



**HAL**  
open science

# On the weak convergence of kernel density estimators in $L_p$ spaces

Gilles Stupfler

► **To cite this version:**

Gilles Stupfler. On the weak convergence of kernel density estimators in  $L_p$  spaces. 2014. hal-00957179v1

**HAL Id: hal-00957179**

**<https://hal.science/hal-00957179v1>**

Preprint submitted on 9 Mar 2014 (v1), last revised 6 Sep 2015 (v2)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# On the weak convergence of kernel density estimators in $L^p$ spaces

Gilles Stupfler

Aix Marseille Université, CERGAM, EA 4225,  
15-19 allée Claude Forbin, 13628 Aix-en-Provence Cedex 1, France

**Abstract.** Since its introduction, the pointwise asymptotic properties of the kernel estimator  $\hat{f}_n$  of a probability density function  $f$  on  $\mathbb{R}^d$ , as well as the asymptotic behaviour of its integrated errors, have been studied in great detail. Its weak convergence in functional spaces, however, is a more difficult problem. In this paper, we show that if  $f_n(x) = \mathbb{E}(\hat{f}_n(x))$  and  $(r_n)$  is any nonrandom sequence of positive real numbers such that  $r_n/\sqrt{n} \rightarrow 0$  then if  $r_n(\hat{f}_n - f_n)$  converges to a Borel measurable weak limit in a weighted  $L^p$  space on  $\mathbb{R}^d$ , with  $1 \leq p < \infty$ , the limit must be 0. We also provide simple conditions for proving or disproving the existence of this Borel measurable weak limit.

**AMS Subject Classifications:** 62G07, 62G20, 60F17.

**Keywords:** Kernel density estimator, weak convergence,  $L^p$  space.

## 1 Introduction

Let  $(X_n)$  be a sequence of independent random copies of a random variable  $X$ , such that  $X$  has a probability density function  $f$  on  $\mathbb{R}^d$ ,  $d \geq 1$ . The Parzen-Rosenblatt estimator of  $f$  is

(Parzen, 1962, Rosenblatt, 1956):

$$\widehat{f}_n(x) = \frac{1}{n} \sum_{i=1}^n K_h(x - X_i)$$

where  $h := h(n) \rightarrow 0$  as  $n \rightarrow \infty$  and  $K_h(u) = h^{-d}K(u/h)$  with  $K$  a probability density function on  $\mathbb{R}^d$ . The quantity  $\widehat{f}_n(x)$  is the empirical estimator of  $f_n(x) = \mathbb{E}(\widehat{f}_n(x)) = \mathbb{E}[K_h(x - X)]$ . It is well-known that  $x \mapsto \sqrt{nh^d}(\widehat{f}_n(x) - f_n(x))$  converges pointwise to a Gaussian distribution provided  $nh^d \rightarrow \infty$  and  $K^2$  is integrable on  $\mathbb{R}^d$ .

Our focus in this paper is rather the study of the convergence properties of the random process  $x \mapsto r_n(\widehat{f}_n(x) - f_n(x))$ , where  $(r_n)$  is a nonrandom sequence of positive real numbers, in  $L^p$  spaces. This is interesting for many purposes, such as examining the asymptotics of the global integrated error of the estimator  $\widehat{f}_n$ , which can then be used to construct asymptotic confidence bands for the function  $f$ . For instance, since  $K$  is a probability density function, the random function  $r_n(\widehat{f}_n - f_n)$  belongs to the space  $L^1(\mathbb{R}^d)$  of Borel measurable functions which are integrable on  $\mathbb{R}^d$ . Besides, if we let  $\mu$  be a nontrivial absolutely continuous measure with respect to the Lebesgue measure on  $\mathbb{R}^d$  and, for any  $p \in [1, \infty)$ ,  $L^p(\mathbb{R}^d, \mu)$  be the space of the Borel measurable functions  $H : \mathbb{R}^d \rightarrow \mathbb{R}$  such that

$$\|H\|_{p,\mu} := \left[ \int_{\mathbb{R}^d} |H(x)|^p d\mu(x) \right]^{1/p} < \infty$$

then, assuming that  $K^p$  is integrable on  $\mathbb{R}^d$ , the random function  $r_n(\widehat{f}_n - f_n)$  belongs to  $L^p(\mathbb{R}^d, \mu)$ . The convergence of  $\|\widehat{f}_n - f_n\|_{p,\mu}$  has been studied by Bickel and Rosenblatt (1973) for  $p = 2$  and Devroye and Györfi (1985) for  $p = 1$ ; a treatment of the general case  $1 \leq p < \infty$  is given in Csörgő and Horváth (1988), Horváth (1991) and Beirlant and Mason (1995).

None of these studies consider the convergence of  $x \mapsto r_n(\widehat{f}_n(x) - f_n(x))$  as a random process taking values in an  $L^p$  space, which is a difficult problem. Recently, Nishiyama (2011) disproved the existence of a nondegenerate weak limit for  $r_n(\widehat{f}_n - f_n)$  in the usual  $L^2(\mathbb{R}^d)$  space provided  $r_n/\sqrt{n} \rightarrow 0$ , thus generalizing a result of Ruymgaart (1998). Our focus in this paper is to generalize the results of Nishiyama (2011) to the convergence of  $r_n(\widehat{f}_n - f_n)$  in the spaces  $L^p(\mathbb{R}^d, \mu)$ , for  $1 \leq p < \infty$ , when  $\mu$  has a bounded Radon-Nikodym derivative. In particular, we shall show that any Borel measurable weak limit of this process in  $L^p(\mathbb{R}^d, \mu)$  is necessarily

0, and that  $\sqrt{nh^d}(\widehat{f}_n - f_n)$  does not converge to a Borel measurable limit in  $L^p(\mathbb{R}^d, \mu)$  under mild conditions on  $f$ ,  $K$  and  $\mu$ . Note that this does not contradict the aforementioned results on the asymptotics of  $\|\widehat{f}_n - f_n\|_{p, \mu}$ , since weak convergence of the norms is necessary but not sufficient for weak convergence in  $L^p(\mathbb{R}^d, \mu)$ .

The outline of the paper is as follows: our main results are stated in Section 2, while statements and proofs of the auxiliary results are deferred to the Appendix.

