PROBABILITY
AND
MATHEMATICAL STATISTICS
Vol. 29, Fasc. 1 (2009), pp. 155–167

MAXIMAL INEQUALITIES FOR U-PROCESSES OF STRONGLY MIXING RANDOM VARIABLES

BY

ALESSIO SANCETTA* (LONDON)

Abstract. Maximal inequalities for U-processes are required in order to achieve a reduction to the first nonvanishing term in their Hoeffding's decomposition, which is the relevant quantity for statistical inference. This paper proves new maximal inequalities under strong mixing for U-processes in some function spaces. As an application we derive a uniform central limit theorem.

2000 AMS Mathematics Subject Classification: Primary: 62E20; Secondary: 60F17.

Key words and phrases: Besov space, Hoeffding decomposition, stochastic equicontinuity, strong mixing, U-process.

1. INTRODUCTION

This paper establishes maximal inequalities for U-processes of arbitrary finite order. A U-process is a U-statistic whose U-kernel belongs to some class of functions. The simplest example is an empirical process, which corresponds to a first order U-process. Many statistical estimators can be written as U-statistics (e.g. quadratic forms) and the extension to a U-process is often considered, especially for nonparametric estimation (e.g. Han [13], Honore and Powell [14], Cavanagh and Sherman [7], Ghosal et al. [12]). Unfortunately, for technical reasons, independent observations are usually assumed, some exceptions being Fan and Li [10], Fan and Ullah [11] and Denker and Keller [8].

We derive uniform bounds for U-processes when the underlying observations are strongly mixing. Because of dependence, well-known results in the literature for U-processes (i.e. Arcones and Giné [3]) do not apply. Some maximal inequalities for U-processes under β mixing have been established by Arcones and Yu [4] using Berbee's coupling method for β mixing sequences, but this approach requires some lengthy technical details and is not applicable in the strong mixing case (see

* I thank a referee for comments that improved the content of the paper and, in particular, the proof of one of the lemmata. I also thank Oliver Linton for suggesting some references.

also Borovkova et al. [6]). Recall that strong mixing is a weaker condition than β mixing (see Rio [16] for details). The goal of the paper is to establish some familiar results of U-processes in the unfamiliar context of strongly mixing random variables. As in Rio [16] we use a representation of functions in some space by means of wavelets (see also Birgé and Massart [5]).

The result of this paper only applies to U-processes indexed by classes of functions in some Besov space, hence in this respect it is less general than the result derived in Arcones and Yu [4]. The main motivation of the paper is the reduction to the first nonvanishing term in Hoeffding's decomposition of a U-process. The first nonvanishing term in a U-process is the one that determines the asymptotic distribution of the process. Hence, for statistical inference it is required that we find such a reduction. To this end, maximal inequalities for U-processes are the necessary technical tool. An example of such application will be given. Since in practice observations might not be independent, the extension to dependent random variables should be pursued. The results are stated in such a way that we can easily bound the reminder terms in a U-process and obtain explicit rates of convergence.

The proof of the result makes use of wavelets representation of functions in some Besov spaces and the idea of Arcones and Giné [3] to rewrite U-statistics in terms of powers of partial sums. Moment inequalities for powers of strongly mixing partial sums can then be applied (see Rio [16]).

The plan for the paper is as follows. Section 2 provides some background definitions and states the result of the paper. Section 3 contains further notation and proves the result.

2. MAXIMAL INEQUALITIES FOR U-PROCESSES UNDER STRONG MIXING

We will use $\|\ldots\|_{p,\mathbb{P}}$ and $\|\ldots\|_{p,\lambda}$ to denote, respectively, the L_p norm with respect to the underlying probability measure $\mathbb{P}$ and the Lebesgue measure λ, while $|\ldots|$ is the Euclidean norm. The symbols $\lesssim$ and $\asymp$ mean inequality and equality up to some finite absolute constant of proportionality. We now turn to the definition of U-processes.

2.1. Definition and notation for U-processes. Consider a stationary sequence of random variables $(X_i)_{i\in\mathbb{Z}}$ with values in $\mathbb{R}$. Let δ_x be the point measure at x, i.e. $\delta_x(A) = 1$ if $x \in A \subset \mathbb{R}$ and $\delta_x(A) = 0$ otherwise. Suppose $f : \mathbb{R}^m \to \mathbb{R}$ is a symmetric function of its arguments. Let $\mathbf{P} := law(X_i)$ for all i. Borrowing the notation from Arcones and Giné [3] and Arcones and Yu [4], define $\pi^{\mathbf{P}}_{k,m}$ as an operator on f such that

$$\pi^{\mathbf{P}}_{k,m} f(x_1, \ldots, x_k) = (\delta_{x_1} - \mathbf{P}) \ldots (\delta_{x_k} - \mathbf{P})\, \mathbf{P}^{m-k} f(X_1, \ldots, X_m), \tag{2.1}$$

where $\mathbf{Q}_1 \ldots \mathbf{Q}_m f = \int \ldots \int f(x_1, \ldots, x_m)\, d\mathbf{Q}_1(x_1) \ldots d\mathbf{Q}_m(x_m)$ for any marginal measure $\mathbf{Q}_k$ ($= \delta_{x_k}, \mathbf{P}$, above) and $\mathbf{P}^{m-k} = \mathbf{P} \ldots \mathbf{P}$ is the $m-k$ product of the marginals. Given a function $g : \mathbb{R}^m \to \mathbb{R}$, we call g a $\mathbf{P}$-*canonical function* if it is symmetric and $\mathbb{E}g\,(x_1, \ldots, x_{m-1}, X_m) = 0$ for any $x_1, \ldots, x_{m-1}$. Then $\pi^{\mathbf{P}}_{k,m} f$ is a $\mathbf{P}$-canonical function in k variables $(k = 1, \ldots, m)$. If f is not symmetric, we may write

$$\mathcal{S}f = \frac{1}{m!} \sum_{1 \leqslant i_1, \ldots, i_m \leqslant n} f\,(x_{i_1}, \ldots, x_{i_m})$$

for its symmetric version. To ease notation assume f is symmetric.

