Uniform in bandwidth consistency of conditional U-statistics

J. Dony* and D.M. Mason[1]

*Mathematics Department, Free University of Brussels (VUB). Pleinlaan 2, B-1050 Brussels, Belgium. e-mail : jdony@vub.ac.be

[1]Food and Resource Economics, University of Delaware. 206 Townsend Hall, Newark, DE 19717. e-mail : davidm@udel.edu

Abstract

In 1991 Stute introduced a class of estimators called conditional U–statistics. They can be seen as a generalization of the Nadaraya-Watson estimator for the regression function, and he proved their strong pointwise consistency to

$$m(\mathbf{t}) := \mathbb{E}[g(Y_1, \ldots, Y_m)|(X_1, \ldots, X_m) = \mathbf{t}], \quad \mathbf{t} \in \mathbb{R}^m.$$

Very recently, Giné and Mason introduced the notion of a local U–process, which generalizes that of a local empirical process, and obtained central limit theorems and laws of the iterated logarithm for this class. We apply the methods developed in Einmahl and Mason (2005) and Giné and Mason (2007a,b) to establish uniform in bandwidth consistency to $m(\mathbf{t})$ of the estimator proposed by Stute.

Keywords. conditional U–statistics, empirical process, kernel estimation, Nadaraya–Watson, regression function, uniform in bandwidth consistency.

1 Introduction and statement of main results

Let (X, Y), $(X_1, Y_1), \ldots, (X_n, Y_n)$ be independent random vectors with common joint density function $f : \mathbb{R} \times \mathbb{R} \to [0, \infty[$, and for a measurable function $\varphi : \mathbb{R}^m \to \mathbb{R}$, consider the regression function

$$m_\varphi(\mathbf{t}) = \mathbb{E}\left[\varphi(Y_1, \ldots, Y_m)|(X_1, \ldots, X_m) = \mathbf{t}\right], \quad \mathbf{t} \in \mathbb{R}^m.$$

Stute [11] introduced a class of estimators for $m_\varphi(\mathbf{t})$, called conditional U–statistics and defined pointwise for $\mathbf{t} \in \mathbb{R}^m$ as

$$\widehat{m}_n(\mathbf{t}; h_n) = \frac{\sum_{(i_1,\ldots,i_m)\in I_n^m} \varphi(Y_{i_1}, \ldots, Y_{i_m}) K\left(\frac{t_1 - X_{i_1}}{h_n}\right) \cdots K\left(\frac{t_m - X_{i_m}}{h_n}\right)}{\sum_{(i_1,\ldots,i_m)\in I_n^m} K\left(\frac{t_1 - X_{i_1}}{h_n}\right) \cdots K\left(\frac{t_m - X_{i_m}}{h_n}\right)}, \tag{1.1}$$

where

$$I_n^m = \{(i_1, \ldots, i_m) : 1 \le i_j \le n, i_j \ne i_l \text{ if } j \ne l\}, \tag{1.2}$$

and $0 < h_n < 1$ goes to zero at a certain rate. Soon afterwards, Sen [10] obtained results on uniform consistency of this estimator. We shall adapt and extend the methods developed in Einmahl and Mason [5] and Giné and Mason [6, 7] to show that under appropriate regularity conditions a much stronger form of consistency holds, namely uniform in bandwidth consistency of $\widehat{m}_n$. This means that with probability 1,

$$\limsup_{n\to\infty} \sup_{\widetilde{a}_n \le h \le b_n} \sup_{\mathbf{t}\in[c,d]^m} |\widehat{m}_n(\mathbf{t}; h) - m_\varphi(\mathbf{t})| = 0, \tag{1.3}$$

for $-\infty < c < d < \infty$ and $\widetilde{a}_n < b_n$, as long as $\widetilde{a}_n \to 0$, $b_n \to 0$ and $b_n/\widetilde{a}_n \to \infty$ at rates depending upon the moments of $\varphi(Y_1, \ldots, Y_m)$. Moreover, we shall show that (1.3) holds uniformly in $\varphi \in \mathcal{F}$ as well. In fact, our results extend those of Einmahl and Mason [5], who treat the case $m = 1$.

We shall infer (1.3) via general uniform in bandwidth results for a specific $U-$statistic process indexed by a class of functions. We define this process in (1.4) below. Towards this end, for $m \le n$, consider a class $\mathcal{F}$ of measurable functions $g : \mathbb{R}^m \to \mathbb{R}$ such that $\mathbb{E}g^2(Y_1, \ldots, Y_m) < \infty$, which satisfies the following conditions $(F.i)-(F.iii)$. First, to avoid measurability problems, we assume that

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

i.e. there exists a countable subclass $\mathcal{F}_0$ of $\mathcal{F}$ such that we can find for any function $g \in \mathcal{F}$ a sequence of functions $g_m \in \mathcal{F}_0$ for which $g_m(z) \to g(z), z \in \mathbb{R}^m$. This condition is discussed in van der Vaart and Wellner [12]. We also assume that $\mathcal{F}$ has a measurable envelope function

$(F.ii)$ $$F(\mathbf{y}) \ge \sup_{g\in\mathcal{F}} |g(\mathbf{y})|, \quad \mathbf{y} \in \mathbb{R}^m.$$

Finally we assume that $\mathcal{F}$ is of VC–type with characteristics A and v ("VC" for Vapnik and Červonenkis), meaning that for some $A \ge 3$ and $v \ge 1$,

$(F.iii)$ $$\mathcal{N}(\mathcal{F}, L_2(Q), \varepsilon) \le \left(\frac{A\|F\|_{L_2(Q)}}{\varepsilon}\right)^v, \quad 0 < \varepsilon \le 2\|F\|_{L_2(Q)},$$

where for $\varepsilon > 0$, $\mathcal{N}(\mathcal{F}, L_2(Q), \varepsilon)$ is defined as the smallest number of $L_2(Q)$ open balls of radius ε required to cover $\mathcal{F}$, and Q is any probability measure on $(\mathbb{R}^m, \mathcal{B})$ such that $\|F\|_{L_2(Q)} < \infty$. (If $(F.iii)$ holds for $\mathcal{F}$, then we say that the VC–type class $\mathcal{F}$ admits the characteristics A and v.)

Let now $K : \mathbb{R} \to \mathbb{R}$ be a kernel function with support contained in $[-1/2, 1/2]$ satisfying

$(K.i)$ $$\sup_{x\in\mathbb{R}} |K(x)| =: \kappa < \infty \quad \text{and} \quad \int K(x)dx = 1.$$

For such kernels, we consider the class of functions $\mathcal{K} := \{hK_h(t - \cdot) : h > 0, t \in \mathbb{R}\}$ and assume that

$$(K.ii) \qquad \mathcal{K} \text{ is pointwise measurable and of VC–type },$$

where as usual $K_h(z) = h^{-1}K(z/h)$, $z \in \mathbb{R}$. Furthermore, let

$$(K.iii) \qquad \widetilde{K}(\mathbf{t}) := \prod_{j=1}^{m} K(t_j)$$

denote the product kernel. Next, if $(S, \mathcal{S})$ is a measurable space, define the general U–statistic with kernel $H : S^k \to \mathbb{R}$ based on S–valued random variables $Z_1, \ldots, Z_n$ as

$$U_n^{(k)}(H) := \frac{(n-k)!}{n!} \sum_{\mathbf{i} \in I_n^k} H(Z_{i_1}, \ldots, Z_{i_k}), \quad 1 \leq k \leq n,$$

where I_n^k is defined as in (1.2) with $m = k$. (Note that we do not require H to be symmetric here.) For a bandwidth $0 < h < 1$ and $g \in \mathcal{F}$, consider the U–kernel

$$G_{g,h,\mathbf{t}}(\mathbf{x}, \mathbf{y}) := g(\mathbf{y})\widetilde{K}_h(\mathbf{t} - \mathbf{x}), \quad \mathbf{x}, \mathbf{y}, \mathbf{t} \in \mathbb{R}^m,$$

and for the sample $(X_1, Y_1), \ldots, (X_n, Y_n)$, define

$$U_n(g, h, \mathbf{t}) := U_n^{(m)}(G_{g,h,\mathbf{t}}) = \frac{(n-m)!}{n!} \sum_{\mathbf{i} \in I_n^m} G_{g,h,\mathbf{t}}(\mathbf{X_i}, \mathbf{Y_i}),$$

where throughout this paper we shall use the notation

$$\begin{aligned} \mathbf{X} &= (X_1, \ldots, X_m) \in \mathbb{R}^m \quad \text{and} \quad \mathbf{X_i} := (X_{i_1}, \ldots, X_{i_k}) \in \mathbb{R}^k, \quad \mathbf{i} \in I_n^k, \\ \mathbf{Y} &= (Y_1, \ldots, Y_m) \in \mathbb{R}^m \quad \text{and} \quad \mathbf{Y_i} := (Y_{i_1}, \ldots, Y_{i_k}) \in \mathbb{R}^k, \quad \mathbf{i} \in I_n^k. \end{aligned}$$

Now introduce the *U–statistic process*

$$u_n(g, h, \mathbf{t}) := \sqrt{n}\{U_n(g, h, \mathbf{t}) - \mathbb{E}U_n(g, h, \mathbf{t})\}. \tag{1.4}$$

We shall establish a strong uniform in bandwidth consistency result for the U–statistic process in (1.4). Theorem 1 gives such a result for bounded classes of functions $\mathcal{F}$, while Theorem 2 is applicable for unbounded classes $\mathcal{F}$ which satisfy a conditional moment condition stated in (1.6) below. In the bounded case, we assume that the envelope function of $\mathcal{F}$ is bounded by some finite constant M, i.e., (1.5) holds.

Theorem 1 *Suppose that the marginal density f_X of X is bounded, and let $a_n = c(\log n/n)^{1/m}$ for $c > 0$. If the class of functions $\mathcal{F}$ is bounded in the sense that for some $0 < M < \infty$,*

$$F(\mathbf{y}) \le M, \quad \mathbf{y} \in \mathbb{R}^m, \tag{1.5}$$

we can infer under the above mentioned assumptions on $\mathcal{F}$ and $\mathcal{K}$ that for all $c > 0$ and $0 < b_0 < 1$ there exists a constant $0 < C < \infty$ such that

$$\limsup_{n\to\infty} \sup_{a_n \le h \le b_0} \sup_{g\in\mathcal{F}} \sup_{\mathbf{t}\in\mathbb{R}^m} \frac{\sqrt{nh^m}|U_n(g,h,\mathbf{t}) - \mathbb{E}U_n(g,h,\mathbf{t})|}{\sqrt{|\log h| \vee \log\log n}} \le C, \quad a.s.$$

Theorem 2 *Suppose that the marginal density f_X of X is bounded, and for $c > 0$ let $a'_n = c((\log n/n)^{1-2/p})^{1/m}$. If $\mathcal{F}$ is unbounded but satisfies for some $p > 2$*

$$\mu_p := \sup_{\mathbf{x}\in\mathbb{R}^m} \mathbb{E}[F^p(\mathbf{Y})|\, \mathbf{X} = \mathbf{x}] < \infty, \tag{1.6}$$

we can infer under the above mentioned assumptions on $\mathcal{F}$ and $\mathcal{K}$ that for all $c > 0$ and $0 < b_0 < 1$ there exists a constant $0 < C' < \infty$ such that,

$$\limsup_{n\to\infty} \sup_{a'_n \le h \le b_0} \sup_{g\in\mathcal{F}} \sup_{\mathbf{t}\in\mathbb{R}^m} \frac{\sqrt{nh^m}|U_n(g,h,\mathbf{t}) - \mathbb{E}U_n(g,h,\mathbf{t})|}{\sqrt{|\log h| \vee \log\log n}} \le C', \quad a.s.$$

