LIMIT THEOREMS AND INEQUALITIES VIA MARTINGALE METHODS

Jean-René Chazottes[1], Christophe Cuny[2], Jérôme Dedecker[3], Xiequan Fan[4]
and Sarah Lemler[5]

Abstract. In this notes, we first give a brief overwiew of martingales methods, from Paul Lévy (1935) untill now, to explain why these methods have become a central tool in probability, statistic and ergodic theory. Next, we present some recent results for/or based on martingales: exponential bounds for super-martingale, concentration inequalities for Lipschitz functionals of dynamical systems, oracle inequalities for the Cox model in a high dimensional setting, and invariance principles for stationary sequences.

Résumé. Dans ces notes, nous faisons d'abord un rapide survol des méthodes de martingales, depuis Paul Lévy (1935) jusqu'à nos nos jours, afin d'expliquer pourquoi ces métodes sont devenues centrales en probabilité, statistique et théorie ergodique. Ensuite, nous présentons des résultats récents sur/ou fondés sur les martingales: des inégalités exponentielles pour les sur-martingales, des inégatés de concentration pour les fonctionnelles lipschitziennes de systèmes dynamiques, des inégalités oracle pour le modèle de Cox en grande dimension, et des principes d'invariances pour les suites stationnaires.

1. Introduction

In this introduction, our goal is twofold. We shall first recall the first developments of the theory of martingales in the field of limit theorems and inequalities. Next we shall briefly explain how these results can be extended to more general sequences. Of course, we shall not give all the references on the subject, this would be a too long exercise, but we shall try to give some essential references from the beginning until now, to explain how martingale methods have become a central tool in probability, statistic and ergodic theory.

In order to write the main results without giving too many notations, we shall consider in this introduction the simple case of stationary sequences. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, let T be a bijective bi-measurable transformation preserving the probability $\mathbb{P}$, and let $\mathcal{I}$ be the σ-algebra of T-invariant sets. Let X_0 be a centered and square integrable random variable with variance σ^2, and define the stationary sequence X_i by $X_i = X_0 \circ T^i$. Given a σ-algebra $\mathcal{F}_0$ such that $\mathcal{F}_0 \subseteq T^{-1}(\mathcal{F}_0)$, we introduce the non-decreasing filtration $\mathcal{F}_i = T^{-i}(\mathcal{F}_0)$, and the tail σ-algebras $\mathcal{F}_{-\infty} = \bigcap_{k \in \mathbb{Z}} \mathcal{F}_k$ and $\mathcal{F}_\infty = \bigvee_{k \in \mathbb{Z}} \mathcal{F}_k$. Let S_n denote the partial sum $S_n = X_1 + \cdots + X_n$.

[1] CPHT, CNRS UMR 7644, École Polytechnique, 91128 Palaiseau cedex, France.

[2] Laboratoire MAS, École Centrale de Paris, Grande Voie des Vignes, 92295 Chatenay-Malabry cedex, France.

[3] Laboratoire MAP5, CNRS UMR 8145, Université Paris-Descartes, Sorbonne Paris Cité, 45 rue des Saints Pères, 75270 Paris cedex 06, France.

[4] LMBA, UMR CNRS 6205, Université de Bretagne Sud, Campus de Tohannic, 56017 Vannes, France

[5] Laboratoire Statistique et Génome, CNRS UMR 8071- USC INRA, Université dEvry Val dEssonne, 23 bvd de France 91 037 vry, France

1.1. The martingale case

Let us first recall some important results when $(X_i)_{i \in \mathbb{Z}}$ is a sequence of martingale differences adapted to the filtration $\mathcal{F}_i$, that is when X_0 is $\mathcal{F}_0$-measurable and $\mathbb{E}(X_0|\mathcal{F}_{-1}) = 0$ almost surely.

Lévy (1935) provided the first generalization of the central limit theorem (CLT) for sequence of independent and identically distributed (iid) random variables to the martingale case. His result writes as follows: if $\mathbb{E}(X_0^2|\mathcal{F}_{-1}) = \sigma^2$ almost surely, then $n^{-1/2} S_n$ converges in distribution to $\mathcal{N}(0, \sigma^2)$.

The strong assumption $\mathbb{E}(X_0^2|\mathcal{F}_{-1}) = \sigma^2$ almost surely has been removed independently by Billingsley (1961) and Ibragimov (1963), who proved that

$$\frac{S_n}{\sqrt{n}} \quad \text{converges in distribution to } \sqrt{\eta} N, \tag{1.1}$$

where $\eta = \mathbb{E}(X_0^2|\mathcal{I})$ and N is a standard Gaussian random variable independent of $\mathcal{I}$. Actually, both proofs were done in the case where $\mathcal{I}$ is $\mathbb{P}$-trivial (the ergodic case) but they remain unchanged in the general situation. As we shall see later on, this result is not only a slight generalization of the iid situation: in many interesting cases, the partial sum of a stationary sequence can be approximated by a martingale with stationary differences, to which the result of Billingsley-Ibramov applies. The CLT for martingales has been next extended to the non stationary case by Brown (1971), who also proved the weak invariance principle for the process $\{S_{[nt]}, t \in [0,1]\}$. A version of Billingsley-Ibragimov's CLT for random variables with values in 2-smooth Banach spaces is given in Woyczinsky (1975).

At the same time, different authors have obtained moments and exponential bounds for martingales. For instance Burkholder (1966, 1973) have proved that, if $\mathbb{E}(|X_0|^p) < \infty$ for some $p \in]1, \infty[$, then there exist two positive constants c_p and C_p depending on p such that

$$c_p \left\| \left(\sum_{k=1}^{n} X_k^2 \right)^{1/2} \right\|_p \leq \left\| \max_{1 \leq k \leq n} |S_k| \right\|_p \leq C_p \left\| \left(\sum_{k=1}^{n} X_k^2 \right)^{1/2} \right\|_p. \tag{1.2}$$

The extension of this famous inequality to the case of continuous martingale is known as the Burkholder-Davis-Gundy inequality (Burkholder Davis and Gundy (1972)). Burkholder's inequality (1.2) is also true in separable Hilbert spaces (using the square of the norm of X_i instead of X_i^2), and the upper bound remains valid in 2-smooth Banach spaces (see Pinelis (1994)).

In the sixties, Hoeffding (1963) and Azuma (1967) proved an exponential bound for the deviation of S_n. In the stationary case, a version of this result writes as follows: if $X_0 \in [Y_0, Y_0 + \ell]$ almost surely, where Y_0 is a $\mathcal{F}_{-1}$-measurable random variable and ℓ is some positive constant, then, for any positive x,

$$\mathbb{P}\Big(\max_{1 \leq k \leq n} |S_k| > x \Big) \leq 2 \exp\Big(-\frac{2x^2}{n\ell^2} \Big). \tag{1.3}$$

This inequality implies in particular that, if $\|X_0\|_\infty < \infty$, then, for any positive x,

$$\mathbb{P}\Big(\max_{1 \leq k \leq n} |S_k| > x \Big) \leq 2 \exp\Big(-\frac{x^2}{2n\|X_0\|_\infty^2} \Big). \tag{1.4}$$

An extension of Inequality (1.4) to 2-smooth Banach spaces is given in Pinelis (1992). Other moments or exponential bounds may be found in the papers by Burkholder (1973), Freedman (1975), Pinelis (1994) or Liu and Watbled (2009).

McDiarmid (1989) has given an interesting application of (1.3) to Lipshitz functions of independent sequences. In the iid case, his inequality writes as follows: let $(\xi_i)_{1 \leq i \leq n}$ be a sequence of iid random variables with values

in $\mathcal{X}$, and let d be some distance on $\mathcal{X}$. Let f be a function from $\mathcal{X}^n$ to $\mathbb{R}$ such that

$$|f(x_1, \ldots, x_n) - f(y_1, \ldots, y_n)| \leq K \sum_{i=1}^{n} d(x_i, y_i).$$

Let $M = \|d(\xi_1, \xi_2)\|_\infty$. Then, for any positive x,

$$\mathbb{P}(f(\xi_1, \ldots, \xi_n) - \mathbb{E}(f(\xi_1, \ldots, \xi_n)) \geq x) \leq \exp\left(-\frac{2x^2}{nM^2K^2}\right). \tag{1.5}$$

McDiarmid (1998) has pointed out a number of applications of Inequality (1.5). This inequality is also an important tool in classification problems (see for instance Freund, Mansour and Shapire (2004)).

Rio (2000) has extended McDiarmid's inequality to a large class of dependent sequences. Collet, Martinez and Schmitt (2002) have proved that McDiarmid's inequality holds when $\xi_i = T^i$ is the i-th iterates of an uniformly expanding map T of $[0, 1]$, and $d(x, y) = |x - y|$.

1.2. The general case

In this section, we shall no longer assume that $\mathbb{E}(X_0|\mathcal{F}_{-1}) = 0$ almost surely, and see how the results of the preceding section can be extended to more general sequences. The main step in this direction is due to Gordin (1969). Gordin noticed that, if

$$\sum_{k=1}^{\infty} \|\mathbb{E}(X_k|\mathcal{F}_0)\|_2 < \infty \quad \text{and} \quad \sum_{k=-\infty}^{0} \|X_k - \mathbb{E}(X_k|\mathcal{F}_0)\|_2 < \infty \tag{1.6}$$

then $X_0 = D_0 + Z - Z \circ T$, where

$$D_0 = \sum_{k \in \mathbb{Z}} (\mathbb{E}(X_k|\mathcal{F}_0) - \mathbb{E}(X_k|\mathcal{F}_{-1})) \quad \text{and} \quad Z = \sum_{k=-\infty}^{0} (X_k - \mathbb{E}(X_k|\mathcal{F}_0)) - \sum_{k=1}^{\infty} \mathbb{E}(X_k|\mathcal{F}_0).$$

Notice that both D_0 and Z are square integrable, that D_0 is $\mathcal{F}_0$-measurable and such that $\mathbb{E}(D_0|\mathcal{F}_{-1}) = 0$. Moreover, setting $D_i = D_0 \circ T^i$, and $M_n = D_1 + \cdots + D_n$, one has $S_n = M_n + Z \circ T - Z \circ T^{n+1}$. Since $n^{-1/2}Z \circ T^{n+1}$ converges in probability (and even almost surely) to zero, it follows from Billingsley-Ibragimov's CLT that (1.1) holds with $\eta = \mathbb{E}(D_0^2|\mathcal{I})$. Gordin's CLT is a major result, which can be applied to many dependent sequences, including mixing sequences, stationary Markov chains (see Gordin and Lifsiz (1978)), and certain dynamical systems (see the paper by Liverani (1996)). As a strinking application, let us mention the paper by Le Borgne (2002) who applied Gordin's CLT to the iterates of ergodic automorphisms of the d-dimensional torus.

