Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Stochastic Processes and their Applications 119 (2009) 1298–1324
stochastic processes and their applications
www.elsevier.com/locate/spa

Asymptotic results for the empirical process of stationary sequences

István Berkes[a], Siegfried Hörmann[b,*], Johannes Schauer[a]

[a] *Institute of Statistics, Graz University of Technology, Steyrergasse 17/IV, 8010 Graz, Austria*
[b] *Department of Mathematics, University of Utah, 155 South 1440 East, Salt Lake City, UT 84112–0090, USA*

Received 24 January 2008; received in revised form 18 June 2008; accepted 20 June 2008
Available online 29 June 2008

Abstract

We prove a strong invariance principle for the two-parameter empirical process of stationary sequences under a new weak dependence assumption. We give several applications of our results.
Published by Elsevier B.V.

MSC: 60F17; 60F15; 62M10

Keywords: Empirical process; Strong approximation; Kiefer process

1. Introduction

Let $\{y_k,\ k \in \mathbb{Z}\}$ be a stationary sequence and let $F(s) = P(y_0 \le s)$ denote its common marginal distribution function. The purpose of the present paper is to study the asymptotic behavior of the empirical process

$$R(s,t) := \sum_{0\le k\le t} (I\{y_k \le s\} - F(s)), \quad s \in \mathbb{R},\ t \ge 0. \tag{1}$$

The process $R(s,t)$ captures several important features of the sequence $\{y_k\}$ and it is one of the basic tools of statistical inference, both parametric and non-parametric, for $\{y_k\}$. We will be interested in the behavior of $R(s,t)$ jointly in s and t, a fact that makes the analysis more technical, but the two-dimensional study of $R(s,t)$ is required in many important statistical

* Corresponding author. Tel.: +1 801 585 7658; fax: +1 801 581 4148.
E-mail address: hormann@math.utah.edu (S. Hörmann).

0304-4149/$ - see front matter. Published by Elsevier B.V.
doi:10.1016/j.spa.2008.06.010

applications, see e.g. Shorack and Wellner [64]. In the case of an independent sequence $\{y_k\}$ the weak limit behavior of $R(s,t)$ was studied first by Müller [50] and Bickel and Wichura [10]; the following basic result is due to Kiefer [44].

Theorem A. *Let* $y_0, y_1, \ldots$ *be i.i.d. uniform* $(0,1)$ *random variables. Then there exists a Gaussian process* $\{K(s,t),\ 0 \le s \le 1,\ 0 \le t < \infty\}$ *with mean* 0 *and covariance* $EK(s,t)K(s',t') = (t \wedge t')(s \wedge s' - ss')$ *such that*

$$\sup_{0\le s\le 1,\, 0\le t\le n} |R(s,t) - K(s,t)| = O(n^{1/3}(\log n)^{2/3}) \quad \text{a.s.} \tag{2}$$

Komlós et al. [45] showed that the error rate in Kiefer's theorem can be improved to $O(\log^2 n)$. It is also known that the rate cannot be better than $O(\log n)$ (cf. Csörgő and Révész [20]). While this means a substantial improvement of (2), Theorem A remains an important result, as one allowing extensions for dependent processes. Note also that to construct the process $K(s,t)$ may require enlarging the probability space of the $\{y_k\}$ or redefining the sequence $\{y_k\}$ on a suitable probability space together with the Gaussian process $K(s,t)$ such that (2) holds. All approximation theorems in the following will be meant in this sense, without explicitly mentioning this fact.

Letting F_n denote the empirical distribution function of the sample $(y_1, \ldots, y_n)$, (2) implies the existence of Brownian bridges $B_1(s), B_2(s), \ldots$ such that

$$\sup_{0\le s\le 1} \left|\sqrt{n}(F_n(s) - s) - B_n(s)\right| = O(n^{-1/6}(\log n)^{2/3}) \qquad \text{a.s.,} \tag{3}$$

improving Donsker's [31] classical invariance principle. Actually, (2) is much more informative than (3): it enables one to prove also strong limit theorems, e.g. laws of the iterated logarithm and fluctuation results for the empirical process $\sqrt{n}(F_n(s) - s)$. Further, as we pointed out above, some important statistical procedures require (2) instead of (3); a typical application is change point problems, see e.g. Bai [2].

For a dependent sequence $\{y_k\}$, the behavior of the empirical process is considerably more complicated than in the i.i.d. case and precise results are known only in a few special cases. For a stationary Gaussian sequence $\{y_k,\ k \in \mathbb{Z}\}$, the empirical process $n^{-1/2}R(s,n)$ $(s \in \mathbb{R})$ converges weakly to a non-degenerate Gaussian process as long as the covariance sequence (r_n) of $\{y_k,\ k \in \mathbb{Z}\}$ decreases sufficiently rapidly. For slowly increasing (r_n) (the critical sequence is $r_n \sim n^{-1}$), a completely different phenomenon takes place: $R(s,n)$ converges weakly, suitably normalized, to a semi-deterministic Gaussian process, see Dehling and Taqqu [27]. Similar results hold for linear processes $\{y_k,\ k \in \mathbb{Z}\}$, see Giraitis and Surgailis [38]. Although this remarkable change of behavior probably holds for a large class of stationary processes, very little is known in this direction. (See the remarks in Berkes and Horváth [7] concerning some nonlinear time series models.) On the other hand, there has been a surge of interest in past years in the empirical processes of nonlinear time series appearing in econometrics and physical sciences, belonging to the weakly dependent type. These processes have nonlinear dynamics given by a stochastic recurrence equation, finite or infinite order. When the correlations of such processes decrease sufficiently rapidly (which is the case in most applications), their empirical process behaves similarly as in the case of a short memory linear process, a fact having important statistical consequences. The basic difficulty in this field is that, despite the simple construction of such processes, the standard theory of weak dependence does not apply for them. The classical approach to weak dependence, developed in the seminal papers of Rosenblatt [58] and Ibragimov

[43], uses the strong mixing property and its variants like β-, ϱ-, ϕ- and ψ-mixing. See Bradley [18] for a comprehensive monograph of mixing theory. Weak invariance principles under strong mixing conditions have been obtained among others by Billingsley [11], Deo [28], Mehra and Rao [49], Rio [56], [57], Withers [68] and Doukhan et al. [33]. The classical mixing conditions are attractive and lead to sharp results, but their scope of applications is rather limited. On the one hand, verifying mixing conditions of the above type is not easy and even when they apply (e.g. for Markov processes), they typically require strong smoothness conditions on the process. For example, even for the AR(1) process

$$X_n = \rho X_{n-1} + \varepsilon_n \quad (|\rho| < 1)$$

with Bernoulli innovations, strong mixing fails to hold (cf. Rosenblatt [62]). Recognizing this fact, an important line of research in probability theory in past years has been to find weak dependence conditions which are strong enough to imply satisfactory asymptotic results, but which are sufficiently general to be satisfied in typical applications. Several conditions of this kind have been found, see Doukhan and Louhichi [32] and Dedecker et al. [22] for recent surveys. For a general overview of the empirical process theory of dependent sequences we refer to Dehling et al. [26]. Yu [74] proved a weak invariance principle for the empirical process of associated sequences. Borovkova et al. [14] consider generalized empirical processes of functionals of absolutely regular processes. Provided the key dependence coefficient decreases sufficiently fast, Ango Nze and Doukhan (cf. [26]) obtain weak convergence to a Gaussian process. For empirical processes related to Gaussian sequences we refer to Csörgő and Mielniczuk [21]. Wu [69] introduced the so-called physical and predictive dependence measures. In [70] he considers the weak convergence of weighted empirical processes under the assumption of causality (see also [71]). For large sample theory of empirical processes generated by long range dependent sequences we refer to Dehling and Taqqu [27]. For an overview and more references see also Giraitis and Surgailis [38] and Koul and Surgailis [46].

Strong invariance principles for empirical processes of the type in Theorem A with dependent data have been far less studied. Berkes and Philipp [8] extended Kiefer's theorem for strong mixing sequences and Berkes and Horváth [6] obtained a similar result for GARCH(p, q) sequence (cf. Bollerslev [13]) under some minor (logarithmic) moment assumptions.

The purpose of the present paper is to study the behavior of the two-parameter empirical process $R(s, t)$ of stationary sequences under a new type of weak dependence condition introduced below. Note that every stationary process $\{y_k, k \in \mathbb{Z}\}$ can be represented, without changing its distribution, as a shift sequence

$$y_k(\omega) = f(T^k \omega), \quad k \in \mathbb{Z}$$

over some probability space $(\Omega, \mathcal{F}, P)$, where $f : \Omega \to \mathbb{R}$ is a measurable function and $T : \Omega \to \Omega$ is a measure-preserving transformation. Actually, most stationary processes in practice can be represented as a shift process of i.i.d. random variables, i.e. they have a representation of the form

$$y_k = f(\ldots, \varepsilon_{k-1}, \varepsilon_k, \varepsilon_{k+1}, \ldots), \tag{4}$$

where $\{\varepsilon_k, k \in \mathbb{Z}\}$ is an i.i.d. sequence and $f : \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}$ is Borel measurable. See Rosenblatt [59–61] for general sufficient criteria for the representation (4). It is easy to see that under mild technical assumptions on the function f, the process $\{y_k, k \in \mathbb{Z}\}$ has the following property:

(A) For any $k \in \mathbb{Z}$ and $m \in \mathbb{N}$ one can find a random variable y_{km} such that we have

$$P(|y_k - y_{km}| \geq \gamma_m) \leq \delta_m \quad (k \in \mathbb{Z},\ m \in \mathbb{N})$$

for some numerical sequences $\gamma_m \to 0$, $\delta_m \to 0$.

(B) For any disjoint intervals $I_1, \ldots, I_r$ of integers and any positive integers $m_1, \ldots, m_r$, the vectors $\{y_{jm_1},\ j \in I_1\}, \ldots, \{y_{jm_r},\ j \in I_r\}$ are independent provided the separation between I_k and I_l is greater than $m_k + m_l$.

Definition 1. A random process $\{y_k,\ k \in \mathbb{Z}\}$ is called *S-mixing* if it satisfies conditions (A) and (B).

In Section 2 various constructions for the y_{km} will be given. The simplest choice (which actually motivated the definition of S-mixing) is

$$y_{km} = f(\ldots, 0, 0, \varepsilon_{k-m}, \ldots, \varepsilon_k, \ldots, \varepsilon_{k+m}, 0, 0, \ldots).$$

Clearly, condition (B) is satisfied. Note that (B) implies that $\{y_{km},\ k \in \mathbb{Z}\}$ is a $2m$-dependent sequence, but this property is not strong enough to prove refined limit theorems for $\{y_k,\ k \in \mathbb{Z}\}$. (We recall that a sequence $\{Z_k\}$ is called m-dependent if for each n the two sets of random variables $\{Z_k,\ k \leq n\}$ and $\{Z_k,\ k > n+m\}$ are independent.) In contrast, S-mixing (here "S" stands for "stationary") will permit us to carry over a large class of limit theorems for independent random variables for $\{y_k,\ k \in \mathbb{Z}\}$. Note that S-mixing does not impose any moment condition on the y_k; for example, it is inherited for the variables $z_k = g(y_k)$ provided that $g : \mathbb{R} \to \mathbb{R}$ is Lipschitz continuous.

If the representation (4) is one-sided, i.e. it has the form

$$y_k = f(\ldots, \varepsilon_{k-1}, \varepsilon_k),$$

then the process $\{y_k,\ k \in \mathbb{Z}\}$ is called causal or non-anticipative. Many popular time series models have a causal representation (cf. [55,65,67]) as an immediate consequence of their "forward" dynamics, for example their definition by a stochastic recurrence equation. However, if we assume e.g., that y_k is some spatial response, causality has no interpretation and may not be realistic. Hence the applicability of S-mixing to this more general model (4) is desirable. We note that causality plays a crucial role in earlier approaches as e.g. in Wu [70] or Ho and Hsing [41].

