Bernoulli **21**(2), 2015, 1231–1237
DOI: 10.3150/14-BEJ603

A tight Gaussian bound for weighted sums of Rademacher random variables

VIDMANTAS KASTYTIS BENTKUS and DAINIUS DZINDZALIETA[1]

[1] *Vilnius University Institute of Mathematics and Informatics, Akademijos 4, Vilnius, Lithuania.*
E-mail: dainiusda@gmail.com

Let $\varepsilon_1, \dots, \varepsilon_n$ be independent identically distributed Rademacher random variables, that is $\mathbb{P}\{\varepsilon_i = \pm 1\} = 1/2$. Let $S_n = a_1\varepsilon_1 + \cdots + a_n\varepsilon_n$, where $\mathbf{a} = (a_1, \dots, a_n) \in \mathbb{R}^n$ is a vector such that $a_1^2 + \cdots + a_n^2 \le 1$. We find the smallest possible constant c in the inequality

$$\mathbb{P}\{S_n \ge x\} \le c\mathbb{P}\{\eta \ge x\} \qquad \text{for all } x \in \mathbb{R},$$

where $\eta \sim N(0,1)$ is a standard normal random variable. This optimal value is equal to

$$c_* = \bigl(4\mathbb{P}\{\eta \ge \sqrt{2}\}\bigr)^{-1} \approx 3.178.$$

Keywords: bounds for tail probabilities; Gaussian; large deviations; optimal constants; random sign; self-normalized sums; Student's statistic; symmetric; tail comparison; weighted Rademachers

1. Introduction

Let $\varepsilon_1, \dots, \varepsilon_n$ be independent identically distributed Rademacher random variables, such that $\mathbb{P}\{\varepsilon_i = \pm 1\} = 1/2$. Let $S_n = a_1\varepsilon_1 + \cdots + a_n\varepsilon_n$, where $\mathbf{a} = (a_1, \dots, a_n) \in \mathbb{R}^n$ is a vector such that $a_1^2 + \cdots + a_n^2 \le 1$.

The main result of the paper is the following theorem.

Theorem 1.1. *Let $\eta \sim N(0,1)$ be a standard normal random variable. Then, for all $x \in \mathbb{R}$,*

$$\mathbb{P}\{S_n \ge x\} \le c\mathbb{P}\{\eta \ge x\}, \tag{1.1}$$

with the constant c equal to

$$c_* := \bigl(4\mathbb{P}\{\eta \ge \sqrt{2}\}\bigr)^{-1} \approx 3.178.$$

The value $c = c_*$ is the best possible since (1.1) becomes equality if $n \ge 2$, $x = \sqrt{2}$ and $S_n = (\varepsilon_1 + \varepsilon_2)/\sqrt{2}$.

Inequality (1.1) was first obtained by Pinelis [4] with $c \approx 4.46$. Bobkov, Götze and Houdré (BGH) [2] gave a simple proof of (1.1) with constant factor $c \approx 12.01$. Their method was to use induction on n together with the inequality

$$\tfrac{1}{2}\mathbb{P}\{\eta \ge A\} + \tfrac{1}{2}\mathbb{P}\{\eta \ge B\} \le \mathbb{P}\{\eta \ge x\} \qquad \text{for all } x \ge \sqrt{3} \text{ and } \tau \in [0,1], \tag{1.2}$$

where $A := \frac{x-\tau}{\sqrt{1-\tau^2}}$ and $B := \frac{x+\tau}{\sqrt{1-\tau^2}}$. Using a method similar to the one in BGH [2], Bentkus [1] proved (1.1) with $c \approx 4.00$ and conjectured that the optimal constant in (1.1) is c_*. Further progress was achieved by Pinelis [5], where (1.1) was proved with $c \approx 1.01c_*$.

