A new covariance inequality and applications

Jérôme DEDECKER* and Paul DOUKHAN†

Abstract

We compare three dependence coefficients expressed in terms of conditional expectations, and we study their behaviour in various situations. Next, we give a new covariance inequality involving the weakest of those coefficients, and we compare this bound to that obtained by Rio (1993) in the strongly mixing case. This new inequality is used to derive sharp limit theorems, such as Donsker's invariance principle and Marcinkiewicz's strong law. As a consequence of a Burkhölder-type inequality, we obtain a deviation inequality for partial sums.

Mathematics Subject Classifications (1991): 60G10, 60F17, 60G48.
Key words: Weak dependence, mixingales, strong mixing, covariance inequalities, weak invariance principle, moment inequalities.
Short title: Covariance inequalities

1 Introduction

To describe the asymptotic behavior of certain time series, many authors have used one of the two following type of dependence: on one hand mixing properties, introduced in this context by Rosenblatt (1956), on an other hand martingales approximations or mixingales, following the works of Gordin (1969, 1973) and McLeisch (1975(a), 1975(b)). Concerning strongly mixing sequences, very deep and elegant results have been established: for recent works, we mention the monographs of Rio (2000) and Bradley (2002). However many classes of time series do not satisfy any mixing condition as it is quoted e.g. in Eberlein and Taqqu (1986) or Doukhan (1994). Conversely, most of such time series enter the scope of mixingales but limit theorems and moment inequalities are more difficult to obtain in this general setting. For instance, we cannot prove any empirical central limit theorem by using only mixingale-type coefficients.

*Laboratoire de Statistique Théorique et Appliquée, Université Paris 6, Site Chevaleret, 13 rue Clisson, 75013 Paris, France. E-mail: dedecker@ccr.jussieu.fr

†LS-CREST and UMR 8088, Université Cergy-Pontoise. E-mail: Paul.Doukhan@ensae.fr

Between those directions, Doukhan and Louhichi (1999) and simultaneously Bickel and Bühlmann (1999) introduced a new idea of weak dependence. The main advantage is that such a kind of dependence contains lots of pertinent examples (cf. Doukhan (2002) and Section 3 below) and can be used in various situations: empirical central limit theorems are proved in Doukhan and Louhichi (1999) and Borovkova, Burton and Dehling (2001), while applications to bootstrap are given by Bickel and Bühlmann (1999) and Ango Nzé, Bühlmann and Doukhan (2002). Such weak dependence conditions are easier to check than mixing properties and allow to cover empirical limit theorems which cannot be achieved *via* mixingales techniques.

Let us describe this type of dependence in more details. Following Coulon-Prieur and Doukhan (2000), we say that a sequence $(X_n)_{n\in\mathbb{Z}}$ of real-valued random variables is s-weakly dependent if there exists a sequence $(\theta_i)_{i\in\mathbb{N}}$ tending to zero at infinity such that: for any positive integer u, any function g from $\mathbb{R}^u$ to $[-1,1]$ and any Lipschitz function f from $\mathbb{R}$ to $[-1,1]$ with Lipschitz coefficient $\mathrm{Lip}(f)$, the following upper bound holds

$$\left|\mathrm{Cov}\,\left(g\big(X_{i_1},\ldots,X_{i_u}\big), f\big(X_{i_u+i}\big)\right)\right| \leq \theta_i \mathrm{Lip}(f) \tag{1.1}$$

for any u-tuple $i_1 \leq i_2 \leq \cdots \leq i_u$. We shall see in Remark 2 of Section 2 that such a coefficient can be expressed in terms of conditional expectations of some functions of the variables, so that it is easily comparable to mixingale-type coefficients. In Section 3 we present a large class of models for which (1.1) holds with a sequence θ_i decreasing to zero as i tends to infinity.

Our purpose in this paper is two-fold. We first compare the s-weak dependence coefficient with both strong mixing and mixingale-type coefficients (cf. Lemma 1, Section 2). Secondly, we establish in Proposition 1 of Section 4 a new covariance inequality involving a mixingale-type coefficient and comparable to that obtained by Rio (1993) in the strongly mixing case. With the help of this inequality, we give sharp versions of certain limit theorems. In Proposition 2 of Section 5, we give an upper bound for the variance of partial sums in terms of mixingale-type coefficients. In Corollary 2 of Section 6, we give three sufficient conditions, in terms of strong mixing, s-weak dependence and mixingale-type coefficients, for the partial sum process of a strictly stationary sequence to converge in distribution to a mixture of Brownian motion. Two of these conditons are new, and may be compared with the help of Lemma 1 to the well known condition of Doukhan, Massart and Rio (1994) for strongly mixing sequences. In the same way, we give in Theorem 3 of Section 6 a new sufficient condition for the partial sums of a s-weak dependent sequence to satisfy a Marcinkiewicz strong law of large numbers, and we compare this condition to that of Rio (1995) for strongly mixing sequences. Finally, we prove in Section 8 a Burkhölder-type inequality for mixingales, and we give an exponential inequality for the deviation of partial sums when the mixingale coefficients decrease with an exponential rate.

2 Three measures of dependence

Notations 1. Let X, Y be real valued random variables. Denote by
– Q_X the generalized inverse of the tail function $x \to \mathbb{P}(|X| > x)$.
– G_X the inverse of $x \to \int_0^x Q_X(u)du$.
– $H_{X,Y}$ the generalized inverse of $x \to \mathbb{E}(|X|\mathbb{1}_{|Y|>x})$.

Definition 1. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, and $\mathcal{M}$ a σ-algebra of $\mathcal{A}$. If $\mathrm{Lip}(g)$ is the lipschitz coefficient of the function g, define the class of functions $\mathcal{L}_1 = \{g : \mathbb{R} \mapsto \mathbb{R}, \|g\|_\infty < \infty, \mathrm{Lip}(g) \leq 1\}$. For any integrable real valued random variable X define

1. $\gamma(\mathcal{M}, X) = \|\mathbb{E}(X|\mathcal{M}) - \mathbb{E}(X)\|_1$.
2. $\theta(\mathcal{M}, X) = \sup\{\|\mathbb{E}(f(X)|\mathcal{M}) - \mathbb{E}(f(X))\|_1,\ f \in \mathcal{L}_1\}$.
3. $\alpha(\mathcal{M}, X) = \sup\{|\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|,\ A \in \mathcal{M}, B \in \sigma(X)\}$.

Let $(X_i)_{i\geq 0}$ be a sequence of integrable real valued random variables and let $(\mathcal{M}_i)_{i\geq 0}$ be a sequence of σ-algebras of $\mathcal{A}$. The sequence of coefficients γ_i is then defined by

$$\gamma_i = \sup_{k\geq 0} \gamma(\mathcal{M}_k, X_{i+k})\,. \tag{2.1}$$

The coefficients θ_i and α_i are defined in the same way.

Remark 1. The coefficient $\gamma(\mathcal{M}, X)$ was introduced by Gordin (1973) (see Theorem 1 of Section 6), and for the $\mathbb{L}^2$-norm by Gordin (1969) and McLeisch (1975a). According to the latter, we say that $\gamma(\mathcal{M}, X)$ is a mixingale-type coefficient.

Remark 2. Let $(X_i)_{i\in\mathbb{Z}}$ be a sequence of integrable random variables and consider the σ-fields $\mathcal{M}_k = \sigma(X_i, i \leq k)$. By homogeneity it is clear that θ_i defined as in (2.1) is the infimum over coefficients such that inequality (1.1) holds.

Remark 3. The usual strong mixing coefficients of the sequence $(X_i)_{i\in\mathbb{Z}}$ are defined by $\alpha'_i = \sup_{k\geq 0}\sup\{|\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|,\ A \in \mathcal{M}_k, B \in \sigma(X_j, j \geq k+i)\}$. In particular α'_i is greater than the coefficient α_i defined as in (2.1). To understand the difference between α_i and α'_i, note that the convergence of α'_i to zero implies that the sequence $(X_i)_{i\in\mathbb{Z}}$ is ergodic (see Notation 3 Section 6 for a definition), which is not true if we only assume that α_i goes to zero. A simple example of a nonergodic sequence for which $\alpha_i = 0$ for $i \geq 2$ is given in Rio (2000) page 67.

