Probab. Theory Relat. Fields (2007) 139:605–635
DOI 10.1007/s00440-007-0055-4

Exact inequalities for sums of asymmetric random variables, with applications

Iosif Pinelis

Received: 25 May 2006 / Revised: 26 December 2006 / Published online: 7 February 2007

Abstract Let $\mathrm{BS}_1, \ldots, \mathrm{BS}_n$ be independent identically distributed random variables each having the standardized Bernoulli distribution with parameter $p \in (0, 1)$. Let $m_*(p) := (1 + p + 2p^2)/(2\sqrt{p - p^2} + 4p^2)$ if $0 < p \leqslant \frac{1}{2}$ and $m_*(p) := 1$ if $\frac{1}{2} \leqslant p < 1$. Let $m \geqslant m_*(p)$. Let f be such a function that f and f'' are nondecreasing and convex. Then it is proved that for all nonnegative numbers $c_1, \ldots, c_n$ one has the inequality

$$\mathsf{E}f(c_1\mathrm{BS}_1 + \cdots + c_n\mathrm{BS}_n) \leqslant \mathsf{E}f\big(s^{(m)}(\mathrm{BS}_1 + \cdots + \mathrm{BS}_n)\big),$$

where $s^{(m)} := \big(\frac{1}{n}\sum_{i=1}^n c_i^{2m}\big)^{\frac{1}{2m}}$. The lower bound $m_*(p)$ on m is exact for each $p \in (0, 1)$. Moreover, $\mathsf{E}f(c_1\mathrm{BS}_1 + \cdots + c_n\mathrm{BS}_n)$ is Schur-concave in $(c_1^{2m}, \ldots, c_n^{2m})$.

A number of corollaries are obtained, including upper bounds on generalized moments and tail probabilities of (super)martingales with differences of bounded asymmetry, and also upper bounds on the maximal function of such (super)martingales. Applications to generalized self-normalized sums and t-statistics are given.

Keywords (super)martingales · Probability inequalities · Generalized moments · Self-normalized sums · t-statistic

Mathematics Subject Classification (2000) Primary: 60E15 · 60G50 · 60G42 · 60G48 · 62F03 · 62F25 · 62G10 · 60G15; Secondary: 60E05 · 62E10 · 62G35

I. Pinelis (✉)
Department of Mathematical Sciences,
Michigan Technological University, Houghton, MI 49931, USA
e-mail: ipinelis@mtu.edu

1 Introduction

Exponential upper bounds, say of the form $\mathrm{e}^{-x^2/(2B^2)}$, on the tails probabilities of sums of independent random variables (r.v.'s) or, more generally, martingales have had many applications in probability and statistics. In the applications, the ratio B/x is usually small.

Such exponential bounds are based on upper bounds on the corresponding exponential moments. However, in comparison with the "ideal", "normal" estimate $1-\Phi(x/B) \sim \frac{B}{x\sqrt{2\pi}}\,\mathrm{e}^{-x^2/(2B^2)}$ of the tail probability (where $\Phi(u) := \mathsf{P}(Z \leqslant u)$, with $Z \sim N(0,1)$), the exponential upper bound $\mathrm{e}^{-x^2/(2B^2)}$ "misses" the usually small factor $\asymp B/x$.

The apparent cause of this deficiency is that the class of exponential moment functions is too small (and so is the class of the power functions). See [10,19–21,24] for more on this subject.

For all $\alpha \geqslant 0$, consider the following much richer classes of functions $f\colon \mathbb{R} \to \mathbb{R}$:

$$\mathcal{H}_+^\alpha := \{f\colon f(x) = \int_{-\infty}^{\infty} (x-t)_+^\alpha\, \mu(\mathrm{d}t) \quad \forall u \in \mathbb{R}\}, \tag{1}$$

where $\mu \geqslant 0$ is a Borel measure, $x_+ := \max(0,x)$, and $x_+^\alpha := (x_+)^\alpha$ for $x \in \mathbb{R}$, $0^0 := 0$; however, the subscript $_+$ will have a different meaning when used with functions or classes of functions (as, for example, in the symbol $\mathcal{H}_+^\alpha$). It is easy to see [21, Proposition 1(ii)] that

$$0 \leqslant \beta < \alpha \quad \text{implies} \quad \mathcal{H}_+^\alpha \subseteq \mathcal{H}_+^\beta. \tag{2}$$

For a characterization of the class $\mathcal{H}_+^\alpha$ for natural α see Lemma 2 in Sect. 4 of this paper (cf. [24, Proposition 1.1]). In particular, for every $t \in \mathbb{R}$, every $\beta \geqslant \alpha$, and every $\lambda > 0$, the functions $u \mapsto (u-t)_+^\beta$ and $u \mapsto \mathrm{e}^{\lambda(u-t)}$ belong to $\mathcal{H}_+^\alpha$.

The following is a special case of Theorem 4 of Pinelis [21]; see also Theorem 3.11 of Pinelis [20].

Theorem 1 *Suppose that $\alpha > 0$, 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*

$$\mathsf{E} f(\xi) \leqslant \mathsf{E} f(\eta) \quad \textit{for all } f \in \mathcal{H}_+^\alpha \tag{3}$$

implies

$$\mathsf{P}(\xi \geqslant x) \leqslant B_{\mathsf{opt}}(x) := \inf_{t\in(-\infty,x)} \frac{\mathsf{E}(\eta-t)_+^\alpha}{(x-t)^\alpha} \tag{4}$$

$$\leqslant c_{\alpha,0}\, \mathsf{P}(\eta \geqslant x) \tag{5}$$

for all real x, where the constant factor

$$c_{\alpha,0} := \Gamma(\alpha+1)(e/\alpha)^{\alpha} \tag{6}$$

is the best possible.

A similar result for the case when $\alpha = 1$ is contained in the book by Shorack and Wellner [29], pp. 797–799.

Note that $c_{\alpha,0} \sim \sqrt{2\pi\alpha}$ as $\alpha \to \infty$ and $c_{\alpha,0} > 1$ for all $\alpha > 0$.

Remark 1 The log-concavity of the tail function $q(u) := \mathsf{P}(\eta \geqslant u)$ in Theorem 1 is needed only for inequality (5) but not for inequality (4). In view of [20, Remark 3.13], the log-concavity requirement can be removed by replacing q in (5) with any (e.g., the least) log-concave majorant of q. However, then the optimality of $c_{\alpha,0}$ is not guaranteed.

Bound $B_{\mathsf{opt}}(x)$ in (4) is obvious but useful in some cases, since it is optimal [20]. Theorem 2.5 of [20] provides a general description of how to compute the optimal bound $B_{\mathsf{opt}}(x)$ effectively, even in a more general setting (the natural condition $x < \sup\,\operatorname{supp}\eta$ was missing in parts (iii) and (iv) of the theorem; thanks are due to V. Bentkus for having drawn my attention to that omission). In [4], some general properties of $B_{\mathsf{opt}}(x)$ are presented, and the description of the calculation of $B_{\mathsf{opt}}(x)$ given by the mentioned Theorem 2.5 of [20] is detailed for $\alpha \in \{1,2,3\}$ and specific families of distributions of the r.v. η: exponential, uniform, normal, Bernoulli, binomial, and Poisson.

Note that, since the class $\mathcal{H}_{+}^{\alpha}$ contains all increasing exponential functions for each $\alpha > 0$, the optimal bound $B_{\mathsf{opt}}(x)$ is also majorized by the standard exponential bound

$$B_{\exp}(x) := \inf_{h \geqslant 0} \mathrm{e}^{-hx}\, \mathsf{E}\,\mathrm{e}^{h\eta}. \tag{7}$$

In particular, since obviously $B_{\exp}(x) \leqslant 1$ for all x, it follows that $B_{\mathsf{opt}}(x) \leqslant 1$ for all x; cf. [4, Lemma 3.1]. Thus, $B_{\exp}(x)$ is better than the bound $c_{\alpha,0}\, \mathsf{P}(\eta \geqslant x)$ in (5) for all small enough x. However, in applications (especially in statistics) it is large values of x that usually are of primary interest, and then the bound $c_{\alpha,0}\, \mathsf{P}(\eta \geqslant x)$ will significantly outperform the exponential bound. For other related developments, see [5,8,20–22,25,26].

In what follows, let $(S_0, S_1, \dots)$ be a sequence of r.v.'s adapted to a nondecreasing sequence of σ-algebras $(H_{\leqslant 0}, H_{\leqslant 1}, \dots)$, with differences

$$X_i := S_i - S_{i-1}, \quad i = 1, 2, \dots .$$

The possible property of $(S_0, S_1, \dots)$ being a (super,sub)martingale will be understood with respect to $(H_{\leqslant 0}, H_{\leqslant 1}, \dots)$.

The following normal domination statement is one of the main results of [24].

Theorem 2 [24] *Suppose that* $(S_0, S_1, \dots)$ *is a supermartingale with* $S_0 \leqslant 0$ *almost surely (a.s.) such that for every* $i = 1, 2, \dots$ *there exist* $H_{\leqslant(i-1)}$*-measurable r.v.'s* A_{i-1} *and* B_{i-1} *and a positive real number* c_i *such that*

$$-A_{i-1} \leqslant X_i \leqslant B_{i-1} \quad \text{and} \tag{8}$$

$$\tfrac{1}{2}(A_{i-1} + B_{i-1}) \leqslant c_i \tag{9}$$

a.s. Then for all $f \in \mathcal{H}_+^5$ *and all* $n = 1, 2, \dots$

$$\mathsf{E} f(S_n) \leqslant \mathsf{E} f(s\sqrt{n} Z), \tag{10}$$

where

$$s := s^{(1)} := \sqrt{\frac{c_1^2 + \cdots + c_n^2}{n}}. \tag{11}$$

Note that inequality (10) for the smaller class of exponential functions in place of the class $\mathcal{H}_+^5$ is due to Hoeffding [13].

Based on Theorem 2 and a more general version of Theorem 1, upper bounds on $\mathsf{E} f(S_n)$ for $f \in \mathcal{H}_+^\beta$ with $\beta \in [0, 5]$ were given in [24], including upper bounds on the tail probabilities $\mathsf{P}(S_n \geqslant x)$, as well as similar bounds on the distribution of $M_n := \max_{0 \leqslant k \leqslant n} S_k$ in place of S_n and the most precise presently known bounds for the measure concentration phenomenon in terms of separately-Lipschitz (or, equivalently, ℓ^1-Lipschitz) functions on product spaces.

Yet, it can be seen that even the best possible normal domination result may be inadequate if the asymmetry of the random summands X_i is significant or if n is not large. In such a case, one may try to use binomial domination instead of normal, as in [1, Theorem 1], [3, Theorem 1.1], and [23, Theorem 2.3]. However, in those theorems it was assumed that all the X_i's are almost surely bounded from above by the same constant. This condition may be too restrictive in certain applications.

In this paper, another approach to the problem of asymmetry is presented. Here we provide binomial upper bounds on generalized moments and tail probabilities for S_n assuming that certain *indices of asymmetry* of the X_i's (rather than the X_i's themselves) are *uniformly* bounded from above. This assumption of bounded asymmetry (in contrast with the uniform boundedness) of the X_i's is rather natural in applications to generalized self-normalized sums and t-statistics; see Sect. 3.

2 Statements of basic results and discussion

Let $\mathcal{C}^2$ denote the class of all twice continuously differentiable functions $f\colon \mathbb{R} \to \mathbb{R}$. Consider the following class of functions:

$$\mathcal{F}_+^3 := \{f \in \mathcal{C}^2 \colon f \text{ and } f'' \text{ are nondecreasing and convex}\}. \tag{12}$$

An equivalent definition would be given by the formula

$$\mathcal{F}^3_+ = \{f \in \mathcal{C}^2 \colon f, f', f'', \text{and } f''' \text{ are nondecreasing}\},$$

where f''' denotes the right derivative of the convex function f''.

For example, functions $x \mapsto a+b\,x+c\,(x-t)^\alpha_+$ and $x \mapsto a+b\,x+c\,e^{\lambda x}$ belong to $\mathcal{F}^3_+$ for all $a \in \mathbb{R}$, $b \geqslant 0$, $c \geqslant 0$, $t \in \mathbb{R}$, $\alpha \geqslant 3$, and $\lambda \geqslant 0$.

It is easy to see that the strict set inclusion $\mathcal{H}^3_+ \subset \mathcal{F}^3_+$ takes place; a related, more nontrivial observation will be presented later as Proposition 2 on page 619.

Remark 2 If a function $f \colon \mathbb{R} \to \mathbb{R}$ is convex and a r.v. X has a finite expectation, then, by Jensen's inequality, $\mathsf{E} f(X)$ always exists in $(-\infty, \infty]$. This remark will be used in this paper (sometimes tacitly) for functions f in the class $\mathcal{F}^3_+$, as well as for other convex functions.

Throughout the paper, unless indicated otherwise, the following notation/assumptions will be used:

$$m \in [1, \infty), \quad p \in (0,1), \quad q := 1-p, \quad \text{and } \mathrm{BS}_1, \dots, \mathrm{BS}_n \overset{\text{i.i.d.}}{\sim} \mathsf{BS}(p), \tag{13}$$

where $\mathsf{BS}(p)$ denotes the standardized Bernoulli distribution with parameter p: for a r.v. BS we let, by definition,

$$\mathrm{BS} \sim \mathsf{BS}(p) \iff \mathsf{P}\left(\mathrm{BS} = \sqrt{\tfrac{q}{p}}\right) = p = 1 - \mathsf{P}\left(\mathrm{BS} = -\sqrt{\tfrac{p}{q}}\right);$$

thus, $\mathsf{BS}(p)$ is a two-point zero-mean unit-variance distribution. In particular, $\mathsf{BS}(\frac{1}{2})$ is the distribution of a Rademacher r.v. ε, with $\mathsf{P}(\varepsilon = \pm 1) = \frac{1}{2}$.

