Electronic Journal of Probability

Vol. 11 (2006), Paper no. 39, pages 1049–1070.

Journal URL
http://www.math.washington.edu/~ejpecp/

On normal domination of (super)martingales

Iosif Pinelis
Department of Mathematical Sciences
Michigan Technological University
Houghton, Michigan 49931, USA
E-mail: ipinelis@math.mtu.edu

Abstract

Let $(S_0, S_1, \dots)$ be a supermartingale relative to a nondecreasing sequence of σ-algebras $(H_{\leqslant 0}, H_{\leqslant 1}, \dots)$, with $S_0 \leqslant 0$ almost surely (a.s.) and differences $X_i := S_i - S_{i-1}$. Suppose that for every $i = 1, 2, \dots$ there exist $H_{\leqslant (i-1)}$-measurable r.v.'s C_{i-1} and D_{i-1} and a positive real number s_i such that $C_{i-1} \leqslant X_i \leqslant D_{i-1}$ and $D_{i-1} - C_{i-1} \leqslant 2s_i$ a.s. Then for all natural n and all functions f satisfying certain convexity conditions

$$\mathsf{E} f(S_n) \leqslant \mathsf{E} f(sZ),$$

where $s := \sqrt{s_1^2 + \cdots + s_n^2}$ and $Z \sim N(0, 1)$. In particular, this implies

$$\mathsf{P}(S_n \geqslant x) \leqslant c_{5,0} \mathsf{P}(sZ \geqslant x) \quad \forall x \in \mathbb{R},$$

where $c_{5,0} = 5!(e/5)^5 = 5.699\dots$. Results for $\max_{0 \leqslant k \leqslant n} S_k$ in place of S_n and for concentration of measure also follow

Key words: supermartingales; martingales; upper bounds ; probability inequalities ;generalized moments

AMS 2000 Subject Classification: Primary 60E15,60J65; Secondary: 60E05,60G15, 60G50, 60J3080H35.

Submitted to EJP on June 1 2006, final version accepted November 15 2006.

1 Introduction

The sharp form,

$$\mathsf{E}f\left(\varepsilon_1 a_1+\cdots+\varepsilon_n a_n\right) \leqslant \mathsf{E}f(Z), \tag{1.1}$$

of Khinchin's inequality (15) for $f(x)=|x|^p$ for the normalized Rademacher sum $\varepsilon_1 a_1+\cdots+\varepsilon_n a_n$, with

$$a_1^2+\cdots+a_n^2=1,$$

was proved by Whittle (1960) (35) for $p \geqslant 3$ and Haagerup (1982) (12) for $p \geqslant 2$; here and elsewhere, the ε_i's are independent Rademacher random variables (r.v.'s), so that $\mathsf{P}(\varepsilon_i=1)=\mathsf{P}(\varepsilon_i=-1)=1/2$ for all i, and $Z \sim N(0,1)$.

For $f(x)=e^{\lambda x}$ $(\lambda \geqslant 0)$, inequality (1.1) follows from Hoeffding (1963) (13), whence

$$\mathsf{P}\left(\varepsilon_1 a_1+\cdots+\varepsilon_n a_n \geqslant x\right) \leqslant \inf_{\lambda \geqslant 0} \frac{\mathsf{E} e^{\lambda Z}}{e^{\lambda x}}=e^{-x^2/2} \quad \forall x \geqslant 0.$$

Since $\mathsf{P}(Z \geqslant x) \sim \frac{1}{x\sqrt{2\pi}} e^{-x^2/2}$ as $x \to \infty$, a factor $\asymp \frac{1}{x}$ is "missing" here. The apparent cause of this deficiency is that the class of the exponential moment functions $f(x)=e^{\lambda x}$ $(\lambda \geqslant 0)$ is too small (and so is the class of the power functions $f(x)=|x|^p$).

Consider the much richer classes $\mathcal{F}_+^{(\alpha)}$ $(\alpha \geqslant 0)$ consisting of all the functions $f\colon \mathbb{R} \to \mathbb{R}$ given by the formula

$$f(x)=\int_{-\infty}^{\infty}(x-t)_+^\alpha\, \mu(dt) \quad \forall x \in \mathbb{R},$$

where $\mu=\mu_f \geqslant 0$ is a Borel measure, $x_+:=\max(0,x)$, $x_+^\alpha:=(x_+)^\alpha$, $0^0:=0$; note that the condition $f\colon \mathbb{R} \to \mathbb{R}$ implies that $\int_{-\infty}^{\infty}(x-t)_+^\alpha\, \mu_f(dt)<\infty$ $\forall x \in \mathbb{R}$, which is equivalent to the requirement that $\int_{-\infty}^{x}(1+|t|)^\alpha\, \mu_f(dt)<\infty$ $\forall x \in \mathbb{R}$.

Define $\mathcal{F}_-^{(\alpha)}$ as the class of all functions of the form $u \mapsto f(-u)$, where $f \in \mathcal{F}_+^{(\alpha)}$. Let $\mathcal{F}^{(\alpha)}:=\{f+g\colon f \in \mathcal{F}_+^{(\alpha)}, g \in \mathcal{F}_-^{(\alpha)}\}$.

It is easy to see (25, Proposition 1(ii)) that

$$0 \leqslant \beta \leqslant \alpha \quad \text{implies} \quad \mathcal{F}_+^{(\alpha)} \subseteq \mathcal{F}_+^{(\beta)}. \tag{1.2}$$

Proposition 1.1. *(29) For natural α, one has $f \in \mathcal{F}_+^{(\alpha)}$ if and only if f has finite derivatives $f^{(0)}:=f, f^{(1)}:=f', \ldots, f^{(\alpha-1)}$ on $\mathbb{R}$ such that $f^{(\alpha-1)}$ is convex on $\mathbb{R}$ and $f^{(j)}(-\infty+)=0$ for all $j=0,1,\ldots,\alpha-1$.*

It follows from Proposition 1.1 that, for every $t \in \mathbb{R}$, every $\beta \geqslant \alpha$, and every $\lambda>0$, the functions $u \mapsto (u-t)_+^\beta$ and $u \mapsto e^{\lambda(u-t)}$ belong to $\mathcal{F}_+^{(\alpha)}$, while the functions $u \mapsto |u-t|^\beta$ and $u \mapsto \cosh\lambda(u-t)$ belong to $\mathcal{F}^{(\alpha)}$.

Eaton (1970) (6) proved the Khinchin-Whittle-Haagerup inequality (1.1) for a class of moment functions, which essentially coincides with the class $\mathcal{F}^{(3)}$, as seen from (22, Proposition A.1). Based on asymptotics, numerics, and a certain related inequality, Eaton (1974) (7) conjectured that the mentioned moment comparison inequality of his implies that

$$\mathsf{P}\left(\varepsilon_1 a_1+\cdots+\varepsilon_n a_n \geqslant x\right) \leqslant \frac{2e^3}{9}\,\frac{1}{x\sqrt{2\pi}} e^{-x^2/2} \quad \forall x>\sqrt{2}.$$

Pinelis (1994) (22) proved the following improvement of this conjecture:

$$\mathsf{P}\left(\varepsilon_1 a_1 + \cdots + \varepsilon_n a_n \geqslant x\right) \leqslant \frac{2e^3}{9}\,\mathsf{P}(Z \geqslant x) \quad \forall x \in \mathbb{R}, \tag{1.3}$$

as well as certain multidimensional extensions of these results.

Later it was realized (Pinelis (1998) (24)) that the reason why it is possible to extract tail comparison inequality (1.3) from the Khinchin-Eaton moment comparison inequality (1.1) for $f \in \mathcal{F}_+^{(3)}$ is that the tail function $x \mapsto \mathsf{P}(Z \geqslant x)$ is log-concave. This realization resulted in a general device, which allows one to extract the optimal tail comparison inequality from an appropriate moment comparison inequality. The following is a special case of Theorem 4 of Pinelis (1999) (25); see also Theorem 3.11 of Pinelis (1998) (24).

Theorem 1.2. *Suppose that $0 \leqslant \beta \leqslant \alpha$, ξ and η are real-valued r.v.'s, and the tail function $u \mapsto \mathsf{P}(\eta \geqslant u)$ is log-concave on $\mathbb{R}$. Then the comparison inequality*

$$\mathsf{E} f(\xi) \leqslant \mathsf{E} f(\eta) \quad \textit{for all } f \in \mathcal{F}_+^{(\alpha)} \tag{1.4}$$

implies

$$\mathsf{E} f(\xi) \leqslant c_{\alpha,\beta}\,\mathsf{E} f(\eta) \quad \textit{for all } f \in \mathcal{F}_+^{(\beta)} \tag{1.5}$$

and, in particular, for all real x,

$$\mathsf{P}(\xi \geqslant x) \leqslant \inf_{f \in \mathcal{F}_+^{(\alpha)}} \frac{\mathsf{E} f(\eta)}{f(x)} \tag{1.6}$$

$$= B_{\mathit{opt}}(x) := \inf_{t \in (-\infty, x)} \frac{\mathsf{E}(\eta - t)_+^\alpha}{(x-t)^\alpha} \tag{1.7}$$

$$\leqslant \min\left(c_{\alpha,0}\,\mathsf{P}(\eta \geqslant x),\ \inf_{h>0} e^{-hx}\,\mathsf{E} e^{h\eta}\right), \tag{1.8}$$

where

$$c_{\alpha,\beta} := \frac{\Gamma(\alpha+1)(e/\alpha)^\alpha}{\Gamma(\beta+1)(e/\beta)^\beta} \quad \textit{if} \quad \beta > 0, \tag{1.9}$$

and $c_{\alpha,\beta}$ is extended by continuity to the case when $\beta = 0$. Moreover, the constant $c_{\alpha,\beta}$ is the best possible in (1.5) *and* (1.8) *(over all pairs (ξ, η) of r.v.'s satisfying* (1.4)*).*

A similar result for the case when $\alpha = 1$ and $\beta = 0$ is contained in the book by Shorack and Wellner (1986) (33), pages 797–799.

