

**LIMIT THEOREMS FOR FUNCTIONALS OF  
MIXING PROCESSES WITH APPLICATIONS TO  
 $U$ -STATISTICS AND DIMENSION ESTIMATION**

SVETLANA BOROVKOVA, ROBERT BURTON, AND HEROLD DEHLING

ABSTRACT. In this paper we develop a general approach for investigating the asymptotic distribution of functionals  $X_n = f((Z_{n+k})_{k \in \mathbf{Z}})$  of absolutely regular stochastic processes  $(Z_n)_{n \in \mathbf{Z}}$ . Such functionals occur naturally as orbits of chaotic dynamical systems, and thus our results can be used to study probabilistic aspects of dynamical systems. We first prove some moment inequalities that are analogous to those for mixing sequences. With their help, several limit theorems can be proved in a rather straightforward manner. We illustrate this by re-proving a central limit theorem of Ibragimov and Linnik. Then we apply our techniques to  $U$ -statistics

$$U_n(h) = \frac{1}{\binom{n}{2}} \sum_{1 \leq i < j \leq n} h(X_i, X_j)$$

with symmetric kernel  $h : \mathbf{R} \times \mathbf{R} \rightarrow \mathbf{R}$ . We prove a law of large numbers, extending results of Aaronson, Burton, Dehling, Gilat, Hill and Weiss for absolutely regular processes. We also prove a central limit theorem under a different set of conditions than the known results of Denker and Keller. As our main application, we establish an invariance principle for  $U$ -processes  $(U_n(h))_h$ , indexed by some class of functions. We finally apply these results to study the asymptotic distribution of estimators of the fractal dimension of the attractor of a dynamical system.

1. INTRODUCTION

In this section we provide some motivation for the research presented in this paper, and present some key examples. We will mainly show how functionals of mixing processes occur naturally in the study of dynamical systems, and how  $U$ -statistics enter in the context of dimension estimation.

1.1. Examples.

**Definition 1.1.** Let  $(\Omega, \mathcal{F}, P)$  be a probability space and let  $(Z_n)_{n \in \mathbf{Z}}$  be a stationary stochastic process.

(i) We call a sequence  $(X_n)_{n \in \mathbf{Z}}$  a two-sided functional (or simply functional) of  $(Z_n)_{n \in \mathbf{Z}}$  if there is a measurable function  $f$  defined on  $\mathbf{R}^{\mathbf{Z}}$  such that

$$(1.1) \quad X_n = f((Z_{n+k})_{k \in \mathbf{Z}}).$$

Note that  $(X_n)_{n \in \mathbf{Z}}$  is necessarily a stationary stochastic process.

---

Received by the editors October 28, 1999 and, in revised form, December 14, 2000.

1991 *Mathematics Subject Classification.* Primary 60F05, 62M10.

Research supported by the Netherlands Organization for Scientific Research (NWO) grant NLS 61-277, NSF grant DMS 96-26575 and NATO collaborative research grant CRG 930819.

(ii) Similarly, if  $(Z_n)_{n \geq 0}$  is one-sided and if

$$(1.2) \quad X_n = f((Z_{n+k})_{k \geq 0})$$

for  $n \geq 0$  and measurable  $f$ , then  $(X_n)_{n \geq 0}$  is called a one-sided functional of  $(Z_n)_{n \geq 0}$ .

Such functionals arise naturally in statistics and in the theory of stochastic processes. For example, an autoregressive process of order  $p$  is defined as a solution to the stochastic difference equation

$$(1.3) \quad X_n = a_1 X_{n-1} + a_2 X_{n-2} + \dots + a_p X_{n-p} + Z_n$$

where  $a_1, \dots, a_p$  are fixed real numbers, and  $(Z_n)_{n \geq 0}$  is an i.i.d. process. Autoregressive processes play an important role in the analysis of time series. If all roots of the characteristic polynomial  $a(z) = a_p z^p + a_{p-1} z^{p-1} + \dots + a_1 z - 1$  lie outside the unit circle, then (1.3) may be iterated backwards to express

$$(1.4) \quad X_n = Z_n + \sum_{k=1}^{\infty} b_k Z_{n-k}$$

so that  $(X_n)_{n \geq 0}$  is a one-sided functional of the i.i.d. process  $(Z_n)_{n \geq 0}$ .

Functionals also arise naturally in the theory of dynamical systems. Thus given  $(X_n)_{n \in \mathbf{Z}}$  representing observables on a dynamical system we may often be able to reverse the process above and find  $(Z_n)_{n \in \mathbf{Z}}$  defined in terms of a partition so that  $(X_n)_{n \in \mathbf{Z}}$  is a functional of  $(Z_n)_{n \in \mathbf{Z}}$ .

If  $(X_n)_{n \in \mathbf{Z}}$  is a stationary process, then there is always a measure-preserving dynamical system  $(\mathcal{X}, \mathcal{A}, \mu, T)$  so that  $(\mathcal{X}, \mathcal{A}, \mu)$  is a probability space and  $T : \mathcal{X} \rightarrow \mathcal{X}$  is invertible, measurable, and measure-preserving so for all  $A \in \mathcal{A}$  we have  $\mu(T^{-1}A) = \mu(A)$ . Further there is a measurable function  $\phi$  defined on  $\mathcal{X}$  so that  $X_n = \phi \circ T^n$ . Assume now that  $(\Omega, \mathcal{F}, P)$  was sufficiently rich and coincides with  $(\mathcal{X}, \mathcal{A}, \mu)$ . If  $Q = \{Q_\alpha : \alpha \in \Lambda\}$  is a measurable partition of  $\Omega$ , then we get another stationary process  $(Z_n)_{n \geq 0}$  with state space  $\Lambda$  defined by  $Z_n(\omega) = \alpha$  if and only if  $\omega \in Q_\alpha$ . Define  $\mathcal{A}_n^m = \sigma(Z_n, Z_{n+1}, \dots, Z_m)$  for  $m \geq n$  and  $\mathcal{A}_I = \sigma(Z_j, j \in I)$ .  $Q$  is a generator for  $T$  if  $A_{\mathbf{Z}} = \mathcal{F}$  up to sets of measure 0. In this case,  $(X_n)_{n \in \mathbf{Z}}$  is a functional of  $(Z_n)_{n \in \mathbf{Z}}$ , so there is an  $f$  such that  $X_k = f((Z_{n+k})_{k \in \mathbf{Z}})$ . The point for any given  $(\Omega, \mathcal{F}, P, T)$  is to find a process  $(Z_n)_{n \in \mathbf{Z}}$  with as little dependence as possible and the function  $f$  to have continuity properties as strong as possible. In most cases we will construct the partition  $Q$  to be finite or countable. If  $(X_n)_{n \in \mathbf{N}}$  is a one-sided process, then the same formalism holds but in that case the transformation  $T : \mathcal{X} \rightarrow \mathcal{X}$  is not invertible and the functional is one-sided.

**Definition 1.2.** A process  $(Z_n)_{n \in \mathbf{Z}}$  is called absolutely regular (also known as weak Bernoulli) if  $\beta_k \rightarrow 0$  where

$$\begin{aligned} \beta_k &= 2 \sup_n \left\{ \sup_{A \in \mathcal{A}_{n+k}^\infty} (P(A|\mathcal{A}_1^n) - P(A)) \right\} \\ &= \sup_n \left\{ \sup \sum_{i=1}^I \sum_{j=1}^J |P(A_i \cap B_j) - P(A_i)P(B_j)| \right\} \end{aligned}$$

where the last supremum is over all finite  $\mathcal{A}_1^n$ -measurable partitions  $(A_1, \dots, A_I)$  and all finite  $\mathcal{A}_{n+k}^\infty$ -measurable partitions  $(B_1, \dots, B_J)$ .

The Ornstein isomorphism theorem of ergodic theory tells us that every process which is a functional of a countable-state weak Bernoulli process is also a functional of a countable state i.i.d. process although the functional may be highly discontinuous.

Other notions of weak dependence are similarly related to the mixing coefficients

$$\begin{aligned} \alpha_k &= \sup_n \sup_{A \in \mathcal{A}_1^n} \sup_{B \in \mathcal{A}_{n+k}^\infty} |P(A \cap B) - P(A)P(B)|, \\ \phi_k &= \sup_n \sup_{A \in \mathcal{A}_1^n} \sup_{B \in \mathcal{A}_{n+k}^\infty} |P(B|A) - P(B)|, \\ \psi_k &= \sup_n \sup_{A \in \mathcal{A}_1^n} \sup_{B \in \mathcal{A}_{n+k}^\infty} \left| \frac{P(A \cap B)}{P(A)P(B)} - 1 \right|. \end{aligned}$$

A process is called strong mixing (respectively uniformly mixing, and  $\psi$ -mixing) if  $\alpha_k \rightarrow 0$  (respectively  $\phi_k \rightarrow 0, \psi_k \rightarrow 0$ ). Since  $\psi_k \geq \phi_k \geq \beta_k \geq \alpha_k$ , we see that  $\psi$ -mixing implies uniform mixing, which again implies absolute regularity, which finally implies strong mixing.

In the rest of this section, we provide some key examples illustrating the above ideas.

**Example 1.1.** Consider the doubling map  $T : [0, 1) \rightarrow [0, 1)$  defined by  $T(x) = 2x \pmod{1}$ . Given an initial value  $X_0$  with uniform distribution on  $[0, 1)$  we generate a stationary stochastic process  $(X_n)_{n \geq 0}$  by  $X_n = T^n(X_0)$ .  $X_0$  may be expressed in binary expansion

$$X_0 = \sum_{n=0}^{\infty} \frac{Z_n}{2^{n+1}}$$

where  $(Z_n)_{n \geq 0}$  are i.i.d. with  $P(Z_n = 0) = P(Z_n = 1) = 1/2$ . The partition in this case is  $Q = \{Q_0 = [0, 1/2); Q_1 = [1/2, 1)\}$ . Thus  $(X_n)$  is a functional of  $(Z_n)$ . The map  $X_0 = f(Z_0, Z_1, \dots)$  is Lipschitz in the sense that if  $a_0 = a'_0, a_1 = a'_1, \dots, a_{n-1} = a'_{n-1}$ , then  $|f(a_0, a_1, \dots) - f(a'_0, a'_1, \dots)| \leq 2^{-n}$ .

**Definition 1.3.** (i) A two-sided functional  $f : \mathbf{R}^{\mathbf{Z}} \rightarrow \mathbf{R}$  is called Lipschitz if there are constants  $C > 0$  and  $0 < \alpha < 1$  such that for any two sequences  $(a_n)_{n \in \mathbf{Z}}$  and  $(a'_n)_{n \in \mathbf{Z}}$  satisfying  $a_n = a'_n$  for  $-l \leq n \leq l$  we have

$$|f((a_n)) - f((a'_n))| \leq C\alpha^l.$$

(ii) Similarly, a one-sided functional  $f : \mathbf{R}^{\mathbf{N}} \rightarrow \mathbf{R}$  is called Lipschitz if  $a_n = a'_n$  for  $0 \leq n \leq l$  implies

$$|f((a_n)) - f((a'_n))| \leq C\alpha^l,$$

again for some constants  $C > 0$  and  $0 < \alpha < 1$ .

If  $f$  is a Lipschitz functional and  $X_n = f((Z_n)_{n \in \mathbf{Z}})$ , then  $(X_n)_{n \in \mathbf{Z}}$  is called a Lipschitz functional of  $(Z_n)_{n \in \mathbf{Z}}$  (similarly in the one-sided case). In many examples, the Lipschitz property is too strong to hold. We thus introduce in the next definition a weaker form of continuity, the so-called  $r$ -approximation condition.

**Definition 1.4.** Let  $r \geq 1$  and define the  $\sigma$ -fields  $\mathcal{A}_{-l}^l = \sigma(Z_{-l}, \dots, Z_l)$ . Suppose  $(a_l)_{l \geq 0}$  are constants with  $a_l \rightarrow 0$ . We say that  $(X_n)_{n \in \mathbf{Z}}$  satisfies the  $r$ -approximation condition if

$$(1.5) \quad E|X_0 - E(X_0 | \mathcal{A}_{-l}^l)|^r \leq a_l.$$

Alternatively, we say that  $(X_n)_{n \in \mathbf{Z}}$  is an  $r$ -approximating functional with constants  $(a_l)_{l \geq 0}$  of  $(Z_n)_{n \in \mathbf{Z}}$ . We make a similar definition in the one-sided case.

**Example 1.2.** Let  $(Z_n)_{n \geq 0}$  be i.i.d. with  $P(Z_n = 0) = P(Z_n = 1) = P(Z_n = 2) = 1/3$ , and define

$$X_0 = f((Z_n)_{n \geq 0}) = \sum_{l=0}^{\infty} \frac{a_{i_l}}{2^{l+1}}$$

where  $i_1 = \inf\{i \geq 0 | Z_i \neq 2\}$  and  $i_{l+1} = \inf\{i > i_l | Z_i \neq 2\}$ , and  $X_k = f((Z_{n+k})_{n \geq 0})$ . Then  $(X_n)_{n \geq 0}$  is a one-sided functional of an i.i.d. process but the functional is not Lipschitz as in Example 1.1. We do have for  $r \geq 1$ , and some  $C > 0, 0 < \alpha < 1$ ,

$$E(|X_0 - E(X_0 | \mathcal{A}_0^n)|^r) \leq C\alpha^n.$$

**Example 1.3** (Regular Continued Fractions). Let  $\mathcal{X} = [0, 1)$  and define the map  $T : \mathcal{X} \rightarrow \mathcal{X}$  by  $T(x) = \frac{1}{x} - \lfloor \frac{1}{x} \rfloor$ . Gauss measure on  $\mathcal{X}$ , given by the density  $f(x) = \frac{1}{\log(2)} \frac{1}{1+x}$  is an invariant distribution for  $T$ . Thus if  $X_0$  is distributed according to the Gauss measure, and if  $X_n = T^n(X_0)$ , then  $(X_n)_{n \geq 0}$  is a stationary stochastic process. Let  $Z_0 = \lfloor \frac{1}{X_0} \rfloor$  and  $Z_n = \lfloor \frac{1}{X_n} \rfloor$ , so that

$$X_0 = f((Z_n)_{n \geq 0}) = \frac{1}{Z_0 + \frac{1}{Z_1 + \frac{1}{Z_2 + \dots}}}$$

The partition in this case is  $Q = \{Q_i : i \geq 1\}$  where  $Q_i = [\frac{1}{i+1}, \frac{1}{i})$ . The functional is Lipschitz as in Example 1.1, but the process is not i.i.d. It is however  $\psi$ -mixing, and thus also absolutely regular.

**Example 1.4** (Piecewise expanding maps of  $[0, 1]$ ). Suppose  $T : [0, 1] \rightarrow [0, 1]$  is piecewise smooth and expanding. This means there are closed intervals  $J_1, \dots, J_N$  such that  $[0, 1] = \bigcup_{j=1}^N J_j$  and  $J_i \cap J_j$  is at most one point for  $i \neq j$ . Further  $T$  restricted to  $J_j$  is monotone,  $C^2$  and satisfies  $|T'(x)| \geq \lambda > 1$  for all  $x$  in the interior of  $J_j$ , for some uniform constant  $\lambda$ . Then  $T$  preserves an absolutely continuous probability measure. If  $(Z_n)_{n \in \mathbf{Z}}$  is the process associated with the above partition  $J_1, \dots, J_N$ , and if  $T$  is totally ergodic, then  $(Z_n)_{n \in \mathbf{Z}}$  is absolutely regular (Hofbauer and Keller [31]).

**Example 1.5** (Automorphisms of the 2-torus). Let  $\Pi^2 = S^1 \times S^1, A = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$ , and define  $T : \Pi^2 \rightarrow \Pi^2$  by

$$T \begin{pmatrix} x \\ y \end{pmatrix} = A \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x + y \\ y \end{pmatrix} \pmod{1}.$$

This defines an automorphism of the torus and preserves the Haar measure, which in this case turns out to be the Lebesgue measure. If  $X_0$  has Lebesgue measure as distribution, then  $X_n = T^n(X_0)$  defines a stationary process. The eigenvalues of  $A$  are  $\frac{1 \pm \sqrt{5}}{2}$ . There is a partition  $Q = \{Q_0, Q_1\}$  where  $Q_0, Q_1$  are rectangles with sides along the eigenvectors. If  $Z_n = \alpha \leftrightarrow X_n \in Q_\alpha$ , then  $Z_n$  is a mixing Markov chain, and hence absolutely regular. Further  $(X_n)_{n \geq 0}$  is a Lipschitz functional of  $(Z_n)_{n \geq 0}$ . This construction applies to any automorphism of  $\Pi^2$  with an eigenvalue which is not a root of unity (Adler and Weiss [2]).

**Example 1.6** (Natural extension of the regular continued fraction). On  $\mathcal{X} = [0, 1) \times [0, 1)$  we define the two-dimensional Gauss measure  $\frac{1}{\log 2} \frac{1}{(1+xy)^2} dx dy$ . Define  $T : \mathcal{X} \rightarrow \mathcal{X}$  by

$$T(x, y) = \left( \frac{1}{x} - Z_0, \frac{1}{Z_0 + y} \right)$$

where  $Z_0(x, y) = \lfloor \frac{1}{x} \rfloor$ . If  $X_0$  has Gauss measure as distribution, and if  $X_n = T^n(X_0)$  for  $n \in \mathbf{Z}$  and  $Z_n = \lfloor \frac{1}{X_n} \rfloor$ , then  $(X_n)_{n \in \mathbf{Z}}$  is a functional of  $(Z_n)_{n \in \mathbf{Z}}$ . This is a two-sided version of Example 1.3 (Nakada, Ito and Tanaka [35]).

**Example 1.7** (Nearest integer continued fraction). In computing continued fractions we can round down as in the regular continued fraction or round up. The nearest integer continued fraction rounds whichever way is closest. In this case  $T$  is naturally defined on  $[-\frac{1}{2}, \frac{1}{2})$ . Let

$$T(x) = \frac{1}{|x|} - \left\lfloor \frac{1}{|x|} + \frac{1}{2} \right\rfloor$$

so that  $x = \frac{\epsilon_0}{Z_0 + T(x)}$  where  $\epsilon_0 = \text{sgn}(x)$  and  $Z_0 = \lfloor \frac{1}{|x|} + \frac{1}{2} \rfloor$ .

Again we have a natural extension. Let  $\phi$  be the positive root of  $\phi^2 + \phi - 1 = 0$ , and let  $\mathcal{X} = [-\frac{1}{2}, 0) \times [0, 1 - \phi) \cup [0, \frac{1}{2}) \times [0, \phi)$ . Define  $\tilde{T} : \mathcal{X} \rightarrow \mathcal{X}$  by

$$\tilde{T}(x, y) = \left( T(x), \frac{1}{Z_0 + \epsilon_0 y} \right).$$

This preserves the measure  $\frac{C}{(1+xy)^2} dx dy$ . The natural partition is  $Q = \{Q_i : i = \pm 2, \pm 3, \dots\}$ ,  $Q_i = \{(x, y) : i = \epsilon_0 Z_0\}$ . The process  $(\epsilon_n Z_n)_{n \in \mathbf{Z}}$  is absolutely regular and  $(T^n(x, y))_{n \in \mathbf{Z}}$  is a Lipschitz functional of  $(\epsilon_n Z_n)_{n \geq 0}$  (Nakada [34]).

**Example 1.8** (Natural extension of Rosen continued fractions). We generalize Example 1.7. Fix an integer  $q \geq 3$  and let  $\lambda_q = 2 \cos(\frac{\pi}{q})$ , and define the map

$$T(x) = \frac{1}{|x|} - \left\lfloor \frac{1}{|x|\lambda} + \frac{1}{2} \right\rfloor \lambda.$$

This amounts to approximating  $\frac{1}{|x|}$  by the nearest multiple of  $\lambda$ . (If  $q = 3$  we get  $\lambda_q = 1$ , and we are back in the nearest integer case.) Then we can express  $x$  as  $x = \frac{\epsilon_0}{Z_0 \lambda + T(x)}$ , where  $\epsilon_0 = \text{sgn}(x)$  and  $Z_0 = \lfloor \frac{1}{|x|\lambda} + \frac{1}{2} \rfloor$ . We can find a finite union of rectangles  $\mathcal{X} \subset [-1, 1] \times [0, 1]$  with probability measure  $\frac{C}{(1+xy)^2} dx dy$  so that the natural extension  $\tilde{T}(x, y) = (Tx, \frac{1}{Z_0 \lambda + \epsilon_0 y})$  is measure-preserving. Again the process  $(\epsilon_n Z_n)_{n \in \mathbf{Z}}$  is known to be absolutely regular (Burton, Kraaikamp and Schmidt [14]).

**Example 1.9.** Suppose we have a diffeomorphism  $T : M \rightarrow M$  of a compact Riemannian manifold  $M$ . A point  $x \in M$  is non-wandering if there is a neighborhood  $U$  of  $x$  so that  $T^n x \notin U$  for  $n \geq 1$ .  $(M, T)$  is called Axiom A if

- $M$  is the closure of the non-wandering points,
- the periodic points are dense,
- $T$  is hyperbolic, i.e. the tangent space of  $DT$  splits into expanding and contracting subspaces.

In this case  $T$  preserves a natural absolutely continuous measure. A theorem of Bowen [11] states that any sufficiently fine finite partition with piecewise smooth

boundaries is a generator and that the resulting sequence  $(Z_n)_{n \in \mathbf{Z}}$  is absolutely regular. One class of examples is the class of automorphisms of the torus.

**Example 1.10.** Fix real numbers  $a, b$  and define  $T : \mathbf{R}^2 \rightarrow \mathbf{R}^2$  by  $T(x, y) = (y, y^2 + b - ax)$ . For some  $a, b$  this map appears to have a fractal attractor. Little is known about this map rigorously. However, extensive simulations have been carried out. Some believe this map to have an ergodic measure and a weak Bernoulli generator.

An independent iterated function system is a stochastic process obtained by composing an i.i.d. sequence of transformations (Barnsley [4]).

**Definition 1.5.** Let  $(\mathcal{X}, d)$  be a bounded metric space and  $(F_n)_{n \geq 0}$  an i.i.d. sequence of transformations  $F_n : \mathcal{X} \rightarrow \mathcal{X}$ . Suppose the diameter of  $\mathcal{X}$  is  $\leq 1$ . Let  $X_0$  be an  $\mathcal{X}$ -valued random variable and define

$$X_n = F_n \circ \dots \circ F_1(X_0).$$

The Markov chain  $(X_n)_{n \geq 0}$  is called an independent iterated function system (abbreviated IIFS).

The following definition appears in Burton and Faris [13].

**Definition 1.6.** An independent iterated function system is called superstable if

1. there is an invariant probability measure  $\mu$  on  $\mathcal{X}$ , i.e. if  $X_0$  has distribution  $\mu$ , then  $X_n$  has the same distribution.
2. there are constants  $K > 0, \gamma > 0$  so that for every pair  $x, y \in \mathcal{X}$  there is a random variable  $W = W_{x,y} < \infty$  a.s. with

$$d(F_n \circ \dots \circ F_1(x), F_n \circ \dots \circ F_1(y)) \leq Ke^{-\gamma(n-W)}$$

**Theorem 1.** A superstable IIFS  $((F_n)_{n \geq 1}, X_0)$  has the property that  $X_n$  converges in distribution to the invariant measure.

*Proof.* Let  $\nu$  denote the distribution of  $X_0$  and let  $Y$  be a random variable with the invariant distribution  $\mu$ . Let  $f$  be bounded and uniformly continuous. Then

$$|Ef(X_n) - Ef(Y)| = |E[f(F_n \circ \dots \circ F_1(X_0)) - f(F_n \circ \dots \circ F_1(Y))]| \rightarrow 0,$$

since  $F_n \circ \dots \circ F_1(X_0) - F_n \circ \dots \circ F_1(Y) \rightarrow 0$ ,  $f$  is uniformly continuous, and using the bounded convergence theorem.  $\square$

Now extend the i.i.d. transformations backwards to get a doubly infinite sequence  $(F_n)_{n \in \mathbf{Z}}$ . Suppose  $(X_n)_{n \in \mathbf{Z}}$  is stationary and satisfies  $X_n = F_n(X_{n-1})$  for  $n \in \mathbf{Z}$ . Assume the superstability condition, then  $(X_n)$  is a functional of  $(F_n)$ , in fact this is a backward one-sided function  $X_k = F((X_{n+k})_{n \leq 0})$  by the superstability condition.

**Theorem 2.** Let  $(F_n)_{n \in \mathbf{Z}}$  and  $(X_n)_{n \in \mathbf{Z}}$  be as above and superstable. Suppose there is a random variable  $W$  which stochastically dominates  $W_{x,y}$ , for all  $x, y$ . Then the  $r$ -approximation property is satisfied with  $a_n = Ke^{-\gamma rn/2} + P(W > n/2)$ .

*Proof.*

$$\begin{aligned} E|X_0 - X'_0|^r &= E|F_0 \circ \dots \circ F_{-n+1}(X_{-n}) - F_0 \circ \dots \circ F_{-n+1}(X'_{-n})|^r \\ &\leq E[\min(K e^{-\gamma(n-W_{x,y})r}, 1)] \\ &= E[\min(K e^{-\gamma(n-W_{x,y})r}, 1) 1_{\{W_{x,y} \leq n/2\}}] + P(W_{x,y} > n/2) \\ &\leq K e^{-\gamma rn/2} + P(W > n/2) \end{aligned}$$

**Example 1.11.** Suppose  $F_n$  is a.s. a strict contraction of  $\mathbf{R}^n$ , i.e. there exists  $\gamma > 0$  so that  $|F_n(x) - F_n(y)| \leq e^{-\gamma}\|x - y\|$ . Then we may take  $W = 0$ , since  $|F_n \circ \dots \circ F_1(x) - F_n \circ \dots \circ F_1(y)| \leq Ce^{-\gamma n}$ . A special case is

$$F_n(x) = \begin{cases} x/3 & \text{with probability } 1/2, \\ 2/3 + x/3 & \text{with probability } 1/2, \end{cases}$$

and the corresponding invariant measure is Lebesgue measure on the standard Cantor set.

**Example 1.12.** Suppose  $F_n$  is Lipschitz a.s. on  $\mathbf{R}^n$ , with Lipschitz constant

$$\|F_n\| = \sup_{x \neq y} \frac{\|F_n(x) - F_n(y)\|}{\|x - y\|}.$$

If  $E\|F_n\| = 2\gamma < 1$ , then  $(F_n)_{n \in \mathbf{Z}}, (X_n)_{n \in \mathbf{Z}}$  is superstable. We get namely

$$\log(\|F_n \circ \dots \circ F_1\|) \leq \frac{1}{n} \sum_{i=1}^n \log \|F_i\| \rightarrow 2\gamma$$

by the strong law of large numbers. Thus we get

$$|F_n \circ \dots \circ F_1(x) - F_n \circ \dots \circ F_1(y)| \leq e^{-\gamma(n-W)}$$

and  $W$  is exponentially dominated, that is  $P(W > n) \leq Ke^{-\beta n}$  for some  $K > 0$  and  $\beta > 0$ . A special case is

$$F_n(x) = \begin{cases} 2x & \text{with probability } 1/3, \\ x/3 & \text{with probability } 1/3, \\ 2/3 + x/3 & \text{with probability } 1/3, \end{cases}$$

and  $2\gamma = 8/9$ .

**Example 1.13** (Kohonen self-organizing neural networks). Let  $N > 1$  be fixed and let

$$\mathcal{X} = \{A \subset [0, 1]^d : \#(A) = N\}.$$

Define on  $\mathcal{X}$  the Hausdorff metric

$$d(A, B) = \min[\max\{|x_i - y_i| : A = (x_1, \dots, x_N), B = (y_1, \dots, y_N)\}].$$

Let  $Z_1, Z_2, \dots$  be i.i.d. uniform on  $[0, 1]$  and  $0 < \alpha < 1$  be fixed. Let  $X_0$  be a random variable with values in  $\mathcal{X}$  and define  $X_{n+1}$  to be the set including  $\alpha Z_{n+1} + (1 - \alpha)x$  where  $x$  is the point in  $X_n$  closest to  $Z_{n+1}$ , not including the point  $x$  but including all other points in  $X_n$ .

This IIFS is used in vector quantization and other applications (see Ritter, Martinez and Schulten [40]). It is superstable and  $W$  may be taken to be exponentially dominated (Burton and Faris [13]).

**1.2.  $U$ -statistics and dimension estimation.** Let  $h : \mathbf{R}^k \rightarrow \mathbf{R}$  be a measurable function, invariant under permutation of its arguments. Then we define the  $U$ -statistic of degree  $m$  corresponding to the kernel  $h$  by

$$U_n = \binom{n}{m}^{-1} \sum_{1 \leq i_1 < \dots < i_m \leq n} h(X_{i_1}, \dots, X_{i_m}),$$

where  $\{X_n\}_{n \in \mathbf{N}}$  is a stationary sequence of  $k$ -dimensional random vectors with common distribution  $F$ . Halmos [27] and Hoeffding [29] introduced  $U$ -statistics as unbiased estimators for functionals of the form

$$\theta(F) = \int_{\mathbf{R}^{k \times m}} h(x_1, \dots, x_m) dF(x_1) \cdots dF(x_m).$$

Already Hoeffding established asymptotic normality of  $U$ -statistics in the case of i.i.d. observations. The study of  $U$ -statistics of dependent observations has attracted attention in the past years. The main motivation for investigations into the asymptotics of dependent  $U$ -statistics has been provided by the problem of dimension estimation.

Estimation of the fractal dimension of a strange attractor from a chaotic time series has attracted considerable attention in the past few years and has become one of the useful tools in the analysis of the underlying dynamics. Though there are several notions of non-integer dimensions, most attention has been devoted to the *correlation dimension*. This is mainly due to the fact that this type of dimension is relatively easy to estimate, and it provides a good measure of the complexity of the dynamics, i.e. the number of active degrees of freedom.

Suppose  $(\mathcal{X}, \mathcal{F}, \mu, T)$  is a dynamical system, where  $\mathcal{X} \subseteq \mathbf{R}^p$  and  $T : \mathcal{X} \rightarrow \mathcal{X}$  is a measurable transformation with invariant probability measure  $\mu$ . We define the *correlation integral*

$$C_\mu(r) = (\mu \times \mu)\{(x, y) : \|x - y\| \leq r\}$$

for  $r \geq 0$ . In many examples it turns out that there exists a constant  $\alpha$  such that

$$(1.6) \quad C_\mu(r) \approx \text{const} \cdot r^\alpha \quad \text{as } r \rightarrow 0.$$

Then the exponent  $\alpha$  is called the *correlation dimension* of  $\mu$ . More formally, one defines the correlation dimension by

$$(1.7) \quad \alpha := \lim_{r \rightarrow 0} \log C(r) / \log r,$$

provided this limit exists. Since all norms on a finite dimensional Euclidean space are equivalent, the correlation dimension is independent of the choice of norm in the definition of  $C(r)$ . For computational convenience, often the maximum norm is taken.

