

# On the asymptotic behaviour of the integrated square error of kernel density estimators with data-dependent bandwidth<sup>☆</sup>

Carlos Tenreiro\*

*Departamento de Matemática, Universidade de Coimbra, Apartado 3008, 3000 Coimbra, Portugal*

Received January 2000; received in revised form December 2000

## Abstract

In this paper, we consider the integrated square error  $J_n = \int \{\hat{f}_n(x) - f(x)\}^2 dx$ , where  $f$  is the common density function of the independent and identically distributed random vectors  $X_1, \dots, X_n$  and  $\hat{f}_n$  is the kernel estimator with a data-dependent bandwidth. Using the approach introduced by Hall (J. Multivariate Anal. 14 (1984) 1), and under some regularity conditions, we derive the  $L_2$  consistency in probability of  $\hat{f}_n$  and we establish an asymptotic expansion in probability and a central limit theorem for  $J_n$ . © 2001 Elsevier Science B.V. All rights reserved

*MSC:* 62G07; 60F05

*Keywords:* Kernel estimators; Integrated square error; Asymptotic distribution; U-statistics

## 1. Introduction

Let  $X_1, \dots, X_n$  be independent  $\mathbb{R}^d$ -valued random vectors, with common density function  $f$ , and let  $f_n$  be the kernel estimator of  $f$  given, for  $x \in \mathbb{R}^d$ , by (cf. Rosenblatt, 1956 and Parzen, 1962)

$$f_n(x) = \frac{1}{n} \sum_{i=1}^n K_{h_n}(x - X_i),$$

where  $K_h(\cdot) = K(\cdot/h)/h^d$ , for  $h > 0$ ,  $K$  is a kernel on  $\mathbb{R}^d$ , i.e., an integrable function such that  $\int_{\mathbb{R}^d} K(u) du = 1$ , and  $(h_n)$  is a sequence of strictly positive real numbers converging to zero as  $n \rightarrow +\infty$ .

In the definition of  $f_n$ , the kernel  $K$  and the bandwidth  $h_n$  enter as unspecified parameters. For a fixed kernel (it is well known that the performance of  $f_n$  is not very sensitive to the choice of the kernel), the bandwidth is usually chosen on the basis of the data. The application of such data-driven procedures leads to

<sup>☆</sup> This work was partially supported by CMUC/FCT and by Projecto Praxis/2/2.1/Mat/400/94.

\* Fax: +351-239-832-568.

*E-mail address:* tenreiro@mat.uc.pt (C. Tenreiro).

an estimator of the form

$$\hat{f}_n(x) = \frac{1}{n} \sum_{i=1}^n K_{A_n}(x - X_i), \quad (1)$$

where  $A_n = A_n(X_1, \dots, X_n)$  is a sequence of measurable bandwidths.

In this paper we consider the integrated square error

$$J_n = \int \{\hat{f}_n(x) - f(x)\}^2 dx, \quad (2)$$

where the integral is over  $\mathbb{R}^d$ , as a measure of the global performance of  $\hat{f}_n$ . Under some regularity conditions on  $f$  and on  $K$ , the usually considered bandwidth selectors are such that  $A_n/h_n - 1 = o_p(1)$  for some deterministic sequence  $(h_n)$ . Assuming this property, we establish an asymptotic expansion in probability for  $J_n$  in terms of the integrated square error:

$$I_n = \int \{f_n(x) - f(x)\}^2 dx, \quad (3)$$

studied by Hall (1984) using  $U$ -statistics techniques (see also Bickel and Rosenblatt, 1973). A such expansion enables us to derive the asymptotic behaviour of  $J_n$  in terms of the corresponding one of  $I_n$ . In particular, the  $L_2$  consistency in probability of  $\hat{f}_n$  is derived, and if an appropriate rate of convergence for  $A_n/h_n - 1$  is available, an asymptotic expansion in probability for  $J_n$ , similar to that given for  $I_n$  by Hall (1984), is obtained. Moreover, if  $A_n/h_n - 1 = O_p(n^{-\alpha})$  for some  $0 < \alpha \leq \frac{1}{2}$ , where  $\alpha$  is close to  $\frac{1}{2}$  (but not necessarily too close), we give a central limit theorem for  $J_n$  and we conclude that its asymptotic distribution depends on  $(A_n)$  only through the deterministic sequence  $(h_n)$ . Bandwidth selectors verifying this condition for the same sequence  $(h_n)$  are then indistinguishable from an integrated square error point of view.

These results, which generalize those obtained by Hall (1984), are presented in Section 3. Some notation and general assumptions on the underlying density function, on the kernel and on the measurable sequence of bandwidths are introduced in Section 2.

A similar study is considered in Liero (1992) for the kernel regression estimator. However, our approach is different from the one adopted by Liero (1992) since, following the methods developed by Bickel and Rosenblatt (1973), this author uses strong approximation principles for empirical processes.

