Boundary Aware Estimators of Integrated Density Derivative Products

Ming-Yen Cheng

Journal of the Royal Statistical Society. Series B (Methodological), Vol. 59, No. 1 (1997), 191-203.

Stable URL:
http://links.jstor.org/sici?sici=0035-9246%281997%2959%3A1%3C191%3ABAEOID%3E2.0.CO%3B2-1

J. R. Statist. Soc. B (1997)
59, *No.* 1, *pp.* 191–203

Boundary Aware Estimators of Integrated Density Derivative Products

By MING-YEN CHENG†

National Chung Cheng University, Chiayi, Taiwan

[Received April 1994. Final revision March 1996]

SUMMARY

Integrated squared density derivatives are important to the plug-in type of bandwidth selector for kernel density estimation. Conventional estimators of these quantities are inefficient when there is a non-smooth boundary in the support of the density. We introduce estimators that utilize density derivative estimators obtained from local polynomial fitting. They retain the rates of convergence in mean-squared error that are familiar from non-boundary cases, and the constant coefficients have similar forms. The estimators and the formula for their asymptotically optimal bandwidths, which depend on integrated products of density derivatives, are applied to automatic bandwidth selection for local linear density estimation. Simulation studies show that the constructed bandwidth rule and the Sheather–Jones bandwidth are competitive in non-boundary cases, but the former overcomes boundary problems whereas the latter does not.

Keywords: BANDWIDTH SELECTION; BOUNDARY EFFECTS; DATA BINNING; LOCAL POLYNOMIAL FITTING; PLUG-IN BANDWIDTH SELECTOR

1. INTRODUCTION

Suppose that $X_1, \ldots, X_n$ is an independent and identically distributed (IID) sample from a population following an unknown density function f. We consider using the sample to estimate the following functional of the density:

$$\theta_{\gamma,\nu} = \int f^{(\gamma)}(x)\, f^{(\nu)}(x)\, \mathrm{d}x,$$

where $\gamma, \nu \geqslant 0$ and $\gamma + \nu$ is an even integer. Throughout this paper, $f^{(\nu)}(x)$ is taken as 0 when x is outside the support of f and as the limit from the right, or left, when x is a left, or right, boundary point. Special cases where $\gamma = \nu$ are crucial to the plug-in type of automatic bandwidth selection for kernel density estimation; see Jones *et al.* (1994). Kernel estimators of these quantities are discussed in Hall and Marron (1987), Jones and Sheather (1991) and Aldershof (1991) among others. When boundaries are present in the support of f, the estimators proposed therein are very inefficient. The main reason is that the non-smoothness of the density at the boundaries introduces an extra bias term which dominates the mean-squared error; see Van Es and Hoogstrate (1994) for detailed discussion.

Instead of the conventional kernel density derivative estimators, we use density derivative estimators resulting from local polynomial fitting to construct estimators

†*Address for correspondence*: Institute of Mathematical Statistics, National Chung Cheng University, Minhsiung, Chiayi 621, Taiwan.
E-mail: mycheng@mthmp.math.ccu.edu.tw

 0035–9246/97/59191

of $\theta_{\gamma,\nu}$. The motivation is that such derivative estimators automatically correct boundary effects. Indeed, they are shown in Cheng *et al.* (1995) to be as efficient as any other linear estimator in a weak minimax sense. Asymptotic properties of the resulting estimators are investigated. They have the same rates of convergence and similar constant coefficients as the estimators of Jones and Sheather (1991) in non-boundary cases and retain the rates of convergence even in boundary cases. The optimal smoothing parameters of the estimators are given so that they, together with the estimators, can be directly adapted for data-driven bandwidth selection.

Binning of the data is essential to the estimators introduced here. The idea of local polynomial fitting arises in the regression setting; see Stone (1977). Binning produces bin counts which can be viewed as responses at the bin centres. Then local polynomial ideas can be introduced in the density estimation context. Local linear regression techniques were applied to estimating distributions and densities by Lejeune and Sarda (1992) without binning the data. In that paper, a local polynomial was fitted to the empirical distribution function by minimizing a kernel weighted L^2-norm. Differentiating the estimated distribution curve gives an estimator of the density. Jones (1993) also mentioned density derivative estimators derived from local polynomial fitting. These estimators of the density and its derivatives are roughly the same as those obtained from our approach; see Section 2 for details.

Bickel and Ritov (1988) gave information bounds for nonparametric estimation of the quantities $\theta_{\gamma,\gamma}$ and provided estimators which attain the best $\sqrt{n}$-convergence. The estimators presented in this paper can achieve the same $\sqrt{n}$-convergence if the density has a larger degree of smoothness than that assumed in Bickel and Ritov (1988). But our estimators need smoothness of the density only in its support whereas Bickel and Ritov (1988) imposed their smoothness conditions over the entire real line. However, no information bounds have been given for classes of densities that are smooth except for having some possible jumps of the density or its derivatives. We conjecture that similar information bounds are available for these more general classes of densities and the estimators presented here will be the corresponding best estimators.

