HAL open science

On the weak convergence of kernel density estimators in Lp spaces

Gilles Stupfler

To cite this version:

Gilles Stupfler. On the weak convergence of kernel density estimators in Lp spaces. Journal of Nonparametric Statistics, 2014, 26 (4), pp.721-735. hal-00957179v2

HAL Id: hal-00957179

https://hal.science/hal-00957179v2

Submitted on 6 Sep 2015

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 $\widehat{f}_n$ 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 $f_n(x) = \mathbb{E}(\widehat{f}_n(x))$ and (r_n) is any nonrandom sequence of positive real numbers such that $r_n/\sqrt{n} \to 0$ then if $r_n(\widehat{f}_n - f_n)$ 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 $h = h(n) \to 0$ as $n \to \infty$ and $K_h(u) = h^{-d}K(u/h)$ with $K : \mathbb{R}^d \to \mathbb{R}$ an integrable function whose integral over $\mathbb{R}^d$ is equal to 1. The random function $x \mapsto \widehat{f}_n(x)$ can be seen as the

empirical counterpart of the function $x \mapsto f_n(x) = \mathbb{E}[\widehat{f}_n(x)] = \mathbb{E}[K_h(x - X)]$ which is well-defined almost everywhere on $\mathbb{R}^d$ and integrable. It is well-known that under some conditions on K, the random process $x \mapsto \sqrt{nh^d}(\widehat{f}_n(x) - f_n(x))$ converges pointwise to a Gaussian distribution provided $nh^d \to \infty$, $f(x) > 0$ and f is continuous at x, see *e.g.* the discussion pp. 1069–1070 in Parzen (1962).

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 $\mathbb{R}^d$. Let further μ 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 $H : \mathbb{R}^d \to \mathbb{R}$ such that

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

Then, assuming for instance that $|K|^p$ is integrable on $\mathbb{R}^d$ and μ has a bounded Radon-Nikodym derivative with respect to the Lebesgue measure, 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).

However, none of these studies consider the convergence of $r_n(\widehat{f}_n - f_n)$ 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 Borel measurable weak limit for $r_n(\widehat{f}_n - f_n)$ in the usual $L^2(\mathbb{R}^d)$ space provided $r_n/\sqrt{n} \to 0$, thus generalising a result of Ruymgaart (1998). Our focus in this paper is to generalise 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 $1 \leq p < \infty$. In particular, we shall show that when μ has a bounded Radon-Nikodym derivative with respect to the Lebesgue measure then 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 μ. 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. A discussion concerning the convergence of the bias term $r_n(f_n - f)$ is given in Section 3. 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 μ 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 $f : L^p(\mathbb{R}^d, \mu) \to \mathbb{R}$ for which there exists $J \geq 1$ such that

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

where g is a continuous bounded real-valued function on $\mathbb{R}^J$ and $H_1, \ldots, H_J$ are bounded elements

of $L^q(\mathbb{R}^d,\mu)$. Observe first that for all $J \geq 1$ and all bounded $H_1,\ldots,H_J \in L^q(\mathbb{R}^d,\mu)$, we have that

$$\forall t_1,\ldots,t_J \in \mathbb{R},\ \forall \varphi \in L^p(\mathbb{R}^d,\mu),\ \sum_{i=1}^{J} t_i T_{H_i}(\varphi) = T_{\sum_{i=1}^{J} t_i H_i}(\varphi).$$

It is thus a consequence of the Cramér-Wold device that the random vectors $(T_{H_1}(G_1),\ldots,T_{H_J}(G_1))$ and $(T_{H_1}(G_2),\ldots,T_{H_J}(G_2))$ have the same distribution. Let ν_1 and ν_2 be the pushforward probability measures on $L^p(\mathbb{R}^d,\mu)$ induced by G_1 and G_2; it is then 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 μ 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 φ, $\psi \in L^p(\mathbb{R}^d,\mu)$ be such that $\varphi \neq \psi$. It is a corollary of the Hahn-Banach theorem that there exists a continuous linear form T on $L^p(\mathbb{R}^d,\mu)$ such that $T(\varphi) \neq T(\psi)$. Since $p \in [1,\infty)$ and μ is $\sigma-$finite, 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, there is a positive integer N such that:

