



**HAL**  
open science

## Hoeffding's inequality for supermartingales

Xiequan Fan, Ion Grama, Quansheng Liu

► **To cite this version:**

Xiequan Fan, Ion Grama, Quansheng Liu. Hoeffding's inequality for supermartingales. *Stochastic Processes and their Applications*, 2012, 122, pp.3545-3559. hal-00905536v2

**HAL Id: hal-00905536**

**<https://hal.science/hal-00905536v2>**

Submitted on 21 Nov 2013

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Hoeffding's inequality for supermartingales

Xiequan Fan\*, Ion Grama and Quansheng Liu

*Université de Bretagne-Sud, LMBA, UMR CNRS 6205,  
Campus de Tohannic, 56017 Vannes, France*

---

## Abstract

We give an extension of Hoeffding's inequality to the case of supermartingales with differences bounded from above. Our inequality strengthens or extends the inequalities of Freedman, Bernstein, Prohorov, Bennett and Nagaev.

*Keywords:* Concentration inequalities; Hoeffding's inequality; Freedman's inequality; Bennett's inequality; martingales; supermartingales

*2000 MSC:* Primary 60G42, 60G40, 60F10; second 60E15, 60G50

---

## 1. Introduction

Let  $\xi_1, \dots, \xi_n$  be a sequence of centered ( $\mathbb{E}\xi_i = 0$ ) random variables such that  $\sigma_i^2 = \mathbb{E}\xi_i^2 < \infty$  and let  $X_n = \sum_{i=1}^n \xi_i$ . Since the seminal papers of Cramér (1938, [10]) and Bernstein (1946, [7]), the estimation of the tail probabilities  $\mathbb{P}(X_n > x)$  for positive  $x$  has attracted much attention. We would like to mention here the celebrated Bennett inequality (1962, cf. (8b) of [2], see also Hoeffding [18]) which states that, for independent and centered random variables  $\xi_i$  satisfying  $\xi_i \leq 1$  and for any  $t > 0$ ,

$$\mathbb{P}(X_n \geq nt) \leq \left( \frac{\sigma^2}{t + \sigma^2} \right)^{n(t + \sigma^2)} e^{nt} \quad (1)$$

$$= \exp \left\{ -nt \left[ \left( 1 + \frac{\sigma^2}{t} \right) \log \left( 1 + \frac{t}{\sigma^2} \right) - 1 \right] \right\}, \quad (2)$$

---

\*Corresponding author.

*E-mail:* fanxiequan@hotmail.com (X. Fan), ion.grama@univ-ubs.fr (I. Grama),  
quansheng.liu@univ-ubs.fr (Q. Liu).

where  $\sigma^2 = \frac{1}{n} \sum_{i=1}^n \sigma_i^2$ . Further, inequalities for the probabilities  $\mathbb{P}(X_n > x)$  have been obtained by Prohorov (1959, [27]), Hoeffding (1963, [18]), Azuma (1967, [1]), Steiger (1967, [29]; 1969, [30]), Freedman (1975, [14]), Nagaev (1979, [22]), Haeusler (1984, [17]), McDiarmid (1989, [21]), Pinelis (1994, [24]), Talagrand (1995, [31]), De La Peña (1999, [11]), Lesigne and Volný (2001, [19]), Nagaev (2003, [23]), Bentkus (2004, [4]), Pinelis (2006, [26]) and Bercu and Touati (2008, [6]) among others.

Most of these results were obtained by an approach based on the use of the exponential Markov's inequality. The challenge for this method is to find a sharp upper bound of the moment generating function  $\phi_i(\lambda) = \mathbb{E}(e^{\lambda \xi_i})$ . Hoeffding [18], Azuma [1] and McDiarmid [21] used the elementary estimation  $\phi_i(\lambda) \leq e^{\lambda^2/2}$ ,  $\lambda \geq 0$ , which holds if  $|\xi_i| \leq 1$ . Better results can be obtained by the following improvement  $\phi_i(\lambda) \leq (e^\lambda - 1 - \lambda)\sigma_i^2$ ,  $\lambda \geq 0$ , which holds for  $\xi_i \leq 1$  (see for example Freedman [14]). Bennett [2] and Hoeffding [18] used a more precise estimation

$$\phi_i(\lambda) \leq \frac{1}{1 + \sigma_i^2} \exp\{-\lambda\sigma_i^2\} + \frac{\sigma_i^2}{1 + \sigma_i^2} \exp\{\lambda\}, \quad \lambda \geq 0, \quad (3)$$

for any  $\xi_i$  satisfying  $\xi_i \leq 1$ . Bennett's estimation (3) is sharp with the equality attained when  $\mathbb{P}(\xi_i = 1) = \frac{\sigma_i^2}{1 + \sigma_i^2}$  and  $\mathbb{P}(\xi_i = -\sigma_i^2) = \frac{1}{1 + \sigma_i^2}$ .

Using (3), Hoeffding improved Bennett's inequality (1) and obtained the following inequality: for independent and centered random variables  $(\xi_i)_{i=1, \dots, n}$  satisfying  $\xi_i \leq 1$  and for any  $0 < t < 1$ ,

$$\mathbb{P}(X_n \geq nt) \leq \left\{ \left(1 + \frac{t}{\sigma^2}\right)^{-\frac{t + \sigma^2}{1 + \sigma^2}} (1 - t)^{-\frac{1 - t}{1 + \sigma^2}} \right\}^n, \quad (4)$$

where  $\sigma^2 = \frac{1}{n} \sum_{i=1}^n \sigma_i^2$  (cf. (2.8) of [18]).

It turns out that, under the stated conditions, Hoeffding's inequality (4) is very tight and improving (4) is a rather difficult task. Significant advances in improving Hoeffding's and Bennett's inequalities have been obtained by several authors. For instance Eaton [13], Pinelis [25] and Talagrand [31] have added to (4) a missing factor of the order  $\frac{\sigma}{\sqrt{n}t}$ . Improvements of the Bennett's inequality (1) can be found in Pinelis [26], where some larger classes of functions are considered instead of the class of exponential functions usually used in Markov's inequality. When  $\xi_i$  are martingale differences,

Bentkus [4] showed that if the conditional variances of  $\xi_i$  are bounded, then  $\mathbb{P}(X_n \geq x) \leq c \mathbb{P}(\sum_{i=1}^n \eta_i \geq x)$ , where  $\eta_i$  are independent and identically distributed Rademacher random variables,  $c = e^2/2 = 3.694\dots$  and  $x$  is a real such that  $\mathbb{P}(\sum_{i=1}^n \eta_i \geq x)$  has a jump at  $x$  (see also [5] for related results). However, to the best of our knowledge, there is no martingale or supermartingale version which reduces exactly to the Hoeffding inequality (4) in the independent case.