## 2 Main results

Let  $p \in [1, \infty)$ . Our first goal is to obtain a simple necessary and sufficient condition to identify the limit of a Borel measurable random process in the space  $L^p(\mathbb{R}^d, \mu)$ . To state such a result, we introduce some notation: let  $q \in (1, \infty]$  be such that  $p^{-1} + q^{-1} = 1$ . For any  $H \in L^q(\mathbb{R}^d, \mu)$ , let  $T_H$  be the continuous linear form on  $L^p(\mathbb{R}^d, \mu)$  defined by

$$\forall G \in L^p(\mathbb{R}^d, \mu), T_H(G) = \int_{\mathbb{R}^d} G(x)H(x)d\mu(x).$$

In the remainder of this paper, we assume that  $\mu$  is a nontrivial absolutely continuous measure with respect to the Lebesgue measure on  $\mathbb{R}^d$ , having a bounded Radon-Nikodym derivative. We have the following result:

**Proposition 1.** *If  $G_1$  and  $G_2$  are two Borel measurable random elements of  $L^p(\mathbb{R}^d, \mu)$ , then the distributions of  $G_1$  and  $G_2$  are equal if and only if for every bounded function  $H \in L^q(\mathbb{R}^d, \mu)$ , the distributions of  $T_H(G_1)$  and  $T_H(G_2)$  are equal.*

**Proof of Proposition 1.** Since for any  $H \in L^q(\mathbb{R}^d, \mu)$ , the map  $T_H$  is a continuous linear form on  $L^p(\mathbb{R}^d, \mu)$ , it is clear that if  $G_1$  and  $G_2$  have equal distributions then  $T_H(G_1)$  and  $T_H(G_2)$  must have equal distributions as well. Conversely, assume that for any bounded function  $H \in L^q(\mathbb{R}^d, \mu)$  the distributions of  $T_H(G_1)$  and  $T_H(G_2)$  are equal. Let  $\mathcal{F}$  be the class of functions  $f : L^p(\mathbb{R}^d, \mu) \rightarrow \mathbb{R}$  for which there exists  $J \geq 1$  such that

$$\forall \varphi \in L^p(\mathbb{R}^d, \mu), f(\varphi) = g(T_{H_1}(\varphi), \dots, T_{H_J}(\varphi))$$

where  $g$  is a continuous bounded real-valued function on  $\mathbb{R}^J$  and  $H_1, \dots, H_J$  are bounded elements of  $L^q(\mathbb{R}^d, \mu)$ . Let  $\nu_1$  and  $\nu_2$  be the pushforward probability measures on  $L^p(\mathbb{R}^d, \mu)$

induced by  $G_1$  and  $G_2$ ; it is clear that

$$\forall f \in \mathcal{F}, \int_{L^p(\mathbb{R}^d, \mu)} f(\varphi) d\nu_1(\varphi) = \int_{L^p(\mathbb{R}^d, \mu)} f(\varphi) d\nu_2(\varphi).$$

Further, in the sense of van der Vaart and Wellner (1996), p.25, the class  $\mathcal{F}$  is a vector lattice of continuous bounded functions on  $L^p(\mathbb{R}^d, \mu)$  which contains the constant functions. Let  $\mathcal{B}(\mathbb{R}^d)$  be the usual Borel  $\sigma$ -algebra on  $\mathbb{R}^d$ : since  $\mu$  is absolutely continuous with respect to the Lebesgue measure and has a bounded Radon-Nikodym derivative, the space  $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \mu)$  is a separable measure space. This makes  $L^p(\mathbb{R}^d, \mu)$  a separable metric space; since  $L^p(\mathbb{R}^d, \mu)$  is complete, it follows that  $G_1$  and  $G_2$  define tight Borel probability measures on  $L^p(\mathbb{R}^d, \mu)$ , see Lemma 1.3.2 p.17 in van der Vaart and Wellner (1996). Thanks to Lemma 1.3.12 p.25 in van der Vaart and Wellner (1996), the proof shall be complete provided we show that the class  $\mathcal{F}$  separates the points of  $L^p(\mathbb{R}^d, \mu)$ .

Let then  $\varphi, \psi \in L^p(\mathbb{R}^d, \mu)$  be such that  $\varphi \neq \psi$ . It is a corollary of Hahn-Banach's theorem that there exists a continuous linear form  $T$  on  $L^p(\mathbb{R}^d, \mu)$  such that  $T(\varphi) \neq T(\psi)$ . Since  $p \in [1, \infty)$ , the dual space of  $L^p(\mathbb{R}^d, \mu)$  is  $L^q(\mathbb{R}^d, \mu)$ , so that  $T = T_h$  for some  $h \in L^q(\mathbb{R}^d, \mu)$ . Especially

$$T_h(\varphi - \psi) = \int_{\mathbb{R}^d} [\varphi(x) - \psi(x)]h(x)d\mu(x) \neq 0.$$

By the dominated convergence theorem, this yields, for some positive integer  $N$ :

$$\left| \int_{\mathbb{R}^d} [\varphi(x) - \psi(x)]h(x)\mathbb{I}_{\{|h(x)| \leq N\}}d\mu(x) \right| \geq \frac{|T_h(\varphi - \psi)|}{2} > 0.$$

Therefore, the function  $H$  defined by  $H(x) = h(x)\mathbb{I}_{\{|h(x)| \leq N\}}$  is a bounded element of  $L^q(\mathbb{R}^d, \mu)$  such that  $T_H(\varphi) \neq T_H(\psi)$ , which completes the proof.  $\blacksquare$

The key idea in this result is that if for any bounded function  $H \in L^q(\mathbb{R}^d, \mu)$ , we have  $T_H(r_n(\widehat{f}_n - f_n)) \rightarrow T_H(G)$  where  $G$  is a Borel measurable element of  $L^p(\mathbb{R}^d, \mu)$ , then the Borel measurable weak limit of  $r_n(\widehat{f}_n - f_n)$  in  $L^p(\mathbb{R}^d, \mu)$ , if it exists, must be  $G$ . In some sense, this generalizes the approach of Nishiyama (2011), where an essential element of the proof is that one can find a countable dense subset of bounded functions  $(e_j)$  in  $L^2(\mathbb{R}^d)$  and characterize the possible weak limits of  $r_n(\widehat{f}_n - f_n)$  by examining the asymptotic properties of  $T_{e_j}(r_n(\widehat{f}_n - f_n))$  for any integer  $j$ .