The classical method to prove limit theorems for weakly dependent random variables is exemplified by the CLT for strong mixing sequences, see e.g. Rosenblatt [58], Billingsley [11] or Ibragimov [43]. This uses a blocking of the variables of the sequence $\{y_k\}$, combined with correlation inequalities, to approximate the characteristic function of normed sums $a_n^{-1}\left(\sum_{k=1}^n y_k - b_n\right)$ of $\{y_k\}$ by the characteristic function of normed sums of independent random variables. This method yields nearly optimal results in the case of the CLT and LIL, but it is not strong enough to prove finer asymptotic results. Much stronger results can be proved by using coupling techniques: if the dependence coefficient between a r.v. X and a σ-algebra $\mathcal{M}$ is small, one can construct a r.v. X^*, independent of $\mathcal{M}$ and having the same distribution as X, such that X and X^* are 'close'. (See Berbee [4], Berkes and Philipp [9] and Bradley [17] for the case of β-, ϕ- and α-mixing, respectively.) This enables one to approximate separated blocks of a weakly dependent sequence (X_n) by independent random variables, leading directly to a large class of limit theorems for (X_n). As noted above, the classical mixing conditions have a rather limited applicability, but equally effective coupling inequalities have been obtained for most of the new weak dependence measures, such as

$$\tau(\mathcal{M}, X) = \int \|F_{X|\mathcal{M}}(t) - F_X(t)\|_1 \mathrm{d}t$$
$$\alpha(\mathcal{M}, X) = \sup_{t\in\mathbb{R}} \|F_{X|\mathcal{M}}(t) - F_X(t)\|_1$$
$$\beta(\mathcal{M}, X) = \|\sup_{t\in\mathbb{R}} F_{X|\mathcal{M}}(t) - F_X(t)\|_1$$
$$\phi(\mathcal{M}, X) = \sup_{t\in\mathbb{R}} \|F_{X|\mathcal{M}}(t) - F_X(t)\|_\infty$$

(see Dedecker et al. [22–24], Rio [56,57]). Here F_X and $F_{X|\mathcal{M}}$ denote, respectively, the distribution function of X, resp. its conditional distribution relative to $\mathcal{M}$. Our S-mixing condition is not directly comparable with the above dependence measures: on the one hand, S-mixing is restricted to a more limited class of processes, namely processes $\{y_k\}$ allowing the representation (4); on the other hand, for such processes its verification is almost immediate (see the examples in Section 3) and it provides the required approximating independent r.v. X^* directly, without coupling inequalities. Actually, S-mixing lies much closer to the predictive dependence measures introduced in Wu [69] which also provide the coupling variables directly, although, as our examples will show, S-mixing leaves more freedom in constructing the approximating independent r.v.'s.

A third approach to weak dependence is martingale approximation, as developed in Gordin [39] and Philipp and Stout [53]. In the context of sequences $\{y_k\}$ of the form (4), particularly complete results have been proved by Wu [69,73]. Again, S-mixing cannot be directly compared to approximate martingale conditions valid for weak dependence sequences: the latter hold for a very large class of processes, but they apply only in the context of partial sums, unlike S-mixing which has no such limitations.

Our paper is organized as follows. In Section 2 we will formulate our results and in Section 3 we will give several applications. In Section 4 we will give the proofs.

2. Results

As discussed in the Introduction, our mixing condition requires approximating r.v.'s y_{km}, $k \in \mathbb{Z}, m \in \mathbb{N}$. In applications, such variables can be constructed in various ways (truncation, substitution, coupling, smoothing); we will discuss various constructions after formulating our theorems. On occasion we will use the notation $a_n \ll b_n$, meaning that $\limsup_{n\to\infty} |a_n/b_n| < \infty$.

For our first result, Theorem 1, we assume that y_0 is uniformly distributed on the unit interval. We define $Y_k(s) = I\{y_k \le s\} - s$, $s \in [0, 1]$.

Theorem 1. *Let $\{y_k,\ k \in \mathbb{Z}\}$ be a stationary S-mixing sequence satisfying* (A) *with $\gamma_m = \delta_m = m^{-A}$, $A > 4$. Assume further that y_0 is uniformly distributed on the unit interval. Then the series*

$$\Gamma(s, s') = \sum_{-\infty<k<\infty} \mathbb{E}\, Y_0(s)Y_k(s') \tag{5}$$

converges absolutely for every choice of parameters $0 \le s, s' \le 1$. Moreover, there exists a two-parameter Gaussian process $K(s, t)$ such that $\mathbb{E}\, K(s, t) = 0$ and $\mathbb{E}\, K(s, t)K(s', t') =$

$(t \wedge t')\,\Gamma(s, s')$ *and for some* $\alpha > 0$ *we have*

$$\sup_{0\le t\le n}\ \sup_{0\le s\le 1} |R(s,t) - K(s,t)| = o(n^{1/2}(\log n)^{-\alpha}) \quad \text{a.s.} \tag{6}$$

In many applications y_0 will not be uniformly distributed, but it has a continuous distribution function $F(s) = P(y_0 \le s)$. Then we consider the random variables $z_k = F(y_k)$, which are uniformly distributed. To check S-mixing for the transformed sequence (z_k) we assume that the distribution function $F(x) = P(y_0 \le x)$ is Lipschitz continuous, i.e. $P(y_0 \in (r, s]) \le L\cdot(s-r)^\theta$ for all $-\infty < r < s < \infty$ and for some $\theta > 0$. Then we have

$$P(|F(y_k) - F(y_{km})| > m^{-A}) \le P\left(|y_k - y_{km}|^\theta > (1/L)\cdot m^{-A}\right).$$

Thus if

$$P(|y_k - y_{km}| > m^{-A/\theta}) \ll m^{-A} \quad \text{with } A > 4,$$

then Theorem 1 applies to the sequence $\{F(y_k),\ k \ge 1\}$. We put

$$\hat{Y}_k(s) = I\{y_k \le s\} - P(y_0 \le s), \quad s \in \mathbb{R}.$$

Theorem 2. *Let* $\{y_k,\ k \in \mathbb{Z}\}$ *be a stationary sequence such that* $P(y_0 \le s)$ *is Lipschitz continuous of order* $\theta > 0$. *Assume that* $\{y_k,\ k \in \mathbb{Z}\}$ *is* S*-mixing and* (A) *holds with* $\gamma_m = m^{-A/\theta}$, $\delta_m = m^{-A}$ *for some* $A > 4$. *Then the series*

$$\hat{\Gamma}(s, s') = \sum_{-\infty<k<\infty} \mathbb{E}\,\hat{Y}_0(s)\hat{Y}_k(s') \tag{7}$$

converges absolutely for every choice of parameters $(s, s') \in \mathbb{R}^2$. *Moreover, there exists a two-parameter Gaussian process* $\hat{K}(s,t)$ *such that* $\mathbb{E}\,\hat{K}(s,t) = 0$ *and* $\mathbb{E}\,\hat{K}(s,t)\hat{K}(s',t') = (t \wedge t')\,\hat{\Gamma}(s, s')$ *and for some* $\alpha > 0$

$$\sup_{0\le t\le n}\ \sup_{s\in\mathbb{R}} |\hat{R}(s,t) - \hat{K}(s,t)| = o\left(n^{1/2}(\log n)^{-\alpha}\right) \quad \text{a.s.} \tag{8}$$

For dependent sequences $\{y_k\}$ one cannot hope to obtain sharp error rates like in Komlós et al. [45] since their quantile transformation techniques depend heavily on independence. However, the rates in (6) and (8) are sufficient to obtain the corresponding weak convergence result.

We formulate now a few corollaries of Theorem 1; analogous results can be obtained for Theorem 2. Let $D[0,1]^2$ denote the Skorokhod space corresponding to functions defined on $[0,1]^2$.

Corollary 1. *Assume that the conditions of Theorem 1 hold and let* $(s,t) \in [0,1]^2$. *Then*

$$\frac{1}{\sqrt{n}} R(s, nt)$$

converges weakly in $D[0,1]^2$ *to some Gaussian process* $\{K(s,t), (s,t) \in [0,1]^2\}$ *with* $\mathbb{E}\,K(s,t) = 0$ *and* $\mathbb{E}\,K(s,t)K(s',t') = (t \wedge t')\,\Gamma(s, s')$ *with covariance function* Γ *given in* (5).

We would like to point out that most of the weak invariance principles mentioned in the Introduction deal only with the weak convergence of $n^{-1/2}R(s,n)$ to some one-parameter Gaussian process, which limits their scope of applications.

Combining (6) with Theorem 2 in Lai [47] we get the following two-dimensional functional law of the iterated logarithm.

Corollary 2. *Assume that the conditions of Theorem 1 hold. Then the sequence*

$$\{(2n\log\log n)^{-1/2}R(s,nt), n \geq 3\}$$

of random functions on $[0,1]^2$ *is relatively compact in the supremum norm and has the unit ball* B *in the reproducing kernel Hilbert space* $H(\Gamma^*)$ *as its set of limits, where* $\Gamma^*(s,s',t,t') = (t\wedge t')\Gamma(s,s')$ *with* Γ *as in (5).*

Construction of the y_{km}

In order to apply Theorems 1 and 2 we have to find approximating random variables y_{km} satisfying our S-mixing condition (A) + (B). Below we will discuss different construction methods.

(I) *Substitution:* For $j > k+m$ and $j < k-m$ replace ε_j with some fixed constant:

$$y_{km} = f(\ldots, c, c, \varepsilon_{k-m}, \ldots, \varepsilon_k, \ldots, \varepsilon_{k+m}, c, c, \ldots). \tag{9}$$

For example, if $y_k = \sum_{j\in\mathbb{Z}} a_j\varepsilon_{k-j}$ is a linear process then taking $c = 0$ gives

$$y_{km} = \sum_{j=-m}^{m} a_j\varepsilon_{k-j}. \tag{10}$$

This substitution method is used by Doukhan and Louhichi [32] to estimate the decay rate of some dependence coefficient in the definition of their weak dependence concept. It is important to note that in general one has to be careful whether these random variables y_{km} are still well defined. For example, the variables in an augmented GARCH(1, 1) model $\{y_k, k\in\mathbb{Z}\}$ permit the explicit representation

$$y_k = \sum_{l=1}^{\infty} g(\varepsilon_{k-l})\prod_{i=1}^{l-1} h(\varepsilon_{k-i}), \tag{11}$$

where g and h are Borel measurable functions with $E\log h(\varepsilon_0) = \mu < 0$ and $E|g(\varepsilon_0)| < \infty$. The series in (11) converges a.s. since by $\mu < 0$ and the law of large numbers there exists an a.s. finite random variable $L > 0$ such that

$$\prod_{i=1}^{l-1} h(\varepsilon_{k-i}) = \exp\left(\sum_{i=1}^{l-1}\log h(\varepsilon_{k-i})\right) < \exp(\mu l/2)$$

for $l \geq L$. However, if $h(0) = 1$, then replacing the random variables ε_{k-i}, $i > m$, with 0 will make the infinite series no longer convergent, since the product in (11) will not generally converge to 0 as $l\to\infty$. For the specific example this unpleasant consequence can be avoided by choosing a constant $c \neq 0$ in (9) such that $|h(c)| < 1$. Nevertheless, if we do not have such a simple explicit form for y_k, it is desirable to have a construction method which assures that y_{km} exists, whenever y_k is well defined.

(II) *Truncation:* Many important linear and nonlinear time series models (including linear processes, ARCH/GARCH type models, etc.) can be represented in the form

$$y_k = T\left(\sum_{l=1}^{\infty} g_l(\varepsilon_{k-j(l)}, \ldots, \varepsilon_k)\right),$$

where $j : \mathbb{N} \to \mathbb{N}$ is non-decreasing, g_l and T are Borel measurable. Setting $t(m) = 0$ if $j(1) > m$ and $t(m) = \max\{n \in \mathbb{N} \mid j(n) \le m\}$ otherwise, gives m-dependent random variables by defining

$$y_{km} = T\left(\sum_{l=1}^{t(m)} g_l(\varepsilon_{k-j(l)}, \ldots, \varepsilon_k)\right).$$

(III) *Coupling:* For each $\ell \ge 1$ we define an i.i.d. sequence $\{\varepsilon_k^{(\ell)},\ k \in \mathbb{Z}\}$ with $\varepsilon_0^{(\ell)} \stackrel{\mathcal{L}}{=} \varepsilon_0$ such that the sequences (ε_k), $(\varepsilon_k^{(1)})$, $(\varepsilon_k^{(2)})$, ... are mutually independent. This is always possible by enlarging the original probability space. Now set

$$y_{km} = f(\ldots, \varepsilon_{k-m-1}^{(k)}, \varepsilon_{k-m}, \ldots, \varepsilon_k, \ldots, \varepsilon_{k+m}, \varepsilon_{k+m+1}^{(k)}, \ldots). \tag{12}$$

The advantage of the coupling method is that the random variables y_{km} have the same marginal distributions as the y_k's. Coupling conditions of this type were first used by Wu [69].