The scope of the paper is to extend the Hoeffding inequality (4) to the case of martingales and supermartingales. Our inequality will recover (4) in the independent case, and in the case of (super)martingales will apply under a very weak constraint on the sum of conditional variances.

The main results of the paper are the following inequalities (see Theorem 2.1 and Remark 2.1). Assume that  $(\xi_i, \mathcal{F}_i)_{i=1, \dots, n}$  are supermartingale differences satisfying  $\xi_i \leq 1$ . Denote by  $\langle X \rangle_k = \sum_{i=1}^k \mathbb{E}(\xi_i^2 | \mathcal{F}_{i-1})$  for  $k = 1, \dots, n$ . Then, for any  $x \geq 0$  and  $v > 0$ ,

$$\begin{aligned} & \mathbb{P}(X_k \geq x \text{ and } \langle X \rangle_k \leq v^2 \text{ for some } k \in [1, n]) \\ & \leq \left\{ \left( \frac{v^2}{x + v^2} \right)^{x+v^2} \left( \frac{n}{n-x} \right)^{n-x} \right\}^{\frac{n}{n+v^2}} \mathbf{1}_{\{x \leq n\}} \end{aligned} \quad (5)$$

$$\leq \exp \left\{ -\frac{x^2}{2(v^2 + \frac{1}{3}x)} \right\}. \quad (6)$$

In the independent case, inequality (5) with  $x = nt$  and  $v^2 = n\sigma^2$  reduces to inequality (4). We will see that the inequalities (5) and (6) strengthen or extend many well-known inequalities obtained by Freedman, De La Peña, Bernstein, Prohorov, Bennett, Hoeffding, Azuma, Nagaev and Haeusler. In particular, if the martingale differences  $(\xi_i, \mathcal{F}_i)_{i=1, \dots, n}$  satisfy  $-b \leq \xi_i \leq 1$  for some constant  $b > 0$ , then we get (see Corollary 2.1), for all  $x \geq 0$ ,

$$\mathbb{P} \left( \max_{1 \leq k \leq n} X_k \geq x \right) \leq \exp \left\{ -\frac{2x^2}{U_n(x, b)} \right\}, \quad (7)$$

where

$$U_n(x, b) = \min \left\{ n(1+b)^2, 4 \left( nb + \frac{1}{3}x \right) \right\}.$$

Notice that inequality (7) is sharper than the usual Azuma-Hoeffding inequality when  $0 < x < \frac{3}{4}n(1-b)^2$ .

Our approach is based on the conjugate distribution technique due to Cramér, and is different from the method used in Hoeffding's original paper [18]. The technique has been developed in Grama and Haeusler [16] to obtain expansions of large deviation for martingales. We refine this technique to get precise upper bounds for tail probabilities, providing a simple and unified approach for improving several well-known inequalities. We also make clear some relations among these inequalities.

Our main results will be presented in Section 2 and proved in Sections 3 and 4.

## 2. Main Results

Assume that we are given a sequence of real supermartingale differences  $(\xi_i, \mathcal{F}_i)_{i=0, \dots, n}$ , defined on some probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ , where  $\xi_0 = 0$  and  $\{\emptyset, \Omega\} = \mathcal{F}_0 \subseteq \dots \subseteq \mathcal{F}_n \subseteq \mathcal{F}$  are increasing  $\sigma$ -fields. So, by definition, we have  $\mathbb{E}(\xi_i | \mathcal{F}_{i-1}) \leq 0$ ,  $i = 1, \dots, n$ . Set

$$X_0 = 0, \quad X_k = \sum_{i=1}^k \xi_i, \quad k = 1, \dots, n. \quad (8)$$

Let  $\langle X \rangle$  be the quadratic characteristic of the supermartingale  $X = (X_k, \mathcal{F}_k)$ :

$$\langle X \rangle_0 = 0, \quad \langle X \rangle_k = \sum_{i=1}^k \mathbb{E}(\xi_i^2 | \mathcal{F}_{i-1}), \quad k = 1, \dots, n. \quad (9)$$

**Theorem 2.1.** *Assume that  $(\xi_i, \mathcal{F}_i)_{i=1, \dots, n}$  are supermartingale differences satisfying  $\xi_i \leq 1$ . Then, for any  $x \geq 0$  and  $v > 0$ ,*

$$\mathbb{P}(X_k \geq x \text{ and } \langle X \rangle_k \leq v^2 \text{ for some } k \in [1, n]) \leq H_n(x, v), \quad (10)$$

where

$$H_n(x, v) = \left\{ \left( \frac{v^2}{x + v^2} \right)^{x+v^2} \left( \frac{n}{n-x} \right)^{n-x} \right\}^{\frac{n}{n+v^2}} \mathbf{1}_{\{x \leq n\}}$$

with the convention that  $(+\infty)^0 = 1$  (which applies when  $x = n$ ).

Because of the obvious inequalities

$$\mathbb{P}(X_n \geq x, \langle X \rangle_n \leq v^2) \quad (11)$$

$$\leq \mathbb{P}\left(\max_{1 \leq k \leq n} X_k \geq x, \langle X \rangle_n \leq v^2\right) \quad (12)$$

$$\leq \mathbb{P}(X_k \geq x \text{ and } \langle X \rangle_k \leq v^2 \text{ for some } k \in [1, n]),$$

the function  $H_n(x, v)$  is also an upper bound of the tail probabilities (11) and (12). Therefore Theorem 2.1 extends Hoeffding's inequality (4) to the case of supermartingales with differences  $\xi_i$  satisfying  $\xi_i \leq 1$ .

The following remark establishes some relations among the well-known bounds of Hoeffding, Freedman, Bennett, Bernstein and De La Peña.

**Remark 2.1.** *For any  $x \geq 0$  and  $v > 0$ , it holds*

$$H_n(x, v) \leq F(x, v) =: \left(\frac{v^2}{x + v^2}\right)^{x+v^2} e^x \quad (13)$$

$$\leq B_1(x, v) =: \exp\left\{-\frac{x^2}{v^2\left(1 + \sqrt{1 + \frac{2x}{3v^2}}\right) + \frac{1}{3}x}\right\} \quad (14)$$

$$\leq B_2(x, v) =: \exp\left\{-\frac{x^2}{2(v^2 + \frac{1}{3}x)}\right\}. \quad (15)$$

Moreover, for any  $x, v > 0$ ,  $H_n(x, v)$  is increasing in  $n$  and

$$\lim_{n \rightarrow \infty} H_n(x, v) = F(x, v). \quad (16)$$

Since  $H_n(x, v) \leq F(x, v)$ , our inequality (10) implies Freedman's inequality for supermartingales [14]. The bounds  $B_1(x, v)$  and  $B_2(x, v)$  are respectively the bounds of Bennett and Bernstein (cf. [2], (8a) and [7]). Note that Bennett and Bernstein obtained their bounds for independent random variables under the Bernstein condition

$$\mathbb{E}|\xi_i|^k \leq \frac{1}{2} k! \left(\frac{1}{3}\right)^{k-2} \mathbb{E}\xi_i^2, \quad \text{for } k \geq 3. \quad (17)$$

We would like to point out that our condition  $\xi_i \leq 1$  does not imply Bernstein condition (17). The bounds  $B_1(x, v)$  and  $B_2(x, v)$  have also been obtained by

De La Peña ([11], (1.2)) for martingale differences  $\xi_i$  satisfying the conditional version of Bernstein's condition (17). Our result shows that the inequalities of Bennett ([2], (8a)), Bernstein [7] and De La Peña ([11], (1.2)) also hold when the (conditional) Bernstein condition is replaced by the condition  $\xi_i \leq 1$ . So Theorem 2.1 refines and completes the inequalities of Bennett, Bernstein and De La Peña for supermartingales with differences bounded from above.

