Electronic Journal of Probability

Electron. J. Probab. **21** (2016), no. 20, 1–19.
ISSN: 1083-6489 DOI: 10.1214/16-EJP4474

Optimal binomial, Poisson, and normal left-tail domination for sums of nonnegative random variables

Iosif Pinelis*

Abstract

Exact upper bounds on the generalized moments $\mathsf{E}\, f(S_n)$ of sums S_n of independent *nonnegative* random variables X_i for certain classes $\mathcal{F}$ of *nonincreasing* functions f are given in terms of (the sums of) the first two moments of the X_i's. These bounds are of the form $\mathsf{E}\, f(\eta)$, where the random variable η is either binomial or Poisson depending on whether n is fixed or not. The classes $\mathcal{F}$ contain, and are much wider than, the class of all decreasing exponential functions. As corollaries of these results, optimal in a certain sense upper bounds on the left-tail probabilities $\mathsf{P}(S_n \leqslant x)$ are presented, for any real x. In fact, more general settings than the ones described above are considered. Exact upper bounds on the exponential moments $\mathsf{E}\exp\{hS_n\}$ for $h < 0$, as well as the corresponding exponential bounds on the left-tail probabilities, were previously obtained by Pinelis and Utev. It is shown that the new bounds on the tails are substantially better.

Keywords: probability inequalities; sums of random variables; submartingales; martingales; upper bounds; generalized moments.
AMS MSC 2010: Primary 60E15, Secondary 60G42; 60G48.
Submitted to EJP on August 11, 2015, final version accepted on February 25, 2016.

1 Introduction

Let $X_1, \dots, X_n$ be independent real-valued random variables (r.v.'s), with

$$S_n := X_1 + \dots + X_n.$$

Exponential upper bounds for S_n go back at least to Bernstein. As the starting point here, one uses the multiplicative property of the exponential function together with the condition of independence of $X_1, \dots, X_n$ to write

$$\mathsf{E}\, e^{hS_n} = \prod_1^n \mathsf{E}\, e^{hX_i} \tag{1.1}$$

for all real h. Then one bounds up each factor $\mathsf{E}\, e^{hX_i}$, thus obtaining an upper bound (say $M_n(h)$) on $\mathsf{E}\, e^{hS_n}$, uses the Markov inequality to write $\mathsf{P}(S_n \geqslant x) \leqslant e^{-hx}\,\mathsf{E}\, e^{hS_n} \leqslant$

*Michigan Technological University, United States of America. E-mail: ipinelis@mtu.edu

$B_n(h,x) := e^{-hx}M_n(h)$ for all real x and all nonnegative real h, and finally tries to minimize $B_n(h,x)$ in $h \geqslant 0$ to obtain an upper bound on the tail probability $\mathsf{P}(S_n \geqslant x)$.

This approach was used and further developed in a large number of papers, including notably the well-known work by Bennett [2] and Hoeffding [11]. Pinelis and Utev [22] offered a general approach to obtaining exact bounds on the exponential moments $\mathsf{E}\, e^{hS_n}$, with a number of particular applications.

Exponential bounds were obtained in more general settings as well, where the r.v.'s $X_1,\dots,X_n$ do not have to be independent or real-valued. It was already mentioned by Hoeffding at the end of Section 2 in [11] that his results remain valid for martingales. Exponential inequalities with optimality properties for vector-valued $X_1,\dots,X_n$ were obtained e.g. in [21, 24] and then used in a large number of papers.

Related to this is work on Rosenthal-type and von Bahr–Esseen-type bounds, that is, bounds on absolute power moments $\mathsf{E}\,|S_n|^p$ of S_n; see e.g. [1, 36, 24, 13, 6, 12, 19, 39, 30, 34, 32].

However, the classes of exponential functions $e^{h\cdot}$ and absolute power functions $|\cdot|^p$ are too narrow in that the resulting bounds on the tails are not as good as one could get in certain settings. It is therefore natural to try to consider wider classes of moment functions and then try to choose the best moment function in such a wider class to obtain a better bound on the tail probability. This approach was used and developed in [9, 10, 23, 25, 4, 31], in particular. The main difficulty one needs to overcome working with such, not necessarily exponential, moment functions is the lack of multiplicative property (1.1).

In some settings, the bounds can be improved if it is known that the r.v.'s $X_1,\dots,X_n$ are nonnegative; see e.g. [13, 6, 12, 19]. However, in such settings the focus has usually been on bounds for the right tail of the distribution of S_n. There has been comparatively little work done concerning the left tail of the distribution of the sum S_n of nonnegative r.v.'s $X_1,\dots,X_n$.

One such result was obtained in [22]. Suppose indeed that the independent r.v.'s $X_1,\dots,X_n$ are nonnegative. Also, suppose here that

$$m := \mathsf{E}\,X_1 + \dots + \mathsf{E}\,X_n > 0 \quad \text{and} \quad s := \mathsf{E}\,X_1^2 + \dots + \mathsf{E}\,X_n^2 < \infty. \tag{1.2}$$

Then [22, Theorem 7] for any $x \in (0,m]$

$$\mathsf{P}(S_n \leqslant x) \leqslant \exp\Big\{-\frac{m^2}{s}\Big(1+\frac{x}{m}\,\ln\frac{x}{em}\Big)\Big\} \leqslant \exp\Big\{-\frac{(x-m)^2}{2s}\Big\} \tag{1.3}$$

(in fact, these inequalities were stated in [22] in the equivalent form for the non-positive r.v.'s $-X_1,\dots,-X_n$). These upper bounds on the tail probability $\mathsf{P}(S_n \leqslant x)$ were based on exact upper bounds on the exponential moments of the sum S_n, which can be written as follows:

$$\mathsf{E}\exp\{hS_n\} \leqslant \mathsf{E}\exp\big\{h\tfrac{s}{m}\Pi_{m^2/s}\big\} \leqslant \mathsf{E}\exp\big\{h\big(m+Z\sqrt{s}\,\big)\big\} \tag{1.4}$$

for all real $h \leqslant 0$. Here and subsequently, for any $\lambda \in (0,\infty)$, let Π_λ and Z stand for any r.v. having the Poisson distribution with parameter $\lambda \in (0,\infty)$ and for any standard normal r.v., respectively. The bounds in (1.3) and (1.4) have certain optimality properties, and they are very simple in form. Yet, they have apparently been little known; in particular, the last bound in (1.3) was rediscovered in [16].

In the present paper, the "Poisson" and "normal" bounds in (1.4) will be extended to a class of moment functions much wider than the "exponential" class (still with the preservation of the optimality property, for each moment function in the wider class). Consequently, the bounds in (1.3) will be much improved. We shall also provide "binomial" upper bounds on the moments and tail probabilities of S_n, which are further improvements of the corresponding "Poisson", and hence "normal", bounds.

2 Summary and discussion

Let $X_1,\dots,X_n$ be nonnegative real-valued r.v.'s. In general, we shall no longer assume that $X_1,\dots,X_n$ are independent; instead, a more general condition, described in the definition below, will be assumed. Moreover, the condition (1.2) will be replaced by a more general one.

Definition 2.1. *Given any* $\mathbf{m}=(m_1,\dots,m_n)$ *and* $\mathbf{s}=(s_1,\dots,s_n)$ *in* $[0,\infty)^n$*, let us say that the r.v.'s* $X_1,\dots,X_n$ *satisfy the* $(\mathbf{m},\mathbf{s})$*-condition if, for some filter* $(\mathcal{A}_0,\dots,\mathcal{A}_{n-1})$ *of sigma-algebras and each* $i\in\overline{1,n}$*, the r.v.* X_i *is* $\mathcal{A}_i$*-measurable,*

$$\mathsf{E}(X_i|\mathcal{A}_{i-1})\geqslant m_i, \quad \textit{and} \quad \mathsf{E}(X_i^2|\mathcal{A}_{i-1})\leqslant s_i. \tag{2.1}$$

Given any nonnegative m *and* s*, let us also say that the* (m,s)*-condition is satisfied if the* $(\mathbf{m},\mathbf{s})$*-condition holds for some* $\mathbf{m}=(m_1,\dots,m_n)$ *and* $\mathbf{s}=(s_1,\dots,s_n)$ *in* $[0,\infty)^n$ *such that*

$$m_1+\dots+m_n\geqslant m \quad \textit{and} \quad s_1+\dots+s_n\leqslant s. \tag{2.2}$$

In the above definition and in what follows, for any α and β in $\mathbb{Z}\cup\{\infty\}$, we let $\overline{\alpha,\beta}:=\{j\in\mathbb{Z}\colon \alpha\leqslant j\leqslant\beta\}$.

The following comments are in order.

- Any independent r.v.'s $X_1,\dots,X_n$ satisfy the $(\mathbf{m},\mathbf{s})$-condition if $\mathsf{E}\,X_i\geqslant m_i$ and $\mathsf{E}\,X_i^2\leqslant s_i$ for each $i\in\overline{1,n}$; if at that (2.2) holds, then the (m,s)-condition holds as well.
- If r.v.'s $X_1,\dots,X_n$ satisfy the $(\mathbf{m},\mathbf{s})$-condition, then the r.v.'s $X_1-m_1,\dots,$ X_n-m_n are submartingale-differences, with respect to the corresponding filter $(\mathcal{A}_0,\dots,\mathcal{A}_{n-1})$.
- If, for some $\mathbf{m}$ and $\mathbf{s}$ in $[0,\infty)^n$, the $(\mathbf{m},\mathbf{s})$-condition is satisfied by some r.v.'s $X_1,\dots,X_n$, then necessarily

  $$s_i\geqslant m_i^2 \quad \text{for all } i\in\overline{1,n}. \tag{2.3}$$

  Moreover, if, for some nonnegative m and s, the (m,s)-condition is satisfied by some r.v.'s $X_1,\dots,X_n$, then necessarily

  $$\tfrac{s}{n}\geqslant\left(\tfrac{m}{n}\right)^2 \quad \text{or, equivalently,} \quad n\geqslant\tfrac{m^2}{s}. \tag{2.4}$$

Definition 2.2. *Given any real numbers* m *and* s *such that* $m>0$ *and* $s\geqslant m^2$ *(cf.* (2.3)*), let* $Y^{m,s}$ *stand for any r.v. such that*

$$\mathsf{E}\,Y^{m,s}=m, \quad \mathsf{E}(Y^{m,s})^2=s, \quad \textit{and} \quad \mathsf{P}(Y^{m,s}\in\{0,\tfrac{s}{m}\})=1;$$

such a r.v. $Y^{m,s}$ *exists, and its distribution is uniquely determined:*

$$\mathsf{P}(Y^{m,s}=\tfrac{s}{m})=1-\mathsf{P}(Y^{m,s}=0)=\tfrac{m^2}{s};$$

moreover, let $Y_1^{m,s},\dots,Y_n^{m,s}$ *denote independent copies of a r.v.* $Y^{m,s}$*. Also, given any* $\mathbf{m}$ *and* $\mathbf{s}$ *in* $(0,\infty)^n$ *such that the condition* (2.3) *holds, we shall always assume the corresponding r.v.'s* $Y^{m_1,s_1},\dots,Y^{m_n,s_n}$ *to be independent.*

Next, let us describe the pertinent classes of generalized moment functions. For any natural j, let $\mathcal{S}^j$ denote the class of all $(j-1)$-times differentiable functions $g\colon\mathbb{R}\to\mathbb{R}$ such that the $(j-1)$th derivative $g^{(j-1)}$ of g has a right-continuous right derivative, which will be denoted here simply by $g^{(j)}$. As usual, we let $g^{(0)}:=g$. Take then any natural

$$k\leqslant j+1$$

and introduce the class of functions

$$\mathcal{F}_+^{k:j} := \big\{ g \in \mathcal{S}^j \colon g^{(i)} \text{ is nondecreasing for each } i \in \overline{k-1,j}\, \big\} \tag{2.5}$$

and, finally, the "reflected" class

$$\mathcal{F}_-^{k:j} := \{ g^- \colon g \in \mathcal{F}_+^{k:j} \}, \tag{2.6}$$

where $g^-(x) := g(-x)$ for all $x \in \mathbb{R}$. It is clear that the class $\mathcal{F}_-^{k:j}$ gets narrower as j increases (with a fixed k), and it gets wider as k increases (with a fixed j).