Our estimators of $\theta_{r,\nu}$ and the formula for their asymptotically optimal bandwidth imply some plug-in data-driven bandwidth selectors in density estimation. An important strength of such bandwidth selectors is that, unlike many conventional bandwidth rules, they yield proper bandwidths even when there are non-smooth boundaries and they are as efficient as conventional selectors in non-boundary cases. To illustrate boundary effects on the conventional bandwidth rules, consider the plug-in bandwidth rule of Sheather and Jones (1991). It is popular from both practical and theoretical viewpoints; see Jones *et al.* (1994) and Sheather (1992) and references therein. In that procedure, $\theta_{\gamma,\gamma}$ is estimated by

$$\tilde{\tilde{\theta}}_{\gamma,\gamma} = \int_{-\infty}^{+\infty} \tilde{f}_h^{(\gamma)}(x)^2 \, \mathrm{d}x$$

or

$$\tilde{\theta}_{\gamma,\gamma} = (-1)^{\gamma} \frac{1}{n} \sum_{i=1}^{n} \tilde{f}_h^{(2\gamma)}(X_i),$$

where $\tilde{f}_h^{(m)}$ is the mth derivative of a kernel density estimator $\tilde{f}_h$. The estimator $\tilde{\theta}_{\gamma,\gamma}$ is motivated by $\theta_{\gamma,\gamma} = (-1)^\gamma \int f^{(2\gamma)} f$ when f is smooth everywhere. However, the previous equality does not hold if there are non-smooth boundaries in the support of f. As for the former, $\tilde{f}_h^{(\gamma)}(x) \neq 0$ for xs that are close to the boundary, if any, but $f^{(\gamma)}(x) = f(x) = 0$. One immediate result of these problems is that the estimators have larger biases and become relatively inefficient. Or intuitively they feel the non-smoothness or discontinuity of the curve and become too large. As a result, the Sheather–Jones procedure selects a bandwidth that is too small for the data. We indeed observed such problems for the Sheather–Jones bandwidth in simulation studies. Such boundary effects do not pertain only to the Sheather–Jones selectors, since most of the existing bandwidth selectors involve using kernel estimators of $\theta_{\gamma,\gamma}$ for various values of γ.

This paper is organized as follows. Estimators of $\theta_{\gamma,\nu}$ together with local polynomial techniques applied to density estimation are discussed in Section 2. Asymptotic properties of the estimators are investigated in Section 3. In Section 4, the results of Section 3 are applied to bandwidth selection for density estimation and a simulation study shows its usefulness in practice.

2. THE ESTIMATORS

First, we discuss estimating density derivatives, with an IID sample, by local polynomial fitting. The idea is to fit some 'response', which is equal to the density plus a random error term, by a local polynomial. An apparent way to create such responses is to bin the data, i.e. divide the support of the density into a set of disjoint intervals and move the observations in each interval to its centre. After this binning, the data are transformed into a set of bin counts. The counts reflect the height of the density at the bin centres: a larger count indicates a higher density. Therefore, these counts can be viewed as the 'responses' at the 'design points'—the bin centres. Precisely, for some positive constant b and each $i = 1, \ldots, g$, define the bin centre as $x_i = L + (i - \frac{1}{2})b$ and the corresponding bin count as

$$c_i = \sum_{j=1}^{n} I_{[x_i - b/2, x_i + b/2)}(X_j).$$

Here, L and g are fixed numbers chosen such that no data points are less than L or greater than $L + gb$.

Each c_i provides information about $f(x_i)$ in the sense that

$$n^{-1} b^{-1} c_i \xrightarrow{\text{P}} b^{-1} \int_{x_i - b/2}^{x_i + b/2} f(u) \, \mathrm{d}u \approx f(x_i), \tag{1}$$

as $n \to \infty$ and $nb \to \infty$. Following the ideas of local polynomial fitting, consider

$$\min_{\beta_0, \ldots, \beta_p} \left[\sum_{i=1}^{g} K\left(\frac{x_i - x}{h} \right) \left\{ n^{-1} b^{-1} c_i - \sum_{j=0}^{p} \beta_j (x_i - x)^j \right\}^2 \right], \tag{2}$$

where K is a non-negative function and h is positive. Denote the solution of the least squares problem (2) as $\hat{b}_j(x), j = 0, 1, \ldots, p$. Then a natural estimator of $f^{(m)}(x)$

is $\hat{f}_h^{(m)}(x) \equiv m!\,\hat{b}_m(x)$, $m = 0, 1, \ldots, p$. When $p = 1$ and $m = 0$, $\hat{f}_h(x)$ is referred to as a local linear estimator of the density $f(x)$.

Define $S_n = (S_{n,i+j-2}(x))_{0 \leqslant i,j \leqslant p+1}$, where

$$S_{n,j}(x) = \sum_{i=1}^{g} K\left(\frac{x_i - x}{h}\right) (x_i - x)^j, \qquad j = 0, 1, \ldots, 2p.$$

Standard weighted least squares calculation shows that

$$\hat{b}_m(x) = \sum_{i=1}^{g} W_m^n\left(\frac{x_i - x}{h}\right) n^{-1} b^{-1} c_i, \tag{3}$$

where $W_m^n(t) = e_{m+1}^{\mathrm{T}} S_n^{-1}(1, ht, \ldots, h^p t^p)^{\mathrm{T}} K(t)$ with e_{m+1} being the $(m+1)$th unit $(p+1)$-vector. The above calculation is straightforward and is given in Fan *et al.* (1993) and Ruppert and Wand (1994).

