



On the weak convergence of the kernel density estimator in the uniform topology

Gilles Stupfler

► To cite this version:

Gilles Stupfler. On the weak convergence of the kernel density estimator in the uniform topology. 2016. hal-01220124v2

HAL Id: hal-01220124

<https://hal.science/hal-01220124v2>

Preprint submitted on 24 Feb 2017

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 the kernel density estimator in the uniform topology

Gilles Stupfler

Aix Marseille Université, CNRS, EHESS, Centrale Marseille, GREQAM UMR 7316,
13002 Marseille, France

Abstract. The pointwise asymptotic properties of the Parzen-Rosenblatt kernel estimator \hat{f}_n of a probability density function f on \mathbb{R}^d have received great attention, and so have its integrated or uniform errors. It has been pointed out in a couple of recent works that the weak convergence of its centered and rescaled versions in a weighted Lebesgue L^p space, $1 \leq p < \infty$, considered to be a difficult problem, is in fact essentially uninteresting in the sense that the only possible Borel measurable weak limit is 0 under very mild conditions. This paper examines the weak convergence of such processes in the uniform topology. Specifically, 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, with probability 1, the sample paths of any tight Borel measurable weak limit in an ℓ^∞ space on \mathbb{R}^d of the process $r_n(\hat{f}_n - f_n)$ must be almost everywhere zero. The particular case when the estimator \hat{f}_n has continuous sample paths is then considered and simple conditions making it possible to examine the actual existence of a weak limit in this framework are provided.

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

Keywords: Kernel density estimator, weak convergence, ℓ^∞ space, tightness.

1 Introduction

The Parzen-Rosenblatt estimator of a probability density function f on \mathbb{R}^d , $d \geq 1$ (Parzen, 1962, Rosenblatt, 1956) is defined as follows:

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

Here, (X_n) is a sequence of independent random copies of a random variable X , such that X has a probability density function f . In particular, we assume that the X_n , $n \geq 1$ are defined on a common probability space and induce Borel measurable maps. The parameter $h = h(n) \rightarrow 0$ as $n \rightarrow \infty$ is called the bandwidth, and we let $K_h(u) = h^{-d}K(u/h)$ for a kernel function $K : \mathbb{R}^d \rightarrow \mathbb{R}$, that is, an integrable function on \mathbb{R}^d with unit integral. The estimator \hat{f}_n is essentially a (possibly modified) version of the histogram whose smoothness is tuned by h and potentially enhanced by the

regularity of the kernel function K . The random function $x \mapsto \widehat{f}_n(x)$ is the empirical counterpart of the function $x \mapsto f_n(x) = \mathbb{E}[f_n(x)] = \mathbb{E}[K_h(x - X)]$ which is well-defined almost everywhere on \mathbb{R}^d and integrable. The pointwise properties of the Parzen-Rosenblatt estimator have been known for a long time: for instance, it is shown in pp.1069–1070 of Parzen (1962) that under some conditions on K , the quantity $\sqrt{nh^d}(\widehat{f}_n(x) - f_n(x))$ converges weakly to a Gaussian distribution provided $nh^d \rightarrow \infty$, $f(x) > 0$ and f is continuous at x .

In this paper, we focus on the weak 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 an ℓ^∞ space on \mathbb{R}^d . In other words, we try to understand the limiting behavior of centered and rescaled versions of \widehat{f}_n in the uniform topology on a (measurable) set. Convergence results of this kind are valuable because they have important corollaries such as uniform convergence properties of the random function $r_n(\widehat{f}_n - f)$: if a nontrivial weak limit can be identified for the process $r_n(\widehat{f}_n - f_n)$ and a suitable condition on the bias term $r_n(f_n - f)$ is further satisfied, then the rate of uniform convergence of the estimator \widehat{f}_n to f shall be exactly r_n .

Specifically, let S be a Borel measurable set in \mathbb{R}^d with positive Lebesgue measure and $\ell^\infty(S)$ be the space of those real-valued functions which are bounded on S :

$$H \in \ell^\infty(S) \Leftrightarrow \|H\|_{\infty,S} := \inf\{C \geq 0 \mid |H(x)| \leq C \text{ for all } x \in S\} < \infty.$$

Assuming that K is bounded on \mathbb{R}^d , it is straightforward that the random function $x \mapsto r_n(\widehat{f}_n(x) - f_n(x))$, $x \in S$, defines a random process belonging to $\ell^\infty(S)$. Clearly, for this random process to converge weakly in $\ell^\infty(S)$ its uniform norm $r_n\|\widehat{f}_n - f_n\|_{\infty,S}$ has to converge weakly in \mathbb{R} , but this is not a sufficient condition. The uniform norm of $r_n(\widehat{f}_n - f_n)$ on S has been studied in many instances in the literature, see for example the early works of Bickel and Rosenblatt (1973), Silverman (1978) and Stute (1982, 1984); Talagrand's inequalities (1994, 1996) and general distributional results on empirical processes (see the monographs by van der Vaart and Wellner, 1996 and van der Vaart, 1998) then sparked renewed interest in this problem, see *e.g.* Einmahl and Mason (2000), Giné and Guillou (2002), Giné *et al.* (2004), Einmahl and Mason (2005) and Dony and Einmahl (2006).