In most practical situations the dynamical system, and thus also the invariant measure  $\mu$ , are unknown and one has to rely on (partial) observations of a finite orbit  $(T^k \omega)_{0 \leq k \leq n}$  of the system. Most models assume that the actual observations are functions of the state. More precisely, one postulates existence of a so-called read-out function  $f : \mathcal{X} \rightarrow \mathbf{R}$  such that  $y_n = f(T^n \omega)$  is observed at time  $n$ . Of course, one cannot hope to get much information about the state  $\omega$  by just observing  $f(\omega)$ . This, however, changes completely if one replaces  $f(\omega)$  by the vector of observations  $f(T^i \omega)$  at  $q$  consecutive time points, i.e.

$$\text{Rec}(\omega) := (f(\omega), f(T\omega), \dots, f(T^{q-1}\omega)).$$

The Takens reconstruction theorem (Takens, [46]) then assures that in generic situations,  $\text{Rec} : \mathcal{X} \rightarrow \mathbf{R}^q$  defines an embedding, provided  $q \geq 2p + 1$ . Consequently we can obtain information about the state space and the dynamics of  $T$  by studying the process of reconstruction vectors

$$X_n = \text{Rec}(T^n \omega), \quad n \geq 0.$$



Among other things, one can show that the correlation dimension of the invariant measure  $\mu$  coincides with that of the marginal distribution  $F$  of  $X_n$ , again provided we are in the generic situation and  $q \geq 2p+1$ . For smaller values of  $q$ , the correlation dimension of  $F$  will equal the embedding dimension.

It is thus of interest to estimate the correlation dimension of the marginal distribution of a stationary stochastic process from a finite sample  $X_1, \dots, X_n$ . A number of procedures for estimating the correlation dimension has been introduced in the literature. Grassberger and Proccacia [26] suggested a procedure of estimating  $\alpha$  which immediately became widely used by mathematicians and applied scientists. Their method is based on the observation that the correlation integral can be written as

$$C(r) = C_F(r) = P(\|X - X'\| \leq r)$$

where  $X$  and  $X'$  are independent copies of  $X_1$ . The natural estimator for the correlation integral from a stationary sequence of reconstruction vectors  $(X_i)_{i=1, \dots, n}$  is then the sample correlation integral

$$(1.8) \quad C_n(r) = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} \mathbf{1}(\|X_i - X_j\| \leq r)$$

(note that  $C_n(r)$  is the proportion of pairs in the sample  $X_1, \dots, X_n$  no more than distance  $r$  apart). The correlation integral  $C_n(r)$  is estimated for a vector of distances  $(r_1, \dots, r_l)$  and the Grassberger-Proccacia estimate for the correlation dimension  $\hat{\alpha}_n^{GP}$  is then obtained by fitting least-squares linear regression of  $\log C_n(r)$  vs.  $\log r$ .

An alternative approach to dimension estimation was proposed by Takens [47]. Assume for a moment that in a neighborhood of  $r = 0$  an exact scaling law holds for the correlation integral, i.e. that

$$(1.9) \quad C(r) = \text{const} \cdot r^\alpha, \quad r \leq r_0,$$

for some  $r_0 > 0$ . Then Takens first considered estimating  $\alpha$  from i.i.d. realizations  $R_i = \|X_i - Y_i\|$  of the distance  $\|X - Y\|$ , where  $X_i$  and  $Y_i$  are independent with distribution  $\mu$ . If (1.9) holds, the conditional distribution of  $U_i := R_i/r_0$  given  $R_i \leq r_0$  is given by

$$P(U_i \leq t | U_i \leq 1) = t^\alpha, \quad \text{for } t \in [0, 1].$$

Then the (conditional) distribution of  $S_i = -\log U_i$  is exponential with parameter  $\alpha$ , i.e.  $S_i$  has density

$$g(s) = \alpha e^{-\alpha s} \cdot \mathbf{1}_{[0, \infty)}(s).$$

Given an i.i.d. sample  $S_1, S_2, \dots, S_N$  of exponentially distributed random variables, the Maximum Likelihood estimator (MLE) of the parameter  $\alpha$  is given by

$$(1.10) \quad \hat{\alpha}_{ML} = \frac{N}{\sum_{i=1}^N S_i}.$$

It turns out that the ML estimator is biased, but that the variant  $\hat{\alpha} = \frac{N-1}{\sum_{i=1}^N S_i}$ , is unbiased, and actually the uniformly minimum variance unbiased (UMVU) estimator. This is a simple consequence of the Lehman-Scheffé lemma, and the fact that  $\sum_{i=1}^N S_i$  is a complete sufficient statistic.

In general, independent realizations of the distances  $\|X - Y\|$  will not be available and thus a modification of the estimator (1.10) becomes necessary. Given a finite

segment  $X_1, \dots, X_n$  of an orbit, we can form  $n(n-1)/2$  pairwise distances  $\|X_i - X_j\|$ . Motivated by the ML estimator (1.10) Takens proposed to use

$$(1.11) \quad \hat{\alpha}_T = - \left( \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} \log \frac{\|X_i - X_j\|}{r_0} \right)^{-1}$$

as estimator for the correlation dimension.

Note that the sample correlation integral  $C_n(r)$  is a  $U$ -statistics of degree 2 with the kernel function  $h(x, y) = \mathbf{1}(\|x - y\| \leq r)$ , and the Takens estimator is the reciprocal of the  $U$ -statistics of degree 2 with the kernel function  $h(x, y) = \log \|x - y\|$ . Since the data from dynamical systems is not i.i.d., this gives the motivation for studying the  $U$ -statistics of dependent observations.

## 2. BASIC DEFINITIONS AND INEQUALITIES

In this section we present the basic techniques of our approach towards limit theorems for functionals of absolutely regular processes. Coupling methods for absolutely regular processes and their functionals play a central role in this. With their help we can show that an  $r$ -approximating functional can be coupled along blocks to a process of independent blocks. Coupling will also be the key technique in establishing new correlation inequalities for functionals.

**2.1. Basic definitions and coupling lemmas.** From now on, we shall assume that the underlying probability space  $(\Omega, \mathcal{F}, P)$  is sufficiently rich to accomodate all new random variables and processes that will be introduced in the course of our investigations. In this way we avoid the need for later extensions of the original probability space. This holds e.g. if, in addition to the original random variables, an independent random variable with a uniform distribution on  $[0, 1]$  exists on  $(\Omega, \mathcal{F}, P)$ . This is the essence of the following lemma of Skorohod [44] (in the formulation of Dudley and Philipp [25]).

**Lemma 2.1.** *Suppose  $(\Omega, \mathcal{F}, P)$  is a probability space,  $S_1, S_2$  are Polish spaces, and  $Q$  is a probability measure on  $S_1 \times S_2$ . Suppose, moreover, that on this probability space there are independent random variables  $X$  and  $U$  where  $U$  is uniformly distributed on  $[0, 1]$ , and  $X : \Omega \rightarrow S_1$  has distribution  $Q_1$ , the first marginal of  $Q$ . Then there exists a function  $g : [0, 1] \times S_1 \rightarrow S_2$  such that  $(X, g(U, X))$  has the distribution  $Q$ .*

First, we shall recall some main coupling results for absolutely regular processes, which will be useful later. The first lemma, due to Berbee [5], states that with high probability any absolutely regular process can be perfectly coupled with a copy whose past is independent of the given process. Here, and in what follows, we denote by a copy of a process another process with the same distribution.

**Lemma 2.2** (Theorem 4.4.7 in Berbee [5]). *Let  $(Z_n)_{n \in \mathbf{Z}}$  be a stationary and absolutely regular stochastic process with mixing coefficients  $(\beta_k)_{k \geq 0}$ . Then there exists a copy  $(Z'_n)_{n \in \mathbf{Z}}$  of  $(Z_n)_{n \in \mathbf{Z}}$  satisfying:*

- (i)  $(Z'_n)_{n < 0}$  is independent of  $(Z_n)_{n \in \mathbf{N}}$ ;
- (ii)  $P(Z'_k = Z_k, Z'_{k+1} = Z_{k+1}, \dots) = 1 - \beta_k$ , for all  $k \geq 0$ .

**Definition 2.3.** Let  $M$  and  $N$  be positive integers,  $(X_n)_{n \in \mathbf{N}}$  a stochastic process, and assume that  $M$  is even. An  $(M, N)$ -blocking of  $(X_n)_{n \in \mathbf{N}}$  is the sequence of

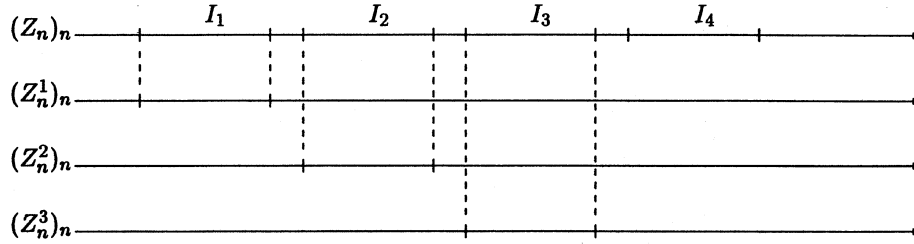


FIGURE 2.1. Coupling as established in Lemma 2.5

blocks  $B_1, B_2, \dots$  of  $N$  consecutive  $X_i$ 's, separated by blocks of length  $M$ . More precisely,

$$B_s = (X_{(s-1)(M+N) + \frac{M}{2} + 1}, \dots, X_{s(M+N) - \frac{M}{2}})$$

for  $s = 1, 2, \dots$ . By  $I_s$  we denote the set of indices in the block  $B_s$ , i.e.  $I_s = \{(s-1)(M+N) + \frac{M}{2} + 1, \dots, s(M+N) - \frac{M}{2}\}$ .

In what follows, the short blocks of length  $M$  will often be discarded. Their only purpose is to serve as separators between the longer blocks to make them almost independent. Our definition of the blocks was chosen symmetrically, i.e. such that the  $s$ th long block of length  $N$  lies in the center of the segment  $\{(s-1)(M+N) + 1, \dots, s(M+N)\}$ . Given a set of indices  $I \subset \mathbf{N}$ , we also define  $X_I := (X_i, i \in I)$ . With this notation we have  $B_s = X_{I_s}$ .

Absolute regularity of a process implies that the sequence of  $(M, N)$ -blocks can be perfectly coupled with the sequence of independent long blocks, which have the same distribution as those of the original process. This result, which is formulated precisely in the following lemma, was obtained independently by several authors. One source is Berbee [5], another is Dehling and Philipp [16], together with an unpublished note by Berger (see Philipp [39]).

**Lemma 2.4** (Theorem 3.4 in Philipp [39]). *Let  $(Z_n)_{n \in \mathbf{N}}$  be a stationary and absolutely regular process with mixing coefficients  $(\beta_k)_{k \geq 0}$ , and with  $(M, N)$ -blocks  $(B_s)_{s \geq 1}$ , where  $M, N$  are positive integers. Then there exists a sequence of independent, identically distributed random vectors  $(B'_s)_{s \geq 1}$ , with the same marginal distributions as  $(B_s)_{s \geq 1}$ , such that*

$$P(B_s = B'_s) = 1 - \beta_M$$

holds for all  $s \in \mathbf{N}$ .

We shall frequently use another variant of the coupling idea, establishing the existence of a sequence of independent copies of the process  $(Z_n)_{n \geq 1}$  that are with high probability perfectly coupled to the original process on precisely one of the long blocks. (See Figure 2.1.)

**Lemma 2.5.** *Let  $(Z_n)_{n \in \mathbf{Z}}$  be stationary and absolutely regular with mixing coefficients  $(\beta_k)$ . Then for all integers  $M, N > 0$  there exists a family of processes  $(Z_n^s)_{n \in \mathbf{Z}}$ ,  $s = 1, 2, \dots$ , such that*

- (i)  $(Z_n^s)_{n \in \mathbf{Z}}$ ,  $s = 1, 2, \dots$ , are independent copies of  $(Z_n)_{n \in \mathbf{Z}}$ ,
- (ii)  $P(Z_n = Z_n^s, \forall n \in I_s) = 1 - \beta_M$  for all  $s = 1, 2, \dots$

*Proof.* By Lemma 2.4 there exist i.i.d. blocks  $(B'_s)_{s \in \mathbf{N}}$  with the same marginal distributions as the  $(M, N)$ -blocks  $(B_s)_{s \in \mathbf{N}}$  of  $(Z_n)_{n \in \mathbf{Z}}$  satisfying  $P(B_s = B'_s) = 1 - \beta_M$ . For a fixed  $s$ , we want to define  $(Z'_n)_{n \in \mathbf{Z}}$  such that it has  $B'_s$  as its  $s$ th  $(M, N)$ -block. As a consequence of Lemma 2.1 this can be done in such a way that (i) is satisfied: let  $(U_s)_{s \in \mathbf{N}}$  be an i.i.d. sequence of random variables uniformly distributed on  $[0, 1]$ . Write  $\mathbf{R}^{\mathbf{Z}} = \mathbf{R}^{I_s} \times \mathbf{R}^{\mathbf{Z} \setminus I_s}$  and let  $Q$  denote the distribution of  $(Z_n)_{n \in \mathbf{Z}}$ . Now the distribution of  $B'_s$  equals the first marginal of  $Q$ , and hence there is a function  $g(U_s, B'_s)$  such that  $(B'_s, g(U_s, B'_s))$  has the same distribution as  $(Z_n)_{n \in \mathbf{Z}}$ . Obviously,  $(Z'_n)_{n \in \mathbf{Z}} := (B'_s, g(U_s, B'_s))$  satisfies the conditions of the lemma. □

**Definition 2.6.** A stochastic process  $(X_n)_{n \in \mathbf{N}}$  is called *nearly regular* if for any  $\epsilon, \delta > 0$  there exists  $M \in \mathbf{N}$  such that for all  $N \in \mathbf{N}$  we can find a sequence  $(B'_s)_{s \geq 1}$  of independent  $\mathbf{R}^N$ -valued random variables satisfying

- (i)  $B'_s$  has the same distribution as  $B_s$ , the  $s$ th  $(M, N)$ -block of  $(X_n)_n$ .
- (ii)  $P(\|B_s - B'_s\| \leq \delta) \geq 1 - \epsilon$ , where  $\|\cdot\|$  denotes the  $L_1$ -norm on  $\mathbf{R}^N$ , that is  $\|x\| = \sum_{i=1}^N |x_i|$ .

The name ‘nearly regular’ refers to the similarity of this property to absolute regularity, in the sense that an absolutely regular process can be perfectly coupled with a process consisting of independent blocks (Lemma 2.4), while near regularity implies closeness to such a process, in an appropriate sense. Such blockwise coupling of weakly dependent processes to an independent process has been introduced in the seminal paper by Berkes and Philipp [8] and has since then been a successful tool in proving various types of limit theorems.

If  $(Z_n)_{n \in \mathbf{Z}}$  is a stationary and absolutely regular process, and  $(X_n)_{n \in \mathbf{Z}}$  is a functional of  $(Z_n)_{n \in \mathbf{Z}}$ , then the process  $(X_n)_{n \in \mathbf{Z}}$  is also stationary and ergodic. However,  $(X_n)$  will in general not be absolutely regular, since each  $X_i$  is a function of the entire process  $(Z_n)_{n \in \mathbf{Z}}$ . On the other hand, if the  $X_i$ ’s can be well approximated by functionals depending on a finite number of  $Z_i$ ’s (and such functionals are absolutely regular), we can expect all asymptotic results, which hold for absolutely regular processes, to hold also for  $(X_n)_{n \in \mathbf{Z}}$ .

In the rest of this section, we want to show that  $r$ -approximating functionals with summable rates are nearly regular. In this and in other applications a different formulation of the  $r$ -approximation condition is required, namely one assuring closeness of  $f((Z_n)_{n \in \mathbf{Z}})$  and  $f((Z'_n)_{n \in \mathbf{Z}})$  in  $L_r$ , if  $(Z_n)_{n \in \mathbf{Z}}$  and  $(Z'_n)_{n \in \mathbf{Z}}$  are two copies of the same process that coincide on a finite segment of indices  $m = -l, \dots, l$ . The next lemma assures that both definitions are essentially the same.

**Lemma 2.7.** (i) Let  $f : \mathbf{R} \rightarrow \mathbf{R}$  be an  $r$ -approximating functional with constants  $(a_l)_{l \geq 0}$ , and let  $(Z'_n)_{n \in \mathbf{Z}}$  be a copy of the process  $(Z_n)_{n \in \mathbf{Z}}$  satisfying  $Z_n = Z'_n$  for  $n = -l, \dots, l$ . If  $X_n = f((Z_{n+k})_{k \in \mathbf{Z}})$  and  $X'_n = f((Z'_{n+k})_{k \in \mathbf{Z}})$  are functionals of  $(Z_n)_{n \in \mathbf{Z}}$  and  $(Z'_n)_{n \in \mathbf{Z}}$  respectively, then

$$E |X_0 - X'_0|^r \leq 2^r a_l,$$

(ii) Let  $f : \mathbf{R} \rightarrow \mathbf{R}$  be a measurable function and suppose that for all copies  $(Z'_n)_{n \in \mathbf{Z}}$  of  $(Z_n)_{n \in \mathbf{Z}}$  satisfying  $Z_n = Z'_n$  for  $n = -l, \dots, l$  we have

$$E |X_0 - X'_0|^r \leq b_l.$$

Then  $f$  is  $r$ -approximating with constants  $(b_l)_l$ .

*Proof.* Define  $u : \mathbf{R}^{2l+1} \rightarrow \mathbf{R}$  by  $u(z_{-l}, \dots, z_l) = E(X_0 | Z_{-l} = z_{-l}, \dots, Z_l = z_l)$ , and note that by (1.5) we get  $E|X_0 - u(Z_{-l}, \dots, Z_l)|^r \leq a_l$ . Thus, using the  $c_r$ -inequalities (see Loéve [33]) we find

$$\begin{aligned} E|X_0 - X'_0|^r &= E|X_0 - u(Z_{-l}, \dots, Z_l) - X'_0 + u(Z'_{-l}, \dots, Z'_l)|^r \\ &\leq 2^{r-1} \{E|X_0 - u(Z_{-l}, \dots, Z_l)|^r + E|X'_0 - u(Z'_{-l}, \dots, Z'_l)|^r\} \\ &\leq 2^r a_l, \end{aligned}$$

where we made use of the fact that  $Z_n = Z'_n$  for  $-l \leq n \leq l$ .

(ii) We define  $V = (Z_k)_{-l \leq k \leq l}$  and  $W = (Z_k)_{k \in \mathbf{Z} \setminus \{-l, \dots, l\}}$ . With these abbreviations, our condition states that whenever  $W' = (Z'_k)_{k \in \mathbf{Z} \setminus \{-l, \dots, l\}}$  is such that  $(V, W')$  has the same distribution as  $(V, W)$ , we have

$$E|f(V, W) - f(V, W')|^r \leq b_l$$

Now let  $U$  be a random variable with uniform distribution on  $[0, 1]$ , and independent of all the previously defined variables. By Lemma 2.1 there exists a function  $g$  such that  $(V, g(V, U))$  has the same distribution as  $(V, W)$  and hence

$$E|f(V, W) - f(V, g(V, U))|^r \leq b_l.$$

Moreover  $E(f(V, g(V, U)) | \mathcal{F}_{-l}^l) = \int_0^1 f(V, g(V, u)) du$  and thus

$$\begin{aligned} E|f(V, W) - E(f(V, W) | \mathcal{F}_{-l}^l)|^r &= E \left| f(V, W) - \int_0^1 f(V, g(V, u)) du \right|^r \\ &\leq E \int_0^1 |f(V, W) - f(V, g(V, u))|^r du \\ &= E|f(V, W) - f(V, g(V, U))|^r \\ &\leq b_l, \end{aligned}$$

where we have made use of the Jensen inequality. □

**Corollary 2.8.** *Let  $f : \mathbf{R} \rightarrow \mathbf{R}$  be an  $r$ -approximating functional with constants  $(a_l)_{l \geq 0}$ , and let  $(Z'_n)_{n \in \mathbf{Z}}$  be a copy of the process  $(Z_n)_{n \in \mathbf{Z}}$  satisfying  $Z_n = Z'_n$  for  $n = -l, \dots, l$ . If  $X_n = f((Z_{n+k})_{k \in \mathbf{Z}})$  and  $X'_n = f((Z'_{n+k})_{k \in \mathbf{Z}})$  are functionals of  $(Z_n)_{n \in \mathbf{Z}}$  and  $(Z'_n)_{n \in \mathbf{Z}}$  respectively, then*

$$(2.1) \quad P(|X_0 - X'_0| \geq (2a_l)^{1/2}) \leq (2a_l)^{1/2}.$$

*Proof.* This is simply an application of Lemma 2.7(i) and the Markov inequality. □

**Theorem 3.** *Let  $(X_n)_{n \in \mathbf{N}}$  be a 1-approximating functional with summable constants  $(a_k)_{k \geq 0}$  of an absolutely regular process with mixing rate  $(\beta_k)_{k \geq 0}$ . Then  $(X_n)_{n \in \mathbf{N}}$  is nearly regular. More precisely, given integers  $K, L, N$ , we can approximate the sequence of  $(K + 2L, N)$ -blocks  $(B_s)_{s \geq 1}$  by a sequence of independent blocks  $(B'_s)_{s \geq 1}$  with the same marginal distribution in such a way that*

$$(2.2) \quad P(\|B_s - B'_s\| \leq 2\alpha_L) \geq 1 - \beta_K - 2\alpha_L,$$

where

$$(2.3) \quad \alpha_L := \left( 2 \sum_{l=L}^{\infty} a_l \right)^{1/2}.$$

(See Figure 2.2.)

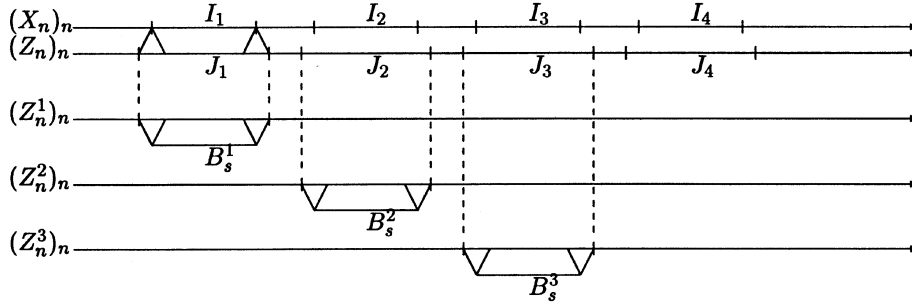


FIGURE 2.2. Coupling ideas in the proof of Theorem 3

*Proof.* Consider the  $(K, 2L + N)$ -blocking of  $(Z_n)_{n \in \mathbf{Z}}$ , and denote the index set of the  $s$ th long block (of length  $2L + N$ ) by  $J_s$ . Let  $(Z_n^s)_{n \in \mathbf{Z}}$ ,  $s = 1, 2, \dots$ , be the independent copies of  $(Z_n)_{n \in \mathbf{Z}}$ , whose existence is given by Lemma 2.5, and satisfying

$$(2.4) \quad \mathbb{P}(Z_{J_s} = Z_{J_s}^s) = 1 - \beta_K.$$

Consider now the  $(K + 2L, N)$ -blocking of  $(X_n)_{n \in \mathbf{N}}$ , and let  $B_s$  denote the  $s$ th block of length  $N$  and  $I_s$  the corresponding index set. Let  $(X_n^s)_{n \in \mathbf{N}}$  be the process obtained by applying the functional  $f$  to  $(Z_n^s)_{n \in \mathbf{Z}}$ , and observe that these are independent copies of  $(X_n)_{n \in \mathbf{N}}$ . Denote by  $B'_s$  the  $s$ th block of the process  $(X_n^s)_{n \in \mathbf{N}}$ , that is

$$B'_s = X_{I_s}^s.$$

One can now directly see that  $B'_s$ ,  $s = 1, 2, \dots$ , are independent, and that  $B'_s$  has the same distribution as  $B_s$ , for any fixed  $s$ . Hence condition (i) of Definition 2.6 is satisfied.

To verify (2.2) we will approximate  $B_s$  (resp.  $B'_s$ ) by a function of  $Z_{J_s}$  (resp.  $Z_{J_s}^s$ ) using the 1-approximation condition, and then apply (2.4). For  $i \in I_s$  we define the function  $u_i : \mathbf{R}^{J_s} \rightarrow \mathbf{R}$  by

$$u_i(z_{J_s}) := \mathbb{E}(X_i | Z_{i-l} = z_{i-l}, \dots, Z_{i+l} = z_{i+l}),$$

where  $l$  is the largest integer such that  $\{i-l, \dots, i+l\} \subset J_s$ . Note that  $l \geq L$ , since  $i \in I_s$  is at least  $L$  away from the boundary of  $J_s$ . Let  $u : \mathbf{R}^{J_s} \rightarrow \mathbf{R}^{I_s}$  be the function with coordinates  $u_i$ . Then, by the 1-approximation condition  $\mathbb{E} |X_i - u_i(Z_{J_s})| \leq a_l$ , and hence

$$\mathbb{E} \|X_{I_s} - u(Z_{J_s})\| \leq 2 \sum_{l=L}^{\infty} a_l.$$

The same inequality holds with  $X$  and  $Z$  replaced by  $X^s$  and  $Z^s$ . By Markov's inequality this implies

$$\mathbb{P}(\|X_{I_s} - u(Z_{J_s})\| \geq \alpha_L) \leq \alpha_L, \quad \mathbb{P}(\|X_{I_s}^s - u(Z_{J_s}^s)\| \geq \alpha_L) \leq \alpha_L,$$

where  $\alpha_L = (2 \sum_{l=L}^{\infty} a_l)^{1/2}$ . Hence

$$\begin{aligned} \mathbb{P}(\|B_s - B'_s\| \geq 2\alpha_L) &\leq \mathbb{P}(\|B_s - u(Z_{J_s})\| \geq \alpha_L) + \mathbb{P}(Z_{J_s} \neq Z_{J_s}^s) \\ &\quad + \mathbb{P}(\|B'_s - u(Z_{J_s}^s)\| \geq \alpha_L) \\ &\leq 2\alpha_L + \beta_K, \end{aligned}$$

thus proving the statement of Theorem 3. □

In what follows, we will often require summability conditions on the sequences  $(a_k)_k$  and  $(\alpha_k)_k$ . We introduce the following notation.

**Definition 2.9.** Let  $(a_k)_{k \geq 0}$  be a sequence of non-negative real numbers. Given  $p, q > 0$ , we say that  $(a_k)_k$  satisfies the  $(p, q)$ -summability condition if

$$(2.5) \quad \sum_{l=0}^{\infty} l^p a_l^q < \infty.$$

**2.2.  $p$ -continuity.** In many situations, we will have to study functions of a given process  $(X_n)_{n \in \mathbf{Z}}$  which itself is again a functional of an absolutely regular process. This occurs e.g. when we investigate the  $U$ -statistic  $\frac{1}{\binom{n}{2}} \sum_{1 \leq i, j \leq n} h(X_i, X_j)$  or the empirical distribution function  $F_n(t) = \frac{1}{n} \sum_{i=1}^n 1_{\{X_i \leq t\}}$ . Under certain continuity conditions on these functions, important properties of  $(X_n)_{n \in \mathbf{Z}}$ , like the  $r$ -approximation property and correlation inequalities are retained.

**Definition 2.10.** Let  $g : \mathbf{R} \rightarrow \mathbf{R}$  be a measurable function, and let  $F$  be some distribution on  $\mathbf{R}$ . We say that  $g$  is  $p$ -continuous with respect to  $F$  if there exists  $\phi : (0, \infty) \rightarrow (0, \infty)$  with  $\phi(\epsilon) = o(1)$  as  $\epsilon \rightarrow 0$  such that

$$(2.6) \quad E(|g(X) - g(X')|^p 1_{\{|X - X'| \leq \epsilon\}}) \leq \phi(\epsilon)$$

holds for all random variables  $X$  and  $X'$  with distribution  $F$ . If the choice of the underlying distribution  $F$  is clearly understood, we simply say that  $g$  is  $p$ -continuous.

**Example 2.1.** (i) Every uniformly continuous function  $u : \mathbf{R} \rightarrow \mathbf{R}$  is  $p$ -continuous. In fact,  $\phi(\epsilon)$  can be chosen to be the modulus of continuity of  $u$ .

(ii) Note that also discontinuous functions can be  $p$ -continuous. Take the indicator function  $u_t(x) = 1_{\{x \leq t\}}$  and a continuous distribution function  $F$ . Then condition (2.6) is automatically satisfied for all  $p$ . Note namely that  $t < x - \epsilon$  and  $|x - x'| \leq \epsilon$  imply that  $t < x'$  and, in the same way,  $t > x + \epsilon$  and  $|x - x'| \leq \epsilon$  imply  $t > x'$ . Hence

$$|1_{\{x \leq t\}} - 1_{\{x' \leq t\}}| 1_{\{|x - x'| \leq \epsilon\}} \leq 1_{\{x - \epsilon \leq t \leq x + \epsilon\}}$$

and thus

$$(2.7) \quad E(|1_{\{X \leq t\}} - 1_{\{X' \leq t\}}|^p 1_{\{|X - X'| \leq \epsilon\}}) \leq F(t + \epsilon) - F(t - \epsilon) = \phi(\epsilon),$$

where  $\phi(\epsilon)$  is some function of  $\epsilon$  which tends to 0 as  $\epsilon \rightarrow 0$  (because of continuity of  $F$ ). If  $F$  has a bounded density  $f$ , then the r.h.s. of (2.7) is bounded by  $\phi(\epsilon) = 2\|f\|_{\infty} \epsilon$ .

The following proposition essentially states that the 1-approximation property is preserved when instantaneous 1-continuous functionals are applied. This result will play a major role in our investigations of the empirical process of functionals of absolutely regular processes.

**Proposition 2.11.** Let  $(X_n)_{n \in \mathbf{Z}}$  be a 1-approximating functional of  $(Z_n)_{n \in \mathbf{Z}}$  with constants  $(a_l)_{l \geq 0}$ , and let  $u : \mathbf{R} \rightarrow \mathbf{R}$  be a function that is 1-continuous with respect to the distribution of  $X_0$ . Then  $(u(X_n))_{n \in \mathbf{Z}}$  is also a 1-approximating functional of  $(Z_n)_{n \in \mathbf{Z}}$  with constants

$$(2.8) \quad a'_l = \phi(\sqrt{2a_l}) + 2\|u(X_0)\|_{2+\delta} (2a_l)^{\frac{1+\delta}{4+2\delta}},$$

provided  $\|u(X_0)\|_{2+\delta} < \infty$ , for some  $\delta > 0$ . If  $u$  is a bounded function, the same holds with

$$(2.9) \quad a'_l = \phi(\sqrt{2a_l}) + 2\|u(X_0)\|_\infty \sqrt{2a_l}$$

replacing (2.8).