Introduce

$$m_*(p) := \begin{cases} \dfrac{1+p+2p^2}{2\left(\sqrt{p-p^2}+2p^2\right)} & \text{if } 0 < p \leqslant \frac{1}{2}, \\ 1 & \text{if } \frac{1}{2} \leqslant p < 1. \end{cases} \tag{14}$$

Later it will be clear that $m_*(p)$ increases from 1 to ∞ as p decreases from $\frac{1}{2}$ to 0 (see the proof of Lemma 17).

Introduce also the notation

$$s^{(m)} := \left(\frac{1}{n}\sum_{i=1}^n c_i^{2m}\right)^{\frac{1}{2m}} \tag{15}$$

for any nonnegative numbers $c_1, \dots, c_n$.

Of the main results of this paper, the following one is perhaps the easiest to state (but not to prove).

Theorem 3 *For any real number*

$$m \geqslant m_*(p), \tag{16}$$

all $f \in \mathcal{F}_+^3$, all natural n, and all nonnegative numbers $c_1, \ldots, c_n$, one has

$$\mathsf{E} f(c_1 \mathrm{BS}_1 + \cdots + c_n \mathrm{BS}_n) \leqslant \mathsf{E} f\left(s^{(m)} \cdot (\mathrm{BS}_1 + \cdots + \mathrm{BS}_n)\right), \tag{17}$$

Moreover, the lower bound $m_*(p)$ on m is exact for each $p \in (0, 1)$.

The exactness of the lower bound $m_*(p)$ is understood here in the following sense: Theorem 3 would be false if the function m_* in (16) were replaced by any other function, say $\tilde{m}_*\colon (0, 1) \to \mathbb{R}$, whose value $\tilde{m}_*(p)$ at any point $p \in (0, 1)$ is less than $m_*(p)$.

The necessary proofs are deferred to Sect. 4.

Remark 3 The general restriction $m \geqslant 1$ in (13) is quite natural. Indeed, if inequality (17) held for some $m \in (0, 1)$ then, taking $c_1 = 1, c_2 = \cdots = c_n = 0$, and letting $n \to \infty$, one would have, by the central limit theorem, the inequality $\mathsf{E} f(\mathrm{BS}_1) \leqslant f(0)$ for all $f \in \mathcal{F}_+^3$, which is false even for $f(x) \equiv \mathrm{e}^x$ or $f(x) \equiv x_+^3$.

Here is a extension of Theorem 3:

Theorem 4 *Suppose that $(S_0, S_1, \ldots)$ is a supermartingale with $S_0 \leqslant 0$ a.s. such that for some natural n and every $i \in \{1, \ldots, n\}$ there are positive $H_{\leqslant(i-1)}$-measurable r.v.'s A_{i-1} and B_{i-1} such that*

$$-A_{i-1} \leqslant X_i \leqslant B_{i-1}, \tag{18}$$

$$\sqrt{A_{i-1}\, B_{i-1}} \leqslant c_i, \quad \textit{and} \tag{19}$$

$$\frac{B_{i-1}}{A_{i-1}} \leqslant \frac{q}{p} \tag{20}$$

a.s., where c_i is a non-random number. Then, for any real number

$$m \geqslant m_*(p) \tag{21}$$

and $f \in \mathcal{F}_+^3$, one has the inequality

$$\mathsf{E} f(S_n) \leqslant \mathsf{E} f\left(s^{(m)} (\mathrm{BS}_1 + \cdots + \mathrm{BS}_n)\right), \tag{22}$$

where $s^{(m)}$ is defined by (15). Moreover, the lower bound $m_*(p)$ on m is exact for each $p \in (0, 1)$.

Condition (20) may be referred to as a bounded-asymmetry condition.

One should compare (19) and (15) with (9) and (11). If $X_{a,b}$ stands for a zero-mean r.v. taking on values in the set $\{-a, b\}$ for some positive a and b, then obviously the half-range $\frac{1}{2}(a+b)$ of $X_{a,b}$ is no less than its standard deviation $\sqrt{ab}$. That is, (9) is more restrictive than (19). On the other hand, one has the inequality $s^{(m)} \geqslant s^{(1)}$ for $m \geqslant 1$. Moreover, the greater the uniform bound $\frac{q}{p}$ on asymmetry in (20) is, the greater m must be according to (21) and hence the more pronounced the inequality $s^{(m)} \geqslant s^{(1)}$ will be. Yet, it will be demonstrated elsewhere that, overall, (19) and (15) work better in certain important statistical applications than (9) and (11). Note also that one can choose the "ideal" value $m = 1$ whenever the asymmetry index $\frac{q}{p}$ does not exceed 1, that is, whenever the X_i's are not skewed to the right.

We shall show that conditions (18), (19), and (20) in Theorem 4 can be replaced by conditions (23), (24), and (25) below. In fact, these two sets of conditions are equivalent to each other in a certain sense; see e.g. Remark 2.4 in [24] and the proof of Theorem 2.3 therein, as well as the proof of Corollary 1 in Sect. 4.1 below.

Corollary 1 *Suppose that* $(S_0, S_1, \dots)$ *is a supermartingale with* $S_0 \leqslant 0$ *a.s. such that for every* $i \in \{1, \dots, n\}$ *there exist non-random positive real numbers* b_i *and* c_i *such that a.s.*

$$X_i \leqslant b_i \text{ a.s.,} \tag{23}$$

$$\mathsf{Var}(X_i | H_{\leqslant(i-1)}) \leqslant c_i^2 \text{ a.s., and} \tag{24}$$

$$\frac{b_i^2}{c_i^2} \leqslant \frac{q}{p}. \tag{25}$$

Then inequality (22) *holds—again for any* $m \geqslant m_*(p)$ *and* $f \in \mathcal{F}_+^3$, *and again with* $s^{(m)}$ *defined by* (15).

Recall the definition of the *Schur majorizarion*: for $\mathbf{a} := (a_1, \dots, a_n)$ and $\mathbf{b} := (b_1, \dots, b_n)$ in $\mathbb{R}^n$, $\mathbf{a} \succcurlyeq \mathbf{b}$ means that $a_1 + \dots + a_n = b_1 + \dots + b_n$ and $a_{[1]} + \dots + a_{[j]} \geqslant b_{[1]} + \dots + b_{[j]}$ for all $j \in \{1, \dots, n\}$, where $a_{[1]} \geqslant \dots \geqslant a_{[n]}$ are the ordered numbers $a_1, \dots, a_n$, from the largest to the smallest. Recall also that a function $\mathcal{Q}\colon [0, \infty)^n \to \mathbb{R}$ is referred to as *Schur-concave* if it reverses the Schur majorization: for any $\mathbf{a}$ and $\mathbf{b}$ in $[0, \infty)^n$ such that $\mathbf{a} \succcurlyeq \mathbf{b}$, one has $\mathcal{Q}(\mathbf{a}) \leqslant \mathcal{Q}(\mathbf{b})$.

Theorems 3 and 4 are contained in the following theorem, which may thus be considered the main result of this paper.

Theorem 5 *The following statements are equivalent to one another.*

(I) $m \geqslant m_*(p)$.

(II) *For all* $f \in \mathcal{F}_+^3$, *all natural* $n \geqslant 2$, *and all nonnegative numbers* $c_1, \dots, c_n$, *one has* (17).

(III) *For every natural* $n \geqslant 2$ *and every function* $f \in \mathcal{F}_+^3$, *the function*

$$[0, \infty)^n \ni (a_1, \dots, a_n) \longmapsto \mathsf{E} f(a_1^{1/(2m)} \mathrm{BS}_1 + \dots + a_n^{1/(2m)} \mathrm{BS}_n) \tag{26}$$

is Schur-concave.

(IV) *Let the S_i's, A_{i-1}'s, B_{i-1}'s, c_i's, and $s^{(m)}$ be as in the statement of Theorem 4. Then one has (22) for all $f \in \mathcal{F}^3_+$.*

The special case of statement (III) of Theorem 5 with $p = \frac{1}{2}$ and $m = 1$ is essentially due to Whittle [31] and Eaton [9].

From the "right-tail" Theorem 5, one can deduce its left-tail and two-tail analogues. Appropriate left-tail and two-tail counterparts of $\mathcal{F}^3_+$ are the following classes of functions:

$$\mathcal{F}^3_- := \{f \in \mathcal{C}^2 \colon f \text{ and } f'' \text{ are nonincreasing and convex}\} \tag{27}$$
$$= \{f \colon \exists g \in \mathcal{F}^3_+ \; \forall x \in \mathbb{R} \; f(x) = g(-x)\} \quad \text{and}$$
$$\mathcal{F}^3 := \{f \in \mathcal{C}^2 \colon f \text{ and } f'' \text{ are convex}\}. \tag{28}$$

Remark 4 Theorem 5 holds with $\mathcal{F}^3_-$ in place of $\mathcal{F}^3_+$ if the restrictions that (i) $(S_0, S_1, \dots)$ is a supermartingale with $S_0 \leqslant 0$ a.s. and (ii) $\dfrac{B_{i-1}}{A_{i-1}} \leqslant \dfrac{q}{p}$ a.s. in Theorem 5 are replaced, respectively, with the following: (i) $(S_0, S_1, \dots)$ is a submartingale with $S_0 \geqslant 0$ a.s. and (ii) $\dfrac{A_{i-1}}{B_{i-1}} \leqslant \dfrac{q}{p}$ a.s. This "left-tail" analogue is a trivial corollary of Theorem 5.

The "two-tail" analogue of Theorem 5 is more difficult to prove. It relies in part on Proposition 1 below, preceded by the following definition.

Definition 1 Let us say that a sequence of functions (f_n) in $\mathcal{C}^2$ converges to a function f in $\mathcal{C}^2$ and write $f_n \to f$ (as $n \to \infty$) if $f_n(x) \uparrow f(x)$ and $f_n''(x) \to f''(x)$ for all real x. (This stronger notion of convergence will make it easier to verify the convergence of relevant expected values; also, it naturally provides for the relevant classes of functions to be closed.)

For any subset $\mathcal{A}$ of $\mathcal{C}^2$, its *closure*—denoted here by $\operatorname{cl}\mathcal{A}$—will be understood here simply as the set of the limits of all sequences in $\mathcal{A}$ that are convergent in $\mathcal{C}^2$. Obviously, $\operatorname{cl}\mathcal{A} \supseteq \mathcal{A}$, for every $\mathcal{A} \subseteq \mathcal{C}^2$.

Obviously, the "two-tail" class $\mathcal{F}^3$ contains both "one-tail" classes $\mathcal{F}^3_+$ and $\mathcal{F}^3_-$. The more informative relation of $\mathcal{F}^3$ to $\mathcal{F}^3_+$ and $\mathcal{F}^3_-$ (on which the proof of Corollary 2 below is partly based) is given by

Proposition 1 *One has $\mathcal{F}^3 = \operatorname{cl}\mathcal{G}^3$, where*

$$\mathcal{G}^3 := \{f \in \mathcal{C}^2 \colon \exists c \geqslant 0 \; \exists f_+ \in \mathcal{F}^3_+ \; \exists f_- \in \mathcal{F}^3_- \; \forall x \in \mathbb{R}$$
$$f(x) = c\,x^2/2 + f_+(x) + f_-(x)\}. \tag{29}$$

However, $\mathcal{F}^3 \neq \mathcal{G}^3$. For example, the function f defined by the formula

$$f(x) := \tfrac{8}{3}\,(1-x)^{3/2}\,\boldsymbol{I}\{x \leqslant 0\} + \left(\tfrac{8}{3} - 4x + x^2 + \tfrac{1}{6}\,x^3 + \tfrac{1}{16}\,x^4\right)\boldsymbol{I}\{x > 0\} \tag{30}$$

(for all $x \in \mathbb{R}$) belongs to $\mathcal{F}^3$ but not to $\mathcal{G}^3$. Here and elsewhere in the paper, $\boldsymbol{I}\{\cdot\}$ denotes the indicator function.

Moreover, functions $x \mapsto a + bx + cx^2 + d\,|x-t|^\alpha$, $x \mapsto \cosh \lambda x$, $x \mapsto \mathrm{e}^{\lambda x}$, $x \mapsto (x-t)_+^\alpha$, and $x \mapsto (t-x)_+^\alpha$ belong to $\mathcal{F}^3$ for all $a \in \mathbb{R}$, $b \in \mathbb{R}$, $c \geqslant 0$, $d \geqslant 0$, $t \in \mathbb{R}$, $\alpha \geqslant 3$, and $\lambda \in \mathbb{R}$. Note also that the classes $\mathcal{F}_+^3$, $\mathcal{F}_-^3$, and $\mathcal{F}^3$ are convex cones; that is, any linear combination with nonnegative coefficients of functions belonging to any one of these classes belongs to the same class.