The proofs of all results are given in Section 4.

## 2. Some notation and general assumptions

The goal of this paper is to describe the asymptotic behaviour of the integrated square error  $J_n$  defined by (2). Our approach is based on the following equalities which give us alternative expressions for the kernel estimators  $f_n$  and  $\hat{f}_n$ :

$$f_n(x) = \frac{1}{nh_n^d} \sum_{i=1}^n W\left(\frac{x - X_i}{h_n}, 1\right) \quad (4)$$

and

$$\hat{f}_n(x) = \frac{1}{nh_n^d} \sum_{i=1}^n W\left(\frac{x - X_i}{h_n}, \frac{A_n}{h_n}\right), \quad (5)$$

for  $x \in \mathbb{R}^d$ , where  $W(u, h) = K_h(u)$  for  $u \in \mathbb{R}^d$  and  $h > 0$ .

Under appropriated conditions on the kernel  $K$  and on  $(A_n)$ , we shall show that  $\hat{f}_n$  can be conveniently approximated by  $f_n$  and that the same occurs for  $J_n$  and  $I_n$  given by (3). Therefore, the asymptotic behaviour of  $J_n$  can be studied from the corresponding one of  $I_n$ .

For  $m \in \mathbb{N}_0$ , let us introduce the set  $\mathcal{K}_b(m)$  of bounded kernels  $K$  on  $\mathbb{R}^d$  of order  $m$ , i.e., such that  $\int \|u\|^m |K(u)| du < \infty$  and  $\int u_1^{a_1} \dots u_d^{a_d} K(u) du = 0$ , whenever  $0 < \sum_{i=1}^d a_i < m$ , with  $a_1, \dots, a_d \in \mathbb{N}_0$  and  $u = (u_1, \dots, u_d)$ , where  $\|\cdot\|$  is the euclidian norm in  $\mathbb{R}^d$ . For  $v \in \mathbb{N}$ , let us denote by  $\mathcal{K}_b^v(m)$  the subset of  $\mathcal{K}_b(m)$  of  $v$ -times continuously differentiable kernels  $K$  for which there exists  $\varepsilon \in ]0, 1[$  such that the functions defined by  $u \rightarrow \|u\|^m \sup_{|h-1| \leq \varepsilon} |\partial^\ell W / \partial h^\ell(u, h)|$ , for  $\ell = 0, 1, \dots, v$ , are bounded and integrable on  $\mathbb{R}^d$ .

The standard normal kernel  $K(u) = (2\pi)^{-d/2} \exp(-\|u\|^2/2)$ ,  $u \in \mathbb{R}^d$ , belongs to  $\mathcal{K}_b^v(2)$  for all  $v \in \mathbb{N}$ , and every  $v$ -times continuously differentiable kernel in  $\mathcal{K}_b(m)$  with compact support belongs to  $\mathcal{K}_b^v(m)$ .

Consider the following assumptions on  $A_n$  denoted by (B).

*Assumptions on  $A_n$  (B)*

There exists a deterministic sequence  $(h_n)$  of strictly positive real numbers satisfying  $h_n \rightarrow 0$  and  $nh_n^d \rightarrow +\infty$ , when  $n \rightarrow +\infty$ , such that

$$\xi_n := \frac{A_n}{h_n} - 1 = o_p(1).$$

If  $A_n$  is chosen such that  $A_n = B_n(X_1, \dots, X_n) a_n$ , with  $(a_n)$  a deterministic sequence of strictly positive real numbers satisfying  $a_n \rightarrow 0$  and  $na_n^d \rightarrow +\infty$ , when  $n \rightarrow +\infty$ , and  $(B_n)$  a sequence of strictly positive measurable functions satisfying  $B_n = B_f + O_p(1/\sqrt{n})$ , for some  $B_f \in ]0, +\infty[$ , the previous conditions are fulfilled with  $h_n = B_f a_n$  and  $\xi_n = O_p(n^{-1/2})$ . This is, for example the case of the bandwidth selection methods using reference distributions or the maximal smoothing principle whenever  $f$  has finite fourth order moments (e.g. Terrel, 1990). A choice of the previous form for  $A_n$  was also considered by Fan (1994) in the context of goodness-of-fit tests based on the kernel density estimator to obtain a invariant test statistic under the null hypothesis of normality. In a real context, with the important previous exceptions, the assumption (B) is satisfied for most bandwidth selectors with  $h_n = h_{\text{MISE}}$  the minimizer of the mean integrated square error (under some regularity conditions on the underlying density function  $f$  and on the kernel  $K$ ). The distinction between bandwidth selectors is usually based on the rate of convergence to 0 of the random sequence  $\xi_n$ :  $\xi_n = O_p(n^{-1/10})$  for the least squares cross-validation and biased cross-validation methods (see Scott and Terrel, 1987; Hall and Marron, 1987);  $\xi_n = O_p(n^{-4/13})$  for the plug-in method of Park and Marron (1990);  $\xi_n = O_p(n^{-5/14})$  for the plug-in method of Sheather and Jones (1991); and  $\xi_n = O_p(n^{-1/2})$  for the plug-in method of Hall et al. (1991), and for the smoothed cross-validation methods of Hall et al. (1992) and Jones et al. (1991). Remark that the previous rates of convergence are not directly comparable since the conditions imposed to  $f$  in each result are not coincident. Hall and Marron (1991) have shown that  $O_p(n^{-1/2})$  is the best achievable rate, i.e., it is impossible to use a data-dependent bandwidth which is closer to  $h_{\text{MISE}}$  than  $n^{-1/2}$  in the previous sense. See also Jones et al. (1996) and Loader (1999) for a survey and some interesting comments on bandwidth selection.

