Ann. Inst. Statist. Math.
Vol. 51, No. 3, 541–569 (1999)

COUNTING BUMPS

RICARDO FRAIMAN[1] AND JEAN MELOCHE[2]

[1] *Centro de Matemática, Universidad de la República, Eduardo Acevedo 1139, Montevideo, 11200, Uruguay and Universidad de San Andrés, Buenos Aires, Argentina*
[2] *Department of Statistics, University of British Columbia, 6356 Agricultural Road, Vancouver, BC, Canada V6T 1Z2*

(Received September 27, 1996; revised December 15, 1997)

Abstract. The number of modes of a density f can be estimated by counting the number of 0-downcrossings of an estimate of the derivative f', but this often results in an overestimate because random fluctuations of the estimate in the neighbourhood of points where f is nearly constant will induce spurious counts. Instead of counting the number of 0-downcrossings, we count the number of "significant" modes by counting the number of downcrossings of an interval $[-\epsilon, \epsilon]$. We obtain consistent estimates and confidence intervals for the number of "significant" modes. By letting ϵ converge slowly to zero, we get consistent estimates of the number of modes. The same approach can be used to estimate the number of critical points of any derivative of a density function, and in particular the number of inflection points.

Key words and phrases: Significant bumps, density estimation, downcrossings, confidence intervals, bandwidth selection.

0. Introduction

In this paper, we consider the problem of estimating the number of modes of a density in a nonparametric framework. This problem has been considered by several authors, including Cox (1966), Good and Gaskins (1980), Silverman (1980, 1981, 1983, 1986) and Hartigan and Hartigan (1985). See also Wong (1982) and Wong and Lane (1983) where density estimation is used for identifying high density clusters. More recently, Donoho (1988) has discussed this problem in the general setting of inference about functionals of an unknown density.

Throughout the text, $\mathcal{D}(\mathcal{I})$ is used to denote the set of all densities on the interval $\mathcal{I}$, $\mathcal{C}(\mathcal{I})$ the set of all bounded and continuous functions on $\mathcal{I}$ and $\mathcal{CD}(\mathcal{I})$ the set of all bounded and continuous densities on $\mathcal{I}$. We say that a mode of f is a maximal closed interval I such that, for all $x \in I$, x is a local maximum of f. (Of course x is said to be a local maximum of f when there exists a neigbourhood $U = U(x)$ such that $f(z) \leq f(x)$ for all $z \in U$.) The number of modes $\rho^* = \rho^*(f)$

of a density is the cardinality of the set of its modes. A flat is a positive length interval on which f is constant.

According to our definition, the uniform density has only one mode and the approach we take allows for such densities with flat parts. This is highly desirable since, as will be shown in Section 1, one cannot determine empirically whether or not a density has such flat parts. This negative result, as well as others we prove in Section 1, are due to the fact that the shape of a density is a discontinuous functional on the space of all densities. However, as shown by Donoho (1988), ρ^* is a lower semi-continuous functional and thus, one-sided inference about ρ^* is possible.

The usual estimate of the number of modes of an unknown density f is the number of modes of some estimate of the density. However, the results obtained by Mammen *et al.* (1992) suggest that such an attempt based on kernel density estimates will fail. Our approach deals directly with that problem.

Donoho (1988) introduced a general method, the neighbourhood procedure, for estimating functionals such as ρ^*. His estimate is the number of modes of the distribution with fewest modes in an ϵ-neighbourhood (in the Kolmogorov distance) of the empirical distribution. By letting ϵ converge to zero slowly while n goes to infinity, this method provides a universally consistent estimate of ρ^*.

Our approach is somewhat related to Donoho's. Instead of looking at the 0-downcrossings of an estimate of f', we look at the ϵ-downcrossings, that is the downcrossings of the interval $[-\epsilon, \epsilon]$. In other words, we only count the number of "substantial" modes $\rho(\mathcal{S}(f,\epsilon))$ of a density estimate. Although it will be formally defined in Section 2, one can think of $\rho(\mathcal{S}(f,\epsilon))$ as the number of modes that can be detected by a computer with a precision of ϵ. This interpretation is useful since, in practice, one is interested in $\rho(\mathcal{S}(f,\epsilon))$—for some fixed and small value of ϵ—rather than in ρ^*. Indeed, when $\rho(\mathcal{S}(f, 0.001)) \neq \rho^*$, the missing modes are not likely to be of any interest. By letting ϵ converge slowly to zero, we get, like Donoho, consistent estimates of ρ^*.

In Section 1, we will show that the existence of a flat part and that the finiteness of the number of modes cannot be empirically verified. Such results are obtained by using the LeCam and Schwartz (1960) necessary and sufficient condition for the existence of consistent estimates. In Section 2, some general consistency results on a method to estimate the number of modes ρ^* based on the number of substantial modes $\rho(\mathcal{S}(f,\epsilon))$ are given. A conservative confidence band for the number of substantial modes $\rho(\mathcal{S}(f,\epsilon))$ is also obtained. Recall that Donoho (1988) has shown that one cannot give a nonparametric confidence band for $\rho^*(f)$. However, in practice, we are mostly interested in the number of substantial modes. In Section 3, we apply the general results of Section 2 to histogram and kernel based estimates of the derivative f'. The results are given for the i.i.d. case and under the assumption that f is smooth mainly for the sake of simplicity. The results of Section 2 can be applied to any consistent estimate of f'; all we need is a bound on the bias $\mathcal{E}\hat{f}' - f'$ and a good exponential inequality for the empirical process. Section 4 is devoted to examples and practical considerations. Finally, Section 5 takes a closer look at the question of bandwidth selection.

1. Estimability

In this section we will show that the functionals

$$\alpha(f) = \begin{cases} 1 & \text{if } f \text{ has a flat part,} \\ 0 & \text{otherwise,} \end{cases}$$

and

$$\beta(f) = \begin{cases} 1 & \text{if } f \text{ has finitely many modes,} \\ 0 & \text{otherwise,} \end{cases}$$

cannot be consistently estimated from a sequence of i.i.d. observations. LeCam and Schwartz (1960) gave a necessary and sufficient condition for the existence of consistent estimates. Roughly speaking, their results imply that if the functional τ has a consistent estimate, τ must be continuous in a dense subset with respect to the topology induced by the $\mathcal{L}^1$ norm

$$\|f\|_1 = \int_{-\infty}^{\infty} f(t)dt.$$

It may be surprising that τ need not actually be continuous but even the mean $\mu(f) = \int x f(x)dx$ does not have this property. The following lemma puts the argument in a context appropriate to our discussion and is proved for the sake of completeness.

LEMMA 1.1. *Let $\mathcal{B}$ be a subset of densities equipped with some norm $\|\cdot\|$ that makes $(\mathcal{B}, \|\cdot\|)$ a complete metric space. Assume that $\|f\|_1 \le c\|f\|$ for some constant c and all $f \in \mathcal{B}$. Let*

$$\phi : \mathcal{B} \mapsto [-K, K]$$

be any bounded characteristic of the densities in $\mathcal{B}$. If ϕ is consistently estimable on $\mathcal{B}$, then there exists a dense subset of points in $\mathcal{B}$ at which ϕ is continuous with respect to the topology induced by the $\|\cdot\|$ norm. Therefore, if ϕ is discontinuous at every point in $\mathcal{B}$, it is not consistently estimable on $\mathcal{B}$.

PROOF. If ϕ is consistently estimable on $\mathcal{B}$, there exists a sequence of statistics $T_n : \mathcal{R}^n \mapsto \mathcal{R}$ such that for all $f \in \mathcal{B}$, if $X_1, X_2, \ldots$ are independent random variables with common density f,

$$T_n(X_1, \ldots, X_n) \xrightarrow{\mathcal{P}_f} \phi(f).$$

Since $|\phi|$ is bounded by K,

$$S_n = T_n I_{|T_n| \le K} + sg(T_n) K I_{|T_n| > K}$$

is also a consistent estimate of ϕ on $\mathcal{B}$, where I_A stands for the indicator function of the set A and $\mathrm{sg}(\cdot)$ for the sign function. Indeed, for all $f \in \mathcal{B}$,

$$\begin{aligned}\mathcal{P}_f\{|S_n - \phi(f)| > \epsilon\} &= \mathcal{P}_f\{|S_n - \phi(f)| > \epsilon, |T_n| \le K\} \\ &\quad + \mathcal{P}_f\{|S_n - \phi(f)| > \epsilon, |T_n| > K\} \\ &\le \mathcal{P}_f\{|T_n - \phi(f)| > \epsilon\} + \mathcal{P}_f\{|K - \phi(f)| > \epsilon, |T_n| > K\} \\ &\le 2\mathcal{P}_f\{|T_n - \phi(f)| > \epsilon\}.\end{aligned}$$

Define $\phi_n : \mathcal{B} \mapsto [-K, K]$ by the equation

$$\phi_n(f) = \mathcal{E}_f S_n(X_1, \ldots, X_n).$$

Because of the boundedness of S_n, for all $f \in \mathcal{B}$

$$\lim_{n\to\infty} \phi_n(f) = \phi(f).$$

In addition, the inequalities

$$\begin{aligned}|\phi_n(f) - \phi_n(g)| &= |\mathcal{E}_f S_n(X_1, \ldots, X_n) - \mathcal{E}_g S_n(X_1, \ldots, X_n)| \\ &= \left| \int_{-\infty}^{+\infty} dx_n \cdots \int_{-\infty}^{+\infty} dx_1 S_n(x_1, \ldots, x_n) \right. \\ &\qquad \left. \cdot \{f(x_1) \cdots f(x_n) - g(x_1) \cdots g(x_n)\} \right| \\ &\le K \int_{-\infty}^{+\infty} dx_n \cdots \int_{-\infty}^{+\infty} dx_1 |f(x_1) \cdots f(x_n) - g(x_1) \cdots g(x_n)| \\ &\le K \int_{-\infty}^{+\infty} dx_n \cdots \int_{-\infty}^{+\infty} dx_1 f(x_1) \\ &\quad \cdot |f(x_2) \cdots f(x_n) - g(x_2) \cdots g(x_n)| \\ &\quad + K \int_{-\infty}^{+\infty} dx_n \cdots \int_{-\infty}^{+\infty} dx_1 |f(x_1) - g(x_1)| g(x_2) \cdots g(x_n) \\ &= K \int_{-\infty}^{+\infty} dx_n \cdots \int_{-\infty}^{+\infty} dx_2 |f(x_2) \cdots f(x_n) - g(x_2) \cdots g(x_n)| \\ &\quad + K\|f - g\|_1 \\ &\le nK\|f - g\|_1 \\ &\le ncK\|f - g\|,\end{aligned}$$

holding for all $f, g \in \mathcal{B}$, we conclude that, for each n, ϕ_n is uniformly continuous with respect to the topology induced by the norm $\|\cdot\|$ on $\mathcal{B}$.

