[J^{(1)}(\hat{t}_1)]). \tag{5.4}$$

By (A.2), through a straightforward calculation,

$$E[J^{(1)}(x)] = \sum_{j=1}^{p} d_j (K_{1h} - K_{2h})(x - t_j) + O(h^{\beta - 1}),$$

uniformly on Ψ. Using this result and replacing x with $\hat{t}_1$, multiplying the result by $I_{[|\hat{t}_1 - t_1| < h^{1+\theta}]} + I_{[|\hat{t}_1 - t_1| \geq h^{1+\theta}]}$, and applying Taylor's theorem to $K_{1h} - K_{2h}$, through a straightforward calculation, (5.4) becomes

$$\begin{aligned} 0 = {} & 2d_1 h^{-2}(\hat{t}_1 - t_1) K_1^{(1)}(0) I_{[|\hat{t}_1 - t_1| < h^{1+\theta}]}(1 + o(1)) \\ & + \sum_{j=1}^{p} d_j (K_{1h} - K_{2h})(\hat{t}_1 - t_j) I_{[|\hat{t}_1 - t_1| \geq h^{1+\theta}]} + O(h^{\beta - 1}) \\ & + (J^{(1)}(\hat{t}_1) - E[J^{(1)}(\hat{t}_1)]). \end{aligned} \tag{5.5}$$

By using partition points z_i on Ψ and the Lipschitz continuity of $K_1^{(1)}$ and $K_2^{(1)}$, through a straightforward calculation,

$$\sup_{|x - t_1| \leq h^{1+\theta}} |(J^{(1)}(x) - E[J^{(1)}(x)]) - (J^{(1)}(\hat{t}_1) - E[J^{(1)}(\hat{t}_1)])| = o_p(n^{-1/2} h^{-3/2}).$$

Combining this result with (5.5) and the property of $\hat{t}_1$ in (3.1), (5.5) becomes

$$\hat{t}_1 - t_1 = [(J^{(1)}(t_1) - E[J^{(1)}(t_1)]) + o_p(n^{-1/2}h^{-3/2}) + O(h^{\beta-1})]/ \\ [-2d_1 h^{-2} K_1^{(1)}(0)(1+o(1))]. \tag{5.6}$$

By the Lindeberg-Levy theorem, through a straightforward calculation,

$$n^{1/2}h^{3/2}(J^{(1)}(t_1) - E[J^{(1)}(t_1)]) \Rightarrow N(0, F_j \int (K_1^{(1)} - K_2^{(1)})^2).$$

Combining this result with (5.6) and (B.4), the proof of the asymptotic normality for $\hat{t}_1 - t_1$ is complete.

The proof of the asymptotic normality for $\hat{d}_1 - d_1$ is now given. By using the above decomposition of $\hat{d}_1 - d_1$ and following essentially the same proof of (5.6),

$$\hat{d}_1 - d_1 = c_\omega(S(t_1) - E[S(t_1)]) + o_p(n^{-1/2}g^{-1/2}) + O(h^{1+\theta}g^{-1} + g^\beta). \tag{5.7}$$

By the Lindeberg-Levy theorem, through a straightforward calculation,

$$(ng)^{1/2}(S(t_1) - E[S(t_1)]) \Rightarrow N(0, F_j \int (K_3 - K_4)^2).$$

Combining this result with (5.7), (B.5), and (B.6), the proof of the asymptotic normality for $\hat{d}_1 - d_1$ is complete.

By (5.6), (5.7), (B.7), and the Cramer-Wold technique, through a straightforward calculation, the asymptotic normality of Λ_1 follows.

We now give the proof for the asymptotic independence between Λ_j, for $j = 1, 2, \ldots, \rho$. By following essentially the same proofs of (5.6) and (5.7), through a straightforward calculation,

$$\hat{t}_j - t_j = [(J^{(1)}(t_j) - E[J^{(1)}(t_j)]) + o_p(n^{-1/2}h^{-3/2}) + O(h^{\beta-1})]/ \\ [-2d_j h^{-2} K_1^{(1)}(0)(1+o(1))], \tag{5.8}$$

$$\hat{d}_j - d_j = c_\omega(S(t_j) - E[S(t_j)]) + o_p(n^{-1/2}g^{-1/2}) + O(h^{1+\theta}g^{-1} + g^\beta). \tag{5.9}$$

Using these two results, (B.7), and the Cramer-Wold technique, then, through a straightforward calculation, Λ_j have joint asymptotic normality, for $j=1,2,\ldots,\rho$. Also, $\mathrm{Cov}[\hat{t}_j - t_j, \hat{t}_k - t_k] = O(n^{-1}h^2)$, $\mathrm{Cov}[\hat{t}_j - t_j, \hat{d}_k - d_k] = O(n^{-1}h)$, $\mathrm{Cov}[\hat{d}_j - d_j, \hat{d}_k - d_k] = O(n^{-1})$, for $j \neq k$. Based on these results, the asymptotic independence between Λ_j follows. Hence the proof of (3.3) is complete, i.e. the proof of Theorem 1 is complete.