### 3. Asymptotic behaviour of $J_n$

Let us denote by  $\mathcal{W}_b(m)$  the set of bounded probability density functions on  $\mathbb{R}^d$  with bounded and continuous partial derivatives up to order  $m$ . For  $f \in \mathcal{W}_b(m)$  and  $\Psi$  a real function on  $\mathbb{R}^d$  such that  $\int \|u\|^m |\Psi(u)| du < \infty$ , let us denote by  $A_\Psi^m f$  the function defined, for  $x \in \mathbb{R}^d$ , by

$$A_\Psi^m f(x) = \frac{(-1)^m}{m!} \sum_{i_1, \dots, i_m=1}^d \int u_{i_1} \dots u_{i_m} \Psi(u) du \frac{\partial^m f}{\partial x_{i_1} \dots \partial x_{i_m}}(x).$$

We present in the following lemma, an asymptotic expansion in probability for the integrated square error  $J_n$  in terms of the integrated square error  $I_n$ . It is proven in Section 4.

**Lemma 1.** *Let us assume that  $f \in \mathcal{W}_b(m)$  and  $K \in \mathcal{K}_b^v(m)$  for some  $m$  and  $v$  in  $\mathbb{N}$ . Under the assumption (B), if the  $m$ -order partial derivatives of  $f$  are square integrable, we have*

$$J_n = I_n + 2(1 - \delta_{1v})\xi_n \left( \frac{1}{nh_n^d} \int K(u)K^{\hat{\circ}}(u) du + h_n^{2m} \int \Delta_K^m f(x)\Delta_{K^{\hat{\circ}}}^m f(x) dx \right) + O_p \left( \xi_n \left( \frac{1}{nh_n^{d/2}} + \frac{1}{\sqrt{nh_n^{-m}}} \right) + \xi_n^v \right) + o_p \left( \xi_n \left( \frac{1}{nh_n^d} + h_n^{2m} \right) \right),$$

where  $\delta_{ij}$  is the kronecker delta and for  $u = (u_1, \dots, u_d) \in \mathbb{R}^d$ ,

$$K^{\hat{\circ}}(u) = -dK(u) - \sum_{i=1}^d u_i \frac{\partial K}{\partial u_i}(u).$$

It is clear now that the asymptotic behaviour of  $J_n$  follows from the corresponding one of  $I_n$  and  $\xi_n$ . If  $f \in \mathcal{W}_b(m)$  and  $K \in \mathcal{K}_b(m)$  for some  $m \in \mathbb{N}$ , and if there exists  $\lambda \in [0, +\infty]$  such that  $\lim_{n \rightarrow +\infty} nh_n^{d+2m} = \lambda$ , we know from Hall (1984) (see also Gouriéroux and Tenreiro, 1996; Tenreiro, 1997 for geometrically  $\beta$ -mixing observations), that the convergence in distribution

$$d_n(I_n - EI_n) \xrightarrow[n \rightarrow +\infty]{d} N(0, 1), \tag{6}$$

occurs with

$$d_n = \begin{cases} nh_n^{d/2}(2\sigma_1^2 + 4\lambda\sigma_2^2)^{-1/2} & \text{if } \lambda \in [0, +\infty[ \\ \sqrt{nh_n^{-m}}(4\sigma_2^2)^{-1/2} & \text{if } \lambda = +\infty, \end{cases}$$

where

$$\sigma_1^2 = \int f^2(x) dx \int \left( \int K(u+z)K(u) du \right)^2 dz$$

and

$$\sigma_2^2 = \int (\Delta_K^m f)^2(x) f(x) dx - \left( \int (\Delta_K^m f)(x) f(x) dx \right)^2.$$

Using Lemma 1 and the fact that  $I_n$  converges in probability to zero as  $n$  tends to infinity, we easily derive the  $L_2$  consistency in probability of the kernel density estimator with data-dependent bandwidth.