Heyde (1975) noticed that Gordin's proof remains valid by assuming only that

$$\text{the sequences} \quad \mathbb{E}(S_n|\mathcal{F}_0) \quad \text{and} \quad S_n - \mathbb{E}(S_n|\mathcal{F}_n) \quad \text{converge in } \mathbb{L}^2 \tag{1.7}$$

and that, under this condition, the process $\{S_{[nt]}, t \in [0, 1]\}$ satisfies both the weak and the strong invariance principle.

Alternatively, it follows from the papers by Hannan (1973) and Heyde (1974), that if

$$\mathbb{E}(X_0|\mathcal{F}_\infty) = X_0 \text{ a. s.}, \quad \mathbb{E}(X_0|\mathcal{F}_{-\infty}) = 0 \text{ a. s. and} \sum_{k \in \mathbb{Z}} \|\mathbb{E}(X_k|\mathcal{F}_0) - \mathbb{E}(X_k|\mathcal{F}_{-1})\|_2 < \infty \tag{1.8}$$

then Gordin's martingale difference D_0 is well defined and square integrable, and $\|S_n - M_n\|_2 = o(\sqrt{n})$, so that the CLT (1.1) holds with $\eta = \mathbb{E}(D_0^2|\mathcal{I})$. Clearly, this condition is weaker than (1.6). Moreover one can

prove that if (1.8) holds then $\{S_{[nt]}, t \in [0,1]\}$ satisfies the weak invariance principle (see Hannan (1979) for the adapted case, *i.e.* when X_0 is $\mathcal{F}_0$-measurable, and Dedecker, Merlevède and Volný (2007) for the non-adapted case and extensions). In a recent paper, Cuny (2012a) has proved that the strong invariance principle also holds as well under (1.8) for random variables with values in a 2-smooth Banach space.

Another major improvment of Gordin's CLT is due to Maxwell and Woodrofe (2000). These authors have proved that, if the condition (1.7) is weakened to

$$\sum_{k=1}^{\infty} \frac{\|\mathbb{E}(S_k|\mathcal{F}_0)\|_2}{k^{3/2}} < \infty \quad \text{and} \quad \sum_{k=1}^{\infty} \frac{\|S_k - \mathbb{E}(S_k|\mathcal{F}_n)\|_2}{k^{3/2}} < \infty \tag{1.9}$$

then there exists a square integrable martingale difference D_0 such that $\|S_n - M_n\|_2 = o(\sqrt{n})$, so that that the CLT (1.1) holds with $\eta = \mathbb{E}(D_0^2|\mathcal{I})$ (in fact Maxwell and Woodroofe have proved the result in the adapted case and the non adapted case (1.9) is due to Volný (2007)). Peligrad and Utev (2005) proved that if (1.9) holds, then $\{S_{[nt]}, t \in [0,1]\}$ satisfies the weak invariance principle (again, the non adapted case is due to Volný (2007)).

Note that necessary and sufficient conditions for the martingale approximation $\|S_n - M_n\|_2 = o(\sqrt{n})$ are given in the papers by Dedecker, Merlevède and Volný (2007) and Zhao and Woodroofe (2008). Further refinements can be found in the paper by Gordin and Peligrad (2011). Finally, Wu (2007) and Dedecker, Doukhan and Merlevède (2012) have obtained rates of convergence in the strong invariance principle, by reinforcing the condition (1.8).

Concerning moments and exponential bounds, let us mention the paper by Peligrad, Utev and Wu (2007). These authors have proved that, for $p \in [2, \infty[$, there exists a positive constant K_p depending on p and a positive constant C such that

$$\frac{\|\max_{1 \le k \le n} |S_k|\|_p}{\sqrt{n}} \le K_p \Big(\|X_0\|_p + C \sum_{k=1}^{n} \frac{\|\mathbb{E}(S_k|\mathcal{F}_0)\|_p}{k^{3/2}} + C \sum_{k=1}^{n} \frac{\|S_k - \mathbb{E}(S_n|\mathcal{F}_k)\|_p}{k^{3/2}} \Big)$$

(the non adapted case is due to Volný (2007)), and that

$$\mathbb{P}\Big(\max_{1 \le k \le n} |S_k| > x \Big) \le 4\sqrt{e} \exp\Big(- \frac{x^2}{2n(\|X_0\|_\infty + C_n)^2} \Big),$$

where

$$C_n = D\Big(\sum_{k=1}^{n} \frac{\|\mathbb{E}(S_k|\mathcal{F}_0)\|_\infty}{k^{3/2}} + \sum_{k=1}^{n} \frac{\|S_k - \mathbb{E}(S_n|\mathcal{F}_k)\|_\infty}{k^{3/2}} \Big)$$

for some positive constant D (the non adapted case is due to Dede (2009)). Rosenthal type inequalities can be found in the paper by Merlevède and Peligrad (2012).

1.3. Organization of the paper

In Section 2, X. Fan presents an extension of Hoeffding's inequality to the case of supermartingales with a new method (see Fan, Grama and Liu (2012) for more details). The approach is based on the technic of conjugate distribution and is different from Hoeffding's method (1963). The results improve on several known inequalities of Bennett (1962), Freedman (1975), Nagaev (1979), Haeusler (1984) and Courbot (1999).

In Section 3, J.-R. Chazottes presents some recent concentration inequalities for a large class of nonuniformly hyperbolic dynamical systems modeled by Young towers (see Chazottes and Gouëzel (2013) for more details). In this context, he shows how to make use of some classical martingale inequalities, namely Azuma-Hoeffding and Rosenthal-Burkholder inequalities.

In Section 4, S. Lemler shows how an appropriate exponential bound for martingale with jumps enables to choose the weights in a Lasso procedure, for the Cox model in a high dimensional setting. As a consequence,

she obtains non asymptotic oracle inequalities for the conditional hazard rate function (see Lemler (2012) for more details).

In Section 5, C. Cuny presents the recent developments of the method of approximation by a martingale initiated by Gordin (1969), which has reached a fairly precise form thanks to the characterizations obtained recently by Zhao and Woodroofe (2008) and Gordin and Peligrad (2011). He also shows how the introduction of the operator $QZ = \mathbb{E}(Z \circ T|\mathcal{F}_0)$ enables to make use of classical tools from ergodic theory. Following this approach, C. Cuny has recently proved (see Cuny (2012a) and Cuny (2012b)) that the almost sure invariance principle holds under the condition (1.8) or under the condition (1.9).

2. EXPONENTIAL INEQUALITIES FOR MARTINGALES

Let $(\xi_i)_{i=1,\ldots,n}$ be a sequence of centered random variables such that $\sigma_i^2 = \mathbb{E}(\xi_i^2) < \infty$. Write $S_n = \sum_{i=1}^n \xi_i$ and $\sigma^2 = \sum_{i=1}^n \sigma_i^2$. Assume that $(\xi_i)_{i=1,\ldots,n}$ are independent and satisfy $|\xi_i| \leq 1$ for all i. Bennett (1962) proved that, for all $x > 0$,

$$\mathbb{P}(S_n \geq x) \quad \leq \quad B(x,\sigma^2) =: \left(\frac{\sigma^2}{x+\sigma^2}\right)^{x+\sigma^2} e^x. \tag{2.1}$$

Bennett's inequality can be improved by a bound depending on n. In fact, Hoeffding (1963) improved Bennett's inequality and showed that if $(\xi_i)_{i=1,\ldots,n}$ are independent and satisfy $\xi_i \leq 1$ for all i, then, for all $x > 0$,

$$\mathbb{P}(S_n \geq x) \quad \leq \quad H_n(x,\sigma^2) =: \left\{\left(\frac{\sigma^2}{x+\sigma^2}\right)^{x+\sigma^2} \left(\frac{n}{n-x}\right)^{n-x}\right\}^{\frac{n}{n+\sigma^2}} \mathbf{1}_{\{x \leq n\}} \tag{2.2}$$

$$\text{and} \quad H_n(x,\sigma^2) \quad \leq \quad B(x,\sigma^2), \tag{2.3}$$

where by convention $\infty^0 = 1$ when $x = n$. Hoeffding's bound $H_n(x,\sigma^2)$ is the best that can be obtained from the classical Bernstein inequality $\mathbb{P}(S_n \geq x) \leq \inf_{\lambda \geq 0} \mathbb{E}e^{\lambda(S_n-x)}$.

Freedman (1975) extended the inequality of Bennett to the case of supermartingales. Let $(\xi_i, \mathcal{F}_i)_{i=1,\ldots,n}$ be a sequence of supermartingale differences, i.e. $\mathbb{E}(\xi_i|\mathcal{F}_{i-1}) \leq 0$. Denote by $\langle S \rangle_k = \sum_{i=1}^k \mathbb{E}(\xi_i^2|\mathcal{F}_{i-1})$. Assume that supermartingale differences $(\xi_i, \mathcal{F}_i)_{i=1,\ldots,n}$ satisfy $\xi_i \leq 1$ for all i. Freedman's inequality states that, for all $x, v > 0$,

$$\mathbb{P}(S_k \geq x \text{ and } \langle S \rangle_k \leq v^2 \text{ for some } k \in [1,n]) \quad \leq \quad B(x,v^2). \tag{2.4}$$

However, one can not obtain the Hoeffding inequality by Freedman's method. In this note, we give the Hoeffding inequality for supermartingales with a new method. The main result of this section is the following theorem.