Remark 4. Let $(X_i)_{i\in\mathbb{Z}}$ be a stationary sequence of integrable random variables and $\mathcal{M}_k = \sigma(X_i, i \leq k)$. Due to the stationarity, the coefficient θ_i defined in (2.1) is equal to $\theta_i = \theta(\mathcal{M}_0, X_i)$. Now if θ_n tends to zero as n tends to infinity then so does $\|\mathbb{E}(f(X_0)|\mathcal{M}_n) - \mathbb{E}(f(X_0))\|_1$ for any Lipschitz function f. Applying the martingale-convergence theorem, we obtain that $\|\mathbb{E}(f(X_0)|\mathcal{M}_{-\infty}) - \mathbb{E}(f(X_0))\|_1 = 0$. This being true for any Lipschitz function, it can be extended to any function f such that $f(X_0)$ belongs to $\mathbb{L}^1$. Combining this result with Birkoff's ergodic theorem, we infer that for any f such that $f(X_0)$ belongs to $\mathbb{L}^1$

$$\frac{1}{n}\sum_{i=1}^{n} f(X_i) \quad \text{converges almost surely to} \quad \mathbb{E}(f(X_0))\,.$$

Of course, this is no longer true if we only assume that $\gamma_i = \gamma(\mathcal{M}_0, X_i)$ tends to zero as n tends to infinity.

The next Lemma shows how to compare these coefficients.

Lemma 1 *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $\mathcal{M}$ be a σ-algebra of $\mathcal{A}$. Let X be an integrable and real valued random variable. For any random variable Y such that $Q_Y \geq Q_X$,*

$$G_Y(\gamma(\mathcal{M}, X)/2) \leq G_Y(\theta(\mathcal{M}, X)/2) \leq 2\alpha(\mathcal{M}, X) \tag{2.2}$$

Analogously, if $(X_i)_{i\geq 0}$ is a sequence of integrable and real-valued random variables, $(\mathcal{M}_i)_{i\geq 0}$ is a sequence of σ-algebra of $\mathcal{A}$ and X is a random variable such that $Q_X \geq \sup_{i\geq 0} Q_{X_i}$, then

$$G_X(\gamma_i/2) \leq G_X(\theta_i/2) \leq 2\alpha_i \tag{2.3}$$

Remark 5. In particular, for any conjugate exponent p and q, we infer from (2.2) that $\theta(\mathcal{M}, X) \leq 2\|X\|_p(2\alpha(\mathcal{M}, X))^{1/q}$. When $p = \infty$, this is a direct consequence of Ibragimov's inequality (1962). In fact, the coefficient $\alpha(\mathcal{M}, X)$ may be defined by $4\alpha(\mathcal{M}, X) = \sup\{\|\mathbb{E}(f(X)|\mathcal{M}) - \mathbb{E}(f(X))\|_1, \|f\|_\infty \leq 1\}$ (see for instance Theorem 4.4 in Bradley (2002)).

Proof of Lemma 1. It is enough to prove (2.2). Clearly $\gamma(\mathcal{M}, X) \leq \theta(\mathcal{M}, X)$. The first inequality is thus proved by using G_Y's monotonicity. In order to prove the second one, there is no loss of generality in assuming that $f \in \mathcal{L}_1$ satisfies $f(0) = 0$. Hence $|f(x)| \leq |x|$ and consequently $G_{f(X)} \geq G_X \geq G_Y$. With G_Y's monotonicity this yields successively,

$$\begin{aligned} G_Y(\theta(\mathcal{M}, X)/2) &= \sup_{f\in\mathcal{L}_1} G_Y(\|\mathbb{E}(f(X)|\mathcal{M}) - \mathbb{E}(f(X))\|_1/2) \\ &\leq \sup_{f\in\mathcal{L}_1} G_{f(X)}(\|\mathbb{E}(f(X)|\mathcal{M}) - \mathbb{E}(f(X))\|_1/2)\,. \end{aligned}$$

The result follows by using Rio's (1993) covariance inequality:

$$\|\mathbb{E}(f(X)|\mathcal{M}) - \mathbb{E}(f(X))\|_1 \leq 2\int_0^{2\alpha(\mathcal{M},X)} Q_{f(X)}(u)du = 2G^{-1}_{f(X)}(2\alpha(\mathcal{M}, X)).$$

3 Examples

Most of the examples of Section 3.1 and 3.2 are studied in Doukhan and Louhichi (1999) and Doukhan (2002).

3.1 Causal Bernoulli shifts

Definition 2. Let $(\xi_i)_{i\in\mathbb{Z}}$ be a stationary sequence of real-valued random variables and H be a measurable function defined on $\mathbb{R}^{\mathbb{N}}$. The stationary sequence $(X_n)_{n\in\mathbb{Z}}$ defined by $X_n = H(\xi_n, \xi_{n-1}, \xi_{n-2}, \ldots)$ is called a causal Bernoulli shift. For such a function H, define the coefficient δ_i by

$$\delta_i = \|H(\xi_0, \xi_{-1}, \xi_{-2}, \ldots) - H(\xi_0, \xi_{-1}, \xi_{-2}, \ldots, \xi_{-i}, 0, 0, \ldots)\|_1. \qquad (3.1)$$

Causal Bernoulli shifts with i.i.d. innovations $(\xi_i)_{i\in\mathbb{Z}}$ satisfy $\theta_i \le 2\delta_i$ (see for instance Rio (1996)). Examples of such situations follows:

• Causal linear process: if $X_n = \sum_{j\ge 0} a_j \xi_{n-j}$ then $\theta_i \le 2\|\xi_0\|_1 \sum_{j\ge i} |a_j|$. In some particular cases we can also obtain an upper bound for the usual strong mixing coefficients α'_i defined in Remark 3: If $a_j = O(j^{-a})$, $\mathbb{E}(|\xi_0|^{1+\delta}) < \infty$ and the distribution of ξ_0 is absolutely continuous then we have $\alpha'_i = O(i^{-(a-2)\delta/(1+\delta)})$ as soon as $a > 2 + 1/\delta$ (see Pham and Tran (1985)). Hence, summability of the series $\sum_{i\ge 0} \alpha'_i$ holds as soon as $a > 3 + 2/\delta$, while summability of $\sum_{i\ge 0} \theta_i$ requires only $a > 2$.

• Other non-Markovian examples of Bernoulli shifts are given in Doukhan (2002). The most striking one is the $ARCH(\infty)$ processes from Giraitis, Kokoszka and Leipus (2000) subject to the recursion $X_t = (a_0 + \sum_{j=1}^{\infty} a_j X_{t-j})\xi_t$. Such models have a stationary representation with chaotic expansion

$$X_t = a_0 \sum_{\ell=1}^{\infty} \sum_{j_1=1}^{\infty} \cdots \sum_{j_\ell=1}^{\infty} a_{j_1} \cdots a_{j_\ell} \xi_{t-j_1} \cdots \xi_{t-(j_1+\cdots+j_\ell)}$$

under the simple assumption $c = \|\xi_0\|_1 \sum_{j=1}^{\infty} |a_j| < 1$. In this case, we have that $\theta_i \le 2c^L + 2\|\xi_0\|_1 (1-c)^{-1} \sum_{j\ge J} |a_j|$ for any $JL \le i$. Indeed, it suffices to approximate the series X_n by $i-$dependent ones $X_{n,i}$ obtained when considering finite sums for which the previous series are subject to the restrictions $\ell \le L$ and $j_1, \ldots, j_\ell \le J$, and to note that $\theta_i \le 2\|X_0 - X_{0,i}\|_1$. This gives rise to various dependence rates: if $a_j = 0$ for large enough $j \ge J$ then $\theta_i = O(c^{i/J})$. If $a_j = O(j^{-b})$ for some $b > 1$, then $\theta_i = O((\ln i/i)^b)$. If $a_j = b^j$ for some $b < 1$, then $\theta_i = O(\exp(-\sqrt{i \ln b \ln c}))$.