According to the Baire category theorem, since $(\mathcal{B}, \|\cdot\|)$ is a complete metric space, and since ϕ is the limit of a sequence of continuous functions, it must be continuous on a dense subset of $\mathcal{B}$.

When proving that a parameter is not consistently estimable, it is often easier to restrict attention to a conveniently chosen subset of densities and to work with a continuous transformation of the parameter. The following trivial lemma can be useful.

LEMMA 1.2. *Assume $\phi \mapsto [-K, K]$ is not consistently estimable on $\mathcal{B}$. If H is a continuous transformation and if $\phi = H \circ \bar{\phi}$ then $\bar{\phi}$ is not consistently estimable on any set containing $\mathcal{B}$.*

LEMMA 1.3. $(\mathcal{CD}([0,1]), \|\cdot\|_\infty)$ *is a complete metric space.*

PROOF. Since $(\mathcal{C}([0,1]), \|\cdot\|_\infty)$ is a complete metric space, it suffices to show that $\mathcal{CD}([0,1])$ is a closed subset of $\mathcal{C}([0,1])$. This follows from Scheffe's Theorem.

Since $\forall f \in \mathcal{CD}([0,1])\ \|f\|_1 \leq \|f\|_\infty$, it follows from Lemma 1.1 that if a bounded functional τ is consistently estimable, it must be continuous with respect to the topology induced by the norm $\|\cdot\|_\infty$ at a dense subset of points in $\mathcal{CD}([0,1])$. However,

LEMMA 1.4. *The functionals α and β are discontinuous with respect to the topology induced by the norm $\|\cdot\|_\infty$ at every point in $\mathcal{CD}([0,1])$.*

PROOF. First consider the functional α. Take any $f \in \mathcal{CD}([0,1])$ with $\alpha(f) = 0$ and define

$$f_n = \begin{cases} \dfrac{f\left(\frac{1}{n}\right)}{\frac{1}{n} f\left(\frac{1}{n}\right) + \int_{1/n}^1 f}, & \text{if } x < \dfrac{1}{n}; \\[2ex] \dfrac{f(x)}{\frac{1}{n} f\left(\frac{1}{n}\right) + \int_{1/n}^1 f}, & \text{otherwise.} \end{cases}$$

On one hand we have

$$|\alpha(f_n) - \alpha(f)| = 1 \qquad \forall n$$

while

$$\|f_n - f\|_\infty \to 0,$$

showing that α is discontinuous at f. Take now any $f \in \mathcal{CD}([0,1])$ with $\alpha(f) = 1$ and put

$$f_n = \frac{f * \eta_{1/n}}{\int_0^1 f * \eta_{1/n}}$$

where η_σ is the normal density with mean zero and variance σ^2. Because the location family of the normal densities is complete (Lehmann (1959), Theorem 4.3.1, p. 132), it follows that f_n has no flat part. On one hand we have

$$|\alpha(f_n) - \alpha(f)| = 1 \qquad \forall n$$

while, since f is uniformly continuous on $[0,1]$,

$$\|f_n - f\|_\infty \to 0,$$

showing that α is discontinuous at f.

Next, consider the functional β. Take any $f \in \mathcal{CD}([0,1])$ with $\beta(f) = 0$ and define f_n to be the linear interpolation of f with knots at $0, \frac{1}{n}, \ldots, 1$ and renormalized so that it integrates to one. The density f_n has a finite number of bumps. On one hand we have

$$|\beta(f_n) - \beta(f)| = 1 \qquad \forall n$$

while, since f is uniformly continuous on $[0,1]$,

$$\|f_n - f\|_\infty \to 0,$$

showing that β is discontinuous at f. Take now any $f \in \mathcal{CD}([0,1])$ with $\beta(f) = 1$ and let $g \in \mathcal{CD}([0,1])$ be any density with $\beta(g) = 0$. Define

$$f_n = \left(1 - \frac{1}{n}\right) f + \frac{1}{n} g.$$

On one hand we have

$$|\beta(f_n) - \beta(f)| = 1 \qquad \forall n$$

while

$$\|f_n - f\|_\infty \to 0,$$

showing that β is discontinuous at f. This proves that β is discontinuous at every point $f \in \mathcal{CD}([0,1])$.

THEOREM 1.5. *If $\mathcal{CD}(\mathcal{I}) \subset \mathcal{B}$, there exist no estimate $\hat{T}_n = \hat{T}_n(X_1, \ldots, X_n)$ such that*

$$\hat{T}_n \xrightarrow{\mathcal{P}_f} \alpha(f) \qquad \forall f \in \mathcal{B}$$

and no estimate $\hat{T}_n = \hat{T}_n(X_1, \ldots, X_n)$ such that

$$\hat{T}_n \xrightarrow{\mathcal{P}_f} \beta(f) \qquad \forall f \in \mathcal{B}.$$

2. General results

In this section, f and g are differentiable functions. Although our argument can be extended to functions with left and right sided derivatives, the exposition is clearer if the unknown functions and their estimates are presumed to be smooth. We use the notation $\hat{f}_n$ to represent an arbitrary differentiable estimate of f, so that $\hat{f}'_n$ is used to estimate f'. The results in this Section can be applied to estimate the number of modes of an unknown density as well as to the number of local maxima of a regression function. In the next section, we will apply the results to kernel based and histogram based estimates of the derivative f'.

To estimate the number of modes of an unknown density f, it is a reasonable idea to use the number of modes of some estimate of the density. The problem with that approach is that in the neighbourhood of points where f is non-zero and nearly constant, the density estimate will typically have many modes. This phenomenon is observed, for example, for uniformly distributed data. The approach we take is to count only the "substantial" modes of the estimate. Formally, instead of counting the number of 0-downcrossings of the derivative $\hat{f}'_n$, we count the number of downcrossings of the interval $[-\epsilon, \epsilon]$. To formalize this approach, we propose the following definitions.

Given any differentiable function f, define

$$\mathcal{S}(f,\epsilon)(x) = \begin{cases} -1, & \text{if } f'(x) < -\epsilon; \\ 0, & \text{if } |f'(x)| \leq \epsilon; \\ 1, & \text{if } f'(x) > \epsilon. \end{cases}$$

It looks very much like the sign function except zero has been replaced by the whole interval $[-\epsilon, \epsilon]$. Given any function s with values in $\{-1, 0, 1\}$ we say that the interval $[a, b]$ is a 0-downcrossing of s if

$$s(a) = 1 \quad \text{and} \quad s(b) = -1.$$

Obviously, 0-downcrossings can overlap and we define $\rho(s)$ as the maximum number of disjoint 0-downcrossings of s. Furthermore, if s_1 and s_2 are two functions with values in $\{-1, 0, 1\}$, we write $s_1 \prec s_2$ whenever each 0-downcrossing of s_1 contains a 0-downcrossing of s_2. Finally, we will use the sup norm:

$$\|f\|_\infty = \sup_x |f(x)|.$$

With these notations, we have

LEMMA 2.1.
i) *If* $s_1 \prec s_2$ *then* $\rho(s_1) \leq \rho(s_2)$.
ii) $s_1 \prec s_2$ *follows from the conditions*

$$s_1(x) = \pm 1 \Rightarrow s_2(x) = \pm 1.$$

iii) *If* $0 < \epsilon < \epsilon'$ *then*

$$\mathcal{S}(f, \epsilon') \prec \mathcal{S}(f, \epsilon)$$

and

$$\rho(\mathcal{S}(f, \epsilon')) \leq \rho(\mathcal{S}(f, \epsilon)).$$

iv) *If* $\|f' - g'\|_\infty < \frac{\epsilon}{4}$ *then*

$$\mathcal{S}\left(f, \frac{3\epsilon}{2}\right) \prec \mathcal{S}(g, \epsilon) \prec \mathcal{S}\left(f, \frac{\epsilon}{2}\right)$$

and

$$\rho\left(\mathcal{S}\left(f,\frac{3\epsilon}{2}\right)\right) \le \rho(\mathcal{S}(g,\epsilon)) \le \rho\left(\mathcal{S}\left(f,\frac{\epsilon}{2}\right)\right).$$

PROOF. The first statement follows directly from the definition of ρ as the number of disjoint 0-downcrossings. For the second, observe that if $s_2(a) = 1$ and $s_2(b) = -1$, the interval $[a,b]$ contains at least one 0-downcrossing of s_2. The third follows from the first two statements, the implications

$$\mathcal{S}(f,\epsilon')(x) = 1 \Rightarrow f'(x) > \epsilon' \Rightarrow f'(x) > \epsilon \Rightarrow \mathcal{S}(f,\epsilon)(x) = 1,$$

and similar ones for -1. Finally, the fourth statement follows from the first two statements, the implications

$$\mathcal{S}\left(f,\frac{3\epsilon}{2}\right)(x) = 1 \Rightarrow f'(x) > \frac{3\epsilon}{2} \Rightarrow g'(x) > \epsilon \Rightarrow \mathcal{S}(g,\epsilon)(x) = 1,$$
$$\Rightarrow g'(x) > \epsilon \Rightarrow f'(x) > \frac{\epsilon}{2} \Rightarrow \mathcal{S}\left(f,\frac{\epsilon}{2}\right)(x) = 1,$$

and similar ones for -1.

It is useful to have a clear idea of what the quantity $\rho(\mathcal{S}(f,\epsilon))$ represents. It is the number of ϵ-downcrossings of the derivative f'. It will coincide with the number of modes $\rho^*(f)$ of f if and only if for each mode $[a,b]$ of f there exists an interval $[c,d] \supset [a,b]$ such that $f'(c) > \epsilon$, $f'(d) < -\epsilon$ and $[c,d]$ contains no other mode than $[a,b]$.

LEMMA 2.2. *For all* $\epsilon > 0$

$$\rho(\mathcal{S}(f,\epsilon)) \le \rho^*(f).$$

PROOF. With each disjoint 0-downcrossing of $\mathcal{S}(f,\epsilon)$ we will associate a distinct mode of f. Assume $[a,c]$ is such a 0-downcrossing of $\mathcal{S}(f,\epsilon)$. Since

$$f'(a) > \epsilon \quad \text{and} \quad f'(c) < -\epsilon,$$

there exist a point $b \in (a,c)$ such that

$$f(b) > \max\{f(a), f(c)\}.$$

Let d be any point in (a,c) at which f attains its maximum. The largest closed interval containing d and on which f is constant is a mode and is included in (a,c).

