

EXACT CONVERSES TO A REVERSE AM—GM INEQUALITY, WITH APPLICATIONS TO SUMS OF INDEPENDENT RANDOM VARIABLES AND (SUPER)MARTINGALES

IOSIF PINELIS

(Communicated by C. P. Niculescu)

Abstract. For every given real value of the ratio $\mu := A_X/G_X > 1$ of the arithmetic and geometric means of a positive random variable X and every real $v > 0$, exact upper bounds on the right- and left-tail probabilities $P(X/G_X \geq v)$ and $P(X/G_X \leq v)$ are obtained, in terms of μ and v . In particular, these bounds imply that $X/G_X \rightarrow 1$ in probability as $A_X/G_X \downarrow 1$. Such a result may be viewed as a converse to a reverse Jensen inequality for the strictly concave function $f = \ln$, whereas the well-known Cantelli and Chebyshev inequalities may be viewed as converses to a reverse Jensen inequality for the strictly concave quadratic function $f(x) \equiv -x^2$. As applications of the mentioned new results, improvements of the Markov, Bernstein–Chernoff, sub-Gaussian, and Bennett–Hoeffding probability inequalities are given.

1. Introduction

Let X be a positive random variable (r.v.). One can define the arithmetic and geometric means of X as follows:

$$A_X := EX \quad \text{and} \quad G_X := \exp E \ln X, \quad (1.1)$$

assuming that EX and $E \ln X$ exist and are finite.

Consider the special case when, for given positive real numbers x_1, \dots, x_n , the distribution of the r.v. X is defined by the formula

$$E f(X) = \frac{1}{n} \sum_{i=1}^n f(x_i) \quad \text{for any function } f: \mathbb{R} \rightarrow \mathbb{R}. \quad (1.2)$$

(So, in the case when the numbers x_1, \dots, x_n are pairwise distinct, any such r.v. X takes each of the values x_1, \dots, x_n with probability $\frac{1}{n}$.) In this case,

$$A_X = EX = \frac{x_1 + \dots + x_n}{n} \quad \text{and} \quad G_X = \exp E \ln X = \sqrt[n]{x_1 \cdots x_n}. \quad (1.3)$$

Mathematics subject classification (2010): 26D15, 60E15.

Keywords and phrases: Arithmetic mean, geometric mean, random variables, inequalities, exact bounds, Jensen inequality, reverse Jensen inequality, converse to a reverse Jensen inequality, duality, Markov inequality, Bernstein–Chernoff inequality, Bennett–Hoeffding inequality.

Thus, the definitions (1.1) of the arithmetic and geometric means of a r.v. X generalize the usual definitions of the arithmetic and geometric means of finitely many positive real numbers.

Since any bounded positive r.v. can be approximated in distribution by uniformly bounded r.v.'s each taking finitely many positive real values with equal probabilities, the exact bounds to be stated in Theorem 2.1 will remain exact in an appropriate sense if one considers only the r.v.'s with such discrete uniform distributions.

The arithmetic mean–geometric mean (AM–GM) inequality

$$A_X \geq G_X \tag{1.4}$$

is a special case (with $f = \ln$) of Jensen's inequality

$$f(EX) \geq Ef(X) \tag{1.5}$$

for concave functions f .

Clearly, if the r.v. X is constant almost surely (a.s.) – that is, if $P(X = c) = 1$ for some real $c > 0$, then the Jensen inequality (1.5) and, in particular, the AM–GM inequality (1.4) turn into the equalities. Therefore, one may expect that, if the r.v. X is close to a constant in some sense, then both sides of the Jensen inequality will be close to each other and, in particular, the arithmetic and geometric means of the r.v. X will be close to each other.

There are indeed a large number of theorems in this vein, called *reverse Jensen inequalities*; see e.g. [3]. Usually, in such theorems the condition of X being close to a constant is that the values of X are in a bounded interval $[m_X, M_X]$, which latter may be thought of as small, with the conclusion that the difference $f(EX) - Ef(X)$ between the left- and right-hand sides of the Jensen inequality (1.5) is small if the interval $[m_X, M_X]$ is small. Somewhat related results were obtained in [10].

Note further that, if the function f is strictly concave, then the equality in (1.5) implies that the r.v. X is a.s. constant. Therefore, it appears natural to inquire whether statements of the following form hold: If the two sides of the Jensen inequality (1.5) with a strictly concave function f are close to each other, then the r.v. X is close to a constant in some sense. Such a statement may be referred to as a *converse to a reverse Jensen inequality*.

Converses to reverse Jensen inequalities are very well known and very widely used in the case when $f(x) \equiv -x^2$. Then the difference between the left- and right-hand sides of (1.5) is $\sigma^2 := \text{Var} X$, the variance of X . In this case, one has Cantelli's inequality

$$P(X - \mu \geq \varepsilon) \vee P(X - \mu \leq -\varepsilon) \leq \frac{\sigma^2}{\sigma^2 + \varepsilon^2} \tag{1.6}$$

and Chebyshev's inequality

$$P(|X - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2} \tag{1.7}$$

for all real $\varepsilon > 0$, where $\mu := EX \in \mathbb{R}$ and $a \vee b := \max(a, b)$. The Cantelli and Chebyshev bounds are exact in their terms. In particular, (1.6) turns into the equality

when $P(X = \mu + \varepsilon) = \varepsilon^2/(\sigma^2 + \varepsilon^2) = 1 - P(X = \mu - \sigma^2/\varepsilon)$ or when $P(X = \mu - \varepsilon) = \varepsilon^2/(\sigma^2 + \varepsilon^2) = 1 - P(X = \mu + \sigma^2/\varepsilon)$, whereas (1.7) turns into the equality when $P(X = \mu + \varepsilon) = P(X = \mu - \varepsilon) = 1/2$.

So, for any given real $\varepsilon > 0$, if $f(x) \equiv -x^2$ and the difference σ^2 between the left- and right-hand sides of (1.5) is small enough, then X deviates from the constant μ with a however small probability. Thus, the Cantelli and Chebyshev inequalities are indeed converses to a reverse Jensen inequality for $f(x) \equiv -x^2$.

