A FOURTH MOMENT INEQUALITY FOR FUNCTIONALS OF STATIONARY PROCESSES

OLIVIER DURIEU,* *Université de Rouen*

Abstract

In this paper, a fourth moment bound for partial sums of functionals of strongly ergodic Markov chains is established. This type of inequality plays an important role in the study of the empirical process invariance principle. This inequality is specially adapted to the technique of Dehling, Durieu, and Volný (2008). The same moment bound can be proved for dynamical systems whose transfer operator has some spectral properties. Examples of applications are given.

Keywords: Stationary process; moment inequality; strongly ergodic Markov chain; dynamical system; empirical distribution; invariance principle

2000 Mathematics Subject Classification: Primary 60G10; 60J10; 60F17; 28D05; 62G20

1. Introduction

Fourth moment bounds for partial sums of stationary processes are a key tool in the study of functional limit theorems. In particular, they play an important role in the investigation of the empirical process invariance principle, which will be the main focus of this paper. Let $(X_i)_{i \geq 0}$ be a stationary process in $\mathbb{R}$, and let $F(t) = \mathrm{P}(X_0 \leq t)$. The empirical process associated with $(X_i)_{i \geq 0}$ is defined by

$$Y_n(t) = \frac{1}{\sqrt{n}} \sum_{i=0}^{n-1} (\mathbf{1}_{(-\infty,t]}(X_i) - F(t)), \qquad t \in \mathbb{R}. \tag{1}$$

We say that $(Y_n)_{n \geq 0}$ satisfies an invariance principle if it converges in distribution to a zero-mean Gaussian process. In general, proofs of invariance principles consist of two parts: a multivariate central limit theorem and tightness. Taking into account the work of Donsker [8] on independent and identically distributed (i.i.d.) sequences, the work of Billingsley [2, Chapter 4] on some weakly dependent processes, and the works which followed, the chaining technique seems to be a suitable way of obtaining the tightness of the process $(Y_n)_{n \geq 0}$. Fourth moment inequalities are a central point in this technique. In many cases (such as the i.i.d. case), if an inequality of the type

$$\mathrm{E}\left(\sum_{i=0}^{n-1} [\mathbf{1}_{(s,t]}(X_i) - (F(t) - F(s))] \right)^4 \leq C(n(t-s) + n^2(t-s)^2) \tag{2}$$

for all $s < t$ is established then the tightness of the process follows. The difficulty is to deal with the sequence of indicator variables $(\mathbf{1}_{(s,t]}(X_i))_{i \geq 0}$. When the process $(X_i)_{i \geq 0}$ has some

mixing properties, such as strong, uniform, or beta mixing, $(\mathbf{1}_{(s,t]}(X_i))_{i \geq 0}$ inherits comparable properties and fourth moment bounds can be established, assuming some regularity conditions on X_0's distribution function (for an overview of this theory, see [6] and the references therein). Collet *et al.* [4] proved a fourth moment inequality for expanding maps of the interval (see also [5]). They used spectral properties of the transfer operator associated on the space of bounded variation functions and the fact that the indicators belong to this space. In other cases, such as some types of Markov chains and dynamical systems (see [10] and [11, Chapter X]), we can have good properties of the Markov operator or the transfer operator on other spaces of functions which do not contain the indicator functions. In Dehling *et al.* [7], a new technique for proving the empirical process invariance principle was developed using approximations of indicators by regular functions. The fourth moment bound of Corollary 1 is well adapted to this situation. The main point to note is that the Banach norm appears only through its logarithm. In Section 2 and Section 3, the fourth moment inequality is stated and proved for strongly ergodic Markov chains. In Section 4 we state the same moment bound for a class of dynamical systems. Section 5 is devoted to examples of applications.

2. Fourth moment inequality for Markov chains

Let $(X_n)_{n \geq 0}$ be a homogeneous Markov chain with a stationary measure ν. Denote by P the associated Markov operator, and denote by E the state space. Consider a Banach space $(\mathcal{B}, \| \cdot \|)$ of ν-measurable functions from E to $\mathbb{R}$, which contains the function $\mathbf{1} = \mathbf{1}_E$. We will assume that the chain is $\mathcal{B}$-geometrically ergodic.

Definition. The Markov chain $(X_n)_{n \geq 0}$ is $\mathcal{B}$-geometrically ergodic or strongly ergodic (with respect to $\mathcal{B}$) if

(a) there exist $\kappa > 0$ and $0 < \theta < 1$ such that, for all $f \in \mathcal{B}$,

$$\| P^n f - \Pi f \| \leq \kappa \theta^n \| f \|, \tag{3}$$

where $\Pi f = \mathrm{E}_\nu(f) \mathbf{1}$;

(b) there exists $p \geq 1$ such that $(\mathcal{B}, \| \cdot \|)$ is continuously included in $(\mathbb{L}^p(\nu), \| \cdot \|_p)$, i.e. there exists a $C > 0$ such that, for all $f \in \mathcal{B}$,

$$\| f \|_p \leq C \| f \|, \tag{4}$$

where $\| f \|_p = (\int |f|^p \, \mathrm{d}\nu)^{1/p}$.