Let $\mathfrak{F}$ be a class of symmetric measurable real functions on $\mathbb{R}^m$. A U-process of order m with U-kernel in $\mathfrak{F}$ is defined as

$$\big(\mathcal{U}_n^{(m)}\,(f)\big)_{f \in \mathfrak{F}} := \left(\frac{(n-m)!}{n!} \sum_{(i_1, \ldots, i_m) \in I_m^n} f\,(X_{i_1}, \ldots, X_{i_m}), f \in \mathfrak{F} \right),$$

where $I_m^n := \{(i_1, \ldots, i_m) : 1 \leqslant i_j \leqslant n, i_j \neq i_k \text{ if } j \neq k\}$ (see Serfling [17] and Arcones and Giné [3] for details on U-statistics and U-processes, respectively). Hence, $\big(\mathcal{U}_n^{(m)}\,(f)\big)_{f \in \mathfrak{F}}$ is a collection of U-statistics. Then $\big(\mathcal{U}_n^{(m)}\,(f)\big)_{f \in \mathfrak{F}}$ has the following Hoeffding's decomposition:

$$\mathcal{U}_n^{(m)}\,(f) = \sum_{k=0}^{m} \binom{m}{k} \mathcal{U}_n^{(k)}(\pi^{\mathbf{P}}_{k,m} f) = \mathbf{P}^m f + \sum_{k=1}^{m} \binom{m}{k} \mathcal{U}_n^{(k)}(\pi^{\mathbf{P}}_{k,m} f), \tag{2.2}$$

where $\pi^{\mathbf{P}}_{k,m} f$ $(k = 1, \ldots, m)$ are $\mathbf{P}$-canonical functions.

2.2. U-kernels in Besov spaces. The U-kernel $f \in \mathfrak{F}$ of the process will be restricted to the Besov space $B_p^{s,\infty}\,(\mathbb{R}^m)$, a smoothness subspace of $L_p\,(\mathbb{R}^m)$. To define this space, let us define the r^{th} difference in the direction of $h \in \mathbb{R}^m$:

$$\triangle_h^r\,(f, x) := \sum_{j=0}^{r} (-1)^{r+j} \binom{r}{j} f\,(x + jh),$$

so that $\triangle_h^1\,(f, x) = f\,(x + h) - f\,(x)$ and higher differences are obtained by induction. The modulus of smoothness of order r of f is given by

$$\omega_r\,(f, t)_p := \sup_{|h| \leqslant t} \big(\int_{\mathbb{R}^m} |\triangle_h^r\,(f, x)|^p\, dx \big)^{1/p}, \quad t > 0.$$

Let $s > 0$ and $r = \lfloor s \rfloor + 1$, where $\lfloor s \rfloor$ is the integer part of s. The Besov space $B_p^{s,\infty}\,(\mathbb{R}^m)$ is defined as the set of all functions in $L_p\,(\mathbb{R}^m)$ such that

$$\big(\int_{\mathbb{R}^m} |\triangle_h^r\,(f, x)|^p\, dx \big)^{1/p} \leqslant M\,|h|^s \tag{2.3}$$

for all $h \in \mathbb{R}^m$ and some finite M. This space is equipped with the seminorm $|f|_{B_p^{s,\infty}} := \sup_{t>0} t^{-s} \omega_r\,(f, t)_p$ and the norm $\|f\|_{B_p^{s,\infty}} := |f|_{B_p^{s,\infty}} + \|f\|_{p,\lambda}$. Note that $|\ldots|_{B_p^{s,\infty}}$ is a seminorm because if f is a polynomial of degree less than r, then $\triangle_h^r\,(f, x) = 0$, implying $|f|_{B_p^{s,\infty}} = 0$. A discussion of Besov spaces and their relation to Sobolev spaces can be found in Adams and Fournier [1].

2.3. Dependence condition. We introduce notation for the weak dependence condition satisfied by the stationary sequence $(X_i)_{i\leqslant k\in\mathbb{Z}}$. Let $\mathcal{F}_k := \sigma(X_i, i \leqslant k)$ and $\mathcal{F}^k := \sigma(X_i, i \geqslant k)$ be the sub-σ-algebras generated by $(X_i)_{i\leqslant k}$ and $(X_i)_{i\geqslant k}$, respectively. We say that $(X_i)_{i\in\mathbb{Z}}$ is α *mixing* if $\lim_n \alpha(\mathcal{F}_0, \mathcal{F}^n) = 0$, where

$$\alpha(\mathcal{F}_0, \mathcal{F}^n) := \sup_{A,B} |\Pr(A\cap B) - \Pr(A)\Pr(B)| = \sup_{A,B} |\mathrm{Cov}(I_A, I_B)|, \tag{2.4}$$

and $A \in \mathcal{F}_0$, $B \in \mathcal{F}^n$ (I_A and I_B are indicator functions of A and B, respectively). We call $\alpha_n := \alpha(\mathcal{F}_0, \mathcal{F}^n)$ the *strong mixing coefficient* of $(X_i)_{i\in\mathbb{Z}}$.

2.4. Statement of the result. We have the following equicontinuity inequality for U-processes.

THEOREM 2.1. *Suppose that $(X_i)_{i\in\mathbb{Z}}$ has strong mixing coefficients satisfying $\sup_{j>0} j^m \alpha_j < \infty$. Let $\mathfrak{F}$ be a class of symmetric functions such that*

$$\mathfrak{F} \subset B_p^{s,\infty}(\mathbb{R}^m) \cap L_2(\mathbb{R}^m),$$

where $s \in (m/p, \infty)$ and $p \in [1,2]$. Then, for all $J \in \mathbb{N}$ and $\gamma > 0$,

$$\| \sup_{\substack{f,g\in\mathfrak{F}\\ \|f-g\|_{2,\lambda}\leqslant\gamma}} |\mathcal{U}_n^{(m)}(f) - \mathcal{U}_n^{(m)}(g)| \|_{2,\mathbb{P}} \lesssim n^{-1/2}(\sup_{f\in\mathfrak{F}} \|f\|_{B_p^{s,\infty}} 2^{J(m/p-s)} + \gamma 2^{Jm/2})$$

and

$$\| \sup_{\substack{f,g\in\mathfrak{F}\\ \|f-g\|_{2,\lambda}\leqslant\gamma}} |\mathcal{U}_n^{(k)}(\pi_{k,m}^{\mathbf{P}} f) - \mathcal{U}_n^{(k)}(\pi_{k,m}^{\mathbf{P}} g)| \|_{2,\mathbb{P}} \lesssim n^{-k/2}(\sup_{f\in\mathfrak{F}} \|f\|_{B_p^{s,\infty}} 2^{J(m/p-s)} + \gamma 2^{Jm/2}).$$