As an example, the function $x \mapsto a + b\,x + c\,e^{-\lambda x}$ belongs to $\mathcal{F}_-^{k:j}$ for any $a \in \mathbb{R}$, $b \leqslant 0$, $c \geqslant 0$, $\lambda \geqslant 0$ (and any natural k and j such that $k \leqslant j+1$). Also, given any $a \in \mathbb{R}$, $b \leqslant 0$, $c \geqslant 0$, and $w \in \mathbb{R}$, the function $x \mapsto a + b\,x + c\,(w-x)_+^\alpha$ belongs to $\mathcal{F}_-^{k:j}$ for any real $\alpha \geqslant j$ (and any natural k and j such that $k \leqslant j+1$); here and elsewhere, as usual, $x_+ := \max(0,x)$ and $x_+^\alpha := (x_+)^\alpha$ for $x \in \mathbb{R}$. Note also that the classes $\mathcal{F}_-^{k:j}$ 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.

Remark 2.3. It is not difficult to see that, if a function f is in the class $\mathcal{F}_-^{k:j}$, then the shifted and/or rescaled function $x \mapsto f(bx+a)$ is also in the same class, for any constants $a \in \mathbb{R}$ and $b \geqslant 0$. That is, these classes of functions are shift- and scale-invariant.

Now we are ready to state the main result of this paper.

Theorem 2.4.

(I) *Let $X_1, \dots, X_n$ be any* nonnegative *r.v.'s satisfying the $(\mathbf{m},\mathbf{s})$-condition for some $\mathbf{m}$ and $\mathbf{s}$ in $(0,\infty)^n$, so that* (2.3) *holds. Then*

$$\mathsf{E}\, f(S_n) \leqslant \mathsf{E}\, f\big(Y^{m_1,s_1} + \dots + Y^{m_n,s_n}\big) \tag{2.7}$$

for all $f \in \mathcal{F}_-^{1:2}$.

(II) *Let $X_1, \dots, X_n$ be any* nonnegative *r.v.'s satisfying the (m,s)-condition for some m and s in $(0,\infty)$, so that* (2.4) *holds. Then*

$$\mathsf{E}\, f(S_n) \leqslant \mathsf{E}\, f\big(Y_1^{\frac{m}{n},\frac{s}{n}} + \dots + Y_n^{\frac{m}{n},\frac{s}{n}}\big) \tag{2.8}$$
$$\leqslant \mathsf{E}\, f\big(\tfrac{s}{m}\Pi_{m^2/s}\big) \tag{2.9}$$
$$\leqslant \mathsf{E}\, f\big(m + Z\sqrt{s}\,\big) \tag{2.10}$$

for all $f \in \mathcal{F}_-^{1:3}$; in fact, (2.10) *and the inequality*

$$\mathsf{E}\, f(S_n) \leqslant \mathsf{E}\, f\big(m + Z\sqrt{s}\,\big) \tag{2.11}$$

both hold for all $f \in \mathcal{F}_-^{1:2}$.

The necessary proofs will be given in Section 3.

Remark 2.5. Under the corresponding conditions given in Theorem 2.4, the expected values in inequalities (2.7)–(2.11) exist (in $\mathbb{R}$ or, at least, in $(-\infty,\infty]$), according to [20, Proposition 5.2, part (i)]. Moreover, the conditions for (2.7)–(2.11) in Theorem 2.4 can be supplemented or relaxed as follows. To describe these extended or relaxed conditions for (2.7)–(2.11), introduce the conditions of equalities in (2.1) and/or (2.2):

$$\mathsf{E}(X_i|\mathcal{A}_{i-1}) = m_i \quad \text{for all } i, \tag{2.12}$$
$$\mathsf{E}(X_i^2|\mathcal{A}_{i-1}) = s_i \quad \text{for all } i, \tag{2.13}$$
$$m_1 + \dots + m_n = m, \tag{2.14}$$
$$s_1 + \dots + s_n = s \tag{2.15}$$

and also conditions

$$\text{the } X_i\text{'s are bounded or } f \geqslant p \text{ for some quadratic polynomial } p, \tag{2.16}$$

$$\mathsf{E}\, X_i^3 < \infty \text{ for all } i. \tag{2.17}$$

Then

(I) inequalities (2.7) and (2.11) hold if any one of the following two conditions holds:

  (i) (2.12) and $f \in \mathcal{F}_-^{2:2}$;
  (ii) (2.12), (2.13), (2.16), and $f \in \mathcal{F}_-^{3:2}$.

(II) inequality (2.8) holds if any one of the following three conditions holds:

  (i) (2.12) and $f \in \mathcal{F}_-^{2:3}$;
  (ii) (2.12), (2.13), (2.16), and $f \in \mathcal{F}_-^{3:3}$;
  (iii) (2.12), (2.13), (2.17), and $f \in \mathcal{F}_-^{4:3}$.

(III) inequality (2.9) holds if any one of the following three conditions holds:

  (i) $f \in \mathcal{F}_-^{2:3}$;
  (ii) (2.16) and $f \in \mathcal{F}_-^{3:3}$;
  (iii) (2.17) and $f \in \mathcal{F}_-^{4:3}$.

(IV) inequality (2.10) holds if any one of the following two conditions holds:

  (i) $f \in \mathcal{F}_-^{2:2}$;
  (ii) (2.16) and $f \in \mathcal{F}_-^{3:2}$.

This remark can be verified similarly to Theorem 2.4.

Obviously, the r.v.'s $Y^{m_1,s_1}, \dots, Y^{m_n,s_n}$ in (2.7) satisfy the $(\mathbf{m}, \mathsf{s})$-condition. So, inequality (2.7) is exact, in the sense that, given any natural n and any $\mathbf{m}$ and s in $(0,\infty)^n$ such that (2.3) holds, the right-hand side of (2.7) is the exact upper bound on its left-hand side. Similarly, given any natural n and any m and s in $(0,\infty)$ such that (2.4) holds, inequality (2.8) is exact.

Proposition 2.6. *Given any m and s in $(0,\infty)$, the Poisson upper bound in* (2.9) *on $\mathsf{E}\, f(S_n)$ is exact $\big($in this case n is not fixed, having only to satisfy* (2.4)$\big)$.

Inequality (2.11) is best possible in the following limited sense, at least. By [20, Corollary 5.9], this inequality holds for all $f \in \mathcal{F}_-^{1:2}$ if and only if it holds for all functions f of the form $f_{w,2}$ for $w \in \mathbb{R}$, where

$$f_{w,\alpha}(x) := (w - x)_+^\alpha. \tag{2.18}$$

Let now positive m and s vary so that $m^2/s \to \infty$, which is the case e.g. when $0 \neq m_1 = m_2 = \cdots$, $0 < s_1 = s_2 = \cdots$, conditions (2.14) and (2.15) hold, and $n \to \infty$. At that, fix any real κ and let $w = m + \kappa\sqrt{s}$. Let $L_{m,s;w} := \mathsf{E}\, f_{w,2}\big(\frac{s}{m}\Pi_{m^2/s}\big)$, which is, according to Proposition 2.6, the exact upper bound on $\mathsf{E}\, f_{w,2}(S_n)$ given m and s. Then $L_{m,s;w} \sim \mathsf{E}\, f_{w,2}\big(m + Z\sqrt{s}\big)$; as usual, $a \sim b$ means that $a/b \to 1$. Indeed, introducing $\tilde{Z} := (\Pi_{m^2/s} - m^2/s)/\sqrt{m^2/s}$, one has $\tilde{Z} \to Z$ in distribution, so that $\frac{1}{s}\, L_{m,s;w} = \mathsf{E}\, f_{\kappa,2}(\tilde{Z}) \to \mathsf{E}\, f_{\kappa,2}(Z) = \frac{1}{s}\, \mathsf{E}\, f_{w,2}\big(m + Z\sqrt{s}\big)$. This convergence is justified, since $f_{\kappa,2}(\tilde{Z})$ is uniformly integrable (as e.g. in [5, Theorem 5.4]), which in turn follows because for any λ and α in $(0,\infty)$ one has $\mathsf{E} \exp \frac{\Pi_\lambda - \lambda}{\sqrt{\lambda}} = \exp\big\{\lambda\big(e^{-1/\sqrt{\lambda}} - 1 + 1/\sqrt{\lambda}\big)\big\} \leqslant \sqrt{e} < \infty$ and $f_{\kappa,\alpha}(x)/e^{-x} \to 0$ as $x \to -\infty$.

