BULLETIN OF THE BRAZILIAN MATHEMATICAL SOCIETY

Bull Braz Math Soc, New Series 37(4), 461-475

Variance and exponential estimates via coupling

Pierre Collet

Abstract. We give a proof of Devroye and exponential inequalities based on a coupling. We mostly deal with the case of continuous random variables with dynamical systems in mind (and their rather special mixing properties) although the approach is more general.

Keywords: concentration inequalities, moments ineqqualities, large deviations.

Mathematical subject classification: 60F10, 37D35, 62G07.

1 Introduction

The central limit theorem and large deviations are very powerful methods to analyse the fluctuations of large sums of random variables. However there are interesting quantities which are not of this form and for which one would like to have an estimation of the fluctuations. Here are two important examples among many others. Consider a real valued discrete time stationary stochastic process (X_n). The power spectrum is defined as the Fourier transform of the correlation function. An estimator of the integral of the power spectrum is the integral of the periodogram given by (in the case where $\mathbf{E}(X_j) = 0$)

$$W_n(u) = \int_0^u \frac{1}{n} \left| \sum_{j=1}^{n} e^{-ijs} X_j \right|^2 ds \ . \tag{1}$$

Under some mild hypothesis (see [1]) it is known that $W_n(u)$ converges almost surely when n tends to infinity, and the limit W is the integral of the Fourier transform of the correlation function. One would like to estimate the size of the random variable

$$I_n = \sup_{u \in [0,2\pi]} \left| W_n(u) - W(u) \right| \ . \tag{2}$$

Received 27 February 2006.

Note that one cannot apply directly the central limit theorem to W_n, and the situation becomes even worse for I_n due to the presence of the supremum.

Another well known example is the kernel density estimate. Assume the random variables X_j have a (common) density g. An often used estimator for g is constructed using a kernel ψ (a non negative integrable function) and is given by

$$g_n(s) = \frac{1}{n\alpha_n} \sum_{j=1}^{n} \psi\left(\frac{s - X_j}{\alpha_n}\right) \tag{3}$$

where the sequence (α_n) of positive numbers converges to zero. Here one would like to estimate for example the L^1 norm of $g - g_n$.

In each of the above examples, one can design a particular method to build an estimation, based for example on correlations (see for example [3], [14], [24]). However for dynamical systems this is sometimes cumbersome due to the fact that mixing properties often depend on the regularity properties of the observables. In these cases it would be interesting to have a general method to approach such questions.

We will describe below one possible method and will formulate the results for a discrete time stochastic process (X_n) with value in $\mathbf{R}^d$. It is convenient to state the following definitions.

Definition 1.1. *For a function K from $\left(\mathbf{R}^d\right)^n$ to $\mathbf{R}$, we define the uniform Lipschitz constant* $\mathrm{Lip}_j(K)$ *with respect to the variable x_j by*

$$\mathrm{Lip}_j(K) = \sup_{u,v} \sup_{x_1,\ldots,\hat{x}_j,\ldots,x_n}$$

$$\frac{\left|K\left(x_1,\ldots,x_{j-1},u,x_{j+1},\ldots,x_n\right) - K\left(x_1,\ldots,x_{j-1},v,x_{j+1},\ldots,x_n\right)\right|}{\|u - v\|}$$

We say that a function K of n variables is componentwise Lipschitz if its uniform Lipschitz constants with respect to all its variables are finite.

Note that if on $\left(\mathbf{R}^d\right)^n$ we use the ℓ^1 norm defined by

$$\left\|(x_1,\ldots,x_n)\right\| = \sum_{j=1}^{n} \left\|x_j\right\| ,$$

then a function is componentwise Lipschitz on $\left(\mathbf{R}^d\right)^n$ if and only if it is Lipschitz. There are however situations where one gets more precise results by using the Lipschitz constants $\mathrm{Lip}_j(K)$ when they depend on the index j.

Definition 1.2. *The stochastic process* (X_n) *satisfies the Devroye inequality if there is a constant* $C > 0$ *such that for any integer n and for any componentwise Lipschitz function* K *of n variables, we have*

$$\mathbf{Var}(\mathcal{K}) \le C \sum_{j=1}^{n} \mathrm{Lip}_j(K)^2$$

where $\mathcal{K}$ *is the random variable*

$$\mathcal{K} = K\big(X_1, \ldots, X_n\big) \ ,$$

and $\mathbf{Var}(\mathcal{K})$ *denotes the variance of* $\mathcal{K}$.