Let us briefly outline our strategy of the proof. For $x \le \sqrt{2}$ Theorem 1.1 follows from the symmetry of S_n. For $x \ge \sqrt{2}$ we consider two cases separately. If $x \in (\sqrt{2}, \sqrt{3})$ and all a_i's are "small" we use Berry–Esseen inequality. Otherwise we use induction on n together with Chebyshev type inequality presented in Lemma 2.1. We remark that the analysis of weighted sums of random variables based on separate study of these two cases has proved recently to be effective idea, see [7]. In [6], this idea was used to obtain asymptotically Gaussian bound

$$\mathbb{P}\{S_n \ge x\} \le \mathbb{P}\{\eta \ge x\}\left(1 + \frac{C}{x}\right),$$

where $C \approx 14.10\ldots$.

A standard application of bounds like (1.1), following Efron [3], is to the Student's statistic and to self-normalized sums. For example, if random variables $X_1, \ldots, X_n$ are independent (not necessary identically distributed), symmetric and not all identically equal to zero, then the statistic

$$T_n = (X_1 + \cdots + X_n)/\sqrt{X_1^2 + \cdots + X_n^2}$$

is sub-Gaussian and

$$\mathbb{P}\{T_n \ge x\} \le c_* \mathbb{P}\{\eta \ge x\} \qquad \text{for all } x \in \mathbb{R}. \tag{1.3}$$

The latter inequality is optimal since it turns into an equality if $n = 2$, $x = \sqrt{2}$ and $X_1 = \varepsilon_1$, $X_2 = \varepsilon_2$. This inequality was previously obtained in [4,5] with constants 4.46 and $\approx 1.01c_*$ in place of c_*.

2. Proofs

In this section, we use the following notation

$$\tau = a_1, \qquad \vartheta = \sqrt{1-\tau^2}, \qquad I(x) = \mathbb{P}\{\eta \ge x\}, \qquad \varphi(x) = -I'(x), \tag{2.1}$$

that is, $I(x)$ is the tail probability for standard normal random variable η and $\varphi(x)$ is the standard normal density. Without loss of generality, we assume that $a_1^2 + \cdots + a_n^2 = 1$ and $a_1 \ge \cdots \ge a_n \ge 0$. Using (2.1) we have $S_n = \tau\varepsilon_1 + \vartheta X$ with $X = (a_2\varepsilon_2 + \cdots + a_n\varepsilon_n)/\vartheta$. The random variable X is symmetric and independent of ε_1. It is easy to check that $\mathbb{E}X^2 = 1$ and

$$\mathbb{P}\{S_n \ge x\} = \tfrac{1}{2}\mathbb{P}\{X \ge A\} + \tfrac{1}{2}\mathbb{P}\{X \ge B\}, \tag{2.2}$$

where $A = \frac{x-\tau}{\vartheta}$ and $B = \frac{x+\tau}{\vartheta}$.

We start with a simple Chebyshev type inequality.

Lemma 2.1. *Let $s > 0$ and $0 \le a \le b$. Then, for any random variable Y we have*

$$a^s \mathbb{P}\big\{|Y| \ge a\big\} + \big(b^s - a^s\big)\mathbb{P}\big\{|Y| \ge b\big\} \le \mathbb{E}|Y|^s. \tag{2.3}$$

If Y is symmetric, then

$$a^s \mathbb{P}\{Y \ge a\} + \big(b^s - a^s\big)\mathbb{P}\{Y \ge b\} \le \mathbb{E}|Y|^s/2. \tag{2.4}$$

Proof. It is clear that (2.3) implies (2.4). To prove (2.3), we use the obvious inequality

$$a^s \mathbb{I}\big\{|Y| \ge a\big\} + \big(b^s - a^s\big)\mathbb{I}\big\{|Y| \ge b\big\} \le |Y|^s, \tag{2.5}$$

where $\mathbb{I}\{E\}$ stands for the indicator function of the event E. Taking expectation, we get (2.3). □

In similarity to (2.3), one can derive a number of inequalities stronger than the standard Chebyshev inequality $\mathbb{P}\{S_n \ge x\} \le 1/(2x^2)$. For example, instead of $\mathbb{P}\{S_n \ge 1\} \le 1/2$ we have the much stronger

$$\mathbb{P}\{S_n \ge 1\} + \mathbb{P}\{S_n \ge \sqrt{2}\} + \mathbb{P}\{S_n \ge \sqrt{3}\} + \cdots \le 1/2.$$