**Theorem 1.** *Under the conditions of Lemma 1, if there exists  $\lambda \in [0, +\infty]$  such that  $\lim_{n \rightarrow +\infty} nh_n^{d+2m} = \lambda$ , we have*

$$J_n \xrightarrow[n \rightarrow +\infty]{p} 0.$$

Under additional assumptions on the rate of convergence of the random sequence  $(\xi_n)$ , we derive now an asymptotic expansion in probability for  $J_n$ .

**Theorem 2.** *Under the conditions of Theorem 1, if*

$$\xi_n^v = o_p \left( \frac{1}{nh_n^d} + h_n^{2m} \right), \tag{7}$$

then

$$J_n = \frac{1}{nh_n^d} \int K^2(u) du + h_n^{2m} \int (\Delta_K^m f)^2(x) dx + o_p \left( \frac{1}{nh_n^d} + h_n^{2m} \right).$$

Remark that if  $h_n \sim Cn^{-\beta}$  for some  $C > 0$  and  $0 < \beta < 1/d$ , and  $n^\alpha \xi_n$  is bounded in probability for some  $0 < \alpha \leq \frac{1}{2}$ , the condition (7) is verified if  $v > \min((1 - \beta d)/\alpha, 2\beta m/\alpha)$ . Most bandwidth selectors satisfy the assumption (B) with  $h_n = h_{\text{MISE}}$  (see the references given in Section 2). In this case, if  $\int (\Delta_K^m f)^2(x) dx \neq 0$  we have

$$h_n = h_{\text{MISE}} \sim \left( \frac{d \int K^2(u) du}{2m \int (\Delta_K^m f)^2(x) dx} \right)^{1/(d+2m)} n^{-1/(d+2m)}$$

and the condition (7) is fulfilled whenever  $v > 2m/(\alpha(d + 2m))$ . Moreover, from Theorem 2

$$J_n = J_{[K, f, m, d]} n^{-2m/(d+2m)} + o_p(n^{-2m/(d+2m)}),$$

with

$$J_{[K, f, m, d]} = (d + 2m) \left[ \frac{1}{d} \int (\Delta_K^m f)^2(x) dx \left( \frac{\int K^2(u) du}{2m} \right)^{2m} \right]^{1/(d+2m)}.$$

The following theorem follows straightforward from Lemma 1. Note that if condition (10) below is satisfied, the asymptotic behaviour of  $J_n$  does not depend on  $\xi_n$ . However, if condition (10) is not satisfied which occurs when  $\xi_n$  converges slowly to zero, the second term  $V_n$  of the expansion (9) below is at least as significant as the first one. In this case the asymptotic behaviour of  $J_n - EI_n$  strongly depends on the bandwidth selector asymptotic behaviour.

**Theorem 3.** *Under the conditions of Theorem 1, if*

$$\xi_n^v = o_p \left( \frac{1}{nh_n^{d/2}} + \frac{1}{\sqrt{nh_n^{-m}}} \right) \quad \text{or} \quad \xi_n^{v-1} = o_p \left( \frac{1}{nh_n^d} + h_n^{2m} \right), \tag{8}$$

for some  $v \in \{2, 3, \dots\}$ , we have

$$J_n - EI_n = d_n^{-1} U_n + V_n + o_p(d_n^{-1} + V_n), \tag{9}$$

where

$$U_n = d_n(I_n - EI_n) \xrightarrow[n \rightarrow +\infty]{d} N(0, 1)$$

and

$$V_n = 2\xi_n \left( \frac{1}{nh_n^d} \int K(u)K^{\hat{v}}(u) du + h_n^{2m} \int \Delta_K^m f(x)\Delta_{K^{\hat{v}}}^m f(x) dx \right).$$

Moreover, if

$$d_n \xi_n \left( \frac{1}{nh_n^d} + h_n^{2m} \right) = o_p(1), \tag{10}$$

then

$$d_n(J_n - EI_n) \xrightarrow[n \rightarrow +\infty]{d} N(0, 1).$$

Remark that if  $n^\alpha \xi_n$  is bounded in probability for some  $0 < \alpha \leq \frac{1}{2}$ , and  $h_n \sim Cn^{-\beta}$ , for some  $C > 0$  and  $0 < \beta < 1/d$ , condition (8) is satisfied whenever  $v > \min((1 - \beta d/2)/\alpha, (\frac{1}{2} + \beta m)/\alpha, 1 + (1 - \beta d)/\alpha, 1 + 2\beta m/\alpha)$ , and the same is true for condition (10) whenever  $\alpha > \frac{1}{2} - \beta m$  for  $\beta \leq 1/(d + 2m)$  and  $\alpha > \beta d/2$  for  $1/(d + 2m) < \beta < 1/d$ . Moreover, from the expansion (9) we conclude that the probability order of convergence of  $J_n - EI_n$  depends on  $\xi_n$  through  $\alpha$  and  $\beta$ . In this case, there exists a function  $\mu(\alpha, \beta)$  such that  $J_n - EI_n = O_p(n^{-\mu(\alpha, \beta)})$ . In the following we describe the behaviour of  $\mu(\alpha, \beta)$  assuming, for simplicity, that

