Statistica Sinica **17**(2007), 289-300

ON THE EXISTENCE AND LIMIT BEHAVIOR OF THE OPTIMAL BANDWIDTH FOR KERNEL DENSITY ESTIMATION

J. E. Chacón, J. Montanero, A. G. Nogales and P. Pérez

Universidad de Extremadura

Abstract: We prove, under mild conditions, the existence of a minimizer of the exact mean integrated square error of a kernel density estimator as a function of the bandwidth. When it exists, we also show some expected limit properties of this optimal bandwidth; in fact, for two common situations, Theorem 3 gives the exact value for the limit. Surprisingly, in some special cases (when using superkernels or the sinc kernel for estimating some classes of densities), this limit is strictly positive and a global zero-bias bandwidth can be chosen.

Key words and phrases: Kernel density estimator, mean integrated squared error, optimal bandwidth, superkernel, sinc kernel.

1. Introduction

Density estimation is one of the most studied topics in nonparametric statistics during the last decades. The problem is to estimate a density function f, given $X_1, \ldots, X_n$, a set of independent $\mathbb{R}$-valued random variables, each having density f. The kernel estimator of f is given by

$$f_{n,h}(x) = \frac{1}{n} \sum_{i=1}^{n} K_h(x - X_i),$$

where the bandwidth h is a positive real number, the kernel K is an integrable function with $\int K(x)dx = 1$, and $K_h(x) = K(x/h)/h$ (see, e.g., Silverman (1986) for an extensive description of the kernel method). The use of the kernel density estimator rests upon the choice of the kernel K and the bandwidth h.

It is usual, for density estimation, to use positive kernels (that is, densities), so that the final estimate we get is itself a density function. In this sense, the differences between several positive kernels are not severe (see Table 6.2 in Scott (1992)), with the Gaussian kernel being the most commonly used. Nevertheless, there is a possible gain in terms of asymptotic rates of convergence if we allow the kernel to take negative values (see Wand and Jones (1995, Section 2.8)).

Unlike the situation with the kernel, bandwidth choice is a very important feature to take into account. We can measure the error of the kernel estimator $f_{n,h}$ through the mean integrated squared error (MISE), which is defined by the expression

$$M_{f,n}(h) = \mathrm{MISE}_f(f_{n,h}) = E_f \int \left[f_{n,h}(x) - f(x)\right]^2 dx, \tag{1}$$

where, here and thereafter, the symbol $\int$ will denote integration over the whole real line, unless otherwise stated. It is sufficient, for the expression (1) to be finite, to suppose that both f and K are square integrable; that is, $f,\, K \in \mathcal{L}_2$. According to (1), a quite natural choice for h is the minimizer of the function $M_{f,n}\colon (0,\infty) \to \mathbb{R}$. Such a minimizer will clearly depend on the sample size n and the density f; we denote it as $h_{0,n}(f)$. The question is: does the MISE always have such a minimizer? In the affirmative case, does $h_{0,n}(f)$ satisfy some typical bandwidth conditions, such as

$$h_{0,n}(f) \to 0 \quad \text{and} \quad nh_{0,n}(f) \to \infty \tag{2}$$

as $n \to \infty$? In general, the answer to these questions is well-known for the asymptotic MISE and the asymptotically optimal bandwidth under enough smoothness assumptions. However, conditions (2) are necessary hypotheses to show that the asymptotic and the exact optimal bandwidth are asymptotically equivalent (see details in Section 3), so our aim is to work with the exact minimizer $h_{0,n}(f)$ and assume conditions as mild and natural as possible.

For this purpose, some preliminary non-asymptotic results are used in Section 2 to provide sufficient conditions for the existence of the optimal bandwidth. Then, under these conditions, the limit behavior of the exact optimal bandwidth sequence is investigated in Section 3, considering also the possibility of superkernels and the sinc kernel, for which the first part of (2) is not verified and a zero-bias bandwidth can be chosen in some special cases. Finally, all the proofs are postponed to Section 4.