We will make use of Lyapunov type bounds with explicit constants for the remainder term in the Central limit theorem. Let $X_1, X_2, \ldots$ be independent random variables such that $\mathbb{E}X_j = 0$ for all j. Set $\beta_j = \mathbb{E}|X_j|^3$. Assume that the sum $Z = X_1 + X_2 + \cdots$ has unit variance. Then there exists an absolute constant, say c_{L}, such that

$$\big|\mathbb{P}\{Z \ge x\} - I(x)\big| \le c_{\mathrm{L}}(\beta_1 + \beta_2 + \cdots). \tag{2.6}$$

It is known that $c_{\mathrm{L}} \le 0.56\ldots$ [8,9]. Note that we actually do not need the best known bound for c_{L}. Even $c_{\mathrm{L}} = 0.958$ suffices to prove Theorem 1.1.

Replacing X_j by $a_j\varepsilon_j$ and using $\beta_j \le \tau a_j^2$ for all j, the inequality (2.6) implies

$$\big|\mathbb{P}\{S_n \ge x\} - I(x)\big| \le c_{\mathrm{L}}\tau. \tag{2.7}$$

Proof of Theorem 1.1. For $x \le \sqrt{2}$ Theorem 1.1 follows from the symmetry of S_n and Chebyshev's inequality (first it was implicitly shown in [1], later in [5]). In the case $x \ge \sqrt{2}$, we argue by induction on n. However, let us first provide a proof of Theorem 1.1 in some special cases where induction fails.

Using the bound (2.7), let us prove Theorem 1.1 under the assumption that

$$\tau \le \tau_{\mathrm{L}} \stackrel{\mathrm{def}}{=} (c_* - 1)I(\sqrt{3})/c_{\mathrm{L}} \quad \text{and} \quad x \le \sqrt{3}. \tag{2.8}$$

Using $c_{\mathrm{L}} = 0.56$, the numerical value of τ_{L} is $0.16\ldots$. In order to prove Theorem 1.1 under the assumption (2.8), note that the inequality (2.7) yields

$$\mathbb{P}\{S_n \ge x\} \le I(x) + \tau c_{\mathrm{L}}. \tag{2.9}$$

If the inequality (2.8) holds, the right-hand side of (2.9) is clearly bounded from above by $c_* I(x)$ for $x \le \sqrt{3}$.

For x and τ such that (2.8) does not hold we use induction on n. If $n = 1$, then we have $S_n = \varepsilon_1$ and Theorem 1.1 is equivalent to the trivial inequality $1/2 \le c_* I(1)$.

Let us assume that Theorem 1.1 holds for $n \le k-1$ and prove it for $n = k$.

Firstly we consider the case $x \ge \sqrt{3}$. We replace S_n by S_k with $X = (a_2\varepsilon_2 + \cdots + a_k\varepsilon_k)/\vartheta$ in (2.2). We can estimate the latter two probabilities in (2.2) applying the induction hypothesis $\mathbb{P}\{X \ge y\} \le c_* I(y)$. We get

$$\mathbb{P}\{S_k \ge x\} \le c_* I(A)/2 + c_* I(B)/2. \tag{2.10}$$

In order to conclude the proof, it suffices to show that the right-hand side of (2.10) is bounded from above by $c_* I(x)$, that is, that the inequality $I(A) + I(B) \le 2I(x)$ holds. As $x \ge \sqrt{3}$ this follows by the inequality (1.2).

