UNIFORM IN BANDWIDTH CONSISTENCY OF KERNEL REGRESSION ESTIMATORS AT A FIXED POINT

JULIA DONY AND UWE EINMAHL

Abstract. We consider pointwise consistency properties of kernel regression function type estimators where the bandwidth sequence in not necessarily deterministic. In some recent papers uniform convergence rates over compact sets have been derived for such estimators via empirical process theory. We now show that it is possible to get optimal results in the pointwise case as well. The main new tool for the present work is a general moment bound for empirical processes which may be of independent interest.

1. Introduction

Let $(X_1, Y_1), (X_2, Y_2), \ldots$ be independent random variables in $\mathbb{R}^d \times \mathbb{R}$ with joint density function f_{XY}, and take $t \in \mathbb{R}^d$ fixed. Let further $\mathcal{F}$ be a class of measurable functions $\varphi : \mathbb{R} \to \mathbb{R}$ with $\mathbb{E}\varphi^2(Y) < \infty$, and consider the regression function $m_\varphi(t) = \mathbb{E}[\varphi(Y)|X = t]$. For any function $\varphi \in \mathcal{F}$ and a bandwidth $0 < h < 1$, define the kernel–type estimator

$$\hat{\varphi}_{n,h}(t) := \frac{1}{nh^d} \sum_{i=1}^{n} \varphi(Y_i) K\Big(\frac{t - X_i}{h}\Big),$$

where K is a kernel function, i.e. K is Borel measurable and $\int K(x)dx = 1$.

Such kernel estimators have been studied for many years, resulting in a considerable amount of literature and references. By choosing $\varphi \equiv 1$, one obtains an estimator for $f_X(t)$, the marginal density of X in $t \in \mathbb{R}^d$. This kernel density estimator (denoted by $\hat{f}_{n,h}(t)$) forms an important special case of the class of kernel estimators $\hat{\varphi}_{n,h}(t)$. It is well–known that for suitable (deterministic) bandwidth sequences h_n going to zero at an appropriate rate and assuming that the density f_X is continuous, one obtains a strongly consistent estimator $\hat{f}_{n,h_n}$ of f_X, i.e. one has with probability 1 that $\hat{f}_{n,h_n}(t) \to f_X(t)$ for all $t \in \mathbb{R}^d$ fixed. For proving such consistency results, one usually writes the difference $\hat{f}_{n,h_n}(t) - f_X(t)$ as the sum of a probabilistic term $\hat{f}_{n,h_n}(t) - \mathbb{E}\hat{f}_{n,h_n}(t)$, and a deterministic term $\mathbb{E}\hat{f}_{n,h_n}(t) - f_X(t)$, the so–called bias. The order of the bias depends on smoothness properties of f_X and of the kernel K, whereas the first (random) term can be studied via techniques based on empirical processes. Hall [14] proved an LIL type result for the probabilistic term corresponding to the kernel density estimator if $d = 1$. The d–dimensional version of this result implies in particular that under suitable conditions on the

The research of J. D. is financed by an FWO-Vlaanderen Postdoctoral Fellowship.
The research of U. E. was partially supported by an FWO–Vlaanderen grant.

bandwidth sequence h_n and the kernel function K, one has

$$\hat{f}_{n,h_n}(t) - \mathbb{E}\hat{f}_{n,h_n}(t) = O\left(\sqrt{\frac{\log\log n}{nh_n^d}}\right) \text{ a.s.} \tag{1.1}$$

Since this is an LIL type result (with corresponding lower bounds), this gives us the precise convergence rate for the pointwise convergence of the probabilistic term.

Deheuvels and Mason [2] later showed that this LIL holds whenever the bandwidth sequence satisfies

$$nh_n^d/\log\log n \to \infty, \qquad \text{as } n \to \infty, \tag{1.2}$$

which is the optimal condition under which (1.1) can hold. The work of [2] is based on a notion of a local empirical process indexed by sets. This was further generalized by Einmahl and Mason [7, 8] who looked at local empirical processes indexed by functions and who established strong invariance principles for such processes. They inferred LIL type results from the strong invariance principles and this not only for density estimators, but also for the Nadaraya–Watson estimator for the regression function and conditional empirical processes.

Recall that the Nadaraya–Watson estimator $\hat{m}_{n,h,\varphi}(t)$ for $m_\varphi(t) = \mathbb{E}[\varphi(Y)|X = t]$, where $\varphi : \mathbb{R} \to \mathbb{R}$ is Borel measurable, is defined as

$$\hat{m}_{n,h,\varphi}(t) := \hat{\varphi}_{n,h}(t)/\hat{f}_{n,h}(t).$$

There is also an LIL for this estimator (see [12] for a first result in this direction) which implies that

$$\hat{m}_{n,h_n,\varphi}(t) - \widehat{\mathbb{E}}\hat{m}_{n,h_n,\varphi}(t) = O\left(\sqrt{\frac{\log\log n}{nh_n^d}}\right) \text{ a.s.,} \tag{1.3}$$

where $\widehat{\mathbb{E}}\hat{m}_{n,h,\varphi}(t) = \mathbb{E}\hat{\varphi}_{n,h}(t)/\mathbb{E}\hat{f}_{n,h}(t)$ is a convenient centering term.
If φ is a bounded function this holds again under the above condition (1.2). If $\mathbb{E}[|\varphi(Y)|^p|X = x]$ is uniformly bounded in a neighborhood of t, where $p > 2$, one needs that $\{h_n^d\}$ is at least of order $O(n^{-1}(\log n)^q)$ for some $q > 2/(p-2)$ (see [7, 8]). From our main result it will actually follow that in this last case $q \geq 2/(p-2)$ is already sufficient.

Some related results have also been obtained for uniform convergence of kernel type estimators on compact subsets or even on $\mathbb{R}^d$. In this case one typically gets a slightly worse convergence rate of the probabilistic term of order $O(\sqrt{\log n/nh_n^d})$ and one needs that the bandwidth sequence satisfies $nh_n^d/\log n \to \infty$ in the bounded case, which is more restrictive than (1.2). For more details see [13, 9, 11] and the references in these papers.

In practice, one has to choose a bandwidth sequence h_n in such a way that the bias and the probabilistic part are reasonably balanced. The optimal choice for h_n then will often depend on some unknown parameter of the distribution which one has to estimate. This can lead to bandwidth sequences depending on the data and the location t. This means that the above results do not apply if one is interested in estimators with such general bandwidth sequences. In [10] "uniform in h" versions of the results in [9, 11] were obtained. This makes it possible to establish consistency of kernel–type estimators when the bandwidth h is allowed to range in an interval which may increase or decrease in length with the sample size. These kinds of results are immediately applicable to proving uniform consistency of kernel–type estimators when the bandwidth h is a function of the data $X_1, \dots, X_n$ or the location $t \in \mathbb{R}^d$.

