Electronic Journal of Probability

Electron. J. Probab. **17** (2012), no. 35, 1–22.
ISSN: 1083-6489 DOI: 10.1214/EJP.v17-2026

An asymptotically Gaussian bound on the Rademacher tails*

Iosif Pinelis†

Abstract

An explicit upper bound on the tail probabilities for the normalized Rademacher sums is given. This bound, which is best possible in a certain sense, is asymptotically equivalent to the corresponding tail probability of the standard normal distribution, thus affirming a longstanding conjecture by Efron. Applications to sums of general centered uniformly bounded independent random variables and to the Student test are presented.

Keywords: probability inequalities; large deviations; Rademacher random variables; sums of independent random variables; Student's test; self-normalized sums; Esscher–Cramér tilt transform; generalized moments; Tchebycheff–Markov systems.
AMS MSC 2010: Primary 60E15, Secondary 60F10; 62G10; 62G15; 60G50; 62G35.
Submitted to EJP on September 30, 2011, final version accepted on May 15, 2012.

1 Introduction, summary, and discussion

Let $\varepsilon_1,\dots,\varepsilon_n$ be independent Rademacher random variables (r.v.'s), so that $\mathsf{P}(\varepsilon_i = 1) = \mathsf{P}(\varepsilon_i = -1) = \frac{1}{2}$ for all i. Let $a_1,\dots,a_n$ be any real numbers such that

$$a_1^2 + \cdots + a_n^2 = 1. \tag{1.1}$$

Let

$$S_n := a_1\varepsilon_1 + \cdots + a_n\varepsilon_n$$

be the corresponding normalized Rademacher sum. Let Z denote a standard normal r.v., with the density function φ, so that $\varphi(x) = \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}$ for all real x.

Upper bounds on the tail probabilities $\mathsf{P}(S_n \geqslant x)$ have been of interest in combinatorics/optimization/operations research; see e.g. [32, 2, 16, 17, 3, 26] and bibliography therein. Other authors, including Bennett [4], Hoeffding [30], and Efron [22], were mainly interested in applications in statistics. The present paper too was motivated in part by statistical applications in [62].

*Supported in part by NSF grant DMS-0805946 and NSA grant H98230-12-1-0237.
†Michigan Technological University, USA. E-mail: ipinelis@mtu.edu

A particular case of a well-known result by Hoeffding [30] is the inequality

$$\mathsf{P}(S_n \geqslant x) \leqslant e^{-x^2/2} \tag{1.2}$$

for all $x \geqslant 0$. Obviously related to this is Khinchin's inequality — see e.g. survey [53]; for other developments, including more recent ones, see e.g. [43, 37, 52, 90]. Papers [65, 73] contain multidimensional analogues of an exact version of Khinchin's inequality, whereas [72] presents their extensions to multi-affine forms in $\varepsilon_1, \dots, \varepsilon_n$ (also known as Rademacher chaoses) with values in a vector space. Latała [42] gave bounds on moments and tails of Gaussian chaoses; Berry–Esseen-type bounds for general chaoses were recently obtained by Mossel, O'Donnell, and Oleszkiewicz [49]. For other kinds of improvements/generalizations of the inequality (1.2) see the recent paper [1] and bibliography there.

While easy to state and prove, bound (1.2) is, as noted by Efron [22], "not sharp enough to be useful in practice". Exponential inequalities such as (1.2) are obtained by finding a suitable upper bound (say $\mathcal{E}(t)$) on the exponential moments $\mathsf{E}\, e^{tS_n}$ and then minimizing the Markov bound $e^{-tx}\mathcal{E}(t)$ on $\mathsf{P}(S_n \geqslant x)$ in $t \geqslant 0$. The best exponential bound of this kind on the standard normal tail probability $\mathsf{P}(Z \geqslant x)$ is $\inf_{t\geqslant 0} e^{-tx}\,\mathsf{E}\, e^{tZ} = e^{-x^2/2}$, for any $x \geqslant 0$. 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 $x \to \infty$; cf. the result by Talagrand [84]. 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. The problem here is that the class of exponential moment functions is too small.

Eaton [19] obtained the moment comparison $\mathsf{E}\, f(S_n) \leqslant \mathsf{E}\, f(Z)$ for a much richer class of moment functions f, which enabled him [20] to derive an upper bound on $\mathsf{P}(S_n \geqslant x)$, which is asymptotic to $c_3\, \mathsf{P}(Z \geqslant x)$ as $x \to \infty$, where

$$c_3 := \tfrac{2e^3}{9} = 4.4634\dots.$$

Eaton further conjectured that $\mathsf{P}(S_n \geqslant x) \leqslant c_3\, \varphi(x)/x$ for $x > \sqrt{2}$. The stronger form of this conjecture,

$$\mathsf{P}(S_n \geqslant x) \leqslant c\, \mathsf{P}(Z \geqslant x) \tag{1.3}$$

for all $x \in \mathbb{R}$ with $c = c_3$ was proved by Pinelis [65], along with a multidimensional extension, which generalized results of Eaton and Efron [18]. Various generalizations and improvements of inequality (1.3) as well as related results were given by Pinelis [66, 67, 70, 74, 76, 77, 79, 57] and Bentkus [6, 7, 9].

Clearly, as pointed out e.g. in [10], the constant c in (1.3) cannot be less than

$$c_* := \frac{\mathsf{P}\left(\frac{1}{\sqrt{2}}(\varepsilon_1 + \varepsilon_2) \geqslant \sqrt{2}\right)}{\mathsf{P}(Z \geqslant \sqrt{2})} = 3.1786\dots, \tag{1.4}$$

which may be compared with c_3. Bobkov, Götze and Houdré (BGH) [11] gave a simple proof of (1.3) with a constant factor $c \approx 12.01$. Their method was based on the Chapman-Kolmogorov identity for the Markov chain (S_n). Such an identity was used, e.g., in [68] concerning a conjecture by Graversen and Peškir [24] on $\max_{k\leqslant n} |S_k|$. Pinelis [78] showed that a modification of the BGH method can be used to obtain inequality (1.3) with a constant factor $c \approx 1.01\, c_* \approx 3.22$. Bentkus and Dzindzalieta [8] recently closed the gap by proving that c_* is indeed the best possible constant factor c in (1.3); they used the Chapman-Kolmogorov identity together with the Berry-Esseen bound and a new extension of the Markov inequality. Bentkus and Dzindzalieta [8] also obtained the inequality

$$\mathsf{P}(S_n \geqslant x) \leqslant \tfrac{1}{4} + \tfrac{1}{8}\left(1 - \sqrt{2 - 2/x^2}\,\right) \quad \text{for } x \in (1, \sqrt{2}\,], \tag{1.5}$$

whereas Holzman and Kleitman [32] proved that $\mathsf{P}(S_n > 1) \leqslant \frac{5}{16}$.

We should also like to mention another kind of result, due to Montgomery-Smith [48], who obtained an upper bound on $\ln \mathsf{P}(S_n \geqslant x)$ and a matching lower bound on $\ln \mathsf{P}(S_n \geqslant Cx)$ for some absolute constant $C > 0$; these bounds depend on $x > 0$ and on the sequence $(a_1, \dots, a_n)$ and differ from each other by no more than an absolute constant factor; the constants were improved by Hitczenko and Kwapien [27]. As was pointed out by the referee, whereas the normal-tail-like bounds obtained in the present paper and its predecessors including [30, 20, 65, 78] will usually work better when the a_i's are fairly balanced, bounds such as the ones obtained in [48] can be advantageous otherwise, when the a_i's significantly differ in magnitude from one another. Indeed, the bounds given in [48] are expressed in terms of an interpolation norm of $(a_1, \dots, a_n)$, which is equivalent (up to a universal constant factor) to an expression based on splitting the a_i's into two groups according to the absolute values of the a_i's. The result of [48] was extended to sums of general independent zero-mean r.v.'s in [29], and the latter work was also motivated in part by that of Latała [41]. The proof in [48] was in part based on an extension of the improvement of Hoffmann-Jørgensen's inequality [31] found by Klass and Nowicki [34]. More recent developments in this direction are given in [35, 36].

In the mentioned paper [22], Efron conjectured that there exists an upper bound on the tail probability $\mathsf{P}(S_n \geqslant x)$ which behaves as the corresponding standard normal tail $\mathsf{P}(Z \geqslant x)$, and he presented certain facts in favor of this conjecture. Efron's conjecture suggests that even the best possible constant factor $c = c_* = 3.17\dots$ in (1.3) is excessive for large x; rather, for such x the ratio of a good bound on $\mathsf{P}(S_n \geqslant x)$ to $\mathsf{P}(Z \geqslant x)$ should be close to 1. Theorem 1.1 below provides such a bound, of simple and explicit form.

Another well-known conjecture, apparently due to Edelman [80, 21], is that

$$\mathsf{P}(S_n \geqslant x) \leqslant \sup_{n \geqslant 1} \mathsf{P}\left(\tfrac{1}{\sqrt{n}}\,(\varepsilon_1 + \dots + \varepsilon_n) \geqslant x\right) \tag{1.6}$$

for all $x \geqslant 0$; that is, the conjecture is that the supremum of $\mathsf{P}(S_n \geqslant x)$ over all finite sequences $(a_1, \dots, a_n)$ satisfying condition (1.1) is the same as that over all such $(a_1, \dots, a_n)$ with equal a_i's; cf. the above discussion concerning the result by Montgomery-Smith [48] vs. normal-tail-like bounds. Conjecture (1.6) was recently disproved; see [92, 59].

Another two known and interesting conjectures are that $\mathsf{P}(S_n > 1) \leqslant \frac{1}{4}$ [32, 2, 26] and that $\mathsf{P}(S_n \geqslant 1) \geqslant \frac{7}{64}$ [13, 28, 51, 89].

The main result of the present paper is

Theorem 1.1. *For all real $x > 0$*

$$\mathsf{P}(S_n \geqslant x) \leqslant Q(x) := \mathsf{P}(Z > x) + \frac{C\varphi(x)}{9 + x^2} < \mathsf{P}(Z > x)\Big(1 + \frac{C}{x}\Big), \tag{1.7}$$

where

$$C := 5\sqrt{2\pi e}\,\mathsf{P}(|Z| < 1) = 14.10\dots. \tag{1.8}$$

Remark 1.2. *The constant factor C is the best possible in the sense that the first inequality in* (1.7) *turns into the equality when $x = n = 1$. It would be of interest to find the optimal value of C if the constant 9 in the denominator in* (1.7) *is replaced by a significantly smaller positive value, say c. Then it could be possible to replace the constant C by a smaller value. At that, the factor $\frac{1}{c+x^2}$ would be decreasing faster than $\frac{1}{9+x^2}$, especially when $x > 0$ is not too large – since the "rate" $\left|\frac{\partial}{\partial x} \ln \frac{1}{c+x^2}\right| = \frac{2}{c/x+x}$ is greater for smaller $c > 0$. However, such a quest appears to entail further significant technical complications. Also, it is an open (and apparently very difficult) problem* *whether the asymptotic rate of decrease of the "extra" term $\frac{C\varphi(x)}{9+x^2}$ as $x \to \infty$ is the best possible one. Such questions appear to be related to the open problems stated at the end of [59]. It is hoped that these matters will be addressed in subsequent studies.*

Using e.g. part (II) of Proposition 3.1 (in Section 3 of this paper), it is easy to see that the ratio of the bound $Q(x)$ in (1.7) to $\mathsf{P}(Z > x)$ increases from ≈ 2.25 to ≈ 3.61 and then decreases to 1 as x increases from 0 to ≈ 2.46 to ∞, respectively. Figure 1 presents a graphical comparison of this ratio, $Q(x)/\,\mathsf{P}(Z > x)$, with

(i) the best possible constant factor $c = c_* \approx 3.18$ in (1.3);

(ii) the level 1, which is asymptotic (as $x \to \infty$) to the ratio of either one of the two bounds in (1.7) to $\mathsf{P}(Z > x)$, and hence, by the central limit theorem, is also asymptotic to the ratio of the supremum of $\mathsf{P}(S_n \geqslant x)$ (over all normalized Rademacher sums S_n) to $\mathsf{P}(Z > x)$;

(iii) the ratio of Hoeffding's bound $e^{-x^2/2}$ to $\mathsf{P}(Z > x)$.