From now on, we shall write $\hat{m}_{n,\varphi}(\mathbf{t}, h)$ for the estimator of the regression function defined in (1.1) to stress the role of $\varphi(\mathbf{y})$. It is clear that $\hat{m}_{n,\varphi}(\mathbf{t}, h)$ can be rewritten for all $\varphi \in \mathcal{F}$ as

$$\hat{m}_{n,\varphi}(\mathbf{t}, h) = \frac{\sum_{\mathbf{i}\in I_n^m} \varphi(\mathbf{Y_i})\widetilde{K}_h(\mathbf{t} - \mathbf{X_i})}{\sum_{\mathbf{i}\in I_n^m} \widetilde{K}_h(\mathbf{t} - \mathbf{X_i})} = \frac{U_n(\varphi, h, \mathbf{t})}{U_n(1, h, \mathbf{t})},$$

where we denote by $U_n(1, h, \mathbf{t})$ the U–statistic $U_n(g, h, \mathbf{t})$ with $g \equiv 1$. To prove the uniform consistency of $\hat{m}_{n,\varphi}(\mathbf{t}, h)$ to $m_\varphi(\mathbf{t})$, we shall consider another, but more appropriate, centering factor than the expectation $\mathbb{E}\hat{m}_{n,\varphi}(\mathbf{t}, h)$, which may not exist or be difficult to compute. Define the centering

$$\widehat{\mathbb{E}}\hat{m}_{n,\varphi}(\mathbf{t}, h) := \frac{\mathbb{E}U_n(\varphi, h, \mathbf{t})}{\mathbb{E}U_n(1, h, \mathbf{t})}. \tag{1.7}$$

This centering permits us to apply Theorems 1 and 2 (depending on whether the class $\mathcal{F}$ is bounded in the sense of (1.5) or unbounded in the sense of (1.6)) to derive results on the convergence rates of the process $\hat{m}_{n,\varphi}(\mathbf{t}, h) - \widehat{\mathbb{E}}\hat{m}_{n,\varphi}(\mathbf{t}, h)$ to zero and the consistency of $\hat{m}_{n,\varphi}(\mathbf{t}, h)$, uniformly in bandwidth.

For any compact interval $I = [c, d]$ with $-\infty < c < d < \infty$ and $\eta > 0$, define $I^\eta = [c - \eta, d + \eta]$ and denote as usual the marginal density function of X by f_X. Then introduce the class of functions defined on the compact subset $J^m = I^\eta \times \ldots \times I^\eta$ of $\mathbb{R}^m$,

$$\mathcal{M} = \{m_\varphi(\cdot)\widetilde{f}(\cdot) : \varphi \in \mathcal{F}\}, \tag{1.8}$$

where the function $\widetilde{f} : \mathbb{R}^m \to \mathbb{R}$ is defined as

$$\widetilde{f}(\mathbf{t}) := \int f(t_1, y_1) \ldots f(t_m, y_m) dy_1 \ldots dy_m = f_X(t_1) \ldots f_X(t_m). \tag{1.9}$$

We have now introduced all the notation that we need to state our results on the uniform consistency of the conditional $U-$statistic estimator proposed by Stute for the general regression function, where this consistency is uniform in $\varphi \in \mathcal{F}$ and in bandwidth as well.

Theorem 3 *Besides being bounded, suppose that the marginal density function f_X of X is continuous and strictly positive on the interval $J = I^\eta$, where $I = [c, d]$ is a compact interval and $\eta > 0$. Assume that the class of functions $\mathcal{M}$ is uniformly equicontinuous. Then it follows that for all sequences $0 < b_n < 1$ with $b_n \to 0$,*

$$\sup_{0<h<b_n} \sup_{\varphi\in\mathcal{F}} \sup_{\mathbf{t}\in I^m} |\widehat{\mathbb{E}}\hat{m}_{n,\varphi}(\mathbf{t}, h) - m_\varphi(\mathbf{t})| = o(1),$$

where $I^m = I \times \ldots \times I$.

Theorem 4 *Besides being bounded, suppose that the marginal density function f_X of X is continuous and strictly positive on the interval $J = I^\eta$, where $I = [c, d]$ is a compact interval and $\eta > 0$. Then it follows under the above mentioned assumptions on $\mathcal{F}$ and $\mathcal{K}$ that for all $c > 0$ and all sequences $0 < b_n < 1$ with $a''_n \le b_n \to 0$, there exists a constant $0 < C'' < \infty$ such that,*

$$\limsup_{n\to\infty} \sup_{a''_n \le h < b_n} \sup_{\varphi\in\mathcal{F}} \sup_{\mathbf{t}\in I^m} \frac{\sqrt{nh^m}|\hat{m}_{n,\varphi}(\mathbf{t}, h) - \widehat{\mathbb{E}}\hat{m}_{n,\varphi}(\mathbf{t}, h)|}{\sqrt{|\log h| \vee \log\log n}} \le C'', \quad a.s.,$$

where $I^m = I \times \ldots \times I$ and a''_n is either a_n or a'_n depending on whether the class $\mathcal{F}$ is bounded or not, i.e. whether (1.5) or (1.6) holds.

The following proposition follows straightforwardly from Theorems 3 and 4.

Proposition 1 *Under the assumptions of Theorems 3 and 4 on f_X and the classes $\mathcal{F}$ and $\mathcal{K}$, it follows that for all sequences $0 < \widetilde{a}_n \le b_n < 1$ satisfying $b_n \to 0$ and $n\widetilde{a}_n / \log n \to \infty$,*

$$\sup_{\widetilde{a}_n \le h < b_n} \sup_{\varphi\in\mathcal{F}} \sup_{\mathbf{t}\in I^m} |\hat{m}_{n,\varphi}(\mathbf{t}, h) - m_\varphi(\mathbf{t})| \longrightarrow 0, \quad a.s., \tag{1.10}$$

where $I^m = I \times \ldots \times I$.

It is readily seen that one can take $\widetilde{a}_n = a'_n$ in the previous proposition and obtain strong uniform consistency of Stute's estimator (1.1) for general bandwidths. However, note that by choosing $\widetilde{a}_n = a_n$, one would only obtain almost sure convergence to a positive constant $\widetilde{c} > 0$ in (1.10).

2 Preliminaries for the proofs of the theorems

Let Ψ be a real valued functional defined on a class of functions $\mathcal{G}$ and g a real valued function defined on $\mathbb{R}^d, d \geq 1$. Occasionally we shall use the notation

$$\|\Psi(G)\|_{\mathcal{G}} = \sup_{G \in \mathcal{G}} |\Psi(G)| \quad \text{and} \quad \|g\|_\infty = \sup_{\mathbf{x} \in \mathbb{R}^d} |g(\mathbf{x})|. \tag{2.1}$$

In the sequel we shall need to symmetrize the functions $G_{g,h,\mathbf{t}}(\cdot,\cdot)$. To do this, we set

$$\bar{G}_{g,h,\mathbf{t}}(\mathbf{x},\mathbf{y}) := (m!)^{-1} \sum_{\sigma \in I_m^m} G_{g,h,\mathbf{t}}(\mathbf{x}_\sigma, \mathbf{y}_\sigma) = (m!)^{-1} \sum_{\sigma \in I_m^m} g(\mathbf{y}_\sigma) \widetilde{K}_h(\mathbf{t} - \mathbf{x}_\sigma),$$

where $\mathbf{z}_\sigma := (z_{\sigma_1}, \ldots, z_{\sigma_m})$. Obviously, the expectation of $G_{g,h,\mathbf{t}}$ remains unchanged after symmetrization, and $U_n^{(m)}(\bar{G}_{g,h,\mathbf{t}}(\cdot,\cdot)) = U_n(g,h,\mathbf{t})$, and thus the U–statistic process in (1.4) may be redefined using the symmetrized kernels, i.e. we consider

$$u_n(g,h,\mathbf{t}) = \sqrt{n}\{U_n^{(m)}(\bar{G}_{g,h,\mathbf{t}}) - \mathbb{E}U_n^{(m)}(\bar{G}_{g,h,\mathbf{t}})\}. \tag{2.2}$$

Moreover, the Hoeffding decomposition tells us that

$$u_n(g,h,\mathbf{t}) = \sqrt{n} \sum_{k=1}^m \binom{m}{k} U_n^{(k)}(\pi_k \bar{G}_{g,h,\mathbf{t}}(\cdot,\cdot)), \tag{2.3}$$

where the k–th Hoeffding projection for the (symmetric) function $L : S^m \times S^m \to \mathbb{R}$ is defined for $\mathbf{x}_k = (x_1, \ldots, x_k) \in S^k$ and $\mathbf{y}_k = (y_1, \ldots, y_k) \in S^k$ as

$$\pi_k L(\mathbf{x}_k, \mathbf{y}_k) := (\delta_{(x_1,y_1)} - P) \times \ldots \times (\delta_{(x_k,y_k)} - P) \times P^{m-k}(L),$$

where P is any probability measure on $(S, \mathcal{S})$. Considering $(X_i, Y_i), i \geq 1$ i.i.d–P and assuming L is in $L_2(P^m)$, this is an orthogonal decomposition, and $\mathbb{E}[\pi_k L(\mathbf{X}_k, \mathbf{Y}_k)|(X_2,Y_2), \ldots, (X_k,Y_k)] = 0$, $k \geq 1$, where we denote $\mathbf{X}_k$ and $\mathbf{Y}_k$ for $(X_1, \ldots, X_k)$ and $(Y_1, \ldots, Y_k)$ respectively. Thus the kernels $\pi_k L$ are canonical for P (or completely degenerate, or completely centered). Also, π_k, $k \geq 1$, are nested projections, i.e., $\pi_k \circ \pi_l = \pi_k$ if $k \leq l$, and

$$\mathbb{E}[(\pi_k L)^2(\mathbf{X}_k, \mathbf{Y}_k)] \leq \mathbb{E}[(L - \mathbb{E}L)^2(\mathbf{X}, \mathbf{Y})] \leq \mathbb{E}L^2(\mathbf{X}, \mathbf{Y}). \tag{2.4}$$

For more details consult de la Peña and Giné [2].

Since we assume $\mathcal{F}$ to be of VC–type with envelope function F, and $\mathcal{K}$ to be of VC–type with envelope κ, it is readily checked (via Lemma A.1 in Einmahl and Mason [4]) that the class of functions on $\mathbb{R}^m \times \mathbb{R}^m$ given by $\{h^m G_{g,h,\mathbf{t}}(\cdot,\cdot) : g \in \mathcal{F}, 0 < h < 1, \mathbf{t} \in \mathbb{R}^m\}$ is of VC–type, as well as the class

$$\mathcal{G} = \{h^m \bar{G}_{g,h,\mathbf{t}}(\cdot,\cdot) : g \in \mathcal{F}, 0 < h < 1, \mathbf{t} \in \mathbb{R}^m\}, \tag{2.5}$$

for which we denote the VC–type characteristics by A_1 and v_1, and the envelope function by

$$\widetilde{F}(\mathbf{y}) \equiv \widetilde{F}(\mathbf{x},\mathbf{y}) = \kappa^m \sum_{\sigma \in I_m^m} F(\mathbf{y}_\sigma), \quad \mathbf{y} \in \mathbb{R}^m. \tag{2.6}$$

(Recall $(F.ii)$ and $(F.iii)$ for terminology.) Next, for $k = 1, \ldots, m$ introduce the classes of functions on $\mathbb{R}^k \times \mathbb{R}^k$,

$$\mathcal{G}^{(k)} = \{h^m \pi_k \bar{G}_{g,h,\mathbf{t}}(\cdot,\cdot) : g \in \mathcal{F}, 0 < h < 1, \mathbf{t} \in \mathbb{R}^m\}. \tag{2.7}$$

Then an argument in Giné and Mason [7] shows that each class $\mathcal{G}^{(k)}$ is of VC–type with characteristics A_1 and v_1 and envelope function $F_k \leq 2^k \|\widetilde{F}\|_\infty$. (See the completion of the proof of Theorem 1 in that paper for more details.)