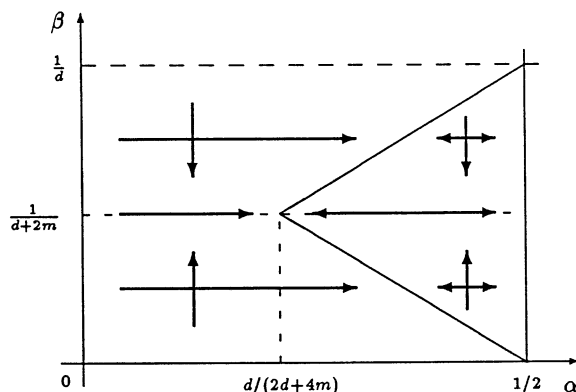


Fig. 1. Behaviour of  $\mu(\alpha, \beta)$ .

condition (8) is fulfilled for all values of  $\alpha$  and  $\beta$ . This is, in particular, true if  $K$  belongs to  $\mathcal{K}_b^v(m)$  for all  $v \in \mathbb{N}$ .

In Fig. 1 above the arrows indicate the increase of  $\mu(\alpha, \beta)$  when  $(\alpha, \beta)$  moves as shown. A double sided arrow indicate that  $\mu(\alpha, \beta)$  is constant along that direction. It is clear that the common practice of considering bandwidth selectors satisfying (B) with  $h_n = h_{\text{MISE}}$  produces a simultaneous minimization of the “asymptotic mean” of  $J_n$  and maximization of the order of convergence of the “asymptotic variance” of  $J_n$ . In this “optimal” case  $\beta = 1/(d + 2m)$  and

$$\mu(\alpha, \beta) = \begin{cases} \alpha + \frac{2m}{d + 2m} & \text{if } 0 < \alpha < \frac{d}{2d + 4m}, \\ \frac{d + 4m}{2d + 4m} & \text{if } \frac{d}{2d + 4m} \leq \alpha \leq \frac{1}{2}. \end{cases}$$

Moreover, if  $\alpha > d/(2d + 4m)$  the asymptotic behaviour of  $J_n$  does not depend on  $\xi_n$ . Bandwidth selectors verifying this condition are then indistinguishable from an integrated square error point of view.

In a real context, if  $K$  is the standard normal kernel ( $m = 2$ ), and (B) is satisfied with  $h_n = h_{\text{MISE}}$ , we have  $\mu(\alpha, \frac{1}{5}) = \frac{9}{10}$ , for  $\frac{1}{10} \leq \alpha \leq \frac{1}{2}$ . This optimal order of convergence is achieved for all the bandwidth selectors referred in Section 2 which satisfy (B) with  $h_n = h_{\text{MISE}}$ .

### 4. Proofs

**Proof of Lemma 1.** For  $x \in \mathbb{R}^d$  let us denote

$$\bar{E}\hat{f}_n(x) = \int K_{A_n}(x - y)f(y) \, dy$$

and consider the following expansion

$$\begin{aligned} J_n &= \int \{\hat{f}_n(x) - \bar{E}\hat{f}_n(x)\}^2 \, dx \\ &+ 2 \int \{\hat{f}_n(x) - \bar{E}\hat{f}_n(x)\} \{\bar{E}\hat{f}_n(x) - E f_n(x)\} \, dx \\ &+ 2 \int \{\hat{f}_n(x) - \bar{E}\hat{f}_n(x)\} \{E f_n(x) - f(x)\} \, dx \end{aligned}$$

$$\begin{aligned}
 &+ \int \{\bar{E}\hat{f}_n(x) - E f_n(x)\}^2 dx \\
 &+ 2 \int \{\bar{E}\hat{f}_n(x) - E f_n(x)\}\{E f_n(x) - f(x)\} dx \\
 &+ \int \{E f_n(x) - f(x)\}^2 dx \\
 &=: J_n^1 + 2J_n^2 + 2J_n^3 + J_n^4 + 2J_n^5 + \int \{E f_n(x) - f(x)\}^2 dx.
 \end{aligned} \tag{11}$$

Each one of the previous terms will be studied in the following propositions.

