Available online at www.sciencedirect.com
ScienceDirect
Stochastic Processes and their Applications 117 (2007) 878–903
stochastic processes and their applications
www.elsevier.com/locate/spa

Probability and moment inequalities for sums of weakly dependent random variables, with applications

Paul Doukhan[a,b], Michael H. Neumann[c,*]

[a] *Laboratoire de Statistique du CREST, Timbre J340, 3, avenue Pierre Larousse, 92240 Malakoff, France*
[b] *SAMOS-MATISSE, Université Paris I, 90 rue de Tolbiac, 75013 Paris, France*
[c] *Friedrich-Schiller-Universität Jena, Institut für Stochastik, Ernst-Abbe-Platz 2, 07743 Jena, Germany*

Received 8 December 2005; received in revised form 3 July 2006; accepted 24 October 2006
Available online 16 November 2006

Abstract

Doukhan and Louhichi [P. Doukhan, S. Louhichi, A new weak dependence condition and application to moment inequalities, Stochastic Process. Appl. 84 (1999) 313–342] introduced a new concept of weak dependence which is more general than mixing. Such conditions are particularly well suited for deriving estimates for the cumulants of sums of random variables. We employ such cumulant estimates to derive inequalities of Bernstein and Rosenthal type which both improve on previous results. Furthermore, we consider several classes of processes and show that they fulfill appropriate weak dependence conditions. We also sketch applications of our inequalities in probability and statistics.
© 2006 Elsevier B.V. All rights reserved.

MSC: 60E15; 62E99

Keywords: Bernstein inequality; Cumulants; Rosenthal inequality; Weak dependence

1. Introduction

For a long time mixing conditions have been the dominating type of conditions for imposing a restriction on the dependence between time series data. They are considered to be useful since they are fulfilled for many classes of processes and since they allow us to derive tools similar to those in the independent case. On the other hand, it turns out that certain classes of

* Corresponding author. Fax: +49 3641 946252.
E-mail addresses: paul.doukhan@ensae.fr (P. Doukhan), mneumann@mathematik.uni-jena.de (M.H. Neumann).

0304-4149/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.spa.2006.10.011

processes which are of interest in statistics are not mixing although a successive forgetting of past states takes place. Typical examples are processes driven by discrete observations as they appear, for example, with model-based time series bootstrap methods. In 1999, Doukhan and Louhichi proposed a new concept of restricting dependence which focuses on covariances rather than the total variation distance between joint distributions and the product of corresponding marginals. It has been shown that this concept is more general than mixing and includes, under natural conditions on the process parameters, essentially all classes of processes of interest in statistics; see for example Ango Nze, Bühlmann and Doukhan [1] for an overview. It became readily apparent that the concept of weak dependence allows in many instances similar tools to be used as in the independent or mixing case. For example, versions of a central limit theorem are derived in Doukhan and Louhichi [11] for sequences of random variables, in Coulon-Prieur and Doukhan [6] for situations as they appear with nonparametric curve estimators, and in Neumann and Paparoditis [22] for general triangular schemes. A first exponential inequality was obtained in Doukhan and Louhichi [11], a Bennett-type inequality in Dedecker and Prieur [8], and a Bernstein-type inequality in Kallabis and Neumann [19].

The concept of weak dependence is particularly suitable for deriving upper estimates for the cumulants of sums of random variables. Such cumulant estimates can serve as a starting point for deriving rather precise approximations of distributions as well as rather tight probability inequalities. The main contributions in this paper are inequalities of Bernstein and Rosenthal type for sums of weakly dependent random variables. In the case of mixing, Bernstein-type inequalities can be easily derived from the well-known Bernstein inequality in the independent case by using coupling arguments; see for example Doukhan [10] and Rio [24]. In the case of weakly dependent random variables, Doukhan and Louhichi [11] proved a first exponential inequality via a combinatorial technique. Unfortunately, rather than the rate of t^2 in the exponent as in the independent case, only a rate of $\sqrt{t}$ was obtained by this approach. Using a new coupling result, Dedecker and Prieur [8] proved a Bennett-type inequality which possibly implies a Bernstein-type inequality with t^2 in the exponent. Using cumulant techniques Kallabis and Neumann [19] derived a Bernstein-type inequality with a leading term of $-t^2/(2\mathrm{var}(X_1 + \cdots + X_n))$ in the exponent, under weak dependence conditions tailor-made for causal processes and with an exponential decay of the coefficients of weak dependence. In this paper, we extend this result to more general conditions of weak dependence, including also noncausal processes and allowing a subexponential decay of the weak dependence coefficients. In Section 4, we discuss several statistical applications of this result. It turns out that certain purposes such as a law of iterated logarithm and a precise asymptotics for nonparametric curve estimators do actually require an exponential inequality with a tight leading term in the exponent.

A second major result is a Rosenthal-type inequality which in particular improves a previous inequality given in Doukhan and Louhichi [11]. Using again cumulant techniques we derive such an inequality with an asymptotically dominating term equal to $p!/(2^{p/2}(p/2)!)(\mathrm{var}(X_1 + \cdots + X_n))^{p/2}$. Such an inequality allows us for example to derive a central limit theorem via the method of moments; see again Section 4.

We present the main results, a Bernstein-type and a Rosenthal-type inequality in the next section. In Section 3, Doukhan and Louhichi's [11] concept of weak dependence is recalled and it is shown that the particular conditions used for our inequalities do actually follow from the usual conditions of weak dependence in many instances. Section 4 contains typical applications of our probability inequalities in probability and statistics. A long list of examples of processes satisfying weak dependence conditions is presented in Section 5. Finally, the proofs of the main theorems and of some auxiliary results of general interest are given in Section 6.

2. A Bernstein-type and a Rosenthal-type inequality for sums of weakly dependent random variables

In this section we will be concerned with probability and moment inequalities for $S_n = X_1 + \cdots + X_n$, where $X_1, \ldots, X_n$ are zero mean random variables which fulfill appropriate weak dependence conditions. Throughout the paper, we denote by σ_n^2 the variance of S_n.

Our first result is a Bernstein-type inequality which generalizes and improves previous inequalities of Doukhan and Louhichi [11] and Kallabis and Neumann [19].

Theorem 1. *Suppose that $X_1, \ldots, X_n$ are real-valued random variables with zero mean, defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Let $\Psi : \mathbb{N}^2 \to \mathbb{N}$ be one of the following functions:*

(a) $\Psi(u, v) = 2v$,
(b) $\Psi(u, v) = u + v$,
(c) $\Psi(u, v) = uv$,
(d) $\Psi(u, v) = \alpha(u + v) + (1 - \alpha)uv$, *for some* $\alpha \in (0, 1)$.

We assume that there exist constants $K, M, L_1, L_2 < \infty$, $\mu, \nu \geq 0$, and a nonincreasing sequence of real coefficients $(\rho(n))_{n\geq 0}$ such that, for all u-tuples $(s_1, \ldots, s_u)$ and all v-tuples $(t_1, \ldots, t_v)$ with $1 \leq s_1 \leq \cdots \leq s_u \leq t_1 \leq \cdots \leq t_v \leq n$ the following inequalities are fulfilled:

$$|\mathrm{cov}(X_{s_1} \cdots X_{s_u}, X_{t_1} \cdots X_{t_v})| \leq K^2 M^{u+v-2}((u + v)!)^{\nu} \Psi(u, v)\rho(t_1 - s_u), \tag{1}$$

where

$$\sum_{s=0}^{\infty}(s + 1)^k \rho(s) \leq L_1 L_2^k (k!)^{\mu} \quad \forall k \geq 0, \tag{2}$$

and

$$\mathbb{E}|X_t|^k \leq (k!)^{\nu} M^k \quad \forall k \geq 0. \tag{3}$$

Then, for all $t \geq 0$,

$$\mathbb{P}(S_n \geq t) \leq \exp\left(-\frac{t^2/2}{A_n + B_n^{1/(\mu+\nu+2)} t^{(2\mu+2\nu+3)/(\mu+\nu+2)}}\right), \tag{4}$$

where A_n can be chosen as any number greater than or equal to σ_n^2 and

$$B_n = 2(K \vee M)L_2\left(\left(\frac{2^{4+\mu+\nu} n K^2 L_1}{A_n}\right) \vee 1\right).$$

Remark 2. (i) Inequality (4) resembles the classical Bernstein inequality for independent random variables. Asymptotically, σ_n^2 is usually of order $O(n)$ and A_n can be chosen equal to σ_n^2 while B_n is usually $O(1)$ and hence negligible. In cases where σ_n^2 is very small or where knowledge of the value of A_n is required for some statistical procedure, it might, however, be better to choose A_n larger than σ_n^2. It follows from (1) and (2) that a rough bound for σ_n^2 is given by

$$\sigma_n^2 \leq 2^{1+\nu} n K^2 \Psi(1, 1) L_1. \tag{5}$$

Hence, taking $A_n = 2^{1+\nu} n K^2 \Psi(1,1) L_1$ we obtain from (4) that

$$\mathbb{P}(S_n \geq t) \leq \exp\left(-\frac{t^2}{C_1 n + C_2 t^{(2\mu+2\nu+3)/(\mu+\nu+2)}}\right), \tag{6}$$

where $C_1 = 2^{2+\nu} K^2 \Psi(1,1) L_1$ and $C_2 = 2B_n^{1/(\mu+\nu+2)}$ with

$$B_n = 2(K \vee M) L_2 \left(\frac{2^{3+\mu}}{\Psi(1,1)} \vee 1\right).$$

Inequality (6) is then more of a Hoeffding-type.

(ii) Based on a Rosenthal-type inequality, Doukhan and Louhichi [11] also proved an exponential inequality for S_n, however, with $\sqrt{t}$ instead of t^2 in the exponent.

(iii) Dedecker and Prieur [8] proved a Bennett-type inequality for weakly dependent random variables. This also implies a Bernstein-type inequality, however, with different constants. In particular, the leading term in the denominator of the exponent differs from σ_n^2. This is a consequence of their method of proof which consists of replacing weakly dependent blocks of random variables by independent ones according to some coupling device (an analogous argument is used in Doukhan [10] for the case of absolute regularity).

(iv) Let $\mathcal{M}$ denote a sub-σ-algebra of $\mathcal{A}$ and $X \in \mathbb{R}^d$ be any random variable on the probability space $(\Omega, \mathcal{A}, \mathbb{P})$. We recall from Dedecker and Prieur [9] that

$$\varphi(\mathcal{M}, X) = \sup\{\|\mathbb{E}(g(X)|\mathcal{M}) - \mathbb{E}g(X)\|_\infty, \operatorname{Lip} g \leq 1\}.$$

Set $\varphi_k(r) = \max_{l \leq k} \frac{1}{l} \sup_{i+r \leq j_1 < j_2 < \cdots < j_l} \varphi(\sigma(\{X_j,\ j \leq i\}), (X_{j_1}, \ldots, X_{j_l}))$. The process $(X_t)_{t\in\mathbb{Z}}$ is called φ-dependent if $\varphi(r) = \sup_{k>0} \varphi_k(r)$ tends to 0 as $r \to \infty$. For an integrable function h and a Lipschitz function k this entails, for $s_1 \leq \cdots \leq s_u \leq t_1 \leq \cdots \leq t_v$, that

$$|\operatorname{Cov}(h(X_{s_1}, \ldots, X_{s_u}), k(X_{t_1}, \ldots, X_{t_v}))| \leq v\mathbb{E}|h(X_{s_1}, \ldots, X_{s_u})| \operatorname{Lip} k \varphi(t_1 - s_u).$$

Let us consider n φ-weakly dependent observations $X_1, \ldots, X_n$. We assume here only that $\mathbb{E}|X_t|^2 \leq 1$ and $\mathbb{P}(|X_t| \leq \sqrt{n}) = 1$ for all $1 \leq t \leq n$. If $\varphi(r) \leq \exp(-ar^b)$ for any $a > 0$, $b \in (0,1)$, then the assumptions of Theorem 1 hold with $\mu = 1/b$ and it is easy to check that there exists a constant C such that the parameter B_n is smaller than $C\sqrt{n}$. The notion of φ-weak dependence is adapted to expanding dynamical systems; see Dedecker and Prieur [9] for more details.