A typical result of [10] is the following asymptotic result concerning the Nadaraya–Watson estimator which holds under certain conditions on the distribution of (X, Y) and the kernel K.

$$\limsup_{n\to\infty} \sup_{(\frac{c\log n}{n})^{\gamma/d}\le h<1} \sup_{t\in I} \sup_{\varphi\in\mathcal{F}} \frac{\sqrt{nh^d}|\hat{m}_{n,h,\varphi}(t) - \widehat{\mathbb{E}}\hat{m}_{n,h,\varphi}(t)|}{\sqrt{|\log h| \vee \log\log n}} < \infty, \quad \text{a.s.,} \tag{1.4}$$

where $\gamma = 1$ or $\gamma = 1-2/p$ depending on whether $\mathcal{F}$ is a bounded class of functions, or if it has an envelope function with a finite p–th moment $(p > 2)$. Here, I is a compact rectangle of $\mathbb{R}^d$ on which f_X has to be bounded and strictly positive. We call this an asymptotic uniform in bandwidth (AUiB) boundedness result. It implies that if one chooses the bandwidth depending on the data and/or the location (as is usually done in practice), one keeps the same order of convergence as the one valid for a deterministic bandwidth sequence, given, for instance, in [9].

The purpose of the present paper is to establish a similar AUiB boundedness result in the "pointwise" setting, where in view of the aforementioned results, one can hope for a slightly smaller order and bigger intervals from which one can choose the bandwidth sequence. Pointwise AUiB boundedness results can be useful in various contexts. In particular, they can be used for deriving consistency results for generalized Hill type estimators introduced in [1]. (See [5] and Chapter 6 in [3].)

2. Main result

Before stating our main result, we have to impose several assumptions on the kernel function, the bandwidth and the class $\mathcal{F}$. These assumptions are mainly technical, and will be listed below.

We first recall some terminology. Let $(\mathcal{X}, \mathcal{A})$ be a measurable space. We say that a class $\mathcal{G}$ of functions $g : \mathcal{X} \to \mathbb{R}$ is *pointwise measurable* if there exists a countable subclass $\mathcal{G}_0$ of $\mathcal{G}$ such that we can find for any function $g \in \mathcal{G}$ a sequence of functions $g_m \in \mathcal{G}_0$ for which $g_m(z) \to g(z), z \in \mathcal{X}$. This property is usually assumed to avoid measurability problems, and is discussed in [17]. Next, we call a class of functions $\mathcal{G}$ with envelope function $G : \mathcal{X} \to [0, \infty]$ a *VC–class*, if $\mathcal{N}(\epsilon, \mathcal{G}) \le C\epsilon^{-\nu}$ for some constants $C, \nu > 0$, that are called the *characteristics* of the class. As usual we define

$$\mathcal{N}(\epsilon, \mathcal{G}) = \sup_Q \mathcal{N}(\epsilon\sqrt{Q(G^2)}, \mathcal{G}, d_Q),$$

where the supremum is taken over all probability measures Q on $(\mathcal{X}, \mathcal{A})$ with $Q(G^2) < \infty$. d_Q is the $L_2(Q)$–metric and $\mathcal{N}(\epsilon, \mathcal{G}, d)$ is the minimal number of d–balls with radius ϵ which are needed to cover the function class $\mathcal{G}$. An envelope function is any $\mathcal{A}$–measurable function $G : \mathcal{X} \to [0, \infty]$ such that

$$\sup_{g \in \mathcal{G}} |g(x)| \le G(x), x \in \mathcal{X}.$$

We are now ready to state the conditions that need to be imposed for our results. Let $\mathcal{F}$ be a class of functions $\varphi : \mathbb{R} \to \mathbb{R}$ satisfiying the following three conditions :

(F.i) $\mathcal{F}$ is a pointwise measurable class,

(F.ii) $\mathcal{F}$ has a (measurable) envelope function $F(y) \ge \sup_{\varphi \in \mathcal{F}} |\varphi(y)|, y \in \mathbb{R}$,

(F.iii) $\mathcal{F}$ is a VC–class.

Next, a kernel function $K : \mathbb{R}^d \to \mathbb{R}$ will be any measurable function satisfying

(K.i) $\|K\|_\infty = \kappa < \infty$ and $\int K(x)dx = 1$,

(K.ii) K has a support contained in $[-1/2, 1/2]^d$,

(K.iii) $\mathcal{K} := \{x \mapsto K(\gamma(t - x)) : \gamma > 0\}$ is a pointwise measurable VC–class of functions from $\mathbb{R}^d$ to $\mathbb{R}$.

Conditions (K.i) and (K.ii) are easy to verify. Many kernels satisfy also condition (K.iii). (See for instance Remark 1 in [8] for some discussion.)

Our main result is then as follows.

Proposition 2.1 (**Pointwise AUiB boundedness of kernel–type estimators**). *Let $\mathcal{F}$ and K satisfy (F) and (K) and assume that the envelope function F of $\mathcal{F}$ satisfies for some $0 < \epsilon < 1$ one of the following conditions on $J := t + [-\epsilon, \epsilon]^d$:*

(A) $\exists\, p > 2 : \sup_{x \in J} \mathbb{E}[F^p(Y)|X = x] =: \mu_p < \infty$.

(B) $\exists\, s > 0 : \sup_{x \in J} \mathbb{E}\left[\exp(sF(Y))|X = x\right] < \infty$.

Then if f_X is bounded on J it follows for any $c > 0$ that

$$\limsup_{n\to\infty} \sup_{a_n \le h \le b_0} \sup_{\varphi \in \mathcal{F}} \frac{\sqrt{nh^d}|\hat{\varphi}_{n,h}(t) - \mathbb{E}\hat{\varphi}_{n,h}(t)|}{\sqrt{\log\log n}} < \infty, \quad a.s., \tag{2.1}$$

where $0 < b_0 < 2\epsilon$ is a positive constant and $a_n^d = cn^{-1}(\log n)^{\frac{2}{p-2}}$ or $a_n^d = cn^{-1}\log\log n$ depending on whether condition (A) or condition (B) holds.

Note that Proposition 2.1(B) is more general than the corresponding result in [10] where one has to assume that the function class $\mathcal{F}$ is bounded. The above Proposition 2.1 provides a result for all classes whose envelope function admits a finite moment generating function. This improvement is possible since in the present case we can use an exponential Bernstein type inequality in terms of a strong second moment of the envelope function. For establishing an AUiB boundedness result uniformly on compact rectangles of $\mathbb{R}^d$ one would need this inequality in terms of the weak second moment. Such an improvement of the Bernstein inequality, however, seems to be only known in the bounded case. This means that, concerning uniform in bandwidth pointwise consistency, there is no distinction between the bounded case and the case where the moment generating function of F is finite. Also in the case where one uses deterministic bandwidth sequences, this result seems to be new.

We conclude this section with formulating two corollaries of our proposition. We first look at the kernel estimator for the density f_X. If the kernel K satistisfies the above conditions, we can conclude for any sequence a_n such that $na_n^d/\log\log n \to \infty$ and $a_n \le b_n$, where $b_n \to 0$,

$$\sup_{a_n \le h \le b_n} |\hat{f}_{n,h}(t) - \mathbb{E}\hat{f}_{n,h}(t)| = O\left(\sqrt{\frac{\log\log n}{na_n^d}}\right) \text{ a.s.,}$$

provided that f_X is bounded on J.