We may now state our first asymptotic result, which identifies the possible Borel measurable limit of  $r_n(\widehat{f}_n - f_n)$  in the space  $L^p(\mathbb{R}^d, \mu)$ :

**Theorem 1.** *Let  $(r_n)$  be a nonrandom sequence of positive real numbers. Assume that  $K^p$  is integrable on  $\mathbb{R}^d$ . If  $r_n/\sqrt{n} \rightarrow 0$  and the random process  $r_n(\widehat{f}_n - f_n)$  converges weakly in  $L^p(\mathbb{R}^d, \mu)$  to a Borel measurable random process  $G$  then  $G = 0$  almost surely.*

**Proof of Theorem 1.** According to Proposition 1, it is enough to show that for every bounded function  $H \in L^q(\mathbb{R}^d, \mu)$ , one has

$$\Delta_n(H) := T_H \left( r_n(\widehat{f}_n - f_n) \right) \rightarrow 0$$

in probability. To this aim, we set  $Z_i(x, h) := K_h(x - X_i) - \mathbb{E}(K_h(x - X_i))$  for  $i = 1, \dots, n$  and we rewrite  $\Delta_n(H)$  as

$$\Delta_n(H) = \frac{r_n}{n} \sum_{i=1}^n W_{n,i} \quad \text{with} \quad W_{n,i} := \int_{\mathbb{R}^d} Z_i(x, h) H(x) d\mu(x).$$

Since the  $Z_i(x, h)$ ,  $i = 1, \dots, n$  are independent copies of the centered random variable  $Z(x, h) := K_h(x - X) - \mathbb{E}(K_h(x - X))$ , the random variables  $W_{n,i}$ ,  $i = 1, \dots, n$  are themselves independent and centered copies of  $W_{n,1}$ . Lemma 1 thus entails for any  $n$

$$\mathbb{E}|\Delta_n(H)|^2 = \left[ \frac{r_n}{\sqrt{n}} \right]^2 \mathbb{E}|W_{n,1}|^2 = \mathcal{O} \left( \left[ \frac{r_n}{\sqrt{n}} \right]^2 \right).$$

A particular consequence of this inequality is that  $\Delta_n(H) \rightarrow 0$  in probability: the proof is complete. ■

We note that Theorem 1 is a generalization of Theorem 2.1 in Nishiyama (2011), which was restricted to the case when  $p = 2$  and  $\mu$  is the Lebesgue measure. This result says that either the limit of the process  $r_n(\widehat{f}_n - f_n)$  exists and is degenerate, or the limit is not Borel measurable. We proceed by giving some insight on what can happen, depending on the rate  $(r_n)$ . Introduce the hypotheses

(H1) The function  $f$  is a bounded function on  $\mathbb{R}^d$ .

(H2) The function  $K$  is a bounded function on  $\mathbb{R}^d$ .

(H3a) The Radon-Nikodym derivative of  $\mu$  with respect to the Lebesgue measure on  $\mathbb{R}^d$  is bounded from below by a positive constant.

(H3b) There exist  $x_0 \in \mathbb{R}^d$  and  $\delta > 0$  such that  $f$  is bounded from below by a positive constant on the Euclidean ball  $B(x_0, \delta)$  with center  $x_0$  and radius  $\delta$ , and  $\mu(B(x_0, \delta/2)) > 0$ .

Hypothesis (H1) was already introduced in Nishiyama (2011), while hypothesis (H2) is classical in kernel density estimation. Hypothesis (H3a) holds especially if  $\mu$  is the Lebesgue measure on  $\mathbb{R}^d$ ; hypothesis (H3b) is true if, for instance, there exists  $x_0 \in \mathbb{R}^d$  such that both  $f$  and the Radon-Nikodym derivative of  $\mu$  with respect to the Lebesgue measure on  $\mathbb{R}^d$  are positive and continuous at  $x_0$ . Stronger versions of this latter condition can be found for instance in Csörgő and Horváth (1988) and Horváth (1991). Starting with the case  $p \geq 2$ , we can state the following result, which contains Theorem 2.2 in Nishiyama (2011):

**Theorem 2.** *Consider the case  $p \geq 2$ . Let  $(r_n)$  be a nonrandom sequence of positive real numbers. Assume that  $K^p$  is integrable on  $\mathbb{R}^d$ ; if  $p > 2$ , assume that (H1) holds and that  $nh^d \rightarrow \infty$ .*

- *If  $r_n/\sqrt{nh^d} \rightarrow 0$ , then  $r_n(\widehat{f}_n - f_n) \rightarrow 0$  in  $L^p(\mathbb{R}^d, \mu)$ .*

*Assume now that (H1) holds. In case  $p = 2$ , assume that (H3a) or (H3b) holds; when  $p > 2$ , assume that (H2) and (H3b) hold.*

- *If  $r_n/\sqrt{nh^d} \rightarrow c \in (0, \infty]$ , then  $r_n(\widehat{f}_n - f_n)$  does not converge to any Borel measurable random element in  $L^p(\mathbb{R}^d, \mu)$ .*

**Proof of Theorem 2.** Consider first the case  $r_n/\sqrt{nh^d} \rightarrow 0$ . The first statement shall be proven provided we show that  $\mathbb{E}(\|r_n(\widehat{f}_n - f_n)\|_{p,\mu}^p) \rightarrow 0$ . To this end, we note that

$$\mathbb{E}(\|r_n(\widehat{f}_n - f_n)\|_{p,\mu}^p) = \left[\frac{r_n}{n}\right]^p \int_{\mathbb{R}^d} \mathbb{E} \left| \sum_{i=1}^n Z_i(x, h) \right|^p d\mu(x). \quad (1)$$

Rosenthal's inequality (Rosenthal, 1970) thus entails

$$\mathbb{E} \left| \sum_{i=1}^n Z_i(x, h) \right|^p \leq B_p \max \left[ n\mathbb{E}|Z(x, h)|^p, (n\mathbb{E}|Z(x, h)|^2)^{p/2} \right] \quad (2)$$

for some positive constant  $B_p$  which only depends on  $p$ . Using together (1), (2) and Lemma 2 entails

$$\mathbb{E}(\|r_n(\widehat{f}_n - f_n)\|_{p,\mu}^p) = \mathcal{O}\left(\left[\frac{r_n}{n}\right]^p \left[nh^{-d(p-1)} + n^{p/2}h^{-dp/2}\right]\right).$$

Since

$$\left[\frac{r_n}{n}\right]^p \left[nh^{-d(p-1)} + n^{p/2}h^{-dp/2}\right] = \left[\frac{r_n}{\sqrt{nh^d}}\right]^p \left[1 + (nh^d)^{1-p/2}\right] \rightarrow 0$$

we obtain  $\mathbb{E}(\|r_n(\widehat{f}_n - f_n)\|_{p,\mu}^p) = o(1)$ , which concludes the proof of the first statement.