Theorem 2.1. *Assume that $(\xi_i, \mathcal{F}_i)_{i=1,\ldots,n}$ be a sequence of supermartingale differences satisfy $\xi_i \leq 1$ for all i. Then, for all $x, v > 0$,*

$$\mathbb{P}(S_k \geq x \text{ and } \langle S \rangle_k \leq v^2 \text{ for some } k \in [1,n]) \quad \leq \quad H_n(x,v^2) \tag{2.5}$$

$$\leq \quad B(x,v^2). \tag{2.6}$$

It is obvious that if $(\xi_i)_{i=1,\ldots,n}$ are independent then our inequality (2.5) implies Hoeffding's inequality (2.2) with $v^2 = \sigma^2$. Hence, we extend the Hoeffding inequality to the case of supermartingales. Note that (2.6) is the Freedman inequality. Thus we improve on Freedman's inequality.

Sketch of the proof of Theorem 2.1. For $0 \leq x \leq n, v > 0$, we define the stopping time $T(x) = \min\{k \in [1, n] :$ $S_k \geq x$ and $\langle S \rangle_k \leq v^2\}$. For $\lambda \geq 0$, we introduce the martingale $(Z_{T \wedge k}, \mathcal{F}_k)_{k=0,...,n}$, where

$$Z_{T \wedge k}(\lambda) = \prod_{i=1}^{T \wedge k} \frac{e^{\lambda \xi_i}}{\mathbb{E}(e^{\lambda \xi_i} | \mathcal{F}_{i-1})}, \qquad Z_0(\lambda) = 1, \tag{2.7}$$

and the *conjugate measure*

$$d\mathbb{P}_\lambda = Z_{T \wedge n}(\lambda) d\mathbb{P}. \tag{2.8}$$

Then $\mathbf{1}_{\{S_k \geq x \text{ and } \langle S \rangle_k \leq v^2 \text{ for some } k \in [1,n]\}} = \sum_{k=1}^{n} \mathbf{1}_{\{T(x)=k\}}$. Using the conjugate measure (2.8), we have

$$
\begin{aligned}
\mathbb{P}\left(S_k \geq x \text{ and } \langle S \rangle_k \leq v^2 \text{ for some } k \in [1, n]\right) &= \mathbb{E}_\lambda Z_{T \wedge n}(\lambda)^{-1} \mathbf{1}_{\{S_k \geq x \text{ and } \langle S \rangle_k \leq v^2 \text{ for some } k \in [1,n]\}} \\
&= \sum_{k=1}^{n} \mathbb{E}_\lambda \exp\left\{-\lambda S_k + \Psi_k(\lambda)\right\} \mathbf{1}_{\{T(x)=k\}},
\end{aligned}
\tag{2.9}
$$

where $\Psi_k(\lambda) = \sum_{i=1}^{k} \log \mathbb{E}(e^{\lambda \xi_i} | \mathcal{F}_{i-1})$. For $T = k$, we can prove that

$$\Psi_k(\lambda) \leq k \, f\left(\lambda, \frac{\langle S \rangle_k}{k}\right) \leq k \, f\left(\lambda, \frac{v^2}{k}\right) \leq n \, f\left(\lambda, \frac{v^2}{n}\right),$$

where $f(\lambda, t) = \log\left(\frac{1}{1+t}e^{-\lambda t} + \frac{t}{1+t}e^{\lambda}\right)$ is increasing and concave for all $t \geq 0$. Hence we get

$$\mathbb{P}\left(S_k \geq x \text{ and } \langle S \rangle_k \leq v^2 \text{ for some } k \in [1, n]\right) \leq \exp\left\{-\lambda x + n \, f\left(\lambda, \frac{v^2}{n}\right)\right\}. \tag{2.10}$$

Optimizing inequality (2.10) in $\lambda > 0$, we obtain the desired inequality (2.5). Inequality (2.6) follows from (2.3).

By a truncation argument, Theorem 2.1 implies the following result for non bounded supermartingale differences.

Corollary 2.2. *Assume that $(\xi_i, \mathcal{F}_i)_{i=1,...,n}$ be a sequence of supermartingale differences. For $y > 0$, define*

$$V_k^2(y) = \sum_{i=1}^{k} \mathbb{E}(\xi_i^2 \mathbf{1}_{\{\xi_i \leq y\}} | \mathcal{F}_{i-1}), \quad k = 1, ..., n.$$

Then, for all $x \geq 0$, $y > 0$ and $v > 0$,

$$\mathbb{P}(S_k \geq x \text{ and } V_k^2(y) \leq v^2 \text{ for some } k \in [1, n]) \leq H_n\left(\frac{x}{y}, \frac{v^2}{y^2}\right) + \mathbb{P}\left(\max_{1 \leq i \leq n} \xi_i > y\right). \tag{2.11}$$

Since $\mathbb{P}(V_n^2(y) > v^2) \leq \mathbb{P}(\langle S \rangle_n > v^2)$, inequality (2.11) improves the corresponding inequalities of Nagaev (1979), Haeusler (1984) and Courbot (1999) in the sense that $B\left(\frac{x}{y}, \frac{v^2}{y^2}\right)$ is replaced by $H_n\left(\frac{x}{y}, \frac{v^2}{y^2}\right)$.

3. MARTINGALE AND CONCENTRATION INEQUALITIES FOR DYNAMICAL SYSTEMS

3.1. Generalities

3.1.1. *Concentration inequalities*

Let Ω be a metric space. A function K on Ω^n is *separately Lipschitz* if, for all i, there exists a constant $\mathrm{Lip}_i(K)$ with

$$|K(x_0,\ldots,x_{i-1},x_i,x_{i+1},\ldots,x_{n-1}) - K(x_0,\ldots,x_{i-1},x_i',x_{i+1},\ldots,x_{n-1})| \le \mathrm{Lip}_i(K)d(x_i,x_i'),$$

for all points $x_1,\ldots,x_n,x_i'$ in Ω. Consider a stationary process $(Z_0,Z_1,\ldots)$ taking values in Ω. We say that this process satisfies an *exponential concentration inequality* if there exists a constant C such that, for any separately Lipschitz function $K(x_0,\ldots,x_{n-1})$, one has

$$\mathbb{E}\big[e^{K(Z_0,\ldots,Z_{n-1})-\mathbb{E}(K(Z_0,\ldots,Z_{n-1}))}\big] \le e^{C\sum_{j=0}^{n-1}\mathrm{Lip}_j(K)^2}. \tag{3.1}$$

Let us stress that this inequality is valid for all n, i.e. the constant C does not depend on the number of variables one is considering. An important consequence of such an inequality is a control on the deviation probabilities: for all $t > 0$,

$$\mathbb{P}\Big(|K(Z_0,\ldots,Z_{n-1}) - \mathbb{E}(K(Z_0,\ldots,Z_{n-1}))| > t\Big) \le 2e^{\frac{-t^2}{4C\sum_{j=0}^{n-1}\mathrm{Lip}_j(K)^2}}. \tag{3.2}$$

This inequality follows from the inequality $\mathbb{P}(Y > t) \le e^{-\lambda t}\mathbb{E}(e^{\lambda Y})$ ($\lambda > 0$) with $Y = K(Z_0,\ldots,Z_{n-1}) - \mathbb{E}(K(Z_0,\ldots,Z_{n-1}))$, then we use inequality (3.1) and optimize over λ by taking $\lambda = t/(2C\sum_{j=0}^{n-1}\mathrm{Lip}_j(K)^2)$.

If $(Z_0,Z_1,\ldots)$ is a sequence of bounded i.i.d. random variables, then (3.1) holds (see e.g. McDiarmid (1989)). One can check that, if $K(Z_0,\ldots,Z_{n-1}) = Z_0 + \cdots + Z_{n-1}$, then (3.2) gives the right scales with respect to the central limit theorem and large deviations.

In some cases, it is not reasonable to hope for such an exponential inequality. One says that $(Z_0,Z_1,\ldots)$ satisfies a *polynomial concentration inequality with moment $Q \ge 2$* if there exists a constant C such that, for any separately Lipschitz function $K(x_0,\ldots,x_{n-1})$, one has

$$\mathbb{E}|K(Z_0,\ldots,Z_{n-1}) - \mathbb{E}(K(Z_0,\ldots,Z_{n-1}))|^Q \le C\left(\sum_{j=0}^{n-1}\mathrm{Lip}_j(K)^2\right)^{Q/2}. \tag{3.3}$$

An important consequence of such an inequality is a control on the deviation probabilities: for all $t > 0$,

$$\mathbb{P}(|K(Z_0,\ldots,Z_{n-1}) - \mathbb{E}(K(Z_0,\ldots,Z_{n-1}))| > t) \le Ct^{-Q}\left(\sum_{j=0}^{n-1}\mathrm{Lip}_j(K)^2\right)^{Q/2}. \tag{3.4}$$

The inequality (3.4) readily follows from (3.3) and the Markov inequality.

Notice that if $(Z_0,Z_1,\ldots)$ is a sequence of i.i.d. random variables with $Z_i \in L^Q$, then (3.3) holds (see Boucheron *et al.* (2005)).

Concentration inequalities are a tool to study in a unified and systematic way the fluctuations of a wide class of functions of random variables $K(Z_0,\ldots,Z_{n-1})$, since the only required condition is that K is separately Lipschitz.

3.1.2. *Dynamical systems as stochastic processes*

We are interested in processes coming from dynamical systems: we consider a map T on a metric space Ω (the "phase space"), and a probability measure μ [1] left invariant by T, i.e. $\mu \circ T^{-1} = \mu$ [2]. The process $(x, Tx, T^2x, \dots)$, where x is distributed following μ, has finite-dimensional marginals given by the measures μ_n on Ω^n given by

$$\mathrm{d}\mu_n(x_0, \dots, x_{n-1}) = \mathrm{d}\mu(x_0)\delta_{x_1=Tx_0} \cdots \delta_{x_{n-1}=Tx_{n-2}}.$$

This is not a product measure but, if the map T is 'sufficiently mixing', one may expect that $T^k x$ is more or less independent of x if k is large, making the process $(x, Tx, \dots)$ look like an independent process to some extent [3]. A natural way of studying the probabilistic properties of such dynamical systems is to look at Birkhoff sums of an observable $f : \Omega \to \mathbb{R}$, namely $f(x) + f(Tx) + \cdots + f(T^n x)$, that is, partial sums of the process $(f(T^k x))_{k \geq 0}$. For a class of nice observables, typically Lipschitz functions, one can prove convergence in law after appropriate scaling, large deviations, etc. We refer the interesting reader to the recent survey by Chazottes (2013).
A general observable based on the observation up to time n is of the form $K(x, Tx, \dots, T^{n-1}x)$. The basic example is of course the Birkhoff sum of some observable f. But many interesting observables do not have such a simple (additive) structure. It is precisely the scope of concentration inequalities to deal with very general observables in a systematic way by using only the fact that they are separately Lipschitz.