Assume now that this density is positive and continuous on J, and $\mathcal{F}$ is a class of functions satisfying conditions (F) and (A). If $\liminf_{n\to\infty} na_n^d/(\log n)^{2/(p-2)} > 0$, and $a_n \le b_n \to 0$, we obtain by a standard argument that

$$\sup_{a_n \le h \le b_n} \sup_{\varphi \in \mathcal{F}} |\hat{m}_{n,h,\varphi}(t) - \widehat{\mathbb{E}}\hat{m}_{n,h,\varphi}(t)| = O\left(\sqrt{\frac{\log\log n}{na_n^d}}\right) \text{ a.s.}$$

As in [10], the proof of Proposition 2.1 is based on the theory of empirical processes. We use again exponential deviation inequalities in combination with certain moment inequalities. The necessary exponential inequalities are available in the literature, but we need a new moment inequality. This will be stated and proved in Section 3. In Section 4 we shall finally prove Proposition 2.1.

3. Moment inequalities

Let $X, X_1, \dots, X_n$ be i.i.d. random variables taking values in a measurable space $(\mathcal{X}, \mathcal{A})$ and let $A \in \mathcal{A}$ be a fixed set. It is our goal to derive a moment inequality for

$$\mathbb{E}\big\|\alpha_n(g \cdot \mathbf{I}_A)\big\|_{\mathcal{G}}$$

where α_n is the empirical process based on $X_1, \dots, X_n$ and $\mathcal{G}$ is a pointwise measurable class of functions $g : \mathcal{X} \to \mathbb{R}$ for which $\mathbb{E}g^2(X)$ exists. Let $G : \mathcal{X} \to [0, \infty]$ be an envelope function for the function class $\mathcal{G}$ and assume that $\mathbb{E}G^2(X) < \infty$. We further assume that $\mathcal{G}$ has the following property:

> *For any sequence of i.i.d. $\mathcal{X}$–valued random variables $Z_1, Z_2, \dots$ it holds that*
>
> $$(\triangle) \qquad \mathbb{E}\Big\|\sum_{i=1}^{k}\{g(Z_i) - \mathbb{E}g(Z_1)\}\Big\|_{\mathcal{G}} \le C_1\sqrt{k}\|G(Z_1)\|_2, \quad 1 \le k \le n,$$
>
> *where $C_1 > 1$ is a constant depending on $\mathcal{G}$ only.*

From Theorem 3.2 below it will follow that VC–classes always have this property. But we first prove our moment inequality.

Theorem 3.1. *Let $\mathcal{G}$ be a pointwise measurable function class satisfying the above assumptions. Then we have for any $A \in \mathcal{A}$,*

$$\mathbb{E}\big\|\alpha_n(g \cdot \mathbf{I}_A)\big\|_{\mathcal{G}} \le 2C_1\|G(X)\mathbf{I}_A(X)\|_2.$$

Proof. Similarly as in [7, 8] we shall use a special representation of the random variables $X_i, i \ge 1$. W.l.o.g. we assume that $0 < \mathbb{P}(A) < 1$. Consider independent random variables $Y_1, Y_2, \dots, Y_1', Y_2', \dots$ such that for all $B \in \mathcal{A}$ and any $i \ge 1$,

$$\mathbb{P}\{Y_i \in B\} = \mathbb{P}\{X \in B | X \in A\} \quad \text{and} \quad \mathbb{P}\{Y_i' \in B\} = \mathbb{P}\{X \in B | X \in A^c\}.$$

Let further $\epsilon_1, \epsilon_2, \ldots$ be independent Bernoulli($\mathbb{P}\{X \in A\}$)–variables, independent of the two other sequences, and set $\nu(n) := \sum_{i=1}^n \epsilon_i$. Finally, define for any $i \geq 1$,

$$X_i^* = \begin{cases} Y_{\nu(i)}, & \text{if } \epsilon_i = 1, \\ Y'_{i-\nu(i)}, & \text{if } \epsilon_i = 0. \end{cases}$$

Then it is easy to see that this leads to a sequence of independent random variables with $X_i^* \stackrel{d}{=} X_i, i \geq 1$. Consequently, it is sufficient to prove the moment bound for the empirical process α_n^* based on the variables $X_i^*, i \geq 1$. Moreover, it is readily seen that

$$\sum_{i=1}^n g(X_i^*)\mathbb{1}_A(X_i^*) = \sum_{i=1}^{\nu(n)} g(Y_i),$$

and also that $\mathbb{E}[g(X^*)\mathbb{1}_A(X^*)] = \mathbb{E}g(Y_1)\mathbb{P}\{X \in A\}$. Consequently, we have that

$$\begin{aligned}
\mathbb{E}\big\|\sqrt{n}\alpha_n^*(g \cdot \mathbb{1}_A)\big\|_{\mathcal{G}} &= \mathbb{E}\Big\|\sum_{i=1}^{\nu(n)} g(Y_i) - n\mathbb{P}\{X \in A\}\mathbb{E}g(Y_1)\Big\|_{\mathcal{G}} \\
&\leq \mathbb{E}\Big\|\sum_{i=1}^{\nu(n)} \big(g(Y_i) - \mathbb{E}g(Y_i)\big)\Big\|_{\mathcal{G}} \\
&\quad + \mathbb{E}\Big\|\sum_{i=1}^{\nu(n)} \mathbb{E}g(Y_i) - n\mathbb{P}\{X \in A\}\mathbb{E}g(Y_1)\Big\|_{\mathcal{G}} \\
&\leq \mathbb{E}\big\|\sqrt{\nu(n)}\widetilde{\alpha}_{\nu(n)}(g)\big\|_{\mathcal{G}} + \mathbb{E}\big|\nu(n) - n\mathbb{P}\{X \in A\}\big| \cdot \sup_{g \in \mathcal{G}} \mathbb{E}|g(Y_1)|,
\end{aligned}$$

where $\widetilde{\alpha}_n(g)$ denotes the empirical process based upon $Y_1 \ldots, Y_n$. Recall that $\nu(n)$ has Binomial($n, \mathbb{P}\{X \in A\}$) distribution so that $\mathbb{E}\nu(n) = n\mathbb{P}\{X \in A\}$, and thus $\mathbb{E}|\nu(n) - n\mathbb{P}\{X \in A\}| \leq \mathrm{Var}(\nu(n))^{1/2} \leq \sqrt{n\mathbb{P}\{X \in A\}}$. Recall also that G is an envelope function of $\mathcal{G}$. Hence, we have that

$$\mathbb{E}\big\|\sqrt{n}\alpha_n^*(g \cdot \mathbb{1}_A)\big\|_{\mathcal{G}} \leq \mathbb{E}\big\|\sqrt{\nu(n)}\widetilde{\alpha}_{\nu(n)}(g)\big\|_{\mathcal{G}} + \sqrt{n\mathbb{P}\{X \in A\}\mathbb{E}G^2(Y_1)}.$$

