Pré-Publicações do Departamento de Matemática
Universidade de Coimbra
Preprint Number 09–33

ON THE EQUIVALENCE OF EXACT AND ASYMPTOTICALLY OPTIMAL BANDWIDTHS FOR KERNEL ESTIMATION OF DENSITY FUNCTIONALS

JOSÉ E. CHACÓN AND CARLOS TENREIRO

Abstract: Given a density f we pose the problem of estimating the density functional $\psi_r = \int f^{(r)} f$ making use of kernel methods. This is a well-known problem but some of its features remained unexplored. We focus on the problem of bandwidth selection. Whereas all the previous studies concentrate on an asymptotically optimal bandwidth here we study the properties of exact, non-asymptotic ones, and relate them with the former. Our main conclusion is that, despite being asymptotically equivalent, for realistic sample sizes much is lost by using the asymptotically optimal bandwidth. In contrast, as a target for data-driven selectors we propose another bandwidth which retains the small sample performance of the exact one.

Keywords: Density functional, exact optimal bandwidth, kernel estimator, normal mixture densities.
AMS Subject Classification (2000): 62G05, 62G20.

1. Introduction

Given a sample $X_1, \dots, X_n$ of independent and identically distributed real random variables with unknown density function $f\colon \mathbb{R} \to \mathbb{R}$, in this paper we focus on the problem of estimating the functional

$$\psi_r = \int f^{(r)}(x) f(x) dx \tag{1}$$

for even r whenever it makes sense and is finite, where $f^{(r)}$ denotes the rth derivative of f. Notice that for such a functional to be finite it suffices, for instance, that both f and $f^{(r)}$ be square integrable.

There exists a wide variety of estimates of these functionals. For instance, van Es (1992) proposes an estimator of ψ_0 based on the spacings of the order statistics and Laurent (1997) and Prakasa Rao (1999), respectively, describe series and wavelet estimates for the problem. However, here we

Received September 3, 2009.
This research has been partially supported by the Spanish Ministerio de Ciencia y Tecnología project MTM2006-06172 (first author) and by the CMUC (Centre for Mathematics, University of Coimbra)/FCT.

will concentrate on kernel estimators,

$$\hat{\psi}_r(g) = \frac{1}{n^2} \sum_{i,j=1}^{n} L_g^{(r)}(X_i - X_j), \tag{2}$$

where L is the kernel function, $g > 0$ is the bandwidth and $L_g^{(r)}$ represents the rth derivative of the function $L_g(x) = L(x/g)/g$, that is, $L_g^{(r)}(x) = L^{(r)}(x/g)/g^{r+1}$. The motivation for this precise type of kernel estimator can be found, for instance, in Wand and Jones (1995).

This problem is also addressed in many other papers. For instance, the case $r = 0$ (estimation of the integral of a squared density) is closely related with the study of rank-based nonparametric statistics, since it appears in the asymptotic variance of the Wilcoxon signed-rank statistic and in the Pitman asymptotic efficiency of the Wilcoxon test relative to the t-test (see Hettmansperger, 1984). The first kernel estimators of ψ_0 date back to at least Bhattacharya and Roussas (1969), Dmitriev and Tarasenko (1973, 1975) and Schuster (1974), but see also Prakasa Rao (1983), Sheather, Hettmansperger and Donald (1994), and references therein. A recent paper on the topic is Giné and Nickl (2008).

The quantities ψ_2, ψ_4 and ψ_6 appear in the expression of the asymptotically optimal bandwidths for histogram, frequency polygon and kernel density estimators (see Scott, 1992). The first papers analyzing the kernel-type estimates of ψ_r for arbitrary r, as a particular case of a more general nonlinear functional, are Dmitriev and Tarasenko (1973) and Levit (1978), but the problem of bandwidth selection for the kernel estimator is considered in Hall and Marron (1987) for the first time, although there the kernel estimate is defined as $n(n-1)^{-1}\{\hat{\psi}_r(g) - n^{-1}L_g^{(r)}(0)\}$, in order to delete the non-stochastic terms in $\hat{\psi}_r(g)$. However, Jones and Sheather (1991) show that indeed the estimator $\hat{\psi}_r(g)$ has improved rates of convergence over the one proposed by Hall and Marron when the bandwidth g is properly chosen. On the other hand, Bickel and Ritov (1988) discuss the information bounds for this nonparametric problem and propose an efficient estimator. References dealing with adaptive kernel procedures include Wu (1995) and Giné and Mason (2008), among others. Multistep kernel estimators are investigated in Aldershof (1991) and also more recently in Tenreiro (2003) and Chacón and Tenreiro (2008).

As usual for real-valued parameters, we will measure the accuracy of the estimator $\hat{\psi}_r(g)$ through its mean squared error (MSE), defined as $\mathrm{MSE}(g) = \mathbb{E}[\{\hat{\psi}_r(g) - \psi_r\}^2]$. In this sense, the optimal bandwidth can be defined to be $g_{\mathrm{MSE}} = \operatorname{argmin}_{g>0} \mathrm{MSE}(g)$. However, it is not clear at all from its definition that such a minimizer exists, and well-experienced researchers in the field take good care not to refer to this bandwidth, but to its asymptotic counterpart (see Jones and Sheather, 1991, or Wand and Jones, 1995). In fact, the typical approach to bandwidth selection starts from considering an asymptotic expansion of the MSE function, say AMSE(g), and considering the asymptotically optimal bandwidth $g_0 = \operatorname{argmin}_{g>0} \text{AMSE}(g)$ as a surrogate for g_{MSE}, which is the exact (i.e., non-asymptotic) one. The study of the asymptotically optimal bandwidth presents no doubts about its existence, and even an explicit formula for it is available. But then another question may be raised: how well does g_0 approximate g_{MSE}? The study of this question leads to the identification of a new bandwidth g_{BA} that annihilates the exact bias of $\hat{\psi}_r(g)$. How well does this new bandwidth approximate g_{MSE} is another question that arises naturally. Therefore, the main purposes of this paper are to present a set of sufficient conditions to the existence of an exact optimal bandwidth and to examine, from an asymptotic and finite sample size point of view, the quality of g_0 and g_{BA} as approximations of the exact optimal bandwidth.

The rest of the paper is organized as follows. In Section 2 we provide mild conditions on the kernel and the density that ensure the existence of an exact optimal bandwidth g_{MSE} and a bias-annihilating bandwidth g_{BA}. In Section 3 we study the asymptotic properties of these bandwidths. In Section 4 we obtain the relative rates of convergence of g_0 and g_{BA} to g_{MSE} and so we quantify the order of these asymptotic approximations. We also establish the order of convergence for $\text{MSE}(g_0) - \text{MSE}(g_{\text{BA}})$ which enables us to compare g_0 and g_{BA} in the sense of the mean squared error. As the results in Section 4 are asymptotic in nature, to assess the quality of the approximations Section 5 contains the case-study of normal mixture densities, for which small- and moderate-sample-size comparisons are made between the three different bandwidths. We will see that for small and moderate sample sizes $\text{MSE}(g_{\text{BA}})$ seems to be much closer to $\text{MSE}(g_{\text{MSE}})$ than $\text{MSE}(g_0)$. In view of these finite sample size results we conclude that bandwidth selectors oriented to g_{BA} should be preferred to the usual ones, which are designed to estimate g_0. All the proofs are deferred to Section 6.

2. Existence of an exact optimal bandwidth

Recall the definitions of ψ_r and $\hat{\psi}_r(g)$ from (1) and (2) in Section 1. The mean squared error (MSE) of the estimator $\hat{\psi}_r(g)$ can be decomposed as $\text{MSE}(g) = B^2(g) + V(g)$, where $B(g)$ and $V(g)$ are the bias and variance of $\hat{\psi}_r(g)$. If we denote

$$\begin{aligned} R_{L,r,g}(f) &= \mathbb{E}L_g^{(r)}(X_1 - X_2) \\ &= \iint L_g^{(r)}(x-y)f(x)f(y)dxdy = \int (L_g^{(r)} * f)(x)f(x)dx, \end{aligned}$$

with $*$ standing for the convolution operator, then it is clear that

$$B(g) = \mathbb{E}\hat{\psi}_r(g) - \psi_r = n^{-1}g^{-r-1}L^{(r)}(0) + (1-n^{-1})R_{L,r,g}(f) - \psi_r. \quad (3)$$

Moreover, using standard U-statistics theory we get that $V(g) = \operatorname{Var}\hat{\psi}_r(g)$ can be written as

$$V(g) = 4(n-2)(n-1)n^{-3}\xi_1 + 2(n-1)n^{-3}\xi_2 - (4n-6)(n-1)n^{-3}\xi_0, \quad (4)$$

where $\xi_0 = \mathbb{E}[L_g^{(r)}(X_1 - X_2)]^2$, $\xi_1 = \mathbb{E}[L_g^{(r)}(X_1 - X_2)L_g^{(r)}(X_1 - X_3)]$ and $\xi_2 = \mathbb{E}[L_g^{(r)}(X_1 - X_2)^2]$. If we denote

$$\begin{aligned} S_{L,r,g}(f) &= \iiint L_g^{(r)}(x-y)L_g^{(r)}(x-z)f(x)f(y)f(z)dxdydz \\ &= \int (L_g^{(r)} * f)(x)^2 f(x)dx, \end{aligned}$$

we just have $\xi_1 = S_{L,r,g}(f)$. Besides, clearly $\xi_0 = R_{L,r,g}(f)^2$ and, using the fact that $L_g^{(r)}(x)^2 = g^{-2r-1}[(L^{(r)})^2]_g$, we can also express $\xi_2 = g^{-2r-1}R_{(L^{(r)})^2,0,g}(f)$.

Combining (3) and (4) with the former representations for ξ_0, ξ_1, ξ_2, we obtain an exact formula for the MSE of the estimator $\hat{\psi}_r(g)$,

$$\begin{aligned} \mathrm{MSE}(g) = &\left\{n^{-1}g^{-r-1}L^{(r)}(0) + (1-n^{-1})R_{L,r,g}(f) - \psi_r\right\}^2 \quad (5) \\ &+ 4(n-2)(n-1)n^{-3}S_{L,r,g}(f) + 2(n-1)n^{-3}g^{-2r-1}R_{(L^{(r)})^2,0,g}(f) \\ &- (4n-6)(n-1)n^{-3}R_{L,r,g}(f)^2. \end{aligned}$$

This exact error formula is analogue of formula (2.2) in Marron and Wand (1992) for kernel density estimators, and will be useful to explore the existence and limit behavior of the optimal bandwidth as well as for the results in Section 5.

