Annales de l'Institut Henri Poincaré - Probabilités et Statistiques
2014, Vol. 50, No. 1, 15–27
DOI: 10.1214/12-AIHP495

ANNALES DE L'INSTITUT HENRI POINCARÉ
PROBABILITÉS ET STATISTIQUES
www.imstat.org/aihp

On the Bennett–Hoeffding inequality

Iosif Pinelis[1]

Department of Mathematical Sciences, Michigan Technological University, Houghton, MI 49931, USA. E-mail: ipinelis@mtu.edu

Received 4 October 2011; accepted 24 May 2012

Abstract. The well-known Bennett–Hoeffding bound for sums of independent random variables is refined, by taking into account positive-part third moments, and at that significantly improved by using, instead of the class of all increasing exponential functions, a much larger class of generalized moment functions. The resulting bounds have certain optimality properties. The results can be extended in a standard manner to (the maximal functions of) (super)martingales. The proof of the main result relies on an apparently new method that may be referred to as infinitesimal spin-off. Parts of the proof also use the method of certificates of positivity in real algebraic geometry.

Résumé. La borne de Bennett–Hoeffding pour des sommes de variables aléatoires indépendantes est précisée, en prenant en compte la partie positive des troisièmes moments et sensiblement améliorée en utilisant, au lieu de la classe de toutes les fonctions exponentielles croissantes, une classe beaucoup plus important de fonctions de moment généralisées. Les limites qui en résultent ont certaines propriétés d'optimalité. Les résultats peuvent être étendus de manière standard pour (les fonctions maximales de) (sur)martingales. La preuve du résultat principal repose sur une méthode apparemment nouvelle. Des éléments de la preuve utilisent également la méthode des certificats de positivité de la géométrie algébrique réelle.

MSC: Primary 60E15; 60G50; secondary 60E07; 60E10; 60G42; 60G48; 60G51

Keywords: Probability inequalities; Sums of independent random variables; Martingales; Supermartingales; Upper bounds; Generalized moments; Lévy processes; Certificates of positivity; Real algebraic geometry

1. Introduction

Let $X_1, \dots, X_n$ be independent random variables (r.v.'s), with the sum $S := X_1 + \dots + X_n$, such that for some real positive constants y and σ and all i one has

$$X_i \le y, \quad \mathsf{E}X_i \le 0, \quad \text{and} \quad \sum \mathsf{E}X_i^2 \le \sigma^2. \tag{BH conds}$$

Essentially, a well-known result by Bennett [1] and Hoeffding [24] provides the exact upper bound on the exponential moments $\mathsf{E}e^{\lambda S}$ ($\lambda > 0$) under (BH conds). To quote Bennett [1]: "Much work has been carried out on [asymptotics] [...]. The majority of this work does not provide estimates for accuracy [...]." The Bennett–Hoeffding (BH) inequality provided such an estimate, and it has been used in hundreds of papers.

The BH inequality has been generalized to include cases when the X_i's are not independent and/or are not real-valued; see e.g. [8–10,13,15,17,20,25,27,28,32,37,39,41,49,50,52]. Yet, there have been few publications that present improvements even in the original case of sums of independent real-valued r.v.'s X_i – especially if one counts only the bounds that are exact in their own terms.

[1] Supported in part by NSF Grant DMS-08-05946 and NSA Grant H98230-12-1-0237.

Pinelis and Utev (PU) (1989) ([38], Theorems 2 and 6) refined the BH bound by also taking into account the sum $\sum \mathsf{E}(X_i)_+^3$ of the positive-part third moments of the X_i's; as usual, we let $x_+ := \max(0, x)$ and $x_+^\alpha := (x_+)^\alpha$. Using the upper bounds on the exponential moments $\mathsf{E}e^{\lambda S}$ and the Markov inequality, one can immediately obtain upper bounds on the tail probability $\mathsf{P}(S \ge x)$; however, such bounds, even the best possible ones, will be missing a factor on the order of $1/x$. This deficiency is caused by the fact that the class – say $\mathcal{E}$ – of all increasing exponential functions is too small.

The exponential class $\mathcal{E}$ is contained, for each $\alpha > 0$, in the much richer class $\mathcal{H}^\alpha$ of functions $f : \mathbb{R} \to \mathbb{R}$ defined as follows:

$$f \in \mathcal{H}^\alpha \quad \iff \quad \begin{array}{l} \text{for some Borel measure } \mu \ge 0 \text{ and all } u \in \mathbb{R} \text{ one has} \\ f(u) = \displaystyle\int_{-\infty}^{\infty} (u-t)_+^\alpha \mu(\mathrm{d}t). \end{array} \tag{$\mathcal{H}^\alpha$}$$

It is easy to see ([43], Proposition 1(ii)) that $0 < \beta < \alpha$ implies $\mathcal{H}^\alpha \subseteq \mathcal{H}^\beta$. Moreover ([45], Proposition 1.1), for natural α, one has $f \in \mathcal{H}^\alpha$ if and only if f has finite derivatives $f^{(0)} := f$, $f^{(1)} := f', \dots, f^{(\alpha-1)}$ on $\mathbb{R}$ such that $f^{(\alpha-1)}$ is convex on $\mathbb{R}$ and $f^{(j)}(-\infty+) = 0$ for $j = 0, 1, \dots, \alpha - 1$.

A class of moment functions similar to $\mathcal{H}^3$ was introduced by Eaton [18,19] (who considered X_i's bounded in absolute value, taking into account only the first moments). This allowed one to restore the missing factor $1/x$. For another approach, see Talagrand [51]. Eaton's idea was further developed in [41,42]. In particular, Pinelis [42] provided a general device allowing one to extract the optimal tail comparison inequality from a generalized moment comparison. Under (BH conds), Bentkus (2002, 2004) [2,4] obtained the exact upper bound (which we shall denote by Be) on the moments $\mathsf{E}f(S)$ for the moment functions f in the class $\mathcal{H}^2$, in place of the class $\mathcal{E}$ of all increasing exponential functions – for a fixed n in [2] and for a freely varying n in [4]; similar results for (continuous-time) martingales that are stochastic integrals were obtained by Klein, Ma and Privault [28].