None of these works though consider the convergence of $r_n(\widehat{f}_n - f_n)$ as a random process taking values in an ℓ^∞ space on \mathbb{R}^d . More broadly, the problem of analyzing the convergence of this process in functional spaces such as L^p spaces on \mathbb{R}^d has long been considered to be difficult. When K^2 is integrable on \mathbb{R}^d , the recent work of Nishiyama (2011) generalized a result of Ruymgaart (1998) by disproving the existence of a nondegenerate Borel measurable weak limit for the process $r_n(\widehat{f}_n - f_n)$ in the $L^2(\mathbb{R}^d)$ space of square-integrable functions on \mathbb{R}^d provided $r_n/\sqrt{n} \rightarrow 0$. The ideas of his paper paved the way for the work of Stupfler (2014) which showed that the same negative conclusion holds in the weighted L^p spaces

$$L^p(\mathbb{R}^d, \mu) := \left\{ H : \mathbb{R}^d \rightarrow \mathbb{R} \mid H \text{ is Borel measurable and } \int_{\mathbb{R}^d} |H(x)|^p d\mu(x) < \infty \right\}$$

for $p \in [1, \infty)$, when K^p is integrable on \mathbb{R}^d and the weighting measure μ is a nontrivial absolutely continuous measure with bounded Radon-Nikodym derivative

with respect to the Lebesgue measure. These results cannot be easily extended to the $\ell^\infty(S)$ space for topological reasons: in particular, both Nishiyama (2011) and Stupfler (2014) use the fact that for p finite, the space $L^p(\mathbb{R}^d, \mu)$ is a separable metric space whose dual space is $L^q(\mathbb{R}^d, \mu)$ for $q = p/(p-1)$. It is well-known that the space $\ell^\infty(S)$ fails to be separable in general and its dual space is more difficult to work with, which causes measurability-related problems for the process $r_n(\hat{f}_n - f_n)$ and makes it very hard to characterize weak convergence to an arbitrary Borel measurable random element in $\ell^\infty(S)$. This is why we start here by introducing a convenient subspace of the dual space of $\ell^\infty(S)$, and we then use it to identify the possible tight Borel measurable weak limits of the process $r_n(\hat{f}_n - f_n)$ in the uniform topology. It is shown in what follows that such a limit must be 0 almost everywhere on S with probability 1. Our conclusion about the possible limits of the process $r_n(\hat{f}_n - f_n)$ appears to be different from what may be obtained when considering other types of convergence, such as the weak convergence of processes constructed using \hat{f}_n and indexed by classes of functions, see the work by van der Vaart (1994) and further developments in *e.g.* Radulović and Wegkamp (2000) and Giné and Nickl (2008), even though these papers also focus on weak convergence to tight (Gaussian) limits. Moreover, we shall highlight that when K is continuous, since the process considered has continuous sample paths, one can show as a corollary of our results that the limit must be 0 everywhere on S and discard the requirement that the weak limit be tight under a further mild condition on S by taking advantage of the particular topology of spaces of continuous functions over compact sets. We finally show how this makes it possible to classify the asymptotic behavior of the process of interest, depending on (r_n) , by using the sharp rates of uniform convergence of \hat{f}_n to f_n obtained in Giné and Guillou (2002). Under a further regularity condition on f , it is then straightforward that our results carry over to the process $r_n(\hat{f}_n - f)$, which is the process of interest in practice, when a classical bias condition is satisfied.

The outline of the paper is as follows: our main results are stated in Section 2 and some concluding remarks, including on possible extensions of our results, are given in Section 3.

2 Main results

In all what follows, we assume that S is a Borel measurable set in \mathbb{R}^d with positive Lebesgue measure and $K : \mathbb{R}^d \rightarrow \mathbb{R}$ is an integrable function with unit integral which is bounded on \mathbb{R}^d . The guiding ideas are those of Nishiyama (2011) and Stupfler (2014): our first result relates the problem of identifying the possible weak limits of the process $r_n(\hat{f}_n - f_n)$ in $\ell^\infty(S)$ to the simpler problem of understanding the weak convergence of sequences of real-valued random variables constructed using this process and a suitably chosen class of continuous linear functionals on $\ell^\infty(S)$. To do so, we start by noting that Borel measurability of the random function $r_n(\hat{f}_n - f_n)$ in $\ell^\infty(S)$ is not clear even though K and the X_i are measurable, because the space $\ell^\infty(S)$ is not separable (see the discussion in Section 1.1 of van der Vaart and Wellner, 1996). In this paper, “weak convergence” thus refers to the notion of weak convergence using outer probabilities (see Definition 1.3.3 p.17 in van der Vaart and Wellner, 1996).

We continue by recalling a few facts about duality in $\ell^\infty(S)$. Since we will investigate the possible tight and Borel measurable weak limits of the process $r_n(\hat{f}_n - f_n)$, it turns out that we need only work with the space of all bounded and Borel measurable functions on S . This space is itself a subspace of $L^\infty(S)$, the space of all Borel measurable functions which are essentially bounded on S :

$$H \in L^\infty(S) \Leftrightarrow \|H\|_{L^\infty(S)} := \inf\{C \geq 0 \mid |H(x)| \leq C \text{ for almost every } x \in S\} < \infty.$$

Let μ_S be the measure on \mathbb{R}^d whose Radon-Nikodym density with respect to the Lebesgue measure is $\mathbb{1}_S$, the indicator of the set S . Then $L^\infty(S) = L^\infty(\mathbb{R}^d, \mu_S)$, the space of the Borel measurable functions on \mathbb{R}^d which are bounded μ_S -almost everywhere. By Theorem 16 p.296 in Dunford and Schwartz (1957), the space $\text{ba}(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \mu_S)$ of the additive, bounded, signed measures on the Borel σ -algebra $\mathcal{B}(\mathbb{R}^d)$ which are absolutely continuous with respect to μ_S is then isometrically isomorphic to the dual space of $L^\infty(S)$. The isomorphism is

$$\nu \in \text{ba}(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \mu_S) \mapsto \left(T_\nu : g \in L^\infty(S) \mapsto \int_{\mathbb{R}^d} g(x) d\nu(x) \right)$$

with the topology on $\text{ba}(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \mu_S)$ being induced by the total variation distance between measures. Because an element of $\text{ba}(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \mu_S)$ may be additive but not countably additive, the whole dual space $\text{ba}(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \mu_S)$ is somewhat inconvenient to work with; in particular, the absolute continuity condition with respect to the (countably additive and σ -finite) measure μ_S is difficult to take advantage of because it does not translate into the existence of a Radon-Nikodym derivative with respect to μ_S . This is why we consider instead the subspace

$$\text{bca}(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \mu_S) = \{\nu \in \text{ba}(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \mu_S) \mid \nu \text{ is countably additive}\}.$$