3.2 Stable Markov chains

Let $(X_n)_{n\ge 0}$ be a stationary Markov chain with value in a metric space (E, d) and satisfying the equation $X_n = F(X_{n-1}, \xi_n)$ for some measurable map F and some i.i.d. sequence $(\xi_i)_{i\geq 0}$. Denote by μ the law of X_0 and by $(X_n^x)_{n\geq 0}$ the chain starting from $X_0^x = x$. If f is a L-Lipschitz function from $\mathbb{E}$ to $\mathbb{R}$, it is easy to see that

$$\|\mathbb{E}(f(X_i)|X_0) - \mathbb{E}(f(X_i))\|_1 \leq L \iint \mathbb{E}(d(X_i^x, X_i^y))\mu(dx)\mu(dy)\,.$$

Consequently, if the Makov chain satisfies $\mathbb{E}(d(X_i^x, X_i^y)) \leq \delta_i d(x,y)$ for some decreasing sequence δ_i, we have that $\theta_i \leq \delta_i \mathbb{E}(d(X_0, X_0'))$ where X_0' is independent and distributed as X_0. Duflo (1997) studied the case where $\mathbb{E}(d(X_1^x, X_1^y)) \leq kd(x,y)$ for some constant $k < 1$, for which $\delta_i = k^i$. We refer to the nice review paper by Diaconis and Friedmann (1999) for various examples of iterative random maps $X_n = F(X_{n-1}, \xi_n)$.

Ango Nzé (1998) obtained geometrical and polynomial mixing rates for functional autoregressive processes $X_n = f(X_{n-1}) + \xi_n$ when the common distribution of the ξ_i has an absolutely continuous component. If this is not true, such a process may not have any mixing property although it is s-weakly dependent (see Example 2 of Section 3.3). Let us give a simple example of a non contracting function f for which the coefficient θ_i decreases with a polynomial rate: for δ in $[0,1[$, C in $]0,1]$ and $S \geq 1$, let $\mathcal{L}(C,\delta)$ be the class of 1-Lipschitz functions f satisfying

$$f(0) = 0 \quad \text{and} \quad |f'(t)| \leq 1 - C(1+|t|)^{-\delta} \quad \text{almost everywhere}$$

and $ARL(C,\delta,S)$ be the class of Markov chains on $\mathbb{R}$ defined by $X_n = f(X_{n-1}) + \xi_n$ where $f \in \mathcal{L}(C,\delta)$ and $\|\xi_0\|_S < \infty$. Dedecker and Rio (2000) proved that for any Markov kernel belonging to $ARL(C,\delta,S)$, there exists an unique invariant probability μ and moreover $\mu(|x|^{S-\delta}) < \infty$. Furhter, if $S > 1+\delta$, the stationary chain is s-weakly dependent with rate $\theta_i = O(n^{(\delta+1-S)/\delta})$.

3.3 Some more precise computations

We now give the precise behaviour of the coefficients γ_i, θ_i and α_i in two simple situations. In the first example, $(X_i)_{i\in\mathbb{Z}}$ is a martingale-difference sequence, $\gamma_i = 0$ while θ_i (and hence α_i) does not even go to zero except if $(X_i)_{i\in\mathbb{Z}}$ is i.i.d. In the second case, $(X_i)_{i\in\mathbb{Z}}$ is an autoregressive process, $\theta_i = \lambda_i = 2^{-i-1}$ while $\alpha_i \equiv 1/4$.

Example 1. Let $(\varepsilon_i)_{i\in\mathbb{Z}}$ be a sequence of i.i.d. mean-zero random variables and Y be an integrable random variable independent of $(\varepsilon_i)_{i\in\mathbb{Z}}$. Consider the strictly stationary sequence $(X_i)_{i\in\mathbb{Z}}$ defined by $X_i = Y\varepsilon_i$ and take $\mathcal{M}_i = \sigma(X_k, k \leq i)$. Since $\mathbb{E}(X_i|\mathcal{M}_{i-1}) = 0$ we infer that $\gamma_i = 0$. Now if θ_i tends to zero, we know from Remark 4 that for any f such that $f(X_0)$ belongs to $\mathbb{L}^1$, the sequence $n^{-1}\sum_{i=1}^n f(X_i)$ converges almost surely to $\mathbb{E}(f(X_0)) = \mathbb{E}(f(Y\varepsilon_0))$. Comparing this limit with that given by the strong law of large numbers, we infer that if θ_i tends to zero, then

$$\mathbb{E}(f(Y\varepsilon_0)) = \int f(Yx)\mathbb{P}^{\varepsilon_0}(dx) \quad \text{almost surely.} \tag{3.2}$$

Taking $f = |\cdot|$ in (3.2) we obtain that $\|\varepsilon_0\|_1(|Y| - \|Y\|_1) = 0$ almost surely, which means that either $\|\varepsilon_0\|_1 = 0$ or $|Y|$ is almost surely constant. In the second case, if Y is not almost surely constant we infer from (3.2) that ε_0 must be symmetric, so that the sequence $(X_i)_{i\in\mathbb{Z}}$ is i.i.d. In any cases, we conclude that θ_i tends to zero if and only if the sequence $(X_i)_{i\in\mathbb{Z}}$ is i.i.d., which is not true in general.

Example 2. Let $(\varepsilon_i)_{i\in\mathbb{Z}}$ be a sequence of independent random variables with common Bernoulli distribution $\mathcal{B}(1/2)$. Consider the linear process $X_i = \sum_{k=0}^{\infty} 2^{-k}\varepsilon_{i-k}$ and define the σ-algebras $\mathcal{M}_i = \sigma(X_k, k \le i)$. For such a process, it is well known that $\alpha_i \equiv 1/4$ (see for instance Doukhan (1994)). To compute γ_i, note that

$$\gamma_i = \|\mathbb{E}(X_i|\mathcal{M}_0) - 1\|_1 = 2^{-i}\Big\|\sum_{k\ge 0} 2^{-k}\Big(\varepsilon_k - \frac{1}{2}\Big)\Big\|_1 = 2^{-i-1}. \tag{3.3}$$

To evaluate θ_i, introduce the variables $V = \sum_{k=0}^{i-1} 2^{-k}\varepsilon_{i-k}$ and $U = \sum_{k=i}^{\infty} 2^{i-k-1}\varepsilon_{i-k}$. Note that U is uniformly distributed over $[0,1]$ and that $X_i = V + 2^{-i+1}U$. Clearly

$$\theta_i = \sup_{f\in\mathcal{L}_1}\Big\|\int f(2^{-i+1}U + v)\mathbb{P}^V(dv) - \mathbb{E}\Big(\int f(2^{-i+1}U + v)\mathbb{P}^V(dv)\Big)\Big\|_1. \tag{3.4}$$

The function $u \to \int f(2^{-i+1}u + v)\mathbb{P}^V(dv)$ being 2^{-i+1}-Lipschitz, we infer from (3.4) that $\theta_i \le 2^{-i+1}\sup_{f\in\mathcal{L}_1}\|f(U) - \mathbb{E}(f(U))\|_1$, or equivalently that

$$\theta_i \le 2^{-i+1}\sup\Big\{\int_0^1 |g(x)|dx,\ g \in \mathcal{L}_1,\ \int_0^1 g(x)dx = 0\Big\}.$$

It is easy to see that the supremum on the right hand side is $1/4$, so that $\theta_i \le 2^{-i-1}$. Since $\theta_i \ge \gamma_i$, we conclude from (3.3) that $\theta_i = \gamma_i = 2^{-i-1}$.