(IV) *Smoothing:* If y_k is integrable, then a further construction for y_{km} is given by

$$y_{km} = E(y_k \mid \mathcal{F}_{k-m,k+m}),$$

where $\mathcal{F}_{a,b}$ denotes the σ-field generated by $\{\varepsilon_j, a \le j \le b\}$. Clearly y_{km} is a function of $\varepsilon_{k-m}, \ldots, \varepsilon_{k+m}$ and it provides the best L_2 approximation of y_k among such functions provided $Ey_k^2 < \infty$. Condition (A) of S-mixing is then an 'in probability' version of the usual definition of near-epoch dependence (NED), thus our method covers stationary sequences satisfying NED. See for example Pötscher and Prucha [54].

3. Applications

In this section we apply our results to several important processes. For the construction of the approximating random variables y_{km} we can now use the special structure of each process. Since our S-mixing concept allows for a variety of construction methods, its verification will be relatively simple in all cases.

3.1. Linear processes

Assume that $y_k = \sum_{j=-\infty}^{\infty} a_j \varepsilon_{k-j}$ with i.i.d. random variables ε_k. If $a_j = 0$ for $j < 0$ (causal case), weak invariance principles have been proved among others by Doukhan and Surgailis [34] (in the short memory case) and by Surgailis [66].

Let y_{km} be given as in (10). Then an inequality of type (A) can easily be obtained. For example, if we assume that $\mathbb{E}\,|\varepsilon_0|^p < \infty$ for some $p > 0$ and $a_k \ll |k|^{-(A+\frac{A}{p}+1)}$ $(k \to \infty)$ we get by the Markov and the Minkowski inequality

$$P(|y_k - y_{km}| > m^{-A}) \le \mathbb{E}\,|\varepsilon_0|^p m^{Ap}\left(\sum_{|k|\ge m} |a_k|\right)^p \ll m^{-A}.$$

The assumption on the decay rate of a_k is a little more restrictive than e.g. in [34] or in [70]. However, the results are not directly comparable since we obtain the strong convergence of the two-parameter empirical process and our results apply to non-causal processes as well. In order to apply Theorem 2 we need conditions ensuring that $F_y(x) = P(y_0 \le x)$ is Lipschitz continuous of some order θ ($F_y \in \mathrm{Lip}_\theta$). A weak invariance principle without smoothness assumptions on the innovations is provided in [25]. It can be easily shown that a sufficient condition for $F_y \in \mathrm{Lip}_\theta$ is $F_\varepsilon \in \mathrm{Lip}_\theta$. In [34] a condition on the characteristic function of ε_0 is required, implying that F_ε is Lipschitz continuous and infinitely often differentiable. Note however, that requiring smoothness conditions for the ε's is not necessary for obtaining $F_y \in \mathrm{Lip}_\theta$. A simple example is when $\varepsilon_k = \pm 1$, each with probability $1/2$. In order that the series defining y_0 converges a.s. we have to require $\sum a_n^2 < \infty$ and without loss of generality we can assume $|a_n| \le 1$. Assume further that

$$\int_{|t|>1} \prod_{n:|a_n|<1/|t|} \mathrm{e}^{-\frac{t^2}{2}a_n^2}\, \mathrm{d}t < \infty.$$

Since

$$|E\mathrm{e}^{\mathrm{i}ty_0}| = \prod_{n=-\infty}^{\infty} |\cos(ta_n)| \le I\{|t| \le 1\} + \prod_{n:|a_n|<1/|t|} \mathrm{e}^{-\frac{t^2}{2}a_n^2} I\{|t| > 1\},$$

we conclude that $\int |E\mathrm{e}^{\mathrm{i}ty_0}|\, \mathrm{d}t < \infty$ and thus F_y has a continuous density (cf. [12, p. 347]). This argument can be easily extended by requiring $|E\mathrm{e}^{\mathrm{i}t\varepsilon}| \le g(t)$ for $|t| \le A$, such that

$$\int_{|t|>A}^{\infty} \prod_{n:|a_n|<A/|t|} g(ta_n)\, \mathrm{d}t < \infty.$$

With the exception of special cases one can say little about the shape of the distribution of y_0 (see e.g. [19, Chapter 3.5]).

3.2. Nonlinear time series

Many important time series models $\{y_k,\ k \in \mathbb{Z}\}$ satisfy a stochastic recurrence equation

$$y_k = G(y_{k-1}, \varepsilon_k), \tag{13}$$

where G is a measurable function and $\{\varepsilon_k, k \in \mathbb{Z}\}$ is an i.i.d. sequence. A typical example is the ARCH(1) model (see Engle [37]). Note that the GARCH(p, q) model is formally not covered by (13), but it can be embedded into a $p + q - 1$ dimensional stationary process satisfying a stochastic recurrence equation similar to (13) (see Bougerol and Picard [15,16]), and thus with suitable changes, our method still works. For further examples, see the discussion in Wu [69] and Shao and Wu [63]. Sufficient conditions for the existence of a stationary solution of (13) were given e.g. by Diaconis and Freedman [29]. They showed that (13) has a unique and stationary solution provided G satisfies the Lipschitz condition

$$|G(x_2, u) - G(x_1, u)| \le K(u)|x_2 - x_1|$$

and

$$E[K(\varepsilon_0)] < \infty, \qquad E[\log K(\varepsilon_0)] < 0 \quad \text{and} \quad E|G(x_0, \varepsilon_0)| < \infty \tag{14}$$

for some $x_0 \in \mathbb{R}$. Iterating (13) yields $y_k = f(\dots, \varepsilon_{k-1}, \varepsilon_k)$ for some measurable function f and it is a natural idea to define y_{km} by truncation, i.e. replacing the ε_j's with 0 for $j < k - m$. However, similarly to the construction of the y_{km} by substitution, truncating the sequence ε_j may ruin the convergence of iterations in (13). To avoid this difficulty we can use the coupling method introduced by Wu [69] and define y_{km} by

$$y_{km} = f(\dots, \varepsilon_{k-m-2}^{(k)}, \varepsilon_{k-m-1}^{(k)}, \varepsilon_{k-m}, \dots, \varepsilon_k),$$

where $\{\varepsilon_k^{(\ell)},\ k \in \mathbb{Z}\}$, $\ell = 1, 2, \dots$ are i.i.d. sequences, independent of each other and of the ε_k's, and having the same distribution as $\{\varepsilon_k,\ k \in \mathbb{Z}\}$. Clearly, the r.v.'s y_{km} satisfy (B). From the results of [63] it follows that under the conditions on G assumed by Diaconis and Freedman there exist $p > 0$ and $0 < \varrho < 1$ such that

$$E|y_k - y_{km}|^p \ll \varrho^m.$$

Thus condition (A) is satisfied with exponentially decreasing γ_n and δ_n and consequently our results hold in this case, too.

3.3. Augmented GARCH sequences

Augmented GARCH sequences, introduced by Duan [35], have been applied with great success in macroeconomics and finance. They include many popular models, for example GARCH [13], AGARCH [30] or EGARCH models [51]. Consider the case of an augmented GARCH(1, 1) sequence $\{y_k,\ k \in \mathbb{Z}\}$ defined by

$$y_k = \sigma_k \varepsilon_k \tag{15}$$

with

$$\Lambda(\sigma_k^2) = c(\varepsilon_{k-1})\, \Lambda(\sigma_{k-1}^2) + g(\varepsilon_{k-1}), \tag{16}$$

where $\{\varepsilon_k,\ k \in \mathbb{Z}\}$ is an i.i.d. sequence and $\Lambda(x)$, $g(x)$ and $c(x)$ are real-valued measurable functions such that $\Lambda^{-1}(x)$ exists. Duan [35] and Aue et al. [1] gave necessary and sufficient conditions for the existence of a unique strictly stationary solution of (15) and (16). If a unique stationary solution exists, we can represent the conditional variance σ_k^2 as

$$\Lambda(\sigma_k^2) = \sum_{j=1}^{\infty} \prod_{i=1}^{j-1} c(\varepsilon_{k-i})\, g(\varepsilon_{k-j}). \tag{17}$$

We can construct the approximating r.v.'s y_{km} by defining

$$\Lambda(\sigma_{km}^2) = \sum_{j=1}^{m} \prod_{i=1}^{j-1} c(\varepsilon_{k-i})\, g(\varepsilon_{k-j}) \tag{18}$$

and

$$y_{km} = \sigma_{km} \varepsilon_k. \tag{19}$$

It is obvious that y_{km} satisfy relation (B). Hörmann [42] used the present dependence concept to prove the functional CLT for the partial sum processes

$$S_n(t) = \sum_{k=1}^{[nt]} h(y_k)$$

under the optimal condition $E|h(y_0)|^2 < \infty$. Similar results can also be obtained for a general class of polynomial GARCH(p, q) sequences (see Berkes et al. [5]).

Assume that the distribution function F_ε of ε is Lipschitz continuous of order θ. Then we get for $s, t \in \mathbb{R}$

$$\begin{aligned} |F_y(t) - F_y(s)| &= |P(y_0 \le t) - P(y_0 \le s)| \\ &\le E|F_\varepsilon(t/\sigma_0) - F_\varepsilon(s/\sigma_0)| \\ &\le C|t-s|^\theta E(\sigma_0^{-\theta}). \end{aligned}$$

Therefore we obtain $F_y \in \mathrm{Lip}_\theta$ if $F_\varepsilon \in \mathrm{Lip}_\theta$ and $E(\sigma_0^{-\theta}) < \infty$.

The following lemma in Hörmann [42] shows that, under logarithmic moment conditions, a polynomial GARCH(1, 1) sequence satisfies (A) with y_{km} defined as in (19). Similar estimates apply for exponential GARCH sequences (see [42, Lemma 7]).

Lemma 1. *Assume that $\Lambda(x) = x^\delta$ with $\delta > 0$ and a strictly stationary solution of (15)–(16) exists. If furthermore*

$$E(\log^+ |g(\varepsilon_0)|)^\mu < \infty \quad \text{and} \quad E(\log^+ |c(\varepsilon_0)|)^\mu < \infty$$

hold for some $\mu > 2$, then for sufficiently small $\alpha > 0$ we have a $C > 0$ such that

$$P\left(|y_k - y_{km}| > \mathrm{e}^{-\alpha m}\right) \le P(|\varepsilon_0| > \mathrm{e}^{\alpha m}) + C\, m^{(2-\mu)/2}.$$

3.4. Linear processes with dependent innovations

Let $\{y_k,\ k \in \mathbb{Z}\}$ be a stationary sequence satisfying conditions (A) and (B) with $\gamma_m = \delta_m = m^{-A_1}$, $A_1 > 1$. Let

$$z_k = \sum_{i=-\infty}^{\infty} a_i\, y_{k-i}$$

be the linear process generated by the y_k and set

$$z_{km} = \sum_{i=-m}^{m} a_i\, y_{k-i,m}.$$

Assume that the following conditions hold:

$$E|y_0|^p < \infty \quad \text{for some } p > 0,$$

$$a_k \ll |k|^{-(A_2 + \frac{A_2}{p} + 1)} \quad A_2 > 0.$$

Then routine computations show that (A) holds with $\gamma_m = \delta_m = m^{-A}$ where $A = \min(A_1 - 1, A_2)$. Condition (B) is also satisfied with the modification that for the independence of the vectors $\{y_{jm_1},\ j \in I_1\}, \ldots, \{y_{jm_r},\ j \in I_r\}$ the separation between I_k and I_l has to be greater than

$4m_k+4m_l$, a difference inconsequential for the validity of our theorems. This shows, for example, that our results apply for an AR(1) processes with augmented GARCH innovations. Processes with dependent innovations play an important role in modeling financial data. (See e.g. [3,40, 48].) Invariance principles for the partial sums of linear processes with dependent innovations have been studied by Wu and Min [72].

4. Proofs

In this section we will prove Theorem 1. The concept of our proof is based on the method of Berkes and Philipp [8]. Since the arguments are rather technical, we outline the main ideas.