Using a Hahn-Jordan decomposition, any element ν of $\text{bca}(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \mu_S)$, which is σ -finite because it is bounded, must have a Radon-Nikodym derivative with respect to μ_S . The particular structure of μ_S then entails that ν must have a Radon-Nikodym derivative with respect to the Lebesgue measure as well, having value 0 everywhere outside S , and we denote it by $d\nu/dx$. With these elements in mind, the following result can be stated:

Proposition 1. *If G_1 and G_2 are two tight Borel measurable random elements of $L^\infty(S)$, then the distributions of G_1 and G_2 are equal if and only if for every $\nu \in \text{bca}(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \mu_S)$ such that $\|d\nu/dx\|_{L^\infty(S)} < \infty$, the distributions of $T_\nu(G_1)$ and $T_\nu(G_2)$ are equal.*

Proof of Proposition 1. For any $\nu \in \text{bca}(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \mu_S)$, the map T_ν is a continuous linear form on $L^\infty(S)$, so if G_1 and G_2 have equal distributions then $T_\nu(G_1)$ and $T_\nu(G_2)$ must have equal distributions as well. Conversely, suppose that for any $\nu \in \text{bca}(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \mu_S)$ such that $\|d\nu/dx\|_{L^\infty(S)} < \infty$, the distributions of $T_\nu(G_1)$ and $T_\nu(G_2)$ are equal. We introduce the class \mathcal{F} of functions $F : L^\infty(S) \rightarrow \mathbb{R}$ for which there exist a positive integer J , a continuous and bounded real-valued function g on \mathbb{R}^J and $\nu_1, \dots, \nu_J \in \text{bca}(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \mu_S)$ having essentially bounded Radon-Nikodym derivatives on S with respect to the Lebesgue measure, such that:

$$\forall \varphi \in L^\infty(S), F(\varphi) = g(T_{\nu_1}(\varphi), \dots, T_{\nu_J}(\varphi)).$$

Because

$$\forall t_1, \dots, t_J \in \mathbb{R}, \forall \nu_1, \dots, \nu_J \in \text{bca}(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \mu_S), \sum_{i=1}^J t_i T_{\nu_i} = T_{\sum_{i=1}^J t_i \nu_i},$$

it comes as a consequence of the Cramér-Wold device that the random vectors $(T_{\nu_1}(G_1), \dots, T_{\nu_J}(G_1))$ and $(T_{\nu_1}(G_2), \dots, T_{\nu_J}(G_2))$ must have the same distribution. If ρ_1 and ρ_2 are the pushforward probability measures on $L^\infty(S)$ induced by G_1 and G_2 , it becomes clear that

$$\forall F \in \mathcal{F}, \int_{L^\infty(S)} F(\varphi) d\rho_1(\varphi) = \int_{L^\infty(S)} F(\varphi) d\rho_2(\varphi).$$

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 $L^\infty(S)$ containing the constant functions. By Lemma 1.3.12 (ii) p.25 in van der Vaart and Wellner (1996), it suffices to prove that the class \mathcal{F} separates the points of $L^\infty(S)$.

Let then $\varphi, \psi \in L^\infty(S)$ be such that $\varphi \neq \psi$. In other words, the Borel measurable set $E = \{\varphi \neq \psi\}$ satisfies $\mu_S(E) > 0$. Since $\mu_S(E \cap [0, N]^d) \uparrow \mu_S(E)$ as $N \rightarrow \infty$, we may find a bounded Borel measurable set F such that $\mu_S(F) > 0$ and $\varphi \neq \psi$ on F . In particular, $\varphi \neq \psi$ on $F \cap S$, which is a Borel measurable set having a positive and finite Lebesgue measure. Let

$$\nu = (\varphi - \psi) \mathbb{1}_F \cdot \mu_S.$$

Then ν is clearly an element of $\text{bca}(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \mu_S)$ with essentially bounded Radon-Nikodym derivative on S and

$$T_\nu(\varphi - \psi) = \int_{F \cap S} [\varphi(x) - \psi(x)]^2 dx > 0$$

which is the desired separation property. The proof is complete. \blacksquare

This result basically makes it possible to identify a tight Borel measurable random element of $\ell^\infty(S)$ up to null sets in S . It is the central tool necessary to prove our first asymptotic result on the possible tight Borel measurable weak limits of the process $r_n(\hat{f}_n - f_n)$ in $\ell^\infty(S)$.

Theorem 1. *Let (r_n) be a nonrandom sequence of positive real numbers. If $r_n/\sqrt{n} \rightarrow 0$ and the random process $r_n(\hat{f}_n - f_n)$ converges weakly in $\ell^\infty(S)$ to a tight Borel measurable random process G then G is almost surely zero almost everywhere on S .*

Proof of Theorem 1. Pick $\nu \in \text{bca}(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \mu_S)$ such that $\|d\nu/dx\|_{L^\infty(S)} < \infty$ and note that the map

$$t \mapsto T_\nu(K_h(\cdot - t)) = \int_S K_h(x - t) \frac{d\nu}{dx}(x) dx$$

is Borel measurable because S is a measurable set and K and $d\nu/dx$ are Borel measurable as well. As a consequence, $T_\nu(K_h(\cdot - X))$ is a Borel measurable real-valued random variable and, by the continuous mapping theorem (see Theorem 1.3.6

p.20 in van der Vaart and Wellner, 1996), the weak convergence of $r_n(\widehat{f}_n - f_n)$ to G implies the following weak convergence of Borel measurable real-valued random variables:

$$\Delta_n(\nu) := T_\nu \left(r_n(\widehat{f}_n - f_n) \right) \rightarrow T_\nu(G) \text{ as } n \rightarrow \infty.$$