Let η denote an arbitrary real-valued r.v. Recalling that for any natural α and any $w \in \mathbb{R}$ the function $f_{w,\alpha}$ belongs to $\mathcal{F}_-^{1:\alpha}$ and applying the Markov inequality, one sees that Theorem 2.4 immediately implies

Corollary 2.7. *Let $X_1,\dots,X_n$ be any* nonnegative *r.v.'s satisfying the (m,s)-condition for some m and s in $(0,\infty)$, so that* (2.4) *holds. Then*

$$\begin{aligned}\mathsf{P}(S_n\leqslant x)&\leqslant P_3\big(\Sigma_{n;m,s};x\big) &(2.19)\\ &\leqslant P_3\big(\Sigma_{\infty;m,s};x\big) &(2.20)\\ &\leqslant P_3\big(m+Z\sqrt{s};x\big); &(2.21)\end{aligned}$$

here and in what follows, x is an arbitrary real number (unless otherwise indicated),

$$\begin{aligned}\Sigma_{n;m,s}&:=Y_1^{\frac{m}{n},\frac{s}{n}}+\dots+Y_n^{\frac{m}{n},\frac{s}{n}}\quad\text{for natural } n, &(2.22)\\ \Sigma_{\infty;m,s}&:=\tfrac{s}{m}\Pi_{m^2/s}, &(2.23)\end{aligned}$$

and

$$P_\alpha(\eta;x):=\inf_{w\in(x,\infty)}\frac{\mathsf{E}(w-\eta)_+^\alpha}{(w-x)^\alpha}$$

for any real $\alpha>0$. Also, the upper bound $P_3\big(m+Z\sqrt{s};x\big)$ on $\mathsf{P}(S_n\leqslant x)$ can be somewhat improved:

$$\mathsf{P}(S_n\leqslant x)\leqslant P_2\big(m+Z\sqrt{s};x\big). \tag{2.24}$$

The computation of $P_\alpha(\eta;x)$ is described (in a somewhat more general setting) in [25, Theorem 2.5]; for normal η, similar considerations were given already in [24, page 363] (those descriptions are given for the right tail of η, so that one will have to make the reflection $x\mapsto -x$ to apply those results). An elaboration of [25, Theorem 2.5] is presented in [28, Proposition 3.2]. Concerning fast and effective calculations of the positive-part moments $\mathsf{E}\,X_+^\alpha$, see [29]. In [3], one can find specific details on the calculation of $P_\alpha(\eta;x)$ for $\alpha\in\{1,2,3\}$ and η with a distribution belonging to a common particular family such as binomial and Poisson.

Let us present here some of those results, which will be useful in this context. Take any real $\alpha>1$ and any r.v. η such that $\mathsf{E}\,\eta_-^\alpha<\infty$; then there exists $\mathsf{E}\,\eta\in(-\infty,\infty]$. Let

$$x_*:=x_*(\eta):=\inf\operatorname{supp}(\eta), \tag{2.25}$$

where $\operatorname{supp}(\eta)$ denotes the support set of (the distribution of) the r.v. η, and

$$\gamma(w):=\gamma(\eta;w):=\frac{\mathsf{E}\,\eta\,(w-\eta)_+^{\alpha-1}}{\mathsf{E}(w-\eta)_+^{\alpha-1}}$$

for $w\in(x_*,\infty)$. Then, by [28, Proposition 3.2], the function γ is continuous and nondecreasing on the interval (x_*,∞) and for every $x\in(x_*,\mathsf{E}\,\eta)$ there exists a unique $w_x=w_{x;\alpha,\eta}\in(x_*,\infty)$ such that

$$\gamma(w_x)=x;$$

in fact, $w_x\in(x,\infty)$. It follows that, for every $x\in(x_*,\mathsf{E}\,\eta)$,

$$\mathcal{E}_{\alpha;x}(w):=\mathsf{E}(w-\eta)_+^{\alpha-1}(\eta-x)\begin{cases}<0 & \text{for } w\in(x_*,w_x),\\ =0 & \text{for } w=w_x,\\ >0 & \text{for } w\in(w_x,\infty);\end{cases} \tag{2.26}$$

in particular, w_x is the only root in (x_*,∞) of the equation

$$\mathcal{E}_{\alpha;x}(w_x)=0. \tag{2.27}$$

Also by [28, Proposition 3.2],

$$P_\alpha(\eta;x)=\begin{cases}\mathsf{P}(\eta\leqslant x)=\mathsf{P}(\eta=x) & \text{for } x\in(-\infty,x_*],\\ \dfrac{\mathsf{E}^\alpha(w_x-\eta)_+^{\alpha-1}}{\mathsf{E}^{\alpha-1}(w_x-\eta)_+^\alpha} & \text{for } x\in(x_*,\mathsf{E}\,\eta),\\ 1 & \text{for } x\in[\mathsf{E}\,\eta,\infty).\end{cases}$$

In particular, the upper bound $P_\alpha(\eta;x)$ on the left-tail probability $\mathsf{P}(\eta\leqslant x)$ is exact for $x\in(-\infty,x_*]$.

Thus, to evaluate $P_\alpha(\eta;x)$ for any real x, it is enough to find w_x (that is, to solve equation (2.27)) for any $x\in(x_*,\mathsf{E}\,\eta)$.

This is especially easy to do if the r.v. η takes values in a lattice, which is the case when η is $\Sigma_{n;m,s}$ or $\Sigma_{\infty;m,s}$, as in Corollary 2.7. Again by [28, Proposition 3.2],

$$P_\alpha(a+b\eta;x)=P_\alpha\big(\eta;\tfrac{x-a}{b}\big)$$

for all real x and a and all $b\in(0,\infty)$. So, the calculation of $P_\alpha(\eta;x)$ for η equal $\Sigma_{n;m,s}$ or $\Sigma_{\infty;m,s}$ reduces to the situation when the r.v. η is integer-valued with $x_*=x_*(\eta)=0$; assume for now that this is the case. In view of (2.19) and (2.20), assume also that $\alpha=3$. Then, by (2.26),

$$\mathcal{E}_{3;x}(w):=a_jw^2-2b_jw+c_j, \tag{2.28}$$

where $x\in(x_*,\mathsf{E}\,\eta)=(0,\mathsf{E}\,\eta)$, $w\in(x_*,\infty)=(0,\infty)$,

$$\begin{aligned} j&:=\lceil w-1\rceil \text{ (so that } j\in\overline{0,\infty} \text{ and } j<w\leqslant j+1),\\ a_j&:=a_{j,x}:=\mathsf{E}(\eta-x)\,\mathrm{I}\{\eta\leqslant j\}\,,\\ b_j&:=b_{j,x}:=\mathsf{E}\,\eta(\eta-x)\,\mathrm{I}\{\eta\leqslant j\}\,,\\ c_j&:=c_{j,x}:=\mathsf{E}\,\eta^2(\eta-x)\,\mathrm{I}\{\eta\leqslant j\}\,.\end{aligned}$$

Therefore and in view of (2.27) and (2.26), for each $x\in(x_*,\mathsf{E}\,\eta)=(0,\mathsf{E}\,\eta)$ one finds w_x as the only root in the interval $(j_x,j_x+1]$ of the quadratic equation

$$a_{j_x}w_x^2-2b_{j_x}w_x+c_{j_x}=0, \tag{2.29}$$

where $j_x:=\min\big\{j\in\overline{0,\infty}\colon a_j\,(j+1)^2-2b_j\,(j+1)+c_j\geqslant 0\big\}$. If $a_{j_x}\neq 0$ then, by (2.26) and (2.28), w_x is the greater of the roots of the above quadratic equation.

The interesting paper [8] presents, for any given $n\in\overline{0,\infty}\cup\{\infty\}$ and $\lambda\in(1,\infty)$, the exact upper bound (say $B_{n,\lambda}$) on $\mathsf{P}(S\leqslant 1)$ under the condition that $S=\sum_{i=1}^n X_i$, where the X_i's are independent r.v.'s such that $0\leqslant X_i\leqslant 1$ for all $i\in\overline{1,n}$ and $\mathsf{E}\,S=\lambda$. $\big($For $\lambda\in[0,1]$, the exact upper bound $B_{n,\lambda}$ is trivial and equals 1; indeed, let X_1 take values 0 and 1 with probabilities $1-\lambda$ and λ, respectively, and let $X_i=0$ for all $i\in\overline{2,n}$.$\big)$ Note that the conditions $0\leqslant X_i\leqslant 1$ for all i and $\mathsf{E}\,S=\lambda$ imply $\sum_i\mathsf{E}\,X_i=\lambda$ and $\sum_i\mathsf{E}\,X_i^2\leqslant\lambda$, which corresponds to the (m,s)-condition with $m=s=\lambda$. So, it makes sense to compare the bound $P_3\big(\Sigma_{n;\lambda,\lambda};1\big)$ in (2.19)–(2.20) with $B_{n,\lambda}$. Graphs of these two bounds and their ratio in the case $n=\infty$ are shown in Figure 1.
The calculations of $P_3\big(\Sigma_{\infty;\lambda,\lambda};1\big)$ here were done in accordance with the above description, containing formulas (2.25)–(2.29); it takes less than 0.3 sec with Mathematica on a standard laptop to produce either of the two graphs in Figure 1. It can be seen that the bound $P_3\big(\Sigma_{\infty;\lambda,\lambda};1\big)$ is not much greater than the optimal bound $B_{\infty,\lambda}$, especially when λ is close to either 1 or ∞; the corresponding comparisons for finite n look similar. On the other hand, our bounds $P_3\big(\Sigma_{n;m,s};x\big)$ hold under much more general conditions: (i) for all $x\in\mathbb{R}$, rather than just for $x=1$; (ii) assuming only the (m,s)-condition (on the

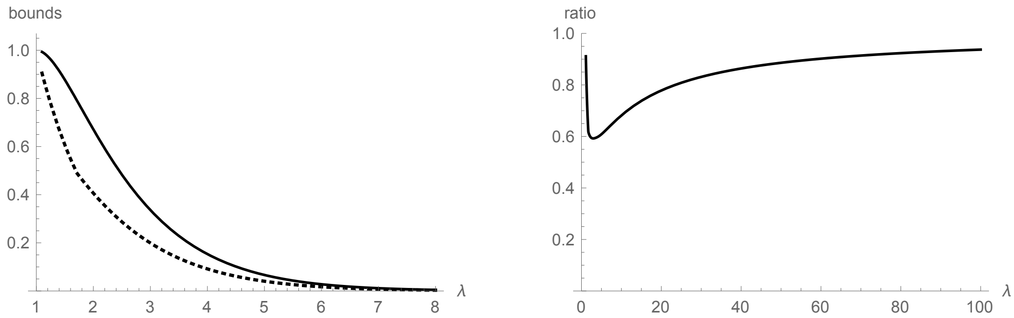

Figure 1: Left panel: graphs $\big\{\big(\lambda, P_3(\Sigma_{\infty;\lambda,\lambda};1)\big)\colon 1.1 \leqslant \lambda \leqslant 8\big\}$ (solid) and $\{(\lambda, B_{\infty,\lambda})\colon 1.1 \leqslant \lambda \leqslant 8\}$ (dotted). Right panel: graph $\big\{\big(\lambda, B_{\infty,\lambda}/P_3(\Sigma_{\infty;\lambda,\lambda};1)\big)\colon 1.1 \leqslant \lambda \leqslant 100\big\}$.

sums of the first and second moments of the X_i's), rather than requiring all the X_i's to be bounded by the constant 1 – which latter also coincides with the value of x chosen in [8]; (iii) assuming the more general dependence conditions.