Let $t_0 = 0$ and $t_k = \exp(k^{1-\varepsilon})$, $k \geq 1$ and $\varepsilon \in (0, 1)$ to be specified later. Further let

$$s_{k_i} = i2^{-[\log k/(2\log 2)]} \qquad \text{for } 0 \leq i \leq d_k = 2^{[\log k/(2\log 2)]} \quad (k \geq 1).$$

In addition set $d_0 = 0$ and $s_{0_0} = 0$. Then

$$\mathcal{G} = \bigcup_{k\geq 0} \{(s_{k_i}, t_k), 0 \leq i \leq d_k\}$$

defines a set of points in $[0, 1] \times [0, \infty)$, which we shall call grid. Note that this construction implies $\{s_{\ell_1}, \ldots, s_{\ell_{d_\ell}}\} \subseteq \{s_{k_1}, \ldots, s_{k_{d_k}}\}$, if $\ell \leq k$. Hence, for every point (s, t) on the grid $\mathcal{G}$, $R(s, t)$ can be written as a telescoping sum of vertical increments $R(s_{k_i}, t_k) - R(s_{k_i}, t_{k-1})$ and horizontal increments $R(s_{k_i}, t_k) - R(s_{k_{i-1}}, t_k)$, where the indices i depend on the point (s, t). The segmentation can be carried out as follows supposing that $(s, t) = (s_{k_i}, t_k)$. We show how to move on the grid from (s, t) to $(0, 0)$ using vertical and horizontal moves, then the increments can easily be obtained. In the first step we want to move vertically on the grid, therefore we check if (s_{k_i}, t_{k-1}) is also a grid point. If it is, we move there and repeat this step starting from (s_{k_i}, t_{k-1}). If we cannot move vertically, we use step two that moves us horizontally from (s_{j_l}, t_j) to $(s_{j_{l-1}}, t_j)$. Then we continue with step one starting from $(s_{j_{l-1}}, t_j)$. Repeating the two steps will lead us to $(0, 0)$.

Of course, the decomposition of (s, t) can also be used to write $K(s, t)$ as sum of increments. Using our dependence condition and a blocking method we get for k sufficiently large, that the distribution of the $\mathbb{R}^{d_k+1}$ valued vector

$$\mathbf{Z}_k = (t_k - t_{k-1})^{-1/2}(R(s_{k_i}, t_k) - R(s_{k_i}, t_{k-1}))_{i=0}^{d_k}$$

is close in distribution to

$$\mathbf{V}_k = (t_k - t_{k-1})^{-1/2}(K(s_{k_i}, t_k) - K(s_{k_i}, t_{k-1}))_{i=0}^{d_k}.$$

This distributional closeness is shown in terms of closeness of the corresponding characteristic functions (Lemma 9). A well known result of Berkes and Philipp [9, Theorem 1] allows us to construct a sequence of independent vectors $\hat{\mathbf{V}}_k$ with $\hat{\mathbf{V}}_k \stackrel{\mathcal{L}}{=} \mathbf{V}_k$ on the same space with the sequence $\mathbf{Z}_k$, such that $\|\mathbf{Z}_k - \hat{\mathbf{V}}_k\|$ is small with high probability. By Lemma 2.11 of Philipp and Dudley [36] we can assume without loss of generality that $\hat{\mathbf{V}}_k = \mathbf{V}_k$, i.e. the sequence $\hat{\mathbf{V}}_k$ can be extended to a process K. Since it also turns out that the horizontal increments are negligible, this implies that we can construct the processes K and R in such a way that they are sufficiently close on the grid $\mathcal{G}$. These results will be derived in Section 3.2. Hence we have to show that the fluctuation of both processes on the rectangles $[s_{k_i}, s_{k_{i+1}}] \times [t_k, t_{k+1}]$ is sufficiently small. The latter issue is treated in the following subsection.

4.1. Increments of the empirical process

Let $\{y_k, k \in \mathbb{Z}\}$ be an S-mixing sequence with approximating random variables y_{km}. We put $Y_{km}(s) = I\{y_{km} \le s\} - s$, $s \in (0, 1)$.

Lemma 2. *Assume that the conditions of Theorem 1 hold. Then there is a constant C_2 such that for any $k \ge 1$ and $0 \le s, t \le 1$*

$$|EY_0(s)Y_k(t)| \le C_2 k^{-A}. \tag{20}$$

Proof. For some natural number $m \le k/2$ write

$$Y_0(s)Y_k(t) = (Y_0(s)Y_k(t) - Y_{0m}(s)Y_{km}(t)) + Y_{0m}(s)Y_{km}(t).$$

By assumption Y_{0m} and Y_{km} are independent. Since all the random variables $|Y_k(t)| \le 1$ and $|Y_{km}(t)| \le 1$ we get

$$\begin{aligned} |EY_0(s)Y_k(t)| &\le |E(Y_0(s)Y_k(t) - Y_{0m}(s)Y_{km}(t))| + |EY_{0m}(s)||EY_{km}(t)| \\ &\le E|Y_0(s) - Y_{0m}(s)| + E|Y_k(t) - Y_{km}(t)| + |EY_{0m}(s)|. \end{aligned} \tag{21}$$

Next observe that

$$E|Y_0(s) - Y_{0m}(s)| = P(Y_0(s) \ne Y_{0m}(s)). \tag{22}$$

Note that $P(Y_0(s) \ne Y_{0m}(s))$ is the probability that y_0 and y_{0m} are on different sides of s. Hence by our assumptions we get

$$\begin{aligned} &P(Y_0(s) \ne Y_{0m}(s)) \\ &\quad \le P(y_0 \in [s - m^{-A}, s + m^{-A}]) + P(|y_0 - y_{0m}| > m^{-A}) \le C_{2,1} m^{-A}. \end{aligned} \tag{23}$$

Also we have by (22), (23) and $EY_0(s) = 0$

$$|EY_{0m}(s)| \le |EY_{0m}(s) - Y_0(s)| + |EY_0(s)| \le C_{2,1} m^{-A}. \tag{24}$$

Now combine (21)–(23) and take $m = \lfloor k/2 \rfloor$. (As usual $\lfloor x \rfloor$ denotes the integer part of the real number x.) $\square$

Remark 1. Lemma 2 implies that the series in (5) converges absolutely.

We define for $0 \le s \le s' \le 1$ the basic increments

$$\begin{aligned} &\bar{Y}_k(s, s') = Y_k(s') - Y_k(s) = I\{s < y_k \le s'\} - (s' - s) \quad \text{and} \\ &\bar{Y}_k'(s, s') = Y_{km}(s') - Y_{km}(s) = I\{s < y_{km} \le s'\} - (s' - s) \quad \text{with } m = \lfloor k^\rho/2 \rfloor, \end{aligned}$$

where $0 < \rho < 1/2$ will be specified later. Our goal is to estimate the increments

$$R(s', t') - R(s, t) = \sum_{1 \le k \le t} \bar{Y}_k(s, s') + \sum_{t < k \le t'} Y_k(s') \quad \text{for } t' > t. \tag{25}$$

Lemma 3. *Assume that the conditions of Theorem 1 are satisfied. Then for $0 \le s \le s' \le 1$ there are constants $C_3, \tau > 0$ such that*

$$E\left|\sum_{k=1}^{N} \bar{Y}_k(s, s')\right|^2 \le C_3 N (s' - s)^\tau,$$

where C_3, τ do not depend on N, s, s'.

Proof. The stationarity of $\{y_k, k \in \mathbb{Z}\}$ implies that

$$E\bar{Y}_k(s,s')\bar{Y}_l(s,s') = E\bar{Y}_1(s,s')\bar{Y}_{l-k+1}(s,s').$$

Using $\bar{Y}_k = \bar{Y}_k(s,s')$ for notational simplicity we obtain

$$E\left|\sum_{k=1}^{N}\bar{Y}_k\right|^2 = N\left(E\bar{Y}_1^2 + 2\sum_{k=2}^{N}E\bar{Y}_1\bar{Y}_k - \frac{2}{N}\sum_{k=2}^{N}(k-1)E\bar{Y}_1\bar{Y}_k\right). \tag{26}$$

Following the proof of Lemma 2 we get that

$$|E\bar{Y}_0(s,s')\bar{Y}_k(s,s')| \le C_{3,1}k^{-A} \quad \text{for all } 0 \le s \le s' \le 1 \tag{27}$$

and some $C_{3,1} > 0$. On the other hand the Cauchy–Schwarz inequality gives

$$|E\bar{Y}_0(s,s')\bar{Y}_k(s,s')| \le E\bar{Y}_0^2(s,s') = (s'-s)\big(1-(s'-s)\big) \le (s'-s). \tag{28}$$

Putting together (27) and (28) we see that

$$|E\bar{Y}_0(s,s')\bar{Y}_k(s,s')| \le C_{3,2}k^{-A(1-\tau)}(s'-s)^{\tau} \tag{29}$$

for some $C_{3,2} > 0$. Choose $\tau > 0$ such that $A(1-\tau) > 1$. Then the desired result follows using (26) and (29) with standard analysis. □

Lemma 4. *Assume that the conditions of Theorem 1 hold. Then there are constants $C_{4,1}, C_{4,2}, C_{4,3}, \eta > 0$ and $\rho \in (0, 1/2)$ such that for all $x > 1$ and for any $0 \le s \le s' \le 1$*

$$P\left(\left|\sum_{k=1}^{N}\bar{Y}_k(s,s')\right| > x\right) \le C_{4,1}\left(\exp\left(-C_{4,2}\frac{x^2}{N(s'-s)^{\eta}}\right) + \exp\left(-C_{4,3}\frac{x}{N^{\rho}}\right) + x^{-(2+\eta)}\right).$$

Proof. We set

$$S_N = \sum_{k=1}^{N}\bar{Y}_k(s,s') \quad \text{and} \quad S'_N = \sum_{k=1}^{N}\bar{Y}'_k(s,s').$$

Again we use $\bar{Y}_k = \bar{Y}_k(s,s')$ and similarly $\bar{Y}'_k = \bar{Y}'_k(s,s')$. Then the Markov and the Minkowski inequalities give for $\kappa \ge 1$

$$\begin{aligned} P\left(|S_N - S'_N| > x\right) &\le x^{-\kappa}E|S_N - S'_N|^{\kappa} \\ &\le x^{-\kappa}\left(\sum_{k=1}^{N}\left(E|\bar{Y}_k - \bar{Y}'_k|^{\kappa}\right)^{1/\kappa}\right)^{\kappa}. \end{aligned}$$

Observe that $|\bar{Y}_k - \bar{Y}'_k| \in \{0, 1\}$. Consequently by (23) we have

$$E|\bar{Y}_k - \bar{Y}'_k|^{\kappa} = E|\bar{Y}_k - \bar{Y}'_k| = P(\bar{Y}_k \ne \bar{Y}'_k) \le C_{4,4}k^{-\rho A}. \tag{30}$$

Since $A > 4$ we can choose ρ close to $1/2$ and $\eta > 0$ such that $\rho A > 2 + \eta$. Then we get

$$P\left(|S_N - S'_N| > x\right) \le C_{4,5}x^{-(2+\eta)}. \tag{31}$$

We know that by definition the variables $\bar{Y}'_k$, $k = 1, \ldots, N$, are $\lfloor N^\rho \rfloor$-dependent. We now define

$$Z_l^{(1)} = \sum_{k=2l\lfloor N^\rho \rfloor+1}^{(2l+1)\lfloor N^\rho \rfloor \wedge N} \bar{Y}'_k \quad 0 \le l \le m,$$

where m is the largest integer such that $2m\lfloor N^\rho \rfloor < N$. Consequently the variables $Z_l^{(1)}$, $0 \le l \le m$, are independent. Furthermore we define

$$Z_l^{(2)} = \sum_{k=(2l+1)\lfloor N^\rho \rfloor+1}^{(2l+2)\lfloor N^\rho \rfloor \wedge N} \bar{Y}'_k \quad 0 \le l \le m.$$

If $(2m+1)\lfloor N^\rho \rfloor \ge N$ then $Z_m^{(2)}$ is 0. Also define $X_l^{(1)}$ just like $Z_l^{(1)}$ with $\bar{Y}_k$ replacing $\bar{Y}'_k$. Remember that we chose $A\rho > 2$, hence we have $A\rho/2 = 1 + \delta$ with $\delta > 0$. The inequality in (30) now implies

$$\begin{aligned}\left(E|Z_l^{(1)} - X_l^{(1)}|^2\right)^{1/2} &\le \sum_{k=2l\lfloor N^\rho \rfloor+1}^{(2l+1)\lfloor N^\rho \rfloor \wedge N} \left(E|\bar{Y}'_k - \bar{Y}_k|^2\right)^{1/2} \\ &\le \sum_{k=2l\lfloor N^\rho \rfloor+1}^{(2l+1)\lfloor N^\rho \rfloor \wedge N} C_{4,4}^{1/2} k^{-\rho A/2} \le C_{4,5} N^{-\rho\delta}(2l)^{-(1+\delta)}.\end{aligned}$$

By using the Minkowski inequality and Lemma 3 we obtain

$$\begin{aligned}E|Z_l^{(1)}|^2 &\le \left(\left(E|X_l^{(1)}|^2\right)^{1/2} + \left(E|Z_l^{(1)} - X_l^{(1)}|^2\right)^{1/2}\right)^2 \\ &\le \left(\left(C_3 N^\rho (s'-s)^\tau\right)^{1/2} + C_{4,6} N^{-\rho\delta} l^{-(1+\delta)}\right)^2.\end{aligned}$$