We now look at the first term. Due to assumption ($\triangle$) and by independence of $\nu(n)$ and the variables $Y_1, Y_2, \ldots$, we can conclude that (note that $\nu(n) \leq n$),

$$\begin{aligned}
\mathbb{E}\big\|\sqrt{\nu(n)}\widetilde{\alpha}_{\nu(n)}(g)\big\|_{\mathcal{G}} &\leq \sum_{k=1}^n \mathbb{E}\big\|\sqrt{k}\widetilde{\alpha}_k(g)\big\|_{\mathcal{G}}\, \mathbb{P}\{\nu(n) = k\} \\
&\leq C_1 \sum_{k=1}^n \sqrt{k\mathbb{E}G^2(Y_1)}\, \mathbb{P}\{\nu(n) = k\} \\
&= C_1\sqrt{\mathbb{E}G^2(Y_1)}\mathbb{E}[\nu^{1/2}(n)] \\
&\leq C_1\sqrt{n\mathbb{P}\{X \in A\}\mathbb{E}G^2(Y_1)},
\end{aligned}$$

where we have used the trivial fact that $\mathbb{E}[\nu^{1/2}(n)] \leq (\mathbb{E}[\nu(n)])^{1/2} = \sqrt{n\mathbb{P}\{X \in A\}}$. Recalling that $\mathbb{E}G^2(Y_1) = \mathbb{E}[G^2(X)\mathbb{1}_A(X)]/\mathbb{P}\{X \in A\}$, we can conclude that

$$\mathbb{E}\big\|\sqrt{n}\alpha_n^*(g \cdot \mathbb{1}_A)\big\|_{\mathcal{G}} \leq 2C_1\sqrt{n\mathbb{E}[G^2(X)\mathbb{1}_A(X)]},$$

proving that the moment bound holds, as claimed. □

The next result shows that condition $(\triangle)$ is satisfied if we have a VC–class.

Theorem 3.2. *Let $\mathcal{G}$ be a pointwise measurable VC–class of functions with envelope function G and characteristics $A, \nu \geq 1$. If $Z, Z_1, Z_2, \ldots$ is a sequence of i.i.d. $\mathcal{X}$–valued random variables satisfying for some $0 < \beta < \infty$, $\mathbb{E}G^2(Z) \leq \beta^2$, then we have for some constant C depending on A, ν only that*

$$\mathbb{E}\Big\|\sum_{i=1}^n (g(Z_i) - \mathbb{E}g(Z))\Big\|_{\mathcal{G}} \leq C\sqrt{n\beta^2}, \quad n \geq 1.$$

Proof. Let $\varepsilon_1, \ldots, \varepsilon_n$ be i.i.d. Rademacher variables which are also independent of $Z_1, \ldots, Z_n$. Then we have,

$$\mathbb{E}\Big\|\sum_{i=1}^n (g(Z_i) - \mathbb{E}g(Z))\Big\|_{\mathcal{G}} \leq 2\mathbb{E}\Big\|\sum_{i=1}^n \varepsilon_i g(Z_i)\Big\|_{\mathcal{G}},$$

and it is sufficient to show that

$$\mathbb{E}\Big\|\sum_{i=1}^n \varepsilon_i g(Z_i)\Big\|_{\mathcal{G}} \leq C'\sqrt{n\beta^2}, \tag{3.1}$$

where C' is a positive constant depending on A and ν only. From the Hoffmann–Jørgensen inequality (see Proposition 6.8. in [16]) it follows that

$$\mathbb{E}\Big\|\sum_{i=1}^n \varepsilon_i g(Z_i)\Big\|_{\mathcal{G}} \leq 6t_0 + 6\mathbb{E}\big[\max_{1\leq i\leq n} G(Z_i)\big], \tag{3.2}$$

where

$$t_0 := \inf_{t>0}\left(\mathbb{P}\Big\{\|\sum_{i=1}^n \varepsilon_i g(Z_i)\|_{\mathcal{G}} > t\Big\} \leq \frac{1}{24}\right).$$

Observing that

$$\mathbb{E}\big[\max_{1\leq i\leq n} G(Z_i)\big] \leq \Big(\mathbb{E}\big[\max_{1\leq i\leq n} G^2(Z_i)\big]\Big)^{1/2} \leq \Big(\mathbb{E}\big[\sum_{i=1}^n G^2(Z_i)\big]\Big)^{1/2} \leq \sqrt{n\beta^2},$$

we see that it is sufficient to show that

$$t_0 \leq C''\sqrt{n\beta^2}. \tag{3.3}$$

Let μ be the distribution of the variable $Z : \Omega \to \mathcal{X}$ and define

$$G_n := \Big\{\mathbf{x} \in \mathcal{X}^n : \sum_{i=1}^n G^2(x_i) \leq 64n\beta^2\Big\}.$$

Note that for any $t > 0$,

$$\begin{aligned}\mathbb{P}\Big\{\|\sum_{i=1}^n \varepsilon_i g(Z_i)\|_{\mathcal{G}} > t\Big\} &= \int_{\mathcal{X}^n} \mathbb{P}\Big\{\|\sum_{i=1}^n \varepsilon_i g(x_i)\|_{\mathcal{G}} > t\Big\}\mu^n(d\mathbf{x}) \\ &\leq \quad \mu^n(G_n^c) + \int_{G_n} \mathbb{P}\Big\{\|\sum_{i=1}^n \varepsilon_i g(x_i)\|_{\mathcal{G}} > t\Big\}\mu^n(d\mathbf{x}) \\ &=: \quad \alpha_1 + \alpha_2.\end{aligned}$$

From Markov's inequality we obtain that $\alpha_1 = \mathbb{P}\big\{\sum_{i=1}^n G^2(Z_i) > 64n\beta^2\big\} \leq 1/64$. To bound α_2, we use a well known inequality of Jain and Marcus [15] which is also stated as Corollary 2.2.8 in [17].