3 Proof of Theorem 1 : the bounded case

We begin with studying the first term of (2.3), namely

$$m\sqrt{n}U_n^{(1)}(\pi_1 \bar{G}_{g,h,\mathbf{t}}(\cdot,\cdot)) = \frac{m}{\sqrt{n}} \sum_{i=1}^n \pi_1 \bar{G}_{g,h,\mathbf{t}}(X_i, Y_i).$$

Linear term of (2.3)

From the definition of the Hoeffding projections and recalling that the sample $(X_1, Y_1), \ldots, (X_n, Y_n)$ is i.i.d., we can say for all $(x, y) \in \mathbb{R}^2$ that

$$\begin{aligned} \pi_1 \bar{G}_{g,h,\mathbf{t}}(x,y) &= \mathbb{E}[\bar{G}_{g,h,\mathbf{t}}((x, X_2, \ldots, X_m), (y, Y_2, \ldots, Y_m))] - \mathbb{E}\bar{G}_{g,h,\mathbf{t}}(\mathbf{X},\mathbf{Y}) \\ &= \mathbb{E}[\bar{G}_{g,h,\mathbf{t}}(\mathbf{X},\mathbf{Y})|(X_1,Y_1) = (x,y)] - \mathbb{E}\bar{G}_{g,h,\mathbf{t}}(\mathbf{X},\mathbf{Y}). \end{aligned}$$

Introduce therefore the function on $\mathbb{R} \times \mathbb{R}$ (for clarity we do not indicate the dependence on m)

$$\begin{aligned} S_{g,h,\mathbf{t}} : \ \mathbb{R} \times \mathbb{R} &\longrightarrow \mathbb{R} \\ (x,y) &\longmapsto mh^m \mathbb{E}[\bar{G}_{g,h,\mathbf{t}}(\mathbf{X},\mathbf{Y})|(X_1,Y_1) = (x,y)]. \end{aligned}$$

Then obviously these functions are symmetric. Using this notation we write

$$mh^m \pi_1 \bar{G}_{g,h,\mathbf{t}}(x,y) = S_{g,h,\mathbf{t}}(x,y) - \mathbb{E}S_{g,h,\mathbf{t}}(X_1,Y_1),$$

and hence for all $g \in \mathcal{F}$, $h \in [a_n, b_0]$ and $\mathbf{t} \in \mathbb{R}^m$, the linear term of the decomposition in (2.3) times h^m is given by

$$\begin{aligned} mh^m \sqrt{n} U_n^{(1)}(\pi_1 \bar{G}_{g,h,\mathbf{t}}) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \{S_{g,h,\mathbf{t}}(X_i,Y_i) - \mathbb{E}S_{g,h,\mathbf{t}}(X_i,Y_i)\} \\ &=: \alpha_n(S_{g,h,\mathbf{t}}), \end{aligned}$$

where this last expression is an empirical process α_n based on the sample $(X_1, Y_1), \ldots, (X_n, Y_n)$ and indexed by the class of functions on $\mathbb{R} \times \mathbb{R}$,

$$\mathcal{S}_n = \{S_{g,h,\mathbf{t}}(\cdot, \cdot) : g \in \mathcal{F}, a_n \le h \le b_0, \mathbf{t} \in \mathbb{R}^m\}.$$

Clearly $\mathcal{S}_n \subset m\mathcal{G}^{(1)}$, and the class $m\mathcal{G}^{(1)}$ has envelope function mF_1, where F_1 is the envelope function of the class $\mathcal{G}^{(1)}$ defined in (2.7). From the above discussion, this class is of VC–type with the same characteristics as $\mathcal{G}$, and therefore, after appropriate identifications of notation, we can apply Theorem 2 of Dony, Einmahl and Mason [3] to conclude that

$$\limsup_{n\to\infty} \sup_{a_n \le h \le b_0} \sup_{g\in\mathcal{F}} \sup_{\mathbf{t}\in\mathbb{R}^m} \frac{m\sqrt{nh^m}|U_n^{(1)}(\pi_1 \bar{G}_{g,h,\mathbf{t}})|}{\sqrt{|\log h| \vee \log\log n}} \le C, \quad \text{a.s.} \tag{3.1}$$

Alternatively, a straightforward modification of the proof of (4.9) below with a_n' replaced by a_n and $\gamma_\ell^{1/p}$ by M, gives (3.1) as well.

The other terms of (2.3)

Our aim now is to show that all the other terms of the Hoeffding decomposition are almost surely bounded or more precisely that for each $k = 2, \ldots, m$,

$$\sup_{a_n \le h \le b_0} \sup_{g\in\mathcal{F}} \sup_{\mathbf{t}\in\mathbb{R}^m} \frac{\binom{m}{k}\sqrt{nh^m}|U_n^{(k)}(\pi_k \bar{G}_{g,h,\mathbf{t}})|}{\sqrt{|\log h| \vee \log\log n}} = O(1), \quad \text{a.s.} \tag{3.2}$$

Since $na_n^m = c^m \log n$, this will be accomplished if we can prove that for each $k = 2, \ldots, m$,

$$\sup_{a_n \le h \le b_0} \sup_{g\in\mathcal{F}} \sup_{\mathbf{t}\in\mathbb{R}^m} \frac{\sqrt{nh^m}|U_n^{(k)}(\pi_k \bar{G}_{g,h,\mathbf{t}})|}{\sqrt{(|\log h| \vee \log\log n)^k}} = O\left(\frac{1}{\sqrt{a_n^m n^{k-1}}}\right), \quad \text{a.s.} \tag{3.3}$$

To obtain uniform in bandwidth convergence rates, we shall need a blocking argument and a decomposition of the interval $[a_n, b_0]$ into smaller intervals. To do this, set $n_\ell = 2^\ell, \ell \ge 0$ and consider the intervals $\mathcal{H}_{\ell,j} := [h_{\ell,j-1}, h_{\ell,j}]$, where the boundaries are given by $h_{\ell,j}^m := 2^j a_{n_\ell}^m$. Setting $L(\ell) = \max\{j : h_{\ell,j} \le 2b_0\}$, observe that

$$[a_{n_\ell}, b_0] \subseteq \bigcup_{\ell=1}^{L(\ell)} \mathcal{H}_{\ell,j} \quad \text{and} \quad L(\ell) \sim \log\left(\frac{n_\ell b_0}{c \log n_\ell}\right) / \log 2, \tag{3.4}$$

implying in particular that $L(\ell) \le 2\log n_\ell$. (This fact will be used repeatedly to finish some important steps of the proofs.) Next, for $1 \le j \le L(\ell)$, consider the class of functions on $\mathbb{R}^m \times \mathbb{R}^m$,

$$\mathcal{G}_{\ell,j} := \{h^m \bar{G}_{g,h,\mathbf{t}}(\cdot, \cdot) : g \in \mathcal{F}, h \in \mathcal{H}_{\ell,j}, \mathbf{t} \in \mathbb{R}^m\},$$

as well as the class on $\mathbb{R}^k \times \mathbb{R}^k$,

$$\mathcal{G}_{\ell,j}^{(k)} := \left\{ \frac{h^m \pi_k \bar{G}_{g,h,\mathbf{t}}(\cdot,\cdot)}{M_k} : g \in \mathcal{F}, h \in \mathcal{H}_{\ell,j}, \mathbf{t} \in \mathbb{R}^m \right\},$$

where $M_k = 2^k \kappa^m M$. Clearly, each class $\mathcal{G}_{\ell,j}$ is of VC–type with the same characteristics and envelope function as $\mathcal{G}$, and $\mathcal{G}_{\ell,j}^{(k)}$ is of VC–type with the same characteristics as $\mathcal{G}^{(k)}$ (and thus as $\mathcal{G}$) with envelope function $M_k^{-1} F_k$, where F_k is the envelope function of $\mathcal{G}^{(k)}$. Notice that from (1.5),

$$M_k \geq \sup_{\mathbf{x},\mathbf{y} \in \mathbb{R}^k} \{ |\pi_k \bar{G}_{g,h,\mathbf{t}}(\mathbf{x},\mathbf{y})| : g \in \mathcal{F}, 0 < h < 1, \mathbf{t} \in \mathbb{R}^m \},$$

and hence each function in $\mathcal{G}_{\ell,j}^{(k)}$ is bounded by 1. Define now for $n_{\ell-1} < n \leq n_\ell$, $\ell = 1, 2, \ldots$

$$\mathcal{U}_n(j,k,\ell) = n_\ell^{-k/2} \sup_{H \in \mathcal{G}_{\ell,j}^{(k)}} \Big| \sum_{\mathbf{i} \in I_n^k} H(\mathbf{X_i}, \mathbf{Y_i}) \Big|. \tag{3.5}$$

From Theorem 4 of Giné and Mason [7] (see Theorem A.1 in the Appendix), we get for $c = 1/2$, $r = 2$ and all $x > 0$ that for any $\ell \geq 1$,

$$\mathbb{P}\Big\{ \max_{n_{\ell-1} < n \leq n_\ell} \mathcal{U}_n(j,k,\ell) > x \Big\} \leq \frac{2}{x} \mathbb{P}\left\{ \mathcal{U}_{n_\ell}(j,k,\ell) > x/2 \right\}^{1/2} \mathbb{E}[\mathcal{U}_{n_\ell}^2(j,k,\ell)]^{1/2}. \tag{3.6}$$

We shall apply an exponential inequality and a moment bound for U–statistics due to respectively de la Peña and Giné [2], and Giné and Mason [7], on the class $\mathcal{G}_{\ell,j}^{(k)}$ to bound (3.6). In order to use these results we must first derive some bounds. Firstly, it is readily checked that

$$\mathcal{U}_n(j,k,\ell) \leq n_\ell^{k/2} \| U_n^{(k)}(\pi_k G) \|_{\mathcal{G}_{\ell,j}^{(k)}}, \tag{3.7}$$

for all $n_{\ell-1} < n \leq n_\ell$. (Recall the notation (2.1).) Secondly, notice that in $(K.i)$, K is assumed to be bounded by κ and has support in $[-1/2, 1/2]$, such that by assumption (1.5) and $M_k = 2^k \kappa^m M$, for $H \in \mathcal{G}_{\ell,j}^{(k)}$ we have by (2.4)

$$\begin{aligned} \mathbb{E} H^2(\mathbf{X}, \mathbf{Y}) &\leq M_k^{-2} h^{2m} \mathbb{E} \bar{G}_{g,h,\mathbf{t}}^2(\mathbf{X}, \mathbf{Y}) \\ &= M_k^{-2} \mathbb{E}\Big[g^2(\mathbf{Y}) \widetilde{K}^2 \Big(\frac{\mathbf{t} - \mathbf{X}}{h} \Big) \Big] \\ &\leq h^m 4^{-k} \|f_X\|_\infty^m. \end{aligned}$$

For $D_m = 4^{-k} \|f_X\|_\infty^m$, this gives us that

$$\sup_{H \in \mathcal{G}_{\ell,j}^{(k)}} \mathbb{E} H^2(\mathbf{X}, \mathbf{Y}) \leq D_m h_{\ell,j}^m =: \sigma_{\ell,j}^2. \tag{3.8}$$

Since $\pi_k \pi_k L = \pi_k L$ for all $k \geq 1$, we can now apply Theorem A.4 to the class $\mathcal{G}_{\ell,j}^{(k)}$ with $\sigma_{\ell,j}^2$ as in (3.8), and obtain easily that for some constant A_k,

$$\mathbb{E}\mathcal{U}_{n_\ell}^2(j,k,\ell) \leq n_\ell^k \mathbb{E}\|U_{n_\ell}^{(k)}(\pi_k H)\|_{\mathcal{G}_{\ell,j}^{(k)}}^2 \leq 2^k A_k h_{\ell,j}^m |\log h_{\ell,j}|^k. \tag{3.9}$$