4 A covariance inequality

Recall that for two real-valued random variables X, Y the functions G_X and $H_{X,Y}$ have been defined in Notations 1 of Section 2. The main result of this paper is the following:

Proposition 1 *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $\mathcal{M}$ be a σ-algebra of $\mathcal{A}$. Let X be an integrable random variable and Y be an $\mathcal{M}$-measurable random variable such that $|XY|$ is integrable. The following inequalities hold*

$$|\mathbb{E}(YX)| \le \int_0^{\|\mathbb{E}(X|\mathcal{M})\|_1} H_{X,Y}(u)du \le \int_0^{\|\mathbb{E}(X|\mathcal{M})\|_1} Q_Y \circ G_X(u)du. \tag{4.1}$$

If furthermore Y is integrable, then

$$|\mathrm{Cov}(Y, X)| \le \int_0^{\gamma(\mathcal{M},X)} Q_Y \circ G_{X-\mathbb{E}(X)}(u)du \le 2\int_0^{\gamma(\mathcal{M},X)/2} Q_Y \circ G_X(u)du. \tag{4.2}$$

Applying Lemma 1, we also have that

$$\int_0^{\gamma(\mathcal{M},X)/2} Q_Y\circ G_X(u)du \le \int_0^{\theta(\mathcal{M},X)/2} Q_Y\circ G_X(u)du \le \int_0^{2\alpha(\mathcal{M},X)} Q_Y(u)Q_X(u)du\,. \qquad (4.3)$$

Remark 6. Combining (4.2) and (4.3) we obtain the inequality

$$|\mathrm{Cov}(Y,X)| \le 2\int_0^{2\alpha(\mathcal{M},X)} Q_Y(u)Q_X(u)du,$$

which was proved by Rio (1993). A converse inequality is given in Theorem (1.1)(b) of the same paper.

Proof of Proposition 1. We start from the inequality

$$|\mathbb{E}(YX)| \le \mathbb{E}(|Y\mathbb{E}(X|\mathcal{M})|) = \int_0^\infty \mathbb{E}(|\mathbb{E}(X|\mathcal{M})|\mathbb{1}_{|Y|>t})dt\,.$$

Clearly we have that $\mathbb{E}(|\mathbb{E}(X|\mathcal{M})|\mathbb{1}_{|Y|>t}) \le \|\mathbb{E}(X|\mathcal{M})\|_1 \wedge \mathbb{E}(|X|\mathbb{1}_{|Y|>t})$. Hence

$$|\mathbb{E}(YX)| \le \int_0^\infty \Big(\int_0^{\|\mathbb{E}(X|\mathcal{M})\|_1} \mathbb{1}_{u<\mathbb{E}(|X|\mathbb{1}_{|Y|>t})}du\Big)dt \le \int_0^{\|\mathbb{E}(X|\mathcal{M})\|_1}\Big(\int_0^\infty \mathbb{1}_{t<H_{X,Y}(u)}dt\Big)du,$$

and the first inequality in (4.1) is proved. In order to prove the second one we use Fréchet's inequality (1957) :

$$\mathbb{E}(|X|\mathbb{1}_{|Y|>t}) \le \int_0^{\mathbb{P}(|Y|>t)} Q_X(u)du. \qquad (4.4)$$

We infer from (4.4) that $H_{X,Y}(u) \le Q_Y\circ G_X(u)$, which yields the second inequality in (4.1).

We now prove (4.2). The first inequality in (4.2) follows directly from (4.1). To prove the second one, note that $Q_{X-\mathbb{E}(X)} \le Q_X + \|X\|_1$ and consequently

$$\int_0^x Q_{X-\mathbb{E}(X)}(u)du \le \int_0^x Q_X(u)du + x\|X\|_1\,. \qquad (4.5)$$

Set $R(x) = \int_0^x Q_X(u)du - x\|X\|_1$. Clearly, R' is non-increasing over $]0,1]$, $R'(\epsilon)\ge 0$ for ϵ small enough and $R'(1)\le 0$. We infer that R is first non-decreasing and next non-increasing, and that for any x in $[0,1]$, $R(x)\ge \min(R(0),R(1))$. Since $\int_0^1 Q_X(u)du = \|X\|_1$, we have that $R(1)=R(0)=0$ and we infer from (4.5) that

$$\int_0^x Q_{X-\mathbb{E}(X)}(u)du \le \int_0^x Q_X(u)du + x\|X\|_1 \le 2\int_0^x Q_X(u)du\,.$$

This implies that $G_{X-\mathbb{E}(X)}(u) \ge G_X(u/2)$ which concludes the proof of (4.2).

To prove (4.3), apply Lemma 1 and set $z = G_X(u)$ in the second integral.

5 Variance of partial sums

Notation 2. For any sequence $(\delta_i)_{i\geq 0}$ of nonnegative numbers define

$$\delta^{-1}(u) = \sum_{i\geq 0} \mathbb{I}_{u<\delta_i}$$

Note that if $(\delta_i)_{i\geq 0}$ is non-increasing, the function $\delta^{-1}(u) = \inf\{k \in \mathbb{N} : \delta_k \leq u\}$ is the generalized inverse of $x \to \delta_{[x]}$, $[\,.\,]$ denoting the integer part. Given $(\delta_i)_{i\geq 0}$ and a random variable X, we introduce the conditions

$$\begin{aligned} \text{for } p > 1, \qquad D(p,\delta,X) \quad &: \quad \int_0^{\|X\|_1} (\delta^{-1}(u))^{p-1}\, Q_X^{p-1} \circ G_X(u)du < \infty. \\ \text{and} \quad D(1,\delta,X) \quad &: \quad \int_0^{\|X\|_1} \ln(1+\delta^{-1}(u))du < \infty. \end{aligned}$$

When $\lambda_i = G_X(\delta_i)$ these conditions are equivalent to

$$\begin{aligned} \text{for } p > 1, \qquad R(p,\lambda,X) \quad &: \quad \int_0^1 (\lambda^{-1}(u))^{p-1} Q_X^p(u)du < \infty. \\ \text{and} \quad R(1,\lambda,X) \quad &: \quad \int_0^1 Q_X(u)\ln(1+\lambda^{-1}(u))du < \infty. \end{aligned}$$

Remark 7. Let $(X_i)_{i\in\mathbb{Z}}$ be a stationary sequence of square integrable random variables and $\mathcal{M}_k = \sigma(X_k)$. Set $S_n = X_1 + \cdots + X_n$. Condition $R(2, 2\alpha, X_0)$ was first introduced by Rio (1993) to control the variance of S_n.

The next lemma gives sufficient conditions for $D(p,\delta,X)$ to hold. The proof will be done in appendix.

Lemma 2 *Let $p > 1$ and $(\delta_i)_{i\geq 0}$ be a non-increasing sequence of nonnegative numbers. Any of the following condition implies $D(p,\delta,X)$*

1. $\mathbb{P}(|X| > x) \leq (c/x)^r$ *for some $r > p$, and* $\sum_{i\geq 0}(i+1)^{p-2}\delta_i^{(r-p)/(r-1)} < \infty$.
2. $\|X\|_r < \infty$ *for some $r > p$, and* $\sum_{i\geq 0} i^{(pr-2r+1)/(r-p)}\delta_i < \infty$.
3. $\mathbb{E}(|X|^p(\ln(1+|X|))^{p-1}) < \infty$ *and $\delta_i = O(a^i)$ for some $a < 1$.*

Moreover $D(1,\delta,X)$ holds if and only if $\sum_{i>0}\delta_i/i < \infty$.

We also need the following comparison lemma, whose proof follows straightforwardly from Lemma 1.

Lemma 3 *Let $(\Omega,\mathcal{A},\mathbb{P})$ be a probability space, $(X_i)_{i\geq 0}$ a sequence of integrable real-valued random variables, $(\mathcal{M}_i)_{i\geq 0}$ a σ-algebra of $\mathcal{A}$, and X a real-valued random variable such that $Q_X \geq \sup_{i\geq 0} Q_{X_i}$. Then*

$$R(p, 2\alpha, X) \Rightarrow D(p, \theta/2, X) \Rightarrow D(p, \gamma/2, X)\,.$$

The first application of Inequality (4.2) is the following control of the variance of partial sums.