In Figure 1, the graph of the latter ratio looks like a steep straight line (and asymptotically, for large x, is a straight line), most of which is outside the vertical range of the picture, thus showing how much the bounds $c_*\,\mathsf{P}(Z \geqslant x)$ and $Q(x)$ improve the Hoeffding bound $e^{-x^2/2}$.

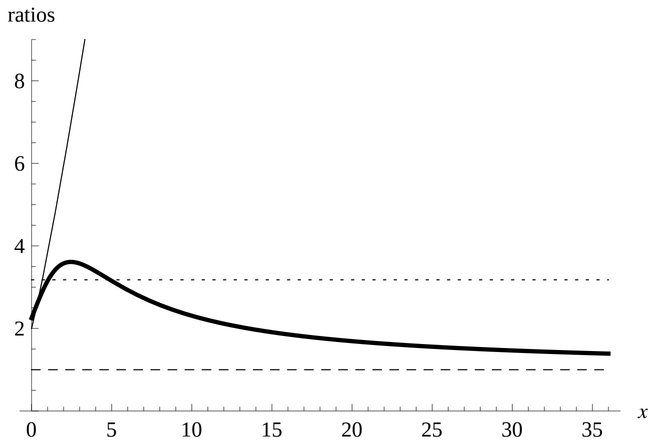

Figure 1: Ratio $Q(x)/\,\mathsf{P}(Z > x)$ (thick solid) compared with the ratio $e^{-x^2/2}/\,\mathsf{P}(Z > x)$ (solid, steeply upwards), as well as with the levels 1 (dashed) and $c_* \approx 3.18$ (dotted)

In view of the main result of Bentkus [5], one immediately obtains the following corollary of Theorem 1.1.

Corollary 1.3. *Let $X, X_1, \dots, X_n$ be independent identically distributed r.v.'s such that $\mathsf{P}(|X| \leqslant 1) = 1$ and $\mathsf{E}\,X = 0$. Then*

$$\mathsf{P}\left(\frac{X_1 + \cdots + X_n}{\sqrt{n}} \geqslant x\right) \leqslant 2\hat{Q}_n(x)$$

for all real $x \geqslant 0$, where $\hat{Q}_n$ is the linear interpolation of the restriction of the function Q to the set $\frac{2}{\sqrt{n}}(\frac{n}{2} - \lfloor \frac{n}{2} \rfloor + \mathbb{Z})$.

Here we shall present just one more application of Theorem 1.1, to the self-normalized sums

$$V_n := \frac{X_1 + \cdots + X_n}{\sqrt{X_1^2 + \cdots + X_n^2}},$$

where, following Efron [22], we assume that the X_i's satisfy the so-called orthant symmetry condition: the joint distribution of $s_1X_1,\dots,s_nX_n$ is the same for any choice of signs $s_1,\dots,s_n\in\{1,-1\}$, 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, $V_n=S_n$ if $X_i=a_i\varepsilon_i$ for all i. It was noted by Efron that (i) Student's statistic T_n is a monotonic function of the so-called self-normalized sum: $T_n=\sqrt{\frac{n-1}{n}}\,V_n/\sqrt{1-V_n^2/n}$ and (ii) the orthant symmetry implies in general that the distribution of V_n is a mixture of the distributions of normalized Rademacher sums S_n. Thus, one obtains

Corollary 1.4. *Theorem 1.1 holds with V_n in place of S_n.*

Note that many of the most significant advances concerning self-normalized sums are rather recent; e.g., a necessary and sufficient condition for their asymptotic normality was obtained only in 1997 by Giné, Götze, and Mason [23].

It appears natural to compare the probability inequalities given in Theorem 1.1 with limit theorems for large deviation probabilities. Most of such theorems, referred to as large deviation principles (LDP's), deal with logarithmic asymptotics, that is, asymptotics of the logarithm of small probabilities; see e.g. [15]. As far as the logarithmic asymptotics is concerned, the mentioned bounds $c_*\,\mathsf{P}(Z\geqslant x)$ and $Q(x)$ and the Hoeffding bound $e^{-x^2/2}$ are all the same: $\ln\big[c_*\,\mathsf{P}(Z\geqslant x)\big]\sim\ln Q(x)\sim\ln e^{-x^2/2}=-x^2/2$ as $x\to\infty$; yet, as we have seen, at least the first two of these bounds are vastly different from the Hoeffding bound, especially from the perspective of statistical practice. Results on the so-called exact asymptotics for large deviations (that is, asymptotics for the small probabilities themselves, rather than for their logarithms) are much fewer; see e.g. [15, Theorem 3.7.4] and [54, Ch. VIII]. Note that the inequalities in (1.7) hold for all $x>0$, and, *a priori*, the summands $a_i\varepsilon_i$ do not have to be identically or nearly identically distributed; cf. conjecture (1.6). In contrast, almost all limit theorems for large deviations in the literature – whether with exact or logarithmic asymptotics – hold only for $x=O(\sqrt{n}\,)$, with n being the number of identically or quasi-identically distributed (usually independent or nearly independent) random summands; the few exceptions here include results of the papers [50, 63, 64, 69, 91] and references therein, where the restriction $x=O(\sqrt{n}\,)$ is not imposed and x is allowed to be arbitrarily large. In general, observe that a limit theorem is a statement on the existence of an inequality, not yet fully specified, as e.g. in "there exists some n_0 such that $|x_n-x|<\varepsilon$ for all $n\geqslant n_0$"; as such, a limit theorem cannot provide a specific bound. Of course, being less specific, limit theorems are applicable to objects of much greater variety and complexity, and limit theorems usually provide valuable initial insight. Yet, it seems natural to suppose that the tendency, say in the studies of large deviation probabilities, will be to proceed from logarithmic asymptotics to asymptotics of the probabilities themselves and then on to exact inequalities. We appear to be largely at the beginning of this process, still struggling even with such comparatively simple objects as the Rademacher sums – the simplicity of which is only comparative, as the discussion around Figure 1 in [78] suggests. However, there have already been a number of big strides made in this direction. For instance, Boucheron, Bousquet, Lugosi, and Massart [12] obtained explicit bounds on moments of general functions of independent r.v.'s; their approach was based on a generalization of Ledoux's entropy method [44, 45], using at that a generalized tensorization inequality due to Latała and Oleszkiewicz [40]. Another, more recent example demonstrating the same tendency is the work by van de Geer [88]. Even more recently, Tropp [86] provided noncommutative generalizations of the Bennett, Bernstein, Chernoff, and Hoeffding bounds – even with explicit and optimal constants; as pointed out in [86], "[a]symptotic theory is less relevant in practice". Yet, as stated above, in the case of Rademacher sums and other related cases significantly more precise bounds can be obtained.

2 Proof of Theorem 1.1: outline

Let us begin the proof with several introductory remarks.

There are many symbols used in the proof. Therefore, let us assume a localization principle for notations: any notations introduced in a section or in a proof of a lemma/sublemma supersede those introduced in preceding sections or proofs. For example, the meaning of the X_i's introduced later in this section differs from that in Section 1.

Without loss of generality (w.l.o.g.), assume that

$$0 \leqslant a_1 \leqslant \ldots \leqslant a_n =: a, \tag{2.1}$$

so that $a = \max_i a_i$. Introduce the numbers

$$u_i := u_{i,x} := xa_i,$$

whence for all $x \geqslant 0$

$$0 \leqslant u_1 \leqslant \ldots \leqslant u_n = xa. \tag{2.2}$$

The proof of Theorem 1.1 is to a large extent based on a careful analysis of the Esscher exponential tilt transform of the r.v. S_n. In introducing and using this transform, Esscher and then Cramér were motivated by applications in actuarial science. Closely related to the Esscher transform is the saddle-point approximation; for a recent development in this area, see [61]. The Esscher tilt has been used extensively in limit theorems for large deviation probabilities, but much less commonly concerning explicit probability inequalities – two rather different in character cases of the latter kind are represented by Raič [81] and Pinelis and Molzon [62]. One may also note that, in deriving LDP's, the exponential tilt is usually employed to get a lower bound on the probability; in contrast, in this paper the tilt is used to obtain the upper bound. One may also note that, whereas in [78, 8] the main difficulty was to deal with moderate values values of x, in the present paper both the moderate and large values of x present significant problems; in a sense, here the consideration depends not just on the value of x itself but, to a greater extent, on the value of the product xa.

The main idea of the proof is to reduce the problem from that on the vector $(a_1, \ldots, a_n)$ of an unbounded dimension n to a set of low-dimensional extremal problems. The first step here is to use exponential tilting to obtain upper bounds on $\mathsf{P}(S_n \geqslant x)$ in term of sums of the form $\sum_i g(u_i)$, which can then be represented as $x^2 \int \tilde{g}\, \mathrm{d}\nu$, where $\tilde{g}(u) := g(u)/u^2$ (for $u \neq 0$),

$$\nu := \frac{1}{x^2} \sum_i u_i^2 \delta_{u_i}, \tag{2.3}$$

and δ_t denotes the Dirac probability measure at point t, so that ν is a probability measure on the interval $[0, xa]$. This step turns the initial finite-dimensional problem into an infinite-dimensional one, involving the measure ν. However, then the well-known Carathéodory principle allows one to reduce the dimension to (at most) $k-1$, where k is the total number of the integrals (with the respect to the measure ν) involved in the extremal problem in hand; see e.g. [58] for recent developments in this direction, and references therein. The above ideas were carried out in the first version of this paper — see [55].

Later, I realized that the systems of integrands one has to deal with in the proof of Theorem 1.1 possess the so-called Tchebycheff and, even, Markov properties; therefore, one can reduce the dimension even further, to about $k/2$, which allows for more effective analyses. It should also be noted that the verification of the Markov property of a finite sequence of functions largely reduces to checking the positivity of several functions of only one variable. Major expositions of the theory of Tchebycheff–Markov systems and its applications are given in the monographs by Karlin and Studden [33] and Kreĭn and Nudel′man [38]; closely related to this theory are certain results in real algebraic geometry, whereby polynomials are "certified" to be positive on a semialgebraic domain by means of an explicit representation, say in terms of sums of squares of polynomials; see e.g. [39, 47]. A brief review of the Tchebycheff and Markov systems of functions, which contains all the definitions and facts necessary for the applications in the present paper, is given in [60]. For the readers' convenience, we shall present here a condensed version of [60] — in Appendix A at the end of this paper.

Even after the just described reductions in dimension, the proof of Theorem 1.1 entails extensive (even if rather routine) calculations, especially symbolic ones.