REMARK 2.1. Clearly, Theorem 2.1 implies

$$\| \sup_{f\in\mathfrak{F}} |\mathcal{U}_n^{(m)}(f)| \|_{2,\mathbb{P}} \lesssim n^{-1/2}(\sup_{f\in\mathfrak{F}} \|f\|_{B_p^{s,\infty}} 2^{J(m/p-s)} + \|f\|_{2,\lambda} 2^{Jm/2}).$$

REMARK 2.2. Note that $B_p^{s,\infty}$ can be embedded into $B_{p'}^{s',\infty}$ as long as $p < p'$ and $s - m/p = s' - m/p' > 0$ (see, e.g., Theorem 2.7.1 in Triebel [18]). Given that the statement of Theorem 2.1 depends on $s - m/p > 0$ only, we could choose $p = 2$ with no loss of generality. To simplify reference to some results to be used, we do not make use of this embedding.

2.5. Application: Donsker theorem for U-processes. As an application of Theorem 2.1 consider a U-process with non-degenerate first term in its Hoeffding's decomposition. Then

$$\sqrt{n}\big(\mathcal{U}_n^{(m)}(f) - \mathbf{P}^m f\big)_{f\in\mathfrak{F}} = \left(m\sqrt{n}\,\mathcal{U}_n^{(1)}(\pi_{1,m}^{\mathbf{P}} f) + \sqrt{n}\sum_{k=2}^{m}\binom{m}{k}\mathcal{U}_n^{(k)}(\pi_{k,m}^{\mathbf{P}} f)\right)_{f\in\mathfrak{F}}.$$

Under the conditions of Theorem 2.1, for $\mathfrak{F} \subset B_p^{s,\infty}(\mathbb{R}^m) \cap L_2(\mathbb{R}^m)$ the U-process is stochastically equicontinuous (setting $\gamma \asymp 2^{J(m/p-m/2-s)}$) and has the same limiting distribution as $\big(m\sqrt{n}\,\mathcal{U}_n^{(1)}(\pi_{1,m}^{\mathbf{P}} f)\big)_{f\in\mathfrak{F}}$ because

$$\sqrt{n}\sum_{k=2}^{m}\binom{m}{k}\big\|\sup_{f\in\mathfrak{F}}|\mathcal{U}_n^{(k)}(\pi_{k,m}^{\mathbf{P}} f)|\big\|_{2,\mathbb{P}} \lesssim n^{-1/2}$$

by Theorem 2.1. By Theorem 2.1 we also know that $\big(m\sqrt{n}\mathcal{U}_n^{(1)}(\pi_{1,m}^{\mathbf{P}} f)\big)_{f\in\mathfrak{F}}$ is stochastically equicontinuous. Then, if $\mathfrak{F}$ is totally bounded with respect to (the metric induced by) $\|\ldots\|_{2,\lambda}$, to show that $\sqrt{n}\big(\mathcal{U}_n^{(m)}(f) - \mathbf{P}^m f\big)_{f\in\mathfrak{F}}$ converges to a Gaussian process with $\|\ldots\|_{2,\lambda}$ continuous sample paths, we only need finite-dimensional convergence of $m\sqrt{n}\,\mathcal{U}_n^{(1)}(\pi_{1,m}^{\mathbf{P}} f)$ (see, e.g., Van der Vaart and Wellner [19], Theorems 1.5.4 and 1.5.7). This follows by an application of the central limit theorem for strongly mixing sequences (e.g., Rio [16], Theorem 4.2). Hence we have easily proved the following

COROLLARY 2.1. *Suppose $\mathfrak{F}$ is a symmetric totally bounded (with respect to $\|\ldots\|_{2,\lambda}$) class of functions in $B_p^{s,\infty}(\mathbb{R}^m) \cap L_2(\mathbb{R}^m)$, where $s \in (m/p,\infty)$ and $p \in [1,2]$. Suppose the strong mixing coefficients satisfy $\sup_{j>0} j^m \alpha_j < \infty$. Then $\sqrt{n}\big(\mathcal{U}_n^{(m)}(f) - \mathbf{P}^m f\big)_{f\in\mathfrak{F}}$ converges weakly to a mean zero Gaussian process $\big(\mathbb{G}(f)\big)_{f\in\mathfrak{F}}$ with a.s. continuous sample paths and covariance function*

$$\begin{aligned}\mathbb{E}\mathbb{G}(f)\,\mathbb{G}(g) = m^2 &\mathrm{Cov}\big(\pi_{1,m}^{\mathbf{P}} f(X_1), \pi_{1,m}^{\mathbf{P}} g(X_1)\big)\\ &+ m^2\sum_{i=1}^{\infty}\big[\mathrm{Cov}\big(\pi_{1,m}^{\mathbf{P}} f(X_1), \pi_{1,m}^{\mathbf{P}} g(X_{1+i})\big)\\ &+ \mathrm{Cov}\big(\pi_{1,m}^{\mathbf{P}} f(X_{1+i}), \pi_{1,m}^{\mathbf{P}} g(X_1)\big)\big].\end{aligned}$$