3.2. Concentration inequalities for a class of nonuniformly hyperbolic systems

In Chazottes and Gouëzel (2013) we obtained concentration inequalities for dynamical systems modeled by the so-called Young towers. We also derived fluctuation bounds for various observables that we shall not present here due to the lack of space.

3.2.1. *Set-up*

In a nutshell, the set up is the following. We consider a map $T : \Omega \circlearrowleft$ which is a nonuniformly hyperbolic system in the sense of L.-S. Young (1998, 1999) : The map is modeled by a Young tower constructed over a hyperbolic base $Y \subset \Omega$. The degree of nonuniformity is measured by the return-time function $R : Y \to \mathbb{Z}^+$ to the base that decays either exponentially or polynomially. Such systems are known to have an SRB measure μ absolutely continuous with respect to the Lebesgue measure m^u.

3.2.2. *Exponential concentration inequality*

Theorem 3.1 (Chazottes and Gouëzel (2013))**.** *Let (Ω, T, μ) be a dynamical system modeled by a Young tower with exponential tails, i.e., $\int \exp(c_0 R)\, \mathrm{d}m^u < \infty$ for some $c_0 > 0$. Then it satisfies an exponential concentration inequality: there exists a constant $C > 0$ such that, for any $n \in \mathbb{N}$, for any separately Lipschitz function $K(x_0, \dots, x_{n-1})$,*

$$\int e^{K(x, Tx, \dots, T^{n-1}x) - \int K(y, Ty, \dots, T^{n-1}y)\, \mathrm{d}\mu(y)}\, \mathrm{d}\mu(x) \leq e^{C \sum_{\ell=0}^{n-1} \mathrm{Lip}_\ell(K)^2}. \tag{3.5}$$

Basic examples to which this theorem applies are subshifts of finite type equipped with a Gibbs measure for a Hölder continuous potential, and Axiom A attractors. These are canonical examples of uniformly hyperbolic dynamical systems. More sophisticated systems encompasses by this theorem are for instance the unimodal map on $\Omega = [-1, 1]$ defined as $x \mapsto 1 - ax^2$, where $a \in [1, 2]$, or Hénon-like attractors. These are examples of nonuniformly hyperbolic dynamical systems. The Hénon attractor results from the iterations of the map $(x, y) \mapsto (1 - ax^2 + y, bx)$, defined on $\Omega = \mathbb{R}^2$, where a, b are positive parameters. For each b sufficiently small, there exists a subset $P_b \subset (2 - \epsilon, 2)$ with positive Lebesgue measure such that for every $a \in P_b$, the map admits a unique SRB measure (see Benedicks and Young (2000)).

---

[1]This probability measure is of course defined on a σ-algebra. Think of the Borel σ-algebra for concreteness.

[2]i.e., $\mu(A) = \mu(T^{-1}A)$ for any measurable set A (T need not be invertible).

[3]Let us stress that the usual mixing coefficients for stochastic processes are generally not suited for our purposes.

3.2.3. *Polynomial concentration inequality*

Theorem 3.2 (Chazottes and Gouëzel (2013))**.** *Let* (Ω, T, μ) *be a dynamical system modeled by a Young tower with a polynomial tail, meaning that, for some* $q \geq 2$, $\int R^q \mathrm{d}m^u < \infty$. *Then it satisfies a polynomial concentration inequality with moment* $2q - 2$, *i.e., there exists a constant* $C > 0$ *such that, for any* $n \in \mathbb{N}$, *for any separately Lipschitz function* $K(x_0, \ldots, x_{n-1})$,

$$\int \left| K(x, Tx, \ldots, T^{n-1}x) - \int K(y, Ty, \ldots, T^{n-1}y) \mathrm{d}\mu(y) \right|^{2q-2} \mathrm{d}\mu(x) \leq C \left(\sum_{\ell=0}^{n-1} \mathrm{Lip}_\ell(K)^2 \right)^{q-1}. \qquad (3.6)$$

The most important example is certainly the so-called Manneville-Pomeau map, a canonical example of a map on $\Omega = [0, 1]$ which is expanding except at $x = 0$. More precisely, let

$$T(x) = \begin{cases} x(1 + 2^\alpha x^\alpha) & \text{if } x \in [0, 1/2) \\ 2x - 1 & \text{if } x \in (1/2, 1] \end{cases}$$

where $\alpha \in (0, 1)$. In this case, the base of the Young tower is $Y = (1/2, 1]$, $m(R = n) \sim c/n^{\frac{1}{\alpha}+1}$ and $\mathrm{d}\mu = h \, \mathrm{d}m$ where the density $h(x) \sim x^{-\alpha}$ as $x \to 0$. One checks easily that (3.6) holds for all $q < \frac{1}{\alpha}$ provided that $\alpha \in (0, 1/2)$.

These results are so far the culminating point in the study of concentration inequalities for dynamical systems. They extend and improve all previous results obtained in Collet, Martinez and Schmitt (2000), Chazottes, Collet and Schmitt (2005a), Chazottes, Collet and Schmitt (2005b) and Chazottes *et al.* (2009). There is still a lot of work to be done (see Chazottes (2013), Section 6).

3.2.4. *Examples of observables*

Let us barely mention some obervables to which we can apply the previous inequalities sucessfully. More details can be found in Collet, Martinez and Schmitt (2000), Chazottes, Collet and Schmitt (2005b) and Chazottes and Gouëzel (2013). For instance, we can study the speed at which the distance between the empirical measure and the SRB measure μ goes to 0. We can also look at the kernel density estimator for maps having absolutely continuous invariant measures. Among other observables we can deal with, let us mention the empirical covariance and the integrated periodogram. Lastly, concentration inequalities can be used to obtain an almost sure central limit theorem from the usual central limit theorem (see Chazottes, Collet and Schmitt (2005b)). This illustrates that even if concentration inequalities are mainly intended to obtain fluctuation *bounds*, they can also be used to get some limit theorems.

3.3. **Strategy of proofs**

Full details can be found in Chazottes and Gouëzel (2013). Our aim here is to give a rough sketch of proofs and highlight the use of martingale inequalities.
The starting point is that one can work in an auxiliary system, the Young tower, instead of the original system (and pull back later the obtained results). *For the sake of notational simplicity, we still denote by* (Ω, T, μ) *this auxiliary dynamical system.* If we start with an invertible map (*e.g.* the Hénon map), we can reduce to a non-invertible Young tower for which one can define the so-called *transfer operator* (see more details below).
Notice that one has to put an appropriate metric on the Young tower with respect to which Lipschitz functions are defined. The projection map going from the Young tower to the original dynamical system is contracting and project Lipschitz functions on the tower to Lipschitz functions on the original phase space.

3.3.1. *Martingale differences*

Fix a separately Lipschitz function $K(x_0, \ldots, x_{n-1})$. We consider it as a function on $\Omega^{\mathbb{N}}$ depending only on the first n coordinates (therefore, we set $\mathrm{Lip}_i(K) = 0$ for $i \geq n$). We endow $\Omega^{\mathbb{N}}$ with the measure μ_∞ limit of

the μ_N (see Chazottes and Gouëzel (2013)) when $N \to \infty$. On $\Omega^{\mathbb{N}}$, let $\mathcal{F}_p$ be the σ-algebra of events depending only on the coordinates $(x_j)_{j \geq p}$ (this is a decreasing sequence of σ-fields).

The very first step is classical: we want to write the function K as a sum of reverse martingale differences with respect to this sequence. Therefore, let $K_p = \mathbb{E}(K|\mathcal{F}_p)$. More precisely,

$$K_p(x_p, x_{p+1}, \ldots) = \mathbb{E}(K|\mathcal{F}_p)(x_p, x_{p+1}, \ldots) = \sum_{T^p(y)=x_p} g^{(p)}(y) K(y, \ldots, T^{p-1}y, x_p, \ldots)$$

where $g^{(p)}$ is the inverse of the jacobian of T^p.

Now let $D_p = K_p - K_{p+1}$. The function D_p is $\mathcal{F}_p$-measurable and $\mathbb{E}(D_p|\mathcal{F}_{p+1}) = 0$. Moreover,

$$K - \mathbb{E}(K) = \sum_{p \geq 0} D_p. \tag{3.7}$$

The basic strategy is to look for good estimates on D_p, then to apply a suitable martingale inequality. The hard part done in Chazottes and Gouëzel (2013) is to obtain good estimates on D_p.

3.3.2. *Exponential case*

The key estimate is the following: there exists $C > 0$ and $0 < \rho < 1$ such that, for any p, one has

$$|D_p| \leq C \sum_{j=0}^{p} \rho^{p-j} \operatorname{Lip}_j(K). \tag{3.8}$$

Next, Hoeffding-Azuma inequality (see e.g. Milman and Schechtman (1986), page 33, or Ledoux (2001)) yields

$$\mathbb{E}\left(e^{\sum_{p=0}^{P-1} D_p}\right) \leq e^{\sum_{p=0}^{P-1} \sup |D_p|^2}.$$

Using Cauchy-Schwarz inequality one easily gets $\sum_{p=0}^{P-1} \sup |D_p|^2 \leq C \sum_j \operatorname{Lip}_j(K)^2$. In view of (3.7), we obtain inequality (3.5).