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

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 $T_H(\varphi) \neq T_H(\psi)$. The proof is complete. ■

A key consequence of this result is that if for any bounded function $H \in L^q(\mathbb{R}^d,\mu)$, we have $T_H(r_n(\widehat{f}_n - f_n)) \to T_H(G)$ 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 generalises 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 then characterise 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 $K \in L^p(\mathbb{R}^d)$. If $r_n/\sqrt{n} \to 0$ 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

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

in probability. We now set $Z_i(x,h) = K_h(x - X_i) - \mathbb{E}(K_h(x - X))$ for $i = 1, \ldots, n$ and $Z(x,h) = K_h(x - X) - \mathbb{E}(K_h(x - X))$. For almost every $x \in \mathbb{R}^d$, $\mathbb{E}(K_h(x - X))$ is well-defined and finite so that the $Z_i(x,h)$, $i = 1, \ldots, n$ are independent copies of the centred random variable $Z(x,h)$. We may rewrite $\Delta_n(H)$ as

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

Observe that the random variables $W_{n,i}(H)$, $i = 1, \ldots, n$ are independent copies of the centred random variable $W_n(H) = \int_{\mathbb{R}^d} Z(x,h)H(x)d\mu(x)$. Lemma 1 thus entails

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

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

■

We point out that Theorem 1 is a generalisation of Theorem 2.1 in Nishiyama (2011), which was restricted to the case when $p = 2$ and μ is the Lebesgue measure. This result says that either the

process $r_n(\widehat{f}_n - f_n)$ converges to a degenerate limit or does not converge to a Borel measurable limit. We proceed by giving some insight on what can happen, depending on the sequence (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 μ 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 δ, 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. In particular, any one of these two conditions ensure that $f_n(x)$ is finite for every x. Hypothesis (H3a) holds in particular if μ 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 μ 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 $K \in L^p(\mathbb{R}^d)$; if $p > 2$, assume that (H1) holds and that $nh^d \to \infty$.*

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

Assume further that in case $p = 2$, condition (H1) holds and that (H3a) or (H3b) holds; when $p > 2$, assume that (H2) and (H3b) hold as well.

- *If $r_n/\sqrt{nh^d} \to c \in (0, \infty]$, 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 $r_n/\sqrt{nh^d} \to 0$. The first statement shall be proven

provided we show that $\mathbb{E}(\|r_n(\widehat{f}_n - f_n)\|_{p,\mu}^p) \to 0$. To this end, we note that

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

Observe that the $Z_i(x,h)$, $i = 1,\ldots,n$ are independent copies of the centred random variable $Z(x,h)$. Moreover, it is a consequence of Lemma 3 that the function $x \mapsto \mathbb{E}|Z(x,h)|^p$ is integrable and thus almost everywhere finite on $\mathbb{R}^d$. By Rosenthal's inequality (Rosenthal, 1970), we obtain for almost every $x \in \mathbb{R}^d$:

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

for some positive constant B_p which only depends on p. Besides, we have that $K \in L^1(\mathbb{R}^d) \cap L^p(\mathbb{R}^d) \subset L^2(\mathbb{R}^d)$ so that we may use together (1), (2) and Lemma 3 to get

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

Since

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

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

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

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

To this aim, we apply Rosenthal's inequality (Rosenthal, 1970) to bound the integrand in equation (1) from below: for almost every $x \in \mathbb{R}^d$,