There is a similar definition for the exponential inequality, namely

Definition 1.3. *The stochastic process* (X_n) *satisfies the exponential inequality if there are constants* $C_1 > 0$ *and* $C_2 > 0$ *such that for any integer n and for any componentwise Lipschitz function* K *of n variables, we have*

$$\mathbf{E}\left(e^{\mathcal{K}-\mathbf{E}\left(\mathcal{K}\right)}\right) \le C_1 e^{C_2 \sum_{j=1}^{n} \mathrm{Lip}_j(K)^2} \ .$$

We now make some remarks about these definitions.

i) One can of course make similar definitions using Hölder constants instead of Lipschitz constants.

ii) In the case where the random variables take values in a finite set, the same definitions are used replacing the Lipschitz constants by the oscillations.

iii) Both inequalities keep the right dependence in n with respect to the case where the observable K is a sum of functions of one variable, namely

$$K\big(x_1, \ldots, x_n\big) = \sum_{j=1}^{n} u\big(x_j\big) \ .$$

In the independent case for example, the variance of a sum is the sum of the variances, and for the exponential estimate one gets immediately a product of n terms (to which one can apply the Hoeffding inequality, see below).

iv) Note that the definitions ask for an estimate valid for any n, not only asymptotically. Of course for small n the bound may be quite pessimistic due to bad constants.

v) The exponential bound may be seen as a rough large deviation result. Note however that it involves observables which are of much more general nature than the observables used in the usual large deviations results and also as mentioned above the bound is required to be not only asymptotic.

vi) In many situations, one exploits the bounds using some Chebyshev's inequality. This leads for example to concentration results.

The Devroye and exponential inequalities have been proved in a variety of situations. They are closely related to the phenomenon of concentration of measures. There are many works for the case of independent variables using a wealth of different techniques. We refer to [12], [19], [5], [18], [27], [22] for details and references.

The first result in the non independent case, in relation with concentration is probably [20] for finite state Markov chains. This was later extended to one dimensional Gibbs sates [21]. The proof is based on information inequalities.

For Φ-mixing processes the results were obtained in [26], and for the Ising model in dimension larger than one at high temperature in [17]. The proof of this last result relies on the Dobrushin uniqueness argument. The more general case of random fields was studied in [8] using coupling. See also [25] for related results.

The exponential inequality for the mixing absolutely continuous invariant measures of piecewise expanding maps of the interval was obtained in [9] using Perron Frobenius operators.

The Devroye inequality for SRB measures of some non uniformly hyperbolic systems was obtained in [6]. This covers unimodal maps, Hénon maps, and more generally systems with Young's tower and exponential mixing (spectral gap). The proof uses Perron Frobenius operators.

The Devroye inequality and higher moment inequalities for the low temperature Ising model were proved in [8]. In this case, it is known that the exponential inequality fails for some observables (in particular the magnetisation, see [16]). The proof is by coupling. Estimates of higher moments were also derived in [8] (see [13] and [11] for analogous results).

We also mention that the case of dynamical systems requires in general special proofs since the mixing conditions depend often on some kind of regularity of the observables.

In the next section we will show how the Devroye and exponential inequalities can be derived from a coupling with some properties of a coupling time. In the last section we will describe processes for which couplings are known to exist which lead to a proof of the inequalities. We will also give an example of application.

2 A coupling argument

It is convenient to assume that the stochastic process (X_m) is defined for all integer (positive and negative) times. Also for a sequence $(x_q, \ldots, x_p)$ (may be infinite), we will use the short hand notation x_q^p.

A time of approximate coupling (in the past) between two infinite sequences $x_{-\infty}^{\infty}$ and $z_{-\infty}^{\infty}$ is defined by

$$T\left(x_{-\infty}^{+\infty}, z_{-\infty}^{+\infty}\right) = \inf\left\{q \geq 1 \,\middle|\, \forall m \leq -q\ ,\ \left|y_m - z_m\right| \leq \frac{1}{(q+m-1)^2}\right\}\ . \tag{4}$$

For two generic sequences this quantity is of course infinite, however this will not be the case under some adequate coupling.

We can now formulate a sufficient condition for the Devroye inequality to hold.

Theorem 2.1. *Let $\left(X_j\right)$ be a mixing stationary stochastic process such that for some (finite) constant $A > 0$ we have almost surely $\|X\|_{l^\infty} \leq A$. Let $\left(Y_j\right)$ and $\left(Z_j\right)$ be two independent copies of the process $\left(X_j\right)$. Then for any componentwise Lipschitz function K of n variables and for any coupling between $\left(Y_j\right)$ and $\left(Z_j\right)$, we have*

$$\mathbf{Var}(\mathcal{K}) \leq 4(2A+1)^2 \mathbf{E}\Big(\mathbf{E}\big(T \big| Y_0, Z_0, Y_1^\infty = Z_1^\infty = X_1^\infty\big)^2\Big) \sum_{r=1}^{n} \mathrm{Lip}_r(K)^2\ .$$

This estimate is of course only useful if one can find a coupling such that the right hand side is finite (note that since K is bounded the left hand side is always finite). We will give an example in the next section.

Proof. For each integer m, we denote by $\mathscr{F}_m$ the sigma algebra generated by the random variables $X_m,\ X_{m+1},\ \ldots$ Since the function K depends only on the variables $x_1, \ldots, x_n$ and is bounded, if the process is mixing we have in the L^2 sense

$$\lim_{j\to\infty} \mathbf{E}\big(\mathcal{K} \mid \mathscr{F}_j\big) = \mathbf{E}\big(\mathcal{K}\big)\ .$$

Indeed, an easy computation leads for any integer N to the identity

$$\mathbf{Var}(\mathcal{K}) = \sum_{j=1}^{N} \mathbf{E}\left(\left[\mathbf{E}\big(\mathcal{K} \mid \mathscr{F}_j\big) - \mathbf{E}\big(\mathcal{K} \mid \mathscr{F}_{j+1}\big)\right]^2\right) + \mathbf{E}\left(\left[\mathbf{E}\big(\mathcal{K} \mid \mathscr{F}_{N+1}\big) - \mathbf{E}\left(\mathcal{K}\right)\right]^2\right)\ .$$