To control the probability term in (3.6), we shall apply an exponential inequality to the same class $\mathcal{G}_{\ell,j}^{(k)}$ (recall that each $H \in \mathcal{G}_{\ell,j}^{(k)}$ is bounded by 1). Setting

$$y^* = C_{1,k}(|\log h_{\ell,j}| \vee \log\log n_\ell)^{k/2} =: C_{1,k}\lambda_{j,k}(\ell), \tag{3.10}$$

where $C_{1,k} < \infty$, Theorem A.6 gives us constants $C_{2,k}, C_{3,k}$ such that for $j = 1, \ldots, L(\ell)$ and any $\rho > 1$,

$$\begin{aligned}\mathbb{P}\left\{\mathcal{U}_{n_\ell}(j,k,\ell) > \rho^{k/2} y^*\right\} &\leq C_{2,k}\exp\left\{-C_{3,k}\rho y^{*2/k}\right\} \\ &\leq \exp\left\{-C_{4,k}\rho \log\log n_\ell\right\}. \end{aligned} \tag{3.11}$$

Then plugging the bounds (3.9) and (3.11) into (3.6), we get for some $C_{5,k} > 0$, any $\rho \geq 2$ and ℓ large enough,

$$\begin{aligned}\mathbb{P}\left\{\max_{n_{\ell-1}<n\leq n_\ell} \mathcal{U}_n(j,k,\ell) > 2\rho^{k/2}y^*\right\} &\leq \frac{(\log n_\ell)^{-\rho\frac{C_{4,k}}{2}}\sqrt{2^k A_k h_{\ell,j}^m |\log h_{\ell,j}|^k}}{C_{1,k}\sqrt{\rho^k(|\log h_{\ell,j}| \vee \log\log n_\ell)^k}} \\ &\leq \sqrt{h_{\ell,j}^m}(\log n_\ell)^{-\rho C_{5,k}}. \end{aligned} \tag{3.12}$$

Finally, note also that

$$n_\ell^{k/2}\|U_n^{(k)}(\pi_k G)\|_{\mathcal{G}_{\ell,j}} \leq C_k M_k \mathcal{U}_n(j,k,\ell), \tag{3.13}$$

for some $C_k > 0$. Therefore by (3.4), for each $k = 2, \ldots, m$ and ℓ large enough,

$$\begin{aligned}\max_{n_{\ell-1}<n\leq n_\ell} A_{n,k} &:= \max_{n_{\ell-1}<n\leq n_\ell}\sup_{a_n\leq h\leq b_0}\sup_{g\in\mathcal{F}}\sup_{\mathbf{t}\in\mathbb{R}^m}\frac{\sqrt{nh^m}|U_n^{(k)}(\pi_k\bar{G}_{g,h,\mathbf{t}})|}{\sqrt{(|\log h| \vee \log\log n)^k}} \\ &\leq \max_{n_{\ell-1}<n\leq n_\ell}\max_{1\leq j\leq L(\ell)}\sup_{h\in\mathcal{H}_{\ell,j}}\sup_{g\in\mathcal{F}}\sup_{\mathbf{t}\in\mathbb{R}^m}\frac{\sqrt{n_\ell h^m}|U_n^{(k)}(\pi_k\bar{G}_{g,h,\mathbf{t}})|}{\sqrt{(|\log h| \vee \log\log n_\ell)^k}} \\ &\leq \frac{C_k M_k}{\sqrt{a_{n_\ell}^m n_\ell^{k-1}}}\max_{n_{\ell-1}<n\leq n_\ell}\max_{1\leq j\leq L(\ell)}\frac{\mathcal{U}_n(j,k,\ell)}{\lambda_{j,k}(\ell)}, \end{aligned}$$

where $\lambda_{j,k}(\ell)$ was defined as in (3.10). Now recall that $h_{\ell,j} \leq 2b_0 < 2$ for $j = 1, \ldots, L(\ell)$ and that $L(\ell) \leq 2\log n_\ell$. Then (3.12) applied with $\rho \geq (2+\delta)/C_{5,k}$, $\delta > 0$ and in combination with the above inequality and the obvious bound $\sqrt{a_n^m n^{k-1}}A_{n,k} \leq \sqrt{a_{n_\ell}^m n_\ell^{k-1}}A_{n,k}$ valid for all

$n_{\ell-1} < n \le n_\ell$, implies for $C_{6,k} \ge 2\rho^{k/2} C_k M_k C_{1,k}$ that for $k = 2, \ldots, m$

$$\begin{aligned}
\mathbb{P}\left\{ \max_{n_{\ell-1}<n\le n_\ell} \sqrt{a_n^m n^{k-1}} A_{n,k} > C_{6,k} \right\} &\le \sum_{j=1}^{L(\ell)} \sqrt{h_{\ell,j}^m} (\log n_\ell)^{-\rho C_{5,k}} \\
&\le L(\ell)\sqrt{2^m} (\log n_\ell)^{-\rho C_{5,k}} \\
&\le \sqrt{2^{m+2}} (\ell \log 2)^{-(1+\delta)}.
\end{aligned}$$

This proves via some elementary bounds and Borel–Cantelli that (3.3) holds, which obviously implies (3.2), and hence completes the proof of Theorem 1.

4 Proof of Theorem 2 : the unbounded case

In case (1.5) is not satisfied, we consider bandwidths lying in the slightly smaller interval $\mathcal{H}'_{n_\ell} = [a'_{n_\ell}, b_0]$ that can be decomposed into the subintervals

$$\mathcal{H}'_{\ell,j} := [h'_{\ell,j-1}, h'_{\ell,j}] \quad \text{with} \quad h'^m_{\ell,j} := 2^j a'^m_{n_\ell}. \tag{4.1}$$

Note that it is straightforward to show that (3.4) remains valid if we replace $h_{\ell,j}$ by $h'_{\ell,j}$. In particular, we still have $L(\ell) \le 2\log n_\ell$ where $L(\ell)$ is now defined as $L(\ell) := \max\{j : h'_{\ell,j} \le 2b_0\}$. Recall that $n_\ell = 2^\ell, \ell \ge 0$ and set for $\ell \ge 1$

$$\gamma_\ell = n_\ell / \log n_\ell. \tag{4.2}$$

For an arbitrary $\varepsilon > 0$ we shall decompose each function in $\mathcal{G}$ as

$$\begin{aligned}
\bar{G}_{g,h,\mathbf{t}}(\mathbf{x},\mathbf{y}) &= \bar{G}_{g,h,\mathbf{t}}(\mathbf{x},\mathbf{y}) \mathbb{I}\{\widetilde{F}(\mathbf{y}) \le \varepsilon \gamma_\ell^{1/p}\} + \bar{G}_{g,h,\mathbf{t}}(\mathbf{x},\mathbf{y}) \mathbb{I}\{\widetilde{F}(\mathbf{y}) > \varepsilon \gamma_\ell^{1/p}\} \\
&=: \bar{G}^{(\ell)}_{g,h,\mathbf{t}}(\mathbf{x},\mathbf{y}) + \widetilde{G}^{(\ell)}_{g,h,\mathbf{t}}(\mathbf{x},\mathbf{y}),
\end{aligned}$$

where $\widetilde{F}(\mathbf{y})$ is the (symmetric) envelope function of the class $\mathcal{G}$ as defined in (2.6). Then $u_n(g,h,\mathbf{t})$ can be decomposed as well for any $n_{\ell-1} < n \le n_\ell$, since from (2.2),

$$\begin{aligned}
u_n(g,h,\mathbf{t}) &= \sqrt{n}\{U_n^{(m)}(\bar{G}^{(\ell)}_{g,h,\mathbf{t}}) - \mathbb{E}U_n^{(m)}(\bar{G}^{(\ell)}_{g,h,\mathbf{t}})\} + \sqrt{n}\{U_n^{(m)}(\widetilde{G}^{(\ell)}_{g,h,\mathbf{t}}) - \mathbb{E}U_n^{(m)}(\widetilde{G}^{(\ell)}_{g,h,\mathbf{t}})\} \\
&=: u_n^{(\ell)}(g,h,\mathbf{t}) + \widetilde{u}_n^{(\ell)}(g,h,\mathbf{t}).
\end{aligned}$$

The term $u_n^{(\ell)}(g,h,\mathbf{t})$ will be called the truncated part and $\widetilde{u}_n^{(\ell)}(g,h,\mathbf{t})$ the remainder part. To prove Theorem 2 we shall apply the Hoeffding decomposition to the truncated part and analyze each of the terms separately, while the remainder part can be treated directly using simple arguments based on standard inequalities. Note for further use that

$$a'^m_{n_\ell} = c^m \gamma_\ell^{2/p-1}, \quad \ell \ge 1. \tag{4.3}$$

4.1 Truncated part

Note that from (2.3) we need to consider the terms of $\sum_{k=1}^{m}\binom{m}{k}U_n^{(k)}(\pi_k\bar{G}_{g,h,\mathbf{t}}^{(\ell)})$. We shall start with the linear term in this decomposition. Following the same reasoning as in the previous section, we can show that $\pi_1\bar{G}_{g,h,\mathbf{t}}^{(\ell)}$ is a centered conditional expectation, and that the first term of (2.3) can be written as an empirical process based upon the sample $(X_1,Y_1),\ldots,(X_n,Y_n)$ and indexed by the class of functions

$$\mathcal{S}'_\ell := \left\{S_{g,h,\mathbf{t}}^{(\ell)}(\cdot,\cdot) : g \in \mathcal{F}, h \in \mathcal{H}'_{n_\ell}, \mathbf{t} \in \mathbb{R}^m\right\},$$

where $\mathcal{H}'_{n_\ell}$ was defined in the beginning of this section, and where

$$S_{g,h,\mathbf{t}}^{(\ell)}(x,y) = mh^m\mathbb{E}\left[\bar{G}_{g,h,\mathbf{t}}^{(\ell)}(\mathbf{X},\mathbf{Y})\big|(X_1,Y_1)=(x,y)\right].$$

To show that $\mathcal{S}'_\ell$ is a VC–class, introduce the class of functions of $(\mathbf{x},\mathbf{y}) \in \mathbb{R}^m \times \mathbb{R}^m$,

$$\mathcal{C} = \left\{h^m\bar{G}_{g,h,\mathbf{t}}(\mathbf{x},\mathbf{y})\mathbb{I}\{\widetilde{F}(\mathbf{y}) \le c\} : g \in \mathcal{F}, 0<h<1, \mathbf{t}\in\mathbb{R}^m, c>0\right\}.$$