In this section, a number of lemmas will be stated, from which Theorem 1.1 will easily follow. Most of these lemmas will be proved in Section 3 – with the exception of Lemmas 2.3 and 2.7, whose proofs are more complicated and will each be presented in a separate section. Each of these two more complicated lemmas is based on a number of sublemmas – which are stated in the corresponding section and used there to prove the lemma. Each of these sublemmas (except for Sublemma 4.1) is a technical statement about one or several smooth functions of one real variable and is proved using the Mathematica implementation of the Tarski algorithm [85, 46, 14]; the proofs of these sublemmas can be found in [56]. It should be quite clear that all such calculations done with an aid of a computer are no less reliable or rigorous than similar, or even less involved, calculations done by hand.

*****

For all $i = 1, \dots, n$, let

$$X_i := a_i \varepsilon_i.$$

Next, let $\tilde X_1, \dots, \tilde X_n$ be any r.v.'s such that

$$\mathsf{E}\, g(\tilde X_1, \dots, \tilde X_n) = \frac{\mathsf{E}\, e^{xS_n} g(X_1, \dots, X_n)}{\mathsf{E}\, e^{xS_n}} \tag{2.4}$$

for all Borel-measurable functions $g\colon \mathbb{R}^n \to \mathbb{R}$. Equivalently, one may require condition (2.4) only for Borel-measurable indicator functions g; clearly, such r.v.'s $\tilde X_i$ do exist. It is also clear that the r.v.'s $\tilde X_i$ are independent. Moreover, for each i the distribution of $\tilde X_i$ is $\big(e^{u_i}\delta_{a_i} + e^{-u_i}\delta_{-a_i}\big)/(e^{u_i} + e^{-u_i})$.

Formula (2.4) presents the mentioned Esscher exponential tilt transform, with the tilting parameter (TP) the same as the x in (1.7); that is, we choose the TP to be the minimizer of $e^{-tx}\,\mathsf{E}\, e^{tZ} = e^{-tx+t^2/2}$ in $t \geqslant 0$ — rather than the minimizer of $e^{-tx}\,\mathsf{E}\, e^{tS_n}$, which latter is usually taken as the TP in limit theorems for large deviations and can thus be expressed only via an implicit function. Our choice of the TP appears to simplify the proof greatly.

In terms of the tilted r.v.'s $\tilde X_1, \dots, \tilde X_n$, introduce now

$$m_x := \sum_i \mathsf{E}\, \tilde X_i = \frac{1}{x} \sum_i u_i \operatorname{th} u_i, \quad s_x := \sqrt{\sum_i \mathsf{Var}\, \tilde X_i} = \frac{1}{x}\sqrt{\sum_i \frac{u_i^2}{\operatorname{ch}^2 u_i}}, \tag{2.5}$$

$$L_x := \frac{1}{s_x^3} \sum_i \mathsf{E}\, |\tilde X_i - \mathsf{E}\, \tilde X_i|^3, \tag{2.6}$$

where $\mathrm{ch} := \cosh$, $\mathrm{sh} := \sinh$, $\mathrm{th} := \tanh$, and $\mathrm{arcch} := \operatorname{arccosh}$ $\big($assuming that $\mathrm{arcch}\, z \geqslant 0$ for all $z \in [1, \infty)$; thus, for each $z \in [1, \infty)$, $\mathrm{arcch}\, z$ is the unique solution $y \geqslant 0$ to the equation $\mathrm{ch}\, y = z\big)$. Let $\overline{F}_n$ and $\overline{\Phi}$ denote, respectively, the tail function of $\tilde{X}_1 + \cdots + \tilde{X}_n$ and the standard normal tail function, so that

$$\overline{F}_n(z) = \mathsf{P}(\tilde{X}_1 + \cdots + \tilde{X}_n \geqslant z) \quad \text{and} \quad \overline{\Phi}(z) = \mathsf{P}(Z \geqslant z)$$

for all real z. Also, let c_{BE} denote the least possible constant in the Berry-Esseen inequality

$$\sup_{z\in\mathbb{R}} \Big|\overline{F}_n(z) - \overline{\Phi}\Big(\frac{z - m_x}{s_x}\Big)\Big| \leqslant c_{\mathrm{BE}} L_x; \tag{2.7}$$

by Shevtsova [83],

$$c_{\mathrm{BE}} \leqslant \tfrac{56}{100};$$

a slightly worse bound, $c_{\mathrm{BE}} \leqslant 0.5606$, is due to Tyurin [87].

Lemma 2.1. *For all $x \geqslant 0$*

$$\mathsf{P}(S_n \geqslant x) \leqslant N(x) + 2c_{\mathrm{BE}} B(x), \tag{2.8}$$

where

$$N(x) := \exp\Big\{\sum_i \ln \mathrm{ch}\, u_i + \frac{x^2 s_x^2}{2} - x m_x + \ln \overline{\Phi}\Big(\frac{x - m_x}{s_x} + x s_x\Big)\Big\}, \tag{2.9}$$

$$B(x) := L_x \exp\Big\{-x^2 + \sum_i \ln \mathrm{ch}\, u_i\Big\}. \tag{2.10}$$

Lemma 2.1 carries out much of the first step in the proof of Theorem 1.1, as mentioned before: using exponential tilting to reduce the original problem, on the vector $(a_1, \ldots, a_n)$ of an unbounded dimension n, to one involving sums of the form $\sum_i g(u_i)$ — recall here the expressions of m_x and s_x in (2.5) in terms of such sums. At this point, only the factor L_x in (2.10) remains to be bound in terms a sum of the form $\sum_i g(u_i)$, which will be done later, in Sublemma 4.1.

Next, introduce the ratio

$$r(x) := \frac{\varphi(x)}{x\overline{\Phi}(x)}, \tag{2.11}$$

which is the inverse Mills ratio at x divided by x. By [71, Proposition 1.2], r is strictly and continuously decreasing from ∞ to 1 on the interval $(0, \infty)$, so that there is a unique root $x_{3/2} \in (0, \infty)$ of the equation

$$r(x_{3/2}) = 3/2;$$

at that,

$$x_{3/2} = 1.03\ldots$$

and

$$1 < r(x) \leqslant \tfrac{3}{2} \quad \text{for } x \geqslant x_{3/2}. \tag{2.12}$$

Introduce also

$$u_* := \tfrac{51}{125} = 0.408 \tag{2.13}$$

and

$$h(x) := \frac{C\varphi(x)}{9 + x^2} \tag{2.14}$$

(cf. (1.7)).

The next two lemmas provide upper bounds on the terms $N(x)$ and $2c_{\mathrm{BE}}B(x)$ in (2.8) — for x large enough; also, in Lemma 2.3, $u_n = \max_i x a_i$ is assumed to be small enough.

Lemma 2.2. *If $x \geqslant x_{3/2}$ then $N(x) \leqslant \overline{\Phi}(x)$.*

Lemma 2.3. *If $x \geqslant \frac{13}{10}$ and $u_n \leqslant u_*$, then $2c_{\mathrm{BE}}B(x) \leqslant h(x)$.*

The proofs of the above two lemmas are comparatively difficult, especially the latter one, of Lemma 2.3, which will take entire Section 4. It is in these two proofs that we use methods of extremal problems (including special tools for Tchebycheff–Markov systems) for measures in a given moment set — to carry out the mentioned reduction from an infinite-dimensional problem to finite, in fact low, dimensions.

In contrast with Lemmas 2.2 and 2.3, the following lemma is easy and to be used just as a quick reference concerning the second inequality in (1.7).

Lemma 2.4. *If $x > 0$ then $h(x) < C\overline{\Phi}(x)/x$.*

Next is another easy lemma, which serves as the induction basis (with $n = 1$) in the proof of Theorem 1.1 below; it is also used in the proof of Lemma 2.8.

Lemma 2.5. *If $x > 0$ then $\mathsf{P}(\varepsilon_1 \geqslant x) \leqslant \overline{\Phi}(x) + h(x)$.*

Now we shall address a case not covered by Lemma 2.3: when u_n is not small enough (and x is still large enough). For this case we adopt an approach, which is based on the Chapman–Kolmogorov identity (2.16) and similar to methods used e.g. in [68, Proof of Proposition 2], [11, Proof of Theorem 4.2], and [78, Proof of Theorem 2]. Clearly, this method is quite different from the combination of the methods of exponential tilting and solving extremal problems for moment sets used — for the small enough values of u_n — in the proofs of Lemmas 2.1–2.3. Consider

$$U := U_{x,a} := \frac{x-a}{\sqrt{1-a^2}} \quad \text{and} \quad V := V_{x,a} := \frac{x+a}{\sqrt{1-a^2}},$$

with a as in (2.1). The following two lemmas provide information about behavior of the two respective terms, $\overline{\Phi}(x)$ and $h(x) = \frac{C\varphi(x)}{9+x^2}$, in the bound $Q(x)$ on $\mathsf{P}(S_n \geqslant x)$ in (1.7). This information will be used to carry the induction step in proof of Theorem 1.1.

Lemma 2.6. *If $x \geqslant \sqrt{3}$ then $\frac{1}{2}\overline{\Phi}(U) + \frac{1}{2}\overline{\Phi}(V) \leqslant \overline{\Phi}(x)$.*

Lemma 2.6 was proved in [11]; cf. also [78, Lemma 5].

Lemma 2.7. *If $x \geqslant \frac{15}{10}$ and $u_n \geqslant u_*$, then $\frac{1}{2}h(U) + \frac{1}{2}h(V) \leqslant h(x)$; recall here that, by (2.2), $a = u_n/x$.*

Thus, Lemmas 2.6 and 2.7 taken together close the gap that was left open in Lemma 2.3 because of the restriction $u_n \leqslant u_*$ there. Still, in Lemmas 2.2, 2.3, 2.6, and 2.7 the value of x was assumed to be large enough. This remaining gap is closed by

Lemma 2.8. *For all $x \in (0, \sqrt{3}\,]$*

$$\mathsf{P}(S_n \geqslant x) \leqslant \overline{\Phi}(x) + h(x). \tag{2.15}$$

A key point in the proof of Lemma 2.8 is using inequality (1.3) with $c = 3.22$, as provided by [78]; we could have used instead the main result of [8], with $c = c_* = 3.17\ldots$, but $c = 3.22$ is enough for our purposes here.

