Journal of
Mathematical
Inequalities
Volume 2, Number 1 (2008), 1–7

ON INEQUALITIES FOR SUMS OF BOUNDED RANDOM VARIABLES

IOSIF PINELIS

(*communicated by G. Peskir*)

Abstract. A new upper bound on $\mathsf{P}(a_1\eta_1 + a_2\eta_2 + \ldots \geqslant x)$ is obtained, where $\eta_1, \eta_2, \ldots$ are independent zero-mean random variables such that $|\eta_i| \leqslant 1$ for all i. A multidimensional analogue of this result and extensions to (super)martingales are presented, as well as an application to self-normalized sums (or, equivalently, to t-statistics).

1. Introduction and summary

Let $\eta_1, \eta_2, \ldots$ be independent (not necessarily identically distributed) zero-mean random variables (r.v.'s) such that $|\eta_i| \leqslant 1$ almost surely (a.s.) for all i; the most important special case is when $\eta_i = \varepsilon_i$ for all i, where the ε_i's are independent Rademacher r.v.'s, so that $\mathsf{P}(\varepsilon_i = 1) = \mathsf{P}(\varepsilon_i = -1) = \frac{1}{2}$ for all i. Let

$$S := a_1\eta_1 + a_2\eta_2 + \ldots;$$

here and in what follows, it is assumed that $a_1, a_2, \ldots$ are any real numbers satisfying the normalization condition

$$a_1^2 + a_2^2 + \cdots = 1. \tag{1}$$

It follows from a result of Hoeffding [11] that

$$\mathsf{P}(S \geqslant x) \leqslant e^{-x^2/2} \quad \text{for all} \quad x \geqslant 0. \tag{2}$$

For $\eta_i = \varepsilon_i$, inequality (1) can be also deduced from the inequality $\mathsf{E}\,|S|^p \leqslant \mathsf{E}\,|Z|^p$ due to Whittle [29] for $p \geqslant 3$ and to Haagerup [10] for $p \geqslant 2$. The general case of any r.v.'s η_i as described above can then be obtained using a result by Hunt [12]; cf. [5, Lemma 2]. With an extra constant factor $e\sqrt{2}$, bound (1) for $\eta_i = \varepsilon_i$ was already given by Khinchin [13]; cf. [14, (1.11)].

Note that the upper bound $e^{-x^2/2}$ in (1) coincides for all $x \geqslant 0$ with the best upper exponential bound, $\inf_{t\geqslant 0} e^{-tx}\,\mathsf{E}\,e^{tZ}$, on the tail probability $\mathsf{P}(Z \geqslant x)$ for a

Mathematics subject classification (2000): 60E15, 60G50, 60G42, 60G48, 26A48, 26D10.

Keywords and phrases: Upper bounds, probability inequalities, bounded random variables, Rademacher random variables, sums of independent random variables, (super)martingales, self-normalized sums, t-statistics.

Paper JMI-02-01

standard normal r.v. Z. Thus, a factor of the order of magnitude of $\frac{1}{x}$ is “missing" in this bound, compared with the asymptotics $\mathsf{P}(Z \geqslant x) \sim \frac{1}{x}\,\varphi(x)$ as $x \to \infty$, where $\varphi(x) := e^{-x^2/2}/\sqrt{2\pi}$ is the density function of Z. (We write $a \sim b$ for $a/b \to 1$.) Now it should be clear that *any exponential* upper bound on the tail probabilities for sums of independent random variables must be missing the $\frac{1}{x}$ factor.

Eaton [5] proved that $\mathsf{P}(|S| \geqslant x) \leqslant 2\,\mathsf{E}(|Z| - t)_+^3/(x-t)^3$ for all x and t such that $0 < t < x$, from which he deduced another upper bound on $\mathsf{P}(|S| \geqslant x)$, which is asymptotic to $c_3\,\mathsf{P}(|Z| \geqslant x)$ as $x \to \infty$, where

$$c_3 := \frac{2e^3}{9} \approx 4.46,$$

and he conjectured that $\mathsf{P}(|S| \geqslant x) \leqslant 2c_3 \frac{1}{x}\,\varphi(x)$ for $x > \sqrt{2}$. The stronger form of this conjecture,

$$\mathsf{P}(S \geqslant x) \leqslant c\,\mathsf{P}(Z \geqslant x) \tag{3}$$

for all $x \in \mathbb{R}$ with $c = c_3$ was proved by Pinelis [16], along with a multidimensional extension. (More exactly, in [16] a two-tail version of inequality (3) was given. The right-tail inequality (3) can be proved quite similarly; alternatively, it follows from general results of [17].) Various generalizations and improvements of inequality (3) as well as related results were given by Pinelis [17, 18, 22, 24, 25, 27] and Bentkus [1, 2, 3].