Corollary 2 *Theorem 5 holds with $\mathcal{F}^3$ in place of $\mathcal{F}_+^3$ if the restrictions* (i) *$m \geqslant m_*(p)$,* (ii) *$(S_0, S_1, \dots)$ is a supermartingale with $S_0 \leqslant 0$ a.s., and* (iii) *$\frac{B_{i-1}}{A_{i-1}} \leqslant \frac{q}{p}$ a.s. in Theorem 5 are replaced, respectively, with the following stronger restrictions*:

(i) *$m \geqslant m_*(p)$ and $p \leqslant \frac{1}{2}$,*
(ii) *$(S_0, S_1, \dots)$ is a martingale with $S_0 = 0$ a.s., and*
(iii) $\max\left(\dfrac{B_{i-1}}{A_{i-1}}, \dfrac{A_{i-1}}{B_{i-1}}\right) \leqslant \dfrac{q}{p}$ a.s.

Using Theorem 4, Corollary 1, and Remark 1 (and in view of definition (12) and Lemma 2 on page 619), one immediately obtains the following bounds on the tail probabilities of S_n, which may be compared with those provided by Corollary 2.2 in [24].

Corollary 3 *Suppose that the conditions of Theorem 4 or the conditions of Corollary 1 hold. Let*

$$T_n := s^{(m)}(\mathrm{BS}_1 + \cdots + \mathrm{BS}_n),$$

where $s^{(m)}$ is still defined by (15). Then for all $m \geqslant m_*(p)$ and all real x

$$\mathsf{P}(S_n \geqslant x) \leqslant \inf_{t \in (-\infty, x)} \frac{\mathsf{E}(T_n - t)_+^3}{(x-t)^3} \tag{31}$$

$$\leqslant c_{3,0}\, \mathsf{P}^{\mathsf{LC}}(T_n \geqslant x), \tag{32}$$

where $x \mapsto \mathsf{P}^{\mathsf{LC}}(T_n \geqslant x)$ is the least log-concave majorant of the function $x \mapsto \mathsf{P}(T_n \geqslant x)$ on $\mathbb{R}$.

According to Remark 1, the upper bound in (31) is majorized by the exponential upper bound (7), so that under the conditions of Corollary 3 one has

$$\mathsf{P}(S_n \geqslant x) \leqslant \mathrm{e}^{-nH} \tag{33}$$

for all real x, where $H := (p+y)\ln\frac{p+y}{p} + (q-y)\ln\frac{q-y}{q}$ if $0 \leqslant y := \frac{x}{n}\frac{\sqrt{pq}}{s^{(m)}} < q$, $H := -\ln p$ if $y = q$, $H := \infty$ if $y > q$, and $H := 0$ if $y < 0$. This exponential upper bound is essentially due to Hoeffding [13].

Note that $\mathsf{P}^{\mathsf{LC}}(T_n \geqslant x) = \mathsf{P}(T_n \geqslant x)$ for all x in the lattice

$$L := \{nb + kh \colon k \in \mathbb{Z}\}$$

generated by the support of the distribution of T_n, where $b := s^{(m)}\sqrt{\frac{q}{p}}$ and $h := s^{(m)}/\sqrt{pq}$. Using also results of [23], one immediately has the following.

Corollary 4 *Under the conditions of Corollary 3,*

$$\mathsf{P}(S_n \geqslant x) \leqslant c_{3,0}\, \mathsf{P}^{\mathsf{Lin,LC}}(T_n \geqslant x + \tfrac{h}{2}) \quad \forall x \in \mathbb{R}, \tag{34}$$

where $x \mapsto \mathsf{P}^{\mathsf{Lin,LC}}(T_n \geqslant x)$ is the least log-concave majorant of the linear interpolation of the tail function $x \mapsto \mathsf{P}(T_n \geqslant x)$ over the lattice L.

The upper bound in (34) usually works better in statistical practice than those in (32) and (33). An explicit formula for $\mathsf{P}^{\mathsf{Lin,LC}}(T_n \geqslant x + \frac{h}{2})$ is given in [23].

Corollary 3 should be compared with Theorem 1.3 of [3], which states that, if (S_i) is a martingale with $S_0 = 0$ satisfying conditions (23) and (24) with the additional restriction

$$b_i = c_i \quad \forall i = 1, \ldots, n,$$

in place of (25), then $\forall x \in \mathbb{R}$

$$\mathsf{P}(S_n \geqslant x) \leqslant c_{3,0}\, \mathsf{P}^{\mathsf{LC}}(T_n \geqslant x) \quad \text{with } T_n = s^{(1)}\cdot(\varepsilon_1 + \cdots + \varepsilon_n). \tag{35}$$

Thus, this result of [3] is a special case of Corollary 3—with $p = \frac{1}{2}$ (so that one may take $m = 1$). By the central limit theorem, inequality (35) implies

$$\mathsf{P}(S_n \geqslant x) \leqslant c_{3,0}\, \mathsf{P}(s^{(1)}\sqrt{n}Z \geqslant x). \tag{36}$$

Inequalities (36) and (35) are extensions of results in [19,20]. A version of inequality (36), with constant factor $1/\mathsf{P}(Z > \sqrt{3}) = 24.01\ldots$ in place of $c_{3,0} = 2e^3/9 = 4.46\ldots$, appeared earlier in [2]. The generalized moments $\mathsf{E} f(T_n)$ in the above upper bounds, where $T_n := s^{(m)} \cdot (\mathrm{BS}_1 + \cdots + \mathrm{BS}_n)$, can be replaced by $\mathsf{E} f(s^{(1)}\sqrt{n}Z)$ provided that $p \geqslant \frac{1}{2}$. That $(S_0, S_1, \ldots)$ is allowed to be a supermartingale (rather than only a martingale) makes it convenient to use the simple but powerful truncation tool; cf. the discussion at the end of Sect. 2 in [24].

We shall also prove the following stronger, "maximal" version of the previous results.

Corollary 5 *One can replace S_n in the left-hand side of inequalities* (31) *and* (34) *by*

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

3 Applications: bounds on self-normalized sums

(Details on the results presented in this section are given in [27].)

Efron [11] considered the so-called self-normalized sum

$$V := \frac{X_1 + \cdots + X_n}{\sqrt{X_1^2 + \cdots + X_n^2}}, \tag{37}$$

assuming that the X_i's satisfy the orthant symmetry condition: the joint distribution of $\delta_1 X_1, \ldots, \delta_n X_n$ is the same for any choice of signs $\delta_1, \ldots, \delta_n \in \{1, -1\}$, so that, in particular, each X_i is symmetric(ally distributed). It suffices that the X_i's be independent and symmetrically (but not necessarily identically) distributed. On the event $\{X_1 = \cdots = X_n = 0\}$, let $V := 0$.

Following Efron [11], note that the conditional distribution of any symmetric r.v. X given $|X|$ is the symmetric distribution on the (at most) two-point set $\{X, -X\}$. Therefore, under the orthant symmetry condition, the distribution of V is the mixture of the distributions of the normalized Khinchin–Rademacher sums $a_1\varepsilon_1 + \cdots + a_n\varepsilon_n$, where the ε_i's are independent of the X_i's and $a_i = X_i/(X_1^2 + \cdots + X_n^2)^{\frac{1}{2}}$, so that $a_1^2 + \cdots + a_n^2 = 1$ (except on the event $\{X_1 = \cdots = X_n = 0\}$, where $a_1 = \cdots = a_n = 0$). Hence, using the well known bound $\mathsf{E}\exp\left\{\lambda\,(a_1\varepsilon_1 + \cdots + a_n\varepsilon_n)\right\} \leqslant \mathrm{e}^{\lambda^2/2}$ (see e.g. the introduction in [24]) one has

$$\mathsf{E}\,\mathrm{e}^{\lambda V} \leqslant \mathsf{E}\,\mathrm{e}^{\lambda Z} \tag{38}$$

for all $\lambda \in \mathbb{R}$, whence

$$\mathsf{P}\,(V \geqslant x) \leqslant \mathrm{e}^{-x^2/2} \quad \forall x \geqslant 0. \tag{39}$$

These results can be easily restated in terms of Student's statistic T, which is a monotonic function of V, as noted by Efron; namely, $T = \sqrt{\frac{n-1}{n}}\, V/\sqrt{1 - V^2/n}$. Inequalities (38) and (39) were improved in [9,10,19] as follows:

$$\mathsf{E} f(V) \leqslant \mathsf{E} f(Z) \quad \forall f \in \mathcal{H}_+^3 \tag{40}$$

and

$$\mathsf{P}\,(V \geqslant x) \leqslant \frac{2\mathrm{e}^3}{9}\,\mathsf{P}(Z \geqslant x) \quad \forall x \in \mathbb{R}. \tag{41}$$

Multivariate analogues of these results, which can be expressed in terms of Hotelling's statistic in place of Student's, were also obtained in [19].

It was pointed out in [19, Theorem 2.8] that, since the normal tail decreases fast, inequality (41) implies that relevant quantiles of V may exceed the corresponding standard normal quantiles only by a relatively small amount, so that one can use (41) rather efficiently to test symmetry even for non-i.i.d. observations.

Here we shall present extensions of inequalities (40) and (41) to the case when the X_i's are not symmetric.

Our basic idea is to represent any zero-mean, possibly asymmetric distribution as an appropriate mixture of two-point zero-mean distributions. Let us assume at first that a zero-mean r.v. X has an everywhere strictly positive density function. Consider the truncated r.v. $X_{a,b} := X\boldsymbol{I}\{a \leqslant X \leqslant b\}$. Then, for every fixed $a \in (-\infty, 0]$, the function $b \mapsto \mathsf{E}\, X_{a,b}$ is continuous and increasing on the interval $[0,\infty)$ from $\mathsf{E}\, X_{a,0} \leqslant 0$ to $\mathsf{E}\, X_{a,\infty} > 0$. Hence, for each $a \in (-\infty, 0]$, there exists a unique value $b \in [0,\infty)$ such that $\mathsf{E}\, X_{a,b} = 0$. Similarly, for each $b \in [0,\infty)$, there exists a unique value $a \in (-\infty, 0]$ such that $\mathsf{E}\, X_{a,b} = 0$. That is, one has a one-to-one correspondence between $a \in (-\infty, 0]$ and $b \in [0,\infty)$ such that $\mathsf{E}\, X_{a,b} = 0$. Denote by $\mathsf{r} := \mathsf{r}_X$ the *reciprocating* function defined on $\mathbb{R}$ and carrying this correspondence, so that

$$\mathsf{E}\, X\boldsymbol{I}\{X \text{ is between } x \text{ and } \mathsf{r}(x)\} = 0 \quad \forall x \in \mathbb{R};$$

the function r is decreasing on $\mathbb{R}$ and such that $\mathsf{r}(\mathsf{r}(x)) = x\ \forall x \in \mathbb{R}$; moreover, $\mathsf{r}(0) = 0$. (Clearly, $\mathsf{r}(x) = -x$ for all real x if the r.v. X is symmetric.) One also has

$$\mathsf{r}(x) = x_-(G(x))\boldsymbol{I}\{x > 0\} + x_+(G(x))\boldsymbol{I}\{x < 0\}, \tag{42}$$

where $x_\pm(h)$ stand for the positive and negative roots x of the equation $G(x) = h$ and, in turn,

$$G(x) := \mathsf{E}\, |X|\boldsymbol{I}\{|X| \leqslant |x|,\ \operatorname{sign} X = \operatorname{sign} x\}. \tag{43}$$

Thus, the set $\{\{x, \mathsf{r}(x)\} \colon x \in \mathbb{R}\}$ of (at-most-)two-point sets constitutes a partition of $\mathbb{R}$. Moreover, the two-point set $\{x, \mathsf{r}(x)\}$ is uniquely determined by the distance $|x - \mathsf{r}(x)|$ between the two points, as well as by the product $|x|\,|\mathsf{r}(x)|$. Now one can see that the conditional distribution of the zero-mean r.v. X given $W := |X - \mathsf{r}(X)|$ (or, equivalently, $Y := |X\,\mathsf{r}(X)|$) is the uniquely determined zero-mean distribution on the two-point set $\{X, \mathsf{r}(X)\}$. Thus, the distribution of the zero-mean r.v. X with an everywhere positive density is represented as a mixture of two-point zero-mean distributions. This mixture is given rather explicitly, provided that the distribution of r.v. X is known.

Thus, one can introduce the following generalized versions of the self-normalized sum (37), which require—instead of the symmetry of independent r.v.'s X_i—only that the X_i's be zero-mean:

$$V_W := \frac{X_1 + \cdots + X_n}{\frac{1}{2}\sqrt{W_1^2 + \cdots + W_n^2}} \quad \text{and} \quad V_{Y,m} := \frac{X_1 + \cdots + X_n}{(Y_1^m + \cdots + Y_n^m)^{\frac{1}{2m}}}, \tag{44}$$

where $m \geqslant 1$,

$$W_i := |X_i - \mathsf{r}_i(X_i)|, \quad \text{and} \quad Y_i := |X_i\,\mathsf{r}_i(X_i)|,$$

and the reciprocating function $\mathsf{r}_i := \mathsf{r}_{X_i}$ is constructed as above, based on the distribution of X_i, for each i, so that the reciprocating functions r_i may be different from one another if the X_i's are not identically distributed. On the event $\{X_1 = \cdots = X_n = 0\}$ (which is the same as either one of events $\{W_1 = \cdots = W_n = 0\}$ and $\{Y_1 = \cdots = Y_n = 0\}$), let $V_W := 0$ and $V_{Y,m} := 0$. Note that $V_W = V_{Y,1} = V$ when the X_i's are symmetric. Logan et al. [16] and Shao [28] obtained limit theorems for the "symmetric" version of $V_{Y,m}$ (with $Y_i = X_i^2$).