This implies that the sequence $\left(\mathbf{E}\left(\mathcal{K} \,\middle|\, \mathscr{F}_j\right) - \mathbf{E}\left(\mathcal{K}\right)\right)$ is Cauchy in L^2 and converges to zero since it converges weakly to zero by the mixing assumption. In particular, we have

$$\mathbf{Var}(\mathcal{K}) = \sum_{j=1}^{\infty} \mathbf{E}\left(\left[\mathbf{E}\left(\mathcal{K} \,\middle|\, \mathscr{F}_j\right) - \mathbf{E}\left(\mathcal{K} \,\middle|\, \mathscr{F}_{j+1}\right)\right]^2\right) \tag{5}$$

We are now going to estimate separately each term of this series. Although K depends only on the variables x_1^n, it is convenient to assume that it depends on all the variables $x_{-\infty}^{+\infty}$, observing that the Lipschitz constants corresponding to variables with indices outside the set $\{1, \ldots, n\}$ vanish. We will denote this extended function again by K (in other words $K(x_{-\infty}^{+\infty}) = K(x_1, \ldots, x_n)$). Using stationarity, it is therefore enough to estimate the quantity (with the notation as before $\mathcal{M} = M(X_{-\infty}^{+\infty})$)

$$\mathbf{E}\left(\left[\mathbf{E}\left(\mathcal{M} \,\middle|\, \mathscr{F}_0\right) - \mathbf{E}\left(\mathcal{M} \,\middle|\, \mathscr{F}_1\right)\right]^2\right) \ , \tag{6}$$

in terms of the Lipschitz constants of the function $M\left(x_{-\infty}^{+\infty}\right) = K\left(\mathscr{S}^{-q} x_{-\infty}^{+\infty}\right)$ where $\mathscr{S}$ is the shift (the index q will become relevant only later on). If we denote by $\left(Y_n\right)$ and $\left(Z_n\right)$ two independent copies of $\left(X_n\right)$, we get

$$\begin{aligned}
&\mathbf{E}\left(\left[\mathbf{E}\left(\mathcal{M} \,\middle|\, \mathscr{F}_0\right) - \mathbf{E}\left(\mathcal{M} \,\middle|\, \mathscr{F}_1\right)\right]^2\right) \\
&\quad = \mathbf{E}\left(\left[\mathbf{E}\left(\mathcal{M} \,\middle|\, X_0^{\infty}\right) - \mathbf{E}\left(\mathcal{M} \,\middle|\, X_1^{\infty}\right]^2\right)\right. \\
&\quad = \mathbf{E}\left(\left[\mathbf{E}\left(\mathcal{M} \,\middle|\, Y_0^{\infty} = X_0^{\infty}\right) - \mathbf{E}\left(\mathcal{M} \,\middle|\, Z_1^{\infty} = X_1^{\infty}\right)\right]^2\right) \\
&\quad = \mathbf{E}\left(\left[\mathbf{E}\left(\mathcal{M} \,\middle|\, Y_0^{\infty} = X_0^{\infty}\right) - \mathbf{E}\left(\mathbf{E}\left(\mathcal{M} \,\middle|\, Z_0\, ,\ Z_1^{\infty} = X_1^{\infty}\right) \,\middle|\, Z_1^{\infty} = X_1^{\infty}\right)\right]^2\right) \\
&\quad \le \mathbf{E}\left(\mathbf{E}\left(\left[\mathbf{E}\left(\mathcal{M} \,\middle|\, Y_0^{\infty} = X_0^{\infty}\right) - \mathbf{E}\left(\mathcal{M} \,\middle|\, Z_0\, ,\ Z_1^{\infty} = X_1^{\infty}\right)\right]^2 \,\middle|\, Z_1^{\infty} = X_1^{\infty}\right)\right) \\
&\quad = \mathbf{E}\left(\mathbf{E}\left(\left[\mathbf{E}\left(\mathcal{M} \,\middle|\, Y_0^{\infty}\right) - \mathbf{E}\left(\mathcal{M} \,\middle|\, Z_0^{\infty}\right)\right]^2 \,\middle|\, Y_1^{\infty} = Z_1^{\infty} = X_1^{\infty}\right)\right) \ .
\end{aligned} \tag{7}$$

Since by hypothesis $\|Y\|_{l^\infty} \le A$, $\|Z\|_{l^\infty} \le A$, and $Y_j = Z_j$ for any $j \ge 1$, we get

$$\begin{aligned}
\left|M\left(Y_{-\infty}^{+\infty}\right) - M\left(Z_{-\infty}^{+\infty}\right)\right| \ &\le \ 2A \sum_{j=-T\left(Y_{-\infty}^{+\infty}, Z_{-\infty}^{+\infty}\right)+1}^{0} \mathrm{Lip}_j(M) \\
&+ \sum_{j=-\infty}^{-T\left(Y_{-\infty}^{+\infty}, Z_{-\infty}^{+\infty}\right)} \frac{\mathrm{Lip}_j(M)}{\left(j + T\left(Y_{-\infty}^{+\infty}, Z_{-\infty}^{+\infty}\right) - 1\right)^2} \ .
\end{aligned}$$