For $\eta_i = \varepsilon_i$, Bobkov, Götze and Houdré (BGH) [4] gave a simple proof of (3) with a constant factor $c \approx 12.01$. Their method was based on a Chapman-Kolmogorov identity. Such an identity was used, e.g., in [19] concerning a conjecture by Graversen and Peškir [9]. In [28], it was shown that a modification of the BGH method can be used to show that for the least possible absolute constant factor c_* in inequality (3) for all $x \in \mathbb{R}$ (again, for $\eta_i = \varepsilon_i$) one has

$$c_* \in [c_0, c_1] \approx [3.18, 3.22], \quad \text{where}$$
$$c_0 := \frac{1}{4\,\mathsf{P}(Z \geqslant \sqrt{2})}, \quad c_1 := c_0 \cdot \Big(1 + \tfrac{1}{250}\big(1 + r(\sqrt{3}\,)\big)\Big) \approx c_0 \cdot 1.01,$$

and $r(x) := \mathsf{P}(Z \geqslant x)/\varphi(x)$ is the inverse Mills ratio.

On the other hand, also for $\eta_i = \varepsilon_i$, Edelman [7] proposed the interesting inequality

$$\mathsf{P}(S \geqslant x) \leqslant \mathsf{P}\,(Z \geqslant x - 1.5/x) \quad \text{for all} \quad x > 0. \tag{4}$$

Employing certain conditioning, Edelman [7] also offered applications of inequality (4) to statistical inference based on Student’s t statistic. Before that, the same conditioning idea (in relation with inequality (2) in place of (4)) was presented by Efron [8] and then by Eaton and Efron [6], in more general settings. The sketch of proof suggested in [7] for inequality (4) required an apparently nontrivial iterative computation procedure, which I have not been able to reproduce within a reasonable amount of computer time, because of rapid deterioration of precision at every step of the iterative procedure.

In this note, a simple proof of the following improvements of inequality (4) is presented.

THEOREM 1. *One has*

$$\mathsf{P}(S \geqslant x) \leqslant \mathsf{P}(Z \geqslant x - \lambda/x) \quad \textit{for all} \quad x > 0, \tag{5}$$

with

$$\lambda = \lambda_3 := \ln c_3 = 1.495\ldots.$$

Moreover, if $\eta_i = \varepsilon_i$ *for all* i*, then inequality* (5) *holds with the better constant*

$$\lambda = \lambda_1 := \ln c_1 = 1.168\ldots.$$

One application of Theorem 1 is to self-normalized sums

$$V := \frac{X_1 + \cdots + X_n}{\sqrt{X_1^2 + \cdots + X_n^2}}, \tag{6}$$

where, following Efron [8], we assume that the X_i 's satisfy the so-called 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 symmetrically distributed. It suffices that the X_i 's be independent and symmetrically (but not necessarily identically) distributed. In particular, $V = S$ if $X_i = a_i \varepsilon_i$ $\forall i$. It was noted by Efron that (i) Student's statistic T is a monotonic function of the self-normalized sum: $T = \sqrt{\frac{n-1}{n}}\, V/\sqrt{1 - V^2/n}$ and (ii) the orthant symmetry implies in general that the distribution of V is a mixture of the distributions of normalized Rademacher sums S. Thus, one obtains

COROLLARY 2. *Inequality* (5) *with* V *in place of* S *holds for* $\lambda = \lambda_1$.

REMARK 1. Another immediate corollary of Theorem 1 is the following two-tail counterpart of (5):

$$\mathsf{P}(|S| \geqslant x) \leqslant \mathsf{P}(|Z| \geqslant x - \lambda/x) \quad \text{for all} \quad x > 0, \tag{7}$$

with $\lambda = \lambda_3 = 1.495\ldots$ for the η_i 's in general, and with $\lambda = \lambda_1 = 1.168\ldots$ when $\eta_i = \varepsilon_i$ for all i.

Moreover, following the lines of proof in [17], [3], or [25], it is easy to see that the upper bounds in (5) and (7) with $\lambda = \lambda_3 = 1.495\ldots$ hold for S_n in place of S, $\forall n$, where (S_i) is a martingale with $S_0 = 0$ a.s. and differences $X_i := S_i - S_{i-1}$ ($i \geqslant 1$) such that $\sum_i \operatorname{ess\,sup} |X_i|^2 \leqslant 1$. Other extensions hold as well; look e.g. in [27] for appearances of the constant $c_{3,0} = 2e^3/9$ together with $\mathsf{P}(Z \geqslant \ldots)$ or $\mathsf{P}(|Z| \geqslant \ldots)$.