In connection with higher order kernel approaches such as those of Gasser *et al.* (1985), it is clear that, as $n \to \infty$ and $b/h \to 0$,

$$S_{n,j}(x) = b^{-1} h^{j+1}\, S_j\{1 + o(1)\},$$

where

$$S_j = \int_{-\infty}^{+\infty} t^j\, K(t)\, \mathrm{d}t, \qquad j = 0, 1, \ldots, 2p.$$

Hence, if we write $S = (S_{i+j-2})_{0 \leqslant i,j \leqslant p+1}$,

$$W_m^n(t) \approx \frac{b}{h^{m+1}}\, e_{m+1}^{\mathrm{T}}\, S^{-1}(1, t, \ldots, t^p)^{\mathrm{T}} K(t), \qquad m = 0, 1, \ldots, p. \tag{4}$$

Therefore, we have the following equivalent kernel representation. For each $m = 0, 1, \ldots, p$,

$$\hat{f}_m(x) \approx \frac{m!}{nh^{m+1}} \sum_{i=1}^{g} K_m^*\left(\frac{x_i - x}{h}\right) c_i, \tag{5}$$

where $K_m^*(t) = e_{m+1}^{\mathrm{T}}\, S^{-1}(1, t, \ldots, t^p)^{\mathrm{T}} K(t)$. Connections between the local polynomial fitting approach and higher order kernel methods were discussed in Müller (1987) and Lejeune (1984). Although it is suggested that the two approaches yield equivalent estimators, advantages of local polynomial estimators include much better interpretability and automatic boundary correction.

When considering boundary cases, the support of f is assumed, without loss of generality, to be $[0, \infty)$. To study asymptotic behaviours of the estimators in the boundary region, we let $x = ch$, $c \geqslant 0$. In that case,

$$S_j = \int_{-c}^{+\infty} t^j K(t)\, \mathrm{d}t, \qquad j = 0, 1, \ldots, 2p$$

and $K_m^*(t) = e_{m+1}^{\mathrm{T}}\, S^{-1}(1, t, \ldots, t^p)^{\mathrm{T}} K(t)\, I_{[-c,\infty)}(t)$.

Lejeune and Sarda (1992) and Jones (1993) employed local polynomial fitting techniques to density estimation in a different manner. Their approach was to fit a local polynomial to the empirical density function by minimizing the L^2-norm, i.e.

$$\min_{\beta_0,\ldots,\beta_p}\left[\int K\left(\frac{u-x}{h}\right)\left\{f_n(u)-\sum_{j=0}^{p}\beta_j(u-x)^j\right\}^2 \mathrm{d}u\right],$$

where

$$f_n(u)=n^{-1}\sum_{i=1}^{n} I(u=X_i),$$

the empirical density function. This yields an estimator of $f^{(m)}$ which is

$$\bar{f}_h^{(m)}(x)=\frac{m!}{nh^{m+1}}\sum_{j=1}^{n} K_m^*\left(\frac{X_i-x}{h}\right), \qquad m=0, 1, \ldots, p. \tag{6}$$

Note that

$$m!\,n^{-1}h^{-(m+1)}\sum_{i=1}^{g} K_m^*\left(\frac{x_i-x}{h}\right)c_i$$

is a binned approximation of $\bar{f}_h^{(m)}(x)$. Hence, expressions (5) and (6) imply that $\bar{f}_h^{(m)}(x)$ and $\hat{f}_h^{(m)}(x)$ are equivalent. Binned implementation of $\hat{f}_h^{(m)}(x)$ is faster than calculating $\hat{f}_h^{(m)}(x)$. Since for each $j=0, 1, \ldots, 2p$, given the weight function K, we can find an explicit formula for

$$S_j=\int_{-c}^{+\infty} t^j K(t)\,\mathrm{d}t$$

whereas

$$S_{n,j}(x)=\sum_{i=1}^{g} K\left(\frac{x_i-x}{h}\right)(x_i-x)^j$$

involves more numerical operations. Proof of the following theorem can be found in Cheng (1994).

Theorem 1. For each fixed $m=0, 1, \ldots, p$, suppose that f and its first m derivatives are bounded and K is bounded. Then, as $n\to\infty$, $h\to 0$, $nh^{2m+1}\to\infty$ and $b/h\to 0$,

$$E\{\hat{f}_h^{(m)}(x)-f^{(m)}(x)\}^2=\left\{\int t^{p+1}K_m^*(t)\,\mathrm{d}t\right\}^2\left\{\frac{m!\,f^{(m)}(x)}{(p+1)!}\right\}^2 h^{2(p+1-m)} + \frac{m!^2 f(x)}{nh^{2m+1}}\int K_m^*(t)^2\,\mathrm{d}t + o\left(h^{2(p+1-m)}+\frac{1}{nh^{2m+1}}\right).$$

Therefore, local polynomial density derivative estimators attain automatic boundary corrections. We propose to estimate $\theta_{\gamma,\nu}$ by

$$\widehat{\theta}_{\gamma,\nu}(a) = b\sum_{i=1}^{g} \widehat{f}_a^{(\gamma)}(x_i)\widehat{f}_a^{(\nu)}(x_i) = \frac{\gamma!\,\nu!}{n^2 b}\sum_{i=1}^{g}\sum_{j=1}^{g}\sum_{k=1}^{g} W_\gamma^n\left(\frac{x_j - x_i}{a}\right) W_\nu^n\left(\frac{x_k - x_i}{a}\right) c_j c_k, \quad (7)$$

where $a > 0$. Note that the bandwidth for $\widehat{\theta}_{\gamma,\nu}(a)$ is a which is different from the bandwidth h when estimating density derivatives. Asymptotic properties of $\widehat{\theta}_{\gamma,\nu}(a)$ are investigated in the following section.