(v) A Bernstein-type inequality with σ_n^2 as a possible leading term in the denominator of the exponent has been derived in Kallabis and Neumann [19] under a weak dependence condition which is tailor-made for causal processes with an exponential decay of the coefficients of weak dependence. The result above is more general and is also applicable to interesting classes of processes where Kallabis and Neumann's [19] inequality does not apply; see Section 5 for a thorough discussion of examples.

(vi) Condition (1) in conjunction with (2) may be interpreted as a weak dependence condition in the sense that the covariances on the left-hand side tend to zero as the time gap between the two blocks of observations increases. Note that the supremum of expression (1) for all u-tuples $(s_1, \ldots, s_u)$ and all v-tuples $(t_1, \ldots, t_v)$ with $1 \leq s_1 \leq \cdots \leq s_u \leq t_1 \leq \cdots \leq t_v < \infty$ such that $t_1 - s_u = r$ is denoted by $C_{u+v}(r)$ in Doukhan and Louhichi's [11] initial paper. Conditions (1) and (2) are typically fulfilled for truncated versions of random variables from many time series models; see also Proposition 8 below. The constant K in (1) is included to possibly take advantage of a sparsity of data as it appears, for example, in nonparametric curve estimation.

(vii) For unbounded random variables, the coefficients $C_p(r)$ may still be bounded by an explicit function of the index p under a weak dependence assumption; see Lemma 10 below. For example, assume that $\mathbb{E}\exp(A|X_t|^a) \le L$ holds for some constants $A, a > 0$, $L < \infty$. Since the inequality $u^p \le p!\mathrm{e}^u$ ($p \in \mathbb{N}$, $u \ge 0$) implies that

$$u^m = (Aa)^{-m/a}(Aau^a)^{m/a} \le (Aa)^{-m/a}(m!)^{1/a}\mathrm{e}^{Au^a} \quad \forall m \in \mathbb{N}$$

we obtain that $\mathbb{E}|X_t|^m \le L(m!)^{1/a}(Aa)^{-m/a}$ holds for all $m \in \mathbb{N}$. Lemma 10 below provides then appropriate estimates for $C_p(r)$. With these bounds we can apply our Theorem 1 to get an exponential inequality of Bernstein-type; see for example Theorem 4.24 on page 102 in Saulis and Statulevicius [27] for a similar result in the case of α-mixing random variables.

A first Rosenthal-type inequality for weakly dependent random variables was derived by Doukhan and Louhichi [11] via direct expansions of the moments of even order. Unfortunately, the variance of the sum did not explicitly show up in their bound. Instead, a rough bound for this expression based on upper estimates was used. Using cumulant bounds in conjunction with Leonov and Shiryaev's formula we are able to obtain a tighter moment inequality which resembles the Rosenthal inequality in the independent case (see Rosenthal [25] and Johnson, Schechtman and Zinn [18] in the independent case, and Theorem 2.12 in Hall and Heyde [17] in the case of martingales).

Theorem 3. *Suppose that $X_1, \dots, X_n$ are real-valued random variables with zero mean, defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Let p be a positive integer. We assume that there exist constants $K, M < \infty$, and a nonincreasing sequence of real coefficients $(\rho(n))_{n\ge 0}$ such that, for all u-tuples $(s_1, \dots, s_u)$ and all v-tuples $(t_1, \dots, t_v)$ with $1 \le s_1 \le \cdots \le s_u \le t_1 \le \cdots \le t_v \le n$ and $u + v \le p$,*

$$|\mathrm{cov}(X_{s_1} \cdots X_{s_u}, X_{t_1} \cdots X_{t_v})| \le K^2 M^{u+v-2} \Psi(u, v)\rho(t_1 - s_u). \tag{7}$$

Furthermore, we assume that

$$\mathbb{E}|X_i|^{p-2} \le M^{p-2}.$$

Then, with $Z \sim \mathcal{N}(0, 1)$,

$$|\mathbb{E}S_n^p - \sigma_n^p \mathbb{E}Z^p| \le B_{p,n} \sum_{1\le u<p/2} A_{u,p} K^{2u}(M \vee K)^{p-2u} n^u,$$

where $B_{p,n} = (p!)^2 2^p \max_{2\le k\le p}\{\rho_{k,n}^{p/k}\}$, $\rho_{k,n} = \sum_{s=0}^{n-1}(s+1)^{k-2}\rho(s)$ and

$$A_{u,p} = \frac{1}{u!} \sum_{k_1+\cdots+k_u=p, k_i\ge 2\,\forall i} \frac{p!}{k_1! \cdots k_u!}.$$

Remark 4. (i) For even p, the above result implies that

$$\mathbb{E}S_n^p \le (p-1)(p-3)\cdots 1\sigma_n^p + B_{p,n} \sum_{1\le u<p/2} A_{u,p} K^{2u}(M \vee K)^{p-2u} n^u,$$

which resembles the classical Rosenthal inequality from the independent case. If $\sup_n B_{p,n} < \infty$ and $\sigma_n^2 \asymp n$, then the first term on the right-hand side is asymptotically dominating, as $n \to \infty$. This term is equal to the p-th moment of a Gaussian random variable with mean 0 and variance σ_n^2.

(ii) Doukhan and Louhichi [11] also obtained a Rosenthal-type inequality, however, essentially with $n \cdot \sum_{k=-\infty}^{\infty} |\mathbb{E}X_0X_k|$ instead of $\mathrm{var}(S_n)$ in the first term.

Remark 5. The inequality from Theorem 3 is well suited for proving a central limit theorem via the method of moments. Assume first that the random variables X_t are uniformly bounded, centered and satisfy condition (7) with $\lim_{s\to\infty} \rho(s)/s^p = 0$, for all $p > 0$. Then

$$\lim_{n\to\infty} \frac{\sigma_n^2}{n} = \sigma^2 = \sum_{k=-\infty}^{\infty} \mathbb{E}X_0X_k$$

is a convergent series, and thus the method of moments implies the central limit theorem,

$$\frac{1}{\sqrt{n}} S_n \xrightarrow{d}_{n\to\infty} \sigma Z.$$

The next section introduces the suitable frame of weak dependence in order to apply those results.

3. Weak dependence

A large class of examples for the assumptions in Theorem 1 to hold is provided by Doukhan and Louhichi [11] with weakly dependent processes. Consider a stationary process $(X_t)_{t\in\mathbb{Z}}$ with values in $\mathbb{R}^d$. Then any real-valued, Lipschitz and bounded function $Y_t = f(X_t)$ will be proved to satisfy the assumptions in the previous theorems if it is weakly dependent. We first recall this notion. Consider a process with values in $\mathbb{R}^d$ endowed with some norm $\|\cdot\|$. Let $h : (\mathbb{R}^d)^u \to \mathbb{R}$ be an arbitrary function. We set

$$\mathrm{Lip}\, h = \sup\left\{ \frac{|h(x_1,\dots,x_u) - h(y_1,\dots,y_u)|}{\|x_1 - y_1\| + \cdots + \|x_u - y_u\|} : (x_1,\dots,x_u) \neq (y_1,\dots,y_u) \right\}.$$

Moreover, Λ denotes the set of functions $h : \mathbb{R}^u \to \mathbb{R}$, for some $u \in \mathbb{N}$, such that $\mathrm{Lip}\, h < \infty$, and $\Lambda^{(1)} = \{h \in \Lambda : \|h\|_\infty \le 1\}$. For each $u \ge 1$, we identify the sets $(\mathbb{R}^d)^u$ and $\mathbb{R}^{du}$.

Definition 6 (*Doukhan and Louhichi [11]*). The sequence $(X_t)_{t\in\mathbb{Z}}$ is called $(\Lambda^{(1)}, \psi, \epsilon)$-weakly dependent if there exists a function $\psi : \mathbb{R}_+^2 \times \mathbb{N}^2 \to \mathbb{R}_+$ and a sequence $\epsilon = (\epsilon_r)_{r\in\mathbb{N}}$ decreasing to zero at infinity such that, for any $g_1, g_2 \in \Lambda^{(1)}$ with $g_1 : \mathbb{R}^{du} \to \mathbb{R}$, $g_2 : \mathbb{R}^{dv} \to \mathbb{R}$ $(u, v \in \mathbb{N})$ and for any u-tuple $(s_1,\dots,s_u)$ and any v-tuple $(t_1,\dots,t_v)$ with $s_1 \le \cdots \le s_u \le s_u + r \le t_1 \le \cdots \le t_v$, the following inequality is fulfilled:

$$|\mathrm{cov}(g_1(X_{s_1},\dots,X_{s_u}), g_2(X_{t_1},\dots,X_{t_v}))| \le \psi(\mathrm{Lip}\, g_1, \mathrm{Lip}\, g_2, u, v)\epsilon_r.$$

Important examples of processes correspond to the following choices of the function ψ:

(a) If $\psi(\mathrm{Lip}\, g_1, \mathrm{Lip}\, g_2, u, v) = v\mathrm{Lip}\, g_2$, then the sequence is called θ-dependent (see [11,7]) and we shall always denote $\epsilon_r = \theta_r$.
(b) If $\psi(\mathrm{Lip}\, g_1, \mathrm{Lip}\, g_2, u, v) = u\mathrm{Lip}\, g_1 + v\mathrm{Lip}\, g_2$, then the sequence is called η-dependent and we shall always denote $\epsilon_r = \eta_r$.
(c) If $\psi(\mathrm{Lip}\, g_1, \mathrm{Lip}\, g_2, u, v) = uv\mathrm{Lip}\, g_1\mathrm{Lip}\, g_2$, then the sequence is called κ-dependent and we shall always denote $\epsilon_r = \kappa_r$.
(d) If $\psi(\mathrm{Lip}\, g_1, \mathrm{Lip}\, g_2, u, v) = u\mathrm{Lip}\, g_1 + v\mathrm{Lip}\, g_2 + uv\mathrm{Lip}\, g_1\mathrm{Lip}\, g_2$, then the sequence is called λ-dependent and we shall always denote $\epsilon_r = \lambda_r$, as in Doukhan and Wintenberger [14]. This case includes the two previous ones.

Remark 7. (i) Assume that $(X_t)_{t\in\mathbb{Z}}$ is an $\mathbb{R}^d$-valued and stationary process which is $(\Lambda^{(1)}, \psi, \epsilon)$-weakly dependent. Then, for any Lipschitz-continuous function $F : \mathbb{R}^d \to \mathbb{R}$ with $\|F\|_\infty = M < \infty$ and $\mathrm{Lip}\, F \le 1$, the process $Y_t = F(X_t)$ is real valued, stationary, and $\|Y_t\|_\infty \le M$. Moreover, it is also $(\Lambda^{(1)}, \psi, \epsilon)$-weakly dependent.

(ii) In the more general case when $\mathrm{Lip}\, F$ possibly exceeds 1 (e.g., if the function F depends on the sample size in a statistical context), then weak dependence still holds where only $\psi(a, b, u, v)$ has to be replaced by $\psi^Y(a, b, u, v) = \psi(a\mathrm{Lip}\, F, b\mathrm{Lip}\, F, u, v)$. For the special cases of η, κ and λ weak dependence conditions, one may re-formulate this as $(Y_t)_{t\in\mathbb{Z}}$ is still an η, κ or λ-weakly dependent sequence but now we have to respectively consider the coefficients

$$\eta_r^Y = \mathrm{Lip}\, F\eta_r, \qquad \kappa_r^Y = \mathrm{Lip}^2 F\kappa_r, \qquad \lambda_r^Y = \max\{\mathrm{Lip}\, F, \mathrm{Lip}^2 F\}\lambda_r.$$

Now we relate the condition in Definition 6 to condition (1) which was also considered in Doukhan and Louhichi [11]. Suppose that $(X_t)_{t\in\mathbb{Z}}$ is a stationary sequence of real-valued random variables with $\|X_t\|_\infty \le M$ which satisfies the condition in Definition 6. To see the connection to condition (1), we consider functions g_1 and g_2 with $g_1(x_1, \ldots, x_u) = \prod_{i=1}^u f(x_i/M)$ and $g_2(x_1, \ldots, x_v) = \prod_{i=1}^v f(x_i/M)$, where $f(u) = u \vee (-1) \wedge 1$. (These functions satisfy $\mathrm{Lip}\, g_i \le 1/M$ and $\|g_i\|_\infty \le 1$.) The covariance in Definition 6 can be expressed as in inequality (1), up to a factor M^{u+v} since $g_1(X_{s_1}, \ldots, X_{s_u}) = X_{s_1} \cdots X_{s_u}/M^u$ and $g_2(X_{t_1}, \ldots, X_{t_v}) = X_{t_1} \cdots X_{t_v}/M^v$.

