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 another 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 Ñé, 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 \(\text{Lip}(f)\), the following upper bound holds

\[
\left|\text{Cov}(g(X_{i_1}, \ldots, X_{i_u}), f(X_{i_u+i}))\right| \leq \theta_i \text{Lip}(f) \tag{1.1}
\]

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 \(\theta_i\) 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 conditions 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 Burkholder-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 $x \rightarrow \mathbb{P}(|X| > x)$.
- $G_X$ the inverse of $x \rightarrow \int_0^x Q_X(u)du$.
- $H_{X,Y}$ the generalized inverse of $x \rightarrow \mathbb{E}(|X|1_{|Y|>x})$.

Definition 1. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, and $\mathcal{M}$ a $\sigma$-algebra of $\mathcal{A}$. If $\text{Lip}(g)$ is the lipschitz coefficient of the function $g$, define the class of functions $\mathcal{L}_1 = \{g : \mathbb{R} \mapsto \mathbb{R}, \|g\|_\infty < \infty, \text{Lip}(g) \leq 1\}$. 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 $\sigma$-algebras of $\mathcal{A}$. The sequence of coefficients $\gamma_i$ is then defined by

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

The coefficients $\theta_i$ and $\alpha_i$ 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 $L^2$-norm by Gordin (1969) and McLeish (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 $\sigma$-fields $\mathcal{M}_k = \sigma(X_i, i \leq k)$. By homogeneity it is clear that $\theta_i$ 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 $\alpha'_i$ is greater than the coefficient $\alpha_i$ defined as in (2.1). To understand the difference between $\alpha_i$ and $\alpha'_i$, note that the convergence of $\alpha'_i$ 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 $\alpha_i$ 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 $\theta_i$ defined in (2.1) is equal to $\theta_i = \theta(\mathcal{M}_0, X_i)$. Now if $\theta_n$ tends to zero as $n$ tends to infinity then so does $\|\mathbb{E}(f(X_0)|\mathcal{M}_i) - \mathbb{E}(f(X_0))\|_1$ 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 $L^1$. Combining this result with Birkoff’s ergodic theorem, we infer that for any $f$ such that $f(X_0)$ belongs to $L^1$

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

Of course, this is no longer true if we only assume that $\gamma_i = \gamma(M_0, X_i)$ 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 $\sigma$-algebra of $\mathcal{A}$. Let $X$ be an integrable and real valued random variable. For any random variable $Y$ such that $Q_Y \geq Q_X$,

$$G_Y(\gamma(M, X)/2) \leq G_Y(\theta(M, X)/2) \leq 2\alpha(M, X) \quad (2.2)$$

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 $\sigma$-algebra of $\mathcal{A}$ and $X$ is a random variable such that $Q_X \geq \sup_{i \geq 0} Q_{X_i}$, then

$$G_X(\gamma_i/2) \leq G_X(\theta_i/2) \leq 2\alpha_i \quad (2.3)$$

**Remark 5.** In particular, for any conjugate exponent $p$ and $q$, we infer from (2.2) that $\theta(M, X) \leq 2\|X\|_p(2\alpha(M, X))^{1/q}$. When $p = \infty$, this is a direct consequence of Ibragimov’s inequality (1962). In fact, the coefficient $\alpha(M, X)$ may be defined by

$$4\alpha(M, X) = \sup\{\|\mathbb{E}(f(X)|M) - \mathbb{E}(f(X))\|_1, \|f\|_{\infty} \leq 1\} \text{ (see for instance Theorem 4.4 in Bradley (2002)).}$$

**Proof of Lemma 1.** It is enough to prove (2.2). Clearly $\gamma(M, X) \leq \theta(M, X)$. 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 $f \in L_1$ 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,

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

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

$$\|\mathbb{E}(f(X)|M) - \mathbb{E}(f(X))\|_1 \leq 2\int_0^{2\alpha(M, X)} Q_{f(X)}(u) du = 2G_{f(X)}^{-1}(2\alpha(M, X)).$$
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 \(\mathbb{R}^n\). 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 \(\delta_i\) by