As we use approximately $2m$ intervals of length around $\lfloor N^\rho \rfloor$ we get $m \sim \frac{1}{2} N^{1-\rho}$. Thus we can show

$$\begin{aligned}\sum_{l=0}^{m} E|Z_l^{(1)}|^2 &\le C_{4,7}\left(N(s'-s)^\tau + N^{\rho/2-\rho\delta}(s'-s)^{\tau/2} + N^{-2\rho\delta}\right) \\ &\le C_{4,8} N\left((s'-s)^{\tau/2} + N^{-(1+2\rho\delta)}\right). \end{aligned} \tag{32}$$

Further it is clear that

$$|Z_l^{(1)}| \le N^\rho \quad \text{for } 0 \le l \le m. \tag{33}$$

With (32) and (33) we can now apply Kolmogorov's exponential bound ([52], Lemma 7.1) to get

$$P\left(\sum_{l=0}^{m} Z_l^{(1)} > x\right) \le \exp\left(-\frac{C_{4,9} x^2}{N(s'-s)^{\tau/2} + N^{-2\rho\delta}}\right) + \exp\left(-\frac{C_{4,10} x}{N^\rho}\right).$$

It can be verified that an analogue inequality holds for $\sum_{l=0}^m Z_l^{(2)}$. As $S'_N = \sum_{l=0}^m (Z_l^{(1)} + Z_l^{(2)})$ we have shown, using (31), that

$$P(S_N > x) \le 2\exp\left(-\frac{C_{4,9} x^2}{N(s'-s)^{\tau/2} + N^{-2\rho\delta}}\right) + 2\exp\left(-\frac{C_{4,10} x}{N^\rho}\right) + C_{4,5} x^{-(2+\eta)}. \tag{34}$$

If $N(s'-s)^{\tau/2} \geq N^{-2\rho\delta}$ the lemma is proven. Otherwise the first term on the right-hand side of (34) is dominated by $2x^{-(2+\eta)}$ which completes the proof. □

Lemma 5. *If the conditions of Theorem 1 are satisfied, then we have for any* $0 \leq z_0 < z \leq 1$, $T > 1$ *and* $\lambda > \max\{(z-z_0)^{\eta/2}, 1/\log T\}$ *positive constants* $C_{5,1}, C_{5,2}, \alpha > 0$ *such that*

$$P\left(\sup_{\substack{z_0 \leq s \leq z \\ 0 \leq t \leq T}} \left|\sum_{k \leq t} \bar{Y}_k(z_o, s)\right| \geq \lambda T^{1/2}\right) \leq C_{5,1}\left(\exp\left(-C_{5,2}\frac{\lambda^2}{(z-z_0)^\eta}\right) + T^{-\alpha}\right),$$

where η *comes from Lemma 4.*

Proof. We use a chaining argument to prove the lemma. We assume without loss of generality that $z_0 = 0$. Let (s,t) be an element in the rectangle $X = [0,z] \times [0,T]$. Then we can write

$$s = z\sum_{i=1}^{\infty} \zeta_i 2^{-i} \quad \text{for } \zeta_i \in \{0,1\} \text{ and thus define } s_v = z\sum_{i=1}^{v} \zeta_i 2^{-i}.$$

In the same way we use

$$t = T\sum_{i=1}^{\infty} \xi_i 2^{-i} \quad \text{for } \xi_i \in \{0,1\} \text{ and define } t_u = T\sum_{i=1}^{u} \xi_i 2^{-i}.$$

Furthermore we set $s_0 = t_0 = 0$. Observe that

$$(s_{v-1}, s_v] \times (t_{u-1}, t_u] \subseteq (zj2^{-v}, z(j+1)2^{-v}] \times (Ti2^{-u}, T(i+1)2^{-u}],$$

where $(j,i) \in \{0,1,\ldots,2^v-1\} \times \{0,1,\ldots,2^u-1\}$ depend on (s,t). Let for any integers $u, v \geq 1$

$$M_{u,v} = \max_{\substack{0 \leq i \leq 2^u-1 \\ 0 \leq j \leq 2^v-1}} \left|\sum_{Ti2^{-u} < k \leq T(i+1)2^{-u}} \bar{Y}_k(zj2^{-v}, z(j+1)2^{-v})\right|.$$

Then we obtain for any $m \in \mathbb{N}$

$$\begin{aligned}\left|\sum_{0<k\leq t} \bar{Y}_k(0,s)\right| &= \left|\sum_{u,v=1}^{m}\sum_{t_{u-1}<k\leq t_u} \bar{Y}_k(s_{v-1},s_v) + \sum_{t_m<k\leq t}\bar{Y}_k(0,s) + \sum_{0<k\leq t_m}\bar{Y}_k(s_m,s)\right| \\ &\leq \sum_{u,v=1}^{m} M_{u,v} + (t-t_m) + \left|\sum_{0<k\leq t_m}\bar{Y}_k(s_m,s)\right| \\ &\leq \sum_{u,v=1}^{m} M_{u,v} + \frac{T}{2^m} + \left|\sum_{0<k\leq t_m}\bar{Y}_k(s_m,s)\right|.\end{aligned}$$

For any $x \geq 0$ we have $\bar{Y}_k(s,s') \leq \bar{Y}_k(s,s'+x) + x$ and since $s - s_m \leq z2^{-m}$ we get

$$\begin{aligned}\left|\sum_{0<k\leq t_m}\bar{Y}_k(s_m,s)\right| &\leq \sum_{u=1}^{m}\left(\left|\sum_{t_{u-1}<k\leq t_u}\bar{Y}_k(s_m, s_m+z2^{-m})\right| + \sum_{t_{u-1}<k\leq t_u}\frac{z}{2^m}\right) \\ &\leq \sum_{u=1}^{m} M_{u,m} + \frac{T}{2^m}.\end{aligned}$$

Altogether this implies that for every $m \geq 1$

$$\left|\sum_{0<k\leq t} \bar{Y}_k(0,s)\right| \leq 2\left(\sum_{u,v=1}^{m} M_{u,v} + \frac{T}{2^m}\right).$$

For $T > 1$ and $\lambda > 0$ we can choose an $m \in \mathbb{N}$ (dependent on T and λ) such that

$$\max\{1, \lambda/2\}2^{m-1} \leq T^{1/2} \leq \lambda 2^{m-1}. \tag{35}$$

We now use $x_\beta := \sum_{u,v=0}^{\infty} 2^{-\beta(u+v)}$ with $\beta > 0$. Consequently we get

$$\begin{aligned} P\left(\sum_{u,v=1}^{m} M_{u,v} + \frac{T}{2^m} > \lambda T^{1/2}\right) &\leq P\left(\sum_{u,v=1}^{m} M_{u,v} > \frac{\lambda}{2} T^{1/2}\right) \\ &\leq P\left(\sum_{u,v=1}^{m} M_{u,v} > \frac{\lambda}{2x_\beta} T^{1/2} \sum_{u,v=1}^{m} 2^{-\beta(u+v)}\right) \\ &\leq \sum_{u,v=1}^{m} P\left(M_{u,v} > \frac{\lambda}{2x_\beta} T^{1/2} 2^{-\beta(u+v)}\right). \end{aligned}$$

By Lemma 4 we have

$$\begin{aligned} &P\left(M_{u,v} > \frac{\lambda}{2x_\beta} T^{1/2} 2^{-\beta(u+v)}\right) \\ &\leq C_{5,3} 2^{u+v}\left[\exp\left(-C_{5,4}\lambda^2 2^{-2\beta(u+v)+u+v\eta} z^{-\eta}\right) + \exp\left(-C_{5,5}\lambda T^{1/2-\rho} 2^{-\beta(u+v)+u\rho}\right)\right. \\ &\quad \left. + \left(\frac{\lambda T^{1/2} 2^{-\beta(u+v)}}{2x_\beta}\right)^{-(2+\eta)}\right] \\ &= C_{5,3} 2^{u+v}\left(s_1(u,v) + s_2(u,v) + s_3(u,v)\right). \end{aligned}$$

We fix a small β. Then if $2\beta < \min\{\eta, 1\}$ we can find a $\delta_1 > 0$ such that

$$\begin{aligned} \sum_{u,v=1}^{m} 2^{u+v} s_1(u,v) &\leq \sum_{u,v=1}^{m} 2^{u+v} \exp\left(-C_{5,6}\lambda^2\left(2^{\delta_1 u} + 2^{\delta_1 v}\right) z^{-\eta}\right) \\ &= \left(\sum_{u=1}^{m} 2^u \exp\left(-C_{5,6}\lambda^2 2^{\delta_1 u} z^{-\eta}\right)\right)^2 \\ &\leq C_{5,7}\exp\left(-C_{5,8}\lambda^2 z^{-\eta}\right). \end{aligned}$$

To drop the dependence of $C_{5,7}$ and $C_{5,8}$ on λ and z we used the property that $\lambda^2 z^{-\eta} \geq 1$.

We now choose β and n such that $n(\beta - \rho) < -1$ and $n(\beta/2 + \rho - 1/2) < -1/2$. We use (35) and $1/\lambda \leq \log T$ to get $2^m \leq 4T^{1/2}/\lambda \leq 4T^{1/2}\log T$. This together with $\mathrm{e}^{-x} \leq c(n)x^{-n}$

for any $x > 0$ implies

$$\begin{aligned}\sum_{u,v=1}^{m} 2^{u+v} s_2(u,v) &\le \sum_{u,v=1}^{m} C_{5,9}\lambda^{-n}2^{n(\beta(u+v)-u\rho)+u+v}T^{-n(1/2-\rho)}\\ &\le C_{5,9}\lambda^{-n}2^{m(\beta n+1)}T^{-n(1/2-\rho)}m\sum_{u=1}^{\infty}2^{u(n\beta-n\rho+1)}\\ &\le C_{5,10}\lambda^{-n}2^{m(\beta n+1)}T^{-n(1/2-\rho)}m\\ &\le C_{5,11}(\log T)^n 4^{\beta n+1}T^{1/2(\beta n+1)}(\log T)^{\beta n+1}T^{-n(1/2-\rho)}\log T\\ &= C_{5,12}(\log T)^{n(1+\beta)+2}T^{n(\beta/2+\rho-1/2)+1/2}\\ &\le C_{5,13}T^{-\delta_2}\end{aligned}$$

for some positive δ_2. Moreover we observe that for $\beta < 1/2 - 1/(2+\eta)$

$$\begin{aligned}\sum_{u,v=1}^{m} 2^{u+v} s_3(u,v) &\le \sum_{u,v=1}^{m} 2^{u+v}\left(\lambda^{-(2+\eta)}T^{-(2+\eta)/2}2^{(\beta(u+v)+1)(2+\eta)}x_\beta^{2+\eta}\right)\\ &\le C_{5,14}(\log T)^{2+\eta}m^2 2^{2m+(2\beta m+1)(2+\eta)}T^{-(2+\eta)/2}\\ &\le C_{5,15}T^{-\delta_3}\end{aligned}$$

for some $\delta_3 > 0$. This completes the proof. $\square$

We now turn to estimating the increments of the Gaussian process $K(s,t)$.

Lemma 6. *Let $K(s,t)$ be a two-parameter Gaussian process with $EK(s,t) = 0$ and $EK(s,t)K(s',t') = (t \wedge t')\Gamma(s,s')$, where $\Gamma(s,s')$ is defined as in Theorem 1. Then there exist constants $C_{6,1}, C_{6,2} > 0$ such that for all $x \ge x_0$, any $0 \le z_0 \le z \le 1$ and $0 \le T_0 \le T$*

$$P\left(\sup_{(s,t)\in I}|K(s,t)-K(z_0,T_0)| \ge x\left(T^{1/2}(z-z_0)^{\tau/2}+|T-T_0|^{1/2}\right)\right)$$
$$\le C_{6,1}\exp(-C_{6,2}x^2),$$

where $I = [z_0,z]\times[T_0,T]$ and τ stems from Lemma 3.