Furthermore, we assume that there exists a constant $N > 0$ such that, for all $f \in \mathcal{B}$ and all $n \in \mathbb{N}$,

$$\| f P^n f \| \leq N \| f \| \| P^n f \|. \tag{5}$$

In particular, this is the case if $(\mathcal{B}, \| \cdot \|)$ is a Banach algebra. In the sequel, with no loss of generality, we assume that $N = 1$.

Strong ergodicity covers a large class of examples (see Section 5). It corresponds to the fact that the Markov transition operator acting on $\mathcal{B}$ has 1 as a simple eigenvalue and the rest of the spectrum is included in a closed ball of radius strictly smaller than 1.

For a function $\varphi \colon E \to \mathbb{R}$, we consider the partial sum

$$S_n(\varphi) = \sum_{i=1}^{n} \varphi(X_i).$$

The aim is to obtain a fourth moment inequality for this partial sum when the function φ belongs to the space $\mathcal{B}$. Our main results are the following.

Theorem 1. *If $(X_n)_{n\geq 0}$ is a $\mathcal{B}$-geometrically ergodic Markov chain with stationary measure ν then, for all $\varphi \in \mathcal{B}$ such that $\mathrm{E}_\nu(\varphi) = 0$ and $\varphi \in \mathbb{L}^4(\nu) \cap \mathbb{L}^{3q}(\nu)$,*

$$
\begin{aligned}
\mathrm{E}_\nu(S_n(\varphi)^4) \leq\ & K[n\|\varphi(X_0)^4\|_1 \log^3(\|\varphi\| + 1) \\
& + n(\|\varphi(X_0)^3\|_q + \|\varphi(X_0)^2\|_q + \|\varphi(X_0)\|_q + \|\varphi(X_0)\|_q^2) \log^2(\|\varphi\| + 1) \\
& + n^2(\|\varphi(X_0)^2\|_1 \log(\|\varphi\| + 1) + \|\varphi(X_0)\|_q)^2],
\end{aligned}
$$

where $1/p + 1/q = 1$ and K is a constant.

As a consequence, the following corollary gives a simpler inequality for the case in which the function φ is bounded.

Corollary 1. *If $(X_n)_{n\geq 0}$ is a $\mathcal{B}$-geometrically ergodic Markov chain with stationary measure ν then, for all $\varphi \in \mathcal{B}$ such that $\mathrm{E}_\nu(\varphi) = 0$, φ is bounded, and $M_\varphi = \max\{1, \sup_x |\varphi(x)|\}$,*

$$
\mathrm{E}_\nu(S_n(\varphi)^4) \leq K M_\varphi^3 [n\|\varphi(X_0)\|_q \log^3(\|\varphi\| + 1) + n^2\|\varphi(X_0)\|_q^2 \log^2(\|\varphi\| + 1)].
$$

Assume that we can prove a multivariate central limit theorem for functions in $\mathcal{B}$. Then, by the technique of Dehling *et al.* [7], if the space $\mathcal{B}$ contains a class of functions approximating the indicators, an empirical process invariance principle follows.

To be complete, we state the following result, which is a corollary of Gordin's theorem [9], from which we can deduce a multivariate central limit theorem.

Proposition 1. *If $(X_n)_{n\geq 0}$ is ergodic and $\mathcal{B}$-geometrically ergodic with $p \geq 2$ in (4), then for all $\varphi \in \mathcal{B}$, $S_n(\varphi - \mathrm{E}_\nu(\varphi))/\sqrt{n}$ converges in distribution to a centred normal law.*

See [11, Theorem IX.2] for sufficient conditions on $\mathcal{B}$ to have the property that strong ergodicity implies ergodicity.

3. Proof of Theorem 1

Let us suppose that the assumptions of Theorem 1 hold. In the sequel, all the expectations are considered with respect to the measure ν and $\mathcal{F}_i$ denotes the σ-algebra generated by X_i.

Let $\varphi \in \mathcal{B}$ with $\mathrm{E}(\varphi) = 0$. As the Markov chain is a stationary process, we have the following bound:

$$
\mathrm{E}(S_n(\varphi)^4) \leq 4!\, n \sum_{i,j,k\geq 0:\ i+j+k\leq n} \mathrm{E}(\varphi(X_0)\varphi(X_i)\varphi(X_{i+j})\varphi(X_{i+j+k})),
$$

which can be decomposed into three sums:

$$
\begin{aligned}
\mathrm{E}(S_n(\varphi)^4) \leq 4!\, n \Bigg(& \sum_{i=1}^{n} \sum_{j,k\leq i} \mathrm{E}(\varphi(X_0)\varphi(X_i)\varphi(X_{i+j})\varphi(X_{i+j+k})) \qquad\qquad (6) \\
& + \sum_{j=1}^{n} \sum_{i,k\leq j} \mathrm{E}(\varphi(X_0)\varphi(X_i)\varphi(X_{i+j})\varphi(X_{i+j+k})) \\
& + \sum_{k=1}^{n} \sum_{i,j\leq k} \mathrm{E}(\varphi(X_0)\varphi(X_i)\varphi(X_{i+j})\varphi(X_{i+j+k})) \Bigg). \qquad (7)
\end{aligned}
$$