**Proposition 1.** For  $u \in \mathbb{R}^d$  and  $h > 0$ , define  $W(u, h) = K(u/h)/h^d$ . If  $K \in \mathcal{X}_b^v(0)$  for some  $v \in \mathbb{N}$ , we have:

(a) For  $x \in \mathbb{R}^d$ ,

$$\hat{f}_n(x) = f_n(x) + \sum_{\ell=1}^{v-1} \zeta_n^\ell \frac{1}{\ell!} \frac{1}{n} \sum_{i=1}^n K_{h_n}^{\partial(\ell)}(x - X_i) + \zeta_n^v \frac{1}{n} \sum_{i=1}^n R_{h_n}(x - X_i, A_n)$$

and

$$\bar{E}\hat{f}_n(x) = E f_n(x) + \sum_{\ell=1}^{v-1} \zeta_n^\ell \frac{1}{\ell!} \int K_{h_n}^{\partial(\ell)}(x - y) f(y) dy + \zeta_n^v \int R_{h_n}(x - y, A_n) f(y) dy,$$

where, for  $u \in \mathbb{R}$  and  $h > 0$ ,

$$K^{\partial(\ell)}(u) = \frac{\partial^\ell W}{\partial h^\ell}(u, 1), \quad \ell = 0, 1, \dots, v - 1$$

and

$$R(u, h) = \frac{1}{(v-1)!} \int_0^1 (1-t)^{v-1} \frac{\partial^v W}{\partial h^v}(u, 1 + t(h-1)) dt.$$

(b) Moreover, the functions  $K^{\partial(\ell)}$  for  $\ell = 0, 1, \dots, v - 1$ , and  $u \rightarrow \sup_{|h-1| \leq \varepsilon} |R(u, h)|$ , where  $\varepsilon > 0$  is given in the definition of  $K$ , are bounded and integrable.

**Proof.** Since  $K$  is  $v$ -times continuously differentiable on  $\mathbb{R}^d$ , we have, from the Taylor expansion,

$$W(u, h) = \sum_{\ell=1}^{v-1} (h-1)^\ell \frac{1}{\ell!} \frac{\partial^\ell W}{\partial h^\ell}(u, 1) + (h-1)^v \frac{1}{(v-1)!} \int_0^1 (1-t)^{v-1} \frac{\partial^v W}{\partial h^v}(u, 1 + t(h-1)) dt,$$

for  $u \in \mathbb{R}^d$  and  $h > 0$ . Thus (a) follows from the equalities (4) and (5) and

$$\bar{E}\hat{f}_n(x) = \frac{1}{h_n^d} \int W\left(\frac{x-y}{h_n}, \frac{A_n}{h_n}\right) f(y) dy.$$

Since  $K \in \mathcal{X}_b^v(0)$ , the boundedness and the integrability of functions given in (b) follows.  $\square$

In all of the following propositions we assume that the assumption (B) is satisfied.

**Proposition 2.** *If  $f$  is bounded and  $K \in \mathcal{K}_b^v(0)$  for some  $v \in \mathbb{N}$ , then*

$$J_n^1 = \int \{f_n(x) - E f_n(x)\}^2 dx + 2(1 - \delta_{1v})\xi_n \frac{1}{nh_n^d} \int K(u)K^{\hat{v}}(u) du + O_p \left( \xi_n \frac{1}{nh_n^{d/2}} + \xi_n^2 \frac{1}{nh_n^d} + \xi_n^v \right),$$

where, for  $u = (u_1, \dots, u_d) \in \mathbb{R}^d$ ,  $K^{\hat{v}}(u) = -dK(u) - \sum_{i=1}^d u_i \partial K / \partial u_i(u)$ .

**Proof.** From Proposition 1(a) and for  $x \in \mathbb{R}^d$ , we have

$$\begin{aligned} \hat{f}_n(x) - \tilde{E} \hat{f}_n(x) &= f_n(x) - E f_n(x) \\ &+ \sum_{\ell=1}^{v-1} \xi_n \frac{1}{\ell!} \frac{1}{n} \sum_{i=1}^n \{K_{h_n}^{\hat{\partial}(\ell)}(x - X_i) - EK_{h_n}^{\hat{\partial}(\ell)}(x - X_0)\} \\ &+ \xi_n^v \frac{1}{n} \sum_{i=1}^n \left\{ R_{h_n}(x - X_i, A_n) - \int R_{h_n}(x - y, A_n) f(y) dy \right\}. \end{aligned} \tag{12}$$

Therefore, from Proposition 1(b) and the convergence in probability of  $A_n/h_n$  to 1, when  $n \rightarrow +\infty$ , we get

$$J_n^1 = \int \{f_n(x) - E f_n(x)\}^2 dx + \sum_{\substack{\ell, \ell' = 0 \\ \ell + \ell' \geq 1}}^{v-1} \xi_n^{\ell + \ell'} \frac{1}{\ell! \ell'!} V_n^{\ell, \ell'} + O_p(\xi_n^v),$$

where

$$V_n^{\ell, \ell'} = \frac{1}{n^2} \sum_{i,j=1}^n \int \{K_{h_n}^{\hat{\partial}(\ell)}(x - X_i) - EK_{h_n}^{\hat{\partial}(\ell)}(x - X_0)\} \{K_{h_n}^{\hat{\partial}(\ell')}(x - X_j) - EK_{h_n}^{\hat{\partial}(\ell')}(x - X_0)\} dx.$$