Since both $\mathcal{G}$ as defined in (2.5) and the class of functions of $\mathbf{y} \in \mathbb{R}^m$ given by $\mathcal{I} = \{\mathbb{I}\{\widetilde{F}(\mathbf{y}) \le c\} : c > 0\}$ are of VC–type (and note that $\mathcal{I}$ has a bounded envelope function), we can apply Lemma A.1 in Einmahl and Mason [4] to conclude that $\mathcal{C}$ is of VC–type as well. Therefore, so is the class of functions $m\mathcal{C}^{(1)}$ on $\mathbb{R}^2$, where $\mathcal{C}^{(1)}$ consists of the π_1-projections of the functions in the class $\mathcal{C}$. Thus we see that $\mathcal{S}'_\ell \subset m\mathcal{C}^{(1)}$ and hence $\mathcal{S}'_\ell$ is of VC–type with the same characteristics as $m\mathcal{C}^{(1)}$. Now, to find an envelope function for $\mathcal{S}'_\ell$, set $\mathbf{t}_j := (t_1,\ldots,t_{j-1},t_{j+1},\ldots,t_m) \in \mathbb{R}^{m-1}$, and $\mathbf{Z}_j(u) := (Z_1,\ldots,Z_{j-1},u,Z_{j+1},\ldots,Z_m) \in \mathbb{R}^m$ for $u \in \mathbb{R}$ and $\mathbf{Z} \in \mathbb{R}^m$. We can then rewrite the function $S_{g,h,\mathbf{t}}^{(\ell)}(x,y) \in \mathcal{S}'_\ell$ as

$$\begin{aligned} S_{g,h,\mathbf{t}}^{(\ell)}(x,y) = {} & K\left(\frac{t_1-x}{h}\right)\mathbb{E}\Big[g(\mathbf{Y}_1(y))\widetilde{K}\Big(\frac{\mathbf{t}_1-\mathbf{X}^*}{h}\Big)\mathbb{I}\{\widetilde{F}(\mathbf{Y}_1(y)) \le \varepsilon\gamma_\ell^{1/p}\}\Big] \\ & + K\Big(\frac{t_2-x}{h}\Big)\mathbb{E}\Big[g(\mathbf{Y}_2(y))\widetilde{K}\Big(\frac{\mathbf{t}_2-\mathbf{X}^*}{h}\Big)\mathbb{I}\{\widetilde{F}(\mathbf{Y}_2(y)) \le \varepsilon\gamma_\ell^{1/p}\}\Big] \\ & + \ldots + K\Big(\frac{t_m-x}{h}\Big)\mathbb{E}\Big[g(\mathbf{Y}_m(y))\widetilde{K}\Big(\frac{\mathbf{t}_m-\mathbf{X}^*}{h}\Big)\mathbb{I}\{\widetilde{F}(\mathbf{Y}_m(y)) \le \varepsilon\gamma_\ell^{1/p}\}\Big], \end{aligned} \tag{4.4}$$

where $\mathbf{X}^* = (X_2,\ldots,X_m) \in \mathbb{R}^{m-1}$ and where (with abuse of notation here) the product kernel in $(K.iii)$ is now defined for $(m-1)$–dimensional vectors, i.e. $\widetilde{K}(\mathbf{u}) = \prod_{i=1}^{m-1}K(u_i)$, $\mathbf{u} \in \mathbb{R}^{m-1}$. Hence, we can bound $S_{g,h,\mathbf{t}}^{(\ell)}(x,y)$ simply as

$$\begin{aligned} |S_{g,h,\mathbf{t}}^{(\ell)}(x,y)| \le {} & \kappa^m\{\mathbb{E}[F(y,Y_2,\ldots,Y_m)] + \mathbb{E}[F(Y_2,y,Y_3,\ldots,Y_m)] \\ & + \ldots + \mathbb{E}[F(Y_2,\ldots,Y_m,y)]\} \\ =: {} & G_m(x,y). \end{aligned}$$

We shall now apply the moment bound in Theorem A.3 to the subclasses

$$\mathcal{S}'_{\ell,j} := \Big\{ S^{(\ell)}_{g,h,\mathbf{t}}(\cdot,\cdot) : g \in \mathcal{F}, h \in \mathcal{H}'_{\ell,j}, \mathbf{t} \in \mathbb{R}^m \Big\}, \quad 1 \le j \le L(\ell),$$

where $\mathcal{H}'_{\ell,j}$ was defined in (4.1). Since $\mathcal{S}'_{\ell,j} \subset \mathcal{S}'_\ell$ for $j = 1, \ldots, L(\ell)$, all these subclasses are of VC–type with the same envelope function and characteristics as the class $m\mathcal{C}^{(1)}$ (which is independent of ℓ), verifying (ii) in the Theorem. For (i), recall that although all the terms of the envelope function $G_m(x,y)$ are different, their expectation is the same. Therefore, denoting $\mathbf{Y}^*$ for $(Y_2, \ldots, Y_m)$ and applying Minkowski's inequality followed by Jensen's inequality, we obtain from assumption (1.6) the following upper bound for the second moment of the envelope function.

$$\begin{aligned}
\mathbb{E}G_m^2(X,Y) &= \kappa^{2m}\mathbb{E}_Y\left\{\mathbb{E}_{\mathbf{Y}^*}[F(Y,Y_2,\ldots,Y_m)] + \mathbb{E}_{\mathbf{Y}^*}[F(Y_2,Y,Y_3,\ldots,Y_m)] \right.\\
&\qquad\qquad \left. + \ldots + \mathbb{E}_{\mathbf{Y}^*}[F(Y_2,\ldots,Y_m,Y)]\right\}^2 \\
&\le m^2\kappa^{2m}\mathbb{E}F^2(Y_1,\ldots,Y_m) \\
&\le m^2\kappa^{2m}\mu_p^{2/p}.
\end{aligned}$$

Note further that by symmetry of $\widetilde{F}$,

$$\mathbb{E}\bar{G}^{(\ell)}_{g,h,\mathbf{t}}(\mathbf{X},\mathbf{Y}) = h^{-m}\mathbb{E}[g(\mathbf{Y})\widetilde{K}\Big(\frac{\mathbf{t}-\mathbf{X}}{h}\Big)\mathbb{I}\{\widetilde{F}(\mathbf{Y}) \le \varepsilon\gamma_\ell^{1/p}\}],$$

such that Jensen's inequality, the change of variable $\mathbf{u} = (\mathbf{t}-\mathbf{x})/h$ and the assumption in (1.6) give the following upper bound for the second moment of any function in $\mathcal{S}'_\ell$:

$$\begin{aligned}
\mathbb{E}(S^{(\ell)}_{g,h,\mathbf{t}}(X,Y))^2 &\le m^2\mathbb{E}\Big[g^2(\mathbf{Y})\widetilde{K}^2\Big(\frac{\mathbf{t}-\mathbf{X}}{h}\Big)\mathbb{I}\{\widetilde{F}(\mathbf{Y}) \le \varepsilon\gamma_\ell^{1/p}\}\Big] \\
&\le m^2\kappa^{2m}h^m\int_{\left[-\frac{1}{2},\frac{1}{2}\right]^m}\mathbb{E}[F^2(\mathbf{Y})\big|\mathbf{X} = \mathbf{t}-h\mathbf{u}]f_X(t_1-hu_1)\ldots f_X(t_m-hu_m)d\mathbf{u} \\
&\le m^2\kappa^{2m}\mu_p^{2/p}\|f_X\|_\infty^m h^m. \qquad (4.5)
\end{aligned}$$

Therefore, with $\beta \equiv m\kappa^m\mu_p^{1/p}(1 \vee \|f_X\|_\infty^m)$, our previous calculations give us that

$$\mathbb{E}G_m^2(X,Y) \le \beta^2 \quad \text{and} \quad \sup_{S \in \mathcal{S}'_{\ell,j}} \mathbb{E}S^2(X,Y) \le \beta^2 h'^m_{\ell,j} =: \sigma^2_{\ell,j},$$

verifying condition (iii) as well. Finally, recall from (2.6) that since $\mathcal{G}$ has envelope function $\widetilde{F}(\mathbf{y})$, it holds for all $x,y \in \mathbb{R}$ that

$$|S^{(\ell)}_{g,h,\mathbf{t}}(x,y)| \le m\mathbb{E}[\widetilde{F}(\mathbf{Y})\mathbb{I}\{\widetilde{F}(\mathbf{Y}) \le \varepsilon\gamma_\ell^{1/p}\}\big|(X_1,Y_1) = (x,y)] \le m\varepsilon\gamma_\ell^{1/p},$$

such that by taking $\varepsilon > 0$ small enough, Theorem A.3 is now applicable, and gives us an absolute constant $A_1 < \infty$ for which

$$\begin{aligned}\mathbb{E}\|\sum_{i=1}^{n_\ell} \epsilon_i S(X_i, Y_i)\|_{\mathcal{S}'_{\ell,j}} &\leq A_1\sqrt{n_\ell h'^m_{\ell,j}|\log h'_{\ell,j}|} \\ &\leq A_1\sqrt{n_\ell h'^m_{\ell,j}(|\log h'_{\ell,j}| \vee \log\log n_\ell)} \\ &=: A_1\lambda'_j(\ell), \end{aligned} \tag{4.6}$$

where $\epsilon_1, \ldots, \epsilon_{n_\ell}$ are independent Rademacher variables, independent of (X_i, Y_i), $1 \leq i \leq n_\ell$. Consequently, applying the exponential inequality of Talagrand [9] to the class $\mathcal{S}'_{\ell,j}$ (see Theorem A.5 in the Appendix) with $M = m\varepsilon\gamma_\ell^{1/p}$, $\sigma^2_{\mathcal{S}'_{\ell,j}} = \beta^2 h'^m_{\ell,j}$ and the moment bound in (4.6), we get for an absolute constant $A_2 < \infty$ and all $t > 0$ that

$$\begin{aligned}&\mathbb{P}\left\{\max_{n_{\ell-1}<n\leq n_\ell} \|\sqrt{n}\alpha_n\|_{\mathcal{S}'_{\ell,j}} \geq C_1(A_1\lambda'_j(\ell) + t)\right\} \\ &\qquad \leq 2\left[\exp\left(-\frac{A_2 t^2}{n_\ell\beta^2 h'^m_{\ell,j}}\right) + \exp\left(-\frac{A_2 t}{m\varepsilon\gamma_\ell^{1/p}}\right)\right]. \end{aligned} \tag{4.7}$$

Towards applying this inequality with $t = \rho\lambda'_j(\ell), \rho > 1$, note that it clearly follows from (4.3) and the definitions of $h'_{\ell,j}$ and $\lambda'_j(\ell)$ that for all $j \geq 0$,

$$\begin{aligned}\frac{\lambda'^2_j(\ell)}{n_\ell h'^m_{\ell,j}} &= |\log h'_{\ell,j}| \vee \log\log n_\ell \geq \log\log n_\ell, \\ \frac{\lambda'^2_j(\ell)}{\gamma_\ell^{2/p}} &= 2^j c^m \log n_\ell(|\log h'_{\ell,j}| \vee \log\log n_\ell) \geq c^m(\log\log n_\ell)^2.\end{aligned}$$

Consequently, (4.7) when applied with $t = \rho\lambda'_j(\ell)$ and any $\rho > 1$ with ℓ large enough, yields for suitable constants A'_2, A''_2 and A_3, the inequality

$$\begin{aligned}&\mathbb{P}\left\{\max_{n_{\ell-1}<n\leq n_\ell} \|\sqrt{n}\alpha_n\|_{\mathcal{S}'_{\ell,j}} \geq C_1(A_1 + \rho)\lambda'_j(\ell)\right\} \\ &\qquad \leq 2\left[\exp\left(-A'_2\rho^2\log\log n_\ell\right) + \exp\left(-A''_2\rho\log\log n_\ell\right)\right] \\ &\qquad \leq 4(\log n_\ell)^{-A_3\rho}. \end{aligned} \tag{4.8}$$

Keeping in mind that $mh^m\sqrt{n}U_n^{(1)}(\pi_1\bar{G}^{(\ell)}_{g,h,\mathbf{t}})$ is an empirical process $\alpha_n(S^{(\ell)}_{g,h,\mathbf{t}})$ indexed by the class $\mathcal{S}'_\ell$, and recalling (3.4), we obtain for $\ell \geq 1$ that,

$$\max_{n_{\ell-1}<n\leq n_\ell} A'_{n,\ell} := \max_{n_{\ell-1}<n\leq n_\ell} \sup_{a'_n\leq h\leq b_0} \sup_{g\in\mathcal{F}} \sup_{\mathbf{t}\in\mathbb{R}^m} \frac{m\sqrt{nh^m}|U_n^{(1)}(\pi_1\bar{G}^{(\ell)}_{g,h,\mathbf{t}})|}{\sqrt{|\log h| \vee \log\log n}}$$

$$
\begin{aligned}
&\le \max_{n_{\ell-1}<n\le n_\ell} \max_{1\le j\le L(\ell)} \sup_{h\in\mathcal{H}'_{\ell,j}} \sup_{g\in\mathcal{F}} \sup_{\mathbf{t}\in\mathbb{R}^m} \frac{2\sqrt{2}\,|\sqrt{n}\alpha_n(S^{(\ell)}_{g,h,\mathbf{t}})|}{\sqrt{n_\ell h'^m_{\ell,j}(|\log h'_{\ell,j}|\vee\log\log n_\ell)}} \\
&\le \max_{n_{\ell-1}<n\le n_\ell} \max_{1\le j\le L(\ell)} \sup_{H\in\mathcal{S}'_{\ell,j}} \frac{3|\sqrt{n}\alpha_n(H)|}{\lambda'_j(\ell)}.
\end{aligned}
$$