*Proof.* Let  $(Z'_n)$  be a copy of  $(Z_n)$ , satisfying  $Z_n = Z'_n$ , for  $-l \leq n \leq l$ , and let  $(X_n)_{n \in \mathbf{Z}}$  and  $(X'_n)_{n \in \mathbf{Z}}$  be the respective functionals. Then the 1-approximation condition and Lemma 2.7(i) imply that

$$\mathbb{E}|X_0 - X'_0| \leq 2a_l.$$

Define the event  $B = \{|X_0 - X'_0| > \sqrt{2a_l}\}$  and note that by Chebychev's inequality we have  $P(B) \leq \sqrt{2a_l}$ . Hence

$$\begin{aligned} \mathbb{E}|u(X_0) - u(X'_0)| &\leq \mathbb{E}(|u(X_0) - u(X'_0)|1_{B^c}) + \mathbb{E}(|u(X_0) - u(X'_0)|1_B) \\ &\leq \phi(\sqrt{2a_l}) + 2\|u(X_0)\|_{2+\delta} (P(B))^{\frac{1+\delta}{2+\delta}} \end{aligned}$$

By Lemma 2.7(ii) this implies the 1-approximation condition for  $(u(X_n))_{n \in \mathbf{N}}$ .  $\square$

In the context of  $U$ -statistics we need a generalization of the  $p$ -continuity condition that is suitable for kernels  $h : \mathbf{R}^2 \rightarrow \mathbf{R}$ . Essentially we require  $h(X, Y)$  and  $h(X', Y)$  to be close whenever  $X$  and  $X'$  are in close  $p$ -mean with respect to a restricted class of probability distributions.

**Definition 2.12.** Let  $(X_n)_{n \in \mathbf{Z}}$  be a stationary stochastic process and let  $h : \mathbf{R}^2 \rightarrow \mathbf{R}$  be a measurable, symmetric kernel. Then we say that  $h$  is  $p$ -continuous if there exists a function  $\phi : (0, \infty) \rightarrow (0, \infty)$  with  $\phi(\epsilon) = o(1)$  as  $\epsilon \rightarrow 0$  such that

$$(2.10) \quad \mathbb{E}|h(X, Y) - h(X', Y)|^p 1_{\{|X - X'| \leq \epsilon\}} \leq \phi(\epsilon),$$

for all random variables  $X, X', Y$  with marginal distribution  $F$  and such that  $(X, Y)$  either has distribution  $F \times F$  or  $P_{X_0, X_k}$  for some  $k \in \mathbf{N}$ .

**Example 2.2.** In the analysis of the Grassberger-Procaccia dimension estimator, the kernel  $h_t(x, y) = 1_{\{|x-y| \leq t\}}$  plays an important role. This kernel is  $p$ -continuous provided

- the distribution functions of  $|X_i - X_j|$  are equicontinuous in  $t$ ,
- $\int \int h_t(x, y) dF(x) dF(y)$  is continuous at  $t$ .

Indeed,  $|X_0 - X_k| \leq t - \epsilon$  and  $|X_0 - X'_0| \leq \epsilon$  imply that  $|X'_0 - X_k| < t$  and, in the same way,  $|X_0 - X_k| > t + \epsilon$  and  $|X_0 - X'_0| \leq \epsilon$  imply  $|X'_0 - X_k| > t$ . Hence

$$\left| 1_{\{|X_0 - X_k| \leq t\}} - 1_{\{|X'_0 - X_k| \leq t\}} \right| 1_{\{|X_0 - X'_0| \leq \epsilon\}} \leq 1_{\{t - \epsilon \leq |X_0 - X_k| \leq t + \epsilon\}}$$

and thus we get

$$\begin{aligned} \mathbb{E} \left| 1_{\{|X_0 - X_k| \leq t\}} - 1_{\{|X'_0 - X_k| \leq t\}} \right|^p 1_{\{|X_0 - X'_0| \leq \epsilon\}} &\leq P(t - \epsilon \leq |X_0 - X_k| \leq t + \epsilon) \\ &\leq \phi(\epsilon) \end{aligned}$$

where  $\phi(\epsilon) = \max(\sup_k P(t - \epsilon \leq |X_0 - X_k| \leq t + \epsilon), P(t - \epsilon \leq |X - Y| \leq t + \epsilon))$ , with  $X, Y$  independent and with distribution function  $F$ .

For (not necessarily symmetric) functions  $g : \mathbf{R}^I \rightarrow \mathbf{R}$  we need an extension of the  $p$ -continuity condition that is contained in the following definition.



**Definition 2.13.** A measurable function  $g : \mathbf{R}^I \rightarrow \mathbf{R}$  is called  $p$ -continuous if there exists a function  $\phi : (0, \infty) \rightarrow (0, \infty)$  with  $\phi(\epsilon) = o(1)$  as  $\epsilon \rightarrow 0$  such that

$$(2.11) \quad E|g(\xi_{I_1}, \xi_{I_2}) - g(\xi_{I_1}, \xi'_{I_2})|^p \mathbf{1}_{\{\|\xi_{I_2} - \xi'_{I_2}\| \leq \epsilon\}} \leq \phi(\epsilon)$$

holds for all disjoint index sets  $I_1$  and  $I_2$  with  $I_1 \cup I_2 = I$  and for all random vectors  $\xi_{I_1}, \xi_{I_2}, \xi'_{I_2}$  such that  $(\xi_{I_1}, \xi_{I_2})$  has distribution  $P_{X_{I_1}, X_{I_2}}$  or  $P_{X_{I_1}} \times P_{X_{I_2}}$  and  $\xi'_{I_2}$  has the same distribution as  $X_{I_2}$ .

**Lemma 2.14.** Let  $h : \mathbf{R}^2 \rightarrow \mathbf{R}$  be a symmetric  $p$ -continuous kernel (in the sense of Definition 2.12) and define

$$g(x_{i_1}, x_{i_2}, x_{i_3}, x_{i_4}) := h(x_{i_1}, x_{i_2})h(x_{i_3}, x_{i_4})$$

(i) If  $h$  is bounded,  $g$  is also  $p$ -continuous.

(ii) If  $E|h(X_{i_1}, X_{i_2})|^p < \infty$ ,  $E|h(X_{i_3}, X_{i_4})|^p < \infty$  and  $E_{X_0} E_{X_k} |h(X_0, X_k)|^p < \infty$ , then  $g$  is  $p/2$ -continuous.

*Proof.* We have to verify (2.11) for any partition of  $I$  into disjoint sets  $I_1, I_2$ . Take e.g.  $I_1 = \{i_1, i_3\}$  and  $I_2 = \{i_2, i_4\}$  (other cases are similar, and left to the reader), and let  $\xi$  have the distributional properties stated in Definition 2.13. Then we can write

$$\begin{aligned} h(\xi_{i_1}, \xi_{i_2})h(\xi_{i_3}, \xi_{i_4}) - h(\xi_{i_1}, \xi'_{i_2})h(\xi_{i_3}, \xi'_{i_4}) &= h(\xi_{i_1}, \xi_{i_2}) (h(\xi_{i_3}, \xi_{i_4}) - h(\xi_{i_3}, \xi'_{i_4})) \\ &\quad + h(\xi_{i_3}, \xi'_{i_4}) (h(\xi_{i_1}, \xi_{i_2}) - h(\xi_{i_1}, \xi'_{i_2})). \end{aligned}$$

From this we get using Hölder's inequality and (2.10)

$$\begin{aligned} &E \left\{ |h(\xi_{i_1}, \xi_{i_2})h(\xi_{i_3}, \xi_{i_4}) - h(\xi_{i_1}, \xi'_{i_2})h(\xi_{i_3}, \xi'_{i_4})|^{p/2} \mathbf{1}_{\{|\xi'_{i_2} - \xi_{i_2}| + |\xi'_{i_4} - \xi_{i_4}| \leq \epsilon\}} \right\} \\ &\leq E \left\{ |h(\xi_{i_1}, \xi_{i_2})|^{p/2} |h(\xi_{i_3}, \xi_{i_4}) - h(\xi_{i_3}, \xi'_{i_4})|^{p/2} \mathbf{1}_{\{|\xi'_{i_4} - \xi_{i_4}| \leq \epsilon\}} \right\} \\ &\quad + E \left\{ |h(\xi_{i_3}, \xi'_{i_4})|^{p/2} \mathbf{1}_{\{|\xi'_{i_4} - \xi_{i_4}| \leq \epsilon\}} |h(\xi_{i_1}, \xi_{i_2}) - h(\xi_{i_1}, \xi'_{i_2})|^{p/2} \mathbf{1}_{\{|\xi'_{i_2} - \xi_{i_2}| \leq \epsilon\}} \right\} \\ &\leq (E|h(\xi_{i_1}, \xi_{i_2})|^p)^{1/2} \left( E|h(\xi_{i_3}, \xi_{i_4}) - h(\xi_{i_3}, \xi'_{i_4})|^p \mathbf{1}_{\{|\xi'_{i_4} - \xi_{i_4}| \leq \epsilon\}} \right)^{1/2} \\ &\quad + \left( E|h(\xi_{i_3}, \xi'_{i_4})|^p \mathbf{1}_{\{|\xi'_{i_4} - \xi_{i_4}| \leq \epsilon\}} \right)^{1/2} \left( E|h(\xi_{i_1}, \xi_{i_2}) - h(\xi_{i_1}, \xi'_{i_2})|^p \mathbf{1}_{\{|\xi'_{i_2} - \xi_{i_2}| \leq \epsilon\}} \right)^{1/2} \\ &\leq M^{1/2} \cdot \phi^{1/2}(\epsilon) + \left( E|h(\xi_{i_3}, \xi'_{i_4})|^p \mathbf{1}_{\{|\xi'_{i_4} - \xi_{i_4}| \leq \epsilon\}} \right)^{1/2} \phi^{1/2}(\epsilon). \end{aligned}$$

Note that both  $(\xi_{i_1}, \xi_{i_2})$  and  $(\xi_{i_3}, \xi_{i_4})$  have one of the distributions for which (2.10) is in force. To bound the remaining term on the r.h.s., we write  $h(\xi_{i_3}, \xi'_{i_4}) = h(\xi_{i_3}, \xi_{i_4}) + (h(\xi_{i_3}, \xi'_{i_4}) - h(\xi_{i_3}, \xi_{i_4}))$  so that

$$\begin{aligned} &E|h(\xi_{i_3}, \xi'_{i_4})|^p \mathbf{1}_{\{|\xi'_{i_4} - \xi_{i_4}| \leq \epsilon\}} \\ &\leq 2^{p-1} (E|h(\xi_{i_3}, \xi_{i_4})|^p + E|h(\xi_{i_3}, \xi'_{i_4}) - h(\xi_{i_3}, \xi_{i_4})|^p \mathbf{1}_{\{|\xi_{i_4} - \xi'_{i_4}| \leq \epsilon\}}) \\ &\leq 2^{p-1} (M + \phi(\epsilon)) \end{aligned}$$

thus proving (ii) of the lemma. Part (i) is simpler, as we can bound both  $h(\xi_{i_1}, \xi_{i_2})$  and  $h(\xi_{i_3}, \xi'_{i_4})$  by the supremum of  $h$ . □

**Lemma 2.15.** Let  $h$  be a  $p$ -continuous kernel and define

$$h_1(x) = \int_{\mathbf{R}} h(x, y) dF(y).$$

Then  $h_1$  is also  $p$ -continuous.

*Proof.* To establish the  $p$ -continuity condition (2.6) for the function  $h_1$ , we take an  $F$ -distributed random variable  $Y$ , independent of  $X, X'$ . Applying Jensen's inequality we then obtain

$$\begin{aligned} E(|h_1(X) - h_1(X')|^p 1_{|X-X'| \leq \epsilon}) &= E(|E_Y(h(X, Y) - h(X', Y))|^p 1_{\{|X-X'| \leq \epsilon\}}) \\ &\leq E_{X, X'} E_Y(|h(X, Y) - h(X', Y)|^p 1_{\{|X-X'| \leq \epsilon\}}) \\ &\leq \phi(\epsilon). \end{aligned}$$

Here we have made use of the fact that  $(X, Y)$  has distribution  $F \times F$  so that (2.10) can be applied. □

**2.3. Correlation inequalities.** Correlation inequalities play a crucial role in all proofs of limit theorems for weakly dependent processes. A well-known example is the inequality

$$(2.12) \quad |E(\xi \cdot \eta) - E\xi \cdot E\eta| \leq 4\alpha_k^{1-1/p-1/q} \|\xi\|_p \|\eta\|_q,$$

valid for strongly mixing processes and random variables  $\xi$  and  $\eta$  that are respectively  $\mathcal{F}_0^n$ - and  $\mathcal{F}_{n+k}^\infty$ -measurable. Here  $\|\xi\|_p$  denotes the  $L_p$ -norm of  $\xi$ , and  $p, q$  have to satisfy  $1/p + 1/q \leq 1$ . This inequality was proved independently by Davydov [15] and Deo [22]. In this section we prove analogous lemmas for functionals of absolutely regular processes.

As absolutely regular processes are also strongly mixing with rates  $\alpha_k \leq \beta_k$ , the correlation inequality (2.12) holds trivially also for absolutely regular processes, with  $\alpha$  replaced by  $\beta$ . It is worth noting that in this case a short proof of (2.12) can be obtained from a coupling argument. We can namely find a copy  $(\xi', \eta')$  of  $(\xi, \eta)$  with the properties

- $\xi'$  is independent of  $\xi, \eta$ ,
- $P(\eta \neq \eta') \leq \beta$ .

Then we get from Hölder's inequality

$$\begin{aligned} |E(\xi\eta) - (E\xi)(E\eta)| &= |E(\xi'\eta') - E(\xi'\eta)| \\ &= |E\{\xi'(\eta' - \eta)1_{\{\eta \neq \eta'\}}\}| \\ &\leq 2\|\xi\|_p \|\eta\|_q (P(\eta \neq \eta'))^{1-1/p-1/q}. \end{aligned}$$

A variant of this technique will be used in all the proofs of correlation inequalities in this section. In the following proposition we construct a suitable coupling.

**Proposition 2.16.** *Let  $(X_n)_{n \in \mathbf{Z}}$  be a 1-approximating functional with constants  $(a_k)$  of an absolutely regular process with mixing coefficients  $(\beta_k)_{k \geq 0}$ . Then, given  $k \in \mathbf{N}$ , there exist copies  $(X'_n)_{n \in \mathbf{Z}}$  and  $(X''_n)_{n \in \mathbf{Z}}$  of  $(X_n)_{n \in \mathbf{Z}}$  with the following properties:*

- (i)  $(X''_n)_{n \in \mathbf{Z}}$  is independent of  $(X_n)_{n \in \mathbf{Z}}$ .
- (ii) There exists a set  $A$  with  $P(A) \geq 1 - \beta_{\lfloor \frac{k}{3} \rfloor}$  such that

$$E(|X_i - X'_i| 1_A) \leq 2a_{i - \lfloor \frac{2k}{3} \rfloor}$$

for all  $i \geq k$ .

- (iii) For all  $i \geq 0$

$$E|X'_{-i} - X''_{-i}| \leq 2a_{i + \lfloor \frac{k}{3} \rfloor}.$$

*Proof.* According to Lemma 2.2, there exists a copy  $(Z'_n)_{n \in \mathbf{Z}}$  of the process  $(Z_n)_{n \in \mathbf{Z}}$ , such that  $(Z'_n)_{n < \lfloor k/3 \rfloor}$  is independent of  $(Z_n)_{n \in \mathbf{Z}}$  and

$$P(Z_i = Z'_i, \text{ for all } i \geq \lfloor \frac{2k}{3} \rfloor) \geq 1 - \beta_{\lfloor k/3 \rfloor}.$$

Let  $(Z''_n)_{n \in \mathbf{Z}}$  be another copy of  $(Z_n)_{n \in \mathbf{Z}}$ , satisfying  $Z''_n = Z'_n$  for  $n \leq k/3$  and such that  $(Z''_n)_{n \in \mathbf{Z}}$  is independent of  $(Z_n)_{n \in \mathbf{Z}}$ . Let  $(X'_n)_{n \in \mathbf{Z}}$  and  $(X''_n)_{n \in \mathbf{Z}}$  be the corresponding functionals, and note that (i) is automatically satisfied. In order to prove (ii), we define for  $i \geq k$  the function

$$u_i(z_{\lfloor \frac{2k}{3} \rfloor}, \dots, z_{2i - \lfloor \frac{2k}{3} \rfloor}) := E(X_i | Z_{\lfloor \frac{2k}{3} \rfloor} = z_{\lfloor \frac{2k}{3} \rfloor}, \dots, Z_{2i - \lfloor \frac{2k}{3} \rfloor} = z_{2i - \lfloor \frac{2k}{3} \rfloor})$$

By the 1-approximation property we get that

$$(2.13) \quad E|X_i - u_i(Z_{\lfloor \frac{2k}{3} \rfloor}, \dots, Z_{2i - \lfloor \frac{2k}{3} \rfloor})| \leq a_{i - \lfloor \frac{2k}{3} \rfloor},$$

$$(2.14) \quad E|X'_i - u_i(Z'_{\lfloor \frac{2k}{3} \rfloor}, \dots, Z'_{2i - \lfloor \frac{2k}{3} \rfloor})| \leq a_{i - \lfloor \frac{2k}{3} \rfloor}.$$

Now define the event  $A = \{Z_i = Z'_i \text{ for all } i \geq \lfloor \frac{2k}{3} \rfloor\}$ . Then combining (2.13) and (2.14), we get

$$\begin{aligned} & E(|X_i - X'_i|1_A) \\ &= E \left| X_i - u_i(Z_{\lfloor \frac{2k}{3} \rfloor}, \dots, Z_{2i - \lfloor \frac{2k}{3} \rfloor}) \right| + E \left| X'_i - u_i(Z'_{\lfloor \frac{2k}{3} \rfloor}, \dots, Z'_{2i - \lfloor \frac{2k}{3} \rfloor}) \right| \\ &\leq 2a_{i - \lfloor \frac{2k}{3} \rfloor}, \end{aligned}$$

which proves (ii). The proof of (iii) follows along the same lines, but is simpler because  $Z''_n = Z'_n$  for  $n \leq k/3$  on the entire probability space.  $\square$

**Corollary 2.17.** *Let  $(X_n)_{n \in \mathbf{Z}}$  be a 1-approximating functional of an absolutely regular process. Then, given  $k \in \mathbf{N}$ , there exist copies  $(X'_n)_{n \in \mathbf{Z}}$  and  $(X''_n)_{n \in \mathbf{Z}}$  of  $(X_n)_{n \in \mathbf{Z}}$  with the following properties:*

- (i)  $(X''_n)_{n \in \mathbf{Z}}$  is independent of  $(X_n)_{n \in \mathbf{Z}}$ .
  - (ii)  $P(\sum_{i=k}^{\infty} |X_i - X'_i| \leq \alpha_{\lfloor k/3 \rfloor}) \geq 1 - \beta_{\lfloor k/3 \rfloor} - \alpha_{\lfloor k/3 \rfloor}$ .
  - (iii)  $P(\sum_{i=0}^{\infty} |X'_i - X''_i| \leq \alpha_{\lfloor k/3 \rfloor}) \geq 1 - \alpha_{\lfloor k/3 \rfloor}$ ,
- where  $(\alpha_n)_{n \geq 1}$  is defined as in (2.3).

*Proof.* We use the same construction as in Proposition 2.16. Then we get from (ii) that

$$E \left( \sum_{i=k}^{\infty} |X_i - X'_i|1_A \right) \leq 2 \sum_{i=k}^{\infty} a_{i - \lfloor \frac{2k}{3} \rfloor} \leq \sum_{i=\lfloor \frac{k}{3} \rfloor}^{\infty} a_i.$$

Now we can apply Markov's inequality, and get

$$\begin{aligned} P \left( \sum_{i=k}^{\infty} |X_i - X'_i| \geq \sqrt{2 \sum_{i=\lfloor \frac{k}{3} \rfloor}^{\infty} a_i} \right) &\leq P \left( \sum_{i=k}^{\infty} |X_i - X'_i|1_A \geq \sqrt{2 \sum_{i=\lfloor \frac{k}{3} \rfloor}^{\infty} a_i} \right) + P(A^c) \\ &\leq \sqrt{2 \sum_{i=\lfloor \frac{k}{3} \rfloor}^{\infty} a_i} + \beta_{k/3}, \end{aligned}$$

thus establishing (ii). The proof of (iii) follows in a similar way from Proposition 2.16(iii).  $\square$

In the proofs of several correlation inequalities the following corollary to Hölder’s inequality will be useful. We have

$$(2.15) \quad E|XY| \leq (E|X|)^{\frac{\delta}{1+\delta}} (\|X\|_{2+\delta})^{\frac{1}{1+\delta}} \|Y\|_{2+\delta},$$

for random variables  $X, Y \in L_{2+\delta}$ , and  $\delta > 0$ . Indeed, (2.15) can be proved by expressing  $|XY|$  as

$$|XY| = |X|^{\frac{\delta}{1+\delta}} |X|^{\frac{1}{1+\delta}} |Y|$$

and then applying Hölder’s inequality with exponents  $p = \frac{1+\delta}{\delta}$ ,  $q = (2 + \delta)(1 + \delta)$ , and  $r = 2 + \delta$ .

For certain fourth moment bounds, we need an analogous inequality for quadruple products, namely

$$(2.16) \quad E|XYZW| \leq (E|X|)^{\frac{\delta}{3+\delta}} (\|X\|_{4+\delta})^{\frac{3}{3+\delta}} \|Y\|_{4+\delta} \|Z\|_{4+\delta} \|W\|_{4+\delta}$$

valid for random variables  $X, Y, Z, W \in L_{4+\delta}$ , where  $\delta > 0$ . This inequality can be proved by writing  $|XYZW| = |X|^{\frac{\delta}{3+\delta}} |X|^{\frac{3}{3+\delta}} |Y||Z||W|$  and using a quintuple Hölder inequality with exponents  $p, q, r, s, t$  given by  $p = \frac{3+\delta}{\delta}$ ,  $q = (4 + \delta)(3 + \delta)/3$ , and  $r = s = t = 4 + \delta$ .

**Lemma 2.18.** *Let  $(X_n)_{n \in \mathbf{N}}$  be a 1-approximating functional with constants  $(a_l)_{l \geq 0}$  of an absolutely regular process with mixing coefficients  $(\beta_k)_{k \geq 0}$ . Then*

(i) *if  $X_0$  is a.s. bounded, we have for all  $i, k \geq 0$*

$$(2.17) \quad |E(X_i X_{i+k}) - (E X_i)(E X_{i+k})| \leq 4\|X_0\|_{\infty} a_{\lfloor \frac{k}{3} \rfloor} + 2\|X_0\|_{\infty}^2 \beta_{\lfloor \frac{k}{3} \rfloor}.$$

(ii) *if  $\|X_0\|_{2+\delta} < \infty$  for some  $\delta > 0$ , we have*

$$(2.18) \quad |E(X_i X_{i+k}) - (E X_i)(E X_{i+k})| \leq 2\|X_0\|_{2+\delta}^2 \left(\beta_{\lfloor \frac{k}{3} \rfloor}\right)^{\frac{\delta}{2+\delta}} + 4 \left(a_{\lfloor \frac{k}{3} \rfloor}\right)^{\frac{\delta}{1+\delta}} \|X_0\|_{2+\delta}^{\frac{2+\delta}{1+\delta}}$$

for all  $i, k \geq 0$ .

*Proof.* Due to stationarity, it suffices to prove (2.17) and (2.18) for  $i = 0$ . Let  $(X'_n)_{n \in \mathbf{Z}}$  and  $(X''_n)_{n \in \mathbf{Z}}$  be copies of  $(X_n)_{n \in \mathbf{Z}}$ , as defined in Proposition 2.16. Thus, the pairs of random variables  $(X_0, X_k)$ ,  $(X'_0, X'_k)$  and  $(X''_0, X''_k)$  have the same distribution. Moreover,  $(X''_0, X''_k)$  is independent of  $(X_0, X_k)$  and the properties (ii) and (iii) of Proposition 2.16 hold. Hence

$$(2.19) \quad \begin{aligned} |E(X_0 X_k) - E X_0 E X_k| &= |E(X'_0 X'_k) - E(X''_0 X''_k)| \\ &\leq E|X'_0(X'_k - X_k)| + E|(X'_0 - X''_0)X_k|. \end{aligned}$$

The first term on the r.h.s. of (2.19) is bounded by

$$\begin{aligned} \|X_0\|_{\infty} E|X'_k - X_k| &\leq \|X_0\|_{\infty} \{E(|X'_k - X_k|1_A) + E(|X'_k - X_k|1_{A^c})\} \\ &\leq \|X_0\|_{\infty} \left\{ 2a_{\lfloor \frac{k}{3} \rfloor} + 2\|X_0\|_{\infty} \beta_{\frac{k}{3}} \right\}. \end{aligned}$$

The second term on the r.h.s. of (2.19) is bounded by

$$\|X_k\|_{\infty} E|X'_0 - X''_0| \leq 2\|X_0\|_{\infty} a_{\lfloor \frac{k}{3} \rfloor},$$

and these two estimates together yield (2.17).

For the proof of (2.18) we use again (2.19). Decompose the first term on the r.h.s. of (2.19) as  $E|X'_0(X'_k - X_k)1_A| + E|X'_0(X'_k - X_k)1_{A^c}|$  and note that by Hölder's inequality

$$\begin{aligned} E|X'_0(X'_k - X_k)1_{A^c}| &\leq \left(\beta_{\lfloor \frac{k}{3} \rfloor}\right)^{\frac{\delta}{2+\delta}} \|X_0\|_{2+\delta} \|X'_k - X_k\|_{2+\delta} \\ &\leq 2 \left(\beta_{\lfloor \frac{k}{3} \rfloor}\right)^{\frac{\delta}{2+\delta}} \|X_0\|_{2+\delta}^2 \end{aligned}$$

Using (2.15) with  $X = (X'_k - X_k)1_A$  and  $Y = X'_0$ , together with Proposition 2.16(ii), we get

$$\begin{aligned} E|X_0(X'_k - X_k)1_A| &\leq \left(2a_{\lfloor \frac{k}{3} \rfloor}\right)^{\frac{\delta}{1+\delta}} (\|X'_k - X_k\|_{2+\delta})^{\frac{1}{1+\delta}} \|X_0\|_{2+\delta} \\ &\leq 2 \left(a_{\lfloor \frac{k}{3} \rfloor}\right)^{\frac{\delta}{1+\delta}} \|X_0\|_{2+\delta}^{\frac{2+\delta}{1+\delta}}. \end{aligned}$$

To bound the second term on the r.h.s. of (2.19), we again make use of (2.15), now with  $X = X'_0 - X''_0$  and  $Y = X_k$ . Then we get

$$\begin{aligned} E|(X_0 - X''_0)X_k| &\leq \left(2a_{\lfloor \frac{k}{3} \rfloor}\right)^{\frac{\delta}{1+\delta}} (\|X_0 - X''_0\|_{2+\delta})^{\frac{1}{1+\delta}} \|X_k\|_{2+\delta} \\ &\leq 2 \left(a_{\lfloor \frac{k}{3} \rfloor}\right)^{\frac{\delta}{1+\delta}} \|X_0\|_{2+\delta}^{\frac{2+\delta}{1+\delta}}. \end{aligned}$$

Combining the last three estimates, we obtain (2.18). □

**Lemma 2.19.** *Let  $(Z_n)_{n \in \mathbf{Z}}$  be an absolutely regular process with mixing coefficients  $(\beta_k)_{k \geq 0}$  and let  $(X_n)_{n \in \mathbf{N}}$  be a 2-approximating functional of  $(Z_n)_{n \in \mathbf{Z}}$  with constants  $(a_k)_{k \geq 0}$ . Then*

(2.20)

$$|E(X_i X_{i+k}) - (EX_i)(EX_{i+k})| \leq 2\|X_0\|_{2+\delta}^2 \beta_{\lfloor k/3 \rfloor}^{\delta/(2+\delta)} + 4(a_{\lfloor k/3 \rfloor})^{1/2} \|X_0\|_2.$$

*Proof.* Due to stationarity, it suffices to prove (2.20) for  $i = 0$ . Let  $(X'_n)_{n \in \mathbf{Z}}$  and  $(X''_n)_{n \in \mathbf{Z}}$  be copies of  $(X_n)_{n \in \mathbf{Z}}$ , as defined in Proposition 2.16. Thus, the pairs of random variables  $(X_0, X_k)$ ,  $(X'_0, X'_k)$  and  $(X''_0, X''_k)$  have the same distribution. As the 2-approximation condition holds, we may replace (ii) and (iii) of Proposition 2.16 by

$$\begin{aligned} E(|X_k - X'_k|^2 1_A) &\leq 4a_{\lfloor \frac{k}{3} \rfloor}, \\ E(|X'_0 - X''_0|^2) &\leq 4a_{\lfloor \frac{k}{3} \rfloor}. \end{aligned}$$

As  $(X''_0, X''_k)$  is independent of  $(X_0, X_k)$ , we get

$$\begin{aligned} |E(X_0 X_k) - E X_0 E X_k| &= |E(X'_0 X'_k) - E(X''_0 X_k)| \\ (2.21) \qquad \qquad \qquad &\leq E|X'_0(X'_k - X_k)| + E|(X'_0 - X''_0)X_k|. \end{aligned}$$

The first term on the r.h.s. of (2.21) is bounded by

$$\begin{aligned} E|X'_0(X'_k - X_k)| &= E(|X'_0||X'_k - X_k|1_A) + E(|X'_0||X_k - X'_k|1_{A^c}) \\ &\leq 2\|X_0\|_2 \left(a_{\lfloor \frac{k}{3} \rfloor}\right)^{1/2} + 2 \left(\beta_{\lfloor \frac{k}{3} \rfloor}\right)^{\frac{\delta}{2+\delta}} \|X_0\|_{2+\delta}^2. \end{aligned}$$

The second term on the r.h.s. of (2.21) is bounded by

$$\|X_k\|_2 \|X'_0 - X''_0\|_2 \leq 2 \|X_0\|_2 \left(a_{\lfloor \frac{k}{3} \rfloor}\right)^{1/2}$$

Combining these bounds, we obtain (2.20). □

**Lemma 2.20.** *Let  $(X_n)_{n \in \mathbf{Z}}$  be a 1-approximating functional of an absolutely regular process. Let  $\xi$  be a  $p_1$ -continuous function of  $(\dots, X_{-1}, X_0)$  and let  $\eta$  be a  $p_2$ -continuous function of  $(X_k, X_{k+1}, \dots)$  with corresponding functions  $\psi_1$  and  $\psi_2$ , where  $p_1, p_2 \geq 1$ . Then*

$$(2.22) \quad \begin{aligned} |\mathbb{E}(\eta \cdot \xi) - \mathbb{E} \eta \cdot \mathbb{E} \xi| &\leq 4 \|\eta\|_p \|\xi\|_q \left(\beta_{\lfloor \frac{k}{3} \rfloor} + \alpha_{\lfloor \frac{k}{3} \rfloor}\right)^{1 - \frac{1}{p} - \frac{1}{q}} \\ &\quad + \psi^{\frac{1}{p_1}}(\alpha_{\lfloor \frac{k}{3} \rfloor}) \|\eta\|_{q_1} + \psi^{\frac{1}{p_2}}(\alpha_{\lfloor \frac{k}{3} \rfloor}) \|\xi\|_{q_2}, \end{aligned}$$

whenever  $\|\eta\|_{\max(p, q_1)} < \infty$ ,  $\|\xi\|_{\max(q, q_2)} < \infty$ , where  $q_1, q_2$  are such that  $\frac{1}{p_1} + \frac{1}{q_1} = 1$ ,  $\frac{1}{p_2} + \frac{1}{q_2} = 1$ , and  $p^{-1} + q^{-1} < 1$ .