We can conclude that for any $\mathbf{x} = (x_1, \ldots, x_n) \in \mathcal{X}^n$ and some absolute constant $c_0 < \infty$,

$$\mathbb{E}\Big\|\sum_{i=1}^n \varepsilon_i g(x_i)\Big\|_{\mathcal{G}} \le \mathbb{E}\Big|\sum_{i=1}^n \varepsilon_i g_0(x_i)\Big| + c_0\sqrt{n}\int_0^\infty \sqrt{\log \mathcal{N}(\epsilon, \mathcal{G}, d_{2,\mathbf{x}})}\, d\epsilon,$$

where $g_0 \in \mathcal{G}$ is arbitrary and $d_{2,\mathbf{x}}^2(g_1, g_2) := n^{-1}\sum_{i=1}^n (g_1(x_i) - g_2(x_i))^2$. Further, it is easy to infer that when $\mathbf{x} \in G_n$,

$$\mathbb{E}\Big|\sum_{i=1}^n \varepsilon_i g_0(x_i)\Big| \le \left(\sum_{i=1}^n g_0^2(x_i)\right)^{1/2} \le 8\sqrt{n\beta^2},$$

and for $g_1, g_2 \in \mathcal{G}$,

$$d_{2,\mathbf{x}}^2(g_1, g_2) \le \frac{2}{n}\sum_{i=1}^n (g_1^2(x_i) + g_2^2(x_i)) \le 256\beta^2.$$

Hence, if $\epsilon > 16\beta$, one needs only one ball of $d_{2,\mathbf{x}}$–radius ϵ to cover the class $\mathcal{G}$. Therefore, $\mathcal{N}(\epsilon, \mathcal{G}, d_{2,\mathbf{x}}) = 1$ whenever $\mathbf{x} \in G_n$ and $\epsilon > 16\beta$.
On the other hand, let $Q_{n,\mathbf{x}}(f) := n^{-1}\sum_{i=1}^n f(x_i)$ and note that $Q_{n,\mathbf{x}}((g_1 - g_2)^2) = d_{2,\mathbf{x}}^2(g_1, g_2)$. Then since $Q_{n,\mathbf{x}}(G^2) \le 256\beta^2$ for $\mathbf{x} \in G_n$, and recalling that $\mathcal{N}(\epsilon, \mathcal{G}) = \sup_Q \mathcal{N}(\epsilon\sqrt{Q(G^2)}, \mathcal{G}, d_Q)$ where d_Q is the $L_2(Q)$–metric, the assumption that $\mathcal{G}$ is a VC–class gives us for any $\mathbf{x} \in G_n$ and whenever $0 < \epsilon \le 16\beta$,

$$\mathcal{N}(\epsilon, \mathcal{G}, d_{2,\mathbf{x}}) \quad\le\quad \mathcal{N}\Big(\frac{\epsilon\sqrt{Q_{n,\mathbf{x}}(G^2)}}{16\beta}, \mathcal{G}, d_{2,\mathbf{x}}\Big) \le \mathcal{N}\Big(\frac{\epsilon}{16\beta}, \mathcal{G}\Big) \le A\epsilon^{-\nu}(16\beta)^\nu.$$

Hence, we have for $\mathbf{x} \in G_n$,

$$\begin{aligned}\int_0^\infty \sqrt{\log \mathcal{N}(\epsilon, \mathcal{G}, d_{2,\mathbf{x}})}\, d\epsilon &= \int_0^{16\beta} \sqrt{\log \mathcal{N}(\epsilon, \mathcal{G}, d_{2,\mathbf{x}})}\, d\epsilon \\ &\le \int_0^{16\beta} \sqrt{\log A(16\beta/\epsilon)^\nu}\, d\epsilon \\ &= \sqrt{\nu}A^{1/\nu}16\beta \int_0^{A^{-1/\nu}} \sqrt{\log\frac{1}{s}}ds,\end{aligned}$$

which can be bounded by $16c\sqrt{\nu}A^{1/\nu}\beta$, with $c := \int_0^1 \sqrt{\log 1/s}\, ds < \infty$. (Recall that $A \ge 1$.) Consequently, by Markov's inequality we have for any $\mathbf{x} \in G_n$:

$$\mathbb{P}\Big\{\|\sum_{i=1}^n \varepsilon_i g(x_i)\|_{\mathcal{G}} > t\Big\} \le t^{-1}\{8\sqrt{n\beta^2} + c_0 c\sqrt{\nu}A^{1/\nu}16\beta\} = \sqrt{64n\beta^2}(1 + 2c_0c_1)/t,$$

with $c_1 = c\sqrt{\nu}A^{1/\nu}$. Taking everything together, we obtain that

$$\mathbb{P}\Big\{\|\sum_{i=1}^n \varepsilon_i g(Z_i)\|_{\mathcal{G}} > t\Big\} \le \frac{1}{64} + \frac{\sqrt{64n\beta^2}}{t}(1 + 2c_0c_1).$$

To finish, recall from (3.3) that we need to find a range for $t > 0$ such that the above probability is bounded by $1/24$. By solving the equation, we see that t should be such that $\sqrt{64n\beta^2}(1 + 2c_0c_1) \le 5t/(3\cdot 2^6)$, or

$$t \ge \frac{3\cdot 2^9\sqrt{n\beta^2}(1 + 2c_0c_1)}{5}.$$

Consequently, by setting $C'' = 3 \cdot 2^7(1 + 2c_0c_1)$ and taking $t \geq C''\sqrt{n\beta^2}$, we have shown that $\mathbb{P}\{\| \sum_{i=1}^n \varepsilon_i g(Z_i)\|_{\mathcal{G}} > t\} \leq \frac{1}{24}$, proving the theorem through (3.3) and the Hoffmann–Jørgensen inequality. □

Note that from our assumptions on $\mathcal{F}$ and the kernel K it follows that the subsequent class of functions on $\mathcal{X} = \mathbb{R}^d \times \mathbb{R}$

$$\mathcal{G} := \left\{ (x, y) \mapsto \varphi(y)K\Big(\frac{t - x}{h}\Big) : \varphi \in \mathcal{F}, h > 0 \right\} \quad \text{is a VC–class} \tag{3.4}$$

with envelope function $G(x, y) = \kappa F(y)$. (Use, for instance, Lemma A.1 in [9].) Consequently, Theorem 3.2 ensures the class $\mathcal{G}$ to satisfy condition ($\triangle$), and thereby all the conditions of Theorem 3.1.

4. Proof of Proposition 2.1

To start the proof of Proposition 2.1, we first show how the process in (2.1) can be expressed in terms of an empirical process indexed by a certain class of functions. To do so, consider the following classes of functions on $\mathcal{X} = \mathbb{R}^d \times \mathbb{R}$ defined by

$$\mathcal{G}_n := \left\{ (x, y) \mapsto \varphi(y)K\Big(\frac{t - x}{h}\Big) : \varphi \in \mathcal{F}, a_n \leq h \leq b_0 \right\},$$

and note that for any $\varphi \in \mathcal{F}$ and $a_n \leq h \leq b_0$,

$$\begin{aligned} \hat{\varphi}_{n,h}(t) - \mathbb{E}\hat{\varphi}_{n,h}(t) &= \frac{1}{nh^d}\Big[\sum_{i=1}^n \varphi(Y_i)K\Big(\frac{t - X_i}{h}\Big) - n\mathbb{E}\varphi(Y)K\Big(\frac{t - X}{h}\Big)\Big] \\ &=: \frac{1}{\sqrt{n}h^d}\alpha_n(g_{\varphi,h}), \end{aligned}$$

where $g_{\varphi,h}(x, y) := \varphi(y)K((t - x)/h)$ and $\alpha_n(g)$ is the empirical process based upon the sample $(X_1, Y_1), \ldots, (X_n, Y_n)$. Further, set $n_k := 2^k, k \geq 0$ and define $h_{k,j}^d := 2^j a_{n_k}^d$. Then $h_{k,0} = a_{n_k}$ and by setting $L(k) := \max\{j : h_{k,j}^d \leq 2b_0^d\}$, it holds that $h_{k,L(k)-1} \leq b_0 < h_{k,L(k)}$ so that $[a_{n_k}, b_0] \subset [h_{k,0}, h_{k,L(k)}]$. Further, consider for $1 \leq j \leq L(k)$ the subclasses