Proposition 8. *Assume that $\|X_t\|_\infty \le M$ and that the real-valued sequence $(X_t)_{t\in\mathbb{Z}}$ is $(\Lambda^{(1)}, \psi, \epsilon)$-weakly dependent. Then*

$$|\mathrm{cov}(X_{s_1} \cdots X_{s_u}, X_{t_1} \cdots X_{t_v})| \le M^{u+v}\psi(M^{-1}, M^{-1}, u, v)\epsilon_{t_1-s_u}. \tag{8}$$

Moreover, if $\epsilon_r = \exp(-ar)$, for some $a > 0$, then we may choose in inequality (2) $\mu = 1$ and $L_1 = L_2 = 1/(1 - \mathrm{e}^{-a})$.
If $\epsilon_r = \exp(-ar^b)$, for some $a > 0$, $b \in (0, 1)$, then we may choose $\mu = 1/b$ and L_1, L_2 appropriately as only depending on a and b.

Remark 9. (i) According to Proposition 8, $(\Lambda^{(1)}, \psi, \epsilon)$-dependence implies condition (1) to be fulfilled with

(a) $\Psi(u, v) = 2v$, $K^2 = M$ and $\rho(r) = \theta_r/2$, under θ-dependence,
(b) $\Psi(u, v) = u + v$, $K^2 = M$ and $\rho(r) = \eta_r$, under η-dependence,
(c) $\Psi(u, v) = uv$, $K = 1$ and $\rho(r) = \kappa_r$, under κ-dependence,
(d) $\Psi(u, v) = (u + v + uv)/2$, $K^2 = M \vee 1$ and $\rho(r) = 2\lambda_r$, under λ-dependence.

(ii) If a vector-valued process $(X_t)_{t\in\mathbb{Z}}$ is an η, κ or λ-weakly dependent sequence and $F : \mathbb{R}^d \to \mathbb{R}$ is a Lipschitz function with $\|F\|_\infty = M < \infty$, then the process $Y_t = F(X_t)$ is real-valued and the relation (1) holds with

(a) $\Psi(u, v) = 2v$, $K^2 = M\mathrm{Lip}\, F$ and $\rho(r) = \theta_r/2$, under θ-dependence,
(b) $\Psi(u, v) = u + v$, $K^2 = M\mathrm{Lip}\, F$ and $\rho(r) = \eta_r$, under η-dependence,
(c) $\Psi(u, v) = uv$, $K = \mathrm{Lip}\, F$ and $\rho(r) = \kappa_r$, under κ-dependence,
(d) $\Psi(u, v) = (u + v + uv)/2$, $K^2 = (M \vee 1)(\mathrm{Lip}^2 F \vee \mathrm{Lip}\, F)$ and $\rho(r) = 2\lambda_r$, under λ-dependence.

Inequality (8) together with the specifications of $\Psi(u, v)$ given in Remark 9 allow us to use the Bernstein-type inequality in Theorem 1 for sums of functions of weakly dependent sequences. Finally, we intend to determine sharp bounds for the coefficients $C_p(r)$ in the case of not necessarily bounded random variables.

Lemma 10. *Assume that the real-valued sequence $(X_t)_{t\in\mathbb{Z}}$ is η, κ or λ-weakly dependent and that $\mathbb{E}|X_t|^m \le M_m$, for some $m \in \mathbb{N}$. Then, for all $p < m$ and according to the type of the weak dependence condition:*

$$C_p(r) \le 2^{p+3} p^2 M_m^{\frac{p-1}{m-1}} \eta_r^{1-\frac{p-1}{m-1}}, \tag{9}$$

$$\le 2^{p+3} p^4 M_m^{\frac{p-2}{m-2}} \kappa_r^{1-\frac{p-2}{m-2}}, \tag{10}$$

$$\le 2^{p+3} p^4 M_m^{\frac{p-1}{m-1}} \lambda_r^{1-\frac{p-1}{m-1}}. \tag{11}$$

Remark 11. This lemma is the essential tool to provide a version of Theorem 3 which yields both a Rosenthal-type moment inequality and a rate of convergence for moments in the central limit theorem. We also note that this result does not require the assumption that the involved random variables are a.s. bounded. In fact, also the use of Theorem 1 does not require such a boundedness; see Remark 2-(vii) above.

4. Applications

In this section we present various applications of the previous results in probability and statistics. A first subsection addresses the basic case of the bounded LIL, while the others, specific to statistics, are concerned with empirical processes and nonparametric curve estimation.

4.1. Bounded LIL

Suppose that $(X_t)_{t\in\mathbb{Z}}$ is a stationary process satisfying the assumptions of Theorem 1 and that $\sigma^2 = \lim_{n\to\infty} \sigma_n^2/n > 0$. Then

$$\limsup_{n\to\infty} \frac{1}{\sigma\sqrt{2n\log\log n}} |S_n| \le 1 \quad \text{a.s.} \tag{12}$$

To prove (12), we select a subsequence $(n_k)_{k\in\mathbb{Z}}$ as $n_k = [a^k]$, for any $a > 1$. We obtain from Theorem 1 that, for $n_k \le n < n_{k+1}$ and any fixed c,

$$\begin{aligned}\mathbb{P}\left(\frac{|S_n|}{\sqrt{2n\sigma^2}} > c\sqrt{\log\log n_k}\right) &\le 2\exp\left(-c^2\log\log n_k \frac{\sigma^2 n}{\sigma_n^2}(1+o(1))\right)\\ &= 2\exp(-c^2\log\log n_k(1+o(1)))\\ &= O(k^{-c^2(1+o(1))}).\end{aligned}$$

This implies by the maximal inequality given in Theorem 2.2 in Móricz, Serfling and Stout [21] that

$$\mathbb{P}\left(\max_{n_k\le n<n_{k+1}} \frac{|S_n|}{\sqrt{2n\sigma^2}} > c\sqrt{\log\log n_k}\right) \le C\, k^{-c'}, \tag{13}$$

where $c' < c^2$ can be chosen arbitrarily close to c^2 and C is an appropriate finite constant; see the remark following the proof of Theorem 2.2 in Móricz, Serfling and Stout [21]. Since $\lim_{k\to\infty} \max_{n_k \le n < n_{k+1}} \frac{\log\log n}{\log\log n_k} = 1$ we conclude from (13) by the Borel–Cantelli lemma that

$$\limsup_{n\to\infty} \frac{1}{\sigma\sqrt{2n\log\log n}} |S_n| \le c \quad \text{a.s.},$$

for any $c > 1$, which in turn implies (12).

4.2. Kernel type density estimation in the supremum norm

Let $(X_t)_{t\in\mathbb{Z}}$ be a stationary η or λ-weakly dependent $\mathbb{R}^d$-valued process. We denote $K : \mathbb{R}^d \to \mathbb{R}$ a Lipschitz function, compactly supported with

$$\int_{\mathbb{R}^d} K(x)\mathrm{d}x = 1.$$

Then, if $h_n \to 0$,

$$\widehat{f}_n(x) = \frac{1}{nh_n^d} \sum_{t=1}^{n} K\left(\frac{X_t - x}{h_n}\right)$$

is an estimator of the marginal density f of X_0, provided this function exists. This section is devoted to derive asymptotic properties of

$$D_n = \sup_{x\in\mathbb{R}^d} |\widehat{f}_n(x) - f(x)|.$$

In the independent case, Giné and Guillou [15, Theorem 2.3] proved the following tight asymptotic bound for the supremum deviation of $\widehat{f}_n$ from its expectation:

$$\sup_{x\in\mathbb{R}^d} |\widehat{f}_n(x) - \mathbb{E}\widehat{f}_n(x)| = O\left(\sqrt{\log n/(nh_n^d)}\right) \quad \text{(a.s.)}. \tag{14}$$

We note that their conditions on the sequence of bandwidths, $h_n \searrow 0$, $nh_n^d/|\log h_n| \to \infty$, $|\log h_n|/\log\log n \to \infty$ and $h_n^d \le ch_{2n}^d$ $(c > 0)$, include our less general condition below.

We first analyze the bias. A Hölder class of β-regular probability densities is given by

$$\mathcal{F}_\infty(\beta, L) = \{f : f \text{ is a density}, |f^{(b)}(x+y) - f^{(b)}(x)| \le L|y|^{\beta-\lfloor\beta\rfloor}, \\ \forall x, y, \forall b = (b_1, \ldots, b_d) \text{ with } b_1 + \cdots + b_d = \lfloor\beta\rfloor\},$$

where $\lfloor\beta\rfloor$ denotes the greatest integer strictly less than β $(\beta > 0)$ and $f^{(b)} = \partial^{\lfloor\beta\rfloor} f/\partial^{b_1}x_1 \cdots \partial^{b_d}x_d$ is a partial derivative. Assume that the density f belongs to $\mathcal{F}_\infty(\beta, L)$ and that K is a kernel of order $\lfloor\beta\rfloor$, i.e., $\int P(x)K(x)\mathrm{d}x = P(0)$ for each polynomial of degree less than or equal to $\lfloor\beta\rfloor$. Then it follows from Taylor's formula that, for each $x \in \mathbb{R}^d$,

$$|\mathbb{E}\widehat{f}_n(x) - f(x)| \le L\int_{\mathbb{R}^d} \|u\|^\beta |K(u)|\mathrm{d}u \cdot \frac{h_n^\beta}{\lfloor\beta\rfloor!}.$$

Furthermore, it is easy to prove that if the joint densities f_k of (X_0, X_k) are bounded, uniformly with respect to $k > 0$, if in addition $nh_n^d \to \infty$ and $\eta_r = O(r^{-a})$ for some $a > 3$ or if $\lambda_r = O(r^{-a})$ for $a > 4$ are fulfilled, then (see [12])

$$\mathrm{var}(\widehat{f_n}(x)) \sim f(x) \int K^2(u)\mathrm{d}u/(nh^d).$$

This section has a double purpose: we first precise a uniform and almost sure convergence rate under weak dependence assumptions in the first subsection, while the second subsection provides precise constants in this asymptotics in the univariate case ($d = 1$). For this, we describe the necessary modifications of the arguments in Butucea and Neumann [5], who stated such a result under an absolute regularity assumption (β-mixing).

4.2.1. Multivariate case, rate of convergence

Using Remark 9 we obtain by Theorem 1 that the result (14) still holds if $h_n \geq Cn^{-c}$ for a constant $C > 0$ and for some $c = c(d)$. We set

$$X_{n,t} = \frac{1}{nh_n^d}\left(K\left(\frac{X_t - x}{h_n}\right) - \mathbb{E}K\left(\frac{X_t - x}{h_n}\right)\right).$$

Then $\widehat{f_n}(x) - \mathbb{E}\widehat{f_n}(x) = X_{n,1} + \cdots + X_{n,n}$. We write $X_{n,t} = F_n(X_t)$, for a function F_n which satisfies $\|F_n\|_\infty \leq 2\|K\|_\infty/(nh_n^d)$ and $\mathrm{Lip}\, F_n \leq \mathrm{Lip}\, K/(nh_n^{d+1})$. As mentioned above, mild conditions on the weak dependence coefficients imply that $\sigma_n^2 = \mathrm{var}(X_{n,1} + \cdots + X_{n,n})$ satisfies $nh_n^d\sigma_n^2 \to_{n\to\infty} f(x)\int K^2(u)\mathrm{d}u$. Now the Bernstein-type inequality from Theorem 1 writes here,

- in the case of η-dependent sequences with $M = \|F_n\|_\infty \sim 1/(nh_n^d)$, $K^2 \sim 1/(n^2h^{2d+1})$ and $B_n \sim 1/(nh_n^{2d+2})$. Now, with $t = C\sqrt{\log n/(nh_n^d)}$, Theorem 1 implies $|\widehat{f_n}(x) - \mathbb{E}\widehat{f_n}(x)| = O(\sqrt{\log n/nh_n^d})$, (a.s.) for each x of $\mathbb{R}^d$ if $h_n \geq n^{-c}$, for a $c < 1/(3d + 4)$. Note that an optimal uniform window width $h_n \sim (\log n/n)^{1/(2\beta+d)}$ yields the tight uniform a.s. rate in the case of $\beta > d + 2$.
- in the case of λ-dependence, with $M = \|F_n\|_\infty \sim 1/(nh_n^d)$, $K \sim 1/(nh_n^{d+1})$ and $B_n \sim 1/(nh_n^{2d+3})$. Now, again with $t = C\sqrt{\log n/(nh_n^d)}$, Theorem 1 implies the same as before if $h_n \geq n^{-c}$, for a $c < 1/(3d + 6)$. The choice of $h_n \sim (\log n/n)^{1/(2\beta+d)}$ yields the tight uniform a.s. rate if $\beta > d + 3$.