3. ASYMPTOTIC PROPERTIES OF ESTIMATORS OF INTEGRATED DENSITY DERIVATIVE PRODUCTS

The density f is said to be in the class F_{p+1}, where p is a non-negative integer, if there is a constant $M > 0$ such that, for any x and y belonging to the support of f,

$$|f^{(p+1)}(x) - f^{(p+1)}(y)| \leqslant M|x - y|. \quad (8)$$

For a more precise version of inequality (8), see Bickel and Ritov (1988). Proof of the following theorem can be found in Cheng (1994).

Theorem 2. Suppose that $f \in F_{p+1}$ with $p + 2 > \gamma + \nu$ and the weight function K is compactly supported with two derivatives. Since $\widehat{\theta}_{\gamma,\nu}(a)$ is symmetric in γ and ν, we assume $\gamma \leqslant \nu$ for clean presentation. Then $\widehat{\theta}_{\gamma,\nu}(a)$ has bias

$$E\{\widehat{\theta}_{\gamma,\nu}(a)\} - \theta_{\gamma,\nu} = \frac{\gamma!\,\nu!}{na^{\gamma+\nu+1}}\int K_\gamma^* K_\nu^* + \frac{(1+\delta_{\gamma\nu})\nu!}{(p+1)!}\, a^{p-\nu+1}\theta_{\gamma,p+1}\left(\int u^{p+1}K_\nu^*\right) + O(n^{-1}a^{-\gamma-\nu}) + O(a^{p-\gamma+1}), \quad (9)$$

and variance

$$\operatorname{var}\{\widehat{\theta}_{\gamma,\nu}(a)\} = \frac{2(\gamma!\,\nu!)^2}{n^2 a^{2(\gamma+\nu)+1}}\left(\int f^2\right)\left\{\int (K_\gamma^* * K_\nu^*)^2\right\} + \frac{4}{n}\left\{\int f(f^{(\gamma+\nu)})^2 - \theta_{\gamma,\nu}^2\right\} + o(n^{-2}a^{-2(\gamma+\nu)-1}) + o(n^{-1}), \quad (10)$$

provided that $n \to \infty$, $a \to 0$, $na^{\gamma+\nu+1} \to \infty$ and $b/a \to 0$.

Remark 1. Consider the case $\gamma = \nu$. Let $k = p - \nu + 1$. Suppose that f is smooth and supported on the entire real line. Then $\widehat{\theta}_{\gamma,\gamma}(a)$ has the same rate of convergence in mean-squared error as the corresponding estimator of Jones and Sheather (1991) based on a kernel of order k. Indeed, if K is the standard normal density, the equivalent kernel K_m^* in approximation (5) is equal to $(m!)^{-1}K^{(m)}$. Then $\widehat{\theta}_{\gamma,\gamma}(a)$ is roughly the same as their estimator $\int \widetilde{f}_a^{(\gamma)}(x)^2\,dx$. Hence theorem 2 is also the first result on binned version estimators of integrated density derivatives in the non-boundary case.

Remark 2. It is worthwhile to mention that our estimators are so effective that even the constant coefficients in the asymptotic mean-squared errors depend on f through the same functionals, whether or not there is any non-smooth boundary.

Corollary 1. Suppose that $\int K_\gamma^* K_\nu^*$ and $\int u^{p+1} K_\nu^*$ have the same sign. Then, if $\theta_{\gamma,p+1} > 0$, the bandwidth which minimizes the asymptotic mean-squared error of $\widehat{\theta}_{\gamma,\nu}(a)$ is

$$a_{\mathrm{AMSE}} = \left\{ \frac{\gamma!\,(p+1)!\,(\gamma+\nu+1) \int K_\gamma^* K_\nu^*}{n(1+\delta_{\gamma\nu})(p-\nu+1)|\theta_{\gamma,p+1}| \int u^{p+1} K_\nu^*} \right\}^{1/(p+\gamma+2)},$$

and the optimal asymptotic mean-squared error is equal to

$$\{(p+\gamma+2)\nu!\}^2 \left(\frac{\gamma! \int K_\gamma^* K_\nu^*}{p-\nu+1} \right)^{2(p-\nu+1)/(p+\gamma+2)} \left\{ \frac{(1+\delta_{\gamma\nu})|\theta_{\gamma,p+1}| \int u^{p+1} K_\nu^*}{(\gamma+\nu+1)(p+1)!} \right\}^{2(\gamma+\nu+1)/(p+\gamma+2)}$$

$$\times\, n^{-2(p-\nu+1)/(p+\gamma+2)} + \frac{4}{n} \left\{ \int f (f^{(\gamma+\nu)})^2 - \theta_{\gamma,\nu}^2 \right\}, \tag{11}$$