By [28, Proposition 3.5],

$$P_\alpha(\eta;x) \uparrow P_\infty(\eta;x) := \inf_{h<0} e^{-hx}\,\mathsf{E}\,e^{h\eta} \tag{2.30}$$

as α increases from 0 to ∞; thus, the bounds $P_\alpha(\eta;x)$ improve on the so-called exponential bounds $P_\infty(\eta;x)$. In particular, letting

$$\lambda := \frac{m^2}{s} \quad\text{and}\quad z := \frac{x-m}{\sqrt{s}},$$

one has $\big($cf. (2.19), (2.20), and (2.24)$\big)$,

$$\begin{aligned}
P_2\big(m+Z\sqrt{s};x\big) &\leqslant P_\infty\big(m+Z\sqrt{s};x\big) = e^{-z^2/2}, && (2.31)\\
P_3\big(\Sigma_{\infty;m,s};x\big) &\leqslant P_\infty\big(\Sigma_{\infty;m,s};x\big) && (2.32)\\
&= \exp\Big\{-\lambda\Big[\Big(1+\frac{z}{\sqrt{\lambda}}\Big)\ln\Big(1+\frac{z}{\sqrt{\lambda}}\Big)-\frac{z}{\sqrt{\lambda}}\Big]\Big\} && (2.33)\\
&\leqslant P_\infty\big(m+Z\sqrt{s};x\big), && (2.34)\\
P_3\big(\Sigma_{n;m,s};x\big) &\leqslant P_\infty\big(\Sigma_{n;m,s};x\big) && (2.35)\\
&= \Big(\frac{\lambda}{\lambda+z\sqrt{\lambda}}\Big)^{\lambda+z\sqrt{\lambda}}\Big(\frac{n-\lambda}{n-\lambda-z\sqrt{\lambda}}\Big)^{n-\lambda-z\sqrt{\lambda}} && (2.36)\\
&\leqslant P_\infty\big(\Sigma_{\infty;m,s};x\big), && (2.37)
\end{aligned}$$

for natural $n \geqslant \lambda$ and $z \in [-\sqrt{\lambda},0)$; for $z=-\sqrt{\lambda}$, the expressions in (2.33) and (2.36) for $P_\infty\big(\Sigma_{n;m,s};x\big)$ and $P_\infty\big(\Sigma_{\infty;m,s};x\big)$ are defined by continuity, as $e^{-\lambda}$ and $(1-\lambda/n)^n$, respectively; inequalities (2.34) and (2.37) follow by (2.30), (2.23), (2.10), (2.22), and (2.9).

The exponential upper bounds (2.31) and (2.35) are the same (up to a shift, rescaling, and reflection $x \mapsto -x$) as Hoeffding's bounds in [11, (2.1) and (2.3)], where they were obtained under an additional condition, which can be stated in terms of the present paper as

$$\mathsf{P}(X_i \leqslant \tfrac{s}{m}) = 1 \text{ for all } i \in \overline{1,n}. \tag{2.38}$$

Note that (2.38), together with the conditions (2.12) and (2.14), implies the second inequalities in (2.1) and (2.2) with $s_i := \frac{s}{m}\,m_i$.

For independent X_i's $\big($but without the additional restriction (2.38)$\big)$, the exponential upper bounds in (2.31) and (2.33) on $\mathsf{P}(S_n \leqslant x)$ — as well as the exact upper bound $\mathsf{E}\, f\big(\frac{s}{m}\Pi_{m^2/s}\big)$ on $\mathsf{E}\, f(S_n)$ for $f(x) \equiv e^{hx}$ with $h < 0$ — were essentially obtained in [22, Theorem 7]. Note two mistakes concerning the latter result: (i) in the proof in [22], $\psi(u)$ should be replaced by $\psi(hu)$ and (ii) what is presented as the proof of Theorem 7 in [22] is in fact that of Theorem 8 therein, and vice versa. Results of [22] seem yet relatively unknown, as the bound $e^{-z^2/2}$ on $\mathsf{P}(S_n \leqslant x)$ appeared later in [16].

By [25, Theorem 3.11] or [26, Theorem 4], with $c_{\alpha,0} := \Gamma(\alpha+1)(e/\alpha)^\alpha$,

$$P_\alpha(\eta; x) \leqslant c_{\alpha,0}\ \mathsf{P}(\eta \leqslant x)$$

provided that the tail function $x \mapsto \mathsf{P}(\eta \leqslant x)$ is log-concave. Combining this result with the Cantelli inequality, one also has the following upper bound on $\mathsf{P}(S_n \leqslant x)$:

$$W(z) := \min\Big(1, \frac{1}{1+z^2},\ c_{2,0}\,\mathsf{P}(Z \leqslant z)\Big);$$

note that $c_{2,0} = e^2/2 = 3.69\dots$. This bound may serve as an easier to compute and deal with approximation to the better bound $P_2\big(m + Z\sqrt{s}; x\big)$.

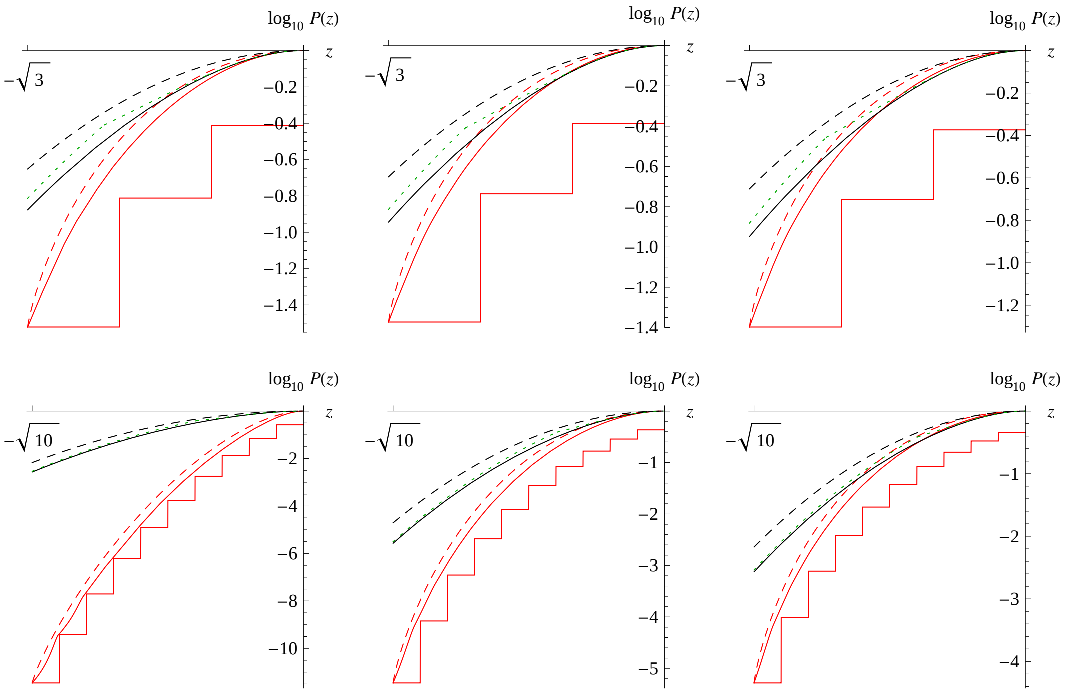

Figure 2: Decimal logarithms of the bounds/tails $P(z)$, for $\lambda = 10$ (first row) and $\lambda = 3$ (second row). The columns correspond to $n = 11$ (left), $n = 30$ (middle), and $n = \infty$ (right).

All the mentioned upper bounds $P(z) := P_\alpha(\eta; x)$ for η equal $\Sigma_{n;m,s}$ or $m + Z\sqrt{s}$ can be fully expressed in terms of z, λ, and n. These bounds are compared graphically in Figure 2 for $\lambda \in \{3, 10\}$, $\alpha \in \{0, 2, 3, \infty\}$, $n \in \{11, 30, \infty\}$, and $z \in \big(-\sqrt{\lambda}, 0\big)$; note that $\mathsf{P}(\Sigma_{n;m,s} \leqslant x) = P_\alpha(\Sigma_{n;m,s}; x) = 0$ if $z < -\sqrt{\lambda}$; here, as is natural, $P_\alpha(\Sigma_{n;m,s}; x)$ is interpreted as the true tail probability $\mathsf{P}(\Sigma_{n;m,s} \leqslant x)$ for $\alpha = 0$. The graphs of $\log_{10} P_\alpha(\Sigma_{n;m,s}; x)$ shown in Figure 2 are red: stepwise for $\alpha = 0$, solid-continuous for $\alpha = 3$, and dashed-continuous for $\alpha = \infty$. The graphs of $\log_{10} P_\alpha\big(m + Z\sqrt{s}; x\big)$ are black:

solid for $\alpha = 2$, and dashed for $\alpha = \infty$. No graphs are shown for $P_\alpha(\Sigma_{n;m,s}; x)$ with $\alpha = 2$, as those are not established bounds; nor is there a graph for $P_\alpha\big(m + Z\sqrt{s}; x\big)$ with $\alpha = 3$, as the better bound with $\alpha = 2$ is available. Also, a graph for $W(z)$ is shown, dotted-green.

It is seen that the bound $P_3\big(\Sigma_{n;m,s}; x\big)$ is close to the true tail probability $\mathsf{P}\,\big(\Sigma_{n;m,s} \leqslant x\big)$, especially for $\lambda = 10$ and $n = 11$, with a zero error at the left end-point $\big(-\sqrt{\lambda}\big)$ of the range of each of the r.v.'s $\big(\Sigma_{n;m,s} - m\big)/\sqrt{s}$, which is in accordance with part (iv)(b) of the mentioned [28, Proposition 3.2]. In the latter case ($\lambda = 10$ and $n = 11$), the bound $P_3\big(\Sigma_{n;m,s}; x\big)$ is over 8 times better near the left-end point of the range than the "normal" exponential bound $e^{-z^2/2}$. However, $P_3\big(\Sigma_{n;m,s}; x\big)$ may be slightly greater for z near 0 than the "normal" better-than-exponential bound $P_2\big(m + Z\sqrt{s}; x\big)$; this is due to the fact the class $\mathcal{F}_-^{1:2}$ is somewhat richer than $\mathcal{F}_-^{1:3}$.

3 Proofs

Proof of Theorem 2.4.