In this paper, we shall extend the mentioned PU exponential bounds from the exponential class $\mathcal{E}$ to the class $\mathcal{H}^3$ of moment functions of S. The relations between the four related bounds – BH, PU, Be, and the bound – say Pin – presented in this paper are illustrated by the following diagram:

$$\begin{tikzcd} \text{BH} \arrow[r, "r"] \arrow[d, "e"] & \text{PU} \arrow[d, "e"] \\ \text{Be} \arrow[r, "pr"] & \text{Pin} \end{tikzcd}$$

In particular, it shows PU to be a refinement (denoted by r) of BH. This refinement is also an improvement, as is obviously the case with any refinement that is exact in its own terms; indeed, the more specific the terms, the better the best possible result is.

The relation of Pin with Be is almost parallel to that of PU with BH. However, the refinement, and hence the improvement, here are only partial (pr) – because the class $\mathcal{H}^3$ (corresponding to Pin) is a bit smaller than $\mathcal{H}^2$ (corresponding to Be), even though, according to [35], Propositions 2.5 and 2.12, $\mathcal{H}^3$ is essentially the largest possible class for Pin, just as $\mathcal{H}^2$ is for Be.

The relations of Be to BH, and of Pin to PU, are pure extensions (e), due to using the larger classes $\mathcal{H}^\alpha$ in place of the smaller class $\mathcal{E}$ of exponential moment functions. Therefore, when applied in order to obtain upper bounds $\inf \mathsf{E}f(S)/f(x)$ on the tail $\mathsf{P}(S \ge x)$, with the inf taken over the corresponding class of moment functions f, these extensions result in improvements.

Various forms of comparison between the bounds BH, PU, Be, and Pin – inequalities, asymptotics, numerics, and graphics are systematically presented in the detailed version of this paper [35], along with other related results.

Compared with the preceding results of the Bennett–Hoeffding type, the results in the present paper require proofs at a significantly higher level of difficulty, with novel ideas. The main idea, described in the proof of Lemma 3.6 on page 23, may be referred to as that of *infinitesimal spin-off*.

A common feature of all the Bennett–Hoeffding type bounds is that they pertain to sums of independent random variables, which by themselves are the most common type of statistics – linear ones, which in turn also serve as the most common approximation to other, nonlinear statistics; see e.g. [7,12,36]. Moreover, the results presented here naturally and effortlessly extend to martingales and supermartingales, and even to their maximal functions, as is done e.g. in [46,47]. In view of the general devices presented in [41] that provide for an automatic Banach-space analogue (with optimal constants) of any exponential bound for real-valued r.v.'s, one may ask whether similar devices exist for moment functions more general than the exponential ones, so that the mentioned results of Eaton, Bentkus, and Pinelis for the classes $\mathcal{H}^\alpha$ be similarly extended to Banach spaces. Other dimensionality reduction devices for sums of random vectors were given in [5,40,44].

Moment functions f of the classes $\mathcal{H}^\alpha$ (especially the functions of the form $x \mapsto (x-t)_+^\alpha$) naturally arise in mathematical finance. For example, $(S-K)_+$ is the value of a call option with the strike price K when the stock price is S.

However, applications of the Bennett–Hoeffding type bounds are mainly in statistics and theoretical probability. In fact, this paper was motivated by certain work on nonuniform bounds (NUB's) of a Berry–Esseen type on the convergence of $\mathsf{P}(S > z\sqrt{n})$ to the corresponding normal tail. Beginning with the classical paper by Nagaev [33], it became clear that Bennett–Hoeffding type bounds on large deviation probabilities play a crucial role in obtaining NUB's. The Nagaev NUB decreases only as fast as $1/z^3$, which naturally corresponds to his assumption of finite third moments of the X_i's. However, if the summands X_i are known to be bounded, the best known (to this author) rate of decrease of the NUB is e^{-cz}, for some constant $c > 0$. Results such as the ones presented in this paper are expected to allow one to get an NUB with a rate of decrease between the normal (e^{-cz^2}) and Poissonian ($\mathrm{e}^{-cz\ln z}$) ones. Hoeffding [24] also described certain applications to U-statistics and other related statistics. Similar applications can be given for the results presented in this paper. Yet another kind of applications is to skewness-corrected self-normalized sums as in [47,48].

There is a rather natural and usual trade-off between the results in this paper and the previous ones: the new bounds are more accurate, but harder to compute (and much harder to prove). Yet, as shown in [35], Section 3.1, these new bounds are quite effectively computable. With currently available standard computing tools, one can very quickly produce entire graphs of such bounds (as they depend on the parameters) – see [35], p. 19.

2. Statements of the main results

As in the Introduction, let $X_1, \dots, X_n$ be independent r.v.'s, with the sum $S = X_1 + \cdots + X_n$. For any $a > 0$ and $\theta > 0$, let Γ_{a^2} and Π_θ stand for any independent r.v.'s such that

$$\Gamma_{a^2} \sim \mathrm{N}\big(0, a^2\big) \quad \text{and} \quad \Pi_\theta \sim \mathrm{Pois}(\theta);$$

that is, Γ_{a^2} has the normal distribution with parameters 0 and a^2, and Π_θ has the Poisson distribution with parameter θ; at that, let Γ_0 and Π_0 be defined as the constant zero r.v. Let also

$$\tilde{\Pi}_\theta := \Pi_\theta - \mathsf{E}\Pi_\theta = \Pi_\theta - \theta.$$

Theorem 2.1. *Let σ, y, and β be any (strictly) positive real numbers such that*

$$\varepsilon := \frac{\beta}{\sigma^2 y} \in (0,1). \tag{2.1}$$

Suppose that the conditions (BH conds) hold, as well as

$$\sum \mathsf{E}(X_i)_+^3 \le \beta. \tag{2.2}$$

Then for all $f \in \mathcal{H}^3$

$$\mathsf{E}f(S) \le \mathsf{E}f\big(\Gamma_{(1-\varepsilon)\sigma^2} + y\tilde{\Pi}_{\varepsilon\sigma^2/y^2}\big). \tag{2.3}$$

Table 1
Attributes of the bounds BH, PU, Be, and Pin

Bound	$\mathcal{F}$	Condition $\sum \mathsf{E}(X_i)_+^3 \le \beta$ imposed?	η
BH	$\mathcal{E}$	no	$y\tilde{\Pi}_{\sigma^2/y^2}$
PU	$\mathcal{E}$	yes	$\Gamma_{(1-\varepsilon)\sigma^2} + y\tilde{\Pi}_{\varepsilon\sigma^2/y^2}$
Be	$\mathcal{H}^2$	no	$y\tilde{\Pi}_{\sigma^2/y^2}$
Pin	$\mathcal{H}^3$	yes	$\Gamma_{(1-\varepsilon)\sigma^2} + y\tilde{\Pi}_{\varepsilon\sigma^2/y^2}$

The proof of Theorem 2.1 will be given in Section 3, where all the necessary proofs are deferred to.