We start by showing that $T_\nu(G) = 0$ almost surely. Because ν is countably additive and σ -finite, Fubini's theorem yields

$$T_\nu(\mathbb{E}(K_h(\cdot - X))) = \mathbb{E}(T_\nu(K_h(\cdot - X))).$$

We may then rewrite $\Delta_n(\nu)$ as a sum of independent and identically distributed centered random variables, as follows:

$$\Delta_n(\nu) = \frac{r_n}{n} \sum_{i=1}^n W_{n,i}(\nu) \text{ with } W_{n,i}(\nu) = T_\nu(K_h(\cdot - X_i)) - \mathbb{E}(T_\nu(K_h(\cdot - X))).$$

A change of variables yields

$$T_\nu(K_h(\cdot - X)) = \int_S K_h(x - X) \frac{d\nu}{dx}(x) dx = \int_{\mathbb{R}^d} K(t) \frac{d\nu}{dx}(X + ht) \mathbb{1}_{\{X+ht \in S\}} dt$$

almost surely. Because $\|d\nu/dx\|_{L^\infty(S)} < \infty$, we get with probability 1:

$$|T_\nu(K_h(\cdot - X))| \leq \|d\nu/dx\|_{L^\infty(S)} \int_{\mathbb{R}^d} |K(t)| dt < \infty.$$

In other words, the random variable $T_\nu(K_h(\cdot - X))$ is almost surely bounded. The triangle inequality thus entails

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

Consequently, $\Delta_n(\nu) \rightarrow 0$ in probability as $n \rightarrow \infty$ and $T_\nu(G) = 0$ almost surely.

Now, because inclusion preserves tightness (see Lemma 14.4 p.257 in Kallenberg, 1997), G also defines a tight Borel measurable random element of $L^\infty(S)$. By Proposition 1, $G = 0$ almost surely in $L^\infty(S)$, which means in particular that G is almost surely zero almost everywhere on S : the proof is complete. \blacksquare

Theorem 1 is an analogue of Theorem 2.1 in Nishiyama (2011) and Theorem 2.2 in Stupfler (2014), which tackled the case of weak convergence in weighted L^p spaces on \mathbb{R}^d , $1 \leq p < \infty$. This result says that either the process $r_n(\widehat{f}_n - f_n)$ converges weakly to an essentially degenerate limit or does not converge weakly to a tight Borel measurable limit.

Remark 1. By Theorem 1.5.4 p.35 in van der Vaart and Wellner (1996), the existence of a tight weak limit for $r_n(\widehat{f}_n - f_n)$ in $\ell^\infty(S)$ is equivalent to asymptotic tightness of this process (van der Vaart and Wellner, 1996, p.21) plus weak convergence of its finite-dimensional marginals. Under some technical conditions on K and continuity of f , joint convergence of the marginals of $r_n(\widehat{f}_n - f_n)$ can be checked when (r_n) has at most order $\sqrt{nh^d}$, by using arguments similar to those of Parzen (1962, Theorems 1A and 2A and discussion on pp.1069-1070); in such a case, Theorem 1 yields that the asymptotic tightness of $r_n(\widehat{f}_n - f_n)$ is equivalent to its weak convergence to a limit essentially equal to 0 on S .

It should be pointed out that on the one hand, because $\ell^\infty(S)$ is not a separable metric space, Borel measurable random elements on this space are not necessarily tight; on the other hand, tightness of a Borel probability measure on a complete metric space such as $\ell^\infty(S)$ is equivalent to separability of this measure (van der Vaart and Wellner, 1996, Lemma 1.3.2 p.17), and nonseparable Borel measures cannot be constructed in the usual Zermelo-Fraenkel system of axioms (van der Vaart and Wellner, 1996, p.24). As a consequence, tightness of the weak limit does not appear to be a very restrictive requirement in practice, and this condition makes it possible to use in the proof of Proposition 1 the very nice characterization of the distribution of a random process on a metric space contained in Lemma 1.3.12 (ii) p.25 in van der Vaart and Wellner (1996), while getting information about a non-tight distribution appears to be difficult, see (i) in this same Lemma. Tightness is consequently a desirable property, encountered in many instances when considering weak convergence in a metric space endowed with a sup-norm, see for example the necessary-and-sufficient conditions for weak convergence in the space $\ell^\infty(S)$ in Section 18 of van der Vaart (1998) and particularly Theorem 18.14 p.261 therein. Other instances where this condition is used include recent works on weak convergence in the uniform topology over a class of functions, see *e.g.* van der Vaart (1994), Radulović and Wegkamp (2000), Mendelson and Zinn (2006), Nickl (2007), Giné and Nickl (2008), Nickl (2009) and Radulović and Wegkamp (2009). It is remarkable that although Theorem 1 implies that the process $x \mapsto r_n(\hat{f}_n(x) - f_n(x))$ cannot have a non-essentially trivial, tight Borel measurable weak limit in the space $\ell^\infty(S)$, the process

$$g \in \mathcal{G} \mapsto \sqrt{n} \int_{\mathbb{R}^d} (\hat{f}_n(x) - f_n(x))g(x)dx,$$

where \mathcal{G} is a suitable class of functions, may actually converge in $\ell^\infty(\mathcal{G})$ to a tight Brownian bridge limit, see for instance Radulović and Wegkamp (2000) and Giné and Nickl (2008).

A case though in which we can strengthen our conclusion about the possible limits of $r_n(\hat{f}_n - f_n)$ is when it takes its values in the space $\mathcal{C}(S)$ of continuous functions on S . In this case, if no point in S is isolated from the point of view of the Lebesgue measure then we should expect the potential weak limit in Theorem 1 to be 0 everywhere on S instead of almost everywhere. Meanwhile, the tightness hypothesis about the weak limit, although fairly mild as mentioned above, can be for instance dropped when S is compact, because the space $\mathcal{C}(S)$ is then a separable and complete subspace of $\ell^\infty(S)$, making any Borel measurable random element tight in this space. Somewhat surprisingly, the tightness requirement can actually also be dropped in the much more general case when S is σ -compact. These two reasons lead us to introduce our next assumption on S :

(H_1) S can be written as the union of countably many compact subsets of \mathbb{R}^d and, for every $x \in S$ and $\varepsilon > 0$, the intersection of S and the Euclidean open ball with center x and radius ε has positive Lebesgue measure.