Now we assume that the kernel K is compactly supported, i.e., $K(x) = 0$ if $\|x\| \geq C$, for some $C < \infty$. Note that the exponential inequality used here allows us to extend the above convergence rates to hold uniformly with respect to $\|x\| \leq n^A$, for an arbitrary norm on $\mathbb{R}^d$ and any $A > 0$. To this end, we consider $\widehat{f_n}(x) - \mathbb{E}\widehat{f_n}(x)$ first on a sequence of increasingly fine grids $\mathcal{X}_n$ for the set $\{x : \|x\| \leq n^A\}$ with a cardinality of order n^γ, for some $\gamma < \infty$. It follows from our Bernstein-type inequality that

$$\mathbb{P}\left(\max_{x\in\mathcal{X}_n} |\widehat{f_n}(x) - \mathbb{E}\widehat{f_n}(x)| > C_\lambda\sqrt{\log n/(nh_n^d)}\right) = O(n^{-\lambda})$$

holds for arbitrary $\lambda < \infty$ and some finite C_λ. Moreover, it follows from the Lipschitz continuity of the kernel function K that

$$|\widehat{f_n}(x) - \widehat{f_n}(y)| \leq Ch_n^{-d}\|x - y\|,$$

for some $C < \infty$. Hence, with $\mathcal{X}_n$ sufficiently dense, we can conclude that

$$\mathbb{P}\left(\sup_{x:\|x\|\leq n^A} |\widehat{f_n}(x) - \mathbb{E}\widehat{f_n}(x)| > C_\lambda\sqrt{\log n/(nh_n^d)}\right) = O(n^{-\lambda}),$$

which implies by the Borel–Cantelli lemma that

$$\sup_{x:\|x\|\le n^A} |\widehat{f}_n(x) - \mathbb{E}\widehat{f}_n(x)| = O\left(\sqrt{\log n/(nh_n^d)}\right) \quad \text{(a.s.)}.$$

In order to prove that convergence holds uniformly over $\mathbb{R}^d$, and then deduce (14) for the case of weak dependence, assume that $\mathbb{E}\|X_0\|^a < \infty$ for some $a > 0$. Then, if A is large enough, $\sup_{\|x\|\ge n^A} f(x) \to_{n\to\infty} 0$ as fast as needed; indeed, the existence of a moment together with an Hölder assumption entail some Riemannian decay of f. The same occurs a.s. for $\widehat{f}_n$. We have that

$$\sum_{n=1}^{\infty} \mathbb{P}\left(\max_{1\le t\le n} \|X_t\| > n^A - Ch_n\right) \le \sum_{n=1}^{\infty} 1 \wedge \frac{\mathbb{E}(\|X_1\|^a + \cdots + \|X_n\|^a)}{(n^A - Ch_n)^a} < \infty,$$

for A sufficiently large. This implies by the Borel–Cantelli Lemma that $\sup_{x:\|x\|\ge n^A} |\widehat{f}_n(x)| \to_{n\to\infty} 0$ a.s. as fast as needed if A is large enough.

Note that for lower order regularities our result does not apply but the Bernstein inequality from Ragache and Wintenberger [23], following Doukhan and Louhichi [11], is still working and the a.s. convergence rate is now obtained only with the $\log n$ term replaced by $\log^{2+2/\mu} n$ under our previous assumptions.

4.2.2. Univariate case, exact asymptotics

In Butucea and Neumann [5], it was shown that asymptotically exact minimax results can be obtained by appropriately tuned kernel density estimators in the case that the real-valued observations ($d = 1$) $X_1, \ldots, X_n$ are mixing, all joint densities p_{X_0,X_k} are uniformly (in k) bounded, the error is measured in the supremum norm and the density is assumed to belong to some Hölder class. Armed with the new Bernstein-type inequality, we can easily generalize these results to the case of weakly dependent observations.

Based on observations $X_1, \ldots, X_n$, we consider the previous kernel estimator $\widehat{f}_n$, where $h_n = C(\log n/n)^{1/(2\beta+1)}$ for $f \in \mathcal{F}_\infty(\beta, L)$. Here the expectations are taken relatively to the parameter $\mathcal{L}$ which is the distribution of the process $(X_t)_{t\in\mathbb{Z}}$, where the marginal distributions of $\mathcal{L}$ have the density f and the process is associated with a certain decay rate of the dependence coefficients. The supremum risk of $\widehat{f}_n$ can be decomposed in a bias and a stochastic term as

$$\mathbb{E}[\|\widehat{f}_n - f\|_\infty] \le \|\mathbb{E}\widehat{f}_n - f\|_\infty + \mathbb{E}[\|\widehat{f}_n - \mathbb{E}\widehat{f}_n\|_\infty].$$

Now it follows, for all $f \in \mathcal{F}_\infty(\beta, L)$, that

$$\begin{aligned}
\|\mathbb{E}\widehat{f}_n - f\|_\infty &\le \sup_x \int \frac{1}{h_n} K\left(\frac{u-x}{h_n}\right)(f(u) - f(x))\mathrm{d}u \\
&\le \sup_x \int K(u)(f(x + h_n u) - f(x))\mathrm{d}u \\
&\le Lh_n^\beta B(\beta), \qquad (15)
\end{aligned}$$

where $B(\beta) = \sup_{g\in\mathcal{F}_\infty(\beta,L)} |\int K(u)(g(u) - g(0))\mathrm{d}u|$.

To analyze the stochastic term, we choose an arbitrary $\epsilon > 0$ and some appropriate $\epsilon' < \infty$. Then, with $\sigma_n = \sqrt{A(\beta, L)/(nh_n)}\|K\|_2\sqrt{2\log(1/h_n)}$, $A(\beta, L) = \max\{g(0) : g \in \mathcal{F}_\infty(\beta, L)\}$,

$$\begin{aligned}
\mathbb{E}[\|\widehat{f}_n - \mathbb{E}\widehat{f}_n\|_\infty] \le\ & (1+\epsilon)\sigma_n \\
& + (1+\epsilon')\sigma_n\, \mathbb{P}(\|\widehat{f}_n - \mathbb{E}\widehat{f}_n\|_\infty > (1+\epsilon)\sigma_n)
\end{aligned}$$

$$+ h_n^{-1} \|K\|_\infty \, \mathbb{P}(\|\widehat{f_n} - \mathbb{E}\widehat{f_n}\|_\infty > (1+\epsilon')\sigma_n)$$
$$= T_1 + T_2 + T_3, \tag{16}$$

say. The term T_1 can be made arbitrarily close to the desired σ_n. Since $\mathrm{var}(\widehat{f_n}(x)) \le A(\beta, L)/(nh_n)\|K\|_2^2(1+o(1))$ we can show, using chaining techniques in conjunction with the Bernstein-type inequality from Theorem 1, that

$$\mathbb{P}\left(\|\widehat{f_n} - \mathbb{E}\widehat{f_n}\|_\infty > (1+\epsilon)\sigma_n\right) = o(1) \tag{17}$$

and

$$\mathbb{P}(\|\widehat{f_n} - \mathbb{E}\widehat{f_n}\|_\infty > (1+\epsilon')\sigma_n) = O(n^{-\lambda}), \tag{18}$$

for arbitrary $\lambda < \infty$, provided ϵ' is sufficiently large. Actually, Butucea and Neumann [5] used a Bernstein-type inequality for mixing random variables in the course of deriving results analogous to (17) and (18), in particular to derive their equations (5.43), (5.45) and (5.52). It can be seen that we can apply our Bernstein-type inequality in the same way which finally leads to (17) and (18); for details see the proof of Theorem 4.2 in Butucea and Neumann [5].

From (15)–(18) we can conclude that

$$\limsup_{n\to\infty} \sup_{f\in\mathcal{F}_\infty(\beta,L)} \{(n/\log n)^{\beta/(2\beta+1)} \mathbb{E}\|\widehat{f_n} - f\|_\infty\}$$
$$\le \left(\frac{n}{\log n}\right)^{\beta/(2\beta+1)} \left\{\sqrt{\frac{A(\beta,L)}{nh_n}}\|K\|_2\sqrt{2\log(1/h_n)} + Lh_n^\beta B(\beta)\right\}. \tag{19}$$

Choosing now the kernel function K and the constant C in the definition of h_n in an optimum manner we can show that the right-hand side of (19) matches the known asymptotic minimax bound in the case of independent data (see [20]) which is also the minimax bound under mixing (see again [5]) and, hence, also the asymptotic risk bound in the more general framework of weak dependence.

4.3. Further applications

Exponential inequalities are quite useful when a large number of random sums has to be simultaneously bounded. Further possible applications in statistics include nonparametric estimation in Barron's classes (see [19]). A possible adaptation of the generic chaining arguments in Talagrand [29] appears very attractive but the use of classes of Lipschitz functions makes it difficult to use our results directly. We think that such applications deserve a detailed study which is, however, well beyond the scope of this paper.

5. Examples

Following Doukhan and Louhichi [11] and Doukhan and Wintenberger [14] we describe here some models which satisfy the previous weak dependence conditions.

- *Gaussian and associated processes* are κ-dependent with

$$\kappa_r = \max_{t\ge r} |\mathrm{cov}(X_0, X_t)|,$$

if they are stationary. Note that such covariances are nonnegative under association.

- *Bernoulli shifts.* Let $H : \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}$ be a measurable function. If the sequence $(\xi_t)_{t\in\mathbb{Z}}$ is independent and identically distributed on the real line, a Bernoulli shift with innovation process $(\xi_t)_{t\in\mathbb{Z}}$ is defined as

  $$X_t = H((\xi_{t-i})_{i\in\mathbb{Z}}), \quad t \in \mathbb{Z}.$$

  This sequence is stationary. The simplest case of an infinitely dependent Bernoulli shift is a moving average process

  $$X_t = \sum_{i=-\infty}^{\infty} a_i \xi_{t-i}. \tag{20}$$

  Bernoulli shifts are η-weakly dependent with $\eta_r \le 2\delta_{[r/2]}$, where

  $$\delta_s \ge \max_{0\le t\le s} \mathbb{E}\|X_0 - \widetilde{X}_0^{(t)}\|, \quad \widetilde{X}_0^{(t)} = H((\xi_i^{(t)})_{i\in\mathbb{Z}}),$$

  with $\xi_i^{(t)} = \xi_i$ if $|i| < t$ and $\xi_i^{(t)} = 0$ if $|i| \ge t$.
- *Chaotic Volterra models.* A Volterra process is a stationary process defined through a convergent Volterra expansion

  $$X_t = v_0 + \sum_{k=1}^{\infty} V_{k;t}, \quad \text{where } V_{k;t} = \sum_{i_1<\cdots<i_k} a_{k;i_1,\ldots,i_k} \xi_{t-i_1} \cdots \xi_{t-i_k}.$$