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

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

- Causal linear process: if \(X_n = \sum_{j \geq 0} a_j \xi_{n-j}\) then \(\theta_i \leq 2\|\xi_0\|_1 \sum_{j \geq i} |a_j|\). In some particular cases we can also obtain an upper bound for the usual strong mixing coefficients \(\alpha_i^\prime\) defined in Remark 3: If \(a_j = O(j^{-\alpha})\), \(E(|\xi_0|^{1+\delta}) < \infty\) and the distribution of \(\xi_0\) is absolutely continuous then we have \(\alpha_i^\prime = O(i^{-(a-2)/(1+\delta)})\) as soon as \(a > 2 + 1/\delta\) (see Pham and Tran (1985)). Hence, summability of the series \(\sum_{i \geq 0} \alpha_i^\prime\) holds as soon as \(a > 3 + 2/\delta\), while summability of \(\sum_{i \geq 0} \theta_i\) 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 \(c = \|\xi_0\|_1 \sum_{j=1}^\infty |a_j| \leq 1\). In this case, we have that \(\theta_i \leq 2c^L + 2\|\xi_0\|_1 (1-c)^{-1} \sum_{j \geq j} |a_j|\) for any \(jL \leq i\). 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 \(\ell \leq L\) and \(j_1, \ldots, j_\ell \leq J\), and to note that \(\theta_i \leq 2\|X_0 - X_{0,i}\|_1\). This gives rise to various dependence rates: if \(a_j = 0\) for large enough \(j \geq J\) 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 \((X_n)_{n \geq 0}\) 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 \(\mu\) 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

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

Consequently, if the Markov chain satisfies \(\mathbb{E}(d(X_i^x, X_i^y)) \leq \delta d(x, y)\) for some decreasing sequence \(\delta_i\), we have that \(\theta_i \leq \delta_i \mathbb{E}(d(X_0, X_0^i))\) where \(X_0^i\) is independent and distributed as \(X_0\). Duflo (1997) studied the case where \(\mathbb{E}(d(X_i^x, X_i^y)) \leq kd(x, y)\) 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.

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 \(\xi_i\) 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 \(\theta_i\) decreases with a polynomial rate: for \(\delta\) 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

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

almost everywhere and \(\text{ARL}(C, \delta, S)\) 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 \(\text{ARL}(C, \delta, S)\), there exists an unique invariant probability \(\mu\) and moreover \(\mu(|x|^{S-\delta}) < \infty\). Furuher, 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 \(\gamma_i, \theta_i\) and \(\alpha_i\) in two simple situations. In the first example, \((X_i)_{i \in \mathbb{Z}}\) is a martingale-difference sequence, \(\gamma_i = 0\) while \(\theta_i\) (and hence \(\alpha_i\)) 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 \(\alpha_i = 1/4\).

**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 \(\theta_i\) 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 \(\theta_i\) tends to zero, then

\[
\mathbb{E}(f(Y \varepsilon_0)) = \int f(Y x) \mathbb{P}^0(dx) \quad \text{almost surely.} \quad (3.2)
\]
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 $\varepsilon_0$ must be symmetric, so that the sequence $(X_i)_{i \in \mathbb{Z}}$ is i.i.d. In any cases, we conclude that $\theta_i$ 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 $\sigma$-algebras $\mathcal{M}_i = \sigma(X_k, k \leq i)$. For such a process, it is well known that $\alpha_i \equiv 1/4$ (see for instance Doukhan (1994)). To compute $\gamma_i$, note that

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

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

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

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 $\theta_i \leq 2^{-i+1} \sup_{f \in \mathcal{L}_1} \| f(U) - \mathbb{E}(f(U)) \|_1$, or equivalently that

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

It is easy to see that the supremum on the right hand side is $1/4$, so that $\theta_i \leq 2^{-i-1}$. Since $\theta_i \geq \gamma_i$, 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 $\sigma$-algebra of $\mathcal{A}$. Let $X$ be an integrable random variable and $Y$ be an $\mathcal{M}$-measurable random variable such that $|X Y|$ is integrable. The following inequalities hold