Remark 1.3. Typically, a log-concave tail function $q(x) := \mathsf{P}(\eta \geqslant x)$ of a r.v. η with $\sup\operatorname{supp}\eta = \infty$ will satisfy the regularity condition

$$\left(\frac{q(x)}{q'(x)}\right)' \to 0 \quad \text{as} \quad x \to \infty. \tag{1.10}$$

It follows from the special case $r = \infty$ of (24, Theorem 4.2) that the constant factor $c_{\alpha,\beta}$ in (1.5) and (1.8) is optimal not only over all pairs (ξ, η) satisfying (1.4), but also for every given r.v. η whose tail function q satisfies condition (1.10) and over all r.v.'s ξ satisfying (1.4). In particular, this is true when η has a normal or exponential distribution. (This remark was prompted by an anonymous referee's question.)

Remark 1.4. As follows from (24, Remark 3.13), a useful point is that the requirement of the log-concavity of the tail function $q(x) := \mathsf{P}(\eta \geqslant x)$ in Theorem 1.2 can be relaxed by replacing q with any (e.g., the least) log-concave majorant of q. However, then the optimality of $c_{\alpha,\beta}$ is not guaranteed.

Note that $c_{3,0} = 2e^3/9$, which is the constant factor in (1.3). Bobkov, Götze, and Houdré (2001) (4) (BGH) obtained a simpler proof of inequality (1.3), but with a constant factor 12.0099... in place of $2e^3/9 = 4.4634\ldots$. To obtain the comparison of the tails of $S_n := \varepsilon_1 a_1 + \cdots + \varepsilon_n a_n$ and Z, BGH used a more direct method, based on the Chapman-Kolmogorov identity for the Markov chain (S_n) (rather than on comparison of generalized moments). Such an identity was used, e.g., in Pinelis (2000) (26) to disprove a conjecture by Graversen and Peškir (1998) (11) on $\max_{k \leqslant n} |S_k|$. In Pinelis (2006) (31), it was shown that a modification of the BGH method can be used to prove that the best constant factor (in place of $2e^3/9$) in inequality (1.3) is in an interval $\approx [3.18, 3.22]$. For related improvements of a result of Edelman (1990) (8), see Pinelis (2006) (30).

Pinelis (1999) (25) obtained the "discrete" improvement of (1.3):

$$\mathsf{P}\left(\varepsilon_1 a_1 + \cdots + \varepsilon_n a_n \geqslant x\right) \leqslant \frac{2e^3}{9}\, \mathsf{P}\left(\frac{1}{\sqrt{n}}(\varepsilon_1 + \cdots + \varepsilon_n) \geqslant x\right) \tag{1.11}$$

for all values x that r.v. $\frac{1}{\sqrt{n}}(\varepsilon_1 + \cdots + \varepsilon_n)$ takes on with nonzero probability.

In this paper, we obtain upper bounds on generalized moments and tails of supermartingales with bounded, possibly asymmetric differences. These bounds are substantially more precise than the corresponding exponential ones and appear to be new even for sums of independent r.v.'s.

2 Domination by normal moments and tails

Throughout, unless specified otherwise, let $(S_0, S_1, \ldots)$ be a supermartingale relative to a nondecreasing sequence $(H_{\leqslant 0}, H_{\leqslant 1}, \ldots)$ of σ-algebras, with $S_0 \leqslant 0$ almost surely (a.s.) and differences $X_i := S_i - S_{i-1}$, $i = 1, 2, \ldots$. Unless specified otherwise, let E_j and Var_j denote the conditional expectation and variance, respectively, given $H_{\leqslant j}$. The following theorem is the basic result in this paper.

Theorem 2.1. *Suppose that for every $i = 1, 2, \ldots$ there exist $H_{\leqslant(i-1)}$-measurable r.v.'s C_{i-1} and D_{i-1} and a positive real number s_i such that*

$$C_{i-1} \leqslant X_i \leqslant D_{i-1} \quad \textit{and} \tag{2.1}$$

$$D_{i-1} - C_{i-1} \leqslant 2s_i \tag{2.2}$$

a.s. Then for all $f \in \mathcal{F}_+^{(5)}$ and all $n = 1, 2, \ldots$

$$\mathsf{E}f(S_n) \leqslant \mathsf{E}f(sZ), \tag{2.3}$$

where

$$s := \sqrt{s_1^2 + \cdots + s_n^2}$$

and $Z \sim N(0, 1)$.

The proofs of this and other statements (wherever a proof is necessary) are deferred to Section 5.

By virtue of Theorem 1.2, one has the following corollary under the conditions of Theorem 2.1.

Corollary 2.2. *For all* $\beta \in [0,5]$*, all* $f \in \mathcal{F}_+^{(\beta)}$*, and all* $n = 1, 2, \dots$

$$\mathsf{E} f(S_n) \leqslant c_{5,\beta}\, \mathsf{E} f(sZ). \tag{2.4}$$

In particular, for all real x*,*

$$\mathsf{P}(S_n \geqslant x) \leqslant \inf_{f\in\mathcal{F}_+^{(5)}} \frac{\mathsf{E} f(sZ)}{f(x)} \tag{2.5}$$

$$= \inf_{t\in(-\infty,x)} \frac{\mathsf{E}(sZ-t)_+^5}{(x-t)^5} \tag{2.6}$$

$$\leqslant \min\left(c_{5,0}\,\mathsf{P}(sZ \geqslant x), \inf_{h>0} e^{-hx}\,\mathsf{E} e^{hsZ}\right) \tag{2.7}$$

$$= \min\left(c_{5,0}\,\overline{\Phi}\left(\frac{x}{s}\right), \exp\left(-\frac{x^2}{2s^2}\right)\right), \tag{2.8}$$

and

$$c_{5,0} = 5!(e/5)^5 = 5.699\dots.$$

(Cf. (28).) In (2.8) and in what follows,

$$\overline{\Phi}(x) := \int_x^\infty \varphi(u)\,\mathrm{d}u, \quad \text{where} \quad \varphi(x) := \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

Remark 2.3. The class $\mathcal{F}_+^{(5)}$ in Theorem 2.1 and hence the constant factors $c_{5,\beta}$ in Corollary 2.2 originate in the crucial Lemma 5.1.2. A natural question is whether the index 5 in $\mathcal{F}_+^{(5)}$ and $c_{5,\beta}$ is the least, and hence the best, possible (recall (1.2)). It can be shown that this value, 5, cannot be replaced by 4. It may be possible to replace 5 by some number α in the interval $(4,5)$. However, in view of the proof of Lemma 5.1.2, it appears that the proof for a (non-integer!) $\alpha \in (4,5)$ in place of 5 would be very difficult, if attainable at all, and its benefits will not be very significant; indeed, for any $\alpha \in (4,5)$ the factor $c_{\alpha,0}$ will be in the rather narrow interval $(c_{4,0}, c_{5,0}) \approx (5.119, 5.699)$; that is, the constant factor $c_{5,0}$ cannot be significantly reduced. Cf. Remark 2.5 below.

Remark 2.4. The upper bound $\exp\left(-\frac{x^2}{2s^2}\right)$ was obtained by Hoeffding (1963) (13) for the case when the C_{i-1}'s and D_{i-1}'s are non-random. The upper bound $c_{5,0}\,\mathsf{P}(sZ \geqslant x) = c_{5,0}\,\overline{\Phi}\left(\frac{x}{s}\right)$ is better than Hoeffding's bound $\exp\left(-\frac{x^2}{2s^2}\right)$ for all $\frac{x}{s} \geqslant 1.89$, and at that $c_{5,0}\,\overline{\Phi}(1.89) = 0.16\dots$, which is significantly greater than 0.05, the standard statistical value. Thus, this improvement is quite relevant for statistics.

Remark 2.5. The upper bound (2.8) – but with a constant factor greater than 427 in place of $c_{5,0} = 5.699\dots$ was obtained in Bentkus (2001) (1) for the case when (S_i) is a martingale. (Bentkus was using direct methods, rather than a generalized moment comparison inequality such as (2.3).) The large value, 427, of the constant factor renders the bound in (1) hardly usable in statistics. Indeed, the upper bound $427\,\overline{\Phi}\left(\frac{x}{s}\right)$ improves the Hoeffding bound $\exp\left(-\frac{x^2}{2s^2}\right)$ only when $\frac{x}{s} > 170$, in which case (in view of (2.8)) one has $\mathsf{P}(S_n \geqslant x) < c_{5,0}\,\overline{\Phi}(170) < 10^{-6200}$.

However, the following improvement of the bound in (1) may in certain instances be even more significant.