Assume now we have a coupling μ_{Y_0,Z_0,X_1^∞} between $\mathbf{P}\big(Y_{-\infty}^{-1}|Y_0, Y_1^\infty = X_1^\infty\big)$ and $\mathbf{P}\big(Z_{-\infty}^{-1}|Z_0, Z_1^\infty = X_1^\infty\big)$, then we can write when $Y_1^\infty = Z_1^\infty = X_1^\infty$

$$\mathbf{E}\big(\mathcal{M}\,\big|\,Y_0^\infty\big) - \mathbf{E}\big(\mathcal{M}\,\big|\,Z_0^\infty\big) = \int \big(M\big(Y_{-\infty}^\infty\big) - M\big(Z_{-\infty}^\infty\big)\big)\,d\mu_{Y_0,Z_0,X_1^\infty}\big(Y_{-\infty}^{-1}, Z_{-\infty}^{-1}\big)$$

and using the above estimation we get

$$\begin{aligned}
&\Big|\mathbf{E}\big(\mathcal{M}\,\big|\,Y_0^\infty\big) - \mathbf{E}\big(\mathcal{M}\,\big|\,Z_0^\infty\big)\Big| \\
&\quad\le \sum_{p=0}^{\infty} \mu_{Y_0,Z_0,X_1^\infty}\big(T = p\big)\left(2A\sum_{j=-p+1}^{0}\mathrm{Lip}_j(M) + \sum_{j=-\infty}^{-p}\frac{\mathrm{Lip}_j(M)}{(j+p-1)^2}\right) \qquad (8)\\
&\quad\le (2A+1)\sum_{p=0}^{\infty}\sum_{j=-\infty}^{-p}\mu_{Y_0,Z_0,X_1^\infty}\big(T \ge p\big)\frac{\mathrm{Lip}_j(M)}{(j+p-1)^2}\,.
\end{aligned}$$

Replacing M by $K \circ \mathscr{S}^q$ and using equations (7) and (5) we get

$$\begin{aligned}
&\mathbf{Var}(\mathcal{K})\\
&\le (2A+1)^2\sum_q \mathbf{E}\left(\left[\sum_{p=0}^{\infty}\sum_{j=-\infty}^{-p}\mu_{Y_0,Z_0,X_1^\infty}\,(T \ge p)\,\frac{\mathrm{Lip}_{j+q}(K)}{(j+p-1)^2}\right]^2\right).
\end{aligned}$$

Using Schwarz inequality we get

$$\begin{aligned}
\mathbf{Var}(\mathcal{K}) \;&\le\; (2A+1)^2\sum_q \mathbf{E}\left(\left[\sum_{p=0}^{\infty}\sum_{j=-\infty}^{-p}\mu_{Y_0,Z_0,X_1^\infty}\big(T \ge p\big)\frac{1}{(j+p-1)^2}\right]\right.\\
&\qquad\left.\times\left[\sum_{p=0}^{\infty}\sum_{j=-\infty}^{-p}\mu_{Y_0,Z_0,X_1^\infty}\big(T \ge p\big)\frac{\mathrm{Lip}_{j+q}(K)^2}{(j+p-1)^2}\right]\right)\\
&\le\; 2(2A+1)^2\sum_q \mathbf{E}\Big(\mathbf{E}\big(T\,\big|\,Y_0, Z_0, Y_1^\infty = Z_1^\infty = X_1^\infty\big)\\
&\qquad\left.\sum_{p=0}^{\infty}\sum_{j=-\infty}^{-p}\mu_{Y_0,Z_0,X_1^\infty}\big(T \ge p\big)\frac{\mathrm{Lip}_{j+q}(K)^2}{(j+p-1)^2}\right)\\
&\le\; 4(2A+1)^2\mathbf{E}\Big(\mathbf{E}\big(T\,\big|\,Y_0, Z_0, Y_1^\infty = Z_1^\infty = X_1^\infty\big)^2\Big)\sum_{r=1}^{n}\mathrm{Lip}_r(K)^2\,.
\end{aligned}$$

□

Note that under similar assumptions on the above coupling, one can derive $(\theta, \mathscr{F}, \psi)$ weak dependence properties in the sense of [13], see also [11].

The exponential inequality can be established under a stronger assumption on the coupling (see also [25] for analogous results).