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

If furthermore $Y$ is integrable, then

$$| \text{Cov}(Y, X) | \leq \int_0^{\gamma(\mathcal{M}, X)/2} Q_Y \circ G_X(u) du \leq 2 \int_0^{\gamma(\mathcal{M}, X)/2} Q_Y \circ G_X(u) du. \quad (4.2)$$
Applying Lemma 1, we also have that
\[
\int_0^{\gamma(M,X)/2} Q_Y \circ G_X(u) du \leq \int_0^{\theta(M,X)/2} Q_Y \circ G_X(u) du \leq \int_0^{2\alpha(M,X)} Q_Y(u)Q_X(u) du. \tag{4.3}
\]

Remark 6. Combining (4.2) and (4.3) we obtain the inequality
\[
|\text{Cov}(Y, X)| \leq 2 \int_0^{2\alpha(M,X)} Q_Y(u)Q_X(u) du,
\]
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
\[
|\mathbb{E}(YX)| \leq \mathbb{E}(|Y|\mathbb{E}(X|M)|) = \int_0^\infty \mathbb{E}(|\mathbb{E}(X|M)|\mathbb{I}_{|Y|>t}) dt.
\]

Clearly we have that \( \mathbb{E}(|\mathbb{E}(X|M)|\mathbb{I}_{|Y|>t}) \leq \|\mathbb{E}(X|M)\|_1 \wedge \mathbb{E}(|X|\mathbb{I}_{|Y|>t}) \). Hence
\[
|\mathbb{E}(YX)| \leq \int_0^\infty \left( \int_0^{\|\mathbb{E}(X|M)\|_1} \mathbb{I}_{u<\mathbb{E}(|X|\mathbb{I}_{|Y|>t})} du \right) dt \leq \int_0^\infty \left( \int_0^\infty \mathbb{I}_{t<\mathbb{E}(X|M)} dt \right) du,
\]

and the first inequality in (4.1) is proved. In order to prove the second one we use Fréchet’s inequality (1957):
\[
\mathbb{E}(|X|\mathbb{I}_{|Y|>t}) \leq \int_0^{\mathbb{P}(|Y|>t)} Q_X(u) du. \tag{4.4}
\]

We infer from (4.4) that \( H_{X,Y}(u) \leq Q_Y \circ G_X(u) \), 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 \( Q_X - \mathbb{E}(X) \leq Q_X + \|X\|_1 \) and consequently
\[
\int_0^x Q_X - \mathbb{E}(X)(u) du \leq \int_0^x Q_X(u) du + x\|X\|_1. \tag{4.5}
\]

Set \( R(x) = \int_0^x Q_X(u) du - x\|X\|_1 \). Clearly, \( R' \) is non-increasing over \([0, 1]\), \( R'(\epsilon) \geq 0 \) for \( \epsilon \) small enough and \( R'(1) \leq 0 \). We infer that \( R \) is first non-decreasing and next non-increasing, and that for any \( x \) in \([0, 1]\), \( R(x) \geq \mathbb{min}(R(0), R(1)) \). 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
\[
\int_0^x Q_X - \mathbb{E}(X)(u) du \leq \int_0^x Q_X(u) du + x\|X\|_1 \leq 2 \int_0^x Q_X(u) du.
\]

This implies that \( G_X - \mathbb{E}(X)(u) \geq G_X(u/2) \) 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 \(x \rightarrow \delta_{[x]}\), \([\cdot]\) denoting the integer part. Given \((\delta_i)_{i \geq 0}\) and a random variable \(X\), we introduce the conditions

for \(p > 1\), \(D(p, \delta, X) : \int_0^{\|X\|_p} (\delta^{-1}(u))^{p-1} \circ G_X(u) \, du < \infty\)
and \(D(1, \delta, X) : \int_0^{\|X\|_1} \ln(1 + \delta^{-1}(u)) \, du < \infty\).