Based on the above lemmas, we can now present

Proof of Theorem 1.1. By definition (2.14) and Lemma 2.4, it is enough to prove inequality (2.15) for all $x > 0$. This can be done by induction on n. Indeed, for $n = 1$ this is Lemma 2.5. Assume now that $n \geqslant 2$. In view of Lemma 2.8, it is enough to prove inequality (2.15) for all $x > \sqrt{3}$. At that, in view of Lemmas 2.1, 2.2, and 2.3, it is enough to consider the case $u_n > u_*$. To do that, write

$$\mathsf{P}(S_n \geqslant x) = \tfrac{1}{2}\,\mathsf{P}(\tilde{S}_{n-1} \geqslant U) + \tfrac{1}{2}\,\mathsf{P}(\tilde{S}_{n-1} \geqslant V), \tag{2.16}$$

where $\tilde{S}_{n-1} := b_1\varepsilon_1 + \cdots + b_{n-1}\varepsilon_{n-1}$, with $b_i := a_i/\sqrt{1-a^2}$. It remains to use the induction hypothesis together with Lemmas 2.6 and 2.7. □

3 Proofs of Lemmas 2.1, 2.2, 2.4, 2.5, and 2.8

Proof of Lemma 2.1. Reading equation (2.4) with $g(X_1, \dots, X_n) = e^{-xS_n} \times \mathbf{I}\{S_n \geqslant x\}$ right-to-left, recalling (2.7), and observing that $\mathsf{E}\, e^{xS_n} = \prod_i \operatorname{ch} u_i$, one has

$$\frac{\mathsf{P}(S_n \geqslant x)}{\mathsf{E}\, e^{xS_n}} = -\int_{[x,\infty)} e^{-xy}\, \mathrm{d}\overline{F}_n(y) = \int_x^\infty x e^{-xy} \big(\overline{F}_n(x) - \overline{F}_n(y)\big)\, \mathrm{d}y \leqslant N_1(x) + B_1(x),$$

where

$$\begin{aligned} N_1(x) &:= \int_x^\infty x e^{-xy} \Big[\overline{\Phi}\Big(\frac{x-m_x}{s_x}\Big) - \overline{\Phi}\Big(\frac{y-m_x}{s_x}\Big)\Big]\, \mathrm{d}y \\ &= \int_x^\infty e^{-xy} \varphi\Big(\frac{y-m_x}{s_x}\Big) \frac{\mathrm{d}y}{s_x} = \frac{N(x)}{\mathsf{E}\, e^{xS_n}} \end{aligned}$$

and

$$B_1(x) := 2c_{\mathrm{BE}}\, L_x \int_x^\infty x e^{-xy}\, \mathrm{d}y = 2c_{\mathrm{BE}}\, L_x\, e^{-x^2} = \frac{2c_{\mathrm{BE}}\, B(x)}{\mathsf{E}\, e^{xS_n}}.$$

Thus, (2.8) follows. □

Now and later in the paper, we need the following special l'Hospital-type rule for monotonicity.

Proposition 3.1. ([75, Propositions 4.1 and 4.3]) *Let $-\infty \leqslant a < b \leqslant \infty$. Let f and g be differentiable functions defined on the interval (a, b). It is assumed that g and g' do not take on the zero value and do not change their respective signs on (a, b).*

- *(I) If $f(a+) = g(a+) = 0$ or $f(b-) = g(b-) = 0$, and if the ratio f'/g' is strictly increasing/decreasing on (a, b), then (respectively) $(f/g)'$ is strictly positive/negative and hence the ratio f/g is strictly increasing/decreasing on (a, b).*
- *(II) If $f(b-) = g(b-) = 0$ and if the ratio f'/g' switches its monotonicity pattern at most once on (a, b) — only from increase to decrease, then the ratio f/g does so.*

Proof of Lemma 2.2. Let us begin this proof by using the well-known fact that the tail function $\overline{\Phi}$ is log-concave. This fact is contained e.g. in [25, 67]. Alternatively, it can be easily obtained using part (I) of Proposition 3.1, since $(\ln \overline{\Phi})' = -\frac{\varphi}{\overline{\Phi}}$. So, one can write

$$\ln \overline{\Phi}(y) \leqslant \ln \overline{\Phi}(x) + (\ln \overline{\Phi})'(x)(y - x) = \ln \overline{\Phi}(x) - x r(x)(y - x),$$

with $y = \frac{x-m_x}{s_x} + x s_x$ (cf. (2.9)) and $r(x)$ defined by (2.11). Therefore and in view of (2.5),

$$\frac{1}{x^2} \ln \frac{N(x)}{\overline{\Phi}(x)} \leqslant \tilde{\mathcal{E}}(r, \nu) := \int_0^{xa} \Big[e(u) + r \cdot \Big(1 - \frac{f(u)}{s_x}\Big)\Big] \nu(\,\mathrm{d}u)$$

(recall (2.1)),

$$e(u) := \frac{\ln \operatorname{ch} u}{u^2} + \frac{1}{2 \operatorname{ch}^2 u} - \frac{\operatorname{th} u}{u} \quad \text{and} \quad f(u) := 1 - \frac{\operatorname{th} u}{u} + \frac{1}{\operatorname{ch}^2 u}$$

for $u \neq 0$, $e(0) := 0$ and $f(0) := 1$, and $r := r(x)$. Note that the probability measure ν on the interval $[0, xa]$ defined by (2.3) satisfies the restriction

$$\int_0^{xa} b \, \mathrm{d}\nu = s_x^2, \quad \text{where} \quad b(u) := \frac{1}{\operatorname{ch}^2 u}. \tag{3.1}$$

Recalling now (2.12), we see that to prove Lemma 2.2 we only need to show that $\tilde{\mathcal{E}}(r, \nu) \leqslant 0$ for all such probability measures ν and all $r \in [1, \frac{3}{2}]$; in fact, since $\tilde{\mathcal{E}}(r, \nu)$ is affine in r, it suffices to consider only $r \in \{1, \frac{3}{2}\}$.

Using Proposition A.3 in Appendix A and the Mathematica command Reduce, one can verify that each of the two systems $(1, -b, f - e)$ and $(1, -b, f)$ is an M_+-system on any interval $[c, d] \subset [0, \infty)$; as mentioned earlier, this verification reduces to checking the positivity of several (Wronskian) functions of only one variable; for the system $(1, -b, f - e)$, this takes about 20 sec on a standard laptop, and about 1 sec for the system $(1, -b, f)$. Since $s_x \in (0, 1]$ and $r \geqslant 1$, the integrand in the integral expression of $\tilde{\mathcal{E}}(r, \nu)$ can be rewritten as $g := r - \frac{1}{\theta}(f - \theta e)$ with $\theta := \frac{s_x}{r} \in (0, 1]$, and so, $(1, -b, -g)$ is an M_+-system on $[0, xa]$, for any $r \geqslant 1$ and any value of s_x. Hence, by Proposition A.4 in Appendix A, the minimum of $\int_0^{xa} (-g) \, \mathrm{d}\nu$, and thus the maximum of $\tilde{\mathcal{E}}(r, \nu)$, over all the probability measures ν on $[0, xa]$ satisfying the restriction $\int_0^{xa} b \, \mathrm{d}\nu = s_x^2$ is attained when the support of ν is a singleton subset (say $\{u\}$) of $[0, xa]$. For this u, one has $s_x = 1/\operatorname{ch} u$, and it now suffices to show that $g(u) = e(u) + r \cdot \big(1 - f(u) \operatorname{ch} u\big) \leqslant 0$ for $r \in \{1, \frac{3}{2}\}$ and $u \in [0, \infty)$; using again the Mathematica command Reduce, it takes about 2 sec to check this in each of the two cases, $r = 1$ and $r = \frac{3}{2}$. □

Proof of Lemma 2.4. Using part (I) of Proposition 3.1, one can see that the ratio $\frac{xh(x)}{\overline{\Phi}(x)}$ is increasing in $x > 0$, from 0 to C. Now the result follows. □

Proof of Lemma 2.5. Observe that the definition (1.8) of C is equivalent to the condition $\overline{\Phi}(1) + h(1) = \frac{1}{2}$ (cf. Remark 1.2). Hence and because $\overline{\Phi} + h$ is decreasing on $(0, \infty)$, one has $\mathsf{P}(\varepsilon_1 \geqslant x) = \frac{1}{2} = \overline{\Phi}(1) + h(1) \leqslant \overline{\Phi}(x) + h(x)$ for all $x \in (0, 1]$. For $x > 1$, one obviously has $\mathsf{P}(\varepsilon_1 \geqslant x) = 0 < \overline{\Phi}(x) + h(x)$. □

Proof of Lemma 2.8. By the symmetry, Chebyshev's inequality, and the main result of [78],

$$\mathsf{P}(S_n \geqslant x) \leqslant \tfrac{1}{2} \, \mathbf{I}\{0 < x \leqslant 1\} + \tfrac{1}{2x^2} \, \mathbf{I}\{1 < x \leqslant \tfrac{13}{10}\} + 3.22 \overline{\Phi}(x) \, \mathbf{I}\{\tfrac{13}{10} < x \leqslant \sqrt{3}\,\}$$

for all $x \in (0, \sqrt{3}\,]$. In particular, for all $x \in (0, 1]$ one has $\mathsf{P}(S_n \geqslant x) \leqslant \frac{1}{2} = \mathsf{P}(\varepsilon_1 \geqslant x) \leqslant \overline{\Phi}(x) + h(x)$, by Lemma 2.5.

Next, let us prove (2.15) for $x \in (1, \frac{13}{10}]$. Write $x^2 \overline{\Phi}(x) = \overline{\Phi}(x)/p(x)$, where $p(x) := 1/x^2$. Note that $\overline{\Phi}(\infty-) = p(\infty-) = 0$ and $\overline{\Phi}'(x)/p'(x) = x^3 \varphi(x)/2$, so that $\overline{\Phi}'/p'$ switches its monotonicity pattern exactly once on $(0, \infty)$, from increase to decrease. Hence, by part (II) of Proposition 3.1, $x^2 \overline{\Phi}(x) = \overline{\Phi}(x)/p(x)$ switches its monotonicity pattern at most once, and at that necessarily from increase to decrease, as x increases from 1 to $\frac{13}{10}$. So, the minimum of $x^2 \overline{\Phi}(x)$ over $x \in [1, \frac{13}{10}]$ is attained at one of the end points of the interval $[1, \frac{13}{10}]$; in fact, the minimum is at $x = 1$. It is also easy to see that the minimum of $x^2 h(x)$ over $x \in [1, \frac{13}{10}]$ is attained at $x = 1$ as well. Thus, $\mathsf{P}(S_n \geqslant x) \leqslant \frac{1}{2x^2} = \frac{1}{2x^2\left(\overline{\Phi}(x) + h(x)\right)} \big(\overline{\Phi}(x) + h(x)\big) \leqslant \frac{1}{2\left(\overline{\Phi}(1) + h(1)\right)} \big(\overline{\Phi}(x) + h(x)\big) = \overline{\Phi}(x) + h(x)$ for $x \in (1, \frac{13}{10}]$.

The case $x \in (\frac{13}{10}, \sqrt{3}\,]$ is similar to the just considered case $x \in (1, \frac{13}{10}]$. Here, using part (II) of Proposition 3.1 again, one can see that $h/\overline{\Phi}$ switches, just once, from increase to decrease on $(0, \infty)$; in particular, $h/\overline{\Phi}$ increases on $(\frac{13}{10}, \sqrt{3}\,]$, because $(h/\overline{\Phi}\,)'(\sqrt{3}) = 0.29\ldots > 0$. So, to complete the proof of Lemma 2.8, it is enough to check that $3.22 \overline{\Phi}(\frac{13}{10}) \leqslant \overline{\Phi}(\frac{13}{10}) + h(\frac{13}{10})$, which is true. □

4 Proof of Lemma 2.3

As was stated earlier, proofs of all sublemmas in this paper (except for Sublemma 4.1 below) can be found in [56].

We shall need the following tight upper bound on the Lyapunov ratio L_x, defined by (2.6):

Sublemma 4.1. *One has*

$$L_x \leqslant \frac{1}{x^3} \sum_i u_i^3 (1+\operatorname{th}^2 u_i) \operatorname{ch} u_i. \tag{4.1}$$

The proof of Sublemma 4.1 will be given at the end of this section.