Proof of Theorem 2.

The proof of (3.4), for $j = 1, 2, \ldots, p$, is the same as that of (3.2). To check (3.4), for $j > p$, let $A^* = \cup_{k=1}^{p}[t_k - h, t_k + h]$. By (3.1),

$$P(\hat{t}_j \in A^* \quad \text{i.o.}) = 0,$$

for $j = p+1, p+2, \ldots, \rho$. Using this result, (2.1), and the result of (5.1), we have

$$\sup_{x \in A^*} E[S(x)] = O(g^{\beta}).$$

Combining this result with the fact that

$$\hat{d}_j = c_s(S(\hat{t}_j) - E[S(\hat{t}_j)]) + c_s E[S(\hat{t}_j)],$$

and utilizing (A.3), (5.2), (B.2), and (B.3), then, through a straightforward calculation, (3.4) follows. Hence, the proof of Theorem 2 is complete.

Acknowledgement

The authors gratefully thank the referees, the associate editor, and the editor for many valuable comments which substantially improved the presentation. The research of the first author was supported by the National Science Council under the contract number NSC81-0208-M007-55.

References

Bickel, P. J. and Rosenblatt, M. (1973). On some global measures of the deviations of density function estimates. Ann. Statist. **1**, 1071-1095.

Cline, D. B. H. and Hart, J. D. (1991). Kernel estimation of densities with discontinuities or discontinuous derivatives. Statistics **22**, 69-84.

van Eeden, C. (1985). Mean integrated squared error of kernel estimators when the density and its derivatives are not necessarily continuous. Ann. Inst. Stat. Math. **37**, 461-472.

van Es, B. (1992). Asymptotics for least squares cross-validation bandwidths in nonsmooth cases. Ann. Statist. **20**, 1647-1657.

Gasser, T. and Müller, H. G. (1979). Kernel estimation of regression functions. In Smoothing Techniques for Curve Estimation. Lecture Notes in Math. **757**, 23-68, Springer-Verlag, New York.

Härdle, W. (1991). Smoothing Techniques: With Implementation in S. Springer Series in Statistics, Springer-Verlag, Berlin.

Hall, P. and Marron, J. S. (1987). Extent to which least-squares cross-validation minimizes integrated square error in nonparametric density estimation. Theory Probab. Related Fields **74**, 567-581.

Johnstone, I. M. and Silverman, B. W. (1990). Speed of estimation in position emission tomography and related inverse problems. Ann. Statist. **18**, 251-280.

Lee, D. (1990). Coping with discontinuities in computer vision: their detection, classification, and measurement. IEEE Transactions on Patter Analysis and Machine Intelligence, **12**, 321-344.

Müller, H. G. (1992). Change-points in nonparametric regression analysis. Ann. Statist. **20**, 737-761.

Prakasa Rao, B. L. S. (1983). Nonparametric Functional Estimation. Academic Press, Orlando, Florida.

Qiu, P., Asano, C. and Li, X. (1991). Estimation of jump regression function. Japan. J. Bull. Inform. & Cybernet. **24**, 197-212.

Silverman, B. W. (1986). Density Estimation for Statistics and Data Analysis. Chapman and Hall, New York.

Silverman, B. W., Jones, M. C., Wilson, J. D. and Nychka, D. W. (1990). A smoothed EM approach to indirect estimation problems, with particular reference to stereology and emission tomography. J. Roy. Statist. Soc. Ser.B **52**, 271-324.

Wu, J. S. and Chu, C. K. (1993). Kernel-type estimators of jump points and values of a regression function. Ann. Statist. **21**, 1545-1566.

Yin, Y. Q. (1988). Detection of the number, locations and magnitudes of jumps. Comm. Statist. Stochastic Models **4**, 445-455.

Institute of Statistics, National Tsing Hua University, Hsinchu 30043, Taiwan.

Institute of Statistical Science, Academia Sinica, Taipei 11529, Taiwan.

(Received February 1992; accepted July 1995)