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

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

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

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

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

When $p=2$ and (H3a) holds, using (1), (4) and (5) entails for some $m>0$

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

so that

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

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

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

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

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

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

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

To this end we write

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

Observe that by the Cauchy-Schwarz inequality,

$$|\mathbb{E}(Z(x,h)Z(y,h))|^2\leq\mathbb{E}|Z(x,h)|^2\mathbb{E}|Z(y,h)|^2.$$

Consequently, $|\mathbb{E}(Z(x,h)Z(y,h))|$ is finite for almost every $(x,y) \in \mathbb{R}^d \times \mathbb{R}^d$ by Lemma 3. We may thus use Lemma 2 to write

$$\begin{aligned}
\mathbb{E}(\|r_n(\widehat{f}_n - f_n)\|_{p,\mu}^{2p}) &= \left[\frac{r_n}{n}\right]^{2p} 2^{p-1}[2^{p-1}[I_{1,n} + I_{2,n}] + I_{3,n}] \qquad (7)\\
\text{with } I_{1,n} &= \int_{\mathbb{R}^d\times\mathbb{R}^d} |n\mathbb{E}(Z(x,h)Z(y,h))|^p \, d\mu(x)d\mu(y),\\
I_{2,n} &= \int_{\mathbb{R}^d\times\mathbb{R}^d} \mathbb{E}\left|\sum_{i=1}^n Z_i(x,h)Z_i(y,h) - \mathbb{E}(Z_i(x,h)Z_i(y,h))\right|^p d\mu(x)d\mu(y)\\
\text{and } I_{3,n} &= \int_{\mathbb{R}^d\times\mathbb{R}^d} \mathbb{E}\left|\sum_{\substack{i,j=1\\ i\neq j}}^n Z_i(x,h)Z_j(y,h)\right|^p d\mu(x)d\mu(y).
\end{aligned}$$

The sequence $I_{1,n}$ is controlled by using the Cauchy-Schwarz inequality and Lemma 3:

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

To control $I_{2,n}$, we let $D_i(x,y,h) = Z_i(x,h)Z_i(y,h) - \mathbb{E}(Z(x,h)Z(y,h))$ for $i = 1,\ldots,n$. For almost every $(x,y) \in \mathbb{R}^d \times \mathbb{R}^d$, the $D_i(x,y,h)$, $i = 1,\ldots,n$ are well-defined independent copies of the centred random variable $D(x,y,h) = Z(x,h)Z(y,h) - \mathbb{E}(Z(x,h)Z(y,h))$. Furthermore, it is a consequence of Lemma 4 that the function $(x,y) \mapsto \mathbb{E}|D(x,y,h)|^p$ is integrable and thus almost everywhere finite on $\mathbb{R}^d \times \mathbb{R}^d$. By Rosenthal's inequality (Rosenthal, 1970), we obtain for almost every $(x,y) \in \mathbb{R}^d \times \mathbb{R}^d$:

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

Using together (9) and Lemma 4 entails

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

Noting that

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

we obtain

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

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 observe that

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

Moreover, the function $(u,v)\mapsto Y(x,y,h,u,v)$ is symmetric and is such that for almost every $(x,y)\in\mathbb{R}^d\times\mathbb{R}^d$,

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

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

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

where B'_p is a positive constant depending only on p. Use Lemma 2 to get

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

Applying Lemma 2 once more and using Lemma 3, we obtain

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

$$\int_{\mathbb{R}^d\times\mathbb{R}^d}\mathbb{E}(\mathbb{E}(|Y(x,y,h,X_1,X_2)|^2|X_1))^{p/2}d\mu(x)d\mu(y) = \mathrm{O}(h^{-d(3p-2)/2}) \tag{13}$$

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

Collecting (11), (12), (13) and (14), we get

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

Since

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

we have that

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

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

$$\mathbb{E}(\|r_n(\widehat{f}_n - f_n)\|_{p,\mu}^{2p}) = \mathrm{O}\left(\left[\frac{r_n}{\sqrt{nh^d}}\right]^{2p} \left[1 + n^{-1}(nh^d)^{1-p/2}\right]\right) = \mathrm{O}(1). \tag{16}$$

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 $\mathbb{E}\|r_n(\widehat{f}_n - f_n)\|_{p,\mu}^p \to 0$, which, in view of (3), is a contradiction. Finally, if $r_n/\sqrt{nh^d} \to \infty$, 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. ■

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 $\mathbb{R}^d$:

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

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 $K \in L^2(\mathbb{R}^d)$ and that (C1) and (C2) hold.*

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

Assume now that (H1), (H2) and (H3b) hold and that $nh^d \to \infty$.