Let us say a few words about (3.8). This estimate is a consequence of the existence of a spectral gap for the *transfer operator* $\mathcal{L}$ when it acts on a suitable Banach space. More precisely, one has $\mathcal{L}u(x) = \sum_{Ty=x} g(y)u(y)$ where g denotes the inverse of the jacobian of T. Let $\mathcal{C}$ be the space of Lipschitz functions on Ω with its canonical norm $\|f\|_{\mathcal{C}} = \sup |f| + \operatorname{Lip}(f)$. One can prove that for a Young tower with exponential tails there exist $C > 0$ and $0 < \rho < 1$ such that $\|\mathcal{L}^k f - \int f \, d\mu\|_{\mathcal{C}} \leq C\rho^k \|f\|_{\mathcal{C}}$. The point is to write $K_p(x_p, x_{p+1}, \ldots)$ as a sum of functions of one variable, by an appropriate telescoping procedure, and then to use the contraction properties of the transfer operator.

3.3.3. *Polynomial case*

For Young towers with polynomial tails, there is no spectral gap for the transfer operator, hence life becomes much more complicated. To control $\mathcal{L}^n f$ one has to rely on Banach algebra techniques to study some renewal sequences of operators entering the decomposition of $\mathcal{L}^n$.

We do not give further details and content ourselves by pointing that the useful martingale inequality to use is the following Rosenthal-Burkholder martingale inequality (see Burkholder (1973) Theorem 21.1 and Inequality (21.5)): for all $Q \geq 2$,

$$\left\|\sum_p D_p\right\|_{L^Q}^Q \leq C\mathbb{E}\left(\left[\sum_p \mathbb{E}(D_p^2|\mathcal{F}_{p+1})\right]^{Q/2}\right) + C \sum_p \mathbb{E}(|D_p|^Q).$$

4. Empirical Bernstein's inequality and applications to high-dimensional survival analysis

4.1. Framework and estimation procedure

In this section, we use an empirical Bernstein's inequality for martingales with jumps, to obtain non-asymptotic oracle inequalities for the conditional hazard rate function. These results should be used to obtain a prognostic on the survival time adjusted on the covariates in a high-dimensional setting. We consider a conditional hazard rate function that rely on a Cox's proportional hazard model. In this particular case, $\lambda_0(t, \boldsymbol{Z_i}) = \alpha_0(t) \exp(f_0(\boldsymbol{Z_i}))$, where $\boldsymbol{Z_i} = (Z_{i,1}, ..., Z_{i,p})^T$ is the vector of covariates of individual i ($i = 1, ..., n$), f_0 is the unknown regression function that we are going to estimate, and α_0 is the baseline hazard function (which is assumed to be known for the sake of simplicity, the general case being treated in Lemler (2012)). We aim at estimating this conditional hazard rate function by the best approximating Cox's model. Since we are in high-dimension, we estimate the unknown parameter f_0 using a weighted Lasso procedure that consists in minimizing the ℓ_1-penalized empirical likelihood. The key argument to prove our results is an empirical Bernstein's inequality for martingales with jumps, where the predictable variation, which is not observable, is replaced by the observable optional variation. This empirical Bernstein's inequality allows us to define a fully data-driven weighted ℓ_1-penalization for the Lasso procedure.

Let us first introduce the notations and the framework. For $i = 1, ..., n$, let N_i be a marked counting process and Y_i a predictable random process in $[0, 1]$. Let $(\mathcal{F}_t)_{t \geq 0}$ be the natural filtration defined by $\mathcal{F}_t = \sigma\{N_i(s), Y_i(s), 0 \leq s \leq t, Z_i, i = 1, ..., n\}$, and let $\Lambda_i(t)$ be the compensator of the process $N_i(t)$ with respect to $(\mathcal{F}_t)_{t \geq t}$, so that $M_i(t) = N_i(t) - \Lambda_i(t)$ is a martingale adapted to $(\mathcal{F}_t)_{t \geq 0}$.

Assumption 4.1. N_i satisfies the Aalen multiplicative intensity model : $\Lambda_i(t) = \int_0^t \lambda_0(s, \boldsymbol{Z_i}) Y_i(s) ds$, for all $t \geq 0$.

Let $\mathbb{F}_M = \{f_1, ..., f_M\}$ where $f_j : \mathbb{R}^p \to \mathbb{R}$ for $j = 1, ..., M$, be a finite set of functions, called a dictionary, where M is large (typically $M >> n$). For any function $f : \mathbb{R}^p \to \mathbb{R}$, let $||f||_{n,\infty} = \max_{1 \leq i \leq n} |f(\boldsymbol{Z_i})|$. We assume that, for all $1 \leq j \leq M$, $||f_j||_{n,\infty}$ is finite almost surely.

We assume that the unknown λ_0 can be well approximated by a function defined for all $\boldsymbol{\beta}$ in $\mathbb{R}^M$ by $\lambda_\beta(t, \boldsymbol{Z_i}) = \alpha_0(t) e^{f_\beta(\boldsymbol{Z_i})}$ where $f_\beta = \sum_{j=1}^M \beta_j f_j$. To estimate the unknown parameter, we consider the log-likelihood defined by

$$C_n(\lambda_\beta) = -\frac{1}{n} \sum_{i=1}^n \left\{ \int_0^\tau \log(\lambda_\beta(t, \boldsymbol{Z_i})) dN_i(t) - \int_0^\tau \lambda_\beta(t, \boldsymbol{Z_i}) Y_i(t) dt \right\}. \tag{4.1}$$

We associate to this estimation criterion, the empirical Kullback divergence defined for all $\boldsymbol{\beta}$ in $\mathbb{R}^M$ by

$$\widetilde{K}_n(\lambda_0, \lambda_\beta) = \frac{1}{n} \sum_{i=1}^n \int_0^\tau \left(\log(\lambda_0(t, \boldsymbol{Z_i})) - \log(\lambda_\beta(t, \boldsymbol{Z_i})) \right) \lambda_0(t, \boldsymbol{Z_i}) Y_i(t) dt$$

$$- \frac{1}{n} \sum_{i=1}^n \int_0^\tau \left(\lambda_0(t, \boldsymbol{Z_i}) - \lambda_\beta(t, \boldsymbol{Z_i}) \right) Y_i(t) dt. \tag{4.2}$$

We consider a weighted Lasso procedure for estimating $\boldsymbol{\beta}$. The Lasso estimator of $\boldsymbol{\beta}$ which minimizes the ℓ_1-penalized empirical likelihood is defined by

$$\hat{\boldsymbol{\beta}}_{\boldsymbol{L}} = \underset{\boldsymbol{\beta} \in \mathbb{R}^M}{\arg \min} \{C_n(\lambda_\beta) + \text{pen}(\boldsymbol{\beta})\}, \text{ with } \text{pen}(\boldsymbol{\beta}) = \sum_{j=1}^M \omega_j |\beta_j|. \tag{4.3}$$

The weights ω_j are positive data-driven weights suitably chosen thanks to the empirical Bernstein's inequality (4.7) (see Theorem 4.3).

4.2. Empirical Bernstein's inequality for martingales with jumps

By definition of the weighted Lasso estimator, we have for all $\boldsymbol{\beta} \in \mathbb{R}^M$

$$C_n(\lambda_{\hat{\boldsymbol{\beta}}_{\boldsymbol{L}}}) + \text{pen}(\hat{\boldsymbol{\beta}}_{\boldsymbol{L}}) \leq C_n(\lambda_{\boldsymbol{\beta}}) + \text{pen}(\boldsymbol{\beta}). \tag{4.4}$$

Using the Doob-Meier decomposition $N_i = M_i + \Lambda_i$, we can easily show that for all $\boldsymbol{\beta} \in \mathbb{R}^M$

$$C_n(\lambda_{\hat{\beta}_L}) - C_n(\lambda_\beta) = \widetilde{K}_n(\lambda_0, \lambda_{\hat{\beta}_L}) - \widetilde{K}_n(\lambda_0, \lambda_\beta) + \sum_{j=1}^M (\hat{\beta}_L - \beta)_j \eta_{n,\tau}(f_j), \tag{4.5}$$

$$\text{where} \quad \eta_{n,\tau}(f_j) = \frac{1}{n}\sum_{i=1}^n \int_0^\tau f_j(\boldsymbol{Z_i})dM_i(t).$$

To obtain non-asymptotic oracle inequalities, we have to control the centered empirical process $\eta_{n,\tau}(f_j)$. Towards that end, we define the predictable variation of $\eta_{n,t}(f_j)$ by

$$V_{n,t}(f_j) = n < \eta_n(f_j) >_t = \frac{1}{n}\sum_{i=1}^n \int_0^t (f_j(\boldsymbol{Z_i}))^2 \lambda_0(t, \boldsymbol{Z_i}) Y_i(s)ds.$$

To control such a process, we usually consider a standard Bernstein's inequality for martingales of the form of (4.6).

Theorem 4.2 (Shorack and Wellner (1986)). *Let $\{M_t\}_{t\geq 0}$ be a locally square integrable martingale with respect to $\{\mathcal{F}_t\}$, $V_t =< M >_t$ and $\Delta M_t = M_t - M_{t^-}$. Suppose that $|\Delta M_t| \leq K$ for all $t > 0$ and some $0 \leq K < \infty$. Then for each $a > 0$, $b > 0$,*

$$\mathbb{P}(M_t \geq a \text{ and } V_t \leq b^2 \text{ for some } t) \leq \exp\left[-\frac{a^2}{2(aK + b^2)}\right]. \tag{4.6}$$

In our case, the predictable variation $V_{n,t}(f_j)$ is not observable (λ_0 is unknown). To override this problem, we replace in the standard Bernstein's inequality, the predictable variation by the observable optional variation of $\eta_{n,t}(f_j)$ defined by

$$\hat{V}_{n,t}(f_j) = n[\eta_n(f_j)]_t = \frac{1}{n}\sum_{i=1}^n \int_0^t (f_j(\boldsymbol{Z_i}))^2 dN_i(s),$$

and which can be seen as an estimator of $V_{n,t}(f_j)$. This procedure has already been considered by Gaïffas and Guilloux (2012). See also Hansen, Reynaud-Bouret and Rivoirard (2012).