2. Existence of the Optimal Bandwidth

Firstly, let us recall an equivalent expression for the MISE. We can write

$$M_{f,n}(h) = B_f(h) + V_{f,n}(h), \tag{3}$$

where

$$B_f(h) = \int \left\{E_f\left[f_{n,h}(x)\right] - f(x)\right\}^2 dx \quad \text{and}$$
$$V_{f,n}(h) = \int \mathrm{Var}_f\left[f_{n,h}(x)\right] dx$$

are called the Integrated Square Bias (ISB) and Integrated Variance (IV), respectively. It is well known (see, e.g., Wand and Jones (1995)) that

$$B_f(h) = R(f) + R(K_h * f) - 2R_{K,h}(f) \quad \text{and} \tag{4}$$

$$V_{f,n}(h) = \frac{R(K)}{nh} - \frac{1}{n}R(K_h * f), \tag{5}$$

where we are using the notation $*$ for convolution, $R(\psi) = \int \psi(x)^2 dx$ for any square integrable function ψ and

$$R_{K,h}(f) := \int (K_h * f)(x) f(x) dx.$$

The following theorem ensures the existence of an optimal bandwidth minimizing the MISE.

Theorem 1. *Let f be a density function in $\mathcal{L}_2$ and K a bounded kernel, continuous at zero and such that*

$$R(K) < 2K(0). \tag{6}$$

Then $M_{f,n}$ has a minimizer on $(0, \infty)$, that is, there exists $h_{0,n}(f) > 0$ such that

$$M_{f,n}(h_{0,n}(f)) \le M_{f,n}(h), \quad \forall h > 0.$$

Condition (6) on the kernel K in the theorem above is quite natural. It is logical, when estimating a density, to use a kernel which attains its maximum at zero because the kernel estimate will assign more weight to the points of the sample. If it does, $R(K) \le \|K\|_\infty < 2\|K\|_\infty = 2K(0)$. Condition (6) has already appeared in the literature; it was a necessary condition to show the optimality of the cross-validation bandwidth selector in Stone (1984).

3. Limit Behavior of the Optimal Bandwidth

Once we have obtained sufficient (and mild) conditions for the exact MISE function to have a minimizer $h_{0,n}(f)$, it is natural to ask if the sequence $\{h_{0,n}(f)\}_n$ of optimal bandwidths satisfies the classical limit conditions $h_{0,n}(f) \to 0$ and $nh_{0,n}(f) \to \infty$ as $n \to \infty$. We will see below that the second one holds quite generally; however, it is a bit surprising that the first one does not necessarily hold.

Theorem 2. *Let f and K be as in Theorem* 1. *Then, the optimal bandwidth sequence satisfies $nh_{0,n}(f) \to \infty$ as $n \to \infty$.*

For the study of the limit behavior of $\{h_{0,n}(f)\}_n$ we need to introduce some extra concepts. If $m_j(K) := \int x^j K(x)dx$, a kernel K is said to be of finite order $k \geq 1$ if

$$m_j(K) = \begin{cases} 0 & \text{for } j = 1, \ldots, k-1, \\ C & \text{for } j = k, \end{cases}$$

with $C \in \mathbb{R} \setminus \{0\}$. For instance, every symmetric density with finite variance is a kernel of order 2. In fact, there is a method for constructing a kernel of arbitrary even order based on a kernel of order 2 (see Schucany and Sommers (1977)), which has been applied to get higher order kernels based on the Gaussian one (Wand and Schucany (1990)). Also, a kernel K is said to be a superkernel if its characteristic function is equal to 1 in a neighborhood of the origin (see Devroye (1992), or Glad, Hjort and Ushakov (2003)). It is possible to find several examples of superkernels, for instance, in Devroye and Lugosi (2001, Chap.17) (see also Example 1 below).

In the following, we use the notation $\varphi_L(t) = \int e^{itx} L(x)dx$, $t \in \mathbb{R}$, for the characteristic function of any function $L \in \mathcal{L}_1$ and, for every density f and every symmetric kernel K (both in $\mathcal{L}_2$), we set

$$\begin{aligned} J_{f,K} &= \sup\{h \geq 0 \colon B_f(h) = 0\}, \\ C_f &= \sup\{r \geq 0 \colon \varphi_f(t) \neq 0 \text{ a.e. for } t \in [0, r]\}, \\ D_f &= \sup\{t \geq 0 \colon \varphi_f(t) \neq 0\}, \\ S_K &= \inf\{t \geq 0 \colon |\varphi_K(t) - 1| \neq 0\}, \\ T_K &= \inf\{r \geq 0 \colon |\varphi_K(t) - 1| \neq 0 \text{ a.e. for } t \geq r\}. \end{aligned}$$

All these exist, with C_f, D_f possibly being infinite, $J_{f,K} \in [0, \infty)$, by part b) of the lemma in Section 4, and S_K, $T_K \in [0, \infty)$, due to the Riemman-Lebesgue Lemma for Fourier transforms (see Kawata (1972, Theorem 2.1.1)). Besides, $C_f \leq D_f$ and $S_K \leq T_K$.