(I) By a standard induction argument (cf. e.g. [27, Lemma 12]), in order to prove part (I) of the theorem, it is enough to show that (2.7) holds for $n = 1$. Moreover, by [20, Corollary 5.9], we may assume that $f = f_{w,2}$ for some $w \in \mathbb{R}$, where $f_{w,2}$ is defined by formula (2.18). So, the proof of part (I) will be complete once it is shown that

$$\mathsf{E}\, f_{w,2}(X) \leqslant \mathsf{E}\, f_{w,2}\big(Y^{m,s}\big) \tag{3.1}$$

whenever the r.v. X is nonnegative, $\mathsf{E}\,X \geqslant m$, $\mathsf{E}\,X^2 \leqslant s$, $w \in \mathbb{R}$, and $0 < m \leqslant \sqrt{s}$. For $w \leqslant 0$, both sides of (3.1) are zero. So, w.l.o.g. $w > 0$. Introduce now $z := \frac{s}{m}$, $v := w \vee z$, and $c := \frac{w}{v}$, and then $g(x) := c^2(v - x)^2$. Then $\mathsf{P}(Y^{m,s} \in \{0, z\}) = 1$, $f_{w,2} \leqslant g$ on $[0, \infty)$ and $f_{w,2} = g$ on $\{0, z\}$, whence $g(Y^{m,s}) = f_{w,2}(Y^{m,s})$ almost surely (a.s.). Note also that $v > 0$ and recall the relations $\mathsf{E}\,X \geqslant m = \mathsf{E}\,Y^{m,s}$ and $\mathsf{E}\,X^2 \leqslant s = \mathsf{E}(Y^{m,s})^2$. Thus,

$$\begin{aligned}
\mathsf{E}\, f_{w,2}(X) \leqslant \mathsf{E}\, g(X) &= c^2(v^2 - 2v\,\mathsf{E}\,X + \mathsf{E}\,X^2) \\
&\leqslant c^2\big(v^2 - 2v\,\mathsf{E}\,Y^{m,s} + \mathsf{E}(Y^{m,s})^2\big) \\
&= \mathsf{E}\, g(Y^{m,s}) = \mathsf{E}\, f_{w,2}(Y^{m,s}),
\end{aligned}$$

which completes the proof of part (I) of Theorem 2.4.

(II) Take any $f \in \mathcal{F}_-^{1:3}$ and consider

$$F_{n,f}(P_1, \dots, P_n) := \mathsf{E}\, f\big(Y^{m_1,s_1} + \dots + Y^{m_n,s_n}\big),$$

the right-hand side of (2.7), where

$$P_i := (m_i, s_i) \tag{3.2}$$

for all i. Note that the function $F_{n,f}$ is symmetric (with respect to all permutations of its n arguments, $P_1, \dots, P_n$). Next, if nonnegative r.v.'s $X_1, \dots, X_n$ satisfy the (m, s)-condition, they satisfy the $(\mathbf{m}, \mathbf{s})$-condition for some $m_1, \dots, m_n, s_1, \dots, s_n$ such that $m_1 + \dots + m_n = m$ and $s_1 + \dots + s_n = s$. So, by (2.7), to prove (2.8) it is enough to show that

$$F_{n,f}(P_1, \dots, P_n) \leqslant F_{n,f}(\bar{P}_n, \dots, \bar{P}_n), \tag{3.3}$$

where $\bar{P}_n := \frac{1}{n}(P_1 + \dots + P_n)$. Here we shall need the following lemma, which establishes a Schur-concavity-like property of the symmetric function $F_{n,f}$.

Lemma 3.1. *For any natural $n \geqslant 2$ and any $t \in [0, 1]$*

$$F_{n,f}(P_1, \dots, P_n) \leqslant F_{n,f}(P_{1+t}, P_{2-t}, P_3 \dots, P_n),$$

where $P_{1+t} := (1 - t)P_1 + tP_2$ and hence $P_{2-t} = tP_1 + (1 - t)P_2$.

The proof of Lemma 3.1 will be given at the end of this section.

Note that $F_{n,f}$ is a function of n points $P_1,\dots,P_n$ in $\mathbb{R}^2$, rather than of n real arguments. If the latter were the case, then Lemma 3.1 together with the well-known Muirhead lemma (see e.g. [15, Lemma 2.B.1]) would immediately imply the Schur-concavity and hence (3.3). However, no appropriate "multidimensional" analogue of the Muirhead lemma seems to exist. Indeed, if one defines the "multivariate" majorization by means of doubly stochastic matrices (in accordance with the Hardy-Littlewood-Polya characterization — see e.g. [15, Theorem 2.B.2]), then the analogue of the Muirhead lemma fails to hold. For example, take $n=3$ and consider the doubly stochastic 3×3 matrices $\big($say A and B_t, for some $t\in[0,1]\big)$ that transform any triple $\tau:=(Q_1,Q_2,Q_3)$ of points in $\mathbb{R}^2$ to (say) $\tilde\tau:=\big(\frac{Q_1+Q_2}{2},\frac{Q_1+Q_3}{2},\frac{Q_2+Q_3}{2}\big)$ and $\tau_t:=\big((1-t)Q_1+tQ_2,tQ_1+(1-t)Q_2,Q_3\big)$, respectively; matrices such as B_t are referred to as T-transform matrices, all of which latter can be written as $C^{-1}B_tC$ for some $t\in[0,1]$ and some permutation matrix C; see e.g. [15, Section 2.B]. Then, if the points Q_1,Q_2,Q_3 are not collinear, already after one application of any matrix B_t with $t\in(0,1)$ to τ one will never be able to get from τ_t to $\tilde\tau$ via any chain of T-transforms, since the points $\frac{Q_1+Q_3}{2}$ and $\frac{Q_2+Q_3}{2}$ do not belong to the convex hull of the set $\{(1-t)Q_1+tQ_2,tQ_1+(1-t)Q_2,Q_3\}$.

We shall verify (3.3) by induction on n. For $n=1$, (3.3) is trivial. Suppose that (3.3) holds for n equal some natural k, and consider $n=k+1$. Introduce $\tilde P_k:=\frac{1}{k+1}\bar P_k+(1-\frac{1}{k+1})P_{k+1}$, $f_{k+1}(x):=\mathsf{E}\,f\big(x+Y^{m_{k+1},s_{k+1}}\big)$, and $g_{k+1}(x):=\mathsf{E}\,f\big(x+Y^{\bar m_{k+1},\bar s_{k+1}}\big)$, where $(\bar m_{k+1},\bar s_{k+1}):=\bar P_{k+1}$. By Remark 2.3, the functions f_{k+1} and g_{k+1} are in $\mathcal{F}_-^{1:3}$. Also,

$$\tfrac{1}{k}\left(\tilde P_k+(k-1)\bar P_k\right)=\bar P_{k+1}. \tag{3.4}$$

It follows that

$$\begin{aligned}
&F_{k+1,f}(P_1,\dots,P_{k+1})\\
&\qquad=\mathsf{E}\,F_{k,f_{k+1}}(P_1,\dots,P_k) &&\big(\text{by the definition of } f_{k+1}\big)\\
&\qquad\leqslant\mathsf{E}\,F_{k,f_{k+1}}(\bar P_k,\dots,\bar P_k) &&\big(\text{by induction}\big)\\
&\qquad=F_{k+1,f}(P_{k+1},\bar P_k,\dots,\bar P_k) &&\big(\text{by the definition of } f_{k+1}\\
& &&\ \text{and the symmetry of } F_{k+1,f}\big)\\
&\qquad\leqslant F_{k+1,f}(\tilde P_k,\bar P_{k+1},\bar P_k,\dots,\bar P_k) &&\big(\text{by Lemma 3.1 with } t=\tfrac{1}{k+1}\big)\\
&\qquad=\mathsf{E}\,F_{k,g_{k+1}}(\tilde P_k,\bar P_k,\dots,\bar P_k) &&\big(\text{by the definition of } g_{k+1}\big)\\
&\qquad\leqslant\mathsf{E}\,F_{k,g_{k+1}}(\bar P_{k+1},\dots,\bar P_{k+1}) &&\big(\text{by induction and (3.4)}\big)\\
&\qquad=F_{k+1,f}(\bar P_{k+1},\dots,\bar P_{k+1}) &&\big(\text{by the definition of } g_{k+1}\big).
\end{aligned}$$

This completes the proof of (2.8), modulo Lemma 3.1.

By an argument similar to that used in the proof of part (I) of Theorem 2.4, it is enough to verify (2.9) and (2.10) for $f=f_{w,3}$, and (2.11) for $f=f_{w,2}$.

In inequality (2.8) with $n+1$ instead of n, take $X_{n+1}=0$ and $X_i=Y_i^{\frac{m}{n},\frac{s}{n}}$ for $i\in\overline{1,n}$; it then follows that the right hand-side of (2.8) is nondecreasing in n, for any fixed positive real m and s. Next, (i) all the r.v.'s in (2.8) and (2.9) are nonnegative, (ii) the function $f_{w,3}$ is continuous and bounded on $[0,\infty)$, and (iii) $Y_1^{\frac{m}{n},\frac{s}{n}}+\dots+Y_n^{\frac{m}{n},\frac{s}{n}}$ converges in distribution to $\frac{s}{m}\Pi_{m^2/s}$ as $n\to\infty$. So, the right hand-side of (2.8) is, not only nondecreasing in n, but also converging to the right hand-side of (2.9) as $n\to\infty$ (for $f=f_{w,3}$). Thus, (2.9) follows.

As for inequality (2.10), it is essentially a special case of (2.11). Indeed, consider the latter inequality with $n\to\infty$ and $X_1=X_1^{(n)},\dots,X_n=X_n^{(n)}$ being independent copies of $c_n\Pi_{\lambda_n}$, where $c_n:=\frac{s-m^2/n}{m}\sim\frac{s}{m}$ and $\lambda_n:=\frac{m^2}{ns-m^2}\sim\frac{m^2}{ns}$. Then the r.v.'s $X_1,\dots,X_n$ satisfy the (m,s)-condition, and S_n converges to $\frac{s}{m}\Pi_{m^2/s}$ in distribution. Therefore, $\mathsf{E}\,f_{w,2}(S_n)\longrightarrow\mathsf{E}\,f_{w,2}\big(\frac{s}{m}\Pi_{m^2/s}\big)$.