and, if $\theta_{\gamma,p+1} < 0$,

$$a_{\mathrm{AMSE}} = \left\{ \frac{\gamma!\,(p+1)! \int K_\gamma^* K_\nu^*}{n(1+\delta_{\gamma\nu})|\theta_{\gamma,p+1}| \int u^{p+1} K_\nu^*} \right\}^{1/(p+\gamma+2)},$$

and the optimal asymptotic mean-squared error is equal to

$$2(\gamma!\nu!)^2 \left(\int f^2 \right) \left\{ \int (K_\gamma^* * K_\nu^*)^2 \right\} \left\{ \frac{(1+\delta_{\gamma\nu})|\theta_{\gamma,p+1}| \int u^{p+1} K_\nu^*}{\gamma!\,(p+1)! \int K_\gamma^* K_\nu^*} \right\}^{2(\gamma+\nu)+1/(p+\gamma+2)} n^{-2(p-\nu)+3/(p+\gamma+2)}$$

$$+ \frac{4}{n} \left\{ \int f (f^{(\gamma+\nu)})^2 - \theta_{\gamma,\nu}^2 \right\}. \tag{12}$$

Remark 3. Expression (12), which is yielded by the Jones–Sheather mean cancellation technique, has a better rate of convergence than expression (11) has. Hence, if a more general class of densities is taken into account, there is a distinction in the rates of convergence between easy and difficult densities.

Remark 4. If $\int K_\gamma^* K_\nu^*$ and $\int u^{p+1} K_\nu^*$ have different signs, then expression (11) holds when $\theta_{\gamma,p+1} < 0$ and expression (12) is valid when $\theta_{\gamma,p+1} > 0$. If $K(t) = (1-t^2)\, I_{[-1,+1]}(t)$, choosing p such that $(p+1-\gamma)/2$ is an odd integer is equivalent to $\int K_\gamma^* K_\nu^*$ having the same sign as $\int u^{p+1} K_\nu^*$. Besides, if f has no important features at the boundaries,

$$\theta_{\gamma,p+1} = (-1)^{(p+1-\gamma)/2} \int (f^{(p+\gamma+1)/2})^2$$

which is negative if and only if $(p+1-\gamma)/2$ is odd. Hence, under such circumstances, we can always obtain expression (12).

Remark 5. From expressions (11) and (12), $E\{\hat{\theta}_{\gamma,\nu}(a_{\mathrm{AMSE}})-\theta_{\gamma,\nu}\}^2$ is asymptotically

$$\sigma^2_{\gamma,\nu}\equiv 4\left\{\int f(f^{(\gamma+\nu)})^2-\theta^2_{\gamma,\nu}\right\}$$

under the condition that $p>\gamma+2\nu$ if $\theta_{\gamma,p+1}>0$ and $p+1>\gamma+2\nu$ if $\theta_{\gamma,p+1}<0$. Note that, when the support of f is unbounded, $(\sigma^2_{\gamma,\gamma})^{-1}$ is the same as the information bound given in Bickel and Ritov (1988). The requirement that p (or $p+1$) $>3\gamma$ for $\hat{\theta}_{\gamma,\gamma}(a)$ to achieve the optimal $\sqrt{n}$-convergence is more than what is needed for the estimator of Bickel and Ritov (1988). However, if $f\in F_{p+1}$ and there is any non-smooth boundary, estimators in Bickel and Ritov (1988) become less efficient but $\hat{\theta}_{\gamma,\gamma}(a)$ still can achieve $\sqrt{n}$-convergence. Given these results, there are two possible generalizations of the information bound results in Bickel and Ritov (1988). One is to consider cases where $\gamma\neq\nu$, and the other is to include densities that are smooth except for having possible discontinuous derivatives on a finite set.

4. APPLICATIONS TO AUTOMATIC BANDWIDTH SELECTION

The results given in corollary 1 are useful for automatic bandwidth selection. Recall from theorem 1 that local linear estimators of $f(x)$ achieve boundary corrections automatically. Hence, in the presence of non-smooth boundaries, local linear density estimators are preferable to conventional kernel density estimators.

The behaviour of $\hat{f}_h(x)$ is mainly decided by its bandwidth h since it controls the amount of smoothing. A proper choice of the bandwidth depends on derivatives of the unknown density. In non-boundary cases, $\hat{f}_h(x)$ is essentially the same as the usual kernel density estimator with kernel K; see equation (5). Hence any bandwidth procedure for kernel density estimation will be appropriate. But, as discussed in Section 1, those bandwidth rules will not be adequate if there are non-smooth boundaries. Since the estimators $\hat{\theta}_{\gamma,\nu}$ are less sensitive to boundary effects, they will hopefully provide a better alternative when replacing the usual pilot estimators in the Sheather–Jones selector. Next we construct a data-based bandwidth for local linear density estimation based on the estimators $\hat{\theta}_{\gamma,\nu}$ and corollary 1. The development is very similar to that of the Sheather–Jones procedure.