Notice that, by definition, $S_K > 0$ for superkernels. Moreover, it is not difficult to show that $S_K = 0$ for any kernel K of finite order, since in this case we have that $\varphi_K(t)$ has a strict local extreme at $t = 0$. Nevertheless, Devroye and Lugosi (2001, p.194), give an example of a kernel K of infinite order (i.e., such that $m_j(K) = 0$ for all $j = 1, 2, \ldots$) with $S_K = 0$.

The following theorem says that, if we use a kernel K with $S_K = 0$ (e.g., a kernel of finite order), then we can assure that the optimal bandwidth sequence converges to 0 with no conditions on the density. However, if we want this property to hold also for superkernels, we have to demand the characteristic function of the density f be strictly positive at arbitrarily large values. This is not a very restrictive condition.

Moreover, if we use a superkernel for estimating a density such that its characteristic function has a compact support, we have two common cases ($S_K = T_K$ or $C_f = D_f$) where we can easily tell what the limit is.

Theorem 3. *Let f and K be as in Theorem 1. The following assertions hold:*

(a) *If either $S_K = 0$ or $D_f = \infty$, then $h_{0,n}(f) \to 0$ as $n \to \infty$. Besides, in either case we have $J_{f,K} = 0$; that is, $B_f(h) > 0$ for all $h > 0$.*

(b) *If either $S_K = T_K$ or $C_f = D_f$, then $h_{0,n}(f) \to S_K/D_f$ as $n \to \infty$. Besides, in either case we have $J_{f,K} = S_K/D_f$.*

Notice that the case $C_f < D_f$ is possible. It happens, for instance, if the set $\{t \geq 0\colon \varphi_f(t) = 0\}$ can be expressed as the union of two or more intervals, separated from each other by sets of positive measure. The analogous case $S_K < T_K$ occurs if the set $\{t \geq 0\colon |\varphi_K(t) - 1| = 0\}$ can be expressed in the same way as before.

Usually, the MISE of a kernel density estimator is asymptotically described under regularity conditions. Namely, if the kernel K is of order k, the density f satisfies some smoothness conditions (involving up to the kth-derivative) and the bandwidth sequence $\{h_n\}_n$ is such that

$$h_n \to 0 \quad \text{and} \quad nh_n \to \infty \tag{7}$$

as $n \to \infty$, then the MISE of the kernel estimator f_{n,h_n} can be expressed as $M_{f,n}(h_n) = A_{f,n}(h_n) + o(h_n^{2k}) + o(n^{-1}h_n^{-1})$, where

$$A_{f,n}(h) := \frac{h^{2k}}{(k!)^2} m_k(K)^2 R(f^{(k)}) + \frac{R(K)}{nh}$$

is called the asymptotic MISE (see Marron and Wand (1992, p.719)). Thus, if we denote by $h_{*,n}(f)$ the minimizer of the function $A_{f,n}(h)$, it is clear that $h_{*,n}(f)$ exists and can be written as $h_{*,n}(f) = c_0(f,K)n^{-1/(2k+1)}$ for a constant $c_0(f,K)$ depending on f and K only. Therefore, it is obvious that it satisfies $h_{*,n}(f) \to 0$ and $nh_{*,n}(f) \to \infty$ as $n \to \infty$. However, the *asymptotic* minimizer $h_{*,n}(f)$ should not be mistaken for the *exact* minimizer $h_{0,n}(f)$; whereas the exact one makes sense with mild assumptions on the kernel and the density (see Theorem 1) and for all $n \in \mathbb{N}$, the use of the asymptotic minimizer rests upon some smoothness conditions and is only valid for large sample sizes. Sharp differences from the exact one can be found for small sample sizes, in some cases (see Marron and Wand (1992)). Nevertheless it is well-known that, under smoothness conditions, if $h_{0,n}(f)$ satisfies (7) we can write