*Proof.* Let  $(X'_n)_{n \in \mathbf{Z}}$  and  $(X''_n)_{n \in \mathbf{Z}}$  be copies of  $(X_n)_{n \in \mathbf{Z}}$ , as defined in Proposition 2.16, and let  $\xi', \eta'$  and  $\xi'', \eta''$  be the respective functions of  $(X'_n)_{n \in \mathbf{Z}}$  and  $(X''_n)_{n \in \mathbf{Z}}$ . Observe that the pairs of random variables  $(\xi, \eta)$ ,  $(\xi', \eta')$  and  $(\xi'', \eta'')$  have the same distribution and that  $(\xi'', \eta'')$  is independent of  $(\xi, \eta)$ . Hence we obtain

$$|\mathbb{E}(\xi \eta) - (\mathbb{E} \xi)(\mathbb{E} \eta)| = |\mathbb{E}(\xi' \eta') - \mathbb{E}(\xi'' \eta'')| \leq \mathbb{E} |(\xi' - \xi) \eta'| + \mathbb{E} |\xi(\eta' - \eta'')|.$$

Now define the events  $B = \sum_{i=0}^{\infty} |X'_{-i} - X''_{-i}| \leq \alpha_{\lfloor k/3 \rfloor}$  and  $D = \sum_{i=k}^{\infty} |X_i - X'_i| \leq \alpha_{\lfloor k/3 \rfloor}$ , and note that  $\mathbb{P}(B) \geq 1 - \alpha_{\lfloor k/3 \rfloor}$  and  $\mathbb{P}(D) \geq 1 - \beta_{\lfloor k/3 \rfloor} - \alpha_{\lfloor k/3 \rfloor}$ . Then, by Hölder's inequality, we obtain  $\mathbb{E} |(\xi' - \xi) \eta'|_{B^c} \leq 2 \|\xi\|_p \|\eta\|_q (\alpha_{\lfloor k/3 \rfloor})^{1 - 1/p - 1/q}$ . Moreover, by  $p$ -continuity we get

$$\begin{aligned} \mathbb{E} |(\xi' - \xi) \eta'|_{B^c} &\leq \|\eta'\|_{q_1} \left( \mathbb{E} \left| (\xi - \xi') 1_{\{\sum_{i=-\infty}^0 X_i - \sum_{i=-\infty}^0 X'_i \leq \alpha_{\lfloor k/3 \rfloor}\}} \right|^{p_1} \right)^{1/p_1} \\ &\leq \|\eta\|_{q_1} (\psi(\alpha_{\lfloor k/3 \rfloor}))^{1/p_1}. \end{aligned}$$

In the same way we can bound  $\mathbb{E} |\xi(\eta' - \eta'')|$ , and thus obtain (2.22). □

In the analysis of empirical processes, we regularly have to estimate fourth moments of partial sums of functionals of absolutely regular processes. In proofs of such estimates, bounds on the covariance of  $X_i$  and  $X_j X_k X_l$ , and of  $X_i X_j$  and  $X_k X_l$ , ( $0 \leq i \leq j \leq k \leq l \leq n$ ), play a crucial role. Such bounds can in principle be derived from Lemma 2.20, as both  $X_i X_j$  and  $X_j X_k X_l$  are 1-continuous functions, if  $X_j \in L_p$ . We obtain namely, using Loève's  $c_r$ -inequalities

$$\begin{aligned} &E \left( |X_i X_j - X'_i X'_j|^p 1_{\{|X_i - X'_i| + |X_j - X'_j| \leq \epsilon\}} \right) \\ &= E \left( |(X_i - X'_i) X_j + X'_i (X_j - X'_j)|^p 1_{\{|X_i - X'_i| + |X_j - X'_j| \leq \epsilon\}} \right) \\ &\leq 2^{p-1} E \left( \{|X_i|^p |X_j - X'_j|^p + |X_i - X'_i|^p |X'_j|^p\} 1_{\{|X_i - X'_i| + |X_j - X'_j| \leq \epsilon\}} \right) \\ &\leq 2^p \epsilon^p E |X_i|^p. \end{aligned}$$

On the other hand, because of the special product structure of these functions, a direct proof of correlation bounds is also possible and moreover yields sharper results, as the next lemma shows.

**Lemma 2.21.** *Let  $(X_n)_{n \in \mathbf{Z}}$  be a 1-approximating functional with constants  $(a_k)_{k \geq 0}$  of an absolutely regular process with mixing coefficients  $(\beta_k)_{k \geq 0}$ . Then, if  $X_0$  is a.s. bounded, we have for all non-negative integers  $i < j < k < l$*

$$(2.23) \quad \begin{aligned} & |E(X_i X_j X_k X_l) - E(X_i)E(X_j X_k X_l)| \\ & \leq \left\{ 6 \left( \beta_{\lfloor \frac{i-j}{3} \rfloor} \right)^{\frac{\delta}{2+\delta}} \|X_0\|_{2+\delta}^2 + 8 \left( a_{\lfloor \frac{i-j}{3} \rfloor} \right)^{\frac{\delta}{1+\delta}} \|X_0\|_{\frac{2+\delta}{1+\delta}}^{\frac{2+\delta}{1+\delta}} \right\} \|X_0\|_{\infty}^2. \end{aligned}$$

$$(2.24) \quad \begin{aligned} & |E(X_i X_j X_k X_l) - E(X_i X_j)E(X_k X_l)| \\ & \leq \left\{ 4 \left( \beta_{\lfloor \frac{k-j}{3} \rfloor} \right)^{\frac{\delta}{2+\delta}} \|X_0\|_{2+\delta}^2 + 8 \left( a_{\lfloor \frac{k-j}{3} \rfloor} \right)^{\frac{\delta}{1+\delta}} \|X_0\|_{\frac{2+\delta}{1+\delta}}^{\frac{2+\delta}{1+\delta}} \right\} \|X_0\|_{\infty}^2. \end{aligned}$$

*Proof.* We first prove (2.23). Due to stationarity, we may assume that  $i = 0$  and that  $0 \leq j \leq k \leq l$ . Let  $(X'_n)_{n \in \mathbf{Z}}$  and  $(X''_n)_{n \in \mathbf{Z}}$  be copies of  $(X_n)_{n \in \mathbf{Z}}$ , as defined in Proposition 2.16, only with  $j$  replacing  $k$  as length of the separating block. Thus, the quadruples of random variables  $(X_0, X_j, X_k, X_l)$ ,  $(X'_0, X'_j, X'_k, X'_l)$  and  $(X''_0, X''_j, X''_k, X''_l)$  have the same distribution. Moreover,  $(X''_0, X''_j, X''_k, X''_l)$  is independent of  $(X_0, X_j, X_k, X_l)$  and the properties (ii) and (iii) of Proposition 2.16 hold. Hence,

$$\begin{aligned} |E(X_0 X_j X_k X_l) - E(X_0)E(X_j X_k X_l)| &= |E(X'_0 X'_j X'_k X'_l) - E(X''_0 X_j X_k X_l)| \\ &\leq E(|X'_0 - X''_0| |X_j X_k X_l|) \\ &\quad + E(|X'_j - X_j| |X'_0 X_k X_l|) \\ &\quad + E(|X'_k - X_k| |X'_0 X'_j X_l|) \\ &\quad + E(|X'_l - X_l| |X'_0 X'_j X'_k|) \\ &=: I_1 + I_2 + I_3 + I_4. \end{aligned}$$

Regarding  $I_1$ , we note that  $E|X'_0 - X''_0| \leq 2a_{\lfloor \frac{j}{3} \rfloor}$  so that by (2.15) we get the estimate

$$I_1 \leq \|X_0\|_{\infty}^2 E(|X'_0 - X''_0| |X_j|) \leq 2 \left( a_{\lfloor \frac{j}{3} \rfloor} \right)^{\frac{\delta}{1+\delta}} \|X_0\|_{\infty}^2 \|X_0\|_{\frac{2+\delta}{1+\delta}}^{\frac{2+\delta}{1+\delta}}.$$

In order to bound  $I_2$ , we decompose  $|(X'_j - X_j)X'_0 X''_k X''_l| = |(X'_j - X_j)X'_0 X''_k X''_l| 1_A + |(X'_j - X_j)X'_0 X''_k X''_l| 1_{A^c}$ . By Hölder's inequality, we get

$$\begin{aligned} E(|(X'_j - X_j)X'_0 X''_k X''_l| 1_{A^c}) &\leq \left( \beta_{\lfloor \frac{j}{3} \rfloor} \right)^{\frac{\delta}{2+\delta}} \|X_0\|_{2+\delta} \|X_j - X'_j\|_{2+\delta} \|X_0\|_{\infty}^2 \\ &\leq 2 \left( \beta_{\lfloor \frac{j}{3} \rfloor} \right)^{\frac{\delta}{2+\delta}} \|X_0\|_{2+\delta}^2 \|X_0\|_{\infty}^2. \end{aligned}$$

Using (2.15) with  $X = (X'_j - X_j)1_A$  and  $Y = X'_0$ , together with Proposition 2.16(ii), we get

$$\begin{aligned} E(|(X'_j - X_j)X'_0 X''_k X''_l| 1_A) &\leq \left( 2a_{\lfloor \frac{j}{3} \rfloor} \right)^{\frac{\delta}{1+\delta}} (\|X'_j - X_j\|_{2+\delta})^{\frac{1}{1+\delta}} \|X_0\|_{2+\delta} \|X_0\|_{\infty}^2 \\ &\leq 2 \left( a_{\lfloor \frac{j}{3} \rfloor} \right)^{\frac{\delta}{1+\delta}} \|X_0\|_{\frac{2+\delta}{1+\delta}}^{\frac{2+\delta}{1+\delta}} \|X_0\|_{\infty}^2. \end{aligned}$$

Putting the last two inequalities together, we obtain

$$I_2 \leq 2 \left( \beta_{\lfloor \frac{j}{3} \rfloor} \right)^{\frac{\delta}{2+\delta}} \|X_0\|_{2+\delta}^2 \|X_0\|_{\infty}^2 + 2 \left( a_{\lfloor \frac{j}{3} \rfloor} \right)^{\frac{\delta}{1+\delta}} \|X_0\|_{\frac{2+\delta}{1+\delta}}^{\frac{2+\delta}{1+\delta}} \|X_0\|_{\infty}^2.$$

For  $I_3$  and  $I_4$  we obtain identical upper bounds, thus proving (2.23). The proof of (2.24) is almost the same, except that now we have two terms of the type  $I_1$  and also two terms of the type  $I_2$ , yielding a slightly different final result.  $\square$

The last of our correlation inequalities extends the result of Lemma 2.21 to the case of unbounded variables. The proof is almost identical, except that now (2.16) is employed to bound the quadruple products.

**Lemma 2.22.** *Let  $(X_n)_{n \in \mathbb{Z}}$  be a 1-approximating functional with constants  $(a_k)_{k \geq 0}$  of an absolutely regular process with mixing coefficients  $(\beta_k)_{k \geq 0}$ . Then, if  $\|X_0\|_{4+\delta}$  is bounded, we have for all non-negative integers  $i < j < k < l$*

$$\begin{aligned}
 & |E(X_i X_j X_k X_l) - E(X_i)E(X_j X_k X_l)| \\
 (2.25) \quad & \leq 6 \left( \beta_{\lfloor \frac{j-i}{3} \rfloor} \right)^{\frac{\delta}{4+\delta}} \|X_0\|_{4+\delta}^4 + 8 \left( a_{\lfloor \frac{j-i}{3} \rfloor} \right)^{\frac{\delta}{3+\delta}} \|X_0\|_{4+\delta}^{\frac{12+3\delta}{3+\delta}},
 \end{aligned}$$

$$\begin{aligned}
 & |E(X_i X_j X_k X_l) - E(X_i X_j)E(X_k X_l)| \\
 (2.26) \quad & \leq 4 \left( \beta_{\lfloor \frac{k-j}{3} \rfloor} \right)^{\frac{\delta}{4+\delta}} \|X_0\|_{4+\delta}^4 + 8 \left( a_{\lfloor \frac{k-j}{3} \rfloor} \right)^{\frac{\delta}{3+\delta}} \|X_0\|_{4+\delta}^{\frac{12+3\delta}{3+\delta}}.
 \end{aligned}$$

*Proof.* We first prove (2.25), and we start as in the proof of Lemma 2.21, obtaining the same upper bound for the l.h.s. of (2.25) by a sum of the four terms  $I_1, I_2, I_3, I_4$ . Regarding  $I_1$ , we note that  $E|X'_0 - X''_0| \leq 2a_{\lfloor \frac{j}{3} \rfloor}$ , so that by (2.16) we get the estimate

$$\begin{aligned}
 I_1 & \leq (E|X'_0 - X''_0|)^{\frac{\delta}{3+\delta}} (\|X'_0 - X''_0\|_{4+\delta})^{\frac{3}{3+\delta}} \|X_0\|_{4+\delta}^3 \\
 & \leq 2^{\frac{\delta}{3+\delta}} \left( a_{\lfloor \frac{j}{3} \rfloor} \right)^{\frac{\delta}{3+\delta}} 2^{\frac{3}{3+\delta}} \|X'_0\|_{4+\delta}^{\frac{3}{3+\delta}} \|X_0\|_{4+\delta}^3 \\
 & = 2 \left( a_{\lfloor \frac{j}{3} \rfloor} \right)^{\frac{\delta}{3+\delta}} \|X_0\|_{4+\delta}^{\frac{12+3\delta}{3+\delta}}.
 \end{aligned}$$

In order to bound  $I_2$ , we decompose  $|(X'_j - X_j)X'_0 X''_k X''_l| = |(X'_j - X_j)X'_0 X''_k X''_l| 1_A + |(X'_j - X_j)X'_0 X''_k X''_l| 1_{A^c}$ . By Hölder's inequality, we obtain

$$\begin{aligned}
 E(|(X'_j - X_j)X'_0 X''_k X''_l| 1_{A^c}) & \leq \left( \beta_{\lfloor \frac{j}{3} \rfloor} \right)^{\frac{\delta}{4+\delta}} \|X'_j - X_j\|_{4+\delta} \|X_0\|_{4+\delta}^3 \\
 & \leq 2 \left( \beta_{\lfloor \frac{j}{3} \rfloor} \right)^{\frac{\delta}{4+\delta}} \|X_0\|_{4+\delta}^4.
 \end{aligned}$$

Using (2.16) with  $X = (X'_j - X_j)1_A, Y = X'_0, Z = X''_k, W = X''_l$ , together with Proposition 2.16(ii), we get

$$\begin{aligned}
 E(|(X'_j - X_j)X'_0 X''_k X''_l| 1_A) & \leq \left( 2a_{\lfloor \frac{j}{3} \rfloor} \right)^{\frac{\delta}{3+\delta}} \left( \|X'_j - X_j\|_{4+\delta} \right)^{\frac{3}{3+\delta}} \|X_0\|_{4+\delta}^3 \\
 & \leq \left( 2a_{\lfloor \frac{j}{3} \rfloor} \right)^{\frac{\delta}{3+\delta}} \|X_0\|_{4+\delta}^{\frac{12+3\delta}{3+\delta}}.
 \end{aligned}$$

Putting the last two inequalities together, we get

$$I_2 \leq 2 \left( \beta_{\lfloor \frac{j}{3} \rfloor} \right)^{\frac{\delta}{4+\delta}} \|X_j\|_{4+\delta}^4 + 2 \left( a_{\lfloor \frac{j}{3} \rfloor} \right)^{\frac{\delta}{3+\delta}} \|X_0\|_{4+\delta}^{\frac{12+3\delta}{3+\delta}}.$$

For  $I_3$  and  $I_4$  we obtain identical upper bounds, thus proving (2.25). The proof of (2.26) is almost the same, except that now we have two terms of the type  $I_1$  and also two terms of the type  $I_2$ , yielding a slightly different final result.  $\square$



**2.4. Moment inequalities.** In this section, we prove some inequalities for second and fourth moments of partial sums  $S_N = X_1 + \dots + X_N$  of 1-approximating functionals of an absolutely regular process. These inequalities are essential tools e.g. for estimating the size of various remainder terms in proofs of central limit theorems.

**Lemma 2.23.** *Let  $(X_k)_{k \in \mathbf{Z}}$  be a 1-approximating functional with constants  $(a_k)_{k \geq 0}$  of an absolutely regular process with mixing coefficients  $(\beta_k)_{k \geq 0}$ . Suppose moreover that  $EX_i = 0$  and that one of the following two conditions holds:*

(i)  $X_0$  is bounded a.s. and  $\sum_{k=0}^\infty (a_k + \beta_k) < \infty$ .

(ii)  $E|X_0|^{2+\delta} < \infty$  and  $\sum_{k=0}^\infty (a_k^{\frac{\delta}{1+\delta}} + \beta_k^{\frac{\delta}{2+\delta}}) < \infty$ .

Then, as  $N \rightarrow \infty$ ,

$$(2.27) \quad \frac{1}{N}ES_N^2 \rightarrow EX_0^2 + 2 \sum_{j=1}^\infty E(X_0X_j)$$

and the sum on the r.h.s. converges absolutely.

*Proof.* Assume that (ii) holds. By stationarity of  $(X_n)_{n \in \mathbf{Z}}$ , we obtain

$$(2.28) \quad \begin{aligned} ES_N^2 &= \sum_{1 \leq i, j \leq N} E(X_iX_j) = NEX_0^2 + 2 \sum_{k=1}^N (N-k)E(X_0X_k) \\ &= N \left( EX_0^2 + 2 \sum_{k=1}^N \left(1 - \frac{k}{N}\right) E(X_0X_k) \right). \end{aligned}$$

As  $EX_i = 0$ , we get using Lemma 2.18 (ii) that

$$|E(X_0X_k)| \leq 2\|X_0\|_{2+\delta}^2 (\beta_{k/3})^{\frac{\delta}{2+\delta}} + 4(a_{k/3})^{\frac{\delta}{1+\delta}} (\|X_0\|_{2+\delta})^{\frac{2+\delta}{1+\delta}}.$$

Thus  $\sum_{k=0}^\infty E(X_0X_k)$  converges absolutely and by the dominated convergence theorem for series we obtain

$$\sum_{k=1}^N \left(1 - \frac{k}{N}\right) E(X_0X_k) \rightarrow \sum_{k=1}^\infty E(X_0X_k)$$

as  $N \rightarrow \infty$ , proving (2.27) under assumption (ii). The proof under (i) is practically the same, now using Lemma 2.18 (i). □

**Lemma 2.24.** *Let  $(X_k)_{k \in \mathbf{Z}}$  be a 1-approximating functional with constants  $(a_k)_{k \geq 0}$  of an absolutely regular process with mixing coefficients  $(\beta_k)_{k \geq 0}$ . Suppose moreover that  $EX_i = 0$  and that one of the following two conditions holds:*

(i)  $X_0$  is bounded a.s. and  $\sum_{k=0}^\infty k^2(a_k + \beta_k) < \infty$ .

(ii)  $E|X_0|^{4+\delta} < \infty$  and  $\sum_{k=0}^\infty k^2(a_k^{\frac{\delta}{3+\delta}} + \beta_k^{\frac{\delta}{4+\delta}}) < \infty$ .

Then there exists a constant  $C$  such that

$$E(S_N^4) \leq CN^2.$$

*Proof.* Assume that (ii) holds. By stationarity of  $(X_n)_{n \in \mathbb{Z}}$  we obtain

$$\begin{aligned}
 ES_N^4 &\leq 4! \sum_{1 \leq i_1 \leq i_2 \leq i_3 \leq i_4 \leq n} |E(X_{i_1} X_{i_2} X_{i_3} X_{i_4})| \\
 (2.29) \quad &\leq 4!N \sum_{\substack{i, j, k \geq 0 \\ i+j+k \leq N}} |E(X_0 X_i X_{i+j} X_{i+j+k})|.
 \end{aligned}$$

By Lemma 2.22, and using  $EX_0 = 0$ , we get the following bounds on the expected values inside the last sum

$$\begin{aligned}
 |E\{X_0(X_i X_{i+j} X_{i+j+k})\}| &\leq 6\beta_{\lfloor \frac{i}{3} \rfloor}^{\frac{\delta}{4+\delta}} \|X_0\|_{4+\delta}^4 + 8a_{\lfloor \frac{i}{3} \rfloor}^{\frac{\delta}{3+\delta}} \|X_0\|_{4+\delta}^{\frac{12+3\delta}{3+\delta}}, \\
 |E\{(X_0 X_i X_{i+j}) X_{i+j+k}\}| &\leq 6\beta_{\lfloor \frac{i}{3} \rfloor}^{\frac{\delta}{4+\delta}} \|X_0\|_{4+\delta}^4 + 8a_{\lfloor \frac{i}{3} \rfloor}^{\frac{\delta}{3+\delta}} \|X_0\|_{4+\delta}^{\frac{12+3\delta}{3+\delta}}.
 \end{aligned}$$

Alternatively, we can split  $E(X_0 X_i X_{i+j} X_{i+j+k})$  up as  $E\{(X_0 X_i)(X_{i+j} X_{i+j+k})\}$ , and apply (2.26) together with (2.18). Then we get

$$\begin{aligned}
 |E\{(X_0 X_i)(X_{i+j} X_{i+j+k})\}| &\leq |E(X_0 X_i)| |E(X_{i+j} X_{i+j+k})| \\
 &\quad + 4\beta_{\lfloor \frac{i}{3} \rfloor}^{\frac{\delta}{4+\delta}} \|X_0\|_{4+\delta}^4 + 8a_{\lfloor \frac{i}{3} \rfloor}^{\frac{\delta}{3+\delta}} \|X_0\|_{4+\delta}^{\frac{12+3\delta}{3+\delta}} \\
 &\leq \left( 2\|X_0\|_{4+\delta}^2 \beta_{\lfloor \frac{i}{3} \rfloor}^{\frac{2+\delta}{4+\delta}} + 2\|X_0\|_{4+\delta}^{\frac{4+\delta}{3+\delta}} a_{\lfloor \frac{i}{3} \rfloor}^{\frac{2+\delta}{4+\delta}} \right) \\
 &\quad \times \left( 2\|X_0\|_{4+\delta}^2 \beta_{\lfloor \frac{k}{3} \rfloor}^{\frac{2+\delta}{4+\delta}} + 2\|X_0\|_{4+\delta}^{\frac{4+\delta}{3+\delta}} a_{\lfloor \frac{k}{3} \rfloor}^{\frac{2+\delta}{4+\delta}} \right) \\
 &\quad + 4\beta_{\lfloor \frac{i}{3} \rfloor}^{\frac{\delta}{4+\delta}} \|X_0\|_{4+\delta}^4 + 8a_{\lfloor \frac{i}{3} \rfloor}^{\frac{\delta}{3+\delta}} \|X_0\|_{4+\delta}^{\frac{12+3\delta}{3+\delta}}.
 \end{aligned}$$

In this way, we have obtained 3 bounds for  $|E(X_0 X_i X_{i+j} X_{i+j+k})|$  that we can use in getting an upper bound for the r.h.s. of (2.29).

$$\begin{aligned}
 (2.30) \quad E(S_N^4) &\leq CN \left\{ \sum_{0 \leq j, k \leq i \leq N} \left( \beta_{\lfloor \frac{i}{3} \rfloor}^{\frac{\delta}{4+\delta}} + a_{\lfloor \frac{i}{3} \rfloor}^{\frac{\delta}{3+\delta}} \right) + \sum_{0 \leq i, j \leq k} \left( \beta_{\lfloor \frac{k}{3} \rfloor}^{\frac{\delta}{4+\delta}} + a_{\lfloor \frac{k}{3} \rfloor}^{\frac{\delta}{3+\delta}} \right) \right. \\
 &\quad + \sum_{0 \leq i, k \leq j \leq N} \left[ \left( \beta_{\lfloor \frac{i}{3} \rfloor}^{\frac{2+\delta}{4+\delta}} + a_{\lfloor \frac{i}{3} \rfloor}^{\frac{2+\delta}{3+\delta}} \right) \left( \beta_{\lfloor \frac{k}{3} \rfloor}^{\frac{2+\delta}{4+\delta}} + a_{\lfloor \frac{k}{3} \rfloor}^{\frac{2+\delta}{3+\delta}} \right) \right. \\
 &\quad \left. \left. + \left( \beta_{\lfloor \frac{j}{3} \rfloor}^{\frac{\delta}{4+\delta}} + a_{\lfloor \frac{j}{3} \rfloor}^{\frac{\delta}{3+\delta}} \right) \right] \right\} \\
 &\leq CN \left\{ 3 \sum_{0 \leq i, k \leq j \leq N} \left( \beta_{\lfloor \frac{j}{3} \rfloor}^{\frac{\delta}{4+\delta}} + a_{\lfloor \frac{j}{3} \rfloor}^{\frac{\delta}{3+\delta}} \right) \right. \\
 &\quad \left. + \sum_{0 \leq i, k \leq j \leq N} \left( \beta_{\lfloor \frac{i}{3} \rfloor}^{\frac{2+\delta}{4+\delta}} + a_{\lfloor \frac{i}{3} \rfloor}^{\frac{2+\delta}{3+\delta}} \right) \left( \beta_{\lfloor \frac{k}{3} \rfloor}^{\frac{2+\delta}{4+\delta}} + a_{\lfloor \frac{k}{3} \rfloor}^{\frac{2+\delta}{3+\delta}} \right) \right\}.
 \end{aligned}$$

Note that

$$\begin{aligned}
 \sum_{0 \leq i, k \leq j \leq N} \left( \beta_{\lfloor j/3 \rfloor}^{\frac{\delta}{4+\delta}} + a_{\lfloor j/3 \rfloor}^{\frac{\delta}{3+\delta}} \right) &\leq \sum_{j=0}^N \sum_{i, k=0}^j \left( \beta_{\lfloor j/3 \rfloor}^{\frac{\delta}{4+\delta}} + a_{\lfloor j/3 \rfloor}^{\frac{\delta}{3+\delta}} \right) \\
 (2.31) \qquad \qquad \qquad &\leq 3 \sum_{j=1}^{\infty} j^2 \left( \beta_j^{\frac{\delta}{4+\delta}} + a_j^{\frac{\delta}{3+\delta}} \right) < \infty,
 \end{aligned}$$

and

$$\begin{aligned}
 \sum_{0 \leq i, k \leq j \leq N} \left( \beta_{\lfloor i/3 \rfloor}^{\frac{2+\delta}{4+\delta}} + a_{\lfloor i/3 \rfloor}^{\frac{2+\delta}{3+\delta}} \right) \left( \beta_{\lfloor k/3 \rfloor}^{\frac{2+\delta}{4+\delta}} + a_{\lfloor k/3 \rfloor}^{\frac{2+\delta}{3+\delta}} \right) \\
 \leq \sum_{j=0}^N \sum_{i, k=0}^j \left( \beta_{\lfloor i/3 \rfloor}^{\frac{2+\delta}{4+\delta}} + a_{\lfloor i/3 \rfloor}^{\frac{2+\delta}{3+\delta}} \right) \left( \beta_{\lfloor k/3 \rfloor}^{\frac{2+\delta}{4+\delta}} + a_{\lfloor k/3 \rfloor}^{\frac{2+\delta}{3+\delta}} \right) \\
 (2.32) \qquad \qquad \qquad \leq N \left[ \sum_{j=0}^{\infty} \left( \beta_j^{\frac{2+\delta}{4+\delta}} + a_j^{\frac{2+\delta}{3+\delta}} \right) \right]^2 < \infty.
 \end{aligned}$$

Combining now (2.31) and (2.32) with (2.30), we have proved the lemma in case (ii). The proof under (i) is almost the same, now using Lemma 2.21 and (2.17) instead of Lemma 2.22 and (2.18). □

### 3. EMPIRICAL PROCESS CLT

In the present section we prove some basic limit theorems for functionals of absolutely regular processes. The main result will be an invariance principle for the empirical process, which will later play an important role in our treatment of  $U$ -processes.

**3.1. Central limit theorem for partial sums.** In this section we shall prove the central limit theorem for partial sums  $S_n = \sum_{k=1}^n X_k$  of functionals of absolutely regular processes. This result is not new, and was already established by Ibragimov and Linnik [32] for the broader class of functionals of strong mixing processes. We provide an independent proof here, mainly as illustration of our general techniques for handling functionals.

**Theorem 4.** *Let  $(X_k)_{k \in \mathbf{Z}}$  be a 1-approximating functional with constants  $(a_k)_{k \geq 0}$  of an absolutely regular process with mixing coefficients  $(\beta_k)_{k \geq 0}$ . Suppose moreover that  $EX_i = 0$ ,  $E|X_0|^{4+\delta} < \infty$  and that*

$$(3.1) \qquad \qquad \qquad \sum_{k=0}^{\infty} k^2 (a_k^{\frac{\delta}{3+\delta}} + \beta_k^{\frac{\delta}{4+\delta}}) < \infty,$$

for some  $\delta > 0$ . Then, as  $n \rightarrow \infty$ ,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i \xrightarrow{d} \mathcal{N}(0, \sigma^2),$$

where  $\sigma^2 = EX_0^2 + 2 \sum_{j=1}^{\infty} E(X_0 X_j)$ . In case  $\sigma^2 = 0$ , we adopt the convention that  $\mathcal{N}(0, 0)$  denotes the point mass at the origin.

If  $X_0$  is bounded, the CLT continues to hold if (3.1) is replaced by the condition that

$$(3.2) \quad \sum_{k=0}^{\infty} k^2 (a_k + \beta_k) < \infty.$$

*Proof.* If  $\sigma^2 = 0$ , Lemma 2.23 implies that  $\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i \rightarrow 0$ , thus proving the theorem in that case. In the rest of the proof we may therefore assume that  $\sigma^2 > 0$ . In order to analyze  $\sum_{i=1}^n X_i$ , we will introduce a blocking with block lengths depending on  $n$ . We then use near regularity to approximate the long blocks by independent blocks, apply the Lyapunov CLT to the sums of the independent blocks and finally show that the contribution of the small separating blocks is negligible.