To prove the second statement, we start by assuming that  $r_n/\sqrt{nh^d} \rightarrow c \in (0, \infty)$  and we show that for any  $p \geq 2$ ,

$$\liminf_{n \rightarrow \infty} \mathbb{E}(\|r_n(\widehat{f}_n - f_n)\|_{p,\mu}^p) > 0. \quad (3)$$

To this aim, we apply Rosenthal's inequality (Rosenthal, 1970) to bound the integrand in equation (1) from below:

$$\mathbb{E}\left|\sum_{i=1}^n Z_i(x, h)\right|^p \geq A_p \max\left[n\mathbb{E}|Z(x, h)|^p, (n\mathbb{E}|Z(x, h)|^2)^{p/2}\right] \geq A_p(n\mathbb{E}|Z(x, h)|^2)^{p/2} \quad (4)$$

for some positive constant  $A_p$  which only depends on  $p$ . Especially,

$$\mathbb{E}(\|r_n(\widehat{f}_n - f_n)\|_{p,\mu}^p) \geq A_p \left[\frac{r_n}{n}\right]^p \int_{\mathbb{R}^d} (n\mathbb{E}|Z(x, h)|^2)^{p/2} d\mu(x).$$

Further, remark that since  $C = \sup_{\mathbb{R}^d} f < \infty$ ,

$$\begin{aligned} \mathbb{E}|Z(x, h)|^2 &= h^{-d} \int_{\mathbb{R}^d} K^2(v) f(x - hv) dv - \left[ \int_{\mathbb{R}^d} K(v) f(x - hv) dv \right]^2 \\ &\geq h^{-d} \int_{\mathbb{R}^d} K^2(v) f(x - hv) dv - C \int_{\mathbb{R}^d} K(v) f(x - hv) dv \end{aligned} \quad (5)$$

where  $C$  is a positive constant. When  $p = 2$  and (H3a) holds, using (1), (4) and (5) entails for some  $m > 0$

$$\mathbb{E}(\|r_n(\widehat{f}_n - f_n)\|_{2,\mu}^2) \geq m \frac{r_n^2}{n} \int_{\mathbb{R}^d} \mathbb{E}|Z(x, h)|^2 dx = m \left[\frac{r_n}{\sqrt{nh^d}}\right]^2 \left[ \int_{\mathbb{R}^d} K^2(v) dv - Ch^d \right]$$

so that

$$\liminf_{n \rightarrow \infty} \mathbb{E}(\|r_n(\widehat{f}_n - f_n)\|_{2,\mu}^2) \geq mc^2 \int_{\mathbb{R}^d} K^2(v) dv > 0$$



which proves (3) in this case. Assume now that (H3b) holds and let  $R > 0$  be so large that  $\int_{B(0,R)} K^2(v)dv > 0$ . For  $n$  large enough, inequality (5) thus entails

$$\begin{aligned} \forall x \in B(x_0, \delta/2), \quad \mathbb{E}|Z(x, h)|^2 &\geq h^{-d} \int_{B(0,R)} K^2(v) f(x - hv) dv - C^2 \\ &\geq mh^{-d} \end{aligned} \quad (6)$$

where  $m$  is a positive constant. Apply (1), (4) and (6) to get for  $n$  large enough

$$\begin{aligned} \mathbb{E}(\|r_n(\widehat{f}_n - f_n)\|_{p,\mu}^p) &\geq \left[\frac{r_n}{n}\right]^p \int_{B(x_0, \delta/2)} (n\mathbb{E}|Z(x, h)|^2)^{p/2} d\mu(x) \\ &\geq m^{p/2} \left[\frac{r_n}{\sqrt{nh^d}}\right]^p \mu(B(x_0, \delta/2)). \end{aligned}$$

As a consequence

$$\liminf_{n \rightarrow \infty} \mathbb{E}(\|r_n(\widehat{f}_n - f_n)\|_{p,\mu}^p) \geq m^{p/2} c^p \mu(B(x_0, \delta/2)) > 0$$

which concludes the proof of (3). We now show the uniform integrability of the sequence of random variables  $(\|r_n(\widehat{f}_n - f_n)\|_{p,\mu}^p)$ . For this purpose, it is enough to prove that

$$\mathbb{E}(\|r_n(\widehat{f}_n - f_n)\|_{p,\mu}^{2p}) = O(1).$$

To this end we note that

$$\mathbb{E}(\|r_n(\widehat{f}_n - f_n)\|_{p,\mu}^{2p}) = \left[\frac{r_n}{n}\right]^{2p} \int_{\mathbb{R}^d \times \mathbb{R}^d} \mathbb{E} \left| \sum_{i,j=1}^n Z_i(x, h) Z_j(y, h) \right|^p d\mu(x) d\mu(y).$$

Hölder's inequality and the triangle inequality entail

$$\mathbb{E}(\|r_n(\widehat{f}_n - f_n)\|_{p,\mu}^{2p}) = \left[\frac{r_n}{n}\right]^{2p} 2^{p-1} [2^{p-1} [I_{1,n} + I_{2,n}] + I_{3,n}] \quad (7)$$

$$\text{with } I_{1,n} = \int_{\mathbb{R}^d \times \mathbb{R}^d} |n\mathbb{E}(Z(x, h)Z(y, h))|^p d\mu(x) d\mu(y),$$

$$I_{2,n} = \int_{\mathbb{R}^d \times \mathbb{R}^d} \mathbb{E} \left| \sum_{i=1}^n Z_i(x, h) Z_i(y, h) - \mathbb{E}(Z_i(x, h) Z_i(y, h)) \right|^p d\mu(x) d\mu(y)$$

$$\text{and } I_{3,n} = \int_{\mathbb{R}^d \times \mathbb{R}^d} \mathbb{E} \left| \sum_{\substack{i,j=1 \\ i \neq j}}^n Z_i(x, h) Z_j(y, h) \right|^p d\mu(x) d\mu(y).$$

The sequence  $I_{1,n}$  is controlled by using the Cauchy-Schwarz inequality and Lemma 2:

$$I_{1,n} \leq n^p \left[ \int_{\mathbb{R}^d} (\mathbb{E}|Z(x, h)|^2)^{p/2} d\mu(x) \right]^2 = O(n^p h^{-dp}). \quad (8)$$

To control  $I_{2,n}$ , we define  $D_i(x, y, h) = Z_i(x, h)Z_i(y, h) - \mathbb{E}(Z_i(x, h)Z_i(y, h))$  for  $i = 1, \dots, n$ ; the  $D_i(x, y, h)$ ,  $i = 1, \dots, n$  are independent copies of the centered random variable  $D(x, y, h) = Z(x, h)Z(y, h) - \mathbb{E}(Z(x, h)Z(y, h))$ . By Rosenthal's inequality (Rosenthal, 1970):

$$\mathbb{E} \left| \sum_{i=1}^n D_i(x, y, h) \right|^p \leq B_p \max \left[ n \mathbb{E}|D(x, y, h)|^p, (n \mathbb{E}|D(x, y, h)|^2)^{p/2} \right]. \quad (9)$$