Proposition 2 *Let $(X_i)_{i\geq 0}$ be a sequence of square integrable random variables and let $\mathcal{M}_i = \sigma(X_i)$. Setting $S_n = X_1 + \cdots + X_n$, we have*

$$\mathrm{Var}(S_n) \leq \sum_{i=1}^{n} \mathrm{Var}(X_i) + 2 \sum_{1\leq j<i\leq n} \int_0^{\gamma_{i-j}/2} Q_{X_i} \circ G_{X_j}(u)du. \tag{5.1}$$

if X is a random variable such that $Q_X \geq \sup_{k\geq 0} Q_{X_k}$, then

$$\mathrm{Var}(S_n) \leq 4n \int_0^{\|X\|_1} ((\gamma/2)^{-1}(u) \wedge n)\, Q_X \circ G_X(u)du\,. \tag{5.2}$$

In particular, if $(X_i)_{i\geq 0}$ is strictly stationary, the sequence $n^{-1}\mathrm{Var}(S_n)$ converges as soon as $D(2, \gamma/2, X_0)$ holds.

Proof of Proposition 2. Inequality (5.1) follows from (4.2) and the decomposition

$$\mathrm{Var}(S_n) = \sum_{k=1}^{n} \mathrm{Var}(X_i) + 2 \sum_{1\leq j<i\leq n} \mathrm{Cov}(X_i, X_j)\,.$$

We now prove (5.2). Since $Q_{X_i} \leq Q_X$ we have that $G_{X_i} \geq G_X$ for any nonnegative integer i. Consequently $Q_{X_i} \circ G_{X_j} \leq Q_X \circ G_X$ and (5.2) follows from (5.1). Finally, if $(X_i)_{i\geq 0}$ is a strictly stationary sequence, Condition $D(2, \gamma/2, X_0)$ ensures that $\sum_{k>0} |\mathrm{Cov}(X_0, X_k)|$ is finite. Applying Césaro's lemma, we conclude that the sequence $n^{-1}\mathrm{Var}(S_n)$ converges.

6 CLT and weak invariance principle

Definition 2. Let T be the shift operator from $\mathbb{R}^{\mathbb{Z}}$ to $\mathbb{R}^{\mathbb{Z}}$: $(T(x))_i = x_{i+1}$. Let $\mathcal{I}$ the σ-algebra of T-invariants elements of $\mathcal{B}(\mathbb{R}^{\mathbb{Z}})$. We say that a strictly stationary sequence $\mathbf{X} = (X_i)_{i\in\mathbb{Z}}$ of real-valued random variables is ergodic if each element of $\mathbf{X}^{-1}(\mathcal{I})$ has measure 0 or 1.

The following theorem is a particular case of a central limit theorem which was first communicated by Gordin at the Vilnius Conference on Probability and Statistics (1973) (a proof may be found in Esseen and Janson (1985)).

Theorem 1 *Let $\mathbf{X} = (X_i)_{i\in\mathbb{Z}}$ be a strictly stationary and ergodic sequence of integrable an centered variables, $\mathcal{M}_i = \sigma(X_j, j \leq i)$ and $S_n = X_1 + \cdots + X_n$. If*

$$G: \quad \sum_{i\geq 0} \gamma_i < \infty \quad \textit{and} \quad \liminf_{n\to\infty} \frac{1}{\sqrt{n}} \|S_n\|_1 < \infty$$

then $n^{-1/2}S_n$ converges in distribution to a normal distribution.

From Samek and Volný (1998), we know that Condition G is not sufficient to obtain the weak invariance principle. The next theorem is due to Dedecker and Rio (2000). For further comments on Condition DR below, see also Dedecker and Merlevède (2002). Contrary to Theorem 1, we do not require $\mathbf{X}$ to be ergodic, and consequently the limit is a mixture of Wiener processes.

Theorem 2 *Let $\mathbf{X} = (X_i)_{i\in\mathbb{Z}}$ be a strictly stationary sequence of square integrable an centered random variables, and let $\mathcal{M}_i = \sigma(X_j, j \le i)$. For any t in $[0,1]$ set $S_n(t) = X_1 + \cdots + X_{[nt]} + (nt - [nt])X_{[nt]+1}$. If*

$$DR: \qquad X_0\mathbb{E}(S_n|\mathcal{M}_0) \quad \text{converges in } \mathbb{L}^1\,,$$

then $\{n^{-1/2}S_n(t), t\in[0,1]\}$ converges in distribution in $(C[0,1], \|\cdot\|_\infty)$ to ηW, where W is a standard Brownian motion independent of η and η is the nonnegative $\mathbf{X}^{-1}(\mathcal{I})$-measurable variable defined by $\eta = \mathbb{E}(X_0^2|\mathbf{X}^{-1}(\mathcal{I})) + 2\sum_{k=1}^{\infty}\mathbb{E}(X_0X_k|\mathbf{X}^{-1}(\mathcal{I}))$.

Applying Proposition 1, we easily get the following result

Corollary 1 *Let $(X_i)_{i\in\mathbb{Z}}$ and $(\mathcal{M}_i)_{i\ge0}$ be as in Theorem 1. The sequences $(\gamma_i)_{i\ge0}$, $(\theta_i)_{i\ge0}$ and $(\alpha_i)_{i\ge0}$ are non-increasing and we have the implications*

$$R(2,2\alpha,X_0) \Rightarrow D(2,\theta/2,X_0) \Rightarrow D(2,\gamma/2,X_0) \Rightarrow DR.$$

Remark 8. The fact that $R(2,2\alpha,X_0)$ implies DR is proved in Dedecker and Rio (2000). For the usual strong mixing coefficients α'_i, the functional central limit theorem has been established under condition $R(2,2\alpha',X_0)$ by Doukhan, Massart and Rio (1994). Note that the latter condition implies that $\mathbf{X}$ is ergodic, so that the limiting process is necessarily Gaussian. Optimality of Condition $R(2,2\alpha',X_0)$ is studied in Bradley (1997, 2002).

Proof of Corollary 1. The two first implications are given in Lemma 2. In order to prove that $D(2,\gamma/2,X_0) \Rightarrow DR$, note that if $\varepsilon_k = \mathrm{sign}(\mathbb{E}(X_k|\mathcal{M}_0))$, we obtain from (4.2) that

$$\sum_{k\ge0}\|X_0\mathbb{E}(X_k|\mathcal{M}_0)\|_1 = \sum_{k\ge0}\mathrm{Cov}(|X_0|\varepsilon_k, X_k) \le 2\int_0^{\|X\|_1}(\gamma/2)^{-1}(u)Q_{X_0}\circ G_{X_0}(u)du\,.$$

7 Marcinkiewicz strong laws

Theorem 3 *Let $(X_i)_{i\in\mathbb{N}}$ be a sequence of integrable random variables, and define $\mathcal{M}_i = \sigma(X_j, 0\le j\le i)$. Let X be a variable such that $Q_X \ge \sup_{k\ge1}Q_{X_k}$. The sequences $(\theta_i)_{i\ge0}$ and $(\alpha_i)_{i\ge0}$ are non-increasing and we have the implication $R(p,2\alpha,X) \Rightarrow D(p,\theta/2,X)$. Further, if $D(p,\theta/2,X)$ holds for some p in $[1,2[$, then $n^{-1/p}\sum_{k=1}^{n}(X_k - \mathbb{E}(X_k))$ converges almost surely to 0 as n goes to infinity.*

Remark 9. The fact that $R(p, 2\alpha, X)$ implies that $n^{-1/p}\sum_{k=1}^{n}(X_k - \mathbb{E}(X_k))$ converges almost surely to 0 has been proved by Rio (1995, 2000).