In the following we will make the following assumptions on the kernel and the density:

- (L1) L is a symmetric kernel with bounded and square integrable derivatives up to order r such that $L^{(r)}$ is continuous at zero with $(-1)^{r/2}L^{(r)}(0) > 0$.
- (D1) The density f has bounded and square integrable derivatives up to order r.

The next result shows that under these mild conditions there is always an exact optimal bandwidth, that is, a bandwidth which minimizes the exact MSE of the kernel estimator. In this sense, it can be considered as the analogue of Theorem 1 in Chacón *et al.* (2007) for kernel density estimators.

Theorem 1. *Under assumptions (L1) and (D1), there exists* $g_{\mathrm{MSE}} = g_{\mathrm{MSE},r,n}(f)$ *such that* $\mathrm{MSE}(g_{\mathrm{MSE}}) \leq \mathrm{MSE}(g)$, *for all* $g > 0$.

Notice that the previous result says nothing about the uniqueness of the optimal bandwidth. Presumably, as in the examples in Marron and Wand (1992) it could be possible to find a situation where the optimal bandwidth is not unique, however we do not pursue this further in this paper.

From an asymptotic point of view, however, it is well-known that the choice of g can be made on the basis of annihilation of the dominant part of the bias (see Section 4 below). We show next that, in fact, for every density f there is a choice of $g = g_{\mathrm{BA}}$ that makes the estimator $\hat{\psi}_r(g)$ unbiased, that is, that annihilates the exact bias, rather than its asymptotic counterpart.

Theorem 2. *Under assumptions (L1) and (D1), there exists* $g_{\mathrm{BA}} = g_{\mathrm{BA},r,n}(f)$ *such that* $B(g_{\mathrm{BA}}) = 0$.

The existence of global bandwidths that make the kernel density estimate unbiased at every point has been shown in Chacón *et al.* (2007). In fact, strictly speaking we cannot consider it an unbiased estimator since such bandwidths depend on the unknown f, but at least we could say that there exists an 'unbiased oracle estimator'. However, only a very special class of density functions allows for this situation, namely the class of densities whose characteristic function has bounded support.

In contrast, in the previous result we show that unbiased oracle kernel estimates of ψ_r (not only asymptotical unbiased) exist under the same mild conditions needed for the existence of the optimal bandwidth. This is a key difference between the problems of estimating the density and the functionals ψ_r.

3. Limit behavior of exact bandwidths

From formula (5) and Lemma 1 in Section 6 below it readily follows that $\mathrm{MSE}(g) \to 0$ for *any* bandwidth sequence $g = g_n$ such that $g \to 0$ and $ng^{r+1} \to \infty$ as $n \to \infty$. Therefore, conditions $g \to 0$ and $ng^{r+1} \to \infty$ are sufficient for $\hat{\psi}_r(g)$ to be consistent. It is natural, then, to wonder if the bandwidths g_{MSE} and g_{BA} also fulfill the previous consistency conditions. We will see that that the second condition holds quite generally but the same is not necessarily true for the first one. This is similar to the situation with the optimal bandwidth for kernel density estimation, as shown in Chacón *et al.* (2007).

Theorem 3. *Under assumptions (L1) and (D1), both* $ng_{\mathrm{MSE}}^{r+1} \to \infty$ *and* $ng_{\mathrm{BA}}^{r+1} \to \infty$ *as* $n \to \infty$.

For the analysis of the limit behaviour of the sequences g_{MSE} and g_{BA} we use the notation $\varphi_F(t) = \int e^{itx}F(x)dx$, $t \in \mathbb{R}$, for the characteristic function of an integrable real function F, and for every density f and every symmetric kernel L, we denote

$$\begin{aligned} C_f &= \sup\{r \geq 0 : \varphi_f(t) \neq 0 \ \text{ a.e. for } t \in [0, r]\}, \\ D_f &= \sup\{t \geq 0 : \varphi_f(t) \neq 0\}, \\ S_L &= \inf\{t \geq 0 : \varphi_L(t) \neq 1\}, \\ T_L &= \inf\{r \geq 0 : \varphi_L(t) \neq 1 \ \text{ a.e. for } t \geq r\} \end{aligned}$$

A detailed discussion about these quantities is presented in Chacón *et al.* (2007). In particular, we remark that all these exist, with C_f, D_f possibly being infinite, $S_L, T_L \in [0, \infty)$, $C_f \leq D_f$ and $S_L \leq T_L$. Notice that, by definition, $S_L > 0$ for superkernels (see Chacón, Montanero and Nogales, 2007).

In the following we show that both the exact optimal bandwidth g_{MSE} and the exact bias-annihilating bandwidth g_{BA} converge to zero under very general conditions. In particular, if L is a kernel of finite order (that is, $|m_\nu|(L) = \int |u^\nu L(u)|du < \infty$ and $m_\nu(L) = \int u^\nu L(u)du \neq 0$ for some even number ν), the convergence to zero takes place with no additional conditions on f other than (D1). The same property occurs in the superkernel case whenever the characteristic function of f has unbounded support.

Theorem 4. *Assume conditions (L1) and (D1). If* $S_L = 0$ *or* $D_f = \infty$ *we have both* $g_{\mathrm{MSE}} \to 0$ *and* $g_{\mathrm{BA}} \to 0$ *as* $n \to \infty$.

In the remaining case $S_L > 0$ and $D_f < \infty$ non-zero limits may occur. In the next example we show that if we use a superkernel and the characteristic function of the density has finite support then any positive number is a possible limit for g_{MSE} or g_{BA}.

Example 1. As in Chacón *et al.* (2007), consider the trapezoidal superkernel given by $L(x) = (\pi x^2)^{-1}[\cos x - \cos(2x)]$ for $x \neq 0$ and $L(0) = 3/(2\pi)$, whose characteristic function is $\varphi_L(t) = I_{[0,1)}(|t|) + (2 - |t|)\, I_{[1,2)}(|t|)$, with $I_A(t)$ standing for the indicator function of the set A, so that $S_L = T_L = 1$. This kernel is symmetric, differentiable of any order, with bounded

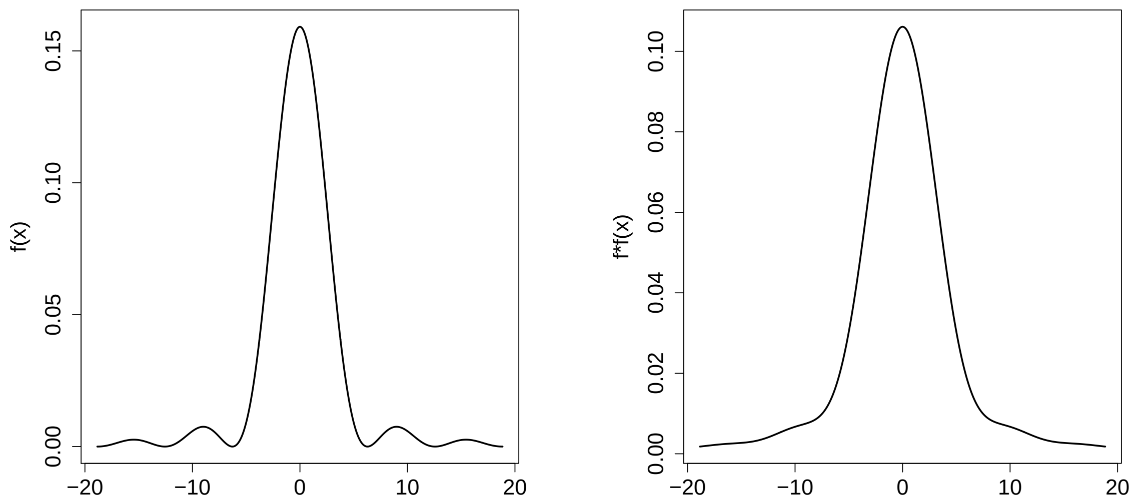

FIGURE 1. *Fejér-de la Vallée-Poussin density (left) and the convolution with itself (right).*

square integrable derivatives, and such that $L^{(r)}(0) = (-1)^{r/2}[\pi(r+1)(r+2)]^{-1}(2^{r+2}-1)$, so that it fulfils condition (L1).

Consider also the Fejér-de la Vallée-Poussin density, defined as $f(x) = (\pi x^2)^{-1}(1-\cos x)$ for $x \neq 0$ and $f(0) = 1/(2\pi)$, and let $f_a(x) = f(x/a)/a$ for any $a > 0$; see Figure 1. This density is differentiable of any order, with bounded square integrable derivatives. The characteristic function of f_a is $\varphi_{f_a}(t) = (1 - a|t|)\, I_{[-1/a,1/a]}(t)$, so that $C_{f_a} = D_{f_a} = 1/a$. Besides, we easily obtain $\psi_r = (-1)^{r/2} 2[\pi a^{r+1}(r+1)(r+2)(r+3)]^{-1}$.

From (13) in Section 6 below we know that $\limsup g \leq a$ for both $g = g_{\text{MSE}}$ and $g = g_{\text{BA}}$. Also, using the formulas for the MSE in the Fourier domain given in Section 6 it is not hard to show that, in this case, for $g \in (0, S_L/D_{f_a}] = (0, a]$ we have

$$B(g) = n^{-1}g^{-r-1}L^{(r)}(0) - n^{-1}\psi_r,$$
$$V(g) = 2(n-1)n^{-3}g^{-2r-1}R_{(L^{(r)})^2,0,g}(f_a) + A,$$

where $A \in \mathbb{R}$ is a constant depending on L, f, r and n, but not on g.

With the formulas for $L^{(r)}(0)$ and ψ_r given above, it is clear that $B(g) \neq 0$ for $g \in (0, a]$, so that it should be $g_{\text{BA}} \geq a$ for every $n \in \mathbb{N}$ and this, together with the upper bound for the limsup, implies that $g_{\text{BA}} \to a$ as $n \to \infty$.