Theorem 2.6. *Suppose that for every $i = 1, 2, \ldots$ there exist a positive $H_{\leqslant(i-1)}$-measurable r.v. D_{i-1} and a positive real number $\hat{s}_i$ such that*

$$X_i \leqslant D_{i-1} \quad \text{and} \tag{2.9}$$

$$\frac{1}{2}\left(D_{i-1} + \frac{\mathsf{Var}_{i-1}X_i}{D_{i-1}}\right) \leqslant \hat{s}_i \tag{2.10}$$

a.s. Let

$$\hat{s} := \sqrt{\hat{s}_1^2 + \cdots + \hat{s}_n^2}. \tag{2.11}$$

Then one has all the inequalities (2.3)–(2.8)*, only with s replaced by $\hat{s}$.*

Remark 2.7. Theorem 2.1 may be considered as a special case Theorem 2.6. Indeed, it can be seen from the proofs of these two theorems (see Lemma 5.1.1 in this paper and Lemma 3.1 in (29)) that one may assume without loss of generality that the supermartingales (S_i) in Theorem 2.1 and 2.6 are actually martingales with $S_0 = 0$. Therefore, to deduce Theorem 2.1 from Theorem 2.6, it is enough to observe that for any r.v. X and constants $c < 0$ and $d > 0$, one has the following implication:

$$\mathsf{E}X = 0 \;\&\; \mathsf{P}(c \leqslant X \leqslant d) = 1 \implies \mathsf{Var}X \leqslant |c|d. \tag{2.12}$$

In turn, implication (2.12) follows from (14) (say), which reduces the situation to that of a r.v. X taking on only two values. Alternatively, in light of the duality result (24, (4)), it is easy to give a direct proof of (2.12). Indeed, $\mathsf{E}X = 0$ and $\mathsf{P}(c \leqslant X \leqslant d) = 1$ imply

$$0 \geqslant \mathsf{E}(X - c)(X - d) = \mathsf{E}X^2 + cd = \mathsf{Var}X - |c|d.$$

However, instead of deducing Theorem 2.1 from Theorem 2.6, we shall go in the opposite direction, proving Theorem 2.6 based on Theorem 2.1.

Thus, Theorem 2.1 is seen as the main result of this paper.

Remark 2.8. The set of conditions (2.9)–(2.10) is equivalent to

$$X_i \leqslant D_{i-1} \quad \text{and} \quad \sigma_*(D_{i-1}, \mathsf{Var}_{i-1}X_i) \leqslant \hat{s}_i$$

a.s., where

$$\sigma_*(d_0, \sigma^2) := \frac{1}{2} \inf_{d \geqslant d_0} \left(d + \frac{\sigma^2}{d}\right) = \min\left(\sigma \vee d_0, \frac{1}{2}\left(d_0 + \frac{\sigma^2}{d_0}\right)\right)$$
$$= \begin{cases} \sigma & \text{if } \sigma \geqslant d_0, \\ \frac{1}{2}\left(d_0 + \frac{\sigma^2}{d_0}\right) & \text{if } \sigma < d_0, \end{cases}$$

for positive σ and d_0. This follows simply because the inequalities $X_i \leqslant D_{i-1}$ and $d \geqslant D_{i-1}$ imply $X_i \leqslant d$.

Thus, in the case when $\mathsf{Var}_{i-1}X_i < D_{i-1}^2$ a.s., conditions (2.9)–(2.10) represent an improvement of condition $D_{i-1}^2 \vee \mathsf{Var}_{i-1}X_i \leqslant \hat{s}_i^2$ a.s., considered in (2; 3). In a certain variety of cases, this improvement may be even more significant than the improvement in the constant factor from 427 to 5.699... *before* the probability sign.

On the other hand, it can be shown that the value $\sigma_*(d_0,\sigma^2)$ is equal or close to the optimal value $s = s_*(d_0,\sigma^2)$, which is the smallest value $s \geqslant 0$ satisfying the inequality $\mathsf{E}(X-t)_+^5 \leqslant \mathsf{E}(s\,Z-t)_+^5$ $\forall t \in \mathbb{R}$, where X is a zero-mean r.v. taking on values d_0 and $-\sigma^2/d_0$ only. If $u := \sigma^2/d_0^2$ is $\geqslant 1$, then $\sigma_*(d_0,\sigma^2) = s_*(d_0,\sigma^2)$. It can be seen that, even if u is as small as 0.1, one has $\sigma_*(d_0,\sigma^2) < 1.13\,s_*(d_0,\sigma^2)$ $\big($whereas $\sigma \vee d_0 = d_0 > 2\,s_*(d_0,\sigma^2)\big)$; if $u = 0.4$, then $\sigma_*(d_0,\sigma^2) < 1.01\,s_*(d_0,\sigma^2)$ $\big($whereas $\sigma \vee d_0 = d_0 > 1.4\,s_*(d_0,\sigma^2)\big)$.

From the "right-tail" bounds stated above, "two-tail" ones immediately follow:

Corollary 2.9. *Let $(S_0,S_1,\dots)$ be a martingale with $S_0 = 0$ a.s. Suppose that conditions* (2.1) *and* (2.2) *hold. Then inequalities* (2.3) *and* (2.4) *hold for all $f \in \mathcal{F}^{(5)}$ and $f \in \mathcal{F}^{(\beta)}$ ($\beta \in [0,5]$), rather than only for all $f \in \mathcal{F}_+^{(5)}$ and $f \in \mathcal{F}_+^{(\beta)}$, respectively.*

Corollary 2.10. *Let $(S_0,S_1,\dots)$ be a martingale with $S_0 = 0$ a.s. Suppose that condition* (2.10) *holds, and condition* (2.9) *holds for $|X_i|$ in place of X_i. Then inequalities* (2.3) *and* (2.4) *with s replaced by $\hat{s}$ hold for all $f \in \mathcal{F}^{(5)}$ and $f \in \mathcal{F}^{(\beta)}$ ($\beta \in [0,5]$).*

That $(S_0,S_1,\dots)$ in Theorems 2.1 and 2.6 is allowed to be a supermartingale (rather than only a martingale) makes it convenient to use the simple but powerful truncation tool. (Such a tool was used, for example, in (20; 21) to prove limit theorems for large deviation probabilities based only on precise enough probability inequalities and without using Cramér's transform, the standard device in the theory of large deviations.) Thus, for instance, one has the following corollary from Theorem 2.6.

Corollary 2.11. *For every $i = 1,2,\dots$, let D_{i-1} be a positive $H_{\leqslant(i-1)}$-measurable r.v. and let $\hat{s}_i$ be a positive real number such that* (2.10) *holds (while* (2.9) *does not have to). Let $\hat{s}$ be still defined by* (2.11).

Then for all real x

$$\mathsf{P}(S_n \geqslant x) \leqslant \mathsf{P}\left(\max_{1\leqslant i\leqslant n} \frac{X_i}{D_{i-1}} \geqslant 1\right) + \min\left(c_{5,0}\,\overline{\Phi}\left(\frac{x}{\hat{s}}\right), \exp\left(-\frac{x^2}{2\hat{s}^2}\right)\right) \tag{2.13}$$

$$\leqslant \sum_{1\leqslant i\leqslant n} \mathsf{P}\left(X_i \geqslant D_{i-1}\right) + \min\left(c_{5,0}\,\overline{\Phi}\left(\frac{x}{\hat{s}}\right), \exp\left(-\frac{x^2}{2\hat{s}^2}\right)\right). \tag{2.14}$$

These bounds are much more precise than the exponential bounds in (10; 9; 19).

3 Maximal inequalities

Introduce

$$M_n := \max_{0\leqslant k\leqslant n} S_k.$$

Theorem 3.1. *The upper bounds on $\mathsf{P}(S_n \geqslant x)$ given in Corollary 2.2 and Theorem 2.6 are also upper bounds on $\mathsf{P}(M_n \geqslant x)$, under the same conditions:* (2.1)–(2.2) *and* (2.9)–(2.10)*, respectively.*

Theorem 3.2. *Let $0 \leqslant \beta \leqslant \alpha$ and $x > t$, and let (S_n) be a martingale or, more generally, a submartingale. Assume, moreover, that $\alpha > 1$. Then, for any natural n,*

$$\mathsf{E}(M_n - x)_+^\beta \leqslant k_{1;\alpha,\beta} \frac{\mathsf{E}(S_n - t)_+^\alpha}{(x-t)^{\alpha-\beta}}, \tag{3.1}$$

where

$$k_{1;\alpha,\beta} := \sup_{\sigma>0} \sigma^{-\beta(\alpha-1)} \left(\int_0^\sigma \frac{\beta s^{\beta-1}\, ds}{1+s} \right)^\alpha \tag{3.2}$$

if $\beta > 0$, and $k_{1;\alpha,0} := 1$. The particular cases of (3.1)*, corresponding to $\beta = 0$ and $\beta = \alpha$, respectively, are Doob's inequalities*

$$\mathsf{P}(M_n \geqslant x) \leqslant \frac{\mathsf{E}(S_n - t)_+^\alpha}{(x-t)^\alpha} \tag{3.3}$$

and

$$\mathsf{E}(M_n)_+^\alpha \leqslant \left(\frac{\alpha}{\alpha-1} \right)^\alpha \mathsf{E}(S_n)_+^\alpha. \tag{3.4}$$

Proposition 3.3. *Let $0 \leqslant \beta < \alpha$, $x > t$, and*

$$k_{\alpha,\beta} := \frac{\beta^\beta (\alpha-\beta)^{\alpha-\beta}}{\alpha^\alpha}. \tag{3.5}$$

Then

$$\forall u \in \mathbb{R} \quad (u-x)_+^\beta \leqslant k_{\alpha,\beta} \frac{(u-t)_+^\alpha}{(x-t)^{\alpha-\beta}}, \tag{3.6}$$

and $k_{\alpha,\beta}$ is the best constant factor here – even under condition $u = 0$. (The values at $\beta = 0$ are understood here as the corresponding limits as $\beta \downarrow 0$.)