Combining condition (H_1), which holds true in most if not all practical applications (for instance if S is an open cube, the closure of an open set, or equal to \mathbb{R}^d), with a continuity assumption about K , we get the following result.

Theorem 2. *Let (r_n) be a nonrandom sequence of positive real numbers. Assume that S satisfies (H_1) and K is a continuous function 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 $\ell^\infty(S)$ to a Borel measurable random process G then $G = 0$ almost surely.*

Proof of Theorem 2. The regularity requirement on K makes it clear that the sample paths of the process $\widehat{f}_n - f_n$ are in fact almost surely continuous on S , in the sense that the event $A_n := \{\widehat{f}_n - f_n \text{ is continuous on } S\}$, although not necessarily measurable, contains a measurable set having probability 1; in particular, A_n has outer probability 1 for every n . Furthermore, because $\ell^\infty(S)$ and $\mathcal{C}(S)$ are complete metric spaces and a uniform limit of continuous functions is continuous, it is clear that the space $\mathcal{C}(S)$ is a closed subspace of $\ell^\infty(S)$: it follows from the Portmanteau theorem (see Theorem 1.3.4 p.18 in van der Vaart and Wellner, 1996) that the probability that G belongs to $\mathcal{C}(S)$ is equal to 1. As a first conclusion, G thus defines a probability measure on $\mathcal{C}(S)$.

We first deal with the case when S is compact. The space $\mathcal{C}(S)$ is then a separable and complete metric space so that any Borel probability measure on $\mathcal{C}(S)$ is tight. It follows that G defines a tight element of $\mathcal{C}(S)$ and thus of $\ell^\infty(S)$ because inclusion preserves tightness (see Lemma 14.4 p.257 in Kallenberg, 1997). By Theorem 1, $G = 0$ almost everywhere on S with probability 1. Finally, because G is continuous on S and (H_1) holds, one concludes that $G = 0$ almost surely on S .

If now S is not compact, notice that for every compact set T contained in S , the restriction map $H \mapsto H|_T$ from $\ell^\infty(S)$ to $\ell^\infty(T)$ is continuous. By the continuous mapping theorem, $r_n(\widehat{f}_n - f_n)$ then converges weakly in $\ell^\infty(T)$ to the restriction $G|_T$ of G on the set T . We thus get $G|_T = 0$ almost surely for any compact subset T of S . The result follows since S is a countable union of compact subsets of \mathbb{R}^d . ■

In particular, when S is an open cube in \mathbb{R}^d , we can infer that any weak limit of $r_n(\widehat{f}_n - f_n)$ in $\ell^\infty(S)$ should be degenerate, although it is known since Stute (1982, 1984) that under additional conditions, the sup-norm of $\widehat{f}_n - f_n$ over S converges almost surely at the rate $v_n := \sqrt{nh^d/|\log h|}$. This last observation suggests that the sequence (v_n) shall play a crucial role in the description of the actual asymptotic behavior of $\widehat{f}_n - f_n$ in $\ell^\infty(S)$. Another consequence of Theorem 2 is that centered and rescaled kernel density estimators cannot converge weakly to a Gaussian process in spaces of continuous functions; see also the introduction of Ruymgaart (1998).

We thus examine in the second part of this work what happens depending on the behavior of (r_n) relatively to (v_n) . The arguments presented in what follows make a heavy use of the exact rates of convergence for the sup-norm of $\widehat{f}_n - f_n$ which were investigated in Giné and Guillou (2002). We introduce the following hypotheses:

(H_2) The set S is σ -compact and its interior S° is dense in S .

(M) The kernel K is a nonnegative, bounded, compactly supported function belonging to the linear span of the nonnegative functions k satisfying the following property: the subgraph $\{(s, u) \in \mathbb{R}^d \times \mathbb{R} \mid k(s) \geq u\}$ of k can be represented as a finite number of Boolean operations among sets of the form $\{(s, u) \in \mathbb{R}^d \times \mathbb{R} \mid p(s, u) \geq \varphi(u)\}$ where p is a polynomial on \mathbb{R}^{d+1} and φ is an arbitrary real function.

(R) The function f is uniformly continuous and the kernel K is continuous on \mathbb{R}^d .

(W) The bandwidth h is such that

$$h \downarrow 0, \frac{nh^d}{|\log h|} \rightarrow \infty, \frac{|\log h|}{\log \log n} \rightarrow \infty \text{ and } nh^d \uparrow \infty.$$

Assumption (H_2) contains condition (H_1) and entails that the supremum of a continuous function over S is also its supremum over its interior S° . Hypothesis (M) is taken from Giné and Guillou (2002) and Giné *et al.* (2004); it is basically a measurability condition ensuring that the class of functions

$$\mathcal{K} = \left\{ x \mapsto K\left(\frac{x-t}{h}\right) : h > 0, t \in \mathbb{R}^d \right\}$$

is a bounded measurable Vapnik-Červonenkis class, which ensures in particular that $\|\hat{f}_n - f_n\|_{\infty, S}$ is a Borel measurable random variable and is thus a key ingredient for the search of uniform rates of convergence for $\hat{f}_n - f_n$. Many kernels, such as the naive kernel or the pyramid kernel, satisfy this assumption, see the discussion p.911 in Giné and Guillou (2002). Regularity condition (R) on f and K especially gives that the sample paths of $\hat{f}_n - f_n$ should be almost surely continuous; notice that a similar condition is also required in Theorem 3.1 of Stute (1984). Condition (W) was already partly introduced in p.87 of Stute (1982) and p.367 of Stute (1984), and in its present form, it is one of the hypotheses necessary for the results of Giné and Guillou (2002) to hold. Related but stronger assumptions are those of Giné *et al.* (2004), p.2574. The following result then holds:

Theorem 3. *Let (r_n) be a nonrandom sequence of positive real numbers. Assume that K satisfies condition (M) , that the density function f is bounded on \mathbb{R}^d and that condition (W) holds.*

(i) *If $r_n/v_n \rightarrow 0$, then $r_n(\hat{f}_n - f_n) \rightarrow 0$ weakly in $\ell^\infty(S)$.*

Assume further that conditions (H_2) and (R) hold, the set S is either \mathbb{R}^d or bounded and the set $\{f > 0\} \cap S^\circ$ is not empty.