On the other hand, it can be shown that $f * f(x) = 2(\pi x^3)^{-1}(x - \sin x)$ for $x \neq 0$ and $f * f(0) = 1/(3\pi)$, so that $f * f$ is a symmetric and decreasing density for $x > 0$, and the same is true for $f_a * f_a$. Therefore the function $g \mapsto R_{(L^{(r)})^2,0,g}(f_a)$ is decreasing since from (9) in Section 6 we can write $R_{(L^{(r)})^2,0,g}(f_a) = 2\int_0^\infty L^{(r)}(u)^2(f_a * f_a)(gu)du$. This implies that for $g \in (0, a]$

the function $\text{MSE}(g) = B^2(g) + V(g)$ is decreasing and so, that $g_{\text{MSE}} \geq a$ for every $n \in \mathbb{N}$, leading to $g_{\text{MSE}} \to a$ as $n \to \infty$.

4. The asymptotically optimal bandwidth

It is well known that the finite sample performance of $\hat{\psi}_r(g)$ depends strongly on the choice of the bandwidth g. In practice, this choice is usually based on the so called asymptotically optimal bandwidth, g_0, that is, the bandwidth that minimizes the main terms of an asymptotic expansion of $\text{MSE}(g)$ when g tends to zero (see Jones and Sheather, 1991). In order to present such an expansion, some additional conditions on the density f and on the kernel L are needed.

(L2) L is a kernel of finite order ν (even); that is, $\nu = \min\{j \in \mathbb{N}, j \geq 1 \colon m_j(L) \neq 0\}$, so that $m_j(L) = 0$ for $j = 1, 2, \ldots, \nu - 1$. Besides, $(-1)^{\nu/2} m_\nu(L) < 0$.

(D2) The density f has bounded and continuous derivatives up to order $r + \nu$.

Under conditions (L1), (L2), (D1) and (D2), if $g \to 0$ the bias and variance of $\hat{\psi}_r(g)$ given by (3) and (4), respectively, admit the asymptotic expansions

$$B(g) = n^{-1} g^{-r-1} L^{(r)}(0) + g^\nu \psi_{r+\nu} m_\nu(L)/\nu! - n^{-1}\psi_r + o(g^\nu) \tag{6}$$

and

$$V(g) = 4n^{-1}\text{Var} f^{(r)}(X_1) + O(n^{-1}g^\nu + n^{-2}g^{-2r-1}).$$

Therefore,

$$\begin{aligned}\text{MSE}(g) = 4n^{-1}\text{Var} f^{(r)}(X_1) + (n^{-1}g^{-r-1}L^{(r)}(0) + g^\nu \psi_{r+\nu} m_\nu(L)/\nu!)^2 \\ + o(n^{-2}g^{-2r-2} + n^{-1}g^{\nu-r-1} + g^{2\nu}),\end{aligned} \tag{7}$$

and the asymptotically optimal bandwidth corresponds to the value of g that makes the dominant term of the bias vanish, that is,

$$g_0 = \left(-\frac{\nu! L^{(r)}(0)}{m_\nu(L)\psi_{r+\nu} n} \right)^{1/(r+\nu+1)}. \tag{8}$$

Notice that the term inside the parenthesis is positive with our assumptions, since we have $(-1)^{r/2}L^{(r)}(0) > 0$, $(-1)^{(r+\nu)/2}\psi_{r+\nu} > 0$ and $(-1)^{\nu/2}m_\nu(L) < 0$.

As the practical choice of g is usually based on this asymptotically optimal bandwidth, g_0, it is natural to wonder if g_0 is a good approximation of the exact optimal bandwidth, g_{MSE}. In the following theorem we establish the asymptotic equivalence between g_0, g_{BA} and g_{MSE}, and also the order of convergence to zero of the relative error $g_0/g_{\mathrm{MSE}} - 1$, $g_{\mathrm{BA}}/g_{\mathrm{MSE}} - 1$ and $g_0/g_{\mathrm{BA}} - 1$.

Theorem 5. *Under assumptions (L1), (L2), (D1) and (D2) we have:*

a) *The bandwidths g_{MSE}, g_{BA} and g_0 are all of the same order; that is,*

$$0 < \liminf\ n^{1/(r+\nu+1)} g_{\mathrm{MSE}} \leq \limsup\ n^{1/(r+\nu+1)} g_{\mathrm{MSE}} < \infty,$$
$$0 < \liminf\ n^{1/(r+\nu+1)} g_{\mathrm{BA}} \leq \limsup\ n^{1/(r+\nu+1)} g_{\mathrm{BA}} < \infty.$$

b) *Additionally, if $\int |u| L^{(r)}(u)^2 du < \infty$ then*

$$g_0/g_{\mathrm{MSE}} \to 1 \quad \textit{and} \quad g_{\mathrm{BA}}/g_{\mathrm{MSE}} \to 1.$$

c) *Moreover, if $|m_{\nu+2}|(L) < \infty$ and f has bounded continuous derivatives up to order $r + \nu + 2$, then there exist constants C, D and E such that*

$$g_0/g_{\mathrm{MSE}} - 1 = C\, n^{-1/(r+\nu+1)}(1 + o(1)),$$
$$g_{\mathrm{BA}}/g_{\mathrm{MSE}} - 1 = D\, n^{-1/(r+\nu+1)}(1 + o(1)),$$
$$g_0/g_{\mathrm{BA}} - 1 = E\, n^{-\min\{r+1,2\}/(r+\nu+1)}(1 + o(1)).$$

From the previous result we see that asymptotically g_0 and g_{BA} approximate g_{MSE} at the same rate. Using a simple Taylor expansion we can prove that $\mathrm{MSE}(g_0)$ and $\mathrm{MSE}(g_{\mathrm{BA}})$ also approximate $\mathrm{MSE}(g_{\mathrm{MSE}})$ at the same rate. In fact, for $g = g_0$ and $g = g_{\mathrm{BA}}$ we have $\mathrm{MSE}(g) - \mathrm{MSE}(g_{\mathrm{MSE}}) = O(n^{-(2\nu+2)/(r+\nu+1)})$. In the next result we restrict our attention to the order of convergence to zero of $\mathrm{MSE}(g_0) - \mathrm{MSE}(g_{\mathrm{BA}})$ which enables us to compare the bandwidths g_0 and g_{BA} in the sense of the mean squared error.

Theorem 6. *Under assumptions (L1), (L2) and (D1), if f has bounded and continuous derivatives up to order $r + \nu + 2$, $|m_{\nu+2}|(L) < \infty$ and $\int |u|^3 L^{(r)}(u)^2 du < \infty$, then there exists a constant Λ such that*

$$\mathrm{MSE}(g_0) - \mathrm{MSE}(g_{\mathrm{BA}}) = \Lambda E\, n^{-\min\{r+2\nu+2, 2\nu+3\}/(r+\nu+1)}(1 + o(1)),$$

where E is the constant appearing in Theorem 5.

Explicit formulas for the constants C, D, E and Λ appearing in Theorems 5 and 6 are given in Section 6. From them we see that $C = D < 0$ and $\Lambda < 0$ for all densities f whenever $r \geq 2$, and also $E < 0$ if the kernel L is such that $(-1)^{\nu/2} m_{\nu+2}(L) < 0$ (which is in particular true for the Gaussian-based kernel L to be used in the next section). Consequently, from an asymptotic point of view we conclude that g_{BA} is not only a better approximation to g_{MSE} than g_0 but is also a better bandwidth than g_0 in the MSE sense because in this case the constant ΛE appearing in Theorem 6 is strictly positive. As we will see in the next section, even for small and moderate sample sizes $\mathrm{MSE}(g_{\mathrm{BA}})$ seems to be much closer to $\mathrm{MSE}(g_{\mathrm{MSE}})$ than $\mathrm{MSE}(g_0)$.

A different situation may occur when $r = 0$. When the kernel L is of order ν, for all densities f satisfying $\int f^{(\nu)} f^2 / \int f^{(\nu)} f - \int f^2 > 0$ (which seems to be true for all sufficiently regular densities although we were not able to prove it) the constants C and D remain negative but in this case C is always bigger than D which implies that $E > 0$. Hence, the asymptotically optimal bandwidth g_0 is a better asymptotic approximation for g_{MSE} than g_{BA}. Also, we can prove that $\Lambda < 0$, so that in the MSE sense it follows that asymptotically g_0 is better than g_{BA} too. Although this is valid asymptotically, we will see in next section that for small and moderate sample sizes g_{BA} may still be preferable to g_0 in some cases.

5. Case study: normal mixture densities

Our goal in this section is to compare the performance of the three bandwidths, g_{MSE}, g_{BA} and g_0, in a non-asymptotic way. To this end we work with the exact MSE formula within the class of normal mixture densities, that is, the class of densities f that can be written as $f(x) = \sum_{\ell=1}^{k} w_\ell \phi_{\sigma_\ell}(x - \mu_\ell)$, where $\phi(x) = (2\pi)^{-1/2} e^{-x^2/2}$ for any $x \in \mathbb{R}$. This class is very rich, containing densities with a wide variety of features, such as kurtosis, skewness, multimodality, etc, and has been previously used for computing exact errors in the context of kernel density estimation (see Marron and Wand, 1992).

We are going to find an explicit formula for the MSE given in (5) in the case where f is the aforementioned normal mixture density and L is the Gaussian-based kernel of even order ν considered in Wand and Schucany (1990), given by $L(x) = \sum_{s=0}^{\nu/2-1} (-1)^s (2^s s!)^{-1} \phi^{(2s)}(x)$. Note that we only need to obtain explicit formulas for $L^{(r)}(0)$, $R_{L,r,g}(f)$, ψ_r, $S_{L,r,g}(f)$ and $R_{(L^{(r)})^2,0,g}(f)$.