3. PROOF

The proof of Theorem 2.1 relies on multidimensional wavelet representation for functions in $B_p^{s,\infty}$. Details on this can be found in the book of Meyer [15] and the review article of DeVore and Lucier [9]. Since $\|f\|_{2,\lambda} < \infty$, $f: \mathbb{R}^m \to \mathbb{R}$

admits the following multiresolution representation via wavelet expansion:

$$f = \sum_{\theta \in \mathbb{Z}^m} b_\theta^f \varphi_\theta + \sum_{j=0}^{\infty} \sum_{\theta \in \Theta_j} a_\theta^f \psi_\theta, \tag{3.1}$$

where $\Theta_j := 2^{-j-1}\mathbb{Z}^m \backslash 2^{-j}\mathbb{Z}^m$ and $\{\varphi_\theta : \theta \in \mathbb{Z}^m\}$, $\{\psi_\theta : \theta \in \Theta_j, j \in \mathbb{N}\}$ are functions which can be chosen to have a compact support in $\mathbb{R}^m$. In particular, $\{\varphi_\theta : \theta \in \mathbb{Z}^m\}$ is a father wavelet, while $\{\psi_\theta : \theta \in \Theta_j, j \geqslant 0\}$ are mother wavelets (see, e.g., Meyer [15], Chapter 3.1). The multidimensional wavelets can be constructed from wavelets on $\mathbb{R}$ by the tensor product method (Meyer [15], Chapter 3.3). Let φ and ψ be wavelets on $\mathbb{R}$ at a resolution level $j = 0$. To ease notation define $\psi^0 := \varphi$ and $\psi^1 := \psi$. Hence,

$$\begin{aligned} \psi_\theta(x_1, \ldots, x_m) &= \sum_{(\epsilon_1, \ldots, \epsilon_m) \in \{0,1\}^m \backslash \{0\}^m} \prod_{k=1}^m 2^{j/2} \psi^{\epsilon_k}(2^j x_k - q_k) \\ &= \sum_{(\epsilon_1, \ldots, \epsilon_m) \in \{0,1\}^m \backslash \{0\}^m} \prod_{k=1}^m 2^{j/2} \psi_{j q_k}^{\epsilon_k}(x_k) \end{aligned} \tag{3.2}$$

when $\theta = 2^{-j}(q_1 - \epsilon_1, \ldots, q_m - \epsilon_m) \in \Theta_j$, $\epsilon_k \in \{0,1\}$ with $\sum_{k=1}^m \epsilon_k \geqslant 1$, and $q_k \in \mathbb{Z}$. (Recall that $\Theta_j := 2^{-j-1}\mathbb{Z}^m \backslash 2^{-j}\mathbb{Z}^m$, so the point $\theta = 2^{-j}(q_1, \ldots, q_m)$ must be excluded.) On the other hand,

$$\varphi_\theta(x_1, \ldots, x_m) = \prod_{k=1}^m 2^{j/2} \varphi(2^j x_k - q_k) \tag{3.3}$$

when $\theta = 2^{-j}(q_1, \ldots, q_m) \in 2^{-j}\mathbb{Z}^m$. The functions φ and ψ are bounded and have compact support. While in (3.1) the father wavelet is computed at a resolution level $j = 0$, in the proof we will need to consider the father wavelet at the resolution level $J > 0$, where J is as in Theorem 2.1. In fact, we recall the following identity:

$$\sum_{\theta \in \mathbb{Z}^m} b_\theta^f \varphi_\theta + \sum_{j=0}^{J} \sum_{\theta \in \Theta_j} a_\theta^f \psi_\theta = \sum_{\theta \in 2^{-J}\mathbb{Z}^m} b_\theta^f \varphi_\theta. \tag{3.4}$$

When $f \in B_p^{s,\infty}$, the wavelets coefficients can be related to $\|f\|_{B_p^{s,\infty}}$ by appropriate choice of the father and mother wavelets φ and ψ. In this case, the wavelets are chosen to be $r = \lfloor s \rfloor + 1$ regular (an index of smoothness for the wavelets; Meyer [15], Chapter 2.2) so that there exists an integer $M < \infty$ (growing linearly in r) such that the support of φ and ψ is in $[(1-M)/2, (1+M)/2]$, implying that the support of $\varphi(2^j x - q_k)$ and $\psi(2^j x - q_k)$ (in terms of x) is in $2^{-j}[q_k + (1-M)/2, q_k + (1+M)/2]$, $q_k \in \mathbb{Z}$. Then, it follows (see Meyer [15], Chapter 6.10, with the aid of Lemma 8.1 in Rio [16]) that if $f \in B_p^{s,\infty}(\mathbb{R}^m)$ and $p \in [1,2]$, then

$$\Big(\sum_{\theta \in \mathbb{Z}^d} |b_\theta^f|^2\Big)^{1/2} \lesssim \|f\|_{B_p^{s,\infty}}$$

and

$$\Big(\sum_{\theta\in\Theta_j} |a_\theta^f|^2\Big)^{1/2} \lesssim \|f\|_{B_p^{s,\infty}}\, 2^{j(m/p-s-m/2)}. \tag{3.5}$$

The goal is to substitute the kernel f with its wavelet representation essentially given by the sum of (3.2), where (3.2) is the sum of products of univariate functions. Hence, the most right-hand side in (3.2) will be used as kernel in the proof. Then, as in Arcones and Giné [2], we will represent the U-process as the product of powers of partial sums. To control these powers of partial sums, we will then use moment inequalities for powers of strongly mixing partial sums (Rio [16]). The proof of Theorem 2.1 relies on a sequence of lemmata that formalize the mentioned ideas.

LEMMA 3.1. *Let $\mathfrak{F}$ be a class of symmetric functions such that*

$$\mathfrak{F} \subset B_p^{s,\infty}(\mathbb{R}^m) \cap L_2(\mathbb{R}^m),$$

where $s > m/p$ and $p \in [1,2]$. Then, for all $J \in \mathbb{N}$, and $\gamma > 0$,

$$\begin{aligned}
\Big\| \sup_{\substack{f,g\in\mathfrak{F}\\ \|f-g\|_{2,\lambda}\leqslant\gamma}} &|\mathcal{U}_n^{(m)}(f) - \mathcal{U}_n^{(m)}(g)|\Big\|_{2,\mathbb{P}} \\
&\leqslant \sup_{f\in\mathfrak{F}} \|f\|_{B_p^{s,\infty}}\, 2^{J(m/p-s)}\,(2^m-1) \\
&\quad\times \max_{j>J} \max_{\epsilon\in\{0,1\}^m\setminus\{0\}^m} \Big\|\mathcal{U}_n^{(m)}\Big(\prod_{k=1}^m \sum_{q_k\in\mathbb{Z}} e_{q_k}\psi_{jq_k}^{\epsilon_k}\Big)\Big\|_{2,\mathbb{P}} \\
&\quad+ \gamma 2^{Jm/2}\Big\|\mathcal{U}_n^{(m)}\Big(\prod_{k=1}^m \sum_{q_k\in\mathbb{Z}} e_{q_k}\psi_{Jq_k}^{0}\Big)\Big\|_{2,\mathbb{P}},
\end{aligned}$$

where $\psi_{jq_k}^{\epsilon_k}$ is as in (3.2), $\epsilon := (\epsilon_1,\ldots,\epsilon_m)$, and $\{(e_{q_k}^{(k)})_{q_k\in\mathbb{Z}}; k=1,\ldots,m\}$ are i.i.d. sequences of Rademacher random variables (i.e. $e_{q_k}^{(k)} \in \{-1,1\}$ such that $\Pr(e_{q_k}^{(k)}=1)=1/2$) independent of each other and independent of $(X_i)_{i\in\mathbb{Z}}$.