LEMMA 2.3. *If* $\rho^*(f) < \infty$ *then there exists an* $\epsilon_0(f) > 0$ *such that for all* $\epsilon < \epsilon_0(f)$

$$\rho(\mathcal{S}(f,\epsilon)) = \rho^*(f).$$

PROOF. With each distinct mode of f we will associate disjoint 0-downcrossings of $\mathcal{S}(f,\epsilon)$. Let $[a_i,b_i]$ $1 \le i \le \rho^*$ be the modes of f and take disjoint open intervals around them:

$$[a_i,b_i] \subset (a_i-\epsilon, b_i+\epsilon) \qquad 1 \le i \le \rho^*.$$

For $1 \le i \le \rho^*$, there exist points $c_i \in (a_i-\epsilon, a_i)$ and $d_i \in (b_i, b_i+\epsilon)$ such that $f'(c_i) > 0$ and $f'(d_i) < 0$. Put

$$\epsilon_0(f) = \min_{1\le i\le \rho^*}\{f'(c_i), -f'(d_i)\}.$$

For all $\epsilon < \epsilon_0(f)$, each of the disjoint intervals $[c_i,d_i]$ are 0-downcrossings of $\mathcal{S}(f,\epsilon)$. The result follows from the inequality of Lemma 2.2.

THEOREM 2.4. *If* $\rho^*(f) < \infty$, *and* $\|f' - \hat{f}_n'\|_\infty < \infty$

$$\mathcal{P}\left\{\rho\left(\mathcal{S}\left(f,\frac{3\epsilon}{2}\right)\right) \le \rho(\mathcal{S}(\hat{f}_n,\epsilon)) \le \rho^*(f)\right\} \ge \mathcal{P}\left\{\|f'-\hat{f}_n'\|_\infty < \frac{\epsilon}{4}\right\}$$

and for all ϵ *sufficiently small*

$$\mathcal{P}\{\rho(\mathcal{S}(\hat{f}_n,\epsilon)) = \rho^*(f)\} \ge \mathcal{P}\left\{\|f'-\hat{f}_n'\|_\infty < \frac{\epsilon}{4}\right\}.$$

PROOF. Using the lemmas,

$$\begin{aligned}
&\mathcal{P}\left\{\|f'-\hat{f}_n'\|_\infty < \frac{\epsilon}{4}\right\}\\
&\le \mathcal{P}\left\{\mathcal{S}\left(f,\frac{3\epsilon}{2}\right) \prec \mathcal{S}(\hat{f}_n,\epsilon) \prec \mathcal{S}\left(f,\frac{\epsilon}{2}\right)\right\}\\
&\le \mathcal{P}\left\{\rho\left(\mathcal{S}\left(f,\frac{3\epsilon}{2}\right)\right) \le \rho(\mathcal{S}(\hat{f}_n,\epsilon)) \le \rho\left(\mathcal{S}\left(f,\frac{\epsilon}{2}\right)\right)\right\}\\
&\le \mathcal{P}\left\{\rho\left(\mathcal{S}\left(f,\frac{3\epsilon}{2}\right)\right) \le \rho(\mathcal{S}(\hat{f}_n,\epsilon)) \le \rho^*(f)\right\}.
\end{aligned}$$

The second inequality follows from Lemma 2.3.

Theorem 2.4 implies that as long as

$$\mathcal{P}\left\{\|f'-\hat{f}_n'\|_\infty > \frac{\epsilon_n}{4}\right\} \to 0,$$

$\rho(\mathcal{S}(\hat{f}_n,\epsilon_n))$ is a consistent estimate of $\rho^*(f)$. Furthermore, if ϵ is small enough that $\rho(\mathcal{S}(f,\frac{3\epsilon}{2})) = \rho^*(f)$, $\rho(\mathcal{S}(\hat{f}_n,\epsilon))$ is a consistent estimate of $\rho^*(f)$. Note that in this case, the estimate is based on a fixed ϵ. As stated in Lemma 2.3, for any given density f, there is such a value for ϵ.

THEOREM 2.5. *If* $\rho^*(f) < \infty$, *and* $\|f' - \hat{f}_n'\|_\infty < \infty$

$$\mathcal{P}\left\{\rho(\mathcal{S}(f,\epsilon)) \in \left[\rho\left(\mathcal{S}\left(\hat{f}_n, \frac{3\epsilon}{2}\right)\right), \rho\left(\mathcal{S}\left(\hat{f}_n, \frac{\epsilon}{2}\right)\right)\right]\right\} \geq \mathcal{P}\left\{\|f' - \hat{f}_n'\|_\infty < \frac{\epsilon}{4}\right\}.$$

PROOF. Using the lemmas,

$$\begin{aligned} &\mathcal{P}\left\{\|f' - \hat{f}_n'\|_\infty < \frac{\epsilon}{4}\right\} \\ &\quad\leq \mathcal{P}\left\{\mathcal{S}\left(\hat{f}_n, \frac{3\epsilon}{2}\right) \prec \mathcal{S}(f,\epsilon) \prec \mathcal{S}\left(\hat{f}_n, \frac{\epsilon}{2}\right)\right\} \\ &\quad\leq \mathcal{P}\left\{\rho\left(\mathcal{S}\left(\hat{f}_n, \frac{3\epsilon}{2}\right)\right) \leq \rho(\mathcal{S}(f,\epsilon)) \leq \rho\left(\mathcal{S}\left(\hat{f}_n, \frac{\epsilon}{2}\right)\right)\right\}. \end{aligned}$$

Theorem 2.5 gives a conservative confidence band for $\rho(\mathcal{S}(f,\epsilon))$ rather than for $\rho^*(f)$. As discussed earlier, $\rho(\mathcal{S}(f,\epsilon))$ is in a sense, the number of "substantial" modes of f. Donoho (1988) proves that one cannot give a non-parametric confidence band for $\rho^*(f)$. Theorem 2.4 gives a lower confidence bound for $\rho^*(f)$, which is possible as shown by Donoho.

The locations of the modes can also be useful. Although we did not look carefully into this matter, the 0-downcrossings of $\mathcal{S}(\hat{f}_n, \epsilon)$ can be used as confidence intervals for them.

The same approach can be used to estimate the number of critical points of any derivative, and in particular, the number of inflection points. Cuevas and Gonzalez Manteiga (1991) proposed a bandwidth selection method based on a known number of inflection points. With our approach, instead of looking at the number of 0-crossings of an estimate of f'', we look at the number of "substantial inflections" or ϵ-crossings. Our arguments can be modified by replacing f' by f'' and one easily verifies that Theorems 2.4 and 2.5 can be proved for the number of inflections points.

3. Histogram and kernel based estimate of f'

We now apply the theorems of the preceding section to histogram and kernel based estimates of the derivative f'. In order to apply our results, all we need is to bound

$$\mathcal{P}\left\{\|\hat{f}_n' - f'\|_\infty \leq \frac{\epsilon}{4}\right\}.$$

We first bound the bias and then use the DKW inequality (see for example Shorack and Wellner (1986), pp. 354–356) on the empirical process. As usual with the estimation of f', the bias

$$\mathcal{E}\hat{f}_n'(x) - f'(x)$$

depends heavily on the degree of smoothness of the unknown derivative f'. To show how smoothness can be handled, we consider two cases:

$$\text{(3.1)} \qquad f \text{ has two derivatives and } \|f''\|_\infty < \infty,$$

and

$$(3.2) \qquad f \text{ has three derivatives and } \|f'''\|_\infty < \infty.$$

For the histogram based estimate, in order to keep the notation simple, we will work on the interval $[0, 1]$. Divide $[0, 1]$ in c_n equal pieces and define

$$A_{n,i} = \left(\frac{i}{c_n}, \frac{i+1}{c_n}\right] \qquad -1 \le i \le c_n,$$

$$p_{n,i}(F) = \int_{A_{n,i}} dF \qquad -1 \le i \le c_n,$$

$$d_{n,i}(F) = c_n^2(p_{n,i}(F) - p_{n,i-1}(F)) \qquad 0 \le i \le c_n$$

and

$$\hat{f}_n'\left(\frac{i+t}{c_n}\right) = (1-t)d_{n,i}(\hat{F}_n) + td_{n,i+1}(\hat{F}_n) \qquad 0 \le t \le 1,$$

where $\hat{F}_n$ is the empirical cumulative distribution function based on the observations. With these definitions, $\hat{f}_n'$ is an appropriate estimate of the derivative f' and the quantity c_n^{-1} plays the role of a bandwidth.

LEMMA 3.1. *If* (3.1) *holds*,

$$\|\mathcal{E}\hat{f}_n' - f'\|_\infty \le \frac{13}{24}\frac{\|f''\|_\infty}{c_n}$$

and if (3.2) *holds*,

$$\|\mathcal{E}\hat{f}_n' - f'\|_\infty \le \frac{5}{24}\frac{\|f'''\|_\infty}{c_n^2}.$$

PROOF. By definition, for all $t \in [0, 1]$,

$$\mathcal{E}d_{n,i}(\hat{F}_n) = c_n^2\left\{\int_{A_{n,i}} dF - \int_{A_{n,i-1}} dF\right\} = \int_0^1 udu \int_{-1}^1 dv f'\left(\frac{i+uv}{c_n}\right)$$

so, by substitution,

$$\begin{aligned}\mathcal{E}\hat{f}_n'\left(\frac{i+t}{c_n}\right) &= (1-t)\mathcal{E}d_{n,i}(\hat{F}_n) + t\mathcal{E}d_{n,i+1}(\hat{F}_n)\\ &= \int_0^1 udu \int_{-1}^1 dv\left\{(1-t)f'\left(\frac{i+uv}{c_n}\right) + tf'\left(\frac{i+1+uv}{c_n}\right)\right\}.\end{aligned}$$

If (3.1) holds, we can perform a one term Taylor expansion and write

$$\begin{aligned}&\left|\mathcal{E}\hat{f}_n'\left(\frac{i+t}{c_n}\right) - f'\left(\frac{i+t}{c_n}\right)\right|\\ &\quad\le (1-t)\int_0^1 udu\int_{-1}^1 dv\left|f'\left(\frac{i+uv}{c_n}\right) - f'\left(\frac{i+t}{c_n}\right)\right|\\ &\qquad + t\int_0^1 udu\int_{-1}^1 dv\left|f'\left(\frac{i+1+uv}{c_n}\right) - f'\left(\frac{i+t}{c_n}\right)\right|\\ &\quad\le \frac{\|f''\|_\infty}{c_n}\int_0^1 udu\int_{-1}^1 dv\{(1-t)|uv-t| + t|uv+1-t|\}\\ &\quad\le \frac{\|f''\|_\infty}{c_n}\frac{13}{24}\end{aligned}$$

since

$$\int_0^1 udu \int_{-1}^1 dv\{(1-t)|uv-t| + t|uv+1-t|\} = \frac{1+2t-4t^3+2t^4}{3}$$

is maximized at $\frac{1}{2}$.