It is interesting to note that from Theorem 2.1 and (14) it follows that

$$\mathbb{P}\left(X_k \geq \frac{x}{3} + v\sqrt{2x} \text{ and } \langle X \rangle_k \leq v^2 \text{ for some } k \in [1, n]\right) \leq e^{-x}, \quad (18)$$

which is another form of Bennett's inequalities (for related bounds we refer to Rio [28] and Bousquet [8]).

If the (super)martingale differences  $(\xi_i, \mathcal{F}_i)_{i=1, \dots, n}$  are in addition bounded from below, our inequality (10) also implies the inequalities (2.1) and (2.6) of Hoeffding [18] as seen from the following corollary.

**Corollary 2.1.** *Assume that  $(\xi_i, \mathcal{F}_i)_{i=1, \dots, n}$  are martingale differences satisfying  $-b \leq \xi_i \leq 1$  for some constant  $b > 0$ . Then, for any  $x \geq 0$ ,*

$$\mathbb{P}\left(\max_{1 \leq k \leq n} X_k \geq x\right) \leq H_n\left(x, \sqrt{nb}\right) \quad (19)$$

and

$$\mathbb{P}\left(\max_{1 \leq k \leq n} X_k \geq x\right) \leq \exp\left\{-\frac{2x^2}{U_n(x, b)}\right\}, \quad (20)$$

where

$$U_n(x, b) = \min\left\{n(1+b)^2, 4\left(nb + \frac{1}{3}x\right)\right\}.$$

The inequalities (19) and (20) remain true for supermartingale differences  $(\xi_i, \mathcal{F}_i)_{i=1, \dots, n}$  satisfying  $-b \leq \xi_i \leq 1$  for some constant  $0 < b \leq 1$ .

In the martingale case, our inequality (19) is a refined version of the inequality (2.1) of Hoeffding [18] in the sense that  $X_n$  is replaced by  $\max_{1 \leq k \leq n} X_k$ . When  $U_n(x, b) = n(1+b)^2$ , inequality (20) is a refined version of the usual Azuma-Hoeffding inequality (cf. [18], (2.6)); when  $0 < x < \frac{3}{4}n(1-b)^2$ , our inequality (20) is sharper than the Azuma-Hoeffding inequality. Related results can be found in Steiger [29], [30], McDiarmid [21], Pinelis [26] and Bentkus [4], [5].

The following result extends an inequality of De La Peña ([11], (1.15)).

**Corollary 2.2.** *Assume that  $(\xi_i, \mathcal{F}_i)_{i=1, \dots, n}$  are supermartingale differences satisfying  $\xi_i \leq 1$ . Then, for any  $x \geq 0$  and  $v > 0$ ,*

$$\begin{aligned} & \mathbb{P}(X_k \geq x \text{ and } \langle X \rangle_k \leq v^2 \text{ for some } k \in [1, n]) \\ & \leq \exp \left\{ -\frac{x}{2} \operatorname{arcsinh} \left( \frac{x}{2v^2} \right) \right\}. \end{aligned} \quad (21)$$

De La Peña [11] obtained the same inequality (21) for martingale differences  $(\xi_i, \mathcal{F}_i)_{i=1, \dots, n}$  under the more restrictive condition that  $|\xi_i| \leq c$  for some constant  $0 < c < \infty$ . In the independent case, the bound in (21) is the Prohorov bound [27]. As was remarked by Hoeffding [18], the right side of (10) is less than the right side of (21). So inequality (10) implies inequality (21).

For unbounded supermartingale differences  $(\xi_i, \mathcal{F}_i)_{i=1, \dots, n}$ , we have the following inequality.

**Corollary 2.3.** *Assume that  $(\xi_i, \mathcal{F}_i)_{i=1, \dots, n}$  are supermartingale differences. Let  $y > 0$  and*

$$V_k^2(y) = \sum_{i=1}^k \mathbb{E}(\xi_i^2 \mathbf{1}_{\{\xi_i \leq y\}} | \mathcal{F}_{i-1}), \quad k = 1, \dots, n. \quad (22)$$

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

$$\begin{aligned} & \mathbb{P}(X_k \geq x \text{ and } V_k^2(y) \leq v^2 \text{ for some } k \in [1, n]) \\ & \leq H_n \left( \frac{x}{y}, \frac{v}{y} \right) + \mathbb{P} \left( \max_{1 \leq i \leq n} \xi_i > y \right). \end{aligned} \quad (23)$$

We notice that inequality (23) improves an inequality of Fuk ([15], (3)). It also extends and improves Nagaev's inequality ([22], (1.55)) which was obtained in the independent case.

Since  $\mathbb{P}(V_n^2(y) > v^2) \leq P(\langle X \rangle_n > v^2)$  and  $H_n(x, v) \leq F(x, v)$ , Corollary 2.3 implies the following inequality due to Courbot [9]:

$$\mathbb{P} \left( \max_{1 \leq k \leq n} X_k \geq x \right) \leq F \left( \frac{x}{y}, \frac{v}{y} \right) + \sum_{i=1}^n \mathbb{P}(\xi_i > y) + \mathbb{P}(\langle X \rangle_n > v^2). \quad (24)$$

A slightly weaker inequality was obtained earlier by Haeusler [17]: in Haeusler's inequality  $F \left( \frac{x}{y}, \frac{v}{y} \right)$  is replaced by a larger bound  $\exp \left\{ \frac{x}{y} \left( 1 - \log \frac{xy}{v^2} \right) \right\}$ . Thus, inequality (23) improves Courbot's and Haeusler's inequalities.



To close this section, we present an extension of the inequalities of Freedman and Bennett under the condition that

$$\mathbb{E}(\xi_i^2 e^{\lambda \xi_i} | \mathcal{F}_{i-1}) \leq e^\lambda E(\xi_i^2 | \mathcal{F}_{i-1}), \quad \text{for any } \lambda \geq 0, \quad (25)$$

which is weaker than the assumption  $\xi_i \leq 1$  used before.