Note that the condition $\varepsilon \in (0,1)$ in (2.1) does not diminish generality, since $\sum \mathsf{E}(X_i)_+^3 \le y \sum \mathsf{E}X_i^2$. Note also that the bound in (2.3) is exact, in the following two senses. First ([35], Proposition 2.5), for any given real $p \in (0,3)$ one cannot replace $\mathcal{H}^3$ in Theorem 2.1 by the larger class $\mathcal{H}^p$. Second ([35], Proposition 2.3), for each $f \in \mathcal{H}^3$ the right-hand side of (2.3) is the supremum of its left-hand side under the restrictions (BH conds) and (2.2); one particular case when the upper bound in (2.3) is attained (in the limit) is when the X_i's are identically distributed (and satisfy certain other conditions); more generally, the nearly extremal distributions of the X_i's have to satisfy some kind of uniform asymptotic negligibility condition or be already close to Gauss–Poisson convolutions.

The mentioned bounds BH, PU, Be, and Pin on the generalized moments of S can each be presented in the following form: for each $f \in \mathcal{F}$,

$$\sup \mathsf{E} f(S) = \mathsf{E} f(\eta),$$

where the sup is taken over all independent X_i's satisfying the conditions (BH conds) (and possibly, depending on the class $\mathcal{F}$, condition (2.2)), and where the class $\mathcal{F}$ of functions and the r.v. η are as in Table 1.

Since in all the mentioned Bennett–Hoeffding type inequalities the X_i's are supposed to be bounded from above, one may say that such X_i's have no right tails. Of course, in applications one would truncate whatever tails the X_i's may have and then apply the Bennett–Hoeffding type bounds to the sum of the truncated random variables, as was done e.g. by Nagaev and Fuk [21,34].

By [42,43], one immediately obtains the following corollary of Theorem 2.1.

Corollary 2.2. *Under the conditions of Theorem 2.1, for all $x \in \mathbb{R}$*

$$\mathsf{P}(S \ge x) \le \frac{2\mathrm{e}^3}{9} \mathsf{P}^{\mathsf{LC}}\big(\Gamma_{(1-\varepsilon)\sigma^2} + y\tilde{\Pi}_{\varepsilon\sigma^2/y^2} \ge x\big), \tag{2.4}$$

where, for any r.v. η, the function $\mathsf{P}^{\mathsf{LC}}(\eta \ge \cdot)$ denotes the least log-concave majorant of the tail function $\mathsf{P}(\eta \ge \cdot)$.

The right-hand side of (2.4) can be effectively bounded from above by using [35], Propositions 3.10 and 3.11; asymptotically, for large $x > 0$, this is done in ([35], (3.26)).

A complete description of the best possible upper bound on the tail $\mathsf{P}(S \ge x)$ given a moment comparison of the form $\mathsf{E} f(S) \le \mathsf{E} f(\eta)$ for a r.v. η and all $f \in \mathcal{H}^\alpha$ is provided in [42], Theorem 2.5; for more on this, see [35], Proposition 3.2. Special cases of η and α are considered in [3,16].

Remark 2.3. *Quite similarly to how it was done e.g. in [46,47], it is easy to extend Theorem 2.1 and Corollary 2.2 to the more general case when the X_i's are the incremental differences of a (discrete-time) (super)martingale and/or replace S by the maximum of the partial sums; cf. e.g. [47], Corollary 5. Let us omit the details.*

3. Proofs

In Section 3.1, we shall first state several lemmas; based on these lemmas, we shall provide a proof of Theorem 2.1. Proofs of the lemmas will be deferred to Section 3.2. Such a structure will allow us to effectively present first the main ideas of the proofs and then the details.

Briefly, the scheme of proof of Theorem 2.1 is as follows. In a standard manner one can reduce the consideration to the case $n = 1$, so that it will remain to prove Lemma 3.6, for one r.v. X. By the definition ($\mathcal{H}^\alpha$), without loss of generality (w.l.o.g.) the moment function $f \in \mathcal{H}^3$ may be assumed to be of the form $(\cdot - w)_+^3$ for some $w \in \mathbb{R}$. Next, by Lemma 3.4, w.l.o.g. the r.v. X may be assumed to be a zero-mean r.v. $X_{a,b}$ taking on only two values $(-a)$ and b, which depend on the sign of w, though. Thus and by a monotonicity property (Lemma 3.5), it will remain to prove (3.10) for $f(\cdot) = (\cdot - w)_+^3$ and $X = X_{a,b}$; this final part of the proof is done by the infinitesimal spin-off method, described on page 23.

3.1. *Statements of lemmas, and the proof of Theorem 2.1*

First here, let us state a few lemmas, from which Theorem 2.1 easily follows. As before, let σ and y be any (strictly) positive real numbers. For any pair of numbers (a, b) such that $a \ge 0$ and $b > 0$, let $X_{a,b}$ denote any zero-mean r.v. with values $(-a)$ and b.

Lemma 3.1. *For all $x \in (-\infty, y]$, one has $x_+^3 \le \frac{y^5}{(y^2+\sigma^2)^2}(x + \sigma^2/y)^2$.*

Lemma 3.2. *Let X be any r.v. such that $X \le y$, $\mathsf{E}X \le 0$, and $\mathsf{E}X^2 \le \sigma^2$. Then*

$$\mathsf{E}X_+^3 \le \frac{y^3\sigma^2}{y^2+\sigma^2}. \tag{3.1}$$

Lemma 3.3. *For any*

$$\beta \in \left(0, \frac{y^3\sigma^2}{y^2+\sigma^2}\right] \tag{3.2}$$

there exists a unique pair $(a, b) \in (0, \infty) \times (0, \infty)$ such that $X_{a,b} \le y$, $\mathsf{E}X_{a,b}^2 = \sigma^2$, and $\mathsf{E}(X_{a,b})_+^3 = \beta$; more specifically, b is the only positive root of equation

$$\sigma^2 b^3 = \beta\big(b^2 + \sigma^2\big) \tag{3.3}$$

and

$$a = \frac{\sigma^2}{b} = \frac{\beta b}{b^3 - \beta}. \tag{3.4}$$

In particular, Lemma 3.3 implies that inequality (3.1) is exact.