If (3.2) holds, we can perform a two term Taylor expansion and write

$$\begin{aligned}
&\left|\mathcal{E}\hat{f}'_n\left(\frac{i+t}{c_n}\right) - f'\left(\frac{i+t}{c_n}\right)\right| \\
&\quad \le (1-t)\int_0^1 udu \int_{-1}^1 dv \left|f'\left(\frac{i+uv}{c_n}\right) - f'\left(\frac{i+t}{c_n}\right)\right| \\
&\quad\quad + t\int_0^1 udu \int_{-1}^1 dv \left|f'\left(\frac{i+1+uv}{c_n}\right) - f'\left(\frac{i+t}{c_n}\right)\right| \\
&\quad \le \frac{\|f'''\|_\infty}{2c_n^2}\int_0^1 udu \int_{-1}^1 dv\{(1-t)(uv-t)^2 + t(uv+1-t)^2\} \\
&\quad \le \frac{\|f'''\|_\infty}{c_n^2}\frac{5}{24}.
\end{aligned}$$

LEMMA 3.2. *Let*

$$\gamma_n = \begin{cases} \dfrac{\epsilon}{4} - \dfrac{13}{24}\dfrac{\|f''\|_\infty}{c_n} & \textit{if} \quad (3.1) \textit{ holds}, \\ \dfrac{\epsilon}{4} - \dfrac{5}{24}\dfrac{\|f'''\|_\infty}{c_n^2} & \textit{if} \quad (3.2) \textit{ holds}, \end{cases} \tag{3.3}$$

with ϵ and c_n chosen so that γ_n is positive. Then

$$\mathcal{P}\left\{\|\hat{f}'_n - f'\|_\infty \le \frac{\epsilon}{4}\right\} \ge 1 - 4\sqrt{2}\exp\left\{-2n\left(\frac{\gamma_n}{4c_n^2}\right)^2\right\}.$$

PROOF. First observe that

$$\begin{aligned}
&\left|\hat{f}'_n\left(\frac{i+t}{c_n}\right) - \mathcal{E}\hat{f}'_n\left(\frac{i+t}{c_n}\right)\right| \\
&\quad\quad |(1-t)(d_{n,i}(\hat{F}_n) - d_{n,i}(F)) + t(d_{n,i+1}(\hat{F}_n) - d_{n,i+1}(F))| \\
&\quad \le \sup_{0\le i\le c_n} |d_{n,i}(\hat{F}_n) - d_{n,i}(F)| \\
&\quad \le 2c_n^2 \sup_{0\le i\le c_n} |p_{n,i}(\hat{F}_n) - p_{n,i}(F)| \\
&\quad \le 4c_n^2\|\hat{F}_n - F\|_\infty
\end{aligned}$$

so that, using Lemma 3.1 we have

$$\begin{aligned}
\mathcal{P}\left\{\|\hat{f}'_n - f'\|_\infty \le \frac{\epsilon}{4}\right\} &\ge \mathcal{P}\{\|\hat{f}'_n - \mathcal{E}\hat{f}'_n\|_\infty \le \gamma_n\} \\
&\ge \mathcal{P}\left\{\|\hat{F}_n - F\|_\infty \le \frac{\gamma_n}{4c_n^2}\right\}.
\end{aligned}$$

The result follows from the DKW inequality.

Thus, as a corollary of Theorems 2.4 and 2.5, we obtain

THEOREM 3.3. *Assume that* $\rho^*(f) < \infty$ *and define* γ_n *as in Equation* (3.3). *Then*

$$\mathcal{P}\left\{\rho\left(\mathcal{S}\left(f, \frac{3\epsilon}{2}\right)\right) \leq \rho(\mathcal{S}(\hat{f}_n, \epsilon)) \leq \rho^*(f)\right\} \geq 1 - 4\sqrt{2}\exp\left\{-2n\left(\frac{\gamma_n}{4c_n^2}\right)^2\right\},$$

$$\mathcal{P}\left\{\rho(\mathcal{S}(f,\epsilon)) \in \left[\rho\left(\mathcal{S}\left(\hat{f}_n, \frac{3\epsilon}{2}\right)\right), \rho\left(\mathcal{S}\left(\hat{f}_n, \frac{\epsilon}{2}\right)\right)\right]\right\}$$
$$\geq 1 - 4\sqrt{2}\exp\left\{-2n\left(\frac{\gamma_n}{4c_n^2}\right)^2\right\}$$

and for all ϵ *sufficiently small*

$$\mathcal{P}\{\rho(\mathcal{S}(\hat{f}_n, \epsilon)) = \rho^*(f)\} \geq 1 - 4\sqrt{2}\exp\left\{-2n\left(\frac{\gamma_n}{4c_n^2}\right)^2\right\}.$$

Notice that the lower bound on the confidence of all these statements is maximized if c_n is set according to ϵ:

$$\frac{1}{c_n(\epsilon)} = \begin{cases} \dfrac{4}{13}\dfrac{\epsilon}{\|f''\|_\infty} & \text{if } (3.1) \text{ holds,} \\ \sqrt{\dfrac{3}{5}\dfrac{\epsilon}{\|f'''\|_\infty}} & \text{if } (3.2) \text{ holds.} \end{cases} \tag{3.4}$$

This concludes our analysis of histogram based estimates. Now we consider kernel density estimates. Our kernel K is a symmetric density, has a integrable second derivative and $\int u^2 K(u)du < \infty$. The kernel estimate of f' is given by

$$\hat{f}_n'(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_n^2}K'\left(\frac{x - X_i}{h_n}\right).$$

LEMMA 3.4 *If* (3.1) *holds,*

$$\|\mathcal{E}\hat{f}_n' - f'\|_\infty \leq h_n\|f''\|_\infty \int |u|K(u)du$$

and if (3.2) *holds,*

$$\|\mathcal{E}\hat{f}_n' - f'\|_\infty \leq \frac{h_n^2}{2}\|f'''\|_\infty \int u^2 K(u)du.$$

PROOF. Using a Taylor expansion we write either

$$|\mathcal{E}\hat{f}_n'(x) - f'(x)| = \left|\int K(u)\{f'(x-h_nu) - f'(x)\}du\right|$$
$$\leq h_n\|f''\|_\infty \int |u|K(u)du$$

if (3.1) holds, or

$$|\mathcal{E}\hat{f}_n'(x) - f'(x)| = \left|\int K(u)\{f'(x-h_nu) - f'(x)du\right|$$
$$\leq \frac{h_n^2}{2}\|f'''\|_\infty \int u^2K(u)du$$

if (3.2) holds.

LEMMA 3.5. *Let*

$$\gamma_n = \begin{cases} \frac{\epsilon}{4} - h_n\|f''\|_\infty \int |u|K(u)du & \text{if} \quad (3.1) \text{ holds}, \\ \frac{\epsilon}{4} - h_n^2\|f'''\|_\infty \int u^2K(u)du & \text{if} \quad (3.2) \text{ holds}, \end{cases} \tag{3.5}$$

with ϵ *and* h_n *chosen so that* γ_n *is positive. Then*

$$\mathcal{P}\left\{\|\hat{f}_n' - f'\|_\infty \leq \frac{\epsilon}{4}\right\} \geq 1 - 4\sqrt{2}\exp\left\{-2n\left(\frac{h_n^2\gamma_n}{\|K''\|_1}\right)^2\right\}.$$

PROOF. First observe that

$$|\hat{f}_n'(x) - \mathcal{E}\hat{f}_n'(x)| = \left|\frac{1}{h_n^2}\int K'\left(\frac{x-t}{h_n}\right)d(\hat{F}_n - F)(t)\right|$$
$$= \left|\frac{1}{h_n^2}\int (\hat{F}_n - F)(t)dK'\left(\frac{x-t}{h_n}\right)\right|$$
$$\leq \frac{\|\hat{F}_n - F\|_\infty}{h_n^2}\int \frac{1}{h_n}\left|K''\left(\frac{x-t}{h_n}\right)\right|dt$$
$$\leq \frac{\|\hat{F}_n - F\|_\infty}{h_n^2}\|K''\|_1$$

so that, using Lemma 3.4 we have

$$\mathcal{P}\left\{\|\hat{f}_n' - f'\|_\infty \leq \frac{\epsilon}{4}\right\} \geq \mathcal{P}\{\|\hat{f}_n' - \mathcal{E}\hat{f}_n'\|_\infty \leq \gamma_n\}$$
$$\geq \mathcal{P}\left\{\|\hat{F}_n - F\|_\infty \leq \frac{h_n^2\gamma_n}{\|K''\|_1}\right\}.$$

The result follows from the DKW inequality.