**Theorem 2.2.** *Assume that  $(\xi_i, \mathcal{F}_i)_{i=1, \dots, n}$  are martingale differences satisfying (25). Then, for any  $x \geq 0$  and  $v > 0$ ,*

$$\mathbb{P}(X_k \geq x \text{ and } \langle X \rangle_k \leq v^2 \text{ for some } k \in [1, n]) \leq F(x, v). \quad (26)$$

Bennett [2] proved (26) in the independent case under the condition that

$$\mathbb{E}|\xi_i|^k \leq \mathbb{E}\xi_i^2, \quad \text{for any } k \geq 3,$$

which is in fact equivalent to  $|\xi_i| \leq 1$ . Taking into account Remark 2.1, we see that (26) recovers the inequalities of Freedman and Bennett under the less restrictive condition (25).

### 3. Proof of Theorems

Let  $(\xi_i, \mathcal{F}_i)_{i=0, \dots, n}$  be the supermartingale differences introduced in the previous section and  $X = (X_k, \mathcal{F}_k)_{k=0, \dots, n}$  be the corresponding supermartingale defined by (8). For any nonnegative number  $\lambda$ , define the *exponential multiplicative martingale*  $Z(\lambda) = (Z_k(\lambda), \mathcal{F}_k)_{k=0, \dots, n}$ , where

$$Z_k(\lambda) = \prod_{i=1}^k \frac{e^{\lambda \xi_i}}{\mathbb{E}(e^{\lambda \xi_i} | \mathcal{F}_{i-1})}, \quad Z_0(\lambda) = 1, \quad \lambda \geq 0.$$

If  $T$  is a stopping time, then  $Z_{T \wedge k}(\lambda)$  is also a martingale, where

$$Z_{T \wedge k}(\lambda) = \prod_{i=1}^{T \wedge k} \frac{e^{\lambda \xi_i}}{\mathbb{E}(e^{\lambda \xi_i} | \mathcal{F}_{i-1})}, \quad Z_0(\lambda) = 1, \quad \lambda \geq 0.$$

Thus, for each nonnegative number  $\lambda$  and each  $k = 1, \dots, n$ , the random variable  $Z_{T \wedge k}(\lambda)$  is a probability density on  $(\Omega, \mathcal{F}, \mathbb{P})$ , i.e.

$$\int Z_{T \wedge k}(\lambda) d\mathbb{P} = \mathbb{E}(Z_{T \wedge k}(\lambda)) = 1.$$

The last observation allows us to introduce, for any nonnegative number  $\lambda$ , the *conjugate probability measure*  $\mathbb{P}_\lambda$  on  $(\Omega, \mathcal{F})$  defined by

$$d\mathbb{P}_\lambda = Z_{T \wedge n}(\lambda) d\mathbb{P}. \quad (27)$$

Throughout the paper, we denote by  $\mathbb{E}_\lambda$  the expectation with respect to  $\mathbb{P}_\lambda$ .

Consider the predictable process  $\Psi(\lambda) = (\Psi_k(\lambda), \mathcal{F}_k)_{k=0, \dots, n}$ , which is called the *cumulant process* and which is related to the supermartingale  $X$  as follows:

$$\Psi_k(\lambda) = \sum_{i=1}^k \log \mathbb{E}(e^{\lambda \xi_i} | \mathcal{F}_{i-1}), \quad 0 \leq k \leq n. \quad (28)$$

We should give a sharp bound for the function  $\Psi_k(\lambda)$ . To this end, we need the following elementary lemma which, in the special case of centered random variables, has been proved by Bennett [2].

**Lemma 3.1.** *If  $\xi$  is a random variable such that  $\xi \leq 1$ ,  $\mathbb{E}\xi \leq 0$  and  $\mathbb{E}\xi^2 = \sigma^2$ , then, for any  $\lambda \geq 0$ ,*

$$\mathbb{E}(e^{\lambda \xi}) \leq \frac{1}{1 + \sigma^2} \exp\{-\lambda \sigma^2\} + \frac{\sigma^2}{1 + \sigma^2} \exp\{\lambda\}. \quad (29)$$

**Proof.** We argue as in Bennett [2]. For  $\lambda = 0$ , inequality (29) is obvious. Fix  $\lambda > 0$  and consider the function

$$\phi(\xi) = a\xi^2 + b\xi + c, \quad \xi \leq 1,$$

where  $a$ ,  $b$  and  $c$  are determined by the conditions

$$\phi(1) = e^\lambda, \quad \phi(-\sigma^2) = \frac{1}{\lambda} \phi'(-\sigma^2) = \exp\{-\lambda \sigma^2\}, \quad \lambda > 0.$$

By simple calculations, we have

$$\begin{aligned} a &= \frac{e^\lambda - e^{-\lambda \sigma^2} - \lambda(1 + \sigma^2)e^{-\lambda \sigma^2}}{(1 + \sigma^2)^2}, \\ b &= \frac{\lambda(1 - \sigma^4)e^{-\lambda \sigma^2} + 2\sigma^2(e^\lambda - e^{-\lambda \sigma^2})}{(1 + \sigma^2)^2} \end{aligned}$$

and

$$c = \frac{\sigma^4 e^\lambda + (1 + 2\sigma^2 + \lambda \sigma^2 + \lambda \sigma^4) e^{-\lambda \sigma^2}}{(1 + \sigma^2)^2}.$$

We now prove that

$$e^{\lambda\xi} \leq \phi(\xi) \quad \text{for any } \xi \leq 1 \text{ and } \lambda > 0, \quad (30)$$

which will imply the assertion of the lemma. For any  $\xi \in \mathbb{R}$ , set

$$f(\xi) = \phi(\xi) - e^{\lambda\xi}.$$

Since  $f(-\sigma^2) = f(1) = 0$ , by Rolle's theorem, there exists some  $\xi_1 \in (-\sigma^2, 1)$  such that  $f'(\xi_1) = 0$ . In the same way, since  $f'(-\sigma^2) = 0$  and  $f'(\xi_1) = 0$ , there exists some  $\xi_2 \in (-\sigma^2, \xi_1)$  such that  $f''(\xi_2) = 0$ . Taking into account that the function  $f''(\xi) = 2a - \lambda^2 e^{\lambda\xi}$  is strictly decreasing, we conclude that  $\xi_2$  is the unique zero point of  $f''(\xi)$ . It follows that  $f(\xi)$  is convex on  $(-\infty, \xi_2]$  and concave on  $[\xi_2, 1]$ , with  $\min_{(-\infty, \xi_2]} f(\xi) = f(-\sigma^2) = 0$  and  $\min_{[\xi_2, 1]} f(\xi) = f(1) = 0$ . Therefore  $\min_{(-\infty, 1]} f(\xi) = 0$ , which implies (30).

