Electron. Commun. Probab. **24** (2019), no. 63, 1–12.
https://doi.org/10.1214/19-ECP269
ISSN: 1083-589X

ELECTRONIC COMMUNICATIONS in PROBABILITY

New insights on concentration inequalities for self-normalized martingales

Bernard Bercu[*] Taieb Touati[†]

Abstract

We propose new concentration inequalities for self-normalized martingales. The main idea is to introduce a suitable weighted sum of the predictable quadratic variation and the total quadratic variation of the martingale. It offers much more flexibility and allows us to improve previous concentration inequalities. Statistical applications on autoregressive process, internal diffusion-limited aggregation process, and online statistical learning are also provided.

Keywords: concentration inequalities; martingales; autoregressive process; statistical learning.
AMS MSC 2010: Primary 60E15; 60G42, Secondary 60G15; 60J80.
Submitted to ECP on June 14, 2019, final version accepted on October 4, 2019.

1 Introduction

Let (M_n) be a locally square integrable real martingale adapted to a filtration $\mathbb{F} = (\mathcal{F}_n)$ with $M_0 = 0$. The predictable quadratic variation and the total quadratic variation of (M_n) are respectively given by

$$<M>_n = \sum_{k=1}^{n} \mathbb{E}[\Delta M_k^2 | \mathcal{F}_{k-1}] \qquad \text{and} \qquad [M]_n = \sum_{k=1}^{n} \Delta M_k^2$$

where $\Delta M_n = M_n - M_{n-1}$ with $<M>_0 = 0$ and $[M]_0 = 0$. Since the pioneer work of Azuma and Hoeffding [1], [16], a wide literature is available on concentration inequalities for martingales. We refer the reader to the recent books [2], [5], [10] where the celebrated Azuma-Hoeffding, Freedman, Bernstein, and De la Peña inequalities are provided. Over the last two decades, there has been a renewed interest in this area of probability. More precisely, extensive studies have been made in order to establish concentration inequalities for (M_n) without boundedness assumptions on its increments [4], [12], [14], [18], [19]. For example, it was established in [4] that for any positive x and y,

$$\mathbb{P}(|M_n| \geqslant x, [M]_n + <M>_n \leqslant y) \leqslant 2 \exp\Big(-\frac{x^2}{2y}\Big). \tag{1.1}$$

We shall improve inequality (1.1) by showing that for any positive x and y,

$$\mathbb{P}(|M_n| \geqslant x, [M]_n + <M>_n \leqslant y) \leqslant 2 \exp\Big(-\frac{8x^2}{9y}\Big). \tag{1.2}$$

[*] Université de Bordeaux, Institut de Mathématiques de Bordeaux, UMR 5251, 351 cours de la libération, 33405 Talence cedex, France. E-mail: bernard.bercu@u-bordeaux.fr

[†] Sorbonne Université, Laboratoire de Probabilités Statistique et Modélisation, UMR 8001, Campus Pierre et Marie Curie, 4 Place Jussieu, 75005 Paris cedex, France. E-mail: taieb.touati@upmc.fr

Moreover, it was proven by Delyon [12] that for any positive x and y,

$$\mathbb{P}(|M_n| \geqslant x, [M]_n + 2 <M>_n \leqslant y) \leqslant 2 \exp\Bigl(-\frac{3x^2}{2y}\Bigr). \tag{1.3}$$

We will show that inequality (1.3) is a special case of a more general result involving a suitable weighted sum of $[M]_n$ and $<M>_n$. Furthermore, it was shown by De la Peña and Pang [11] that for any positive x,

$$\mathbb{P}\Bigl(\frac{|M_n|}{\sqrt{[M]_n + <M>_n + \mathbb{E}[M_n^2]}} \geqslant x\sqrt{\frac{3}{2}}\Bigr) \leqslant \Bigl(\frac{2}{3}\Bigr)^{1/3} x^{-2/3} \exp\Bigl(-\frac{x^2}{2}\Bigr). \tag{1.4}$$

We shall improve inequality (1.4) by using of the tailor-made normalization

$$S_n(a) = [M]_n + c(a) <M>_n, \tag{1.5}$$

where for any $a > 1/8$,

$$c(a) = \frac{2(1 - 2a + 2\sqrt{a(a+1)})}{8a - 1}. \tag{1.6}$$

The novelty of our approach is that $S_n(a)$ is a suitable weighted sum $<M>_n$ and $[M]_n$. For small values of n, the behavior of $<M>_n$ may be totally different from that of $[M]_n$. Consequently, our approach provides interesting concentration inequalities in many situations where $<M>_n \neq [M]_n$. The paper is organised as follows. Section 2 is devoted to our new concentration inequalities for self-normalized martingales which improve some previous results of Bercu and Touati [4], Delyon [12] and De la Peña and Pang [11]. Section 3 deals with statistical applications on autoregressive process, internal diffusion-limited aggregation process, and online statistical learning. All technical proofs are postponed to Sections 4 and 5.

2 Main results

Our first result holds without any additional assumption on (M_n).

Theorem 2.1. *Let (M_n) be a locally square integrable real martingale. Then, as soon as $a > 1/8$, we have for any positive x and y,*

$$\mathbb{P}(|M_n| \geqslant x, S_n(a) \leqslant y) \leqslant 2 \exp\Bigl(-\frac{x^2}{2ay}\Bigr), \tag{2.1}$$

where $S_n(a) = [M]_n + c(a) <M>_n$ and $c(a)$ is given by (1.6).

Remark 2.2. The function c is positive, strictly convex and $c(a) \sim 1/2a$ as a tends to infinity. Special values are given in Table 1.

Table 1: Special values of the function $c(a)$

a	$9/55$	$4/21$	$9/40$	$25/96$	$1/3$	$9/16$	$49/72$	$4/5$
c(a)	10	6	4	3	2	1	$4/5$	$2/3$

In the special case where $<M>_n = [M]_n$, $S_n(a)$ reduces to $S_n(a) = (1 + c(a)) <M>_n$ and the best choice of a is clearly the one that minimizes $aS_n(a) = a(1 + c(a)) <M>_n$, that is $a = 1/3$.

Remark 2.3. On the one hand, $c(a) = 1$ if and only if $a = 9/16$. Replacing the value $a = 9/16$ into (2.1) immediately leads to (1.2) as $S_n(a) = [M]_n + <M>_n$. On the other hand, $c(a) = 2$ if and only if $a = 1/3$. Hence, in this special case, $S_n(a) = [M]_n + 2 <M>_n$ and we find again (1.3) by taking the value $a = 1/3$ into (2.1).