Singh (1979) used a similar argument to prove that the uniform continuity and the boundedness of f' are a necessary and sufficient condition for the strong uniform consistency of the kernel density estimate. As a corollary of Theorems 2.4 and 2.5, we obtain

THEOREM 3.6. *Assume that $\rho^*(f) < \infty$ and define γ_n as in equation* (3.5). *Then*

$$\mathcal{P}\left\{\rho\left(\mathcal{S}\left(f, \frac{3\epsilon}{2}\right)\right) \leq \rho(\mathcal{S}(\hat{f}_n, \epsilon)) \leq \rho^*(f)\right\} \geq 1 - 4\sqrt{2}\exp\left\{-2n\left(\frac{h_n^2\gamma_n}{\|K''\|_1}\right)^2\right\},$$

$$\mathcal{P}\left\{\rho(\mathcal{S}(f,\epsilon)) \in \left[\rho\left(\mathcal{S}\left(\hat{f}_n, \frac{3\epsilon}{2}\right)\right), \rho\left(\mathcal{S}\left(\hat{f}_n, \frac{\epsilon}{2}\right)\right)\right]\right\}$$
$$\geq 1 - 4\sqrt{2}\exp\left\{-2n\left(\frac{h_n^2\gamma_n}{\|K''\|_1}\right)^2\right\}$$

and for all ϵ sufficiently small

$$\mathcal{P}\{\rho(\mathcal{S}(\hat{f}_n, \epsilon)) = \rho^*(f)\} \geq 1 - 4\sqrt{2}\exp\left\{-2n\left(\frac{h_n^2\gamma_n}{\|K''\|_1}\right)^2\right\}.$$

Notice that the lower bound on the confidence of all these statements is maximized if h_n is set according to ϵ:

$$h_n(\epsilon) = \begin{cases} \dfrac{1}{6}\dfrac{\epsilon}{\|f''\|_\infty \int |u|K(u)du} & \text{if } (3.1) \text{ holds,} \\ \sqrt{\dfrac{1}{8}\dfrac{\epsilon}{\|f'''\|_\infty \int u^2K(u)du}} & \text{if } (3.2) \text{ holds.} \end{cases} \tag{3.6}$$

4. Remarks and simulations

Remark 0. All our statements can be made in terms of a different notion of "substantial" mode. Donoho (1988), for example, introduces such an alternate notion based on a Kolmogorov distance from the empirical cumulative distribution function. Our notion of a "substantial" mode has a simple interpretation and the computations required to provide estimates or confidence intervals are very easy. Furthermore, its mathematical analysis is elementary. In the context of nonparametric regression, Heckman (1992) defines a "substantial" bump if the regression function increases and decreases on long enough contiguous intervals.

Remark 1. Although $\rho^*(f) \geq 1$ for any density f, it may be the case that $\rho(\mathcal{S}(f,\epsilon)) = 0$. Indeed, $\rho(\mathcal{S}(f,\epsilon))$ is the number of modes that can be bracketed in disjoint intervals $[a, b]$ such that $f'(a) > \epsilon$ and $f'(b) < -\epsilon$, so that

$$\rho(\mathcal{S}(f, \|f'\|_\infty)) = 0$$

always holds. The leftmost and rightmost modes can be handled a little differently by requiring only that $f'(b) < -\epsilon$ for the leftmost and $f'(a) > \epsilon$ for the rightmost. Although this modification can be used in practice to make sure that $\rho(\mathcal{S}(f,\epsilon)) \geq 1$, we did not follow this approach because it complicates the presentation.

Remark 2. The confidence interval

$$\left[\rho\left(\mathcal{S}\left(\hat{f}_n, \frac{3\epsilon}{2}\right)\right), \rho\left(\mathcal{S}\left(\hat{f}_n, \frac{\epsilon}{2}\right)\right)\right]$$

does not behave as an "ordinary" confidence interval. As n grows to infinity, it converges to

$$\left[\rho\left(\mathcal{S}\left(f, \frac{3\epsilon}{2}\right)\right), \rho\left(\mathcal{S}\left(f, \frac{\epsilon}{2}\right)\right)\right]$$

which may or may not collapse. It will reduce to a single point only if

$$\rho\left(\mathcal{S}\left(f, \frac{3\epsilon}{2}\right)\right) = \rho\left(\mathcal{S}\left(f, \frac{\epsilon}{2}\right)\right).$$

For example (Fig. 0), with $\epsilon = 0.05$ and

$$f_0(x) = \frac{2}{10}\eta(x+5) + \frac{5}{10}\eta(x) + \frac{3}{10}\eta(x-5)$$

where η is the standard normal density, we have

$$\rho\left(\mathcal{S}\left(f, \frac{3\epsilon}{2}\right)\right) = 1,$$

$$\rho(\mathcal{S}(f,\epsilon)) = 2,$$

and

$$\rho\left(\mathcal{S}\left(f, \frac{\epsilon}{2}\right)\right) = 3.$$

A confidence interval with "ordinary" behaviour can be given at the expense of the introduction of an extra parameter. Indeed, the arguments of Lemma 2.1 and Theorem 2.5 can easily be modified into

THEOREM 4.1. *If* $\rho^*(f) < \infty$ *and* $\|f' - \hat{f}_n'\|_\infty < \infty$

$$\mathcal{P}\{\rho(\mathcal{S}(f,\epsilon)) \in [\rho(\mathcal{S}(\hat{f}_n, \epsilon+\gamma)), \rho(\mathcal{S}(\hat{f}_n, \epsilon-\gamma))]\} \geq \mathcal{P}\left\{\|f' - \hat{f}_n'\|_\infty < \frac{\gamma}{2}\right\}.$$

Theorem 2.5 is the particular case where $\gamma = \epsilon/2$. If γ_n converges to zero slowly enough that

$$\mathcal{P}\left\{\|f' - \hat{f}_n'\|_\infty < \frac{\gamma_n}{2}\right\} \to 1,$$

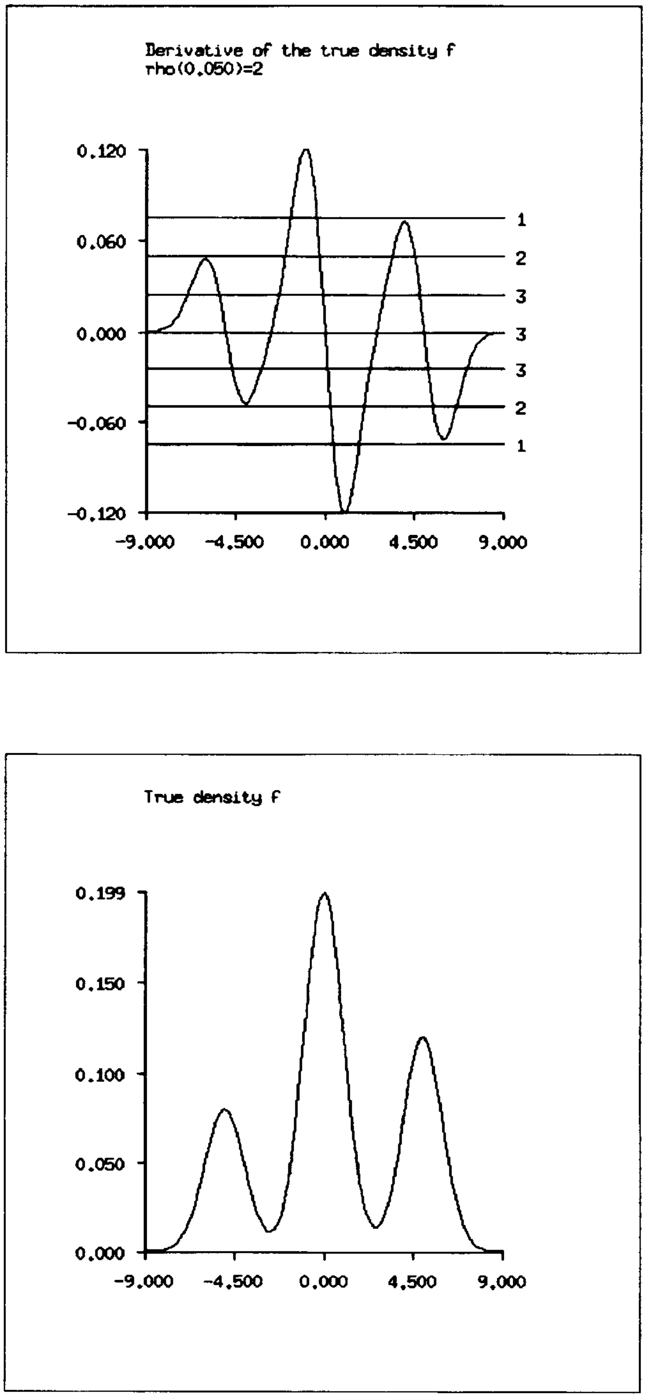

Derivative of the true density f
rho(0.050)=2
0.120
0.060
0.000
-0.060
-0.120
-9.000
-4.500
0.000
4.500
9.000
1
2
3
3
3
2
1
True density f
0.199
0.150
0.100
0.050
0.000
-9.000
-4.500
0.000
4.500
9.000

Fig. 0.

the confidence interval

$$[\rho(\mathcal{S}(\hat{f}_n, \epsilon + \gamma_n)), \rho(\mathcal{S}(\hat{f}_n, \epsilon - \gamma_n))]$$

will collapse to the single point $\rho(\mathcal{S}(f, \epsilon))$. We chose not to follow this line because it requires a better understanding of the behaviour of

$$\mathcal{P}\{\|f' - \hat{f}'_n\|_\infty < \alpha\}.$$

Remark 3. Notice that Theorems 2.4, 2.5 and 4.1 depend in no way on the stochastic structure of the observations. All they do is provide lower bounds on the confidence of various statements in terms of

$$\mathcal{P}\{\|f' - \hat{f}'_n\|_\infty < \alpha\}.$$

Furthermore, in Section 3, in order to reduce this bound to one depending on

$$\mathcal{P}\{\|\hat{F}_n - F\|_\infty \leq \beta\},$$

the only assumption made is that the observations are identically distributed. To this point, nothing is said about the dependence structure of the observations, so that our results can be extended without too much effort to any case for which

$$\mathcal{P}\{\|\hat{F}_n - F\|_\infty \leq \beta\} \to 1.$$

Remark 4. The Lemmas of Section 3 provide a bound on

$$\mathcal{P}\left\{\|f' - \hat{f}'_n\|_\infty < \frac{\epsilon}{4}\right\}$$

that is required to apply Theorems 2.4 and 2.5. The bounds we use show that the probability of coverage converges to 1 exponentially fast as n grows to infinity. Unfortunately, the approach we chose is not tight enough and the conservative level we get from our bounds is close to zero, even for large sample sizes. For example, assume that (3.2) holds and that the kernel is the standard normal density. Substituting the bandwidth that maximizes our lower bound, this lower bound becomes

$$1 - 4\sqrt{2}\exp\left\{-2.14n\left(\frac{\epsilon}{8}\right)^4 \frac{1}{\|f'''\|_\infty^2}\right\}$$

and in order to get a bound of $100(1-\alpha)\%$, n must be quite large:

$$n \geq 1914\frac{\|f'''\|_\infty^2}{\epsilon^4}\log\left(\frac{4\sqrt{2}}{\alpha}\right).$$

Table 1 shows the right hand side of the last inequality for different values of ϵ and levels with $\|f'''\|_\infty = 1$. In practice, however, we had very good results with 100 observations and we believe that a direct evaluation of

$$\mathcal{P}\left\{\rho(\mathcal{S}(f,\epsilon)) \in \left[\rho\left(\mathcal{S}\left(\hat{f}_n, \frac{3\epsilon}{2}\right)\right), \rho\left(\mathcal{S}\left(\hat{f}_n, \frac{\epsilon}{2}\right)\right)\right]\right\}$$

Table 1.