To study the terms $\mathrm{E}(\varphi(X_0)\varphi(X_i)\varphi(X_{i+j})\varphi(X_{i+j+k}))$, we will consider three cases according to the greatest integer between i, j, and k. First, however, we note that in all cases, by Hölder's inequality, we have

$$|\mathrm{E}(\varphi(X_0)\varphi(X_i)\varphi(X_{i+j})\varphi(X_{i+j+k}))| \leq \|\varphi(X_0)\|_4^4 \leq \|\varphi(X_0)^4\|_1. \tag{8}$$

Furthermore, let n_0 be a positive integer such that

$$-\frac{\log(\|\varphi\|+1)}{\log\theta} < n_0 \leq -\frac{\log(\|\varphi\|+1)}{\log\theta} + 1.$$

Note that $\theta^{n_0}\|\varphi\| \leq 1$.

Case 1: $i, j \leq k$. Here we use the properties of the Markov operator P on the space $\mathcal{B}$ to obtain another majoration. Successively applying Hölder's inequality, (3), and (4), we obtain

$$\begin{aligned}
|\mathrm{E}(\varphi(X_0)&\varphi(X_i)\varphi(X_{i+j})\varphi(X_{i+j+k}))| \\
&= |\mathrm{E}(\varphi(X_0)\varphi(X_i)\varphi(X_{i+j})(\mathrm{E}(\varphi(X_{i+j+k}) \mid \mathcal{F}_{i+j}) - \mathrm{E}(\varphi(X_{i+j+k}))))| \\
&\leq \|\varphi(X_0)\varphi(X_i)\varphi(X_{i+j})\|_q \|P^k\varphi(X_0) - \Pi\varphi(X_0)\|_p \\
&\leq \|\varphi(X_0)\|_{3q}^3 C\|P^k\varphi - \Pi\varphi\| \\
&\leq \|\varphi(X_0)^3\|_q C\kappa\theta^k\|\varphi\|.
\end{aligned} \tag{9}$$

Now, for sum (7), using (8) for the $n_0 - 1$ first terms and (9) for the others, we obtain

$$\begin{aligned}
\sum_{k=1}^{n} \sum_{i,j\leq k} &\mathrm{E}(\varphi(X_0)\varphi(X_i)\varphi(X_{i+j})\varphi(X_{i+j+k})) \\
&\leq \sum_{k=1}^{n_0-1} k^2\|\varphi(X_0)^4\|_1 + \sum_{k=n_0}^{n} k^2 C\kappa\|\varphi(X_0)^3\|_q \theta^k\|\varphi\| \\
&\leq (n_0-1)^3\|\varphi(X_0)^4\|_1 + C\kappa\|\varphi(X_0)^3\|_q \sum_{k=n_0}^{n} k^2\theta^{k-n_0}.
\end{aligned}$$

There exists a constant C_1 which depends only on θ such that

$$\begin{aligned}
\sum_{k=n_0}^{n} k^2\theta^{k-n_0} &\leq \sum_{k\geq 2}(k+n_0-2)^2\theta^{k-2} \\
&\leq \sum_{k\geq 2} k^2\theta^{k-2} + 2(n_0-2)\sum_{k\geq 2} k\theta^{k-2} + (n_0-2)^2\sum_{k\geq 2}\theta^{k-2}k \\
&\leq C_1(n_0-1)^2,
\end{aligned}$$

because the three series converge. Thus, writing $C_2 = -1/\log\theta$, we obtain $n_0 - 1 \leq C_2\log(\|\varphi\|+1)$ and

$$\begin{aligned}
\sum_{k=1}^{n} \sum_{i,j\leq k} &\mathrm{E}(\varphi(X_0)\varphi(X_i)\varphi(X_{i+j})\varphi(X_{i+j+k})) \\
&\leq C_2^3\|\varphi(X_0)^4\|_1 \log^3(\|\varphi\|+1) + C_3\|\varphi(X_0)^3\|_q \log^2(\|\varphi\|+1),
\end{aligned} \tag{10}$$

where $C_3 = C\kappa C_1 C_2^2$.

Case 2: $i, k \leq j$. We can decompose the expectation as follows:

$$| \mathrm{E}(\varphi(X_0)\varphi(X_i)\varphi(X_{i+j})\varphi(X_{i+j+k}))|$$
$$\leq | \mathrm{E}(\varphi(X_0)\varphi(X_i)(\mathrm{E}(\varphi(X_{i+j}) \, \mathrm{E}(\varphi(X_{i+j+k}) \mid \mathcal{F}_{i+j}) \mid \mathcal{F}_i) - \mathrm{E}(\varphi(X_0)\varphi(X_k))))|$$
$$+ | \mathrm{E}(\varphi(X_0)\varphi(X_i)) \, \mathrm{E}(\varphi(X_0)\varphi(X_k))|. \tag{11}$$

On the right-hand side, let us call $I_{i,j,k}$ the first term and $II_{i,k}$ the second term.