Using together (9) and Lemma 3 entails

$$I_{2,n} = O(nh^{-2d(p-1)} + n^{p/2}h^{-d(3p-2)/2}).$$

Since  $nh^d \rightarrow \infty$  and

$$nh^{-2d(p-1)} + n^{p/2}h^{-d(3p-2)/2} = n^{p/2}h^{-d(3p-2)/2} \left[ 1 + (nh^d)^{1-p/2} \right]$$

we obtain

$$I_{2,n} = O(n^{p/2}h^{-d(3p-2)/2}). \quad (10)$$

We now turn to controlling  $I_{3,n}$ . To this end, we define

$$\begin{aligned} Y(x, y, h, u, v) &= [K_h(x - u) - \mathbb{E}(K_h(x - X))][K_h(y - v) - \mathbb{E}(K_h(y - X))] \\ &+ [K_h(y - u) - \mathbb{E}(K_h(y - X))][K_h(x - v) - \mathbb{E}(K_h(x - X))] \end{aligned}$$

and we note that for any  $x, y \in \mathbb{R}^d$ , we have

$$\sum_{\substack{i,j=1 \\ i \neq j}}^n Z_i(x, h)Z_j(y, h) = \sum_{1 \leq i < j \leq n} Y(x, y, h, X_i, X_j).$$

Moreover, the function  $(u, v) \mapsto Y(x, y, h, u, v)$  is symmetric and is such that

$$\mathbb{E}|Y(x, y, h, X_1, X_2)|^p < \infty \quad \text{and} \quad \mathbb{E}(Y(x, y, h, X_1, X_2)|X_1) = 0 \quad \text{almost surely.}$$

An analogue of Rosenthal's inequality for symmetric statistics (see *e.g.* Theorem 4 in Ibragimov and Sharakhmetov, 1999) thus yields

$$\begin{aligned} & \mathbb{E} \left| \sum_{1 \leq i < j \leq n} Z_i(x, h) Z_j(y, h) \right|^p \\ & \leq B'_p \max \left[ n^2 \mathbb{E} |Y(x, y, h, X_1, X_2)|^p, n^{p/2+1} \mathbb{E} (\mathbb{E}([Y(x, y, h, X_1, X_2)]^2 | X_1))^{p/2}, \right. \\ & \quad \left. n^p (\mathbb{E} |Y(x, y, h, X_1, X_2)|^2)^{p/2} \right] \end{aligned} \quad (11)$$

where  $B'_p$  is a positive constant depending only on  $p$ . Use Hölder's inequality to obtain

$$\begin{aligned} |Y(x, y, h, X_1, X_2)|^p & \leq 2^{p-1} [|Z_1(x, h)|^p |Z_2(y, h)|^p + |Z_1(y, h)|^p |Z_2(x, h)|^p], \\ \mathbb{E}([Y(x, y, h, X_1, X_2)]^2 | X_1) & \leq 2 [|Z_1(x, h)|^2 \mathbb{E}|Z_2(y, h)|^2 + |Z_1(y, h)|^2 \mathbb{E}|Z_2(x, h)|^2] \\ \text{and } \mathbb{E}|Y(x, y, h, X_1, X_2)|^2 & \leq 2 [\mathbb{E}|Z_1(x, h)|^2 \mathbb{E}|Z_2(y, h)|^2 + \mathbb{E}|Z_2(x, h)|^2 \mathbb{E}|Z_1(y, h)|^2]. \end{aligned}$$

Applying Hölder's inequality once more and using Lemma 2, we obtain

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} \mathbb{E}|Y(x, y, h, X_1, X_2)|^p d\mu(x) d\mu(y) = O(n^2 h^{-2d(p-1)}), \quad (12)$$

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} \mathbb{E}(\mathbb{E}([Y(x, y, h, X_1, X_2)]^2 | X_1))^{p/2} d\mu(x) d\mu(y) = O(n^{p/2+1} h^{-d(3p-2)/2}) \quad (13)$$

$$\text{and } \int_{\mathbb{R}^d \times \mathbb{R}^d} (\mathbb{E}|Y(x, y, h, X_1, X_2)|^2)^{p/2} d\mu(x) d\mu(y) = O(n^p h^{-dp}). \quad (14)$$

Since  $nh^d \rightarrow \infty$  and

$$n^2 h^{-2d(p-1)} + n^{p/2+1} h^{-d(3p-2)/2} + n^p h^{-dp} = n^p h^{-dp} \left[ 1 + (nh^d)^{2-p} + (nh^d)^{1-p/2} \right]$$

we get, collecting (11), (12), (13) and (14), the bound

$$I_{3,n} = O(n^p h^{-dp}). \quad (15)$$

Use together (7), (8), (10) and (15) to obtain

$$\mathbb{E}(\|r_n(\widehat{f}_n - f_n)\|_{p,\mu}^{2p}) = O\left(\left[\frac{r_n}{\sqrt{nh^d}}\right]^{2p} [1 + h^d]\right) = O(1).$$

Now, if  $r_n(\widehat{f}_n - f_n)$  did converge in distribution to a Borel measurable random element in  $L^p(\mathbb{R}^d, \mu)$ , then the weak limit would be 0 by Theorem 1. This would entail that the sequence

of random variables  $(\|r_n(\widehat{f}_n - f_n)\|_{p,\mu}^p)$  converges to 0 in probability; since this sequence is uniformly integrable, this implies  $\mathbb{E}\|r_n(\widehat{f}_n - f_n)\|_{p,\mu}^p \rightarrow 0$ , which, in view of (3), is a contradiction. Finally, if  $r_n/\sqrt{nh^d} \rightarrow \infty$ , the existence of a weak Borel measurable limit for  $r_n(\widehat{f}_n - f_n)$  in  $L^p(\mathbb{R}^d, \mu)$  would entail, by Slutsky's lemma, that  $\|\sqrt{nh^d}(\widehat{f}_n - f_n)\|_{p,\mu}$  converges in probability to 0, which is a contradiction. The proof is complete.  $\blacksquare$

We now turn to the case  $p \in [1, 2)$  which is different since Rosenthal-type inequalities cannot be applied. We introduce the following classical assumption in kernel density estimation:

(C1) The function  $K$  has a compact support.

We also introduce an integrability condition:

(C2) There exist  $\alpha \geq 0$  and  $R > 0$  such that if  $\|\cdot\|$  is the standard Euclidean norm on  $\mathbb{R}^d$ :

$$\sup_{x \in \mathbb{R}^d} f(x) [\mathbb{I}_{\{\|x\| \leq 2R\}} + \|x\|^\alpha \mathbb{I}_{\{\|x\| > R/2\}}] < \infty \quad \text{and} \quad \int_{\|x\| > R} \|x\|^{-\alpha/2} d\mu(x) < \infty.$$

If  $f$  is bounded on  $\mathbb{R}^d$ , then hypothesis (C2) is for instance satisfied if  $x \mapsto \|x\|^{2d+\varepsilon} f(x)$  is bounded in a neighborhood of infinity for some  $\varepsilon > 0$ : this latter condition can also be found in Horváth and Zitikis (2004). We can now state the analogue of Theorem 2 in the case  $p \in [1, 2)$ .