Theorem 4.3. *For any $x > 0$ and c_1, c_2, c_3 some positive constants, we have*

$$\mathbb{P}\left[|\eta_{n,t}(f_j)| \geq \left(c_1\sqrt{\frac{x + \hat{\ell}_{n,x}(f_j)}{n}\hat{V}_{n,t}(f_j)} + c_2\frac{x + 1 + \hat{\ell}_{n,x}(f_j)}{n}\right)||f_j||_{n,\infty}\right] \leq c_3 e^{-x}, \tag{4.7}$$

$$\text{where} \quad \hat{\ell}_{n,x}(f_j) = 2\log\log\left(\frac{6en\hat{V}_{n,t}(f_j) + 56ex||f_j||_{n,\infty}^2}{24||f_j||_{n,\infty}^2} \vee e\right).$$

Choice of the weights : We choose the data-driven weights for $j = 1, ..., M$ as

$$\omega_j = \left(c_1\sqrt{\frac{x + \log M + \hat{\ell}_{n,x}(f_j)}{n}\hat{V}_{n,\tau}(f_j)} + c_2\frac{x + 1 + \log M + \hat{\ell}_{n,x}(f_j)}{n}\right)||f_j||_{n,\infty}. \tag{4.8}$$

4.3. Oracle inequalities for the Lasso for the conditional hazard rate

For this choice of weights, we introduce the following set $\mathcal{A} = \bigcap\limits_{j=1}^{M} \left\{ |\eta_{n,\tau}(f_j)| \leq \omega_j \right\}$. On $\mathcal{A}$,

$$\left| \sum_{j=1}^{M} (\hat{\beta}_L - \beta)_j \eta_{n,\tau}(f_j) \right| \leq \sum_{j=1}^{M} \omega_j |(\hat{\beta}_L - \beta)_j| \text{ and } \mathbb{P}(\mathcal{A}^c) \leq \sum_{j=1}^{M} \mathbb{P}(|\eta_{n,\tau}(f_j)| > \omega_j) \leq c_3 e^{-x}.$$

From (4.4) and (4.5), we deduce that

$$\widetilde{K}_n(\lambda_0, \lambda_{\hat{\beta}_L}) \leq \widetilde{K}_n(\lambda_0, \lambda_\beta) + \sum_{j=1}^{M} (\hat{\beta}_L - \beta)_j \eta_{n,\tau}(f_j) + \sum_{j=1}^{M} \omega_j |\beta|_j - \sum_{j=1}^{M} \omega_j |\hat{\beta}_L|_j. \qquad (4.9)$$

We finally obtain the following non-asymptotic oracle inequality with a slow rate of convergence of order $\sqrt{\log M/n}$.

Theorem 4.4. *Let $A > 0$ be some numerical positive constant and $x > 0$ be fixed. Then, with probability larger than $1 - Ae^{-x}$*

$$\widetilde{K}_n(\lambda_0, \lambda_{\hat{\beta}_L}) \leq \inf_{\beta \in \mathbb{R}^M} \{\widetilde{K}_n(\lambda_0, \lambda_\beta) + 2\operatorname{pen}(\beta)\}, \qquad (4.10)$$

with $\operatorname{pen}(\beta)$ *defined by (4.3) and (4.8).*

Under the classical restricted eigenvalue condition **RE** and other assumptions, we can also obtain some non-asymptotic oracle inequalities with a fast rate of convergence of order $\log M/n$ and some results in variable selection (see Lemler (2012) for more details).

Sketch of the proof of Theorem 4.3. We only detail the steps to prove Equation (4.7). Let us denote by $U_{n,t}$ and $H_i(f_j)$ the quantities

$$U_{n,t}(f_j) = \frac{1}{n} \sum_{i=1}^{n} \int_0^t H_i(f_j) dM_i(s) \quad \text{and} \quad H_i(f_j) := \frac{f_j(\boldsymbol{Z_i})}{\max\limits_{1 \leq i \leq n} |f_j(\boldsymbol{Z_i})|}.$$

Since $H_i(f_j)$ is a bounded predictable process with respect to $\mathcal{F}_t$, $U_{n,t}(f_j)$ is a square integrable martingale. Its predictable variation and its optional variation are respectively given by

$$\vartheta_{n,t}(f_j) = n < U_n, (f_j) >_t = \frac{1}{n} \sum_{i=1}^{n} \int_0^t (f_j(\boldsymbol{Z_i}))^2 d\Lambda_i(s) \quad \text{and} \quad \hat{\vartheta}_{n,t}(f_j) = n[U_n(f_j)]_t = \frac{1}{n} \sum_{i=1}^{n} \int_0^t (f_j(\boldsymbol{Z_i}))^2 dN_i(s).$$

The proof relies on the three following steps :

Step1 : We prove first that

$$\mathbb{P}\left[U_{n,t}(f_j) \geq \sqrt{\frac{2\omega \vartheta_{n,t}(f_j)x}{vn}} + \frac{x}{3n}, v < \vartheta_{n,t}(f_j) \leq \omega \right] \leq e^{-x}.$$

Step 2 : Step 2 consists in replacing $\vartheta_{n,t}(f_j)$ by the observable $\hat{\vartheta}_{n,t}(f_j)$ in Step 1. It follows that

$$\mathbb{P}\left[U_{n,t}(f_j) \geq 2\sqrt{\frac{\omega x}{vn} \hat{\vartheta}_{n,t}(f_j)} + \left(2\sqrt{\frac{\omega}{v}\left(\frac{\omega}{v} + \frac{1}{3}\right)} + \frac{1}{3}\right)\frac{x}{n}, v \leq \vartheta_{n,t}(f_j) < \omega \right] \leq 3e^{-x}. \qquad (4.11)$$

Step 3 : Finally, Step 3 is devoting to remove the event $\{v \leq \vartheta_{n,t}(f_j) < \omega\}$ from Inequality (4.11). $\qquad \square$

5. Approximation by a martingale

Since its introduction by Gordin (1969), the martingale method has attracted number of probabilists. It was originally designed for the Central Limit Theorem (CLT) but it has been successfully used for almost any limit theorem (the Weak Invariance Principle (WIP) and its quenched version, the Law of the Iterated Logarithm and its functional versions, the Marcinkiewicz-Zygmund strong law of large numbers...)

When applied to the CLT or WIP problem, the martingale method has reached a fairly precise form thanks to the characterizations obtained recently by Zhao and Woodroofe (2008) (CLT case) and Gordin and Peligrad (2011) (CLT and WIP cases). Their characterization, of theoretical interest, proved to be useful in applications as well, especially concerning the characterization by Gordin and Peligrad (2011).

The martingale method applies to "different" situations : stationary processes (adapted or not), non-invertible dynamical systems, functionals of Markov chain. In this note, for the sake of clarity, we shall only be concerned with *adapted* stationary processes. As we recall below, this case is actually equivalent to considering functionals of Markov chains. The case of non-adapted processes may be treated similarly (see for instance Volný (2007) or Cuny (2012a)). The adaptation of the results mentioned below to the setting of non-invertible dynamical systems needs more care since what we really obtain in that case is a "reverse" martingale approximation. We shall consider only real-valued processes, but some results extend to Hilbert space-valued processes.

We use the same notations as in the introduction. We want to study the process $(X_n = X_0 \circ T^n)_{n \in \mathbb{Z}}$, which is adapted to the non-decreasing filtration $(\mathcal{F}_n = T^{-n}(\mathcal{F}_0))_{n \in \mathbb{Z}}$, i.e. X_0 is $\mathcal{F}_0$-measurable. We assume that $X_0 \in L^p(\Omega, \mathcal{F}_0, \mathbb{P})$, for some $p \geq 1$. For simplicity, we assume T to be ergodic.

Let us define an operator Q on $L^1(\Omega, \mathcal{F}_0, \mathbb{P})$ by $QZ = \mathbb{E}(Z \circ T | \mathcal{F}_0)$. The operator Q is a positive contraction of every $L^r(\Omega, \mathcal{F}_0, \mathbb{P})$, $r \geq 1$, hence it is a Markov operator. It turns out that this operator allows to see our process $(X_n)_{n \in \mathbb{Z}}$ as a functional of a Markov chain, see Cuny and Volný (2012).

Another advantage of this operator is that it allows to translate projective conditions in terms of Q, hence to make use of classical facts from ergodic theory of operators. For instance, the martingale-coboundary decomposition of Gordin is easily characterized as follows, which has been observed by Volný (1993) (the non-adapted case is also considered there). Volný worked under the regularity condition $\mathbb{E}(X_0 | \mathcal{F}_{-\infty}) = 0$ a.s. but it is actually not needed.

Proposition 5.1. *Let $p \geq 1$ and let $X_0 \in L^p(\Omega, \mathcal{F}_0, \mathbb{P})$. The following are equivalent*

(i) $X_0 = D_0 + Z - Z \circ T^{-1}$ *with* $D_0, Z \in L^p(\Omega, \mathcal{F}_0, \mathbb{P})$ *and* $\mathbb{E}(D_0 | \mathcal{F}_{-1}) = 0$ *almost surely ;*

(ii) $\sup_{n \geq 1} \|\mathbb{E}(S_n | \mathcal{F}_0)\|_p < \infty$.

Proof. $(i) \Rightarrow (ii)$ is obvious. Now, (ii) reads: $\sup_{n \geq 1} \|QX_0 + \cdots + Q^n X_0\|_p < \infty$. Hence, by a result of Browder (1958, Lemma 5) when $1 < p < \infty$, and by Theorem 7 of Lin and Sine (1983) when $p = 1$, there exists $Y \in L^p(\Omega, \mathcal{F}_0, \mathbb{P})$ such that $X_0 = (I - Q)Y$, and (i) follows by taking $D_0 = Y - \mathbb{E}(Y | \mathcal{F}_{-1}) = Y - (QY) \circ T^{-1}$ and $Z = -QY$. $\qquad\square$

Our goal now will be to explain that, in several situations, to prove a limit theorem by mean of a martingale approximation it is enough to prove it when we have the martingale-coboundary decomposition above, provided that we have a maximal inequality of a certain type (adapted to the limit theorem under consideration). This approach has been used explicitly in Cuny (2012a) and Cuny (2012b) and implicitly, for instance, in Jiang and Wu (2003), Cuny and Volný (2012) or Cuny and Merlevède (2012).