	50%	80%	90%
$\epsilon = 1.5$	918	1,264	7,723
$\epsilon = 1.0$	4,644	6,398	7,723
$\epsilon = 0.5$	74,294	102,355	123,581

or a better bound than the one we have used, would provide useful estimates of the confidence level of our intervals. Research in that direction is under way.

Remark 5. Even if our bounds have little practical value, it should be noted that equations (3.4) and (3.6) suggest that a fixed amount of smoothing be made, irrespective of the sample size. This is not really surprising since for that bandwidth and all smaller ones, f and $\mathcal{E}\hat{f}_n'(x)$ have the same number of "substantial" modes (see the next section). In order to select the bandwidth by using estimates of $\|f''\|_\infty$ and $\|f'''\|_\infty$ together with equations (3.4) and (3.6), the smoothness of the unknown density would have to be known. In our experience, the bandwidths selected by ordinary least squares cross-validation are appropriate for use with the kernel based estimate of f'. In fact, our estimates do not seem to be too sensitive to the bandwidth; this results from eliminating the spurious 0-downcrossings of $\hat{f}_n'$ by using a positive ϵ. Various automatic bandwidth selection procedures lead to more or less the same intervals.

Remark 6. The choice of ϵ is left to the analyst. A very large value for ϵ will restrict the attention to very "substantial" modes. Naturally, one expects that for small values of ϵ more observations will be required to identify the modes and this is reflected in the exponential bounds we have provided. One difficulty with the choice of ϵ is that the notion of "substantial" depends on the scale of the underlying density. Indeed, the modes counted in $\rho(\mathcal{S}(f_1, \epsilon))$ correspond to those counted in $\rho(\mathcal{S}(f_a, a^2\epsilon))$ where $f_a(x) = af(ax)$. For that reason, we recommend using values of ϵ around 1 once the data has been rescaled to the interval $[0, 1]$. The analyst can get a better idea of what $\epsilon = 1$ corresponds to in terms of the size of the modes by experimenting with known densities on the interval $[0, 1]$.

Remark 7. The next three figures are typical of simulation results. The first two graphs display the underlying true density and its derivative. The bottom two graphs display kernel based estimates of f and f' respectively, using the bandwidth selected by ordinary least squares cross-validation for a sample of size $n = 100$. The horizontal lines are

$$y = -\epsilon = -0.05, \qquad y = 0, \qquad y = \epsilon = 0.05$$

used to count the number of downcrossings. The true densities are

$$f_1(x) = \frac{1}{4}\frac{1}{.7}\eta\left(\frac{x+5}{.7}\right) + \frac{1}{2}\eta(x) + \frac{1}{4}\frac{1}{.7}\eta\left(\frac{x-5}{.7}\right)$$

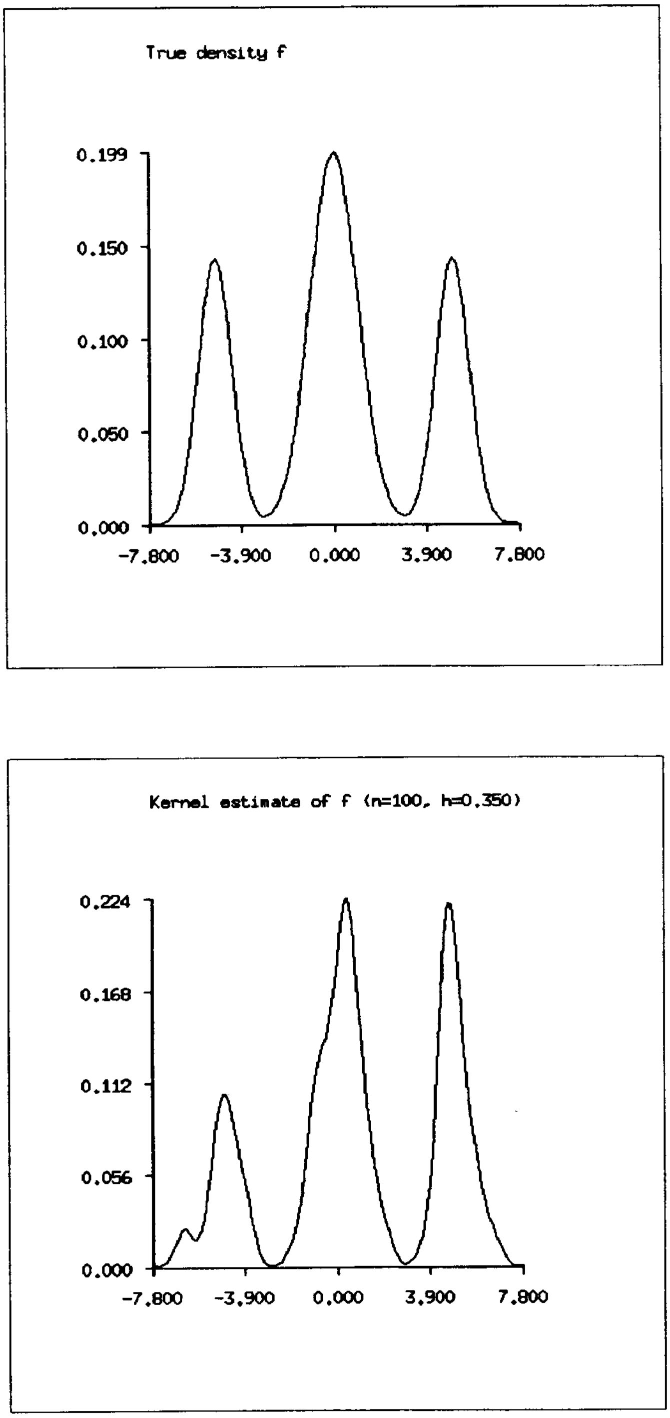
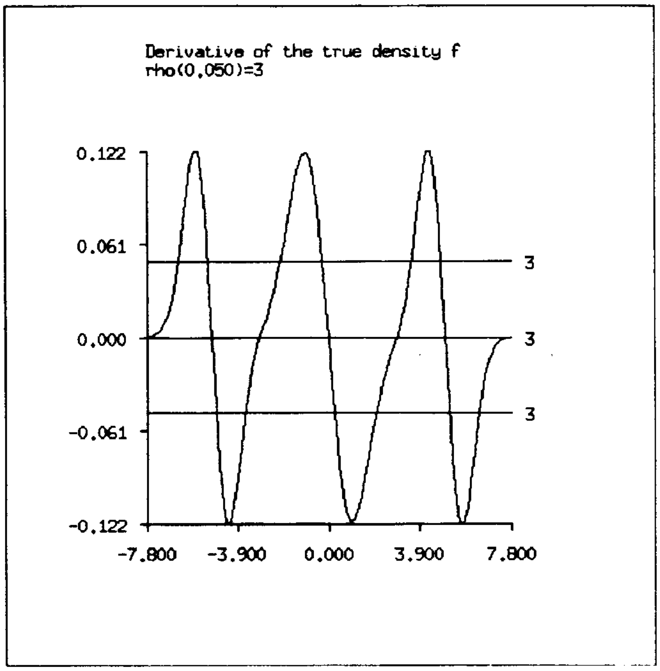

True density f
0.199
0.150
0.100
0.050
0.000
-7.800
-3.900
0.000
3.900
7.800
Kernel estimate of f (n=100, h=0.350)
0.224
0.168
0.112
0.056
0.000
-7.800
-3.900
0.000
3.900
7.800

Fig. 1.



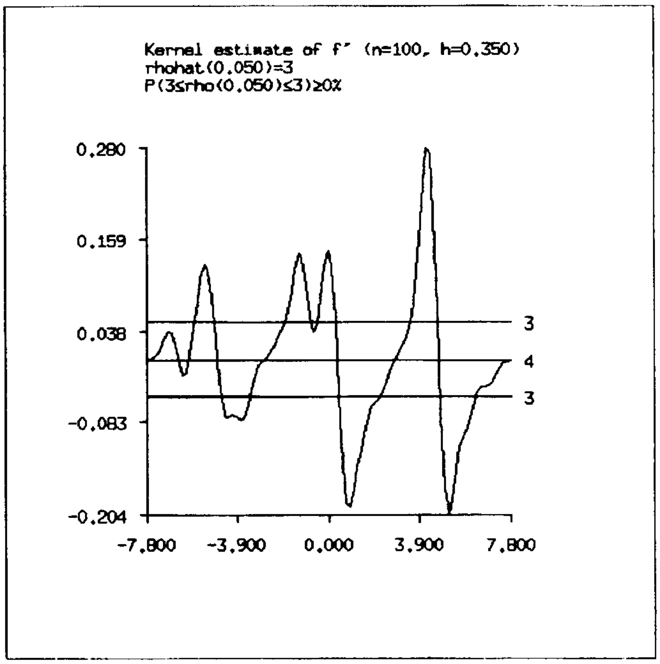
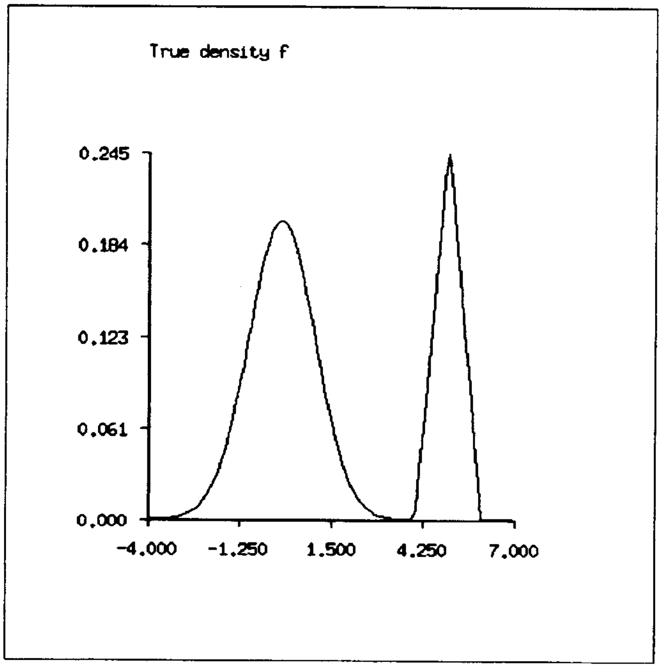


Fig. 1. (continued).