Lemma 3.4. *Fix any $w \in \mathbb{R}$, $y > 0$, $\sigma > 0$, and β satisfying condition (3.2), and let (a, b) be the unique pair of numbers described in Lemma 3.3. Then*

$$\begin{aligned}
&\sup\big\{\mathsf{E}(X-w)_+^3\colon\ X \le y, \mathsf{E}X = 0, \mathsf{E}X^2 = \sigma^2, \mathsf{E}X_+^3 = \beta\big\} \\
&\quad = \max\big\{\mathsf{E}(X-w)_+^3\colon\ X \le y, \mathsf{E}X = 0, \mathsf{E}X^2 = \sigma^2, \mathsf{E}X_+^3 = \beta\big\} && (3.5)\\
&\quad = \max\big\{\mathsf{E}(X-w)_+^3\colon\ X \le y, \mathsf{E}X \le 0, \mathsf{E}X^2 \le \sigma^2, \mathsf{E}X_+^3 \le \beta\big\} && (3.6)\\
&\quad = \begin{cases} \mathsf{E}(X_{a,b}-w)_+^3 & \text{if } w \le 0, \\ \mathsf{E}(X_{\tilde a,\tilde b}-w)_+^3 & \text{if } w \ge 0, \end{cases} && (3.7)
\end{aligned}$$

where

$$\tilde b := y \quad \text{and} \quad \tilde a := \frac{\beta y}{y^3 - \beta} \tag{3.8}$$

(*cf.* (3.4)). *At that*, $\tilde{a} > 0$, $X_{\tilde{a},\tilde{b}} \le y$, $\mathsf{E}X_{\tilde{a},\tilde{b}} = 0$, *and* $\mathsf{E}(X_{\tilde{a},\tilde{b}})_+^3 = \beta$, *but one can only say that* $\mathsf{E}X_{\tilde{a},\tilde{b}}^2 \le \sigma^2$, *and the latter inequality is strict if* $\beta \ne \frac{y^3\sigma^2}{y^2+\sigma^2}$.

Lemma 3.5. *Let* $\sigma_0, \beta_0, \sigma, \beta$ *be any real numbers such that* $0 \le \sigma_0 \le \sigma$, $0 \le \beta_0 \le \beta$, $\beta_0 \le \sigma_0^2 y$, *and* $\beta \le \sigma^2 y$. *Then*

$$\mathsf{E}f(\Gamma_{\sigma_0^2-\beta_0/y} + y\tilde{\Pi}_{\beta_0/y^3}) \le \mathsf{E}f(\Gamma_{\sigma^2-\beta/y} + y\tilde{\Pi}_{\beta/y^3}) \tag{3.9}$$

for all $f \in \mathcal{H}^2$, *and hence for all* $f \in \mathcal{H}^3$.

Lemma 3.6. *Let* X *be any r.v. such that* $X \le y$, $\mathsf{E}X \le 0$, $\mathsf{E}X^2 \le \sigma^2$, *and* $\mathsf{E}X_+^3 \le \beta$, *where* β *satisfies condition* (3.2). *Then for all* $f \in \mathcal{H}^3$

$$\mathsf{E}f(X) \le \mathsf{E}f(\Gamma_{\sigma^2-\beta/y} + y\tilde{\Pi}_{\beta/y^3}). \tag{3.10}$$

Proof of Theorem 2.1. Let $\sigma_i^2 := \mathsf{E}X_i^2$, $\beta_i := \mathsf{E}(X_i)_+^3$, $\sigma_0^2 := \sum_{i=1}^n \sigma_i^2$, $\beta_0 := \sum_{i=1}^n \beta_i$, $Y_i := \Gamma_{\sigma_i^2-\beta_i/y} + y\tilde{\Pi}_{\beta_i/y^3}$, and $T := \sum_{i=1}^n Y_i$; at that, assume the Y_i's to be independent. Then, by a standard argument (cf. e.g. the proof of [45], Theorem 2.1) based on Lemma 3.6,

$$\mathsf{E}f(S) \le \mathsf{E}f(T) = \mathsf{E}f(\Gamma_{\sigma_0^2-\beta_0/y} + y\tilde{\Pi}_{\beta_0/y^3}) \quad \text{for all } f \in \mathcal{H}^3.$$

On the other hand, it is clear from (BH conds) and (2.2) that $0 \le \sigma_0^2 \le \sigma^2$ and $0 \le \beta_0 \le \beta$; next, $\beta_i \le \sigma_i^2 y$ for all $i = 1, \dots, n$ and hence $\beta_0 \le \sigma_0^2 y$; also, by (2.1), $\sigma^2 - \beta/y = (1-\varepsilon)\sigma^2$ and $\beta/y^3 = \varepsilon\sigma^2/y^2$. It remains to use Lemma 3.5. □

3.2. *Proofs of the lemmas*

Proof of Lemma 3.1. This follows because $\frac{x^3}{(x+\sigma^2/y)^2}$ is nondecreasing in $x \in [0, y]$ from 0 to $\frac{y^3}{(y+\sigma^2/y)^2} = \frac{y^5}{(y^2+\sigma^2)^2}$. □

Proof of Lemma 3.2. This follows by Lemma 3.1:

$$\mathsf{E}X_+^3 \le \frac{y^5}{(y^2+\sigma^2)^2}\big(\mathsf{E}X^2 + (\sigma^2/y)^2\big) \le \frac{y^3\sigma^2}{y^2+\sigma^2}. \qquad \square$$