Consequently, recalling once again that $L(\ell)\le 2\log n_\ell$, we can infer from (4.8) that for some constant $C_5(\rho)\ge 3C_1(A_1+\rho)$,

$$
\begin{aligned}
\mathbb{P}\left\{\max_{n_{\ell-1}<n\le n_\ell} A'_{n,\ell} > C_5(\rho)\right\} &\le \sum_{j=1}^{L(\ell)} \mathbb{P}\left\{\max_{n_{\ell-1}<n\le n_\ell} \|\sqrt{n}\alpha_n\|_{\mathcal{S}'_{\ell,j}} > C_1(A_1+\rho)\lambda'_j(\ell)\right\} \\
&\le 8(\log n_\ell)^{1-A_3\rho}.
\end{aligned}
$$

The Borel–Cantelli lemma when combined with this inequality for $\rho\ge(2+\delta)/A_3$, $\delta>0$ and with the choice $n_\ell=2^\ell$, establish for some $C'<\infty$ and with probability one, that

$$
\limsup_{\ell\to\infty} \max_{n_{\ell-1}<n\le n_\ell} \sup_{a'_n\le h\le b_0} \sup_{g\in\mathcal{F}} \sup_{\mathbf{t}\in\mathbb{R}^m} \frac{m\sqrt{nh^m}|U_n^{(1)}(\pi_1\bar{G}^{(\ell)}_{g,h,\mathbf{t}})|}{\sqrt{|\log h|\vee\log\log n}} \le C', \tag{4.9}
$$

finishing the study of the first term in (2.3). We now show that all the other terms of (2.3) are asymptotically bounded or go to zero at the proper rate, which will be obtained if we can prove that for $k=2,\ldots,m$ and with probability one,

$$
\max_{n_{\ell-1}<n\le n_\ell} \sup_{a'_n\le h\le b_0} \sup_{g\in\mathcal{F}} \sup_{\mathbf{t}\in\mathbb{R}^m} \frac{\sqrt{nh^m}|U_n^{(k)}(\pi_k\bar{G}^{(\ell)}_{g,h,\mathbf{t}})|}{\sqrt{|\log h|\vee\log\log n}} = O(\gamma_\ell^{1-k/2}). \tag{4.10}
$$

Analogously to the bounded case, we start by defining the classes of functions on $\mathbb{R}^m\times\mathbb{R}^m$ and $\mathbb{R}^k\times\mathbb{R}^k$,

$$
\begin{aligned}
\mathcal{G}'_{\ell,j} &:= \left\{h^m\bar{G}^{(\ell)}_{g,h,\mathbf{t}}(\cdot,\cdot) : g\in\mathcal{F}, h\in\mathcal{H}'_{\ell,j}, \mathbf{t}\in\mathbb{R}^m\right\}, \\
\mathcal{G}'^{(k)}_{\ell,j} &:= \left\{h^m(\pi_k\bar{G}^{(\ell)}_{g,h,\mathbf{t}})(\cdot,\cdot)/(2^k\varepsilon\gamma_\ell^{1/p}) : g\in\mathcal{F}, h\in\mathcal{H}'_{\ell,j}, \mathbf{t}\in\mathbb{R}^m\right\}.
\end{aligned}
$$

Then it is easily verified that these classes are of VC–type with characteristics that are independent of ℓ, and with envelope functions $\widetilde{F}$ and $(2^k\varepsilon\gamma_\ell^{1/p})^{-1}F_k$ respectively. The function $\widetilde{F}$ is defined as in (2.6) and F_k is determined just as in the proof of Theorem 1 in Giné and Mason [7]. Note that, in the same spirit as (3.5) and (3.7), by setting

$$
\mathcal{U}'_n(j,k,\ell) := \sup_{H\in\mathcal{G}'^{(k)}_{\ell,j}} \Big|\frac{1}{n_\ell^{k/2}}\sum_{\mathbf{i}\in I_n^k} H(\mathbf{X_i},\mathbf{Y_i})\Big|, \quad n_{\ell-1}<n\le n_\ell,
$$

we have for all $k = 2, \dots, m$ and $n_{\ell-1} < n \leq n_\ell$,

$$\mathcal{U}'_n(j,k,\ell) \leq n_\ell^{k/2} \|U_n^{(k)}(\pi_k G)\|_{\mathcal{G}'^{(k)}_{\ell,j}}.$$

Consequently, applying Theorem A.1 with $c = 1/2$ and $r = 2$, gives us precisely (3.6) with $\mathcal{U}_n(j,k,\ell)$ and $\mathcal{U}_{n_\ell}(j,k,\ell)$ replaced by $\mathcal{U}'_n(j,k,\ell)$ and $\mathcal{U}'_{n_\ell}(j,k,\ell)$ respectively. Therefore the same methodology as in the bounded case will be applied. Note also that, as held for all the functions in $\mathcal{G}^{(k)}_{\ell,j}$, the functions in $\mathcal{G}'^{(k)}_{\ell,j}$ are bounded by 1, and have second moments that can be bounded by $h^m D_m$ for a suitable D_m by arguing as in (4.5) and (3.8). Consequently, the expression in (3.8) is satisfied for functions in $\mathcal{G}'^{(k)}_{\ell,j}$ as well, i.e.

$$\sup_{H \in \mathcal{G}'^{(k)}_{\ell,j}} \mathbb{E}H^2(\mathbf{X},\mathbf{Y}) \leq D_m h'^m_{\ell,j} =: \sigma'^2_{\ell,j}.$$

Hence, all the conditions for Theorems A.4 and A.6 are satisfied, so that after some obvious identifications and modifications, the second part of the proof of Theorem 1 (and (3.12) in particular) gives us for all $j = 1, \dots, L(\ell)$ and any $\rho > 2$,

$$\mathbb{P}\left\{ \max_{n_{\ell-1} < n \leq n_\ell} \mathcal{U}'_n(j,k,\ell) > 2\rho^{k/2} y'^* \right\} \leq \sqrt{h'^m_{\ell,j}} (\log n_\ell)^{-C_{7,k}\rho}, \tag{4.11}$$

with $y'^* = C'_{1,k} \lambda'_{j,k}(\ell)$, and where $\lambda'_{j,k}(\ell)$ is defined as in (3.10) with $h_{\ell,j}$ replaced by $h'_{\ell,j}$. Now, to finish the proof of (4.10), note that similarly to (3.13), for some $C_k > 0$,

$$n_\ell^{k/2} \|U_n^{(k)}(\pi_k G)\|_{\mathcal{G}'_{\ell,j}} \leq 2^k C_k \varepsilon \gamma_\ell^{1/p} \mathcal{U}'_n(j,k,\ell).$$

This gives that

$$\begin{aligned} \max_{n_{\ell-1} < n \leq n_\ell} A'_{n,\ell,k} &:= \max_{n_{\ell-1} < n \leq n_\ell} \sup_{a'_n \leq h \leq b_0} \sup_{g \in \mathcal{F}} \sup_{\mathbf{t} \in \mathbb{R}^m} \frac{\sqrt{nh^m} |U_n^{(k)}(\pi_k \bar{G}^{(\ell)}_{g,h,\mathbf{t}})|}{\sqrt{(|\log h| \vee \log\log n)^k}} \\ &\leq \frac{2^k c_k \varepsilon \gamma_\ell^{1/p}}{\sqrt{a'^m_{n_\ell} n_\ell^{k-1}}} \max_{n_{\ell-1} < n \leq n_\ell} \max_{1 \leq j \leq L(\ell)} \frac{\mathcal{U}'_n(j,k,\ell)}{\lambda'_{j,k}(\ell)}. \end{aligned}$$

From (4.3) we see now that $\gamma_\ell^{2/p} / a'^m_{n_\ell} n_\ell^{k-1} = c^{-m} n_\ell^{2-k} / \log n_\ell$. Therefore by choosing $C_{8,k} > 2^{k+1} c^{-m/2} \varepsilon c_k C'_{1,k} ((2+\delta)/C_{7,k})^{k/2}$ and noting that $h'_{\ell,j} < 2$ for all $j = 1, \dots, L(\ell)$, we can infer from (4.11) that

$$\mathbb{P}\Big\{ \max_{n_{\ell-1} < n \leq n_\ell} \sqrt{\frac{\log n}{n^{2-k}}} A'_{n,\ell,k} > C_{8,k} \Big\} \leq \sqrt{2^{m+1}} (\log n_\ell)^{-(1+\delta)}.$$

This implies immediately via Borel–Cantelli that for all $k = 2, \dots, m$ and $\ell \geq 1$,

$$\max_{n_{\ell-1} < n \leq n_\ell} \sup_{a'_n \leq h \leq b_0} \sup_{g \in \mathcal{F}} \sup_{\mathbf{t} \in \mathbb{R}^m} \frac{\sqrt{nh^m} |U_n^{(k)}(\pi_k \bar{G}^{(\ell)}_{g,h,\mathbf{t}})|}{\sqrt{(|\log h| \vee \log\log n)^k}} = O\left(\sqrt{\frac{n_\ell^{2-k}}{\log n_\ell}} \right), \quad \text{a.s.,}$$

which obviously implies (4.10). Finally, recalling the Hoeffding decomposition (2.3), this implies together with (4.9) that with probability one,

$$\limsup_{\ell\to\infty} \max_{n_{\ell-1}<n\le n_\ell} \sup_{a'_n\le h\le b_0} \sup_{g\in\mathcal{F}} \sup_{\mathbf{t}\in\mathbb{R}^m} \frac{\sqrt{nh^m}|U_n^{(m)}(\bar{G}^{(\ell)}_{g,h,\mathbf{t}}) - \mathbb{E}U_n^{(m)}(\bar{G}^{(\ell)}_{g,h,\mathbf{t}})|}{\sqrt{|\log h| \vee \log\log n}} \le C''. \tag{4.12}$$

4.2 Remainder part

Consider now the remainder process $\widetilde{u}_n^{(\ell)}(g,h,\mathbf{t})$ based on the unbounded (symmetric) U–kernel given by

$$\widetilde{G}^{(\ell)}_{g,h,\mathbf{t}}(\mathbf{x},\mathbf{y}) := \bar{G}_{g,h,\mathbf{t}}(\mathbf{x},\mathbf{y})\mathbb{I}\{\widetilde{F}(\mathbf{y}) > \varepsilon\gamma_\ell^{1/p}\},$$

where we defined γ_ℓ as in (4.2). We shall show that this U–process is asymptotically negligible at the rate given in Theorem 2. More precisely, we shall prove that as $\ell\to\infty$,

$$\max_{n_{\ell-1}<n\le n_\ell} \sup_{a'_n\le h\le b_0} \sup_{g\in\mathcal{F}} \sup_{\mathbf{t}\in\mathbb{R}^m} \frac{\sqrt{nh^m}|U_n^{(m)}(\widetilde{G}^{(\ell)}_{g,h,\mathbf{t}}) - \mathbb{E}U_n^{(m)}(\widetilde{G}^{(\ell)}_{g,h,\mathbf{t}})|}{\sqrt{|\log h| \vee \log\log n}} = o(1), \quad \text{a.s.} \tag{4.13}$$