In the remaining part of the proof, we can assume that $x \in (\sqrt{2}, \sqrt{3})$ and $\tau \ge \tau_{\mathrm{L}}$. In this case in order to prove Theorem 1.1, we have to improve the arguments used to estimate the right-hand side of (2.2). This is achieved by applying the Chebyshev type inequalities of Lemma 2.1. By Lemma 2.1, for any symmetric X such that $\mathbb{E}X^2 = 1$, and $0 \le A \le B$, we have

$$A^2\mathbb{P}\{X \ge A\} + \big(B^2 - A^2\big)\mathbb{P}\{X \ge B\} \le 1/2. \tag{2.11}$$

By (2.1), we can rewrite (2.11) as

$$(x-\tau)^2\mathbb{P}\{X \ge A\} + 4x\tau\mathbb{P}\{X \ge B\} \le \vartheta^2/2. \tag{2.12}$$

For $x \in (\sqrt{2}, \sqrt{3})$ and $\tau \ge \tau_{\mathrm{L}}$ we consider the cases

$$\text{(i) } (x-\tau)^2 \ge 4x\tau \quad \text{and} \quad \text{(ii) } (x-\tau)^2 \le 4x\tau$$

separately. We denote the sets of points (x, τ) such that $x \in (\sqrt{2}, \sqrt{3})$, $\tau \ge \tau_{\mathrm{L}}$ and (i) or (ii) holds by E_1 and E_2, respectively.

(i) Using (2.2), (2.12) and the induction hypothesis we get

$$\mathbb{P}\{S_k \ge x\} \le \frac{D\mathbb{P}\{X \ge B\} + \vartheta^2/2}{2(x-\tau)^2} \le \frac{c_* D I(B) + \vartheta^2/2}{2(x-\tau)^2}, \tag{2.13}$$

where $X = (a_2\varepsilon_2 + \cdots + a_k\varepsilon_k)/\vartheta$ and $D = (x-\tau)^2 - 4x\tau$.

In order to finish the proof of Theorem 1.1 (in this case) it suffices to show that the right-hand side of (2.13) is bounded above by $c_* I(x)$. In other words, we have to check that the function

$$f(x,\tau) \equiv f \stackrel{\text{def}}{=} \big((x-\tau)^2 - 4x\tau\big)c_* I(B) - 2c_*(x-\tau)^2 I(x) + \vartheta^2/2, \tag{2.14}$$

is negative on E_1, where $B = (x+\tau)/\vartheta$.

By Lemma 2.5 below, we have

$$f(x,\tau) \le f(\sqrt{3}, \tau) =: g(\tau). \tag{2.15}$$

Since $\tau \le (3 - 2\sqrt{2})x$ the inequality $f \le 0$ on E_1 follows from Lemma 2.2, below.

(ii) Using (2.2), (2.12) and induction hypothesis we get

$$\mathbb{P}\{S_k \geq x\} \leq \frac{C\mathbb{P}\{X \geq A\} + \vartheta^2/2}{8x\tau} \leq \frac{C/(2A^2) + \vartheta^2/2}{8x\tau}, \tag{2.16}$$

where $X = (a_2\varepsilon_2 + \cdots + a_k\varepsilon_k)/\vartheta$ and $C = 4x\tau - (x-\tau)^2$.

In order to finish the proof (in this case) it suffices to show that the right-hand side of (2.16) is bounded above by $c_* I(x)$. In other words, we have to check that

$$C/\big(2A^2\big) + \vartheta^2/2 \leq 8x\tau c_* I(x) \qquad \text{on } E_2. \tag{2.17}$$

Recalling that $C = 4x\tau - (x-\tau)^2$, $A = (x-\tau)/\vartheta$, inequality (2.17) is equivalent to

$$h \stackrel{\text{def}}{=} \frac{1-\tau^2}{(x-\tau)^2} - 4c_* I(x) \leq 0 \qquad \text{on } E_2. \tag{2.18}$$

Inequality (2.18) follows from Lemma 2.6, below. The proof of Theorem 1.1 is complete. □

Lemma 2.2. *The function g defined by* (2.15) *is negative for all $\tau \in [\tau_{\mathrm{L}}, (3-2\sqrt{2})\sqrt{3}]$.*

Lemma 2.3. $I'(B) \geq \vartheta I'(x)$ *on E_1.*

Lemma 2.4. $I(B) \geq I(x) + I'(x)\tau$ *on E_1.*

Lemma 2.5. *The partial derivative $\partial_x f$ of the function f defined by* (2.14) *is positive on E_1.*