For any even $r_1, r_2 \in \mathbb{N}$, $\mu_1, \mu_2, \mu_3 \in \mathbb{R}$ and $\sigma_1 > 0, \sigma_2 > 0, \sigma_3 > 0$, write

$$\tilde{\mu} = \Big\{ \sigma_1^{-2}\sigma_2^{-2}(\mu_1 - \mu_2)^2 + \sigma_1^{-2}\sigma_3^{-2}(\mu_1 - \mu_3)^2 + \sigma_2^{-2}\sigma_3^{-2}(\mu_2 - \mu_3)^2 \Big\}^{1/2},$$

$$\tilde{\sigma} = \Big\{ \sigma_1^{-2} + \sigma_2^{-2} + \sigma_3^{-2} \Big\}^{1/2}, \quad \tilde{\tilde{\mu}} = \tilde{\sigma}^{-2} \Big\{ \sigma_1^{-2}\mu_1 + \sigma_2^{-2}\mu_2 + \sigma_3^{-2}\mu_3 \Big\}.$$

and $\mu_k^{\dagger} = \mu_k - \tilde{\tilde{\mu}}$. Then, for $\boldsymbol{\mu} = (\mu_1, \mu_2, \mu_3)$ and $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \sigma_3)$ let us denote

$$I_{r_1,r_2}(\boldsymbol{\mu};\boldsymbol{\sigma}) = (2\pi)^{-1/2}\phi_{\tilde{\sigma}}(\tilde{\mu})(\sigma_1\sigma_2\sigma_3)^{-1}\sum_{j_1=0}^{r_1}\sum_{j_2=0}^{r_2}\mathrm{OF}(j_1+j_2)$$
$$\times\binom{r_1}{j_1}\binom{r_2}{j_2}H_{r_1-j_1}(\sigma_1^{-1}\mu_1^\dagger)H_{r_2-j_2}(\sigma_2^{-1}\mu_2^\dagger)\sigma_1^{-r_1-j_1}\sigma_2^{-r_2-j_2}\tilde{\sigma}^{-j_1-j_2},$$

where for any $p\in\mathbb{N}$ we write $\mathrm{OF}(2p) = (2p-1)(2p-3)\cdots 3\cdot 1 = (2p)!(2^p p!)^{-1}$, $\mathrm{OF}(2p+1) = 0$ and $H_p(x)$ the pth Hermite polynomial, defined by $H_p(x) = (-1)^p\phi^{(p)}(x)/\phi(x)$.

Theorem 7. *For $L(x) = \sum_{s=0}^{\nu/2-1}(-1)^s(2^s s!)^{-1}\,\phi^{(2s)}(x)$ and $f(x) = \sum_{\ell=1}^k w_\ell\,\phi_{\sigma_\ell}(x-\mu_\ell)$ we have*

$$B(g) = (-1)^{r/2}n^{-1}g^{-r-1}(2\pi)^{-1/2}\sum_{s=0}^{\nu/2-1}(-1)^s(2^s s!)^{-1}\mathrm{OF}(2s+r)$$
$$+\sum_{\ell,\ell'=1}^{k}w_\ell w_{\ell'}\Big\{(1-n^{-1})\sum_{s=0}^{\nu/2-1}(-1)^s(2^s s!)^{-1}\phi^{(2s+r)}_{\sigma_{\ell\ell'}(g)}(\mu_{\ell\ell'}) - \phi^{(r)}_{\sigma_{\ell\ell'}}(\mu_{\ell\ell'})\Big\}$$

$$V(g) = 4(n-2)(n-1)n^{-3}\sum_{\ell_1,\ell_2,\ell_3=1}^{k}w_{\ell_1}w_{\ell_2}w_{\ell_3}$$
$$\times\sum_{s,s'=0}^{\nu/2-1}(-1)^{s+s'}(2^{s+s'}s!s'!)^{-1}I_{2s+r,2s'+r}(\mu_{\ell_1},\mu_{\ell_2},\mu_{\ell_3};\sigma_{\ell_1}(g),\sigma_{\ell_2}(g),\sigma_{\ell_3})$$
$$+2(n-1)n^{-3}\sum_{\ell,\ell'=1}^{k}w_\ell w_{\ell'}\sum_{s,s'=0}^{\nu/2-1}(-1)^{s+s'}(2^{s+s'}s!s'!)^{-1}$$
$$\times I_{2s+r,2s'+r}(0,0,\mu_{\ell\ell'};g,g,\sigma_{\ell\ell'})$$
$$-(4n-6)(n-1)n^{-3}\Big\{\sum_{\ell,\ell'=1}^{k}w_\ell w_{\ell'}\sum_{s=0}^{\nu/2-1}(-1)^s(2^s s!)^{-1}\phi^{(2s+r)}_{\sigma_{\ell\ell'}(g)}(\mu_{\ell\ell'})\Big\}^2,$$

where $\mu_{\ell\ell'} = \mu_\ell - \mu_{\ell'}$ and $\sigma^2_{\ell\ell'} = \sigma^2_\ell + \sigma^2_{\ell'}$ for $\ell,\ell' = 1,2,\ldots,k$ and for any $\sigma>0$ we write $\sigma(g) = (\sigma^2+g^2)^{1/2}$.

For L and f as given in the previous theorem we can also write

$$g_0 = \Big|(2^{\nu-1}/\pi)^{1/2}(\nu/2)!\sum_{s=0}^{\nu/2-1}\mathrm{OF}(2s+r)(2^s s!)^{-1}\psi_{r+\nu}^{-1}n^{-1}\Big|^{1/(r+\nu+1)}$$

with

$$\psi_{r+\nu} = \sum_{\ell,\ell'=1}^{k} w_\ell w_{\ell'} \phi_{\sigma_{\ell\ell'}}^{(r+\nu)}(\mu_{\ell\ell'})$$

(see Section 6). The previous theorem allows us to compute $g_{\rm MSE}$ and $g_{\rm BA}$ numerically, and therefore to compare the exact errors of the three bandwidth sequences, $\mathrm{MSE}(g_{\rm MSE})$, $\mathrm{MSE}(g_{\rm BA})$ and $\mathrm{MSE}(g_0)$, in a non-asymptotic way. In the following we restrict our attention to the case $\nu = 2$. However, similar results were observed for all the considered higher kernel orders.

To analyze the finite sample behaviour of $g_{\rm MSE}$, $g_{\rm BA}$ and g_0 we use some of the 15 normal mixture densities introduced in Marron and Wand (1992). Precisely, we focus on their normal mixture densities #1, #7 and #12, corresponding to the cases where the difficulty in estimating the density itself is low, medium and high. In Figure 2 (left column) we show the relative efficiencies $[\mathrm{MSE}(g_{\rm MSE})/\mathrm{MSE}(g)]^{1/2}$ for $g = g_{\rm BA}$ (solid lines) and $g = g_0$ (dashed lines) against $\log_{10}(n)$ for $r = 0, 2, 4, 6$. As expected, for each of $g = g_{\rm BA}$ and $g = g_0$ the efficiency graphs are naturally placed in descending order as r increases, that is, for $g = g_{\rm BA}$ the top solid curve in each plot corresponds to $r = 0$ and the bottom solid curve corresponds to $r = 6$, and similarly for $g = g_0$. This reflects the fact that the approximations to $g_{\rm MSE}$ given by $g_{\rm BA}$ and g_0 get worse (in the MSE sense) as the degree of derivative r increases, as predicted by the asymptotic theory (see Theorems 5 and 6 above).

However, even though both $\mathrm{MSE}(g_{\rm BA})$ and $\mathrm{MSE}(g_0)$ exhibit the same relative order of convergence to $\mathrm{MSE}(g_{\rm MSE})$, we can see in the left column of Figure 2 that for small and moderate sample sizes there are striking differences between $g_{\rm BA}$ and g_0. Whereas for $n \geq 10$ and the cases $r = 0, 2, 4, 6$ represented in Figure 2 the efficiency of $g_{\rm BA}$ is always greater than 90%, showing that the loss in changing the goal from $g_{\rm MSE}$ to $g_{\rm BA}$ is nearly negligible, in some cases Figure 2 shows that the use of the bandwidth g_0 may lead to a very disappointing performance of the estimator.

The situation for low and medium density estimation difficulty (densities #1 and #7 above) is very similar: g_0 is even more efficient than $g_{\rm BA}$ for density #1 when $r = 0$, and it is also quite acceptable for $r = 2$, but for $r \geq 4$ the efficiency of g_0 decays rapidly, and it is already lower than 70% (for $r = 4$) or 50% (for $r = 6$) for sample size $n = 100$. This effect is even more dramatic for the case of a difficult-to-estimate density as #12: for sample size $n = 100$ the efficiency of g_0 is about 60% for $r = 0$ and it is lower than 10% for $r \geq 2$.

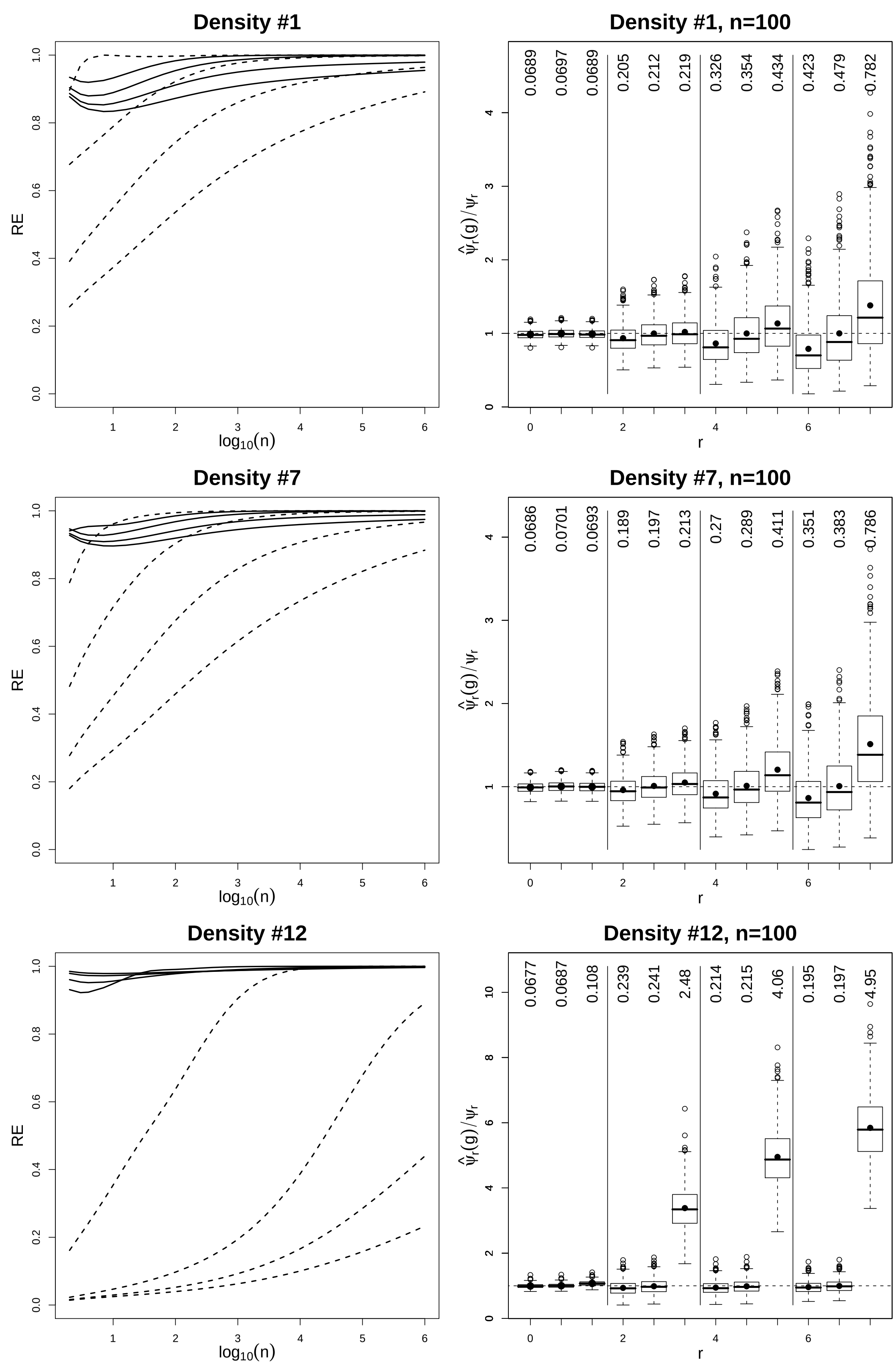

Figure 2. *Relative efficiencies (left) and distribution boxplots (right) for the estimator* $\hat{\psi}_r(g)$.