Assume from now on that $X_0 \in L^2(\Omega, \mathcal{F}_0, \mathbb{P})$. We say that X_0 or $(X_n)_{n \in \mathbb{Z}}$ admits a martingale approximation of type (CLT), (WIP) or (ASIP) if there exists $D_0 \in L^2(\Omega, \mathcal{F}_0, \mathbb{P})$ with $\mathbb{E}(D_0 | \mathcal{F}_{-1}) = 0$ a.s. such that, writing

$M_n = D_0 \circ T + \cdots + D_0 \circ T^n,$

$$\mathbb{E}((S_n - M_n)^2) = o(n) \qquad\qquad\qquad \text{(CLT)}$$

$$\mathbb{E}(\max_{1 \leq k \leq n}(S_k - M_k)^2) = o(n) \qquad\qquad \text{(WIP)}$$

$$|S_n - M_n| = o(\sqrt{n \log \log n}) \qquad \mathbb{P}\text{-a.s.} \qquad \text{(ASIP)}.$$

The notation (ASIP) stands for the almost sure invariance principle. The above martingale approximations are unique. If X_0 admits a martingale approximation of one of the above types then it satisfies the corresponding limit theorem.

An important fact is that the set of $X_0 \in L^2(\Omega, \mathcal{F}_0, \mathbb{P})$ admitting a martingale approximation (of any of the above type) is a vector space, containing $(I - Q)L^2(\Omega, \mathcal{F}_0, \mathbb{P})$, hence stable by Q.

We would like to study the validity of some limit theorems under the Hannan condition (1.8) and/or the Maxwell-Woodroofe condition (1.9) mentioned in the introduction. In terms of the operator Q those conditions read

$$\|X_0\|_{H_2} := \sum_{n \geq 0} \|Q^n X_0 - (Q^{n+1}X_0) \circ T^{-1}\|_2 < \infty \,; \qquad (5.1)$$

$$\|X_0\|_{MW_2} := \|X_0\|_2 + \sum_{n \geq 1} \frac{\|QX_0 + \cdots + Q^n X_0\|_2}{n^{3/2}} < \infty \,. \qquad (5.2)$$

Let us consider the spaces $H_2 := \{X_0 \in L^2(\Omega, \mathcal{F}_0, \mathbb{P}) \; : \; \mathbb{E}_{-\infty}(X_0) = 0 \quad \text{and} \quad \|X_0\|_{H_2} < \infty\}$ and $MW_2 := \{X_0 \in L^2(\Omega, \mathcal{F}_0, \mathbb{P}) \; : \; \|X_0\|_{MW_2} < \infty\}$. It is not hard to prove that those spaces are Banach spaces and that Q induces a contraction of H_2 and of MW_2.

Moreover, for every $X_0 \in H_2$, $\|Q^n X_0\|_{H_2} \xrightarrow[n \to +\infty]{} 0$. With little effort one can also prove that, for every $X_0 \in MW_2$, $\|QX_0 + \cdots + Q^n X_0\|_{MW_2}/n \xrightarrow[n \to +\infty]{} 0$.

By the mean ergodic theorem (see e.g. Theorems 1.2 and 1.3 p. 73 of Krengel (1985)), we have, noticing that Q has no fixed points neither on H_2 nor on MW_2,

$$H_2 = \overline{(I-Q)H_2}^{H_2} \qquad \text{and} \qquad MW_2 = \overline{(I-Q)MW_2}^{MW_2} \qquad (5.3)$$

5.1. The approximating martingale.

We first have to find D_0. According to Gordin and Peligrad (2001), if there exists a martingale approximation of type (CLT), necessarily

$$D_0 = \lim_{n \to +\infty} \frac{1}{n} \sum_{k=1}^{n} \sum_{\ell=0}^{k-1} (Q^k X_0 - (Q^{k+1}X_0) \circ T^{-1}) \,, \qquad (5.4)$$

where the limit holds in $L^2(\Omega, \mathbb{P})$. Set $\mathcal{D}(X_0) := D_0$, whenever the above limit exists. Then, $\mathcal{D}$ is an *unbounded* operator on $L^2(\Omega, \mathbb{P})$.

As one may expect (from the proof of Proposition 5.1), for every $Y \in L^2(\Omega, \mathcal{F}_0, \mathbb{P})$, $\mathcal{D}((I-Q)Y) = Y - (QY) \circ T^{-1}$.

The operator $\mathcal{D}$ is well-defined (i.e. bounded) on H_2 (this is easily verified) and on MW_2, by the results of Gordin and Peligrad (2011). For some problems, it may be useful to have an other form of $\mathcal{D}$, more adapted to the conditions (5.1) or (5.2).

For every $X_0 \in H_2$ we have, with convergence in $L^2(\Omega, \mathbb{P})$ (and $\mathbb{P}$-a.s.),

$$\mathcal{D}X_0 = \sum_{n \geq 0} \left(Q^n X_0 - (Q^{n+1} X_0) \circ T^{-1} \right) \qquad \text{and} \qquad \|\mathcal{D}X_0\|_2 \leq \|X_0\|_{H_2} . \tag{5.5}$$

For every $X_0 \in MW_2$, we have (see Cuny and Merlevède (2012)), with convergence in $L^2(\Omega, \mathbb{P})$ (and $\mathbb{P}$-a.s.)

$$\mathcal{D}X_0 = \sum_{n \geq 0} \sum_{k \geq n} \frac{Q^k X_0 - (Q^{k+1} X_0) \circ T^{-1}}{k+1} \qquad \text{and} \qquad \|\mathcal{D}X_0\|_2 \leq C\|X_0\|_{MW_2}, \tag{5.6}$$

for a universal constant $C > 0$. It is not hard to see that the inner sum above converges as soon as $X_0 \in L^2(\Omega, \mathbb{P})$. One can prove that if the representation (5.5) holds, then (5.6) holds as well, and that (5.6) implies (5.4).

5.2. Some maximal inequalities.

Before proving the martingale approximation properties, we shall need the corresponding estimate with a big "O" instead of a little "o". For this we need some maximal inequalities. We start with the martingale case. We have

Proposition 5.2. *Let* $D_0 \in L^2(\Omega, \mathcal{F}_0, \mathbb{P})$ *with* $\mathbb{E}(D_0|\mathcal{F}_{-1}) = 0$ $\mathbb{P}$*-a.s. Write* $M_n := \sum_{k=1}^{n} D_0 \circ T^k$. *We have*

$$\sup_{n \geq 1} \frac{\| \max_{1 \leq k \leq n} |M_k| \|_2}{\sqrt{n}} \leq 2\|D_0\|_2 ;$$

$$\left\| \sup_{n \geq 1} \frac{|M_n|}{\sqrt{nL(L(n))}} \right\|_1 \leq C\|D_0\|_2 , \tag{5.7}$$

where $L(n) = \max(1, \log n)$ *and* $C > 0$ *is a universal constant.*

The first estimate is nothing but Doob's maximal inequality. The second one has been proven in Cuny (2012a) (the maximal function actually lies in any $L^p(\Omega, \mathbb{P})$, $1 \leq p < 2$). Both inequalities hold without ergodicity. The maximal inequality (5.7) seems to be new in the martingale setting. In the iid case, it may be found for instance in the paper by Ledoux and Talagrand (1988). To emphasize the usefulness of such inequality we mention, that thanks to (5.7), in order to prove the law of the iterated logarithm for martingales with stationary ergodic increments in L^2, it suffices to prove it for martingales with *bounded* increments (this follows from a Banach principle argument).

In the next proposition, $\mathcal{X}$ stands either for H_2 or for MW_2. For $X_0 \in \mathcal{X}$, we denote $D_0 = \mathcal{D}X_0$ and $M_n = \sum_{k=1}^{n} D_0 \circ T^k$.

Proposition 5.3. *Assume that* $X_0 \in \mathcal{X}$. *We have*

$$\sup_{n \geq 1} \frac{\| \max_{1 \leq k \leq n} |S_k - M_k| \|_2}{\sqrt{n}} \leq C\|X_0\|_{\mathcal{X}} ; \tag{5.8}$$

$$\left\| \sup_{n \geq 1} \frac{|S_n - M_n|}{\sqrt{nL(L(n))}} \right\|_1 \leq C\|X_0\|_{\mathcal{X}} , \tag{5.9}$$

where $C > 0$ *is a universal constant.*

The estimate (5.9) is proved in Cuny (2012b), under the Maxwell-Woodroofe condition. For the other estimates, at first, notice that to prove (5.8) or (5.9), it suffices to treat separately $(S_n)_{n\geq 1}$ and $(M_n)_{n\geq 1}$. For the martingale part both estimates follow from Proposition 5.2 combined with (5.5) or (5.6). For $(S_n)_{n\geq 1}$ itself, (5.8) follows from Theorem 1 (iii) of Wu (2007) under the Hannan condition and from Proposition 2.3 of Peligrad and Utev (2005) under the Maxwell-Woodroofe condition. The estimate (5.9) for $(S_n)_{n\geq 1}$ is proved in Cuny (2012a) under the Hannan condition.

5.3. The conclusion.

It follows from (5.8) and a Banach principle argument, that the set of $X_0 \in \mathcal{X}$, for which (WIP) holds, is closed in $\mathcal{X}$. But (WIP) holds on $(I - Q)\mathcal{X}$, hence, by (5.3), it holds on $\mathcal{X}$ too.

Similarly, by (5.9), the set of $X_0 \in \mathcal{X}$, for which (ASIP) holds, is closed in $\mathcal{X}$ and we conclude as above.

References

[1] Azuma, K. *Weighted sums of certain dependent random variables*, Thoku Math. J. 19 (1967) 357-367.

[2] Benedicks, M. and Young L.-S. *Markov extensions and decay of correlations for certain Hénon maps*, Géométrie complexe et systèmes dynamiques (Orsay, 1995). Astérisque 261 (2000) 13–56.

[3] Bennett, G. *Probability inequalities for the sum of independent random variables*, J. Amer. Statist. Assoc. 57 (1962) 33–45.

[4] Billingsley, P. *The Lindeberg-Lévy theorem for martingales*, Proc. Amer. Math. Soc. 12 (1961) 788-792.