(ii) *If $r_n/v_n \rightarrow c \in (0, \infty]$, then $r_n(\hat{f}_n - f_n)$ does not converge weakly to any Borel measurable random element in $\ell^\infty(S)$.*

Proof of Theorem 3. To show (i), apply first Theorem 2.3 in Giné and Guillou (2002) to get

$$\limsup_{n \rightarrow \infty} v_n \|\hat{f}_n - f_n\|_{\infty, \mathbb{R}^d} \leq C \text{ almost surely}$$

where C is a nonnegative finite constant. Since $r_n/v_n \rightarrow 0$ this entails

$$\lim_{n \rightarrow \infty} r_n \|\hat{f}_n - f_n\|_{\infty, \mathbb{R}^d} = 0 \text{ almost surely.}$$

But clearly

$$r_n \|\hat{f}_n - f_n\|_{\infty, S} \leq r_n \|\hat{f}_n - f_n\|_{\infty, \mathbb{R}^d}$$

and thus

$$\lim_{n \rightarrow \infty} r_n \|\widehat{f}_n - f_n\|_{\infty, S} = 0 \text{ almost surely}$$

of which a consequence is $r_n(\widehat{f}_n - f_n) \rightarrow 0$ weakly in $\ell^\infty(S)$.

Point (ii) is shown by recalling that because (H_2) holds, then

$$\|\widehat{f}_n - f_n\|_{\infty, S} = \|\widehat{f}_n - f_n\|_{\infty, S^\circ} \text{ almost surely.}$$

Applying Proposition 3.1 in Giné and Guillou (2002) when S° is bounded or Theorem 3.3 in Giné and Guillou (2002) when it is equal to \mathbb{R}^d , we obtain

$$\lim_{n \rightarrow \infty} \frac{r_n \|\widehat{f}_n - f_n\|_{\infty, S}}{\sqrt{2d \|f\|_{\infty, S} \int_{\mathbb{R}^d} K^2}} = c > 0 \text{ almost surely.}$$

It follows that $r_n \|\widehat{f}_n - f_n\|_{\infty, S}$ has a positive (possibly infinite) almost sure limit; if $r_n(\widehat{f}_n - f_n)$ converged weakly to a Borel measurable random element in $\ell^\infty(S)$ then this limit would be almost surely 0 by Theorem 2 and therefore $r_n \|\widehat{f}_n - f_n\|_{\infty, S}$ would converge to 0 in probability, which is a contradiction. The proof is complete. \blacksquare

Theorem 3, which offers a full classification of the asymptotic behavior of $r_n(\widehat{f}_n - f_n)$ in $\ell^\infty(S)$ depending on the rate (r_n) , is the counterpart of Theorem 2.2 in Nishiyama (2011) and Theorems 2.3 and 2.4 in Stupfler (2014) for the uniform topology. We conclude Section 2 with several remarks about this last result.

Remark 2. In Theorem 3, the uniform continuity condition on f contained in (R) can be relaxed, as mentioned in Giné and Guillou (2002), by assuming that the set $\{f > 0\}$ is open, f is continuous and bounded on this set, and $f(x)$ converges to 0 as $\|x\| \rightarrow \infty$. This makes it possible to apply Theorem 3 to the uniform distribution or the exponential distribution.

Remark 3. It is interesting to note that the rate r_n for which the weak behavior of $r_n(\widehat{f}_n - f_n)$ becomes nontrivial in $\ell^\infty(S)$ is $v_n = \sqrt{nh^d}/|\log h|$ which is asymptotically smaller than the rate $\sqrt{nh^d}$ playing an analogue role in $L^p(\mathbb{R}^d, \mu_S)$ when p is finite (Nishiyama, 2011 and Stupfler, 2014). While it is a well-known fact that uniform rates of convergence usually feature a logarithmic penalty term, it also suggests that one cannot easily deduce the limiting behavior of $r_n(\widehat{f}_n - f_n)$ in $\ell^\infty(S)$ from its behavior in an $L^p(\mathbb{R}^d, \mu_S)$ space for p finite. Indeed, write tentatively

$$\forall p \in [1, \infty), \quad r_n \|\widehat{f}_n - f_n\|_{p, \mu_S} \leq r_n \|\widehat{f}_n - f_n\|_{\infty, S} |S|^{1/p}$$

when S has a finite Lebesgue measure $|S|$. Then it is known that under very mild hypotheses the left-hand side does not converge to 0 in probability if and only if r_n has at least order $\sqrt{nh^d}$, see *e.g.* the proof of Theorems 2.3 and 2.4 in Stupfler (2014). The strongest conclusion one can reach from this inequality is thus that $\sqrt{nh^d} \|\widehat{f}_n - f_n\|_{\infty, S}$ cannot converge to 0 in probability. In other words, applying Theorem 2, the best we can infer is that $\sqrt{nh^d}(\widehat{f}_n - f_n)$ cannot converge weakly to a Borel measurable element in $\ell^\infty(S)$, which is not informative enough since we know from Stute (1982, 1984) that the rate of convergence of $\|\widehat{f}_n - f_n\|_{\infty, S}$ in \mathbb{R} is strictly smaller than $\sqrt{nh^d}$.