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

for \(p > 1\), \(R(p, \lambda, X) : \int_0^{1} (\lambda^{-1}(u))^{p-1} Q_X(u) \, du < \infty\)
and \(R(1, \lambda, X) : \int_0^{1} Q_X(u) \ln(1 + \lambda^{-1}(u)) \, du < \infty\).

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 \(\sigma\)-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

\[
\text{Var}(S_n) \leq \sum_{i=1}^{n} \text{Var}(X_i) + 2 \sum_{1 \leq j < i \leq n} \int_{0}^{\gamma_{i-j}/2} Q_{X_i} \circ G_{X_j}(u)du. \tag{5.1}
\]

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

\[
\text{Var}(S_n) \leq 4n \int_{0}^{|X|_1} ((\gamma/2)^{-1}(u) \wedge n) Q_X \circ G_X(u)du. \tag{5.2}
\]

In particular, if \((X_i)_{i \geq 0}\) is strictly stationary, the sequence \(n^{-1}\text{Var}(S_n)\) 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

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

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 \(\sum_{k>0} |\text{Cov}(X_0, X_k)|\) is finite. Applying Cesaro’s lemma, we conclude that the sequence \(n^{-1}\text{Var}(S_n)\) 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 \(\sigma\)-algebra of \(T\)-invariants elements of \(B(\mathbb{R}^\mathbb{Z})\). We say that a strictly stationary sequence \(X = (X_i)_{i \in \mathbb{Z}}\) of real-valued random variables is ergodic if each element of \(X^{-1}(\mathcal{I})\) 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 \(X = (X_i)_{i \in \mathbb{Z}}\) 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

\[
G : \sum_{i \geq 0} \gamma_i < \infty \quad \text{and} \quad \lim inf_{n \to \infty} \frac{1}{\sqrt{n}} ||S_n||_1 < \infty
\]

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 $X$ to be ergodic, and consequently the limit is a mixture of Wiener processes.

**Theorem 2** Let $X = (X_i)_{i \in \mathbb{Z}}$ be a strictly stationary sequence of square integrable an centered random variables, and let $M_i = \sigma(X_j; j \leq i)$. For any $t$ in $[0,1]$ set $S_n(t) = X_1 + \cdots + X_{[nt]} + (nt - [nt])X_{[nt]+1}$. If $DR: \quad X_0E(S_n[M_0] \rightarrow \mathbb{L}^1$, \[ DR : \quad X_0E(S_n[M_0] \rightarrow \mathbb{L}^1, \]

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

Applying Proposition 1, we easily get the following result

**Corollary 1** Let $(X_i)_{i \in \mathbb{Z}}$ and $(M_i)_{i \geq 0}$ be as in Theorem 1. The sequences $(\gamma_i)_{i \geq 0}$, $(\theta_i)_{i \geq 0}$ and $(\alpha_i)_{i \geq 0}$ 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 $\alpha_i'$, 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 $X$ 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 $\varepsilon_k = \text{sign}(E(X_k[M_0]))$, we obtain from (4.2) that

\[ \sum_{k \geq 0} ||X_0E(X_k[M_0])||_1 = \sum_{k \geq 0} \text{Cov}(|X_0|\varepsilon_k, X_k) \leq 2 \int_0^{||X||} \int_0^{(\gamma/2)^{-1}} (u) Q_{X_0} \circ G_{X_0}(u) du. \]