In this paper, we shall provide converses to reverse Jensen inequalities for $f = \ln$, that is, converses to reverse AM–GM inequalities. This case appears to be the next in importance after the Chebyshev–Cantelli “quadratic” case of $f(x) \equiv -x^2$ – see the applications to the so-called exponential bounds on the tails of the distributions of sums of independent r.v.’s in Section 3; here one may also note e.g. [2, Lemma 3.9]. Just as the Cantelli and Chebyshev bounds, our bounds are exact in their own terms. However, the case of $f = \ln$ is much more difficult than that of $f(x) \equiv -x^2$.

2. Basic results and discussion

The main result of this paper is as follows.

THEOREM 2.1. *Let X be a positive r.v. with finite EX and $E\ln X$. Suppose that $P(X = c) < 1$ for each real c , so that*

$$\mu := \frac{A_X}{G_X} > 1. \tag{2.1}$$

Then

(I)

$$P\left(\frac{X}{G_X} \geq v\right) \leq p_{\mu,v} := p_v := \frac{\mu - z_v}{v - z_v} \in (0, 1) \text{ for each } v \in (\mu, \infty) \tag{2.2}$$

and

$$P\left(\frac{X}{G_X} \leq v\right) \leq p_v \in (0, 1) \text{ for each } v \in (0, 1), \tag{2.3}$$

where, for each $v \in (\mu, \infty)$, $z_v = z_{\mu,v}$ is the only root $z \in (0, 1)$ of the equation

$$F(z) := F_{\mu,v}(z) := (v - \mu) \ln z + (\mu - z) \ln v = 0 \tag{2.4}$$

and, for each $v \in (0, 1)$, $z_v = z_{\mu,v}$ is the only root $z \in (\mu, \infty)$ of equation (2.4).

(II) *For each $v \in (\mu, \infty)$ and for each $v \in (0, 1)$, the upper bound p_v in the corresponding inequalities in (2.2) and (2.3) is exact, as it is attained when*

$$P(X = v) = p_v = 1 - P(X = z_v), \tag{2.5}$$

and for such a r.v. X the condition $A_X/G_X = \mu$ holds – cf. (2.1).

(III) We have

$$\rho_\mu(v) := \sup_{A_X/G_X=\mu} P\left(\frac{X}{G_X} \geq v\right) = 1 \quad \text{for each } v \in (-\infty, \mu] \quad (2.6)$$

and

$$\lambda_\mu(v) := \sup_{A_X/G_X=\mu} P\left(\frac{X}{G_X} \leq v\right) = 1 \quad \text{for each } v \in [1, \infty), \quad (2.7)$$

where $\sup_{A_X/G_X=\mu}$ denotes the supremum over all positive r.v.'s X with finite EX and $E\ln X$ and with $A_X/G_X = \mu$. In particular, for each $v \in [1, \mu]$, the exact upper bound on either one of the two tail probabilities, $P\left(\frac{X}{G_X} \geq v\right)$ and $P\left(\frac{X}{G_X} \leq v\right)$, is 1; it is not attained, though.

(IV) One also has the following simple (but not exact) upper bounds on $P\left(\frac{X}{G_X} \geq v\right)$ and $P\left(\frac{X}{G_X} \leq v\right)$:

$$P\left(\frac{X}{G_X} \geq v\right) \leq q_{\mu,v} := q_v := \min\left(1, \frac{\mu - 1}{v - 1 - \ln v}\right) \quad \text{for each } v \in [1, \infty) \quad (2.8)$$

(with $q_1 := 1$) and

$$P\left(\frac{X}{G_X} \leq v\right) \leq q_v \quad \text{for each } v \in (0, \mu]. \quad (2.9)$$

(V) The condition $\mu := A_X/G_X$ in (2.1) can be replaced by the $A_X/G_X \leq \mu$.

REMARK 2.2. Part (II) of Theorem 2.1 shows that, as in the cases of the Cantelli and Chebyshev inequalities, the “extreme” r.v.’s X providing the attainment in our inequalities (2.2) and (2.3) take only two values.

REMARK 2.3. Inequalities (2.8) and (2.9) imply concentration of the r.v. X near its (say) geometric mean G_X when the arithmetic mean A_X is close to G_X . More precisely, we have $X/G_X \rightarrow 1$ in probability as $\mu = A_X/G_X \downarrow 1$. Thus, Theorem 2.1 indeed provides converses to the reverse Jensen inequality for $f = \ln$.

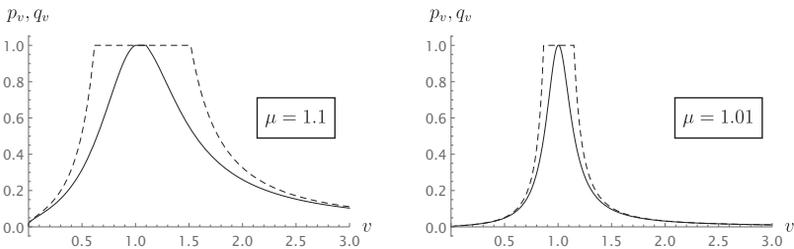


Figure 1: Graphs $\{(v, p_v) : 0 < v < 3\}$ (solid) and $\{(v, q_v) : 0 < v < 3\}$ (dashed) for $\mu = 1.1$ (left) and $\mu = 1.01$ (right), with $p_v := 1$ for $v \in [1, \mu]$.

Remark 2.3 is illustrated in Figure 1.