Given integers  $K = K_n, L = L_n$  and  $N = N_n$  (to be chosen later), we introduce the  $(K + 2L, N)$ -blocking  $(B_s)_{s=1}^{p_n}$  of the random variables  $X_1, \dots, X_n$ . Here  $p_n$  denotes the index of the last block fully contained in  $X_1, \dots, X_n$ , i.e.

$$p_n = \left\lfloor \frac{n}{K + 2L + N} \right\rfloor.$$

By Theorem 3 there exists a sequence of i.i.d. blocks  $(B'_s)_{s=1}^{p_n}$  with the same block-wise marginal distribution as  $(B_s)_{s=1}^{p_n}$  and such that

$$(3.3) \quad P(\|B_s - B'_s\| \geq 2\alpha_L) \leq 2\alpha_L + \beta_L.$$

Let  $I_s$  denote the index set of the  $s$ th  $(K + 2L, N)$  block and let  $J_s$  denote the index set of the  $s$ th separating block (of length  $K + 2L$ ). Denote by  $i_n$  the index of the last long block fully contained in  $1, 2, \dots, n$ . Define

$$U_s = \sum_{i \in I_s} X_i \quad \text{and} \quad V_s = \sum_{i \in I_s} X'_i$$

where  $B' = (X'_i)_{i \in I_s}$ . Then we can decompose  $\sum_{i=1}^n X_i$  as follows

$$(3.4) \quad \begin{aligned} \sum_{i=1}^n X_i &= \sum_{s=1}^{p_n} \sum_{i \in I_s} X_i + \sum_{s=1}^{p_n} \sum_{i \in J_s} X_i + \sum_{i=i_n+1}^n X_i \\ &= \sum_{s=1}^{p_n} U_s + \sum_{s=1}^{p_n} \sum_{i \in J_s} X_i + \sum_{i=i_n+1}^n X_i \\ &= \sum_{s=1}^{p_n} V_s + \sum_{s=1}^{p_n} (U_s - V_s) + \sum_{s=1}^{p_n} \sum_{i \in J_s} X_i + \sum_{i=i_n+1}^n X_i. \end{aligned}$$

We will prove that the first term on the r.h.s. is asymptotically normal and that the remaining terms are negligible.

At this point we choose the block lengths in such a way that the following hold:

$$(3.5) \quad K_n, L_n, N_n \rightarrow \infty,$$

$$(3.6) \quad \frac{K_n + L_n}{N_n} \rightarrow 0,$$

$$(3.7) \quad \frac{N_n}{n} \rightarrow 0,$$

$$(3.8) \quad \frac{n}{N_n} (\beta_{K_n} + \alpha_{L_n}) \rightarrow 0.$$

Such a choice is always possible, because  $\beta_k + \alpha_k \rightarrow 0$  as  $k \rightarrow \infty$ . The above conditions then imply that  $p_n \cdot \frac{N_n}{n} \rightarrow 1$  and hence that

$$p_n(K_n + L_n)/n \rightarrow 0.$$

Regarding  $\sum_{j=1}^{p_n} V_s$ , we note that  $V_1, \dots, V_{p_n}$  are independent identically distributed random variables with  $EV_1 = 0$  and  $\sigma_N^2 = \frac{1}{N} \text{Var}(V_1) \rightarrow \sigma^2$  as  $n \rightarrow \infty$ . Moreover, the Lyapunov condition for the CLT is satisfied with  $\delta = 2$  because

$$\frac{1}{(\sum_{s=1}^{p_n} \text{Var}(V_s))^2} \sum_{s=1}^{p_n} EV_s^4 = \frac{p_n E\left(\sum_{i=1}^N X_i\right)^4}{p_n^2 N^2 \sigma_N^4} = O\left(\frac{1}{p_n \sigma_N^4}\right)$$

and  $p_n \rightarrow \infty, \sigma_N \rightarrow \sigma$ . Hence we get  $\frac{1}{\sqrt{p_n \text{Var}(V_1)}} \sum_{s=1}^{p_n} V_s \rightarrow N(0, 1)$ , and since  $p_n \text{Var}(V_1) = N p_n \sigma_N^2 \sim n \sigma^2$  that

$$\frac{1}{\sqrt{n}} \sum_{s=1}^{p_n} V_s \rightarrow N(0, \sigma^2).$$

To bound the second term on the r.h.s. of (3.4), we note that

$$|U_s - V_s| \leq \sum_{i \in I_s} |X_i - X'_i| = |B_s - B'_s|.$$

Hence we get from (3.3) that

$$\begin{aligned} P\left(\frac{1}{\sqrt{n}} \sum_{s=1}^{p_n} |U_s - V_s| \geq \frac{p_n \alpha_L}{\sqrt{n}}\right) &= P\left(\sum_{s=1}^{p_n} |U_s - V_s| \geq p_n \alpha_L\right) \\ &\leq \sum_{s=1}^{p_n} P(|U_s - V_s| \geq \alpha_L) \\ &\leq p_n (\beta_K + 2\alpha_L). \end{aligned}$$

Thus by (3.8) we find that  $\frac{1}{\sqrt{n}} \sum_{s=1}^{p_n} |U_s - V_s| \rightarrow 0$  in probability.

Regarding the last term, note that there exists a constant  $C$  such that

$$E\left(\sum_{i \in I} X_i\right)^2 \leq C \#I.$$

This can be proved using similar arguments as in the proof of Lemma 2.23. Hence we get

$$E\left(\frac{1}{\sqrt{n}} \sum_{s=1}^{p_n} \sum_{i \in J_s} X_i + \sum_{i=i_n+1}^n X_i\right)^2 \leq C(p_n(K + 2L) + N)/n \rightarrow 0.$$

Now we can apply Slutsky's lemma to obtain the CLT. □

**3.2. Empirical process indexed by functions.** Let  $(X_n)_{n \geq 1}$  be a stationary ergodic process of  $\mathbf{R}$ -valued random variables with marginal distribution function  $F(t) := P(X_1 \leq t)$ . We define the empirical distribution function (e.d.f.)  $F_n : \mathbf{R} \rightarrow$

$\mathbf{R}$  by

$$\begin{aligned} F_n(t) &= \frac{1}{n} \#\{1 \leq i \leq n : X_i \leq t\} \\ &= \frac{1}{n} \sum_{i=1}^n 1_{(-\infty, t]}(X_i). \end{aligned}$$

The e.d.f. gives the fraction of observations in the sample that is  $\leq t$  and can hence be considered as the sample analogue and natural estimator of the true distribution function  $F(t)$ .

By the ergodic theorem  $F_n(t)$  converges to  $F(t)$  almost surely, and this even holds uniformly over all  $t \in \mathbf{R}$  (Glivenko-Cantelli lemma). Hence it is natural to study asymptotic properties of the normalized differences

$$W_n(t) = \sqrt{n}(F_n(t) - F(t)),$$

also called empirical process. Donsker (1953) proved the invariance principle for the empirical process in the case of i.i.d. observations, showing weak convergence of  $W_n(t)$  towards a Gaussian process  $W(t)$  and thereby confirming an earlier conjecture of Doob [24]. This result was extended to functionals of uniformly mixing processes by Billingsley [9], and to strongly mixing processes by Deo [22]. Berkes and Philipp [7] and Philipp [39] obtained further results for the empirical distribution function of strong mixing sequences, such as almost sure invariance principle and law of iterated logarithm.

Note that  $F_n(t)$  is the distribution function of the distribution placing mass  $\frac{1}{n}$  in each of the points  $X_i, 1 \leq i \leq n$ . Thus we can write  $F_n(t) = \int 1_{(-\infty, t]}(s) dF_n(s)$ , and hence

$$W_n(t) = \sqrt{n} \left( \int 1_{(-\infty, t]}(s) dF_n(s) - \int 1_{(-\infty, t]}(s) dF(s) \right).$$

This representation motivates the study of the (generalized) empirical process

$$\begin{aligned} W_n(t) &= \sqrt{n} \int u_t(s) d(F_n(s) - F(s)) \\ (3.9) \quad &= \sqrt{n} \left( \frac{1}{n} \sum_{i=1}^n u_t(X_i) - \int u_t(x) dF(x) \right) \end{aligned}$$

where  $\mathcal{U} = \{u_t, t \in [0, t_0]\}$  is some one-parameter class of functions. We will later specifically be interested in the functions  $u_t(x) = P(|Y - x| \leq t)$ , but at this point we will consider a general class  $\mathcal{U}$ , satisfying the following two properties

$$(3.10) \quad 0 \leq u_t(x) \leq 1, \text{ for all } t \in [0, t_0],$$

$$(3.11) \quad u_s(x) \leq u_t(x), \text{ for } 0 \leq s \leq t \leq t_0,$$

both for all  $x \in \mathbf{R}$ .

Without loss of generality we shall assume that  $t_0 = 1$  and define

$$G(t) = \mathbf{E} u_t(X_0) \text{ and } g_t(x) = u_t(x) - G(t).$$

We assume that  $G$  is Lipschitz continuous on  $[0, 1]$ , i.e. that there exists a constant  $C$  such that for all  $t, s \in [0, 1]$

$$(3.12) \quad |G(t) - G(s)| \leq C|t - s|$$

We view  $\{W_n(t), 0 \leq t \leq 1\}$  as a random element of  $D[0, 1]$ , the space of right-continuous functions on  $[0, 1]$  with left limits, equipped with the Skorohod topology.

**Theorem 5.** *Let  $(X_n)_{n \in \mathbf{N}}$  be a 1-approximating functional with constants  $(a_k)_{k \geq 0}$  of an absolutely regular process with mixing coefficients  $(\beta_k)_{k \geq 0}$  satisfying*

$$(3.13) \quad \sum_{k=0}^{\infty} k^2 \beta_k^{\delta/(2+\delta)} < \infty,$$

for some  $0 < \delta < 1$ . Let  $\mathcal{U}$  be a class of functions satisfying (3.10) and (3.11), and assume that  $u_t, t \in [0, 1]$ , are 1-continuous in the sense of (2.6), with the same  $\phi$  for all  $t$ . Assume moreover that

$$(3.14) \quad \sum_{k=0}^{\infty} k^2 a_k^{\frac{\delta}{2+2\delta}} < \infty \text{ and } \sum_{k=0}^{\infty} k^2 (\phi(\sqrt{2a_k}))^{\delta/(1+\delta)} < \infty.$$

Assume moreover that the function  $G(t)$  satisfies the condition (3.12).

Then the empirical process  $\{W_n(t), t \in [0, 1]\}_{n \in \mathbf{N}}$ , defined in (3.9), converges weakly in  $D[0, 1]$  to the mean-zero Gaussian process  $\{W(t), t \in [0, 1]\}$  with covariance structure

$$(3.15) \quad \begin{aligned} \text{Cov}(W(s), W(t)) &= \text{Cov}(u_s(X_1), u_t(X_1)) \\ &+ \sum_{k=1}^{\infty} \text{Cov}(u_s(X_1), u_t(X_{k+1})) + \sum_{k=1}^{\infty} \text{Cov}(u_s(X_{k+1}), u_t(X_1)). \end{aligned}$$

Moreover, the series on the r.h.s. of (3.15) converges absolutely, and the limiting process  $W$  has continuous sample paths with probability 1.

The key tool in the proof of Theorem 5 is the following lemma bounding fourth moments of partial sums  $S_n = \sum_{i=1}^n Y_i$  of functionals of an absolutely regular process. This lemma sharpens the statement of Lemma 2.24 by providing explicit bounds in terms of moments of the  $Y_i$ 's. In turn, we have to require boundedness of the  $Y_i$ 's.

**Lemma 3.1.** *Let  $(Y_k)_{k \in \mathbf{Z}}$  be a 1-approximating functional with constants  $(a_k)_{k \geq 0}$  of an absolutely regular process with mixing coefficients  $(\beta_k)_{k \geq 0}$ . Suppose moreover that  $|Y_k| \leq 1, EY_k = 0$  and*

$$\sum_{k=0}^{\infty} k^2 \left( a_k^{\frac{\delta}{1+\delta}} + \beta_k^{\frac{\delta}{2+\delta}} \right) < \infty.$$

Then for  $S_n = Y_1 + \dots + Y_n$  we have

$$E S_n^4 \leq C \left( n \|Y_0\|_{\frac{2+\delta}{1+\delta}}^{\frac{2+\delta}{1+\delta}} + n^2 \|Y_0\|_{\frac{4+2\delta}{1+\delta}}^{\frac{4+2\delta}{1+\delta}} \right),$$

for some constant  $C$  depending only on  $(a_k)_{k \geq 0}$  and  $(\beta_k)_{k \geq 0}$ .

*Proof.* We proceed in the same way as in the proof of Lemma 2.24. By stationarity of  $(Y_n)_{n \in \mathbf{Z}}$  we obtain

$$E S_n^4 \leq 4!n \sum_{\substack{i,j,k \geq 0 \\ i+j+k \leq n}} E(Y_0 Y_i Y_{i+j} Y_{i+j+k}).$$

By Lemma 2.21 and using  $EY_0 = 0$ , we get the following bounds on the expected values inside the last sum:

$$\begin{aligned}
 |E\{Y_0(Y_i Y_{i+j} Y_{i+j+k})\}| &\leq 6\left(\beta_{\lfloor \frac{i}{3} \rfloor}\right)^{\frac{\delta}{2+\delta}} \|Y_0\|_{2+\delta}^2 + 8\left(a_{\lfloor \frac{i}{3} \rfloor}\right)^{\frac{\delta}{1+\delta}} \|Y_0\|_{2+\delta}^{\frac{2+\delta}{1+\delta}}, \\
 |E\{(Y_0 Y_i Y_{i+j} Y_{i+j+k})\}| &\leq 6\left(\beta_{\lfloor \frac{k}{3} \rfloor}\right)^{\frac{\delta}{2+\delta}} \|Y_0\|_{2+\delta}^2 + 8\left(a_{\lfloor \frac{k}{3} \rfloor}\right)^{\frac{\delta}{1+\delta}} \|Y_0\|_{2+\delta}^{\frac{2+\delta}{1+\delta}}.
 \end{aligned}$$

Alternatively, we can split  $E(Y_0 Y_i Y_{i+j} Y_{i+j+k})$  up as  $E\{(Y_0 Y_i)(Y_{i+j} Y_{i+j+k})\}$  and apply (2.24) together with (2.18) to get the bound

$$\begin{aligned}
 |E\{Y_0 Y_i Y_{i+j} Y_{i+j+k}\}| &\leq E(Y_0 Y_i) E(Y_{i+j} Y_{i+j+k}) \\
 &\quad + 4\left(\beta_{\lfloor \frac{i}{3} \rfloor}\right)^{\frac{\delta}{2+\delta}} \|Y_0\|_{2+\delta}^2 + 8\left(a_{\lfloor \frac{i}{3} \rfloor}\right)^{\frac{\delta}{1+\delta}} \|Y_0\|_{2+\delta}^{\frac{2+\delta}{1+\delta}} \\
 &\leq \left(2\|Y_0\|_{2+\delta}^2 \left(\beta_{\lfloor \frac{i}{3} \rfloor}\right)^{\frac{\delta}{2+\delta}} + 4\left(a_{\lfloor \frac{i}{3} \rfloor}\right)^{\frac{\delta}{1+\delta}} \|Y_0\|_{2+\delta}^{\frac{2+\delta}{1+\delta}}\right) \\
 &\quad \times \left(2\|Y_0\|_{2+\delta}^2 \left(\beta_{\lfloor \frac{k}{3} \rfloor}\right)^{\frac{\delta}{2+\delta}} + 4\left(a_{\lfloor \frac{k}{3} \rfloor}\right)^{\frac{\delta}{1+\delta}} \|Y_0\|_{2+\delta}^{\frac{2+\delta}{1+\delta}}\right) \\
 &\quad + 4\left(\beta_{\lfloor \frac{i}{3} \rfloor}\right)^{\frac{\delta}{2+\delta}} \|Y_0\|_{2+\delta}^2 + 8\left(a_{\lfloor \frac{i}{3} \rfloor}\right)^{\frac{\delta}{1+\delta}} \|Y_0\|_{2+\delta}^{\frac{2+\delta}{1+\delta}}.
 \end{aligned}$$

Any of the 3 bounds on  $E(Y_0 Y_i Y_{i+j} Y_{i+j+k})$  thus obtained can be used in getting an upper bound on the terms in the expansion of  $ES_n^4$ . In the same way as in the proof of Lemma 2.24 we get then

$$\begin{aligned}
 ES_n^4 &\leq Cn \left\{ \sum_{0 \leq j, k \leq i \leq n} \left( \beta_{\lfloor \frac{i}{3} \rfloor}^{\frac{\delta}{2+\delta}} \|Y_0\|_{2+\delta}^2 + a_{\lfloor \frac{i}{3} \rfloor}^{\frac{\delta}{1+\delta}} \|Y_0\|_{2+\delta}^{\frac{2+\delta}{1+\delta}} \right) \right. \\
 &\quad + \sum_{0 \leq i, k \leq j \leq n} \left( \beta_{\lfloor \frac{i}{3} \rfloor}^{\frac{\delta}{2+\delta}} \|Y_0\|_{2+\delta}^2 + a_{\lfloor \frac{i}{3} \rfloor}^{\frac{\delta}{1+\delta}} \|Y_0\|_{2+\delta}^{\frac{2+\delta}{1+\delta}} \right) \\
 &\quad \left. \times \left( \beta_{\lfloor \frac{i}{3} \rfloor}^{\frac{\delta}{2+\delta}} \|Y_0\|_{2+\delta}^2 + a_{\lfloor \frac{i}{3} \rfloor}^{\frac{\delta}{1+\delta}} \|Y_0\|_{2+\delta}^{\frac{2+\delta}{1+\delta}} \right) \right\},
 \end{aligned}
 \tag{3.16}$$

where  $C$  is some universal constant. Now we note that

$$\begin{aligned}
 \sum_{0 \leq j, k \leq i \leq n} \beta_{\lfloor \frac{i}{3} \rfloor}^{\frac{\delta}{2+\delta}} &\leq \sum_{i=1}^n i^2 \beta_{\lfloor \frac{i}{3} \rfloor}^{\frac{\delta}{2+\delta}} < \infty, \\
 \sum_{0 \leq j, k \leq i \leq n} a_{\lfloor \frac{i}{3} \rfloor}^{\frac{\delta}{1+\delta}} &\leq \sum_{i=1}^n i^2 a_{\lfloor \frac{i}{3} \rfloor}^{\frac{\delta}{1+\delta}} < \infty,
 \end{aligned}$$

by the assumptions of the lemma. Hence

$$\begin{aligned}
 \sum_{0 \leq j, k \leq i \leq n} \left( \beta_{\lfloor \frac{i}{3} \rfloor}^{\frac{\delta}{2+\delta}} \|Y_0\|_{2+\delta}^2 + a_{\lfloor \frac{i}{3} \rfloor}^{\frac{\delta}{1+\delta}} \|Y_0\|_{2+\delta}^{\frac{2+\delta}{1+\delta}} \right) &\leq \|Y_0\|_{2+\delta}^{\frac{2+\delta}{1+\delta}} \sum_{0 \leq j, k \leq i \leq n} \left( \beta_{\lfloor \frac{i}{3} \rfloor}^{\frac{\delta}{2+\delta}} + a_{\lfloor \frac{i}{3} \rfloor}^{\frac{\delta}{1+\delta}} \right) \\
 &\leq C \|Y_0\|_{2+\delta}^{\frac{2+\delta}{1+\delta}},
 \end{aligned}
 \tag{3.17}$$



as  $(2 + \delta)/(1 + \delta) \leq 2$  and  $\|Y_0\|_{2+\delta} \leq 1$ . Concerning the second term on the r.h.s. of (3.16) we obtain

$$\begin{aligned} & \sum_{0 \leq i, k \leq j \leq n} \left( \beta_{\lfloor \frac{i}{3} \rfloor}^{\frac{\delta}{2+\delta}} \|Y_0\|_{2+\delta}^2 + a_{\lfloor \frac{i}{3} \rfloor}^{\frac{\delta}{1+\delta}} \|Y_0\|_{2+\delta}^{\frac{2+\delta}{1+\delta}} \right) \times \left( \beta_{\lfloor \frac{k}{3} \rfloor}^{\frac{\delta}{2+\delta}} \|Y_0\|_{2+\delta}^2 + a_{\lfloor \frac{k}{3} \rfloor}^{\frac{\delta}{1+\delta}} \|Y_0\|_{2+\delta}^{\frac{2+\delta}{1+\delta}} \right) \\ & \leq n \left\{ \sum_{i=1}^{\infty} \left( \beta_{\lfloor \frac{i}{3} \rfloor}^{\frac{\delta}{2+\delta}} \|Y_0\|_{2+\delta}^2 + a_{\lfloor \frac{i}{3} \rfloor}^{\frac{\delta}{1+\delta}} \|Y_0\|_{2+\delta}^{\frac{2+\delta}{1+\delta}} \right) \right\}^2 \\ & \leq Cn \|Y_0\|_{2+\delta}^{\frac{4+2\delta}{1+\delta}}, \end{aligned}$$

where again we have used the fact that  $(2 + \delta)/(1 + \delta) \leq 2$  and  $\|Y_0\|_{2+\delta} \leq 1$ . The last inequality, together with (3.16) and (3.17) proves the statement of the lemma.  $\square$

*Proof of Theorem 5.* First consider, for  $t \in [0, 1]$  fixed,  $W_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^n g_t(X_i)$ . The random variables  $g_t(X_i)$  are centered and bounded, and themselves again functionals of an absolutely regular process  $(Z_n)_{n \in \mathbf{Z}}$  satisfying, according to Proposition 2.11, the 1-approximation condition. Moreover, the summability condition (3.14) on the corresponding sequence  $(\phi(\alpha_l))$  is that required by Theorem 4. Hence, Theorem 4 yields that for fixed  $t$ ,  $W_n(t)$  is asymptotically normal with mean 0 and variance

$$\sigma^2 = E(g_t(X_0))^2 + 2 \sum_{k=1}^{\infty} \text{Cov}(g_t(X_0), g_t(X_k)).$$

Furthermore, an application of the Cramer-Wold device yields that for any  $(t_1, \dots, t_k)$  the vector  $(W_n(t_1), \dots, W_n(t_k))$  approaches a  $k$ -dimensional normal distribution centered at the origin, and the covariances of these limit distributions are those given by (3.16).

It remains to show that the sequence of processes  $\{W_n(t), t \in [0, 1]\}_{n \in \mathbf{N}}$  is tight. According to Theorem 15.5 in Billingsley [9], tightness follows if we can show that for all  $\epsilon, \eta > 0$  there exists a  $\tau \in (0, 1)$  such that

$$(3.18) \quad P\left( \sup_{s \leq t \leq s+\tau} |W_n(t) - W_n(s)| \geq \epsilon \right) \leq \eta\tau.$$

In order to prove (3.18), we fix  $\epsilon$  and  $\eta$  and consider the difference

$$W_n(t) - W_n(s) = \frac{1}{\sqrt{n}} \sum_{i=1}^n (g_t(X_i) - g_s(X_i)).$$

First we bound the fourth moment of the sum on the r.h.s. with the help of the previous lemma. For this, note that since  $|g_t(x) - g_s(x)| \leq 1$ , we have for  $p \geq 1$

$$E|g_t(X_0) - g_s(X_0)|^p \leq E|g_t(X_0) - g_s(X_0)| \leq 2(G(t) - G(s)).$$

Hence, for any  $\delta > 0$ ,

$$\|g_t(X_0) - g_s(X_0)\|_{2+\delta} \leq C(G(t) - G(s))^{\frac{1}{2+\delta}}.$$

For  $0 < \delta < 2$ , we thus get

$$\begin{aligned} (\|g_t(X_0) - g_s(X_0)\|_{2+\delta})^4 & \leq C(G(t) - G(s))^{\frac{4}{2+\delta}} \\ & = C(G(t) - G(s))^{1+\delta_1} \end{aligned}$$

for some  $\delta_1 > 0$ . Now we can apply Lemma 3.1 to get

$$E \left\{ \sum_{i=1}^n (g_t(X_i) - g_s(X_i)) \right\}^4 \leq C_0 \left( n^2 (G(t) - G(s))^{1+\delta_1} + n (G(t) - G(s))^{\frac{1+\delta_1}{2}} \right)$$

and

(3.19)

$$E \{ W_n(t) - W_n(s) \}^4 \leq C_1 (G(t) - G(s))^{1+\delta_1} + \frac{1}{n} (G(t) - G(s))^{\frac{1+\delta_1}{2}}.$$

Now fix  $r \in [0, 1]$ , choose some number  $h \geq \epsilon/n^{\frac{2}{1+\delta_1}}$  and consider  $m$  consecutive intervals  $[r + (i - 1)h, r + ih]$  of length  $h$  in  $[0, 1]$ . By condition (3.12) on  $G(t)$  we have that

$$\max_{i \leq m} (G(s + ih) - G(s + (i - 1)h)) \leq Ch.$$

From (3.19) we get the following bound on the fourth moment of the increments of  $W_n$  on the interval  $[r + ih, r + (i + k)h]$ :

$$\begin{aligned} E (W_n(s + (i + k)h) - W_n(s + ih))^4 &\leq K_1 \left( (kh)^{1+\delta_1} + \frac{1}{n} (kh)^{\frac{1+\delta_1}{2}} \right) \\ (3.20) \qquad \qquad \qquad &\leq \frac{2K_1}{\epsilon} (kh)^{1+\delta_1}, \end{aligned}$$

because  $k \geq 1$ , and  $h$  was taken in such a way that  $h^{(1+\delta_1)/2} > \frac{\epsilon^{(1+\delta_1)/2}}{n} > \frac{\epsilon}{n}$  (we assume that  $\epsilon < 1$ ). By (3.20) and Theorem 12.2 in Billingsley [9] we have the following maximal inequality

$$(3.21) \qquad P \left( \max_{i \leq m} |W_n(r + ih) - W_n(r)| \geq \lambda \right) \leq \frac{K_2 (mh)^{1+\delta_1}}{\epsilon \lambda^4},$$

where the constant  $K_2$  depends only on the  $a_k$ 's and the  $\beta_k$ 's. Denote  $G_n(t) = \frac{1}{n} \sum_{i=1}^n u_t(X_i)$  and take  $s, t \in [0, 1]$  with  $s \leq t < s + h$ . Then by monotonicity of  $G_n$  and  $G$  we have

$$\begin{aligned} W_n(t) - W_n(s) &= \sqrt{n} (G_n(t) - G(t) - G_n(s) + G(s)) \\ &\leq \sqrt{n} (G_n(s + h) - G_n(s) - G(s + h) + G(s) \\ &\quad + G(s + h) - G(t)) \\ &\leq \sqrt{n} |G_n(s + h) - G(s + h) - G_n(s) + G(s)| \\ &\quad + \sqrt{n} |G(s + h) - G(t)| \\ (3.22) \qquad \qquad &\leq |W_n(s + h) - W_n(s)| + Ch\sqrt{n} \end{aligned}$$

and similarly

$$\begin{aligned} W_n(s) - W_n(t) &= \sqrt{n} (G_n(s) - G_n(t) + G(t) - G(s)) \\ &\leq \sqrt{n} (G(t) - G(s)) \\ (3.23) \qquad \qquad &\leq Ch\sqrt{n} \leq Ch\sqrt{n} + |W_n(s + h) - W_n(s)|. \end{aligned}$$

Together, from (3.22) and (3.23) we obtain

$$(3.24) \qquad |W_n(t) - W_n(s)| \leq |W_n(s + h) - W_n(s)| + Ch\sqrt{n}.$$

From (3.24) it follows that

$$\sup_{r \leq t \leq r+mh} |W_n(t) - W_n(r)| \leq 3 \max_{i \leq m} |W_n(r + ih) - W_n(r)| + Ch\sqrt{n}.$$

Now let  $h$  be such that  $h < \epsilon/C\sqrt{n}$ . Then by the last inequality and (3.21) we get

$$P\left(\sup_{r \leq t \leq r+mh} |W_n(t) - W_n(r)| \geq 4\epsilon\right) \leq \frac{K_2(mh)^{1+\delta_1}}{\epsilon^5}.$$

Now let  $\tau < (\frac{\eta\epsilon^5}{K_2})^{1/\delta_1}$  be given and choose  $h$  satisfying  $\epsilon/n^{\frac{2}{1+\delta_1}} \leq h < \epsilon/(Cn^{1/2})$  and an integer  $m$  such that  $\tau = mh$ . Then

$$P\left(\sup_{r \leq t \leq r+\tau} |W_n(t) - W_n(r)| \geq 4\epsilon\right) < \eta\tau.$$

The existence of  $m$  and  $h$  satisfying  $\tau = mh$  as well as the above restrictions follows if we can find an integer  $m$  satisfying

$$(3.25) \quad \frac{C\tau\sqrt{n}}{\epsilon} < m \leq \frac{\tau n^{2/(1+\delta_1)}}{\epsilon},$$

which is indeed true for sufficiently large  $n$ . This completes the proof of the theorem. □

#### 4. $U$ -STATISTICS LIMIT THEOREMS

In this section we will study the asymptotic behavior of  $U$ -statistics when the underlying process  $(X_n)_{n \geq 1}$  is a functional of an absolutely regular process. For notational brevity, we shall restrict the attention to bivariate  $U$ -statistics, i.e. to

$$U_n(h) = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} h(X_i, X_j),$$

where  $h : \mathbf{R}^k \times \mathbf{R}^k \rightarrow \mathbf{R}$  is a measurable function, symmetric in the sense that  $h(x, y) = h(y, x)$ . The extension to the case  $m > 2$  is however straightforward. Moreover, and also without loss of generality, we will assume that the underlying process is real-valued. The results that we obtain here are a weak law of large numbers and a central limit theorem, both for  $p$ -continuous kernels.

**4.1. Law of large numbers.** The history of the law of large numbers for  $U$ -statistics dates back to Hoeffding [30] and Berk [6] who proved for i.i.d. observations  $(X_n)_{n \geq 1}$  with marginal distribution  $F$  that

$$(4.1) \quad U_n \xrightarrow{a.s.} \int \int h(x, y) dF(x) dF(y)$$

provided  $h \in L_1(F \times F)$ . Aaronson, Burton, Dehling, Gilat, Hill and Weiss [1] investigated whether (4.1) would continue to hold for ergodic processes  $(X_n)_{n \geq 1}$ . By means of counterexamples, they could show that this is false, unless extra conditions are imposed. Specifically, (4.1) holds if one of the following conditions is satisfied

- $h(x, y)$  is bounded and  $F \times F$ -almost everywhere continuous,
- $h(x, y)$  is bounded and the process  $(X_n)_{n \geq 1}$  is absolutely regular.

Under the same conditions, but with boundedness of  $h(x, y)$  replaced by uniform integrability of the set  $\{h(X_i, X_j) : i, j \geq 1\}$ , Borovkova, Burton and Dehling [12] established a weak law of large numbers, i.e. convergence in (4.1) in probability.

The conditions in Aaronson et al. are quite restrictive because one either needs a.e. continuity of the kernel or absolute regularity of the underlying process. In this section, we will prove a  $U$ -statistic weak law of large numbers under conditions that are in between the two sets of conditions mentioned above, and that can be verified for examples arising in dynamical systems.