Our conclusion is that g_0 is indeed a bad surrogate for $g_{\rm MSE}$, especially for $r \geq 4$. This is quite a striking conclusion, since g_0 is the usual target bandwidth for plug-in bandwidth selection methods for the estimation of ψ_r.

In the right column of Figure 2 we show the boxplots for the distribution of $\hat{\psi}_r(g)/\psi_r$ based on 500 generated samples of size $n = 100$. In each graph we have vertical lines dividing the cases according to $r = 0, 2, 4, 6$ and, for each of these cases, we have three boxplots corresponding to the use of the theoretical $g = g_{\rm MSE}$, $g = g_{\rm BA}$ and $g = g_0$ in the estimator, from left to right. We have also added a solid circle to each boxplot indicating the sample mean of the distribution and a number on top with the square root of the sample MSE of $\hat{\psi}_r(g)/\psi_r$.

The boxplots show the reasons for the bad efficiency results of g_0. Although this bandwidth is meant to annihilate the asymptotically dominant bias term, it looks like g_0 does not get close to this goal for moderate sample sizes, since $\hat{\psi}_r(g_0)$ clearly overestimates ψ_r in mean, especially for $r \geq 4$. Moreover, this occasionally large bias does not come with a reduction in variance, since in fact $\hat{\psi}_r(g_0)$ is more variable than the other two estimators. Both effects (in bias and variance) are highly stressed for the case of density #12. In contrast, it is possible to observe how the estimator using the bandwidth $g_{\rm BA}$ is unbiased, as it should be by definition, at the expense of only a slightly increase of variance over $g_{\rm MSE}$. Nevertheless, the distributions of the estimator with $g_{\rm MSE}$ or $g_{\rm BA}$ are very similar.

In view of the results in this simulation study it is clear that bandwidth selectors oriented to $g_{\rm BA}$ should give raise to much better performance than the usual ones, which are designed to estimate g_0. This is a subject that we intend to study in detail in the future.

Remark 1. For the case $\nu = 2$ the exact MSE formula for normal mixture densities can be found in Aldershof's thesis (1991). However, we would like to highlight that although this result was known before, its consequences (as extracted from Figure 2) had not been fully explored yet.

6. Proofs

We start by presenting some properties of $R_{L,r,g}(f)$ and $S_{L,r,g}(f)$ as functions of g. Let us denote $\psi_{r,s} = \int f^{(r)} f^{(s)} f$.

Lemma 1. *Under assumptions (L1) and (D1), we have:*

a) *The function $g \mapsto R_{L,r,g}(f)$ is continuous and such that $\lim_{g\to 0} R_{L,r,g}(f) = \psi_r \int L$ and $\lim_{g\to\infty} g^{r+1} R_{L,r,g}(f) = L^{(r)}(0)$.*

b) *The function* $g \mapsto S_{L,r,g}(f)$ *is continuous and such that* $\lim_{g\to 0} S_{L,r,g}(f) = \psi_{r,r}(\int L)^2$ *and* $\lim_{g\to\infty} g^{2r+2} S_{L,r,g}(f) = L^{(r)}(0)^2$.

Proof. Using the fact that $L^{(j)}$ and $f^{(j)}$ are bounded and square integrable, for $j = 0, 1, \ldots, r$, and the same tools as in Hall and Marron (1987), it is straightforward to check that we can write

$$R_{L,r,g}(f) = \int L(u)\big(f^{(r)} * \bar{f}\big)(gu)du, \tag{9}$$

with $\bar{f}(x) = f(-x)$. Therefore, as $L \in L_1$ the continuity and the first limit in part *a*) follow from the Dominated Convergence Theorem (DCT) and the boundedness and continuity of the convolution product of square integrable functions, together with the fact that $(f^{(r)} * \bar{f})(0) = \psi_r$. For the second limit, using again the DCT, together with the boundedness and continuity of $L^{(r)}$ at zero, we obtain

$$\lim_{g\to\infty} g^{r+1} R_{L,r,g}(f) = \lim_{g\to\infty} \iint L^{(r)}(\tfrac{x-y}{g}) f(x) f(y) dx dy = L^{(r)}(0)$$

as stated.

The proof of part *b*) can be obtained in a similar way. For the first limit we start by writing

$$\begin{aligned} S_{L,r,g}(f) &= \iiint L(u)L(v) f^{(r)}(gu - x) f^{(r)}(gv - x)\bar{f}(x) dx du dv \qquad (10)\\ &= \iint L(u)L(v)\big(f^{(r)} \odot f^{(r)} \odot \bar{f}\big)(gu, gv) du dv, \end{aligned}$$

where we are denoting

$$(\alpha \odot \beta \odot \gamma)(y, z) = \int \alpha(y - x)\beta(z - x)\gamma(x) dx.$$

Reasoning as in the proof of Theorem 21.33 in Hewitt and Stromberg (1965), for $\alpha, \beta, \gamma \in L_3$ it can be shown that $\alpha \odot \beta \odot \gamma$ is a bounded continuous function. Consequently, as $f, f^{(r)} \in L_3$ (since they are bounded and square integrable), we get the stated limit by using again the DCT, together with the fact that $(f^{(r)} \odot f^{(r)} \odot \bar{f})(0, 0) = \psi_{r,r}$. The second limit again follows from a direct application of the DCT, since

$$\begin{aligned} \lim_{g\to\infty} g^{2r+2} S_{L,r,g}(f) &= \lim_{g\to\infty} \iiint L^{(r)}(\tfrac{x-y}{g}) L^{(r)}(\tfrac{x-z}{g}) f(x) f(y) f(z) dx dy dz \\ &= L^{(r)}(0)^2. \end{aligned}$$

□

Proof of Theorem 1. From the previous lemma, together with (3) and (4), we conclude that $B^2(g)$ and $V(g)$ are continuous functions such that

$$\lim_{g\to 0} B^2(g) = \infty, \quad \lim_{g\to\infty} B^2(g) = \psi_r^2, \quad \lim_{g\to 0} V(g) = \infty, \quad \lim_{g\to\infty} g^{2r+2}V(g) = 0.$$

So the MSE function, which equals $\mathrm{MSE}(g) = B^2(g)+V(g)$, is a continuous function such that

$$\lim_{g\to 0} \mathrm{MSE}(g) = \infty, \qquad \lim_{g\to\infty} \mathrm{MSE}(g) = \psi_r^2.$$

Therefore, to show that there exist a value $g_{\mathrm{MSE}} = g_{\mathrm{MSE},r,n}(f)$ minimizing the MSE function, it suffices to show that, for big enough g_*, we have $\mathrm{MSE}(g_*) < \psi_r^2$. So if we define

$$D(g) = \mathrm{MSE}(g) - \psi_r^2 = [B^2(g) - \psi_r^2] + V(g),$$

all that we need to show is that, for some $\rho > 0$, we have $\lim_{g\to\infty} g^\rho D(g) < 0$.

But using the previous lemma we have

$$\lim_{g\to\infty} g^{2r+2}[B^2(g) - \psi_r^2] = -\operatorname{sig}\big(\psi_r L^{(r)}(0)\big)\cdot\infty.$$

As our assumptions imply that $\operatorname{sig}\psi_r = \operatorname{sig} L^{(r)}(0) = (-1)^{r/2}$, it immediately follows that $\lim_{g\to\infty} g^{2r+2}[B^2(g) - \psi_r^2] = -\infty$. This, together with the limit properties of the variance allows us to conclude that $\lim_{g\to\infty} g^{2r+2}D(g) = -\infty$ and so the proof is complete. □

Proof of Theorem 2. From the previous lemma, together with (3), we know that $B(g)$ is a continuous function such that

$$\lim_{g\to 0} B(g) = (\operatorname{sig} L^{(r)}(0))\cdot\infty, \qquad \lim_{g\to\infty} B(g) = -\psi_r.$$

Again, our assumptions imply that $\operatorname{sig}\big(-\psi_r\cdot L^{(r)}(0)\big) = -1$, which yields the proof using Bolzano's theorem. □

Proof of Theorem 3. For $g = g_{\mathrm{MSE}}$ or $g = g_{\mathrm{BA}}$, suppose that ng^{r+1} does not converge to infinity. Then ng^{r+1} has a subsequence which is upper bounded by some positive constant C. Therefore, along that subsequence we have $g \to 0$.

For $g = g_{\mathrm{MSE}}$ this implies that

$$\begin{aligned}\limsup_{n\to\infty} \mathrm{MSE}(g_{\mathrm{MSE}}) &\geq \limsup_{n\to\infty} B^2(g_{\mathrm{MSE}}) = \limsup_{n\to\infty} (n^{-1}g_{\mathrm{MSE}}^{-r-1}L^{(r)}(0))^2 \\ &\geq (L^{(r)}(0)/C)^2 > 0,\end{aligned}$$

which contradicts the fact that $0 \leq \mathrm{MSE}(g_{\mathrm{MSE}}) \leq \mathrm{MSE}(n^{-1/(r+2)}) \to 0$ that follows from (5) together with the previous lemma.