- *If $r_n/\sqrt{nh^d} \to c \in (0,\infty]$, 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 $\mathbb{E}(\|r_n(\widehat{f}_n - f_n)\|_{1,\mu}) \to 0$. 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 $Z_i(x,h)$, $i = 1,\ldots,n$ are independent copies of the centred 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 $J_n = \mathrm{O}(1)$. A change of variables yields

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

Pick now $\alpha \geq 0$ and $R > 0$ as in (C2) and write

$$\begin{aligned} J_n &= J_{n,1} + J_{n,2} \qquad (17) \\ \text{with } J_{n,1} &= \int_{\|x\|\leq R} \sqrt{\int_{\mathbb{R}^d} K^2(v) f(x-hv) dv}\, d\mu(x) \\ \text{and } J_{n,2} &= \int_{\|x\|>R} \sqrt{\int_{\mathbb{R}^d} K^2(v) f(x-hv) dv}\, d\mu(x). \end{aligned}$$

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

$$J_{n,1} \leq \mu(B(0,R)) \sqrt{\int_{\mathbb{R}^d} K^2(v) dv \sup_{\|x\|\leq 2R} f(x)} < \infty \qquad (18)$$

and

$$\begin{aligned} J_{n,2} &\leq \sqrt{\sup_{\|x\|>R/2} \|x\|^\alpha f(x)} \int_{\|x\|>R} \sqrt{\int_{\mathbb{R}^d} K^2(v)\|x-hv\|^{-\alpha} dv}\, d\mu(x) \\ &\leq 2^{\alpha/2} \sqrt{\int_{\mathbb{R}^d} K^2(v)dv \sup_{\|x\|>R/2} \|x\|^\alpha f(x)} \int_{\|x\|>R} \|x\|^{-\alpha/2} d\mu(x) < \infty. \end{aligned} \tag{19}$$

Collecting (17), (18) and (19) proves that $J_n = \mathrm{O}(1)$, which is the desired result. To show the first statement for any $p \in (1,2)$, note that for all $1 \leq p_0 < p_1 < \infty$ 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

$$\forall H \in L^{p_0}(\mathbb{R}^d, \mu) \cap L^{p_1}(\mathbb{R}^d, \mu),\ \|H\|_{p_\theta,\mu} \leq \|H\|_{p_0,\mu}^{1-\theta} \|H\|_{p_1,\mu}^{\theta}. \tag{20}$$

Use then inequality (20) 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 $\|r_n(\widehat{f}_n - f_n)\|_{1,\mu} \to 0$ and (by Theorem 2) $\|r_n(\widehat{f}_n - f_n)\|_{2,\mu} \to 0$ in probability, we obtain $\|r_n(\widehat{f}_n - f_n)\|_{p,\mu} \to 0$ in probability, which completes the proof of the first statement.

We now turn to the second statement. Observe first that $K \in L^r(\mathbb{R}^d)$ for all $r \geq 1$ because K is a bounded integrable function. We may thus apply inequality (20) with $p_0 = p$, $p_\theta = 2$ and $p_1 = 2+\delta$ for an arbitrary $\delta > 0$ 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 $r_n/\sqrt{nh^d} \to c \in (0,\infty)$, it holds that

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

see (16) in the proof of Theorem 2. Consequently $\|r_n(\widehat{f}_n - f_n)\|_{2+\delta,\mu} = \mathrm{O}_{\mathbb{P}}(1)$ and thus

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

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. Therefore $\|r_n(\widehat{f}_n - f_n)\|_{p,\mu} \to 0$ in probability and thus $\|r_n(\widehat{f}_n - f_n)\|_{2,\mu} \to 0$ in probability as well, which is a contradiction. We can finally handle the case $r_n/\sqrt{nh^d} \to \infty$ as in the proof of Theorem 2: the proof is complete. ■

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

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

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

We close this section by mentioning that in the case $d = 1$, these results are very different from those which can be obtained for the empirical cumulative distribution process

$$\widehat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{\{X_i \leq x\}}.$$