Since  $b \geq 0$  and  $\mathbb{E}\xi \leq 0$ , from (30), it follows that, for any  $\lambda > 0$ ,

$$\mathbb{E}(e^{\lambda\xi}) \leq a\sigma^2 + c = \frac{1}{1 + \sigma^2} \exp\{-\lambda\sigma^2\} + \frac{\sigma^2}{1 + \sigma^2} \exp\{\lambda\}.$$

This completes the proof of Lemma 3.1.  $\square$

The following technical lemma is from Hoeffding [18] (see Lemma 3 therein and its proof). For reader's convenience, we shall give a proof following [18].

**Lemma 3.2.** *For any  $\lambda \geq 0$  and  $t \geq 0$ , let*

$$f(\lambda, t) = \log \left( \frac{1}{1+t} \exp\{-\lambda t\} + \frac{t}{1+t} \exp\{\lambda\} \right). \quad (31)$$

*Then  $\frac{\partial}{\partial t} f(\lambda, t) > 0$  and  $\frac{\partial^2}{\partial t^2} f(\lambda, t) < 0$  for any  $\lambda > 0$  and  $t \geq 0$ .*

**Proof.** Denote

$$g(y) = \frac{e^{-\lambda y} + y - 1}{y}, \quad y \geq 1.$$

Then  $f(\lambda, t) = \lambda + \log g(1+t)$ . By straightforward calculation, we have, for any  $y \geq 1$ ,

$$g'(y) = \frac{e^{-\lambda y}(e^{\lambda y} - 1 - \lambda y)}{y^2} > 0$$

and

$$g''(y) = -\frac{2e^{-\lambda y}}{y^3} \left( e^{\lambda y} - 1 - \lambda y - \frac{\lambda^2 y^2}{2} \right) < 0.$$

Since  $g(y) > 0$  for  $y = t + 1 \geq 1$ ,

$$\frac{\partial}{\partial t} f(\lambda, t) = \frac{g'(y)}{g(y)} \quad \text{and} \quad \frac{\partial^2}{\partial t^2} f(\lambda, t) = \frac{g''(y)g(y) - g'(y)^2}{g(y)^2},$$

it follows that  $\frac{\partial}{\partial t} f(\lambda, t) > 0$  and  $\frac{\partial^2}{\partial t^2} f(\lambda, t) < 0$  for all  $\lambda, t > 0$ .  $\square$

**Lemma 3.3.** *Assume that  $(\xi_i, \mathcal{F}_i)_{i=1, \dots, n}$  are supermartingale differences satisfying  $\xi_i \leq 1$ . Then, for any  $\lambda \geq 0$  and  $k = 1, \dots, n$ ,*

$$\Psi_k(\lambda) \leq kf \left( \lambda, \frac{\langle X \rangle_k}{k} \right). \quad (32)$$

**Proof.** For  $\lambda = 0$ , inequality (32) is obvious. By Lemma 3.1, we have, for any  $\lambda > 0$ ,

$$\mathbb{E}(e^{\lambda \xi_i} | \mathcal{F}_{i-1}) \leq \frac{\exp \{-\lambda \mathbb{E}(\xi_i^2 | \mathcal{F}_{i-1})\}}{1 + \mathbb{E}(\xi_i^2 | \mathcal{F}_{i-1})} + \frac{\mathbb{E}(\xi_i^2 | \mathcal{F}_{i-1})}{1 + \mathbb{E}(\xi_i^2 | \mathcal{F}_{i-1})} \exp\{\lambda\}.$$

Therefore, using (31) with  $t = \mathbb{E}(\xi_i^2 | \mathcal{F}_{i-1})$ , we get

$$\log \mathbb{E}(e^{\lambda \xi_i} | \mathcal{F}_{i-1}) \leq f(\lambda, \mathbb{E}(\xi_i^2 | \mathcal{F}_{i-1})). \quad (33)$$

By Lemma 3.2, for fixed  $\lambda > 0$ , the function  $f(\lambda, t)$  has a negative second derivative in  $t$ . Hence,  $f(\lambda, t)$  is concave in  $t \geq 0$ , and therefore, by Jensen's inequality,

$$\sum_{i=1}^k f(\lambda, \mathbb{E}(\xi_i^2 | \mathcal{F}_{i-1})) = k \sum_{i=1}^k \frac{1}{k} f(\lambda, \mathbb{E}(\xi_i^2 | \mathcal{F}_{i-1})) \leq kf \left( \lambda, \frac{\langle X \rangle_k}{k} \right). \quad (34)$$

Combining (33) and (34), we obtain

$$\Psi_k(\lambda) = \sum_{i=1}^k \log \mathbb{E}(e^{\lambda \xi_i} | \mathcal{F}_{i-1}) \leq kf \left( \lambda, \frac{\langle X \rangle_k}{k} \right).$$

This completes the proof of Lemma 3.3.  $\square$

**Proof of Theorem 2.1.** For any  $0 \leq x \leq n$ , define the stopping time

$$T(x) = \min\{k \in [1, n] : X_k \geq x \text{ and } \langle X \rangle_k \leq v^2\}, \quad (35)$$

with the convention that  $\min\{\emptyset\} = 0$ . Then

$$\mathbf{1}\{X_k \geq x \text{ and } \langle X \rangle_k \leq v^2 \text{ for some } k \in [1, n]\} = \sum_{k=1}^n \mathbf{1}\{T(x) = k\}.$$

Using the change of measure (27), we have, for any  $0 \leq x \leq n$ ,  $v > 0$  and  $\lambda \geq 0$ ,

$$\begin{aligned} & \mathbb{P}(X_k \geq x \text{ and } \langle X \rangle_k \leq v^2 \text{ for some } k \in [1, n]) \\ &= \mathbb{E}_\lambda Z_{T \wedge n}(\lambda)^{-1} \mathbf{1}\{X_k \geq x \text{ and } \langle X \rangle_k \leq v^2 \text{ for some } k \in [1, n]\} \\ &= \sum_{k=1}^n \mathbb{E}_\lambda \exp\{-\lambda X_{T \wedge n} + \Psi_{T \wedge n}(\lambda)\} \mathbf{1}\{T(x)=k\} \\ &= \sum_{k=1}^n \mathbb{E}_\lambda \exp\{-\lambda X_k + \Psi_k(\lambda)\} \mathbf{1}\{T(x)=k\} \\ &\leq \sum_{k=1}^n \mathbb{E}_\lambda \exp\{-\lambda x + \Psi_k(\lambda)\} \mathbf{1}\{T(x)=k\}. \end{aligned} \quad (36)$$