Proof. We define $Z(s,t) := K(z_0+s(z-z_0), T_0+t(T-T_0)) - K(z_0,T_0)$ for $(s,t)\in[0,1]^2$. Then clearly

$$\sup_{(s,t)\in I}|K(s,t)-K(z_0,T_0)| = \sup_{(s,t)\in[0,1]^2}|Z(s,t)|.$$

We observe that $\Gamma(s,s') = \Gamma(s',s)$ and thus

$$\begin{aligned}E|K(s,t)-K(s',t)|^2 &= t\left(\Gamma(s,s)+\Gamma(s',s')-2\Gamma(s,s')\right)\\ &= t\sum_{k\in\mathbb{Z}}E\bar{Y}_0(s,s')\bar{Y}_k(s,s').\end{aligned}$$

Hence by (29) we get

$$E|K(s,t)-K(s',t)|^2 \le C_{6,3}t|s'-s|^\tau.$$

Combining this observation with the definition of $Z(s,t)$ we infer that

$$E|Z(s,t)-Z(s',t)|^2 \le C_{6,3}T|s'-s|^\tau(z-z_0)^\tau. \tag{36}$$

Lemma 2 shows that $\Gamma(s, s')$ is uniformly bounded. Thus

$$E|Z(s,t) - Z(s,t')|^2 \le C_{6,4}|t' - t|(T - T_0). \tag{37}$$

Next observe that by the Minkowski inequality

$$E|Z(s,t)|^2 \le \left(E^{1/2}|Z(s,t) - Z(0,t)|^2 + E^{1/2}|Z(0,t) - Z(0,0)|^2\right)^2.$$

Together with (36) and (37) this implies

$$\sup_{(s,t)\in[0,1]^2} E|Z(s,t)|^2 \le C_{6,5}\left(T^{1/2}(z - z_0)^{\tau/2} + (T - T_0)^{1/2}\right)^2. \tag{38}$$

Combining (36)–(38) with Lemma 2 in Lai [47] completes the proof. □

We partition the set $[0,1] \times [0,\infty)$ in rectangles $[s_{k_i}, s_{k_{i+1}}] \times [t_k, t_{k+1}]$ where $(s_{k_i}, t_k) \in \mathcal{G}$, where $\mathcal{G}$ is the grid defined at the beginning of Section 4.

Lemma 7 shows that in order to prove Theorem 1 it suffices to construct a Gaussian process $K(s,t)$ with the covariance function given in Theorem 1 which satisfies for some $\gamma_1 > 0$

$$\max_{0\le i\le d_k-1} |R(s_{k_i}, t_k) - K(s_{k_i}, t_k)| \overset{\text{a.s.}}{=} \mathcal{O}\left(t_k^{1/2}(\log t_k)^{-\gamma_1}\right). \tag{39}$$

That is, it suffices to show that $K(s,t)$ and $R(s,t)$ are close to each other on the grid $\mathcal{G}$.

Lemma 7. *Let $\hat{R}(i,k)$ denote the maximal fluctuation of $R(s,t)$ over the rectangle $[s_{k_i}, s_{k_{i+1}}] \times [t_k, t_{k+1}]$. Similarly define for $K(s,t)$ the random variable $\hat{K}(i,k)$. Then there is a $\gamma_0 > 0$ such that*

$$\max_{0\le i\le d_k-1} \hat{R}(i,k) \overset{a.s.}{=} \mathcal{O}\left(t_k^{1/2}(\log t_k)^{-\gamma_0}\right).$$

The same estimate applies for $\hat{K}(i,k)$.

Proof. Observe that (25) implies

$$\begin{aligned}\max_{0\le i\le d_k-1} \hat{R}(i,k) &\le 2 \max_{0\le i\le d_k-1} \sup_{\substack{s_{k_i}\le s\le s_{k_{i+1}}\\ t_k\le t\le t_{k+1}}} |R(s,t) - R(s_{k_i}, t_k)| \\ &\le 2 \max_{0\le i\le d_k-1} \sup_{s_{k_i}\le s\le s_{k_{i+1}}} \left|\sum_{1\le l\le t_k} \bar{Y}_l(s_k, s)\right| + 2 \sup_{\substack{s\in[0,1]\\ t_k\le t\le t_{k+1}}} \left|\sum_{t_k<l\le t} \bar{Y}_l(0,s)\right|.\end{aligned}$$

By Lemma 5 we obtain

$$\begin{aligned}&P\left(\max_{0\le i\le d_k-1} \sup_{s_{k_i}\le s\le s_{k_{i+1}}} \left|\sum_{1\le l\le t_k} \bar{Y}_l(s_{k_i}, s)\right| \ge t_k^{1/2}(\log t_k)^{-\eta/8}\right) \\ &\quad\le C_{7,1}k^{1/2}\left(\exp\left(-C_{7,2}\frac{(\log t_k)^{-\eta/4}}{k^{-\eta/2}}\right) + t_k^{-\alpha}\right) \le C_{7,4}k^{-2},\end{aligned} \tag{40}$$

where we used $d_k \sim k^{1/2}$. Using

$$t_{k+1} - t_k \sim (1-\varepsilon)(\log t_k)^{-\varepsilon/(1-\varepsilon)} t_k \tag{41}$$

which follows from the mean-value theorem we get from some easy estimates

$$P\left(\sup_{\substack{s\in[0,1]\\ t_k\le t\le t_{k+1}}}\left|\sum_{t_k<l\le t}\bar{Y}_l(0,s)\right|\ge t_k^{1/2}(\log t_k)^{-\varepsilon/4}\right)\le C_{7,5}k^{-2}.$$

Application of the Borel–Cantelli lemma finishes the proof of the first proposition. The second part of the lemma can be tackled similarly by using Lemma 6. □

4.2. *Construction of the approximating Gaussian process*

We define the following increments in the parameters s and t:

$$\begin{aligned}\Delta_l^{(j)}&=R(s_{l_j},t_l)-R(s_{l_j},t_{l-1})\quad\text{for } l\ge 1\quad\text{and}\\ B_l^{(j)}&=R(s_{l_j},t_l)-R(s_{l_m},t_l)\quad\text{for } l\ge 1, m=\max\{j-1,0\}.\end{aligned}$$

If (s,t_k) is an element of the grid $\mathcal{G}$, $R(s,t_k)$ can be represented as a sum of the above defined increments, i.e. there are constants m_l, j_l depending on s such that

$$R(s,t_k)=\sum_{l=1}^{k}\left(\delta_l B_l^{(j_l)}+\Delta_l^{(m_l)}\right),\tag{42}$$

where $\delta_l=\delta_l(s)\in\{0,1\}$. Similarly to $\Delta_l^{(j)}$ and $B_l^{(j)}$ we define the increments of $K(s,t)$ as $\hat{\Delta}_l^{(j)}$ and $\hat{B}_l^{(j)}$. Thus we get a representation for $K(s,t_k)$ analogous to (42):

$$K(s,t_k)=\sum_{l=1}^{k}\left(\delta_l \hat{B}_l^{(j_l)}+\hat{\Delta}_l^{(m_l)}\right).\tag{43}$$

Choosing $\varepsilon/(1-\varepsilon)$ smaller than $\eta/8$ we get by (40) and the Borel–Cantelli lemma some $\gamma_2>0$ such that for $k\to\infty$

$$\begin{aligned}\left|\sum_{l=1}^{k}\delta_l B_l^{(j_l)}\right|&\le\sum_{l=1}^{k}\max_{0\le i\le d_l-1}\sup_{s_{l_i}\le s\le s_{l_{i+1}}}\left|\sum_{1\le j\le t_l}\bar{Y}_j(s_{l_i},s)\right|\\ &\ll\sum_{l=1}^{k}t_l^{1/2}(\log t_l)^{-\eta/8}\quad\text{a.s.}\\ &\ll t_k^{1/2}(\log t_k)^{-\gamma_2}.\end{aligned}$$

By similar arguments we can show an analogous result for the process $K(s,t)$. Consequently in view of (39) the representations in (42) and (43) imply that Theorem 1 will be proved if we succeed in constructing the approximating Gaussian process such that for any $s=s_{k_i}$, $i=1,\dots,d_k$ the sum of t-increments

$$\left|\sum_{l=1}^{k}\left(\Delta_l^{(m_l)}-\hat{\Delta}_l^{(m_l)}\right)\right|$$

is not too large. Specifically Theorem 1 follows from

$$\sum_{l=1}^{k} \max_{0\le i\le d_l} \left|\left(R(s_{l_i},t_l)-R(s_{l_i},t_{l-1})\right)-\left(K(s_{l_i},t_l)-K(s_{l_i},t_{l-1})\right)\right|$$
$$\le c_6 t_k^{1/2}(\log t_k)^{-\gamma_3} \quad \text{a.s.} \tag{44}$$

for some $\gamma_3 > 0$ for $k \to \infty$. To show (44) we need some more lemmas.

If $j \in \{t_{l-1}+1,\ldots,t_l\}$ we define the random variables

$$\hat{y}_j = y_{jm}, \quad \text{with } m = \lfloor t_l^{\rho}/2 \rfloor \tag{45}$$

for some $0 < \rho < 1/2$. Additionally we set

$$\hat{Y}_j(s) = I\{\hat{y}_j \le s\} - P(\hat{y}_j \le s)$$

and for $p_{l-1} = \lfloor t_l^{\rho} \rfloor$ we divide the interval $I_l = \{t_{l-1}+p_{l-1}+1,\ldots,t_l\}$ into blocks $I_{l_1}, J_{l_1}, I_{l_2}, J_{l_2}, \ldots, I_{l_n}, J_{l_n}$, where $|I_{l_k}| = \lfloor |I_l|^{\rho^*} \rfloor$ for some $\rho < \rho^* < 1/2$ and $|J_{l_k}| = \lfloor t_l^{\rho} \rfloor$. The last blocks may be incomplete and of course $n = n(l)$. Then we get

$$\sum_{j=t_{l-1}+p_{l-1}+1}^{t_l} \hat{Y}_j(s) = \sum_{k=1}^{n}\sum_{j\in I_{l_k}} \hat{Y}_j(s) + \sum_{k=1}^{n}\sum_{j\in J_{l_k}} \hat{Y}_j(s) =: \sum_{k=1}^{n} T_{l_k}(s) + \sum_{k=1}^{n} T'_{l_k}(s).$$

We now introduce the vector

$$\mathbf{T}_{l_k} := \left(T_{l_k}(s_{l_0}),\ldots,T_{l_k}(s_{l_{d_l}})\right).$$

Observe that n is proportional to $|I_l|^{1-\rho^*}$ and by definition $\mathbf{T}_{l_1}, \mathbf{T}_{l_2}\ldots,\mathbf{T}_{l_n}$ is an $\mathbb{R}^{d_l+1}$ valued independent sequence with $E\mathbf{T}_{l_1} = \mathbf{0}$. We also set

$$\boldsymbol{\xi}_{l_k} = \frac{\mathbf{T}_{l_k}}{|I_{l_1}|^{1/2}} \quad k = 1,\ldots,n.$$

Lemma 8. *We set* $\operatorname{Var}\boldsymbol{\xi}_{l_1} = \boldsymbol{\Sigma}_l = \left(\Sigma_l(s_{l_i},s_{l_j})\right)_{i,j=0}^{d_l}$. *Under the conditions of Theorem 1 there exists a constant C_8 such that*

$$\sup_{0\le i,j\le d_l} \left|\Sigma_l(s_{l_i},s_{l_j}) - \Gamma(s_{l_i},s_{l_j})\right| \le C_8 |I_{l_1}|^{-1} \quad \textit{for } l \ge 1.$$

Proof. Using the stationarity of $\{Y_k(s), k\in\mathbb{Z}\}$ little algebra shows that

$$\frac{1}{N-M}\,\mathbb{E}\left(\sum_{M<k,m\le N} Y_k(s)Y_m(s')\right) = \sum_{|k|<(N-M)} \mathbb{E}\,Y_0(s)Y_k(s')$$
$$-\frac{1}{N-M}\sum_{k=1}^{(N-M)-1} k(\mathbb{E}\,Y_0(s)Y_k(s') + \mathbb{E}\,Y_k(s)Y_0(s')) \quad (M<N).$$

Hence we may write

$$\Gamma(s,s') = \frac{1}{|I_{l_1}|}\,\mathbb{E}\left(\sum_{k,m\in I_{l_1}} Y_k(s)Y_m(s')\right)$$

$$+\frac{1}{|I_{l_1}|}\sum_{k=1}^{|I_{l_1}|-1} k(\mathbb{E}\, Y_0(s)Y_k(s') + \mathbb{E}\, Y_k(s)Y_0(s')) + \sum_{|k|=|I_{l_1}|}^{\infty} \mathbb{E}\, Y_0(s)Y_k(s')$$

$$= \frac{1}{|I_{l_1}|}\mathbb{E}\left(\sum_{k,m\in I_{l_1}} Y_k(s)Y_m(s')\right) + O\left(|I_{l_1}|^{-1}\right) \quad (l\to\infty), \tag{46}$$

where (46) follows from Lemma 2. (Note that O is uniformly in $0 \le s, s' \le 1$.) Consequently we have

$$\begin{aligned}
&|\Sigma_l(s_{l_i}, s_{l_j}) - \Gamma(s_{l_i}, s_{l_j})| \\
&\quad \le \frac{1}{|I_{l_1}|}\sum_{k,m\in I_{l_1}} \mathbb{E}\, |\hat{Y}_k(s_{l_i})\hat{Y}_m(s_{l_j}) - Y_k(s_{l_i})Y_m(s_{l_j})| + O\left(|I_{l_1}|^{-1}\right).
\end{aligned}$$

By (24) and (45) we infer for $k, m \in I_{l_1}$

$$\begin{aligned}
\mathbb{E}\, |\hat{Y}_k(s)\hat{Y}_m(s') - Y_k(s)Y_m(s')| &\le \mathbb{E}\, |\hat{Y}_m(s') - Y_m(s')| + \mathbb{E}\, |\hat{Y}_k(s) - Y_k(s)| \\
&\le C_{8,1} t_l^{-A\rho}.
\end{aligned}$$

Eq. (41) yields $|I_{l_1}| = \mathcal{O}\left(t_l^{\rho^*} l^{-\varepsilon\rho^*}\right)$ and this finishes the proof. $\square$

We set $\boldsymbol{\Gamma}_l = ((\Gamma(s_{l_i}, s_{l_j})))_{i,j=0}^{d_l}$ and denote $\|A\|_\infty = \sup_{i,j} |a_{ij}|$ for some matrix $A = ((a_{ij}))$. Since $\Gamma(s, s')$ is a bounded function we infer by the last lemma that $\sup_l \|\boldsymbol{\Sigma}_l\|_\infty < \infty$.