Our second result for self-normalized martingales is as follows.

Theorem 2.4. *Let (M_n) be a locally square integrable real martingale. Then, as soon as $a > 1/8$, we have for any positive x and y,*

$$\mathbb{P}\left(\frac{|M_n|}{S_n(a)} \geqslant x, S_n(a) \geqslant y\right) \leqslant 2\exp\Bigl(-\frac{x^2y}{2a}\Bigr) \tag{2.2}$$

where $S_n(a) = [M]_n + c(a){<}M{>}_n$ and $c(a)$ is given by (1.6)*. Moreover, we have for any positive x,*

$$\mathbb{P}\Bigl(\frac{|M_n|}{S_n(a)} \geqslant x\Bigr) \leqslant 2\inf_{p>1}\left(\mathbb{E}\Bigl[\exp\Bigl(-\frac{(p-1)x^2S_n(a)}{2a}\Bigr)\Bigr]\right)^{1/p}. \tag{2.3}$$

Remark 2.5. In the case $a = 9/16$, we find from (2.2) and (2.3) that for any positive x and y,

$$\mathbb{P}\left(\frac{|M_n|}{[M]_n + {<}M{>}_n} \geqslant x, [M]_n + {<}M{>}_n \geqslant y\right) \leqslant 2\exp\Bigl(-\frac{8x^2y}{9}\Bigr),$$

$$\mathbb{P}\Bigl(\frac{|M_n|}{[M]_n + {<}M{>}_n} \geqslant x\Bigr) \leqslant 2\inf_{p>1}\left(\mathbb{E}\Bigl[\exp\Bigl(-\frac{8(p-1)x^2}{9}\Bigl([M]_n + {<}M{>}_n\Bigr)\Bigr)\Bigr]\right)^{1/p}.$$

Similar concentration inequalities for self-normalized martingales can be obtained for $a = 1/3$. In addition, via the same lines as in the proof of Theorem 2.4, we find that for any positive x and y,

$$\mathbb{P}\left(\frac{|M_n|}{{<}M{>}_n} \geqslant x, c(a){<}M{>}_n \geqslant [M]_n + y\right) \leqslant 2\exp\Bigl(-\frac{x^2y}{2ac^2(a)}\Bigr), \tag{2.4}$$

$$\mathbb{P}\Bigl(\frac{|M_n|}{{<}M{>}_n} \geqslant x, [M]_n \leqslant c(a)y{<}M{>}_n\Bigr) \leqslant 2\inf_{p>1}\left(\mathbb{E}\Bigl[\exp\Bigl(-\frac{(p-1)x^2{<}M{>}_n}{2ac(a)(1+y)}\Bigr)\Bigr]\right)^{1/p}. \tag{2.5}$$

Our third result deals with missing factors in exponential inequalities for self-normalized martingales with upper bounds independent of $[M]_n$ or ${<}M{>}_n$.

Theorem 2.6. *Let (M_n) be a locally square integrable real martingale. Assume that $\mathbb{E}[|M_n|^p] < \infty$ for some $p \geqslant 2$. Then, as soon as $a > 1/8$, we have for any positive x,*

$$\mathbb{P}\Bigl(\frac{|M_n|}{\sqrt{aS_n(a) + (\mathbb{E}[|M_n|^p])^{2/p}}} \geqslant \frac{x}{\sqrt{B_q}}\Bigr) \leqslant C_q x^{-B_q}\exp\Bigl(-\frac{x^2}{2}\Bigr) \tag{2.6}$$

where $q = p/(p-1)$ is the Hölder conjugate exponent of p,

$$B_q = \frac{q}{2q-1} \qquad \textit{and} \qquad C_q = \Bigl(\frac{q}{2q-1}\Bigr)^{B_q/2}.$$

In particular, for $p = 2$, we have for any positive x,

$$\mathbb{P}\Bigl(\frac{|M_n|}{\sqrt{aS_n(a) + \mathbb{E}[M_n^2]}} \geqslant x\sqrt{\frac{3}{2}}\Bigr) \leqslant \Bigl(\frac{2}{3}\Bigr)^{1/3} x^{-2/3}\exp\Bigl(-\frac{x^2}{2}\Bigr). \tag{2.7}$$

Remark 2.7. In the case $a = 9/16$, we deduce from (2.7) that for any positive x,

$$\mathbb{P}\Bigl(\frac{|M_n|}{\sqrt{a([M]_n + {<}M{>}_n) + \mathbb{E}[M_n^2]}} \geqslant x\sqrt{\frac{3}{2}}\Bigr) \leqslant \Bigl(\frac{2}{3}\Bigr)^{1/3} x^{-2/3}\exp\Bigl(-\frac{x^2}{2}\Bigr).$$

Since $a < 1$, this inequality clearly leads to (1.4). Consequently, in the case $a = 9/16$, (2.7) provides a tighter upper bound than inequality (1.4). Moreover, if $a = 1/3$, we obtain from (2.7) that for any positive x,

$$\mathbb{P}\Bigl(\frac{|M_n|}{\sqrt{[M]_n + 2\,{<}M{>}_n + 3\mathbb{E}[M_n^2]}} \geqslant \frac{x}{\sqrt{2}}\Bigr) \leqslant \Bigl(\frac{2}{3}\Bigr)^{1/3} x^{-2/3}\exp\Bigl(-\frac{x^2}{2}\Bigr).$$

Proof. The proofs are given in Sections 4 and 5. □

3 Statistical applications

3.1 Autoregressive process

Consider the first-order autoregressive process given, for all $n \geqslant 1$, by

$$X_n = \theta X_{n-1} + \varepsilon_n \tag{3.1}$$

where X_n and ε_n are the observation and the driven noise of the process, respectively. Assume that (ε_n) is a sequence of independent random variables sharing the same $\mathcal{N}(0,\sigma^2)$ distribution where $\sigma^2 > 0$. The process is said to be stable if $|\theta| < 1$, unstable if $|\theta| = 1$, and explosive if $|\theta| > 1$. We estimate θ by the standard least-squares estimator given, for all $n \geqslant 1$, by

$$\widehat{\theta}_n = \frac{\sum_{k=1}^n X_{k-1}X_k}{\sum_{k=1}^n X_{k-1}^2}. \tag{3.2}$$