**Theorem 3.** *Consider the case  $p \in [1, 2)$ . Let  $(r_n)$  be a nonrandom sequence of positive real numbers. Assume that  $K^2$  is integrable on  $\mathbb{R}^d$  and that (C1) and (C2) hold.*

- *If  $r_n/\sqrt{nh^d} \rightarrow 0$ , then  $r_n(\widehat{f}_n - f_n) \rightarrow 0$  in  $L^p(\mathbb{R}^d, \mu)$ .*

*Assume now that for some  $\delta > 0$ ,  $K^{2+\delta}$  is integrable on  $\mathbb{R}^d$ , that (H1), (H2) and (H3b) hold and that  $nh^d \rightarrow \infty$ .*

- *If  $r_n/\sqrt{nh^d} \rightarrow c \in (0, \infty]$ , then  $r_n(\widehat{f}_n - f_n)$  does not converge to any Borel measurable random element in  $L^p(\mathbb{R}^d, \mu)$ .*

**Proof of Theorem 3.** We start by proving the first statement in the case  $p = 1$ ; to this end, we shall show that  $\mathbb{E}(r_n\|\widehat{f}_n - f_n\|_{1,\mu}) \rightarrow 0$ . Note that by the Cauchy-Schwarz inequality,

$$\mathbb{E}(r_n\|\widehat{f}_n - f_n\|_{1,\mu}) \leq \frac{r_n}{\sqrt{nh^d}} \int_{\mathbb{R}^d} \sqrt{h^d \mathbb{E} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_i(x, h) \right|^2} d\mu(x).$$

Since the  $Z_i(x, h)$ ,  $i = 1, \dots, n$  are independent copies of the centered random variable  $Z(x, h)$  which is such that  $\mathbb{E}|Z(x, h)|^2 \leq \mathbb{E}|K_h^2(x - X)|$ , we get

$$\mathbb{E}(r_n \|\widehat{f}_n - f_n\|_{1, \mu}) \leq \frac{r_n}{\sqrt{nh^d}} J_n \quad \text{with} \quad J_n := \int_{\mathbb{R}^d} \sqrt{h^d \mathbb{E}|K_h^2(x - X)|} d\mu(x).$$

It is then enough to show that  $J_n = \mathcal{O}(1)$ . A change of variables yields

$$J_n = \int_{\mathbb{R}^d} \sqrt{\int_{\mathbb{R}^d} K^2(v) f(x - hv) dv} d\mu(x).$$

Write now

$$\begin{aligned} J_n &= J_{n,1} + J_{n,2} & (16) \\ \text{with } J_{n,1} &= \int_{\|x\| \leq R} \sqrt{\int_{\mathbb{R}^d} K^2(v) f(x - hv) dv} d\mu(x) \\ \text{and } J_{n,2} &= \int_{\|x\| > R} \sqrt{\int_{\mathbb{R}^d} K^2(v) f(x - hv) dv} d\mu(x). \end{aligned}$$

Since  $K$  has a compact support, we have for  $n$  large enough

$$J_{n,1} \leq \sqrt{\int_{\mathbb{R}^d} K^2(v) dv} \sup_{\|x\| \leq 2R} f(x) \int_{\|x\| \leq R/2} d\mu(x) < \infty \quad (17)$$

and

$$\begin{aligned} J_{n,2} &\leq \sqrt{\sup_{\|x\| > R/2} \|x\|^\alpha f(x)} \int_{\|x\| > R} \sqrt{\int_{\mathbb{R}^d} K^2(v) \|x - hv\|^{-\alpha} dv} d\mu(x) \\ &\leq 2^{\alpha/2} \sqrt{\int_{\mathbb{R}^d} K^2(v) dv} \sup_{\|x\| > R/2} \|x\|^\alpha f(x) \int_{\|x\| > R} \|x\|^{-\alpha/2} d\mu(x) < \infty. \end{aligned} \quad (18)$$

Collecting (16), (17) and (18) proves that  $J_n = \mathcal{O}(1)$ , which is the desired result. To show the first statement for any  $p \in (1, 2)$ , note that for every  $1 \leq p_0 < p_1 < \infty$  and every  $p_\theta \in (p_0, p_1)$ , if  $\theta \in (0, 1)$  is such that

$$\frac{1}{p_\theta} = \frac{1 - \theta}{p_0} + \frac{\theta}{p_1}$$

then Hölder's inequality entails

$$\forall H \in L^{p_0}(\mathbb{R}^d, \mu) \cap L^{p_1}(\mathbb{R}^d, \mu), \quad \|H\|_{p_\theta, \mu} \leq \|H\|_{p_0, \mu}^{1-\theta} \|H\|_{p_1, \mu}^\theta. \quad (19)$$

Use then inequality (19) with  $p_0 = 1$ ,  $p_\theta = p$  and  $p_1 = 2$  to obtain

$$\|r_n(\widehat{f}_n - f_n)\|_{p,\mu} \leq \|r_n(\widehat{f}_n - f_n)\|_{1,\mu}^{1-\theta} \|r_n(\widehat{f}_n - f_n)\|_{2,\mu}^\theta$$

for some  $\theta \in (0, 1)$ . Since  $\|r_n(\widehat{f}_n - f_n)\|_{1,\mu} \rightarrow 0$  and (by Theorem 2)  $\|r_n(\widehat{f}_n - f_n)\|_{2,\mu} \rightarrow 0$  in probability, we obtain  $\|r_n(\widehat{f}_n - f_n)\|_{p,\mu} \rightarrow 0$  in probability, which completes the proof of the first statement.