Thus, it remains to prove (2.11), for $f = f_{w,2}$. If at that $w \leqslant 0$, then the left-hand side of (2.11) is zero, while its right-hand side is nonnegative. Therefore and by rescaling, w.l.o.g. $w = 1$. Also, as in the proof of part (I) of Theorem 2.4, w.l.o.g. $n = 1$. Thus, also in view of (2.7) and (2.4), to complete the proof of Theorem 2.4, it suffices to show that

$$\delta(m) := \delta(m,k) := \frac{\mathsf{E}\left(1 - m - kmZ\right)_+^2 - \mathsf{E}\left(1 - Y^{m,\,k^2m^2}\right)_+^2}{2\big(k^2m^2 + (m-1)^2\big)} \geqslant 0$$

for all $m \in (0,\infty)$ and $k \in (1,\infty)$. Take indeed any $k \in (1,\infty)$. Note that

$$\delta'(m)\, k^2 \left(k^2m^2 + (m-1)^2\right)^2 = \begin{cases} (D\delta)_1(m) \text{ if } m \in (0, 1/k^2], \\ (D\delta)_2(m) \text{ if } m \in [1/k^2, \infty), \end{cases}$$

where

$$(D\delta)_1(m) := k^2m(1-m) - k^5m^2\varphi\Big(\frac{m-1}{km}\Big),$$
$$(D\delta)_2(m) := (k^2-1)(k^2m - 1 + m) - k^5m^2\varphi\Big(\frac{m-1}{km}\Big),$$

and φ is the standard normal density function. Next, for $m \in (0, 1/k^2]$ one has $m(1-m) > 0$ and

$$\frac{\mathrm{d}}{\mathrm{d}m}\Big(\frac{(D\delta)_1(m)}{m(1-m)}\Big) = -\frac{k^3\left(k^2m^2 + (m-1)^2\right)}{(1-m)^2m^2}\,\varphi\Big(\frac{m-1}{km}\Big) < 0;$$

so, $(D\delta)_1$ — and hence δ' — may change in sign on the interval $(0, 1/k^2]$ at most once, and only from $+$ to $-$. Similarly, for $m \in (1/k^2, \infty)$ one has $k^2m - 1 + m > k^2m - 1 > 0$ and

$$\frac{\mathrm{d}}{\mathrm{d}m}\Big(\frac{(D\delta)_2(m)}{k^2m - 1 + m}\Big) = -\frac{k^3\left(k^2m - 1\right)\left(k^2m^2 + (m-1)^2\right)}{m\left(k^2m - 1 + m\right)^2}\,\varphi\Big(\frac{m-1}{km}\Big) < 0;$$

so, $(D\delta)_2$ — and hence δ' — may change in sign on the interval $[1/k^2, \infty)$ at most once, and only from $+$ to $-$. Thus, δ' may change in sign on the interval $(0,\infty)$ at most once, and only from $+$ to $-$. It follows that $\delta(m) \geqslant \delta(0+) \wedge \delta(\infty-)$ for all $m \in (0,\infty)$. So, to complete the proof of Theorem 2.4, it remains to check that $\delta(0+) \wedge \delta(\infty-) \geqslant 0$. In fact, one can see that $\delta(0+) = 0$ and

$$2\delta(\infty-) = q(t) := \mathsf{P}(Z > t) - \frac{t\,\varphi(t)}{t^2+1} > 0, \tag{3.5}$$

with $t := 1/k > 0$. The inequality in (3.5) is well known; see e.g. [37, (19) for ϕ_2]; alternatively, it follows because $q'(t) = -\frac{2\varphi(t)}{(t^2+1)^2} < 0$ and $q(\infty-) = 0$. This completes the entire proof of Theorem 2.4, modulo Lemma 3.1. □

Proof of Lemma 3.1. W.l.o.g. $n = 2$ — cf. e.g. the first equality in the big display following (3.4). Also, by the symmetry under permutations, w.l.o.g. $t \in [0, \frac12]$. Moreover, w.l.o.g. $t \neq \frac12$; here and elsewhere we are using (sometimes tacitly) a version of continuity relevant in a given context. So, it suffices to show that $G'(t) \geqslant 0$ for all $t \in [0, \frac12)$, where

$$G(t) := G_{P_1,P_2}(t) := F_{2,f}(P_{1+t}, P_{2-t}). \tag{3.6}$$

Actually, it is enough to show that

$$G'(0) \overset{(?)}{\geqslant} 0, \tag{3.7}$$

because for any $\tau \in [0, \frac12)$ and $s := \frac{t-\tau}{1-2\tau}$, one has $P_{1+t} = (1-s)P_{1+\tau} + sP_{2-\tau}$ and $P_{2-t} = sP_{1+\tau} + (1-s)P_{2-\tau}$, whence $G_{P_1,P_2}(t) = G_{P_{1+\tau},P_{2-\tau}}(s)$ and $G'_{P_1,P_2}(\tau) = G'_{P_{1+\tau},P_{2-\tau}}(0)/(1-2\tau)$. Next — cf. the proof of part (I) of Theorem 2.4 — w.l.o.g. $f(x) = (w-x)_+^3$ for some $w \in \mathbb{R}$ and all $x \in \mathbb{R}$. Thus,

$$G(t) = \mathsf{E}(w - Y^{m_{1+t},s_{1+t}} - Y^{m_{2-t},s_{2-t}})_+^3,$$

where $(m_u, s_u) := P_u$ for any u. If $w \leqslant 0$ then $G(t) = 0$ for all t, so that there is nothing to prove. Therefore, by rescaling, w.l.o.g. $w = 1$. So, in view of Definition 2.2, $G(t)$ can be expressed in terms of the variables t, a, p, b, q only, where

$$a := \frac{s_1}{m_1} > 0, \quad b := \frac{s_2}{m_2} > 0, \quad p := \frac{m_1^2}{s_1} \in (0,1], \quad q := \frac{m_2^2}{s_2} \in (0,1]. \tag{3.8}$$

By the symmetry relation $G_{P_1,P_2}(t) = G_{P_2,P_1}(t)$ (and continuity), w.l.o.g. $0 < b < a$, so that $0 < b < a < a + b$. Thus, it suffices to consider the following four cases:

(C_0) $1 \in (a + b, \infty)$;

(C_1) $1 \in (a, a + b)$;

(C_2) $1 \in (b, a)$;

(C_3) $1 \in (0, b)$;

at that, with each case it is assumed $0 < b < a$ and $0 < p, q < 1$. In each of the cases (C_k) with $k \in \{0, 1, 2, 3\}$, the expression

$$D_k := A_k G'(0), \tag{3.9}$$

is a polynomial in a, b, p, q, where

$$A_0 := 1,\ A_1 := A_3 := a^2b^2, \text{ and } A_2 := a^2. \tag{3.10}$$

Therefore, to finish the proof of inequality (3.7) and thus that of Lemma 3.1, it remains to verify the following lemma. □

Lemma 3.2. *In each of the cases (C_k) with $k \in \{0, 1, 2, 3\}$, the polynomial D_k in a, b, p, q, defined by* (3.9) *and* (3.10)*, is nonnegative for all p and q in $(0, 1)$.*

Proof of Lemma 3.2. For each k, D_k is a polynomial and the conditions that define the case C_k are polynomial (in fact, affine) inequalities. So, the verification that D_k is nonnegative in each of the cases C_k can be done in a completely algorithmic manner, due to the well-known Tarski theory [38, 14, 7]. This theory is implemented in Mathematica via `Reduce` and other related commands. Thus, the Mathematica command `Reduce[der0 < 0 && case0]` (where `der0` and `case0` stand for D_0 and $\big[(C_0)\ \&\ 0 < b < a\ \&\ p \in (0,1)\ \&\ q \in (0,1)\big]$, respectively) outputs `False` (in about 0.3 sec on a standard desktop), which means that indeed $D_0 \geqslant 0$ in the case (C_0). Cases (C_1), (C_2), and (C_3) can be treated quite similarly, with Mathematica execution times of about 5.4 sec, 0.65 sec, and 0.04 sec, respectively.

Details of the corresponding calculations can be found in Mathematica notebook solution-tarsky.nb and its pdf copy solution-tarsky.pdf in the folder Mathematica in the zip file LeftTailBounds.zip posted at the SelectedWorks site `works.bepress.com/iosif-pinelis/7/download/`. The symbols `der0`, ..., `der3` in the mentioned Mathematica notebook correspond to $D_0, \dots, D_3$ defined by formulas (3.9)–(3.10) in the paper.

This completes the proof of Lemma 3.2, which appears no less reliable than computations done "by hand"; cf. e.g. the views of Okounkov [17, page 35], Voevodsky [35], and Odlyzko [18] on computer-assisted proofs. □

However, as Okounkov [17] notes in his interview, "perhaps we should not be dependent on commercial software here". Indeed, details of the execution of the Mathematica command `Reduce[]` are not open to examination.

Therefore, in addition to the above proof, in the next section an alternative proof of Lemma 3.2 is provided, which relies, instead of the Mathematica command `Reduce`, on the Redlog package of the computer algebra system Reduce; both Reduce and Redlog are open-source and freely distributed (`http://www.redlog.eu/`).

Yet another proof of Lemma 3.2 is given in Section 5 of the arXiv version [33] of this paper. That proof, which is very long, uses only standard tools of calculus and also such a standard tool of algebra as the resultant.

4 Alternative proof of Lemma 3.2

Recall that, for each $k \in \{0,1,2,3\}$, D_k is a polynomial in a,b,p,q. For each $k \in \{0,1,2,3\}$, in the case (C_k), the quadruple (a,b,p,q) belongs to the set

$$\Omega_k := \omega_k \times (0,1)^2, \tag{4.1}$$

where

$$\begin{aligned}
\omega_0 :=& \{(a,b) \in \mathbb{R}^2 \colon 0 < b < a < a+b < 1\} \\
=& \{(a,b) \in \mathbb{R}^2 \colon b < a < a+b < 1\}, \\
\omega_1 :=& \{(a,b) \in \mathbb{R}^2 \colon 0 < b < a < 1 < a+b\} \\
=& \{(a,b) \in \mathbb{R}^2 \colon b < a < 1 < a+b\}, \\
\omega_2 :=& \{(a,b) \in \mathbb{R}^2 \colon 0 < b < 1 < a\}, \\
\omega_3 :=& \{(a,b) \in \mathbb{R}^2 \colon 1 < b < a\}.
\end{aligned} \tag{4.2}$$

For each $k \in \{0,1,2,3\}$, let $\bar{\omega}_k$ denote the topological closure of ω_k, so that $\bar{\omega}_k$ is defined by the system of non-strict inequalities corresponding to the strict inequalities defining the set ω_k.

We shall use notation such as the following:

$$D_{k;p=\delta} := D_k\big|_{p=\delta}, \quad D_{k;q=\varepsilon} := D_k\big|_{q=\varepsilon}, \quad D_{k;p=\delta,q=\varepsilon} := D_k\big|_{p=\delta,q=\varepsilon}; \tag{4.3}$$

sometimes in such notation we shall use, instead of D_k, a modified version $\tilde{D}_k$ of D_k, which differs from D_k by a factor which is manifestly positive in the corresponding context.

> In files pertaining to the mentioned package Redlog, we shall use notations such as `der0`, ..., `der3`, `der0p0`, ..., `der3p1q1` for $D_0, \dots, D_3, D_{0;p=0}, \dots, D_{3;p=1,q=1}$, respectively, or possibly for $\tilde{D}.$ in place of $D.$.

Unfortunately, for polynomials in several variables the mentioned package Redlog is either much slower than Mathematica (as in the cases of the polynomials D_0 and D_3 in (3.9)) or unable to complete the desired verification of the nonnegativity (as in the cases of the polynomials D_1 and D_2 in (3.9)). I have also tried another well-known open-source program, QEPCAD B (Quantifier Elimination by Partial Cylindrical Algebraic Decomposition, Version B), but it crashes even where Redlog eventually produces the result.

More specifically, Redlog verifies the nonnegativity of the polynomials D_0 and D_3 (in cases (C_0) and (C_3)) in about 107 min and 0.45 sec, respectively; details on this can be found in the .log files der0.log and der3.log and in the corresponding .png files der0.png and der3.png.