Proof of Lemma 3.3. Take any β satisfying condition (3.2). Let $f(x) := \sigma^2 x^{3/2} - \beta(x+\sigma^2)$. Then $f(0) = -\beta\sigma^2 < 0$, $f(y^2) = \sigma^2 y^3 - \beta(y^2+\sigma^2) \ge 0$ by (3.2), and the function f is convex on $[0,\infty)$. Hence, f has exactly one positive root, say x_*, and at that $x_* \in (0, y^2]$. Let $b := x_*^{1/2}$, so that $b \in (0, y]$ and b is the only positive root of equation (3.3). Letting now $a := \sigma^2/b$, one has $X_{a,b} \le y$, $\mathsf{E}X_{a,b} = 0$, $\mathsf{E}X_{a,b}^2 = ab = \sigma^2$, and $\mathsf{E}(X_{a,b})_+^3 = \frac{ab^3}{a+b} = \frac{\sigma^2 b^3}{\sigma^2+b^2} = \beta$, by (3.3). It also follows that $a = \frac{\beta b}{b^3-\beta}$. Finally, the uniqueness of the pair (a, b) follows from the uniqueness of the positive root b of equation (3.3). □

Proof of Lemma 3.4. Let X be any r.v. such that $X \le y$, $\mathsf{E}X \le 0$, $\mathsf{E}X^2 \le \sigma^2$, and $\mathsf{E}X_+^3 \le \beta$. Let us consider separately the following possible cases: $w \le -a$, $-a \le w \le 0$, and $w \ge 0$.

Case 1: $w \le -a$. In this case, it is obvious that $w < 0$, and we claim that

$$f_1(x) := A_0 + A_1 x + A_2 x^2 + A_3 x_+^3 - (x-w)_+^3 \ge 0 \tag{3.11}$$

for all $x \in \mathbb{R}$, where

$$2A_0 := \frac{2a^3b}{3a+b} - w^3, \qquad A_1 := 3\frac{b(a^2+w^2)+a(3w^2-a^2)}{3a+b},$$
$$A_2 := -3\frac{(a+b)w+2a(a+w)}{3a+b}, \qquad A_3 := \frac{(a+b)^3}{b^2(3a+b)}$$

are obviously nonnegative constants. To verify the inequality in (3.11), note first that

$$f_1(x) = \begin{cases} \dfrac{a^2(b-x)^2(x(a+3b)+2ab)}{b^2(3a+b)} \geq 0 & \text{if } x \geq 0, \\ \dfrac{(a+x)^2(2ab-x(3a+b))}{3a+b} \geq 0 & \text{if } w \leq x < 0. \end{cases}$$

It remains to verify (3.11) for $x \in (-\infty, w)$, in which latter case $f_1''(x)\frac{3a+b}{2} = -6a(a+w) - 3w(a+b) \geq 0$, so that f_1 is convex on $(-\infty, w]$. Also, $f_1(w)(3a+b) = (a+w)^2(2ab-(3a+b)w) \geq 0$ and $f_1'(w)(3a+b) = -3(a+w)(a(a+w)-ab+(2a+b)w) \leq 0$. Thus, (3.11) holds for $x \in (-\infty, w)$ as well. Moreover, one can check that $f_1(x) = 0$ for $x \in \{-a, b\}$.

Therefore and because $\mathsf{E}X \leq 0 = \mathsf{E}X_{a,b}$, $\mathsf{E}X^2 \leq \sigma^2 = \mathsf{E}X_{a,b}^2$, and $\mathsf{E}X_+^3 \leq \beta = \mathsf{E}(X_{a,b})_+^3$, one has

$$\begin{aligned}\mathsf{E}(X-w)_+^3 &\leq A_0 + A_1\mathsf{E}X + A_2\mathsf{E}X^2 + A_3\mathsf{E}X_+^3 \\ &\leq A_0 + A_1\mathsf{E}X_{a,b} + A_2\mathsf{E}X_{a,b}^2 + A_3\mathsf{E}(X_{a,b})_+^3 = \mathsf{E}(X_{a,b}-w)_+^3.\end{aligned}$$

Case 2: $-a \leq w \leq 0$. For this case, the counterpart of (3.11) is that

$$f_2(x) := \lambda_2(x+a)^2 + \lambda_3 x_+^3 - (x-w)_+^3 \geq 0 \tag{3.12}$$

for all $x \in \mathbb{R}$, where

$$\lambda_2 := \frac{-3w(b-w)^2}{(a+b)(3a+b)} \quad \text{and} \quad \lambda_3 := \frac{(b-w)^2(2(w+a)+a+b)}{b^2(3a+b)}$$

are obviously nonnegative constants. As in Case 1, let us consider here the three subcases, according as $x \geq 0$, $w \leq x < 0$, or $x < w$.

For $x \geq 0$, one has $f_2(x)b^2(a+b)(3a+b) = -w(b-x)^2(p_0(w)+(a+b)p_1(w)x)$, where $p_0(w) := 3a^2b^2 - 6a^2bw - (4ab+b^2)w^2$ and $p_1(w) := 6ab + 3(b-a)w - 2w^2$. So, in this subcase, it is enough to show that $p_0(w) \geq 0$ and $p_1(w) \geq 0$ for $w \in [-a, 0]$, which follows because p_0 and p_1 are concave, with $p_0(-a) = 2a^3b + 2a^2b^2 \geq 0$, $p_0(0) = 3a^2b^2 \geq 0$, $p_1(-a) = a^2 + 3ab \geq 0$, and $p_1(0) = 6ab \geq 0$.

For $x \in [w, 0)$, note that $f_{22}(x) := f_2(x)(a+b)(3a+b)$ is a third-degree polynomial, with the leading coefficient $-3a^2 - 4ab - b^2 < 0$. So, on any interval f_{22} may change in its direction of convexity at most once, and only from convexity to concavity (when moving left to right). At that, $f_{22}(w) = -3(b-w)^2w(a+w)^2 \geq 0$ and $f_{22}'(w) = -6(b-w)^2w(a+w) \geq 0$. So, in this subcase, it is enough to show that $f_{22}(0) \geq 0$, which follows because $f_{22}(0) = -bwp_{22}(w)$, where $p_{22}(w) := 3a^2b - 6a^2w - (4a+b)w^2$ is concave in w, with $p_{22}(-a) = 2a^3 + 2a^2b \geq 0$ and $p_{22}(0) = 3a^2b \geq 0$.

To complete the proof of inequality (3.12) in Case 2, it remains to note that in the subcase $x < w$ one has $f_2(x)(a+b)(3a+b) = -3(b-w)^2w(a+x)^2 \geq 0$.

Moreover, $f_2(x) = 0$ for $x \in \{-a, b\}$.