Theorem 3.4. *Under the same conditions:* (2.1)–(2.2) *and* (2.9)–(2.10)*, respectively, the upper bounds given in Corollary 2.2 and Theorem 2.6 hold with M_n in place of S_n and the constant factor $\dfrac{k_{1;5,\beta}}{k_{5,\beta}}\, c_{5,\beta}$ in place of $c_{5,\beta}$.*

Similarly, results of (29) can be extended.

Remark 3.5. Note that

$$\int_0^\sigma \frac{\beta s^{\beta-1}\, ds}{1+s} = \sigma^\beta {}_2F_1(\beta, 1; 1+\beta; -\sigma) = \beta\sigma^\beta \int_0^1 (1-u)^{\beta-1}(1+\sigma u)^{-\beta}\, du,$$

where ${}_2F_1$ is a hypergeometric function. Note also that, for $\beta \in (0, \alpha)$, there is some $\sigma_{\alpha,\beta} \in (0, \infty)$ such that the expression under the sup sign in (3.2) is increasing in $\sigma \in (0, \sigma_{\alpha,\beta})$ and decreasing in $\sigma \in (\sigma_{\alpha,\beta}, \infty)$; this can be seen from the proof of Proposition 3.8. Thus, the sup is attained at the unique point $\sigma_{\alpha,\beta}$.

Proposition 3.6. *Let α and β be as in Theorem 3.2. Then*

$$k_{1;\alpha,\beta} \leqslant k_{2;\alpha,\beta} := \frac{\Gamma(1+\beta)\Gamma(\alpha-\beta)}{\Gamma(\alpha)}. \tag{3.7}$$

Remark 3.7.
$$k_{2;\alpha,0} = k_{\alpha,0} = 1 = k_{1;\alpha,0}.$$

Proposition 3.8. *Let α and β be as in Theorem 3.2. Then*
$$k_{1;\alpha,\beta} \leqslant k_{3;\alpha,\beta} := k_{\alpha,\beta}\left(\frac{\alpha}{\alpha-1}\right)^{\alpha}, \tag{3.8}$$
where $k_{\alpha,\beta}$ is defined by (3.5).

Proposition 3.9. *Let $\alpha > 1$. Then*
$$k_{1;\alpha,\alpha} = k_{3;\alpha,\alpha} = \left(\frac{\alpha}{\alpha-1}\right)^{\alpha}. \tag{3.9}$$

Corollary 3.10. *Let α and β be as in Theorem 3.2. Then*
$$k_{\alpha,\beta} \leqslant k_{1;\alpha,\beta} \leqslant k_{2;\alpha,\beta} \wedge k_{3;\alpha,\beta}; \tag{3.10}$$
at that
$$k_{\alpha,0} = k_{1;\alpha,0} = k_{2;\alpha,0} = 1, \tag{3.11}$$
while
$$k_{1;\alpha,\alpha} = k_{3;\alpha,\alpha} = \left(\frac{\alpha}{\alpha-1}\right)^{\alpha} > k_{\alpha,\alpha} = 1. \tag{3.12}$$

4 Concentration inequalities for separately Lipschitz functions

Definition 4.1. Let us say that a real-valued function g of n (not necessarily real-valued) arguments is *separately Lipschitz* if it satisfies a Lipschitz type condition in each of its arguments:
$$|g(x_1, \dots, x_{i-1}, \tilde{x}_i, x_{i+1}, \dots, x_n) - g(x_1, \dots, x_n)| \leqslant \rho_i(\tilde{x}_i, x_i) < \infty \tag{4.1}$$
for all i and all $x_1, \dots, x_n, \tilde{x}_i$, where $\rho_i(\tilde{x}_i, x_i)$ depends only on i, $\tilde{x}_i$, and x_i. Let the *radius* of the separately Lipschitz function g be defined as
$$r := \sqrt{r_1^2 + \cdots + r_n^2},$$
where
$$r_i := \frac{1}{2}\sup_{\tilde{x}_i, x_i} \rho_i(\tilde{x}_i, x_i). \tag{4.2}$$

The concentration inequalities given in this section follow from martingale inequalities given in Section 2. The proofs here are based on the improvements given in (20) and (32) of the method of Yurinskiĭ(1974) (36); cf. (17; 18) and (1).

Papers (36), (20), and (32) deal mainly with separately Lipschitz function g of the form
$$g(x_1, \dots, x_n) = \|x_1 + \cdots + x_n\|,$$
where the x_i's are vectors in a normed space; however, it was already understood there that the methods would work for much more general functions g – see e.g. (32, Remark 1). In a similar

fashion, various concentration inequalities for general functions g were obtained in (17; 18) and (1).

Suppose that a r.v. Y with a finite mean can be represented as a real-valued Borel function g of independent (not necessarily real-valued) r.v.'s $X_1, \dots, X_n$:

$$Y = g(X_1, \dots, X_n).$$

Theorem 4.2. *If g is separately Lipschitz with a radius $r > 0$, then*

$$\mathsf{E} f(Y - \mathsf{E}Y) \leqslant \mathsf{E} f(rZ) \quad \text{for all } f \in \mathcal{F}^{(5)} \quad \text{and} \tag{4.3}$$

$$\mathsf{E} f(Y - \mathsf{E}Y) \leqslant c_{5,\beta}\, \mathsf{E} f(rZ) \quad \text{for all } \beta \in [0,5] \text{ and all } f \in \mathcal{F}^{(\beta)}, \tag{4.4}$$

where $Z \sim N(0,1)$. In particular, for all real x,

$$\mathsf{P}(Y - \mathsf{E}Y \geqslant x) \leqslant c_{5,0}\, \mathsf{P}(rZ \geqslant x). \tag{4.5}$$

Inequality (4.5) – but with a constant factor greater than 427 in place of $c_{5,0} = 5.699\dots$ was obtained in Bentkus (2001) (1); cf. Remark 2.5.

The already comparatively weak separately-Lipschitz condition assumed in Theorem 4.2 can be further relaxed, as follows.

Theorem 4.3. *Inequalities* (4.3), (4.4), *and* (4.5) *will hold if the separately-Lipschitz condition of Theorem 4.2 is relaxed so that r_i is replaced by*

$$\hat{r}_i := \frac{1}{2} \sup_{x_1,\dots,x_i,\tilde{x}_i} |\mathsf{E} g(x_1, \dots, x_{i-1}, \tilde{x}_i, X_{i+1}, \dots, X_n) - \mathsf{E} g(x_1, \dots, x_i, X_{i+1}, \dots, X_n)|, \tag{4.6}$$

for every i. Note that $\hat{r}_i \leqslant r_i$ for all i.

Theorems 4.2 and 4.3 are based on Theorem 2.1 (and also on the mentioned improvements of Yurinskiĭ's method). Using Theorem 2.6 instead of Theorem 2.1, one can obtain the following improvement of these results.

Theorem 4.4. *Suppose that*

$$\begin{aligned} \Xi_i(x_1, \dots, x_{i-1}, x_i) &:= \mathsf{E} g(x_1, \dots, x_{i-1}, x_i, X_{i+1}, \dots, X_n) \quad (4.7)\\ &\quad - \mathsf{E} g(x_1, \dots, x_{i-1}, X_i, X_{i+1}, \dots, X_n)\\ &\leqslant D_{i-1}(x_1, \dots, x_{i-1}), \quad (4.8) \end{aligned}$$

and

$$\frac{1}{2}\left(D_{i-1}(x_1, \dots, x_{i-1}) + \frac{\mathsf{E}\Xi_i(x_1, \dots, x_{i-1}, X_i)^2}{D_{i-1}(x_1, \dots, x_{i-1})} \right) \leqslant s_i, \tag{4.9}$$

for all i and all $x_1, \dots, x_{i-1}, x_i$, where $D_{i-1}(x_1, \dots, x_{i-1}) > 0$ depends only on i and $x_1, \dots, x_{i-1}$, and s_i depends only on i. Let

$$s := \sqrt{s_1^2 + \dots + s_n^2}.$$

Then inequalities (4.3), (4.4), *and* (4.5) *will hold if r is replaced there by s.*

Remark 4.5. Under the conditions of Theorem 4.2 or Theorem 4.3, bound (4.5) can be replaced by any one of the better bounds (2.6)–(2.8) with s in the latter bounds replaced by r. Similarly, under the conditions of Theorem 4.4, bound (4.5) can be replaced by any one of the bounds (2.6)–(2.8).

The next proposition shows how to obtain good upper bounds on $\Xi_i(x_1,\dots,$
$x_{i-1},x_i)$ and $\mathsf{E}\Xi_i(x_1,\dots,x_{i-1},X_i)^2$, to be used in Theorem 4.4.