Proof of Lemma 3.1. Using the notation above, define

$$\Pi_J f := \sum_{\theta\in\mathbb{Z}^m} b_\theta^f \varphi_\theta + \sum_{j=0}^J \sum_{\theta\in\Theta_j} a_\theta^f \psi_\theta.$$

Clearly,

$$\begin{aligned}
|\mathcal{U}_n^{(m)}(f) - \mathcal{U}_n^{(m)}(g)| \\
&\leqslant |\mathcal{U}_n^{(m)}(f) - \mathcal{U}_n^{(m)}(\Pi_J f)| + |\mathcal{U}_n^{(m)}(g) - \mathcal{U}_n^{(m)}(\Pi_J g)| \\
&\quad+ |\mathcal{U}_n^{(m)}(\Pi_J f) - \mathcal{U}_n^{(m)}(\Pi_J g)|.
\end{aligned}$$

Then

$$\begin{aligned}&\Big\| \sup_{\substack{f,g\in\mathfrak{F}\\ \|f-g\|_{2,\lambda}\leqslant\gamma}} |\mathcal{U}_n^{(m)}(f)-\mathcal{U}_n^{(m)}(g)|\Big\|_{2,\mathbb{P}}\\ &\leqslant 2\Big\|\sup_{f\in\mathfrak{F}}|\mathcal{U}_n^{(m)}(f)-\mathcal{U}_n^{(m)}(\Pi_J f)|\Big\|_{2,\mathbb{P}}+2\Big\|\sup_{\|f\|_{2,\lambda}\leqslant\gamma}|\mathcal{U}_n^{(m)}(\Pi_J f)|\Big\|_{2,\mathbb{P}}\\ &=\mathrm{I}+\mathrm{II}.\end{aligned}$$

Control over I. By the Cauchy–Schwarz inequality,

$$\begin{aligned}|\mathcal{U}_n^{(m)}(f)-\mathcal{U}_n^{(m)}(\Pi_J f)| &= \Big|\sum_{j>J}\sum_{\theta\in\Theta_j} a_\theta^f\mathcal{U}_n^{(m)}(\psi_\theta)\Big|\\ &\leqslant \sum_{j>J}\Big(\sum_{\theta\in\Theta_j}|a_\theta^f|^2\Big)^{1/2}\Big(\sum_{\theta\in\Theta_j}|\mathcal{U}_n^{(m)}(\psi_\theta)|^2\Big)^{1/2}.\end{aligned}$$

Therefore, from (3.5), we have

$$\begin{aligned}(3.6)\quad &\Big\|\sup_{f\in\mathfrak{F}}|\mathcal{U}_n^{(m)}(f)-\mathcal{U}_n^{(m)}(\Pi_M f)|\Big\|_{2,\mathbb{P}}\\ &\leqslant \sup_{f\in\mathfrak{F}}\|f\|_{B_p^{s,\infty}}\sum_{j>J}2^{j(m/p-s)}\Big(\sum_{\theta\in\Theta_j}2^{-jm}\mathbb{E}|\mathcal{U}_n^{(m)}(\psi_\theta)|^2\Big)^{1/2}.\end{aligned}$$

Let $(e_{q_k})_{q_k\in\mathbb{Z}}$ $(k=1,\dots,m)$ be as in the statement of the lemma. Then, from (3.2), we get

$$\begin{aligned}&\sum_{\theta\in\Theta_j}2^{-jm}\mathbb{E}|\mathcal{U}_n^{(m)}(\psi_\theta)|^2\\ &=\Big(\sum_{(q_1,\dots,q_m)\in\mathbb{Z}^m}\Big)\Big(\sum_{(\epsilon_1,\dots,\epsilon_m)\in\{0,1\}^m\setminus\{0\}^m}\Big)2^{-jm}\mathbb{E}\Big|\mathcal{U}_n^{(m)}\Big(\prod_{k=1}^m 2^{j/2}\psi_{jq_k}^{\epsilon_k}\Big)\Big|^2,\end{aligned}$$

which (by using (3.2), the sum over θ is equal to the sum over $(q_1,\dots,q_m)$) is equal to

$$\sum_{(\epsilon_1,\dots,\epsilon_m)\in\{0,1\}^m\setminus\{0\}^m}\mathbb{E}\Big|\mathcal{U}_n^{(m)}\Big(\prod_{k=1}^m\sum_{q_k\in\mathbb{Z}}e_{q_k}^{(k)}\psi_{jq_k}^{\epsilon_k}\Big)\Big|^2,$$

by Lemma 3.2 (stated at the end of the proof) and noting that the 2^{-jm} simplifies with the $2^{jm/2}$ inside the squared absolute value. Hence, substituting the last display in (3.6), we obtain

$$
\begin{aligned}
\mathrm{I} &\leqslant \sup_{f\in\mathfrak{F}} \|f\|_{B_p^{s,\infty}} \sum_{j>J} 2^{j(m/p-s)} \\
&\quad \times \sum_{(\epsilon_1,\dots,\epsilon_m)\in\{0,1\}^m\setminus\{0\}^m} \Big\|\mathcal{U}_n^{(m)}\Big(\prod_{k=1}^m \sum_{q_k\in\mathbb{Z}} e_{q_k}^{(k)}\psi_{jq_k}^{\epsilon_k}\Big)\Big\|_{2,\mathbb{P}} \\
&\lesssim \sup_{f\in\mathfrak{F}} \|f\|_{B_p^{s,\infty}} 2^{J(m/p-s)}\,(2^m-1) \\
&\quad \times \max_{j>J} \max_{\epsilon\in\{0,1\}^m\setminus\{0\}^m} \Big\|\mathcal{U}_n^{(m)}\Big(\prod_{k=1}^m \sum_{q_k\in\mathbb{Z}} e_{q_k}^{(k)}\psi_{jq_k}^{\epsilon_k}\Big)\Big\|_{2,\mathbb{P}}
\end{aligned}
$$

because $\sum_{j>J} 2^{j(m/p-s)} \asymp 2^{J(m/p-s)}$ when $m/p-s<0$, and there are 2^m-1 elements in the sum over $(\epsilon_1,\dots,\epsilon_m)$.