Set

$$\mathbf{X}_l = n^{-1/2}\sum_{k=1}^{n} \boldsymbol{\xi}_{l_k}$$

and denote by $\langle\cdot \mid \cdot\rangle$ the inner product. Further let $\|\cdot\|$ denote the Euclidian norm.

Lemma 9. *Let $\|u\| \le K \exp\left(l^{1/2}\right)$ for some absolute number K. Then there exist constants $C_{9,1}$, $C_{9,2}$ such that*

$$|E \exp(i\langle\mathbf{u}, \mathbf{X}_l\rangle) - \exp(-1/2\langle\mathbf{u}, \boldsymbol{\Gamma}_l\mathbf{u}\rangle)| \le C_{9,1}\exp\left(-C_{9,2} l^{1-\varepsilon}\right)\|\mathbf{u}\|^2,$$

where ε comes from the definition of t_l.

Proof. For a matrix $A \in \mathbb{R}^{d\times d}$ and $\mathbf{u} \in \mathbb{R}^d$ we get

$$|\langle\mathbf{u}, A\mathbf{u}\rangle| \le \|\mathbf{u}\|\|A\mathbf{u}\| \le \|\mathbf{u}\|^2\|A\| \le d\|\mathbf{u}\|^2\|A\|_\infty.$$

Consequently we get by Lemma 8 and the mean-value theorem that

$$\begin{aligned}
|\exp(-1/2\langle\mathbf{u}, \boldsymbol{\Gamma}_l\mathbf{u}\rangle) - \exp(-1/2\langle\mathbf{u}, \boldsymbol{\Sigma}_l\mathbf{u}\rangle)| &\le |\langle\mathbf{u}, (\boldsymbol{\Gamma}_l - \boldsymbol{\Sigma}_l)\mathbf{u}\rangle| \\
&\le C_{9,3}|I_{l_1}|^{-1}\|\mathbf{u}\|^2 d_l \\
&\le C_{9,4}\exp(-C_{9,5} l^{1-\varepsilon})\|\mathbf{u}\|^2.
\end{aligned} \tag{47}$$

Assume for the moment that the vectors $\boldsymbol{\xi}_{l_k} = \left(\xi_{l_k}(s_{l_0}), \ldots, \xi_{l_k}(s_{l_{d_l}})\right)$ for $1 \le k \le n$ are not only independent but also have the same distribution, then we get

$$E \exp\left(i \langle \mathbf{u}, \mathbf{X}_l \rangle\right) = \left(E \exp\left(i n^{-1/2} \sum_{j=0}^{d_l} u_j \xi_{l_1}(s_{l_j})\right)\right)^n .$$

Some routine analysis shows that $|\exp(ix) - (1 + ix - x^2/2)| \le |x|^3/6$. Thus there exists some $\Theta = \Theta(\mathbf{u}, l)$ with $|\Theta| \le 1$ such that

$$E \exp\left(i n^{-1/2} \sum_{j=0}^{d_l} u_j \xi_{l_1}(s_{l_j})\right) = 1 - \frac{1}{2n}\langle \mathbf{u}, \boldsymbol{\Sigma}_l \mathbf{u}\rangle + \frac{\Theta}{6n^{3/2}} E \left| \sum_{j=0}^{d_l} u_j \xi_{l_1}(s_{l_j})\right|^3 .$$

From the Cauchy–Schwarz inequality and from $|\xi_{l_k}(s)| \le |I_{l_1}|^{1/2}$ we infer that

$$\begin{aligned} E \left| \sum_{j=0}^{d_l} u_j \xi_{l_1}(s_{l_j})\right|^3 &\le |I_{l_1}|^{1/2}(d_l + 1)^{1/2} \|\mathbf{u}\| E \left| \sum_{j=0}^{d_l} u_j \xi_{l_1}(s_{l_j})\right|^2 \\ &\le |I_{l_1}|^{1/2}(d_l + 1)^{1/2} \|\mathbf{u}\| \langle \mathbf{u}, \boldsymbol{\Sigma}_l \mathbf{u}\rangle \\ &\le C_{9,6} |I_l|^{\rho^*/2} d_l^{3/2} \|\mathbf{u}\|^3. \end{aligned}$$

Since $n \sim |I_l|^{1-\rho^*}$ we can find a $\Theta' = \Theta'(\mathbf{u}, l)$ within the complex unit circle such that

$$\begin{aligned} E \exp\left(i n^{-1/2} \sum_{j=0}^{d_l} u_j \xi_{l_1}(s_{l_j})\right) &= 1 - \frac{1}{2n}\langle \mathbf{u}, \boldsymbol{\Sigma}_l \mathbf{u}\rangle + \Theta' \frac{C_{9,7}}{6} |I_l|^{2\rho^* - 3/2} d_l^{3/2} \|\mathbf{u}\|^3 \\ &=: 1 - \frac{1}{2n}\langle \mathbf{u}, \boldsymbol{\Sigma}_l \mathbf{u}\rangle + r(l, \mathbf{u}). \end{aligned}$$

The relation $|(1-t)^r - \exp(-rt)| \le t/2$ holds for $0 \le t \le 1$ and $r \ge 1$. For $\langle \mathbf{u}, \boldsymbol{\Sigma}_l \mathbf{u}\rangle \le 2n$ we then get

$$\left| \exp(-1/2\langle \mathbf{u}, \boldsymbol{\Sigma}_l \mathbf{u}\rangle) - \left(1 - \frac{1}{2n}\langle \mathbf{u}, \boldsymbol{\Sigma}_l \mathbf{u}\rangle\right)^n \right| \le \frac{1}{4n}\langle \mathbf{u}, \boldsymbol{\Sigma}_l \mathbf{u}\rangle. \tag{48}$$

Again assuming $\langle \mathbf{u}, \boldsymbol{\Sigma}_l \mathbf{u}\rangle \le 2n$ we obtain using $|z^n - w^n| \le n|z - w|$ for $z, w \in \mathbb{C}$, $|z|, |w| \le 1$,

$$\left| \left(1 - \frac{1}{2n}\langle \mathbf{u}, \boldsymbol{\Sigma}_l \mathbf{u}\rangle\right)^n - \left(1 - \frac{1}{2n}\langle \mathbf{u}, \boldsymbol{\Sigma}_l \mathbf{u}\rangle + r(l, \mathbf{u})\right)^n \right| \le n|r(l, \mathbf{u})|, \tag{49}$$

because both terms on the left-hand side are within the complex unit circle (one according to our assumptions, the other as it is a characteristic function).

If the $\boldsymbol{\xi}_{l_k} = \left(\xi_{l_k}(s_{l_0}), \ldots, \xi_{l_k}(s_{l_{d_l}})\right)$ are not identically distributed, the estimates used for the first block I_{l_1} in Lemmas 8 and 9 are still valid for the blocks $I_{l_2}, \ldots, I_{l_n}$. Replacing the inequality $|z^n - w^n| \le n|z - w|$ by $|\prod_{j=1}^n z_j - \prod_{j=1}^n w_j| \le \sum_{j=1}^n |z_j - w_j|$, we get the statement in the general case.

Putting together Eqs. (47)–(49) with respect to the restrictions for $\|\mathbf{u}\|$ and the value of n we conclude the proof. □

To complete the proof of Theorem 1, we need the following result of Berkes and Philipp [9].

Lemma 10. *Let* $\{\mathbf{X}_l, l \geq 1\}$ *be a sequence of independent* $\mathbb{R}^{d_l}$, $d_l \geq 1$, *valued random vectors with characteristic functions* $f_l(\mathbf{u})$, $\mathbf{u} \in \mathbb{R}^{d_l}$, *and let* $\{G_l, l \geq 1\}$ *be a sequence of probability distributions on* $\mathbb{R}^{d_l}$ *with characteristic functions* $g_l(\mathbf{u})$, $\mathbf{u} \in \mathbb{R}^{d_l}$. *Suppose that for some non-negative numbers* λ_l, δ_l *and* $W_l \geq 10^8 d_l$

$$|f_l(\mathbf{u}) - g_l(\mathbf{u})| \leq \lambda_l$$

for all $\mathbf{u}$ *with* $\|\mathbf{u}\| \leq W_l$ *and*

$$G_l(\mathbf{u} : \|\mathbf{u}\| > W_l/4) \leq \delta_l.$$

Then without changing its distribution we can redefine the sequence $\{\mathbf{X}_l, l \geq 1\}$ *on a richer probability space together with a sequence* $\{\mathbf{Y}_l, l \geq 1\}$ *of independent random variables such that* $\mathbf{Y}_l \stackrel{\mathcal{L}}{=} G_l$ *and*

$$P(\|\mathbf{X}_l - \mathbf{Y}_l\| \geq \alpha_l) \leq \alpha_l \quad \textit{for } l \in \mathbb{N},$$

where $\alpha_1 = 1$ *and*

$$\alpha_l = 16 d_l W_l^{-1} \log W_l + 4\lambda_l^{1/2} W_l^{d_l} + \delta_l \quad \textit{for } l \geq 2.$$

Proof of Theorem 1. Let $f_l(\mathbf{u})$ be the characteristic function of $\mathbf{X}_l$ and $g_l(\mathbf{u})$ the characteristic function of a d_l-dimensional Gaussian vector $\mathbf{G}_l = (G_l(1), \ldots, G_l(d_l))$ with covariance matrix $\mathrm{Var}(\mathbf{G}_l) = \mathbf{\Gamma}$. As $\Gamma(s, s')$ is a bounded function we get by choosing $W_l = \exp(c_1 l^{\varepsilon})$ with some positive constant c_1 that

$$\begin{aligned} P(\|\mathbf{G}_l\| > W_l/4) &\leq P\left(\max_{1 \leq i \leq d_l} |G_l(i)| > W_l/(4d_l)\right) \\ &\leq c_2 d_l \exp\left(-c_3 (W_l/d_l)^2\right) \\ &\leq c_4 \exp(-c_5 l^{-\varepsilon}). \end{aligned} \tag{50}$$

With the help of Lemmas 9, 10 and (50) we can redefine the sequence $\{\mathbf{X}_l\}$ on a richer probability space together with a sequence of independent Gaussian vectors $\{\mathbf{Y}_l\}$ with covariance matrix $\mathrm{Var}(\mathbf{Y}_l) = \mathbf{\Gamma}_l$ such that

$$P\left(\|\mathbf{X}_l - \mathbf{Y}_l\| \geq c_6 \exp(-c_7 l^{\varepsilon})\right) \leq c_6 \exp(-c_7 l^{\varepsilon}).$$

We set

$$\mathbf{Z}_l = (t_l - t_{l-1})^{-1/2} \left(R(s_{l_i}, t_l) - R(s_{l_i}, t_{l-1})\right)_{i=0}^{d_l} \quad \text{and}$$

$$\mathbf{V}_l = (t_l - t_{l-1})^{-1/2} \left(K(s_{l_i}, t_l) - K(s_{l_i}, t_{l-1})\right)_{i=0}^{d_l}.$$