For any function ψ on the real line, denote $\int\psi(t)^2\,\mathrm{d}t$ as $R(\psi)$. The mean integrated squared error (MISE) of the local linear density estimator $\hat{f}_h$ is asymptotically

$$\frac{h^4}{4}\left(\int u^2K\right)^2\theta_{2,2}+\frac{1}{nh}R(K),$$

as $n\rightarrow\infty$, $h\rightarrow 0$, $nh\rightarrow\infty$ and $b/h\rightarrow 0$; see Cheng (1994). Its minimizer with respect to h is

$$h_*=\left\{R(K)\bigg/\left(\int u^2K\right)^2\theta_{2,2}\right\}^{1/5}n^{-1/5}, \tag{13}$$

which is referred to as the asymptotically optimal bandwidth for local linear density estimation. Note that h_* is unknown since it depends on the quantity $\theta_{2,2}$.

To estimate $\theta_{2,2}$ we take $p = 3$ and obtain $\hat{\theta}_{2,2}(a)$. At this stage, a proper choice of the bandwidth a is needed. According to corollary 1, the asymptotically optimal bandwidth for $\hat{\theta}_{2,2}(a)$ is

$$a_* = \left\{ 24\chi\, R(K_2^*) \Big/ \theta_{2,4} \int u^4 K_2^* \right\}^{1/7} n^{-1/7}, \tag{14}$$

where

$$\chi = \begin{cases} -1, & \text{if } \theta_{2,4} < 0, \\ 5/2, & \text{if } \theta_{2,4} > 0. \end{cases}$$

From equations (13) and (14),

$$a_* = C(K)\, D(f) h_*^{5/7}, \tag{15}$$

where

$$C(K) = \left\{ \frac{24\, R(K_2^*) \left(\int u^2 K \right)^2}{R(K) \int u^4 K_2^*} \right\}^{1/7}, \qquad D(f) = \left(\frac{\chi \theta_{2,2}}{\theta_{2,4}} \right)^{1/7}.$$

Following the ideas of plug-in estimation of h_*, replace $\theta_{2,2}$ of equation (13) by $\hat{\theta}_{2,2}\{a(h)\}$ and find the solution in h of the equation

$$h = \left[R(K) \Big/ \left(\int u^2 K \right)^2 \hat{\theta}_{2,2}\{a(h)\} \right]^{1/5} n^{-1/5}, \tag{16}$$

where

$$a(h) = C(K)\, D(f) h^{5/7}. \tag{17}$$

Here $D(f)$ involves $\theta_{2,2}$ and $\theta_{2,4}$ which are again unknown, but it can be estimated by some reference value through a scale parametric model of f. Let η_1 be a fixed density function, e.g. the standard normal density, that has been normalized so that some measure of scale such as the standard deviation is equal to 1. It can be verified that $D(\eta_\lambda) = \lambda^{2/7} D(\eta_1)$, where $\eta_\lambda(\cdot) = \lambda^{-1} \eta_1(\cdot/\lambda)$. Hence, from equation (17), set

$$a_\lambda(h) = C(K)\, D(\eta_1) \lambda^{2/7} h^{5/7}.$$

Our plug-in bandwidth selector, denoted as $\hat{h}_1$, is defined to be the solution of the equation analogous to equation (16) with $a(h)$ replaced by $a_{\hat{\lambda}}(h)$, where $\hat{\lambda}$ is a $\sqrt{n}$-consistent estimate of λ.

$\hat{h}_1$ involves only one-step estimation of $\theta_{\gamma,\nu}$, i.e. $D(f)$ of equation (17) is directly estimated by a reference value. Conventional plug-in procedures use reference values at one stage later. For example, in the Sheather–Jones procedure, the bandwidth in

$\tilde{\theta}_{2,2}$ contains unknown quantities $\theta_{2,2}$ and $\theta_{3,3}$. They are estimated by some kernel estimators whose bandwidths depend on appropriate reference values.

We do not implement two steps of estimation in $\hat{h}_1$ since that requires the estimate $\hat{\theta}_{2,4}$ which would make the procedure disadvantageous from the following two aspects. First, $\hat{\theta}_{2,4}$ requires computing the inverses of g matrices each of size 6×6 (see equations (3) and (7)), which will make the implementation extremely slow. Second, $\hat{\theta}_{2,4}$ involves $\hat{f}^{(4)}(x)$, which inherits a large variability from the data; hence it makes the bandwidth procedure unstable. An alternative to doing two estimation steps is to keep everything in the Sheather–Jones rule unchanged except replacing $\tilde{\theta}_{2,2}$ by $\hat{\theta}_{2,2}$. The resulting bandwidth is denoted as $\hat{h}_2$ and the original Sheather–Jones bandwidth as $\hat{h}_{\text{SJPI}}$.

A simulation study was conducted to examine the performance of $\hat{h}_1$, $\hat{h}_2$ and $\hat{h}_{\text{SJPI}}$ in practice. 10 independent samples each of size 100 were drawn from the χ^2-distribution with 2 degrees of freedom. For each sample, the bandwidths were computed and local linear density estimates were implemented based on them. Fig. 1 depicts the local linear density estimates using bandwidths $\hat{h}_{\text{SJPI}}$ and $\hat{h}_1$.

The conventional bandwidth $\hat{h}_{\text{SJPI}}$ is too small and yields undersmoothed estimates of the χ_2^2-density, which is discontinuous at 0. The proposed bandwidth rule $\hat{h}_1$ achieves boundary correction and produces reasonably good estimates of the underlying density. Estimates based on $\hat{h}_2$ are not shown here, but they are close to those based on $\hat{h}_1$. Similar results were observed from a separate simulation in the uniform[0,1] case.