$$h_{0,n}(f) = h_{*,n}(f) + o(n^{-\frac{1}{2k+1}}), \tag{8}$$

which shows that the sequences $\{h_{0,n}(f)\}_n$ and $\{h_{*,n}(f)\}_n$ are of the same order (see, for instance, Hall and Marron (1991, p.160)). But we want to emphasize that, for arriving at (8), we need the optimal bandwidth sequence to exist and satisfy (7), as in Theorems 2 and 3 with no smoothness assumptions on the density. Moreover, if the smoothness conditions on f should fail, then $h_{*,n}(f)$ is not of the same order as above, as shown in van Eeden (1985) (see also van Es (1991)).

Theorem 3 also gives us some insight into what could be achieved with the use of superkernels. Let K be a superkernel such that $S_K = T_K$, and suppose that $D_f < \infty$. Then, Theorem 3 states that $h_{0,n}(f) \to S_K/D_f > 0$ as $n \to \infty$. Moreover, as $J_{f,K} = S_K/D_f > 0$, for any fixed $h_\star \in (0, J_{f,K}]$ (not depending on n), we have $E_f[f_{n,h_\star}(x)] = f(x)$ for almost every $x \in \mathbb{R}$ and for all $n \in \mathbb{N}$. Thus, in that case we can write

$$M_{f,n}(h_\star) = V_{f,n}(h_\star) = \frac{1}{n}\left[\frac{R(K)}{h_\star} - R(K_{h_\star} * f)\right] = O(n^{-1}).$$

This is the best rate that can be achieved in nonparametric density estimation, whether of kernel type or not, as shown in Boyd and Steele (1978). In fact, the case $D_f < \infty$ is the only one in which the kernel estimate can achieve this rate, as shown in Davis (1977).

The previous bandwidth $h_\star$ is called a global "zero-bias bandwidth" for almost every $x \in \mathbb{R}$. In a similar way, Sain and Scott (2002) show, for non-negative kernels, the existence of a local zero-bias bandwidth $h_0(x)$, not varying with n, for every x in the region where f is convex.

Notice that the existence of this zero-bias bandwidth in the previous situation shows a remarkable difference from the case when a nonnegative kernel is used. Yamato (1972) shows that, when using a positive kernel, no global zero-bias bandwidth can be found for any density.

The next example illustrates that it is possible to exhibit cases where the optimal bandwidth sequence converges to any previously chosen positive value.

Example 1. Consider the Fejér-de la Vallée-Poussin density $f(x) = (\pi x^2)^{-1}(1-\cos x)$, $x \in \mathbb{R}$, and let $f_a(x) = f(x/a)/a$ for any $a > 0$. The characteristic function of f_a is $\varphi_{f_a}(t) = (1-a|t|)\cdot I_{[-1/a,1/a]}(t)$, $t \in \mathbb{R}$, and $C_{f_a} = D_{f_a} = 1/a$. Besides, let us choose the trapezoidal superkernel given by $K(x) = (\pi x^2)^{-1}[\cos x - \cos(2x)]$, $x \in \mathbb{R}$, which has characteristic function $\varphi_K(t) = I_{[0,1)}(|t|) + (2-|t|)\cdot I_{[1,2)}(|t|)$, $t \in \mathbb{R}$, so that $S_K = T_K = 1$. The previous theorem states that $h_{0,n}(f_a) \to a$, as $n \to \infty$.

In fact, for a class of densities g with constant D_g, say $D_g = 1$, we get a zero-bias bandwidth for the trapezoidal superkernel K if we set $h = S_K = 1$ (independently of the sample size n). For instance, such a class is the location family of densities $\{g_{(\mu)}: \mu \in \mathbb{R}\}$, with $g_{(\mu)}(x) = f(x-\mu)$, $x, \mu \in \mathbb{R}$; Theorem 10.5.2 in Kawata (1972) yields a way to construct even a nonparametric family of densities g with $D_g = 1$. Figure 1 illustrates this situation: the top graph is the density f (thick line) and an estimate (thin line) for $n = 100$ using the trapezoidal superkernel; the middle and bottom left graphs show 50 estimates (grey lines) for the same sample size using the trapezoidal superkernel and the Sheather-Jones method (Sheather and Jones (1991)), respectively, while the middle and bottom right graphs exhibit the mean graphs (dashed lines) of these 50 estimates.