**Theorem 6.** *Let  $(X_n)_{n \in \mathbf{N}}$  be a 1-approximating functional of a stationary and absolutely regular process with summable constants  $(a_l)_{l \geq 0}$ . Let  $F$  denote the distribution of  $X_1$ . Suppose, moreover, that  $h : \mathbf{R}^2 \rightarrow \mathbf{R}$  is a measurable symmetric function satisfying the 1-Lipschitz condition (2.10), and that the family of random variables  $\{h(X_i, X_j) : i, j \geq 1\}$  is uniformly integrable. Then*

$$U_n(h) = \frac{1}{n(n-1)} \sum_{1 \leq i \neq j \leq n} h(X_i, X_j) \longrightarrow \int_{\mathbf{R}^2} h(x, y) dF(x) dF(y)$$

in probability, as  $n \rightarrow \infty$ .

*Proof.* Let  $\epsilon > 0$  be given. Then uniform integrability of  $\{h(X_i, X_j) : i, j \geq 1\}$  implies that there exists  $\delta > 0$  such that

$$(4.2) \quad E|h(X_i, X_j)1_B| \leq \epsilon$$

holds for all measurable sets  $B$  with  $P(B) < \delta$ . Choose moreover  $\delta$  so small that  $\phi(\delta) < \epsilon$ , where  $\phi$  is the function arising in the formulation of the 1-continuity condition (2.10). According to Theorem 3 the process  $(X_n)_{n \in \mathbf{N}}$  is nearly regular, i.e. there exists an integer  $m$  such that the  $(m, N)$ -blocks  $(B_s)_{s \geq 1}$  can be coupled to an i.i.d. sequence of blocks  $(B'_s)_{s \geq 1}$  with the same marginals and in such a way that

$$(4.3) \quad P(\|B_s - B'_s\| \geq \delta) \leq \delta/2.$$

Now take  $N$  so big that  $\frac{m}{m+N} < \epsilon$ . Note that, if  $n$  is the sample size, the number of full  $(m, N)$ -blocks in the sample is  $p = \lfloor \frac{n}{m+N} \rfloor$ .

In the rest of the proof we will show that the random variables in the small separating blocks of length  $m$  can be neglected and that the error introduced by replacing  $B_s$  by  $B'_s$  is negligible. The main term will then be a  $U$ -statistic with independent vector valued inputs  $(B'_s)_{s \geq 1}$  that can be treated by Hoeffding's classical  $U$ -statistic law of large numbers. To this end we define a new kernel  $H : \mathbf{R}^N \times \mathbf{R}^N \rightarrow \mathbf{R}$  by

$$H(\xi, \eta) := \frac{1}{N^2} \sum_{1 \leq i, j \leq N} h(x_i, y_j)$$

where  $\xi = (x_1, \dots, x_N)$  and  $\eta = (y_1, \dots, y_N)$ . From (4.2) we can infer that, for  $k \neq l$ ,

$$(4.4) \quad E|H(B_k, B_l)|1_B \leq \epsilon$$

for all sets  $B$  with  $P(B) < \delta$ .

Independence of  $B'_k$  and  $B'_l$  implies that  $EH(B'_k, B'_l) = \int \int h(x, y) dF(x) dF(y) =: \theta(F)$  for all  $k \neq l$ . Thus by the  $U$ -statistics law of large numbers for independent observations

$$(4.5) \quad \frac{1}{p(p-1)} \sum_{1 \leq k \neq l \leq p} H(B'_k, B'_l) \rightarrow \theta(F)$$

almost surely and in  $L_1$ .

We denote by  $I_s$  the set of indices in the block  $B_s$ , and by  $J_s$  the set of indices between the blocks  $B_s$  and  $B_{s+1}$ . To estimate the difference between  $\sum H(B_k, B_l)$

and  $\sum H(B'_k, B'_l)$ , we write

$$\begin{aligned}
 |H(B_k, B_l) - H(B'_k, B'_l)| &= \frac{1}{N^2} \sum_{i \in I_k} \sum_{j \in I_l} |h(X_i, X_j) - h(X'_i, X'_j)| \\
 &\leq \frac{1}{N^2} \sum_{i \in I_k} \sum_{j \in I_l} |h(X_i, X_j) - h(X'_i, X'_j)| \\
 (4.6) \qquad &+ \frac{1}{N^2} \sum_{i \in I_k} \sum_{j \in I_l} |h(X'_i, X_j) - h(X'_i, X'_j)|.
 \end{aligned}$$

Note that  $|X_i - X'_i| \leq \|B_k - B'_k\|$  holds for  $i \in I_k$ . Hence by the 1-continuity property (2.10) of  $h(x, y)$  we get  $E|h(X_i, X_j) - h(X'_i, X'_j)|1_{\{\|B_k - B'_k\| \leq \delta, \|B_l - B'_l\| \leq \delta\}} \leq \phi(\delta)$ , for all  $i \in I_k, j \in I_l$ . Similarly,  $E|h(X'_i, X_j) - h(X'_i, X'_j)|1_{\{\|B_k - B'_k\| \leq \delta, \|B_l - B'_l\| \leq \delta\}} \leq \phi(\delta)$ , again for all  $i \in I_k, j \in I_l$ . Note that in both cases, (2.10) may be applied—in the first case because the joint distribution of  $(X_i, X_j)$  is a 2-dimensional marginal of the underlying process and in the second case, because  $X'_i$  is independent of  $X'_j$ . These inequalities together with (4.6) imply

$$E |H(B_k, B_l) - H(B'_k, B'_l)| 1_{\{\|B_k - B'_k\| \leq \delta, \|B_l - B'_l\| \leq \delta\}} \leq 2\phi(\delta) \leq 2\epsilon.$$

By (4.3) we have  $P(\|B_k - B'_k\| \geq \delta \text{ or } \|B_l - B'_l\| \geq \delta) \leq \delta$ , and thus (4.4) implies

$$E (|H(B_k, B_l) - H(B'_k, B'_l)| 1_{\{\|B_k - B'_k\| \geq \delta \text{ or } \|B_l - B'_l\| \geq \delta\}}) \leq \epsilon.$$

Taking the last two inequalities together and averaging over all indices  $(k, l)$  with  $1 \leq k \neq l \leq p$ , we obtain the following estimate for the difference between the  $U$ -statistic based on  $(B_k)_k$  and the one based on  $(B'_k)_k$ :

$$(4.7) \quad E \left| \frac{1}{p(p-1)} \sum_{1 \leq k \neq l \leq p} H(B_k, B_l) - \frac{1}{p(p-1)} \sum_{1 \leq k \neq l \leq p} H(B'_k, B'_l) \right| \leq 3\epsilon.$$

Moreover  $|\frac{1}{p(p-1)} - \frac{N^2}{n(n-1)}| \leq \frac{2\epsilon}{p(p-1)}$  for  $p$  large enough, and thus

$$E \left| \frac{1}{p(p-1)} \sum_{1 \leq k \neq l \leq p} H(B_k, B_l) - \frac{N^2}{n(n-1)} \sum_{1 \leq k \neq l \leq p} H(B_k, B_l) \right| \leq 2C_0\epsilon,$$

where  $C_0 = \sup_{k,l} E|H(B_k, B_l)| \leq \sup_{i,j} E|h(X_i, X_j)| < \infty$ . This last estimate together with (4.7) and  $L_1$ -convergence in (4.5) shows that for  $n$  large enough we get

$$(4.8) \quad E \left| \frac{N^2}{n(n-1)} \sum_{1 \leq k \neq l \leq p} H(B_k, B_l) - \theta(F) \right| \leq C\epsilon.$$

It remains to show that the original  $U$ -statistic  $U_n(h)$  is close to the block  $U$ -statistic  $\frac{N^2}{n(n-1)} \sum_{1 \leq k \neq l \leq p} H(B_k, B_l)$ . Let  $i_p$  denote the last index in the block  $B_p$ . Then

we can decompose  $U_n(h)$  as follows:

$$\begin{aligned} \sum_{1 \leq i \neq j \leq n} h(X_i, X_j) &= \sum_{1 \leq k \neq l \leq p} \sum_{i \in I_k, j \in I_l} h(X_i, X_j) + \sum_{k=1}^p \sum_{i, j \in I_k, i \neq j} h(X_i, X_j) \\ &+ 2 \sum_{i \in J_k, j \in I_l, j = ni + 1}^{n_l + N} h(X_i, X_j) + \sum_{1 \leq k \neq l \leq p} \sum_{i \in J_k, j \in J_l} h(X_i, X_j) \\ &+ \sum_{k=1}^p \sum_{i, j \in J_k, i \neq j} h(X_i, X_j) + \sum_{i=i_p+1}^n \sum_{j=1}^n h(X_i, X_j) \\ &+ \sum_{i=1}^{i_p} \sum_{j=i_p+1}^n h(X_i, X_j) \end{aligned}$$

Note that the first term on the r.h.s. equals  $N^2 \sum_{1 \leq k \neq l \leq p} H(B_k, B_l)$ . A careful study of the index set now shows that

$$\begin{aligned} E \left| \sum_{1 \leq i \neq j \leq n} h(X_i, X_j) - \sum_{1 \leq k \neq l \leq p} \sum_{i \in I_k, j \in I_l} h(X_i, X_j) \right| \\ \leq C_0(pN^2 + 2p^2mN + p^2mN + p^2m^2 + 2n(m + N)) \end{aligned}$$

where  $C_0 = \sup_{i,j} E|h(X_i, X_j)|$ . As  $p \leq n/N$  and  $m \leq \epsilon N$ , the r.h.s. of the above inequality is bounded by  $C(\epsilon + N/n)n^2$  and hence

$$E \left| \frac{1}{n(n-1)} \sum_{1 \leq i \neq j \leq n} h(X_i, X_j) - \frac{1}{n(n-1)} \sum_{1 \leq k \neq l \leq p} N^2 H(B_k, B_l) \right| \leq C\epsilon$$

for  $n$  large enough. This, together with (4.8), proves the theorem. □

**4.2. Central limit theorem.** The central limit theorem for  $U$ -statistics was first established by Hoeffding [29], in the case where the underlying observations  $(X_n)_{n \in \mathbf{N}}$  are independent and identically distributed. Yoshihara [49] proved the CLT in the absolutely regular case, and Denker and Keller [20], [21] extended this to functionals of absolutely regular processes. Here we shall apply our techniques to the latter case, and prove a CLT for  $U$ -statistics of functionals of absolutely regular processes under a set of conditions different from those in Denker and Keller [21].

A crucial tool in our analysis are estimates that bound the difference between integrals with respect to the joint distribution and the product distribution of disjoint sets of random variables,  $X_{i_1}, \dots, X_{i_j}$  and  $X_{i_{j+1}}, \dots, X_{i_l}$ . Such a bound was first established by Yoshihara [49] for absolutely regular processes  $(X_n)_{n \geq 1}$ , and we quote his result here for later reference. Let  $i_1 < i_2 < \dots < i_l$  be integers and let  $g : \mathbf{R}^l \rightarrow \mathbf{R}$  be a measurable function. For  $j \in \{1, \dots, l-1\}$  we denote by  $E_{X_{i_1}, \dots, X_{i_j}} g(X_{i_1}, \dots, X_{i_j}, \dots, X_{i_l})$  the expectation of  $g(X_{i_1}, \dots, X_{i_j}, \dots, X_{i_l})$  taken with respect to the random variables  $X_{i_1}, \dots, X_{i_j}$ , with the remaining variables kept fixed.

**Lemma 4.1** (Yoshihara). *Let  $(X_n)_{n \geq 1}$  be an absolutely regular process with mixing coefficients  $(\beta_n)_{n \geq 0}$ , let  $1 \leq i_1 < i_2 < \dots < i_l$  be integers and  $j \in \{1, \dots, l-1\}$ .*

Let  $g : \mathbf{R}^l \rightarrow \mathbf{R}$  be a measurable function satisfying

$$\begin{aligned} \left( E_{X_{i_1}, \dots, X_{i_l}} |g(X_{i_1}, \dots, X_{i_l})|^r \right)^{1/r} &\leq M, \\ \left( E_{X_{i_1}, \dots, X_{i_j}} E_{X_{i_{j+1}}, \dots, X_{i_l}} |g(X_{i_1}, \dots, X_{i_l})|^r \right)^{1/r} &\leq M, \end{aligned}$$

for some  $M > 0$ ,  $r > 1$ . Then

$$\left| E_{X_{i_1}, \dots, X_{i_l}} g(X_{i_1}, \dots, X_{i_l}) - E_{X_{i_1}, \dots, X_{i_j}} E_{X_{i_{j+1}}, \dots, X_{i_l}} g(X_{i_1}, \dots, X_{i_l}) \right| \leq 4M\beta_d^{1/s},$$

where  $d = |i_{j+1} - i_j|$  and  $\frac{1}{r} + \frac{1}{s} = 1$ .

**Lemma 4.2.** Let  $(X_n)_{n \in \mathbf{N}}$  be a 1-approximating functional with constants  $(a_k)_{k \geq 1}$  of an absolutely regular process with mixing coefficients  $(\beta_k)_{k \geq 1}$ , and let  $h(x, y)$  be 1-continuous in the sense of (2.10). Assume that for some constants  $r > 1$  and  $M < \infty$  we have

$$(4.9) \quad (E_{X_0, X_k} |h(X_0, X_k)|^r)^{1/r} \leq M,$$

$$(4.10) \quad (E_{X_0} E_{X_k} |h(X_0, X_k)|^r)^{1/r} \leq M,$$

for all  $k \in \mathbf{N}$ . Then

$$(4.11) \quad |Eh(X_0, X_k) - E_{X_0} E_{X_k} h(X_0, X_k)| \leq 4M(\beta_{\lfloor k/3 \rfloor} + \alpha_{\lfloor k/3 \rfloor})^{1/s} + 2\phi(\alpha_{\lfloor k/3 \rfloor}),$$

where  $s$  is such that  $\frac{1}{r} + \frac{1}{s} = 1$ .

*Proof.* Let  $(X'_n)_{n \in \mathbf{Z}}$  and  $(X''_n)_{n \in \mathbf{Z}}$  be copies of  $(X_n)_{n \in \mathbf{Z}}$ , as defined in Corollary 2.17. Thus, the pairs of random variables  $(X_0, X_k)$ ,  $(X'_0, X'_k)$  and  $(X''_0, X''_k)$  have the same distribution. Moreover,  $(X_0, X_k)$  is independent of  $(X''_0, X''_k)$  and the properties (ii) and (iii) of Corollary 2.17 hold. Then we can write the l.h.s. of (4.11) as follows

$$\begin{aligned} |E[h(X''_0, X_k) - h(X'_0, X'_k)]| &\leq E|h(X''_0, X_k) - h(X''_0, X'_k)| \\ &\quad + E|h(X''_0, X'_k) - h(X'_0, X'_k)|. \end{aligned}$$

Define the events  $B = \{|X_k - X'_k| \leq \alpha_{\lfloor k/3 \rfloor}\}$  and  $D = \{|X'_0 - X''_0| \leq \alpha_{\lfloor k/3 \rfloor}\}$ , and note that  $P(B) \geq 1 - \beta_{\lfloor k/3 \rfloor} - \alpha_{\lfloor k/3 \rfloor}$  and  $P(D) \geq 1 - \alpha_{\lfloor k/3 \rfloor}$ . Then

$$E|h(X''_0, X_k) - h(X''_0, X'_k)|1_B \leq \phi(\alpha_{\lfloor k/3 \rfloor})$$

by the 1-continuity property of  $h$ . Here we have made use of the fact that  $X''_0$  is independent of  $X_k$  so that (2.10) holds. By Hölder's inequality we get

$$E|h(X''_0, X_k) - h(X''_0, X'_k)|1_{B^c} \leq 2M(P(B^c))^{1/s} \leq 2M(\beta_{\lfloor k/3 \rfloor} + \alpha_{\lfloor k/3 \rfloor})^{1/s}.$$

Moreover,

$$E|h(X''_0, X'_k) - h(X'_0, X'_k)|1_D \leq \phi(\alpha_{\lfloor k/3 \rfloor})$$

where we have used (2.10) which holds here because  $(X'_n)_{n \in \mathbf{N}}$  is a copy of  $(X_n)_{n \in \mathbf{Z}}$ . Again, by Hölder's inequality we get

$$E|h(X''_0, X_k) - h(X''_0, X'_k)|1_{D^c} \leq 2M(P(D^c))^{1/s} \leq 2M(\alpha_{\lfloor k/3 \rfloor})^{1/s}.$$

The last four inequalities together prove the assertion of the lemma. □

The next lemma generalizes (4.11) to functions  $g : \mathbf{R}^l \rightarrow \mathbf{R}$ . Note that symmetry is not required in this case.

**Lemma 4.3.** *Let  $(X_n)_{n \in \mathbf{Z}}$  be a 1-approximating functional with constants  $(a_k)_{k \geq 0}$  of an absolutely regular process with mixing coefficients  $(\beta_k)$ , and let  $g : \mathbf{R}^l \rightarrow \mathbf{R}$  be 1-continuous in the sense of (2.11). Moreover, let  $i_1 < i_2 < \dots < i_l$  be integers and assume that for some  $M < \infty$  and  $j \in \{1, \dots, l-1\}$*

$$(4.12) \quad (\mathbf{E} |g(X_{i_1}, \dots, X_{i_l})|^r)^{1/r} \leq M,$$

$$(4.13) \quad \left( \mathbf{E}_{X_{i_1}, \dots, X_{i_j}} \mathbf{E}_{X_{i_{j+1}}, \dots, X_{i_l}} |g(X_{i_1}, \dots, X_{i_l})|^r \right)^{1/r} \leq M,$$

then

$$(4.14) \quad \begin{aligned} & \left| \mathbf{E} g(X_{i_1}, \dots, X_{i_l}) - \mathbf{E}_{X_{i_1}, \dots, X_{i_j}} \mathbf{E}_{X_{i_{j+1}}, \dots, X_{i_l}} g(X_{i_1}, \dots, X_{i_l}) \right| \\ & \leq 4M(\beta_{\lfloor k/3 \rfloor} + \alpha_{\lfloor k/3 \rfloor})^{1/s} + 2\phi(\alpha_{\lfloor k/3 \rfloor}), \end{aligned}$$

where  $k = |i_{j+1} - i_j|$  and  $\frac{1}{r} + \frac{1}{s} = 1$ .

*Proof.* Without loss of generality we may assume that  $i_j = 0$  and  $i_{j+1} = k$ . Let  $(X'_n)_{n \in \mathbf{Z}}$  and  $(X''_n)_{n \in \mathbf{Z}}$  be copies of  $(X_n)_{n \in \mathbf{Z}}$  as defined in Corollary 2.17. Thus, the  $l$ -tuples  $(X_{i_1}, \dots, X_{i_l})$ ,  $(X'_{i_1}, \dots, X'_{i_l})$ ,  $(X''_{i_1}, \dots, X''_{i_l})$  have the same distributions, and  $(X_{i_1}, \dots, X_{i_l})$  is independent of  $(X''_{i_1}, \dots, X''_{i_l})$ . Moreover, properties (ii) and (iii) of Corollary 2.17 hold. Then we can rewrite and bound the l.h.s. of (4.14) as

$$\begin{aligned} & \left| \mathbf{E} g(X''_{i_1}, \dots, X''_{i_j}, X_{i_{j+1}}, \dots, X_{i_l}) - \mathbf{E} g(X'_{i_1}, \dots, X'_{i_l}) \right| \\ & \leq \left| \mathbf{E} g(X''_{i_1}, \dots, X''_{i_j}, X_{i_{j+1}}, \dots, X_{i_l}) - \mathbf{E} g(X''_{i_1}, \dots, X''_{i_j}, X'_{i_{j+1}}, \dots, X'_{i_l}) \right| \\ & \quad + \left| \mathbf{E} g(X''_{i_1}, \dots, X''_{i_j}, X'_{i_{j+1}}, \dots, X'_{i_l}) - \mathbf{E} g(X'_{i_1}, \dots, X'_{i_l}) \right|. \end{aligned}$$

Define the events  $B = \{\sum_{\nu=j+1}^l |X_{i_\nu} - X'_{i_\nu}| \leq \alpha_{\lfloor \frac{k}{3} \rfloor}\}$  and  $D = \{\sum_{\nu=1}^j |X'_{i_\nu} - X''_{i_\nu}| \leq \alpha_{\lfloor \frac{k}{3} \rfloor}\}$  and note that  $P(B) \geq 1 - \beta_{\lfloor \frac{k}{3} \rfloor} - \alpha_{\lfloor \frac{k}{3} \rfloor}$  and  $P(D) \geq 1 - \alpha_{\lfloor \frac{k}{3} \rfloor}$ . Now,

$$\mathbf{E} \left| g(X''_{i_1}, \dots, X''_{i_j}, X_{i_{j+1}}, \dots, X_{i_l}) - g(X''_{i_1}, \dots, X''_{i_j}, X'_{i_{j+1}}, \dots, X'_{i_l}) \right| 1_B \leq \phi(\alpha_{\lfloor \frac{k}{3} \rfloor}),$$

by the 1-continuity property of  $g$ . Here we have made use of the fact that  $X''_{i_1}, \dots, X''_{i_j}$  is independent of  $X_{i_{j+1}}, \dots, X_{i_l}$  so that (2.11) may be applied. By Hölder's inequality we get

$$\begin{aligned} & \mathbf{E} \left| g(X''_{i_1}, \dots, X''_{i_j}, X_{i_{j+1}}, \dots, X_{i_l}) - g(X''_{i_1}, \dots, X''_{i_j}, X'_{i_{j+1}}, \dots, X'_{i_l}) \right| 1_{B^c} \\ & \leq 2M (P(B^c))^{1/s} \leq 2M \left( \beta_{\lfloor \frac{k}{3} \rfloor} + \alpha_{\lfloor \frac{k}{3} \rfloor} \right)^{1/s}. \end{aligned}$$

In the same way, we find by the 1-continuity property of  $g$

$$\mathbf{E} \left| g(X''_{i_1}, \dots, X''_{i_j}, X'_{i_{j+1}}, \dots, X'_{i_l}) - g(X'_{i_1}, \dots, X'_{i_j}, X'_{i_{j+1}}, \dots, X'_{i_l}) \right| 1_D \leq \phi(\alpha_{\lfloor \frac{k}{3} \rfloor}),$$

In this case we may apply (2.11) because  $X'_{i_1}, \dots, X'_{i_j}, X'_{i_{j+1}}, \dots, X'_{i_l}$  is a copy of  $X_{i_1}, \dots, X_{i_l}$ . Again, using Hölder's inequality, we obtain

$$\begin{aligned} & \mathbf{E} \left| g(X''_{i_1}, \dots, X''_{i_j}, X_{i_{j+1}}, \dots, X_{i_l}) - g(X'_{i_1}, \dots, X'_{i_j}, X'_{i_{j+1}}, \dots, X'_{i_l}) \right| 1_{D^c} \\ & \leq 2M \left( \alpha_{\lfloor \frac{k}{3} \rfloor} \right)^{1/s}. \end{aligned}$$

Putting all inequalities together, we finally obtain (4.14). □



**Theorem 7.** *Let  $(X_n)_{n \in \mathbf{N}}$  be a 1-approximating functional of an absolutely regular process with mixing coefficients  $(\beta_k)$ , and let  $h$  be a bounded 1-continuous kernel. Suppose that the sequences  $(\beta_k)_{k \geq 1}$ ,  $(\alpha_k)_{k \geq 1}$  and  $(\phi(\alpha_k))_{k \geq 1}$ , satisfy the following summability condition:*

$$(4.15) \quad \sum_{k=1}^{\infty} k^2(\beta_k + \alpha_k + \phi(\alpha_k)) < \infty.$$

Then the series

$$(4.16) \quad \sigma^2 = \text{Var}(h_1(X_0))^2 + 2 \sum_{k=1}^{\infty} \text{Cov}(h_1(X_0), h_1(X_k))$$

converges absolutely and, as  $n \rightarrow \infty$ ,

$$\sqrt{n}(U_n - \theta) \xrightarrow{d} \mathcal{N}(0, 4\sigma^2).$$

*Proof.* We make use of the Hoeffding decomposition for  $U$ -statistics,

$$U_n = \theta(F) + \frac{2}{n} \sum_{i=1}^n (h_1(X_i) - \theta(F)) + \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} J(X_i, X_j)$$

where  $\theta(F) = \int h(x, y)dF(x)dF(y)$ ,  $h_1(x) = \int h(x, y)dF(y)$  and  $J(x, y) = h(x, y) - h_1(x) - h_1(y) + \theta$  (see Hoeffding [29]). The random variables  $g(X_i) = h_1(X_i) - \theta(F)$  are bounded and have mean zero. As  $h$  is 1-continuous, Proposition 2.11 states that  $(g(X_n))_{n \geq 1}$  is itself a 1-approximating functional of  $(Z_n)_{n \geq 1}$ . Moreover, the summability condition (4.15) on the corresponding sequence  $(\phi(\alpha_i))$  is that required by Theorem 4. Hence, from Theorem 4 it follows that

$$\frac{2}{\sqrt{n}} \sum_{i=1}^n (h_1(X_i) - \theta(F)) \longrightarrow \mathcal{N}(0, 4\sigma^2),$$

where  $\sigma^2$  is given by (4.16). Then the statement of the theorem will follow, if we can prove that  $\sqrt{n}R_n \rightarrow 0$  in probability, where  $R_n$  is defined by

$$R_n = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} J(X_i, X_j).$$

That this is indeed the case is established in the following lemma. □

**Lemma 4.4.** *Under the conditions of Theorem 7,*

$$(4.17) \quad \sup_n \mathbb{E} \left\{ \frac{1}{n} \sum_{1 \leq i < j \leq n} J(X_i, X_j) \right\}^2 < \infty,$$

and hence  $\sqrt{n}R_n \rightarrow 0$  in probability, as  $n \rightarrow \infty$ .

*Proof.* We can write

$$\begin{aligned}
 \mathbb{E} \left( \sum_{1 \leq i < j \leq n} J(X_i, X_j) \right)^2 &= \sum_{1 \leq i_1 < j_1 \leq n} \sum_{1 \leq i_2 < j_2 \leq n} \mathbb{E} (J(X_{i_1}, X_{j_1}) J(X_{i_2}, X_{j_2})) \\
 (4.18) \qquad &= \sum_{\substack{1 \leq i_1 < j_1, i_2 < j_2 \leq n \\ i_1 \neq i_2 \text{ or } j_1 \neq j_2}} \mathbb{E} (J(X_{i_1}, X_{j_1}) J(X_{i_2}, X_{j_2})) \\
 &\quad + \sum_{\substack{1 \leq i_1 < j_1, i_2 < j_2 \leq n \\ i_1 = i_2 \text{ and } j_1 = j_2}} \mathbb{E} (J(X_{i_1}, X_{j_1}) J(X_{i_2}, X_{j_2})).
 \end{aligned}$$

Note that  $J(x, y)$  is a degenerate kernel, i.e. that  $\int J(x, y) dF(x) = 0$  for all  $y \in \mathbf{R}$ . Hence

$$E_{X_{j_2}} (J(X_{i_1}, X_{i_2}) J(X_{j_1}, X_{j_2})) = \int_{\mathbf{R}} J(X_{i_1}, X_{j_1}) J(X_{i_2}, y) dF(y) = 0,$$

and similarly for  $E_{X_{i_1}}, E_{X_{i_2}}, E_{X_{j_1}}$ . As both  $h$  and  $h_1$  are bounded and 1-continuous, this also holds for  $g(x_1, x_2, x_3, x_4) = J(x_1, x_2) J(x_3, x_4)$ . Thus may apply Lemma 4.3 with  $r = \infty$  and get for  $i_1 \leq i_2 \leq j_1 < j_2$

$$\begin{aligned}
 &|\mathbb{E} (J(X_{i_1}, X_{j_1}) J(X_{i_2}, X_{j_2})) - E_{X_{i_1}, X_{i_2}, X_{j_1}} E_{X_{j_2}} (J(X_{i_1}, X_{j_1}) J(X_{i_2}, X_{j_2}))| \\
 &\leq 16(\beta_k + \alpha_k) + 2\phi(\alpha_k),
 \end{aligned}$$

where  $k = \lfloor |j_2 - j_1|/3 \rfloor$ .

In the elements of the first sum on the r.h.s of (4.18) at least one index is different from all others, say,  $j_2$ , and suppose  $i_1 \leq i_2 \leq j_1 < j_2$ . Let  $d_i$  be the  $i$ th largest difference between consecutive indices. If  $d_1 = j_2 - j_1$  then

$$|\mathbb{E} (J(X_{i_1}, X_{j_1}) J(X_{i_2}, X_{j_2}))| \leq 16(\beta_{\lfloor d_1/3 \rfloor} + \alpha_{\lfloor d_1/3 \rfloor}) + 2\phi(\alpha_{\lfloor d_1/3 \rfloor}).$$

Then

$$\begin{aligned}
 &\sum_{\substack{1 \leq i_1 \leq i_2 \leq j_1 < j_2 \leq n \\ d_1 = j_2 - j_1}} |\mathbb{E} (J(X_{i_1}, X_{j_1}) J(X_{i_2}, X_{j_2}))| \\
 &\leq 16 \sum_{\substack{1 \leq i_1 \leq i_2 \leq j_1 < j_2 \leq n \\ d_1 = j_2 - j_1}} (\beta_{\lfloor d_1/3 \rfloor} + \alpha_{\lfloor d_1/3 \rfloor} + \phi(\alpha_{\lfloor d_1/3 \rfloor})) \\
 &\leq 16n \sum_{k=1}^n k^2 (\beta_k + \alpha_k + \phi(\alpha_k)).
 \end{aligned}$$

If  $d_1$  is not  $j_2 - j_1$  we apply Lemma 4.3 twice to obtain

$$\begin{aligned}
 \mathbb{E} [J(X_{i_1}, X_{j_1}) J(X_{i_1}, X_{j_2})] &\leq 16(\beta_{\lfloor d_1/3 \rfloor} + \alpha_{\lfloor d_1/3 \rfloor}) + 2\phi(\alpha_{\lfloor d_1/3 \rfloor}) \\
 &\quad + 16(\beta_{\lfloor d_2/3 \rfloor} + \alpha_{\lfloor d_2/3 \rfloor}) + 2\phi(\alpha_{\lfloor d_2/3 \rfloor})
 \end{aligned}$$

and then

$$\begin{aligned} & \sum_{\substack{1 \leq i_1 \leq i_2 \leq j_1 < j_2 \leq n \\ d_1 \neq j_2 - j_1}} \mathbb{E}[J(X_{i_1}, X_{j_1}) J(X_{i_2}, X_{j_2})] \\ & \leq 16 \sum_{\substack{1 \leq i_1 \leq i_2 \leq j_1 < j_2 \leq n \\ d_1 \neq j_2 - j_1}} (\beta_{\lfloor d_1/3 \rfloor} + \alpha_{\lfloor d_1/3 \rfloor} + \phi(\alpha_{\lfloor d_1/3 \rfloor})) \\ & \quad + (\beta_{\lfloor d_2/3 \rfloor} + \alpha_{\lfloor d_2/3 \rfloor} + \phi(\alpha_{\lfloor d_2/3 \rfloor})) \\ & \leq 32n^2 \sum_{k=1}^n k(\beta_k + \alpha_k + \phi(\alpha_k)). \end{aligned}$$

Estimating the sums in the other cases in the same way we get that the first sum is bounded by

$$\sum_{\substack{1 \leq i_1 < j_1, i_2 < j_2 \leq n \\ i_1 \neq i_2 \text{ or } j_1 \neq j_2}} \mathbb{E}[J(X_{i_1}, X_{j_1}) J(X_{i_2}, X_{j_2})] \leq Cn^2 \sum_{k=1}^n k(\beta_k + \alpha_k + \phi(\alpha_k)).$$

In the second sum there are at most  $n^2$  terms, all are bounded, hence

$$\sum_{1 \leq i_1 \neq j_1 \leq n} \mathbb{E}[J(X_{i_1}, X_{j_1}) J(X_{i_1}, X_{j_1})] \leq Cn^2.$$

Combining all the estimates above with the summability conditions on  $\beta_k, \alpha_k$  and  $\phi_k$ , we get that

$$\sum_{1 \leq i_1 < j_1 \leq n} \sum_{1 \leq i_2 < j_2 \leq n} \mathbb{E}[J(X_{i_1}, X_{j_1}) J(X_{i_2}, X_{j_2})] \leq Cn^2.$$

From here, we obtain directly (4.17), which again implies weak convergence of  $\sqrt{n}R_n$  to 0. □

### 5. U-PROCESSES

This section is devoted to the study of the empirical distribution function of data sets of the type  $\{h(X_i, X_j), 1 \leq i < j \leq n\}$ , for some symmetric kernel  $h : \mathbf{R}^2 \rightarrow \mathbf{R}$ . Our main result will be an invariance principle for the empirical process when the underlying observations are functionals of an absolutely regular process.