  Here v_0 is a constant and $(a_{k;i_1,\ldots,i_k})_{(i_1,\ldots,i_k)\in\mathbb{Z}^k}$ are real numbers for each $k \ge 1$. Let $p \ge 1$. Then this expression converges in L^p, provided that $\mathbb{E}|\xi_0|^p < \infty$ and the weights satisfy $\sum_{k=1}^{\infty} \sum_{i_1<\cdots<i_k} |a_{k;i_1,\ldots,i_k}|^p < \infty$. Those processes are η-dependent since δ_s from above is now the tail of the previous series. The forthcoming examples contain some particular models of this type.
- *LARCH(∞) models.* A vast literature is devoted to the study of conditionally heteroscedastic models; see Giraitis, Leipus and Surgailis [16]. It was shown in Doukhan, Teyssière and Winant [13] that a simple equation in terms of a vector-valued process allows a unified treatment of those models. Let $(\xi_t)_{t\in\mathbb{Z}}$ be an iid sequence of random $d \times m$ matrices, $(a_j)_{j\in\mathbb{N}}$ be a sequence of $m \times d$ matrices, and a be a vector in $\mathbb{R}^m$. A vector valued LARCH(∞) model is a solution of the recurrence equation

  $$X_t = \xi_t \left(a + \sum_{j=1}^{\infty} a_j X_{t-j} \right). \tag{21}$$

  Below we provide sufficient conditions for the following chaotic expansion

  $$X_t = \xi_t \left(a + \sum_{k=1}^{\infty} \sum_{j_1,\ldots,j_k\ge 1} a_{j_1} \xi_{t-j_1} a_{j_2} \xi_{t-j_1-j_2} \cdots a_{j_k} \xi_{t-j_1-\cdots-j_k} a \right). \tag{22}$$

  Such vector-valued LARCH(∞) models include a large variety of models, for example
  . *Bilinear models*, $X_t = \zeta_t(\alpha + \sum_{j=1}^{\infty} \alpha_j X_{t-j}) + \beta + \sum_{j=1}^{\infty} \beta_j X_{t-j}$, where the variables are real-valued and ζ_t is the innovation. Expansion (22) coincides then with the chaotic expansion in Giraitis, Leipus and Surgailis [16].
  . *GARCH (p, q) models*,

  $$\begin{cases} r_t = \sigma_t \varepsilon_t \\ \sigma_t^2 = \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 + \gamma_0 + \sum_{j=1}^{q} \gamma_j r_{t-j}^2 \end{cases}$$

where $\beta_i \geq 0$, $\gamma => 0$ and $\gamma_j \geq 0$ ($j \geq 1$), and the variables ε_t are centered at expectation; see Giraitis, Leipus and Surgailis [16].

. *ARCH*(∞) *processes*, given by $\begin{cases} r_t = \sigma_t \varepsilon_t \\ \sigma_t^2 = \beta_0 + \sum_{j=1}^{\infty} \beta_j \sigma_{t-j}^2 \end{cases}$.

Now we turn to the general case given by Eq. (21). Assume that $\lambda = \sum_{j\geq 1} \|a_j\| \mathbb{E}\|\xi_0\| < 1$. Then a stationary solution of Eq. (21) in L^1 is given as (22). With $A(x) = \sum_{j\geq x} \|a_j\|$, this solution is θ-weakly dependent with

$$\theta_r = \mathbb{E}\|\xi_0\| \left(\mathbb{E}\|\xi_0\| \sum_{k=1}^{r-1} k\lambda^{k-1} A\left(\frac{r}{k}\right) + \frac{\lambda^r}{1-\lambda} \right) \|a\|.$$

There exist some constants $K > 0$ and $b, C > 0$ such that

$$\theta_r \leq \begin{cases} K \dfrac{(\log(r))^{b\vee 1}}{r^b}, & \text{under Riemannian decay, } A(x) \leq Cx^{-b}, \\ K(q \vee \lambda)^{\sqrt{r}}, & \text{under geometric decay, } A(x) \leq Cq^x, q < 1. \end{cases}$$

- *Noncausal LARCH*(∞) *models* (now a_j is defined for $j \neq 0$) allow the same results of existence (only replace summation for $j > 0$ by summation for $j \neq 0$) and dependence is now of the η type with

$$\eta_r = \mathbb{E}\|\xi_0\| \left(\mathbb{E}\|\xi_0\| \sum_{0\leq 2k<r} k\lambda^{k-1} A\left(\frac{r}{2k}\right) + \frac{\lambda^{r/2}}{1-\lambda} \right) \|a\|, \quad A(x) = \sum_{|j|\geq x} \|a_j\|.$$

- *Stable Markov processes* are even θ-weakly dependent.

  We consider a stationary sequence satisfying a recurrence equation

$$X_t = F(X_{t-1}, \ldots, X_{t-d}, \xi_t),$$

where the sequence $(\xi_t)_{t\in\mathbb{Z}}$ is iid. In this case, $Y_t = (X_t, \ldots, X_{t-d+1})'$ is a Markov chain such that $Y_t = M(Y_{t-1}, \xi_t)$ with

$$M(x_1, \ldots, x_d, \xi) = (F(x_1, \ldots, x_d, \xi), x_1, \ldots, x_{d-1}).$$

Then $\mathbb{E}\|F(x,\xi) - F(y,\xi)\| \leq a\|x - y\|$ if $a = (\sum_{i=1}^d a_i)^{1/d} < 1$, where $a_i \geq 0$ are such that $\mathbb{E}|M(x,\xi_0) - M(y,\xi_0)| \leq \sum_{i=1}^d a_i |x_i - y_i|$, and $\|(x_1, \ldots, x_d)'\| = \max_{1\leq i\leq d} a^{i-1}|x_i|$ denotes a norm on $\mathbb{R}^d$.

In this setting, it is simple to derive that θ-dependence holds with $\theta_r = O(a^r)$ for:

. *Functional AR models*, $X_t = r(X_{t-1}, \ldots, X_{t-d}) + \xi_t$, if $\mathbb{E}|\xi_0| < \infty$ and $|r(u_1, \ldots, u_d) - r(v_1, \ldots, v_d)| \leq \sum_{i=1}^d a_i |u_i - v_i|$, for some $a_1, \ldots, a_d \geq 0$ with $a = (\sum_{i=1}^d a_i)^{1/d} < 1$.

. *ARCH-type processes*. With $d = 1$, let $M(u,z) = A(u) + B(u)z$, for suitable Lipschitz functions A, B. The corresponding iterative model satisfies the previous relation with

$$a = \mathrm{Lip}\,(A) + \mathbb{E}|\xi_0| \mathrm{Lip}\,(B) < 1.$$

Examples of such Markov processes are nonlinear AR(1) processes (case $B \equiv 1$), stochastic volatility models (case $A \equiv 0$), or classic ARCH(1) models (case $A(u) = \alpha u$, $B(u) = \sqrt{\beta + \gamma u^2}$ with $\alpha, \beta, \gamma \geq 0$, but here we also need $\mathbb{E}\xi_0 = 0$ and $\mathbb{E}\xi_0^2 < \infty$).

. *Branching type models*. Set $\xi_t = (\xi_t^{(1)}, \ldots, \xi_t^{(D)})'$, $d = 1$, and $D \geq 2$. Let now $A_1, \ldots, A_D$ be Lipschitz functions $\mathbb{R} \to \mathbb{R}$, and:

$$M(u, (z^{(1)}, \ldots, z^{(D)})) = \sum_{j=1}^{D} A_j(u) z^{(j)}, \quad (u, z^{(1)}, \ldots, z^{(D)}) \in \mathbb{R}^{D+1}.$$

For such kernels, we also require $a = \sum_{j=1}^{D} \mathrm{Lip}\,(A_j)\mathbb{E}|\xi_0^{(j)}| < 1$.

- *Compound processes* may be λ-dependent. Instead of independence, assume that the sequence $(\xi_t)_{t\in Z}$ is stationary and η-dependent with coefficients $(\rho_{\xi,r})_{r\geq 0}$. Then, for example, linear processes (20) are then η-weakly dependent with $\eta_r = \eta_{\xi,r/2} + \delta_{r/2}$ and $\delta_{r/2}$ according to the definition after Eq. (20). Such hereditary properties of weak dependence are unknown under mixing. We present here the results from Doukhan and Wintenberger [14] who sharpen analogous results from Borovkova, Burton and Dehling [4].

We now focus on specific examples of two sided linear sequences

$$X_t = \sum_{i\in\mathbb{Z}} b_i Y_{t-i}$$

with dependent inputs $(Y_t)_{t\in\mathbb{Z}}$. Let us denote by $(Y_t)_{t\in\mathbb{Z}}$ a weakly dependent innovation process. The coefficient λ is proved to be very useful to study Bernoulli shifts with such stationary innovations as in Lemma 12 below. Let $H : \mathbb{R}^{(\mathbb{Z})} \to \mathbb{R}$ be a measurable function on the space of finitely supported real valued sequences. We define $X_t = H(Y_{t-i}, i \in \mathbb{Z})$. Such models are proved to exhibit either λ or η-weak dependence properties. We set here $\|x\| = \sup_{i\in\mathbb{Z}} |x_i|$. In order to study weak dependence properties of $(X_t)_{t\in\mathbb{Z}}$, we assume that $H : \mathbb{R}^{(\mathbb{Z})} \to \mathbb{R}$ is such that for each $s \in \mathbb{Z}$, if $x, y \in \mathbb{R}^{(\mathbb{Z})}$ satisfy $x_i = y_i$ for each index $i \neq s$,

$$|H(x) - H(y)| \leq b_s |x_s - y_s|. \tag{23}$$

The following lemma proves both the existence and the weak dependence properties of such models.

Lemma 12 (*Doukhan and Wintenberger [14]*). *Let $(Y_t)_{t\in\mathbb{Z}}$ be a strictly stationary process with a finite moment of order $m > 2$ and let H satisfy the condition (23) for $l = 0$ and some nonnegative sequence $(b_s)_{s\in\mathbb{Z}}$ such that $\sum_j b_j < \infty$. Then,*

- *the process $X_n = H(Y_{n-j}, j \in \mathbb{Z}) := \lim_{I\to\infty} H\left(Y_{n-j}\mathbf{1}_{\{|j|\leq I\}}, j \in \mathbb{Z}\right)$ is a strongly stationary process with finite moments of order m.*
- *if the input process $(Y_t)_{t\in\mathbb{Z}}$ is λ-weakly dependent (the weak dependence coefficients are denoted $\lambda_{Y,r}$), then $(X_t)_{t\in\mathbb{Z}}$ is λ-weakly dependent with*

$$\lambda_k \leq \inf_{2r\leq k} \left[2\sum_{|i|\geq r} b_i \|Y_0\|_1 + (2r+1)^2 L^2 \lambda_{Y,k-2r} \right].$$

- *if the input process $(Y_t)_{t\in\mathbb{Z}}$ is η-weakly dependent (the weak dependence coefficients are denoted $\eta_{Y,r}$) then $(X_t)_{t\in\mathbb{Z}}$ is η-weakly dependent and*

$$\eta_k \leq \inf_{2r\leq k} \left[2\sum_{|i|\geq r} b_i \|Y_0\|_1 + (2r+1) L \eta_{Y,k-2r} \right].$$

Beyond linear functions one may think to noncausal ARCH(∞) inputs (with bounded inputs) and Doukhan and Wintenberger [14] consider more general examples for which H does not satisfy Eq. (23) as polynomial Volterra models.

6. Proofs

6.1. Proofs of the main theorems

First, note that it is not possible to adapt the classic method of proving the Bernstein inequality in the independent case since it makes heavy use of the independence; see Bennett [2, pp. 33–45].

One possible approach to proving an exponential inequality involves replacing blocks of weakly dependent random variables by independent ones and then applying an available inequality from the independent case. This was recently done by Dedecker and Prieur [8] who derived a Bennett inequality which then possibly implies a Bernstein-type inequality, however, with constants in the exponent different to ours. In particular, the leading term (asymptotically, as $n \to \infty$) in the denominator of the exponent will then differ from $\mathrm{var}(S_n)$ which is possible in our case. We also note here that an abstract presentation of cumulant techniques involving *Umbral Calculus* is presented in Rota and Chen [26].