Lemma 2.6. *The function h defined by* (2.18) *is negative on E_2.*

Proof of Lemma 2.2. Since $g(\tau_{\mathrm{L}}) < 0$ it is sufficient to show that g is a decreasing function for $\tau_{\mathrm{L}} \leq \tau \leq (3-2\sqrt{2})\sqrt{3}$. Note that

$$g(\tau) = \big((\sqrt{3}-\tau)^2 - 4\sqrt{3}\tau\big)c_* I(B) + \big(1-\tau^2\big)/2 - 2c_*(\sqrt{3}-\tau)^2 I(\sqrt{3})$$

and

$$\begin{aligned} g'(\tau) = {} & (2\tau - 6\sqrt{3})c_* I(B) - \big((\sqrt{3}-\tau)^2 - 4\sqrt{3}\tau\big)c_*\varphi(B)(1+\tau\sqrt{3})\vartheta^{-3} \\ & - \tau + 4c_*(\sqrt{3}-\tau)I(\sqrt{3}), \end{aligned}$$

where φ is the standard normal distribution. Hence

$$g'(\tau) \leq w(\tau) \stackrel{\text{def}}{=} (2\tau - 6\sqrt{3})c_* I(B) - \tau + 4c_*(\sqrt{3}-\tau)I(\sqrt{3}).$$

Note that the value of B in previous three displayed formulas should also be computed with $x = \sqrt{3}$. Using Lemma 2.4, we get

$$g'(\tau) \leq -2c_*(\sqrt{3}+\tau)I(\sqrt{3}) + 2c_*\tau(3\sqrt{3}-\tau)\varphi(\sqrt{3}) - \tau \stackrel{\text{def}}{=} Q(\tau)$$

with

$$Q(\tau)=-\alpha\tau^2+\beta\tau-\gamma,\qquad \alpha=0.56\ldots,\qquad \beta=1.67\ldots,\qquad \gamma=0.45\ldots.$$

Clearly, Q is negative on the interval $[\tau_{\mathrm{L}},(3-2\sqrt{2})\sqrt{3}]$. It follows that g' is negative, and g is decreasing on $[\tau_{\mathrm{L}},(3-2\sqrt{2})\sqrt{3}]$. □

Proof of Lemma 2.3. Since $I'=-\varphi$ by (2.1), the inequality $I'(B)\ge\vartheta I'(x)$ is equivalent to

$$u(\tau)\overset{\text{def}}{=}\bigl(1-\tau^2\bigr)\exp\biggl\{\frac{(x+\tau)^2}{1-\tau^2}-x^2\biggr\}-1\ge 0.$$

Since $u(0)=0$, it suffices to check that $u'\ge 0$. Elementary calculations show that $u'\ge 0$ is equivalent to the trivial inequality $x+\tau^2x+\tau x^2+\tau^3\ge 0$. □

Proof of Lemma 2.4. Let $g(\tau)=I(B)$. Then the inequality $I(B)\ge I(x)+I'(x)\tau$ turns into $g(\tau)\ge g(0)+g'(0)\tau$. The latter inequality holds provided that $g''(\tau)\ge 0$. Next, it is easy to see that $g'(\tau)=-\varphi(B)B'$ and $g''(\tau)=(BB'^2-B'')\varphi(B)$. Hence, to verify that $g''(\tau)\ge 0$ we verify that $BB'^2-B''\ge 0$. This last inequality is equivalent to $-2+2x^2+x^3\tau+x^2\tau^2+x\tau+2x\tau^3+3\tau^2\ge 0$, which holds since $x\ge 1$. The proof of Lemma 2.4 is complete. □

Proof of Lemma 2.5. We have

$$\partial_x f=2(x-3\tau)c_*I(B)+Dc_*I'(B)/\vartheta-4c_*(x-\tau)I(x)-2c_*(x-\tau)^2I'(x).$$