**5.1. Motivation and examples.** One of the main motivations for our study of  $U$ -statistics of absolutely regular processes was the empirical correlation integral

$$C_n(r) = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} 1_{\{\|X_i - X_j\| \leq r\}}.$$

So far we have always viewed  $r$  as a fixed parameter, in which case  $C_n(r)$  is an ordinary  $U$ -statistic. However, already the Grassberger-Procaccia estimator for the dimension estimation involves the sample correlation integral evaluated at several points  $r_1, \dots, r_m$ . Therefore, in this section we will view  $C_n(r)$  as a stochastic process, indexed by  $r \in [0, r_0]$ , for some  $r_0 > 0$  and we will study the asymptotic properties of this stochastic process.

**Definition 5.1.** Let  $(X_n)_{n \geq 1}$  be a stationary stochastic process and let  $h : \mathbf{R}^m \rightarrow \mathbf{R}$  be a measurable symmetric function. Then

$$U_n(t) := \frac{1}{\binom{n}{m}} \sum_{1 \leq i_1 < \dots < i_m \leq n} \mathbf{1}_{\{h(X_{i_1}, \dots, X_{i_m}) \leq t\}}, \quad t \in \mathbf{R},$$

is called a  $U$ -statistic empirical distribution function or an empirical distribution function of  $U$ -statistic type.

Observe that  $U_n(t)$  is the distribution function of the empirical  $U$ -statistic distribution  $\frac{1}{\binom{n}{m}} \sum \delta_{h(X_{i_1}, \dots, X_{i_m})}$ , obtained by placing equal mass at each of the points  $h(X_{i_1}, \dots, X_{i_m})$ ,  $1 \leq i_1 < \dots < i_m \leq n$ . Also note that for fixed  $t$ ,  $U_n(t)$  is simply a  $U$ -statistic with kernel  $\mathbf{1}_{\{h(X_{i_1}, \dots, X_{i_m}) \leq t\}}$ .

**Example 5.1.** (i) For  $m = 1$  and  $h(x) = x$ , the corresponding  $U$ -statistic empirical distribution function is just the ordinary empirical distribution function

$$F_n(t) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{X_i \leq t\}}$$

of the data  $X_1, \dots, X_n$ .

(ii) The sample correlation integral is a  $U$ -statistic empirical distribution function with kernel  $h(x, y) = \|x - y\|$ .

(iii) This occurs also in many other applications, for instance, in the analysis of the archaeological data, so-called “ley hunting”. Suppose  $X_1, \dots, X_n$  are observations from an unknown distribution  $G$  on  $\mathbf{R}^2$ , and we are interested in testing randomness against presence of some collinearities in the data. For this Broadbent and Heaton (see Silverman and Brown [43]) suggested the following approach: denote  $\alpha(x, y, z) =$  “the largest angle of the triangle  $xyz$ ”, and study the statistics

$$T_n(\epsilon) = \binom{n}{3}^{-1} \sum_{\substack{(i,j,k) \\ 1 \leq i < j < k \leq n}} \mathbf{1}(\alpha(X_i, X_j, X_k) > \pi - \epsilon)$$

which is a  $U$ -statistic of degree 3 estimating

$$\theta_G(\epsilon) = P(\alpha(X, Y, Z) > \pi - \epsilon),$$

where  $X, Y, Z$  are chosen independently according to the distribution  $G$ . In this case we are also interested in the behaviour of  $T_n(\epsilon)$  not for a fixed  $\epsilon$ , but for  $\epsilon$  on some interval  $(0, \epsilon_0]$ .

The  $U$ -statistic empirical distribution function is the natural estimator for

$$U(t) = P(h(Y_1, \dots, Y_m) \leq t)$$

where  $Y_1, \dots, Y_m$  are independent random variables, with the same marginal distribution function as the process  $(X_n)_{n \geq 1}$ . In the case of independent observations,  $U_n(t)$  has many nice properties. Among other things,  $U_n(t)$  is unbiased, moreover the minimum variance unbiased estimator of  $U(t)$  and also consistent. The next theorem shows that consistency also holds in the stationary, ergodic case. In the formulation of this and of following theorems, we will restrict ourselves to bivariate  $U$ -statistics, i.e. to the case  $m = 2$ .

**Theorem 8.** *Let  $(X_n)_{n \in \mathbf{Z}}$  be a stationary ergodic process, and let  $h : \mathbf{R}^2 \rightarrow R$  be a kernel satisfying  $\int \int 1_{\{h(x,y)=t\}} dF(x)dF(y) = 0$  for all  $t$ . Then*

$$\sup_{t \in \mathbf{R}} |U_n(t) - U(t)| \rightarrow 0$$

as  $n \rightarrow \infty$ .

*Proof.* This result can be established in the same way as the Glivenko-Cantelli lemma for the ordinary empirical distribution function. The only difference is that the  $U$ -statistic ergodic theorem of Aaronson et al. [1] has to be employed to get pointwise convergence of  $U_n(t)$  towards  $U(t)$ .  $\square$

In many statistical applications, approximations for the distribution of the process  $\{U_n(t) - U(t); t \in \mathbf{R}\}$  are required. This is e.g. the case if one wants to determine confidence bands for  $\{U(t); t \in \mathbf{R}\}$ . This motivates the study of the empirical process of  $U$ -statistics structure  $\{W_n(t); t \in \mathbf{R}\}$ , defined by

$$W_n(t) = \sqrt{n}(U_n(t) - U(t)).$$

Alternatively,  $W_n(t)$  is also called a  $U$ -process. Weak convergence of the empirical processes of  $U$ -statistics structure to a Gaussian process was shown by Silverman [42] and Serfling [41] for the case of i.i.d. random variables. Further properties of  $U$ -processes were investigated by Dehling, Denker and Philipp [17], Helmers, Janssen and Serfling [28], Nolan and Pollard [36], and others. For weakly dependent observations, weak convergence of  $U$ -processes has been established by Arcones and Yu [3] and Borovkova [10]. Their results hold for absolutely regular processes. In this section we will establish an invariance principle for functionals of an absolutely regular process.

As noted above, for fixed  $t$ ,  $U_n(t)$  is simply a  $U$ -statistic, with kernel given by

$$(5.1) \quad h(x, y; t) = 1_{\{h(x,y) \leq t\}}.$$

More generally, one can study  $U$ -processes indexed by any class of kernels  $\mathcal{H}$ . In what follows we will consider 1-dimensional classes of kernels that can be parametrized by a real parameter  $t \in [0, 1]$ . We will denote the kernels by  $h(x, y; t)$ , but these are not necessarily of the form (5.1).

Let

$$\mathcal{H} = \{h(x, y; t) : t \in [0, 1]\}$$

denote a class of kernel functions, and let  $(X_n)_{n \in \mathbf{Z}}$  be a stationary stochastic process. We then define the processes  $(\{U_n(t), t \in [0, 1]\})_{n \geq 1}$ ,  $\{U(t), t \in [0, 1]\}$  and  $(\{W_n(t), t \in [0, 1]\})_{n \geq 1}$  by

$$\begin{aligned} U_n(t) &= \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} h(X_i, X_j, t), \\ U(t) &= \int h(x, y, t) dF(x) dF(y), \\ W_n(t) &= \sqrt{n}(U_n(t) - U(t)). \end{aligned}$$

In our analysis, we need several assumptions on the class of kernels. We first assume that for all  $x, y \in \mathbf{R}$  and  $t \in [0, 1]$  we have

- (A1)  $0 \leq h(x, y, t) \leq 1, \quad h(x, y, 0) = 0,$
- (A2)  $h(x, y, t)$  is increasing in  $t$ , for fixed  $x, y \in \mathbf{R}$ .

Define

$$h_1(x, t) = \int_{\mathbf{R}^k} h(x, y, t) dF(y).$$

Note that, if the kernel function  $h$  satisfies the conditions (A1) and (A2), then the function  $h_1$  satisfies analogous conditions, namely  $0 \leq h_1(x, t) \leq 1$ ,  $h_1(x, 0) = 0$  and it is increasing in  $t$ .

Moreover, we have to impose uniform Lipschitz conditions on  $E_{X_0, X_k} h(X_0, X_k, t)$  and  $E_{X_0} E_{X_k} h(X_0, X_k, t)$ . We assume that there exists a constant  $C > 0$  such that

$$(5.2) \quad |U(t) - U(s)| \leq C|t - s|,$$

$$(5.3) \quad |(E h(X_0, X_k, t) - E h(X_0, X_k, s))| \leq C|t - s|$$

hold for all  $s, t \in [0, 1]$  and  $k \geq 1$ .

In this section we want to study the behaviour of the  $U$ -process called  $(\{W_n(t), 0 \leq t \leq 1\})_{n \geq 1}$  when the underlying sequence is absolutely regular, resp. a functional of an absolutely regular process. Under some regularity conditions, we prove weak convergence of the  $U$ -process to a Gaussian process with mean 0 and covariance structure

$$(5.4) \quad E(W(s)W(t)) = 4 \operatorname{Cov}(h_1(X_1, s)h_1(X_1, t)) + 4 \left\{ \sum_{k=1}^{\infty} \operatorname{Cov}(h_1(X_1, s)h_1(X_{k+1}, t)) + \operatorname{Cov}(h_1(X_1, t)h_1(X_{k+1}, s)) \right\}.$$

We can now formulate our main result.

**Theorem 9.** *Let  $(X_n)_{n \in \mathbf{Z}}$  be a 1-approximating functional with constants  $(\alpha_k)_{k \geq 0}$  of an absolutely regular process with mixing rate  $(\beta_k)_{k \geq 0}$ . Suppose that the class of kernels  $\mathcal{H}$  satisfies (A1) and (A2), that  $h_t$  is uniformly 1-continuous (i.e. that (2.10) holds with the same  $\phi$ -function for all  $t \in [0, 1]$ ) and that (5.2) and (5.3) hold. Moreover assume that the following summability conditions hold:*

$$(5.5) \quad \sum_{k=0}^{\infty} k^2 (\alpha_k + \beta_k)^{1/2} < \infty,$$

$$(5.6) \quad \sum_{k=0}^{\infty} k^5 \phi(\sqrt{\alpha_k}) < \infty,$$

in addition to (3.13) and (3.14). Then the process  $(W_n = \sqrt{n}(U_n(t) - U(t)), t \in [0, 1])_{n \in \mathbf{N}}$  converges weakly in  $D[0, 1]$  to the mean-zero Gaussian process  $(W(t), t \in [0, 1])$  with covariance structure given by (5.4). Moreover, the series (5.4) converges absolutely, and the limit process  $W$  has continuous sample paths on  $[0, 1]$  with probability 1.

*Proof.* According to Hoeffding’s projection method,  $U_n(t)$  can be decomposed as

$$U_n(t) = U(t) + \frac{2}{n} \sum_{i=1}^n (h_1(X_i, t) - U(t)) + R_n(t),$$

where

$$R_n(t) = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} (h(X_i, X_j, t) - h_1(X_i, t) - h_1(X_j, t) + U(t))$$

is the remainder of  $U_n(t)$ . Then the process  $W_n(t)$  can be written as

$$W_n(t) = \frac{2}{\sqrt{n}} \sum_{i=1}^n (h_1(X_i, t) - U(t)) + \sqrt{n}R_n(t) = W'_n(t) + \sqrt{n}R_n(t),$$

where

$$(5.7) \quad W'_n(t) = \frac{2}{\sqrt{n}} \sum_{i=1}^n (h_1(X_i, t) - U(t)).$$

Note that the process  $\{W'_n(t), 0 \leq t \leq 1\}$  has the form of an empirical process indexed by the functions  $h_1(x, t)$ . Thus, Theorem 5 implies that  $W'_n$  converges weakly in  $D[0, 1]$  to the mean-zero Gaussian process  $W$  with covariance structure given by (5.4).

By Slutsky's theorem, this establishes weak convergence of  $(W_n)_{n \geq 1}$  to the same limit, provided we can show that the remainder terms  $\{\sqrt{n}R_n(t); 0 \leq t \leq 1\}$  converge in probability to 0. This convergence of the remainder term to 0 is the essential part of the proof of weak convergence of  $U$ -processes. We present the proof of it in the rest of this section.

**5.2.  $U$ -process chaining lemma.** The first lemma establishes conditions under which an arbitrary sequence of processes converges to zero in probability uniformly over an interval. Since it does not specify the form of the process or any dependence structure, this lemma can be applied in a more general context and may be of independent interest.

In this lemma we shall use the ordinary Lipschitz-continuity condition, so we shall first remind the reader of the definition of a Lipschitz-continuous function on  $[0, 1]$ . A function  $f$  on  $[0, 1]$  is called Lipschitz-continuous if for all  $s, t \in [0, 1]$  there is a constant  $C$  such that

$$|f(t) - f(s)| \leq C|t - s|.$$

In what follows  $C, C_1, C_2, \dots$  denote some positive constants.

**Lemma 5.2.** *Let  $\{X_n(t), t \in [0, 1]\}_{n \in \mathbf{N}}$  be a sequence of  $\mathbf{R}$ -valued stochastic processes with  $X_n(0) = 0$  a.s. Suppose that for some positive constants  $\alpha, \beta, \gamma$  with  $\alpha \leq 1$  and  $\gamma < 1$  the following two conditions hold*

(i) *There exists a Lipschitz-continuous function  $f$  on  $[0, 1]$  such that*

$$(5.8) \quad \mathbb{E} |X_n(t) - X_n(s)|^2 \leq \frac{|f(t) - f(s)|^\gamma}{n^\beta}, \quad \forall s, t \in [0, 1].$$

(ii) *There exist a Lipschitz-continuous function  $g$ , a monotone Lipschitz-continuous function  $\Lambda$  and stochastic processes  $\{Y_n(t), t \in [0, 1]\}_{n \geq 1}$  satisfying*

$$(5.9) \quad \mathbb{E} |Y_n(t) - Y_n(s)|^r \leq C_1 [(g(t) - g(s))^{1+h} + \frac{1}{n} (g(t) - g(s))^{\frac{1+h}{2}}], \quad \forall s, t \in [0, 1]$$

for some  $r > 0, 0 < h < 1$ , such that for all  $s, t, \delta$  with  $0 \leq s \leq t < s + \delta \leq 1$ :

$$(5.10) \quad |X_n(t) - X_n(s)| \leq |X_n(s + \delta) - X_n(s)| + |Y_n(s + \delta) - Y_n(s)| + |\Lambda(s + \delta) - \Lambda(s)| \cdot n^\alpha.$$

Then

$$\sup_{0 \leq t \leq 1} |X_n(t)| \longrightarrow 0 \text{ in probability,}$$

whenever  $\alpha(1 - \gamma) < \beta$ .

*Proof.* For  $k = 0, 1, 2, \dots, K$  (we shall specify  $K$  later) we introduce refining partitions of  $[0, 1]$  into  $2^k$  subintervals,  $0 = s_0^{(k)} < s_1^{(k)} < \dots < s_{2^k}^{(k)} = 1$ , in such a way that for any  $i = 1, 2, \dots, 2^k$

$$(5.11) \quad |s_i^{(k)} - s_{i-1}^{(k)}| \leq 2^{-k}.$$

Then we get, due to Lipschitz-continuity of  $\Lambda$  that

$$(5.12) \quad |\Lambda(s_i^{(k)}) - \Lambda(s_{i-1}^{(k)})| \leq C \cdot 2^{-k}.$$

For  $t \in [0, 1]$  we denote by  $i_k(t)$  the index of that point of the  $k$ th partition that is closest to  $t$  from the left, i.e. such that  $s_{i_k(t)}^{(k)} \leq t < s_{i_k(t)+1}^{(k)}$ . Note that  $s_{i_k(t)}^{(k)} \leq s_{i_{k+1}(t)}^{(k+1)}$ , as the  $(k+1)$ th partition is a refinement of the  $k$ th. In this way we have obtained a chain

$$0 = s_{i_0(t)}^{(0)} \leq s_{i_1(t)}^{(1)} \leq \dots \leq s_{i_K(t)}^{(K)} \leq t < s_{i_K(t)+1}^{(K)},$$

and we can write

$$X_n(t) = \sum_{k=1}^K \left( X_n(s_{i_k(t)}^{(k)}) - X_n(s_{i_{k-1}(t)}^{(k-1)}) \right) + \left( X_n(t) - X_n(s_{i_K(t)}^{(K)}) \right).$$

Then

$$\begin{aligned} \sup_{0 \leq t \leq 1} |X_n(t)| &\leq \max_{i_1(t)} |X_n(s_{i_1(t)}^{(1)}) - X_n(s_{i_0(t)}^{(0)})| \\ &\quad + \max_{i_1(t), i_2(t)} |X_n(s_{i_2(t)}^{(2)}) - X_n(s_{i_1(t)}^{(1)})| + \dots \\ &\quad + \max_{i_{K-1}(t), i_K(t)} |X_n(s_{i_K(t)}^{(K)}) - X_n(s_{i_{K-1}(t)}^{(K-1)})| \\ &\quad + \sup_{s_{i_K(t)}^{(K)} \leq t < s_{i_K(t)+1}^{(K)}} |X_n(t) - X_n(s_{i_K(t)}^{(K)})|. \end{aligned}$$

By (5.10) and (5.12)

$$\begin{aligned} \sup_{s_{i_K(t)}^{(K)} \leq t < s_{i_K(t)+1}^{(K)}} |X_n(t) - X_n(s_{i_K(t)}^{(K)})| &\leq \max_{i_K(t)} |X_n(s_{i_K(t)+1}^{(K)}) - X_n(s_{i_K(t)}^{(K)})| \\ &\quad + \max_{i_K(t)} |Y_n(s_{i_K(t)+1}^{(K)}) - Y_n(s_{i_K(t)}^{(K)})| \\ &\quad + C \cdot 2^{-K} n^\alpha. \end{aligned}$$

Now, let  $\epsilon > 0$  be given and take  $K = \lceil \alpha \log_2 n + \log_2 \frac{2C}{\epsilon} \rceil + 1$ , so that  $C \cdot 2^{-K} n^\alpha < \frac{\epsilon}{2}$ .



Then, as  $\sum_{k=1}^{\infty} \frac{1}{(k+2)^2} < \frac{1}{2}$ ,

$$\begin{aligned}
 & \mathbb{P} \left( \sup_{0 \leq t \leq 1} |X_n(t)| > \epsilon \right) \\
 & \leq \sum_{k=1}^K \mathbb{P} \left( \max_{i_{k-1}(t), i_k(t)} |X_n(s_{i_k}^{(k)}) - X_n(s_{i_{k-1}}^{(k-1)})| > \frac{\epsilon}{(k+2)^2} \right) \\
 & \quad + \mathbb{P} \left( \max_{i_K(t)} |X_n(s_{i_K}^{(K)}) - X_n(s_{i_K}^{(K)})| > \frac{\epsilon}{(K+3)^2} \right) \\
 (5.13) \quad & \quad + \mathbb{P} \left( \max_{i_K(t)} |Y_n(s_{i_K}^{(K)}) - Y_n(s_{i_K}^{(K)})| > \frac{\epsilon}{(K+4)^2} \right).
 \end{aligned}$$

As the partitions are nested, we have that either  $s_{i_{k-1}(t)}^{(k-1)} = s_{i_k(t)}^{(k)}$  or  $s_{i_{k-1}(t)}^{(k-1)} = s_{i_k(t)+1}^{(k)}$ . Hence by Chebyshev's inequality, (5.8) and (5.11)

$$\begin{aligned}
 (5.14) \quad & \mathbb{P} \left( \max_{i_{k-1}(t), i_k(t)} |X_n(s_{i_k}^{(k)}) - X_n(s_{i_{k-1}}^{(k-1)})| > \frac{\epsilon}{(k+2)^2} \right) \\
 & \leq \sum_{i=1}^{2^k} \mathbb{P} \left( |X_n(s_i^{(k)}) - X_n(s_{i-1}^{(k)})| > \frac{\epsilon}{(k+2)^2} \right) \\
 & \leq \sum_{i=1}^{2^k} \frac{|f(s_i^{(k)}) - f(s_{i-1}^{(k)})| \gamma (k+2)^4}{n^\beta \epsilon^2} \\
 & \leq \frac{C_1 (k+2)^4}{n^\beta \epsilon^2} \sum_{i=1}^{2^k} |s_i^{(k)} - s_{i-1}^{(k)}|^\gamma \leq \frac{C_2 (k+2)^4}{n^\beta \epsilon^2} \cdot 2^{k(1-\gamma)}.
 \end{aligned}$$

In the same way we get

$$(5.15) \quad \mathbb{P} \left( \max_{i_K(t)} |X_n(s_{i_K}^{(K)}) - X_n(s_{i_K}^{(K)})| > \frac{\epsilon}{(K+3)^2} \right) \leq \frac{C_3 (K+3)^4}{n^\beta \epsilon^2} \cdot 2^{K(1-\gamma)}.$$

Similarly, by Markov's inequality and (5.9) we have

$$\begin{aligned}
 (5.16) \quad & \mathbb{P} \left( \max_{i_K(t)} |Y_n(s_{i_K}^{(K)}) - Y_n(s_{i_K}^{(K)})| > \frac{\epsilon}{(K+4)^2} \right) \\
 & \leq \sum_{i=1}^{2^K} \mathbb{P} \left( |Y_n(s_i^{(K)}) - Y_n(s_{i-1}^{(K)})| > \frac{\epsilon}{(K+4)^2} \right) \\
 & \leq \frac{(K+4)^{2r}}{\epsilon^r} \sum_{i=1}^{2^K} E \left| Y_n(s_i^{(K)}) - Y_n(s_{i-1}^{(K)}) \right|^r \\
 & \leq C \frac{(K+4)^{2r}}{\epsilon^r} 2^K \left( \left( \frac{1}{2^K} \right)^{1+h} + \frac{1}{n} \left( \frac{1}{2^K} \right)^{\frac{1+h}{2}} \right) \\
 & \leq \frac{C(K+4)^{2r}}{\epsilon^r} \cdot 2^{-Kh},
 \end{aligned}$$

as  $\frac{1}{n} \leq C \left(\frac{1}{2^K}\right)^{\frac{1+h}{2}}$  by our choice of  $K$  and the conditions imposed on  $h$  and  $\alpha$ . Finally, combining (5.14), (5.15) and (5.16) with conditions on  $\alpha, \beta, \gamma$  we obtain

$$\begin{aligned} \mathbb{P} \left( \sup_{0 \leq t \leq 1} |X_n(t)| > \epsilon \right) &\leq \frac{C_5 K (K+3)^4}{n^\beta \epsilon^2} \cdot 2^{K(1-\gamma)} + \frac{C_4 (K+4)^{2r}}{\epsilon^r} \cdot 2^{-Kh} \\ &\leq \frac{C_6 (K+3)^5}{\epsilon^2} \cdot n^{\alpha(1-\gamma)-\beta} + \frac{C_7 (K+4)^{2r}}{\epsilon^r} \cdot n^{-\alpha h} \rightarrow 0, \end{aligned}$$

because  $(K+3)^p = (\alpha \log_2 n + \log_2 \frac{2C}{\epsilon} + 4)^p = o(n^\tau)$  for any  $\tau, p > 0$ . □

Chaining techniques as explored in the proof of Lemma 5.2 are well-known and widely applied in the theory of empirical processes. In distinction to known results, Lemma 5.2 allows for an extra random term  $Y_n$  in the bound (5.10). This makes Lemma 5.2 applicable to  $U$ -processes where  $Y_n$  will be the first order term in the Hoeffding expansion. The second distinctive feature is the fact that only second moment bounds are required, thus avoiding technically involved higher moment calculations. This is made possible by the extra  $n^{-\beta}$  term in (5.8), an idea exploited earlier by Dehling and Taqqu [18] in their proof of the invariance principle for the empirical process of long-range dependent sequences.

**5.3.  $U$ -processes for functionals.** In this section we will finish the proof of Theorem 9. What remains to be done is to establish the following result.

**Lemma 5.3.** *Let the condition of Theorem 9 be satisfied. Then*

$$\sup_{t \in [0,1]} \sqrt{n} |R_n(t)| \rightarrow 0 \text{ in probability, as } n \rightarrow \infty.$$

In the proof of Lemma 5.3 we will make use of Lemma 5.2, which requires a bound on the increments of  $R_n$ . We formulate this bound as a separate lemma.

**Lemma 5.4.** *Under the conditions of Theorem 9, there exists a positive constant  $C$  such that*

$$(5.17) \quad \mathbb{E} (R_n(t) - R_n(s))^2 \leq \frac{C |t - s|^{1/2}}{n^2}$$

for all  $s, t \in [0, 1]$ .

*Proof.* Note that  $R_n(t) - R_n(s)$  can be written as

$$R_n(t) - R_n(s) = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} J(X_i, X_j)$$

where  $J(x, y) = J_{s,t}(x, y) = (h(x, y, t) - h(x, y, s)) - (h_1(x, t) - h_1(x, s)) - (h_1(y, t) - h_1(y, s)) + (U(t) - U(s))$  (in what follows we will drop the index on  $J(x, y)$  hoping that this will not give rise to confusion). Then the l.h.s. of (5.17) becomes

$$E(R_n(t) - R_n(s))^2 = \frac{4}{n^2(n-1)^2} \sum_{\substack{1 \leq i_1 < j_1 \leq n \\ 1 \leq i_2 < j_2 \leq n}} E(J(X_{i_1}, X_{j_1})J(X_{i_2}, X_{j_2})).$$

Note that  $J$  is a degenerate kernel and that hence  $E_{X_{j_2}} J(X_{i_1}, X_{j_1})J(X_{i_2}, X_{j_2}) = 0$  provided  $j_2$  is different from the other three indices, and likewise for any other index occurring only once in  $\{i_1, j_2, i_2, j_2\}$ .