Our proofs of Theorems 1 and 3 are based on a result of Bentkus and Rudzkis [3] which we quote here for the reader's convenience. Let ξ be an arbitrary real-valued random variable with $\mathbb{E}\xi = 0$ and finite moments of all orders. The k-th cumulant of ξ is defined as

$$\Gamma_k(\xi) = \frac{1}{i^k}\frac{\mathrm{d}^k}{\mathrm{d}t^k}\log \mathbb{E}\mathrm{e}^{\mathrm{i}t\xi}\Big|_{t=0}.$$

If there exist $\gamma \geq 1$, $\sigma^2 > 0$ and $B \geq 0$ such that

$$|\Gamma_k(\xi)| \leq \left(\frac{k!}{2}\right)^{\gamma} \sigma^2 B^{k-2} \quad \text{for all } k = 2, 3, \ldots,$$

then, for all $t \geq 0$,

$$\mathbb{P}(\xi \geq t) \leq \exp\left(-\frac{t^2/2}{\sigma^2 + B^{1/\gamma} t^{(2\gamma-1)/\gamma}}\right). \tag{24}$$

Note that the quotation of this result in Lemma 2.4 in the monograph by Saulis and Statulevicius [27, p. 19] contains a typo; it was correctly stated and proved in the initial paper, by Bentkus and Rudzkis [3, Lemma 2.1].

Before we proceed with the calculations, we recall some notions needed in the course of the proof. It follows from the definition of the cumulants that

$$\Gamma_k(S_n) = \sum_{1 \leq t_1, \ldots, t_k \leq n} \Gamma(X_{t_1}, \ldots, X_{t_k}), \tag{25}$$

where

$$\Gamma(X_{t_1}, \ldots, X_{t_k}) = \frac{1}{i^k}\frac{\partial^k}{\partial u_{t_1} \ldots \partial u_{t_k}} \log \mathbb{E}\mathrm{e}^{\mathrm{i}(u_1 X_1 + \cdots + u_n X_n)}\Big|_{u_1 = \cdots = u_n = 0}$$

are mixed cumulants. For any random variable Y with finite expectation, we define $\overline{Y} = Y - \mathbb{E}Y$. For $1 \leq t_1 \leq \cdots \leq t_k \leq n$, define so-called centred moments as $\overline{\mathbb{E}}(X_{t_1}, \ldots, X_{t_k}) = \mathbb{E}[X_{t_1} X_{t_2} \cdots \overline{X_{t_{k-1}}\, \overline{X_{t_k}}}]$ ($\overline{\mathbb{E}}(X_{t_1}) = \mathbb{E}X_{t_1}$). Statulevicius [28, Lemma 3] has shown that, for $1 \leq t_1 \leq \cdots \leq t_k \leq n$, the mixed cumulants can be expressed in terms of centred moments as

$$\Gamma(X_{t_1}, \ldots, X_{t_k}) = \sum_{\nu=1}^{k} (-1)^{\nu-1} \sum_{\bigcup_{p=1}^{\nu} I_p = I} N_\nu(I_1, \ldots, I_\nu) \prod_{p=1}^{\nu} \overline{\mathbb{E}} X_{I_p}, \tag{26}$$

where $\sum_{\bigcup_{p=1}^{\nu} I_p = I}$ denotes the summation over all unordered partitions in disjoint subsets $I_1, \ldots, I_\nu$ of the set $I = \{1, \ldots, k\}$; see also Eq. (1.63) in Saulis and Statulevicius [27], as a more easily available reference. Given such a partition, $\overline{\mathbb{E}} X_{I_p}$ stands for $\overline{\mathbb{E}}(X_{t_{i_1^{(p)}}}, \ldots, X_{t_{i_{k_p}^{(p)}}})$ if

$I_p = \{i_1^{(p)}, \ldots, i_{k_p}^{(p)}\}$ with $i_1^{(p)} < \cdots < i_{k_p}^{(p)}$. We arrange the subsets in the partitions such that $i_1^{(1)} < \cdots < i_1^{(\nu)}$. $N_\nu(I_1, \ldots, I_\nu)$ are certain nonnegative integers defined as follows. Let, for $i \in I$, $n_i(I_1, \ldots, I_\nu) = \#\{p : i_1^{(p)} < i < i_{k_p}^{(p)}\}$. Then

$$N_1(I) = 1$$

and, for $\nu \geq 2$,

$$N_\nu(I_1, \ldots, I_\nu) = \prod_{p=2}^{\nu} n_{i_1^{(p)}}(I_1, \ldots, I_\nu);$$

see Eqs. (4.36) and (4.37) in Saulis and Statulevicius [27, p. 80]. According to this, it follows that $N_\nu(I_1, \ldots, I_\nu) \neq 0$ if and only if $\{I_1, \ldots, I_\nu\}$ is connected, that is, $n_{i_1^{(p)}}(I_1, \ldots, I_\nu) > 0$ for all $p = 2, \ldots, \nu$. Furthermore, we have that

$$\sum_{\nu=1}^{k} \sum_{\bigcup_{p=1}^{\nu} I_p = I} N_\nu(I_1, \ldots, I_\nu) = (k-1)!; \tag{27}$$

see [27, Eq. (4.43)].

As a first step to deriving estimates for the cumulants of $S_n = X_1 + \cdots + X_n$, we derive estimates for the centred moments.

Lemma 13. *Suppose that $X_1, \ldots, X_n$ are zero mean random variables satisfying condition (1) from Theorem 1, for $u + v \leq k$. Furthermore, assume that $\mathbb{E}|X_i|^{k-2} \leq ((k-2)!)^\nu M^{k-2}$. Then, for $i \in \{1, \ldots, k-1\}$,*

$$|\overline{\mathbb{E}}(X_{t_1}, \ldots, X_{t_k})| \leq 2^k (k!)^\nu K^2 M^{k-2} \rho(t_{i+1} - t_i).$$

Proof. For the four cases (a)–(d), most parts of this proof are the same. Accordingly, we distinguish between them only when we apply condition (1), and at the end of this proof when certain upper estimates are summed up.

For $t_1 \leq \cdots \leq t_k$, $k \in \mathbb{N}$, we define the short-hand notation $Y_k = X_{t_k}$ and, for $1 \leq j < k$, $Y_j = X_{t_j} \overline{X_{t_{j+1}} \cdots \overline{X_{t_{k-1}} \overline{X_{t_k}}}}$.

Elementary calculations show, for $1 \leq j \leq i < k$, that

$$\begin{aligned} Y_j &= X_{t_j} Y_{j+1} - X_{t_j} \mathbb{E}[Y_{j+1}] \\ &= \cdots \\ &= X_{t_j} \cdots X_{t_i} Y_{i+1} - \sum_{l=j}^{i} X_{t_j} \cdots X_{t_l} \mathbb{E}[Y_{l+1}] \\ &= X_{t_j} \cdots X_{t_i} \overline{Y}_{i+1} - \sum_{l=j}^{i-1} X_{t_j} \cdots X_{t_l} \mathbb{E}[Y_{l+1}]. \end{aligned} \tag{28}$$

Since $\mathbb{E}X_{t_k} = 0$, in the special case of $i = k-1$ this becomes

$$Y_j = X_{t_j} \cdots X_{t_k} - \sum_{l=j}^{k-2} X_{t_j} \cdots X_{t_l} \mathbb{E}[Y_{l+1}]. \tag{29}$$

Without making use of the weak dependence assumption, we conclude recursively, for $3 \le j < k$, that

$$\mathbb{E}|Y_j| \le 2^{k-j}((k-j+1)!)^{\nu} M^{k-j+1}. \tag{30}$$

Hence, we obtain again in conjunction with (29) that

$$\begin{aligned}
C_{j,i} &:= |\mathrm{cov}(X_{t_j} \cdots X_{t_i}, Y_{i+1})| \\
&\le |\mathrm{cov}(X_{t_j} \cdots X_{t_i}, X_{t_{i+1}} \cdots X_{t_k})| \\
&\quad + \sum_{l=i+1}^{k-2} |\mathrm{cov}(X_{t_j} \cdots X_{t_i}, X_{t_{i+1}} \cdots X_{t_l} \mathbb{E}[Y_{l+1}])| \\
&\le K^2((k-j+1)!)^{\nu} M^{k-j-1} \Psi(i-j+1, k-i)\rho(t_{i+1}-t_i) \\
&\quad + \sum_{l=i+1}^{k-2} K^2((l-j+1)!)^{\nu} M^{l-j-1} \Psi(i-j+1, l-i)\rho(t_{i+1}-t_i) \\
&\quad \cdot 2^{k-l-1}((k-l)!)^{\nu} M^{k-l} \\
&\le K^2((k-j+1)!)^{\nu} M^{k-j-1} \Bigg\{ \Psi(i-j+1, k-i) \\
&\quad + \sum_{l=i+1}^{k-2} \Psi(i-j+1, l-i) 2^{k-l-1} \Bigg\} \rho(t_{i+1}-t_i).
\end{aligned} \tag{31}$$

Before we turn to estimating $|\overline{\mathbb{E}}(X_{t_1}, \ldots, X_{t_k})|$, we estimate the term in curly braces on the right-hand side of (31) in the four cases (a)–(d).

(a) If $\Psi(u, v) = 2v$, then

$$\begin{aligned}
&\Bigg\{ \Psi(i-j+1, k-i) + \sum_{l=i+1}^{k-2} \Psi(i-j+1, l-i) 2^{k-l-1} \Bigg\} \\
&= 2(k-i) + 2 \sum_{l=i+1}^{k-2} (l-i) 2^{k-l-1} \\
&\le 2^{k-i-1} \sum_{l'=1}^{\infty} l' 2^{1-l'} = 2^{k-i-1} \frac{\mathrm{d}}{\mathrm{d}p} \left(\frac{1}{1-p} \right) \Bigg|_{p=1/2} \\
&= 2^{k-i+1} =: \lambda_{j,i}^{(a)}.
\end{aligned} \tag{32}$$

(b) If $\Psi(u, v) = u + v$, then

$$\begin{aligned}
&\Bigg\{ \Psi(i-j+1, k-i) + \sum_{l=i+1}^{k-2} \Psi(i-j+1, l-i) 2^{k-l-1} \Bigg\} \\
&= (k-j+1) + \sum_{l=i+1}^{k-2} (l-j+1) 2^{k-l-1} \\
&= (k-j+1) \sum_{l=1}^{\infty} 2^{-l} + \sum_{l=i+1}^{k-2} (l-j+1) 2^{k-l-1}
\end{aligned}$$

$$\leq \sum_{l'=k-j+1}^{\infty} l' 2^{-l'+(k-j)} + \sum_{l'=i-j+2}^{k-j-1} l' 2^{-l'+(k-j)}$$
$$\leq 2^{k-j-1} \sum_{l'=i-j+2}^{\infty} l' 2^{1-l'}$$
$$= (i-j+3)2^{k-i-1} =: \lambda_{j,i}^{(b)}. \tag{33}$$

Here the last equation follows from

$$\sum_{l'=i-j+2}^{\infty} l' 2^{1-l'} = \frac{\mathrm{d}}{\mathrm{d}p} \left(\sum_{l'=i-j+2}^{\infty} p^{l'} \right) \Bigg|_{p=1/2}$$
$$= \frac{\mathrm{d}}{\mathrm{d}p} \left(\frac{p^{i-j+2}}{1-p} \right) \Bigg|_{p=1/2} = (i-j+3)2^{j-i}.$$

(c) If $\Psi(u, v) = uv$, then we obtain by (32) that

$$\left\{ \Psi(i-j+1, k-i) + \sum_{l=i+1}^{k-2} \Psi(i-j+1, l-i) 2^{k-l-1} \right\}$$
$$= (i-j+1) \left((k-i) + \sum_{l=i+1}^{k-2} (l-i) 2^{k-l-1} \right)$$
$$\leq (i-j+1) 2^{k-i} =: \lambda_{j,i}^{(c)}. \tag{34}$$

(d) If $\Psi(u, v) = \alpha(u+v) + (1-\alpha)uv$, then we obtain immediately by (33) and (34) that

$$\left\{ \Psi(i-j+1, k-i) + \sum_{l=i+1}^{k-2} \Psi(i-j+1, l-i) 2^{k-l-1} \right\}$$
$$\leq \alpha \lambda_{j,i}^{(b)} + (1-\alpha) \lambda_{j,i}^{(c)} =: \lambda_{j,i}^{(d)}. \tag{35}$$