Similarly, for $g = g_{\mathrm{BA}}$ we would obtain that

$$\begin{aligned}0 = B^2(g_{\mathrm{BA}}) &= \limsup_{n\to\infty} B^2(g_{\mathrm{BA}}) \\ &= \limsup_{n\to\infty} (n^{-1} g_{\mathrm{BA}}^{-r-1} L^{(r)}(0))^2 \geq (L^{(r)}(0)/C)^2 > 0,\end{aligned}$$

so that the result also follows by contradiction. □

Proof of Theorem 4. Let us prove the result for g_{MSE}. Denote by $\Lambda_{f,L}$ the set of accumulation points of the sequence (g_{MSE}). Take $0 < \lambda \in \Lambda_{f,L}$ and (g_{n_k}) a subsequence of (g_{MSE}) such that $\lambda = \lim_{k\to\infty} g_{n_k}$. Writing $B(g;n)$ and $\mathrm{MSE}(g;n)$ for $B(g)$ and $\mathrm{MSE}(g)$, respectively, from equalities (3) and (4) we get that, for fixed $g > 0$,

$$\lim_{n\to\infty} \mathrm{MSE}(g;n) = \lim_{n\to\infty} B^2(g;n) = [R_{L,r,g}(f) - \psi_r]^2,$$

so that using Lemma 1 and Theorem 3, we obtain

$$\begin{aligned}0 &= \lim_{g\to 0}[R_{L,r,g}(f) - \psi_r]^2 = \lim_{g\to 0}\lim_{k\to\infty} B^2(g;n_k) = \lim_{g\to 0}\lim_{k\to\infty} \mathrm{MSE}(g;n_k) \\ &\geq \lim_{k\to\infty} \mathrm{MSE}(g_{n_k};n_k) \geq \lim_{k\to\infty} B^2(g_{n_k};n_k) = [R_{L,r,\lambda}(f) - \psi_r]^2.\end{aligned}$$

Therefore

$$\begin{aligned}\Lambda_{f,L} &\subset \{\lambda \geq 0 : R_{L,r,\lambda}(f) = \psi_r\} \\ &= \{\lambda \geq 0 : \textstyle\int_0^\infty t^r |\varphi_f(t)|^2 [1 - \varphi_L(t\lambda)] dt = 0\},\end{aligned} \tag{11}$$

since from Parseval's formula, together with $\varphi_{f^{(r)}}(t) = (it)^r \varphi_f(t)$ (see Butzer and Nessel, 1971, Proposition 5.2.19) we easily get that

$$\psi_r = (-1)^{r/2}\pi^{-1} \int_0^\infty t^r |\varphi_f(t)|^2 dt$$

and

$$R_{L,r,\lambda}(f) = (-1)^{r/2}\pi^{-1} \int_0^\infty t^r |\varphi_f(t)|^2 \varphi_L(t\lambda) dt.$$

Additionally, we also have

$$\Lambda_{f,L} \subset \left[0, \min\left(\frac{S_L}{C_f}, \frac{T_L}{D_f}\right)\right]. \tag{12}$$

This is because in fact, if $\lambda > 0$ is such that $\lambda \in \Lambda_{f,L}$, from (11) we have

$$\int_0^{C_f} t^r |\varphi_f(t)|^2 [1 - \varphi_L(t\lambda)] dt = 0 \quad \text{and} \quad \int_{T_L/\lambda}^\infty t^r |\varphi_f(t)|^2 [1 - \varphi_L(t\lambda)] dt = 0.$$

Taking into account that φ_L is a real function (for L being symmetric) and such that $1 - \varphi_L(t\lambda) \geq 0$, from the first equality we conclude that $\varphi_L(s) = 1$ for all $0 \leq s \leq \lambda C_f$, and then $S_L \geq \lambda C_f$, that is, $\lambda \leq S_L/C_f$.

From the second equality we have $\varphi_f(t) = 0$ for all $t \geq T_L/\lambda$, and then $D_f \leq T_L/\lambda$, that is, $\lambda \leq T_L/D_f$.

From (12) we finally get

$$0 \leq \limsup_{n\to\infty} g_{\mathrm{MSE}} \leq \min\left(\frac{S_L}{C_f}, \frac{T_L}{D_f}\right), \tag{13}$$

which concludes the proof for g_{MSE}.

Similarly, notice that any λ being an accumulation point of g_{BA} the equality $R_{L,r,\lambda}(f) - \psi_r = 0$ should also hold, due to Theorem 3, the continuity properties in Lemma 1 and the fact that $B(g_{\mathrm{BA}}) = 0$. Consequently, (13) is also true for g_{BA} and so the desired result. □

As a tool for the proof of Theorem 5 we will need the following lemma, which follows directly from expressions (9) and (10) for $R_{L,r,g}(f)$ and $S_{L,r,g}(f)$, respectively, the differentiation theorem under the integral sign and standard Taylor expansions.

Lemma 2. *Under assumptions (L1), (L2), (D1) and (D2) we have:*

a) *The function $g \mapsto R_{L,r,g}(f)$ is differentiable with*

$$R_{L,r,g}(f) = \psi_r + g^\nu \psi_{r+\nu} m_\nu(L)/\nu! + o(g^\nu),$$
$$dR_{L,r,g}(f)/dg = g^{\nu-1}\psi_{r+\nu} m_\nu(L)/(\nu-1)! + o(g^{\nu-1}).$$

Additionally, if $|m_{\nu+2}|(L) < \infty$ and f has bounded continuous derivatives up to order $r+\nu+2$, the previous residual term $o(g^\nu)$ may be replaced by $g^{\nu+2}\psi_{r+\nu+2} m_{\nu+2}(L)/(\nu+2)! + o(g^{\nu+2})$.

b) *If $\int |u| L^{(r)}(u)^2 du < \infty$, the function $g \mapsto R_{(L^{(r)})^2,0,g}(f)$ is differentiable and such that $dR_{(L^{(r)})^2,0,g}(f)/dg = o(1)$.*

c) *The function $g \mapsto S_{L,r,g}(f)$ is differentiable and such that*

$$dS_{L,r,g}(f)/dg = 2g^{\nu-1}\psi_{r+\nu,r} m_\nu(L)/(\nu-1)! + o(g^{\nu-1}).$$

Proof of Theorem 5. a) From expansion (7) and taking for g the asymptotically optimal bandwidth (8), we easily get

$$n^{2\nu/(r+\nu+1)}\left(\mathrm{MSE}(g_0) - 4n^{-1}\mathrm{Var} f^{(r)}(X_1)\right) = o(1)$$

and then, as $\mathrm{MSE}(g_{\mathrm{MSE}}) \leq \mathrm{MSE}(g_0)$,

$$\limsup n^{2\nu/(r+\nu+1)}\left(\mathrm{MSE}(g_{\mathrm{MSE}}) - 4n^{-1}\mathrm{Var} f^{(r)}(X_1)\right) < \infty. \tag{14}$$

Moreover, using the fact that $g_{\mathrm{MSE}} \to 0$ (that follows from Theorem 4, due to condition (L2)), from expansion (7) we also get

$$
\begin{aligned}
&n^{2\nu/(r+\nu+1)} \Big(\mathrm{MSE}(g_{\mathrm{MSE}}) - 4n^{-1}\mathrm{Var} f^{(r)}(X_1)\Big) \\
&= \Big((n^{1/(r+\nu+1)} g_{\mathrm{MSE}})^{-r-1} L^{(r)}(0) + (n^{1/(r+\nu+1)} g_{\mathrm{MSE}})^{\nu} m_\nu(L)\psi_{r+\nu}/\nu!\Big)^2 \\
&+ o\Big((n^{1/(r+\nu+1)} g_{\mathrm{MSE}})^{-2r-2} + (n^{1/(r+\nu+1)} g_{\mathrm{MSE}})^{\nu-r-1} + (n^{1/(r+\nu+1)} g_{\mathrm{MSE}})^{2\nu}\Big),
\end{aligned}
$$

which contradicts (14) if $\liminf\ n^{1/(r+\nu+1)} g_{\mathrm{MSE}} = 0$ or $\limsup\ n^{1/(r+\nu+1)} g_{\mathrm{MSE}} = \infty$. Therefore the proof for g_{MSE} is complete. The proof for g_{BA} can be obtained in a similar way by noting that, using (6),

$$
\begin{aligned}
0 &= n^{\nu/(r+\nu+1)} B(g_{\mathrm{BA}}) \\
&= \{(n^{1/(r+\nu+1)} g_{\mathrm{BA}})^{-r-1} L^{(r)}(0) + (n^{1/(r+\nu+1)} g_{\mathrm{BA}})^{\nu} m_\nu(L)\psi_{r+\nu}/\nu!\}(1+o(1)).
\end{aligned}
$$

b) From Lemma 2 and equalities (3) and (4) the functions $B(g)$ and $V(g)$, and therefore $\mathrm{MSE}(g)$, are differentiable with

$$
B'(g) = -(r+1)n^{-1}g^{-r-2}L^{(r)}(0) + \nu g^{\nu-1}\psi_{r+\nu} m_\nu(L)/\nu! + o(g^{\nu-1}) \tag{15}
$$

and

$$
V'(g) = 2c_{1,r}\, n^{-1} g^{\nu-1} m_\nu(L)/\nu! - 2c_{2,r}\, n^{-2} g^{-2r-2} + o(n^{-1}g^{\nu-1} + n^{-2}g^{-2r-2}), \tag{16}
$$

with $c_{1,r} = 4\nu(\psi_{r+\nu,r} - \psi_{r+\nu}\psi_r)$ and $c_{2,r} = (2r+1)\psi_0 \int (L^{(r)})^2$.