Control over II. Note that $\Pi_J f$ is the projection of f onto the space spanned by the father wavelet φ_θ at the J^{th} resolution level, i.e.

$$
\Pi_J f = \sum_{\theta\in 2^{-J}\mathbb{Z}^m} b_\theta^f \varphi_\theta,
$$

so that, by the Cauchy–Schwarz inequality,

$$
\begin{aligned}
|\mathcal{U}_n^{(m)}(\Pi_J f)| &= \Big|\sum_{\theta\in 2^{-J}\mathbb{Z}^m} b_\theta^f \mathcal{U}_n^{(m)}(\varphi_\theta)\Big| \\
&\leqslant \Big(\sum_{\theta\in 2^{-J}\mathbb{Z}^m} |b_\theta^f|^2\Big)^{1/2}\Big(\sum_{\theta\in 2^{-J}\mathbb{Z}^m} |\mathcal{U}_n^{(m)}(\varphi_\theta)|^2\Big)^{1/2}.
\end{aligned}
$$

Hence, using the same notation and argument as in the control over I, together with the last display and (3.3), we have

$$
\begin{aligned}
\mathrm{II} &\leqslant \gamma\Big(\sum_{\theta\in 2^{-J}\mathbb{Z}^m} \mathbb{E}|\mathcal{U}_n^{(m)}(\varphi_\theta)|^2\Big)^{1/2} \\
&= \gamma\Big\|\mathcal{U}_n^{(m)}\Big(\prod_{k=1}^m \sum_{q_k\in\mathbb{Z}} e_{q_k}^{(k)} 2^{J/2}\psi_{Jq_k}^0\Big)\Big\|_{2,\mathbb{P}},
\end{aligned}
$$

because $\big(\sum_{\theta\in 2^{-J}\mathbb{Z}^m} |b_\theta^f|^2\big)^{1/2} \leqslant \gamma$ if $\|f\|_{2,\lambda}\leqslant\gamma$, by (3.1) and (3.4) (recall that the wavelets are orthonormal functions with respect to the Lebesgue measure λ). ∎

The following is used in the previous proof.

LEMMA 3.2. *Suppose that* $\{(e_{q_k}^{(k)})_{q_k\in\mathbb{Z}}; k=1,\dots,m\}$ *are i.i.d. sequences of Rademacher random variables* (see Lemma 3.1) *independent of each other and independent of* $(X_i)_{i\in\mathbb{Z}}$. *Then*

$$
\sum_{(q_1,\dots,q_m)\in\mathbb{Z}^m} \mathbb{E}\Big|\mathcal{U}_n^{(m)}\Big(\prod_{k=1}^m \psi_{jq_k}^{\epsilon_k}\Big)\Big|^2 = \mathbb{E}\Big|\mathcal{U}_n^{(m)}\Big(\prod_{k=1}^m \sum_{q_k\in\mathbb{Z}} e_{q_k}^{(k)}\psi_{jq_k}^{\epsilon_k}\Big)\Big|^2.
$$

Proof of Lemma 3.2. By the definition of the U-process we have

$$\mathbb{E}\Big|\frac{n!}{(n-m)!}\mathcal{U}_n^{(m)}\Big(\prod_{k=1}^{m}\sum_{q_k\in\mathbb{Z}} e_{q_k}^{(k)}\psi_{jq_k}^{\epsilon_k}\Big)\Big|^2 = \mathbb{E}\Big|\sum_{(i_1,\dots,i_m)\in I_m^n}\prod_{k=1}^{m}\sum_{q_k\in\mathbb{Z}} e_{q_k}^{(k)}\psi_{jq_k}^{\epsilon_k}(X_{i_k})\Big|^2.$$

Hence,

$$\begin{aligned}
\mathbb{E}\Big|&\sum_{(i_1,\dots,i_m)\in I_m^n}\prod_{k=1}^{m}\sum_{q_k\in\mathbb{Z}} e_{q_k}^{(k)}\psi_{jq_k}^{\epsilon_k}(X_{i_k})\Big|^2\\
&= \mathbb{E}\Big|\sum_{(q_1,\dots,q_m)\in\mathbb{Z}^m} e_{q_1}^{(1)}\dots e_{q_m}^{(m)}\sum_{(i_1,\dots,i_m)\in I_m^n}\prod_{k=1}^{m}\psi_{jq_k}^{\epsilon_k}(X_{i_k})\Big|^2\\
&= \sum_{(q_1,\dots,q_m,q_1',\dots,q_m')\in\mathbb{Z}^{2m}} \mathbb{E}\Big[e_{q_1}^{(1)}e_{q_1'}^{(1)}\dots e_{q_m}^{(m)}e_{q_m'}^{(m)}\\
&\quad\times\Big(\sum_{(i_1,\dots,i_m)\in I_m^n}\prod_{k=1}^{m}\psi_{jq_k'}^{\epsilon_k}(X_{i_k})\sum_{(i_1,\dots,i_m)\in I_m^n}\prod_{k=1}^{m}\psi_{jq_k'}^{\epsilon_k}(X_{i_k})\Big)\Big]\\
&= \sum_{(q_1,\dots,q_m,q_1',\dots,q_m')\in\mathbb{Z}^{2m}} \mathbb{E}e_{q_1}^{(1)}e_{q_1'}^{(1)}\dots e_{q_m}^{(m)}e_{q_m'}^{(m)}\\
&\quad\times\mathbb{E}\Big(\sum_{(i_1,\dots,i_m)\in I_m^n}\prod_{k=1}^{m}\psi_{jq_k'}^{\epsilon_k}(X_{i_k})\sum_{(i_1,\dots,i_m)\in I_m^n}\prod_{k=1}^{m}\psi_{jq_k'}^{\epsilon_k}(X_{i_k})\Big),
\end{aligned}$$