It follows that

$$\begin{aligned}\mathsf{E}(X-w)_+^3 &\leq \lambda_2\mathsf{E}(X+a)^2 + \lambda_3\mathsf{E}X_+^3 \\ &\leq \lambda_2\mathsf{E}(X_{a,b}+a)^2 + \lambda_3\mathsf{E}(X_{a,b})_+^3 = \mathsf{E}(X_{a,b}-w)_+^3.\end{aligned}$$

Case 3: $w \geq 0$. Then

$$f_3(x) := \frac{(y-w)_+^3}{y^3} x_+^3 - (x-w)_+^3 \geq 0$$

for all $x \in (-\infty, y]$, since $\frac{(x-w)^3}{x^3}$ is nondecreasing in $x \in [w, \infty)$ for each $w \geq 0$; moreover, it is obvious that $f_3(x) = 0$ for $x \in (-\infty, 0] \cup \{y\}$.

Further, $y^3 \geq b^3 > \frac{ab^3}{a+b} = \beta$; hence, again by (3.8), $\tilde{a} > 0$. It follows that $f_3(x) = (x-w)_+^3$ for $x \in \{-\tilde{a}, \tilde{b}\}$. Moreover, $\mathsf{E}(X_{\tilde{a},\tilde{b}})_+^3 = \beta$. Thus,

$$\mathsf{E}(X-w)_+^3 \leq \frac{(y-w)_+^3}{y^3} \mathsf{E}X_+^3 \leq \frac{(y-w)_+^3}{y^3} \mathsf{E}(X_{\tilde{a},\tilde{b}})_+^3 = \mathsf{E}(X_{\tilde{a},\tilde{b}} - w)_+^3.$$

Moreover,

$$\mathsf{E}X_{\tilde{a},\tilde{b}}^2 = \tilde{a}\tilde{b} = \frac{\beta y^2}{y^3 - \beta} \leq \frac{\beta b^2}{b^3 - \beta} = ab = \sigma^2; \tag{3.13}$$

the inequality here takes place because $\frac{\beta u^2}{u^3 - \beta}$ decreases in $u > \beta^{1/3}$, while, as shown, $y^3 \geq b^3 > \beta$; the inequality in (3.13) is strict if $\beta \neq \frac{y^3\sigma^2}{y^2+\sigma^2}$ (because then $\frac{b^3\sigma^2}{b^2+\sigma^2} = \beta < \frac{y^3\sigma^2}{y^2+\sigma^2}$, and hence $b < y$).

Thus, in all the three cases, one has equality (3.7). Moreover, in the case $w \leq 0$ the maximum in (3.5) is attained and equals $\mathsf{E}(X_{a,b} - w)_+^3$, since $X_{a,b} \leq y$, $\mathsf{E}X_{a,b} = 0$, $\mathsf{E}X_{a,b}^2 = \sigma^2$, and $\mathsf{E}(X_{a,b})_+^3 = \beta$. The last sentence of Lemma 3.4 has also been proved.

To complete the proof of the lemma, it remains to show that in the case $w \geq 0$ the maxima in (3.5) and (3.6) are attained and equal $\mathsf{E}(X_{\tilde{a},\tilde{b}} - w)_+^3$; the same last sentence of Lemma 3.4 shows that in this case the max in (3.5) is *not* attained *at* $X = X_{\tilde{a},\tilde{b}}$ if $\beta \neq \frac{y^3\sigma^2}{y^2+\sigma^2}$ – because then $\mathsf{E}X_{\tilde{a},\tilde{b}}^2 < \sigma^2$.

Thus, it suffices to construct a r.v., say X_v, such that $\mathsf{E}(X_v - w)_+^3 = \mathsf{E}(X_{\tilde{a},\tilde{b}} - w)_+^3$, while $X_v \leq y$, $\mathsf{E}X_v = 0$, $\mathsf{E}X_v^2 = \sigma^2$, and $\mathsf{E}(X_v)_+^3 = \beta$. One way to satisfy all these conditions is to let $X_v \sim \tilde{p}\delta_y + q_1\delta_{-a_1} + r_1\delta_v$, where v is close enough to $-\infty$, $r_1 := -\Delta q$, $q_1 := \tilde{q} + \Delta q$, $a_1 := \tilde{a} + \Delta a$, $\tilde{p} := \beta/y^3$, $\tilde{q} := 1 - \tilde{p}$, $\Delta q := -\frac{\tilde{q}d^2}{d^2 + \tilde{q}(v+\tilde{a})^2}$, $\Delta a := \frac{d^2}{\tilde{q}(v+\tilde{a})}$, $d := \sqrt{\sigma^2 - \tilde{a}\tilde{b}} = \sqrt{\sigma^2 - \tilde{a}y}$, and $\tilde{a}$ and $\tilde{b}$ are given by (3.8). □

Proof of Lemma 3.5. In view of the relation $\mathcal{H}^3 \subseteq \mathcal{H}^2$, definition ($\mathcal{H}^\alpha$), and the Fubini theorem, it is enough to prove inequality (3.9) for all functions of the form $u \mapsto (u-w)_+^2$ for $w \in \mathbb{R}$. By rescaling, w.l.o.g. $y = 1$. Further, r.v. $\Gamma_{\sigma^2 - \beta_0} + \tilde{\Pi}_{\beta_0}$ equals in distribution $\Gamma + \Gamma_{\sigma_0^2 - \beta_0} + \tilde{\Pi}_{\beta_0}$, where Γ is any r.v. such that $\Gamma \sim \mathrm{N}(0, \sigma^2 - \sigma_0^2)$ and Γ is independent of $\Gamma_{\sigma_0^2 - \beta_0}$ and $\tilde{\Pi}_{\beta_0}$. Now, conditioning on $\Gamma_{\sigma_0^2 - \beta_0}$ and $\tilde{\Pi}_{\beta_0}$ and using Jensen's inequality, one has $\mathsf{E}(\Gamma_{\sigma_0^2 - \beta_0} + \tilde{\Pi}_{\beta_0} - w)_+^2 \leq \mathsf{E}(\Gamma_{\sigma^2 - \beta_0} + \tilde{\Pi}_{\beta_0} - w)_+^2$ for all $w \in \mathbb{R}$, so that w.l.o.g. $\sigma_0 = \sigma$ and $\beta_0 < \beta \leq \sigma^2$. Moreover, r.v.'s $\Gamma_{\sigma^2 - \beta_0} + \tilde{\Pi}_{\beta_0}$ and $\Gamma_{\sigma^2 - \beta} + \tilde{\Pi}_{\beta}$ equal in distribution $\Gamma_{d^2} + W$ and $\tilde{\Pi}_{d^2} + W$, respectively, where $d := (\beta - \beta_0)^{1/2}$ and W is any r.v. which is independent of Γ_{d^2} and $\tilde{\Pi}_{d^2}$ and equals $\Gamma_{\sigma^2 - \beta} + \tilde{\Pi}_{\beta_0}$ in distribution. Thus, by conditioning on W, it suffices to prove that