These constructions can be extended to the general case of any zero-mean r.v. X, absolutely continuous or not. Here, one can use randomization (by means of a r.v. uniformly distributed in interval $(0,1)$) to deal with the atoms of the distribution of r.v. X, and a modification of the inverse functions $x_\pm(h)$ to deal with the intervals on which the distribution function of X and hence the function G are constant. Namely, in general $\mathsf{r}(X)$ is replaced by $\mathsf{r}(X,U)$, where U is a r.v. uniformly distributed in interval $(0,1)$ and independent of X and, for $x \in \mathbb{R}$ and $u \in (0,1)$,

$$\mathsf{r}(x,u) := \begin{cases} x_-(G(x-) + u \cdot (G(x) - G(x-))) & \text{if } x \in [0,\infty), \\ x_+(G(x+) + u \cdot (G(x) - G(x+))) & \text{if } x \in (-\infty,0], \end{cases}$$
$$x_+(h) := \inf\{x \in [0,\infty] \colon G(x) \geqslant h\},$$
$$x_-(h) := \sup\{x \in [-\infty,0] \colon G(x) \geqslant h\}.$$

It follows that, conditionally on the Y_i's (or, equivalently, on the W_i's), the X_i's are independent zero-mean r.v.'s, and the conditional distribution of each X_i is supported by the two-point set $\{X_i, \mathsf{r}(X_i)\}$. Thus, the distribution of each X_i is a mixture of zero-mean two-point distributions on the sets of the form $\{x_i, \mathsf{r}_i(x_i)\}$ for $x_i \in [0,\infty)$. Let us now use Theorem 4 and Corollary 3 with each supermartingale-difference X_i there replaced by a zero-mean r.v. taking on two values, $(-A_{i-1})$ and B_{i-1}, where

$$A_{i-1} := -\frac{\mathsf{r}_i(x_i)}{(y_1^m + \cdots + y_n^m)^{\frac{1}{2m}}}, \quad B_{i-1} := \frac{x_i}{(y_1^m + \cdots + y_n^m)^{\frac{1}{2m}}},$$

$y_i := |x_i \mathsf{r}_i(x_i)|$, and $x_i \in [0,\infty)$; at that, let $c_i := \sqrt{A_{i-1}B_{i-1}}$. Then one obtains

Corollary 6 *Suppose that for some $p \in (0,1)$ and all $i \in \{1,\ldots,n\}$*

$$\frac{X_i}{|\mathsf{r}_i(X_i)|} \boldsymbol{I}\{X_i > 0\} \leqslant \frac{q}{p} \text{ a.s.} \tag{45}$$

Then for all $m \geqslant m_*(p)$

$$\mathsf{E} f(V_{Y,m}) \leqslant \mathsf{E} f(T_n) \quad \forall f \in \mathcal{F}_+^3 \quad \text{and} \tag{46}$$
$$\mathsf{P}(V_{Y,m} \geqslant x) \leqslant c_{3,0}\, \mathsf{P}^{\mathsf{LC}}(T_n \geqslant x) \quad \forall x \in \mathbb{R}, \tag{47}$$

where T_n and $\mathsf{P}^{\mathsf{LC}}(T_n \geqslant x)$ have the same meaning as in Corollary 3, with $s^{(m)} = n^{-1/(2m)}$, and, in accordance with (6), $c_{3,0} = 2e^3/9 = 4.4634\ldots$.

Similarly, using the W_i's instead of the Y_i's and results of [24] instead of Theorem 4 and Corollary 3 of this paper, one has

Corollary 7

$$\mathsf{E} f(V_W) \leqslant \mathsf{E} f(Z) \quad \forall f \in \mathcal{H}^5_+, \quad \textit{whence} \tag{48}$$

$$\mathsf{P}(V_W \geqslant x) \leqslant c_{5,0}\,\mathsf{P}(Z \geqslant x) \quad \forall x \in \mathbb{R}, \tag{49}$$

where, in accordance with (6), $c_{5,0} = 5!(e/5)^5 = 5.699\ldots$.

Of course, one can replace the upper bound in inequalities like (47) by either of the more precise (but slightly less transparent and more difficult to compute) upper bounds given in (31) and (34).

Condition (45) is likely to hold when the X_i's are bounded i.i.d. r.v.'s.

Note that the reciprocating function r depends on the (usually unknown in statistics) distribution of the underlying r.v. X. However, if, e.g. the X_i's constitute an i.i.d. sample, then the function G defined by (43) can be estimated based on the sample, so that one can estimate the reciprocating function r. Thus, replacing $X_1+\cdots+X_n$ in the numerators of V_W and $V_{Y,m}$ by $X_1+\cdots+X_n-n\theta$, one obtains approximate pivots to be used to construct approximate confidence intervals or, equivalently, approximate tests for an unknown mean θ. One can also use bootstrap to estimate the distributions of such pivots.

Inequalities presented in Corollaries 6 and 7 may be compared with limit theorems for large deviations, such as the ones given in [14,28]. In addition to the obvious distinction that limit theorems provide asymptotics, rather than certain and explicit bounds, one may notice that in most cases the limit theorems are obtained for i.i.d. X_i's. Therefore, the inequalities may turn useful for small and/or heteroscedastic samples, as they provide universal conservative bounds, which may be rather accurate. Another advantage of our approach is that it provides one with generalized moment comparison inequalities (such as (46) or (48)) for rich classes of generalized moment functions f. Exponential and power-moment inequalities for self-normalized sums were obtained in [7].

4 Proofs

4.1 Statements of lemmas and proofs of the main results

The proofs of the main results are preceded in this section by some definitions and a series of lemmas. At least one of them (Lemma 12) may be of independent interest. The proofs of the lemmas are deferred further to Sect. 4.2.

Let us introduce more classes of functions, in addition to the classes $\mathcal{H}^3_+$, $\mathcal{F}^3_+$, $\mathcal{F}^3_-$, $\mathcal{F}^3$, and $\mathcal{G}^3$ (recall (1), (12), (27), (28), and (29)):

$$\mathcal{G}^3_+ := \{f\colon \exists a \in \mathbb{R}, b \in \mathbb{R}, h \in \mathcal{H}^3_+ \ \forall x \in \mathbb{R} \quad f(x) = a + b\,x + h(x)\}; \tag{50}$$

$$\mathcal{G}^3_{++} := \{f\colon \exists a \in \mathbb{R}, b \geqslant 0, h \in \mathcal{H}^3_+ \ \forall x \in \mathbb{R} \quad f(x) = a + b\,x + h(x)\}. \tag{51}$$

Remark 5 It is not difficult to see that, if a function f is in $\mathcal{F}^3_+$ or any other defined above class of functions, then the shifted function $x \mapsto f(x+a)$ is also in the same class, for any real constant a. That is, all these classes of functions are shift-invariant.

Our first lemma is very simple and probably well known, but I have not been able to find it in the literature. So, it is stated (and proved) here for the readers' convenience and easy reference.

Lemma 1 *Suppose that a function $g\colon \mathbb{R} \to \mathbb{R}$ is convex and such that there exists a finite limit $g(-\infty) := \lim_{x\to-\infty} g(x)$; in particular, the latter condition will obviously be the case if g is nonnegative and nondecreasing on $\mathbb{R}$. Then $g'(-\infty) = 0$, where g' is the right derivative of g.*

Lemma 2 *If α is a natural number then $\mathcal{H}^\alpha_+$ coincides with the class $\tilde{\mathcal{H}}^\alpha_+$ of all functions $f\colon \mathbb{R} \to \mathbb{R}$ such that the derivative $f^{(\alpha-1)}$ is everywhere finite and convex, and $f^{(0)}(-\infty) = \cdots = f^{(\alpha-1)}(-\infty) = 0$. Moreover, if $f \in \mathcal{H}^\alpha_+$, then all the functions $f^{(0)}, \dots, f^{(\alpha-1)}$ are nonnegative.*

Lemma 3 *Let $f\colon \mathbb{R} \to \mathbb{R}$ be a function such that f'' is finite, nonnegative, nondecreasing, and convex, with $f''(-\infty) = 0$. Then $f \in \operatorname{cl} \mathcal{G}^3_+$. If, moreover, f is nondecreasing, then $f \in \operatorname{cl} \mathcal{G}^3_{++}$.*

Lemma 4 *One has $\mathcal{F}^3_+ = \operatorname{cl} \mathcal{G}^3_{++}$.*

Lemma 5 *One has $\mathcal{F}^3 = \operatorname{cl} \mathcal{G}^3$, where $\mathcal{G}^3$ is defined by* (29).

Lemma 6 *If $f \in \mathcal{F}^3_+$, then either $f(x) = O(x)$ as $x \to \infty$ or*
$\liminf_{x\to\infty} f(x)/x^2 \in (0,\infty]$.

Lemma 7 *If $f \in \mathcal{F}^3_+$, then $f(x) = O(|x|)$ as $x \to -\infty$.*

Lemma 8 $\mathcal{F}^3 \neq \mathcal{G}^3$.

Proposition 2 *There exists a function $g \in \mathcal{F}^3_+ \setminus \mathcal{G}^3_+$. (For example, one can let $g := f'$, where f is defined by* (30).*) Since $\mathcal{G}^3_{++} \subseteq \mathcal{G}^3_+$, it follows that $\mathcal{F}^3_+ \neq \mathcal{G}^3_{++}$.* (This proposition complements Lemmas 4 and 8.)

The following two lemmas are essentially well known. Their statements (and proofs) are given here for easy reference.

Lemma 9 (Cf. e.g. [15] and [3, Lemma 4.3].) *Let X be a r.v. such that $\mathsf{E}\,X \leqslant 0$ and $-a \leqslant X \leqslant b$ a.s. for some positive real numbers a and b. Let $\mathrm{BS} \sim \mathsf{BS}(p)$ with*

$$p := \frac{a}{b+a}.$$

Then

$$\mathsf{E}\,f(X) \leqslant \mathsf{E}\,f(\sqrt{a\,b}\,\mathrm{BS})$$

for any nondecreasing convex function f, and hence for any function $f \in \mathcal{F}^3_+$.

Lemma 10 *If X is a zero-mean r.v., then $\mathsf{E} f(cX)$ is nondecreasing in $c \geqslant 0$ for any convex function f and hence for any $f \in \mathcal{F}^3_+$.*

Lemma 11 *Let $\mathrm{BS}(p) \sim \mathsf{BS}(p)$. Then $e(p) := \mathsf{E} f(\mathrm{BS}(p))$ is nonincreasing in $p \in (0,1)$ for any $f \in \mathcal{H}^2_+$ and so, by (2), for any $f \in \mathcal{H}^3_+$, whence, by Lemma 4, for any $f \in \mathcal{F}^3_+$.*

The extension from Theorem 3 to Theorem 4 is based in part on the following simple lemma, which may be of independent interest.

Lemma 12 *Suppose that $(S_0, S_1, \dots)$ is a supermartingale with $S_0 \leqslant 0$ a.s. such that for every $i \in \{1, \dots, n\}$ one has (18), (19), and (20). Then*

$$\mathsf{E} f(S_n) \leqslant \mathsf{E} f(c_1 \mathrm{BS}_1 + \cdots + c_n \mathrm{BS}_n) \tag{52}$$

for any $f \in \mathcal{H}^2_+$, and so, by (2), for any $f \in \mathcal{H}^3_+$, whence, by Lemma 4 and Lebesgue's dominated convergence theorem, for any $f \in \mathcal{F}^3_+$.

For $m \geqslant 1$, introduce

$$\begin{aligned} p_* := p_*(m) &:= \frac{2m + 1 - \sqrt{4(m-1)(m+2)+1}}{4(2m-1)} \\ &= \frac{2}{(2m-1)\left(2m+1+\sqrt{4(m-1)(m+2)+1}\right)}, \end{aligned} \tag{53}$$

so that $p_* \in (0, \frac{1}{2}]$. Introduce also

$$\begin{aligned} \delta_1(u,c,p,m) &:= 2c(1-c^{2m-2})u + 2pc(1-c^{2m-1}) + c^2(1-c^{2m-3}); \\ \delta_2(u,c,p,m) &:= (1-p)(1-c^{2m-1})u^2 \\ &\quad + 2c(1-c^{2m-2})u + 2pc(1-c^{2m-1}) + c^2(1-c^{2m-3}); \\ \delta_3(u,c,p,m) &:= -c^{2m-1}u^2 - 2(c^{2m-1} - cp + c^{2m}p)\,u \\ &\quad + \big((2c + c^2 - 2c^{2m} - c^{2m+1})p - c^{2m-1}\big); \\ \delta_4(u,c,p,m) &:= (1-c^{2m-1})\,p\,(1+c+u)^2. \end{aligned}$$

Lemma 13 *For any given pair (p,m) such that $m \geqslant 1$ and $p \in (0,1)$, the following two statements are equivalent to each other.*

(i) *For every every function $f \in \mathcal{H}^3_+$, the function*

$$[0,\infty)^2 \ni (a_1, a_2) \longmapsto \mathsf{E} f\big(a_1^{1/(2m)} \mathrm{BS}_1 + a_2^{1/(2m)} \mathrm{BS}_2\big),$$

is Schur-concave, where $\mathrm{BS}_i \overset{\text{i.i.d.}}{\sim} \mathsf{BS}(p)$, $i = 1, 2$.