which (by independence of the Rademacher r.v.'s and the X's) is equal to

$$\sum_{(q_1,\dots,q_m)\in\mathbb{Z}^m}\mathbb{E}\Big|\sum_{(i_1,\dots,i_m)\in I_m^n}\prod_{k=1}^{m}\psi_{jq_k'}^{\epsilon_k}(X_{i_k})\Big|^2,$$

because the Rademacher variables are independent of each other and have variance one. The term on the right-hand side of the last equality is, by definition, equal to

$$\sum_{(q_1,\dots,q_m)\in\mathbb{Z}^m}\mathbb{E}\Big|\frac{n!}{(n-m)!}\mathcal{U}_n^{(m)}\Big(\prod_{k=1}^{m}\psi_{jq_k}^{\epsilon_k}\Big)\Big|^2. \quad\blacksquare$$

The U-statistics on the right-hand side of the bound of Lemma 3.1 can be bounded using the fact that the U-kernel is the product of functions on $\mathbb{R}$.

LEMMA 3.3. *Define*

$$\phi_k(x_k) := \sum_{q_k\in\mathbb{Z}} e_{q_k}^{(k)}\psi_{jq_k}^{\epsilon_k}(x_k),\qquad k=1,\dots,m, \tag{3.7}$$

where $\psi_{jq_k}^{\epsilon_k}$ *and* $(e_{q_k}^{(k)})_{q_k\in\mathbb{Z}}$ *are as in Lemma* 3.1. *Then*

$$\left\|\mathcal{U}_n^{(k)}\left(\pi_{k,m}^{\mathbf{P}}\prod_{k=1}^{m}\phi_k\right)\right\|_{2,\mathbb{P}} \asymp (\mathbf{P}\phi_1)^{m-k}\left[\sum_{1\leqslant i_1,\dots,i_{2k}\leqslant n}\mathbb{E}\left|\prod_{s=1}^{2k}\frac{(1-\mathbf{P})\,\phi_{i_s}(X_{i_s})}{n}\right|\right]^{1/2}.$$

P r o o f o f L e m m a 3.3. Let $\kappa_m^{\phi}(x_1,\dots,x_m) := \phi(x_1)\dots\phi(x_m)$ for some bounded function ϕ. Then

$$\begin{aligned}\pi_{k,m}^{\mathbf{P}}\kappa_m^{\phi}(x_1,\dots,x_k) &= (\delta_{x_1}-\mathbf{P})\dots(\delta_{x_k}-\mathbf{P})\,\mathbf{P}^{m-k}\kappa_m^{\phi}(X_1,\dots,X_m)\\ &= (\delta_{x_1}-\mathbf{P})\,\phi(X_1)\dots(\delta_{x_k}-\mathbf{P})\,\phi(X_k)\,\big(\mathbf{P}\phi(X_1)\big)^{m-k}\end{aligned}$$

and

$$\pi_{k,m}^{\mathbf{P}}\kappa_m^{\phi}(X_1,\dots,X_k) = \kappa_k^{(1-\mathbf{P})\phi}(X_1,\dots,X_k)\,\big(\mathbf{P}\phi(X_1)\big)^{m-k}. \tag{3.8}$$

Define

$$\tilde{\mathcal{U}}_n^{(k)} := \binom{n}{k}\mathcal{U}_n^{(k)},$$

noting that $n^{-k}\tilde{\mathcal{U}}_n^{(k)} \asymp \mathcal{U}_n^{(k)}$. Then

$$\begin{aligned}\mathbb{E}|\tilde{\mathcal{U}}_n^{(k)}(\kappa_k^{(1-\mathbf{P})\phi})|^2 &= \mathbb{E}\Big|\sum_{(i_1,\dots,i_k)\in I_k^n}\Big[\prod_{s=1}^{k}(1-\mathbf{P})\,\phi(X_{i_s})\Big]\Big|^2\\ &\leqslant \sum_{(i_1,\dots,i_{2k})\in\{1,\dots,n\}^{2k}}\mathbb{E}\Big|\prod_{s=1}^{2k}(1-\mathbf{P})\,\phi(X_{i_s})\Big|,\end{aligned}$$

where we have expanded the square and taken absolute values so that the terms in the summation are positive allowing us to change the indices of summation from I_{2k}^n (recall we are taking squares) to $\{1,\dots,n\}^{2k}$ ($I_{2k}^n \subset \{1,\dots,n\}^{2k}$). This inequality together with (3.8) gives the result. ∎

The last final step is to bound the k^{th} power of the partial sum of the function in (3.7).

LEMMA 3.4. *Suppose that the strong mixing coefficients of* $(X_i)_{i\in\mathbb{Z}}$ *satisfy* $\sup_{j>0} j^k\alpha_j < \infty$. *Then*

$$\sum_{1\leqslant i_1,\dots,i_{2k}\leqslant n}\mathbb{E}\left|\prod_{s=1}^{2k}\frac{(1-\mathbf{P})\,\phi_{i_s}(X_{i_s})}{n}\right| \lesssim n^{-k},$$

where ϕ_i *is as in Lemma* 3.3.