The definition of the $\mathbf{X}_l$ assures that $\|\mathbf{X}_l - \mathbf{Z}_l\|$ is small. In fact, using arguments akin to our previous considerations, we can show that

$$P\left(\|\mathbf{X}_l - \mathbf{Z}_l\| \geq \exp(-c_8 l^{\varepsilon})\right) \leq c_9 l^{-2}. \tag{51}$$

The Borel–Cantelli lemma then implies that for some constant $c_{10} > 0$ and for all $l \geq l_0(\omega)$

$$\|\mathbf{X}_l - \mathbf{Z}_l\| \leq c_{10} \exp(-c_8 l^{\varepsilon}).$$

By the definition of $\mathbf{V}_l$ we have

$$\{\mathbf{Y}_l, l \geq 1\} \stackrel{\mathcal{L}}{=} \{\mathbf{V}_l, l \geq 1\}.$$

By enlarging the probability space (see Lemma 2.11 in [36]) we can get

$$\{\mathbf{Y}_l, l \geq 1\} = \{\mathbf{V}_l, l \geq 1\}.$$

Altogether we have shown that

$$\max_{0\leq i\leq d_l} \left|\left(R(s_{l_i}, t_l) - R(s_{l_i}, t_{l-1})\right) - \left(K(s_{l_i}, t_l) - K(s_{l_i}, t_{l-1})\right)\right|$$
$$\leq c_{11}(t_l - t_{l-1})^{1/2} \exp(-c_{12} l^{\varepsilon}) \quad \text{a.s. for } l \to \infty.$$

This shows (44) and thus completes the proof of Theorem 1. □

References

[1] A. Aue, I. Berkes, L. Horváth, Strong approximation for the sums of squares of augmented GARCH sequences, Bernoulli 12 (2006) 583–608.
[2] J. Bai, Weak convergence of the sequential empirical process of residuals in ARMA models, Ann. Statist. 22 (1994) 2051–2061.
[3] R. Baillie, C.-F. Chung, M. Tieslau, Analyzing inflation by the fractionally integrated ARFIMA-GARCH Model, J. Appl. Econom. 11 (1996) 23–40.
[4] H.C.P. Berbee, Random walks with stationary increments and renewal theory, in: Mathematical Centre Tracts, vol. 112, Mathematisch Centrum, Amsterdam, 1979.
[5] I. Berkes, S. Hörmann, L. Horváth, The functional central limit theorem for a family of GARCH observations with applications, Stat. Probab. Lett. (in press).
[6] I. Berkes, L. Horváth, Strong approximation of the empirical process of GARCH sequences, Ann. Appl. Probab. 11 (2001) 789–809.
[7] I. Berkes, L. Horváth, Asymptotic results for long memory LARCH sequences, Ann. Appl. Probab. 13 (2003) 641–668.
[8] I. Berkes, W. Philipp, An almost sure invariance principle for the empirical distribution function of mixing random variables, Z. Wahrsch. Verw. Gebiete 41 (1977) 115–137.
[9] I. Berkes, W. Philipp, Approximation theorems for independent and weakly dependent random vectors, Ann. Probab. 7 (1979) 29–54.
[10] P. Bickel, M. Wichura, Convergence criteria for multiparameter stochastic processes and some applications, Ann. Math. Statist. 42 (1971) 1656–1670.
[11] P. Billingsley, Convergence of Probability Measures, Wiley, 1968.
[12] P. Billingsley, Probability and Measure, Wiley, 1995.
[13] T. Bollerslev, Generalized autoregressive conditional heteroskedasticity, J. Econometrics 31 (1986) 307–327.
[14] S. Borovkova, R. Buron, H. Dehling, Limit theorems for functionals of mixing processes with applications to U-Statistics, Trans. Amer. Math. Soc. 353 (2001) 4261–4318.
[15] P. Bougerol, N. Picard, Strict stationarity of generalized autoregressive processes, Ann. Probab. 20 (1992) 1714–1730.
[16] P. Bougerol, N. Picard, Stationarity of GARCH processes and of some nonnegative time series, J. Econometrics 52 (1992) 115–127.
[17] R.C. Bradley, Approximation theorems for strongly mixing random variables, Michigan Math. J. 30 (1983) 69–81.
[18] R.C. Bradley, Introduction to Strong Mixing Conditions, Vol. I–III, Kendrick Press, 2007.
[19] L. Breiman, Probability, Addison Wesley, 1968.
[20] M. Csörgő, P. Révész, Strong Approximations in Probability and Statistics, Academic Press, 1981.
[21] S. Csörgő, J. Mielniczuk, The empirical process of a short-range dependent stationary sequence under Gaussian subordination, Prob. Theory Related Fields 104 (1996) 15–25.
[22] J. Dedecker, P. Doukhan, G. Lang, J.R. León, S. Louhichi, C. Prieur, Weak dependence with examples and applications, in: Lecture Notes in Statistics, vol. 190, Springer, 2007.
[23] J. Dedecker, C. Prieur, Coupling for τ-dependent sequences and applications, J. Theoret. Probab. 17 (2004) 861–885.
[24] J. Dedecker, C. Prieur, New dependence coefficients. Examples and applications to statistics, Probab. Theory Related Fields 132 (2005) 203–236.
[25] J. Dedecker, C. Prieur, An empirical central limit theorem for dependent sequences, Stochast. Process. Appl. 117 (2007) 121–142.

[26] H. Dehling, T. Mikosch, M. Sørensen, Empirical Process Techniques for Dependent Data, Birkäuser, 2002.
[27] H. Dehling, M. Taqqu, The empirical process of some long-range dependent sequences with an application to U-statistics, Ann. Statist. 17 (1989) 1767–1783.
[28] C. Deo, A note on empirical processes of strong-mixing sequences, Ann. Probab. 1 (1973) 870–875.
[29] P. Diaconis, D. Freedman, Iterated random functions, SIAM Rev. 41 (1999) 45–76.
[30] Z. Ding, R. Engle, C.W.J. Granger, A long memory property of stock market returns and a new model, J. Emp. Finance 1 (1993) 83–106.
[31] M. Donsker, Justification and extension of Doob's heuristic approach to the Kolmogorov–Smirnov theorems, Ann. Math. Stat. 23 (1952) 277–281.
[32] P. Doukhan, S. Louhichi, A new weak dependence condition and applications to moment inequalities, Stochast. Process. Appl. 84 (1999) 313–342.
[33] P. Doukhan, P. Massart, E. Rio, The invariance principle for the empirical measure of a weakly dependent process, Ann. Inst. H. Poincaré B 31 (1995) 393–427.
[34] P. Doukhan, D. Surgailis, Functional central limit theorem for the empirical process of short memory linear processes, C. R. Acad. Sci. Paris 326 (1998) 87–92.
[35] J.C. Duan, Augmented GARCH(p, q) process and its diffusion limit, J. Econometrics 79 (1997) 97–127.
[36] R. Dudley, W. Philipp, Invariance principles for sums of Banach space valued random elements and empirical processes, Z. Wahrsch. Verw. Gebiete 62 (1983) 509–552.
[37] R.F. Engle, Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, Econometrica 50 (1982) 987–1007.
[38] L. Giraitis, D. Surgailis, The reduction principle for the empirical process of a long memory linear process, in: Empirical Process Techniques for Dependent Data, Birkhäuser, Boston, 2002, pp. 241–255.
[39] M. Gordin, The central limit theorem for stationary processes, Dokl. Akad. Nauk SSSR 188 (1969) 739–741.
[40] M. Hauser, R. Kunst, Forecasting high-frequency financial data with the ARFIMA-ARCH model, J. Forecasting 20 (2001) 501–518.
[41] H. Ho, T. Hsing, Limit theorems for functionals of moving averages, Ann. Probab. 25 (1997) 1636–1669.
[42] S. Hörmann, Augmented GARCH sequences: Dependence structure and asymptotics, Bernoulli 14 (2008) 543–561.
[43] I.A. Ibragimov, Some limit theorems for stationary processes, Teor. Verojatnost. Primenen. 7 (1962) 361–392 (in Russian).
[44] J. Kiefer, Skorohod embedding of multivariate rv's and the sample df, Z. Wahrsch. Verw. Gebiete 24 (1972) 1–35.
[45] P. Komlós, J. Major, G. Tusnády, An approximation of partial sums of independent RV's and the sample DF, I. Z. Wahrsch. Verw. Gebiete 32 (1975) 111–131.
[46] H. Koul, D. Surgailis, Asymptotic expansion of the empirical process of long memory moving averages, in: Empirical Process Techniques for Dependent Data, Birkhäuser, Boston, 2002, pp. 213–239.
[47] T. Lai, Reproducing kernel Hilbert spaces and the law of the iterated logarithm for Gaussian processes, Z. Wahrsch. Verw. Gebiete 29 (1974) 7–19.
[48] D. Lien, Y. Tse, Forecasting the Nikkei spot index with fractional cointegration, J. Forecasting 18 (1999) 259–273.
[49] K. Mehra, M. Sudhakara Rao, Weak convergence of generalized empirical processes relative to d_q under strong mixing, Ann. Probab. 3 (1975) 979–991.
[50] D.W. Müller, On Glivenko–Cantelli convergence, Z. Wahrsch. Verw. Gebiete 16 (1970) 195–210.
[51] D.B. Nelson, Conditional heteroskedasticity in asset returns: A new approach, Econometrica 59 (1991) 347–370.
[52] V.V. Petrov, Limit Theorems of Probability Theory, Oxford University Press, 1995.
[53] W. Philipp, W.F. Stout, Almost sure invariance principles for partial sums of weakly dependent random variables, Mem. Amer. Math. Soc. 161 (1975).
[54] B.M. Pötscher, I.R. Prucha, Dynamic Nonlinear Econometric Models. Asymptotic Theory, Springer, 1997.
[55] M. Priestley, Nonlinear and Nonstationary Time Series Analysis, Academic Press, 1988.
[56] E. Rio, Processus empriques absoluments réguliers et entropie universelle, Probab. Theory Related Fields 111 (1998) 585–608.
[57] E. Rio, Théorie asymptotique des processus aléatoires faiblement dépendants, in: Collection Mathématiques & Applications, vol. 31, Springer, Berlin, 2000.
[58] M. Rosenblatt, A central limit theorem and a strong mixing condition, Proc. Natl. Acad. Sci. USA 42 (1956) 43–47.
[59] M. Rosenblatt, Stationary processes as shifts of functions of independent random variables, J. Math. Mech. 8 (1959) 665–681.
[60] M. Rosenblatt, Independence and dependence, in: Proc. 4th Berkeley Sympos. Math. Stat. and Prob., vol. II, 1961, pp. 431–443.
[61] M. Rosenblatt, Markov Processes. Structure and Asymptotic Behavior, Springer, 1971.

[62] M. Rosenblatt, Stationary Sequences and Random Fields, Birkhäuser, 1985.
[63] X. Shao, W.B. Wu, Limit theorems for iterated random functions, J. Appl. Probab. 41 (2004) 425–436.
[64] G. Shorack, J. Wellner, Empirical Processes with Applications to Statistics, Wiley, 1986.
[65] R. Stine, Nonlinear time series, in: Encyclopedia of Statistical Sciences, Wiley, 1997.
[66] D. Surgailis, Stable limits of empirical processes of moving averages with infinite variance, Stochast. Process. Appl. 100 (2002) 255–274.
[67] H. Tong, Non-linear Time Series: A Dynamical System Approach, Oxford University Press, 1990.
[68] C. Withers, Convergence of empirical processes of mixing rv's on [0, 1], Ann. Statist. 5 (1975) 1101–1108.
[69] W. Wu, Nonlinear system theory: Another look at dependence, Proc. Natl. Acad. Sci. USA 102 (2005) 14150–14154.
[70] W. Wu, Empirical processes of stationary sequences, Statist. Sinica 18 (2008) 313–333.
[71] W. Wu, Oscillation of empirical distribution functions under dependence, in: IMS Lecture Notes-Monograph Series, High Dimensional Probab. 51 (2006) 53–61.
[72] W. Wu, W. Min, On linear processes with dependent innovations, Stochast. Process. Appl. 115 (2005) 939–958.
[73] W. Wu, Strong invariance principles for dependent random variables, Ann. Probab. 35 (2007) 2294–2320.
[74] H. Yu, A Glivenko-Cantelli lemma and weak convergence for empirical processes of associated sequences, Probab. Theory Related Fields 95 (1993) 357–370.