Proof of Theorem 3. The first implication is given in Lemma 2. Now, setting $\lambda_i = G_X(\theta_i/2)$, Condition $D(p, \theta/2, X)$ is equivalent to $R(p, \lambda, X)$. The latter condition is the same as in Rio (2000), Corollary (3.1), with λ in place of α. In fact, the proof of Theorem 2 is the same as that of Rio's corollary (cf. Rio (2000) pages 57-60). This comes from the fact that the truncation $\bar{X}_i$ used by Rio is an 1-Lipschitz function of X_i. Consequently the coefficients θ_i of the sequence $(\bar{X}_i)_{i\in\mathbb{N}}$ are smaller or equal to that of $(X_i)_{i\in\mathbb{N}}$. The only tool we need is a maximal inequality similar to Corollary 2.4 in Peligrad (1999) or Theorem 3.2 in Rio (2000).

Proposition 3 *Let $(X_i)_{i\in\mathbb{N}}$ be a sequence of square integrable random variables, and $\mathcal{M}_i = \sigma(X_j, 0 \le j \le i)$. Let X be a random variable such that $Q_X \ge \sup_{k\ge 1} Q_{X_k}$. Let $\lambda_i = G_X(\gamma_i/2)$, $S_n = \sum_{k=1}^{n} X_k - \mathbb{E}(X_k)$ and $S_n^* = \max(0, \ldots, S_n)$. For any positive integer p and any positive real x we have*

$$\mathbb{P}(S_n^* \ge 2x) \le \frac{4}{x^2}\sum_{k=1}^{n}\int_0^1 (\lambda^{-1}(u) \wedge p) Q_X^2(u)du + \frac{4}{x}\sum_{k=1}^{n}\int_0^{\lambda_p} Q_X(u)du\,. \tag{7.1}$$

Proof of Proposition 3. As noted by Rio (2000), page 55, It suffices to consider the case $x = 1$. Indeed, for any positive real x consider the sequences $(X_i/x)_{i\in\mathbb{Z}}$ and $(\gamma_i/x)_{i\ge 0}$, the variable X/x and the functions $Q_{X/x}$ and $G_{X/x}$ given by $Q_{X/x}(u) = Q_X(u)/x$ and $G_{X/x}(u) = G_X(xu)$. The coefficient λ_i of the sequence $(X_i/x)_{i\in\mathbb{Z}}$ is given by $G_{X/x}(\gamma_i/2x) = G_X(\gamma_i/2)$ and is the same as that of $(X_i)_{i\in\mathbb{Z}}$. By homogeneity, it is enough to prove (7.1) for $x = 1$.

The end of the proof follows Rio (2000), pages 55-57, by noting that:

1. Let Y be any $\mathcal{M}_{k-p}$-measurable random variable such that $\|Y\|_\infty \le 1$. By (4.2) and the fact that $Q_Y \circ G_{X_k} \le Q_Y \circ G_X$, we have

$$|\mathrm{Cov}(Y, X_k)| \le 2\int_0^{\gamma_p/2} Q_Y \circ G_X(u)du \le 2\int_0^{\lambda_p} Q_X(u)du\,.$$

2. Let Z be any $\mathcal{M}_i$-measurable random variable such that $|Z| \le |X_i|$. By (4.2) and the fact that $Q_Z \circ G_{X_k} \le Q_Z \circ G_X$, we have

$$|\mathrm{Cov}(Z, X_k)| \le 2\int_0^{\gamma_{k-i}/2} Q_X \circ G_X(u)du = 2\int_0^{\lambda_{k-i}} Q_X^2(u)du\,.$$

8 Burkhölder's inequality

The next result extends Theorem 2.5 of Rio (2000) to non-stationary sequences.

Proposition 4 *Let $(X_i)_{i\in\mathbb{N}}$ be a sequence of centered and square integrable random variables, and $\mathcal{M}_i = \sigma(X_j, 0 \le j \le i)$. Define $S_n = X_1 + \cdots + X_n$ and*

$$b_{i,n} = \max_{i\le l\le n} \Big\| X_i \sum_{k=i}^{l} \mathbb{E}(X_k|\mathcal{M}_i)\Big\|_{p/2} .$$

For any $p \ge 2$, the following inequality holds

$$\|S_n\|_p \le \Big(2p\sum_{i=1}^{n} b_{i,n}\Big)^{1/2} . \tag{8.1}$$

Proof of Proposition 4. We proceed as in Rio (2000) pages 46-47. For any t in $[0,1]$ and $p \ge 2$, let $h_n(t) = \|S_{n-1} + tX_n\|_p^p$. Our induction hypothesis at step $n-1$ is the following: for any $k < n$

$$h_k(t) \le (2p)^{p/2}\Big(\sum_{i=1}^{k-1} b_{i,k} + tb_{k,k}\Big)^{p/2} .$$

Clearly, this assumption is true at step 1. Assuming that it holds for $n-1$, we have to check it at step n. Setting $G(i,n,t) = X_i(t\mathbb{E}(X_n|\mathcal{M}_i) + \sum_{k=i}^{n-1}\mathbb{E}(X_k|\mathcal{M}_i))$ and applying Theorem (2.3) in Rio (2000) with $\psi(x) = |x|^p$, we get

$$\frac{h_n(t)}{p^2} \le \sum_{i=1}^{n-1}\int_0^1 \mathbb{E}\big(|S_{i-1} + sX_i|^{p-2}G(i,n,t)\big)ds + \int_0^t \mathbb{E}(|S_{n-1} + sX_n|^{p-2}X_n^2)ds . \tag{8.2}$$

Note that the function $t \to \mathbb{E}(|G(i,n,t)|^{p/2})$ is convex, so that for any t in $[0,1]$, $\mathbb{E}(|G(i,n,t)|^{p/2}) \le \mathbb{E}(|G(i,n,0)|^{p/2}) \vee \mathbb{E}(|G(i,n,1)|^{p/2}) \le b_{i,n}^{p/2}$. Applying Hölder's inequality, we obtain

$$\mathbb{E}\big(|S_{i-1} + sX_i|^{p-2}G(i,n,t)\big) \le (h_i(s))^{(p-2)/p}\,\|G(i,n,t)\|_{p/2} \le (h_i(s))^{(p-2)/p}\, b_{i,n} .$$

This bound together with (8.2) and the induction hypothesis yields

$$\begin{aligned} h_n(t) &\le p^2\Big(\sum_{i=1}^{n-1} b_{i,n}\int_0^1 (h_i(s))^{(p-2)/p}ds + b_{n,n}\int_0^1 (h_n(s))^{(p-2)/p}ds\Big) \\ &\le p^2\Big(\sum_{i=1}^{n-1}(2p)^{\frac{p}{2}-1}b_{i,n}\int_0^1\Big(\sum_{j=1}^{i} b_{j,n} + sb_{i,n}\Big)^{\frac{p}{2}-1}ds + b_{n,n}\int_0^1 (h_n(s))^{1-\frac{2}{p}}ds\Big) . \end{aligned}$$

Integrating with respect to s we find

$$b_{i,n}\int_0^1\Big(\sum_{j=1}^{i} b_{j,n} + sb_{i,n}\Big)^{\frac{p}{2}-1}ds = \frac{2}{p}\Big(\sum_{j=1}^{i} b_{j,n}\Big)^{\frac{p}{2}} - \frac{2}{p}\Big(\sum_{j=1}^{i-1} b_{j,n}\Big)^{\frac{p}{2}} ,$$

and summing in j we finally obtain

$$h_n(t) \le \Big(2p \sum_{j=1}^{n-1} b_{j,n}\Big)^{\frac{p}{2}} + p^2 b_{n,n} \int_0^1 (h_n(s))^{1-\frac{2}{p}} ds\,. \tag{8.3}$$

Clearly the function $u(t) = (2p)^{p/2}(b_{1,n} + \cdots + t b_{n,n})^{p/2}$ solves the equation associated to Inequality (8.3). A classical argument ensures that $h_n(t) \le u(t)$ which concludes the proof.