Since $\mathrm{E}(\varphi P^k \varphi(X_0)) = \mathrm{E}(\varphi(X_0) \, \mathrm{E}(\varphi(X_k) \mid \mathcal{F}_0)) = \mathrm{E}(\varphi(X_0)\varphi(X_k))$, we have

$$I_{i,j,k} \leq \|\varphi(X_0)\varphi(X_i)\|_q \|P^j(\varphi P^k \varphi)(X_0) - \Pi(\varphi P^k \varphi)(X_0)\|_p$$
$$\leq \|\varphi(X_0)\|_{2q}^2 C \|P^j(\varphi P^k \varphi) - \Pi(\varphi P^k \varphi)\|$$
$$\leq C \|\varphi(X_0)^2\|_q \kappa \theta^j \|\varphi P^k \varphi\|$$

and, by assumption (5),
$$\|\varphi P^k \varphi\| \leq \|\varphi\| \|P^k \varphi\| \leq \kappa \theta^k \|\varphi\|^2.$$

Therefore,
$$I_{i,j,k} \leq C\kappa^2 \|\varphi(X_0)^2\|_q \theta^{j+k} \|\varphi\|^2 \leq C\kappa^2 \|\varphi(X_0)^2\|_q \theta^j \|\varphi\|^2. \tag{12}$$

Now, thanks to decomposition (11) (also using inequality (8)), for large enough n,

$$\sum_{j=1}^n \sum_{i,k \leq j} \mathrm{E}(\varphi(X_0)\varphi(X_i)\varphi(X_{i+j})\varphi(X_{i+j+k}))$$
$$\leq \sum_{j=1}^{2n_0-2} j^2 \|\varphi(X_0)^4\|_1 + \sum_{j=2n_0-1}^n \sum_{i,k \leq j} (I_{i,j,k} + II_{i,k})$$
$$\leq 8(n_0-1)^3 \|\varphi(X_0)^4\|_1 + \sum_{j=2n_0-1}^n j^2 I_{i,j,k} + \sum_{j=1}^n \sum_{i,k \leq j} II_{i,k},$$

where n_0 has been defined previously.

Inequality (12) and $\theta^{n_0} \|\varphi\| \leq 1$ imply that

$$\sum_{j=2n_0-1}^n j^2 I_{i,j,k} \leq \sum_{j=2n_0-1}^n j^2 C\kappa^2 \|\varphi(X_0)^2\|_q \theta^j \|\varphi\|^2$$
$$\leq C\kappa^2 \|\varphi(X_0)^2\|_q \sum_{j=2n_0-1}^n j^2 \theta^{j-2n_0}.$$

As before, there exists a constant C_4 depending on θ such that

$$\sum_{j=2n_0-1}^n j^2 \theta^{j-2n_0} \leq C_4(n_0-1)^2.$$

So,
$$\sum_{j=2n_0-1}^n j^2 I_{i,j,k} \leq C\kappa^2 \|\varphi(X_0)^2\|_q C_4(n_0-1)^2.$$

For the third term, we have

$$\sum_{j=1}^{n} \sum_{i,k \leq j} II_{i,k} \leq n \left(\sum_{i=1}^{n} |\operatorname{E}(\varphi(X_0)\varphi(X_i))| \right) \left(\sum_{k=1}^{n} |\operatorname{E}(\varphi(X_0)\varphi(X_k))| \right).$$

We can see that

$$\begin{aligned}
|\operatorname{E}(\varphi(X_0)\varphi(X_i))| &\leq \|\varphi(X_0)\|_q \|P^i\varphi(X_0) - \Pi\varphi(X_0)\|_p \\
&\leq C\|\varphi(X_0)\|_q \|P^i\varphi - \Pi\varphi\| \\
&\leq C\kappa \|\varphi(X_0)\|_q \theta^i \|\varphi\|
\end{aligned} \tag{13}$$

and, in the same way,

$$|\operatorname{E}(\varphi(X_0)\varphi(X_k))| \leq C\kappa \|\varphi(X_0)\|_q \theta^k \|\varphi\|. \tag{14}$$

Alternatively, by Hölder's inequality,

$$|\operatorname{E}(\varphi(X_0)\varphi(X_i))| \leq \|\varphi(X_0)^2\|_1 \quad \text{and} \quad |\operatorname{E}(\varphi(X_0)\varphi(X_k))| \leq \|\varphi(X_0)^2\|_1. \tag{15}$$

Thus, by (13), (14), and (15),

$$\begin{aligned}
\sum_{i=1}^{n} |\operatorname{E}(\varphi(X_0)\varphi(X_i))| &\leq \sum_{i=1}^{n_0-1} \|\varphi(X_0)^2\|_1 + \sum_{i=n_0}^{n} C\kappa \|\varphi(X_0)\|_q \theta^i \|\varphi\| \\
&\leq (n_0 - 1)\|\varphi(X_0)^2\|_1 + C\kappa \|\varphi(X_0)\|_q \sum_{i=n_0}^{n} \theta^{i-n_0} \\
&\leq (n_0 - 1)\|\varphi(X_0)^2\|_1 + C_5 \|\varphi(X_0)\|_q,
\end{aligned}$$

where $C_5 = C\kappa \sum_{i \geq 0} \theta^i < \infty$ and

$$\sum_{k=1}^{n} |\operatorname{E}(\varphi(X_0)\varphi(X_k))| \leq (n_0 - 1)\|\varphi(X_0)^2\|_1 + C_5 \|\varphi(X_0)\|_q.$$