(ii) *for all* $u \in \mathbb{R}$ *and* $c \in (0,1)$*, one has the inequalities*

$$\delta_1(u,c,p,m)\, \boldsymbol{I}\{u \geqslant 0\} \geqslant 0;$$
$$\delta_2(u,c,p,m)\, \boldsymbol{I}\{-c \leqslant u \leqslant 0\} \geqslant 0;$$
$$\delta_3(u,c,p,m)\, \boldsymbol{I}\{-1 \leqslant u \leqslant -c\} \geqslant 0.$$

Lemma 14 *For all* $u \geqslant 0$*,* $c \in (0,1)$*,* $p \in [p_*,1)$*, and* $m \geqslant 1$*, one has* $\delta_1(u) := \delta_1(u,c,p,m) \geqslant 0$.

Lemma 15 *For all* $u \in [-c,0]$*,* $c \in (0,1)$*,* $p \in [p_*,1)$*, and* $m \geqslant 1$*, one has* $\delta_2(u) := \delta_2(u,c,p,m) \geqslant 0$.

Lemma 16 *For all* $u \in [-1,-c]$*,* $c \in (0,1)$*,* $p \in [p_*,1)$*, and* $m \geqslant 1$*, one has* $\delta_3(u) := \delta_3(u,c,p,m) \geqslant 0$.

Lemma 17 (Recall (14) and (53).) *For* $p \in (0,1)$ *and* $m \geqslant 1$*, one has*

$$m \geqslant m_*(p) \iff p \geqslant p_*(m).$$

Lemma 18 *In the context of Theorem 5, implication* (II) $\Longrightarrow$ (I) *is true.*

Proof of Corollary 1 According to Theorem 2.1 in [23] (cf. [1,3,4]), conditions (23) and (24) imply that

$$\mathsf{E} f(S_n) \leqslant \mathsf{E} f(Z_1 + \cdots + Z_n) \quad \forall f \in \mathcal{H}_+^2, \tag{54}$$

where $Z_1, \ldots, Z_n$ are independent zero-mean r.v.'s such that each Z_i takes on only two values, $B_{i-1} := b_i$ and $-A_{i-1} := -c_i^2/b_i$, so that condition (18) of Theorem 4 is satisfied with Z_i in place of X_i, and at that conditions (19) and (20) hold (the latter one by virtue of (25)). Thus and in view of inequality (54), set inclusion (2), and Theorem 4, one has the inequality (22) of Corollary 1 for all $f \in \mathcal{H}_+^3$, and so, by Lemma 4 and Lebesgue's dominated convergence theorem, for all $f \in \mathcal{F}_+^3$. This completes the proof of Corollary 1. (Note: Theorem 4, used in this proof, follows immediately from Theorem 5, which will be proved next, without using Corollary 1.) □

Proof of Theorem 5 It suffices to prove the implications

$$\text{(I)} \Longrightarrow \text{(III)} \Longrightarrow \text{(II)} \Longrightarrow \text{(I)} \quad \text{and}$$
$$\text{(II)} \Longrightarrow \text{(IV)} \Longrightarrow \text{(II)}. \tag{55}$$

(I) $\Longrightarrow$ (III): Suppose that condition $m \geqslant m_*(p)$ of item (I) takes place. By Lemma 17, this condition is equivalent to $p \geqslant p_*(m)$. Now statement (III) for $n = 2$ with $\mathcal{H}_+^3$ in place of $\mathcal{F}_+^3$ follows from Lemmas 13, 14, 15, and 16. Hence, by the definition (51) of $\mathcal{G}_{++}^3$ and Lemma 4, one has (III) for $n = 2$ (recall that the r.v.'s BS_i each take on only finitely many, namely two, values).

Now, in view of Remark 5 on page 619, by conditioning on all of the r.v.'s $\mathrm{BS}_1, \dots, \mathrm{BS}_n$ except any given two of them, (III) for any natural $n \geqslant 2$ follows from the well-known result by Muirhead [18] (see, e.g., [17, Remark B.1 of Chap. 2]), which states that a function of n nonnegative arguments is Schur-concave iff it is Schur-concave in any two of its arguments.

(III) $\Longrightarrow$ (II): Let here $a_1 := c_1^{2m}, \dots, a_n := c_n^{2m}$ and $\overline{a} := (a_1 + \cdots + a_n)/n$, so that $s^{(m)} = \overline{a}^{\,1/(2m)}$. Note that $(a_1, \dots, a_n) \succcurlyeq (\underbrace{\overline{a}, \dots, \overline{a}}_{n})$. Now implication (III) $\Longrightarrow$ (II) follows.

(II) $\Longrightarrow$ (I): This implication is true by Lemma 18.

(II) $\Longrightarrow$ (IV): This implication follows immediately from Lemma 12.

(IV) $\Longrightarrow$ (II): This implication is trivial. □

Proof of Proposition 1 This follows immediately from Lemmas 5 and 8. □

Proof of Corollary 2 It suffices to prove the same implications, (55), as in the proof of Theorem 5, only with the changes stated in the formulation of Corollary 2. Below, all these implications are understood in the context of Corollary 2. The proofs of most of these implications are similar to their proofs in the context of Theorem 5. Below, only the most significant changes are described.

(I) $\Longrightarrow$ (III): To prove this implication, in view of Theorem 5, Remark 4, and Proposition 1, it suffices to verify that the function (26) is Schur-concave when $f(x) = x^2$ (for all real x). Also (cf. the proof of implication (I) $\Longrightarrow$ (III) in the proof of Theorem 5), one may assume that $n = 2$. Thus, it suffices to verify that, for any given $m \geqslant 1$, the expression

$$\mathsf{E}(\mathrm{BS}_1 \cos^{1/m}\theta + \mathrm{BS}_2 \sin^{1/m}\theta)^2 = \cos^{2/m}\theta + \sin^{2/m}\theta$$

is nondecreasing in $\theta \in [0, \pi/4]$. But this is easy to see.

(II) $\Longrightarrow$ (I): This implication follows from Theorem 5, Remark 4, and the observation that both classes $\mathcal{F}_+^3$ and $\mathcal{F}_-^3$ are contained in $\mathcal{F}^3$.

(II) $\Longrightarrow$ (IV): In view of Theorem 5, Remark 4, Proposition 1, and Lebesgue's dominated convergence theorem, it suffices to verify that inequality (22) holds when $f(x) = x^2$ for all real x and $(S_0, \dots, S_n)$ is a martingale as described in the formulation of Corollary 2. Note that inequality (52) holds for the function $f_0(x) := x_+^2$ in place of f, since $f_0 \in \mathcal{H}_+^2$. By symmetry, (52) also holds for the function $\tilde{f}_0(x) := (-x)_+^2$ in place of f, given that $(S_0, \dots, S_n)$ is a martingale as described. Thus, (52) holds when $f(x) = x^2\big(= x_+^2 + (-x)_+^2\big)$ for all real x. It remains to note that

$$\mathsf{E}(c_1\mathrm{BS}_1 + \cdots + c_n\mathrm{BS}_n)^2 = n\,(s^{(1)})^2 \leqslant n\,(s^{(m)})^2 = \mathsf{E}(s^{(m)}(\mathrm{BS}_1 + \cdots + \mathrm{BS}_n))^2,$$

so that one does have inequality (22) when $f(x) = x^2$ for all real x. □

Proof of Corollary 5 In the case when (S_i) is a martingale, one can obtain Corollary 5 similarly to Corollary 3, using the Doob inequality $\mathsf{P}(M_n \geqslant x) \leqslant \mathsf{E}(S_n - t)_+^3/(x - t)^3$ for the convex function $u \mapsto f_t(u) := (u - t)_+^3$ and $t < x$.

It remains to observe that any supermartingale-differences X_i satisfying condition (18) are majorized by certain martingale-differences $\tilde{X}_i$ satisfying (18) (in place of X_i) as well. Such a reduction is similar to that in the proof of Lemma 3.1 in [23]. Namely, for each $i = 1, 2, \dots$ let

$$\tilde{X}_i := (1-\gamma_{i-1})X_i + \gamma_{i-1}B_{i-1}, \quad \text{where } \gamma_{i-1} := \frac{\mathsf{E}_{i-1} X_i}{\mathsf{E}_{i-1} X_i - B_{i-1}}$$

and E_j denotes the conditional expectation given $H_{\leqslant j}$; then γ_{i-1} is $H_{\leqslant(i-1)}$-measurable; hence, $\mathsf{E}_{i-1} \tilde{X}_i = 0$ a.s.; the conditions $B_{i-1} > 0$ and $\mathsf{E}_{i-1} X_i \leqslant 0$ a.s. imply that $\gamma_{i-1} \in [0, 1)$ a.s.; hence, condition (18) yields $\tilde{X}_i \in [X_i, B_{i-1}] \subseteq [-A_{i-1}, B_{i-1}]$ a.s. Thus, the $\tilde{X}_i$'s are martingale-differences satisfying the same condition (18), while $\tilde{X}_i \geqslant X_i$ a.s. for each i, so that $S_n \leqslant \tilde{X}_1 + \cdots + \tilde{X}_n$ a.s. □

4.2 Proofs of the lemmas

Symbolic calculations in the proofs of some lemmas (especially Lemmas 13, 15, 17, and 18) are rather involved and better done with Mathematica or similar software.

Proof of Lemma 1 The convexity of g implies $g(0) - g(x) \geqslant g'(x)(-x)$ and $g(2x) - g(x) \geqslant g'(x)x$, so that $|g'(x)||x| \leqslant \max(|g(0) - g(x)|, |g(2x) - g(x)|)$ for all $x \in \mathbb{R}$. Letting now $x \to -\infty$ and using the existence of the finite limit $g(-\infty)$, one has $g'(-\infty) = 0$. □

Proof of Lemma 2 This lemma was stated essentially as Proposition 1.1 in [23]. The proof given here is a little more detailed. Assume first that $f \in \mathcal{H}_+^\alpha$, so that $f(x) = \int (x-t)_+^\alpha \, \mathrm{d}\mu(t)$ for a function $\mu \in \mathcal{M}_+^\alpha$, whence $f^{(\alpha-1)}(x) = \alpha! \int (x - t)_+ \, \mathrm{d}\mu(t)$ is convex as a limit of linear combinations with nonnegative coefficients of convex functions $x \mapsto (x-t)_+$. The conditions $f^{(0)}(-\infty) = \cdots = f^{(\alpha-1)}(-\infty) = 0$ follow by Lebesgue's dominated convergence theorem. Thus, $f \in \tilde{\mathcal{H}}_+^\alpha$. Moreover, it is clear that all the functions $f^{(0)}, \dots, f^{(\alpha-1)}$ are nonnegative.

Assume now that $f \in \tilde{\mathcal{H}}_+^\alpha$. Consider first the case $\alpha = 1$. Then f is convex and $f(-\infty) = 0$. Hence, by Lemma 1, one has $f'(-\infty) = 0$. Therefore,

$$f(x) = \int_{-\infty}^{x} f'(u)\,\mathrm{d}u = \int_{-\infty}^{x} \mathrm{d}u \int_{-\infty}^{u} \mathrm{d}f'(v) = \int (x-v)_+ \,\mathrm{d}\,\mu(v),$$

by Fubini's theorem, where $\mu := f'$; thus, $f \in \mathcal{H}_+^1$. The case of any natural $\alpha \geqslant 2$ can now be treated by induction, in a similar manner. Indeed, if $f \in \tilde{\mathcal{H}}_+^\alpha$ for a natural $\alpha \geqslant 2$, then $f' \in \tilde{\mathcal{H}}_+^{\alpha-1}$, by the definition of $\tilde{\mathcal{H}}_+^\alpha$. Hence, for a function $\mu \in \mathcal{M}_+^\alpha$,

$$f(x)=\int\limits_{-\infty}^{x} f'(u)\,\mathrm{d}u=\int\limits_{-\infty}^{x}\mathrm{d}u\int(u-v)_+^{\alpha-1}\,\mathrm{d}\mu(v)=\int(x-v)_+^{\alpha}\,\mathrm{d}\,\mu(v)/\alpha,$$

so that $f\in\mathcal{H}_+^{\alpha}$. □

Proof of Lemma 3 Let f be any function satisfying the conditions of Lemma 3. For any $y\in\mathbb{R}$, introduce then the functions defined by the formulas

$$f_{2,y}(x):=\big(f''(y)+f'''(y)(x-y)\big)_+\,\boldsymbol{I}\{x\leqslant y\}+f''(x)\,\boldsymbol{I}\{x>y\}\,; \tag{56}$$

$$f_y(x):=\Big(f(y)+f'(y)(x-y)+\int\limits_{x}^{y} f_{2,y}(u)\,(u-x)\,\mathrm{d}u\Big)\,\boldsymbol{I}\{x\leqslant y\}$$
$$+f(x)\,\boldsymbol{I}\{x>y\} \tag{57}$$

for all real x. Here f''' denotes the right derivative of the convex function f'', so that f''' is nondecreasing. Since $f''(y)\geqslant 0$, the function $f_{2,y}$ is continuous, whence it is seen that

$$f_y''=f_{2,y}. \tag{58}$$

Because f'' is convex, one has

$$f''(x)\geqslant f''(y)+f'''(y)(x-y) \tag{59}$$

for all y and x; also, it is given that f'' is nonnegative; it follows that

$$f''\geqslant f_{2,y}. \tag{60}$$

Observe that, moreover, the family of functions $(f_{2,y})$ is nonincreasing in $y\in\mathbb{R}$. Indeed, let y and y_1 be any real numbers such that $y_1<y$. Then $f_{2,y_1}=f''\geqslant f_{2,y}$ on $[y_1,\infty)$, in view of (56) and (60). Recalling (59) and the fact that f''' is nondecreasing, one has the inequalities $f''(y_1)\geqslant f''(y)+f'''(y)(y_1-y)$ and $f'''(y_1)(x-y_1)\geqslant f'''(y)(x-y_1)$ for all $x\leqslant y_1$; adding these inequalities, one sees that $f''(y_1)+f'''(y_1)(x-y_1)\geqslant f''(y)+f'''(y)(x-y)$. It follows, in view of (56), that $f_{2,y_1}\geqslant f_{2,y}$ on the interval $(-\infty,y_1]$ as well, and hence on the entire real line.