Using Lemma 3.3, we deduce, for any  $0 \leq x \leq n$ ,  $v > 0$  and  $\lambda \geq 0$ ,

$$\begin{aligned} & \mathbb{P}(X_k \geq x \text{ and } \langle X \rangle_k \leq v^2 \text{ for some } k \in [1, n]) \\ &\leq \sum_{k=1}^n \mathbb{E}_\lambda \exp\left\{-\lambda x + kf\left(\lambda, \frac{\langle X \rangle_k}{k}\right)\right\} \mathbf{1}\{T(x)=k\}. \end{aligned} \quad (37)$$

By Lemma 3.2,  $f(\lambda, t)$  is increasing in  $t \geq 0$ . Therefore

$$\begin{aligned} & \mathbb{P}(X_k \geq x \text{ and } \langle X \rangle_k \leq v^2 \text{ for some } k \in [1, n]) \\ &\leq \sum_{k=1}^n \mathbb{E}_\lambda \exp\left\{-\lambda x + kf\left(\lambda, \frac{v^2}{k}\right)\right\} \mathbf{1}\{T(x)=k\}. \end{aligned} \quad (38)$$

As  $f(\lambda, 0) = 0$  and  $f(\lambda, t)$  is concave in  $t \geq 0$  (see Lemma 3.2), the function  $f(\lambda, t)/t$  is decreasing in  $t \geq 0$  for any  $\lambda \geq 0$ . Hence, we have, for any

$0 \leq x \leq n$ ,  $v > 0$  and  $\lambda \geq 0$ ,

$$\begin{aligned}
& \mathbb{P}(X_k \geq x \text{ and } \langle X \rangle_k \leq v^2 \text{ for some } k \in [1, n]) \\
& \leq \exp \left\{ -\lambda x + n f \left( \lambda, \frac{v^2}{n} \right) \right\} \mathbb{E}_\lambda \sum_{k=1}^n \mathbf{1}_{\{T(x)=k\}} \\
& \leq \exp \left\{ -\lambda x + n f \left( \lambda, \frac{v^2}{n} \right) \right\}.
\end{aligned} \tag{39}$$

Since the function in (39) attains its minimum at

$$\lambda = \lambda(x) = \frac{1}{1 + v^2/n} \log \frac{1 + x/v^2}{1 - x/n}, \tag{40}$$

inserting  $\lambda = \lambda(x)$  in (39), we obtain, for any  $0 \leq x \leq n$  and  $v > 0$ ,

$$\mathbb{P} \left( X_k \geq x \text{ and } \langle X \rangle_k \leq v^2 \text{ for some } k \in [1, n] \right) \leq H_n(x, v),$$

where

$$H_n(x, v) = \inf_{\lambda \geq 0} \exp \left\{ -\lambda x + n f \left( \lambda, \frac{v^2}{n} \right) \right\}, \tag{41}$$

which gives the bound (10).  $\square$

**Proof of Remark 2.1.** We will use the function  $f(\lambda, t)$  defined by (31). Since  $\frac{\partial^2}{\partial t^2} f(\lambda, t) \leq 0$  for any  $t \geq 0$  and  $\lambda \geq 0$ , it holds

$$f(\lambda, t) \leq f(\lambda, 0) + \frac{\partial}{\partial t} f(\lambda, 0) t = (e^\lambda - 1 - \lambda) t, \quad t, \lambda \geq 0. \tag{42}$$

Hence, using (41), for any  $x \geq 0$  and  $v > 0$ ,

$$\begin{aligned}
H_n(x, v) & \leq \inf_{\lambda \geq 0} \exp \left\{ -\lambda x + (e^\lambda - 1 - \lambda) v^2 \right\} \\
& = \left( \frac{v^2}{x + v^2} \right)^{x+v^2} e^x,
\end{aligned} \tag{43}$$

which proves (13). Using the inequality

$$(e^\lambda - 1 - \lambda) v^2 \leq \frac{\lambda^2 v^2}{2(1 - \frac{1}{3}\lambda)}, \quad \text{for any } \lambda, v \geq 0,$$

we get, for any  $x \geq 0$  and  $v > 0$ ,

$$\begin{aligned} \left(\frac{v^2}{x+v^2}\right)^{x+v^2} e^x &\leq \inf_{3>\lambda\geq 0} \exp\left\{-\lambda x + \frac{\lambda^2 v^2}{2(1-\frac{1}{3}\lambda)}\right\} \\ &= \exp\left\{-\frac{x^2}{v^2\left(1+\sqrt{1+\frac{2x}{3v^2}}\right)+\frac{1}{3}x}\right\} \\ &\leq \exp\left\{-\frac{x^2}{2(v^2+\frac{1}{3}x)}\right\}, \end{aligned}$$

where the last inequality follows from the fact  $\sqrt{1+\frac{2x}{3v^2}} \leq 1+\frac{x}{3v^2}$ . This proves (14) and (15).

Since  $f(\lambda, t)/t$  is decreasing in  $t \geq 0$  for any  $\lambda \geq 0$ , from (41), we find that  $H_n(x, v)$  is increasing in  $n$ . Taking into account that  $\lim_{n \rightarrow \infty} \left(\frac{n}{n-x}\right)^{n-x} = e^x$ , we obtain (16). This completes the proof of Remark 2.1.  $\square$

**Lemma 3.4.** *Assume that  $(\xi_i, \mathcal{F}_i)_{i=1, \dots, n}$  are martingale differences satisfying*

$$\mathbb{E}(\xi_i^2 e^{\lambda \xi_i} | \mathcal{F}_{i-1}) \leq e^\lambda \mathbb{E}(\xi_i^2 | \mathcal{F}_{i-1})$$

for any  $\lambda \geq 0$ . Then, for any  $\lambda \geq 0$  and  $k = 1, \dots, n$ ,

$$\Psi_k(\lambda) \leq (e^\lambda - 1 - \lambda) \langle X \rangle_k.$$

**Proof.** Denote  $\psi_i(\lambda) = \log E(e^{\lambda \xi_i} | \mathcal{F}_{i-1})$ ,  $\lambda \geq 0$ . Since  $\psi_i(0) = 0$  and  $\psi_i'(0) = E(\xi_i | \mathcal{F}_{i-1}) = 0$ , by Leibniz-Newton formula, it holds

$$\psi_i(\lambda) = \int_0^\lambda \psi_i'(y) dy = \int_0^\lambda \int_0^y \psi_i''(t) dt dy.$$

Therefore for any  $\lambda \geq 0$  and  $k = 1, \dots, n$ ,

$$\Psi_k(\lambda) = \sum_{i=1}^k \psi_i(\lambda) = \sum_{i=1}^k \int_0^\lambda \int_0^y \psi_i''(t) dt dy. \quad (44)$$