7 Marcinkiewicz strong laws

**Theorem 3** Let $(X_i)_{i \in \mathbb{N}}$ be a sequence of integrable random variables, and define $M_i = \sigma(X_j; 0 \leq j \leq i)$. Let $X$ be a variable such that $Q_X \geq \sup_{k \geq 1} Q_{X_k}$. The sequences $(\theta_i)_{i \geq 0}$ and $(\alpha_i)_{i \geq 0}$ 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 $n^{-1/p} \sum_{k=1}^n \left( X_k - E(X_k) \right)$ 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 $\lambda$ in place of $\alpha$. 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 $\theta_i$ of the sequence $(\bar{X}_i)_{i \in N}$ are smaller or equal to that of $(X_i)_{i \in N}$. 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 $(X_i)_{i \in N}$ be a sequence of square integrable random variables, and $\mathcal{M}_i = \sigma(X_j, 0 \leq j \leq i)$. Let $X$ be a random variable such that $Q_X \geq \sup_{k \geq 1} Q_{X_k}$. 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

$$\mathbb{P}(S_n^* \geq 2x) \leq \frac{4}{x^2} \sum_{k=1}^{n} \int_{0}^{1} (\lambda^{-1}(u) \wedge p) Q_X^2(u) du + \frac{4}{x} \sum_{k=1}^{n} \int_{0}^{\lambda_k} Q_X(u) du. \quad (7.1)$$

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 $(\gamma_i/x)_{i \geq 0}$, 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 $\lambda_i$ 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 $\|Y\|_\infty \leq 1$. By (4.2) and the fact that $Q_Y \circ G_{X_k} \leq Q_Y \circ G_X$, we have

$$|\text{Cov}(Y, X_k)| \leq 2 \int_{0}^{\gamma_{p/2}} Q_Y \circ G_X(u) du \leq 2 \int_{0}^{\lambda_p} Q_X(u) du.$$

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

$$|\text{Cov}(Z, X_k)| \leq 2 \int_{0}^{\gamma_{i/2}} Q_X \circ G_X(u) du = 2 \int_{0}^{\lambda_{i/2}} Q_X^2(u) du.$$

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 \(\mathcal{M}_i = \sigma(X_j, 0 \leq j \leq i)\). Define \(S_n = X_1 + \cdots + X_n\) and

\[
b_{i,n} = \max_{i \leq t \leq n} \left\| X_i \sum_{k=i}^{l} \mathbb{E}(X_k | \mathcal{M}_i) \right\|_{p/2}.
\]

For any \(p \geq 2\), the following inequality holds

\[
\|S_n\|_p \leq \left( 2p \sum_{i=1}^{n} b_{i,n} \right)^{1/2}.
\]

(8.1)

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

\[
h_k(t) \leq (2p)^{p/2} \left( \sum_{i=1}^{k-1} b_{i,k} + tb_{k,k} \right)^{p/2}.
\]

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

\[
\frac{h_n(t)}{p^2} \leq \sum_{i=1}^{n-1} \int_0^1 \mathbb{E}(|S_{i-1} + sX_i|^{p-2}G(i,n,t))ds + \int_0^t \mathbb{E}(|S_{n-1} + sX_n|^{p-2}X_n^2)ds.
\]

(8.2)

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

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

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

\[
h_n(t) \leq p^2 \left( \sum_{i=1}^{n-1} b_{i,n} \int_0^1 (h_i(s))^{(p-2)/p}ds + b_{n,n} \int_0^1 (h_n(s))^{(p-2)/p}ds \right)
\]

\[
\leq p^2 \left( \sum_{i=1}^{n-1} (2p)^{\frac{p-2}{p}} b_{i,n} \int_0^1 \left( \sum_{j=1}^{i} b_{j,n} + s b_{i,n} \right)^{\frac{p-2}{p}}ds + b_{n,n} \int_0^1 (h_n(s))^{1-\frac{2}{p}}ds \right).
\]

Integrating with respect to \(s\) we find

\[
b_{i,n} \int_0^1 \left( \sum_{j=1}^{i} b_{j,n} + s b_{i,n} \right)^{\frac{p-2}{p}}ds = \frac{2}{p} \left( \sum_{j=1}^{i} b_{j,n} \right)^{\frac{p}{r}} - \frac{2}{p} \left( \sum_{j=1}^{i-1} b_{j,n} \right)^{\frac{p}{r}},
\]