Using degenerated U-statistics techniques (see Hall, 1984) and the fact that  $K^{\hat{\partial}(\ell)}$  for  $\ell = 0, 1, \dots, v - 1$ , are bounded and integrable, we obtain

$$V_n^{\ell, \ell'} = \frac{1}{nh_n^d} \int K^{\hat{\partial}(\ell)}(u)K^{\hat{\partial}(\ell')}(u) du + \frac{1}{nh_n^{d/2}} U_n^{\ell, \ell'} + o_p \left( \frac{1}{nh_n^{d/2}} \right),$$

where  $U_n^{\ell, \ell'}$  is asymptotically normal. Then,

$$J_n^1 = \int \{f_n(x) - E f_n(x)\}^2 dx + 2(1 - \delta_{1v})\xi_n \frac{1}{nh_n^d} \int K(u)K^{\hat{v}}(u) du + O_p \left( \xi_n \frac{1}{nh_n^{d/2}} + \xi_n^2 \frac{1}{nh_n^d} + \xi_n^v \right),$$

where  $K^{\hat{v}} = K^{\hat{\partial}(1)}$ . Finally, it suffices to note that  $\partial W / \partial h(u, h) = -dK_h(u)/h - \sum_{i=1}^d u_i \partial K_h / \partial u_i(u)/h$ , for  $u = (u_1, \dots, u_d) \in \mathbb{R}^d$  and  $h > 0$ .  $\square$

**Proposition 3.** *If  $f \in \mathcal{W}_b^m(m)$  and  $K \in \mathcal{K}_b^v(m)$  for some  $m \in \mathbb{N}$  and  $v \in \mathbb{N}$ , then*

$$J_n^2 = O_p \left( \xi_n \frac{1}{\sqrt{nh_n^{-m}}} + \xi_n^v \right).$$



**Proof.** From Proposition 1 we deduce

$$J_n^2 = \sum_{\ell, \ell'=1}^{v-1} \xi_n^{\ell+\ell'} \frac{1}{\ell! \ell'!} \frac{1}{n} \sum_{i=1}^n \int \{K_{h_n}^{\partial(\ell)}(x - X_i) - EK_{h_n}^{\partial(\ell)}(x - X_0)\} \psi_{n, \ell'}(x) dx + O_p(\xi_n^v),$$

where  $\psi_{n, \ell}(x) = \int K_{h_n}^{\partial(\ell)}(x - y) f(y) dy$ , for  $\ell = 1, \dots, v - 1$ .

On the other hand, as for  $f \in \mathcal{W}_b(m)$  and  $K \in \mathcal{K}_b^v(m)$  we have

$$\sup_{n \in \mathbb{N}} \sup_{x \in \mathbb{R}^d} |\psi_{n, \ell}(x)| = O(h_n^m), \tag{13}$$

the conclusion follows.  $\square$

**Proposition 4.** Under the conditions of Proposition 3 we have

$$J_n^3 = \int \{f_n(x) - Ef_n(x)\} \{Ef_n(x) - f(x)\} dx + O_p \left( \xi_n \frac{1}{\sqrt{nh_n^{-m}}} + \xi_n^v h_n^m \right).$$

**Proof.** From (12) and from the equality

$$\sup_{n \in \mathbb{N}} \sup_{x \in \mathbb{R}^d} |Ef_n(x) - f(x)| = O(h_n^m), \tag{14}$$

which is valid for  $f \in \mathcal{W}_b(m)$  and  $K \in \mathcal{K}_b(m)$ , we get

$$\begin{aligned} J_n^3 &= \int \{f_n(x) - Ef_n(x)\} \{Ef_n(x) - f(x)\} dx \\ &+ \sum_{\ell=1}^{v-1} \xi_n^{\ell} \frac{1}{\ell!} \frac{1}{n} \sum_{i=1}^n \int \{K_{h_n}^{\partial(\ell)}(x - X_i) - EK_{h_n}^{\partial(\ell)}(x - X_0)\} \{Ef_n(x) - f(x)\} dx + O_p(\xi_n^v h_n^m) \\ &= \int \{f_n(x) - Ef_n(x)\} \{Ef_n(x) - f(x)\} dx + O_p \left( \xi_n \frac{1}{\sqrt{nh_n^{-m}}} + \xi_n^v h_n^m \right). \quad \square \end{aligned}$$

**Proposition 5.** Under the conditions of Proposition 3 we have

$$J_n^4 = O_p(\xi_n^2 h_n^{2m} + \xi_n^{v+1}).$$

**Proof.** Follows from Proposition 1 and (13).  $\square$

**Proposition 6.** Under the conditions of Proposition 3, if the  $m$ -order partial derivatives of  $f$  are square integrable we have

$$J_n^5 = (1 - \delta_{1v}) \xi_n h_n^{2m} \int \Delta_K^m f(x) \Delta_{K^c}^m f(x) dx + o_p(\xi_n h_n^{2m}) + O_p(\xi_n^v h_n^m).$$