By Sublemma 4.1 and the definition (2.10) of $B(x)$,

$$B(x) \leqslant \tfrac{1}{x} e^{-x^2+\tilde{J}}, \tag{4.2}$$

where

$$\tilde{J} := \tilde{J}(x,\nu) := x^2 \int \ell \, \mathrm{d}\nu + \ln \int k \, \mathrm{d}\nu,$$

$$k(u) := u(1+\operatorname{th}^2 u) \operatorname{ch} u, \quad \ell(u) := \tfrac{\ln \operatorname{ch} u}{u^2} \text{ for } u \neq 0,$$

$\ell(0) := \frac{1}{2}$, and ν is the probability measure on the interval $[0, u_*]$ defined by (2.3), so that ν satisfies the restriction (3.1). To obtain the upper bound $h(x)$ on $2c_{\mathrm{BE}}B(x)$ as stated in Lemma 2.3, we shall maximize $\tilde{J}(x,\nu)$ over all such probability measures ν. To do so, let us first maximize $\int k \, \mathrm{d}\nu$ given values of the integrals $\int 1 \, \mathrm{d}\nu (=1)$, $\int b \, \mathrm{d}\nu (= s_x^2$, as in (3.1)), and $\int \ell \, \mathrm{d}\nu$.

Noting that $(\ln \operatorname{ch})'' = \operatorname{th}' = \operatorname{sech}^2$ and applying (twice) the special lHospital-type rule for monotonicity given by part (I) of Proposition 3.1, one sees that

$$\ell' < 0 \quad \text{on} \quad (0, \infty). \tag{4.3}$$

Sublemma 4.2. [56] *The sequence* $(g_0, g_1, g_2, g_3) := (1, -b, -\ell, k)$ *is an* M_+*-system on* $[0, u_*]$*; here one may want to recall Definition A.2 in Appendix A.*

A proof of Sublemma 4.2 can be found in [56], where it is based on the statement in [60] reproduced as Proposition A.3 in Appendix A here.

So, by Proposition A.4 (with $n = 2$ and $m = 1$ there), it suffices to consider measures ν of the form $\nu = (1-t)\delta_u + t\delta_{u_*}$ for some $t \in [0,1]$ and $u \in [0, u_*]$. For such ν,

$$\tilde{J}(x,\nu) = J(t,u) := J_x(t,u) := x^2 \cdot \big((1-t)\ell(u) + t\ell(u_*)\big) + \ln\big((1-t)k(u) + tk(u_*)\big).$$

Thus, we need to maximize $J(t,u)$ over all $(t,u) \in [0,1] \times [0,u_*]$; clearly, this maximum is attained. For all $(t,u) \in (0,1) \times [0,u_*)$,

$$\frac{\partial J(t,u)}{\partial t} \, \frac{(1-t)k(u) + tk(u_*)}{u_* - u} = \frac{\big(k(u_*) - k(u)\big) + \tau\,\big(\ell(u_*) - \ell(u)\big)}{u_* - u}$$
$$= k'(w) + \tau \ell'(w), \tag{4.4}$$

$$\frac{\partial J(t,u)}{\partial u} \, \frac{(1-t)k(u) + tk(u_*)}{1-t} = k'(u) + \tau \ell'(u), \tag{4.5}$$

where $\tau := x^2 \cdot \big((1-t)k(u) + tk(u_*)\big)$ and w is some number such that $u < w < u_*$ (whose existence follows by the mean-value theorem). So, if the maximum of J over the set $[0,1] \times [0,u_*]$ is attained at some point $(t,u) \in (0,1) \times (0,u_*)$, then at this point one has $\frac{\partial J}{\partial t} = 0 = \frac{\partial J}{\partial u}$, whence, by (4.4), (4.5), and (4.3), $\frac{k'(w)}{\ell'(w)} = -\tau = \frac{k'(u)}{\ell'(u)}$ while $u_* > w > u \geqslant 0$, which contradicts

Sublemma 4.3. [56] *The function* $\rho := \frac{k'}{\ell'}$ *is strictly increasing on the interval* $[0, u_*]$ *(by continuity, we let* $\rho(0) := \rho(0+) = -\infty$*).*

Also, no maximum of J is attained at any point $(t, u) \in (0, 1) \times \{0\}$, because at any such point the right-hand side of (4.5) is $k'(0) + \tau\ell'(0) = 1 + \tau \cdot 0 > 0$, whereas the left-hand side of (4.5) must be $\leqslant 0$. Thus, the maximum can be attained at some point $(t, u) \in [0, 1]\times[0, u_*]$ only if either $t \in \{0, 1\}$ or $u = u_*$. Therefore the maximizing measure ν must be concentrated at one point, say u, of the interval $[0, u_*]$. Together with (4.2), this shows that

$$B(x) \leqslant \sup_{u\in[0,u_*]} \frac{1}{x}\, e^{-x^2+J_0(x,u)},$$

where

$$J_0(x, u) := J_x(0, u) = x^2 \cdot \ell(u) + \ln k(u).$$

So, Lemma 2.3 reduces now to the following statement:

$$\Lambda(x, u) := J_0(x, u) - \frac{x^2}{2} - \ln x + \ln(9 + x^2) - K \overset{(?)}{\leqslant} 0 \tag{4.6}$$

for all $(x, u) \in [\frac{13}{10}, \infty) \times [0, u_*]$, where

$$K := \ln \frac{C}{2\sqrt{2\pi}\, c_{\mathrm{BE}}}.$$

Thus, one may want to maximize Λ in $u \in [0, u_*]$. Towards that end, observe that for all $u > 0$

$$\frac{1}{-\ell'(u)}\,\frac{\partial\Lambda}{\partial u} = \gamma(u) - x^2,$$

where

$$\gamma := -\frac{k'}{k\ell'} = -\rho\,\frac{1}{k};$$

so, the partial derivative of Λ in $u > 0$ equals $\gamma(u) - x^2$ in sign. On the other hand, the function $\frac{1}{k}$ is positive and strictly decreasing and, in view of Sublemma 4.3, the function $(-\rho)$ is so as well (on the interval $[0, u_*]$). It follows that the function γ too is positive and strictly decreasing on $(0, u_*]$; at that, $\gamma(0+) = \infty$.

Introduce now

$$x_* := \sqrt{\gamma(u_*)} = 7.39\ldots. \tag{4.7}$$

By the mentioned properties of the function γ, for each $x \in (0, x_*]$ one has $\gamma(u) \geqslant x^2$ for all $u \in [0, u_*]$ and hence $\Lambda(x, u)$ increases in $u \in [0, u_*]$, so that $\Lambda(x, u) \leqslant \Lambda(x, u_*)$ for all $u \in [0, u_*]$. Since the derivative of $\Lambda(x, u_*)$ in x is a rather simple rational function, it is easy to see that $\Lambda(x, u_*) \leqslant 0$ for all $x \geqslant \frac{13}{10}$. So, inequality (4.6) holds for all $(x, u) \in [\frac{13}{10}, x_*] \times [0, u_*]$.

It remains to prove (4.6) for each $x \in [x_*, \infty)$ (and all $u \in [0, u_*]$). For each such x, there is a unique $u_x \in [0, u_*]$ such that $\gamma(u) - x^2$ and hence $\frac{\partial\Lambda}{\partial u}$ are opposite to $u - u_x$ in sign, and so, $\Lambda(x, u) \leqslant \Lambda(x, u_x)$ for all $u \in [0, u_*]$.

Since, by (4.3), the function ℓ is strictly and continuously decreasing on $[0, \infty)$, there is a unique inverse function $\ell^{-1}\colon (0, \frac{1}{2}] \mapsto [0, \infty)$. Now introduce

$$\tilde{J}_0(x, \lambda) := J_0\big(x, \ell^{-1}(\lambda)\big) = x^2\lambda + \ln \tilde{k}(\lambda), \quad \text{where} \quad \tilde{k} := k \circ \ell^{-1}$$

and $\lambda \in [\ell(u_*), \ell(0)] = [\ell(u_*), \frac{1}{2}]$. Next, observe that $(\ln \tilde{k})' = -\gamma\circ\ell^{-1}$, which is decreasing on $[\ell(u_*), \frac{1}{2}]$, because γ and ℓ (and hence ℓ^{-1}) are decreasing. It follows that the function $\ln \tilde{k}$ is concave on $[\ell(u_*), \frac{1}{2}]$, and so, $\tilde{J}_0(x, \lambda)$ is concave in $\lambda \in [\ell(u_*), \frac{1}{2}]$ – for each real x. At this point, we need

Sublemma 4.4. [56] *If* $u \in (0, u_*]$ *then* $\gamma(u) > \frac{6}{u^2}$.

By (4.7) and Sublemma 4.4, if $u = \frac{\sqrt{6}}{x}$ and $x \geqslant x_*$, then $u \in (0, u_*]$ and $\gamma(\frac{\sqrt{6}}{x}) > x^2 = \gamma(u_x)$, which in turn implies that $\frac{\sqrt{6}}{x} < u_x$, $\ell(\frac{\sqrt{6}}{x}) > \ell(u_x)$, and $(\ln \tilde{k})'\big(\ell(\frac{\sqrt{6}}{x})\big) < (\ln \tilde{k})'\big(\ell(u_x)\big) = -\gamma(u_x) = -x^2$ $\big($since γ, ℓ, and $(\ln \tilde{k})'$ are decreasing$\big)$; so, for all $\lambda \in [\ell(u_*), \frac12]$, $\frac{\partial \tilde{J}_0}{\partial \lambda}\big(x, \ell(\frac{\sqrt{6}}{x})\big) < \frac{\partial \tilde{J}_0}{\partial \lambda}\big(x, \ell(u_x)\big) = 0$; therefore and by the concavity of $\tilde{J}_0(x, \lambda)$ in λ,

$$\tilde{J}_0(x, \lambda) \leqslant \tilde{J}_0\big(x, \ell(\tfrac{\sqrt{6}}{x})\big) + \tfrac{\partial \tilde{J}_0}{\partial \lambda}\big(x, \ell(\tfrac{\sqrt{6}}{x})\big)\big(\lambda - \ell(\tfrac{\sqrt{6}}{x})\big) \leqslant \hat{J}_0(x, \tfrac{\sqrt{6}}{x})$$

for all $\lambda \in [\ell(u_*), \frac12]$, where

$$\hat{J}_0(x, u) := J_0(x, u) + \big(x^2 - \gamma(u)\big)\big(\ell(u_*) - \ell(u)\big).$$

Thus, in view of (4.6), Lemma 2.3 reduces to the inequality $\hat{J}_0(x, \frac{\sqrt{6}}{x}) - \frac{x^2}{2} - \ln x + \ln(9 + x^2) - K \leqslant 0$ for all $x \geqslant x_*$, where we change the variable once again, from x to u, by the formula $x = \frac{\sqrt{6}}{u}$. So, Lemma 2.3 reduces to

Sublemma 4.5. [56] *For all* $u \in (0, u_*]$

$$\tilde{\Lambda}(u) := \hat{J}_0(\tfrac{\sqrt{6}}{u}, u) - \tfrac{3}{u^2} - \ln \tfrac{\sqrt{6}}{u} + \ln(9 + \tfrac{6}{u^2}) \leqslant K.$$