From these expansions together with (6), part *a*) of this result and equation $\mathrm{MSE}'(g_{\mathrm{MSE}}) = 2B(g_{\mathrm{MSE}})B'(g_{\mathrm{MSE}}) + V'(g_{\mathrm{MSE}}) = 0$ we obtain

$$
\begin{aligned}
&n g_{\mathrm{MSE}}^{r+1} B(g_{\mathrm{MSE}})\, n g_{\mathrm{MSE}}^{r+2} B'(g_{\mathrm{MSE}}) \\
&\qquad = -n^2 g_{\mathrm{MSE}}^{2r+3} V'(g_{\mathrm{MSE}})/2 \\
&\qquad = -c_{1,r}\, n g_{\mathrm{MSE}}^{2r+\nu+2} m_\nu(L)/\nu! + c_{2,r}\, g_{\mathrm{MSE}} + o(n g_{\mathrm{MSE}}^{2r+\nu+2} + g_{\mathrm{MSE}})
\end{aligned} \tag{17}
$$

where

$$
n g_{\mathrm{MSE}}^{r+1} B(g_{\mathrm{MSE}}) = L^{(r)}(0) + n g_{\mathrm{MSE}}^{r+\nu+1} \psi_{r+\nu} m_\nu(L)/\nu! + o(1) \tag{18}
$$

and

$$
\begin{aligned}
&n g_{\mathrm{MSE}}^{r+2} B'(g_{\mathrm{MSE}}) \\
&= -(r+1)L^{(r)}(0) + \nu\, n g_{\mathrm{MSE}}^{r+\nu+1} \psi_{r+\nu} m_\nu(L)/\nu! + o(1) \\
&= (-1)^{r/2+1}\{(r+1)|L^{(r)}(0)| + \nu\, n g_{\mathrm{MSE}}^{r+\nu+1} |\psi_{r+\nu}||m_\nu(L)|/\nu!\} + o(1).
\end{aligned}
$$

is such that $\liminf n g_{\mathrm{MSE}}^{r+2} |B'(g_{\mathrm{MSE}})| > 0$. Therefore, from (17) we finally get

$$
L^{(r)}(0) + n g_{\mathrm{MSE}}^{r+\nu+1} \psi_{r+\nu} m_\nu(L)/\nu! = o(1), \tag{19}
$$

that concludes the proof for g_{MSE}. Also, notice that from $0 = ng_{\mathrm{BA}}^{r+1}B(g_{\mathrm{BA}})$ and (6) we obtain the same formula as in (19) with g_{BA} instead of g_{MSE} and thus the limit $g_0/g_{\mathrm{BA}} \to 1$ and, consequently, $g_{\mathrm{BA}}/g_{\mathrm{MSE}} \to 1$.

c) Using the fact that f has a bounded derivative of order $r+\nu+2$, from Lemma 2 we know that the residual term $o(g^\nu)$ appearing in the expansion of $B(g)$ can be replaced by $O(g^{\nu+2})$. This enables us to improve the order of convergence of the residual term in equation (18) which can be replaced by $O(g_{\mathrm{MSE}}^2) = o(g_{\mathrm{MSE}})$. Using again equation (17) and the fact that $ng_{\mathrm{MSE}}^{r+2}B'(g_{\mathrm{MSE}}) = -c_{3,r}(1+o(1))$, where $c_{3,r} = (r+\nu+1)L^{(r)}(0)$, we get

$$\begin{aligned} &L^{(r)}(0) + ng_{\mathrm{MSE}}^{r+\nu+1}\psi_{r+\nu}m_\nu(L)/\nu! \\ &\quad = \ c_{1,r}c_{3,r}^{-1}\,ng_{\mathrm{MSE}}^{2r+\nu+2}m_\nu(L)/\nu! - c_{2,r}c_{3,r}^{-1}\,g_{\mathrm{MSE}} + o(g_{\mathrm{MSE}}). \end{aligned}$$

Taking into account that g_0 satisfies the equality

$$L^{(r)}(0) + ng_0^{r+\nu+1}\psi_{r+\nu}m_\nu(L)/\nu! = 0, \tag{20}$$

for some $\bar{g}$ between g_0 and g_{MSE} we have

$$\begin{aligned} &n(r+\nu+1)\bar{g}^{r+\nu}(g_0/g_{\mathrm{MSE}} - 1)\psi_{r+\nu}m_\nu(L)/\nu! \\ &\quad = \ -c_{1,r}c_{3,r}^{-1}\,ng_{\mathrm{MSE}}^{2r+\nu+1}m_\nu(L)/\nu! + c_{2,r}c_{3,r}^{-1} + o(1). \end{aligned}$$

In order to conclude it suffices to remark that $n^{1/(r+\nu+1)}\bar{g} = c_{0,r}(1+o(1))$ where $c_{0,r}^{r+\nu+1} = -\nu!L^{(r)}(0)/(m_\nu(L)\psi_{r+\nu})$. Therefore, the announced convergence for $g_0/g_{\mathrm{MSE}} - 1$ takes place with

$$C = C_{L,r,\nu}(f) = -c_{0,r}c_{3,r}^{-2}\{c_{2,r} + 4\nu(\psi_{\nu,0}\psi_\nu^{-1} - \psi_0)L(0)\delta_{r0}\},$$

where δ_{r0} is the Kronecker symbol, that is, $\delta_{r0} = 1$ for $r=0$ and $\delta_{r0} = 0$ for $r \neq 0$.

On the other hand, starting from $0 = ng_{\mathrm{BA}}^{r+1}B(g_{\mathrm{BA}})$ and using (6) with the residual term $o(g^\nu)$ replaced by $O(g^{\nu+2})$ we come to

$$L^{(r)}(0) + ng_{\mathrm{BA}}^{r+\nu+1}\psi_{r+\nu}m_\nu(L)/\nu! = g_{\mathrm{BA}}^{r+1}\psi_r + O(ng_{\mathrm{BA}}^{r+\nu+3}). \tag{21}$$

Reasoning as before we conclude that the announced convergence for $g_{\mathrm{BA}}/g_{\mathrm{MSE}} - 1$ takes place with

$$D = D_{L,r,\nu}(f) = -c_{0,r}c_{3,r}^{-2}\{c_{2,r} + (4\nu(\psi_{\nu,0}\psi_\nu^{-1} - \psi_0) + (\nu+1)\psi_0)L(0)\delta_{r0}\}.$$

Finally, using the fact that f has a bounded continuous derivative of order $r+\nu+2$, from Lemma 2 we know that the residual term $o(g^\nu)$ appearing in the expansion of $B(g)$ can be replaced by $g^{\nu+2}\psi_{r+\nu+2}m_{\nu+2}(L)/(\nu+2)! + o(g^{\nu+2})$ which enables us write the residual term in equation (21) more precisely. Together with equation (20) we conclude that the announced convergence for $g_0/g_{\rm BA}-1$ takes place with

$$E = E_{L,r,\nu}(f) = -c_{0,r}c_{3,r}^{-1}\{c_{0,r}^{r+\nu+2}\psi_{r+\nu+2}m_{\nu+2}(L)/(\nu+2)!(1-\delta_{r0})-\psi_0\delta_{r0}\}.$$

□

The orders of convergence for the higher order derivatives of $R_{L,r,g}(f)$, $R_{(L^{(r)})^2,0,g}(f)$ and $S_{L,r,g}(f)$ given in the next lemma will be used in the proof of Theorem 6. They follow directly from expressions (9) and (10), the differentiation theorem under the integral sign and standard Taylor expansions.

Lemma 3. *Under assumptions (L1), (L2) and (D1), if f has bounded and continuous derivatives up to order $r+\nu+2$, $|m_{\nu+2}|(L)<\infty$ and $\int |u|^3 L^{(r)}(u)^2du<\infty$, then the functions $g\mapsto R_{L,r,g}(f)$, $g\mapsto R_{(L^{(r)})^2,0,g}(f)$ and $g\mapsto S_{L,r,g}(f)$ are three-times differentiable with*

$$\begin{aligned} d^2R_{L,r,g}(f)/dg^2 &= O(g^{\nu-2}), \quad d^3R_{L,r,g}(f)/dg^3 = O(g^{\nu-3}),\\ d^2R_{(L^{(r)})^2,0,g}(f)/dg^2 &= O(1), \quad d^3R_{(L^{(r)})^2,0,g}(f)/dg^3 = O(1),\\ d^2S_{L,r,g}(f)/dg^2 &= O(g^{\nu-2}), \quad d^3S_{L,r,g}(f)/dg^3 = O(g^{\nu-3}). \end{aligned}$$

Proof of Theorem 6. From Lemmas 2 and 3 and equalities (3) and (4) the functions $B(g)$ and $V(g)$, and therefore $\mathrm{MSE}(g)$, are three-times differentiable with

$$B''(g) = O(n^{-1}g^{-r-3}+g^{\nu-2}), \quad B'''(g) = O(n^{-1}g^{-r-4}+g^{\nu-3})$$

and

$$V''(g) = O(n^{-2}g^{-2r-3}+n^{-1}g^{\nu-2}), \quad V'''(g) = O(n^{-2}g^{-2r-4}+n^{-1}g^{\nu-3}).$$

Moreover, a Taylor expansion for $g\mapsto\mathrm{MSE}(g)$ around $g=g_{\rm BA}$ leads to

$$\begin{aligned} \mathrm{MSE}(g_0)-\mathrm{MSE}(g_{\rm BA}) &= \mathrm{MSE}'(g_{\rm BA})g_{\rm BA}(g_0/g_{\rm BA}-1)\\ &+ \mathrm{MSE}''(g_{\rm BA})g_{\rm BA}^2(g_0/g_{\rm BA}-1)^2/2+\mathrm{MSE}'''(\tilde g)g_{\rm BA}^3(g_0/g_{\rm BA}-1)^3/3!, \end{aligned}$$

for some $\tilde g$ between g_0 and $g_{\rm BA}$. Taking into account that $B(g_{\rm BA})=0$ and $n^{1/(r+\nu+1)}g_{\rm BA}=c_{0,r}(1+o(1))$, from the previous orders of convergence for $B''(g)$, $B'''(g)$, $V''(g)$ and $V'''(g)$, the expansions (15) and (16) for $B'(g)$ and $V'(g)$, respectively, and Theorem 5.c), we get

$$\mathrm{MSE}'(g_{\rm BA})g_{\rm BA} = c_{0,r}^{-2r-1}d_r\,n^{-(2\nu+1)/(r+\nu+1)}(1+o(1)),$$

$$\mathrm{MSE}''(g_{\rm BA})g_{\rm BA}^2(g_0/g_{\rm BA}-1) = 2c_{0,r}^{-2r-2}c_{3,r}^2E\,n^{-\min(r+2\nu+1,2\nu+2)/(r+\nu+1)}(1+o(1))$$

and

$$\mathrm{MSE}'''(\tilde g)g_{\rm BA}^3(g_0/g_{\rm BA}-1)^2 = O(n^{-(2\nu+2)/(r+\nu+1)}),$$

where $d_r = -2\{c_{2,r} + 4\nu(\psi_{\nu,0}\psi_\nu^{-1} - \psi_0)L(0)\delta_{r0}\}$ and the constants $c_{0,r}, c_{2,r}$ and $c_{3,r}$ are defined in the proof of Theorem 5. Therefore, from Theorem 5.c), the announced convergence for $\mathrm{MSE}(g_0) - \mathrm{MSE}(g_{\mathrm{BA}})$ will take place with $\Lambda = \Lambda_{L,r,\nu}(f) = c_{0,r}^{-2r-2}\{c_{0,r}d_r + c_{3,r}^2 E\delta_{r0}\}$. □