$$\mathsf{E}(\Gamma_{d^2} - w)_+^2 \leq \mathsf{E}(\tilde{\Pi}_{d^2} - w)_+^2 \tag{3.14}$$

for all $d > 0$ and $w \in \mathbb{R}$. Note that Γ_{d^2} and $\tilde{\Pi}_{d^2}$ are limits in distribution of $U_n := \sum_{i=1}^n U_{i;n}$ and $V_n := \sum_{i=1}^n V_{i;n}$, respectively, as $n \to \infty$, where the $U_{i;n}$'s are i.i.d. copies of $X_{d/\sqrt{n}, d/\sqrt{n}}$ and the $V_{i;n}$'s are i.i.d. copies of $X_{d^2/n,1}$. By [2,4], one has (3.14) with U_n and V_n in place of Γ_{d^2} and $\tilde{\Pi}_{d^2}$, respectively, provided that $n \geq d^2$ (so that $X_{d/\sqrt{n}, d/\sqrt{n}} \leq 1$).

Finally, it is clear that, for each $w \in \mathbb{R}$, $(x-w)_+^2 = \mathrm{o}(\mathrm{e}^x)$ as $x \to \infty$. Hence and in view of the Bennett–Hoeffding exponential bound, for each $w \in \mathbb{R}$ the sequences of r.v.'s $((U_n - w)_+^2)$ and $((V_n - w)_+^2)$ are uniformly integrable. Now (3.14) follows by a limit transition; see e.g. [6], Theorem 5.4. □

Proof of Lemma 3.6. In view of definition ($\mathcal{H}^\alpha$) and the Fubini theorem, it is enough to prove inequality (3.9) for all functions of the form $u \mapsto (u-w)_+^3$ for $w \in \mathbb{R}$. By Lemma 3.4, $\mathsf{E}(X-w)_+^3 \le \mathsf{E}(X_{a,b}-w)_+^3$ for some a and b such that $a>0$, $b>0$, $X_{a,b} \le y$, $\mathsf{E}X_{a,b}^2 = ab \le \sigma^2$, and $\mathsf{E}(X_{a,b})_+^3 = \beta$; at that, one of course also has $\mathsf{E}X_{a,b}=0$. So, if one could prove inequality (3.10) with $X_{a,b}$ in place of X and ab in place of σ^2, then it would remain to refer to Lemma 3.5. Thus, w.l.o.g. one has $X = X_{a_0,b_0}$ for some positive a_0 and b_0, and at that

$$b_0 \le y, \quad \mathsf{E}X_{a_0,b_0}^2 = a_0 b_0 = \sigma^2 \quad \text{and} \quad \mathsf{E}(X_{a_0,b_0})_+^3 = \frac{a_0 b_0^3}{b_0 + a_0} = \beta.$$

By rescaling, w.l.o.g.

$$y = 1, \quad \text{whence } b_0 \le 1.$$

The main idea of the proof of Lemma 3.6 may be referred to as that of *infinitesimal spin-off*, and it may be informally described as follows. Starting with the r.v. X_{a_0,b_0}, decrease both a_0 and b_0 simultaneously by infinitesimal amounts (say Δa and Δb) so that $\mathsf{E}(X_{a_0,b_0}-w)_+^3 \le \mathsf{E}(X_{a,b}+X_{\Delta_1,\Delta_1}+X_{\Delta_2,1}-w)_+^3$ for all $w\in\mathbb{R}$, where the r.v.'s $X_{a,b}, X_{\Delta_1,\Delta_1}, X_{\Delta_2,1}$ are independent, $a = a_0 - \Delta a$ and $b = b_0 - \Delta b$, and Δ_1 and Δ_2 are infinitesimal positive numbers which, together with Δa and Δb, are chosen in such a manner that $\mathsf{E}X_{a,b}^2 + \mathsf{E}X_{\Delta_1,\Delta_1}^2 + \mathsf{E}X_{\Delta_2,1}^2$ and $\mathsf{E}(X_{a,b})_+^3 + \mathsf{E}(X_{\Delta_1,\Delta_1})_+^3 + \mathsf{E}(X_{\Delta_2,1})_+^3$ match (exactly or closely enough) $\mathsf{E}X_{a_0,b_0}^2$ and $\mathsf{E}(X_{a_0,b_0})_+^3$, respectively. Continue decreasing a and b while "spinning off" the current infinitesimal r.v.'s X_{Δ_1,Δ_1} and $X_{\Delta_2,1}$, at that keeping the balance of the total variance and the sum of the positive-part third moments, as described above. Stop when $X_{a,b}=0$ almost surely, that is, when a or b is decreased to 0 (if ever); one can see that such a termination point is indeed attainable. Then the sum of all the symmetric independent infinitesimal spin-offs X_{Δ_1,Δ_1} will have a centered Gaussian distribution, while the sum of the highly asymmetric $X_{\Delta_2,1}$'s with the infinitesimal Δ_2's will give a centered Poisson component. At that, the balances of the variances and positive-part third moments will each be kept (the infinitesimal X_{Δ_1,Δ_1}'s will provide in the limit a total zero contribution to the latter of the two balances).