Using integration-by-parts/Fubini's theorem as in the proof of Lemma 2, one can verify that for any function $g\in\mathcal{C}^2$ and all real y and x

$$g(x)=\Big(g(y)+g'(y)(x-y)+\int\limits_{x}^{y} g''(u)\,(u-x)\,\mathrm{d}u\Big)\,\boldsymbol{I}\{x\leqslant y\}$$
$$+g(x)\,\boldsymbol{I}\{x>y\}\,. \tag{61}$$

By (58), for any real w one has $f_w \in \mathcal{C}^2$, so that one can substitute f_w for g in (61). In fact, let us do so for $w \in \{y, y_1\}$, again assuming that $y_1 < y$. At that, by (57), one has $f_{y_1} = f = f_y$ on the interval $[y, \infty)$ and hence $f_{y_1}(y) = f_y(y)$ and $f'_{y_1}(y) = f'_y(y)$. Now, since the family of functions $(f''_y)_{y\in\mathbb{R}} = (f_{2,y})_{y\in\mathbb{R}}$ is nonincreasing, one can see that the family $(f_y)_{y\in\mathbb{R}}$ is nonincreasing as well. Next, since $\boldsymbol{I}\{x \leqslant y\} \to 0$ and $\boldsymbol{I}\{x > y\} \to 1$ for each x as $y \to -\infty$, one concludes, in view of (58), that for any decreasing sequence (y_n) in $\mathbb{R}$ converging to $-\infty$ one has $f_{y_n} \to f$, in the sense of Definition 1.

It remains to verify that for every real y one has $f_y \in \mathcal{G}^3_+$ and, moreover, $f_y \in \mathcal{G}^3_{++}$ in the case when f is known to be nondecreasing. Observe that

$$f_{2,y}(x) = f'''(y)\,(x - z)\,\boldsymbol{I}\{z \leqslant x \leqslant y\} + f''(x)\,\boldsymbol{I}\{x > y\}\,, \tag{62}$$

where

$$z := y\,\boldsymbol{I}\big\{f'''(y) = 0\big\} + (y - f''(y)/f'''(y))\,\boldsymbol{I}\big\{f'''(y) \neq 0\big\}\,.$$

Indeed, f''' is nonnegative and nondecreasing (since f'' is nondecreasing and convex). Hence, in the case when $f'''(y) = 0$, one has $f''' = 0$ on the entire interval $(-\infty, y]$. This and the condition $f''(-\infty) = 0$ implies $f'' = 0$ on the entire interval $(-\infty, y]$, so that $f''(y) = 0$. Now one sees that expressions (56) and (62) both equal $f''(x)\,\boldsymbol{I}\{x > y\}$ in the case when $f'''(y) = 0$. In the other case, when $f'''(y) \neq 0$, one has $f'''(y) > 0$ (since f''' is nonnegative). Also, here $f''(y) + f'''(y)(x - y) = f'''(y)\,(x - z)$, whence (62) again follows.

Now, for the right derivative $f'_{2,y}$ of $f_{2,y}$, (62) yields

$$f'_{2,y}(x) = f'''(y)\,\boldsymbol{I}\{z \leqslant x \leqslant y\} + f'''(x)\,\boldsymbol{I}\{x > y\}\,.$$

Since f''' is nonnegative and nondecreasing, it follows now that $f'_{2,y}$ is nondecreasing. Therefore, $f_{2,y}$ is convex. That is, by (58), f''_y is convex. Also, (62) and (58) show that $f''_y = 0$ on the interval $(-\infty, z]$. This means that

$$f_y(x) = a + b\,x \text{ for some real constants } a \text{ and } b \text{ and all } x \leqslant z. \tag{63}$$

Let now

$$h_y(x) := f_y(x) - (a + b\,x)$$

for all real x. Then $h''_y = f''_y$ is convex. Moreover, $h_y = 0$ on the interval $(-\infty, z]$, so that $h_y(-\infty) = h'_y(-\infty) = h''_y(-\infty) = 0$. By Lemma 2, $h_y \in \mathcal{H}^3_+$. Thus, $f_y \in \mathcal{G}^3_+$.

If, moreover, f is nondecreasing, then $f' \geqslant 0$. Hence and because (in view of (60)) $f''(u) \geqslant f_{2,y}(u) \geqslant f_{2,y}(u)\,\boldsymbol{I}\{u > x\}$ for all x, y, and u, one has

$$f_y'(x) = \Big(f'(y) - \int_{-\infty}^{y} f_{2,y}(u)\,\boldsymbol{I}\{u > x\}\,\mathrm{d}u\Big)\,\boldsymbol{I}\{x \leqslant y\} + f'(x)\,\boldsymbol{I}\{x > y\}$$
$$\geqslant f'(-\infty)\,\boldsymbol{I}\{x \leqslant y\} + f'(x)\,\boldsymbol{I}\{x > y\} \geqslant 0$$

for all x. Now (63) implies $b \geqslant 0$. Since $h_y \in \mathcal{H}_+^3$, one finally sees that $f_y \in \mathcal{G}_{++}^3$. □

Proof of Lemma 4 First note that $\operatorname{cl}\mathcal{F}_+^3 = \mathcal{F}_+^3$, because the pointwise convergence preserves both the monotonicity and the convexity.

Next, take any $f \in \mathcal{G}_{++}^3$, so that

$$f(x) = a + b\,x + \int (x-t)_+^3\,\mathrm{d}\mu(t)$$

for all x, where $a \in \mathbb{R}$, $b \geqslant 0$, and μ is nondecreasing and $\int (-t)_+^3\,\mathrm{d}\mu(t) < \infty$. It follows that f is nondecreasing and convex, since the functions $x \mapsto (x-t)_+^3$ are so. Similarly, f'' is nondecreasing and convex, since $f''(x) = 6\int (x-t)_+\,\mathrm{d}\mu(t)$. That is, $f \in \mathcal{F}_+^3$ for any $f \in \mathcal{G}_{++}^3$, so that $\mathcal{G}_{++}^3 \subseteq \mathcal{F}_+^3$, whence $\operatorname{cl}\mathcal{G}_{++}^3 \subseteq \operatorname{cl}\mathcal{F}_+^3 = \mathcal{F}_+^3$.

It remains to show that $\mathcal{F}_+^3 \subseteq \operatorname{cl}\mathcal{G}_{++}^3$. Take any $f \in \mathcal{F}_+^3$. Then, by definition (12), f and f'' are nondecreasing and convex. Hence, f' is nonnegative, nondecreasing, and convex. Now Lemma 1 yields $f''(-\infty) = 0$. Also, f'' is nonnegative, since f is convex. Thus, by Lemma 3, $f \in \operatorname{cl}\mathcal{G}_{++}^3$. □

Proof of Lemma 5 First note that $\operatorname{cl}\mathcal{F}^3 = \mathcal{F}^3$, because the pointwise convergence preserves the convexity.

Next, it is trivial that $\mathcal{G}^3 \subseteq \mathcal{F}^3$, whence $\operatorname{cl}\mathcal{G}^3 \subseteq \operatorname{cl}\mathcal{F}^3 = \mathcal{F}^3$.

It remains to show that $\mathcal{F}^3 \subseteq \operatorname{cl}\mathcal{G}^3$. Take any $f \in \mathcal{F}^3$. Then, by definition (28), f and f'' are convex. The latter condition implies that at least one of the following three cases must take place: f'' is nondecreasing on $\mathbb{R}$ or f'' is nonincreasing on $\mathbb{R}$ or f'' switches from nonincreasing to nondecreasing.

Case 1: f'' is nondecreasing on $\mathbb{R}$. Since f is convex, $f'' \geqslant 0$ on $\mathbb{R}$. Hence, there exists the limit $c := f''(-\infty) \in [0,\infty)$. Let

$$g(x) := f(x) - c\,x^2/2$$

for all real x. Then $g'' = f'' - c = f'' - f''(-\infty) \geqslant 0$, since f'' is nondecreasing. Also, $g'' = f'' - c$ is nondecreasing and convex, since f'' is so. In addition, $g''(-\infty) = 0$. Therefore, by Lemma 3, $g \in \operatorname{cl}\mathcal{G}_+^3$. That is, there exists a sequence of functions (g_n) such that $g_n \to g$ and

$$g_n(x) = a_n + b_n\,x + h_n(x)$$

for all real x, where, for each n, a_n and b_n are real constants and h_n is a function in $\mathcal{H}^3_+$; then $h_n \in \mathcal{F}^3_+$ (because $\mathcal{H}^3_+ \subseteq \mathcal{F}^3_+$, as seen, for example, from the proof of Lemma 2). Let now

$$f_n(x) := c\,x^2/2 + g_n(x) = c\,x^2/2 + a_n + b_n\,x + h_n(x)$$

for all x and n. Then $g_n \to g$ implies $f_n \to f$. Moreover, for every n one has $f_n \in \mathcal{G}^3$. Indeed, if $b_n < 0$, then the function h_n belongs to $\mathcal{F}^3_+$ and the function $x \mapsto a_n + b_n\,x$ belongs to $\mathcal{F}^3_-$; and if $b_n \geqslant 0$, then the function $x \mapsto a_n + b_n\,x + h_n(x)$ belongs to $\mathcal{F}^3_+$ and the function $x \mapsto 0$ belongs to $\mathcal{F}^3_-$. Thus, $f \in \operatorname{cl}\mathcal{G}^3$ for any $f \in \mathcal{F}^3$ satisfying the condition of Case 1.

Case 2: f'' is nonincreasing on $\mathbb{R}$. This case reduces to Case 1 by considering the function $x \mapsto \tilde{f}(x) := f(-x)$ in place of f. Indeed, if $f \in \mathcal{F}^3$ and f'' is nonincreasing, then $\tilde{f} \in \mathcal{F}^3$ and $\tilde{f}''$ is nondecreasing. Moreover, $f_n \in \mathcal{G}^3 \iff \tilde{f}_n \in \mathcal{G}^3$, where $\tilde{f}_n(x) := f_n(-x)$ for all real x.

Case 3: There exists some real x_0 such that f'' is nonincreasing on $(-\infty, x_0]$ and nondecreasing on $[x_0, \infty)$. Here without loss of generality (w.l.o.g.) $x_0 = 0$. Let

$$g(x) := f(x) - f'(0)x - f''(0)x^2/2$$

for all real x, so that $g(0) = f(0)$ and $g'(0) = g''(0) = 0$; moreover, $g'' = f'' - f''(0)$ is convex on $\mathbb{R}$ (since f'' is so), nonincreasing on $(-\infty, 0]$, and nondecreasing on $[0, \infty)$, whence $g'' \geqslant 0$. Let, for all real x,

$$h_+(x) := g''(x)\boldsymbol{I}\{x > 0\} \quad \text{and} \quad h_-(x) := g''(x)\boldsymbol{I}\{x \leqslant 0\},$$

so that $h_+ + h_- = g''$, $h_\pm \geqslant 0$, h_+ is nondecreasing, and h_- is nonincreasing. Also, h_+ and h_- are convex, since g'' is convex and $g'' \geqslant g''(0) = 0$. Let further, for all real x,

$$H_+(x) := \int (x-t)_+\, h_+(t)\, \mathrm{d}t \quad \text{and} \quad H_-(x) := \int (t-x)_+\, h_-(t)\, \mathrm{d}t,$$

so that $H_\pm(0) = 0$,

$$H'_+(x) = \int \boldsymbol{I}\{x > t\}\; h_+(t)\, \mathrm{d}t = \int_0^x g''(t)\, \mathrm{d}t\, \boldsymbol{I}\{x > 0\} = g'(x)\boldsymbol{I}\{x > 0\};$$

$$H''_+(x) = g''(x)\boldsymbol{I}\{x > 0\} = h_+(x);$$

$$H'_-(x) = -\int \boldsymbol{I}\{t > x\}\; h_-(t)\, \mathrm{d}t = -\int_x^0 g''(t)\, \mathrm{d}t\, \boldsymbol{I}\{x \leqslant 0\} = g'(x)\boldsymbol{I}\{x \leqslant 0\};$$

$$H''_-(x) = g''(x)\boldsymbol{I}\{x \leqslant 0\} = h_-(x).$$

It follows that $H'_+ + H'_- = g'$, whence for all real x one has $H_+(x) + H_-(x) = g(x) - g(0)$, that is,

$$f(x) = f(0) + f'(0)x + f''(0)x^2/2 + H_+(x) + H_-(x). \tag{64}$$

Also, $H''_+ = h_+$ is nonnegative, nondecreasing, and convex, and hence H_+ is also convex. Also, since $h_+ \geqslant 0$, the first expression for $H'_+(x)$ above shows that $H'_+ \geqslant 0$. Thus, H_+ and H''_+ are nondecreasing and convex; that is, $H_+ \in \mathcal{F}^3_+$. Similarly, $H_- \in \mathcal{F}^3_-$.