Furthermore, using the dimensionality reduction device given in [23], one immediately obtains a multi-dimensional generalization of (7):

$$\mathsf{P}\left(\|\eta_1 \mathbf{x}_1 + \eta_2 \mathbf{x}_2 + \ldots\| \geqslant x\right) \leqslant \mathsf{P}(|Z| \geqslant x - \lambda/x) \quad \forall x > 0,$$

where $\mathbf{x}_1, \mathbf{x}_2, \dots$ are any non-random vectors in a Hilbert space $(H, \|\cdot\|)$ such that $\|\mathbf{x}_1\|^2 + \|\mathbf{x}_2\|^2 + \cdots = 1$; here, again, $\lambda = \lambda_3 = 1.495\dots$ in general, and $\lambda = \lambda_1 = 1.168\dots$ when $\eta_i = \varepsilon_i$ for all i. (Cf. [27, Remark 1.4].)

2. Proofs

Theorem 1 follows immediately from Proposition 1 (below) and inequalities (2) and (3) (with $c = c_3$ in general and with $c = c_1$ when $\eta_i = \varepsilon_i$ for all i).

PROPOSITION 1. *For all $c \in [c_0, c_3]$ and all $x > 0$,*

$$\min\left(e^{-x^2/2}, c\,\mathsf{P}(Z \geqslant x)\right) \leqslant \mathsf{P}\left(Z \geqslant x - (\ln c)/x\right). \tag{8}$$

Note that the constant $\ln c$ is the best possible in inequality (8); indeed, using l'Hospital's rule for limits it is easy to see that $c\,\mathsf{P}(Z \geqslant x) \sim \mathsf{P}\left(Z \geqslant x - (\ln c)/x\right)$ as $x \to \infty$. Moreover, (8) will hold (for all $x > 0$) even for some (but not all) $c \in (0, c_0)$ and for some (but not all) $c \in (c_3, \infty)$. However, Proposition 1 as presented here will be sufficient for the purposes of this paper.

The proof of Proposition 1 is based on the following three lemmas.

LEMMA 1.
(i) *For each $\lambda \in \mathbb{R}$ there exists a unique point $x_*(\lambda) \in \mathbb{R}$ such that*

$$\min\left(e^{-x^2/2}, e^{\lambda}\,\mathsf{P}(Z \geqslant x)\right) = \begin{cases} e^{-x^2/2} & \text{if} \quad x \leqslant x_*(\lambda), \\ e^{\lambda}\,\mathsf{P}(Z \geqslant x) & \text{if} \quad x \geqslant x_*(\lambda). \end{cases}$$

(ii) *For each $x \in \mathbb{R}$, let $\lambda_*(x) := \ln\left(e^{-x^2/2}/\,\mathsf{P}(Z \geqslant x)\right)$, so that $\lambda = \lambda_*(x)$ is the only solution of equation $e^{-x^2/2} = e^{\lambda}\,\mathsf{P}(Z \geqslant x)$. Then $x = x_*(\lambda) \iff \lambda = \lambda_*(x)$, for all real λ and x. Moreover, the function λ_* is increasing and concave.*

LEMMA 2. *For any $\lambda > 0$ and $b > 0$, one has $\inf_{x \in (0,b]} r_1(x) = \min\left(1, r_1(b)\right)$, where $r_1(x) := e^{x^2/2}\,\mathsf{P}(Z \geqslant x - \lambda/x)$.*

LEMMA 3. *For any $\lambda > 0$ and $b > 0$, one has $\inf_{x \in [b,\infty)} r_2(x) = \min\left(1, r_2(b)\right)$, where $r_2(x) := \mathsf{P}(Z \geqslant x - \lambda/x)/\left(e^{\lambda}\,\mathsf{P}(Z \geqslant x)\right)$.*

The proofs of these lemmas will be given at the end of the paper. They (as well as the proof of Proposition 1) make use of the following l'Hospital-type rules for monotonicity:

PROPOSITION 2. *Let $-\infty \leqslant a < b \leqslant \infty$. Let f and g be real-valued differentiable functions, defined on the interval (a, b), such that g and g' do not take on the zero value on (a, b). Let*

$$r := f/g \quad \textit{and} \quad \rho := f'/g'.$$

(i) [26, Corollary 3.1] *If ρ is monotonic on (a,b), then r may switch at most once on (a,b) – either from increase to decrease or vice versa.*