It can indeed be shown that if F denotes the cumulative distribution function of X then the random process $\sqrt{n}(\widehat{F}_n - F)$ converges to a centred Gaussian limit in any $L^2(\mathbb{R}, \mu)$ space provided μ is a finite measure on $\mathbb{R}$, see Theorem 1.8.4 and Example 1.8.6 in van der Vaart and Wellner (1996). Analogue results on the convergence of this empirical process in $L^p(\mathbb{R}, \mu)$, $p \in [1, \infty)$, may be obtained by using necessary and sufficient conditions such as the ones presented in van der Vaart and Wellner (1996), p.92 and Theorem 10.10 in Ledoux and Talagrand (1991).

3 The bias term

Our main results consider the centred random process $r_n(\widehat{f}_n - f_n)$. Of course, in statistical applications, the process of interest would rather be $r_n(\widehat{f}_n - f)$; observe that if the norm $\|r_n(f_n - f)\|_{p,\mu}$ of the bias term $r_n(f_n - f)$ converges to 0, then our main results also hold with $r_n(\widehat{f}_n - f_n)$ replaced by $r_n(\widehat{f}_n - f)$. When $d = 1$, the pointwise behavior of this bias term is well documented. It can be

analysed for instance under the assumption that f is $m \geq 2$ times continuously differentiable and K is such that

$$\forall j \in \{1, \ldots, m-1\}, \int_{\mathbb{R}} v^j K(v) dv = 0 \text{ and } \int_{\mathbb{R}} |v^m K(v)| dv < \infty.$$

See for instance Parzen (1962) when $m = 2$: in this case, it turns out that the bias term asymptotically vanishes when $r_n h^2 \to 0$. Interestingly, this is still true for the integrated bias term $\|r_n(f_n - f)\|_{p,\mu}$ for any $p \in [1, \infty)$ under some further assumptions on f, see Lemma 9 in Csörgő and Horváth (1988). When $p = 2$ and $m > 2$, this condition can be weakened to $r_n h^m \to 0$ provided the $m-$th order derivative of f is square integrable, see Theorem 24.2 p.346 in van der Vaart (1998) and the discussion in Nishiyama (2011).

Our aim here is to state a result on the integrated bias term $\|r_n(f_n - f)\|_{p,\mu}$ for an arbitrary dimension d under analogue assumptions on f and K. In order to achieve this, we need some additional notation. For $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$ and $\alpha = (\alpha_1, \ldots, \alpha_d) \in \mathbb{N}^d$, we introduce the multi-index notation

$$x^\alpha = \prod_{i=1}^d x_i^{\alpha_i}, \ \alpha! = \prod_{i=1}^d \alpha_i! \text{ and } |\alpha| = \sum_{i=1}^d \alpha_i.$$

We then introduce the class $\mathcal{K}_m$ of the integrable functions $K : \mathbb{R}^d \to \mathbb{R}$ which integrate to 1 over $\mathbb{R}^d$ and are such that

$$\forall \alpha \in \mathbb{N}^d, \qquad |\alpha| \in \{1, \ldots, m-1\} \Rightarrow \int_{\mathbb{R}^d} v^\alpha K(v) dv = 0$$
$$\text{and } |\alpha| = m \Rightarrow \int_{\mathbb{R}^d} |v^\alpha K(v)| dv < \infty.$$

We denote the space of the m times continuously differentiable real-valued functions on $\mathbb{R}^d$ by $\mathcal{C}^m(\mathbb{R}^d)$. Furthermore, for all $f \in \mathcal{C}^m(\mathbb{R}^d)$ and $k \leq m$, we denote the $k-$th order partial derivatives of f by

$$\partial_\alpha f = \frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1} \cdots \partial x_d^{\alpha_d}}, \ |\alpha| = k.$$

Our result is the following:

Proposition 2. *Assume that $f \in \mathcal{C}^m(\mathbb{R}^d)$ is such that all its $m-$th order derivatives belong to $L^p(\mathbb{R}^d)$. Then for all $K \in \mathcal{K}_m$ there exists a positive constant C such that for n large enough:*

$$\|f_n - f\|_{p,\mu} \leq C h^m.$$

Proof of Proposition 2. The proof is largely based on the proof of Theorem 24.1 p.345 in van der Vaart (1998). For all x, $v \in \mathbb{R}^d$, Taylor's formula with the remainder in an integral form entails:

$$\begin{aligned} f(x-hv)-f(x) &= \sum_{k=1}^{m-1}(-1)^k h^k \sum_{|\alpha|=k} \frac{\partial_\alpha f(x)}{\alpha!} v^\alpha + R_m(x, x-hv) \\ \text{where } R_m(a,b) &= m \sum_{|\alpha|=m} \frac{(b-a)^\alpha}{\alpha!} \int_0^1 (1-t)^{m-1} \partial_\alpha f(a+t(b-a))dt. \end{aligned}$$

Since $K \in \mathcal{K}_m$, we may thus write

$$(f_n - f)(x) = \int_{\mathbb{R}^d} [f(x-hv) - f(x)]K(v)dv = \int_{\mathbb{R}^d} R_m(x, x-hv)K(v)dv.$$

By Lemma 2, it follows that there exists a positive constant c_1 such that

$$\frac{\|f_n - f\|_{p,\mu}^p}{h^{pm}} \leq c_1 \sum_{|\alpha|=m} \int_{\mathbb{R}^d} \left| \int_{\mathbb{R}^d} \int_0^1 (1-t)^{m-1} \partial_\alpha f(x-thv) v^\alpha K(v) dt\, dv \right|^p d\mu(x). \tag{21}$$

Let $q \in (1, \infty]$ be such that $p^{-1} + q^{-1} = 1$. For any $\alpha \in \mathbb{N}^d$ such that $|\alpha| = m$, Hölder's inequality applied to the functions $(t,v) \mapsto (1-t)^{m-1}|\partial_\alpha f(x-thv)||v^\alpha K(v)|^{1/p}$ and $(t,v) \mapsto |v^\alpha K(v)|^{1/q}$ (in the case $p = 1$ and $q = \infty$, we set $1/q = 0$) entails

$$\begin{aligned} & \int_{\mathbb{R}^d} \left| \int_{\mathbb{R}^d} \int_0^1 (1-t)^{m-1} \partial_\alpha f(x-thv) v^\alpha K(v) dt\, dv \right|^p d\mu(x) \\ \leq & \left[\int_{\mathbb{R}^d} |v^\alpha K(v)| dv \right]^{p/q} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \int_0^1 (1-t)^{p(m-1)} |\partial_\alpha f(x-thv)|^p |v^\alpha K(v)| dt\, dv\, d\mu(x) \\ \leq & \; c_2 \left[\int_{\mathbb{R}^d} |v^\alpha K(v)| dv \right]^{1+p/q} \left[\int_{\mathbb{R}^d} |\partial_\alpha f(x)|^p dx \right] = c_3 < \infty \end{aligned} \tag{22}$$

where c_2 and c_3 are positive constants. Combining (21) and (22) and summing over α yields the desired result. ■

A consequence of Proposition 2 is that if $r_n h^m \to 0$, then $\|r_n(f_n - f)\|_{p,\mu} \to 0$. This generalises the aforementioned results of Csörgő and Horváth (1988) and van der Vaart (1998).

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

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

Proof of Lemma 1. Let $B = \sup_{\mathbb{R}^d} |\varphi|$. Remark that by a change of variables:

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

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

The second result is a classical consequence of Hölder's inequality and the triangle inequality for sums of real numbers.