$$\mathcal{G}_{k,j} := \left\{ (x, y) \mapsto \varphi(y)K\Big(\frac{t - x}{h}\Big) : \varphi \in \mathcal{F}, h_{k,j-1} \leq h \leq h_{k,j} \right\},$$

and note that $\mathcal{G}_{n_k} \subset \bigcup_{j=1}^{L(k)} \mathcal{G}_{k,j}$. Since a_n is eventually non–decreasing, we have for all $n_{k-1} < n \leq n_k$ and for any $\varphi \in \mathcal{F}$ if $k \geq 1$ is large enough,

$$\begin{aligned} \sup_{a_n \leq h \leq b_0} \frac{\sqrt{nh^d}|\hat{\varphi}_{n,h}(t) - \mathbb{E}\hat{\varphi}_{n,h}(t)|}{\sqrt{\log\log n}} &\leq \sup_{a_{n_k} \leq h \leq b_0} \frac{\|\alpha_n(g)\|_{\mathcal{G}_{n_k}}}{\sqrt{h^d \log\log n}} \\ &\leq \max_{1 \leq j \leq L(k)} \frac{2\sqrt{2}\|\sqrt{n}\alpha_n(g)\|_{\mathcal{G}_{k,j}}}{\sqrt{n_k h_{k,j}^d \log\log n_k}}. \end{aligned} \tag{4.1}$$

Recall further that the class $\mathcal{G}$ in (3.4) is a VC–class with envelope $G(x, y) = \kappa F(y)$. This of course implies that also $\mathcal{G}_{k,j}$ is a VC–class for this envelope function and with characteristics independent of k and j.

In view of Theorem 3.2, the assumptions of Theorem 3.1 are satisfied and we have for any subset A of $\mathbb{R}^d \times \mathbb{R}$,

$$\mathbb{E}\big\|\alpha_{n_k}(g \cdot \mathbb{I}_A)\big\|_{\mathcal{G}_{k,j}} \le 2C'\|G(X,Y)\mathbb{I}_A(X,Y)\|_2,$$

where C' is a positive constant (depending on the characteristics of $\mathcal{G}$ only). Setting $A_{k,j} = \{t + [-h_{k,j}/2, h_{k,j}/2]^d\} \times \mathbb{R}$, we can conclude that

$$\mathbb{E}\|\alpha_{n_k}(g)\|_{\mathcal{G}_{k,j}} = \mathbb{E}\big\|\alpha_{n_k}(g \cdot \mathbb{I}_{A_{k,j}})\big\|_{\mathcal{G}_{k,j}} \le C'\left(\mathbb{E}G^2_{k,j}(X,Y)\right)^{1/2}, \tag{4.2}$$

where $G_{k,j}(x,y) = \kappa F(y)\mathbb{I}\{x \in t + [-h_{k,j}/2, h_{k,j}/2]^d\}$ is another envelope function for $\mathcal{G}_{k,j}$.

4.1. **Proof under condition (A).** We use the empirical process version of a recent Fuk–Nagaev type inequality in Banach spaces. (See Theorem 3.1 in [6].) By a slight misuse of notation, we also write α_n for the empirical process based on the sample $Z_1, \ldots, Z_n$ in a general measurable space $(\mathcal{X}, \mathcal{A})$. (We shall apply this inequality on $\mathcal{X} = \mathbb{R}^d \times \mathbb{R}$ and with $Z_i = (X_i, Y_i), i \ge 1$.)

Fact 4.1 (**Fuk–Nagaev type inequality**). *Let $Z, Z_1, \ldots, Z_n$ be i.i.d. $\mathcal{X}$–valued random variables and consider a pointwise measurable class $\mathcal{G}$ of functions $g : \mathcal{X} \to \mathbb{R}$ with envelope function G. Assume that for some $p > 2$, $\mathbb{E}G^p(Z) < \infty$. Then we have for $0 < \eta \le 1, \delta > 0$ and any $t > 0$,*

$$\begin{aligned}&\mathbb{P}\left\{\max_{1\le k\le n} \|\sqrt{k}\alpha_k(g)\|_{\mathcal{G}} \ge (1+\eta)\beta_n + t\right\}\\ &\qquad \le \exp\left(-\frac{t^2}{(2+\delta)n\sigma^2}\right) + nC_2\mathbb{E}G^p(Z)|/t^p,\end{aligned}$$

where $\sigma^2 = \sup_{g\in\mathcal{G}} \mathbb{E}g^2(Z)$, $\beta_n = \mathbb{E}\|\sqrt{n}\alpha_n(g)\|_{\mathcal{G}}$ and C_2 is a positive constant depending on η, δ and p.

Let $\kappa = \sup_{x\in\mathbb{R}^d} |K(x)|$ and recall that by (K.ii) the support of K lies in $[-1/2, 1/2]^d$. Recall further that f_X is bounded on $J = t + [-\epsilon, \epsilon]^d$. Then for any $g_{\varphi,h} \in \mathcal{G}_{k,j}$ with k large enough such that $h_{k,j} \le 2\epsilon$, it holds that

$$\begin{aligned}\mathbb{E}(g_{\varphi,h}(X,Y))^2 &= \mathbb{E}\Big[\varphi^2(Y)K^2\Big(\frac{t-X}{h}\Big)\Big]\\ &\le h^d\kappa^2 \int_{[-\frac12,\frac12]^d} \mathbb{E}\left[F^2(Y)|X = t - uh\right] f_X(t-uh)du\\ &\le \kappa^2\|f_X\|_J \int_{[-\frac{h_{k,j}}{2},\frac{h_{k,j}}{2}]^d} \mathbb{E}\left[F^p(Y)|X = t + x\right]^{2/p} dx,\end{aligned}$$

which on account of (A) implies that for k sufficiently large,

$$\sup_{g\in\mathcal{G}_{k,j}} \mathbb{E}g^2(X,Y) \le h^d_{k,j}\kappa^2\|f_X\|_J\mu_p^{2/p} =: \sigma^2_{k,j}.$$

In the same way, we obtain that

$$\mathbb{E}\|g(X,Y)\|^q_{\mathcal{G}_{k,j}} \le \mathbb{E}G^q_{k,j}(X,Y) \le C_q\sigma^2_{k,j}, \quad 2 \le q \le p,$$

where $C_q = \kappa^{q-2}\mu_p^{(q-2)/p}$. This gives in particular that $\mathbb{E}G^2_{k,j}(X,Y) \le \sigma^2_{k,j}$ and we can infer from (4.2),

$$\mathbb{E}\|\sqrt{n_k}\alpha_{n_k}(g)\|_{\mathcal{G}_{k,j}} \le C'\sqrt{n_k\sigma^2_{k,j}}. \tag{4.3}$$