We now turn to the second statement: apply inequality (19) with  $p_0 = p$ ,  $p_\theta = 2$  and  $p_1 = 2 + \delta$  to obtain

$$\|r_n(\widehat{f}_n - f_n)\|_{2,\mu} \leq \|r_n(\widehat{f}_n - f_n)\|_{p,\mu}^{1-\theta} \|r_n(\widehat{f}_n - f_n)\|_{2+\delta,\mu}^\theta$$

for some  $\theta \in (0, 1)$ . In the case  $r_n/\sqrt{nh^d} \rightarrow c \in (0, \infty)$ , it holds that

$$\mathbb{E}\|r_n(\widehat{f}_n - f_n)\|_{2+\delta,\mu}^{4+2\delta} = O(1),$$

see the proof of Theorem 2. Consequently  $\|r_n(\widehat{f}_n - f_n)\|_{2+\delta,\mu} = O_{\mathbb{P}}(1)$  and thus

$$\|r_n(\widehat{f}_n - f_n)\|_{2,\mu} = O_{\mathbb{P}}\left(\|r_n(\widehat{f}_n - f_n)\|_{p,\mu}^{1-\theta}\right).$$

Hence if we had  $\|r_n(\widehat{f}_n - f_n)\|_{p,\mu} \rightarrow 0$  in probability, then we would have  $\|r_n(\widehat{f}_n - f_n)\|_{2,\mu} \rightarrow 0$  in probability as well, which is a contradiction. We can finally handle the case  $r_n/\sqrt{nh^d} \rightarrow \infty$  as in the proof of Theorem 2: the proof is complete.  $\blacksquare$

**Remark.** When  $\mu$  is a finite measure on  $\mathbb{R}^d$ , Hölder's inequality entails for any  $p \in [1, 2)$

$$r_n\|\widehat{f}_n - f_n\|_{p,\mu} \leq r_n\|\widehat{f}_n - f_n\|_{2,\mu} \left[\mu(\mathbb{R}^d)\right]^{1-p/2}$$

so that by Theorem 2, we have  $r_n\|\widehat{f}_n - f_n\|_{p,\mu} \rightarrow 0$  provided  $r_n/\sqrt{nh^d} \rightarrow 0$ . Hypotheses (C1) and (C2) thus need not be satisfied in this particular setting for the first statement of Theorem 3 to hold.

## Appendix: preliminary results and their proofs

Recall from the proof of Theorem 1 the notation  $Z(x, h) = K_h(x - X) - \mathbb{E}(K_h(x - X))$ . The first result is a simple bound we shall use in the proof of Theorem 1.

**Lemma 1.** *Let  $\varphi$  be a bounded Borel measurable function on  $\mathbb{R}^d$ . Then there exists a positive constant  $C$  such that for any  $n$ ,*

$$\left| \int_{\mathbb{R}^d} Z(x, h) \varphi(x) dx \right| \leq C \text{ almost surely.}$$

**Proof of Lemma 1.** Let  $B = \sup_{\mathbb{R}^d} |\varphi|$ . Remark that

$$\left| \int_{\mathbb{R}^d} K_h(x - X) \varphi(x) dx \right| \leq \int_{\mathbb{R}^d} K(v) |\varphi(X + hv)| dv \leq B \text{ almost surely.}$$

Using the triangle inequality concludes the proof of the result. ■

The second result is a technical result we shall use repeatedly in the proof of the first part of Theorem 2.

**Lemma 2.** *Let  $\varphi$  be a bounded Borel measurable function on  $\mathbb{R}^d$  and  $p, q \in [1, \infty)$ . Assume that  $K^p$  is integrable on  $\mathbb{R}^d$ ; if either  $p \notin \{1, 2\}$  or  $q > 1$ , assume further that (H1) holds. Then we have*

$$\int_{\mathbb{R}^d} (\mathbb{E}|Z(x, h)|^p)^q \varphi(x) dx = O(h^{-dq(p-1)}).$$

**Proof of Lemma 2.** The case  $p = q = 1$  is an immediate consequence of Lemma 1. When  $p = 2$  and  $q = 1$ , we have

$$\mathbb{E}|Z(x, h)|^2 \leq \mathbb{E}(K_h^2(x - X)). \quad (20)$$

Let  $B = \sup_{\mathbb{R}^d} |\varphi|$  and note that for any  $r \geq 1$ , a change of variables entails

$$\int_{\mathbb{R}^d} \mathbb{E}(K_h^r(x - X)) \varphi(x) dx \leq B h^{-d(r-1)} \int_{\mathbb{R}^d} K^r(v) dv = O(h^{-d(r-1)}). \quad (21)$$

It is then a consequence of (20) and (21) with  $r = 2$  that

$$\int_{\mathbb{R}^d} \mathbb{E}|Z(x, h)|^2 \varphi(x) dx \leq \int_{\mathbb{R}^d} \mathbb{E}(K_h^2(x - X)) \varphi(x) dx = O(h^{-d})$$

which is the result in this case. When  $p \notin \{1, 2\}$  or  $q > 1$ , we start by remarking that Hölder's inequality yields

$$\begin{aligned} (\mathbb{E}|Z(x, h)|^p)^q &\leq 2^{q(p-1)} [\mathbb{E}(K_h^p(x - X)) + (\mathbb{E}(K_h(x - X)))^p]^q \\ &\leq 2^{pq-1} [(\mathbb{E}(K_h^p(x - X)))^q + (\mathbb{E}(K_h(x - X)))^{pq}]. \end{aligned}$$

The proof shall then be complete provided we prove that for any  $r \geq 1$  and  $s > 1$ , we have

$$\int_{\mathbb{R}^d} (\mathbb{E}(K_h^r(x - X)))^s \varphi(x) dx = O(h^{-ds(r-1)}). \quad (22)$$

To this end, note that since  $C = \sup_{\mathbb{R}^d} f < \infty$ ,

$$\begin{aligned} (\mathbb{E}(K_h^r(x - X)))^{s-1} &= \left[ h^{-d(r-1)} \int_{\mathbb{R}^d} K^r(v) f(x - hv) dv \right]^{s-1} \\ &\leq C^{s-1} \left[ \int_{\mathbb{R}^d} K^r(v) dv \right]^{s-1} h^{-d(r-1)(s-1)}. \end{aligned}$$

Consequently, there exists a constant  $C' > 0$  such that

$$\int_{\mathbb{R}^d} (\mathbb{E}(K_h^r(x - X)))^s \varphi(x) dx \leq C' h^{-d(r-1)(s-1)} \int_{\mathbb{R}^d} \mathbb{E}(K_h^r(x - X)) dx.$$

Using (21) yields (22) and completes the proof.  $\blacksquare$

Recall the notation  $D(x, y, h) = Z(x, h)Z(y, h) - \mathbb{E}(Z(x, h)Z(y, h))$  from the proof of Theorem 2. The third lemma is a key bound to prove the second part of this result.