At least in the case when the distribution of the r.v. X/G_X is highly concentrated (that is, when μ is close to 1), the simple bound q_v on the tails of the distribution of the r.v. X/G_X is not too far from the exact bound p_v when v is somewhat close to 1 but $q_v < 1$ (so that v is not too close to 1). More precisely, we have the following proposition:

PROPOSITION 2.4. *Suppose that μ and v both go to 1 in any way such that $v \in (\mu, \infty) \cup (0, 1)$ and q_v is less than 1. Then*

$$p_v - \frac{q_v}{1 + q_v} \rightarrow 0. \tag{2.10}$$

PROPOSITION 2.5. *For $v \in (\mu, \infty) \cup (0, 1)$, one has the following expression for the root z_v of equation (2.4):*

$$z_v = z_{\mu,v} = -\tilde{z}_v W_{\kappa(v)} \left(-\frac{e^{-\mu/\tilde{z}_v}}{\tilde{z}_v} \right), \tag{2.11}$$

where

$$\tilde{z}_v := \frac{v - \mu}{\ln v}, \tag{2.12}$$

$$\kappa(v) := \begin{cases} 0 & \text{if } v \in (\mu, \infty), \\ -1 & \text{if } v \in (0, 1), \end{cases}$$

and W_k is the k th branch of Lambert’s W function [4], so that

(i) for all $t \in (-1, \infty)$ and $u \in (-1/e, \infty)$, we have $te^t = u \iff t = W_0(u)$;

(ii) for all $t \in (-\infty, -1)$ and $u \in (-1/e, 0)$, we have $te^t = u \iff t = W_{-1}(u)$.

3. Applications: Improvements of Markov’s bound and exponential bounds on the tails of the distributions of sums of independent r.v.’s and (super)martingales

3.1. Improvements of the Markov bound and of the Bernstein–Chernoff bound

By Markov’s inequality, with μ as in Theorem 2.1,

$$P \left(\frac{X}{G_X} \geq v \right) \leq \frac{\mu}{v} \tag{3.1}$$

for all real $v > 0$ (this inequality is nontrivial only if $v > \mu$).

The bound $p_{\mu,v} = \frac{\mu - z_v}{v - z_v}$ in (2.2) is a (best possible) improvement of the Markov bound $\frac{\mu}{v}$ in (3.1) – because $z_v (< 1) < \mu < v$. Even though Markov’s inequality is well-known (and easy to see) to be exact in its terms, the just mentioned improvement

has been possible by taking into account that the geometric mean of the r.v. X/G_X is 1. This improvement over Markov’s inequality may be dramatic in some cases. Indeed, when e.g. $\mu(> 1)$ is close to 1 while $\nu(> 1)$ is not close to 1, then even the suboptimal bound $\frac{\mu - 1}{\nu - 1 - \ln \nu}$ in (2.8) will be much less than the Markov bound $\frac{\mu}{\nu}$. Similarly, inequality (2.3) is a best possible, and in some settings dramatic, improvement of the corresponding left-tail Markov inequality.

Take now any r.v. Y with

$$EY = 0, \tag{3.2}$$

any real number y , and any positive real number λ . The so-called Bernstein–Chernoff inequality

$$P(Y \geq y) \leq \frac{E e^{\lambda Y}}{e^{\lambda y}} \tag{3.3}$$

is a particular case of Markov’s inequality (3.1), with

$$X := e^{\lambda Y} \quad \text{and} \quad \nu := e^{\lambda y}. \tag{3.4}$$

Also, the condition (3.2) implies that here

$$G_X = 1.$$

Actually, the Bernstein–Chernoff inequality (3.3) is, not only a special case of Markov’s inequality (3.1), but of course also a restatement of (3.1). In particular, just as Markov’s inequality (3.1) does not take into account the fact that the geometric mean of X/G_X is 1, the Bernstein–Chernoff inequality (3.3) does not take condition (3.2) into account.

Therefore, one can use Theorem 2.1 to improve, not only Markov’s inequality (3.1), but also its equivalent, the Bernstein–Chernoff inequality (3.3).

When the r.v. Y has an additional structure, one can obtain an upper bound $B(\lambda)$ on $E e^{\lambda Y}$, and then $\inf_{\lambda \geq 0} e^{-\lambda y} B(\lambda)$ will be an upper bound – referred to as an exponential bound – on the tail probability $P(Y \geq y)$. A general approach to obtaining best possible exponential bounds of this kind, along with a number of specific results, in the case when Y is the sum of independent r.v.’s was presented in [12]. Details on what has been said in this paragraph are provided in the following two subsections.

3.2. Improvements of the exponential bound in the sub-Gaussian case

Suppose that

$$Y = Y_1 + \dots + Y_n, \tag{3.5}$$

where Y_1, \dots, Y_n are independent zero-mean r.v.’s.

In this subsection, we will consider the particularly simple case when the Y_i ’s are sub-Gaussian, that is, when

$$E e^{\lambda Y_i} \leq e^{\lambda^2 \sigma_i^2 / 2} \tag{3.6}$$

for some positive real numbers $\sigma_1, \dots, \sigma_n$, all $i \in [n] := \{1, \dots, n\}$, and real $\lambda \geq 0$. If $Y_i \sim N(0, \sigma_i^2)$ for all $i \in [n]$, then the sub-Gaussianity condition (3.6) holds with the

equality sign. Also, for instance, (3.6) holds when $|Y_i| \leq \sigma_i$ for all $i \in [n]$; cf. e.g. [6, inequality (4.16)].

The constants σ_i^2 in (3.6) are referred to as (obviously, never unique) sub-Gaussian proxy variances of the corresponding r.v.'s Y_i . Clearly then,

$$\sigma^2 := \sigma_1^2 + \dots + \sigma_n^2 \tag{3.7}$$

is a sub-Gaussian proxy variance of the sum Y :

$$\mathbb{E} e^{\lambda Y} \leq e^{\lambda^2 \sigma^2 / 2} \tag{3.8}$$

for all real $\lambda \geq 0$.

Take any real $y \geq 0$. Then, by (3.3),

$$\mathbb{P}(Y \geq y) \leq \inf_{\lambda \geq 0} \frac{e^{\lambda^2 \sigma^2 / 2}}{e^{\lambda y}} = \frac{e^{\lambda_y^2 \sigma^2 / 2}}{e^{\lambda_y y}} = P_1(t) := e^{-t^2 / 2}, \tag{3.9}$$

where

$$\lambda_y := y / \sigma^2, \quad t := y / \sigma,$$

and $\sigma := \sqrt{\sigma^2}$.