It remains, in this section, to present

Proof of Sublemma 4.1. Observe that $L_x = (xs_x)^{-3} \sum_i u_i^3(1 - \operatorname{th}^4 u_i)$. So, inequality (4.1) means exactly that

$$\sum_i u_i^3(1 - \operatorname{th}^4 u_i) - s_x^3 \sum_i u_i^3(1 + \operatorname{th}^2 u_i) \operatorname{ch} u_i = \sum_i u_i^2 g(u_i) \leqslant 0 \tag{4.8}$$

for all u_i's in the interval $[0, u_*]$ such that $\sum_i u_i^2 = x^2$ and $\sum_i \frac{u_i^2}{\operatorname{ch}^2 u_i} = x^2 s_x^2$, where

$$g(u) := u(1 - \operatorname{th}^4 u) - s_x^3 u(1 + \operatorname{th}^2 u) \operatorname{ch} u = u\Big(2 - \frac{1}{\operatorname{ch}^2 u}\Big)\Big(\frac{1}{\operatorname{ch}^2 u} - s_x^3 \operatorname{ch} u\Big).$$

Next, the object $\sum_i u_i^2 g(u_i)$ in (4.8) with the restrictions $\sum_i u_i^2 = x^2$ and $\sum_i \frac{u_i^2}{\operatorname{ch}^2 u_i} = x^2 s_x^2$ can be rewritten as $x^2 \mathsf{E}\, h(Y)$ given $\mathsf{E}\, Y = s_x^2$, where $h(\cdot) := h_a(\cdot)$ $\big($as in (4.9) below$\big)$ with $a = s_x^3$ and Y is a r.v. with the distribution $\nu := \frac{1}{x^2} \sum_i u_i^2 \delta_{v_i}$, with $v_i := \frac{1}{\operatorname{ch}^2 u_i}$; note that one always has $s_x \in (0, 1]$ and ν is indeed a probability measure due to the restriction $\sum_i u_i^2 = x^2$. So, by Subsublemma 4.6 below and Jensen's inequality, $x^{-2} \sum_i u_i^2 g(u_i) = \mathsf{E}\, h(Y) \leqslant h(\mathsf{E}\, Y) = h(s_x^2) = 0$, which proves the inequality in (4.8) and hence that in (4.1). □

Subsublemma 4.6. [56] *For each* $a \in [0, 1]$*, the function*

$$(0, 1] \ni v \mapsto h_a(v) := \operatorname{arcch}(\tfrac{1}{\sqrt{v}})(2 - v)(v - \tfrac{a}{\sqrt{v}}) \tag{4.9}$$

is concave.

5 Proof of Lemma 2.7

This proof could be somewhat simplified using the mentioned result (1.5); however, let us present an independent proof here, which is not much more complicated. Let

$$\Delta := \Delta(x, u) := \tfrac{\sqrt{2\pi}}{C}\big[\tfrac12 h\big(U_{x,u/x}\big) + \tfrac12 h\big(V_{x,u/x}\big) - h(x)\big].$$

We have to show that $\Delta(x,u) \leqslant 0$ for all pairs (x,u) in the set

$$P := \{(x,u) \in [\tfrac{15}{10},\infty) \times [u_*,\infty) \colon u < x\};$$

the condition $u < x$ here corresponds to the condition $a = a_n < 1$.

The idea of the proof of Lemma 2.7 is, essentially, to fix a value of x and then differentiate $\Delta(x,u)$ twice with respect to certain functions of u (which may be different for different fixed values of x) so that the sign of the resulting generalized second partial derivative of $\Delta(x,u)$ in u be comparatively easy to determine. In other words, we establish a generalized convexity pattern for $\Delta(x,u)$ in u.

Toward this end, introduce first the set

$$\tilde{P} := \{(x,u) \in [\tfrac{15}{10},\infty) \times [\tfrac{4}{10},\infty) \colon u < x\},$$

which is slightly larger than P; recall here (2.13). Then we shall consider the mentioned generalized first and second partial derivatives of $\Delta(x,u)$ in u:

$$\Delta_1 := \Delta_1(x,u) := F_1(x,u)\,\frac{\partial \Delta}{\partial u} \quad \text{and} \tag{5.1}$$

$$\Delta_2 := \Delta_2(x,u) := F_2(x,u)\frac{\partial \Delta_1}{\partial u}, \tag{5.2}$$

where $F_1(x,u)$ and $F_2(x,u)$ are certain expressions, to be defined soon, such that

$$F_1(x,u)(u-1) > 0 \text{ and } F_2(x,u) > 0 \text{ for all } (x,u) \in \tilde{P} \text{ with } u \neq 1. \tag{5.3}$$

Moreover, we shall show that $F_1(x,u)$ and $F_2(x,u)$ are such that the Δ_1 and Δ_2 as in (5.1) and (5.2) possess the following properties:

$$\Delta_2 > 0 \text{ on } \tilde{P}, \tag{5.4}$$

$$\Delta_1(x, x-) = -\tfrac{1}{2} < 0 \text{ for } x > 0, \tag{5.5}$$

and, furthermore, one has the following sublemmas, proved in [56]:

Sublemma 5.1. [56] $\Delta_1(x, \frac{4}{10}) > 0$ *for all* $x \in [\frac{15}{10},\infty)$

and

Sublemma 5.2. [56] $\Delta(x,u_*) < 0$ *for all* $x \in [\frac{15}{10},\infty)$.

It will then follow by (5.2), (5.3), and (5.4) that $\Delta_1(x,u)$ increases in $u \in [\frac{4}{10},1)$ and in $u \in (1,x)$ for each $x \in [\frac{15}{10},\infty)$, whence, by (5.5), $\Delta_1(x,u) < 0$ for all $(x,u) \in \tilde{P}$ such that $u > 1$ and, by Sublemma 5.1, $\Delta_1(x,u) > 0$ for all $(x,u) \in \tilde{P}$ such that $u < 1$. Thus, for all points $(x,u) \in \tilde{P}$ with $u \neq 1$ one has $\Delta_1(x,u)(u-1) < 0$ and hence, by (5.1) and (5.3), $\Delta(x,u)$ decreases in $u \in [\frac{4}{10},x)$ for each $x \in [\frac{15}{10},\infty)$. Using now Sublemma 5.2 and recalling that $u_* > \frac{4}{10}$, one concludes that $\Delta < 0$ on P, which yields Lemma 2.7.

It remains to present F_1 and F_2 such that (5.3), (5.4), and (5.5) hold indeed. Let

$$F_1(x,u) := \exp\left\{\frac{\left(u-x^2\right)^2}{2\left(x^2-u^2\right)}\right\} \frac{\left(x^2-u^2\right) p_2(x,u)^2}{(u-1)x^2\left(x^2-u\right)p_1(x,u)} \quad \text{and} \tag{5.6}$$

$$F_2(x,u) := \exp\left\{\frac{2ux^2}{x^2-u^2}\right\} \frac{(u-1)^2(x-u)^2(u+x)^2\left(x^2-u\right)^2 p_1(x,u)^2 p_3(x,u)^3}{p_2(x,u)}, \tag{5.7}$$

where

$$\left.\begin{aligned} p_1(x,u) &:= x^2(11+x^2) - (10u^2 + 2ux^2),\\ p_2(x,u) &:= x^2(9+x^2) - (8u^2 + 2ux^2),\\ p_3(x,u) &:= x^2(9+x^2) - (8u^2 - 2ux^2). \end{aligned}\right\} \tag{5.8}$$

Using e.g. the Mathematica command Reduce, one can see that on the set $\tilde{P}$ the polynomials p_1, p_2, and p_3 are positive. Note also that $u < x < x^2$ for all $(x,u) \in \tilde{P}$. So, (5.3) holds.

Next, with definitions (5.1), (5.2), (5.6), and (5.7) in place, it turns out that $\Delta_2(x,u)$ is a polynomial in (x,u) (of degree 24 in x, and 14 in u). Using Reduce again, one verifies (5.4).

Finally, it is straightforward (even if somewhat tedious) to check (5.5).

A Tchebycheff–Markov systems

For a nonnegative integer n, let $g_0,\dots,g_n$ be (real-valued) continuous functions on an interval $[a,b]$ for some a and b such that $-\infty < a < b < \infty$. Let $\mathcal{M}$ denote the set of all (nonnegative) Borel measures on $[a,b]$. Take any point $\mathbf{c} = (c_0,\dots,c_n) \in \mathbb{R}^{n+1}$ such that

$$\mathcal{M}_{\mathbf{c}} := \Big\{\mu \in \mathcal{M} \colon \int_a^b g_i \,\mathrm{d}\mu = c_i \quad \text{for all} \quad i \in \overline{0,n}\Big\} \neq \emptyset; \tag{A.1}$$

here and in what follows, for any m and n in $\mathbb{Z} \cup \{\infty\}$ we let $\overline{m,n} := \{j \in \mathbb{Z} \colon m \leqslant j \leqslant n\}$.

Definition A.1. *The sequence $(g_0,\dots,g_n)$ of functions is a T-system if the restrictions of these $n+1$ functions to any subset of $[a,b]$ of cardinality $n+1$ are linearly independent. If, for each $k \in \overline{0,n}$, the initial subsequence $(g_0,\dots,g_k)$ of the sequence $(g_0,\dots,g_n)$ is a T-system, then $(g_0,\dots,g_n)$ is said to be an M-system (where M refers to Markov).*

Let $(g_0,\dots,g_n)$ be a T-system on $[a,b]$. Let $\det\big(g_i(x_j)\big)_0^n$ denote the determinant of the matrix $\big(g_i(x_j) \colon i \in \overline{0,n},\ j \in \overline{0,n}\big)$. This determinant is continuous in $(x_0,\dots,x_n)$ in the (convex) simplex (say Σ) defined by the inequalities $a \leqslant x_0 < \dots < x_n \leqslant b$ and does not vanish anywhere on Σ. So, $\det\big(g_i(x_j)\big)_0^n$ is constant in sign on Σ.

Definition A.2. *The sequence $(g_0,\dots,g_n)$ is said to be a T_+-system on $[a,b]$ if $\det\big(g_i(x_j)\big)_0^n > 0$ for all $(x_0,\dots,x_n) \in \Sigma$. If $(g_0,\dots,g_k)$ is a T_+-system on $[a,b]$ for each $k \in \overline{0,n}$, then the sequence $(g_0,\dots,g_n)$ is said to be an M_+-system on $[a,b]$.*

In the case when the functions $g_0,\dots,g_n$ are n times differentiable at a point $x \in (a,b)$, consider also the *Wronskians*

$$W_0^k(x) := \det\big(g_i^{(j)}(x)\big)_0^k,$$

where $k \in \overline{0,n}$ and $g_i^{(j)}$ is the jth derivative of g_i, with $g_i^{(0)} := g_i$; in particular, $W_0^0(x) = g_0(x)$.

Proposition A.3. *Suppose that the functions $g_0,\dots,g_n$ are (still continuous on $[a,b]$ and) n times differentiable on (a,b). Then, for the sequence $(g_0,\dots,g_n)$ to be an M_+-system on $[a,b]$, it is necessary that $W_0^k \geqslant 0$ on (a,b) for all $k \in \overline{0,n}$, and it is sufficient that $u_0 > 0$ on $[a,b]$ and $W_0^k > 0$ on (a,b) for all $k \in \overline{1,n}$.*

Thus, verifying the M_+-property largely reduces to checking the positivity of several functions of only one variable.

A special case of Proposition A.3 (with $n=1$ and $g_0 = 1$) is the following well-known fact: if a function g_1 is continuous on $[a,b]$ and has a positive derivative on (a,b), then g_1 is (strictly) increasing on $[a,b]$; vice versa, if g_1 is increasing on $[a,b]$, then the derivative of g_1 (if exists) must be nonnegative on (a,b).