Now we obtain from (28) that

$$|\mathbb{E}[Y_j]| \leq C_{j,i} + \sum_{l=j}^{i-1} |\mathbb{E}[X_{t_j} \cdots X_{t_l}]| \cdot |\mathbb{E}[Y_{l+1}]|.$$

Therefore, we obtain recursively that

$$|\overline{\mathbb{E}}(X_{t_1}, \ldots, X_{t_k})| = |\mathbb{E}[Y_1]|$$
$$\leq C_{1,i} + \sum_{l=1}^{i-1} (l!)^{\nu} M^l |\mathbb{E}[Y_{l+1}]|$$
$$\leq \cdots \leq C_{1,i} + \sum_{1 \leq l_1 \leq i-1} (l_1!)^{\nu} M^{l_1} C_{l_1+1,i}$$
$$+ \sum_{1 \leq l_1 < l_2 \leq i-1} (l_2!)^{\nu} M^{l_2} C_{l_2+1,i}$$
$$+ \cdots + \sum_{1 \leq l_1 < \cdots < l_{i-1} \leq i-1} ((i-1)!))^{\nu} M^{i-1} C_{i,i}. \tag{36}$$

At this point we have to distinguish again between the four cases (a)–(d). From (36), (31) and (32)–(35) we obtain the common upper estimate

$$|\overline{\mathbb{E}}(X_{t_1},\dots,X_{t_k})| \le K^2(k!)^\nu M^{k-2}\rho(t_{i+1}-t_i) \times \left\{ \lambda_{1,i}^{(\delta)} + \sum_{1\le l_1\le i-1} \lambda_{l_1+1,i}^{(\delta)} + \sum_{1\le l_1<l_2\le i-1} \lambda_{l_2+1,i}^{(\delta)} + \cdots + \sum_{1\le l_1<\cdots<l_{i-1}\le i-1} \lambda_{l_{i-1}+1,i}^{(\delta)} \right\}, \tag{37}$$

where $\delta = a, b, c, d$ refers to the four different cases. Now it remains to estimate the term in curly braces on the right-hand side of (37).

(a) If $\Psi(u, v) = 2v$, then

$$\{\dots\} = 2^{k-i+1} \sum_{l=0}^{i-1} \binom{i-1}{l} = 2^k. \tag{38}$$

(c) If $\Psi(u, v) = uv$, then we obtain

$$\begin{aligned}\{\dots\} &= \left\{ \lambda_{1,i}^{(c)} + \sum_{1\le l_1\le i-1} \lambda_{l_1+1,i}^{(c)} + \sum_{1\le l_1<l_2\le i-1} \lambda_{l_2+1,i}^{(c)} + \cdots + \sum_{1\le l_1<\cdots<l_{i-1}\le i-1} \lambda_{l_{i-1}+1,i}^{(c)} \right\} \\ &= 2^{k-i} \left(i + \sum_{1\le l_1\le i-1} (i-l_1) + \sum_{1\le l_1<l_2\le i-1} (i-l_2) + \cdots + \sum_{1\le l_1<\cdots<l_{i-1}\le i-1} (i-l_{i-1}) \right).\end{aligned}$$

Since

$$\begin{aligned}\sum_{1\le l_1<\cdots<l_m\le i-1} (i-l_m) &= i\binom{i-1}{m} - \sum_{l=m}^{i-1} l\binom{l-1}{m-1} \\ &= (m+1)\binom{i}{m+1} - m\binom{i}{m+1} = \binom{i}{m+1}\end{aligned}$$

we get

$$\{\dots\} = 2^{k-i}\left(i + \sum_{m=1}^{i-1} \binom{i}{m+1} \right) = 2^{k-i} \sum_{m=1}^{i} \binom{i}{m} < 2^k. \tag{39}$$

(b) If $\Psi(u, v) = u + v$, then we obtain analogously to (39) that

$$\{\dots\} = 2^{k-i-1}\left((i+2) + \sum_{1\le l_1\le i-1} (i-l_1+2) \right.$$

$$+\sum_{1\le l_1<l_2\le i-1}(i-l_2+2)+\cdots+\sum_{1\le l_1<\cdots<l_{i-1}\le i-1}(i-l_{i-1}+2)\Bigg)$$
$$<2^{k-i-1}\left(\sum_{m=1}^{i}\binom{i}{m}+2\sum_{m=0}^{i-1}\binom{i-1}{m}\right)$$
$$<2^k. \tag{40}$$

(d) If $\Psi(u,v)=\alpha(u+v)+(1-\alpha)uv$, then we easily obtain from $\lambda_{j,i}^{(d)}=\alpha\lambda_{j,i}^{(b)}+(1-\alpha)\lambda_{j,i}^{(c)}$, (40) and (39) that

$$\{\ldots\}<2^k. \tag{41}$$

The assertion of the lemma follows now from (37) and (38)–(41). □

Eqs. (25) and (26) and the result of Lemma 13 can now be used to derive estimates for the cumulants of S_n.

Lemma 14. *Suppose that the assertions of Lemma 13 are fulfilled. Then, for $k\ge 2$,*

$$|\Gamma_k(S_n)|\le n(k!)^{2+\nu}2^kK^2(K\vee M)^{k-2}\sum_{s=0}^{n-1}(s+1)^{k-2}\rho(s).$$

Proof. We deviate from the proof of similar results in Saulis and Statulevicius [27] since we are not able to follow all of their arguments. In particular, we cannot verify their Eq. (4.55) on page 94 which is crucial for their approach.

From (25) we obtain that

$$|\Gamma_k(S_n)|\le k!\sum_{1\le t_1\le\cdots\le t_k\le n}|\Gamma(X_{t_1},\ldots,X_{t_k})|. \tag{42}$$

According to (26) and Lemma 13, we have, for $1\le t_1\le\cdots\le t_k\le n$, that

$$|\Gamma(X_{t_1},\ldots,X_{t_k})|$$
$$\le\sum_{\nu=1}^{k}\sum_{\bigcup_{p=1}^{\nu}I_p=I}N_\nu(I_1,\ldots,I_\nu)\prod_{p=1}^{\nu}|\overline{\mathbb{E}}(X_{I_p})|$$
$$\le\sum_{\nu=1}^{k}\sum_{\bigcup_{p=1}^{\nu}I_p=I}N_\nu(I_1,\ldots,I_\nu)\prod_{p=1}^{\nu}2^{k_p}(k_p!)^\nu K^2M^{k_p-2}\min_{1<j\le k_p}\rho(t_{i_j^{(p)}}-t_{i_{j-1}^{(p)}}).$$

Note that we have, for any connected partition,

$$\max_{1\le p\le\nu}\max_{1<j\le k_p}\{t_{i_j^{(p)}}-t_{i_{j-1}^{(p)}}\}\ge\max_{1<i\le k}\{t_i-t_{i-1}\}.$$

Since $N_\nu(I_1,\ldots,I_\nu)=0$ if $\{I_1,\ldots,I_\nu\}$ is not connected we therefore obtain, in conjunction with (27), that

$$|\Gamma(X_{t_1},\ldots,X_{t_k})|$$
$$\le\sum_{\nu=1}^{k}\sum_{\bigcup_{p=1}^{\nu}I_p=I}N_\nu(I_1,\ldots,I_\nu)2^k(k!)^\nu K^2(K\vee M)^{k-2}\min_{1<i\le k}\rho(t_i-t_{i-1})$$

$$\leq (k-1)!2^k(k!)^\nu K^2(K\vee M)^{k-2}\min_{1<i\leq k}\rho(t_i-t_{i-1}).$$

This implies that

$$\sum_{1\leq t_1\leq\cdots\leq t_k\leq n}|\Gamma(X_{t_1},\ldots,X_{t_k})|$$

$$\leq n(k-1)!2^k(k!)^\nu K^2(K\vee M)^{k-2}\sum_{s_2,\ldots,s_k=0}^{\infty}\min_{2\leq i\leq k}\rho(s_i). \tag{43}$$

Since $\#\{(s_2,\ldots,s_k):0\leq s_i\leq s,\max\{s_2,\ldots,s_k\}=s\}\leq(k-1)(s+1)^{k-2}$ we obtain that

$$\sum_{s_2,\ldots,s_k=0}^{\infty}\min_{2\leq i\leq k}\rho(s_i)\leq(k-1)\sum_{s=0}^{\infty}(s+1)^{k-2}\rho(s),$$

this, with (42) and (43), yields the assertion of the lemma. □

Proof of Theorem 1. From Lemma 14 we obtain, for $k\geq 3$, that

$$|\Gamma_k(S_n)|\leq n(k!)^{2+\mu+\nu}2^kK^2L_1((K\vee M)L_2)^{k-2}$$
$$\leq\left(\frac{k!}{2}\right)^{2+\mu+\nu}2^{4+\mu+\nu}nK^2L_1\left(2(K\vee M)L_2\right)^{k-2},$$

which implies that

$$|\Gamma_k(S_n)|\leq\left(\frac{k!}{2}\right)^{2+\mu+\nu}A_nB_n^{k-2}$$

holds for all $k\geq 2$. The assertion of the theorem follows now from (24). □

Proof of Theorem 3. Recall that we have $\Gamma_1(S_n)=0$ and $\Gamma_2(S_n)=\sigma_n^2$. Therefore, Leonov and Shiryaev's formula (see [27], formula (1.53) on page 11) is written as

$$\mathbb{E}S_n^p=\sum_{u=1}^{[p/2]}\frac{1}{u!}\sum_{k_1+\cdots+k_u=p}\frac{p!}{k_1!\cdots k_u!}\Gamma_{k_1}(S_n)\cdots\Gamma_{k_u}(S_n). \tag{44}$$

Note that $\Gamma_1(S_n)=0$ implies that the inner sums can be reduced to indices such that $k_i\geq 2$ for all i. If p is an even number, then the summand with $u=p/2$ on the right-hand side of (44) is equal to

$$\frac{p!}{2^{p/2}(p/2)!}(\Gamma_2(S_n))^{p/2}=\mathbb{E}Z^p\sigma_n^p.$$

According to Lemma 14, we have, for $2\leq k\leq p$, that

$$|\Gamma_k(S_n)|\leq n(k!)^2 2^k\rho_{k,n}K^2L_1((M\vee K)L_2)^{k-2}.$$

Applying Hölder's inequality to the Gamma function Γ we see that $(k!)^{p/k}\leq p!$. Hence, we obtain that

$$\sum_{1\leq u<p/2}\frac{1}{u!}\sum_{k_1+\cdots+k_u=p,k_i\geq 2\,\forall i}\frac{p!}{k_1!\cdots k_u!}\Gamma_{k_1}(S_n)\cdots\Gamma_{k_u}(S_n)$$
$$\leq B_{p,n}\sum_{1\leq u<p/2}A_{u,p}K^{2u}(M\vee K)^{p-2u}n^u. \tag{45}$$

The assertion follows now from (44) and (45). □

6.2. Proofs of some auxiliary results

Proof of Proposition 8. Inequality (8) is obvious.