Since, by Jensen's inequality,  $\mathbb{E}(e^{t\xi_i} | \mathcal{F}_{i-1}) \geq 1$ , we get, for any  $t \geq 0$ ,

$$\begin{aligned} \psi_i''(t) &= \frac{\mathbb{E}(\xi_i^2 e^{t\xi_i} | \mathcal{F}_{i-1})}{\mathbb{E}(e^{t\xi_i} | \mathcal{F}_{i-1})} - \frac{\mathbb{E}(\xi_i e^{t\xi_i} | \mathcal{F}_{i-1})^2}{\mathbb{E}(e^{t\xi_i} | \mathcal{F}_{i-1})^2} \\ &\leq \mathbb{E}(\xi_i^2 e^{t\xi_i} | \mathcal{F}_{i-1}) \\ &\leq e^t \mathbb{E}(\xi_i^2 | \mathcal{F}_{i-1}). \end{aligned} \quad (45)$$

Inserting (45) into (44), we obtain

$$\begin{aligned}\Psi_k(\lambda) &\leq \sum_{i=1}^k \int_0^\lambda \int_0^y e^t \mathbb{E}(\xi_i^2 | \mathcal{F}_{i-1}) dt dy \\ &= (e^\lambda - 1 - \lambda) \langle X \rangle_k.\end{aligned}$$

This completes the proof of Lemma 3.4.  $\square$

**Proof of Theorem 2.2.** By (36) and Lemma 3.4, we obtain, for any  $x \geq 0$ ,  $v > 0$  and  $\lambda \geq 0$ ,

$$\begin{aligned}&\mathbb{P}(X_k \geq x \text{ and } \langle X \rangle_k \leq v^2 \text{ for some } k \in [1, n]) \\ &\leq \sum_{k=1}^n \mathbb{E}_\lambda \exp \{-\lambda x + \Psi_k(\lambda)\} \mathbf{1}_{\{T(x)=k\}} \\ &\leq \sum_{k=1}^n \mathbb{E}_\lambda \exp \{-\lambda x + (e^\lambda - 1 - \lambda) \langle X \rangle_k\} \mathbf{1}_{\{T(x)=k\}} \\ &\leq \sum_{k=1}^n \mathbb{E}_\lambda \exp \{-\lambda x + (e^\lambda - 1 - \lambda) v^2\} \mathbf{1}_{\{T(x)=k\}} \\ &\leq \exp \{-\lambda x + (e^\lambda - 1 - \lambda) v^2\}.\end{aligned}\tag{46}$$

Since the function in (46) attains its minimum at

$$\lambda = \lambda(x) = \log \left( 1 + \frac{x}{v^2} \right),\tag{47}$$

inserting  $\lambda = \lambda(x)$  in (46), we have, for any  $x \geq 0$  and  $v > 0$ ,

$$\begin{aligned}&\mathbb{P}(X_k \geq x \text{ and } \langle X \rangle_k \leq v^2 \text{ for some } k \in [1, n]) \\ &\leq F(x, v) = \left( \frac{v^2}{x + v^2} \right)^{x+v^2} e^x.\end{aligned}$$

This completes the proof of Theorem 2.2.  $\square$

#### 4. Proof of Corollaries

We use Theorem 2.1 to prove Corollaries 2.1 and 2.3.



**Proof of Corollary 2.1.** As  $-b \leq \xi_i \leq 1$ , we have  $-\xi_i \leq b$  and  $1-\xi_i \geq 0$ , so that  $-\xi_i(1-\xi_i) \leq b(1-\xi_i)$ . When  $\mathbb{E}(\xi_i|\mathcal{F}_{i-1}) = 0$  or  $\mathbb{E}(\xi_i|\mathcal{F}_{i-1}) \leq 0$  and  $0 < b \leq 1$ , it follows that

$$\begin{aligned}\mathbb{E}(\xi_i^2|\mathcal{F}_{i-1}) &= \mathbb{E}(-\xi_i(1-\xi_i)|\mathcal{F}_{i-1}) + \mathbb{E}(\xi_i|\mathcal{F}_{i-1}) \\ &\leq b + (1-b)\mathbb{E}(\xi_i|\mathcal{F}_{i-1}) \\ &\leq b.\end{aligned}\tag{48}$$

Therefore  $\langle X \rangle_n \leq nb$ . Hence, using Theorem 2.1, we have, for any  $0 \leq x \leq n$ ,

$$\begin{aligned}\mathbb{P}\left(\max_{1 \leq k \leq n} X_k \geq x\right) &\leq \sup_{v^2 \leq nb} \mathbb{P}\left(X_k \geq x \text{ and } \langle X \rangle_k \leq v^2 \text{ for some } k \in [1, n]\right) \\ &\leq \sup_{v^2 \leq nb} H_n(x, v) \\ &= H_n\left(x, \sqrt{nb}\right),\end{aligned}\tag{49}$$

which obtains inequality (19). Using (15), we get, for any  $x \geq 0$ ,

$$H_n\left(x, \sqrt{nb}\right) \leq \exp\left\{-\frac{x^2}{2(nb + \frac{1}{3}x)}\right\}.\tag{50}$$

From (41), we have, for any  $x \geq 0$ ,

$$H_n\left(x, \sqrt{nb}\right) = \inf_{\lambda \geq 0} \exp\{-\lambda x + nf(\lambda, b)\}.\tag{51}$$

With the notations  $z = \lambda(1+b)$  and  $p = \frac{b}{1+b}$ , we obtain

$$f(\lambda, b) = g(z) = -zp + \log(1 - p + pe^z).$$

Since  $g(0) = g'(0) = 0$ ,

$$g'(z) = -p + \frac{p}{p + (1-p)e^{-z}}$$

and

$$g''(z) = \frac{p(1-p)e^{-z}}{(p + (1-p)e^{-z})^2} \leq \frac{1}{4},$$

we have

$$f(\lambda, b) = g(z) \leq \frac{1}{8}z^2 = \frac{1}{8}\lambda^2(1+b)^2.$$

Returning to (51), we deduce, for any  $x \geq 0$ ,

$$\begin{aligned}
H_n(x, \sqrt{nb}) &= \inf_{\lambda \geq 0} \exp \{-\lambda x + nf(\lambda, b)\} \\
&\leq \inf_{\lambda \geq 0} \exp \left\{ -\lambda x + \frac{1}{8} \lambda^2 n (1+b)^2 \right\} \\
&= \exp \left\{ -\frac{2x^2}{n(1+b)^2} \right\}.
\end{aligned} \tag{52}$$

Combining (50) and (52), we obtain (20).  $\square$

**Proof of Corollary 2.3.** For  $y > 0$  and  $k = 1, \dots, n$ , set

$$\begin{aligned}
X'_k &= \sum_{i=1}^k \xi_i \mathbf{1}_{\{\xi_i \leq y\}}, & X''_k &= \sum_{i=1}^k \xi_i \mathbf{1}_{\{\xi_i > y\}}, \\
X_k &= X'_k + X''_k & \text{and} & & V_k^2(y) &= \sum_{i=1}^k \mathbb{E}(\xi_i^2 \mathbf{1}_{\{\xi_i \leq y\}} | \mathcal{F}_{i-1}).
\end{aligned}$$