Finally,

$$\begin{aligned}
\sum_{j=1}^{n} \sum_{i,k \leq j} &\operatorname{E}(\varphi(X_0)\varphi(X_i)\varphi(X_{i+j})\varphi(X_{i+j+k})) \\
&\leq 8C_2^3 \|\varphi(X_0)^4\|_1 \log^3(\|\varphi\| + 1) + C_6 \|\varphi(X_0)^2\|_q \log^2(\|\varphi\| + 1) \\
&\quad + n(C_2 \|\varphi(X_0)^2\|_1 \log(\|\varphi\| + 1) + C_5 \|\varphi(X_0)\|_q)^2,
\end{aligned} \tag{16}$$

where $C_6 = C\kappa^2 C_4 C_2^2$.

Case 3: $j, k \leq i$. Three uses of the operator properties give

$$\begin{aligned}
|\operatorname{E}(\varphi(X_0)&\varphi(X_i)\varphi(X_{i+j})\varphi(X_{i+j+k}))| \\
&= |\operatorname{E}(\varphi(X_0) \operatorname{E}(\varphi(X_i) \operatorname{E}(\varphi(X_{i+j}) \operatorname{E}(\varphi(X_{i+j+k}) \mid \mathcal{F}_{i+j}) \mid \mathcal{F}_i) \mid \mathcal{F}_0))| \\
&= |\operatorname{E}(\varphi(X_0)[\operatorname{E}(\varphi(X_i) \operatorname{E}(\varphi(X_{i+j}) \operatorname{E}(\varphi(X_{i+j+k}) \mid \mathcal{F}_{i+j}) \mid \mathcal{F}_i) \mid \mathcal{F}_0) \\
&\qquad - \operatorname{E}(\varphi(X_i)\varphi(X_{i+j})\varphi(X_{i+j+k}))])| \\
&\leq \|\varphi(X_0)\|_q \|P^i(\varphi P^j(\varphi P^k \varphi))(X_0) - \Pi(\varphi P^j(\varphi P^k \varphi))(X_0)\|_p \\
&\leq \|\varphi(X_0)\|_q C\kappa \theta^i \|\varphi P^j(\varphi P^k \varphi)\|
\end{aligned} \tag{17}$$

and

$$\|\varphi P^j(\varphi P^k\varphi)\| \leq \|\varphi\|\|P^j(\varphi P^k\varphi)\|$$
$$\leq \|\varphi\|(\|P^j(\varphi P^k\varphi) - \Pi\varphi P^k\varphi\| + |\,\mathrm{E}(\varphi(X_0)\varphi(X_k))|)$$
$$\leq \|\varphi\|(\kappa\theta^j\|\varphi P^k\varphi\| + |\,\mathrm{E}(\varphi(X_0)\varphi(X_k))|)$$
$$\leq \kappa^2\theta^{j+k}\|\varphi\|^3 + C\kappa\|\varphi(X_0)\|_q\theta^k\|\varphi\|^2, \tag{18}$$

where we used inequality (14) in the last line.

From (17) and (18), we derive

$$|\,\mathrm{E}(\varphi(X_0)\varphi(X_i)\varphi(X_{i+j})\varphi(X_{i+j+k}))|$$
$$\leq C\kappa^3\|\varphi(X_0)\|_q\theta^{i+j+k}\|\varphi\|^3 + C^2\kappa^2\|\varphi(X_0)\|_q^2\theta^{i+k}\|\varphi\|^2$$
$$\leq C_7\theta^i\|\varphi(X_0)\|_q\|\varphi\|^2(\|\varphi\| + \|\varphi(X_0)\|_q),$$

where $C_7 = \max\{C\kappa^3, C^2\kappa^2\}$.

With this last inequality and (8), the sum in (6) can be bounded in the same way as before. We use the integer n_0 to obtain

$$\sum_{i=1}^{n}\sum_{j,k\leq i}\mathrm{E}(\varphi(X_0)\varphi(X_i)\varphi(X_{i+j})\varphi(X_{i+j+k}))$$
$$\leq \sum_{i=1}^{3n_0-3}i^2\|\varphi(X_0)^4\|_1 + C_7\sum_{i=3n_0-2}^{n}i^2\theta^i\|\varphi(X_0)\|_q\|\varphi\|^2(\|\varphi\| + \|\varphi(X_0)\|_q)$$
$$\leq 27(n_0-1)^3\|\varphi(X_0)^4\|_1 + C_7(\|\varphi(X_0)\|_q + \|\varphi(X_0)\|_q^2)\sum_{i=3n_0-2}^{n}i^2\theta^{i-3n_0}.$$