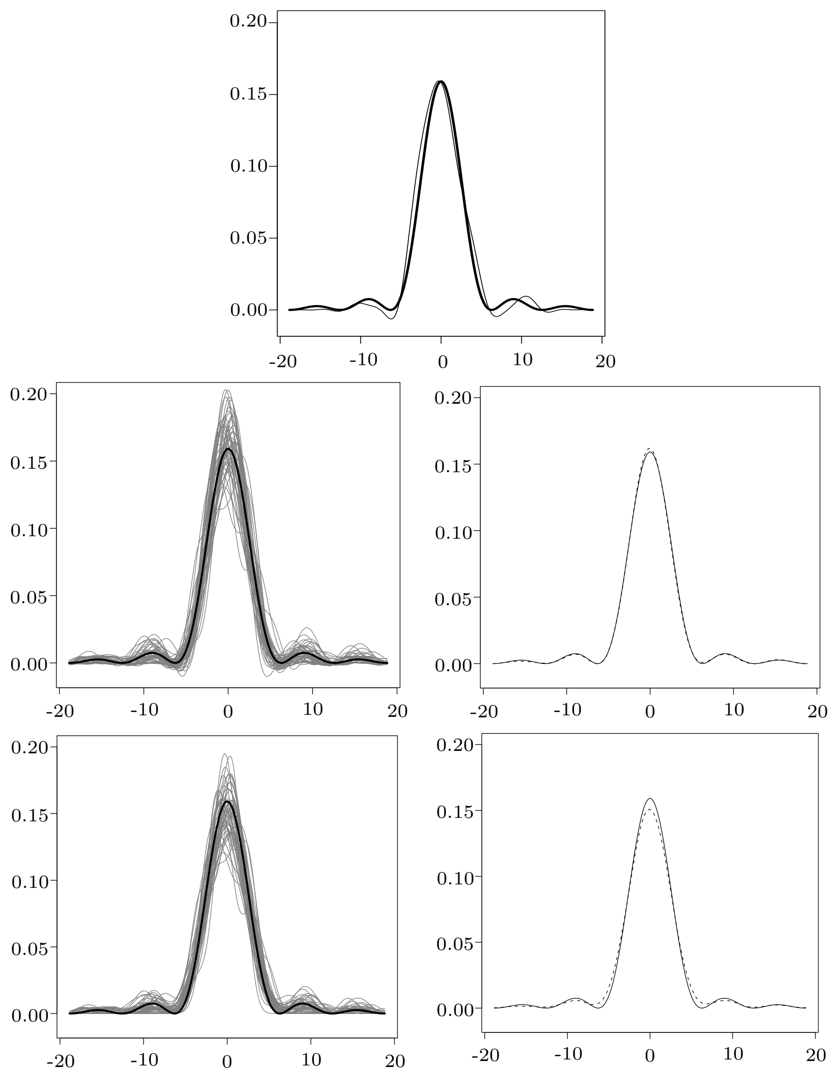

Figure 1. Estimations of the Fejér-de la Vallée-Poussin density.

The so-called "sinc kernel" is $K(x) = \sin x/(\pi x)$, $x \in \mathbb{R}$. In fact, it is not a kernel, since it is not integrable. However, the sinc kernel belongs to $\mathcal{L}_2$ and satisfies VP $\int K(x) = 1$, where VP $\int$ stands for $\lim_{C\to\infty} \int_{-C}^{C}$ and, therefore, it can be considered a "quasi-kernel". The sinc kernel does satisfy some optimality conditions in the $\mathcal{L}_2$ sense (see Davis (1977))) and is sometimes used in kernel density estimation. Based on a kernel density estimator with the sinc kernel, it is possible to construct another estimator, positive and integrating to 1, with smaller MISE, as shown in Glad, Hjort and Ushakov (2003). The principal value of the characteristic function of the sinc kernel is equal to 1 in a neighborhood of the origin, as in the superkernel case, so it is natural to wonder if the optimal bandwidth sequence converges to 0. The answer is, again, no. In fact, we can find examples where the optimal bandwidth sequence converges to any previously chosen positive value, as we did in the superkernel case.