Applying the Fuk–Nagaev type inequality (see Fact 4.1) with $\eta = \delta = 1$, we find that for all $x > 0$,

$$\begin{aligned}\mathbb{P}\Big\{ & \max_{n_{k-1}<n\le n_k} \|\sqrt{n}\alpha_n(g)\|_{\mathcal{G}_{k,j}} \ge x + 2C'\sqrt{n_k\sigma_{k,j}^2}\Big\} \\ &\le \exp\Big(-\frac{x^2}{3n_k\sigma_{k,j}^2}\Big) + C_2 x^{-p} n_k \mathbb{E}G_{k,j}^p(X,Y) \\ &\le \exp\Big(-\frac{x^2}{3c_1 n_k h_{k,j}^d}\Big) + c_2 x^{-p} n_k h_{k,j}^d,\end{aligned}$$

where $c_1 = \kappa^2\|f_X\|_J \mu_p^{2/p}$ and $c_2 = C_2\kappa^p\mu_p\|f_X\|_J$. Taking $x = \rho\sqrt{n_k h_{k,j}^d \log\log n_k}$ for $\rho > 0$ and recalling the condition on a_n, we get for large enough k,

$$\begin{aligned}\mathbb{P}\Big\{ & \max_{n_{k-1}<n\le n_k} \|\sqrt{n}\alpha_n(g)\|_{\mathcal{G}_{k,j}} \ge \rho\sqrt{n_k h_{k,j}^d \log\log n_k}\Big\} \\ &\le \exp\Big(-\frac{\rho^2\log\log n_k}{4c_1}\Big) + \frac{c_3(\rho)(n_k h_{k,j}^d)^{1-p/2}}{(\log\log n_k)^{p/2}} \\ &\le (\log n_k)^{-\frac{\rho^2}{4c_1}} + \frac{c_4(\rho)2^{-j(\frac{p}{2}-1)}}{\log n_k(\log\log n_k)^{p/2}}.\end{aligned}$$

Finally, it is not too difficult to see that $L(k) \le 2\log n_k$ for any $k \ge 1$ (since $\epsilon < 1$). Therefore, recalling the empirical process representation in (4.1), this implies immediately that for k large enough,

$$\begin{aligned}\mathbb{P}\Big\{ & \max_{n_{k-1}<n\le n_k} \sup_{a_n\le h\le b_0} \sup_{\varphi\in\mathcal{F}} \frac{\sqrt{nh^d}|\hat{\varphi}_{n,h}(t) - \mathbb{E}\hat{\varphi}_{n,h}(t)|}{\sqrt{\log\log n}} > 2\sqrt{2}\rho\Big\} \\ &\le \sum_{j=1}^{L(k)} \mathbb{P}\Big\{\max_{n_{k-1}<n\le n_k} \|\sqrt{n}\alpha_n(g)\|_{\mathcal{G}_{k,j}} \ge \rho\sqrt{n_k h_{k,j}^d\log\log n_k}\Big\} \\ &\le 2(\log n_k)^{1-\frac{\rho^2}{4c_1}} + \frac{c_4(\rho)}{\log n_k(\log\log n_k)^{p/2}}\,\frac{1-2^{-(\frac{p}{2}-1)L(k)}}{1-2^{-(\frac{p}{2}-1)}} \\ &\le 2(\log n_k)^{1-\frac{\rho^2}{4c_1}} + \frac{c_4(\rho)(1+\xi)}{\log n_k(\log\log n_k)^{p/2}},\end{aligned}$$

for some $\xi > 0$. Finally, note that for any $\delta > 0$,

$$\sum_{k=1}^{\infty}(k(\log k)^{1+\delta})^{-1} < \infty$$

so that by taking ρ large enough so that $\rho > \sqrt{8c_1}$, the previous calculations yield (2.1) via the Borel–Cantelli lemma by summing in k. □

4.2. **Proof under condition (B).** To start with the proof of Proposition 2.1 in this case, set

$$\mu_p := \sup_{x\in J}\mathbb{E}\left[F^p(Y)|X=x\right], \quad p \ge 1,$$

so that obviously μ_p is finite for all $p \ge 1$. We consider the function classes $\mathcal{G}_{k,j}$ exactly as in the previous case and consider the envelope functions

$$G_{k,j}(x,y) = \kappa F(y)\mathbb{1}\Big\{x \in\ t + \big[-\frac{h_{k,j}}{2}, \frac{h_{k,j}}{2}\big]^d\Big\}.$$

Our proof is similar to the one under condition (A). The main difference is that we now use another exponential inequality which follows from a result of Yurinskii. (See Theorem 3.3.1 and (3.3.7) in [18].)

Fact 4.2 (Bernstein type inequality). *Let $Z, Z_1, \dots, Z_n$ be i.i.d. $\mathcal{X}$–valued random variables and consider a pointwise measurable class $\mathcal{G}$ of functions $g : \mathcal{X} \to \mathbb{R}$ with envelope function G. Assume that for some $H > 0$,*

$$\mathbb{E}G^m(Z) \leq \frac{m!}{2}\sigma^2 H^{m-2}, \quad m \geq 2,$$

where $\sigma^2 \geq \mathbb{E}G^2(Z)$. Then we have for any $t > 0$,

$$\begin{aligned}
&\mathbb{P}\left\{\max_{1\leq k\leq n} \|\sqrt{k}\alpha_k(g)\|_{\mathcal{G}} \geq \beta_n + t\right\} \\
&\qquad \leq \ \exp\left(-\frac{t^2}{2n\sigma^2 + 2tH}\right) \ \leq \ \exp\left(-\frac{t^2}{4n\sigma^2}\right) \vee \exp\left(-\frac{t}{4H}\right),
\end{aligned}$$

where $\beta_n = \mathbb{E}\|\sqrt{n}\alpha_n(g)\|_{\mathcal{G}}$.

Condition (B) implies that for some $s > 0$, and uniformly on $x \in J$,

$$\mathbb{E}[\exp(sF(Y))|X = x] \leq M + 1 < \infty,$$

where $M > 0$. Hence, a simple Taylor expansion in combination with the monotone convergence theorem yields that for all $x \in J$,

$$\sum_{m=1}^{\infty} \frac{s^m}{m!}\, \mathbb{E}[F^m(Y)|X = x] \leq M,$$

so that we can bound μ_p for any $p \geq 2$ as

$$\mu_p = \sup_{x\in J} \mathbb{E}\left[F^p(Y)|X = x\right] \leq \frac{p!M}{s^p}, \quad p \geq 2.$$

In particular, $\mu_2 \leq 2M/s^2$. Furthermore, we obtain in the same way as in the previous case that

$$\mathbb{E}G^2_{k,j}(X, Y) \leq 2Ms^{-2}h^d_{k,j}\kappa^2\|f_X\|_J =: \sigma^2_{k,j}.$$