Since  $E(\xi_i | \mathcal{F}_{i-1}) \leq 0$  implies  $E(\xi_i \mathbf{1}_{\{\xi_i \leq y\}} | \mathcal{F}_{i-1}) \leq 0$ ,  $X'_k$  is a sum of supermartingale differences. Now, for any  $y > 0$ ,  $0 \leq x \leq ny$  and  $v > 0$ ,

$$\begin{aligned}
&\mathbb{P}(X_k \geq x \text{ and } V_k^2(y) \leq v^2 \text{ for some } k \in [1, n]) \\
&= \mathbb{P}(X'_k + X''_k \geq x \text{ and } V_k^2(y) \leq v^2 \text{ for some } k \in [1, n]) \\
&\leq \mathbb{P}(X'_k \geq x \text{ and } V_k^2(y) \leq v^2 \text{ for some } k \in [1, n]) \\
&\quad + \mathbb{P}(X''_k > 0 \text{ and } V_k^2(y) \leq v^2 \text{ for some } k \in [1, n]) \\
&\leq \mathbb{P}\left(\frac{X'_k}{y} \geq \frac{x}{y} \text{ and } \frac{V_k^2(y)}{y^2} \leq \frac{v^2}{y^2} \text{ for some } k \in [1, n]\right) \\
&\quad + \mathbb{P}\left(\max_{1 \leq i \leq n} \xi_i > y\right).
\end{aligned}$$

Applying Theorem 2.1 to  $X_k = \frac{X'_k}{y}$ , we obtain Corollary 2.3.  $\square$

## Acknowledgements

We would like to thank the two referees for their helpful remarks and suggestions.

- [1] Azuma, K., 1967. Weighted sum of certain independent random variables. *Tohoku Math. J.* **19**, No. 3, 357–367.
- [2] Bennett, G., 1962. Probability inequalities for the sum of independent random variables, *J. Amer. Statist. Assoc.* **57**, No. 297, 33–45.
- [3] Bentkus, V., 1997. An Edgeworth expansion for symmetric statistics, *Ann. Statist.* **25**, No. 2, 851–896.
- [4] Bentkus, V., 2004. On Hoeffding’s inequality, *Ann. Probab.* **32**, No. 2, 1650–1673.
- [5] Bentkus, V., Kalosha, N. and van Zuijlen, M., 2006. On domination of tail probabilities of (super)martingales: explicit bounds, *Lithuanian Math. J.* **46**, No. 1, 1–43.
- [6] Bercu, B. and Touati, A., 2008. Exponential inequalities for self-normalized martingales with applications. *Ann. Appl. Probab.* **18**, 1848–1869.
- [7] Bernstein, S., 1946. *The Theory of Probabilities* (Russian). Moscow, Leningrad.
- [8] Bousquet, O., 2002. A Bennett concentration inequality and its application to suprema of empirical processes. *C. R. Acad. Sci. Paris sér. I* **334**, 495–500.
- [9] Courbot, B., 1999. Rates of convergence in the functional CLT for martingales. *C. R. Acad. Sci. Paris* **328**, 509–513.
- [10] Cramér, H., 1938. Sur un nouveau théorème-limite de la théorie des probabilités. *Actualite’s Sci. Indust.* **736**, 5–23.
- [11] De La Peña, V. H., 1999. A general class of exponential inequalities for martingales and ratios. *Ann. Probab.* **27**, No. 1, 537–564.
- [12] Dzhaparidze, K., and van Zanten, J.H., 2001. On Bernstein-type inequalities for martingales. *Stochastic Process. Appl.* **93**, 109–117.
- [13] Eaton, M. L., 1974. A probability inequality for linear combination of bounded random variables. *Ann. Statist.* **2**, No. 3, 609–614.

- [14] Freedman, D. A., 1975. On tail probabilities for martingales. *Ann. Probab.* **3**, No. 1, 100–118.
- [15] Fuk, D. X., 1973. Some probabilistic inequalities for martingales. *Siberian. Math. J.* **14**, No. 1, 185–193.
- [16] Grama, I. and Haeusler, E., 2000. Large deviations for martingales via Cramer’s method. *Stochastic Process. Appl.* **85**, 279–293.
- [17] Haeusler, E., 1984. An exact rate of convergence in the functional central limit theorem for special martingale difference arrays. *Probab. Theory Relat. Fields* **65**, No. 4, 523–534.
- [18] Hoeffding, W., 1963. Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58**, 13–30.
- [19] Lesigne, E. and Volný, D., 2001. Large deviations for martingales. *Stochastic Process. Appl.* **96**, 143–159.
- [20] Liu, Q. and Watbled, F., 2009. Exponential inequalities for martingales and asymptotic properties of the free energy of directed polymers in a random environment. *Stochastic Process. Appl.* **119**, 3101–3132.
- [21] McDiarmid, C., 1989. On the method of bounded differences. *Surveys in combi.*
- [22] Nagaev, S. V., 1979. Large deviations of sums of independent random variabls. *Ann. Probab.* **7**, No. 5, 745–789.
- [23] Nagaev, S. V., 2003. On probability and moment inequalities for supermartingales and martingales. *Acta. Appl. Math.* **79**, 35–46.
- [24] Pinelis, I., 1994. Optimum bounds for the distributions of martingales in Banach spaces. *Ann. Probab.* **22**, 1679–1706.
- [25] Pinelis, I., 1994. Extremal probabilistic problems and Hotelling’s  $T^2$  test under a symmetry assumption. *Ann. Statist.* **22**, No. 4, 357–368.
- [26] Pinelis, I., 2006. Binomial upper bounds on generalized moments and tail probabilities of (super)martingales with differences bounded from above. *High Dimensional probab.* **51**, 33–52.

- [27] Prohorov, Yu. V., 1959. An extremal problem in probability theory. *Theor. Probability Appl.* **4**, 201–203.
- [28] Rio, E., 2002. A Bennett type inequality for maxima of empirical processes. *Ann. Inst. H. Poincaré Probab. Statist.* **6**, 1053–1057.
- [29] Steiger, W. T., 1967. Some Kolmogoroff-type inequalities for bounded random variables. *Biometrika* **54**, 641–647.
- [30] Steiger, W. T., 1969. A best possible Kolmogoroff-type inequality for martingales and a characteristic property. *Ann. Math. Statist.* **40**, No. 3, 764–769.
- [31] Talagrand, M., 1995. The missing factor in Hoeffding’s inequalities. *Ann. Inst. H. Poincaré Probab. Statist.* **31**, 689–702.