Using Theorem 2.1, one can immediately improve the upper bound $e^{-t^2 / 2}$ on $\mathbb{P}(Y \geq y)$ in (3.9):

PROPOSITION 3.1. *For all real $y > 0$,*

$$\mathbb{P}(Y \geq y) \leq P_2(t) := p_{\mu_t, \nu_t} = \frac{\mu_t - z_{\mu_t, \nu_t}}{\nu_t - z_{\mu_t, \nu_t}} < \frac{\mu_t}{\nu_t} = e^{-t^2 / 2} = P_1(t), \tag{3.10}$$

where

$$\mu_t := e^{\lambda_y^2 \sigma^2 / 2} = e^{y^2 / (2\sigma^2)} = e^{t^2 / 2}, \quad \nu_t := e^{\lambda_y y} = e^{y^2 / \sigma^2} = e^{t^2}, \tag{3.11}$$

and $z_{\mu, \nu}$ is as defined in part (I) of Theorem 2.1 or, equivalently, as in formula (2.11).

Concerning the inequality in (3.10), recall the reasoning in the second paragraph of Subsection 3.1.

So, the bound $P_2(t)$ on $\mathbb{P}(Y \geq y)$ in (3.10) improves the bound $P_1(t)$ in (3.9) for all real $t > 0$ or, equivalently, for all real $y > 0$. To get the bound $P_2(t)$, we borrowed the minimizer λ_y of the bound $\frac{e^{\lambda^2 \sigma^2 / 2}}{e^{\lambda y}}$ on $\mathbb{P}(Y \geq y)$ and used λ_y in the definitions of μ_t and ν_t in (3.11). While this choice of λ is optimal for the Markov bound $\frac{e^{\lambda^2 \sigma^2 / 2}}{e^{\lambda y}}$, it will not be optimal for the better bound of the form $p_{\mu, \nu}$ based on Theorem 2.1.

So, we can improve the bound $P_2(t)$ on $\mathbb{P}(Y \geq y)$ – and thus further improve the bound $P_1(t)$ – by avoiding the mentioned borrowing, as follows:

PROPOSITION 3.2. For all real $y > 0$,

$$P(Y \geq y) \leq P_{\text{opt}}(y) := P_{\text{opt}}(\sigma, y) := \inf_{\lambda > 0} p_{\mu_\lambda, v_y}(\lambda), \tag{3.12}$$

where

$$\mu_\lambda := e^{\lambda^2 \sigma^2 / 2} \quad \text{and} \quad v_y(\lambda) := e^{\lambda y}. \tag{3.13}$$

The drawback of the optimal bound $P_{\text{opt}}(y)$ is that its expression in (3.12) is implicit; also, in distinction with the simpler bounds $P_1(t)$ and $P_2(t)$, $P_{\text{opt}}(y) = P_{\text{opt}}(\sigma, y)$ will depend on σ, y not only through the simple ratio $t = y/\sigma$.

On the other hand, clearly we can use the simple bound $q_{\mu, v}$ in (2.8) to immediately get the following:

PROPOSITION 3.3. For all real $y > 0$,

$$P(Y \geq y) \leq P_3(t) := q_{\mu_t, v_t} = \min\left(1, \frac{\mu_t - 1}{v_t - 1 - \ln v_t}\right) = \min\left(1, \frac{e^{t^2/2} - 1}{e^{t^2} - 1 - t^2}\right), \tag{3.14}$$

where μ_t and v_t are as in (3.11).

We see that the bound $P_3(t)$ is quite explicit and almost as simple as the bound $P_1(t) = e^{-t^2/2}$ in (3.9). Moreover, a simple algebra shows that $P_3(t) < P_1(t)$ (for a real $t > 0$) if and only if $1 + t^2 < e^{t^2/2}$, that is, if and only if

$$t > t_* := \sqrt{-2W_{-1}(-1/(2\sqrt{e})) - 1} = 1.585\dots,$$

where, as in Proposition 2.5, W_k denotes the k th branch of Lambert’s W function. Also, $P_1(t_*) = P_3(t_*) = 0.284\dots$, which is substantially greater than commonly used values of the level of significance in statistical testing. So, the bound $P_3(t)$ is an improvement of the bound $P_1(t)$ for values of t relevant in statistics.

(Parts of) the graphs of the ratios of the bounds $P_2(t)$ in (3.10), $P_3(t)$ in (3.14), and $P_{\text{opt}}(\sigma, \sigma t)$ in (3.12) with $\sigma = 6$ to the baseline sub-Gaussian bound $P_1(t)$ in (3.9) are shown in Figure 2.

3.3. Improvements of the Bennett–Hoeffding exponential bound

It is seen from Figure 2 that the new bounds P_2 and P_3 , and even the optimal bound P_{opt} , provide only relatively limited improvements over the baseline sub-Gaussian bound P_1 .

In this subsection, it will be shown that the corresponding improvements over the well-known and widely used Bennett–Hoeffding exponential bound can be arbitrarily large (in a relative sense) in certain settings.

Here it is still assumed that (3.5) holds, with independent zero-mean r.v.’s Y_1, \dots, Y_n . However, instead of the sub-Gaussian condition (3.6), we now assume that

$$Y_i \leq b$$

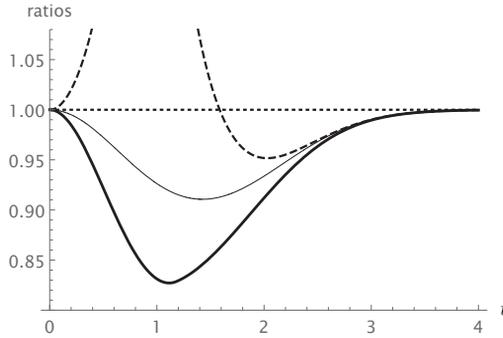


Figure 2: Graphs $\{(t, P_2(t)/P_1(t)) : 0 < t < 4\}$ (thin), $\{(t, P_3(t)/P_1(t)) : 0 < t < 4, P_3(t)/P_1(t) < 1.08\}$ (dashed), and $\{(t, P_{\text{opt}}(6, 6t)/P_1(t)) : 0 < t < 4\}$ (thick).

for some real $b > 0$ and all $i \in [n]$. We will also use notation (3.7), but now with

$$\sigma_i^2 := \text{Var } Y_i = \mathbb{E} Y_i^2,$$

rather with σ_i^2 being a sub-Gaussian proxy variance of Y_i .