It is well-known that whatever the value of θ is, $\widehat{\theta}_n$ converges almost surely to θ. Moreover, White [20] has shown that in the stable case $|\theta| < 1$,

$$\sqrt{n}\big(\widehat{\theta}_n - \theta\big) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1-\theta^2),$$

while in the explosive case $|\theta| > 1$ with initial value $X_0 = 0$,

$$|\theta|^n\big(\widehat{\theta}_n - \theta\big) \xrightarrow{\mathcal{L}} (\theta^2 - 1)\mathcal{C}$$

where $\mathcal{C}$ stands for the Cauchy distribution. Furthermore, in the stable case $|\theta| < 1$, it was proven in [3] that the sequence $(\widehat{\theta}_n)$ satisfies a large deviation principle with a convex-concave rate function. A fairly simple concentration inequality for the estimator $\widehat{\theta}_n$ was established in [4], whatever the value of θ is. More precisely, assume that X_0 is independent of (ε_n) with $\mathcal{N}(0,\tau^2)$ distribution where $\tau^2 \geqslant \sigma^2$. Then, for all $n \geqslant 1$ and for any positive x, we have

$$\mathbb{P}(|\widehat{\theta}_n - \theta| \geqslant x) \leqslant 2\exp\Big(-\frac{nx^2}{2(1+y_x)}\Big) \tag{3.3}$$

where y_x is the unique positive solution of the equation $h(y_x) = x^2$ and h is the function given, for any positive x, by $h(x) = (1+x)\log(1+x) - x$. It follows from (3.3) that, as soon as $0 < x < 1/2$,

$$\mathbb{P}\big(|\widehat{\theta}_n - \theta| \geqslant x\big) \leqslant 2\exp\Big(-\frac{nx^2}{2(1+2x)}\Big).$$

The situation in which (ε_n) is not normally distributed, is much more difficult to handle. If (ε_n) is a sequence of independent and identically distributed random variables, uniformly bounded with symmetric distribution, we can use De la Peña's inequality [9] for self-normalized conditionally symmetric martingales, to prove concentration inequalities for the least-squares estimator, see [2]. Our motivation is to establish concentration inequalities for $\widehat{\theta}_n$ in the situation where the distribution of (ε_n) is non-symmetric.

Corollary 3.1. *Assume that (ε_n) is a sequence of independent and identically distributed random variables such that, for all $n \geqslant 1$,*

$$\varepsilon_n = \begin{cases} 2q & \text{with probability} \quad p, \\ -2p & \text{with probability} \quad q, \end{cases}$$

where $p \in]0,1/2]$ and $q = 1-p$. Moreover, assume that X_0 is independent of (ε_n) with $|X_0| \geqslant 2p$. Then, for any $a > 1/8$ and for any x in the interval $[0, \sqrt{ad(a)}]$, we have

$$\mathbb{P}\big(|\widehat{\theta}_n - \theta| \geqslant x\big) \leqslant 2\exp\Big(-\frac{np^2x^2}{ad(a)}\Big) \qquad \text{where} \qquad d(a) = \frac{4\big(q^2 + pqc(a)\big)^2}{\big(p^2 + pqc(a)\big)}. \tag{3.4}$$

Remark 3.2. In the symmetric case $p = 1/2$, we clearly have from (3.6) that $<M>_n= [M]_n$, $S_n(a) = (1 + c(a)) <M>_n$ and $d(a)$ reduces to $d(a) = 1 + c(a)$. Hence, if $a = 1/3$, $c(a) = 2$ and $d(a) = 3$. Consequently, we deduce from (3.4) that for any x in $[0, 1]$,

$$\mathbb{P}\big(|\widehat{\theta}_n - \theta| \geqslant x\big) \leqslant 2\exp\Big(-\frac{nx^2}{4}\Big).$$

Moreover, in the nonsymmetric case $p \neq 1/2$, we always have $<M>_n\neq [M]_n$. For example, if $p = 1/3$ and $a = 9/16$, $c(a) = 1$ and $d(a) = 16/3$ which implies that $ad(a) = 3$. Therefore, we obtain from (3.4) that for any x in $[0, \sqrt{3}]$,

$$\mathbb{P}\big(|\widehat{\theta}_n - \theta| \geqslant x\big) \leqslant 2\exp\Big(-\frac{nx^2}{27}\Big).$$

Proof. It immediately follows from (3.1) together with (3.2) that for all $n \geqslant 1$,

$$\widehat{\theta}_n - \theta = \sigma^2\frac{M_n}{<M>_n} \tag{3.5}$$

where $\sigma^2 = 4pq$ and (M_n) is the locally square integrable real martingale given by

$$M_n = \sum_{k=1}^n X_{k-1}\varepsilon_k, \qquad <M>_n= \sigma^2\sum_{k=1}^n X_{k-1}^2, \qquad [M]_n = \sum_{k=1}^n X_{k-1}^2\varepsilon_k^2. \tag{3.6}$$

We clearly have $\big(c(a) + r\big) <M>_n\leqslant S_n(a) \leqslant \big(c(a) + r^{-1}\big) <M>_n$ with $r = p/q$. Hence, we obtain from (2.3) that for any $a > 1/8$ and for any positive x,

$$\mathbb{P}\big(|M_n| \geqslant xS_n(a)\big) \leqslant 2\left(\mathbb{E}\Big[\exp\Big(-\frac{x^2S_n(a)}{2a}\Big)\Big]\right)^{1/2} \tag{3.7}$$

which implies via (3.5) that

$$\mathbb{P}\big(|\widehat{\theta}_n - \theta| \geqslant x\big) \leqslant 2\left(\mathbb{E}\Big[\exp\Big(-\frac{x^2 <M>_n}{2a\sigma^2 d(a)}\Big)\Big]\right)^{1/2} \tag{3.8}$$

where $d(a)$ is given by (3.4). It only remains to find a suitable upper-bound for the Laplace transform of $<M>_n$. We have from (3.1) that $X_n^2 = \theta^2X_{n-1}^2 + 2\theta X_{n-1}\varepsilon_n + \varepsilon_n^2$. Hence, if $\mathcal{F}_n = \sigma(X_0, \ldots, X_n)$, we obtain that for any real t and for all $n \geqslant 1$,

$$\mathbb{E}[\exp(tX_n^2)|\mathcal{F}_{n-1}] = \exp(t\theta^2X_{n-1}^2)\Lambda_{n-1}(t) \tag{3.9}$$

where

$$\Lambda_{n-1}(t) = p\exp\big(4tq^2 + 4\theta tqX_{n-1}\big) + q\exp\big(4tp^2 - 4\theta tpX_{n-1}\big). \tag{3.10}$$