Example 2. For the sinc kernel, it is possible to obtain an exact formula for the MISE. Formula (2.2) in Davis (1977) states that

$$M_{f,n}(h) = R(f) + \frac{1}{n\pi h} - \frac{n+1}{n\pi}\int_0^{\frac{1}{h}} |\varphi_f(t)|^2 dt. \tag{9}$$

Let f_a, $a > 0$, be as in Example 1. Then, using (9), we obtain that

$$M_{f_a,n}(h) = \begin{cases} \frac{n(h-a)^3+3a^2h-a^3}{3an\pi h^3} & \text{for } h \geq a \\ \frac{1}{n\pi h} - \frac{1}{3an\pi} & \text{for } h \leq a. \end{cases}$$

Moreover, $M_{f_a,n}(h)$ has a unique minimizer $h_{0,n}(f_a) = a(n+1+\sqrt{n+1})/n$ that satisfies $h_{0,n}(f_a) \to a$, as $n \to \infty$.

4. Proofs

It is easy to show that, if we write $\bar{\Psi}(u) = \Psi(-u)$ for every real function Ψ, then $R(K_h * f) = R_{K*\bar{K},h}(f)$. Thus, combining (3) with (4) and (5), we come to

$$M_{f,n}(h) = R(f) + \frac{R(K)}{nh} + \frac{n-1}{n} R_{K*\bar{K},h}(f) - 2R_{K,h}(f). \tag{10}$$

Notice that, although $R_{K,h}(f)$ is defined for $h > 0$, it is possible to write

$$R_{K,h}(f) = \int K(u)(f * \bar{f})(hu)du, \tag{11}$$

and this expression also makes sense for $h = 0$ (in fact, for any $h \in \mathbb{R}$), with $R_{K,0}(f) = R(f)$.

Lemma. *Let f be a density function and K a kernel, both belonging to $\mathcal{L}_2$.*

(a) *The function* $h \mapsto R_{K,h}(f)$ *is continuous and satisfies* $\lim_{h\to 0} R_{K,h}(f) = R(f)$ *and* $\lim_{h\to\infty} R_{K,h}(f) = 0$.

(b) *The functions* B_f, $V_{f,n}$ *and* $M_{f,n}$ *are continuous on* $(0,\infty)$ *and satisfy* $\lim_{h\to 0} B_f(h) = 0$, $\lim_{h\to\infty} B_f(h) = R(f)$, $\lim_{h\to 0} V_{f,n}(h) = \infty$, $\lim_{h\to\infty} V_{f,n}(h) = 0$, $\lim_{h\to 0} M_{f,n}(h) = \infty$ *and* $\lim_{h\to\infty} M_{f,n}(h) = R(f)$.

(c) *If, in addition,* K *is bounded and continuous at zero, then* $\lim_{h\to\infty} hR_{K,h}(f) = K(0)$ *and* $\lim_{h\to\infty} h[M_{f,n}(h) - R(f)] = R(K) - 2K(0)$.

Proof. Part (b) is a straightforward consequence of part (a). To show part (a), recall that, as $f \in \mathcal{L}_2$, we have that $f * \bar{f}$ is a continuous function such that $(f * \bar{f})(x) \to 0$ as $|x| \to \infty$ (see Rudin (1990, p.4)) and, therefore, $f * \bar{f}$ is bounded. Hence, in view of (11), the continuity of the function $h \mapsto R_{K,h}(f)$ and its limits follow directly from the Dominated Convergence Theorem.

For the first limit of part (c), using again the Dominated Convergence Theorem, we can exchange the limit and the integral to get

$$\lim_{h\to\infty} hR_{K,h}(f) = \lim_{h\to\infty} \iint K\left(\frac{x-y}{h}\right) f(x)f(y)dxdy = K(0),$$

as desired. The second limit in part c) follows from the first one and the fact that $K * \bar{K}$ is a continuous function such that $(K * \bar{K})(0) = R(K)$.

Proof of Theorem 1. By hypothesis, we have

$$\lim_{h\to\infty} h\left[M_{f,n}(h) - R(f)\right] = R(K) - 2K(0) < 0.$$

Hence, we can choose $h > 0$ big enough so that $M_{f,n}(h) < R(f)$. This fact, together with the properties of the MISE function shown in the lemma, gives the proof.

Next, we give the proofs of the results about the limit behavior of the optimal bandwidth sequence.