It follows e.g. from [12, Theorem 2] that, under the above conditions on Y, Y_1, \dots, Y_n , the best possible upper bound on $\mathbb{E} e^{\lambda Y}$ is given by the inequality

$$\mathbb{E} e^{\lambda Y} \leq \mu_{\sigma, b}(\lambda) := \exp \left\{ \frac{\sigma^2}{b^2} (e^{\lambda b} - 1 - \lambda b) \right\},$$

for each real $\lambda \geq 0$. Thus, we have the Markov bound on $\mathbb{P}(Y \geq y)$:

$$\mathbb{P}(Y \geq y) \leq \frac{\mu_{\sigma, b}(\lambda)}{v_y(\lambda)},$$

where $v_y(\lambda) = e^{\lambda y}$, as in (3.13). Minimizing the latter bound on $\mathbb{P}(Y \geq y)$ in $\lambda \geq 0$, we get

$$\mathbb{P}(Y \geq y) \leq P_{\text{BH}}(y, \sigma, b) := \frac{\mu_{\sigma, b}(\lambda_{y, \sigma, b})}{v_y(\lambda_{y, \sigma, b})}, \tag{3.15}$$

where

$$\lambda_{y, \sigma, b} := \frac{1}{b} \ln \left(1 + \frac{by}{\sigma^2} \right),$$

so that

$$\mu_{\sigma, b}(\lambda_{y, \sigma, b}) = \exp \left\{ \frac{y}{b} \left[1 - \frac{\sigma^2}{by} \ln \left(1 + \frac{by}{\sigma^2} \right) \right] \right\} \tag{3.16}$$

and

$$v_y(\lambda_{y, \sigma, b}) = \exp \left\{ \frac{y}{b} \ln \left(1 + \frac{by}{\sigma^2} \right) \right\}. \tag{3.17}$$

The bound $P_{\text{BH}}(y, \sigma, b)$ on $\mathbb{P}(Y \geq y)$ in (3.15) is the famous and widely used Bennett [1]–Hoeffding [6] bound.

Since the Bennett–Hoeffding (BH) bound is a species of the Markov bound, it can be improved using Theorem 2.1, just as the sub-Gaussian bound was improved in Propositions 3.1, 3.2, and 3.3 of Subsection 3.2. Here we will only consider the simplest of such improvements of the BH bound, based on (2.8) (cf. (3.14)), even though this improvement is not the best possible:

PROPOSITION 3.4. *For all real $y > 0$,*

$$P(Y \geq y) \leq P_{\text{BH},1}(y, \sigma, b) := \min \left(1, \frac{\mu_{\sigma,b}(\lambda_{y,\sigma,b}) - 1}{v_y(\lambda_{y,\sigma,b}) - 1 - \ln v_y(\lambda_{y,\sigma,b})} \right), \tag{3.18}$$

where $\mu_{\sigma,b}(\lambda_{y,\sigma,b})$ and $v_y(\lambda_{y,\sigma,b})$ are as in (3.16) and (3.17).

Suppose now that

$$\frac{y}{b} = c \quad \text{and} \quad \frac{\sigma^2}{by} \leq e^{-C/c},$$

where c and C are positive real numbers. Then $\mu_{\sigma,b}(\lambda_{y,\sigma,b}) \leq \exp\{\frac{y}{b}\} = e^c$ and $v_y(\lambda_{y,\sigma,b}) \geq e^C$, so that

$$\frac{P_{\text{BH},1}(y, \sigma, b)}{P_{\text{BH}}(y, \sigma, b)} \leq \frac{e^c - 1}{e^c} \frac{e^C}{e^C - 1 - C},$$

because $\frac{\mu-1}{\mu}$ is increasing in $\mu > 0$ and $\frac{v}{v-1-\ln v}$ is decreasing in $v > 1$. So, the ratio $P_{\text{BH},1}(y, \sigma, b)/P_{\text{BH}}(y, \sigma, b)$ of the improved BH bound $P_{\text{BH},1}(y, \sigma, b)$ to the original BH bound $P_{\text{BH}}(y, \sigma, b)$ can be however small if c is small enough and C is bounded away from 0. Conditions with C not small and c not large arise in settings when possibly heavy tails of the distributions of the Y_i 's must be appropriately truncated – see e.g. [5, 11].

The results of Subsections 3.2 and 3.3 can be extended to the case when the Y_i 's are (super)martingale-differences; cf. e.g. [8, Section 8].

4. Proofs

Proof of Theorem 2.1. This proof is implicitly based on a duality argument; see e.g. [7, 9].

Note that the probabilities $P(\frac{X}{G_X} \geq v)$ and $P(\frac{X}{G_X} \leq v)$ will not change if we replace there X by X/u , for any positive real u . So, without loss of generality we may and shall assume that $G_X = 1$, that is,

$$E \ln X = 0, \tag{4.1}$$

so that the probabilities $P(\frac{X}{G_X} \geq v)$ and $P(\frac{X}{G_X} \leq v)$ become simply $P(X \geq v)$ and $P(X \leq v)$.

Take now any positive real v and any positive real $z \neq v$, and for all real $x > 0$ let

$$g(x) := ax - b \ln x + c,$$

where

$$a := a(z) := \frac{1/v}{h(r)}, \quad b := b(z) := az, \quad c := c(z) := az \ln \frac{z}{e},$$

$$h(r) := 1 - r + r \ln r, \quad r := z/v.$$

Note that the function h is convex on $(0, \infty)$, with $h(1) = 0 = h'(1)$. So, $h > 0$ on $(0, \infty) \setminus \{1\}$ and hence $a > 0$ and $b > 0$. Therefore, the function g is convex on $(0, \infty)$. Moreover,

$$g(z) = g'(z) = 0 \quad \text{and} \quad g(v) = 1.$$

So, if $0 < z < v$, then $g(x) \geq I\{x \geq v\}$ for all real $x > 0$, where $I\{\cdot\}$ denotes the indicator. Hence, in view of (4.1),

$$P(X \geq v) \leq E g(X) = aEX - bE \ln X + c = a\mu + c$$

$$= R_z(v) := \frac{\mu - z + z \ln z}{v - z + z \ln z - z \ln v} \quad \text{if} \quad 0 < z < v. \tag{4.2}$$

Similarly, if $0 < v < z$, then $g(x) \geq I\{x \leq v\}$ for all real $x > 0$, whence

$$P(X \leq v) \leq R_z(v) \quad \text{if} \quad 0 < v < z. \tag{4.3}$$

Recalling the conditions $\mu > 1$ in (2.1) and $q_1 = 1$ in the statement of part (IV) of Theorem 2.1, as well as the fact that no probability can exceed 1, and then substituting 1 for z in (4.2) and (4.3), we get part (IV) of Theorem 2.1.