With $A_{k,j} = t + [-h_{k,j}/2, h_{k,j}/2]^d$, it then easily follows that for any $m \geq 1$ and k large enough (so that $A_{k,j} \subseteq J$),

$$\begin{aligned}
\mathbb{E}G^m_{k,j}(X, Y) &= \int \mathbb{E}[G^m_{k,j}(X, Y)|X = x]f_X(x)dx \\
&= \kappa^m \int_{A_{k,j}} \mathbb{E}\big[F^m(Y)|X = x\big] f_X(x)dx \\
&\leq \kappa^m \mu_m h^d_{k,j}\|f_X\|_J \\
&\leq \frac{m!}{2}\sigma^2_{k,j}\kappa^{m-2}s^{-(m-2)}.
\end{aligned}$$

Using the same argument as in the previous case, we can find suitable constants $A_1, A_2 > 0$ such that

$$\mathbb{E}\big\|\sqrt{n}_k\alpha_{n_k}(g)\big\|_{\mathcal{G}_{k,j}} \leq A_1\sqrt{n_k\sigma^2_{k,j}} \leq A_2\sqrt{n_k h^d_{k,j}}.$$

Hence all the conditions of the above Bernstein type inequality (= Fact 4.2) are satisfied for k large enough with $H = \kappa/s$ and $\beta_n^2 = O(n_k h_{k,j}^d) = o(n_k h_{k,j}^d \log\log n_k)$. This gives us for all $1 \le j \le L(k)$ and $\rho > 0$ (note that $n_k h_{j,k}^d \ge \log\log n_k$),

$$\begin{aligned}
&\mathbb{P}\Big\{ \max_{n_{k-1}<n\le n_k} \|\sqrt{n}\alpha_n(g)\|_{\mathcal{G}_{k,j}} \ge \rho\sqrt{n_k h_{k,j}^d \log\log n_k}\Big\} \\
&\le \quad \exp\Big(-\frac{\rho^2 s^2 \log\log n_k}{8M\kappa^2\|f_X\|_J}\Big) + \exp\Big(-\frac{\rho s\sqrt{n_k h_{k,j}^d \log\log n_k}}{4\kappa}\Big) \\
&\le \quad (\log n_k)^{-A_3\rho^2} + \exp\Big(-\rho s\sqrt{2^j}\log\log n_k/4\kappa\Big) \\
&\le \quad (\log n_k)^{-A_3\rho^2} + (\log n_k)^{-A_4\rho},
\end{aligned}$$

with $A_3 = s^2/8M\kappa^2\|f_X\|_J$ and $A_4 = s/2\sqrt{2}\kappa$. Consequently, by the empirical process representation in (4.1), we have for any positive constant $\rho < \infty$ (recall also that $L(k) \le 2\log n_k, k \ge 1$) that

$$\begin{aligned}
&\mathbb{P}\Big\{ \max_{n_{k-1}<n\le n_k} \sup_{a_n\le h\le b_0} \sup_{\varphi\in\mathcal{F}} \frac{\sqrt{nh^d}|\hat{\varphi}_{n,h}(t) - \mathbb{E}\hat{\varphi}_{n,h}(t)|}{\sqrt{8\log\log n}} > \rho\Big\} \\
&\le \quad \sum_{j=1}^{L(k)} \mathbb{P}\Big\{ \max_{n_{k-1}<n\le n_k} \|\sqrt{n}\alpha_n(g)\|_{\mathcal{G}_{k,j}} \ge \rho\sqrt{n_k h_{k,j}^d \log\log n_k}\Big\} \\
&\le \quad 2(\log n_k)^{1-A_3\rho^2} + 2(\log n_k)^{1-A_4\rho}.
\end{aligned}$$

Finally, by taking ρ large enough such that $\rho > \sqrt{2/A_3} \vee 2/A_4$, the result follows from the Borel–Cantelli lemma by summing in k. □

References

[1] Csörgő, S., Deheuvels, P. and Mason, D.M. (1985). Kernel estimates of the tail index of a distribution, *Ann. Stat.* **13,** 1050–1077.

[2] Deheuvels, P. and Mason, D.M. (1994). Functional laws of the iterated logarithm for local empirical processes indexed by sets. *Ann. Probab.* **22,** 1619–1661.

[3] Dony, J. (2008). *Nonparametric regression estimation : An empirical process approach to the uniform in bandwidth consistency of kernel–type estimators and conditional U–statistics.* Ph.D. Thesis, Vrije Universiteit Brussel, Belgium.

[4] Dony, J., Einmahl, U. and Mason, D.M. (2006). Uniform in bandwidth consistency of local polynomial regression function estimators. *Austr. J. Stat.* **35,** 105–120.

[5] Dony, J. and Mason, D.M. (2009). Uniform in bandwidth consistency of kernel estimators of the tail index. *In progress.*

[6] Einmahl, U. and Li, D. (2008). Characterization of LIL behavior in Banach space. *Trans. Am. Math. Soc.* **360**, 6677-6693.

[7] Einmahl, U. and Mason, D.M. (1997). Gaussian approximation of local empirical processes indexed by functions. *Prob. Th. Rel.Fields* **107,** 283–311.

[8] Einmahl, U. and Mason, D.M. (1998). Strong approximations to the local empirical process. *Progr. Probab.* **43,** 75–92.

[9] Einmahl, U. and Mason, D.M. (2000). An empirical process approach to the uniform consistency of kernel-type function estimators. *J. Theor. Probab.* **13,** 1-37.

[10] Einmahl, U. and Mason, D.M. (2005). Uniform in bandwidth consistency of kernel-type function estimators. *Ann. Stat.* **33,** 1380–1403.

[11] Giné, E. and Guillou, A. (2002). Rates of strong consistency for multivariate kernel density estimators. *Ann. Inst. H. Poincaré.* **38,** 907-921.

[12] Haerdle, W. (1984). A law of the iterated logarithm for nonparametric regression function estimators. *Ann. Statist.* **12** 624-635.

[13] Haerdle, W., Janssen, P. and Serfling, R. (1988). Strong uniform consistency rates for estimators of conditional functionals. *Ann. Statist.* **16**, 1428-1449.

[14] Hall, P. (1981). Laws of the iterated logarithm for nonparametric density estimators. *Z. Wahrsch. verw. Gebiete.* **56,** 47–61.

[15] Jain, M.C. and Marcus, M.B. (1978). Continuity of sub–Gaussian processes, Dekker, New York. *Advances in Probability* **4,** 81–196.

[16] Ledoux, M. and Talagrand, M. (1991). *Probability on Banach Spaces.* Springer-Verlag, Berlin.

[17] van der Vaart, A.W. and Wellner, J.A. (1996). *Weak convergence and empirical processes. With applications to statistics.* Springer Series in Statistics. Springer–Verlag, New York.

[18] Yurinskii, V.V. (1995). *Sums and Gaussian vectors.* Lecture Notes in Mathematics **1617** Springer–Verlag, Berlin.

Department of Mathematics, Free University of Brussels (VUB), Pleinlaan 2, B-1050 Brussels, Belgium

E-mail address: jdony@vub.ac.be

E-mail address: ueinmahl@vub.ac.be