Proof of Theorem 7. As noted previously, to obtain an explicit formula for the MSE function we just need to provide explicit formulas for $L^{(r)}(0)$, $R_{L,r,g}(f)$, ψ_r, $S_{L,r,g}(f)$ and $R_{(L^{(r)})^2,0,g}(f)$. From Fact C.1.6 in Appendix C in Wand and Jones (1995) we already have

$$\begin{aligned} L^{(r)}(0) &= \sum_{s=0}^{\nu/2-1} (-1)^s(2^s s!)^{-1}\,(-1)^{(2s+r)/2}(2\pi)^{-1/2}\mathrm{OF}(2s+r) \\ &= (-1)^{r/2}(2\pi)^{-1/2}\sum_{s=0}^{\nu/2-1}(2^s s!)^{-1}\,\mathrm{OF}(2s+r). \end{aligned}$$

Also, using Fact C.1.12 there, taking into account that r is even,

$$\begin{aligned} \psi_r &= \int f^{(r)}(x)f(x)dx = \sum_{\ell,\ell'=1}^{k} w_\ell w_{\ell'} \int \phi_{\sigma_\ell}^{(r)}(x-\mu_\ell)\phi_{\sigma'_\ell}(x-\mu_{\ell'})dx \\ &= \sum_{\ell,\ell'=1}^{k} w_\ell w_{\ell'} \phi_{\sigma_{\ell\ell'}}^{(r)}(\mu_{\ell\ell'}). \end{aligned}$$

And from the same result and Fact C.1.11 we have

$$\begin{aligned} R_{L,r,g}(f) &= \int (L_g^{(r)} * f)f \\ &= \sum_{\ell,\ell'=1}^{k} w_\ell w_{\ell'} \sum_{s=0}^{\nu/2-1}(-1)^s(2^s s!)^{-1}\int \big[\phi_g^{(2s+r)} * \phi_{\sigma_\ell}(\cdot-\mu_\ell)\big]\phi_{\sigma_{\ell'}}(\cdot-\mu_{\ell'}) \\ &= \sum_{\ell,\ell'=1}^{k} w_\ell w_{\ell'} \sum_{s=0}^{\nu/2-1}(-1)^s(2^s s!)^{-1}\int \phi_{\sigma_\ell(g)}^{(2s+r)}(x-\mu_\ell)\phi_{\sigma_{\ell'}}(x-\mu_{\ell'})dx \\ &= \sum_{\ell,\ell'=1}^{k} w_\ell w_{\ell'} \sum_{s=0}^{\nu/2-1}(-1)^s(2^s s!)^{-1}\,\phi_{\sigma_{\ell\ell'}(g)}^{(2s+r)}(\mu_{\ell\ell'}) \end{aligned}$$

and so, the formula for the bias is complete.

On the other hand, Theorem 6.1 in Aldershof *et al.* (1995) with $m = 3$ and $r_3 = 0$ states that

$$\int \phi_{\sigma_1}^{(r_1)}(x-\mu_1)\phi_{\sigma_2}^{(r_2)}(x-\mu_2)\phi_{\sigma_3}(x-\mu_3)dx = I_{r_1,r_2}(\boldsymbol{\mu};\boldsymbol{\sigma}). \tag{22}$$

But we have $L_g^{(r)} * f = \sum_{\ell=1}^k w_\ell \sum_{s=0}^{\nu/2-1}(-1)^s(2^s s!)^{-1}\,\phi_{\sigma_\ell(g)}^{(2s+r)}(\cdot-\mu_\ell)$ so that from (22) we obtain

$$\begin{aligned}
S_{L,r,g}(f) &= \int (L_g^{(r)} * f)^2 f \\
&= \sum_{\ell_1,\ell_2,\ell_3=1}^{k} w_{\ell_1}w_{\ell_2}w_{\ell_3}\sum_{s,s'=0}^{\nu/2-1}(-1)^{s+s'}(2^{s+s'}s!s'!)^{-1} \\
&\qquad \times \int \phi_{\sigma_{\ell_1}(g)}^{(2s+r)}(x-\mu_{\ell_1})\phi_{\sigma_{\ell_2}(g)}^{(2s'+r)}(x-\mu_{\ell_2})\phi_{\sigma_{\ell_3}}(x-\mu_{\ell_3})dx \\
&= \sum_{\ell_1,\ell_2,\ell_3=1}^{k} w_{\ell_1}w_{\ell_2}w_{\ell_3}\sum_{s,s'=0}^{\nu/2-1}(-1)^{s+s'}(2^{s+s'}s!s'!)^{-1} \\
&\qquad \times I_{2s+r,2s'+r}(\mu_{\ell_1},\mu_{\ell_2},\mu_{\ell_3};\sigma_{\ell_1}(g),\sigma_{\ell_2}(g),\sigma_{\ell_3}).
\end{aligned}$$

For the remaining term, it is easy to check that

$$f * \bar{f}(z) = \sum_{\ell,\ell'=1}^{k} w_\ell w_{\ell'}\phi_{\sigma_{\ell\ell'}}(z-\mu_{\ell\ell'}).$$

Also, we know that

$$\begin{aligned}
g^{-2r-1}R_{(L^{(r)})^2,0,g}(f) &= \iint (L_g^{(r)})^2(x-y)f(x)f(y)dxdy \\
&= \int (L_g^{(r)})^2(z)(f*\bar{f})(z)dz
\end{aligned}$$

so that in the normal mixture case we have

$$\begin{aligned}
&g^{-2r-1}R_{(L^{(r)})^2,0,g}(f) \\
&= \sum_{\ell,\ell'=1}^{k} w_\ell w_{\ell'}\sum_{s,s'=0}^{\nu/2-1}(-1)^{s+s'}(2^{s+s'}s!s'!)^{-1}\int \phi_g^{(2s+r)}(z)\phi_g^{(2s'+r)}(z)\phi_{\sigma_{\ell\ell'}}(z-\mu_{\ell\ell'})dz
\end{aligned}$$

so we conclude again by using formula (22) with $\mu_1 = \mu_2 = 0$, $\mu_3 = \mu_{\ell\ell'}$ and $\sigma_1 = \sigma_2 = g$, $\sigma_3 = \sigma_{\ell\ell'}$. □

References

Aldershof, B. (1991) *Estimation of Integrated Squared Density Derivatives.* Ph.D. thesis, University of North Carolina, Chapel Hill.

Bhattacharya, G. K. and Roussas, G. G. (1969) Estimation of a certain functional of a probability density function. *Skand. Aktuarietidskr*, **1969**, 201–206.

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

Butzer, P.L. and Nessel, R.J. (1971) *Fourier Analysis and Approximation, Vol. 1.* Birkhäuser Verlag, Basel.

Chacón, J.E., Montanero, J. and Nogales, A.G. (2007) A note on kernel density estimation at a parametric rate. *J. Nonparametr. Stat.*, **19**, 13–21.

Chacón, J.E., Montanero, J., Nogales, A.G. and Pérez, P. (2007) On the existence and limit behavior of the optimal bandwidth for kernel density estimation. *Statist. Sinica*, **17**, 289–300.

Chacón, J.E. and Tenreiro, C. (2008) A data-based method for choosing the number of pilot stages for plug-in bandwidth selection. *Submitted.*

Dmitriev, Y.G. and Tarasenko, F.P. (1973) On the estimation of functionals of the probability density and its derivatives. *Theory Probab. Appl.*, **18**, 628–633.

Dmitriev, Y.G. and Tarasenko, F.P. (1975) On a class of non-parametric estimates of non-linear functionals of density. *Theory Probab. Appl.*, **19**, 390–394.

Giné, E. and Mason D.M. (2008) Uniform in bandwidth estimation of integral functionals of the density function. *Scand. J. Statist.*, **35**, 739–761.

Giné, E. and Nickl, R. (2008) A simple adaptive estimator of the integrated square of a density. *Bernoulli*, **14**, 47–61.

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

Hettmansperger, T.P. (1984) *Statistical Inference Based on Ranks.* Wiley, New York.

Hewitt, E. and Stromberg, K. (1965) *Real and Abstract Analysis.* Springer-Verlag, Berlin.

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.

Laurent, B. (1997) Estimation of integral functionals of a density and its derivatives. *Bernoulli*, **3**, 181–211.

Levit, B. Y. (1978) Asymptotically efficient estimation of nonlinear functionals. *Problems Inform. Transmission*, **14**, 65–72.

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

Prakasa Rao, B.L.S. (1983) *Nonparametric Functional Estimation.* Academic Press, New York.

Prakasa Rao, B.L.S. (1999) Estimation of the integrated squared density derivatives by wavelets. *Bull. Inform. Cybernet.*, **31**, 47–65.

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

Schuster, E.F. (1974) On the rate of convergence of an estimate of a functional of a probability density. *Scand. Actuar. J.*, **1974**, 103–107.

Sheather, S.J., Hettmansperger, T.P. and Donald, M.R. (1994) Data-based bandwidth selection for kernel estimators of the integral of $f^2(x)$. *Scand. J. Statist.*, **21**, 265–275.

Tenreiro, C. (2003) On the asymptotic normality of multistage integrated density derivatives kernel estimators. *Statist. Probab. Lett.*, **64**, 311–322.

van Es, B. (1992) Estimating functionals related to a density by a class of statistics based on spacings. *Scand. J. Statist.*, **19**, 61–72.

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.

Wu, T.-J. (1995) Adaptive root n estimates of integrated squared density derivatives. *Ann. Statist.*, **23**, 1474–1495.

José E. Chacón
Departamento de Matemáticas, Universidad de Extremadura, Spain
E-mail address: jechacon@unex.es
URL: http://kolmogorov.unex.es/~jechacon/

Carlos Tenreiro
CMUC, Department of Mathematics, University of Coimbra, 3001–454 Coimbra, Portugal
E-mail address: tenreiro@mat.uc.pt
URL: http://www.mat.uc.pt/~tenreiro/