Proposition 4.6. *If g is separately Lipschitz so that* (4.1) *holds, then for all i and all $x_1,\dots,x_{i-1}$,*

$$\mathsf{E}\Xi_i(x_1,\dots,x_{i-1},X_i)^2 \leqslant \inf_{x_i} \mathsf{E}\rho_i(X_i,x_i)^2 \leqslant \mathsf{E}\rho_i(X_i,\mathsf{E}X_i)^2; \tag{4.10}$$

it is assumed that the function ρ_i is measurable in an appropriate sense; for the second inequality in (4.10), *it is also assumed that an appropriately defined expectation $\mathsf{E}X_i$ exists, for all i. If, moreover, the function g is convex in each of its arguments, then for all i and all $x_1,\dots,x_i$,*

$$\Xi_i(x_1,\dots,x_{i-1},x_i) \leqslant \rho_i(x_i,\mathsf{E}X_i). \tag{4.11}$$

Remark 4.7. We do not require that ρ_i be a metric. However, the smallest possible ρ_i, which is the supremum of the left-hand side of (4.1) over all $x_1,\dots,x_{i-1}$,
$x_{i+1},\dots,x_n$, is necessarily a pseudo-metric. Note also that, for r_i defined by (4.2),

$$\rho_i(x_i,\mathsf{E}X_i) = \rho_i(x_i,0) \leqslant r_i$$

for all x_i, provided e.g. the additional conditions that (i) $\rho_i(x_i,\tilde{x}_i) = \|x_i-\tilde{x}_i\|_i$ for some semi-norms $\|\cdot\|_i$ and all i, x_i and $\tilde{x}_i$; (ii) X_i is symmetrically distributed; and (iii) x_i belongs to the support of the distribution of X_i.

Corollary 4.8. *Let here $X_1,\dots,X_n$ be independent r.v.'s with values in a separable Banach space with norm $\|\cdot\|$, and let*

$$Y := \|X_1+\dots+X_n\|.$$

Suppose that, a.s.,

$$\|X_i-\mathsf{E}X_i\| \leqslant d_i \tag{4.12}$$

and

$$\frac{1}{2}\left(d_i+\frac{\mathsf{E}\|X_i-\mathsf{E}X_i\|^2}{d_i}\right) \leqslant s_i, \tag{4.13}$$

for all i, where $d_i>0$ and $s_i>0$ are non-random constants. Let

$$s := \sqrt{s_1^2+\dots+s_n^2}.$$

Then inequalities (4.3), (4.4), *and* (4.5) *will hold if r is replaced there by s.*

Concerning exponential bounds for sums of independent B-valued r.v.'s and for martingales in 2-smooth spaces, see (23).

The separately-Lipschitz condition (4.1) is obviously equivalent the ℓ^1-like Lipschitz condition

$$|g(\tilde{x}_1,\dots,\tilde{x}_n)-g(x_1,\dots,x_n)| \leqslant \sum_{i=1}^n \rho_i(\tilde{x}_i,x_i) < \infty \tag{4.14}$$

for all $x_1, \dots, x_n, \tilde{x}_1, \dots, \tilde{x}_n$ (provided that each ρ_i is the smallest possible and hence a pseudo-metric, as indicated in Remark 4.7). A particular case of the ℓ^1-like pseudo-metric is the widely used (especially in combinatorics and computer science (17; 18)) Hamming distance. The upper bounds presented in this section are substantially more precise than exponential bounds such as ones found in (17; 18); cf. Remark 2.4.

There is a great amount of literature on the measure concentration phenomenon, including treatments in terms of metrics other than ℓ^1; refer e.g. to Talagrand (34), Ledoux (16), Dembo (5), and Bobkov, Houdré and Götze (4).

5 Proofs

5.1 Proofs for Section 2

Let us first observe that Theorem 2.1 can be easily reduced to the case when (S_n) is a martingale. This is implied by the following lemma, which is obvious and stated here for the convenience of reference.

Lemma 5.1.1. *Let (S_n) be a supermartingale as in Theorem 2.1, so that conditions* (2.1) *and* (2.2) *are satisfied. Let*

$$\tilde{X}_i := X_i - \mathsf{E}_{i-1}X_i, \quad \tilde{C}_{i-1} := C_{i-1} - \mathsf{E}_{i-1}X_i, \quad \text{and} \quad \tilde{D}_{i-1} := D_{i-1} - \mathsf{E}_{i-1}X_i.$$

Then $\tilde{X}_i$ is $H_{\leqslant i}$-measurable, $\tilde{C}_{i-1}$ and $\tilde{D}_{i-1}$ are $H_{\leqslant (i-1)}$-measurable, and one has

$$X_i \leqslant \tilde{X}_i, \ \mathsf{E}_{i-1}\tilde{X}_i = 0, \ \tilde{C}_{i-1} \leqslant \tilde{X}_i \leqslant \tilde{D}_{i-1}, \text{ and } \tilde{D}_{i-1} - \tilde{C}_{i-1} \leqslant 2s_i. \ \text{a.s.}$$

Proof of Theorem 2.1. The proof is similar to the proof of Theorem 2.1 in (29) but based on the crucial Lemma 5.1.2 below, in place of Lemma 3.2 in (29). Also, one has to refer here to Lemma 5.1.1 instead of Lemma 3.1 in (29). Indeed, by Lemma 5.1.1, one may assume that $\mathsf{E}_{i-1}X_i = 0$ for all i. Let $Z_1, \dots, Z_n$ be independent standard normal r.v.'s, which are also independent of the X_i's, and let

$$R_i := X_1 + \cdots + X_i + s_{i+1}Z_{i+1} + \cdots + s_nZ_n.$$

Let $\tilde{\mathsf{E}}_i$ denote the conditional expectation given $X_1, \dots, X_{i-1}, Z_{i+1}, \dots, Z_n$. Note that, for all $i = 1, \dots, n$, one has $\tilde{\mathsf{E}}_iX_i = \mathsf{E}_{i-1}X_i = 0$; moreover, $R_i - X_i = X_1 + \cdots + X_{i-1} + s_{i+1}Z_{i+1} + \cdots + s_nZ_n$ is a function of $X_1, \dots, X_{i-1}, Z_{i+1}, \dots, Z_n$. Hence, by Lemma 5.1.2, for any $f \in \mathcal{F}_+^{(5)}$, $\tilde{f}_i(x) := f(R_i - X_i + x)$, and all $i = 1, \dots, n$, one has $\tilde{\mathsf{E}}_if(R_i) = \tilde{\mathsf{E}}_i\tilde{f}_i(X_i) \leqslant \tilde{\mathsf{E}}_i\tilde{f}_i(s_iZ_i) = \tilde{\mathsf{E}}_if(R_{i-1})$, whence $\mathsf{E}f(S_n) \leqslant \mathsf{E}f(R_n) \leqslant \mathsf{E}f(R_0) = \mathsf{E}f(s\,Z)$ (the first inequality here follows because $S_0 \leqslant 0$ a.s. and any function f in $\mathcal{F}_+^{(5)}$ is nondecreasing). □

Lemma 5.1.2. *Let X be a r.v. such that $\mathsf{E}X = 0$ and $c \leqslant X \leqslant d$ a.s. for some real constants c and d (whence $c \leqslant 0$ and $d \geqslant 0$). Let $Z \sim N(0,1)$. Then for all $f \in \mathcal{F}_+^{(5)}$*

$$\mathsf{E}f(X) \leqslant \mathsf{E}f(\tfrac{d-c}{2}\,Z). \tag{5.1}$$

Proof. This proof is rather long. Let $\mathcal{X}_{c,d}$ be the set of all r.v.'s X such that $\mathsf{E}X = 0$ and $c \leqslant X \leqslant d$ a.s. Without loss of generality (w.l.o.g.), $f = f_t$ for some $t \in \mathbb{R}$, where

$$f_t(x) := (x-t)_+^5.$$

In view of (14) (say), for any given real t, a maximum of $\mathsf{E}f_t(X)$ over all r.v.'s X in $\mathcal{X}_{c,d}$ is attained when X takes on only two values, say a and b, in the interval $[c,d]$. Since the function f_t is convex, it then follows that w.l.o.g. $a = c$ and $b = d$. $\big($Indeed, one can prove that $\mathsf{E}f_t(\sigma Z)$ is non-decreasing in $\sigma > 0$ by an application of Jensen's inequality.$\big)$ Moreover, by rescaling, w.l.o.g. $d - c = 2$. In other words, then one has the following:

$$X = \begin{cases} 2r & \text{with probability } 1-r, \\ 2r-2 & \text{with probability } r, \end{cases}$$

for some $r \in [0,1]$. Now the right-hand side of inequality (5.1) can be written as

$$\mathsf{E}f_t(Z) = R(t) := P(t)\varphi(t) - Q(t)\overline{\Phi}(t), \tag{5.2}$$

where

$$P(t) := 8 + 9t^2 + t^4 \quad \text{and} \quad Q(t) := t(15 + 10t^2 + t^4),$$

and its left-hand side as

$$\mathsf{E}f_t(X) = L(r,t) := r(2r-2-t)_+^5 + (1-r)(2r-t)_+^5, \tag{5.3}$$

so that (5.1) is reduced to the inequality

$$L(r,t) \leqslant R(t) \tag{5.4}$$

for all $r \in [0,1]$ and all real t.

Note that (5.4) is trivial for $t \geqslant 2r$, because then $L(r,t) = 0 \leqslant \mathsf{E}f_t(Z) = R(t)$.