As in this special case, the proof of Proposition A.3 in general can be based on the mean-value theorem; cf. e.g. [33, Theorem 1.1 of Chapter XI], which states that the requirement for W_0^k to be strictly positive on the *closed* interval $[a,b]$ for all $k \in \overline{0,n}$

is equivalent to a condition somewhat stronger than being an M_+-system on $[a,b]$; in connection with this, one may also want to look at [38, Theorem IV.5.2]. Note that, in the applications to the proofs of Lemmas 2.2 and 2.3 of this paper, the relevant Wronskians vanish at the left endpoint of the interval.

The proof of Proposition A.3 can be obtained by induction on n using the recursive formulas for the determinants $\det\big(g_i(x_j)\big)_0^n$ and W_0^n as displayed right above [33, (5.5) in Chapter VIII] and in [33, (5.6) in Chapter VIII], where we use g_i in place of ψ_i.

Proposition A.4. *Suppose that $(g_0,\dots,g_{n+1})$ is an M_+-system on $[a,b]$ or, more generally, each of the sequences $(g_0,\dots,g_n)$ and $(g_0,\dots,g_{n+1})$ is a T_+-system on $[a,b]$. Suppose also that condition* (A.1) *holds. Let $m:=\lfloor\frac{n+1}{2}\rfloor$. Then one has the following.*

(I) *The maximum (respectively, the minimum) of $\int_a^b g_{n+1}\,\mathrm{d}\mu$ over all $\mu\in\mathcal{M}_{\mathbf{c}}$ is attained at a unique measure $\mu_{\max}$ (respectively, $\mu_{\min}$) in $\mathcal{M}_{\mathbf{c}}$. Moreover, the measures $\mu_{\max}$ and $\mu_{\min}$ do not depend on the choice of g_{n+1}, as long as g_{n+1} is such that $(g_0,\dots,g_{n+1})$ is a T_+-system on $[a,b]$.*

(II) *There exist subsets $X_{\max}$ and $X_{\min}$ of $[a,b]$ such that $X_{\max}\supseteq\operatorname{supp}\mu_{\max}$, $X_{\min}\supseteq\operatorname{supp}\mu_{\min}$, and*

  (a) *if $n=2m$ then $\operatorname{card}X_{\max}=\operatorname{card}X_{\min}=m+1$, $X_{\max}\ni b$, and $X_{\min}\ni a$;*

  (b) *if $n=2m-1$ then $\operatorname{card}X_{\max}=m+1$, $\operatorname{card}X_{\min}=m$, and $X_{\max}\supseteq\{a,b\}$.*

To illustrate Proposition A.4, one may consider the simplest two special cases when the conditions of the proposition hold and its conclusion is obvious:

(i) $n=0$, $g_0(x)\equiv1$, g_1 is increasing on $[a,b]$, and $c_0\geqslant0$; then $\operatorname{supp}\mu_{\max}\subseteq\{b\}$ and $\operatorname{supp}\mu_{\min}\subseteq\{a\}$; in fact, $\mu_{\max}=c_0\delta_b$ and $\mu_{\min}=c_0\delta_a$; here and in what follows, δ_x denotes the Dirac probability measure at point x.

(ii) $n=1$, $g_0(x)\equiv1$, $g_1(x)\equiv x$, g_2 is strictly convex on $[a,b]$, $c_0\geqslant0$, and $c_1\in[c_0a,c_0b]$; then $\operatorname{supp}\mu_{\max}\subseteq\{a,b\}$ and $\operatorname{card}\operatorname{supp}\mu_{\min}\leqslant1$; in fact, $\mu_{\max}=\frac{c_0b-c_1}{b-a}\delta_a+\frac{c_1-c_0a}{b-a}\delta_b$, and $\mu_{\min}=c_0\delta_{c_1/c_0}$ if $c_0>0$ and $\mu_{\min}=0$ if $c_0=0$.

These examples also show that the T-property of systems of functions can be considered as generalized monotonicity/convexity; see e.g. [82] and bibliography there.

Proof of Proposition A.4. Consider two cases, depending on whether $\mathbf{c}$ is strictly or singularly positive; in equivalent geometric terms, this means, respectively, that $\mathbf{c}$ belongs to the interior or the boundary of the smallest closed convex cone containing the subset $\{(g_0(x),\dots,g_n(x))\colon x\in[a,b]\}$ of $\mathbb{R}^{n+1}$ [38, Theorem IV.6.1].

In the first case, when $\mathbf{c}$ is strictly positive, both statements of Proposition A.4 follow by [38, Theorem IV.1.1]; at that, one should let $X_{\max}=\operatorname{supp}\mu_{\max}$ and $X_{\min}=\operatorname{supp}\mu_{\min}$. (The condition that $\mathbf{c}$ be strictly positive appears to be missing in the statement of the latter theorem; cf. [33, Theorem 1.1 of Chapter 1.1].)

In the other case, when $\mathbf{c}$ is singularly positive, use [38, Theorem III.4.1], which states that in this case the set $\mathcal{M}_{\mathbf{c}}$ consists of a single measure (say μ_*), and its support set $X_*:=\operatorname{supp}\mu_*$ is of an index $\leqslant n$; that is, $\ell_-+2\ell+\ell_+\leqslant n$, where ℓ_-, ℓ, and ℓ_+ stand for the cardinalities of the intersections of X_* with the sets $\{a\}$, (a,b), and $\{b\}$. It remains to show that this condition on the index of X_* implies that there exist subsets $X_{\max}$ and $X_{\min}$ of $[a,b]$ satisfying the conditions (IIa) and (IIb) of Proposition A.4 and such that $X_{\max}\cap X_{\min}\supseteq X_*$.

If $n=2m$ then $\operatorname{card}(X_*\cap(a,b))=\ell\leqslant\lfloor\frac{2m-\ell_--\ell_+}{2}\rfloor\leqslant\lfloor\frac{2m-\ell_-}{2}\rfloor=m-\ell_-$; so, $\operatorname{card}(X_*\cup\{b\})\leqslant\ell_-+(m-\ell_-)+1=m+1$. Adding now to the set $X_*\cup\{b\}$ any $m+1-\operatorname{card}(X_*\cup\{b\})$

points of the complement of $X_* \cup \{b\}$ to $[a,b]$, one obtains a subset $X_{\max}$ of $[a,b]$ such that $X_{\max} \supseteq X_*$, $X_{\max} \ni b$, and $\operatorname{card} X_{\max} = m+1$. Similarly, there exists a subset $X_{\min}$ of $[a,b]$ such that $X_{\min} \supseteq X_*$, $X_{\min} \ni a$, and $\operatorname{card} X_{\min} = m+1$.

If $n = 2m-1$ then $\operatorname{card}(X_* \cap (a,b)) = \ell \leqslant \lfloor \frac{2m-1-\ell_- -\ell_+}{2} \rfloor \leqslant m-1$ and hence $\operatorname{card}(X_* \cup \{a,b\}) \leqslant 1 + (m-1) + 1 = m+1$. So, there exists a subset $X_{\max}$ of $[a,b]$ such that $X_{\max} \supseteq X_*$, $X_{\max} \supseteq \{a,b\}$, and $\operatorname{card} X_{\max} = m+1$. One also has $\operatorname{card} X_* = \ell_- + \ell + \ell_+ \leqslant \lfloor \frac{2m-1+\ell_- +\ell_+}{2} \rfloor \leqslant \lfloor \frac{2m+1}{2} \rfloor = m$. So, there exists a subset $X_{\min}$ of $[a,b]$ such that $X_{\min} \supseteq X_*$ and $\operatorname{card} X_{\min} = m$. □

References

[1] Sergei N. Antonov and Victor M. Kruglov, *Sharpened versions of a Kolmogorov's inequality*, Statist. Probab. Lett. **80** (2010), no. 3-4, 155–160. MR-2575440

[2] A. Ben-Tal, A. Nemirovski, and C. Roos, *Robust solutions of uncertain quadratic and conic quadratic problems*, SIAM J. Optim. **13** (2002), no. 2, 535–560 (electronic). MR-1951034

[3] Aharon Ben-Tal and Arkadi Nemirovski, *On safe tractable approximations of chance-constrained linear matrix inequalities*, Math. Oper. Res. **34** (2009), no. 1, 1–25. MR-2542986

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

[5] V. Bentkus, *An inequality for large deviation probabilities of sums of bounded i.i.d. random variables*, Liet. Mat. Rink. **41** (2001), no. 2, 144–153. MR-1851123

[6] ______, *A remark on the inequalities of Bernstein, Prokhorov, Bennett, Hoeffding, and Talagrand*, Liet. Mat. Rink. **42** (2002), no. 3, 332–342. MR-1947624

[7] ______, *An inequality for tail probabilities of martingales with differences bounded from one side*, J. Theoret. Probab. **16** (2003), no. 1, 161–173. MR-1956826

[8] V. Bentkus and D. Dzindzalieta, *A tight Gaussian bound for weighted sums of Rademacher random variables (preprint)*.

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

[10] ______, *On measure concentration for separately Lipschitz functions in product spaces*, Israel J. Math. **158** (2007), 1–17. MR-2342455

[11] Sergey G. Bobkov, Friedrich Götze, and Christian Houdré, *On Gaussian and Bernoulli covariance representations*, Bernoulli **7** (2001), no. 3, 439–451. MR-1836739

[12] 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

[13] D. L. Burkholder, *Independent sequences with the Stein property*, Ann. Math. Statist. **39** (1968), 1282–1288. MR-0228045

[14] 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

[15] Amir Dembo and Ofer Zeitouni, *Large deviations techniques and applications*, second ed., Applications of Mathematics (New York), vol. 38, Springer-Verlag, New York, 1998. MR-1619036

[16] Kürşad Derinkuyu and Mustafa Ç. Pınar, *On the S-procedure and some variants*, Math. Methods Oper. Res. **64** (2006), no. 1, 55–77. MR-2264772

[17] Kürşad Derinkuyu, Mustafa Ç. Pınar, and Ahmet Camcı, *An improved probability bound for the approximate S-Lemma*, Oper. Res. Lett. **35** (2007), no. 6, 743–746. MR-2361043

[18] M. L. Eaton and Bradley Efron, *Hotelling's T^2 test under symmetry conditions*, J. Amer. Statist. Assoc. **65** (1970), 702–711. MR-0269021

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

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

[21] D. Edelman, *Private communication, 1994*.