Corollary 2 *Let $(X_i)_{i\in\mathbb{N}}$ and $(\mathcal{M}_i)_{i\in\mathbb{N}}$ be as in Proposition 4. Let X be any random variable such that $Q_X \ge \sup_{k\ge 1} Q_{X_k}$. This sequence of coefficients is non-increasing and for $p \ge 2$ we have the inequality*

$$\|S_n\|_p \le \sqrt{2pn}\Big(\int_0^{\|X\|_1} (\gamma^{-1}(u) \wedge n)^{p/2} Q_X^{p-1} \circ G_X(u) du\Big)^{1/p}.$$

Proof of Corollary 2. Let $q = p/(p-2)$. By duality there exists Y such that $\|Y\|_q = 1$ and

$$b_{i,n} \le \sum_{k=i}^{n} \mathbb{E}(|Y X_i \mathbb{E}(X_k|\mathcal{M}_i)|)\,.$$

Let $\lambda_i = G_X(\gamma_i)$. Applying (4.1) and Fréchet's inequality (1957), we obtain

$$b_{i,n} \le \sum_{k=i}^{n} \int_0^{\gamma_{k-i}} Q_{YX_i} \circ G_X(u) du \le \sum_{k=i}^{n} \int_0^{\lambda_{k-i}} Q_Y(u) Q_X^2(u) du\,.$$

Using the duality once more, we get

$$b_{i,n}^{p/2} \le \int_0^1 (\lambda^{-1}(u) \wedge n)^{p/2} Q_X^p(u) du \le \int_0^{\|X\|_1} (\gamma^{-1}(u) \wedge n)^{p/2} Q_X^{p-1} \circ G_X(u) du\,.$$

The result follows.

Corollary 3 *Let $(X_i)_{i\in\mathbb{N}}$, $(\mathcal{M}_i)_{i\in\mathbb{N}}$ be as in Proposition 4. Assume that the sequence $(X_i)_{i\in\mathbb{N}}$ is uniformly bounded by M and that there exist $c > 0$ and a in $]0,1[$ such that $\gamma_i \le Mca^i$. The following inequality holds*

$$\mathbb{P}(|S_n| > x) \le C(a,c) \exp\Big(\frac{-x\sqrt{\ln(1/a)}}{\sqrt{n}eM}\Big)\,,$$

where the constant $C(a,c)$ is defined by

$$C(a,c) = u(c/a) \;\; \textit{with} \;\; u(x) = \exp(2e^{-1}\sqrt{x}) 1\!\!1_{x\le e^2} + x 1\!\!1_{x>e^2}\,.$$

Assume now that θ_i is such that $\theta_i \le 2Mca^i$. For any K-Lipschitz function f and $S_n(f) = \sum_{i=1}^n f(X_i) - \mathbb{E}(f(X_i))$ we have

$$\mathbb{P}(|S_n(f)| > x) \le C(a,c) \exp\Big(\frac{-x\sqrt{\ln(1/a)}}{\sqrt{n}2eKM}\Big).$$

Proof of Corollary 4. Set $\lambda_i = \gamma_i/M$. Applying first Markov's inequality and then Corollary 2, we obtain

$$\mathbb{P}(|S_n| > x) \le \Big(\frac{\|S_n\|_p}{x}\Big)^p \le \Big(\frac{\sqrt{2pn}M}{x}\Big)^p \int_0^1 (\lambda^{-1}(u))^{p/2} du\,. \tag{8.4}$$

By assumption $\lambda_{[x]} \le ca^{x-1}$. Setting $u = ca^{x-1}$ we get

$$\int_0^1 (\lambda^{-1}(u))^{p/2} du \le \frac{c\ln(1/a)}{a} \int_0^\infty x^{p/2} a^x dx \le \frac{c}{a}\Big(\frac{\sqrt{p}}{\sqrt{2\ln(1/a)}}\Big)^p$$

This bound together with (8.4) yields

$$\mathbb{P}(|S_n| > x) \le \min\left(1, \frac{c}{a}\Big(\frac{\sqrt{n}pM}{x\sqrt{\ln(1/a)}}\Big)^p\right).$$

Set $b = \sqrt{n}M(x\sqrt{\ln(1/a)})^{-1}$. The function $p \to ca^{-1}(bp)^p$ has an unique minimum over $[2,\infty[$ at point $\min(2, 1/be)$. It follows that $\mathbb{P}(|S_n| > x) \le h(1/be)$, where h is the function from $\mathbb{R}_+$ to $\mathbb{R}_+$ defined by $h(y) = 1 \wedge (ca^{-1}(2/ey)^2 \mathbb{1}_{y<2} + ca^{-1}e^{-y}\mathbb{1}_{y\ge 2})$. The result follows by noting that $h(y) \le u(c/a)\exp(-y)$. To prove the second point, note that $\|f(X_i) - \mathbb{E}(f(X_i))\|_\infty \le 2KM$ and that, by definition of θ_i, $\sup_{k\ge 0} \|\mathbb{E}(f(X_{i+k})|\mathcal{M}_k) - \mathbb{E}(f(X_i + k)))\|_1 \le K\theta_i$.

9 Appendix

Proof of Lemma 2. We proceed as in Rio (2000). For any function f we have that $f(\delta^{-1}(u)) = \sum_{i=0}^\infty f(i+1)\mathbb{1}_{\delta_{i+1}\le u<\delta_i}$. Assume that $f(0) = 0$. Since we can write $f(i+1) = \sum_{j=0}^i f(j+1) - f(j)$, we infer that

$$f(\delta^{-1}(u)) \;=\; \sum_{j=0}^\infty (f(j+1) - f(j))\mathbb{1}_{u<\delta_j} \tag{9.1}$$

The last assertion of Lemma 2 follows by taking $f(x) = \ln(1+x)$.

Proof of 1. Since $\mathbb{P}(|X| > x) \le (c/x)^r$ we easily get that

$$\int_0^x Q_X(u)du \le \frac{c(r-1)}{r} x^{(r-1)/r} \quad \text{and then} \quad G_X(u) \ge \Big(\frac{ur}{c(r-1)}\Big)^{r/(r-1)}.$$

Set $C_p = 1 \vee (p-1)$ and $K_{p,r} = C_p c(c - cr^{-1})^{(p-1)/(r-1)}$, and apply (9.1) with $f(x) = x^{p-1}$. Since $(i+1)^{p-1} - i^{p-1} \le C_p(i+1)^{p-2}$, we obtain

$$\begin{aligned}
\int_0^{\|X\|_1} (\delta^{-1}(u))^{p-1} Q_X^{p-1} \circ G_X(u)du &\le C_p \sum_{i\ge 0}(i+1)^{p-2} \int_0^{\delta_i} Q^{p-1} \circ G_X(u)du \\
&\le K_{p,r} \sum_{i\ge 0}(i+1)^{p-2} \int_0^{\delta_i} u^{(1-p)/(r-1)} du \\
&\le \frac{K_{p,r}(r-1)}{r-p} \sum_{i\ge 0}(i+2)^{p-2}\delta_i^{(r-p)/(r-1)}.
\end{aligned}$$

Proof of 2. Note first that $\int_0^{\|X\|_1} Q_X^{r-1} \circ G_X(u) du = \int_0^1 Q_X^r(u) du = \mathbb{E}(|X|^r)$. Applying Hölder's inequality, we obtain that

$$\Big(\int_0^{\|X\|_1} (\delta^{-1}(u))^{p-1} Q_X^{p-1} \circ G_X(u) du\Big)^{r-1} \leq \|X\|_r^{rp-r} \Big(\int_0^{\|X\|_1} (\delta^{-1}(u))^{(p-1)(r-1)/(r-p)} du\Big)^{r-p}.$$