To prove part (I) of Theorem 2.1, consider separately the two cases: $v \in (\mu, \infty)$ and $v \in (0, 1)$.

If $v \in (\mu, \infty)$, then the function $F: (0, \infty) \rightarrow \mathbb{R}$ is concave, with $F(0+) = -\infty < 0$ and $F(1) = (\mu - 1) \ln v > 0$ (since $v > \mu > 1$). So, indeed there is exactly one root $z = z_v \in (0, 1)$ of equation (2.4). Next, from the equality $F(z_v) = 0$ we get $\ln z_v = (\mu - z_v) \ln(v) / (\mu - v)$. Substituting this expression for $\ln z_v$ into the expression for $R_z(v)$ in (4.2) and recalling the definition of p_v in (2.2), we get

$$R_{z_v}(v) = p_v. \tag{4.4}$$

Therefore and because here

$$0 < z_v < 1 < \mu < v, \tag{4.5}$$

we see that the inequality in (2.2) follows by (4.2).

The case $v \in (0, 1)$ is similar (to the case $v \in (\mu, \infty)$). Indeed, if $v \in (0, 1)$, then the function $F: (0, \infty) \rightarrow \mathbb{R}$ is convex, with $F(\mu) = (v - \mu) \ln \mu < 0$ (since $\mu > 1$) and $F(\infty-) = \infty > 0$ (since $v > \mu > 1$). So, indeed there is exactly one root $z = z_v \in (\mu, \infty)$ of equation (2.4). Of course, equality (4.4) holds for $v \in (0, 1)$ as well. Therefore and because here

$$0 < v < 1 < \mu < z_v, \tag{4.6}$$

we see that the inequality in (2.3) follows by (4.3).

Also, in view of (4.5) and (4.6), in either one of the cases $v \in (\mu, \infty)$ and $v \in (0, 1)$, μ is strictly between v and z_v , whence $p_v = \frac{\mu - z_v}{v - z_v} \in (0, 1)$.

Thus, part (I) of Theorem 2.1 is proved.

To prove part (II) of Theorem 2.1, note first that, in view of the just proved inclusion $p_v \in (0, 1)$, there does exist a r.v. X as in (2.5). For such a r.v. X , we have $A_X = EX = \mu$ and $\ln G_X = E \ln X = \frac{F(z_v)}{v - z_v} = 0$, by the definition of z_v , so that the condition $A_X/G_X = \mu$ holds. Also, again in view of (4.5) and (4.6), we have $z_v < v$ if $v \in (\mu, \infty)$, and $z_v > v$ if $v \in (0, 1)$. So, for any r.v. X as in (2.5), the inequalities in (2.2) and (2.3) turn into the equalities; that is, the upper bound p_v in the inequalities in (2.2) and (2.3) is exact, as it is attained for X as in (2.5). This proves part (II) of Theorem 2.1.

Next, consider part (III) of Theorem 2.1. Note that the function ρ_μ is nonincreasing on \mathbb{R} and $\rho_\mu \leq 1$ on \mathbb{R} . Also, by part (II) of Theorem 2.1 and the definition of p_v in (2.2), for $v \in (\mu, \infty)$ we have $\rho_\mu(v) = p_v \rightarrow 1$ as $v \downarrow \mu$, because $\mu > 1$ and $z_v < 1$. So, $\rho_\mu(\mu+) = 1$ and hence $1 \geq \rho_\mu(v) \geq \rho_\mu(\mu+) = 1$ for all $v \in (-\infty, \mu]$. This proves (2.6).

Further, the function λ_μ is nondecreasing on \mathbb{R} and $\lambda_\mu \leq 1$ on \mathbb{R} . Also, by part (II) of Theorem 2.1, for $v \in (0, 1)$ we have $\lambda_\mu(v) = p_v$. Let now $v \uparrow 1$. Then $|(v - \mu) \ln z_v|$ is bounded away from 0, because $\mu > 1$ and $z_v > \mu$. So, in view of (2.4) and the condition $F(z_v) = 0$, $|(\mu - z_v) \ln v|$ is bounded away from 0. So, for $v \uparrow 1$ we have $z_v \rightarrow \infty$ and hence $\lambda_\mu(v) = p_v \rightarrow 1$, again by the definition of p_v in (2.2). Therefore, $\lambda_\mu(1-) = 1$ and hence $1 \geq \lambda_\mu(v) \geq \lambda_\mu(1-) = 1$ for all $v \in [1, \infty)$. This proves (2.7).

Concerning the last, non-attainment clause in part (III) of Theorem 2.1: If $P(\frac{X}{G_X} \geq v) = 1$ for some $v \in [1, \mu]$, then $P(\frac{X}{G_X} \geq 1) = 1$, which implies that $P(X = G_X) = 1$, which contradicts the inequality in (2.1). Similarly, if $P(\frac{X}{G_X} \leq v)$ for some $v \in [1, \mu]$, then $P(\frac{X}{G_X} \leq \mu) = 1$, which implies that $\frac{A_X}{G_X} = E \frac{X}{G_X} \leq \mu$, with the strict inequality $\frac{A_X}{G_X} < \mu$ (contradicting the definition of μ in (2.1)) unless $P(\frac{X}{G_X} = \mu) = 1$. But the latter equality implies $P(X = c) = 1$ for some real $c > 0$, which contradicts the inequality in (2.1) (since the function \ln is strictly concave).