The sum is bounded by the corresponding series, which is finite (majoration by $C_8(n_0-1)^2$, where C_8 depends only on θ). So, we can conclude the study of case 3 with

$$\sum_{i=1}^{n}\sum_{j,k\leq i}\mathrm{E}(\varphi(X_0)\varphi(X_i)\varphi(X_{i+j})\varphi(X_{i+j+k}))$$
$$\leq 27C_2^3\|\varphi(X_0)^4\|_1\log^3(\|\varphi\| + 1) + C_7C_8C_2^2(\|\varphi(X_0)\|_q + \|\varphi(X_0)\|_q^2)\log^2(\|\varphi\| + 1). \tag{19}$$

To conclude, let K be the maximum of all the constants appearing in (10), (16), and (19),

$$\mathrm{E}(S_n(\varphi)^4)$$
$$\leq 4!\,K[n\|\varphi(X_0)^4\|_1\log^3(\|\varphi\| + 1)$$
$$+ n(\|\varphi(X_0)^3\|_q + \|\varphi(X_0)^2\|_q + \|\varphi(X_0)\|_q + \|\varphi(X_0)\|_q^2)\log^2(\|\varphi\| + 1)$$
$$+ n^2(\|\varphi(X_0)^2\|_1\log(\|\varphi\| + 1) + \|\varphi(X_0)\|_q)^2].$$

4. Fourth moment inequality for dynamical systems

In Section 2 we dealt with homogeneous Markov chains through their operators. As usual, the techniques can be applied to dynamical systems, using the transfer operator. Here we state

the result for dynamical systems, but the proof (which is essentially the same as in Section 3) is left to the reader.

Let $(\Omega, \mathcal{A}, \mu)$ be a probability space, and let T be a measurable measure preserving transformation (i.e. for all $A \in \mathcal{A}$, $\mu(T^{-1}A) = \mu(A)$). Let us consider the Perron–Frobenius operator (or the transfer operator) of T, $P \colon \mathbb{L}^1(\mu) \to \mathbb{L}^1(\mu)$ defined by

$$\int_\Omega Pf(x)g(x)\,\mathrm{d}\mu(x) = \int_\Omega f(x)g \circ T(x)\,\mathrm{d}\mu(x)$$

for all $f \in \mathbb{L}^1(\mu)$ and $g \in \mathbb{L}^\infty(\mu)$.

As in the Markov case, we assume that there exists a Banach space $(\mathcal{B}, \|\cdot\|)$ of μ-measurable functions from Ω to $\mathbb{R}$ which contains $\mathbf{1}$ and satisfies (5), and that P verifies the same assumptions, i.e.

(i) there exist $\kappa > 0$ and $0 < \theta < 1$ such that, for all $f \in \mathcal{B}$,

$$\|P^n f - \Pi f\| \le \kappa \theta^n \|f\|,$$

where $\Pi f = \mathrm{E}_\mu(f)\,\mathbf{1}$;

(ii) there exists $p \ge 1$ such that $(\mathcal{B}, \|\cdot\|)$ is continuously included in $(\mathbb{L}^p(\mu), \|\cdot\|_p)$.

Again, assertion (i) follows from some quasicompactness of the Perron–Frobenius operator; see [1, Chapter 1] and [11, Chapter II].

Under these assumptions, we have the same fourth moment bound, stated as follows.

Theorem 2. *For all* $f \in \mathcal{B}$ *such that* $\mathrm{E}_\mu(f) = 0$, f *is bounded, and* $M_f = \max\{1, \sup_x |f(x)|\}$,

$$\mathrm{E}_\mu\left(\left(\sum_{i=1}^n f \circ T^i\right)^4\right) \le K M_f^3 [n\|f\|_q \log^3(\|f\| + 1) + n^2 \|f\|_q^2 \log^2(\|f\| + 1)],$$

where $1/p + 1/q = 1$.

5. Applications

In this section we give some examples where the fourth moment inequality applies and then leads to some empirical process invariance principles.

5.1. Uniform ergodicity

Let $(X_n)_{n \ge 0}$ be a Markov chain on the state space E. Denote by $(\mathcal{B}^\infty, \|\cdot\|_\infty)$ the space of bounded measurable functions from E to $\mathbb{R}$ provided with the uniform norm. We say that the Markov chain $(X_n)_{n \ge 0}$ is *uniformly ergodic* if it is $\mathcal{B}^\infty$-geometrically ergodic. This condition is equivalent to the fact that the process satisfies the so-called Doeblin's condition (see [12, Chapter 16]).

In this situation, if X_0 has a distribution function which is regular enough, our fourth moment bound (Corollary 1) implies inequality (2). Then tightness follows and the empirical process invariance principle will follow from the multivariate central limit theorem. Note that this result has already been proved in [2].

Many examples of uniformly ergodic Markov chains are given in [12, Chapter 16], such as T-chains on compact spaces. Another example is given by the Knudsen gas model (see, e.g. [14]).

5.2. Expanding maps

For a large class of expanding maps, empirical process invariance principles have already been established in [4] and [5]. Expanding maps are an example of dynamical systems on an interval. Such transformations are studied in [3], as continuous fraction expansions, β-transformations, and Gauss maps. In many cases, we can show that the Perron–Frobenius operator admits a spectral gap on the space of bounded variation (BV) functions. Since the indicator functions belong to BV functions, Theorem 2 with a condition on the initial distribution proves inequality (2).