Recall that for all $g\in\mathcal{F}$, $h\in[a'_n,b_0]$ and $\mathbf{t},\mathbf{x}\in\mathbb{R}^m$, $\widetilde{F}(\mathbf{y}) \ge h^m|\bar{G}_{g,h,\mathbf{t}}(\mathbf{x},\mathbf{y})|$, so from the symmetry of $\widetilde{F}$, it holds that

$$|U_n^{(m)}(\widetilde{G}^{(\ell)}_{g,h,\mathbf{t}})| \le h^{-m}U_n^{(m)}\left(\widetilde{F}\cdot\mathbb{I}\{\widetilde{F} > \varepsilon\gamma_\ell^{1/p}\}\right),$$

where $U_n^{(m)}(\widetilde{F}\cdot\mathbb{I}\{\widetilde{F} > \varepsilon\gamma_\ell^{1/p}\})$ is a U–statistic based on the positive and symmetric kernel $\mathbf{y}\to\widetilde{F}(\mathbf{y})\mathbb{I}\{\widetilde{F}(\mathbf{y}) > \varepsilon\gamma_\ell^{1/p}\}$. Recalling that $a_n'^m = c^m(\log n/n)^{1-2/p}$, we obtain easily that for all $g\in\mathcal{F}$, $h\in[a'_n,b_0]$, $\mathbf{t}\in\mathbb{R}^m$ and some $C>0$,

$$\begin{aligned}\max_{n_{\ell-1}<n\le n_\ell} \frac{\sqrt{nh^m}|U_n^{(m)}(\widetilde{G}^{(\ell)}_{g,h,\mathbf{t}})|}{\sqrt{|\log h| \vee \log\log n}} &\le \frac{\sqrt{n_\ell}\, U_{n_\ell}^{(m)}\left(\widetilde{F}\cdot\mathbb{I}\{\widetilde{F} > \varepsilon\gamma_\ell^{1/p}\}\right)}{\sqrt{a_{n_\ell}'^m(|\log a'_{n_\ell}| \vee \log\log n_\ell)}} \\ &\le C\gamma_\ell^{1-1/p}U_{n_\ell}^{(m)}\left(\widetilde{F}\cdot\mathbb{I}\{\widetilde{F} > \varepsilon\gamma_\ell^{1/p}\}\right).\end{aligned}$$

Arguing in the same way, since a U–statistic is an unbiased estimator of its kernel, we get that uniformly in $g\in\mathcal{F}$, $h\in[a'_n,b_0]$ and $\mathbf{t}\in\mathbb{R}^m$,

$$\begin{aligned}\max_{n_{\ell-1}<n\le n_\ell} \frac{\sqrt{nh^m}|\mathbb{E}U_n^{(m)}(\widetilde{G}^{(\ell)}_{g,h,\mathbf{t}})|}{\sqrt{|\log h| \vee \log\log n}} &\le C\gamma_\ell^{1-1/p}\mathbb{E}U_{n_\ell}^{(m)}\left(\widetilde{F}\cdot\mathbb{I}\{\widetilde{F} > \varepsilon\gamma_\ell^{1/p}\}\right) \\ &\le C'\mathbb{E}[\widetilde{F}^p(\mathbf{Y})\mathbb{I}\{\widetilde{F}(\mathbf{Y}) > \varepsilon\gamma_\ell^{1/p}\}].\end{aligned} \tag{4.14}$$

From (4.14) we see that as $\ell \to \infty$,

$$\max_{n_{\ell-1}<n\le n_\ell} \sup_{a'_n\le h\le b_0} \sup_{g\in\mathcal{F}} \sup_{\mathbf{t}\in\mathbb{R}^m} \frac{\sqrt{nh^m}|\mathbb{E}U_n^{(m)}(\widetilde{G}_{g,h,\mathbf{t}}^{(\ell)})|}{\sqrt{|\log h|\vee\log\log n}} = o(1). \tag{4.15}$$

Thus to finish the proof of (4.13) it suffices to show that

$$U_{n_\ell}^{(m)}(\widetilde{F}\cdot\mathbf{1}\{\widetilde{F}>\varepsilon\gamma_\ell^{1/p}\}) = o(\gamma_\ell^{1/p-1}), \quad \text{a.s.} \tag{4.16}$$

First note that from Chebyshev's inequality and a well–known inequality for the variance of a U–statistic (see Theorem 5.2 of Hoeffding [8]) we get for any $\delta>0$,

$$\begin{aligned}
&\mathbb{P}\left\{\left|U_{n_\ell}^{(m)}\left(\widetilde{F}\cdot\mathbf{1}\{\widetilde{F}>\varepsilon\gamma_\ell^{1/p}\}\right)-\mathbb{E}U_{n_\ell}^{(m)}\left(\widetilde{F}\cdot\mathbf{1}\{\widetilde{F}>\varepsilon\gamma_\ell^{1/p}\}\right)\right|>\delta\gamma_\ell^{-(1-1/p)}\right\}\\
&\qquad\le\ \delta^{-2}\gamma_\ell^{2-2/p}\mathrm{Var}\left(U_{n_\ell}^{(m)}\left(\widetilde{F}\cdot\mathbf{1}\{\widetilde{F}>\varepsilon\gamma_\ell^{1/p}\}\right)\right)\\
&\qquad\le\ m\delta^{-2}\frac{n_\ell^{1-2/p}}{(\log n_\ell)^{2-2/p}}\mathbb{E}[\widetilde{F}^2(\mathbf{Y})\mathbf{1}\{\widetilde{F}(\mathbf{Y})>\varepsilon\gamma_\ell^{1/p}\}].
\end{aligned} \tag{4.17}$$

Next, in order to establish the finite convergence of the series of the above probabilities, we split the indicator function $\mathbf{1}\{\widetilde{F}(\mathbf{Y})>\varepsilon\gamma_\ell^{1/p}\}$ into two distinct parts determined by whether $\widetilde{F}(\mathbf{Y})>n_\ell^{1/p}$ or $\varepsilon\gamma_\ell^{1/p}<\widetilde{F}(\mathbf{Y})\le n_\ell^{1/p}$, and consider the corresponding second moments in (4.17) separately. In the second case, note that from (1.6) and (2.6), $\mathbb{E}\widetilde{F}^p(\mathbf{Y})\le\mu_p\kappa^{pm}(m!)^p$, and observe that since $p>2$ and $n_\ell=2^\ell$,

$$\sum_{\ell=1}^{\infty}\frac{n_\ell^{1-2/p}}{(\log n_\ell)^{2-2/p}}\mathbb{E}[\widetilde{F}^2(\mathbf{Y})\mathbf{1}\{\widetilde{F}(\mathbf{Y})>n_\ell^{1/p}\}]\le\mathbb{E}[\widetilde{F}^p(\mathbf{Y})]\sum_{\ell=1}^{\infty}(\log n_\ell)^{-(2-2/p)}<\infty.$$

To handle the first case, we shall need the following fact from Einmahl and Mason [4].

Fact 1 *Let $(c_n)_{n\ge1}$ be a sequence of positive constants such that $c_n/n^{1/s}\nearrow\infty$ for $s>0$, and let Z be a random variable satisfying $\sum_{n=1}^\infty\mathbb{P}\{|Z|>c_n\}<\infty$. Then we have for any $q>s$,*

$$\sum_{k=1}^{\infty}2^k\mathbb{E}[|Z|^q\mathbf{1}\{|Z|\le c_{2^k}\}]/(c_{2^k})^q<\infty.$$

Setting $c_n=n^{1/p}$ into Fact 1, we conclude from this inequality that for $p<s<r\le 2p$,

$$\sum_{\ell=1}^{\infty}\frac{n_\ell^{1-2/p}}{(\log n_\ell)^{2-2/p}}\mathbb{E}[\widetilde{F}^2(\mathbf{Y})\mathbf{1}\{\varepsilon\gamma_\ell^{1/p}<\widetilde{F}(\mathbf{Y})\le n_\ell^{1/p}\}]$$

$$\leq \sum_{\ell=1}^{\infty} \frac{\varepsilon^{r-2}}{(\log n_\ell)^{2-r/p}} \frac{n_\ell \mathbb{E}[\widetilde{F}^r(\mathbf{Y})\mathbb{I}\{\widetilde{F}(\mathbf{Y}) \leq n_\ell^{1/p}\}]}{n_\ell^{r/p}} < \infty.$$

Finally, note that the bound leading to (4.14) implies that

$$\gamma_\ell^{1-1/p} \mathbb{E} U_{n_\ell}^{(m)}\big(\widetilde{F} \cdot \mathbb{I}\{\widetilde{F} > \varepsilon \gamma_\ell^{1/p}\}\big) = o(1).$$

Consequently, the above results together with (4.17) imply via Borel-Cantelli and the arbitrary choice of $\delta > 0$ that (4.16) holds, which when combined with (4.15) and (4.14) completes the proof of (4.13). This also finishes the proof of Theorem 2 since we have already established the result in (4.12).

5 Proof of Theorem 3 : uniform consistency of $\hat{m}_n(\mathbf{t}, h)$ to $m_\varphi(\mathbf{t})$

Theorem 3 is essentially a consequence of Theorem A.2 in the Appendix. Recall that a U–statistic with U–kernel H is an unbiased estimator of $\mathbb{E}H$. Writing $d\mathbf{x}$ and $d\mathbf{y}$ for $dx_1 dx_2 \dots dx_m$ and $dy_1 dy_2 \dots dy_m$ respectively, we see that

$$\mathbb{E} U_n(1, h, \mathbf{t}) = \int \widetilde{K}_h(\mathbf{t} - \mathbf{x}) f(x_1, y_1) \cdots f(x_m, y_m) d\mathbf{x} d\mathbf{y} = \widetilde{f} * \widetilde{K}_h(\mathbf{t}),$$

where the function $\widetilde{f} : \mathbb{R}^m \to \mathbb{R}$ is defined in (1.9). Since we assume f_X to be continuous on $J = I^\eta$, the function $\widetilde{f}$ is continuous on $J^m = J \times \ldots \times J$. Therefore we can infer from Theorem A.2 that

$$\sup_{0<h<b_n} \sup_{\mathbf{t} \in I^m} |\mathbb{E} U_n(1, h, \mathbf{t}) - \widetilde{f}(\mathbf{t})| \longrightarrow 0, \tag{5.1}$$

for all sequences of positive constants $b_n \to 0$, and where $I^m = I \times \ldots \times I$. In the same way, notice that

$$\begin{aligned} \mathbb{E} U_n(\varphi, h, \mathbf{t}) &= \int \varphi(\mathbf{y}) \widetilde{K}_h(\mathbf{t} - \mathbf{x}) f(x_1, y_1) \cdots f(x_m, y_m) d\mathbf{x} d\mathbf{y} \\ &= \Big\{ \mathbb{E}\left[\varphi(\mathbf{Y}) | \mathbf{X} = \cdot\right] \ \widetilde{f}(\cdot) \Big\} * \widetilde{K}_h(\mathbf{t}). \end{aligned}$$

Hence, Theorem A.2 applied to the class of functions $\mathcal{M}$ as defined in (1.8) gives that

$$\sup_{0<h<b_n} \sup_{\varphi \in \mathcal{F}} \sup_{\mathbf{t} \in I^m} |\mathbb{E} U_n(\varphi, h, \mathbf{t}) - m_\varphi(\mathbf{t}) \widetilde{f}(\mathbf{t})| \longrightarrow 0. \tag{5.2}$$

Keeping in mind the definition of $\widehat{\mathbb{E}} \hat{m}_{n,\varphi}(\mathbf{t}, h)$ in (1.7), it is clear that since f_X is bounded away from zero on J, (5.1) and (5.2) imply that

$$\sup_{0<h<b_n} \sup_{\varphi \in \mathcal{F}} \sup_{\mathbf{t} \in I^m} |\widehat{\mathbb{E}} \hat{m}_{n,\varphi}(\mathbf{t}, h) - m_\varphi(\mathbf{t})| = o(1),$$

finishing the proof of Theorem 3.