It follows from the so-called Kearns-Saul's inequality given in Lemma 2.36, page 37 of [2] that for any real s,

$$p\exp(qs) + q\exp(-ps) \leqslant \exp\Big(\frac{\varphi(p)s^2}{4}\Big), \tag{3.11}$$

where $\varphi(p) = (q - p)/\log(q/p) \in [0, 1/2]$. Then, we deduce from (3.10) and (3.11) with $s = 4\theta tX_{n-1}$ that for any $t \leqslant 0$, $\Lambda_{n-1}(t) \leqslant \exp\big(4tp^2 + 4\varphi(p)t^2\theta^2X_{n-1}^2\big)$ leading to

$$\mathbb{E}[\exp(tX_n^2)|\mathcal{F}_{n-1}] \leqslant \exp\big(4tp^2 + t\theta^2X_{n-1}^2(1 + 4\varphi(p)t)\big). \tag{3.12}$$

As soon as $t \in [-1/2, 0]$, we get from (3.12) that $\mathbb{E}[\exp(tX_n^2)|\mathcal{F}_{n-1}] \leqslant \exp(4tp^2)$. Consequently, for any $t \in [-1/2\sigma^2, 0]$ and for all $n \geqslant 1$,

$$\mathbb{E}[\exp(t <M>_n)] \leqslant \mathbb{E}[\exp(t <M>_{n-1})]\exp(4tp^2\sigma^2) \leqslant \exp(4ntp^2\sigma^2) \tag{3.13}$$

as $|X_0| \geqslant 2p$. Therefore, it follows from (3.8) and (3.13) that for any $x \in [0, \sqrt{ad(a)}]$,

$$\mathbb{P}\big(|\widehat{\theta}_n - \theta| \geqslant x\big) \leqslant 2\exp\Big(-\frac{np^2x^2}{ad(a)}\Big)$$

which achieves the proof of Corollary 3.1. □

3.2 Internal diffusion-limited aggregation process

Our second application deals with the internal diffusion-limited aggregation process. This aggregation process, first introduced in Mathematics by Diaconis and Fulton [13], is a cluster growth model in $\mathbb{Z}^d$ where explorers, starting from the origin at time 0, are travelling as a simple random walk on $\mathbb{Z}^d$ until they reach an uninhabited site that is added to the cluster. In the special case $d = 1$, the cluster is an interval $A(n) = [L_n, R_n]$ which, properly normalized, converges almost surely to $[-1, 1]$. In dimension $d \geqslant 2$, Lawler, Bramson and Griffeath [17] have shown that the limit shape of the cluster is a sphere. We shall restrict our attention on the one-dimensional internal diffusion-limited aggregation process. Consider the simple random walk on the integer number line $\mathbb{Z}$ starting from the origin at time 0. At each step, the explorer moves to the right $+1$ or to the left -1 with equal probability $1/2$. Let $(A(n))$ be the sequence of random subsets of $\mathbb{Z}$, recursively defined as follows: $A(0) = \{0\}$ and, for all $n \geqslant 0$,

$$A(n+1) = \begin{cases} A(n) \cup \{L_n - 1\} \\ A(n) \cup \{R_n + 1\} \end{cases}$$

if the explorer leaves $A(n)$ by the left side or by the right side, respectively, where L_n and R_n stand for being the minimum and the maximum of $A(n) = \{L_n, L_n+1, \dots, R_n-1, R_n\}$. The random set $A(n)$ is characterized by $X_n = L_n + R_n$ as $R_n - L_n = n$. One can observe that L_n and R_n correspond to the number of negative and positive sites of $A(n)$, respectively. It was proven in [13] that

$$\lim_{n\to\infty} \frac{X_n}{n} = 0 \qquad \text{a.s.}$$

and

$$\frac{X_n}{\sqrt{n}} \xrightarrow{\mathcal{L}} \mathcal{N}\Big(0, \frac{1}{3}\Big).$$

It is possible to prove from Azuma-Hoeffding's inequality [2] that for any positive x,

$$\mathbb{P}\Big(\frac{|X_n|}{n} \geqslant x\Big) \leqslant 2\exp\Big(-\frac{3}{8}nx^2\Big). \tag{3.14}$$

Our goal is to improve this inequality with a suitable use of Theorems 2.1 and 2.6.

Corollary 3.3. *For any a in the interval $]1/8, 9/16]$ and for any positive x, we have*

$$\mathbb{P}\Big(\frac{|X_n|}{n} \geqslant x\Big) \leqslant 2\exp\Big(-\frac{nx^2}{2ac_n(a)}\Big) \tag{3.15}$$

and

$$\mathbb{P}\Big(\frac{|X_n|}{\sqrt{n}} \geqslant x\Big) \leqslant (d_n(a))^{1/3}x^{-2/3}\exp\Big(-\frac{x^2}{3d_n(a)}\Big) \tag{3.16}$$

where

$$c_n(a) = \Big(\frac{2n+1}{n+1}\Big)\Big(\frac{3+c(a)}{6}\Big) + \Big(\frac{n(1+c(a))+2c(a)}{(n+1)^2}\Big), \qquad d_n(a) = c_n(a) + \Big(\frac{n+2}{3n}\Big). \tag{3.17}$$

Remark 3.4. The calculation of $c_n(a)$ and $d_n(a)$ is quite straightforward. As a matter of fact, if $a = 1/3$, $c(a) = 2$ and it immediately follows from (3.17) that for all $n \geqslant 1$ $c_n(a) \leqslant 3$ and $d_n(a) \leqslant 4$. We can deduce from (3.15) that for any positive x,

$$\mathbb{P}\Big(\frac{|X_n|}{n} \geqslant x\Big) \leqslant 2\exp\Big(-\frac{nx^2}{2}\Big)$$

which clearly outperforms inequality (3.14). In addition, (3.16) implies that for any positive x,

$$\mathbb{P}\Big(\frac{|X_n|}{\sqrt{n}} \geqslant x\Big) \leqslant \Big(\frac{2}{x}\Big)^{2/3}\exp\Big(-\frac{x^2}{12}\Big).$$

Moreover, if $a = 25/96$, $c(a) = 3$ and we obtain from (3.17) that for all $n \geqslant 1$, $c_n(a) \leqslant 7/2$ and $d_n(a) \leqslant 9/2$. We find from (3.15) that for any positive x,

$$\mathbb{P}\Big(\frac{|X_n|}{n} \geqslant x\Big) \leqslant 2\exp\Big(-\frac{96nx^2}{175}\Big).$$