Gouëzel [10] gave an example of an expanding map of the interval which has a spectral gap on the space of Lipschitz functions but not on the space of BV functions. For Gouëzel's example, Theorem 2 holds on the space of Lipschitz functions.

5.3. Subshifts

Let E be a finite set, and let $\mathcal{E} = E^{\mathbb{N}}$. The metric d defined on $\mathcal{E}$ is

$$d(x, y) = 2^{-\inf\{k \geq 0:\, x_k \neq y_k\}}.$$

Let $A = (a(i, j))_{i, j \in E}$ be a matrix with coefficients in $\{0, 1\}$, and let

$$\Omega = \{x \in \mathcal{E} : a(x_k, x_{k+1}) = 1 \text{ for all } k \geq 0\}.$$

Write T for the shift operator on Ω, i.e. $(Tx)_k = x_{k+1}$ for all $k \geq 0$. Denote by $\mathcal{B}$ the space of complex-valued functions on Ω, which are Lipschitz continuous with respect to the metric d. The norm on $\mathcal{B}$ is $\| \cdot \| = \| \cdot \|_\infty + m(\cdot)$, where

$$m(f) = \sup\left\{\frac{|f(x) - f(y)|}{d(x, y)}, \ x \neq y\right\}. \tag{20}$$

Note that, since (Ω, d) is compact, $\mathcal{B} \subset \mathbb{L}^\infty$. The Ruelle–Perron–Frobenius theorem shows that the transfer operator P has some quasicompact properties on $\mathcal{B}$. See [11, Theorem XII.6] or [13]. If 1 is the only eigenvalue of modulus 1 and if it is simple, then conditions (i) and (ii) hold and Theorem 2 is satisfied. If f is a Lipschitz continuous function on Ω then, by [7], an empirical process invariance principle is satisfied for the process $(f \circ T^i)_{i \geq 0}$.

5.4. Linear processes

Let $(A, \| \cdot \|_A)$ be a separable Banach space. Let $(a_i)_{i \geq 0}$ be a sequence of linear form on A such that $\sum_{i \geq 0} |a_i| < \infty$, where $|a| = \sup\{|a(x)|, \ \|x\|_A \leq 1\}$. Let $(\xi_i)_{i \in \mathbb{Z}}$ be an i.i.d. sequence of bounded B-valued random variables, where B is a compact subset of A. We define the $\mathbb{R}$-valued linear process

$$X_k = \sum_{i \geq 0} a_i(\xi_{k-i}).$$

If A is a finite set, linear processes can be viewed as functionals of subshifts. Here, in the general case, we use a slightly different metric. Assume that there exist $\rho < 1$ and $C > 0$ such that $|a_i| \leq C\rho^i$ for all $i \geq 0$. We defined on $B^{\mathbb{N}}$ the metric

$$d(x, y) = \sum_{i \geq 0} \rho^i \|x_i - y_i\|_A,$$

where $x = (x_i)_{i \geq 0}$ and $y = (y_i)_{i \geq 0}$. Denote by $Y_k = (\dots, \xi_{-1}, \xi_0)$. Then $(Y_k)_{k \geq 0}$ is a Markov chain on $B^{\mathbb{N}}$, and we can show that $(Y_k)_{k \geq 0}$ satisfies (3) on the space $\mathcal{B}$ of Lipschitz continuous

functions from $B^{\mathbb{N}}$ to $\mathbb{R}$. Indeed, if $f \in \mathcal{B}$,

$$
\begin{aligned}
|\operatorname{E}(f(Y_k) \mid Y_0 = y) &- \operatorname{E}(f(Y_k) \mid Y_0 = x)| \\
&= |\operatorname{E}(f(\ldots, y_0, \xi_1, \ldots, \xi_k) - f(\ldots, x_0, \xi_1, \ldots, \xi_k))| \\
&\leq C\rho^k \|y - x\|_A,
\end{aligned}
$$

and then (3) holds. Hence, $(Y_k)_{k \in \mathbb{Z}}$ is strongly ergodic with respect to the space of Lipschitz functions on $B^{\mathbb{N}}$ and Theorem 1 holds. It is clear that $f(x) = \sum_{i \geq 0} a_i(x_i)$ is a Lipschitz function (on $B^{\mathbb{N}}$), and so a fourth moment bound holds for $(X_k)_{k \geq 0}$ on the space of Lipschitz functions (on $\mathbb{R}$).

5.5. Random iterative Lipschitz models

Let G be a semigroup of Lipschitz transformations of a metric space (E, d), and let $\mathcal{G}$ be a σ-algebra on G. We assume that the action of G on E is measurable.