Lemma 2. *Let $m \geq 2$ be an integer and $x_1, \ldots, x_m$ be real numbers. For any $p \in [1, \infty)$, we have that*

$$\left| \sum_{i=1}^m x_i \right|^p \leq m^{p-1} \sum_{i=1}^m |x_i|^p.$$

Proof of Lemma 2. When $p = 1$, this is just the triangle inequality. When $p \in (1, \infty)$, apply Hölder's inequality to the functions $\varphi : i \mapsto x_i$ and $\psi : i \mapsto 1$ on the space $\{1, \ldots, n\}$ endowed with the counting measure. ■

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

Lemma 3. *Let φ be a nonnegative bounded Borel measurable function on $\mathbb{R}^d$ and p, $q \in [1, \infty)$. Assume that $K \in L^p(\mathbb{R}^d)$; if either $p \notin \{1, 2\}$ or $q > 1$, assume further that (H1) holds. Then the function $x \mapsto (\mathbb{E}|Z(x,h)|^p)^q$ is integrable on $\mathbb{R}^d$ and we have*

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

Proof of Lemma 3. The case $p = q = 1$ is an immediate consequence of Lemma 1. When $p = 2$ and $q = 1$, we have

$$\int_{\mathbb{R}^d} \mathbb{E}|Z(x,h)|^2 \varphi(x)dx \leq \int_{\mathbb{R}^d} \mathbb{E}(K_h^2(x-X))\varphi(x)dx. \tag{23}$$

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

$$\int_{\mathbb{R}^d} \mathbb{E}|K_h(x-X)|^r \varphi(x)dx \leq Bh^{-d(r-1)} \int_{\mathbb{R}^d} |K(v)|^r dv = \mathrm{O}(h^{-d(r-1)}) \tag{24}$$

provided $K \in L^r(\mathbb{R}^d)$. In this case, the result is then a consequence of (23) and (24) with $r = 2$. When $p \notin \{1, 2\}$ or $q > 1$, we start by remarking that Lemma 2 yields

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

The proof shall then be complete provided we prove that for any $r \geq 1$ and $s > 1$, if $K \in L^r(\mathbb{R}^d)$ then $x \mapsto (\mathbb{E}|K_h(x-X)|^r)^s$ is integrable on $\mathbb{R}^d$ and we have

$$\int_{\mathbb{R}^d} (\mathbb{E}|K_h(x-X)|^r)^s \varphi(x)dx = \mathrm{O}(h^{-ds(r-1)}). \tag{25}$$

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

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

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

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

Using (24) yields (25) and completes the proof. ■

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 last lemma is a key bound to prove the second part of this result.

Lemma 4. *Let φ be a nonnegative bounded Borel measurable function on $\mathbb{R}^d$ and p, $q \in [1, \infty)$. Assume that $K \in L^p(\mathbb{R}^d)$ and (H1) holds; if either $p \notin \{1, 2\}$ or $q > 1$, assume further that (H2) holds. Then the function $(x,y) \mapsto (\mathbb{E}|D(x,y,h)|^p)^q$ is integrable on $\mathbb{R}^d \times \mathbb{R}^d$ and we have*

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

Proof of Lemma 4. In the case $p = q = 1$, the result is an immediate consequence of the triangle inequality and Lemma 1:

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

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

$$\mathbb{E}|D(x,y,h)|^2 \leq \mathbb{E}|Z(x,h)Z(y,h)|^2$$

because $D(x,y,h)$ is a centred random variable. As a consequence

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} \mathbb{E}|D(x,y,h)|^2\varphi(x)\varphi(y)dx\,dy \leq \mathbb{E}\left[\int_{\mathbb{R}^d} |Z(x,h)|^2\varphi(x)dx\right]^2. \tag{26}$$

Besides, if $B = \sup_{\mathbb{R}^d} \varphi$ then Lemma 2 and a change of variables yield

$$\begin{aligned}\int_{\mathbb{R}^d} |Z(x,h)|^2\varphi(x)dx &\leq 2\int_{\mathbb{R}^d} \left[K_h^2(x-X) + [\mathbb{E}(K_h(x-X))]^2\right]\varphi(x)dx \\ &\leq 2\left[Bh^{-d}\int_{\mathbb{R}^d} K^2(v)dv + \int_{\mathbb{R}^d} (\mathbb{E}|K_h(x-X)|)^2\varphi(x)dx\right]\end{aligned}$$

almost surely. The result thus follows from (25) (see the proof of Lemma 3) and (26). When $p \notin \{1,2\}$ or $q > 1$, we use Lemma 2 to get