P r o o f o f L e m m a 3.4. At first we show that

$$\sup_{x\in\mathbb{R}} |\phi_i(x)| = \sup_{x\in\mathbb{R}} \Big| \sum_{q_i\in\mathbb{Z}} e^{(i)}_{q_i} \psi^{\epsilon_i}(2^j x - q_i) \Big| \lesssim 1,$$

recalling the definition of $\phi_i(x)$ in Lemma 3.3 and using the notation in (3.2). In the remarks about wavelets we mentioned that there is a positive integer $M < \infty$ (depending linearly on the index of regularity r) such that

$$\begin{aligned} &\psi^{\epsilon_i}(2^j x - q_i) \lesssim 1 \qquad \text{if } x \in 2^{-j}\left[q_i + (1-M)/2, q_i + (1+M)/2\right], \\ &\psi^{\epsilon_i}(2^j x - q_i) = 0 \qquad \text{otherwise.} \end{aligned}$$

Hence,

$$|\phi_i(x)| = \Big| \sum_{q_i\in\mathbb{Z}} e^{(i)}_{q_i} \psi^{\epsilon_i}(2^j x - q_i) \Big| \leqslant \sum_{q_i\in\mathbb{Z}} |\psi^{\epsilon_i}(2^j x - q_i)| \leqslant M \sup_{x\in\mathbb{R}} |\psi^{\epsilon_i}(x)| \lesssim 1,$$

because the wavelets are bounded and, for arbitrary but fixed x, there are at most M non-zero elements in the summation of the above display. Clearly,

$$\sum_{1\leqslant i_1,\dots,i_{2k}\leqslant n} \mathbb{E}\Big| \prod_{s=1}^{2k} \frac{(1-\mathbf{P})\,\phi_{i_s}(X_{i_s})}{n} \Big| \asymp \sum_{1\leqslant i_1\leqslant\dots\leqslant i_{2k}\leqslant n} \mathbb{E}\Big| \prod_{s=1}^{2k} \frac{(1-\mathbf{P})\,\phi_{i_s}(X_{i_s})}{n} \Big|$$

(e.g. eq. (2.16) in Rio [16]). Then, for bounded ϕ_i, the right-hand side of the above display is of order n^{-k} if $\sup_{j>0} j^k \alpha_j < \infty$ (see Rio [16], eq. (2.16), (2.23) and Lemma 2.2). ∎

Since $\pi^{\mathbf{P}}_{k,m}$ does not affect the wavelets coefficients, Lemmata 3.1, 3.3 and 3.4 imply the lemma from which Theorem 2.1 follows as a corollary.

LEMMA 3.5. *Under the conditions of Theorem* 2.1,

$$\Big\| \sup_{\substack{f,g\in\mathfrak{F} \\ \|f-g\|_{2,\lambda}\leqslant\gamma}} \left|\mathcal{U}^{(k)}_n\left(\pi^{\mathbf{P}}_{k,m}(f-g)\right)\right| \Big\|_{2,\mathbb{P}} \lesssim n^{-k/2} \big(\sup_{f\in\mathfrak{F}} \|f\|_{B^{s,\infty}_p} 2^{J(m/p-s)} + \gamma 2^{Jm/2}\big).$$

We can prove Theorem 2.1.

P r o o f o f T h e o r e m 2.1. From (2.2) we deduce

$$\Big\| \sup_{\substack{f,g\in\mathfrak{F} \\ \|f-g\|_{2,\lambda}\leqslant\gamma}} \left|(1-\mathbf{P}^m)\,\mathcal{U}^{(m)}_n(f-g)\right| \Big\|_{2,\mathbb{P}} \leqslant \sum_{k=1}^{m} \binom{m}{k} \Big\| \sup_{\substack{f,g\in\mathfrak{F} \\ \|f-g\|_{2,\lambda}\leqslant\gamma}} \left|\mathcal{U}^{(k)}_n\left(\pi^{\mathbf{P}}_{k,m}(f-g)\right)\right| \Big\|_{2,\mathbb{P}}.$$

Hence, applying Lemma 3.5 to each term in the summation gives the result. ∎

REFERENCES

[1] R. A. Adams and J. J. F. Fournier, *Sobolev Spaces*, Academic Press, Amsterdam 2003.
[2] M. A. Arcones and E. Giné, *On the bootstrap of U and V statistics*, Ann. Statist. 20 (1992), pp. 655–674.
[3] M. A. Arcones and E. Giné, *Limit theorems for U-processes*, Ann. Probab. 21 (1993), pp. 1494–1542.
[4] M. A. Arcones and B. Yu, *Central limit theorems for empirical and U-processes of stationary mixing sequences*, J. Theoret. Probab. 7 (1994), pp. 47–71.
[5] L. Birgé and P. Massart, *An adaptive compression algorithm in Besov spaces*, Constr. Approx. 16 (2000), pp. 1–36.
[6] S. Borovkova, R. Burton and H. Dehling, *Limit theorems for functionals of mixing processes with applications to U-statistics and dimension estimation*, Trans. Amer. Math. Soc. 353 (2001), pp. 4261–4318.
[7] C. Cavanagh and R. P. Sherman, *Rank estimators for monotonic index models*, J. Econometrics 84 (1998), pp. 351–381.
[8] M. Denker and G. Keller, *Rigorous statistical procedures for data from dynamical systems*, J. Statist. Phys. 44 (1986), pp. 67–93.
[9] R. A. DeVore and B. J. Lucier, *Wavelets*, Acta Numer. 1 (1992), pp. 1–56.
[10] Y. Fan and Q. Li, *Central limit theorem for degenerate U-statistics of absolutely regular processes with applications to model specification testing*, J. Nonparametr. Stat. 10 (1999), pp. 245–271.
[11] Y. Fan and A. Ullah, *On goodness-of-fit tests for weakly dependent processes using kernel methods*, J. Nonparametr. Stat. 11 (1999), pp. 337–360.
[12] S. Ghosal, A. Sen and A. W. Van der Vaart, *Testing monotonicity of regression*, Ann. Statist. 28 (2000), pp. 1054–1082.
[13] A. K. Han, *A non-parametric analysis of transformations*, J. Econometrics 35 (1987), pp. 191–209.
[14] B. E. Honore and J. L. Powell, *Pairwise difference estimators of censored and truncated regression models*, J. Econometrics 64 (1994), pp. 241–278.
[15] Y. Meyer, *Wavelets and Operators*, Cambridge University Press, Cambridge 1992.
[16] E. Rio, *Théorie asymptotique des processus aléatoires faiblement dépendants*, Springer, Paris 2000.
[17] J. R. Serfling, *Approximation Theorems of Mathematical Statistics*, Wiley, New York 1980.
[18] H. Triebel, *Theory of Function Spaces*, Birkhäuser, Basel 1983.
[19] A. Van der Vaart and J. A. Wellner, *Weak Convergence of Empirical Processes*, Springer, New York 2000.

FIRST, BNP Paribas
10 Harewood Avenue, London NW1 6AA
E-mail: asancetta@gmail.com

Received on 5.3.2008;
revised version on 15.9.2008