Therefore, it remains to consider two cases: $(r,t) \in B$ and $(r,t) \in C$, where

$$B := \{(r,t)\colon 0 \leqslant r \leqslant 1, t \leqslant 2r-2\} \quad \text{and} \quad C := \{(r,t)\colon 0 \leqslant r \leqslant 1, 2r-2 \leqslant t \leqslant 2r\}.$$

Case 1 $(r,t) \in B$. Note that in this case $t \leqslant 0$ and, by (5.3),

$$L(r,t) = r(2r-2-t)^5 + (1-r)(2r-t)^5.$$

For $t \neq 0$, one has the identity

$$\frac{Q(t)^2}{\varphi(t)}\,\partial_t\left(\frac{R(t)-L(r,t)}{Q(t)}\right) = Q_2(r,t) := \frac{Q_1(r,t)}{\varphi(t)} - 120, \tag{5.5}$$

where

$$Q_1(r,t) := Q'(t)L(r,t) - Q(t)\,\partial_t L(r,t),$$

which is a polynomial in r and t. Note that

$$\partial_r Q_2(r,t) = \frac{\partial_r Q_1(r,t)}{\varphi(t)} \quad \text{and} \quad \partial_t Q_2(r,t) = \frac{20\,Q(t)}{\varphi(t)} d(r,t),$$

where

$$d(r,t) := \frac{tQ_1(r,t) + \partial_t Q_1(r,t)}{20\,Q(t)}.$$

Therefore, the critical points of Q_2 in the interior $\operatorname{int} B$ of domain B are the solutions (r,t) of the system of polynomial equations

$$\begin{cases} d(r,t) = 0, \\ \partial_r Q_1(r,t) = 0. \end{cases}$$

Further, $d(r, 2r-2-u)$ is a polynomial in r and u, of degree 2 in r; moreover, for $(r,t) \in \operatorname{int} B$, one has $t < 2r-2$, so that, in terms of $u := 2r-2-t > 0$,

$$d(r,t) = 0 \text{ if and only if } r = r_1(u) \text{ or } r = r_2(u),$$

where

$$r_1(u) := \frac{1+u/2}{1+u} \in (0,1) \quad \text{and} \quad r_2(u) := \frac{2+2u+u^2/2}{2+2u+u^2} \in (0,1).$$

Using the Sturm theorem or the convenient command **Reduce** of Mathematica, one can see that the only solution $u = u_1 > 0$ of the algebraic equation $\partial_r Q_1(r,t)|_{r=r_1(u),t=2r_1(u)-2-u} = 0$ is $0.284\dots$, and $Q_2(r,t)|_{r=r_1(u_1),t=2r_1(u_1)-2-u_1} < 0$. As for the equation $\partial_r Q_1(r,t)|_{r=r_2(u),t=2r_2(u)-2-u} = 0$, it has no solutions $u > 0$.

Thus, $Q_2 < 0$ at the only critical point $(r,t) = \big(r_1(u_1), 2r_1(u_1) - 2 - u_1\big)$ of Q_2 in $\operatorname{int} B$.

Next, with $u \geqslant 0$,

$$Q_2(r,t)|_{r=0,t=2r-2-u} = -20\left(6 + \frac{(2+u)^5}{\varphi(2+u)}\left(7 + 4\,u + u^2\right)\right) < 0.$$

Similarly, with $u \geqslant 0$,

$$Q_2(r,t)|_{r=1,t=2r-2-u} = -20\left(6 + \frac{u^5(3+u^2)}{\varphi(u)}\right) < 0.$$

Now consider the function

$$q_2(r) := Q_2(r,t)|_{t=2r-2}.$$

Then $\varphi(2r-2)q_2'(r)$ is a polynomial, whose only root $r = r_3 \in (0,1)$ is $0.865\dots$. But $q_2(r_3) < 0$. Therefore, $Q_2 < 0$ at the only critical point of Q_2 in the relative interior of the boundary $t = 2r-2$ of domain B.

Thus, as far as the sign of Q_2 on B is concerned, it remains to consider the behavior of Q_2 as $t \to -\infty$, which is as follows: $Q_2(r,t) \sim 20(2r-1)^2 t^7/\varphi(t) \to -\infty < 0$ for every $r \neq 1/2$ and $Q_2(r,t) \sim 40t^5/\varphi(t) \to -\infty < 0$ for $r = 1/2$.

(As usual, $a \sim b$ means $a/b \to 1$.)

We conclude that $Q_2 < 0$ on B. Hence, in view of (5.5), the ratio $\frac{R(t)-L(r,t)}{Q(t)}$ is decreasing in t on B (letting here $\frac{R(t)-L(r,t)}{Q(t)} := -\infty$ at $(r,t) = (1,0)$).

Next, note that $\varphi(t)$ and $1-\overline{\Phi}(t)$ are $o(1/|t|^p)$ for every $p > 0$ as $t \to -\infty$. Hence, in view of (5.2), one has the following as $t \to -\infty$: $R(t) - L(r,t) = -Q(t) - L(r,t) + o(1) \sim -10(2r-1)^2 t^3 \to \infty$ for every $r \neq 1/2$ and $R(t) - L(r,t) \sim -10t \to \infty$ for $r = 1/2$.

Hence, $\frac{R(t)-L(r,t)}{Q(t)} < 0$ for each $r \in (0,1)$ and all $t < 0$ with large enough $|t|$. Since $\frac{R(t)-L(r,t)}{Q(t)}$ is decreasing in t on B, one has $\frac{R(t)-L(r,t)}{Q(t)} < 0$ on B, whence $L(r,t) \leqslant R(t)$ on B (because $Q(t) \leqslant 0$ on B).

It remains to consider

Case 2 $(r,t) \in C$. Here, letting $v := 2r - t$, one has $0 \leqslant v \leqslant 2$, and, by (5.3),

$$L(r,t) = (1-r)(2r-t)^5.$$

Let us use here notation introduced in the above consideration of Case 1. Then

$$d(r,t)|_{t=2r-v} = -(1-r)v^3 \left(1 - \frac{v}{2} r\right) < 0$$

for $(r,t) = (r, 2r-v) \in \operatorname{int} C$. This implies that Q_2 has no critical points in $\operatorname{int} C$.

Next, with $v > 0$,

$$Q_2(r,t)|_{r=0,t=2r-v} = -20\left(6 + \frac{v^5(3+v^2)}{\varphi(t)}\right) < 0.$$

On the boundaries $r = 1$ and $t = 2r$ of C, one has $Q_2 = -120 < 0$. The boundary $t = 2r - 2$ of C is common with B, and it was shown above that $Q_2 < 0$ on that boundary as well.

Thus, $Q_2 < 0$ on C. It follows by (5.5) that the ratio $\frac{R(t)-L(r,t)}{Q(t)}$ is decreasing in t on $C_- := \{(r,t) \in C \colon t < 0\}$ and on $C_+ := \{(r,t) \in C \colon t > 0\}$.

Hence, just as on B, one has that $L(r,t) \leqslant R(t)$ on C_-.

Moreover, $\frac{R(t)-L(r,t)}{Q(t)} = \frac{R(t)}{Q(t)} > 0$ for $t = 2r$, since $Q > 0$ on C_+. Because $\frac{R(t)-L(r,t)}{Q(t)}$ is decreasing in t on C_+, one has $\frac{R(t)-L(r,t)}{Q(t)} > 0$ on C_+ and hence $L(r,t) < R(t)$ on C_+.

One concludes that $L(r,t) \leqslant R(t)$ on the entire set C. □

Proof of Theorem 2.6. This proof is similar to the proof of Theorem 2.1 in (29) and Theorem 2.1 of this paper, but based on the following lemma, instead of Lemma 3.2 in (29) or Lemma 5.1.2. (Here one has also to refer to Lemma 3.1 in (29), rather than to Lemma 5.1.1.) □

Lemma 5.1.3. *Suppose that X is a r.v. such that $\mathsf{E}X = 0$, $X \leqslant d$ a.s., and $\mathsf{E}X^2 \leqslant \sigma^2$, for some positive constants d and σ. Let*

$$s := \frac{1}{2}\left(d + \frac{\sigma^2}{d}\right)$$

and $Z \sim N(0,1)$. Then for all $f \in \mathcal{F}^{(5)}$

$$\mathsf{E}f(X) \leqslant \mathsf{E}f(sZ). \tag{5.6}$$

Proof. In view of (1.2), one has $\mathcal{F}^{(5)} \subseteq \mathcal{F}^{(2)}$. Therefore, by Lemma 3.2 in (29), one may assume without loss of generality that here $X = d \cdot X_a$, where $a = \sigma^2/d^2$. Now it is seen that Lemma 5.1.3 follows from Lemma 5.1.2. ☐

5.2 Proofs for Section 3

Proof of Theorem 3.1. Lemma 5.1.1 and Lemma 3.1 in (29) reduce Theorem 3.1 to the case when (S_n) is a martingale, and then Theorem 3.1 follows by Doob's inequality (3.3), Theorems 2.1 and 2.6, and inequality (1.8). ☐

Proof of Theorem 3.2. For every $y > t$, by Doob's inequality,

$$\mathsf{P}(M_n \geqslant y) \leqslant \frac{\mathsf{E}(S_n - t)_+ I\{M_n \geqslant y\}}{y - t}.$$

Hence, letting

$$J(u) := \int_x^u \frac{\beta(y-x)^{\beta-1}}{y-t}\,\mathrm{d}y\; I\{u > x\} \quad \text{and} \quad \alpha' := \frac{\alpha}{\alpha-1}, \tag{5.7}$$

and using Fubini's theorem, one has

$$\begin{aligned}
\mathsf{E}(M_n - x)_+^\beta &= \int_x^\infty \beta(y-x)^{\beta-1}\mathsf{P}(M_n \geqslant y)\,\mathrm{d}y \\
&\leqslant \int_x^\infty \beta(y-x)^{\beta-1}\frac{\mathsf{E}(S_n - t)_+ I\{M_n \geqslant y\}}{y-t}\,\mathrm{d}y \\
&= \mathsf{E}\int_x^\infty \beta(y-x)^{\beta-1}\frac{(S_n - t)_+ I\{M_n \geqslant y\}}{y-t}\,\mathrm{d}y \\
&= \mathsf{E}(S_n - t)_+ J(M_n) \\
&\leqslant \left(\mathsf{E}(S_n - t)_+^\alpha\right)^{1/\alpha} \left(\mathsf{E}J(M_n)^{\alpha'}\right)^{1/\alpha'},
\end{aligned} \tag{5.8}$$

by Hölder's inequality.