[5] Boucheron, S., Bousquet, O., Lugosi, G. and Massart, P. *Moment inequalities for functions of independent random variables*, Ann. Probab. 33 (2005) 514–560.

[6] Browder, F. E. *On the iteration of transformations in noncompact minimal dynamical systems*, Proc. Amer. Math. Soc. 9 (1958) 773-780.

[7] Brown, B. M. *Martingale central limit theorems*, Ann. Math. Statist. 42 (1966) 5966.

[8] Burkholder, D. L. *Martingale transforms*, Ann. Math. Statist. 37 (1971) 1494-1504.

[9] Burkholder, D. L. *Distribution function inequalities for martingales*, Ann. Probab. 1 (1973) 19-42.

[10] Burkholder, D. L., Davis, B. J. and Gundy, R. F. *Integral inequalities for convex functions of operators on martingales. Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability* (Univ. California, Berkeley, Calif., 1970/1971), Vol. II: Probability theory, pp. 223240. Univ. California Press, Berkeley, Calif., (1972).

[11] Chazottes, J.-R. *Fluctuations of observables in dynamical systems: from limit theorems to concentration inequalities*, to appear in a volume in honor of V. Afraimovich (2013), http://arxiv.org/abs/1201.3833.

[12] Chazottes, J.-R., Collet, P., Redig, F. and Verbitskiy E. *A concentration inequality for interval maps with an indifferent fixed point*, Ergod. Th. Dynam. Systems 29 (2009) 1097–1117.

[13] Chazottes, J.-R. and Gouëzel, S. *Optimal concentration inequalities for dynamical systems*, to appear in Commun. Math. Phys. (2013), 47 pages, http://arxiv.org/abs/1111.0849.

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

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

[16] Collet, P. *Variance and exponential estimates via coupling*, Bull. Braz. Math. Soc. (N.S.) 37 (2006) 461–475.

[17] Collet, P., Martinez, S. and Schmitt, B. *Exponential inequalities for dynamical measures of expanding maps of the interval*, Probab. Theory Related Fields 123 (2002) 301-322.

[18] Courbot, B. *Rates of convergence in the functional CLT for martingales*, C. R. Acad. Sci. Paris 328 (1999) 509–513.

[19] Cuny, C. *ASIP for martingales in 2-smooth Banach spaces. Application to stationary sequences*, (2012a) http://hal.archives-ouvertes.fr/hal-00745651

[20] Cuny, C. *An almost sure invariance principle under the Maxwell-Woodroofe condition*, in progress. (2012b).

[21] Cuny, C. and Merlevède, F. *On martingale approximations and the quenched weak invariance principle*, (2012) http://arxiv.org/abs/1202.2964

[22] Cuny, C and Volný, D. *A quenched invariance principle for stationary processes*, (2012) http://arxiv.org/abs/1202.4875

[23] Dede, S. *Moderate deviations for stationary sequences of Hilbert-valued bounded random variables*, J. Math. Anal. Appl. 349 (2009) 374-394.

[24] Dedecker, J., Doukhan, P. and Merlevède, F. *Rates of convergence in the strong invariance principle under projective criteria*, Electron. J. Probab. 17 (2012) 1–31.

[25] Dedecker, J., Merlevède, F. and Volný, D. *On the weak invariance principle for non adapted sequences under projective criteria*, J. Theoret. Probab. 20 (2007) 971–1004.

[26] Fan, X., Grama, I. and Liu, Q. *Hoeffding's inequality for supermartingales*, Stochastic Process. Appl. 122 (2012) 3545–3559.

[27] Freedman, D. A. *On tail probabilities for martingales*, Ann. Probability 3 (1975) 100-118.

[28] Freund, Y., Mansour, Y. and Schapire, R. E. *Generalization bounds for averaged classifiers*, Ann. Statist. 32 (2004) 1698–1722.

[29] Gaïffas, S. and Guilloux, A. *High-dimensional additive hazard models and the Lasso*. Electron. J. Stat. 6 (2012) 522-546.

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

[31] Gordin, M. I. and Lifsic, B. A. *Central limit theorem for stationary Markov processes*, (Russian) Dokl. Akad. Nauk SSSR 239 (1978) 766-767.

[32] Gordin, M. I. and Peligrad, M. *On the functional central limit theorem via martingale approximation*, Bernoulli 17 (2011) 424–440.

[33] Haeusler, E. *An exact rate of convergence in the functional central limit theorem for special martingale difference arrays*, Probab. Theory Relat. Fields 65 (1984) 523–534.

[34] Hannan, E. J. *Central limit theorems for time series regression*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete 26 (1973) 157-170.

[35] Hannan E. J. *The central limit theorem for time series regression*, Stochastic Process. Appl. 9 (1979) 281-289.

[36] Hansen, N. R., Reynaud-Bouret, P. and Rivoirard, V. *Lasso and probabilistic inequalities for multivariate point processes*, Work in progress, personnal communication, (2012).

[37] Heyde, C. C. *On the central limit theorem for stationary processes*, Z. Wahrscheinlichkietstheorie und Verw. Gebiete 30 (1974), 315-320.

[38] Heyde, C. C. *On the central limit theorem and iterated logarithm law for stationary processes*, Bull. Austral. Math. Soc. 12 (1975) 1-8.

[39] Hoeffding, W. *Probability inequalities for sums of bounded random variables*, J. Amer. Statist. Assoc. 58 (1963) 13-30.

[40] Le Borgne S. *Limit theorems for non-hyperbolic automorphisms of the torus*, Israel J. Math. 109 (1999) 61–73.

[41] Ibragimov, I. A. *A central limit theorem for a class of dependent random variables*, Theory Probab. Appl. 8 (1963) 83-89.

[42] Jiang, Y. and Wu, L. *Hilbertian invariance principle for empirical process associated with a Markov process*, Chinese Ann. Math. Ser. B 24 (2003) 1-16.

[43] Krengel, U. Ergodic theorems, de Gruyter Studies in Mathematics, 6. Walter de Gruyter & Co., Berlin, (1985).

[44] Ledoux, M. *The concentration of measure phenomenon*, Mathematical Surveys and Monographs 89. American Mathematical Society, (2001). Second printing (2005).

[45] Ledoux, M. and Talagrand, M. *Characterization of the law of the iterated logarithm in Banach spaces*, Ann. Probab. 16 (1988) 1242–1264.

[46] Lemler, S. *Oracle inequalities for the Lasso for the conditional hazard rate in a high-dimensional setting*, (2012) `http://arxiv.org/abs/1206.5628`

[47] Lévy, P. *Propriétés asymptotiques des sommes de variables aléatoires indépendantes ou enchaînées*, J. Math. Pures Appl. Ser. 8 (1935) 347-402.

[48] Lin, M. and Sine, R. *Ergodic theory and the functional equation* $(IT)x = y$. J. Operator Theory 10 (1983) 153166.

[49] Liu, Q. and Watbled, F. *Exponential inequalities for martingales and asymptotic properties of the free energy of directed polymers in a random environment*, Stochastic Process. Appl. 119 (2009) 3101-3132.

[50] Liverani, C. *Central limit theorem for deterministic systems*, International Conference on Dynamical Systems, 56-75, Pitman Res. Notes Math. Ser., 362, Longman, Harlow, (1996).

[51] Maxwell, M. and Woodroofe, M. *Central limit theorems for additive functionals of Markov chains*, Ann. Probab. 28 (2000) 713-724.

[52] McDiarmid, C. *On the method of bounded differences*, Surveys in combinatorics, 148-188, London Math. Soc. Lecture Note Ser., 141, Cambridge Univ. Press, Cambridge, (1989).

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

[54] Nagaev, S. V. *Large deviations of sums of independent random variabels*, Ann. Probab. 7 (1979) No. 5, 745–789.

[55] Merlevède, F. and Peligrad, M. *Rosenthal inequalities for martingales and stationary sequences and examples*, to appear in Ann. Probab. (2012) `http://arxiv.org/abs/1103.3242`

[56] Milman, V. and Schechtman, G. *Asymptotic theory of finite-dimensional normed spaces*, Lect. Notes in Math. 1200 Springer (1986).

[57] Peligrad, M. and Utev, S. *A new maximal inequality and invariance principle for stationary sequences*, Ann. Probab. 33 (2005) 798-815.

[58] Peligrad, M., Utev, S. and Wu, W. B. *A maximal $\mathbb{L}^p$-inequality for stationary sequences and its applications*, Proc. Amer. Math. Soc. 135 (2007) 541-550

[59] Pinelis, I. *An approach to inequalities for the distributions of infinite-dimensional martingales. Probability in Banach spaces*, 8, 128-134, Progr. Probab., 30, Birkhäuser Boston, Boston, MA, (1992).

[60] Pinelis, I. *Optimum bounds for the distributions of martingales in Banach spaces*. Ann. Probab. 22 (1994) 1679-1706.

[61] Rio, E. *Inégalités de Hoeffding pour les fonctions lipschitziennes de suites dépendantes* (French) C. R. Acad. Sci. Paris S'er. I Math. 330 (2000) 905-908.

[62] Shorack, G. R. and Wellner, J. A. *Empirical processes with applications to statistics*. Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons Inc. New York, (1986).

[63] Volný, D. *Approximating martingales and the central limit theorem for strictly stationary processes*. Stochastic Process. Appl. 44 (1993) 41-74.

[64] Volný, D. *A nonadapted version of the invariance principle of Peligrad and Utev*, C. R. Math. Acad. Sci. Paris 345 (2007) 167-169.

[65] Woyczyński, W.A. *A central limit theorem for martingales in Banach spaces*. Bull. Acad. Polon. Sci. Sr. Sci. Math. Astronom. Phys. 23 (1975) 917–920.

[66] Wu, W. B. *Strong invariance principles for dependent random variables* Ann. Probab. 35 (2007) 2294–2320.

[67] Young, L.-S. *Statistical properties of dynamical systems with some hyperbolicity*, Ann. of Math. (2) 147 (1998) 585–650.

[68] Young, L.-S. *Recurrence times and rates of mixing*, Israel J. Math. 110 (1999) 153–188.

[69] Zhao, O. and Woodroofe, M. *On martingale approximations*, Ann. Appl. Probab. 18 (2008) 1831–1847.