Let $(g_n)_{n \geq 1}$ be an i.i.d. sequence of G-valued random variables with distribution η. Let X_0 be an E-valued random variable independent of $(g_n)_{n \geq 1}$. We consider the Markov chain $(X_n)_{n \geq 0}$ defined by

$$
X_n = g_n(X_{n-1})
$$

with transition operator P defined by

$$
Pf(x) = \int_G f(g(x)) \, d\eta(g).
$$

We say that η is contracting if

$$
\limsup_n \left\{ \int_G \frac{d(g(x), g(y))}{d(x, y)} \, d\eta^{*n}(g) : x, y \in E, \ x \neq y \right\}^{1/n} < 1,
$$

where η^{*n} denotes the distribution of $g_n \circ \cdots \circ g_1$.

Assume that (E, d) is compact. Let $\mathcal{B}_0$ be the space of $\mathbb{C}$-valued Lipschitz continuous functions on E provided with the norm $\| \cdot \|_0 = \| \cdot \|_\infty + m(\cdot)$, where m is defined as in (20). It is shown in [11, Theorem X.3] that if η is contracting then there exists a unique P-invariant measure on E and $(X_n)_{n \geq 0}$ is $\mathcal{B}_0$-geometrically ergodic with respect to this measure. Then a fourth moment bound holds and, thanks to Dehling *et al.* [7], an empirical invariance principle follows. One example of application is given by products of invertible random matrices (see [11, Section X.4]).

In the case where (E, d) is not compact but every closed ball in E is compact, we can have a similar result but with another Banach space (see [11, Theorem X.4]). Here, the Banach space is the space $\mathcal{B}_1$ of locally Lipschitz functions with weight. These are the $\mathbb{C}$-valued functions f such that

$$
m_1(f) = \sup \left\{ \frac{|f(x) - f(y)|}{d(x, y) p(x) p(y)} : x \neq y \right\} < \infty,
$$

where $p(x) = 1 + d(x, x_0)$ for a fixed $x_0 \in E$. The norm is

$$
\|f\|_1 = \sup \left\{ \frac{|f(x)|}{p(x)^2} : x \in E \right\} + m_1(f).
$$

As an example, we mention a large class of autoregressive models.

5.6. Autoregressive models

The process $(X_n)_{n \geq 0} \subset \mathbb{R}^d$ is called autoregressive with initial value $X_0 \in \mathbb{R}^d$ if it satisfies, for all $n \geq 1$,

$$X_n = A X_{n-1} + Y_n,$$

where $A \in \mathcal{M}(\mathbb{R}^d)$ and $(Y_n)_{n \geq 1} \subset \mathbb{R}^d$ is an i.i.d. sequence of random variables, independent of X_0. See [11, Theorem X.16] for conditions under which $(X_n)_{n \in \mathbb{N}}$ is $\mathcal{B}_1$-geometrically ergodic.

References

[1] BALADI, V. (2000). *Positive Transfer Operators and Decay of Correlations* (Adv. Ser. Nonlinear Dynamics **16**). World Scientific, River Edge, NJ.

[2] BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. John Wiley, New York.

[3] BROISE, A. (1996). Transformations dilatantes de l'intervalle et théorèmes limites. Études spectrales d'opérateurs de transfert et applications. *Astérisque* **238**, 1–109.

[4] COLLET, P., MARTINEZ, S. AND SCHMITT, B. (2004). Asymptotic distribution of tests for expanding maps of the interval. *Ergodic Theory Dynam. Systems* **24**, 707–722.

[5] DEDECKER, J. AND PRIEUR, C. (2007). An empirical central limit theorem for dependent sequences. *Stoch. Process. Appl.* **117**, 121–142.

[6] DEHLING, H. AND PHILIPP, W. (2002). Empirical process techniques for dependent data. In *Empirical Process Techniques for Dependent Data*, Birkhäuser, Boston, MA, pp. 3–113.

[7] DEHLING, H., DURIEU, O. AND VOLNÝ, D. (2008). New techniques for empirical processes of dependent data. Preprint. Available at http://arxiv.org/abs/0806.2941.

[8] DONSKER, M. D. (1952). Justification and extension of Doob's heuristic approach to the Komogorov–Smirnov theorems. *Ann. Math. Statist.* **23**, 277–281.

[9] GORDIN, M. I. (1969). The central limit theorem for stationary processes. *Dokl. Akad. Nauk SSSR* **188**, 739–741 (in Russian). English translation: *Soviet Math. Dokl.* **10**, 1174–1176.

[10] GOUËZEL, S. (2008). An interval map with a spectral gap on Lipschitz functions, but not on bounded variation functions. Preprint. Available at http://arxiv.org/abs/0809.0658v1.

[11] HENNION, H. AND HERVÉ, L. (2001). *Limit Theorems for Markov Chains and Stochastic Properties of Dynamical Systems by Quasi-compactness* (Lecture Notes Math. **1766**). Springer, Berlin.

[12] MEYN, S. P. AND TWEEDIE, R. L. (1993). *Markov Chains and Stochastic Stability*. Springer, London.

[13] PARRY, W. AND POLLICOTT, M. (1990). Zeta functions and the periodic orbit structure of hyperbolic dynamics. *Astérisque* **187–188**, 268.

[14] PÈNE, F. (2005). Rate of convergence in the multidimensional central limit theorem for stationary processes. Application to the Knudsen gas and to the Sinai billiard. *Ann. Appl. Prob.* **15**, 2331–2392.