Note also that $f''(0) \geqslant 0$, since f belongs to $\mathcal{F}^3$ and is hence convex. If $f'(0) < 0$, then the function H_+ belongs to $\mathcal{F}^3_+$ and the function $x \mapsto f(0) + f'(0)x + H_-(x)$ belongs to $\mathcal{F}^3_-$; and if $f'(0) \geqslant 0$, then the function $x \mapsto f(0) + f'(0)x + H_+(x)$ belongs to $\mathcal{F}^3_+$ and the function H_- belongs to $\mathcal{F}^3_-$. Thus, (64) implies that $f \in \mathcal{G}^3 \subseteq \operatorname{cl}\mathcal{G}^3$ for any $f \in \mathcal{F}^3$ satisfying the condition of Case 3.

One concludes that, in all cases $f \in \mathcal{F}^3$ implies $f \in \operatorname{cl}\mathcal{G}^3$. That is, $\mathcal{F}^3 \subseteq \operatorname{cl}\mathcal{G}^3$ indeed. □

Proof of Lemma 6 Let $f \in \mathcal{F}^3_+$.

Case 1: $f'' = 0$ on $\mathbb{R}$. Then there exist real a and b such that $f(x) = a + b\,x$ for all real x, so that $f(x) = O(x)$ as $x \to \infty$.

Case 2: there exists some $t \in \mathbb{R}$ such that $f''(t) \neq 0$. By (12), f'' is nonnegative (because f is convex) and nondecreasing. Hence, $f''(t) > 0$ and $f''(x) \geqslant f''(t)$ for all $x \geqslant t$. It follows that $f(x) \geqslant f(t) + f'(t)\,(x-t) + f''(t)\,(x-t)^2/2$ for all $x \geqslant t$, whence $\liminf_{x\to\infty} f(x)/x^2 \geqslant f''(t)/2 > 0$. □

Proof of Lemma 7 Let $f \in \mathcal{F}^3_+$. Since f is nondecreasing, one has $f(x) \leqslant f(0)$ for all $x \leqslant 0$. On the other hand, $f(x) \geqslant f(0) + f'(0)x$ for all real x, since f convex. It follows that $|f(x)| \leqslant |f(0)| + |f'(0)x|$ for all $x \leqslant 0$, so that $f(x) = O(|x|)$ as $x \to -\infty$. □

Proof of Lemma 8 Let f be the function defined by (30). Then it is easy to see that $f \in \mathcal{F}^3$.

Suppose that $f \in \mathcal{G}^3$. Then, by (29), there exist $c \geqslant 0$, $f_+ \in \mathcal{F}^3_+$, and $f_- \in \mathcal{F}^3_-$ such that for all real x

$$f(x) = c\,x^2/2 + f_+(x) + f_-(x).$$

Let $x \to -\infty$. Then, by Lemma 7, $f_+(x) = O(|x|)$, while by Lemma 6, either $f_-(x) = O(|x|)$ or $\liminf_{x\to-\infty} f_-(x)/x^2 > 0$. It follows that either $f(x) = O(|x|)$ as $x \to -\infty$ or $\liminf_{x\to-\infty} f(x)/x^2 > 0$. However, neither of these two alternatives is compatible with the fact that $f(x) = \frac{8}{3}\,(1-x)^{3/2}$ for $x \leqslant 0$. Thus, $f \in \mathcal{F}^3 \setminus \mathcal{G}^3$. □

Proof of Proposition 2 Let $g := f'$, where f is defined by (30). Then it is easy to see that $g \in \mathcal{F}^3_+$.

Suppose that $g \in \mathcal{G}^3_+$. Then, by (50), there exist real a and b and $h \in \mathcal{H}^3_+$ such that for all real x one has $g(x) = a + b\,x + h(x)$ and hence $g'(x) = b + h'(x)$.

By Lemma 2, $h(-\infty) = h'(-\infty) = 0$. Hence, $b = g'(-\infty)$. But, by inspection, $g'(-\infty) = 0$. It follows that $b = 0$, and so, $g(-\infty) = a + h(-\infty) = a \in \mathbb{R}$. This contradicts the fact that $g(-\infty) = -\infty$. □

Proof of Lemma 9 Write

$$X = \frac{X+a}{b+a}\, b + \frac{b-X}{b+a}\, (-a).$$

Let f be any nondecreasing convex function. The convexity (together with the condition $-a \leqslant X \leqslant b$ a.s.) implies that

$$f(X) \leqslant \frac{X+a}{b+a} f(b) + \frac{b-X}{b+a} f(-a) = \mathsf{E} f(\sqrt{a\,b}\,\mathrm{BS}) + \frac{f(b)-f(-a)}{b+a}\, X$$

a.s. Now, since f is nondecreasing and $\mathsf{E}\,X \leqslant 0$, the lemma follows. □

Proof of Lemma 10 Since f is convex, the function $[0,\infty) \ni c \mapsto f(c\,X)$ is convex as well. Hence, the function $[0,\infty) \ni c \mapsto g(c) := \mathsf{E} f(c\,X)$ is convex. Since $\mathsf{E}\,X = 0$, one has $g(c) \geqslant g(0)$ for all real c, by Jensen's inequality. Therefore, the right derivative of g is nonnegative at 0 and hence on $[0,\infty)$. Now the lemma follows. □

Proof of Lemma 11 In view of the definition of the class $\mathcal{H}^2_+$, it suffices to verify the statement of the lemma for all functions f of the form $f_t(x) := (x-t)^2_+$, for all real t, so that $e(p) = \mathsf{E}(\mathrm{BS}(p) - t)^2_+$. Then

$$e'(p) = \left(\sqrt{\tfrac{q}{p}} - t\right)\left(-\sqrt{\tfrac{p}{q}} - t\right) \boldsymbol{I}\left\{-\sqrt{\tfrac{p}{q}} < t < \sqrt{\tfrac{q}{p}}\right\} \leqslant 0.$$

□

Proof of Lemma 12 The proof is rather standard; cf. e.g. the corresponding proofs in [9,30,19,12,20,6]. W.l.o.g., the BS_i's are independent of the S_i's. For $i = 0, 1, \ldots, n$ and $f \in \mathcal{H}^2_+$, introduce

$$F_i := \mathsf{E} f\left(S_i + c_{i+1}\mathrm{BS}_{i+1} + \cdots + c_n\mathrm{BS}_n\right).$$

Recall that, by Remark 5, the classes $\mathcal{H}^\alpha_+$ are invariant with respect to the shifts. Hence, by Lemmas 9, 10, and 11,

$$\begin{aligned}
F_i &= \mathsf{E}\,\mathsf{E}_{i-1} f(S_{i-1} + X_i + c_{i+1}\mathrm{BS}_{i+1} + \cdots + c_n\mathrm{BS}_n)\\
&\leqslant \mathsf{E}\,\mathsf{E}_{i-1} f(S_{i-1} + \sqrt{A_{i-1}B_{i-1}}\,\tilde{\mathrm{BS}}_i + c_{i+1}\mathrm{BS}_{i+1} + \cdots + c_n\mathrm{BS}_n)\\
&\leqslant \mathsf{E}\,\mathsf{E}_{i-1} f(S_{i-1} + c_i\,\tilde{\mathrm{BS}}_i + c_{i+1}\mathrm{BS}_{i+1} + \cdots + c_n\mathrm{BS}_n)\\
&\leqslant \mathsf{E}\,\mathsf{E}_{i-1} f(S_{i-1} + c_i\mathrm{BS}_i + c_{i+1}\mathrm{BS}_{i+1} + \cdots + c_n\mathrm{BS}_n)\\
&= F_{i-1}
\end{aligned}$$

for $i = 1, \dots, n$, where E_{i-1} denotes the conditional expectation given the σ-algebra G_{i-1} generated by $H_{\leqslant(i-1)}$ and $(\mathrm{BS}_{i+1}, \dots, \mathrm{BS}_n)$, and the conditional distribution of $\tilde{\mathrm{BS}}_i$ given G_{i-1} is $\mathsf{BS}(p_i)$, with

$$p_i := \frac{A_{i-1}}{B_{i-1} + A_{i-1}},$$

so that $p_i \geqslant p$, according to (20). Hence,

$$\mathsf{E} f(S_n) = F_n \leqslant F_0 \leqslant \mathsf{E} f(c_1 \mathrm{BS}_1 + \cdots + c_n \mathrm{BS}_n);$$

the last inequality follows because $S_0 \leqslant 0$ a.s. and all functions $f \in \mathcal{H}^2_+$ are nondecreasing. □

Proof of Lemma 13 Statement (i) is equivalent to the following: for every $t \in \mathbb{R}$, the function

$$[0, \infty)^2 \ni (a_1, a_2) \longmapsto e_{t,p,m}(a_1, a_2) := \mathsf{E} f_t \left(a_1^{1/(2m)} \mathrm{BS}_1 + a_2^{1/(2m)} \mathrm{BS}_2 \right)$$

is Schur-concave, where $\mathrm{BS}_i \overset{\text{i.i.d.}}{\sim} \mathsf{BS}(p)$ and

$$f_t(x) := \tfrac{1}{3}\, (x - t)_+^3.$$

Using the homogeneity property

$$e_{t,p,m}(\lambda a_1, \lambda a_2) = \lambda^{3/(2m)}\, e_{\lambda^{-1/(2m)}\, t,p,m}(a_1, a_2)$$

for every $\lambda > 0$, one may assume w.l.o.g. that $a_1 + a_2 = 1$, so that $a_1 = \cos^2\theta$ and $a_2 = \sin^2\theta$, for some $\theta \in [0, \pi/2]$; moreover, in view of the same homogeneity property, one may replace here the i.i.d. standardized Bernoulli r.v.'s BS_1 and BS_2 with i.i.d. *centered* Bernoulli r.v.'s BC_1 and BC_2, such that

$$\mathsf{P}(\mathrm{BC}_i = 1 - p) = p = 1 - \mathsf{P}(\mathrm{BC}_i = -p), \quad i = 1, 2.$$

Therefore, statement (i) is equivalent to

$$\Delta_{p,m}(\theta, t) := \partial_\theta \left(\mathsf{E} f_t(\mathrm{BC}_1 \cos^{1/m}\theta + \mathrm{BC}_2 \sin^{1/m}\theta) \right) \tag{65}$$

being nonnegative for all $\theta \in [0, \pi/4]$ and all $t \in \mathbb{R}$ (where $\partial_\theta := \partial/\partial\theta$). Substituting $(-u - p\cos^{1/m}\theta - p\sin^{1/m}\theta)$ for t $\big($after the differentiation of $\mathsf{E} f_t(\mathrm{BC}_1 \cos^{1/m}\theta + \mathrm{BC}_2 \sin^{1/m}\theta)$ in $\theta\big)$, one sees that statement (i) is equivalent to

$$\begin{aligned}\Delta_{1,p,m}(\theta,u) &:= \frac{m}{(1-p)p}\cos^{1-1/m}\theta\ \sin^{1-1/m}\theta \\ &\quad\times\Delta_{p,m}\big(\theta,-u-p\cos^{1/m}\theta-p\sin^{1/m}\theta\big) \\ &= \Delta_{2,p,m}(\cos^{1/m}\theta,\sin^{1/m}\theta,u) \end{aligned} \tag{66}$$

being nonnegative for all $\theta\in[0,\pi/4]$ and $u\in\mathbb{R}$, where

$$\begin{aligned}\Delta_{2,p,m}(c_1,c_2,u) := &-(c_1^{2m-1}-c_2^{2m-1})\,q\,u_+^2-(c_2^{2m-1}q+c_1^{2m-1}p)\,(c_1+u)_+^2 \\ &+(c_1^{2m-1}q+c_2^{2m-1}p)\,(c_2+u)_+^2 \\ &+(c_1^{2m-1}-c_2^{2m-1})\,p\,(c_1+c_2+u)_+^2. \end{aligned} \tag{67}$$

Now, in view of the homogeneity relation

$$\Delta_{2,p,m}(c_1,c_2,u)=c_1^{2m+1}\,\Delta_{2,p,m}(1,c_2/c_1,u/c_1),\quad \text{where } 0<c_2<c_1,$$

statement (i) reduces to $\Delta_{2,p,m}(1,c,u)$ being nonnegative for all $c\in(0,1)$ and $u\in\mathbb{R}$.