Observe that for all real u

$$J(u) \leqslant c^{1/\alpha}(u-x)_+^{\beta/\alpha'}, \quad \text{where} \quad c := \frac{k_{1;\alpha,\beta}}{(x-t)^{\alpha-\beta}}. \tag{5.9}$$

Indeed, introducing new variables $\sigma := \frac{u-x}{x-t}$ and $s := \frac{y-x}{x-t}$, one can see that, for $u > x$,

$$\begin{aligned}
J(u) &= (x-t)^{\beta-1}\int_0^\sigma \frac{\beta s^{\beta-1}\,\mathrm{d}s}{1+s} \quad \text{and} \\
c^{1/\alpha}(u-x)_+^{\beta/\alpha'} &= k_{1;\alpha,\beta}^{1/\alpha}\sigma^{\beta(1-1/\alpha)}(x-t)^{\beta-1},
\end{aligned}$$

so that (5.9) follows, in view of (3.2).

Now (5.8) and (5.9) imply (3.1).

That (3.3) is the particular case of (3.1) corresponding to $\beta = 0$ follows because $k_{1;\alpha,0}$ was defined as 1. Finally, that (3.4) is the particular case of (3.1) corresponding to $\beta = \alpha$ follows by Proposition 3.9, to be proved later in this paper. ☐

Proof of Proposition 3.3. Elementary calculus. □

Proof of Theorem 3.4. This is similar to the proof Theorem 3.1, but relies on inequality (3.1) in place of Doob's inequality (3.3). It also utilizes Theorems 2.1 and 2.6 and inequality

$$\inf_{t\in(-\infty,x)} \frac{\mathsf{E}(Z-t)_+^\alpha}{(x-t)^{\alpha-\beta}} \leqslant \frac{c_{\alpha,\beta}}{k_{\alpha,\beta}}\,\mathsf{E}(Z-x)_+^\beta \quad (\alpha>\beta\geqslant 0),$$

which in turn follows from the proof of Theorem 3.11 in (24); cf. the second inequality in (24), identities (25) and (26), the second inequality in (23), and the definition in (16) there. □

Proof of Proposition 3.6. Introduce

$$f(\sigma,\alpha,\beta,\gamma) := \sigma^{-\beta(\alpha-\gamma)/\gamma}\left(\int_0^\sigma \frac{\beta s^{\beta-1}\,ds}{(1+s)^\gamma}\right)^{\alpha/\gamma},$$
$$K(\alpha,\beta,\gamma) := \sup_{\sigma>0} f(\sigma,\alpha,\beta,\gamma).$$

Then $\sigma^{-\beta/\alpha}f(\sigma,\alpha,\beta,\gamma)^{1/\alpha} = (\mathsf{E}Y^\gamma)^{1/\gamma}$, where $Y := \frac{1}{1+S}$ and S is a r.v. with density $s\mapsto \sigma^{-\beta}\beta s^{\beta-1}I\{0<s<\sigma\}$. Hence, $f(\sigma,\alpha,\beta,\gamma)$ is non-decreasing in γ, and then so is $K(\alpha,\beta,\gamma)$. Therefore,

$$k_{1;\alpha,\beta} = K(\alpha,\beta,1) \leqslant K(\alpha,\beta,\alpha) = k_{2;\alpha,\beta}.$$

□

Proof of Proposition 3.8. W.l.o.g., $0<\beta<\alpha$. By (3.2),

$$k_{1;\alpha,\beta} = \sup_{\sigma>0} r(\sigma)^\alpha, \tag{5.10}$$

where

$$r(\sigma) := \frac{f(\sigma)}{g(\sigma)}, \quad f(\sigma) := \int_0^\sigma \frac{\beta s^{\beta-1}\,ds}{1+s}, \quad \text{and} \quad g(s) := \sigma^{\beta(1-1/\alpha)}.$$

Note that the monotonicity pattern of

$$r_1(\sigma) := \frac{f'(\sigma)}{g'(\sigma)} = \frac{\alpha}{\alpha-1}\frac{\sigma^{\beta/\alpha}}{1+\sigma} \tag{5.11}$$

on $(0,\infty)$ is $\nearrow\searrow$; that is, there exists some $\sigma_1(\alpha,\beta)\in(0,\infty)$ such that $r_1\nearrow$ (is increasing) on $(0,\sigma_1(\alpha,\beta))$ and $r_1\searrow$ (is decreasing) on $(\sigma_1(\alpha,\beta),\infty)$; namely, here

$$\sigma_1(\alpha,\beta) = \frac{\beta}{\alpha-\beta}. \tag{5.12}$$

Also, $gg'>0$ on $(0,\infty)$. Hence, it follows from (27, Proposition 1.9) that r has one of these monotonicity patterns on $(0,\infty)$: $\nearrow$ or $\searrow$ or $\nearrow\searrow$ or $\searrow\nearrow$ or $\searrow\nearrow\searrow$. However, $r(\sigma)$ is positive on $(0,\infty)$ and converges to 0 when $\sigma\downarrow 0$ as well as when $\sigma\to\infty$. This leaves only one possible pattern for r: $\nearrow\searrow$. Hence, there is some $\sigma(\alpha,\beta) \in (0,\infty)$, at which r attains its maximum on $(0,\infty)$; moreover, $r'(\sigma(\alpha,\beta)) = 0$, which is equivalent to $r(\sigma(\alpha,\beta)) = r_1(\sigma(\alpha,\beta))$. Thus,

$$k_{1;\alpha,\beta} = \sup_{\sigma>0} r(\sigma)^\alpha = r(\sigma(\alpha,\beta))^\alpha = r_1(\sigma(\alpha,\beta))^\alpha \leqslant \sup_{\sigma>0} r_1(\sigma)^\alpha$$
$$= r_1(\sigma_1(\alpha,\beta))^\alpha = k_{3;\alpha,\beta},$$

in view of (5.11), (5.12), and the definition in (3.8). □

Proof of Proposition 3.9. In the case $\beta = \alpha > 1$, the function r_1 given by (5.11) is increasing on $(0,\infty)$ to $r_1(\infty) = \frac{\alpha}{\alpha-1}$. Hence, so does r, according to (27, Proposition 1.1) and l'Hospital's rule rule for limits. Now Proposition 3.9 follows in view of (5.10). □

Proof of Corollary 3.10. The second inequality in (3.10) follows by Propositions 3.6 and 3.8. Equalities (3.11) follow from the definitions. The first two equalities in (3.12) follow by Proposition 3.9, while the third equality in (3.12) follows from the definition.

It remains to prove the first inequality in (3.10). Suppose the contrary: $k_{\alpha,\beta} > k_{1;\alpha,\beta}$. Then Theorem 3.2 (with $S_i = 0$ a.s. $\forall i$) will imply that inequality (3.6) with $u = 0$ holds with constant factor $k_{1;\alpha,\beta}$ in place of $k_{\alpha,\beta}$, which contradicts Proposition 3.3, according to which $k_{\alpha,\beta}$ is the best constant factor in (3.6), even for $u = 0$. □

5.3 Proofs for Section 4

The proofs here are based on the improvements given in (20) and (32) of the method of Yurinskiĭ(1974) (36); cf. (17; 18) and (1).

For a r.v. Y as in Theorem 4.2, consider the martingale expansion

$$Y - \mathsf{E}Y = \xi_1 + \cdots + \xi_n,$$

of $Y - \mathsf{E}Y$ with the martingale-differences

$$\xi_i := \mathsf{E}_i Y - \mathsf{E}_{i-1} Y, \tag{5.13}$$

where E_i and Var_i denote, respectively, the conditional expectation and variance given the σ-algebra (say $H_{\leqslant i}$) generated by $(X_1,\dots,X_i)$. For each i pick an arbitrary non-random x_i, and introduce the r.v.

$$\eta_i := Y - \tilde{Y}_i, \quad \text{where} \quad \tilde{Y}_i := g(X_1,\dots,X_{i-1},x_i,X_{i+1},\dots,X_n). \tag{5.14}$$

Proof of Theorems 4.2 and 4.3. Note that, for the function Ξ_i defined by (4.7), one has $\Xi_i(X_1,\dots,X_i) = \xi_i$ a.s., where ξ_i is defined by (5.13). It follows from (5.13) that

$$C_{i-1} \leqslant \xi_i \leqslant D_{i-1} \quad \text{and} \quad D_{i-1} - C_{i-1} \leqslant 2\hat{r}_i \leqslant 2r_i, \tag{5.15}$$

where r_i and $\hat{r}_i$ are given by (4.2) and (4.6), and

$$C_{i-1} := \inf_{x_i} \mathsf{E}_{i-1}(-\eta_i) = \inf_{x_i} \mathsf{E}_{i-1}\tilde{Y}_i - \mathsf{E}_{i-1}Y \quad \text{and}$$
$$D_{i-1} := \sup_{x_i} \mathsf{E}_{i-1}(-\eta_i) = \sup_{x_i} \mathsf{E}_{i-1}\tilde{Y}_i - \mathsf{E}_{i-1}Y$$

are $H_{\leqslant(i-1)}$-measurable. Now Theorem 4.3 – and hence Theorem 4.2 – follow by Theorem 2.1 and Corollary 2.2. □