13
and summing in \( j \) we finally obtain
\[
h_n(t) \leq \left(2p \sum_{j=1}^{n-1} b_{j,n} \right)^{\frac{p}{2}} + p^2 b_{n,n} \int_0^1 (h_n(s))^{1-\frac{p}{2}} ds. \tag{8.3}
\]

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

**Corollary 2** Let \((X_i)_{i \in \mathbb{N}}\) and \((M_i)_{i \in \mathbb{N}}\) be as in Proposition 4. Let \( X \) be any random variable such that \( Q_X \geq \sup_{k \geq 1} Q_{X_k} \). This sequence of coefficients is non-increasing and for \( p \geq 2 \) we have the inequality
\[
\|S_n\|_p \leq \sqrt{2pn} \left( \int_0^{\|X\|_1} (\gamma^{-1}(u) \land n)^{p/2} Q_X^{p-1} \circ G_X(u) du \right)^{1/p}.
\]

**Proof of Corollary 2.** Let \( q = p/(p-2) \). By duality there exists \( Y \) such that \( \|Y\|_q = 1 \) and
\[
b_{i,n} \leq \sum_{k=1}^n \mathbb{E}(\|YX_k|\mathbb{E}(X_k|M_i))\right).
\]

Let \( \lambda_i = G_X(\gamma_i) \). Applying (4.1) and Fréchet’s inequality (1957), we obtain
\[
b_{i,n} \leq \sum_{k=1}^n \int_0^{\gamma_{k-1}} Q_{YX_i} \circ G_X(u) du \leq \sum_{k=1}^n \int_0^{\lambda_{k-1}} Q_{Y} Q_X^{p-1}(u) du.
\]

Using the duality once more, we get
\[
b_{i,n}^{p/2} \leq \int_0^1 (\lambda^{-1}(u) \land n)^{p/2} Q_X(u) du \leq \int_0^{\|X\|_1} (\gamma^{-1}(u) \land n)^{p/2} Q_X^{p-1} \circ G_X(u) du.
\]

The result follows.

**Corollary 3** Let \((X_i)_{i \in \mathbb{N}}, (M_i)_{i \in \mathbb{N}}\) 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 \( \gamma_i \leq Mca^i \). The following inequality holds
\[
\mathbb{P}(\|S_n\| > x) \leq C(a,c) \exp\left(-x\sqrt{\frac{\ln(1/a)}{\sqrt{neM}}}\right),
\]
where the constant \( C(a,c) \) is defined by
\[
C(a,c) = u(c/a) \quad \text{with} \quad u(x) = \exp(2e^{-1}\sqrt{x})\mathbb{1}_{x \leq e^2} + x\mathbb{1}_{x > e^2}.
\]

Assume now that \( \theta_i \) is such that \( \theta_i \leq 2Mca^i \). For any \( K \)-Lipschitz function \( f \) and \( S_n(f) = \sum_{i=1}^n f(X_i) - \mathbb{E}(f(X_i)) \) we have
\[
\mathbb{P}(\|S_n(f)\| > x) \leq C(a,c) \exp\left(-x\sqrt{\frac{\ln(1/a)}{\sqrt{n2eKM}}}\right).
\]

14
Proof of Corollary 4. Set \( \lambda_i = \gamma_i / M \). Applying first Markov’s inequality and then Corollary 2, we obtain
\[
P(|S_n| > x) \leq \left( \frac{\|S_n\|_p}{x} \right)^p \leq \left( \frac{\sqrt{2pM}}{x} \right)^p \int_0^1 (\lambda^{-1}(u))^{p/2} du.
\] (8.4)

By assumption \( \lambda_{[x]} \leq ca^{-1} \). Setting \( u = ca^{-1} \) we get
\[
\int_0^1 (\lambda^{-1}(u))^{p/2} du \leq \frac{c \ln(1/a)}{a} \int_0^\infty x^{p/2} a^x dx \leq \frac{c}{a} \left( \frac{\sqrt{p}}{2 \ln(1/a)} \right)^p
\]
This bound together with (8.4) yields
\[
P(|S_n| > x) \leq \min \left( 1, \frac{c}{a} \left( \frac{\sqrt{npM}}{x \ln(1/a)} \right)^p \right).
\]