Remark 4. While Theorem 3 says that we cannot expect $v_n(\hat{f}_n - f_n)$ to converge to a nondegenerate limit in $\ell^\infty(S)$, it does not mean that finding uniform results that have practical implications is impossible. For instance, Theorem 3.1 in Bickel and Rosenblatt (1973) in the case $d = 1$ and the approximation results in Propositions 3.1 and 3.2 in Chernozhukov *et al.* (2014) suggest that, under certain regularity hypotheses, the problem of finding constants A and B such that the difference

$$2d|\log h| \left[\frac{v_n \|\hat{f}_n - f_n\|_{\infty, S}}{\sqrt{2d\|f\|_{\infty, S} \int_{\mathbb{R}^d} K^2}} - 1 \right] - A \log(|\log h| \vee e) - B$$

converges weakly to a nondegenerate weak limit has a solution; when $d = 1$, the result of Bickel and Rosenblatt (1973) actually gives A and B such that

$$2|\log h| \left[\sup_{x \in [0,1]} \left| \frac{v_n(\hat{f}_n(x) - f_n(x))}{\sqrt{2f(x) \int_{\mathbb{R}^d} K^2}} \right| - 1 \right] - A \log(|\log h| \vee e) - B$$

converges weakly to a nondegenerate, explicit limit. In other words, it appears that the correct way to obtain asymptotic uniform confidence bands on f is to look directly at the (possibly weighted) supremum of $v_n|\hat{f}_n - f_n|$ over S instead of working on the weak behavior of $v_n(\hat{f}_n - f_n)$ in $\ell^\infty(S)$.

3 Concluding remarks

In this paper, we examined the weak behavior of centered and rescaled versions $r_n(\hat{f}_n - f_n)$ of the Parzen-Rosenblatt density estimator \hat{f}_n in ℓ^∞ spaces on \mathbb{R}^d . In particular, we showed that under mild conditions, any Borel measurable weak limit of this process is equal to 0, although the exact almost sure asymptotics for uniform norms of $\hat{f}_n - f_n$ are known to be nontrivial. Interestingly, our results are similar to the negative results of Stupfler (2014) regarding the weak behavior of $r_n(\hat{f}_n - f_n)$ in L^p spaces on \mathbb{R}^d for p finite; besides, the basic idea in the case of L^p spaces, which was to understand the weak behavior of this process through the weak behavior of a suitable collection of its integrals, can actually also be used successfully in the ℓ^∞ space, because of the particular structure of its dual space. In other words, although an L^p space for $1 \leq p < \infty$ and an ℓ^∞ space are structurally very different from each other, their dual spaces have enough common characteristics to ensure that the weak convergence properties of $\hat{f}_n - f_n$ can be examined in the same way.

Moreover, the presented technique may be applied to other types of density estimators to analyze their weak behavior in functional spaces. Consider for instance the wavelet density estimator on \mathbb{R} (Doukhan and León, 1990, Kerkycharian and Picard, 1992), that is:

$$\tilde{f}_n(x) := \sum_{k \in \mathbb{Z}} \hat{\alpha}_{j_n, k} 2^{j_n/2} \Phi(2^{j_n} x - k) \quad \text{where} \quad \hat{\alpha}_{j_n, k} := \frac{1}{n} \sum_{i=1}^n 2^{j_n/2} \Phi(2^{j_n} X_i - k).$$

Here $j_n \uparrow \infty$ is a sequence of integers; the function Φ is a square-integrable function such that $\{\Phi(\cdot - k), k \in \mathbb{Z}\}$ is an orthonormal system in $L^2(\mathbb{R})$ and moreover, the

(functional) linear spaces defined by induction by

$$V_0 = \left\{ g(x) = \sum_{k \in \mathbb{Z}} c_k \Phi(x - k), (c_k) \text{ square-summable} \right\}$$

$$\text{and } \forall j \geq 1, V_j = \{h(x) = g(2x), g \in V_{j-1}\}$$

are nested and such that their union is dense in $L^2(\mathbb{R})$. Then clearly

$$r_n \left(\tilde{f}_n(x) - \mathbb{E}(\tilde{f}_n(x)) \right) = \frac{1}{n} \sum_{i=1}^n Y_{n,i}(x)$$

$$\text{where } Y_{n,i}(x) = \sum_{k \in \mathbb{Z}} [\Phi(2^{j_n} X_i - k) - \mathbb{E}(\Phi(2^{j_n} X_i - k))] 2^{j_n} \Phi(2^{j_n} x - k).$$

As a consequence, if $\nu \in \text{bca}(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), dx)$, we may write

$$V(r_n, \nu) := \text{Var} \left(T_\nu \left(r_n (\tilde{f}_n - \mathbb{E}(\tilde{f}_n)) \right) \right) = \frac{r_n^2}{n} \mathbb{E} \left(\left[\int_{\mathbb{R}} Y_{n,1}(x) \frac{d\nu}{dx}(x) dx \right]^2 \right).$$

After straightforward computations, we get

$$V(r_n, \nu) = \frac{r_n^2}{n} \mathbb{E} \left(\left[\int_{\mathbb{R}} \{S(2^{j_n} X_1, y) - \mathbb{E}(S(2^{j_n} X_1, y))\} \frac{d\nu}{dx} \left(\frac{y}{2^{j_n}} \right) dy \right]^2 \right)$$

$$\text{with } S(x, y) = \sum_{k \in \mathbb{Z}} \Phi(x - k) \Phi(y - k).$$

If moreover Φ is bounded and compactly supported, then $|S(x, y)| \leq Q(y - x)$ where $Q : \mathbb{R} \rightarrow \mathbb{R}_+$ is bounded and compactly supported, see Lemma 8.6 in Härdle *et al.* (1998). Because Q is then integrable, this entails

$$V(r_n, \nu) \leq \frac{r_n^2}{n} \left[2 \|d\nu/dx\|_{\infty, \mathbb{R}} \int_{\mathbb{R}} Q(y) dy \right]^2 = O \left(\left[\frac{r_n}{\sqrt{n}} \right]^2 \right),$$

which is a bound similar to the one we had found in the proof of Theorem 1 in the case of the kernel density estimator. It appears then that Theorem 1 holds for wavelet density estimators on \mathbb{R} as well; in other words, any centered and rescaled version of the wavelet density estimator, if it converges to a tight Borel measurable weak limit in $\ell^\infty(\mathbb{R})$, must in fact converge essentially to 0. It is known though that the exact rate of almost sure convergence of the uniform norm $\|\tilde{f}_n - \mathbb{E}(\tilde{f}_n)\|_{\infty, \mathbb{R}}$ is $\sqrt{n/(j_n 2^{j_n})}$ under certain regularity conditions, see Theorem 2 in Giné and Nickl (2009). Classifying the weak behavior of $r_n(\tilde{f}_n - \mathbb{E}(\tilde{f}_n))$ in $\ell^\infty(\mathbb{R})$ can then likely be done as in Theorem 3 of the present paper for \hat{f}_n . The method of proof presented here seems therefore flexible enough to apply to, and yield the same results for, other density estimators than the Parzen-Rosenblatt estimator.