Theorem 2.2. *Let $\left(X_j\right)$ be a stationary mixing stochastic process such that for some (finite) constant $A > 0$ we have almost surely $\|X\|_{l^\infty} \le A$. Let $\left(Y_j\right)$ and $\left(Z_j\right)$ be two independent copies of the process $\left(X_j\right)$. Then for any componentwise Lipschitz function K of n variables and for any coupling between $\left(Y_j\right)$ and $\left(Z_j\right)$, we have*

$$\mathbf{E}\left(e^{\mathcal{K}-\mathbf{E}(\mathcal{K})}\right) \le e^{2(2A+1)^2\left(1+\left\|\mathbf{E}\left(T^2\middle|Y_0,Z_0,Y_1^\infty=Z_1^\infty=X_1^\infty\right)\right\|_{L^\infty}\right)^2 \sum_{r=1}^n \mathrm{Lip}_r(K)^2} .$$

Proof. The proof is rather similar to the proof of the previous Theorem. For any finite integer N we have since $\mathcal{K}$ does not depend on $X_{-\infty}^0$

$$\begin{aligned}\mathbf{E}\left(e^{\mathcal{K}-\mathbf{E}(\mathcal{K})}\right) &= \mathbf{E}\left(\prod_{q=1}^{N-1} e^{\mathbf{E}(\mathcal{K}|\mathscr{F}_q)-\mathbf{E}(\mathcal{K}|\mathscr{F}_{q+1})} \quad e^{\mathbf{E}(\mathcal{K}|\mathscr{F}_N)-\mathbf{E}(\mathcal{K})}\right) ,\\ &= \mathbf{E}\left(e^{\mathbf{E}(\mathcal{K}|\mathscr{F}_N)-\mathbf{E}(\mathcal{K})} \prod_{q=1}^{N-1} \mathbf{E}\left(e^{\mathbf{E}(\mathcal{K}|\mathscr{F}_q)-\mathbf{E}(\mathcal{K}|\mathscr{F}_{q+1})}\middle|\mathscr{F}_q\right)\right) .\end{aligned}$$

Using Hoeffding's inequality (see [12]) we get

$$\mathbf{E}\left(e^{\mathcal{K}-\mathbf{E}(\mathcal{K})}\right) \le \mathbf{E}\left(e^{\mathbf{E}(\mathcal{K}|\mathscr{F}_N)-\mathbf{E}(\mathcal{K})} \prod_{q=0}^{N-1} e^{\mathrm{Osc}_{X_q}\mathbf{E}(\mathcal{K}|\mathscr{F}_q)^2/8}\right) . \tag{9}$$

As before, by stationarity it is enough to estimate $\mathrm{Osc}_{X_0}\mathbf{E}(\mathcal{M}|\mathscr{F}_0)$. This is given by equation (8), and we get using Schwarz inequality

$$\begin{aligned}&\mathrm{Osc}_{X_0}\mathbf{E}(\mathcal{M}|\mathscr{F}_0)\\ &\le (2A+1)\sum_{p=0}^{\infty}\sum_{j=-\infty}^{-p} \frac{\sqrt{(p+1)\,\mu_{Y_0,Z_0,X_1^\infty}\left(T \ge p\right)}}{(j+p-1)}\\ &\qquad \frac{\mathrm{Lip}_j(M)\sqrt{\mu_{Y_0,Z_0,X_1^\infty}\left(T \ge p\right)}}{(j+p-1)\sqrt{p+1}}\\ &\le 2(2A+1)\Big(1+\mathbf{E}\left(T^2\middle|Y_0,Z_0,Y_1^\infty=Z_1^\infty\right)\Big)^{1/2}\end{aligned}$$

$$\times \left(\sum_{p=0}^{\infty} \sum_{j=-\infty}^{-p} \frac{\text{Lip}_j(M)^2 \, \mu_{Y_0, Z_0, X_1^\infty}\big(T \geq p\big)}{(j+p-1)^2(p+1)} \right)^{1/2}$$

$$\leq \; 4(2A+1)\Big(1 + \big\| \mathbf{E}\big(T^2 \big| Y_0, Z_0, Y_1^\infty = Z_1^\infty\big) \big\|_{L^\infty}\Big) \left(\sum_{j=-\infty}^{0} \frac{\text{Lip}_j(M)^2}{(1-j)^2} \right)^{1/2}$$

since for $j \leq -p$ we have $(j+p-1)^2(p+1)^2 \geq (1-j)^2$. It now follows from (9) that for any integer N we have

$$\mathbf{E}\left(e^{\mathcal{K} - \mathbf{E}(\mathcal{K})}\right)$$

$$\leq \prod_{q=0}^{N-1} e^{2(2A+1)^2 \left(1 + \left\| \mathbf{E}\left(T^2 \middle| Y_0, Z_0, Y_1^\infty = Z_1^\infty\right) \right\|_{L^\infty}\right)^2 \sum_{j=-\infty}^{0} \frac{\text{Lip}_{j+q}(L)^2}{(1-j)^2}} \mathbf{E}\left(e^{\mathbf{E}(\mathcal{K}|\mathscr{F}_N) - \mathbf{E}(\mathcal{K})}\right) ,$$

$$\leq \; e^{2(2A+1)^2 \left(1 + \left\| \mathbf{E}\left(T^2 \middle| Y_0, Z_0, Y_1^\infty = Z_1^\infty\right) \right\|_{L^\infty}\right)^2 \sum_r \text{Lip}_r(L)^2} \mathbf{E}\left(e^{\mathbf{E}(\mathcal{K}|\mathscr{F}_N) - \mathbf{E}(\mathcal{K})}\right) .$$

Since K is bounded, we have

$$\left| \mathbf{E}\left(e^{\mathbf{E}(\mathcal{K}|\mathscr{F}_N) - \mathbf{E}(\mathcal{K})}\right) - 1 \right| \leq e^{2\|K\|_{L^\infty}} \mathbf{E}\Big(\big(\mathcal{K}|\mathscr{F}_N) - \mathbf{E}(\mathcal{K})\big)^2\Big)$$

and as explained before, it follows from the mixing property that the right hand side tends to zero when N tends to infinity. □

Estimations for higher moments in the case where the exponential inequality fails can be obtained by similar coupling techniques using the Marcinkiewicz-Zygmund inequality (see [8]) or the Dedecker Doukhan inequality [10], [13], [11]. See also [23] for related results.