True density f
0.245
0.184
0.123
0.061
0.000
-4.000
-1.250
1.500
4.250
7.000

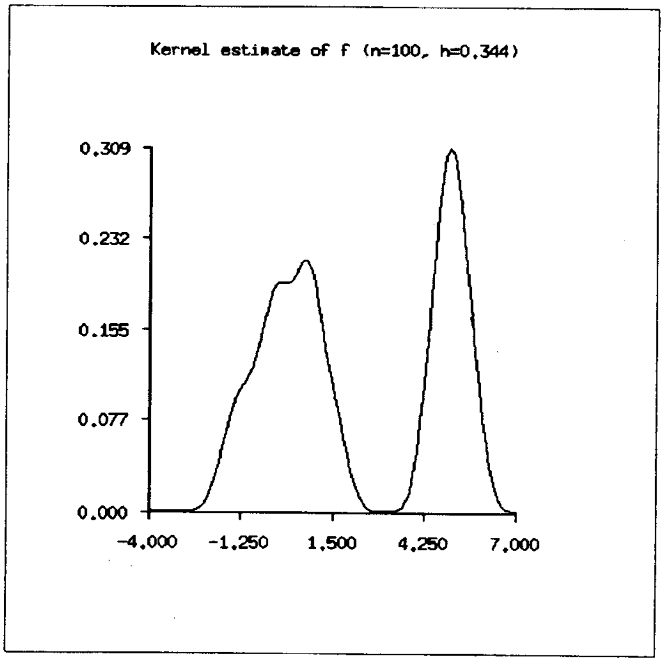
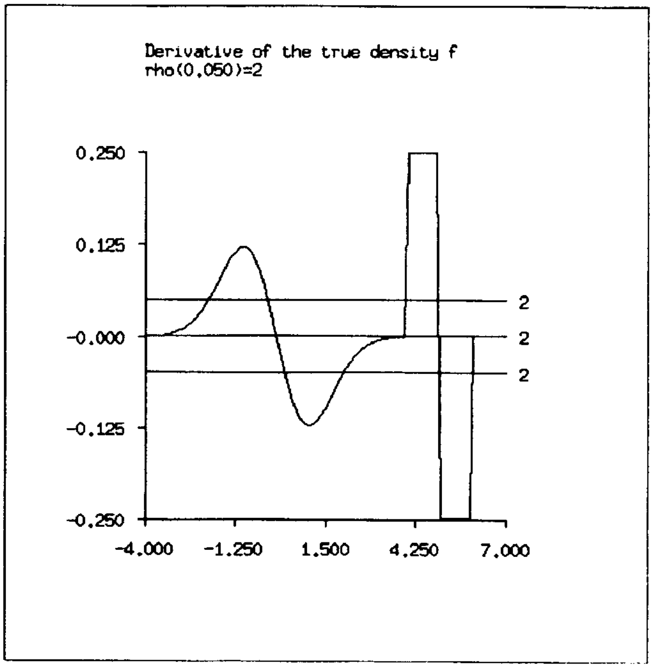

Kernel estimate of f (n=100, h=0.344)
0.309
0.232
0.155
0.077
0.000
-4.000
-1.250
1.500
4.250
7.000

Fig. 2.



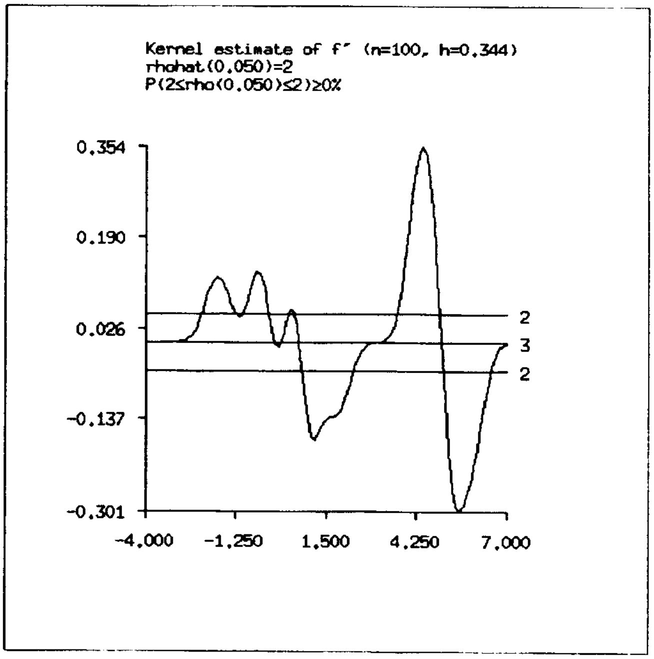
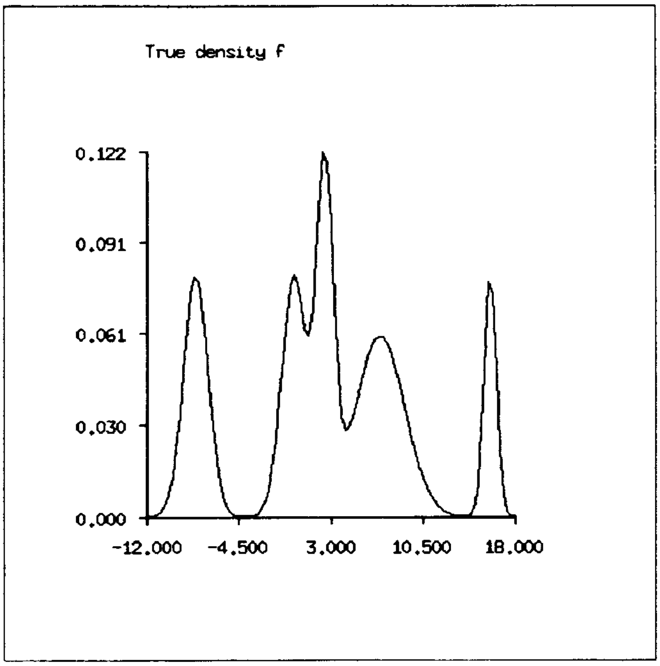


Fig. 2. (continued).


True density f
0.122
0.091
0.061
0.030
0.000
-12.000
-4.500
3.000
10.500
18.000

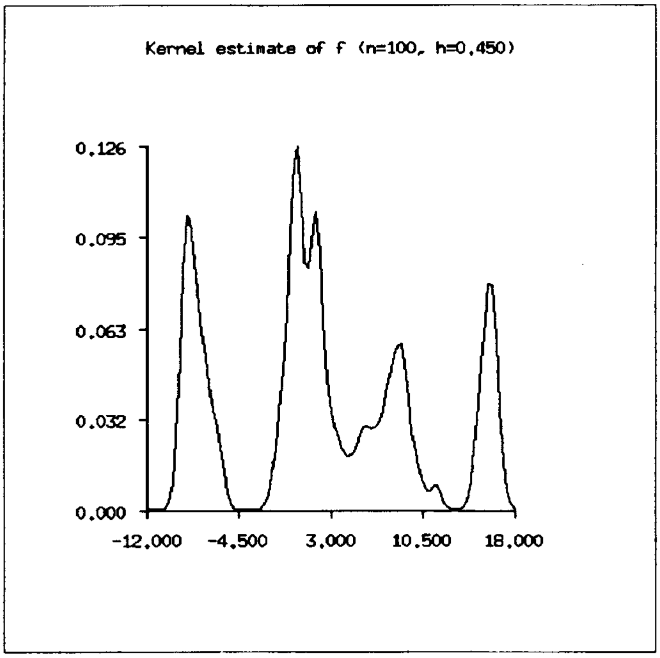
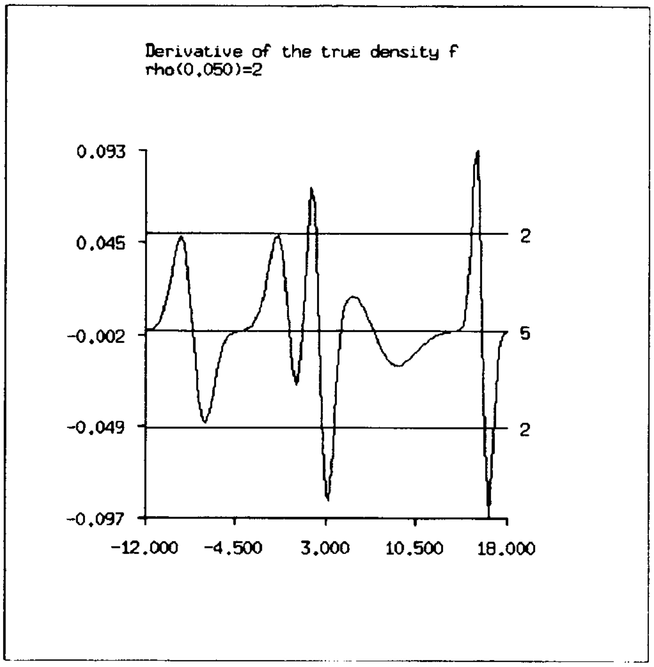

Kernel estimate of f (n=100, h=0.450)
0.126
0.095
0.063
0.032
0.000
-12.000
-4.500
3.000
10.500
18.000

Fig. 3.