[22] Bradley Efron, *Student's t-test under symmetry conditions*, J. Amer. Statist. Assoc. **64** (1969), 1278–1302. MR-0251826

[23] Evarist Giné, Friedrich Götze, and David M. Mason, *When is the Student t-statistic asymptotically standard normal?*, Ann. Probab. **25** (1997), no. 3, 1514–1531. MR-1457629

[24] S. E. Graversen and G. Peškir, *Extremal problems in the maximal inequalities of Khintchine*, Math. Proc. Cambridge Philos. Soc. **123** (1998), no. 1, 169–177. MR-1474873

[25] Richard L. Hall, Marek Kanter, and Michael D. Perlman, *Inequalities for the probability content of a rotated square and related convolutions*, Ann. Probab. **8** (1980), no. 4, 802–813. MR-577317

[26] Jean-Baptiste Hiriart-Urruty, *A new series of conjectures and open questions in optimization and matrix analysis*, ESAIM Control Optim. Calc. Var. **15** (2009), no. 2, 454–470. MR-2513094

[27] Paweł Hitczenko and Stanisław Kwapień, *On the Rademacher series*, Probability in Banach spaces, 9 (Sandjberg, 1993), Progr. Probab., vol. 35, Birkhäuser Boston, Boston, MA, 1994, pp. 31–36. MR-1308508

[28] ______, *On the Rademacher series*, Probability in Banach spaces, 9 (Sandjberg, 1993), Progr. Probab., vol. 35, Birkhäuser Boston, Boston, MA, 1994, pp. 31–36. MR-1308508

[29] Paweł Hitczenko and Stephen Montgomery-Smith, *Measuring the magnitude of sums of independent random variables*, Ann. Probab. **29** (2001), no. 1, 447–466. MR-1825159

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

[31] Jørgen Hoffmann-Jørgensen, *Sums of independent Banach space valued random variables*, Studia Math. **52** (1974), 159–186. MR-0356155

[32] Ron Holzman and Daniel J. Kleitman, *On the product of sign vectors and unit vectors*, Combinatorica **12** (1992), no. 3, 303–316. MR-1195893

[33] Samuel Karlin and William J. Studden, *Tchebycheff systems: With applications in analysis and statistics*, Pure and Applied Mathematics, Vol. XV, Interscience Publishers John Wiley & Sons, New York-London-Sydney, 1966. MR-0204922

[34] Michael J. Klass and Krzysztof Nowicki, *An improvement of Hoffmann-Jørgensen's inequality*, Ann. Probab. **28** (2000), no. 2, 851–862. MR-1782275

[35] ______, *Uniformly accurate quantile bounds via the truncated moment generating function: the symmetric case*, Electron. J. Probab. **12** (2007), no. 47, 1276–1298 (electronic). MR-2346512

[36] ______, *Uniformly accurate quantile bounds for sums of arbitrary independent random variables*, J. Theoret. Probab. **23** (2010), no. 4, 1068–1091. MR-2735737

[37] H. König and S. Kwapień, *Best Khintchine type inequalities for sums of independent, rotationally invariant random vectors*, Positivity **5** (2001), no. 2, 115–152. MR-1825172

[38] M. G. Kreĭn and A. A. Nudel′man, *The Markov moment problem and extremal problems*, American Mathematical Society, Providence, R.I., 1977, Ideas and problems of P. L. Čebyšev and A. A. Markov and their further development, Translated from the Russian by D. Louvish, Translations of Mathematical Monographs, Vol. 50. MR-0458081

[39] Jean Bernard Lasserre, *Moments, positive polynomials and their applications*, Imperial College Press Optimization Series, vol. 1, Imperial College Press, London, 2010. MR-2589247

[40] R. Latała and K. Oleszkiewicz, *Between Sobolev and Poincaré*, Geometric aspects of functional analysis, Lecture Notes in Math., vol. 1745, Springer, Berlin, 2000, pp. 147–168. MR-1796718

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

[42] ______, *Estimates of moments and tails of Gaussian chaoses*, Ann. Probab. **34** (2006), no. 6, 2315–2331. MR-2294983

[43] Rafał Latała and Krzysztof Oleszkiewicz, *On the best constant in the Khinchin-Kahane inequality*, Studia Math. **109** (1994), no. 1, 101–104. MR-1267715

[44] Michel Ledoux, *On Talagrand's deviation inequalities for product measures*, ESAIM Probab. Statist. **1** (1995/97), 63–87 (electronic). MR-1399224

[45] ______, *The concentration of measure phenomenon*, Mathematical Surveys and Monographs, vol. 89, American Mathematical Society, Providence, RI, 2001. MR-1849347

[46] 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

[47] Murray Marshall, *Positive polynomials and sums of squares*, Mathematical Surveys and Monographs, vol. 146, American Mathematical Society, Providence, RI, 2008. MR-2383959

[48] S. J. Montgomery-Smith, *The distribution of Rademacher sums*, Proc. Amer. Math. Soc. **109** (1990), no. 2, 517–522. MR-1013975

[49] Elchanan Mossel, Ryan O'Donnell, and Krzysztof Oleszkiewicz, *Noise stability of functions with low influences: invariance and optimality*, Ann. of Math. (2) **171** (2010), no. 1, 295–341. MR-2630040

[50] A. V. Nagaev, *Probabilities of large deviations of sums of independent random variables (Doctor of Science Thesis, Tashkent 1970).*

[51] Krzysztof Oleszkiewicz, *On the Stein property of Rademacher sequences*, Probab. Math. Statist. **16** (1996), no. 1, 127–130. MR-1407938

[52] ______, *On a nonsymmetric version of the Khinchine-Kahane inequality*, Stochastic inequalities and applications, Progr. Probab., vol. 56, Birkhäuser, Basel, 2003, pp. 157–168. MR-2073432

[53] G. Peshkir and A. N. Shiryaev, *Khinchin inequalities and a martingale extension of the sphere of their action*, Uspekhi Mat. Nauk **50** (1995), no. 5(305), 3–62. MR-1365047

[54] V. V. Petrov, *Sums of independent random variables*, Springer-Verlag, New York, 1975, Translated from the Russian by A. A. Brown, Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 82. MR-0388499

[55] I. Pinelis, *An asymptotically Gaussian bound on the Rademacher tails, preprint, version 1, `http://arxiv.org/pdf/1007.2137v1.pdf`.*

[56] ______, *An asymptotically Gaussian bound on the Rademacher tails, preprint, version 3, `http://arxiv.org/pdf/1007.2137v3.pdf`.*

[57] ______, *On the Bennett-Hoeffding inequality (preprint), arXiv:0902.4058v1 [math.PR].*

[58] ______, *On the extreme points of moments sets, preprint, `http://arxiv.org/find/all/1/au:+pinelis/0/1/0/all/0/1`.*

[59] ______, *On the supremum of the tails of normalized sums of independent Rademacher random variables, preprint, `http://arxiv.org/find/all/1/au:+pinelis/0/1/0/all/0/1`.*

[60] ______, *Tchebycheff systems and extremal problems for generalized moments: a brief survey, preprint, `http://arxiv.org/find/all/1/au:+pinelis/0/1/0/all/0/1`.*

[61] ______, *Exponential deficiency of convolutions of densities, published online, DOI 10.1051/ps/2010010*, ESAIM: Probability and Statistics (2011).

[62] I. Pinelis and R. Molzon, *Berry-Esséen bounds for general nonlinear statistics, with applications to Pearson's and non-central Student's and Hotelling's (preprint), arXiv:0906.0177v1 [math.ST].*

[63] I. F. Pinelis, *A problem on large deviations in a space of trajectories*, Theory Probab. Appl. **26** (1981), no. 1, 69–84.

[64] I. F. Pinelis, *Asymptotic equivalence of the probabilities of large deviations for sums and maximum of independent random variables*, Limit theorems of probability theory, Trudy Inst. Mat., vol. 5, "Nauka" Sibirsk. Otdel., Novosibirsk, 1985, pp. 144–173, 176. MR-821760

[65] 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

[66] ______, *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

[67] ______, *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

[68] ______, *On exact maximal Khinchine inequalities*, High dimensional probability, II (Seattle, WA, 1999), Progr. Probab., vol. 47, Birkhäuser Boston, Boston, MA, 2000, pp. 49–63. MR-1857314

[69] ______, *Exact asymptotics for large deviation probabilities, with applications*, Modeling uncertainty, Internat. Ser. Oper. Res. Management Sci., vol. 46, Kluwer Acad. Publ., Boston, MA, 2002, pp. 57–93. MR-1893275

[70] ______, *L'Hospital type rules for monotonicity: applications to probability inequalities for sums of bounded random variables*, JIPAM. J. Inequal. Pure Appl. Math. **3** (2002), no. 1, Article 7, 9 pp. (electronic). MR-1888922

[71] ______, *Monotonicity properties of the relative error of a Padé approximation for Mills' ratio*, JIPAM. J. Inequal. Pure Appl. Math. **3** (2002), no. 2, Article 20, 8 pp. (electronic). MR-1906389

[72] ______, *Spherically symmetric functions with a convex second derivative and applications to extremal probabilistic problems*, Math. Inequal. Appl. **5** (2002), no. 1, 7–26. MR-1880267

[73] ______, *Dimensionality reduction in extremal problems for moments of linear combinations of vectors with random coefficients*, Stochastic inequalities and applications, Progr. Probab., vol. 56, Birkhäuser, Basel, 2003, pp. 169–185. MR-2073433

[74] ______, *Binomial upper bounds on generalized moments and tail probabilities of (super)martingales with differences bounded from above*, High dimensional probability, IMS Lecture Notes Monogr. Ser., vol. 51, Inst. Math. Statist., Beachwood, OH, 2006, pp. 33–52. MR-2387759

[75] ______, *On l'Hospital-type rules for monotonicity*, JIPAM. J. Inequal. Pure Appl. Math. **7** (2006), no. 2, Article 40, 19 pp. (electronic). MR-2221321

[76] ______, *On normal domination of (super)martingales*, Electron. J. Probab. **11** (2006), no. 39, 1049–1070. MR-2268536

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

[78] ______, *Toward the best constant factor for the Rademacher-Gaussian tail comparison*, ESAIM Probab. Stat. **11** (2007), 412–426. MR-2339301

[79] ______, *On inequalities for sums of bounded random variables*, J. Math. Inequal. **2** (2008), no. 1, 1–7. MR-2453629

[80] S. Portnoy, *Private communication, 1991*.

[81] Martin Raič, *CLT-related large deviation bounds based on Stein's method*, Adv. in Appl. Probab. **39** (2007), no. 3, 731–752. MR-2357379

[82] Moshe Shaked and J. George Shanthikumar, *Stochastic orders*, Springer Series in Statistics, Springer, New York, 2007. MR-2265633

[83] I. G. Shevtsova, *Refinement of estimates for the rate of convergence in Lyapunov's theorem*, Dokl. Akad. Nauk **435** (2010), no. 1, 26–28. MR-2790498

[84] Michel Talagrand, *The missing factor in Hoeffding's inequalities*, Ann. Inst. H. Poincaré Probab. Statist. **31** (1995), no. 4, 689–702. MR-1355613

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

[86] Joel A. Tropp, *User-friendly tail bounds for sums of random matrices (preprint), arXiv:1004.4389v7 [math.PR]*.

[87] I. Tyurin, *New estimates of the convergence rate in the Lyapunov theorem (preprint, arXiv:0912.0726v1 [math.PR])*.

[88] Sara A. van de Geer, *On non-asymptotic bounds for estimation in generalized linear models with highly correlated design*, Asymptotics: particles, processes and inverse problems, IMS Lecture Notes Monogr. Ser., vol. 55, Inst. Math. Statist., Beachwood, OH, 2007, pp. 121–134. MR-2459935

[89] Mark Veraar, *A note on optimal probability lower bounds for centered random variables*, Colloq. Math. **113** (2008), no. 2, 231–240. MR-2425084

[90] ______, *On Khintchine inequalities with a weight*, Proc. Amer. Math. Soc. **138** (2010), no. 11, 4119–4121. MR-2679633

[91] Vladimir Vinogradov, *Refined large deviation limit theorems*, Pitman Research Notes in Mathematics Series, vol. 315, Longman Scientific & Technical, Harlow, 1994. MR-1312369

[92] A. V. Zhubr, *On one extremal problem for N-cube*, To appear, 2012.

Acknowledgments. I am pleased to thank the referee for a careful reading of the paper and useful suggestions about the exposition.