3 Examples and applications

Couplings satisfying the condition

$$\big\| \mathbf{E}\big(T^2 \big| Y_0, Z_0, Y_1^\infty = Z_1^\infty\big) \big\|_{L^\infty} < \infty \tag{10}$$

have been obtained for piecewise expanding maps of the interval (taken as $[0, 1]$). We recall that these are maps f such that there exists an increasing sequence $a_0 = 0 < a_1 \ldots < a_k = 1$ such that on each subinterval $]a_j, a_{j+1}[$ $(0 \leq j < k)$, the map is continuous and monotone and the graph extends to a C^2 map in a neighborhood of $]a_j, a_{j+1}[$. Moreover we assume that there is a constant $s > 1$

such that in all intervals $]a_j, a_{j+1}[$ the modulus of the slope (of some iterate) is larger than one. If the transformation has a dense orbit, it is known that there is a unique absolutely continuous invariant probability measure. We refer to [15] or [4] for more details and properties of these dynamical systems. Using the density of these invariant measures one can define a Markov chain describing the backward orbits of the dynamical system (see [2]). Although these chains have rather singular transition probabilities (each transition probability is atomic with a finite number of atoms less than or equal to $k-1$), it was proved in [2] that there are two constants $C > 1$ and $\rho < 1$ such that for any x and y in $[0, 1]$, except may be for a finite number of points, there is a coupling $\mu_{x,y}$ such that

$$\mu_{x,y}\Big(\sup_{q>n} \rho^{-q}|X_{-q} - Y_{-q}| > C^{-1}\Big) \leq C\rho^n \ ,$$

where (X_{-n}) and (Y_{-n}) are the processes of preimages of x and y, namely $f(X_{-n}) = X_{-n+1}$. It then follows easily that condition (10) is satisfied.

Several applications have already been developed for these inequalities, we mention in particular the estimation of the fluctuations of the empirical covariance, of the empirical integrated periodogram, of the empirical measure, and of the kernel density estimate. One can also use these estimates to prove almost sure central limit theorems (with estimates on the velocity of convergence), and concentration. We refer to [12] and [7] for details and references. We will derive below an estimate for the fluctuation of the kernel density estimator, and show a relation with concentration.

Assume we have a real valued discrete time stationary stochastic process (X_n) which satisfies the Devroye inequality. Assume also that the random variables have a (common) density g, and consider a kernel density estimate g_n (see 3) with a kernel ψ which is a Lipschitz function with compact support. In order to estimate the variance of $\|g - g_n\|_{L^1}$, we consider the function

$$K\big(x_1, \ldots, x_n\big) = \int \left| \frac{1}{n\alpha_n} \sum_{j=1}^{n} \psi\left(\frac{s - x_j}{\alpha_n}\right) - g(s) \right| ds \ .$$

It is easy to see that K is Lipschitz and moreover

$$\mathrm{Lip}_j(K) \leq \frac{1}{n\alpha_n} M(\psi) \ ,$$

where

$$M(\psi) = \sup_{x,y} \frac{1}{|x-y|} \int \big|\psi(t-x) - \psi(t-y)\big| dt < \infty \ .$$

Applying Devroye's inequality (1.2) we get

$$\mathbf{Var}\big(\|g - g_n\|_{L^1}\big) \leq C\frac{1}{n\alpha_n^2}M(\psi)^2 \ .$$

One would like now to use the Chebyshev inequality in order to estimate the fluctuation of $\|g - g_n\|_{L^1}$. For this purpose, one needs to estimate $\mathbf{E}\big(\|g - g_n\|_{L^1}\big)$. There are several ways to obtain such an estimate, here we will use again the Devroye inequality. We refer to [9] and [7] for estimates using correlations. We will from now on assume that the random variables (X_n) are bounded, namely there are constants $a > 0$ and b such that almost surely $X_n \in [b, a + b]$. We also assume that the kernel ψ has support in the interval $[-c, c]$ for some $c > 0$. Using Schwarz inequality, we get

$$\mathbf{E}\big(\|g - g_n\|_{L^1}\big)^2 \leq \int_{b-c\alpha_n}^{b+a+c\alpha_n} \mathbf{E}\big((g_n(s) - g(s))^2\big)ds \ .$$

As mentioned before, if information on correlations are available, they can be used at this point, and they eventually give better estimates than what is obtained below(see [14], [24] and references therein for results in these directions). Here we will derive a rough estimate applying again the Devroye inequality (1.2) to $g_n(s) - g(s)$ for each fixed s. Defining the function K (for fixed s) by

$$K(x_1, \ldots, x_n) = \left| \frac{1}{n\alpha_n}\sum_{j=1}^{n} \psi\left(\frac{s - x_j}{\alpha_n}\right) - g(s) \right| \ ,$$

we get for any $s \in \mathbf{R}$

$$\mathrm{Lip}_j(K) \leq \frac{\mathrm{Lip}(\psi)}{n\alpha_n^2} \ .$$

Therefore applying the Devroye inequality we get

$$\mathbf{Var}\big(g(s) - g_n(s)\big) \leq C\frac{1}{n\alpha_n^4}\mathrm{Lip}_j(\psi)^2 \ .$$