**Lemma 3.** *Let  $\varphi$  be a bounded Borel measurable function on  $\mathbb{R}^d$  and  $p, q \in [1, \infty)$ . Assume that  $K^p$  is integrable on  $\mathbb{R}^d$ ; if either  $p \notin \{1, 2\}$  or  $q > 1$ , assume further that (H1) and (H2) hold. Then we have*

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} (\mathbb{E}|D(x, y, h)|^p)^q \varphi(x) \varphi(y) dx dy = O(h^{-d[2q(p-1)+q-1]}).$$

**Proof of Lemma 3.** In the case  $p = q = 1$ , the result is an immediate consequence of the triangle inequality and Lemma 1:

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} \mathbb{E}|D(x, y, h)| \varphi(x) \varphi(y) dx dy \leq 2 \mathbb{E} \left[ \int_{\mathbb{R}^d} Z(x, h) \varphi(x) dx \right]^2 = O(1).$$

When  $p = 2$  and  $q = 1$ , we have

$$\mathbb{E}|D(x, y, h)|^2 \leq \mathbb{E}(K_h^2(x - X)K_h^2(y - X)).$$

As a consequence

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} \mathbb{E}|D(x, y, h)|^2 \varphi(x) \varphi(y) dx dy \leq \mathbb{E} \left[ \int_{\mathbb{R}^d} K_h^2(x - X) \varphi(x) dx \right]^2.$$



Letting  $B = \sup_{\mathbb{R}^d} |\varphi|$  and using a change of variables entails

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} \mathbb{E}|D(x, y, h)|^2 \varphi(x) \varphi(y) dx dy \leq B^2 \left[ \int_{\mathbb{R}^d} K^2(v) dv \right]^2 h^{-2d} = O(h^{-2d})$$

which is the result in this case. When  $p \notin \{1, 2\}$  or  $q > 1$ , Hölder's inequality yields

$$\begin{aligned} (\mathbb{E}|D(x, y, h)|^p)^q &\leq 2^{q(p-1)} [\mathbb{E}|Z(x, h)Z(y, h)|^p + (\mathbb{E}|Z(x, h)Z(y, h)|)^p]^q \\ &\leq 2^{pq-1} [(\mathbb{E}|Z(x, h)Z(y, h)|^p)^q + (\mathbb{E}|Z(x, h)Z(y, h)|)^{pq}]. \end{aligned}$$

The result shall then be proven provided we show that for any  $r \geq 1$  and  $s > 1$ , we have

$$\int_{\mathbb{R}^d} (\mathbb{E}|Z(x, h)Z(y, h)|^r)^s \varphi(x) \varphi(y) dx dy = O(h^{-d[2s(r-1)+s-1]}). \quad (23)$$

To this aim, we use Hölder's inequality to get

$$\begin{aligned} (\mathbb{E}|Z(x, h)Z(y, h)|^r)^s &\leq 4^{s(r-1)} [\mathbb{E}(K_h^r(x-X)K_h^r(y-X))] \\ &\quad + \mathbb{E}(K_h^r(x-X))(\mathbb{E}(K_h^r(y-X)))^r \\ &\quad + \mathbb{E}(K_h^r(y-X))(\mathbb{E}(K_h^r(x-X)))^r \\ &\quad + (\mathbb{E}(K_h^r(x-X)))^r(\mathbb{E}(K_h^r(y-X)))^r \\ &\leq 4^{rs-1} [(\mathbb{E}(K_h^r(x-X)K_h^r(y-X)))^s \\ &\quad + (\mathbb{E}(K_h^r(x-X)))^s(\mathbb{E}(K_h^r(y-X)))^{rs} \\ &\quad + (\mathbb{E}(K_h^r(y-X)))^s(\mathbb{E}(K_h^r(x-X)))^{rs} \\ &\quad + (\mathbb{E}(K_h^r(x-X)))^{rs}(\mathbb{E}(K_h^r(y-X)))^{rs}]. \end{aligned} \quad (24)$$

Note further that if  $C = \sup_{\mathbb{R}^d} (K^r f) < \infty$  then

$$\begin{aligned} (\mathbb{E}(K_h^r(x-X)K_h^r(y-X)))^{s-1} &\leq C^{s-1} h^{-dr(s-1)} \left[ \int_{\mathbb{R}^d} K_h^r(x-u) du \right]^{s-1} \\ &= C^{s-1} \left[ \int_{\mathbb{R}^d} K^r(v) dv \right]^{s-1} h^{-d(2r-1)(s-1)}. \end{aligned} \quad (25)$$

Moreover,

$$\int_{\mathbb{R}^d} \mathbb{E}(K_h^r(x-X)K_h^r(y-X)) \varphi(x) \varphi(y) dx dy \leq B^2 \mathbb{E} \left[ \int_{\mathbb{R}^d} K_h^r(x-X) dx \right]^2 = O(h^{-2d(r-1)}). \quad (26)$$

Collecting (25) and (26) entails for any  $r \geq 1$  and  $s > 1$ :

$$\int_{\mathbb{R}^d} (\mathbb{E}(K_h^r(x-X)K_h^r(y-X)))^s \varphi(x) \varphi(y) dx dy = O(h^{-d[2s(r-1)+s-1]}). \quad (27)$$

Using together (24), (27) and Lemma 1 yields (23): the proof is complete.  $\blacksquare$

## References

- Beirlant, J., Mason, D.M. (1995) On the asymptotic normality of  $L_p$ -norms of empirical functionals, *Mathematical Methods of Statistics* **4**: 1–19.
- Bickel, P.J., Rosenblatt, M. (1973) On some global measures of the deviations of density function estimates, *Annals of Statistics* **1**: 1071–1095.
- Csörgő, M., Horváth, L. (1988) Central limit theorems for  $L_p$ -norms of density estimators, *Probability Theory and Related Fields* **80**: 269–291.
- Devroye, L., Györfi, L. (1985) *Nonparametric density estimation: the  $L_1$  view*, John Wiley and Sons, New York.
- Horváth, L. (1991) On  $L_p$ -norms of multivariate density estimators, *Annals of Statistics* **19**(4): 1933–1949.
- Horváth, L., Zitikis, R. (2004) Asymptotics of the  $L_p$ -norms of density estimators in the first-order autoregressive models, *Statistics and Probability Letters* **66**: 91–103.
- Ibragimov, R., Sharakhmetov, S. (1999) Analogues of Khintchine, Marcinkiewicz-Zygmund and Rosenthal inequalities for symmetric statistics, *Scandinavian Journal of Statistics* **26**: 621–633.
- Nishiyama, Y. (2011) Impossibility of weak convergence of kernel density estimators to a non-degenerate law in  $L_2(\mathbb{R}^d)$ , *Journal of Nonparametric Statistics* **23**(1): 129–135.
- Parzen, E. (1962) On estimation of a probability density function and mode, *Annals of Mathematical Statistics* **33**(3): 1065–1076.
- Rosenblatt, M. (1956) Remarks on some nonparametric estimates of a density function, *Annals of Mathematical Statistics* **27**(3): 832–837.
- Rosenthal, H.P. (1970) On the subspaces of  $L^p$  ( $p > 2$ ) spanned by sequences of independent random variables, *Israel Journal of Mathematics* **8**: 273–303.
- Ruymgaart, F.H. (1998) A note on weak convergence of density estimators in Hilbert spaces, *Statistics* **30**: 331–343.

Van der Vaart, A.W., Wellner, J.A. (1996) *Weak convergence and empirical processes with applications to statistics*, Springer.