It remains to note that

$$\Delta_{2,p,m}(1,c,u)=\begin{cases}\delta_1(u,c,p,m) & \text{if } u\geqslant 0,\\ \delta_2(u,c,p,m) & \text{if } u\in[-c,0],\\ \delta_3(u,c,p,m) & \text{if } u\in[-1,-c],\\ \delta_4(u,c,p,m) & \text{if } u\in[-1-c,-1],\\ 0 & \text{otherwise,}\end{cases} \tag{68}$$

and $\delta_4(u,c,p,m)$ is manifestly nonnegative for all $c\in(0,1)$, $u\in\mathbb{R}$, $p\in(0,1)$, and $m>1$. □

Proof of Lemma 14 Note that $\delta_1(u)\geqslant\delta_1(0)$ for $u\geqslant 0$, $c\in(0,1)$, and $m\geqslant 1$. Next, (68) shows that $\delta_1(0)=\delta_2(0)$. Now the lemma follows immediately from Lemma 15 (which will be proved next, without using Lemma 14). □

Proof of Lemma 15 W.l.o.g., $m>1$. Note that

$$\partial_p\delta_2(u,c,p,m)=(2c-u^2)(1-c^{2m-1})>0$$

for $u\in[-c,0]$, $c\in(0,1)$, and $m>1$, so that w.l.o.g.

$$p=p_*=p_*(m).$$

Next, $\delta_2(u)$ is a convex quadratic polynomial, whose minimum over all $u\in\mathbb{R}$ is attained at

$$u=u_*(c,p,m):=-\frac{c(1-c^{2m-2})}{(1-c^{2m-1})(1-p)}.$$

Hence, it suffices to show that $\gamma(c,p_*)$ is nonnegative for all $c \in [0,1]$ and $m > 1$, where

$$\gamma(c,p) := \delta_2(u_*(c,p,m),c,p,m)\,\frac{(1-c^{2m-1})(1-p)}{c} = -2\big(1-c^{2m-1}\big)^2 p^2 \\ +(1-c^{2m-1})(2-c+c^{2m-2}-2c^{2m-1})p - c^{2m-2}(1-c)^2.$$

The main idea in the proof of this lemma is to replace here the entry of p^2 with the equivalent (for $p = p_*$), first-degree in p polynomial expression according to the identity

$$p_*^2 = \frac{(4m^2-1)p_* - 1}{2(2m-1)^2} \tag{69}$$

(which follows from (53)), to obtain

$$(2m-1)^2\gamma(c,p_*) = f(c,p_*),$$

where

$$f(c,p) := (1-c^{2m-1})^2 - (2m-1)^2(1-c)^2c^{2m-2} \\ +p\,(1-c^{2m-1})(2m-1)\big((2m-3)(1-c^{2m-1}) \\ -(2m-1)\,c\,(1-c^{2m-3})\big).$$

It suffices to show that $f(c,p) \geqslant 0$ for all $p \in (0,1)$, $m > 1$, and $c \in (0,1)$. Introduce

$$g(c) := f(c,p)/c^{2m};$$
$$g_1(c) := g'(c)/c^{2m-3};$$
$$g_2(c) := g_1'(c)\,c^2;$$
$$g_3(c) := g_2'(c)/c^{1-2m};$$
$$g_4(c) := g_3'(c)/c^{1-2m};$$
$$g_5(c) := g_4'(c)/c^{2m-4}.$$

Then, letting

$$s := m-1 > 0,$$

one has $g_5'(c) = 8c^{-1-2s}\,s\,(1+s)(1+2s)^2(1+4s)\big(1-p+4s^2p\big) > 0$ for all $c \in (0,1)$, and so, g_5 is increasing on $(0,1)$ to

$$g_5(1) = -16\,s\,(1+s)(1+2s)^2(1-p+s+4s^2p) < 0.$$

Hence, $g_5 < 0$ on $(0, 1)$, so that g_4 is decreasing on $(0, 1)$ to

$$g_4(1) = 8s(1 + 2s)^2(1 - p + s + 4s^2 p) > 0.$$

Hence, $g_4 > 0$ on $(0, 1)$. Since $g_3(1) = g_2(1) = g_1(1) = g(1) = 0$, it follows successively that $g_3 < 0$, $g_2 > 0$, $g_1 < 0$, and $g > 0$ on $(0, 1)$. This completes the proof of Lemma 15. □

Proof of Lemma 16 This follows because $\delta_3(u)$ is concave in u, $\delta_3(-1) = c^2(1 - c^{2m-1})p \geqslant 0$, and $\delta_3(-c) = \delta_2(-c) \geqslant 0$, where the latter equality and inequality follow immediately from (68) and Lemma 15, respectively. □

Proof of Lemma 17 It is clear from the second expression for $p_*(m)$ in (53) that $p_*(m)$ decreases continuously from $\frac{1}{2}$ to 0 as m increases from 1 to ∞. Also, one can verify that $m_*(p_*(m)) = m$ for all $m \geqslant 1$ (here one may use identity (69)).

If now $p \in (0, p_*(m))$, then $p = p_*(m_1)$ for some $m_1 > m$, whence $m < m_1 = m_*(p_*(m_1)) = m_*(p)$.

It remains to consider the condition $p \geqslant p_*(m)$. If at that $p > \frac{1}{2}$, then $m \geqslant 1 = m_*(p)$, by (14). If, however, $p \in [p_*(m), \frac{1}{2}]$, then $p = p_*(m_1)$ for some $m_1 \in [1, m]$, whence $m \geqslant m_1 = m_*(p_*(m_1)) = m_*(p)$. □

Proof of Lemma 18 Suppose, to the contrary, that statement (II) of Theorem 5 is true, while $m < m_*(p)$. Then, by Lemma 17, one has

$$p < p_* := p_*(m);$$

then, in particular, one has $0 < p < \frac{1}{2}$. Introduce

$$u_p := -\frac{2^{1-\frac{1}{2m}}(m-1)}{(2m-1)(1-p)}.$$

In view of the elementary inequality $p_* \leqslant 1/(2m-1)$ for $m \geqslant 1$ and the condition $p < p_*$, one has

$$p < \frac{1}{2m-1},$$

which implies that

$$-2^{-\frac{1}{2m}} < u_p \leqslant 0.$$

Taking into account these bounds on p and u_p and recalling (66)–(67), one can see that

$$\partial_\theta \Delta_{1,p,m}(\theta, u_p)|_{\theta=\pi/4} = \frac{2^{1-1/(2m)}(2m-1)(p-p_*)(p-p_{**})}{m(1-p)}, \tag{70}$$

$$\text{where} \quad p_{**} := p_{**}(m) := \frac{2m+1+\sqrt{4(m-1)(m+2)+1}}{4(2m-1)} > p_*.$$

Because of the assumption $p < p_*$, identity (70) yields

$$\partial_\theta \Delta_{1,p,m}(\theta, u_p)|_{\theta=\pi/4} > 0.$$

On the other hand,

$$\partial_\theta \Delta_{1,p,m}(\theta, u)|_{\theta=\pi/4} = \frac{m\, 2^{1/m-1}}{(1-p)p}\, \partial_\theta \Delta_{p,m}(\theta, -u - 2^{1-1/(2m)}\, p)|_{\theta=\pi/4}$$

for all real u; this follows from (66), in view of the fact that the derivatives of $\cos^{1-1/m}\theta\ \sin^{1-1/m}\theta$ and $\cos^{1/m}\theta + \sin^{1/m}\theta$ in θ at $\theta = \pi/4$ are zero. Hence,

$$\partial_\theta \Delta_{p,m}(\theta, t_p)|_{\theta=\pi/4} > 0 \quad \text{for } t_p := -u_p - 2^{1-1/(2m)}\, p.$$

Note also that $\Delta_{p,m}(\pi/4, t) = 0$ for all real t. Therefore, $\Delta_{p,m}(\theta, t_p) < 0$ for all θ in a left neighborhood of $\pi/4$. Now (65) implies that $\pi/4$ is not a point of maximum in θ of $\mathsf{E}\, f_t(\mathrm{BC}_1\ \cos^{1/m}\theta + \mathrm{BC}_2\ \sin^{1/m}\theta)$ for $t = t_p$. Hence, in view of the homogeneity argument used in the proof of Lemma 13, $\pi/4$ is not a point of maximum in θ of $\mathsf{E}\, f_t(\mathrm{BS}_1\ \cos^{1/m}\theta + \mathrm{BS}_2\ \sin^{1/m}\theta)$ for $t = t_p/\sqrt{pq}$. But, for any $t \in \mathbb{R}$, one has $f_t \in \mathcal{H}^3_+ \subseteq \mathcal{F}^3_+$ (where the set inclusion follows by Lemma 2). Thus, one obtains a contradiction with the assumed statement (II) of Theorem 5. □

Acknowledgements I thank an anonymous referee for a thorough reading of the manuscript and suggestions for improvement of the presentation.

References

1. Bentkus, V.: A remark on the inequalities of Bernstein, Prokhorov, Bennett, Hoeffding, Talagrand. Lithuanian Math. J. **42**, 262–269 (2002)
2. Bentkus, V.: An inequality for tail probabilities of martingales with differences bounded from one side. J. Theoret. Probab. **16**, 161–173 (2003)
3. Bentkus, V.: On Hoeffding's inequalities. Ann. Probab. **32**, 1650–1673 (2004)
4. Bentkus, V., Kalosha, N., Zuijlen, M. van : On domination of tail probabilities of (super)martingales: explicit bounds. Liet. Mat. Rink. **46**(1), 3–54 (2006)
5. Bobkov, S.G., Götze, F., Houdré, C.: On Gaussian and Bernoulli covariance representations. Bernoulli **7**, 439–451 (2001)
6. Peña, V. H. de la , Ibragimov, R., Sharakhmetov, S.: On extremal distributions and sharp L_p-bounds for sums of multilinear forms. Ann. Probab. **31**, 630–675 (2003)
7. Peña, V. H. de la , Klass M., J., Lai T., L.: Self-normalized processes: exponential inequalities, moment bounds and iterated logarithm laws. Ann. Probab. **32**, 1902–1933 (2005)
8. Dufour, J.-M., Hallin, M.: Improved Eaton bounds for linear combinations of bounded random variables, with statistical applications. J. Am. Stat. Assoc. **88**, 1026–1033 (1993)
9. Eaton, M.L.: A note on symmetric Bernoulli random variables. Ann. Math. Stat. **41**, 1223–1226 (1970)
10. Eaton, M.L.: A probability inequality for linear combinations of bounded random variables. Ann. Stat. **2**, 609–614 (1974)
11. Efron, B.: Student's t test under symmetry conditions. J. Am. Stat. Assoc. **64**, 1278–1302 (1969)
12. Figiel, T., Hitczenko, P., Johnson, W.B., Schechtman, G., Zinn, J.: Extremal properties of Rademacher functions with applications to the Khintchine and Rosenthal inequalities. Trans. Am. Math. Soc. **349**, 997–1027 (1997)

13. Hoeffding, W.: Probability inequalities for sums of bounded random variables. J. Am. Stat. Assoc. **58**, 13–30 (1963)
14. Jing, B.-Y., Shao, Q.-M., Zhou, W.: Saddlepoint approximation for Student's *t*-statistic with no moment conditions. Ann. Stat. **32**, 2679–2711 (2004)
15. Hunt, G.A.: An inequality in probability theory. Proc. Am. Math. Soc. **6**, 506–510 (1955)
16. Logan, B.F., Mallows, C.L., Rice, S.O., Shepp, L.A.: Limit distributions of self-normalized sums. Ann. Probab. **1**, 788–809 (1973)
17. Marshall, A.W., Olkin, I.: Inequalities: theory of majorization and its applications. Mathematics in Science and Engineering, Vol. 143. Academic-Harcourt Brace Jovanovich Publishers, New York-London (1979)
18. Muirhead, R.F.: Some methods applicable to identities and inequalities of symmetric algebraic functions of *n* letters. Proc. Edinb. Math. Soc. **21**, 144–157 (1903)
19. Pinelis, I.: Extremal probabilistic problems and Hotelling's T^2 test under a symmetry condition. Ann. Stat. **22**(1), 357–368 (1994)
20. Pinelis, I.: Optimal tail comparison based on comparison of moments. High dimensional probability (Oberwolfach, 1996), 297–314, Progr. Probab., **43**, Birkhäuser, Basel (1998)
21. Pinelis, I.: 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**, Am. Math. Soc., Providence, RI (1999)
22. Pinelis, I.: Dimensionality reduction in extremal problems for moments of linear combinations of vectors with random coefficients. Stochastic inequalities, applications, 169–185, Progr. Probab., 56, Birkhäuser, Basel (2003)
23. Pinelis, I.: Binomial upper bounds on generalized moments and tail probabilities of (super)martingales with differences bounded from above. In: IMS Lecture Notes–Monograph Series, High Dimensional Probability Vol. 51 (2006) 33–52. Institute of Mathematical Statistics (2006). doi: 10.1214/074921706000000743. http://www.arxiv.org/abs/math.PR/0512301
24. Pinelis, I.: On normal domination of (super)martingales. Electr. J. Probab. **11** Paper 40, 1049-1070 (2006). http://www.math.washington.edu/~ejpecp/include/getdoc.php?id=3724&+ article=1648&mode=pdf
25. Pinelis, I.: On inequalities for sums of bounded random variables. Preprint (2006), http://www.arxiv.org/abs/math.PR/0603030
26. Pinelis, I.: Toward the best constant factor for the Rademacher–Gaussian tail comparison. Preprint (2006), http://www.arxiv.org/abs/math.PR/0605340
27. Pinelis, I.: Student's *t*-test without symmetry conditions. Preprint (2006), http://www.arxiv.org/abs/math.ST/0606160
28. Shao, Q.-M.: Self-normalized large deviations. Ann. Probab. **25**, 285–328 (1997)
29. Shorack, G.R., Wellner, J.A.: Empirical Processes with Applications to Statistics. Wiley, New York (1986)
30. Utev, S.A.: Extremal problems in moment inequalities. (Russian) Limit theorems of probability theory, Trudy Inst. Mat. **5**, 56–75, 175. "Nauka" Sibirsk. Otdel., Novosibirsk (1985)
31. Whittle, P.: Bounds for the moments of linear and quadratic forms in independent variables. Teor. Verojatnost. i Primenen. **5**, 331–335 (1960)