(ii) [26, Proposition 4.3] *At this point suppose in addition that $f(b-) = g(b-) = 0$; suppose also that $\rho \nearrow\searrow$ on (a,b) – that is, for some $c \in [a,b]$, $\rho \nearrow$ (ρ is increasing) on (a,c) and $\rho \searrow$ on (c,b). Then $r \nearrow\searrow$ on (a,b). Moreover, if $\rho \nearrow$ on (a,b) then r is so. (Note that the symbol $\nearrow\searrow$ has a slightly different meaning in [26]: there c cannot equal a or b.)*

Alternatively, instead of Proposition 2, one can use results of [20].

Proof of Proposition 1. By part (ii) of Lemma 1, $\lambda_*(x)$ increases from $\lambda_0 := \ln c_0 = 1.156\dots$ to $\lambda_3 = \ln c_3 = 1.495\dots$ as x increases from $x_0 := x_*(\lambda_0) = 0.670\dots$ to $x_3 := x_*(\lambda_3) = 1.312\dots$; moreover, by the concavity of the function λ_*,

$$\lambda_*(x) \geqslant \lambda_{**}(x) := \lambda_0 + \frac{\lambda_3 - \lambda_0}{x_3 - x_0}(x - x_0) \quad \text{for all } x \in [x_0, x_3]. \tag{9}$$

Consider now the ratios $r_3 := f/g$ and $\rho := f'/g'$, where $f(x) := \mathsf{P}(Z \geqslant x - \lambda_{**}(x)/x)$ and $g(x) := e^{-x^2/2}$. Then ρ has the form $R_1 e^{R_2}$, where R_1 and R_2 are certain rational functions; hence, ρ' has the same form, whence it is straightforward to see that $\rho' > 0$ and hence $\rho \nearrow$ on $(x_0, x_{3/2})$, and $\rho' < 0$ and hence $\rho \searrow$ on $(x_{3/2}, x_3)$, where $x_{3/2} = 0.718\dots$ is the only root of ρ' in the interval (x_0, x_3). So, by part (i) of Proposition 2, on each of the intervals $(x_0, x_{3/2})$ and $(x_{3/2}, x_3)$, the ratio r_3 may switch at most once from increase to decrease or vice versa. However, $r_3'(x_0) = -0.080\dots < 0$, $r_3'(x_{3/2}) = -0.093\dots < 0$, and $r_3'(x_3) = -0.023\dots < 0$. Therefore, $r_3 \searrow$ on each of the intervals $(x_0, x_{3/2})$ and $(x_{3/2}, x_3)$, and hence $r_3 \searrow$ on the entire interval (x_0, x_3). Since $r_3(x_3) = 1.020\dots > 1$, one has $r_3 > 1$ on the interval $[x_0, x_3]$. Hence, in view of (9), $\mathsf{P}(Z \geqslant x - \lambda_*(x)/x) > e^{-x^2/2}$ for all $x \in [x_0, x_3]$; that is, $r_1(x_*(\lambda)) > 1$ for all $\lambda \in [\lambda_0, \lambda_3]$. It remains to refer to part (i) of Lemma 1 and Lemmas 2 and 3. □

Proof of Lemma 1. Apparently, this lemma is essentially known. A quick way to prove it is to observe that the Mills ratio $\sqrt{2\pi}e^{x^2/2}\,\mathsf{P}(Z \geqslant x)$ is identical to $\int_0^\infty e^{-xu-u^2/2}\,du$, and hence decreasing (to 0) and log-convex (in $x \in \mathbb{R}$). Alternatively, one can use here the mentioned l'Hospital-type rule for monotonicity. (Cf. [21, 22].) □

Proof of Lemma 2. Write $r_1 = f/g$ and consider the ratio $\rho := f'/g'$, where $f(x) := \mathsf{P}(Z \geqslant x - \lambda/x)$ and $g(x) := e^{-x^2/2}$. Then

$$\rho(x) = \frac{\lambda + x^2}{\sqrt{2\pi}\,x^3\,e^{\frac{\lambda^2}{2x^2} - \lambda}} \quad \text{and} \quad \rho'(x) = \frac{\lambda^3 - (3-\lambda)\lambda x^2 - x^4}{\sqrt{2\pi}\,x^6\,e^{\frac{\lambda^2}{2x^2} - \lambda}},$$

so that ρ' changes sign from $+$ to $-$ on $(0,\infty)$ and hence $\rho \nearrow\searrow$ on $(0,\infty)$. So, by Proposition 2, $r_1 \nearrow\searrow$ on $(0,\infty)$. Also, $r_1(0+) = 1$. Now Lemma 2 follows. □