> The .log and .png files mentioned in this section are in the folder Reduce(Redlog) in the zip file LeftTailBounds.zip at the SelectedWorks site `works.bepress.com/iosif-pinelis/7/download/`.

These execution times, 107 min and 0.45 sec, may be compared with the corresponding ones for Mathematica, mentioned in the proof of Lemma 3.2 in the preceding section: 0.3 sec and 0.04 sec).

To verify the nonnegativity of the polynomials D_1 and D_2 with Redlog, each of these two verification problems has to be reduced, by a human, to a series (or rather a tree) of simpler problems, as presented below.

Lemma 4.1. *In the case* (C_1)*, the polynomial* D_1 *in* a, b, p, q *is nonnegative for all* p *and* q *in* $(0, 1)$ *– that is,* $D_1 \geqslant 0$ *for all* $(a, b, p, q) \in \Omega_1$*.*

Proof. Assume indeed in this proof that $(a, b, p, q) \in \Omega_1$, unless otherwise stated. One has

$$\begin{aligned} D_1 = a^5b^2p - 2a^4b^3p + a^3b^4p + a^4p^2 - 3a^5p^2 + 3a^6p^2 - a^7p^2 \\ - 2a^3bp^2 + 6a^4bp^2 - 6a^5bp^2 + 2a^6bp^2 + 3a^3b^2p^2 - 3a^4b^2p^2 \\ + 2a^4b^3p^2 - a^3b^4p^2 + a^4b^3q - 2a^3b^4q + a^2b^5q + 2a^2b^2pq \\ - 6a^3b^2pq + 6a^4b^2pq - 2a^5b^2pq - 6a^2b^3pq + 6a^2b^4pq - 2a^2b^5pq \\ - 2ab^3q^2 + 3a^2b^3q^2 - a^4b^3q^2 + b^4q^2 + 6ab^4q^2 - 3a^2b^4q^2 + 2a^3b^4q^2 \\ - 3b^5q^2 - 6ab^5q^2 + 3b^6q^2 + 2ab^6q^2 - b^7q^2. \end{aligned} \tag{4.4}$$

Consider

$$\begin{aligned} \frac{\partial_p^2 D_1\, \partial_q^2 D_1 - (\partial_p \partial_q D_1)^2}{4a^3(a-b)^2b^3} &= \det{}_1 \\ &:= -2 + 9a - 15a^2 + 10a^3 - 3a^5 + a^6 + 9b - 33ab + 48a^2b - 36a^3b + 15a^4b \\ &\quad - 3a^5b - 15b^2 + 48ab^2 - 54a^2b^2 + 24a^3b^2 - 3a^4b^2 + 10b^3 - 36ab^3 + 24a^2b^3 \\ &\quad - 7a^3b^3 + 15ab^4 - 3a^2b^4 - 3b^5 - 3ab^5 + b^6; \end{aligned} \tag{4.5}$$

here and in the sequel, ∂_α denotes, as usual, the partial differentiation in α. Using the mentioned package Redlog, we see that $\det_1 < 0$ on ω_1; this takes about 0.5 sec; see details in the files der1det.log and der1det.png.

Hence, the determinant of the Hessian matrix of D_1 with respect to p and q is negative for all $(a, b, p, q) \in \Omega_1$. It follows that D_1 is saddle-like in p and q, and so, for each fixed $(a, b) \in \omega_1$, the minimum of the polynomial D_1 in $(p, q) \in [0, 1]^2$ is not attained at any point $(p, q) \in (0, 1)^2$; therefore, this minimum is attained at some point (p, q) on the boundary of the unit square $[0, 1]^2$.

Consider then each of the four boundary subcases of Case 1: $p = 0$, $p = 1$, $q = 0$, and $q = 1$. Using Redlog, we see that $D_{1;p=0} \geqslant 0$ for $(a, b, q) \in \omega_1 \times (0, 1)$ (execution time ≈ 0.25 sec; details in the files der1p0.log and der1p0.png) and $D_{1;q=0} \geqslant 0$ for $(a, b, p) \in \omega_1 \times (0, 1)$ (execution time ≈ 0.25 sec; details in the files der1q0.log and der1q0.png).

The subcases $p = 1$ and $q = 1$ require more care. Recall notation (4.3).

To consider the subcase $p = 1$, assume that $(a, b) \in \omega_1$ and $q \in (0, 1)$. In view of (4.4),

$$\begin{aligned} D_{1;p=1} = a^3(a - 3a^2 + 3a^3 - a^4 - 2b + 6ab - 6a^2b + 2a^3b + 3b^2 - 3ab^2 + a^2b^2) \\ + a^2b^2(2 - 6a + 6a^2 - 2a^3 - 6b + a^2b + 6b^2 - 2ab^2 - b^3)q \\ - b^3(2a - 3a^2 + a^4 - b - 6ab + 3a^2b - 2a^3b + 3b^2 + 6ab^2 - 3b^3 - 2ab^3 + b^4)q^2. \end{aligned} \tag{4.6}$$

Using Redlog, we see (in about 0.16 sec) that

$$D_{1;p=1}^{002} := \frac{\partial_q^2 D_{1;p=1}}{2b^3} = 3a^2(1-b) - (2a-b)(1-b)^3 + a^3(2b-a) \geqslant 0 \tag{4.7}$$

and (in about 0.8 sec) that

$$\begin{aligned} D_{1;p=1,q=1}^{001} &:= \frac{\partial_q D_{1;p=1}\big|_{q=1}}{b^2} \\ &= a^4(6 - b - 2a) - 2(1-b)^3b(2a-b) - 2a^3(3-b^2) + a^2(2-b^3) \leqslant 0 \end{aligned} \tag{4.8}$$

(details in the files der1p1.log and der1p1.png; notations `D002der1p1` and `D001der1p1q1` there correspond to $D^{002}_{1;p=1}$ and $D^{001}_{1;p=1,q=1}$, respectively). So, $D_{1;p=1}$ is convex and decreasing in q. At that,

$$D_{1;p=1}\big|_{q=1} = (a-b)^2[a^2(1-a)^3 + (1-b)^3b^2] > 0. \tag{4.9}$$

We conclude that indeed $D_{1;p=1} \geqslant 0$.

To complete the proof of Lemma 4.1, it remains to consider the subcase $q = 1$. Expanding $D_{1;q=1}$ in powers of p, one has

$$D_{1;q=1} = \psi(p) := Ap^2 + Bp + C,$$

where

$$\begin{aligned} A &:= -a^3(a^4 - 2a^3b - 3a^3 + 6a^2b + 3a^2 - 2ab^3 + 3ab^2 - 6ab - a + b^4 \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad - 3b^2 + 2b), \\ B &:= -a^2b^2(a^3 + 2a^2b - 6a^2 - ab^2 + 6a + 2b^3 - 6b^2 + 6b - 2), \\ C &:= b^3(a^2b^2 - 3a^2b + 3a^2 + 2ab^3 - 6ab^2 + 6ab - 2a - b^4 + 3b^3 - 3b^2 + b). \end{aligned} \tag{4.10}$$

Let

$$\begin{aligned} \text{discr} :=& \frac{B^2 - 4AC}{a^3b^3(a-b)^2} \\ =& 8 - 36a + 60a^2 - 44a^3 + 12a^4 - 36b + 132ab - 180a^2b + 108a^3b - 24a^4b \\ &+ a^5b + 60b^2 - 192ab^2 + 192a^2b^2 - 84a^3b^2 + 10a^4b^2 - 40b^3 + 144ab^3 \\ &- 84a^2b^3 + 21a^3b^3 - 60ab^4 + 12a^2b^4 + 12b^5 + 12ab^5 - 4b^6, \\ d_1 :=& \frac{\psi'(0)}{a^2b^2} = \frac{B}{a^2b^2} = 2 - 6a + 6a^2 - a^3 - 6b - 2a^2b + 6b^2 + ab^2 - 2b^3, \\ d_2 :=& \frac{\psi'(1)}{a^2(a-b)} = \frac{2A+B}{a^2(a-b)} = 2a - 6a^2 + 6a^3 - 2a^4 - 2b + 6ab - 6a^2b + 2a^3b \\ &\qquad\qquad\qquad\qquad\qquad + 6b^2 - 6ab^2 + a^2b^2 - 6b^3 + 3ab^3 + 2b^4. \end{aligned}$$

Note that discr equals in sign the discriminant of the quadratic polynomial $\psi(p)$. Therefore, discr > 0 if and only if $\psi(p)$ takes both positive and negative values as p varies from $-\infty$ to ∞. Using Redlog, we see that (i) $A \geqslant 0$ (≈ 0.16 sec execution time); (ii) the conjunction of the conditions discr > 0, $d_1 < 0$, and $b < 1/2$ never takes place over the set ω_1 (≈ 3.1 sec execution time); and (iii) the conjunction of the conditions discr > 0, $d_2 > 0$, and $b > 1/2$ never takes place over the set ω_1 (≈ 25.5 min execution time); details are in the files der1q1.log, der1q1-top.png (for the first 10 Redlog commands), and der1q1-bottom.png (for the last 10 Redlog commands); in those files, `AA` stands for A/a^3, with A as in (4.10). So, over the set ω_1 one has the following: (i') the function ψ is convex; (ii') if $b < 1/2$ and ψ changes sign over $\mathbb{R}$, then $\psi'(0) \geqslant 0$ and hence $\psi(p)$ is nondecreasing in $p \in [0,1]$; and (iii') if $b > 1/2$ and ψ changes sign over $\mathbb{R}$, then $\psi'(1) \leqslant 0$ and hence $\psi(p)$ is nonincreasing in $p \in [0,1]$. Thus, in view of the continuity of $D_{1;q=1} = \psi(p)$ in b, it remains to verify that $\tilde{D}_{1;q=1,p=0} := \psi(0)/b^3 = C/b^3$ and $\tilde{D}_{1;q=1,p=1} = \psi(1)/(a-b)^2 = (A+B+C)/(a-b)^2$ are both nonnegative (over ω_1). For $\tilde{D}_{1;q=1,p=0}$ this is checked by Redlog in about 0.05 sec (details in files der1q1.log and der1q1-bottom.png), whereas $\tilde{D}_{1;q=1,p=1} = a^2(1-a)^3 + b^2(1-b)^3$ is manifestly positive (over ω_1).