It improves the above inequality for $a = 1/3$. Finally, we deduce from (3.16) that

$$\mathbb{P}\Big(\frac{|X_n|}{\sqrt{n}} \geqslant x\Big) \leqslant \Big(\frac{3}{\sqrt{2}x}\Big)^{2/3} \exp\Big(-\frac{2x^2}{27}\Big).$$

Proof. It follows from a stopping time argument for gambler's ruin that for all $n \geqslant 1$, $X_n = X_{n-1} + \xi_n$ where the distribution of ξ_n given $\mathcal{F}_{n-1}$ is a Rademacher $\mathcal{R}(p_n)$ distribution with

$$p_n = \frac{(n+1-X_{n-1})}{2(n+1)}.$$

Hence, we clearly have

$$\mathbb{E}[X_n|\mathcal{F}_{n-1}] = X_{n-1} + \mathbb{E}[\xi_n|\mathcal{F}_{n-1}] = \Big(\frac{n}{n+1}\Big)X_{n-1} \tag{3.18}$$

and

$$\mathbb{E}[X_n^2|\mathcal{F}_{n-1}] = X_{n-1}^2 + 2X_{n-1}\mathbb{E}[\xi_n|\mathcal{F}_{n-1}] + 1 = 1 + \Big(\frac{n-1}{n+1}\Big)X_{n-1}^2. \tag{3.19}$$

Let (M_n) be the sequence defined by $M_n = (n+1)X_n$. We immediately deduce from (3.18) and (3.19) that (M_n) is a locally square integrable real martingale such that

$$<M>_n = \sum_{k=1}^{n}(k+1)^2 - \sum_{k=1}^{n} X_{k-1}^2.$$

Moreover, for all $n \geqslant 1$, $|X_n| \leqslant n$. Hence,

$$[M]_n = \sum_{k=1}^{n}((k+1)X_k - kX_{k-1})^2 = \sum_{k=1}^{n}(k\xi_k + X_k)^2 \leqslant 3\sum_{k=1}^{n} k^2 + \sum_{k=1}^{n} X_k^2.$$

One can observe that we always have $<M>_n \neq [M]_n$. In addition,

$$S_n(a) \leqslant (3 + c(a))\sum_{k=1}^{n} k^2 + (1 - c(a))\sum_{k=1}^{n-1} X_k^2 + X_n^2 + c(a)n(n+2). \tag{3.20}$$

For any $a \in]1/8, 9/16]$, $c(a) \geqslant 1$. Therefore, we obtain from (3.20) that

$$S_n(a) \leqslant (3 + c(a))\sum_{k=1}^{n} k^2 + n(n + c(a)(n+2)) \leqslant n(n+1)^2 c_n(a) \tag{3.21}$$

where $c_n(a)$ is given by (3.17). Hence, it follows from (2.1) with $y = n(n+1)^2 c_n(a)$ that for any $a \in]1/8, 9/16]$ and for any positive x,

$$\mathbb{P}\Big(\frac{|X_n|}{n} \geqslant x\Big) = \mathbb{P}\big(|M_n| \geqslant xn(n+1), S_n(a) \leqslant y\big) \leqslant 2\exp\Big(-\frac{nx^2}{2ac_n(a)}\Big),$$

which is exactly inequality (3.15). Furthermore, we can deduce from identity (3.19) that for all $n \geqslant 1$,

$$\mathbb{E}[X_n^2] = \frac{(n+2)}{3} \qquad \text{and} \qquad \mathbb{E}[M_n^2] = \frac{(n+1)^2(n+2)}{3}. \tag{3.22}$$

Finally, we find from (2.7) together with (3.17), (3.21) and (3.22) that for any $a \in]1/8, 9/16]$ and for any positive x,

$$\mathbb{P}\Big(\frac{|X_n|}{\sqrt{nd_n(a)}} \geqslant x\sqrt{\frac{3}{2}}\Big) \leqslant \Big(\frac{2}{3}\Big)^{1/3} x^{-2/3} \exp\Big(-\frac{x^2}{2}\Big)$$

which clearly leads to (3.16), completing the proof of Corollary 3.3. □

3.3 Online statistical learning

Our third application is devoted to the study of the statistical risk of hypothesis during an online learning process using concentration inequalities for martingales. We refer the reader to the survey of Cesa-Bianchi and Lugosi [8] for a rather exhaustive description of the underlying theory concerning online learning. Our approach is based on the contributions of Cesa-Bianchi *et al.* [6], [7] dealing with the statistical risk of hypothesis in the situation where the ensemble of hypotheses is produced by training a learning algorithm incrementally on a data set of independent and identically distributed random variables. Their bounds rely on Freedman concentration inequality for martingales [15]. Consider the task of predicting a sequence in an online manner with inputs and outputs taking values in some abstract measurable spaces $\mathcal{X}$ and $\mathcal{Y}$, respectively. We call hypothesis H, the classifier or regressor generated by a learning algorithm after training. The predictive performance of hypothesis H is evaluated by the theoretical risk denoted $R(H)$, which is the expected loss on a realisation $(X,Y) \in \mathcal{X} \times \mathcal{Y}$ drawn from the underlying distribution

$$R(H) = \mathbb{E}[\ell(H(X),Y)]$$

where ℓ is a nonnegative and bounded loss function. For the sake of simplicity, we assume that ℓ is bounded by 1. Denote by $\mathcal{S}_n = \{(X_1,Y_1),\ldots,(X_n,Y_n)\}$ a training data set of independent random variables sharing the same unknown distribution as (X,Y). Our goal is to predict $Y_{n+1} \in \mathcal{Y}$ given $X_{n+1} \in \mathcal{X}$, on the basis of $\mathcal{S}_n$. Let $\mathcal{H}_n = \{H_0, H_1, \ldots, H_{n-1}\}$ be a finite ensemble of hypotheses generated by an online learning algorithm where the initial hypothesis H_0 is arbitrarily chosen.The empirical risk and the average risk associated with $\mathcal{H}_n$ and the training data set $\mathcal{S}_n$ are respectively given by

$$\widehat{R}_n = \frac{1}{n}\sum_{k=1}^{n} \ell(H_{k-1}(X_k),Y_k) \qquad \text{and} \qquad R_n = \frac{1}{n}\sum_{k=1}^{n} R(H_{k-1}). \tag{3.23}$$

Our bound on the average risk R_n is as follows.