Combining the above estimates, we get

$$\mathbf{E}\big(\|g - g_n\|_{L^1}^2\big) \leq \frac{\mathcal{O}(1)}{n\alpha_n^4} + \int_{b-c\,\alpha_n}^{b+a+c\,\alpha_n} \Big[\mathbf{E}\big(g(s) - g_n(s)\big)\Big]^2 ds \ .$$

For the last term, we have

$$\mathbf{E}\big(g(s) - g_n(s)\big) = \int \psi(v)\big(g(s + \alpha_n v) - g(s)\big)dv.$$

If g is known to be Lipschitz for example, we get

$$\left|\mathbf{E}\big(g(s) - g_n(s)\big)\right| \leq \mathcal{O}(1)\alpha_n,$$

and finally using the optimal choice for the bound $\alpha_n = \mathcal{O}(1)n^{-1/6}$ we obtain the following result.

Theorem 3.1. *Assume the stationary mixing process (X_n) satisfies the Devroye inequality, and the random variable X_0 is almost surely bounded with a Lipschitz density g. Let g_n be the kernel density estimate* (3), *with ψ non negative, of integral one, Lipschitz with compact support. Then*

$$\mathbf{E}\big(\|g - g_n\|_{L^1}^2\big) \leq \frac{\mathcal{O}(1)}{n^{1/3}} \ .$$

As we already mentioned, the Devroye and exponential estimates have many interesting consequences. Another one is the so called concentration of measure. Although this is a well known result, for the convenience of the reader we now give a simple example. For E a subset of $\mathbf{R}^n$ and ϵ a positive number, we denote by E_ϵ the ϵ neighborhood of E (in the l^1 norm) defined by

$$E_\epsilon = \Big\{ \big(x_1, \ldots, x_n\big) \in \mathbf{R}^n \,\big|\, \exists\, (y_1, \ldots, y_n) \in E \\ \text{such that } n^{-1}\sum_{j=1}^{n} \left|x_j - y_j\right| \leq \epsilon \Big\}.$$

Theorem 3.2. *Let (X_m) be a real valued mixing process satisfying the Devroye inequality* (1.2). *Let E be a measurable subset of $\mathbf{R}^n$ and let $\mathcal{T}_E$ be the subset of the probability space defined by*

$$\mathcal{T}_E = \Big\{ \, (X_1, \ldots, X_n) \in E \Big\} \ .$$

Assume that $\mathbf{P}\left(\mathcal{T}_E\right) > 0$, then for any $\epsilon > 0$, the complement $\mathcal{T}_{E_\epsilon}^c$ of $\mathcal{T}_{E_\epsilon}$ satisfies

$$\mathbf{P}\left(\mathcal{T}_{E_\epsilon}^c\right) \leq \left(1 + \frac{1}{\mathbf{P}\left(\mathcal{T}_E\right)}\right) \frac{C}{n\epsilon^2}$$

where C is the constant in the Devroye inequality (1.2).

Roughly speaking this means that in large dimension, if we have a set of positive measure, a small neighborhood is almost of full measure.

Proof. Consider the function

$$K\left(x_1, \ldots, x_n\right) = \inf_{(y_1,\ldots,y_n)\in E} n^{-1} \sum_{j=1}^{n} \left|x_j - y_j\right| .$$

This function is componentwise Lipschitz, and the Lipschitz constants are all bounded by $1/n$. Applying the Devroye inequality, we conclude that

$$\mathbf{Var}(\mathcal{K}) \le \frac{C}{n} .$$

In order to apply the Chebyshev inequality, we need to estimate $\mathbf{E}(\mathcal{K})$. We observe that $K\left(x_1^n\right) = 0$ if $\left(x_1^n\right) \in E$, therefore

$$\mathbf{E}(\mathcal{K})^2 \mathbf{P}\left(\mathcal{T}_E\right) \le \mathbf{E}\Big(\chi_{\mathcal{T}_E}\left(\mathcal{K} - \mathbf{E}(\mathcal{K})\right)^2\Big) \le \mathbf{Var}(\mathcal{K}) \le \frac{C}{n} .$$

Combining the two estimates, we get

$$\mathbf{E}(\mathcal{K}^2) \le \left(1 + \frac{1}{\mathbf{P}\left(\mathcal{T}_E\right)}\right) \frac{C}{n} .$$

We now observe that $\mathcal{T}_E^c = \left\{\mathcal{K} > \epsilon\right\}$ and the result follows using Chebyshev's inequality. $\square$

Acknowledgements. This paper was written following a talk at the ninth Brazilian school of probability. The author is grateful to A.Galves for his kind invitation to the meeting and for his encouragements to write this paper.

References

[1] P.J. Brockwell and R.A. Davis, *Time Series: Theory and Methods*. Springer-Verlag, New York, 1996.