Thus, for each $v \in [1, \mu]$, the exact upper bound, 1, on either one of the two tail probabilities, $P(\frac{X}{G_X} \geq v)$ and $P(\frac{X}{G_X} \leq v)$, is not attained.

Finally, concerning part (V) of Theorem 2.1: Given only the condition $A_X/G_X \leq \mu$ (which means that $EX \leq \mu$ when (4.1) is assumed), the second equality sign in (4.2) can be replaced by \leq , since $a > 0$. So, the inequality $P(X \geq v) \leq R_z(v)$ will continue to hold when $0 < z < v$. Similarly, (4.3) will continue to hold.

Theorem 2.1 is now completely proved. \square

Proof of Proposition 2.4. For brevity, let

$$p := p_v, \quad q := q_v, \quad z := z_v,$$

and then

$$\delta_\mu := \mu - 1 \downarrow 0, \quad \delta_v := v - 1 \rightarrow 0, \quad \delta_z := z - 1,$$

so that, by (2.4), (2.2), (2.8), and the condition $q_v < 1$,

$$(\delta_v - \delta_\mu) \ln(1 + \delta_z) + (\delta_\mu - \delta_z) \ln(1 + \delta_v) = 0, \tag{4.7}$$

$$p = \frac{\delta_\mu - \delta_z}{\delta_v - \delta_z}, \tag{4.8}$$

and

$$q = \frac{\delta_\mu}{\delta_v - \ln(1 + \delta_v)} \sim \frac{\delta_\mu}{\delta_v^2/2},$$

whence

$$\delta_\mu \sim q\delta_v^2/2 = o(\delta_v). \tag{4.9}$$

Therefore, $\delta_v - \delta_\mu \sim \delta_v \sim \ln(1 + \delta_v)$ and hence (4.7) implies $\ln(1 + \delta_z) \sim \delta_z - \delta_\mu$. Since $\delta_\mu \rightarrow 0$, it follows that $\delta_z \rightarrow 0$. (Otherwise, without loss of generality we have $\ln(1 + \delta_z) \sim \delta_z$, which implies $\delta_z \rightarrow 0$, since $\ln(1 + u) < u$ for all $u \in (-1, \infty) \setminus \{0\}$ and $\ln(1 + u) \not\sim u$ as $u \downarrow -1$ or $u \rightarrow \infty$.)

Now (4.7) and (4.9) yield

$$\left(\delta_v - \frac{q\delta_v^2}{2+o(1)}\right) \left(\delta_z - \frac{\delta_z^2}{2+o(1)}\right) + \left(\frac{q\delta_v^2}{2+o(1)} - \delta_z\right) \left(\delta_v - \frac{\delta_v^2}{2+o(1)}\right) = 0, \tag{4.10}$$

which simplifies to

$$\left(1 + \frac{(1-q)\delta_v}{2+o(1)}\right) \left(\delta_z - \frac{\delta_z^2}{2+o(1)}\right) + \left(\frac{q\delta_v^2}{2+o(1)} - \delta_z\right) = 0 \tag{4.11}$$

and then to

$$\delta_z^2 - (1-q)(1+o(1))\delta_v\delta_z - q(1+o(1))\delta_z^2 = 0. \tag{4.12}$$

Also, by part (I) of Theorem 2.1, $\delta_z\delta_v < 0$. So, “solving” the “quadratic” equation (4.12), we get

$$\begin{aligned} \frac{\delta_z}{\delta_v} &= \frac{(1-q)(1+o(1)) - \sqrt{(1-q)^2(1+o(1)) + 4q(1+o(1))}}{2} \\ &= \frac{(1-q)(1+o(1)) - (1+q)(1+o(1))}{2} = -q + o(1). \end{aligned}$$

Now (2.10) follows by (4.8) and (4.9). \square

The proof of Proposition 2.5 is based in part on the following lemmas.

LEMMA 4.1. *We have $\tilde{z}_v \in (0, \infty)$ for all $v \in (\mu, \infty) \cup (0, 1)$. Also,*

$$\tilde{z}_v \begin{cases} < v & \text{if } v \in (\mu, \infty), \\ > v & \text{if } v \in (0, 1). \end{cases} \tag{4.13}$$

LEMMA 4.2. For $v \in (\mu, \infty) \cup (0, 1)$,

$$H_\mu(v) := \frac{G_\mu(v)}{v-1} > 0, \tag{4.14}$$

where

$$G_\mu(v) := F(\tilde{z}_v) = \mu \ln v + (v - \mu) \ln \frac{v - \mu}{\ln v} + \mu - v,$$

with \tilde{z}_v as in (2.12).

LEMMA 4.3. We have $\tilde{z}_v \in (0, \infty)$ for all $v \in (\mu, \infty) \cup (0, 1)$. Also,

$$z_v \begin{cases} < \tilde{z}_v & \text{if } v \in (\mu, \infty), \\ > \tilde{z}_v & \text{if } v \in (0, 1). \end{cases} \tag{4.15}$$

Proof of Lemma 4.1. That $\tilde{z}_v \in (0, \infty)$ for all $v \in (\mu, \infty) \cup (0, 1)$ follows immediately from the definition of \tilde{z}_v in (2.12) and the condition $\mu > 1$. Next, for each $v \in (\mu, \infty) \cup (0, 1)$, each of the two inequalities in (4.13) can be rewritten as $l(v) > 0$, where

$$l(v) := v \ln v - v + \mu.$$

The function l is convex on $(0, \infty)$, with $l(1) = \mu - 1 > 0$ and $l'(1) = 0$. So, $l(v) > 0$ for $v \in (0, 1) \cup (1, \infty)$ and hence for $v \in (\mu, \infty) \cup (0, 1)$, which completes the proof of Lemma 4.1. \square

Proof of Lemma 4.2. Note first that the partial derivative of $G_\mu(v)$ in μ is $\ln v - \ln \tilde{z}_v$. So, by Lemma 4.1, $G_\mu(v)$ is increasing in $\mu \in [1, v]$ if $v \in (1, \infty)$ and decreasing in $\mu \in [1, \infty)$ if $v \in (0, 1)$. It follows that $G_\mu(v) > G_1(v)$ if $v \in (\mu, \infty)$ and $G_\mu(v) < G_1(v)$ if $v \in (0, 1)$.