Proof of Theorem 4.4. This proof is similar to that of Theorems 4.2 and 4.3, but based on Theorem 2.6 (in place of Theorem 2.1) and Corollary 2.2. (Note that $\mathsf{E}\Xi_i(x_1,\dots,x_{i-1},X_i)^2$ is the same as conditional expectation $\mathsf{E}_{i-1}\xi_i^2$ given that $X_1 = x_1,\dots,X_{i-1} = x_{i-1}$.) $\square$

Proof of Proposition 4.6. For each i,

$$\xi_i = \mathsf{E}_i\eta_i - \mathsf{E}_{i-1}\eta_i, \tag{5.16}$$

because $\mathsf{E}_i\tilde{Y}_i = \mathsf{E}_{i-1}\tilde{Y}_i$, in view of the independence of the X_i's. By (4.1), for any given x_i,

$$|\eta_i| \leqslant \rho_i(X_i, x_i) \tag{5.17}$$

a.s. It follows from (5.16) and (5.17) that, for any x_i,

$$\begin{aligned}\mathsf{E}_{i-1}\xi_i^2 = \mathsf{E}_{i-1}(\mathsf{E}_i\eta_i - \mathsf{E}_{i-1}\eta_i)^2 = \mathsf{Var}_{i-1}(\mathsf{E}_i\eta_i) \leqslant \mathsf{E}_{i-1}(\mathsf{E}_i\eta_i)^2 \leqslant \mathsf{E}_{i-1}\mathsf{E}_i\eta_i^2 \\ = \mathsf{E}_{i-1}\eta_i^2 \leqslant \mathsf{E}_{i-1}\rho_i(X_i,x_i)^2 = \mathsf{E}\rho_i(X_i,x_i)^2,\end{aligned}$$

which proves (4.10).

To prove (4.11), suppose in addition that the function g is convex in each of its arguments, as stated in the second part of Proposition 4.6. Let $\tilde{\mathsf{E}}_i$ denote the conditional expectation given $X_1,\dots,X_{i-1},X_{i+1},\dots,X_n$. Then, by Jensen's inequality, one has for all i

$$\begin{aligned}\mathsf{E}_{i-1}Y = \mathsf{E}_{i-1}\tilde{\mathsf{E}}_iY = \mathsf{E}_{i-1}\tilde{\mathsf{E}}_ig(X_1,\dots,X_n) \\ \geqslant \mathsf{E}_{i-1}g(X_1,\dots,X_{i-1},\tilde{\mathsf{E}}_iX_i,X_{i+1},\dots,X_n) \\ = \mathsf{E}_{i-1}g(X_1,\dots,X_{i-1},\mathsf{E}X_i,X_{i+1},\dots,X_n) = \mathsf{E}_{i-1}\tilde{Y},\end{aligned}$$

in view of (5.14), if x_i is chosen to coincide with $\mathsf{E}X_i$; hence,

$$\mathsf{E}_{i-1}\eta_i = \mathsf{E}_{i-1}Y - \mathsf{E}_{i-1}\tilde{Y} \geqslant 0.$$

This and formulas (5.16) and (5.17) imply that

$$\xi_i \leqslant \mathsf{E}_i\eta_i \leqslant \rho_i(X_i, \mathsf{E}X_i),$$

which is equivalent to (4.11). $\square$

Proof of Corollary 4.8. This follows immediately from Theorem 4.4 and Proposition 4.6, with $\rho_i(\tilde{x}_i, x_i) = \|\tilde{x}_i - x_i\|$. $\square$

References

[1] Bentkus, V. (2001). On measure concentration for separately Lipschitz functions in product spaces. *Israel J. Math.* To appear.

[2] Bentkus, V. (2003) An inequality for tail probabilities of martingales with differences bounded from one side. *J. Theoret. Probab.* **16**, 161–173.

[3] Bentkus, V. (2004) On Hoeffding's inequalities. *Ann. Probab.* **32**, 1650–1673.

[4] Bobkov, S. G., Götze, F. and Houdré, C. (2001) On Gaussian and Bernoulli covariance representations. *Bernoulli* **7**, 439–451.

[5] Dembo, A. (1997) Information inequalities and concentration of measure. *Ann. Probab.* **25**, 927–939.

[6] Eaton, M. L. (1970). A note on symmetric Bernoulli random variables. *Ann. Math. Statist.* **41**, 1223–1226.

[7] Eaton, M. L. (1974). A probability inequality for linear combinations of bounded random variables. *Ann. Statist.* **2**, 609–614.

[8] Edelman, D. (1990). An inequality of optimal order for the tail probabilities of the T statistic under symmetry. *J. Amer. Statist. Assoc.* **85**, 120–122.

[9] Fuk, D. H. (1971). Certain probabilistic inequalities for martingales. *Siberian Math. J.* **14**, 131–137.

[10] Fuk, D. H. and Nagaev, S. V. (1971). Probabilistic inequalities for sums of independent random variables. (Russian. English summary) *Teor. Verojatnost. i Primenen.* **16**, 660–675.

[11] Graversen, S. E.; Peškir, G. (1998). Extremal problems in the maximal inequalities of Khintchine. *Math. Proc. Cambridge Philos. Soc.* **123**, 169–177.

[12] Haagerup, U. (1982). The best constants in the Khinchine inequality. *Studia Math.* **70**, 231–283.

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

[14] Karr, A. F. (1983). Extreme points of certain sets of probability measures, with applications. *Math. Oper. Res.* **8**, 1, 74–85.

[15] Khinchin, A. (1923). Über dyadische Brüche. *Math. Z.* **18**, 109–116.

[16] Ledoux, M. (1999). Concentration of measure and logarithmic Sobolev inequalities. *Séminaire de Probabilités, XXXIII, 120–216, Lecture Notes in Math.* **1709**, 120–216, Springer, Berlin.

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

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

[19] Nagaev, S. V. (1979). Large deviations of sums of independent random variables. *Ann. Probab.* **7**, 745–789.

[20] Pinelis, I. F. (1981). Limit theorems on large deviations for sums of independent random variables with Cramer's condition violated. (Russian) Deposited at VINITI (All-Russian Institute of Scientific and Technical Information for All-Union Institute of Scientific and Technical Information), No. 1674-81Dep., 94 pages.

[21] Pinelis, I. F. (1985) Asymptotic equivalence of the probabilities of large deviations for sums and maximum of independent random variables. (Russian) *Limit theorems of probability theory, 144–173, 176, Trudy Inst. Mat.,* **5**, *"Nauka" Sibirsk. Otdel., Novosibirsk.*

[22] Pinelis, I. (1994). Extremal probabilistic problems and Hotelling's T^2 test under a symmetry condition. *Ann. Statist.* **22**, 1, 357–368.

[23] Pinelis, I. (1994). Optimum bounds for the distributions of martingales in Banach spaces. *Ann. Probab.* **22**, 4, 1679–1706.

[24] Pinelis, I. (1998). Optimal tail comparison based on comparison of moments. *High dimensional probability (Oberwolfach, 1996), 297–314, Progr. Probab.,* ***43****, Birkhäuser, Basel.*

[25] Pinelis, I. (1999). Fractional sums and integrals of r-concave tails and applications to comparison probability inequalities *Advances in stochastic inequalities (Atlanta, GA, 1997), 149–168, Contemp. Math.,* ***234****, Amer. Math. Soc., Providence, RI.*

[26] Pinelis, I. (2000). On exact maximal Khinchine inequalities. *High dimensional probability, II (Seattle, WA, 1999), 49–63, Progr. Probab.,* ***47****, Birkhäuser Boston, Boston, MA.*

[27] Pinelis, I. (2001). L'Hospital type rules for oscillation, with applications. *JIPAM. J. Inequal. Pure Appl. Math.* **2**, 3, Article 33, 24 pp. (electronic).

[28] Pinelis, I. (2002). L'Hospital type rules for monotonicity: applications to probability inequalities for sums of bounded random variables. *JIPAM. J. Inequal. Pure Appl. Math.* **3**, 1, Article 7, 9 pp. (electronic).

[29] Pinelis, I. (2005). Binomial upper bounds on generalized moments and tail probabilities of (super)martingales with differences bounded from above. *Preprint,* `http://arxiv.org/abs/math.PR/0512301`. To appear in *Proc. conf. on high dimensional probability held in NM, 2005.*

[30] Pinelis, I. (2006). On inequalities for sums of bounded random variables. *Preprint,* `http://arxiv.org/abs/math.PR/ 0603030`.

[31] Pinelis, I. (2006). Toward the best constant factor for the Rademacher-Gaussian tail comparison. *Preprint,* `http://arxiv.org/abs/ math.PR/0605340`.

[32] Pinelis, I.; Sakhanenko, A. I. (1985). Remarks on inequalities for probabilities of large deviations. *Theory Probab. Appl.* **30**, 143–148.

[33] Shorack, G. R. and Wellner, J. A. (1986). *Empirical Processes with Applications to Statistics.* Wiley, New York.

[34] Talagrand, M. (1995). Concentration of measure and isoperimetric inequalities in product spaces. *Inst. Hautes Études Sci. Publ. Math.* No. 81, 73–205.

[35] Whittle, P. (1960). Bounds for the moments of linear and quadratic forms in independent variables. *Teor. Verojatnost. i Primenen.* **5**, 331–335.

[36] Yurinskii, V. V. (Jurinskiĭ, V. V.) (1974). Exponential estimates for large deviations. (Russian) *Teor. Verojatnost. i Primenen.* **19**, 152–154.