Corollary 3.5. *Let $\mathcal{H}_n = \{H_0, H_1, ..., H_{n-1}\}$ be a finite ensemble of hypotheses generated by a learning algorithm. Then, for any a in the interval $]1/8, 9/16]$ and for any positive x, we have*

$$\mathbb{P}(R_n \geqslant \widehat{R}_n + x) \leqslant \exp\Big(-\frac{nx^2}{2a(1+c(a)V_n)}\Big), \tag{3.24}$$

where

$$V_n = \frac{1}{n}\sum_{k=1}^{n} \mathbb{E}[\ell^2(H_{k-1}(X),Y)]. \tag{3.25}$$

In other words, for any $0 < \delta \leqslant 1$,

$$\mathbb{P}\Big(R_n \geqslant \widehat{R}_n + \sqrt{\frac{2a(1+c(a)V_n)\log(1/\delta)}{n}}\Big) \leqslant \delta. \tag{3.26}$$

Moreover, denote $m(a) = \max(4(1+c(a)), c^2(a))/2$. Then, for any $0 < \delta \leqslant 1$ and for all integer $n \geqslant am(a)\log(1/\delta)$, we also have

$$\mathbb{P}\Big(R_n \geqslant \widehat{R}_n + \frac{ac(a)\log(1/\delta)}{n} + \sqrt{\frac{a\,\Delta_n(a)\log(1/\delta)}{n}}\Big) \leqslant \delta \tag{3.27}$$

where $\Delta_n(a) = 2 + 2c(a)\widehat{R}_n + ac^2(a)\log(1/\delta)/n$.

Remark 3.6. On the one hand, (3.26) improves the deviation inequality given in Proposition 1 of Cesa-Bianchi, Conconi and Gentile [6],

$$\mathbb{P}\Big(R_n \geqslant \widehat{R}_n + \sqrt{\frac{2\log(1/\delta)}{n}}\Big) \leqslant \delta,$$

as V_n is always smaller than 1. On the other hand, (3.27) is drastically more accurate than the deviation inequality given in Proposition 2 of Cesa-Bianchi and Gentile [7],

$$\mathbb{P}\Big(R_n \geqslant \widehat{R}_n + \frac{36}{n}\log\Big(\frac{n\widehat{R}_n+3}{\delta}\Big) + 2\sqrt{\frac{\widehat{R}_n}{n}\log\Big(\frac{n\widehat{R}_n+3}{\delta}\Big)}\Big) \leqslant \delta. \tag{3.28}$$

Indeed, one can observe that the right-hand sides of (3.27) and (3.28) are increasing functions of $\widehat{R}_n$. The smallest value in (3.28) for $\widehat{R}_n = 0$ is given by $36\log(3/\delta)/n$. Consequently, inequality (3.28) is only effective for $n \geqslant 36\log(3/\delta)$, which implies that n must always be greater than 40. For example, if $\delta = 1/5$, it is necessary to assume that $n \geqslant 36\log(15)$, that is $n \geqslant 98$. If $a = 1/3$, then $c(a) = 2$ and $m(a) = 6$. Consequently, inequality (3.27) is interesting as soon as $n \geqslant -2\log(\delta)$. For example, if $\delta = 1/5$, it is necessary to assume that $n \geqslant 4$. For instance, if $\delta = 1/5$, $n = 100$ and $a = 1/3$, the smallest values in (3.27) and (3.28) are respectively given by 0.220 and 0.975. Finally, for all values of δ, n and a, one can easily check that (3.27) is always sharper than (3.28).

Proof. Let (M_n) be the locally square integrable real martingale given by

$$M_n = \sum_{k=1}^{n}\big(R(H_{k-1}) - \ell(H_{k-1}(X_k), Y_k)\big), \tag{3.29}$$

where we recall that $R(H) = \mathbb{E}[\ell(H(X), Y)]$. We clearly have

$$<M>_n = \sum_{k=1}^{n}\big(\mathbb{E}[\ell^2(H_{k-1}(X), Y))] - R^2(H_{k-1})\big), \quad [M]_n = \sum_{k=1}^{n}\big(R(H_{k-1}) - \ell(H_{k-1}(X_k), Y_k)\big)^2.$$

Consequently, for any $a \in]1/8, 9/16]$,

$$S_n(a) \leqslant (1 - c(a))\sum_{k=1}^{n} R^2(H_{k-1}) + \sum_{k=1}^{n} \ell^2(H_{k-1}(X_k), Y_k) + c(a)\sum_{k=1}^{n}\mathbb{E}[\ell^2(H_{k-1}(X), Y)]$$

Hence, as $c(a) \geqslant 1$ and ℓ is bounded by 1, we obtain from (3.25) that $S_n(a) \leqslant n(1 + c(a)V_n)$. Therefore, it follows from (2.1) with $y = n(1 + c(a)V_n)$ that for any $a \in]1/8, 9/16]$ and for any positive x,

$$\mathbb{P}\Big(\frac{M_n}{n} \geqslant x\Big) \leqslant \exp\Big(-\frac{nx^2}{2a(1 + c(a)V_n)}\Big). \tag{3.30}$$

However, we clearly have from (3.29) that $M_n = n\big(R_n - \widehat{R}_n\big)$. Hence, (3.30) immediately implies (3.24) and (3.26). It only remains to prove (3.27). Since ℓ is bounded by 1, we obtain from (3.25) that $V_n \leqslant R_n$. Consequently, (3.26) ensures that for any $0 < \delta \leqslant 1$,

$$\mathbb{P}\big(\Phi_a(R_n) \geqslant \widehat{R}_n\big) \leqslant \delta \tag{3.31}$$

where the function Φ_a is defined, for all x in $[0, 1]$, by

$$\Phi_a(x) = x - \sqrt{\frac{2a(1 + c(a)x)\log(1/\delta)}{n}}.$$

It is not hard to see that, as soon as $n \geqslant am(a)\log(1/\delta)$ with $m(a) = \max(4(1 + c(a)), c^2(a))/2$, Φ_a is a strictly convex and increasing function on $[0, 1]$. Then, Φ_a is invertible and it follows from straightforward calculations that

$$\Phi_a^{-1}(x) = x + \frac{ac(a)\log(1/\delta)}{n} + \sqrt{\frac{a\log(1/\delta)}{n}\Big(2 + 2c(a)x + \frac{ac^2(a)\log(1/\delta)}{n}\Big)}.$$

Finally, we immediately obtain from (3.31) that

$$\mathbb{P}\big(\Phi_a(R_n) \geqslant \widehat{R}_n\big) = \mathbb{P}\big(R_n \geqslant \Phi_a^{-1}(\widehat{R}_n)\big) \leqslant \delta \tag{3.32}$$

which is exactly inequality (3.27), completing the proof of Corollary 3.5. □

4 Two keystone lemmas

Our first lemma deals with a sharp upper bound on the Hermite generating function associated with a centered random variable X.