Proof of Theorem 2. First, we can prove that the sequence of minimum MISE values goes to 0; that is,

$$\lim_{n\to\infty} \min_{h>0} M_{f,n}(h) = \lim_{n\to\infty} M_{f,n}(h_{0,n}(f)) = 0. \tag{12}$$

To show this, let us suppose that $M_{f,n}(h_{0,n}(f)) \not\to 0$ as $n \to \infty$. Then, there is some $\varepsilon > 0$ and some subsequence of functions (which we denote the same) such that $M_{f,n}(h) \geq M_{f,n}(h_{0,n}(f)) > \varepsilon$ for all $n \in \mathbb{N}$ and $h > 0$. Since it is clear from (4) and (5) that $M_{f,n}(h) \to B_f(h)$ pointwise as $n \to \infty$, it should be that $B_f(h) \geq \varepsilon$ for all $h > 0$, which contradicts the lemma above. Thus, it has to be $M_{f,n}(h_{0,n}(f)) \to 0$ as $n \to \infty$.

Now, suppose that $nh_{0,n}(f) \not\to \infty$ as $n \to \infty$. Then there exists a constant $C > 0$ and a subsequence $\{n_k h_{0,n_k}(f)\}_{k\in\mathbb{N}}$, such that

$$n_k h_{0,n_k}(f) \le C, \quad \forall k \in \mathbb{N}; \tag{13}$$

hence, $h_{0,n_k}(f) \to 0$ as $k \to \infty$. Using (10), together with (12) and the limits of the lemma above, we get that

$$\lim_{k\to\infty} \frac{R(K)}{n_k h_{0,n_k}(f)} = 0.$$

Thus $n_k h_{0,n_k}(f) \to \infty$ as $k \to \infty$, which contradicts (13).

Proof of Theorem 3. The theorem follows easily from the chain of inequalities:

$$\frac{S_K}{D_f} \le \inf_{n\in\mathbb{N}} h_{0,n}(f) \le \limsup_{n\to\infty} h_{0,n}(f) \le J_{f,K} \le \min\left\{\frac{S_K}{C_f}, \frac{T_K}{D_f}\right\}.$$

To prove the first one, notice that we have $B_f(h) = 0$ for all $h \in (0, S_K/D_f]$ because, if $D_f \le S_K/h$, then using the Parseval Identity,

$$\begin{aligned} 0 \le B_f(h) &= \frac{1}{\pi}\int_0^\infty |\varphi_f(t)|^2 |\varphi_K(th) - 1|^2 dt \\ &\le \frac{1}{\pi}\int_0^{\frac{S_K}{h}} |\varphi_f(t)|^2 |\varphi_K(th) - 1|^2 dt + \frac{1}{\pi}\int_{D_f}^\infty |\varphi_f(t)|^2 |\varphi_K(th) - 1|^2 dt = 0, \end{aligned}$$

with the last equality due to the fact that $|\varphi_K(th) - 1| = 0$ for all $t \in [0, S_K/h]$ and $\varphi_f(t) = 0$ for all $t \ge D_f$, by definition of S_K and D_f, respectively. Besides, for $h \in (0, S_K/D_f]$, using again the Parseval Identity, the integrated variance satisfies

$$\begin{aligned} V_{f,n}(h) &= \frac{R(K)}{nh} - \frac{1}{n\pi}\int_0^\infty |\varphi_f(t)|^2 |\varphi_K(th)|^2 dt \\ &= \frac{R(K)}{nh} - \frac{1}{n\pi}\int_0^{D_f} |\varphi_f(t)|^2 dt. \end{aligned}$$

Thus, for $h \in (0, S_K/D_f]$, the MISE can be written as

$$M_{f,n}(h) = \frac{R(K)}{nh} - \frac{R(f)}{n},$$

which is a decreasing function in h. It then has to be that $h_{0,n}(f) \ge S_K/D_f$ for all $n \in \mathbb{N}$.

For the last inequality, if $h > S_K/C_f$ from the definition of S_K there is some nonempty interval $I \subset [S_K/h, C_f]$ such that $|\varphi_K(th) - 1| \ne 0$ for all $t \in I$. Then

$$B_f(h) \ge \frac{1}{\pi}\int_I |\varphi_f(t)|^2 |\varphi_K(th) - 1|^2 dt > 0$$

and, therefore, it must be that $J_{f,K} \leq S_K/C_f$. A completely analogous reasoning can be used to show that $B_f(h) > 0$ for all $h > T_K/D_f$, so we also have $J_{f,K} \leq T_K/D_f$.