We further investigated the bandwidths by another simulation study. The density was χ_1^2, χ_2^2, χ_3^2 or χ_5^2. The sample size was 100 or 1000. For χ_1^2, χ_3^2 and χ_5^2, the asymptotically optimal bandwidth h_* does not exist, since the quantity $\theta_{2,2}$ is infinite. Hence the bandwidths rules $\hat{h}_1$, $\hat{h}_2$ and $\hat{h}_{\text{SJPI}}$ were examined by MISEs of local linear density estimators based on them. Each MISE was estimated as the average of 1000 realizations of the integrated squared error, the values of which were calculated by the trapezoidal rule from corresponding errors evaluated on a grid of 600 equally spaced points. Table 1 summarizes the results.

Under the χ_2^2-distribution, the proposed bandwidths achieve a substantial reduction in MISE (about 30% when $n = 100$ and 50% when $n = 1000$) compared with $\hat{h}_{\text{SJPI}}$. In the other cases, the three bandwidths give very close MISE values. Indeed, all are too small when applied to χ_1^2-samples. The reason is that the χ_1^2-

TABLE 1

Approximate MISEs of local linear estimators based on $\hat{h}_{\text{SJPI}}$, $\hat{h}_1$ or $\hat{h}_2$, for χ^2-samples of size 100 or 1000

Degrees of freedom	*Sample size*	*MISE($\hat{h}_{\text{SJPI}}$)*	*MISE($\hat{h}_1$)*	*MISE($\hat{h}_2$)*
1	100	0.08365	0.08719	0.08797
1	1000	0.04753	0.04821	0.04787
2	100	0.00796	0.00541	0.00578
2	1000	0.00174	0.00083	0.00088
3	100	0.00525	0.00552	0.00566
3	1000	0.00109	0.00109	0.00109
5	100	0.00286	0.00298	0.00307
5	1000	0.00051	0.00052	0.00052