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

The result shall then be proven provided we show that for any r, $s \geq 1$ the function $x \mapsto (\mathbb{E}|Z(x,h)Z(y,h)|^r)^s$ is integrable and we have that

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} (\mathbb{E}|Z(x,h)Z(y,h)|^r)^s\varphi(x)\varphi(y)dx\,dy = \mathrm{O}(h^{-d[2s(r-1)+s-1]}). \tag{27}$$

To this aim, we use Lemma 2 to get

$$\begin{aligned}
(\mathbb{E}|Z(x,h)Z(y,h)|^r)^s &\leq 4^{s(r-1)}\,[\mathbb{E}|K_h(x-X)K_h(y-X)|^r \\
&+ \mathbb{E}|K_h(x-X)|^r(\mathbb{E}|K_h(y-X)|)^r \\
&+ \mathbb{E}|K_h(y-X)|^r(\mathbb{E}|K_h(x-X)|)^r \\
&+ (\mathbb{E}|K_h(x-X)|)^r(\mathbb{E}|K_h(y-X)|)^r]^s \\
&\leq 4^{rs-1}\,[(\mathbb{E}|K_h(x-X)K_h(y-X)|^r)^s \\
&+ (\mathbb{E}|K_h(x-X)|^r)^s(\mathbb{E}|K_h(y-X)|)^{rs} \\
&+ (\mathbb{E}|K_h(y-X)|^r)^s(\mathbb{E}|K_h(x-X)|)^{rs} \\
&+ (\mathbb{E}|K_h(x-X)|)^{rs}(\mathbb{E}|K_h(y-X)|)^{rs}]\,. \qquad (28)
\end{aligned}$$

Note further that since $C = \sup_{\mathbb{R}^d}(|K|^r f) < \infty$ then for all $s > 1$

$$\begin{aligned}
(\mathbb{E}|K_h(x-X)K_h(y-X)|^r)^{s-1} &\leq C^{s-1}h^{-dr(s-1)}\left[\int_{\mathbb{R}^d}|K_h(x-u)|^r du\right]^{s-1} \\
&= C^{s-1}\left[\int_{\mathbb{R}^d}|K(v)|^r dv\right]^{s-1} h^{-d(2r-1)(s-1)}. \qquad (29)
\end{aligned}$$

Moreover,

$$\begin{aligned}
\int_{\mathbb{R}^d\times\mathbb{R}^d}\mathbb{E}|K_h(x-X)K_h(y-X)|^r\varphi(x)\varphi(y)dx\,dy &\leq B^2h^{-2d(r-1)}\left[\int_{\mathbb{R}^d}|K(v)|^r dv\right]^2 \\
&= \mathrm{O}(h^{-2d(r-1)}). \qquad (30)
\end{aligned}$$

Collecting (29) and (30) entails:

$$\int_{\mathbb{R}^d\times\mathbb{R}^d}(\mathbb{E}|K_h(x-X)K_h(y-X)|^r)^s\varphi(x)\varphi(y)dx\,dy = \mathrm{O}(h^{-d[2s(r-1)+s-1]}). \qquad (31)$$

Using (25) along with (28) and (31) yields (27): the proof is complete. ■

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*, New York: John Wiley and Sons.

Horváth, L. (1991) On L_p-norms of multivariate density estimators, *Annals of Statistics* **19**: 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.

Ledoux, M., Talagrand, M. (1991) *Probability in Banach spaces: Isoperimetry and processes*, Springer–Verlag, Berlin.

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**: 129–135.

Parzen, E. (1962) On estimation of a probability density function and mode, *Annals of Mathematical Statistics* **33**: 1065–1076.

Rosenblatt, M. (1956) Remarks on some nonparametric estimates of a density function, *Annals of Mathematical Statistics* **27**: 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. (1998) *Asymptotic statistics*, Cambridge: Cambridge University Press.

Van der Vaart, A.W., Wellner, J.A. (1996) *Weak convergence and empirical processes with applica-*

tions to statistics, New York: Springer–Verlag.