We now order the indices  $i_1, j_1, i_2, j_2$  and denote by  $d_i, i = 1, 2, 3$ , the  $i$ th largest difference between two consecutive indices. Suppose for a moment that

$i_1 \leq i_2 \leq j_1 \leq j_2$ —other orderings may be treated in a similar way. Then we have to distinguish between two cases, namely (i)  $d_1 = j_2 - j_1$  (or  $d_1 = i_2 - i_1$ ) and (ii)  $d_1 = j_1 - i_2$ . We begin with case (i), assuming specifically that  $d_1 = j_2 - j_1$ . Let  $d_0 = d_0(n, s, t)$  denote an integer whose specific value will be chosen later. Then if  $d_1 \geq d_0$ , we apply Lemma 4.3 with  $s = t = \frac{1}{2}$  to obtain

$$|E(J(X_{i_1}, X_{j_1})J(X_{i_2}, X_{j_2}))| \leq M_{s,t} \left( \alpha_{\lfloor \frac{d_1}{3} \rfloor} + \beta_{\lfloor \frac{d_1}{3} \rfloor} \right)^{\frac{1}{2}} + \phi(\alpha_{\lfloor \frac{d_1}{3} \rfloor}^{1/2})$$

where

$$M_{s,t} = \max \left\{ (E|J(X_{i_1}, X_{j_1})J(X_{i_2}, X_{j_2})|^2)^{1/2}; \right. \\ \left. (E_{X_{j_2}} E_{X_{i_1} X_{j_1} X_{i_2}} |J(X_{i_1}, X_{j_1})J(X_{i_2}, X_{j_2})|^2)^{1/2} \right\}.$$

Recall that  $h(x, y, t)$ ,  $h_1(x, t)$  and  $U(t)$  are all increasing functions of  $t$ , with values between 0 and 1 and that  $Eh_1(X, t) = U(t)$ . Thus

$$E(J(X_{i_1}, X_{j_1}))^2 \leq 4^2 \{E(h(X_{i_1}, X_{j_1}, t) - h(X_{i_1}, X_{j_1}, s)) \\ + 2 E(h(X_{i_1}, t) - h(X_{i_1}, s)) + (U(t) - U(s))\} \\ \leq 4^2 \{E(h(X_{i_1}, X_{j_1}, t) - h(X_{i_1}, X_{j_1}, s)) \\ + 3(U(t) - U(s))^2\}$$

and similarly

$$E_{X_{i_1} X_{j_1}} (J(X_{i_1}, X_{j_1}))^2 \leq 4^3(U(t) - U(s)).$$

Thus by the conditions imposed on  $h$  and  $U$ , we get that

$$M_{s,t} \leq C|t - s|^{1/2}.$$

If  $d_1 \leq d_0$ , we use the fact that  $|J(x, y)| \leq 2$  to get the bound

$$(5.18) \quad E|J(X_{i_1}, X_{j_1})J(X_{i_2}, X_{j_2})| \leq 2E|J(X_{i_1}, X_{j_1})| \leq C|t - s|.$$

Putting these together, we obtain

$$\sum_{\substack{1 \leq i_1 \leq i_2 \leq j_1 \leq j_2 \leq n \\ d_1 = j_2 - j_1}} E|J(X_{i_1}, X_{j_1})J(X_{i_2}, X_{j_2})| \\ = \sum_{\substack{1 \leq i_1 \leq i_2 \leq j_1 \leq j_2 \leq n \\ d_1 = j_2 - j_1; d_1 \leq d_0}} E|J(X_{i_1}, X_{j_1})J(X_{i_2}, X_{j_2})| \\ + \sum_{\substack{1 \leq i_1 \leq i_2 \leq j_1 \leq j_2 \leq n \\ d_1 = j_2 - j_1; d_1 > d_0}} E|J(X_{i_1}, X_{j_1})J(X_{i_2}, X_{j_2})| \\ \leq C \sum_{\substack{1 \leq i_1 \leq i_2 \leq j_1 \leq j_2 \leq n \\ d_1 = j_2 - j_1; d_1 \leq d_0}} |t - s| \\ + C \sum_{\substack{1 \leq i_1 \leq i_2 \leq j_1 \leq j_2 \leq n \\ d_1 = j_2 - j_1; d_1 > d_0}} \left( \left( \alpha_{\lfloor \frac{d_1}{3} \rfloor} + \beta_{\lfloor \frac{d_1}{3} \rfloor} \right)^{1/2} |s - t|^{1/2} + \phi(\alpha_{\lfloor \frac{d_1}{3} \rfloor}^{1/2}) \right) \\ \leq Cn \left\{ d_0^3 |t - s| + \sum_{d=1}^n d^2 \left( \alpha_{\lfloor \frac{d}{3} \rfloor} + \beta_{\lfloor \frac{d}{3} \rfloor} \right)^{1/2} |s - t|^{1/2} + \sum_{d=d_0}^n d^2 \phi(\alpha_{\lfloor \frac{d}{3} \rfloor}^{1/2}) \right\}$$

Note that (5.6) implies that

$$\begin{aligned} \sum_{d=d_0}^{\infty} d^2 \phi(\alpha_{\lfloor d/3 \rfloor}^{1/2}) &\leq C \sum_{d=d_0}^{\infty} d^2 \phi(\alpha_d^{1/2}) \\ &\leq \frac{C}{d_0^3} \sum_{d=d_0}^{\infty} d^5 \phi(\alpha_d^{1/2}) \leq C/d_0^3. \end{aligned}$$

Thus, choosing  $d_0 = |t - s|^{-1/6}$ , we may conclude that

$$(5.19) \quad \sum_{\substack{1 \leq i_1 \leq i_2 \leq j_1 \leq j_2 \leq n \\ d_1 = j_2 - j_1}} \mathbb{E} |J(X_{i_1}, X_{j_1})J(X_{i_2}, X_{j_2})| \leq Cn|t - s|^{1/2}.$$

We now turn to case (ii), i.e.  $d_1 = j_1 - i_2$ . Again, let  $d_0 = d_0(n, s, t)$  denote an integer whose value will be chosen later. Then, if  $d_1 > d_0$ , we apply Lemma 4.3 again to get

$$\begin{aligned} |\mathbb{E} J(X_{i_1}, X_{j_1})J(X_{i_2}, X_{j_2})| &\leq |\mathbb{E}_{X_{i_1}, X_{i_2}} \mathbb{E}_{X_{j_1}, X_{j_2}} J(X_{i_1}, X_{j_1})J(X_{i_2}, X_{j_2})| \\ &\quad + M_{s,t} \left( \alpha_{\lfloor \frac{d_1}{3} \rfloor} + \beta_{\lfloor \frac{d_1}{3} \rfloor} \right)^{1/2} + \phi(\alpha_{\lfloor \frac{d_1}{3} \rfloor}^{1/2}) \end{aligned}$$

where again as above we find that  $M_{s,t} \leq |t - s|^{1/2}$ . If  $d_1 < d_0$ , we use the estimate (5.18). Then we obtain as in case (i) the estimate

$$\begin{aligned} &\sum_{\substack{1 \leq i_1 \leq i_2 \leq j_1 \leq j_2 \leq n \\ d_1 = j_1 - i_1}} |\mathbb{E} J(X_{i_1}, X_{j_1})J(X_{i_2}, X_{j_2})| \\ &\leq \sum_{\substack{1 \leq i_1 \leq i_2 \leq j_1 \leq j_2 \leq n \\ d_1 = j_1 - i_1}} |\mathbb{E}_{X_{i_1}, X_{i_2}} \mathbb{E}_{X_{j_1}, X_{j_2}} J(X_{i_1}, X_{j_1})J(X_{i_2}, X_{j_2})| + Cn|t - s|^{1/2} \end{aligned}$$

In treating the sum on the r.h.s., we have to distinguish the cases where  $d_2 = i_2 - i_1$  and where  $d_2 = j_2 - j_1$ . As both yield the same result, we may assume that  $d_2 = j_2 - j_1$ . Then again, if  $d_2 \geq d_0$ , we apply Lemma 4.3 to find

$$|\mathbb{E}_{X_{i_1}, X_{i_2}} \mathbb{E}_{X_{j_1}, X_{j_2}} J(X_{i_1}, X_{j_1})J(X_{i_2}, X_{j_2})| \leq M_{s,t} \left( \alpha_{\lfloor \frac{d_2}{3} \rfloor} + \beta_{\lfloor \frac{d_2}{3} \rfloor} \right)^{1/2} + \phi(\alpha_{\lfloor \frac{d_2}{3} \rfloor}^{1/2}).$$

If  $d_2 < d_0$ , we apply again (5.18). Put together, we find

$$\begin{aligned} &\sum_{\substack{1 \leq i_1 \leq i_2 \leq j_1 \leq j_2 \leq n \\ d_1 = j_1 - i_1; d_2 = j_2 - j_1}} |\mathbb{E}_{X_{i_1}, X_{i_2}} \mathbb{E}_{X_{j_1}, X_{j_2}} J(X_{i_1}, X_{j_1})J(X_{i_2}, X_{j_2})| \\ &\leq C \sum_{\substack{1 \leq i_1 \leq i_2 \leq j_1 \leq j_2 \leq n \\ d_1 = j_1 - i_1; d_2 = j_2 - j_1; d_2 \geq d_0}} \left( M_{s,t} \left( \alpha_{\lfloor \frac{d_2}{3} \rfloor} + \beta_{\lfloor \frac{d_2}{3} \rfloor} \right)^{1/2} + \phi(\alpha_{\lfloor \frac{d_2}{3} \rfloor}^{1/2}) \right) \\ &\quad + \sum_{\substack{1 \leq i_1 \leq i_2 \leq j_1 \leq j_2 \leq n \\ d_1 = j_1 - i_1; d_2 = j_2 - j_1; d_2 < d_0}} (C|t - s|) + Cn|t - s|^{1/2} \end{aligned}$$

where again  $M_{s,t} \leq |t - s|^{1/2}$ . The first of the three terms on the r.h.s. is bounded

by

$$C \sum_{d=1}^n n^2 d \left( \alpha_{\lfloor \frac{d}{3} \rfloor} + \beta_{\lfloor \frac{d}{3} \rfloor} \right)^{1/2} |t - s|^{1/2} \leq Cn^2 |t - s|^{1/2}.$$

We now choose  $d_0 = |t - s|^{-1/4}$ , and get the following upper bound on the second term:

$$\begin{aligned} \sum_{\substack{1 \leq i_1 \leq i_2 \leq j_1 \leq j_2 \leq n \\ d_1 = j_1 - i_1; d_2 = j_2 - j_1; d_2 \geq d_0}} \phi(\alpha_{\lfloor \frac{d_2}{3} \rfloor}^{1/2}) &\leq \sum_{d=d_0}^n n^2 d \phi(\alpha_{\lfloor \frac{d}{3} \rfloor}^{1/2}) \\ &= n^2 \sum_{d=d_0}^n d \phi(\alpha_{\lfloor \frac{d}{3} \rfloor}^{1/2}) \leq \frac{1}{d_0^2} = n^2 |t - s|^{1/2}. \end{aligned}$$

Finally, we obtain

$$\sum_{\substack{1 \leq i_1 \leq i_2 \leq j_1 \leq j_2 \leq n \\ d_1 = j_1 - i_1; d_2 = j_2 - j_1; d_2 < d_0}} (C|t - s|) \leq n^2 d_0^2 |t - s| \leq Cn^2 |t - s|^{1/2}.$$

All these estimates together prove the statement of the lemma. □

*Proof of Lemma 5.3.* In our proof, we want to employ Lemma 5.2 with  $X_n(t) = \sqrt{n}R_n(t)$ . Observe that, according to (5.17), condition (5.8) of Lemma 5.2 holds with  $\gamma = \frac{1}{2}$  and  $\beta = 1$ . Moreover, we get for any  $s, t, \delta > 0$  with  $0 \leq s \leq t \leq s + \delta \leq 1$  we have

$$\begin{aligned} R_n(t) - R_n(s) &= \frac{1}{n(n-1)} \sum_{1 \leq i \neq j \leq n} [(h(X_i, X_j, t) - h(X_i, X_j, s)) \\ &\quad - (h_1(X_i, t) - h_1(X_i, s)) - (h_1(X_j, t) - h_1(X_j, s)) \\ &\quad + (U(t) - U(s))] \\ &\leq \frac{1}{n(n-1)} \sum_{1 \leq i \neq j \leq n} [(h(X_i, X_j, s + \delta) - h(X_i, X_j, s)) \\ &\quad - (h_1(X_i, s + \delta) - h_1(X_i, s)) - (h_1(X_j, s + \delta) - h_1(X_j, s)) \\ &\quad + (U(s + \delta) - U(s))] \\ &\quad + \frac{1}{n(n-1)} \sum_{1 \leq i \neq j \leq n} [(h_1(X_i, s + \delta) - h_1(X_i, s)) \\ &\quad + (h_1(X_j, s + \delta) - h_1(X_j, s))] \\ &\leq |R_n(s + \delta) - R_n(s)| + \frac{2}{n} \left| \sum_{i=1}^n [(h_1(X_i, s + \delta) - U(s + \delta)) \right. \\ &\quad \left. - (h_1(X_i, s) - U(s))] \right| + 2|U(s + \delta) - U(s)| \\ &\leq |R_n(s + \delta) - R_n(s)| + \frac{1}{\sqrt{n}} |W'_n(s + \delta) - W'_n(s)| \\ &\quad + 2|U(s + \delta) - U(s)| \end{aligned}$$

and

$$\begin{aligned}
 R_n(s) - R_n(t) &= \frac{1}{n(n-1)} \sum_{1 \leq i \neq j \leq n} [(h(X_i, X_j, s) - h(X_i, X_j, t)) \\
 &\quad - (h_1(X_i, s) - h_1(X_i, t)) - (h_1(X_j, s) - h_1(X_j, t)) \\
 &\quad + (U(s) - U(t))] \\
 &\leq |R_n(s + \delta) - R_n(s)| \\
 &\quad + \frac{1}{n(n-1)} \sum_{1 \leq i \neq j \leq n} [(h_1(X_i, t) - h_1(X_i, s)) \\
 &\quad + (h_1(X_j, t) - h_1(X_j, s)) + (h_1(X_j, s + \delta) - h_1(X_j, s)) \\
 &\quad + (h_1(X_i, s + \delta) - h_1(X_i, s))] \\
 &\leq \frac{4}{n} \sum_{i=1}^n [(h_1(X_i, s + \delta) - U(s + \delta)) - (h_1(X_i, s) - U(s))] \\
 &\quad + 4|U(s + h) - U(s)| \\
 &\leq |R_n(s + \delta) - R_n(s)| + \frac{2}{\sqrt{n}} |W'_n(s + \delta) - W'_n(s)| \\
 &\quad + 4|U(s + h) - U(s)|,
 \end{aligned}$$

from which it follows that

$$\begin{aligned}
 \sqrt{n}|R_n(t) - R_n(s)| &\leq \sqrt{n}|R_n(s + \delta) - R_n(s)| + 2|W'_n(s + \delta) - W'_n(s)| \\
 (5.20) \qquad \qquad \qquad &\quad + 4\sqrt{n}|U(s + \delta) - U(s)|.
 \end{aligned}$$

where

$$W'_n(t) = \frac{2}{\sqrt{n}} \sum_{i=1}^n [h_1(X_i, t) - U(t)],$$

as in (5.7). Note that (5.20) is the representation (5.10) of Lemma 5.2 for the process  $\sqrt{n}R_n(t)$ .

Finally, condition (5.9) of Lemma 5.2 holds for  $W'_n(t)$  with  $r = 4$  by an application of the fourth moment bound (3.19). Thus, by applying Lemma 5.2 to the process  $\sqrt{n}R_n(t)$ , the statement of the lemma follows.  $\square$

### 6. APPLICATIONS TO DIMENSION ESTIMATION

In this section we shall study applications of Theorem 9 to dimension estimation. This involves verification of the conditions on the kernel function as well as on the mixing behavior of the process.

We assume that the data sequence  $(X_n)_{n \geq 0}$  is generated by a weak Bernoulli dynamical system. We assume that the conditions regarding the mixing rate  $(\beta_k)_{k \geq 0}$  and the 1-approximation rate  $(a_k)$  are satisfied (see e.g. Section 1). We denote by  $F$  the marginal distribution of the sequence and we assume that  $F$  has compact support. In this section we will drop the assumption that the  $X_n$ 's are real-valued, noting that the results of the previous sections are then still valid.

When estimating correlation dimensions by the Grassberger-Procaccia method, we have to study the sample correlation integrals

$$U_n(t) = \frac{2}{n(n-1)} \sum_{1 \leq i, j \leq n} 1_{\{\|X_i - X_j\| \leq t\}}$$

which is a  $U$ -process with kernels  $h(x, y, t)$ ,  $0 \leq t \leq r_0$ , given by

$$h(x, y, t) = 1_{\{\|x-y\| \leq t\}}.$$

Without loss of generality we will and may assume that  $r_0 = 1$ . The terms of the Hoeffding decomposition are given by

$$\begin{aligned} h_1(x, t) &= \int 1_{\{\|x-y\| \leq t\}} dF(y) = P(\|x - Y\| \leq t), \\ U(t) &= \int \int 1_{\{\|x-y\| \leq t\}} dF(x) dF(y) = P(\|X - Y\| \leq t), \end{aligned}$$

where  $X$  and  $Y$  are independent random variables with distribution  $F$ .

As the monotonicity and boundedness conditions (A1) and (A2) on the class of kernels  $\{1_{\{\|x-y\| \leq t\}}, 0 \leq t \leq 1\}$  are automatically satisfied, it remains to verify (5.2) and (5.3) as well as the uniform 1-approximation condition. Using the arguments of Example 2.2, the kernels  $h(x, y, t)$ ,  $0 \leq t \leq 1$ , are uniformly 1-continuous with modulus of continuity  $\phi(\epsilon)$  if

$$\begin{aligned} P(t - \epsilon \leq \|X_0 - X_k\| \leq t + \epsilon) &\leq \phi(\epsilon), \\ P(t - \epsilon \leq \|X - Y\| \leq t + \epsilon) &\leq \phi(\epsilon) \end{aligned}$$

holds for all  $0 \leq t \leq 1$ ,  $\epsilon > 0$ ,  $k \geq 1$ , where  $X$  and  $Y$  are independent with distribution  $F$ . These inequalities again follow, with  $\phi(\epsilon) = 2C\epsilon$ , if we can show that

$$(6.1) \quad P(s \leq \|X_0 - X_k\| \leq t) \leq 2C|t - s|,$$

$$(6.2) \quad P(s \leq \|X - Y\| \leq t) \leq 2C|t - s|$$

holds for all  $s, t \in [0, 1]$ ,  $s \leq t$ . As (6.1) and (6.2) also imply (5.2) and (5.3), it suffices to establish (6.1) and (6.2).

In general, (6.2) will be easier to verify as it only involves the product distribution  $F \times F$ . If  $F$  has a bounded density with respect to Lebesgue measure,  $\|X - y\|$  also has a bounded density and thus

$$P(s \leq \|X - Y\| \leq t) \leq M|t - s|$$

where  $M$  is an upper bound on the density of  $\|X - Y\|$ .

Alternatively, (6.2) holds if the correlation integral obeys an exact scaling law, i.e. if

$$U(t) = ct^\alpha, \quad 0 \leq t \leq t_0,$$

where the exponent  $\alpha$  satisfies  $\alpha \geq 1$ . Then, according to the mean value theorem, we have for  $0 \leq s \leq t \leq t_0$

$$U(t) - U(s) = t^\alpha - s^\alpha = (t - s)\alpha\xi^{\alpha-1},$$

where  $\xi \in (s, t)$ . Note that  $\alpha\xi^{\alpha-1}$  is bounded as  $\alpha \geq 1$  so that (6.2) holds.

In general, (6.1) will be much harder to verify as it involves the joint distribution of  $X_0$  and  $X_k$ . If the pairs  $(X_0, X_k)$  have joint densities that are uniformly bounded by some constant  $M$ , then (6.1) holds.

For many dynamical systems joint densities will not exist and we have to verify (6.1) in some different way. One example where this is possible is the doubling map  $tx = 2x \pmod{1}$  on the unit interval  $[0, 1]$ . The invariant measure in this case is Lebesgue measure on  $[0, 1]$ .

Expressing  $X_0$  in the dyadic expansion as  $X_0 = 0.a_1a_2 \dots = \sum_{i=1}^{\infty} \frac{a_i}{2^i}$ , we obtain  $X_1 = 0.a_2a_3 \dots$ , and thus  $X_0 = \frac{X_1}{2} + \frac{a_1}{2}$ . Hence

$$X_1 - X_0 = \frac{X_1}{2} - \frac{a_1}{2}.$$

Now, conditionally given  $a_1 = 0$ , the difference  $X_1 - X_0$  has a uniform distribution on  $[0, \frac{1}{2}]$ . Conditionally given  $a_1 = 1$ , the corresponding distribution is uniform on  $[-\frac{1}{2}, 0]$ . Put together, we see that  $X_1 - X_0$  has a uniform distribution on  $[-\frac{1}{2}, \frac{1}{2}]$  so that  $P(s \leq |X_1 - X_0| \leq t) = 2|t - s|$  for  $0 \leq s \leq t \leq \frac{1}{2}$ .

In a similar way we can treat the general distance  $X_k - X_0$ . In this case  $X_0 = \frac{X_k}{2^k} + \sum_{i=1}^k \frac{a_i}{2^i}$  and hence

$$X_k - X_0 = X_k(1 - \frac{1}{2^k}) - \sum_{i=1}^k \frac{a_i}{2^i}.$$

Now, the first term on the r.h.s. has a uniform distribution on  $[0, 1 - \frac{1}{2^k}]$ . Conditionally given fixed values of  $a_1, \dots, a_k$ , the difference  $X_k - X_0$  is thus uniformly distributed on  $[-\sum_{i=1}^k \frac{a_i}{2^i}, -\sum_{i=1}^k \frac{a_i}{2^i} + 1 - \frac{1}{2^k}]$ . For  $k \geq 2$ , these intervals overlap. A short analysis shows that the interval  $[\frac{j-1}{2^k}, \frac{j}{2^k}]$  is covered by  $2^k - j$  intervals of the type  $[-\sum_{i=1}^k \frac{a_i}{2^i}, -\sum_{i=1}^k \frac{a_i}{2^i} + 1 - \frac{1}{2^k}]$  for various values of  $a_1, \dots, a_k$ . Hence the difference  $X_k - X_0$  has density  $(1 - \frac{j}{2^k}) / (1 - \frac{1}{2^k}) = \frac{2^k - j}{2^k - 1}$  on the interval  $[\frac{j-1}{2^k}, \frac{j}{2^k}]$ ,  $j = 1, 2, \dots, 2^k$ , and symmetrically on the negative half-line. This density is bounded by 1 and thus  $P(s \leq |X_k - X_0| \leq t) \leq 2|t - s|$ , for any  $0 \leq s \leq t \leq 1$ .

A more complicated example for which the condition (6.1) can also be verified directly, is the torus automorphism on the 2-dimensional torus  $\tau^2 = S^1 \times S^1$ , given by

$$T\mathbf{x} = A\mathbf{x} \bmod 1, \quad \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \text{ with } A = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}.$$

Let  $\mathbf{X}_0 = \begin{pmatrix} X_0^1 \\ X_0^2 \end{pmatrix}$  be uniformly distributed on  $\tau^2$ , i.e.  $X_0^{1,2} \sim U[0, 1]$ . Then

$$\mathbf{X}_k = A^k \mathbf{X}_0 \bmod 1 = \begin{pmatrix} f_k & f_{k-1} \\ f_{k-1} & f_{k-2} \end{pmatrix} \times \begin{pmatrix} X_0^1 \\ X_0^2 \end{pmatrix} \bmod 1,$$

where  $f_i$  are Fibonacci numbers, and

$$\begin{aligned} \mathbf{X}_k - \mathbf{X}_0 &= (A^k - I)\mathbf{X}_0 \bmod 1 = \begin{pmatrix} f_k - 1 & f_{k-1} \\ f_{k-1} & f_{k-2} - 1 \end{pmatrix} \times \begin{pmatrix} X_0^1 \\ X_0^2 \end{pmatrix} \bmod 1 \\ (6.3) \quad &= \begin{pmatrix} (f_k - 1)X_0^1 + f_{k-1}X_0^2 \\ f_{k-1}X_0^1 + (f_{k-2} - 1)X_0^2 \end{pmatrix} \bmod 1. \end{aligned}$$

Both  $X_0^1$  and  $X_0^2$  are distributed uniformly on  $[0, 1]$ , and, since all entries of the matrix in (6.3) are integers and everything is taken modulo 1, both coordinates of the vector  $\mathbf{X}_k - \mathbf{X}_0$  are also distributed uniformly on  $[0, 1]$ . Suppose for norm in  $\mathbf{R}^2$  is taken maximum norm (which is the usual choice in these applications), or either of the coordinates. Then, for all  $t, s \in [0, 1]$

$$P(s \leq \|\mathbf{X}_k - \mathbf{X}_0\| \leq t) = |t - s|$$

and the condition (6.1) is satisfied in this example as well with equality and  $C = 1$ .



## REFERENCES

1. Aaronson, J., Burton, R., Dehling, H., Gilat, D., Hill, T., Weiss, B. (1996). Strong laws for  $L$ - and  $U$ -statistics. *Trans. Amer. Math. Soc.* **348**, 2845-2865. MR **97b**:60047
2. Adler, R. and Weiss, B. (1970) Similarity of Automorphisms of the Torus. *Memoirs Amer. Math. Soc.*, No. 98. MR **41**:1966
3. Arcones, M.A., Yu, B. (1994). Central limit theorems for empirical processes and  $U$ -processes of stationary mixing sequences. *J. Theor. Probab.* **7**, 47-71. MR **95a**:60024
4. Barnsley, M. (1993) *Fractals Everywhere*, 2nd ed., Academic Press, Orlando FL. MR **94h**:58101
5. Berbee, H.C.P. (1979). *Random walks with stationary increments and renewal theory*. Mathematical Centre Tracts **112**, Mathematisch Centrum, Amsterdam. MR **81e**:60093
6. Berk, R.H. (1966). Limiting behavior of posterior distributions when the model is incorrect. *Ann. Math. Statist.* **37**, 51-58. MR **32**:6603
7. Berkes, I., Philipp, W. (1977). An almost sure invariance principle for the empirical distribution function of mixing random variables. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **41**, 115-137. MR **57**:4276
8. Berkes, I., Philipp, W. (1979). Approximation theorems for independent and weakly dependent random vectors. *Ann. Probab.* **7**, 29-54. MR **80k**:60008
9. Billingsley, P. (1968). *Convergence of Probability Measures*. John Wiley and Sons, NY. MR **38**:1718
10. Borovkova, S. (1995). Weak convergence of the empirical process of  $U$ -statistics structure for dependent observations. *Theor. Stoch. Proc.* **2** (18), 115-124.
11. Bowen, R. (1975). Smooth partitions of Anosov diffeomorphisms are weak Bernoulli. *Israel J. Math.* **21**, 95 - 100. MR **52**:6786
12. Borovkova, S., Burton, R.M., Dehling, H.G. (1999). Consistency of the Takens estimator for the correlation dimension. *The Annals of Applied Probability* **9**, 376-390. MR **2000g**:60035
13. Burton, R. and Faris, W. (1996) A self-organizing cluster process. *Ann. Appl. Probab* **6**, 1232-1247. MR **97k**:60268
14. Burton, R., C. Kraaikamp, and Schmidt, T. (2000) Natural Extensions for the Rosen Fractions. *Trans. Amer. Math. Soc.* **352**, 1277-1298. MR **2000j**:11123
15. Davydov, Yu.A. (1970). The invariance principle for stationary processes. *Th. Prob. Appl.* **15**, 487-498. (Russian orig., MR **44**:1102)
16. Dehling, H., Philipp, W. (1982). Almost sure invariance principles for weakly dependent vector-valued random variables. *Ann. Probab.* **10**, 689-701. MR **83m**:60011
17. Dehling, H., Denker, M., Philipp, W. (1987). The almost sure invariance principle for the empirical process of  $U$ -statistic structure. *Ann. Inst. Henri Poincaré* **23**, 121-134. MR **88i**:60061
18. Dehling, H. and Taqqu, M.S. (1989). The empirical process of some long-range dependent sequences with an application to  $U$ -statistics. *The Annals of Statistics* **17**, 1767-1783. MR **91c**:60025
19. Denker, M., Grillenberger, C., Siegmund, K. (1976) *Ergodic Theory on Compact Spaces*. Lecture Notes in Mathematics **527**, Springer Verlag, Berlin. MR **56**:15879
20. Denker, M., Keller, G. (1983) On  $U$ -Statistics and v. Mises' Statistics for Weakly Dependent Processes. *Z. Wahrscheinlichkeitstheorie verw. Geb.* **64**, 505-522. MR **85e**:60044
21. Denker, M., Keller, G. (1986). Rigorous statistical procedures for data from dynamical systems. *J. Stat. Phys.* **44**, 67-93. MR **87k**:58149
22. Deo, C.M. (1973). A note on empirical processes of strong-mixing sequences. *Ann. Probab.* **1**, 870-875. MR **50**:8631
23. Donsker, M.D. (1951). An invariance principle for certain probability limit theorems. *Mem. Amer. Math. Soc.* **6**. MR **12**:723a
24. Doob, J. L. (1949) Heuristic approach to the Kolmogorov-Smirnov theorems. *Ann. Math. Statistics* **20**, 393-403. MR **11**:43a
25. Dudley, R. M., Philipp, W. (1983). Invariance principles for sums of Banach space valued random elements and empirical processes. *Z. Wahrsch. Verw. Gebiete* **62**, 509-552. MR **84g**:60012
26. Grassberger, P., Procaccia, I. (1983). Characterization of strange attractors. *Phys. Rev. Lett.* **50**, 346-349. MR **84k**:58141
27. Halmos, P. A. (1946). The theory of unbiased estimation. *Ann. Math. Statist.* **17**, 34-54. MR **7**:463g

28. Helmers, R., Janssen, P., Serfling, R. (1988). Glivenko-Cantelli properties of some generalized empirical df's and strong convergence of generalized  $L$ -statistics. *Probab. Th. Rel. Fields* **79**, 75-93. MR **89h**:60046
29. Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Statist.* **19**, 293-325. MR **10**:134g
30. Hoeffding, W. (1961). The strong law of large numbers for  $U$ -statistics. *University of North Carolina Mimeo Report* No. 302.
31. Hofbauer, F., Keller, G. (1982). Ergodic properties of invariant measures for piecewise monotonic transformations. *Math. Z.* **180**, 119-140. MR **83h**:28028
32. Ibragimov, I.A., Linnik, Yu.V. (1971). *Independent and stationary sequences of random variables*. Wolters-Noordhoff Publishing, Groningen. MR **48**:1287
33. Loève, M. (1977). *Probability Theory. I*, 4th edition, Springer Verlag, Berlin. MR **58**:31324a
34. Nakada, H. (1981). Metrical theory for a class of continued fraction transformations and their natural extensions. *Tokyo J. Math.* **4**, 399-426. MR **83k**:10095
35. Nakada, H., S. Ito, and Tanaka, S. (1977). On the invariant measure for the transformations associated with some real continued fractions. *Keio Engineering Reports* **30**, 159-175. MR **58**:16574
36. Nolan, D., Pollard, D. (1988). Functional limit theorems for  $U$ -processes. *Ann. Prob.* **16**, 1291-1298. MR **89g**:60123
37. Petersen, K. (1983). *Ergodic Theory*, Cambridge University Press, Cambridge. MR **87i**:28002
38. Philipp, W. (1977). A functional law of the iterated logarithm for empirical distribution functions of weakly dependent random variables. *Ann. Prob.* **5**, 319-350. MR **56**:1397
39. Philipp, W. (1986). Invariance principles for independent and weakly dependent random variables. in: *Dependence in probability and statistics*. (Eberlein, E., Taqqu, M., eds) Progress in Probability and Statistics **11**, 225-269. Birkhäuser Boston. MR **89a**:60093
40. Ritter, H., T. Martinez, and K. Schulten (1992). *Neural Computation and Self-Organizing Maps*. Addison-Wesley Reading, MA.
41. Serfling, R. (1984). Generalized  $L$ -,  $M$ - and  $R$ -statistics. *Ann. Statist.* **12**, 76-86. MR **85i**:62018
42. Silverman, B. (1983). Convergence of a class of empirical distribution functions of dependent random variables. *Ann. Prob.* **11**, 745-751 MR **84m**:60033
43. Silverman, B., Brown, T. (1978). Short distances, flat triangles and Poisson limits. *J. Appl. Prob.* **15**, 815-825. MR **80c**:60042
44. Skorohod, A.V. (1956). Limit theorems for stochastic processes. *Theor. Probab. Appl.* **21**, 628-632.
45. Strassen, V., Dudley, R.M. (1969). The central limit theorem and  $\varepsilon$ -entropy. In: *Lecture Notes in Mathematics* **89**, 224-231, Springer-Verlag, Berlin.
46. Takens, F. (1981). Detecting strange attractors in turbulence. In: *Dynamical systems and turbulence*. Lecture Notes in Mathematics **898**, 336-381. Springer-Verlag. MR **83i**:58065
47. Takens, F. (1985). On the numerical determination of the dimension of the attractor. In: *Dynamical Systems and Bifurcations*. Lecture Notes in Mathematics **1125**, 99-106. Springer-Verlag. MR **86f**:58043
48. Withers, C.S. (1975). Convergence of empirical processes of mixing random variables. *Ann. Statist.* **3**, 1101-1108. MR **52**:15593
49. Yoshihara K. (1976). Limiting behaviour of  $U$ -statistics for stationary, absolutely regular processes. *Z. Wahr. Verw. Geb.* **35**, 237-252. MR **54**:6221

ITS-SSOR, DELFT UNIVERSITY OF TECHNOLOGY, MEKELWEG 4, 2628 CD DELFT, THE NETHERLANDS

*E-mail address:* S.A.Borovkova@its.tudelft.nl

DEPARTMENT OF MATHEMATICS, OREGON STATE UNIVERSITY, KIDDER HALL 368, CORVALLIS OREGON 97331

*E-mail address:* burton@math.orst.edu

FAKULTÄT FÜR MATHEMATIK, RUHR-UNIVERSITÄT BOCHUM, UNIVERSITÄTSSTRASSE 150, D-44780 BOCHUM, GERMANY

*E-mail address:* herold.dehling@ruhr-uni-bochum.de