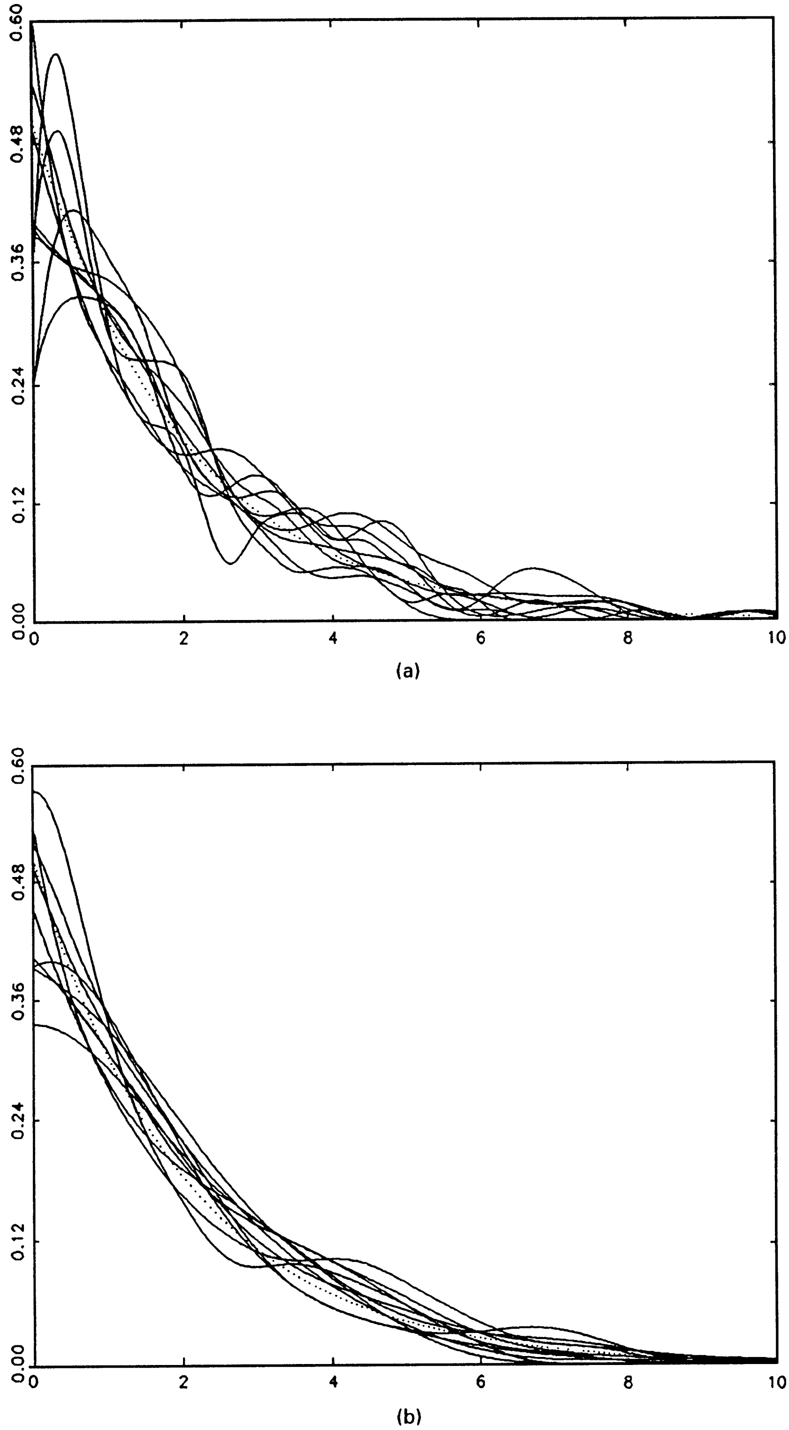

Fig. 1. Local linear density estimates (———) of the χ_2^2-density (·········) using different data-based bandwidths: (a) Sheather–Jones bandwidth $\widehat{h}_{SJPI}$; (b) proposed bandwidth $\widehat{h}_1$ (sample size $n = 100$)

density has a sharp spike at the boundary which makes automatic bandwidth selection extremely difficult. As for χ_3^2- and χ_5^2-samples, the performances of the bandwidth procedures are about the same and yield reasonably good estimates of the density.

We conclude from our simulation results that

(a) $\hat{h}_1$ improves $\hat{h}_{SJPI}$ when the underlying densities have non-smooth boundaries but without tough features there,
(b) $\hat{h}_1, \hat{h}_2$ and $\hat{h}_{SJPI}$ are competitive when there is either no important feature near the boundaries or no boundary and
(c) there is no apparent advantage of $\hat{h}_2$ over $\hat{h}_1$.

Asymptotic properties of $\hat{h}_1$ were studied in Cheng (1995). Ruppert *et al.* (1995) developed a bandwidth selector, called $\hat{h}_{STE}$, for local linear regression. The procedure utilized parallel estimators of $\hat{\theta}_{\gamma,\nu}(a)$ and its performance was demonstrated by simulation studies. We conjecture that a theoretical justification for this bandwidth selector is possible and is analogous to that for $\hat{h}_1$.

ACKNOWLEDGEMENTS

The author would like to thank a referee for helpful suggestions including the bandwidth rule $\hat{h}_2$. This research is part of the author's thesis under the supervision of Professor J. S. Marron at the University of North Carolina.

REFERENCES

Aldershof, B. (1991) Estimation of integrated squared density derivatives. *Mimeo 2053*. Institute of Statistics, University of North Carolina, Chapel Hill.

Bickel, P. J. and Ritov, Y. (1988) Estimating integrated squared density derivatives: sharp best order of convergence estimates. *Sankhya* A, **50**, 381–393.

Cheng, M.-Y. (1994) On boundary effects of smooth curve estimators. *Mimeo 2319*. Institute of Statistics, University of North Carolina, Chapel Hill.

——(1995) A bandwidth selector for local linear density estimators. *Ann. Statist.*, to be published.

Cheng, M.-Y., Fan, J. and Marron, J. S. (1995) On automatic boundary corrections. Submitted to *Ann. Statist.*

Fan, J., Gasser, T., Gijbels, I., Brockmann, M. and Engel, J. (1993) Local polynomial fitting: a standard for nonparametric regression. *Mimeo 2302*. Institute of Statistics, University of North Carolina, Chapel Hill.

Gasser, T., Müller, H.-G. and Mammitzsch, V. (1985) Kernels for nonparametric curve estimation. *J. R. Statist. Soc.* B, **47**, 238–252.

Hall, P. and Marron, J. S. (1987) Estimation of integrated squared density derivatives. *Statist. Probab. Lett.*, **6**, 109–115.

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

Jones, M. C., Marron, J. S. and Sheather, S. J. (1994) Progress in data-based bandwidth selection for kernel density estimation. *Mimeo 2088*. Institute of Statistics, University of North Carolina, Chapel Hill.

Jones, M. C. and Sheather, S. J. (1991) Using non-stochastic terms to advantage in kernel-based estimation of integrated squared density derivatives. *Statist. Probab. Lett.*, **11**, 511–514.

Lejeune, M. (1984) Optimization in non-parametric regression. In *COMPSTAT 1984—Proc. 6th Symp. Computing in Statistics* (eds T. Havranek, Z. Sidak and M. Novak), pp. 421–426. Heidelberg: Physica.

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

Müller, H. G. (1987) Weighted local regression and kernel methods for nonparametric curve fitting. *J. Am. Statist. Ass.*, **82**, 231–238.

Ruppert, D., Sheather, S. J. and Wand, M. P. (1995) An effective bandwidth selector for local least squares regression. *J. Am. Statist. Ass.*, **90**, 1257–1270.

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

Sheather, S. J. (1992) The performance of six popular bandwidth selection methods on some real data sets. *Comput. Statist.*, **7**, 225–250.

Sheather, S. J. and Jones, M. C. (1991) A reliable data-based bandwidth selection method for kernel density estimation. *J. R. Statist. Soc.* B, **53**, 683–690.

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

Van Es, A. J. and Hoogstrate, A. J. (1994) Kernel estimators of integrated squared density derivatives in non-smooth cases. In *Asymptotic Statistics: Proc. 5th Prague Symp.* (eds P. Mandl and M. Hušková), pp. 313–224. Heidelberg: Physica.