**Proof.** From Proposition 1 and (14) we have

$$J_n^5 = \sum_{\ell=1}^{v-1} \xi_n^{\ell} \frac{1}{\ell!} V_{n, \ell} + O_p(\xi_n^v h_n^m),$$

where, for  $\ell = 1, \dots, v - 1$ ,

$$\begin{aligned} V_{n,\ell} &= \int \int K_{h_n}^{\delta(\ell)}(x-y) f(y) \, dy \{E f_n(x) - f(x)\} \, dx \\ &= h_n^{2m} \int \Delta_K^m f(x) \Delta_{K^{\delta(\ell)}}^m f(x) \, dx + o_p(h_n^{2m}). \quad \square \end{aligned}$$

Now, the asymptotic expansion for  $J_n$  given in Lemma 1 is a trivial consequence of (11) and of the previous propositions. The proof of Lemma 1 is then completed.  $\square$

**Proof of Theorem 2.** From Lemma 1 we have

$$J_n = I_n + o_p\left(\frac{1}{nh_n^d} + h_n^{2m}\right).$$

The result follows now from the convergence in distribution (6) and the usual expansion (cf. Bosq and Lecoutre, 1987, p. 80)

$$EI_n = \frac{1}{nh_n^d} \int K^2(u) \, du + h_n^{2m} \int (\Delta_K^m f)^2(x) \, dx + O\left(\frac{1}{n}\right) + o(h_n^{2m}). \quad \square$$

## Acknowledgements

The author is grateful to a referee for helpful comments.

## References

- Bickel, P.J., Rosenblatt, M., 1973. On some global measures of the deviations of density function estimates. *Ann. Statist.* 1, 1071–1095.
- Bosq, D., Lecoutre, J.-P., 1987. *Théorie de l'Estimation Fonctionnelle*. Economica, Paris.
- Fan, Y., 1994. Testing the goodness of fit of a parametric density function by kernel method. *Econom. Theory* 10, 316–356.
- Gouriéroux, C., Tenreiro, C., 1996. Local power properties of kernel based goodness of fit tests, Preprint 9617, Departamento de Matemática, Universidade de Coimbra. *J. Multivariate Anal.*, to appear.
- Hall, P., 1984. Central limit theorem for integrated square error properties of multivariate nonparametric density estimators. *J. Multivariate Anal.* 14, 1–16.
- Hall, P., Marron, J.S., 1987. Extent to which least-squares cross-validation minimises integrated square error in nonparametric density estimation. *Probab. Theory Related Fields* 74, 567–581.
- Hall, P., Marron, J.S., 1991. Lower bound for bandwidth selection in density estimation. *Probab. Theory Related Fields* 90, 149–173.
- Hall, P., Marron, J.S., Park, B.U., 1992. Smoothed cross-validation. *Probab. Theory Related Fields* 92, 1–20.
- Hall, P., Sheather, S.J., Jones, M.C., Marron, J.S., 1991. On optimal data-based bandwidth selection in kernel density estimation. *Biometrika* 78, 263–269.
- Jones, M.C., Marron, J.S., Park, B.U., 1991. A simple root  $n$  bandwidth selector. *Ann. Statist.* 19, 1919–1932.
- Jones, M.C., Marron, J.S., Sheather, S.J., 1996. A brief survey of bandwidth selection for density estimation. *J. Amer. Statist. Assoc.* 91, 401–407.
- Liero, H., 1992. Asymptotic normality of a weighted integrated squared error of kernel regression estimates with data-dependent bandwidth. *J. Statist. Plann. Inference* 30, 307–325.
- Loader, C.R., 1999. Bandwidth selection: classical or plug-in? *Ann. Statist.* 27, 415–438.
- Park, B.U., Marron, J.S., 1990. Comparison of data-driven bandwidth selectors. *J. Amer. Statist. Assoc.* 85, 66–72.
- Parzen, E., 1962. On estimation of a probability density function and mode. *Ann. Math. Statist.* 33, 1065–1076.
- Rosenblatt, M., 1956. Remarks on some non-parametric estimates of a density function. *Ann. Math. Statist.* 27, 832–837.
- Scott, D.W., Terrel, G.R., 1987. Biased and unbiased cross-validation in density estimation. *J. Amer. Statist. Assoc.* 82, 1131–1146.
- Sheather, S.J., Jones, M.C., 1991. A reliable data-based bandwidth selection method for kernel density estimation. *J. Roy. Statist. Soc. Ser. B* 53, 683–690.
- Tenreiro, C., 1997. Loi asymptotique des erreurs quadratiques intégrées des estimateurs à noyau de la densité et de la régression sous des conditions de dépendance. *Portugaliae Math.* 54, 187–213.
- Terrel, G.R., 1990. The maximal smoothing principle in density estimation. *J. Amer. Statist. Assoc.* 85, 470–477.