Proof of Lemma 3. Write $r_2 = f/g$ and consider the ratio $\rho := f'/g'$, where $f(x) := \mathsf{P}(Z \geqslant x - \lambda/x)$ and $g(x) := e^{\lambda}\,\mathsf{P}(Z \geqslant x)$ for $x > 0$. Then

$$\rho(x) = e^{-\frac{\lambda^2}{2x^2}}(1 + \tfrac{\lambda}{x^2}) \quad \text{and} \quad \rho'(x) = \big(\lambda^2 - (2-\lambda)x^2\big)\,\lambda\, x^{-5}\, e^{-\frac{\lambda^2}{2x^2}},$$

so that $\rho \nearrow\searrow$ on $(0,\infty)$. By Proposition 2, one now has $r_2 \nearrow\searrow$ on $(0,\infty)$. Also, by l'Hospital's rule for limits, $r_2(\infty-) = \rho(\infty-) = 1$. Now Lemma 3 follows. □

REFERENCES

[1] V. BENTKUS, *A remark on the inequalities of Bernstein, Prokhorov, Bennett, Hoeffding, and Talagrand*, Lithuanian Math. J., **42**, (2002), 262–269.

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

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

[4] S. G. BOBKOV, F. GÖTZE, C. HOUDRÉ, *On Gaussian and Bernoulli covariance representations*, Bernoulli, **7**, (2002), 439–451.

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

[6] M. L. EATON, B. EFRON, *Hotelling's T^2 test under symmetry conditions*, J. Amer. Statist. Assoc., **65**, (1970), 702–711.

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

[8] B. EFRON, *Student's t test under symmetry conditions*, J. Amer. Statist. Assoc., **64**, (1969), 1278–1302.

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

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

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

[12] G. A. HUNT, *An inequality in probability theory*, Proc. Amer. Math. Soc., **6**, (1955), 506–510.

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

[14] G. PEŠKIR, A. N. SHIRYAEV, *The Inequalities of Khintchine and Expanding Sphere of Their Action*, Russian Math. Surveys, **50**, (1995), 849–904.

[15] I. PINELIS, *Extremal probabilistic problems and Hotelling's T^2 test under symmetry condition*, (1991), Preprint, *URL:* `http://arxiv.org/abs/math/0701806`.

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

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

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

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

[20] I. PINELIS, *L'Hospital type rules for oscillation, with applications*, J. Inequal. Pure Appl. Math., **2**, (2001), no. 3, Article 33, 24 pp. (electronic).

[21] I. PINELIS, *Monotonicity properties of the relative error of a Padé approximation for Mills' ratio*, J. Inequal. Pure Appl. Math., **3**, (2002), no. 2, Art. 20, 8 pp. (electronic).

[22] I. PINELIS, *L'Hospital type rules for monotonicity: applications to probability inequalities for sums of bounded random variables*, J. Inequal. Pure Appl. Math., **3**, (2002), no. 1, Article 7, 9 pp. (electronic).

[23] I. PINELIS, *Dimensionality reduction in extremal problems for moments of linear combinations of vectors with random coefficients*, in Stochastic inequalities and applications, Progr. Probab., Birkhäuser, Basel, **56**, (2003), 169–185.

[24] I. PINELIS, *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, Institute of Mathematical Statistics, **51**, (2006). DOI: 10.1214/074921706000000743. *URL:* http://arxiv.org/abs/math.PR/0512301.

[25] I. PINELIS, *On normal domination of (super)martingales*, Electronic J. Probab., **11**, (2006), Paper 39, 1049–1070. *URL:* http://www.math.washington.edu/~ejpecp/viewarticle.php?id=1648&layout=abstract.

[26] I. PINELIS, *On l'Hospital-type rules for monotonicity*, J. Inequal. Pure Appl. Math., **7**, (2006), no. 2, Article 40 (electronic).

[27] I. PINELIS, (2006). *Exact inequalities for sums of asymmetric random variables, with applications*, Probab. Theory Related Fields **139**, (2007), no. 3-4, 605–635.

[28] I. PINELIS, (2007). *Toward the best constant factor for the Rademacher-Gaussian tail comparison*, ESAIM Probab. Stat., **11**, (2007), 412–426 (electronic).

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

(Received June 26, 2006)

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