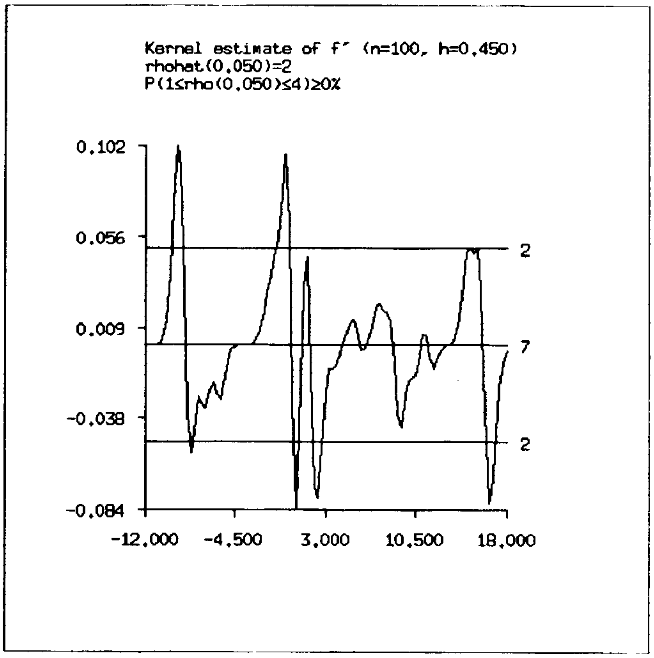


Fig. 3. (continued).

for Fig. 1,

$$f_2(x) = \frac{1}{2}\eta(x) + \frac{1}{2}(1 - |x-5|)1\{|x-5| < 1\}$$

for Fig. 2 and

$$f_3(x) = \frac{1}{5}\eta(x) + \frac{1}{5}\frac{1}{.7}\eta\left(\frac{x-2.5}{.7}\right) + \frac{1}{10}\frac{1}{.5}\eta\left(\frac{x-16}{.5}\right) + \frac{3}{10}\frac{1}{2}\eta\left(\frac{x-7}{2}\right) + \frac{1}{5}\eta(x+8)$$

for Fig. 3, where η is the standard normal density. Although the graphs of the true densities show strongly delineated modes, the large difference between the horizontal and vertical scales of all the figures should be taken into account.

Remark 8. To better assess the performance of our confidence intervals, we performed three simulations where the true underlying density of the observations was f_1, f_2 and f_3 respectively. In both cases, the bandwidth was selected by ordinary least squares cross-validation and we used 100 samples of size $n = 50, 100, 200$. The observed coverage probabilities are reported in Tables 2 and 3.

Table 2.

$\epsilon = 0.01$

	f_1	f_2	f_3
$n = 50$	0.83	0.35	0.24
$n = 100$	0.84	0.31	0.41
$n = 200$	0.85	0.28	0.54

Table 3.

$\epsilon = 0.05$

	f_1	f_2	f_3
$n = 50$	0.91	0.67	0.57
$n = 100$	0.94	0.67	0.80
$n = 200$	0.95	0.65	0.92

It is not surprising to see the observed coverage probabilities increase with ϵ: identifying large bumps is an easier task than identifying small ones. The observed coverage probabilities increase more or less with n as expected. Comparing the result for f_1 and f_3 shows that a large number of modes makes the problem more difficult. Inspection of the results has shown that the poor results with f_2 are due to the poor performance of cross-validation in selecting a proper bandwidth for not so smooth densities. In many cases, the selected bandwidth was from 2 to 5 times smaller than the actual optimal bandwidth.

In other words, in order to count the number of bumps, there is no reason to select a bandwidth smaller than

$$h_c = \frac{\epsilon_0(f)}{4\|f''\|_\infty \int |u|K(u)du}.$$

Although this critical value is of no practical value as it depends on the unknown quantities $\epsilon_0(f)$ and $\|f''\|_\infty$ it follows that bias is not the main problem. This is a nice feature of the problem of bump counting and, more generally, of the problem of identifying the shape: the bandwidth does not need to converge to zero, it suffices that it be small enough.

To take one more step in that direction, consider the quantity $\rho(\mathcal{S}(\mathcal{E}\hat{f}_h, \epsilon))$ to be the parameter of interest. Theorem 2.5 can be applied to the functions $\hat{f}_n$ and $\mathcal{E}\hat{f}_h$ and by using the argument we used in Lemma 3.5, we get

THEOREM 5.2. *If* $\rho^*(f) < \infty$,

$$\begin{aligned} &\mathcal{P}\left\{\rho(\mathcal{S}(\mathcal{E}\hat{f}_h, \epsilon)) \in \left[\rho\left(\mathcal{S}\left(\hat{f}_n, \frac{3\epsilon}{2}\right)\right), \rho\left(\mathcal{S}\left(\hat{f}_n, \frac{\epsilon}{2}\right)\right)\right]\right\} \\ &\geq \mathcal{P}\left\{\|\mathcal{E}\hat{f}'_h - \hat{f}'_n\|_\infty < \frac{\epsilon}{4}\right\} \\ &\geq 1 - 4\sqrt{2}\exp\left\{-2n\left(\frac{h^2\epsilon}{4\|K''\|_1}\right)^2\right\}. \end{aligned}$$

This inequality provides a better explanation of the results of our simulations than the one provided by the previous inequalities. Since in practice, one always uses a specific value for h—though a random one—it can be argued that Theorem 5.2 is of real practical value. In addition, the bound is good enough to be practically useful. For a fixed bandwidth h, the interval

$$\left[\rho\left(\mathcal{S}\left(\hat{f}_n, \frac{3\epsilon}{2}\right)\right), \rho\left(\mathcal{S}\left(\hat{f}_n, \frac{\epsilon}{2}\right)\right)\right]$$

is an interval for $\rho(\mathcal{S}(\mathcal{E}\hat{f}_h, \epsilon))$ more that it is one for $\rho(\mathcal{S}(f, \epsilon))$ or for $\rho^*(f)$. Using it as an interval for the latter parameters is justified only in view of the comments we have made in this section.

Acknowledgements

We would like to thank a referee for valuable comments and interesting suggestions which improved the presentation of the paper.

REFERENCES

Cox, D. R. (1966). Notes on the analysis of mixed frequency distributions, *British J. Math. Statist. Psych.*, **19**, 39–47.

5. Bandwidth selection

Bandwidth selection is one of the most difficult and important questions in density estimation. The bandwidth must constitute a compromise to the opposing goals of making the bias small, which requires a small bandwidth, and making the variation small, which requires a large one. Although many automatic procedures are discussed in the literature, there does not appear to be a general consensus on how one should proceed. In fact, many authors have already pointed out that the bandwidth should be selected according to the role the density estimate is meant to fulfil.

Density estimation is often used as a tool of exploration. The shape of the density underlying the observations and its features, such as modes, are what the data analyst often seeks. With that in mind, we argue that it is not necessary to let the bandwidth converge to zero because, for a small enough bandwidth h, the functions $\mathcal{E}\hat{f}_h$ and f display the same characteristics. In particular, they have the same number of modes. We write $\mathcal{E}\hat{f}_h$ instead of $\mathcal{E}\hat{f}_n$ to emphasize that this expectation depends on h and not on n. For the normal kernel,

$$\rho(\mathcal{S}(\mathcal{E}\hat{f}_h, \epsilon))$$

is a non increasing function of h that converges to $\rho(\mathcal{S}(f, \epsilon))$ as h converges to zero. This is shown in Silverman (1981). In fact, for any kernel, assume that $\rho^*(f) < \infty$ and that equation (3.1) holds so that

$$\|\mathcal{E}\hat{f}_h' - f'\|_\infty \leq h_n \|f''\|_\infty \int |u| K(u) du.$$

According to Lemmas 2.1, 2.2 and 2.3, if $\frac{3}{2}\epsilon < \epsilon_0(f)$,

$$\|\mathcal{E}\hat{f}_h' - f'\| < \frac{\epsilon}{4} \Rightarrow \rho^*(f) = \rho\left(\mathcal{S}\left(f, \frac{3\epsilon}{2}\right)\right) \leq \rho(\mathcal{S}(\mathcal{E}\hat{f}_h, \epsilon)) \leq \rho\left(\mathcal{S}\left(f, \frac{\epsilon}{2}\right)\right) \leq \rho^*(f)$$

so that for $\frac{3}{2}\epsilon < \epsilon_0(f)$ we have

$$\|\mathcal{E}\hat{f}_h' - f'\| < \frac{\epsilon}{4} \Rightarrow \rho^*(f) = \rho(\mathcal{S}(\mathcal{E}\hat{f}_h, \epsilon)).$$

Therefore,

THEOREM 5.1. *Assume that $\rho^*(f) < \infty$ and that $\frac{3}{2}\epsilon < \epsilon_0(f)$. If*

$$h_n \leq \frac{\epsilon}{4\|f''\|_\infty \int |u| K(u) du}$$

then

$$\rho^*(f) = \rho(\mathcal{S}(\mathcal{E}\hat{f}_h, \epsilon)).$$

Cuevas, A. and Gonzalez Manteiga, W. (1991). Data-driven smoothing based on convexity properties, *Nonparametric Functional Estimation and Related Topics*, *NATO Adv. Sci. Inst. Ser. C. Math. Phys. Sci.* (ed. G. Roussas), **335**, 225–240, Kluwer, The Netherlands.

Donoho, D. L. (1988). One-sided inference about functionals of a density, *Ann. Statist.*, **16**, 1390–1420.

Good, I. J. and Gaskins, R. A. (1980). Density estimation and bump-hunting by the penalized likelihood method exemplified by scattering and meteorite data, *J. Amer. Statist. Assoc.*, **75**, 42–56.

Hartigan, J. A. and Hartigan, P. M. (1985). The dip test of unimodality, *Ann. Statist.*, **13**, 70–84.

Heckman, N. E. (1992). Bump hunting in regression analysis, *Statist. Probab. Lett.*, **14**, 141–152.

LeCam, L. and Schwartz, L. (1960). A necessary and sufficient condition for the existence of consistent estimates, *Ann. Math. Statist.*, **31**, 140–150.

Lehmann, E. L. (1959). *Testing Statistical Hypotheses*, Wiley, New York.

Mammen, E., Marron, J. S. and Fisher, N. I. (1992). Some asymptotics for multimodality test based on kernel density estimates, *Probab. Theory Related Fields*, **91**, 115–132.

Shorack, G. R. and Wellner, J. A. (1986). *Empirical Processes with Application to Statistics*, Wiley, New York.

Silverman, B. W. (1980). Comments on Good and Gaskins, *J. Amer. Statist. Assoc.*, **75**, 67–68.

Silverman, B. W. (1981). Using kernel density estimates to investigate multimodality, *J. Roy. Statist. Soc. Ser. B*, **43**, 97–99.

Silverman, B. W. (1983). Some properties of a test for multimodality based on kernel density estimates, *Probability, Statistics and Analysis* (eds. J. F. C. Kingman and G. E. H. Reuter), 248–259, Cambridge University Press, Cambridge.

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

Singh, R. S. (1979). On necessary and sufficient conditions for uniform strong consistency of estimators of a density and its derivatives, *J. Multivariate Anal.*, **9**, 157–164.

Wong, M. A. (1982). A hybrid clustering method for identifying high-density clusters, *J. Amer. Statist. Assoc.*, **77**, 841–847.

Wong, M. A. and Lane, T. (1983). A k-th nearest neighbour clustering procedure, *J. Roy. Statist. Soc. Ser. B*, **45**, 362–368.