6 Proof of Theorem 4 : convergence rates of the conditional U–statistic $\hat{m}_{n,\varphi}(\mathbf{t}, h)$

Observe that

$$\begin{aligned}
|\hat{m}_{n,\varphi}(\mathbf{t}, h) - \widehat{\mathbb{E}}\hat{m}_{n,\varphi}(\mathbf{t}, h)| &= \left|\frac{U_n(\varphi, h, \mathbf{t})}{U_n(1, h, \mathbf{t})} - \frac{\mathbb{E}U_n(\varphi, h, \mathbf{t})}{\mathbb{E}U_n(1, h, \mathbf{t})}\right| \\
&\leq \frac{|U_n(\varphi, h, \mathbf{t}) - \mathbb{E}U_n(\varphi, h, \mathbf{t})|}{|U_n(1, h, \mathbf{t})|} \\
&\qquad + \frac{|\mathbb{E}U_n(\varphi, h, \mathbf{t})| \cdot |U_n(1, h, \mathbf{t}) - \mathbb{E}U_n(1, h, \mathbf{t})|}{|U_n(1, h, \mathbf{t})| \cdot |\mathbb{E}U_n(1, h, \mathbf{t})|} \\
&=: (\mathrm{I}) + (\mathrm{II}).
\end{aligned}$$

From Theorem 1, (5.1) and f_X bounded away from zero on J we get for some $\xi_1, \xi_2 > 0$ and c large enough in $a_n = c(\log n/n)^{1/m}$,

$$\liminf_{n\to\infty} \sup_{a_n \leq h < b_n} \sup_{\mathbf{t}\in I^m} |U_n(1, h, \mathbf{t})| = \xi_1 > 0, \quad \text{a.s.},$$

and for n large enough,

$$\sup_{a_n \leq h < b_n} \sup_{\mathbf{t}\in I^m} |\mathbb{E}U_n(1, h, \mathbf{t})| = \xi_2 > 0.$$

Further, for a''_n be either a_n or a'_n, we obtain readily from the assumptions (1.5) or (1.6) on the envelope function that

$$\sup_{a''_n \leq h < b_n} \sup_{\varphi\in\mathcal{F}} \sup_{\mathbf{t}\in I^m} |\mathbb{E}U_n(\varphi, h, \mathbf{t})| = O(1).$$

Hence, we can now use Theorem 1 to handle (II), while for (I), depending on whether the class $\mathcal{F}$ satisfies (1.5) or (1.6), we apply Theorem 1 or Theorem 2 respectively. Taking everything together we conclude that for c large enough and some $C'' > 0$, with probability one,

$$\begin{aligned}
&\limsup_{n\to\infty} \sup_{a''_n \leq h < b_n} \sup_{\varphi\in\mathcal{F}} \sup_{\mathbf{t}\in I^m} \frac{\sqrt{nh^m}|\hat{m}_{n,\varphi}(\mathbf{t}, h) - \widehat{\mathbb{E}}\hat{m}_{n,\varphi}(\mathbf{t}, h)|}{\sqrt{|\log h| \vee \log\log n}} \\
&\leq \limsup_{n\to\infty} \sup_{a''_n \leq h < b_n} \sup_{\varphi\in\mathcal{F}} \sup_{\mathbf{t}\in I^m} \frac{\sqrt{nh^m}(\mathrm{I})}{\sqrt{|\log h| \vee \log\log n}} \\
&\qquad + \limsup_{n\to\infty} \sup_{a''_n \leq h < b_n} \sup_{\varphi\in\mathcal{F}} \sup_{\mathbf{t}\in I^m} \frac{\sqrt{nh^m}(\mathrm{II})}{\sqrt{|\log h| \vee \log\log n}} \\
&\leq C'',
\end{aligned}$$

proving the assertion of Theorem 4.

A Appendix

The first result below is stated as Theorem 4 in Giné and Mason [7], and is essentially a consequence of a martingale inequality due to Brown [1]. The second Theorem is a generalization of Bochner's lemma.

Theorem A.1 (Theorem 4 of Giné and Mason, 2007b) *Let $X_1, X_2, \dots$ be i.i.d. S-valued with probability law P. Let $\mathcal{H}$ be a P–separable collection of measurable functions $f : S^k \to \mathbb{R}$ and assume that $\mathcal{H}$ is P–canonical (which means that every f in $\mathcal{H}$ is P–canonical). Further assume that $\mathbb{E}\|f(X_1, \dots, X_k)\|_{\mathcal{H}}^r < \infty$ for some $r > 1$, and let s be the conjugate of r. Then, with S_n defined as*

$$S_n = \sup_{f \in \mathcal{H}} \Big| \sum_{\mathbf{i} \in I_n^k} f(X_{i_1}, \dots, X_{i_k}) \Big|, \quad n \geq k,$$

we have for all $x > 0$ and $0 < c < 1$,

$$\mathbb{P}\Big\{ \max_{k \leq m \leq n} S_m > x \Big\} \leq \frac{\mathbb{P}\{S_n > cx\}^{1/s} (\mathbb{E} S_n^r)^{1/r}}{x(1-c)}.$$

Theorem A.2 *Let $I = [a, b]$ be a compact interval. Suppose that $\mathcal{H}$ is a uniformly equicontinuous family of real valued functions φ on $J = [a - \eta, b + \eta]^d$ for some $d \geq 1$ and $\eta > 0$. Further assume that K is an L_1–kernel with support in $[-1/2, 1/2]^d$ satisfying $\int_{\mathbb{R}^d} K(\mathbf{u}) d\mathbf{u} = 1$. Then uniformly in $\varphi \in \mathcal{H}$ and for any sequence of positive constants $b_n \to 0$,*

$$\sup_{0 < h < b_n} \sup_{\mathbf{z} \in I^d} |\varphi * K_h(\mathbf{z}) - \varphi(\mathbf{z})| \longrightarrow 0, \quad \text{as } n \to \infty,$$

where $K_h(\mathbf{z}) = h^{-d} K(\mathbf{z}/h)$ and

$$\varphi * K_h(\mathbf{z}) := h^{-d} \int_{\mathbb{R}^d} \varphi(\mathbf{x}) K\left(\frac{\mathbf{z} - \mathbf{x}}{h}\right) d\mathbf{x}.$$

A.1 Moment bounds

Theorem A.3 (Proposition 1 of Einmahl and Mason, 2005) *Let $\mathcal{G}$ be a pointwise measurable class of bounded functions with envelope function G such that for some constants $C, \nu \geq 1$ and $0 < \sigma \leq \beta$, the following conditions hold:*

(i) $\mathbb{E} G^2(X) \leq \beta^2$;

(ii) $\mathcal{N}(\epsilon, \mathcal{G}) \leq C \epsilon^{-\nu}, \quad 0 < \epsilon < 1$;

(iii) $\sigma_0^2 := \sup_{g \in \mathcal{G}} \mathbb{E} g^2(X) \leq \sigma^2$;

(iv) $\sup_{g \in \mathcal{G}} \|g\|_\infty \leq \frac{1}{4\sqrt{\nu}} \sqrt{n\sigma^2 / \log(C_1 \beta / \sigma)}$, *where* $C_1 = C^{1/\nu} \vee e$.

Then we have for some absolute constant A,

$$\mathbb{E}\|\sum_{i=1}^{n}\varepsilon_i g(X_i)\|_{\mathcal{G}} \leq A\sqrt{\nu n\sigma^2\log(C_1\beta/\sigma)},$$

where $\varepsilon_1,\ldots,\varepsilon_n$ are i.i.d Rademacher variables independent of $X_1,\ldots,X_n$.

Theorem A.4 (Corollary 1 of Giné and Mason, 2007b) *Let $\mathcal{F}$ be a collection of measurable functions $f : S^m \to \mathbb{R}$, symmetric in their entries with absolute values bounded by $M > 0$, and let P be any probability measure on $(S, \mathcal{S})$ (with X_i i.i.d–P). Assume that $\mathcal{F}$ is of VC–type with envelope function $F \equiv M$ and with characteristics A and v. Then for every $m \in \mathbb{N}$, $A \geq e^m, v \geq 1$ there exist constants $C_1 := C_1(m, A, v, M)$ and $C_2 = C_2(m, A, v, M)$ such that for $k = 1,\ldots,m$,*

$$n^k\mathbb{E}\|U_n^{(k)}(\pi_k f)\|_{\mathcal{F}}^2 \leq C_1^2 2^k\sigma^2\left(\log\frac{A}{\sigma}\right)^k,$$

assuming $n\sigma^2 \geq C_2\log(A/\sigma)$, where σ^2 is any number satisfying

$$\|P^m f^2\|_{\mathcal{F}} \leq \sigma^2 \leq M^2.$$

A.2 Exponential inequalities

Theorem A.5 (Talagrand, 1994) *Let $\mathcal{G}$ be a pointwise measurable class of functions satisfying*

$$\|g\|_\infty \leq M < \infty, \quad g \in \mathcal{G}.$$

Then we have for all $t > 0$,

$$\mathbb{P}\left\{\max_{1\leq m\leq n}\|\sqrt{m}\alpha_m\|_{\mathcal{G}} \geq A_1(\mathbb{E}\Big\|\sum_{i=1}^{n}\varepsilon_i g(X_i)\Big\|_{\mathcal{G}} + t)\right\}$$

$$\leq\ 2\left\{\exp\left(-\frac{A_2 t^2}{n\sigma_{\mathcal{G}}^2}\right) + \exp\left(-\frac{A_2 t}{M}\right)\right\},$$

where $\sigma_{\mathcal{G}}^2 = \sup_{g\in\mathcal{G}} Var(g(X))$ and A_1, A_2 are universal constants.

We now state the exponential inequality that will permit us to control the probability term in (3.6), and which is stated as Theorem 5.3.14 in de la Peña and Giné [2].

Theorem A.6 (Theorem 5.3.14 of de la Penã and Giné, 1999) *Let $\mathcal{H}$ be a VC–subgraph class of uniformly bounded measurable real valued kernels H on $(S^m, \mathcal{S}^m)$, symmetric in their entries. Then for each $1 \leq k \leq m$ there exist constants $c_k, d_k \in]0,\infty[$ such that, for all $n \geq m$ and $t > 0$,*

$$\left\{\|n^{k/2}U_n^{(k)}(\pi_k H)\|_{\mathcal{H}} > t\right\} \leq c_k\exp\{-d_k t^{2/k}\}.$$

Acknowledgements. We would like to thank Uwe Einmahl for pointing out an oversight and for very helpful suggestions. David Mason's research was partially supported by an NSF Grant. Julia Dony's research is financed by a PhD grant from the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT Vlaanderen).

References

[1] Brown, B. (1971). Martingale central limit theorems. *Ann. Math. Statist.* **42**, 59–66.

[2] de la Peña, V. H. and Giné, E. (1999). *Decoupling. From dependence to independence. Randomly stopped processes. U–statistics and processes. Martingales and beyond.* Probability and its Applications. Springer–Verlag, New York.

[3] Dony, J., Einmahl, U. and Mason, D.M. (2006). Uniform in bandwidth consistency of local polynomial regression function estimators. *Austr. J. Stat.* **35:2&3**, 105–120.

[4] 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.

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

[6] Giné, E. and Mason, D. M. (2007a). On local U–statistic processes and the estimation of densities of functions of several sample variables. To appear in *Ann. Stat.*

[7] Giné, E. and Mason, D.M. (2007b). Laws of the iterated logarithm for the local U–statistic process. To appear in *J. Theor. Probab.*

[8] Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Statistics* **19**, 293–325.

[9] Talagrand, M. (1994). Sharper bounds for Gaussian and empirical processes. *Ann. Probab.* **22,** 28–76.

[10] Sen, A. (1994). Uniform strong consistency rates for conditional U–statistics. *Sankhyā* **56, Series A,** 179–194.

[11] Stute, W. (1991). Conditional U–statistics. *Ann. Probab.* **19:2**, 812–825.

[12] van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes with Applications to Statistics,* Springer–Verlag, New York.