Now, apply (9.1) with $f(x) = x^q$ and $q = (p-1)(r-1)/(r-p)$. Noting that $(i+1)^q - i^q \leq (1 \vee q)(i+1)^{q-1}$, we infer that

$$\int_0^{\|X\|_1} (\delta^{-1}(u))^{(p-1)(r-1)/(r-p)} du \quad \leq \quad (1 \vee q) \sum_{i \geq 0} (i+1)^{(pr-2r+1)/(r-p)} \delta_i \,.$$

Proof of 3. Let $\tau_i = \delta_i / \|X\|_1$ and U be a random variable uniformly distributed over $[0, 1]$. We have

$$\begin{aligned} \int_0^{\|X\|_1} (\delta^{-1}(u))^{p-1} Q_X^{p-1} \circ G_X(u) du \quad &= \quad \int_0^1 (\tau^{-1}(u))^{p-1} Q_X^{p-1} \circ G_X(u\|X\|_1) du \\ &= \quad \mathbb{E}((\tau^{-1}(U))^{p-1} Q_X^{p-1} \circ G_X(U\|X\|_1)) \,. \end{aligned}$$

Let ϕ be the function defined on $\mathbb{R}^+$ by $\phi(x) = x(\ln(1+x))^{p-1}$. Denote by ϕ^* its Young's transform. Applying Young's inequality, we have that

$$\mathbb{E}((\tau^{-1}(U))^{p-1} Q_X^{p-1} \circ G_X(U\|X\|_1)) \leq 2\|(\tau^{-1}(U))^{p-1}\|_{\phi^*} \|Q_X^{p-1} \circ G_X(U\|X\|_1)\|_\phi$$

Here, note that $\|Q_X \circ G_X(U\|X\|_1)\|_\phi$ is finite as soon as

$$\int_0^{\|X\|_1} Q_X^{p-1} \circ G_X(u)(\ln(1 + Q_X^{p-1} \circ G_X(u)))^{p-1} du < \infty \,.$$

Setting $z = G_X(u)$, we obtain the condition

$$\int_0^1 Q_X^p(u)(\ln(1 + Q_X^{p-1}(u)))^{p-1} du < \infty \,. \tag{9.2}$$

Since both $\ln(1 + |x|^{p-1}) \leq \ln(2) + (p-1)\ln(1+|x|)$ and $Q_X(U)$ has the same distribution as $|X|$, we infer that (9.2) holds as soon as $\mathbb{E}(|X|^p(\ln(1+|X|))^{p-1})$ is finite. It remains to control $\|(\tau^{-1}(U))^{p-1}\|_{\phi^*}$. Arguing as in Rio (2000) page 17, we see that $\|(\tau^{-1}(U))^{p-1}\|_{\phi^*}$ is finite as soon as there exists $c > 0$ such that

$$\sum_{i \geq 0} \tau_i \, \phi'^{-1}((i+1)^{p-1}/c^{p-1}) < \infty \,. \tag{9.3}$$

Since ϕ'^{-1} has the same behaviour as $x \to \exp(x^{1/(p-1)})$ as x goes to infinity, we can always find $c > 0$ such that (9.3) holds provided that $\delta_i = O(a^i)$ for some $a < 1$.

References

[1] Ango Nzé, P. (1998). Critères d'ergodicité géométrique ou arithmétique de modèles linéaires perturbés à représentation markovienne, *C. R. Acad. Sci. Paris Sér. 1* **326** 371-376.

[2] Ango Nzé, P. Bühlman, P. and Doukhan, P. (2000). Nonparametric Regression under non-mixing weak dependence, *Technical Report, Department of Statistics, ETH Zürich.*

[3] Bickel, P. and Bühlmann, P. (1999). A new mixing notion and functional central limit theorems for a sieve bootstrap in time series. *Bernoulli* **5-3**, 413-446.

[4] Bradley, R. C. (1997). On quantiles and the central limit question for strongly mixing sequences, *J. Theoret. Probab.* **10** 507-555.

[5] Bradley, R. C. (2002). *Introduction to Strong Mixing Conditions, Volume 1.* Technical Report, Department of Mathematics, I. U. Bloomington.

[6] Coulon-Prieur, C. and Doukhan, P. (2000). A triangular CLT for weakly dependent sequences. *Statist. Probab. Lett.* **47** 61-68.

[7] Dedecker, J. and Rio, E. (2000). On the functional central limit theorem for stationary processes, *Ann. Inst. H. Poincaré Probab. Statist.* **36** 1-34.

[8] Dedecker, J. and Merlevède, F. (2002). Necessary and sufficient conditions for the conditional central limit theorem, *to appear in Ann. Probab.*

[9] Diaconis, P. and Freedmann, D. (1999). Iterated random functions, *SIAM Review* **41** 45-76.

[10] Doukhan, P. (1994). *Mixing: properties and examples.* Lecture Notes in Statist. **85**, Springer-Verlag.

[11] Doukhan, P. (2002). Limit theorems for stationary sequences. *To appear in Long Range Dependence, Theory and Applications, Doukhan, P. Oppenheim, G. and Taqqu, M. S. editors.* Birkhäuser.

[12] Doukhan, P. and Louhichi, S. (1999). A new weak dependence condition and application to moment inequalities, *Stochastic Process. Appl.* **84** 313-343.

[13] Doukhan, P. Massart, P. and Rio, E. (1994). The functional central limit theorem for strongly mixing processes, *Ann. Inst. H. Poincaré Probab. Statist.* **30** 63-82.

[14] Duflo, M. (1997). *Random Iterative Models.* Springer, New-York.

[15] Eberlein, E. and Taqqu, M. S. (1986). *Progress in Probability an Statistics.* **11**. *Eberlein, E. and Taqqu, M. S. editors.* Birkhäuser.

[16] Esseen, C. G. and Janson, S. (1985). On moment conditions for sums of independent variables and martingale differences, *Stochastic Process. Appl.* **19** 173-182.

[17] Fréchet, M. (1957). Sur la distance de deux lois de probabilité, *C.R. Acad. Sci. Paris*, **244** 689-692.

[18] Giraitis, L. Kokoszka, P. and Leipus, R. (2000). Stationary ARCH models: dependence structure and Central Limit Theorem, *Econometric Theory* **16** 3-22.

[19] Gordin, M. I. (1969). The central limit theorem for stationary processes, *Dokl. Akad. Nauk SSSR* **188** 739-741.

[20] Gordin, M. I. (1973). Abstracts of Communication, T.1: A-K, International Conference on Probability Theory, Vilnius.

[21] Ibragimov, I. A. (1962). Some limit theorems for stationary processes, *Theor. Probab. Appl.* **7** 349-382.

[22] McLeish, D. L. (1975a). A maximal inequality and dependent strong laws, *Ann. Probab.* **3** 829-839.

[23] McLeish, D. L. (1975b). Invariance Principles for dependent Variables. *Z. Wahrsch. Verw. Gebiete*, **32** 165-178.

[24] Peligrad, M. (1999) Convergence of stopped sums of weakly dependent random variables. *Elect. J. Probab.* **4** 1-13.

[25] Pham, T. D., Tran, L. T. (1985) Some mixing properties of time series models, *Stochastic Process. Appl.* **19** 297-303.

[26] Rio, E. (1993). Covariance inequalities for strongly mixing processes, *Ann. Inst. H. Poincaré Probab. Statist.*, **29** 587-597.

[27] Rio, E. (1995). A maximal inequality and dependent Marcinkiewicz-Zygmund strong laws, *Ann. Probab.* **23** 918-937.

[28] Rio, E. (1996) Sur le théorème de Berry-Esseen pour les suites faiblement dépendantes, *Probab. Theory Relat. Fields* **104** 255-282.

[29] Rio, E. (2000). *Théorie asymptotique des processus aléatoires faiblement dépendants.* Collection Mathématiques & Applications **31**. Springer, Berlin.

[30] Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition, *Proc. Nat. Acad. Sci. U. S. A.* **42** 43-47.

[31] Samek, P. and Volný, D. (2000). On the invariance principle and the law of iterated logarithm for stationary processes, *Mathematical physics and stochastic analysis,* 424-438, Word Sci. Publishing.