If $\rho(s) = \exp(-as)$, then

$$\sum_{s=0}^{\infty}(s+1)^k \mathrm{e}^{-as} \le \sum_{s=0}^{\infty}(s+1)\cdots(s+k)\mathrm{e}^{-as} = \frac{\mathrm{d}^k}{\mathrm{d}p^k}\left(\frac{1}{1-p}\right)\Big|_{p=\mathrm{e}^{-a}} = k!\frac{1}{(1-\mathrm{e}^{-a})^{k+1}}.$$

If $\rho(s) = \exp(-as^b)$, we have, for $k = 0$,

$$\sum_{s=0}^{\infty}\exp(-as^b) \le 1 + \int_0^{\infty}\exp(-au^b)\mathrm{d}u = 1 + \frac{1}{ba^{1/b}}\Gamma\left(\frac{1}{b}\right)$$

and, for $k \ge 1$,

$$\sum_{s=0}^{\infty}(s+1)^k\exp(-as^b) \le 1 + 2^k\mathrm{e}^{-a} + \sum_{s=2}^{\infty}(s+1)^k\exp(-as^b).$$

The last term on the right-hand side can be further estimated by

$$\int_1^{\infty}(u+2)^k\exp(-au^b)\mathrm{d}u \le 3^k\int_0^{\infty}u^k\exp(-au^b)\mathrm{d}u \le 3^k\frac{1}{ba^{(k+1)/b}}\Gamma\left(\frac{k+1}{b}\right).$$

Furthermore, it follows from Stirling's formula that

$$\begin{aligned}&\Gamma\left(\frac{k+1}{b}\right)\\ &\le \sqrt{2\pi}\exp\left(\frac{1}{12(\frac{k+1}{b}-1)}\right)\left(\frac{k+1}{b}-1\right)^{\frac{k+1}{b}-\frac{1}{2}}\exp\left(-\left(\frac{k+1}{b}-1\right)\right)\\ &\le \exp\left(1-\frac{1}{12(\frac{k+1}{b}-1)}\right)\left(\frac{1}{b}\right)^{\frac{k+1}{b}-\frac{1}{2}}(\sqrt{2\pi}(k+1)^{k+3/2}\mathrm{e}^{-(k+1)})^{1/b}\\ &\le C^k((k+1)!)^{1/b} \le C'^k(k!)^{1/b},\end{aligned}$$

for some appropriate $C, C' < \infty$. Putting these bounds together we obtain the assertion. □

Proof of Lemma 10. Set $g_T(x) = x \vee T \wedge (-T)$ and $\overline{X}_i = g_T(X_i) - \mathbb{E}g_T(X_i)$, for some $T > 0$. Then

$$|\mathrm{cov}(X_{s_1}\cdots X_{s_u}, X_{s_{u+1}}\cdots X_{s_p})| \le \sum_{j=0}^{p}|A_j|,$$

where $A_0 = \mathrm{cov}(\overline{X}_{s_1}\cdots\overline{X}_{s_u}, \overline{X}_{s_{u+1}}\cdots\overline{X}_{s_p})$, and for $1 \le j \le u$,

$A_j = \operatorname{cov}(X_{s_1}\cdots X_{s_{j-1}}(X_{s_j} - \overline{X}_{s_j})\overline{X}_{s_{j+1}}\cdots\overline{X}_{s_u}, \overline{X}_{s_{u+1}}\cdots\overline{X}_{s_p})$, and if $u < j \le p$, $A_j = \operatorname{cov}(X_{s_1}\cdots X_{s_u}, X_{s_{u+1}}\cdots X_{s_{j-1}}(X_{s_j} - \overline{X}_{s_j})\overline{X}_{s_{j+1}}\cdots\overline{X}_{s_p})$. First we bound $|A_j|$, for $1 \le j \le p$. We obtain by Hölder's inequality, with $\frac{p-1}{m} + \frac{1}{m'} = 1$, that

$$|A_j| \le 2\prod_{i<j}\|X_{s_i}\|_m \|X_{s_j} - \overline{X}_{s_j}\|_{m'}\prod_{i>j}\|\overline{X}_{s_i}\|_m.$$

From Jensen's inequality we get $\|\overline{X}_{s_i}\|_m \le 2\|g_T(X_{s_i})\|_m \le 2\|X_{s_i}\|_m \le 2M_m^{1/m}$. Furthermore, we have $\mathbb{E}|X_{s_j} - g_T(X_{s_j})|^{m'} \le \mathbb{E}[|X_{s_j}|^{m'} I(|X_{s_j}| > T)] \le T^{m'-m}M_m$ and $|\mathbb{E}g_T(X_{s_j})| = |\mathbb{E}(X_{s_j} - g_T(X_{s_j}))| \le \|X_{s_j} - g_T(X_{s_j})\|_{m'}$. This implies $\|X_{s_j} - \overline{X}_{s_j}\|_{m'} \le \|X_{s_j} - g_T(X_{s_j})\|_{m'} + |\mathbb{E}g_T(X_{s_j})| \le 2T^{1-m/m'}M_m^{1/m'}$. Therefore, we obtain that

$$\begin{aligned}|A_j| &= 2^{p-j+1}M_m^{\frac{p-1}{m}}\|X_{s_i} - \overline{X}_{s_i}\|_{m'}\\ &\le 2^{p-j+2}M_m^{\frac{p-1}{m}+\frac{1}{m'}}T^{1-\frac{m}{m'}} = 2^{p-j+2}M_m T^{p-m},\end{aligned}$$

which implies

$$\sum_{j=1}^{p}|A_j| \le 2^{p+2}M_m T^{p-m}. \tag{46}$$

Now we bound $|A_0|$. Coming back to the definition of weak dependence, we write it as $\operatorname{cov}(h(X_{s_1},\dots X_{s_u}), k(X_{s_{u+1}},\dots X_{s_p}))$ where $h(x_1,\dots,x_u) = \prod_{i=1}^{u}(g_T(x_i) - \mathbb{E}g_T(X_{s_i}))$, and $k(x_{u+1},\dots,x_p) = \prod_{i=u+1}^{p}(g_T(x_i) - \mathbb{E}g_T(X_{s_i}))$. Hence $\|h\|_\infty \le 2^u T^u$, $\|k\|_\infty \le 2^{p-u}T^{p-u}$, $\operatorname{Lip} h \le u2^{u-1}T^{u-1}$ and $\operatorname{Lip} k \le (p-u)2^{p-u-1}T^{p-u-1}$ and, according to the weak dependence in use, we get

$$\begin{aligned}|\operatorname{cov}(\overline{X}_{s_1}\cdots\overline{X}_{s_u}, \overline{X}_{s_{u+1}}\cdots\overline{X}_{s_p})| &\le (u^2(2T)^{u-1}(2T)^{p-u}\\ &\quad + (p-u)^2(2T)^{p-u-1}(2T)^u)\eta_r\\ &\le p^2 2^{p-1}T^{p-1}\eta_r,\\ |\operatorname{cov}(\overline{X}_{s_1}\cdots\overline{X}_{s_u}, \overline{X}_{s_{u+1}}\cdots\overline{X}_{s_p})| &\le (u(p-u))^2(2T)^{p-2}\kappa_r\\ &\le \frac{p^4}{4}2^{p-2}T^{p-2}\kappa_r = p^4 2^{p-4}T^{p-2}\kappa_r\end{aligned}$$

or

$$|\operatorname{cov}(\overline{X}_{s_1}\cdots\overline{X}_{s_u}, \overline{X}_{s_{u+1}}\cdots\overline{X}_{s_p})| \le p^2\left(\frac{p^2}{8}\vee 1\right)2^{p-1}T^{p-2}(T\vee 1)\lambda_r.$$

To conclude, we equilibrate the bounds for the term $|A_0|$ with that of the other terms $|A_j|$.

- Under η-dependence, set $T = (\frac{M_m}{\eta_r})^{\frac{1}{m-1}}$. Then we obtain

$$C_p(r) \le 2^{p+3}p^2 M_m^{\frac{p-1}{m-1}}\eta_r^{1-\frac{p-1}{m-1}}.$$

- Under κ-dependence, set $T = (\frac{M_m}{\kappa_r})^{\frac{1}{m-2}}$. Then we obtain

$$C_p(r) \le 2^{p+3}p^4 M_m^{\frac{p-2}{m-2}}\kappa_r^{1-\frac{p-2}{m-2}}.$$

- Under λ-dependence, set $T = (\frac{M_m}{\eta_r})^{\frac{1}{m-1}}$. Then we obtain

$$C_p(r) \le 2^{p+3} p^4 M_m^{\frac{p-1}{m-1}} \lambda_r^{1-\frac{p-1}{m-1}}. \quad \square$$

Acknowledgements

We thank two anonymous referees for their careful reading of the manuscript and many helpful comments.

References

[1] P. Ango Nze, P. Bühlmann, P. Doukhan, Nonparametric regression estimation under weak dependence beyond mixing and association, Ann. Statist. 30 (2002) 397–430.
[2] G. Bennett, Probability inequalities for the sum of independent random variables, J. Amer. Statist. Assoc. 57 (1962) 33–45.
[3] R. Bentkus, R. Rudzkis, On exponential estimates of the distribution of random variables, Lithuanian Math. J. 20 (1980) 15–30 (in Russian).
[4] S.A. Borovkova, R.M. Burton, H.G. Dehling, Limit theorems for functionals of mixing processes with applications to U-statistics and dimension estimation, Trans. Amer. Math. Soc. 353 (2001) 4261–4318.
[5] C. Butucea, M.H. Neumann, Exact asymptotics for estimating the marginal density of discretely observed diffusion processes, Bernoulli 11 (2005) 411–444.
[6] C. Coulon-Prieur, P. Doukhan, A triangular CLT for weakly dependent sequences, Statist. Probab. Lett. 47 (2000) 61–68.
[7] J. Dedecker, P. Doukhan, A new covariance inequality and applications, Stochastic Process. Appl. 106 (2003) 63–80.
[8] J. Dedecker, C. Prieur, Coupling for τ-dependent sequences and applications, J. Theoret. Probab. 17 (2004) 861–885.
[9] J. Dedecker, C. Prieur, New dependence coefficients. Examples and applications to statistics, Probab. Theory Related Fields 132 (2005) 203–236.
[10] P. Doukhan, Mixing: Properties and Examples, in: Lecture Notes in Statistics, vol. 85, Springer-Verlag, 1994.
[11] P. Doukhan, S. Louhichi, A new weak dependence condition and application to moment inequalities, Stochastic Process. Appl. 84 (1999) 313–342.
[12] P. Doukhan, S. Louhichi, Functional estimation for weakly dependent stationary time series, Scand. J. Statist. 28 (2001) 325–342.
[13] P. Doukhan, G. Teyssière, P. Winant, Vector valued LARCH infinity processes, in: P. Bertail, P. Doukhan, P. Soulier (Eds.), Dependence in Probability and Statistics, Springer, New York, 2005.
[14] P. Doukhan, O. Wintenberger, A central limit theorem under non causal weak dependence conditions and sharp moment assumptions, 2005. preprint.
[15] E. Giné, A. Guillou, Rates of strong uniform consistency for multivariate kernel density estimators, Ann. Inst. H. Poincaré 38 (2002) 907–921.
[16] L. Giraitis, L. Leipus, D. Surgailis, Recent advances in ARCH modelling, in: G. Teyssière, A. Kirman (Eds.), Long Memory in Economics, Springer Verlag, 2003.
[17] P. Hall, C.C. Heyde, Martingale Limit Theory and its Applications, Academic Press, New York, 1980.
[18] B.W. Johnson, G. Schechtman, J. Zinn, Best constants in moment inequalities for linear combinations of independent and exchangeable random variables, Ann. Probab. 13 (1985) 234–253.
[19] R.S. Kallabis, M.H. Neumann, An exponential inequality under weak dependence, Bernoulli 12 (2006) 333–350.
[20] A.P. Korostelev, M. Nussbaum, The asymptotic minimax constant for sup-norm loss in nonparametric density estimation, Bernoulli 5 (1999) 1099–1118.
[21] F.A. Móricz, J.A. Serfling, W.F. Stout, Moment and probability bounds with quasi-superadditive structure for the maximum partial sum, Ann. Probab. 10 (1982) 1032–1040.
[22] M.H. Neumann, E. Paparoditis, Goodness-of-fit tests for Markovian time series models, Technical Report No. 16/2005, Department of Mathematics and Statistics, University of Cyprus, 2005.
[23] N. Ragache, O. Wintenberger, Convergence rates for density estimates of weakly dependent time series, SAMOS, Paris 1, 2005. preprint.

[24] E. Rio, Théorie asymptotique pour des processus aléatoires faiblement dépendants, in: SMAI, Mathématiques et Applications, vol. 31, Springer, 2000.
[25] H.P. Rosenthal, On the subspaces of $L^p(p > 2)$ spanned by sequences of independent random variables, Israel J. Math. 8 (1970) 273–303.
[26] G.C. Rota, J. Chen, On the combinatorics of cumulants, J. Combin. Theory Ser. A 91 (2000) 283–304.
[27] L. Saulis, V.A. Statulevicius, Limit Theorems for Large Deviations, Mathematics and its Applications, Kluwer, 1991.
[28] V.A. Statulevicius, Limit theorems for random functions, Lithuanian Math. J. 10 (1970) 583–592 (in Russian).
[29] M. Talagrand, The Generic Chaining, Springer-Verlag, 2005.