So, to complete the proof of Lemma 4.2, it is enough to show that

$$H(v) := H_1(v) = \frac{\ln v}{v-1} + \ln \frac{v-1}{\ln v} - 1 \stackrel{(?)}{>} 0 \quad \text{if } v \in (0, 1) \cup (1, \infty). \tag{4.16}$$

We have

$$H'(v) := \frac{(v-1-\ln v)(v \ln v - v + 1)}{(v-1)^2 v \ln v},$$

which is easily seen to be of the same sign as $v-1$ for all $v \in (0, 1) \cup (1, \infty)$. So, $H(v)$ is decreasing in $v \in (0, 1)$ and increasing in $v \in (1, \infty)$. Also, $H(1+) = H(1-) = 0$. Thus, (4.16) is true, which completes the proof of Lemma 4.2. \square

Proof of Lemma 4.3. Consider first the case $v \in (\mu, \infty)$. Then, as was noted in the proof of part (I) of Theorem 2.1, the function $F: (0, \infty) \rightarrow \mathbb{R}$ is concave. Also, $F(v) = 0$ and, by the definition of z_v in part (I) of Theorem 2.1, $F(z_v) = 0$. Further, by Lemma 4.2, $F(\tilde{z}_v) > 0$. Therefore and in view of the concavity of F , \tilde{z}_v is strictly between v and z_v . But, by Lemma 4.1, here $\tilde{z}_v < v$. So, the first inequality in (4.15) is proved.

The second case, with $v \in (0, 1)$, is treated similarly. In this case, the function $F: (0, \infty) \rightarrow \mathbb{R}$ is convex and, by Lemma 4.2, $F(\tilde{z}_v) < 0$. Here we still have $F(v) = 0$ and $F(z_v) = 0$, whence again \tilde{z}_v is strictly between v and z_v . But, by Lemma 4.1, here $\tilde{z}_v > v$. So, the second inequality in (4.15) is proved as well. \square

Proof of Proposition 2.5. Take any $v \in (\mu, \infty) \cup (0, 1)$. By the definition of z_v in part (I) of Theorem 2.1, $F(z_v) = 0$, that is, $(v - \mu) \ln z_v + (\mu - z_v) \ln v = 0$. Dividing the latter equality by $v - \mu$ and recalling the definition of \tilde{z}_v in (2.12), rewrite the defining condition on z_v as

$$\ln z_v - \frac{z_v}{\tilde{z}_v} = -\frac{\mu}{\tilde{z}_v}. \quad (4.17)$$

Exponentiating both sides of (4.17) and then dividing the resulting expressions by $-\tilde{z}_v$, rewrite (4.17) as

$$t_v e^{t_v} = u_v,$$

where

$$t_v := -\frac{z_v}{\tilde{z}_v} \quad \text{and} \quad u_v := -\frac{e^{-\mu/\tilde{z}_v}}{\tilde{z}_v}.$$

Note also that $t e^t \in (-1/e, 0)$ for $t \in (-\infty, -1) \cup (-1, 0)$. So, in view of the description of the branches W_0 and W_{-1} of Lambert's W function given at the end of the statement of Proposition 2.5, it remains to check that $t_v \in (-1, 0)$ if $v \in (\mu, \infty)$ and $t_v \in (-\infty, -1)$ if $v \in (0, 1)$; but these conditions on t_v follow immediately by Lemma 4.3. Proposition 2.5 is proved. \square

REFERENCES

- [1] G. BENNETT, *Probability inequalities for the sum of independent random variables*, J. Amer. Statist. Assoc., **57** (297): 33–45, 1962.
- [2] S. M. BUCKLEY, *Estimates for operator norms on weighted spaces and reverse Jensen inequalities*, Trans. Amer. Math. Soc., **340** (1): 253–272, 1993.
- [3] I. BUDIMIR, S. S. DRAGOMIR, AND J. PEČARIĆ, *Further reverse results for Jensen's discrete inequality and applications in information theory*, JIPAM. J. Inequal. Pure Appl. Math., **2** (1): Article 5, 14, 2001.
- [4] R. M. CORLESS, G. H. GONNET, D. E. G. HARE, D. J. JEFFREY, AND D. E. KNUTH, *On the Lambert W function*, Adv. Comput. Math., **5** (4): 329–359, 1996.
- [5] C. C. HEYDE, *On large deviation problems for sums of random variables which are not attracted to the normal law*, Ann. Math. Statist., **38**: 1575–1578, 1967.
- [6] W. HOEFFDING, *Probability inequalities for sums of bounded random variables*, J. Amer. Statist. Assoc., **58**: 13–30, 1963.
- [7] J. H. B. KEMPERMAN, *On the role of duality in the theory of moments*, In Semi-infinite programming and applications (Austin, Tex., 1981), volume 215 of Lecture Notes in Econom. and Math. Systems, pages 63–92. Springer, Berlin, 1983.
- [8] I. PINELIS, *Optimum bounds for the distributions of martingales in Banach spaces*, Ann. Probab., **22** (4): 1679–1706, 1994.
- [9] I. PINELIS, *Optimal tail comparison based on comparison of moments*, In High dimensional probability (Oberwolfach, 1996), volume 43 of Progr. Probab., pages 297–314, Birkhäuser, Basel, 1998.
- [10] I. PINELIS, *Exact upper and lower bounds on the difference between the arithmetic and geometric means*, Bull. Aust. Math. Soc., **92** (1): 149–158, 2015.

- [11] I. F. PINELIS, *A problem on large deviations in a space of trajectories*, Theory Probab. Appl., **26** (1): 69–84, 1981.
- [12] I. F. PINELIS AND S. A. UTEV, *Sharp exponential estimates for sums of independent random variables*, Theory Probab. Appl., **34** (2): 340–346, 1989.

(Received February 9, 2021)

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