This completes the proof of Lemma 4.1. □

Lemma 4.2. *In the case (C_2), the polynomial D_2 in a, b, p, q is nonnegative for all p and q in $(0,1)$ – that is, $D_2 \geqslant 0$ for all $(a,b,p,q) \in \Omega_1$.*

Proof. Assume indeed in this proof that $(a,b,p,q)\in\Omega_2$, unless otherwise stated. One has

$$D_2 = a^2p - 3a^3p + 3a^4p - 2a^4bp + a^3b^2p + 3a^3p^2 - 3a^4p^2 + 2a^4bp^2 - a^3b^2p^2 - 2abq \\ + 3a^2bq + b^2q - 3a^2b^2q + a^2b^3q - 6a^2bpq + 6a^2b^2pq - 2a^2b^3pq + 6ab^2q^2 - 3b^3q^2 - 6ab^3q^2 \\ + 3b^4q^2 + 2ab^4q^2 - b^5q^2. \quad (4.11)$$

Using Redlog (details in the files der2.log, der2-top.png, and der2-bottom.png), we see that

$$\tfrac{1}{2}\,\partial_a^2 D_2 = p - 9ap + 18a^2p - 12a^2bp + 3ab^2p + 9ap^2 - 18a^2p^2 + 12a^2bp^2 - 3ab^2p^2 \\ + 3bq - 3b^2q + b^3q - 6bpq + 6b^2pq - 2b^3pq \geqslant 0 \quad (4.12)$$

on Ω_2 (execution time ≈ 22.6 min) – so that D_2 is convex in a,

$$D_{2;a=1} = p - 2bp + b^2p + 2bp^2 - b^2p^2 + bq - 2b^2q + b^3q - 6bpq + 6b^2pq - 2b^3pq + 6b^2q^2 \\ - 9b^3q^2 + 5b^4q^2 - b^5q^2 \geqslant 0 \quad (4.13)$$

for b,p,q in $(0,1)$ (execution time ≈ 1.2 sec), and

$$\partial_a D_2\big|_{a=1} = 5p - 8bp + 3b^2p - 3p^2 + 8bp^2 - 3b^2p^2 + 4bq - 6b^2q + 2b^3q - 12bpq + 12b^2pq \\ - 4b^3pq + 6b^2q^2 - 6b^3q^2 + 2b^4q^2 \geqslant 0 \quad (4.14)$$

for b,p,q in $(0,1)$ (execution time ≈ 1.2 sec); the symbols `der2DDa`, `der2a1`, and `Dder2a1` in the mentioned Redlog files stand for $\frac{1}{2}\,\partial_a^2 D_2$, $D_{2;a=1}$, and $\partial_a D_2\big|_{a=1}$, respectively. To complete the proof of Lemma 4.2, it remains to recall the definition (4.1). □

Lemma 3.2 follows immediately from Lemmas 4.1 and 4.2 and the nonnegativity of D_0 and D_3, mentioned in the beginning of this section.

References

[1] Bengt von Bahr and Carl-Gustav Esseen, *Inequalities for the rth absolute moment of a sum of random variables,* $1 \leqslant r \leqslant 2$, Ann. Math. Statist **36** (1965), 299–303. MR-0170407

[2] George Bennett, *Probability inequalities for the sum of independent random variables*, J. Amer. Statist. Assoc. **57** (1962), no. 297, 33–45.

[3] V. Bentkus, N. Kalosha, and M. van Zuijlen, *On domination of tail probabilities of (super)martingales: explicit bounds*, Liet. Mat. Rink. **46** (2006), no. 1, 3–54. MR-2251440

[4] Vidmantas Bentkus, *On Hoeffding's inequalities*, Ann. Probab. **32** (2004), no. 2, 1650–1673. MR-2060313

[5] Patrick Billingsley, *Convergence of probability measures*, John Wiley & Sons Inc., New York, 1968. MR-0233396

[6] Stéphane Boucheron, Olivier Bousquet, Gábor Lugosi, and Pascal Massart, *Moment inequalities for functions of independent random variables*, Ann. Probab. **33** (2005), no. 2, 514–560. MR-2123200

[7] George E. Collins, *Quantifier elimination for real closed fields by cylindrical algebraic decomposition*, Quantifier elimination and cylindrical algebraic decomposition (Linz, 1993), Texts Monogr. Symbol. Comput., Springer, Vienna, 1998, pp. 85–121. MR-1634190

[8] Christopher R. Dance, *An inequality for the sum of independent bounded random variables*, J. Theoret. Probab. **27** (2014), no. 2, 358–369. MR-3195817

[9] Morris L. Eaton, *A note on symmetric Bernoulli random variables*, Ann. Math. Statist. **41** (1970), 1223–1226. MR-0268930

[10] Morris L. Eaton, *A probability inequality for linear combinations of bounded random variables*, Ann. Statist. **2** (1974), 609–613.

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

[12] R. Ibragimov and Sh. Sharakhmetov, *The best constant in the Rosenthal inequality for nonnegative random variables*, Statist. Probab. Lett. **55** (2001), no. 4, 367–376. MR-1877641

[13] Rafał Latała, *Estimation of moments of sums of independent real random variables*, Ann. Probab. **25** (1997), no. 3, 1502–1513. MR-1457628

[14] S. Łojasiewicz, *Sur les ensembles semi-analytiques*, Actes du Congrès International des Mathématiciens (Nice, 1970), Tome 2, Gauthier-Villars, Paris, 1971, pp. 237–241. MR-0425152

[15] Albert W. Marshall and Ingram Olkin, *Inequalities: theory of majorization and its applications*, Mathematics in Science and Engineering, vol. 143, Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York, 1979. MR-552278

[16] Andreas Maurer, *A bound on the deviation probability for sums of non-negative random variables*, JIPAM. J. Inequal. Pure Appl. Math. **4** (2003), no. 1, Article 15, 6 pp. (electronic). MR-1965995

[17] Vicente Muñoz and Ulf Persson, *Interviews with three Fields medallists: Andrei Okounkov*, Newsletter of the European Mathematical Society (2006, December), no. 62, 34–35, `http://www.ams.org/notices/200703/comm-fields-interviews.pdf`, pp. 1–2.

[18] Andrew Odlyzko, *Review: Experimental Mathematics in Action*, Amer. Math. Monthly **118** (2011), no. 10, 946–951, `http://dx.doi.org/10.4169/amer.math.monthly.118.10.946`.

[19] Adam Osękowski, *Sharp inequalities for sums of nonnegative random variables and for a martingale conditional square function*, ALEA Lat. Am. J. Probab. Math. Stat. **7** (2010), 243–256. MR-2672787

[20] I. Pinelis, *Convex cones of generalized multiply monotone functions and the dual cones*, ArXiv e-prints (2016), arXiv:1501.06599v2 [math.CA], to appear in Banach Journal of Mathematical Analysis.

[21] I. F. Pinelis and A. I. Sakhanenko, *Remarks on inequalities for probabilities of large deviations*, Theory Probab. Appl. **30** (1985), no. 1, 143–148.

[22] I. F. Pinelis and S. A. Utev, *Sharp exponential estimates for sums of independent random variables*, Theory Probab. Appl. **34** (1989), no. 2, 340–346. MR-1005745

[23] Iosif Pinelis, *Extremal probabilistic problems and Hotelling's T^2 test under a symmetry condition*, Ann. Statist. **22** (1994), no. 1, 357–368. MR-1272088

[24] Iosif Pinelis, *Optimum bounds for the distributions of martingales in Banach spaces*, Ann. Probab. **22** (1994), no. 4, 1679–1706. MR-1331198

[25] Iosif Pinelis, *Optimal tail comparison based on comparison of moments*, High dimensional probability (Oberwolfach, 1996), Progr. Probab., vol. 43, Birkhäuser, Basel, 1998, pp. 297–314. MR-1652335

[26] Iosif Pinelis, *Fractional sums and integrals of r-concave tails and applications to comparison probability inequalities*, Advances in stochastic inequalities (Atlanta, GA, 1997), Contemp. Math., vol. 234, Amer. Math. Soc., Providence, RI, 1999, pp. 149–168. MR-1694770

[27] Iosif Pinelis, *Exact inequalities for sums of asymmetric random variables, with applications*, Probab. Theory Related Fields **139** (2007), no. 3–4, 605–635. MR-2322709

[28] Iosif Pinelis, *On the Bennett-Hoeffding inequality*, `http://arxiv.org/abs/0902.4058`; a shorter version appeared in [31], 2009.

[29] Iosif Pinelis, *Positive-part moments via the Fourier–Laplace transform*, J. Theor. Probab. **24** (2011), 409–421.

[30] Iosif Pinelis, *Exact Rosenthal-type inequalities for $p = 3$, and related results*, Statistics & Probability Letters **83** (2013), no. 12, 2634–2637.

[31] Iosif Pinelis, *On the Bennett–Hoeffding inequality*, Annales de l'Institut Henri Poincaré, Probabilités et Statistiques **50** (2014), no. 1, 15–27.

[32] Iosif Pinelis, *Best possible bounds of the von Bahr–Esseen type*, Ann. Funct. Anal. **6** (2015), no. 4, 1–29. MR-3365979

[33] Iosif Pinelis, *Exact binomial, Poisson, and Gaussian bounds for the left tails of sums of nonnegative random variables*, Version 3, http://arxiv.org/abs/1503.06482, 2015.

[34] Iosif Pinelis, *Exact Rosenthal-type bounds*, Ann. Probab. **43** (2015), no. 5, 2511–2544. MR-3395468

[35] Julie Rehmeyer, *Voevodsky's Mathematical Revolution*, Scientific American (2013), no. October 1, http://blogs.scientificamerican.com/guest-blog/2013/10/01/voevodskys-mathematical-revolution/.

[36] Haskell P. Rosenthal, *On the subspaces of* L^p $(p > 2)$ *spanned by sequences of independent random variables*, Israel J. Math. **8** (1970), 273–303. MR-0271721

[37] L. R. Shenton, *Inequalities for the normal integral including a new continued fraction*, Biometrika **41** (1954), 177–189. MR-0061785

[38] Alfred Tarski, *A Decision Method for Elementary Algebra and Geometry*, RAND Corporation, Santa Monica, Calif., 1948. MR-0028796

[39] I. S. Tyurin, *Some optimal bounds in the central limit theorem using zero biasing*, Statist. Probab. Lett. **82** (2012), no. 3, 514–518. MR-2887466

Electronic Journal of Probability
Electronic Communications in Probability

Advantages of publishing in EJP-ECP

- Very high standards
- Free for authors, free for readers
- Quick publication (no backlog)
- Secure publication (LOCKSS[1])
- Easy interface (EJMS[2])

Economical model of EJP-ECP

- Non profit, sponsored by IMS[3], BS[4] , ProjectEuclid[5]
- Purely electronic

Help keep the journal free and vigorous

- Donate to the IMS open access fund[6] (click here to donate!)
- Submit your best articles to EJP-ECP
- Choose EJP-ECP over for-profit journals

[1]LOCKSS: Lots of Copies Keep Stuff Safe http://www.lockss.org/
[2]EJMS: Electronic Journal Management System http://www.vtex.lt/en/ejms.html
[3]IMS: Institute of Mathematical Statistics http://www.imstat.org/
[4]BS: Bernoulli Society http://www.bernoulli-society.org/
[5]Project Euclid: https://projecteuclid.org/
[6]IMS Open Access Fund: http://www.imstat.org/publications/open.htm