Set \( b = \sqrt{nM} (x \sqrt{\ln(1/a)})^{-1} \). The function \( p \to ca^{-1}(bp)^p \) has an unique minimum over \([2, \infty[\) at point \( \min(2, 1/be) \). It follows that \( P(|S_n| > x) \leq h(1/be) \), where \( h \) is the function from \( \mathbb{R}_+ \) to \( \mathbb{R}_+ \) defined by \( h(y) = 1 \wedge (ca^{-1}/ey)^2 \mathbb{I}_{y<2} + ca^{-1}e^{-y} \mathbb{I}_{y \geq 2} \).

The result follows by noting that \( h(y) \leq u(c/a) \exp(-y) \). To prove the second point, note that \( \|f(X_i) - \mathbb{E}(f(X_i))\|_\infty \leq 2KM \) and that, by definition of \( \theta_i \), \( \sup_{k \geq 0} \|\mathbb{E}(f(X_{i+k})|\mathcal{M}_k) - \mathbb{E}(f(X_{i+k}))\|_1 \leq K \theta_i \).

9 Appendix

Proof of Lemma 2. We proceed as in Rio (2000). For any function \( f \) we have that \( f(\delta^{-1}(u)) = \sum_{i=0}^\infty f(i+1) \mathbb{I}_{\delta_{i+1} \leq u < \delta_i} \). Assume that \( f(0) = 0 \). Since we can write \( f(i+1) = \sum_{j=0}^\infty f(j+1) - f(j) \), we infer that
\[
f(\delta^{-1}(u)) = \sum_{j=0}^\infty (f(j+1) - f(j)) \mathbb{I}_{u < \delta_j}
\] (9.1)

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

Proof of 1. Since \( P(|X| > x) \leq (c/x)^r \) we easily get that
\[
\int_0^x Q_X(u) du \leq \frac{c(r-1)}{r} x^{(r-1)/r} \quad \text{and then} \quad G_X(u) \geq \left( \frac{ur}{c(r-1)} \right)^{r/(r-1)}.
\]

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 \( (i+1)^{p-1} - i^{p-1} \leq C_p (i+1)^{p-2} \), we obtain
\[
\int_0^{||X||_1} (\delta^{-1}(u))^{p-1} Q_X^{-1} \circ G_X(u) du \leq C_p \sum_{i \geq 0} (i+1)^{p-2} \int_0^{\delta_i} Q^{p-1} \circ G_X(u) du
\]
\[
\leq K_{p,r} \sum_{i \geq 0} (i+1)^{p-2} \int_0^{\delta_i} u^{(1-p)/(r-1)} du
\]
\[
\leq K_{p,r} \frac{(r-1)}{r-p} \sum_{i \geq 0} (i+2)^{p-2} \delta_i^{(r-p)/(r-1)}
\]
Proof of 2. Note first that \( \int_0^{\|X\|_1} Q_X^{-1} \circ G_X(u) du = \int_0^1 Q_X^{-1} = \mathbb{E}(|X|^r) \). Applying Hölder’s inequality, we obtain that

\[
\left( \int_0^{\|X\|_1} (\delta^{-1}(u))^{p-1} Q_X^{-1} \circ G_X(u) du \right)^{r-1} \leq \|X\|_{r^p} \left( \int_0^{\|X\|_1} (\delta^{-1}(u))^{(p-1)(r-1)/(r-p)} du \right)^{r-p}.
\]

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

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

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

\[
\int_0^{\|X\|_1} (\delta^{-1}(u))^{p-1} Q_X^{-1} \circ G_X(u) du = \int_0^1 (\tau^{-1}(u))^{p-1} Q_X^{-1} \circ G_X(u\|X\|_1) du
= \mathbb{E}((\tau^{-1}(U))^{p-1} Q_X^{-1} \circ G_X(U\|X\|_1))
\]

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

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

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

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

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 \( \phi^{-1} \) 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