We have to show that $\partial_x f\ge 0$ on E_1. Using Lemma 2.3, we can reduce this to the inequality

$$2(x-3\tau)I(B)-(x+\tau)^2I'(x)-4(x-\tau)I(x)\ge 0. \tag{2.19}$$

On E_1 we have that $0\le\tau\le(3-2\sqrt{2})x$, so $x-3\tau\ge x-3(3-2\sqrt{2})x=(6\sqrt{2}-8)x>0$. By Lemma 2.4 we have that left-hand side of (2.19) is bigger than

$$\begin{aligned}&2(x-3\tau)\bigl(I(x)+I'(x)\tau\bigr)-(x+\tau)I'(x)-4(x-\tau)I(x)\\&\quad=-2(x+\tau)I(x)-\bigl(x^2+7\tau^2\bigr)I'(x).\end{aligned}$$

Inequality (2.19) follows by the inequality $-(x^2+7\tau^2)I'(x)\ge\alpha x(x+\tau)\varphi(x)>2(x+\tau)I(x)$ on E_1 with $\alpha=4\sqrt{14}-14$, where the second inequality follows from the fact that $\varphi(x)x/I(x)$ increases for $x>0$ and is larger than $2/\alpha$ for $x=\sqrt{2}$. The proof of Lemma 2.5 is complete. □

Proof of Lemma 2.6. It is easy to see that the function h attains its maximal value at $\tau=1/x$. Hence, it suffices to check (2.18) with $\tau=1/x$, that is, that for $\sqrt{2}\le x\le\sqrt{3}$ the inequality $g(x)\overset{\text{def}}{=}1-4c_*(x^2-1)I(x)\le 0$ holds. Using $4c_*I(\sqrt{2})=1$, we have $g(\sqrt{2})=0$ and $g(\sqrt{3})<0$. Next, $g'(x)=-8c_*xI(x)+4c_*(x^2-1)\varphi(x)$, so $g'(\sqrt{2})<0$ and $g'(\sqrt{3})>0$. We have that $g''(x)=4c_*((5-x^2)x\varphi(x)-2I(x))$. Since $I(x)\le\varphi(x)/x$ we have that $g''(x)\ge 4c_*((5-x^2)x\varphi(x)-2\varphi(x)/x)=4c_*\varphi(x)/x((5-x^2)x^2-2)\ge 8c_*\varphi(x)/x>0$ for $x\in(\sqrt{2},\sqrt{3})$. The proof of Lemma 2.6 is complete. □

Acknowledgements

This research was funded by a grant (No. MIP-47/2010) from the Research Council of Lithuania.

References

[1] Bentkus, V. (2007). On measure concentration for separately Lipschitz functions in product spaces. *Israel J. Math*. **158** 1–17. MR2342455

[2] Bobkov, S.G., Götze, F. and Houdré, C. (2001). On Gaussian and Bernoulli covariance representations. *Bernoulli* **7** 439–451. MR1836739

[3] Efron, B. (1969). Student's t-test under symmetry conditions. *J. Amer. Statist. Assoc*. **64** 1278–1302. MR0251826

[4] Pinelis, I. (1994). Extremal probabilistic problems and Hotelling's T^2 test under a symmetry condition. *Ann. Statist*. **22** 357–368. MR1272088

[5] Pinelis, I. (2007). Toward the best constant factor for the Rademacher–Gaussian tail comparison. *ESAIM Probab. Stat*. **11** 412–426. MR2339301

[6] Pinelis, I. (2012). An asymptotically Gaussian bound on the Rademacher tails. *Electron. J. Probab*. **17** no. 35, 22. MR2924368

[7] Rudelson, M. and Vershynin, R. (2009). Smallest singular value of a random rectangular matrix. *Comm. Pure Appl. Math*. **62** 1707–1739. MR2569075

[8] Shevtsova, I.G. (2010). Refinement of estimates for the rate of convergence in Lyapunov's theorem. *Dokl. Akad. Nauk* **435** 26–28. MR2790498

[9] Tyurin, I.S. (2011). Improvement of the remainder in the Lyapunov theorem. *Teor. Veroyatn. Primen*. **56** 808–811. MR3137072

Received February 2013 and revised January 2014