Lemma 4.1. *Let X be a square integrable random variable with zero mean and variance σ^2. For all $t \in \mathbb{R}$, denote*

$$L(t) = \mathbb{E}\Big[\exp\Big(tX - \frac{at^2}{2}X^2\Big)\Big] \tag{4.1}$$

with $a > 1/8$. Then, for all $t \in \mathbb{R}$,

$$L(t) \leqslant 1 + \frac{b(a)t^2}{2}\sigma^2 \qquad \text{where} \qquad b(a) = \frac{2a(1-2a+2\sqrt{a(a+1)})}{8a-1}. \tag{4.2}$$

Proof. In order to simplify the notation, denote $b = b(a)$. The proof of Lemma 4.1 relies on the following Hermite inequality, see also Proposition 12 in Delyon [12] for the special value $a = 1/3$. For all $x \in \mathbb{R}$, we have

$$\exp\Big(x - \frac{ax^2}{2}\Big) \leqslant 1 + x + \frac{bx^2}{2}. \tag{4.3}$$

As a matter of fact, let

$$\varphi_a(x) = \log\Big(1 + x + \frac{bx^2}{2}\Big) - x + \frac{ax^2}{2}. \tag{4.4}$$

It is of course necessary to assume that $b > 1/2$ which ensures that $1 + x + bx^2/2$ is positive whatever the value of x is. We clearly have

$$\varphi_a'(x) = \Big(1 + x + \frac{bx^2}{2}\Big)^{-1} x P_{a,b}(x), \tag{4.5}$$

where the second degree polynomial $P_{a,b}$ is given by

$$P_{a,b}(x) = \frac{abx^2}{2} + \frac{(2a-b)x}{2} + a + b - 1.$$

Hereafter, assume that $a > 1/8$ and $b \neq 1-a$. The unique positive root of the discriminant of $P_{a,b}$ is gven by $b = b(a)$. Consequently, as $\varphi_a'(0) = 0$ and $\varphi_a(0) = 0$, we deduce from (4.5) that the function φ_a reaches its minimum at $x = 0$ and we find that for all $x \in \mathbb{R}$, $\varphi_a(x) \geqslant 0$ which immediately leads to (4.3). Therefore, we obtain from (4.3) that for all $t \in \mathbb{R}$,

$$L(t) = \mathbb{E}\Big[\exp\Big(tX - \frac{at^2}{2}X^2\Big)\Big] \leqslant \mathbb{E}\Big[1 + tX + \frac{bt^2X^2}{2}\Big] = 1 + \frac{bt^2}{2}\sigma^2,$$

which is exactly what we wanted to prove. □

Our second exponential supermartingale lemma is as follows.

Lemma 4.2. *Let (M_n) be a locally square integrable real martingale. For all $t \in \mathbb{R}$ and $n \geqslant 0$, denote*

$$V_n(t) = \exp\Big(tM_n - \frac{at^2}{2}[M]_n - \frac{b(a)t^2}{2} {<}M{>}_n\Big) \tag{4.6}$$

with $a > 1/8$. Then, $(V_n(t))$ is a positive supermartingale such that $\mathbb{E}[V_n(t)] \leqslant 1$.

Proof. The proof follows from Lemma 4.1 together with standard arguments, see [4] page 1860. □

5 Proofs of the main results

Proof of Theorem 2.1. For any positive x and y, let $A_n = \big\{|M_n| \geqslant x, aS_n(a) \leqslant y\big\}$. We have the decomposition $A_n = A_n^+ \cup A_n^-$ where $A_n^+ = \big\{M_n \geqslant x, aS_n(a) \leqslant y\big\}$ and $A_n^- = \big\{M_n \leqslant -x, aS_n(a) \leqslant y\big\}$. It follows from Markov's inequality together with Lemma 4.2

that for all positive t,

$$\begin{aligned}\mathbb{P}(A_n^+) &\leqslant \mathbb{E}\Big[\exp\Big(tM_n - tx\Big)\mathrm{I}_{A_n^+}\Big] \leqslant \mathbb{E}\Big[\exp\Big(tM_n - \frac{t^2}{2}aS_n(a)\Big)\exp\Big(\frac{t^2}{2}aS_n(a) - tx\Big)\mathrm{I}_{A_n^+}\Big],\\ &\leqslant \exp\Big(\frac{t^2y}{2} - tx\Big)\mathbb{E}[V_n(t)] \leqslant \exp\Big(\frac{t^2y}{2} - tx\Big).\end{aligned}$$

Hence, by taking the optimal value $t = x/y$ in the above inequality, we find that

$$\mathbb{P}(A_n^+) \leqslant \exp\Big(-\frac{x^2}{2y}\Big).$$

We also obtain the same upper bound for $\mathbb{P}(A_n^-)$ which ensures that

$$\mathbb{P}(A_n) \leqslant 2\exp\Big(-\frac{x^2}{2y}\Big). \tag{5.1}$$

Finally, inequality (5.1) clearly leads to (2.1) replacing y by ay. □

Proof of Theorem 2.4. For any positive x and y, let $B_n = \big\{|M_n| \geqslant xS_n(a), S_n(a) \geqslant y\big\} = B_n^+ \cup B_n^-$ where $B_n^+ = \big\{M_n \geqslant xS_n(a), S_n(a) \geqslant y\big\}$ and $B_n^- = \big\{M_n \leqslant -xS_n(a), S_n(a) \geqslant y\big\}$. Proceeding as in the proof of Theorem 2.1, we have that for all $0 < t < 2x/a$,

$$\begin{aligned}\mathbb{P}(B_n^+) &\leqslant \mathbb{E}\Big[\exp\Big(tM_n - txS_n(a)\Big)\mathrm{I}_{B_n^+}\Big] &(5.2)\\ &\leqslant \mathbb{E}\Big[\exp\Big(tM_n - \frac{t^2}{2}aS_n(a)\Big)\exp\Big(\frac{t}{2}(ta-2x)S_n(a)\Big)\mathrm{I}_{B_n^+}\Big],\\ &\leqslant \exp\Big(\frac{t}{2}(ta-2x)y\Big)\mathbb{E}[V_n(t)] \leqslant \exp\Big(\frac{t}{2}(ta-2x)y\Big). &(5.3)\end{aligned}$$