Acknowledgments

The author acknowledges both anonymous referees for their helpful and stimulating comments which led to significant enhancements of this article.

References

- Bickel, P.J., Rosenblatt, M. (1973). On some global measures of the deviations of density function estimates, *The Annals of Statistics* **1**(6): 1071–1095.
- Chernozhukov, V., Chetverikov, D., Kato, K. (2014). Gaussian approximation of suprema of empirical processes, *The Annals of Statistics* **42**(4): 1564–1597.
- Dony, J., Einmahl, U. (2006). Weighted uniform consistency of kernel density estimators with general bandwidth sequences, *Electronic Journal of Probability* **11**: 844–859.
- Doukhan, P., León, J.R. (1990). Déviation quadratique d’estimateurs de densité par projections orthogonales, *Comptes Rendus de l’Académie des Sciences - Series I - Mathematics* **310**: 425–430.
- Dunford, N., Schwartz, J. (1957). *Linear Operators Part I: General Theory*, John Wiley and Sons, New York.
- Einmahl, U., Mason, D.M. (2000). An empirical process approach to the uniform consistency of kernel-type function estimators, *Journal of Theoretical Probability* **13**(1): 1–37.
- Einmahl, U., Mason, D.M. (2005). Uniform in bandwidth consistency of kernel-type function estimators, *The Annals of Statistics* **33**(3): 1380–1403.
- Giné, E., Guillou, A. (2002). Rates of strong uniform consistency for multivariate kernel density estimators, *Annales de l’Institut Henri Poincaré (B) Probability and Statistics* **38**(6): 907–921.
- Giné, E., Koltchinskii, V., Zinn, J. (2004). Weighted uniform consistency of kernel density estimators, *The Annals of Probability* **32**(3B): 2570–2605.
- Giné, E., Nickl, R. (2008). Uniform central limit theorems for kernel density estimators, *Probability Theory and Related Fields* **141**(3): 333–387.
- Giné, E., Nickl, R. (2009). Uniform central limit theorems for wavelet density estimators, *The Annals of Probability* **37**(4): 1605–1646.
- Härdle, W., Kerkycharian, G., Picard, D., Tsybakov, A. (1998). *Wavelets, Approximation, and Statistical Applications*, Lecture Notes in Statistics, Volume 129, Springer, New York.
- Kallenberg, O. (1997). *Foundations of Modern Probability*, Springer, New York.
- Kerkycharian, G., Picard, D. (1992). Density estimation in Besov spaces, *Statistics and Probability Letters* **13**(1): 15–24.
- Khasminskii, R.Z. (1978). A lower bound on the risk of non-parametric estimates of densities in the uniform metric, *Theory of Probability & Its Applications* **23**(4): 794–798.
- Korostelev, A., Nussbaum, M. (1999). The asymptotic minimax constant for sup-norm loss in nonparametric density estimation, *Bernoulli* **5**(6): 1099–1118.
- Mendelson, S., Zinn, J. (2006). Modified empirical CLT’s under only pre-gaussian conditions. In *IMS Lecture Notes-Monograph Series: High Dimensional Probability*, Volume 51, pp.173–184.

- Nickl, R. (2007). Donsker-type theorems for nonparametric maximum likelihood estimators, *Probability Theory and Related Fields* **138**(3): 411–449.
- Nickl, R. (2009). On convergence and convolutions of random signed measures, *Journal of Theoretical Probability* **22**(1): 38–56.
- 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.
- Radulović, D., Wegkamp, M. (2000). Weak convergence of smoothed empirical processes: beyond Donsker classes. In: Giné, E., Mason, D.M., Wellner, J.A. (eds.) High dimensional probability II, Progress in Probability, Volume 47, pp.89–105, Birkhäuser, Boston.
- Radulović, D., Wegkamp, M. (2009). Uniform central limit theorems for pregaussian classes of functions. In IMS Collections: High Dimensional Probability V: The Luminy Volume, Volume 5, pp.84–102.
- Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function, *Annals of Mathematical Statistics* **27**(3): 832–837.
- Ruymgaart, F.H. (1998). A note on weak convergence of density estimators in Hilbert spaces, *Statistics* **30**(4): 331–343.
- Silverman, B.V. (1978). Weak and strong uniform consistency of the kernel estimate of a density and its derivatives, *The Annals of Statistics* **6**(1): 177–184.
- Stupfler, G. (2014). On the weak convergence of kernel density estimators in L^p spaces, *Journal of Nonparametric Statistics* **26**(4): 721–735.
- Stute, W. (1982). A law of the logarithm for kernel density estimators, *The Annals of Probability* **10**(2): 414–422.
- Stute, W. (1984). The oscillation behavior of empirical processes: the multivariate case, *The Annals of Probability* **12**(2): 361–379.
- Talagrand, M. (1994). Sharper bounds for Gaussian and empirical processes, *The Annals of Probability* **22**(1): 28–76.
- Talagrand, M. (1996). New concentration inequalities in product spaces, *Inventiones Mathematicae* **126**: 505–563.
- Van der Vaart, A.W. (1994). Weak convergence of smoothed empirical processes, *Scandinavian Journal of Statistics* **21**(4): 501–504.
- Van der Vaart, A.W. (1998). *Asymptotic Statistics*, Cambridge University Press, Cambridge.
- Van der Vaart, A.W., Wellner, J.A. (1996). *Weak Convergence and Empirical Processes with Applications to Statistics*, Springer-Verlag, New York.