Finally, to show the third inequality, set $l_f = \limsup_{n\to\infty} h_{0,n}(f)$ and suppose that $l_f > J_{f,K}$. Then there exists a subsequence $\{h_{0,n_k}\}_k$ such that $\lim_{k\to\infty} h_{0,n_k}(f) = l_f$. If $l_f = \infty$, the lemma allows us to conclude that $B_f(h_{0,n_k}(f)) \to R(f) > 0$ as $k \to \infty$. If $l_f < \infty$ then, as $k \to \infty$, $B_f(h_{0,n_k}(f)) \to B_f(l_f)$, which is also strictly positive since we are supposing that $l_f > J_{f,K}$. In any case, we have that $\lim_{k\to\infty} B_f(h_{0,n_k}(f)) > 0$, which contradicts (12).

Acknowledgement

The authors would like to thank both referees, an associate editor and Professor Wang for several comments and suggestions that have led to a sensible improvement of the paper. This research is supported by Spanish Ministerio de Ciencia y Tecnología project BFM2002-01217.

References

Boyd, D. W. and Steele, J. M. (1978). Lower bounds for nonparametric density estimation rates. *Ann. Statist.* **6**, 932-934.

Davis, K. B. (1977). Mean integrated square error properties of density estimates. *Ann. Statist.* **5**, 530-535.

Devroye, L. (1992). A note on the usefulness of superkernels in density estimation. *Ann. Statist.* **20**, 2037-2056.

Devroye, L. and Lugosi, G. (2001). *Combinatorial Methods in Density Estimation*. Springer-Verlag, New York.

Glad, I. K., Hjort, N. L. and Ushakov, N. G. (2003). Correction of density estimators that are not densities. *Scand. J. Statist.* **30**, 415-427.

Hall, P. and Marron, J. S. (1991). Lower bounds for bandwidth selection in density estimation. *Probab. Theory Related Fields* **90**, 149-163.

Kawata, T. (1972). *Fourier Analysis in Probability Theory*. Academic Press, New York.

Marron, J. S. and Wand, M. P. (1992). Exact mean integrated square error. *Ann. Statist.* **20**, 712-736.

Rudin, W. (1990). *Fourier Analysis on Groups*. Wiley, New York.

Sain, S. R. and Scott, D. W. (2002). Zero-bias locally adaptive density estimators. *Scand. J. Statist.* **29**, 441-460.

Schucany, W. R. and Sommers, J. P. (1977). Improvement of kernel-type density estimators. *J. Amer. Statist. Assoc.* **72**, 420-423.

Scott, D. W. (1992). *Multivariate Density Estimation: Theory, Practice and Visualization*. Wiley, New York.

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

Silverman, B. W. (1986). *Density Estimation for Statstics and Data Analysis*. Chapman and Hall, London.

Stone, C. J. (1984). An asymptotically optimal window selection rule for kernel density estimates. *Ann. Statist.* **12**, 1286-1297.

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

van Es, A. J. (1991). *Aspects of Nonparametric Density Estimation, CWI Tract* 77. CWI, Amsterdam.

Wand, M. P. and Jones, M. C. (1995). *Kernel Smoothing*. Chapman and Hall, London.

Wand, M. P. and Schucany, W. R. (1990). Gaussian-based kernels. *Canad. J. Statist.* **18**, 197-204.

Yamato, H. (1972). Some statistical properties of estimators of density and distribution functions. *Bull. Math. Statist.* **15**, 113-131.

Dpto. de Matemáticas, Universidad de Extremadura, Avda. de Elvas, s/n, 06071-Badajoz, Spain.

E-mail: jechacon@unex.es

Dpto. de Matemáticas, Universidad de Extremadura, Avda. de Elvas, s/n, 06071-Badajoz, Spain.

E-mail: jmf@unex.es

Dpto. de Matemáticas, Universidad de Extremadura, Avda. de Elvas, s/n, 06071-Badajoz, Spain.

E-mail: nogales@unex.es

Dpto. de Matemáticas, Universidad de Extremadura, Avda. de Elvas, s/n, 06071-Badajoz, Spain.

E-mail: paloma@unex.es

(Received July 2004; accepted August 2005)