[2] A.D. Barbour, R.M. Gerrard and G.Reinert, Iterates of expanding maps. Probab. Theory Related Fields, **116** (2000), 151–180.

[3] J.-M. Bardet, P. Doukhan and J.R. Leon, A uniform central limit theorem for the periodogram and its applications to Whittle parametric estimation for weakly dependent time series. http://www.crest.fr/pageperso/doukhan/publications.htm.

[4] A. Broise, *Transformations dilatantes de l'intervalle et théorèmes limites, Études spectrales d'opérateurs de transfert et applications*. Astérisque, **238** (1996).

[5] S. Boucheron, O. Bousquet and G. Lugosi, Concentration inequalities. In *Advanced Lectures in Machine Learning*, 208–240 O. Bousquet, U.v. Luxburg and G. Rätsch editors, Springer (2004).

[6] J.-R. Chazottes, P. Collet and B.Schmitt, Devroye Inequality for a class of non-uniformly hyperbolic dynamical systems. Nonlinearity, **18** (2005), 2323–2340.

[7] J.-R. Chazottes, P. Collet and B.Schmitt, Statistical Consequences of Devroye inequality for processes. Applications to a class of non-uniformly hyperbolic dynamical systems. Nonlinearity, **18** (2005), 2341–2364.

[8] J.-R. Chazottes, P. Collet, C. Külske and F. Redig, Deviation inequalities via coupling for stochastic processes and random fields. Preprint 2005. math.PR/0503483.

[9] P. Collet, S. Martinez and B. Schmitt, Exponential inequalities for dynamical measures of expanding maps of the interval. Probab. Th. and Relat. Fields, **123** (2002), 301–322.

[10] J. Dedecker and P. Doukhan, A new covariance inequality and applications. Stochastic Process. Appl., **106** (2003), 63–80.

[11] J. Dedecker and C. Prieur, New dependence coefficients. Examples and applications to statistics. Probab. Theory and Relat. Fields, **132** (2005), 203–236.

[12] L. Devroye, Exponential inequalities in nonparametric estimation. In *Nonparametric functional estimation and related topics* (Spetses, 1990), 31–44, NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci., **335**, Kluwer Acad. Publ., Dordrecht (1991).

[13] P. Doukhan and S. Louhichi, A new weak dependence condition and applications to moment inequalities. Stochastic Process. Appl., **84** (1999), 313–342.

[14] P. Doukhan and S. Louhichi, Estimation de la densité d'une suite faiblement dépendante. C.R. Acad. Sci. Paris Sér. I Math., **327** (1998), 989–992. Functional estimation of a density under a new weak dependence condition. Scand. J. Statist., **28** (2001), 325–341.

[15] F. Hofbauer and G. Keller, Ergodic properties of invariant measures for piecewise monotonic transformations. Math. Z., **180** (1982), 119–140.

[16] D. Ioffe, Exact large deviation bounds up to T_c for the Ising model in two dimensions. Probab. Theory Related Fields, **102** (1995), 313–330.

[17] C. Külske, Concentration inequalities for functions of Gibbs fields with application to diffraction and random Gibbs measures. Comm. Math. Phys., **239** (2003), 29–51.

[18] M. Ledoux, *The concentration of measure phenomenon*. Mathematical Surveys and Monographs, **89**. American Mathematical Society, Providence, RI (2001).

[19] C. McDiarmid, Concentration. Probabilistic methods for algorithmic discrete mathematics, 195–248, Algorithms Combin., **16** (1998), Springer, Berlin.

[20] K. Marton, A measure concentration inequality for contracting Markov chains. Geom. Funct. Anal., **6** (1996), 556–571.

[21] K. Marton, Measure concentration for a class of random processes. Probab. Theory Relat. Fields, **110** (1998), 427–439.

[22] P. Massart, *Concentration inequalities and model selection*. Ecole d'été de Probabilités de Saint-Flour 2003. Lecture Notes in Mathematics, Springer (to appear in 2006).

[23] V. Maume-Deschamps, Concentration inequalities and estimation of conditional probabilities. To appear in Lect. Notes in Statistics.

[24] C. Prieur, Estimation de la densité invariante de systèmes dynamiques en dimension 1. C.R. Acad. Sci. Paris Sér. I Math., **332** (2001), 761–764. Density estimation for one-dimensional dynamical systems. ESAIM Probab. Statist., **5** (2001), 51–76.

[25] E. Rio, Inégalités de Hoeffding pour les fonctions Lipschitziennes de suites dépendantes. C. R. Acad. Sci. Paris Sér. I Math., **330** (2000), 905–908.

[26] P. Samson, Concentration of measure inequalities for Markov chains and Φ mixing processes. Ann. Probab., **28** (2000), 416–461.

[27] M. Talagrand, Concentration of measure and isoperimetric inequalities in product spaces. Inst. Hautes Etudes Sci. Publ. Math., **81** (1995), 73–205. New concentration inequalities in product spaces. Invent. Math., **126** (1996), 505–563. A new look at independence. Ann. Probab., **24** (1996), 1–34.

Pierre Collet
Centre de Physique Théorique
CNRS UMR 7644
Ecole Polytechnique
F-91128 Palaiseau Cedex
FRANCE

E-mail: collet@cpht.polytechnique.fr