Consequently, we find from (5.3) with the particular choice $t = x/a$ that

$$\mathbb{P}(B_n^+) \leqslant \exp\Big(-\frac{x^2y}{2a}\Big). \tag{5.4}$$

The same upper bound holds for $\mathbb{P}(B_n^-)$ which clearly implies (2.2). Furthermore, for any positive x, let $C_n = \big\{|M_n| \geqslant xS_n(a)\big\} = C_n^+ \cup C_n^-$ where $C_n^+ = \big\{M_n \geqslant xS_n(a)\big\}$ and $C_n^- = \big\{M_n \leqslant -xS_n(a)\big\}$. By Holder's inequality, we have for all positive t and $q > 1$,

$$\begin{aligned}\mathbb{P}(C_n^+) &\leqslant \mathbb{E}\Big[\exp\Big(\frac{t}{q}M_n - \frac{tx}{q}S_n(a)\Big)\mathrm{I}_{C_n^+}\Big],\\ &\leqslant \mathbb{E}\Big[\big(V_n(t)\big)^{1/q}\exp\Big(\frac{t}{2q}(ta-2x)S_n(a)\Big)\Big],\\ &\leqslant \Big(\mathbb{E}\Big[\exp\Big(\frac{tp}{2q}(ta-2x)S_n(a)\Big)\Big]\Big)^{1/p}. &(5.5)\end{aligned}$$

Consequently, as $p/q = p-1$, we deduce from (5.5) with the optimal value $t = x/a$ that

$$\mathbb{P}(C_n^+) \leqslant \inf_{p>1}\Big(\mathbb{E}\Big[\exp\Big(-\frac{(p-1)x^2S_n(a)}{2a}\Big)\Big]\Big)^{1/p}.$$

We find the same upper bound for $\mathbb{P}(C_n^-)$, completing the proof of Theorem 2.4. □

Proof of Theorem 2.6. We already saw from Lemma 4.2 that for all $t \in \mathbb{R}$,

$$\mathbb{E}\Big[\exp\Big(tA_n - \frac{t^2}{2}B_n^2\Big)\Big] \leqslant 1$$

where $A_n = M_n$ and $B_n^2 = a[M]_n + b(a)<M>_n$. It means that the pair of random variables (A_n, B_n) safisties the canonical assumption in [11]. Theorem 2.6 follows from Theorem 2.1 in [11]. □

References

[1] Kazuoki Azuma, *Weighted sums of certain dependent random variables*, Tohoku Math. J. **19** (1967), 357–367. MR-0221571

[2] Bernard Bercu, Bernard Delyon, and Emmanuel Rio, *Concentration inequalities for sums and martingales*, SpringerBriefs in Mathematics, Springer, Cham, 2015. MR-3363542

[3] Bernard Bercu, Fabrice Gamboa, and Alain Rouault, *Large deviations for quadratic forms of stationary Gaussian processes*, Stochastic Process. Appl. **71** (1997), no. 1, 75–90. MR-1480640

[4] Bernard Bercu and Abderrahmen Touati, *Exponential inequalities for self-normalized martingales with applications*, Ann. Appl. Probab. **18** (2008), no. 5, 1848–1869. MR-2462551

[5] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart, *Concentration inequalities*, Oxford University Press, Oxford, 2013, A nonasymptotic theory of independence. MR-3185193

[6] Nicolò Cesa-Bianchi, Alex Conconi, and Claudio Gentile, *On the generalization ability of on-line learning algorithms*, IEEE Trans. Inform. Theory **50** (2004), no. 9, 2050–2057. MR-2097190

[7] Nicolò Cesa-Bianchi and Claudio Gentile, *Improved risk tail bounds for on-line algorithms*, IEEE Trans. Inform. Theory **54** (2008), no. 1, 386–390. MR-2446761

[8] Nicolò Cesa-Bianchi and Gábor Lugosi, *Prediction, learning, and games*, Cambridge University Press, Cambridge, 2006. MR-2409394

[9] Victor H. de la Peña, *A general class of exponential inequalities for martingales and ratios*, Ann. Probab. **27** (1999), no. 1, 537–564. MR-1681153

[10] Victor H. de la Peña, Michael J. Klass, and Tze Leung Lai, *Pseudo-maximization and self-normalized processes*, Probab. Surv. **4** (2007), 172–192. MR-2368950

[11] Victor H. de la Peña and Guodong Pang, *Exponential inequalities for self-normalized processes with applications*, Electron. Commun. Probab. **14** (2009), 372–381. MR-2545288

[12] Bernard Delyon, *Exponential inequalities for sums of weakly dependent variables*, Electron. J. Probab. **14** (2009), no. 28, 752–779. MR-2495559

[13] Persi Diaconis and William Fulton, *A growth model, a game, an algebra, Lagrange inversion, and characteristic classes*, Rend. Sem. Mat. Univ. Politec. Torino **49** (1991), no. 1, 95–119 (1993). MR-1218674

[14] Xiequan Fan, Ion Grama, and Quansheng Liu, *Exponential inequalities for martingales with applications*, Electron. J. Probab. **20** (2015), no. 1, 1–22. MR-3311214

[15] David A. Freedman, *On tail probabilities for martingales*, Ann. Probability **3** (1975), 100–118. MR-0380971

[16] Wassily Hoeffding, *Probability inequalities for sums of bounded random variables*, J. Amer. Statist. Assoc. **58** (1963), 13–30. MR-0144363

[17] Gregory F. Lawler, Maury Bramson, and David Griffeath, *Internal diffusion limited aggregation*, Ann. Probab. **20** (1992), no. 4, 2117–2140. MR-1188055

[18] Iosif Pinelis, *On the Bennett-Hoeffding inequality*, Ann. Inst. Henri Poincaré Probab. Stat. **50** (2014), no. 1, 15–27. MR-3161520

[19] Emmanuel Rio, *Extensions of the Hoeffding-Azuma inequalities*, Electron. Commun. Probab. **18** (2013), no. 54, 6. MR-3078017

[20] John S. White, *The limiting distribution of the serial correlation coefficient in the explosive case*, Ann. Math. Statist. **29** (1958), 1188–1197. MR-0100952

Acknowledgments. The authors would like to thank the two anonymous referees for their useful comments and suggestions which help to improve the presentation of the paper substantially. This work is dedicaded to Professor Abderrahmen Touati who was one of the founding fathers of this type of inequalities. Abdou was very much appreciated both, for his scientific rigor and for his human qualities. It is an honor and a duty for us to make his knowledge and principles live after him.