To formalize this idea, introduce a family of r.v.'s of the form

$$\eta_b := X_{a(b),b} + \xi_{\tau(b)} \quad \text{for } b \in [\varepsilon, b_0],$$

where

$$\varepsilon := \frac{b_0^2}{b_0+a_0} = \frac{\beta}{\sigma^2},$$

$$a(b) := \frac{b}{\varepsilon}(b-\varepsilon), \qquad \tau(b) := a_0 b_0 - a(b) b,$$

$$\xi_t := W_{(1-\varepsilon)t} + \tilde{\Pi}_{\varepsilon t}, \qquad \tilde{\Pi}_s := \Pi_s - \mathsf{E}\Pi_s,$$

$W.$ is a standard Wiener process, $\Pi.$ is a Poisson process with intensity 1, and $X_{a(b),b}, W., \Pi.$ are independent for each $b\in[\varepsilon,b_0]$. Note that $\varepsilon\in(0,b_0)\subseteq(0,1)$; also, τ is decreasing and hence nonnegative on the interval $[\varepsilon,b_0]$, since $a(b)b$ is increasing in $b\in[\varepsilon,b_0]$ and $a(b_0)=a_0$.

Let further

$$\mathcal{E}(b) := \mathcal{E}(b,w) := \mathsf{E}(\eta_b - w)_+^3 = \frac{b\mathsf{E}(\xi_{\tau(b)} - a(b) - w)_+^3 + a(b)\mathsf{E}(\xi_{\tau(b)} + b - w)_+^3}{b + a(b)}.$$

Since $a(b_0)=a_0$ and $a(\varepsilon)=0$, one has $X_{a(\varepsilon),\varepsilon}=0$. Thus, Lemma 3.6 is reduced to the inequality $\mathcal{E}(\varepsilon)\ge\mathcal{E}(b_0)$. Note that $\mathcal{E}(b)$ is continuous in $b\in[\varepsilon,b_0]$; this follows because of the uniform integrability (cf. the last paragraph in the proof of Lemma 3.5). So, it is enough to show that the left derivative $\mathcal{E}'(b)$ of $\mathcal{E}(b)$ is no greater than 0 for all $b\in(\varepsilon,b_0)$. To compute this derivative, one can use the following

Lemma 3.7. *Consider any function* $f:(\varepsilon,b_0)\times\mathbb{R}\ni(b,x)\mapsto f(b,x)\in\mathbb{R}$ *such that* $|f''_{bb}(b,x)|+|f''_{bx}(b,x)|+|f''_{xx}(b,x)|+|f'_x(b,x)|\le C_f e^{|x|}$ *and* $|f''_{xx}(b,x_1)-f''_{xx}(b,x_2)|\le C_f|x_1-x_2|(e^{|x_1|}+e^{|x_2|})$ *for some constant* C_f*, all* $b\in(\varepsilon,b_0)$*, and all* x,x_1,x_2 *in* $\mathbb{R}$*. Then for all* $b\in(\varepsilon,b_0)$

$$\lim_{h\downarrow 0}\frac{\mathsf{E}f(b-h,\xi_{\tau(b-h)})-\mathsf{E}f(b,\xi_{\tau(b)})}{-h}=\mathsf{E}F_f(b,\xi_{\tau(b)}),$$

where

$$F_f(b,x):=f'_b(b,x)+\left(\frac{(1-\varepsilon)f''_{xx}(b,x)}{2}+\varepsilon\big(f(b,x+1)-f(b,x)-f'_x(b,x)\big)\right)\tau'(b).$$

The proof of this lemma involves little more than routine Taylor expansions; for details, see Lemma 4.10 in [35] and its proof therein; cf. the well-known formula for the Lévy process generator, e.g. [26], Theorem 19.10.

By Lemma 3.7, for all $b\in(\varepsilon,b_0)$ one has $\mathcal{E}'(b)=\mathsf{E}G(b,\xi_{\tau(b)}-w)$, where

$$\begin{aligned}
&G(b,x):=\left(\frac{b}{b+a(b)}\right)'_b\big(f_1(b,x)-f_2(b,x)\big)+\frac{bF_{f_1}(b,x)+a(b)F_{f_2}(b,x)}{b+a(b)},\\
&f_1(b,x):=\big(x-a(b)\big)^3_+,\qquad f_2(b,x):=(x+b)^3_+.
\end{aligned}\tag{3.15}$$

Thus, it suffices to show that $G(b,u)\le 0$ for all $b\in(\varepsilon,b_0)$ and $u\in\mathbb{R}$. Observe now that $(\frac{b}{b+a(b)})'_b=-\frac{1}{b+a(b)}$, $a'(b)=1+\frac{2a(b)}{b}$, $\tau'(b)=-(3a(b)+b)$, and $\varepsilon=\frac{b^2}{b+a(b)}$. Substituting into (3.15) these expressions of $(\frac{b}{b+a(b)})'_b$, $a'(b)$, $\tau'(b)$, and ε in terms of only b and $a(b)$, one has $(b+a(b))^2G(b,u)=-\tilde G(a(b),b,-u)$, where

$$\begin{aligned}
\tilde G(a,b,t):={}&\big(a+b-3ab^3-b^4\big)(-a-t)^3_+\\
&-\big(a+b+3a^2b^2+ab^3\big)(b-t)^3_+\\
&+b^2(3a+b)\big(b(1-a-t)^3_++a(1+b-t)^3_+\big)\\
&+3\big(2a^2+3ab+b^2-3ab^3-b^4\big)(-a-t)^2_+\\
&-3a\big(a+b+3ab^2+b^3\big)(b-t)^2_+\\
&+3(3a+b)\big(a+b-b^2\big)\big(b(-a-t)_++a(b-t)_+\big).
\end{aligned}$$

To complete the proof of the theorem, it is enough to show that $\tilde G(a,b,t)\ge 0$ for all $a>0$, $b\in(0,1]$, and $t\in\mathbb{R}$. At that, w.l.o.g. $t<1+b$, since $\tilde G(a,b,t)=0$ for all $a>0$, $b\in(0,1]$, and $t\ge 1+b$. Next, one has either $a+b\le 1$ or $a+b>1$. In the first case, $-a\le b\le 1-a\le 1+b$, while in the second case $-a\le 1-a\le b\le 1+b$. Therefore, it remains to verify that $\tilde G(a,b,t)\ge 0$ in each of the following 8 (sub)cases:

Case 10: $a>0$ and $b>0$ and $a+b\le 1$ and $t\le -a$;
Case 11: $a>0$ and $b>0$ and $a+b\le 1$ and $-a\le t\le b$;
Case 12: $a>0$ and $b>0$ and $a+b\le 1$ and $b\le t\le 1-a$;
Case 13: $a>0$ and $b>0$ and $a+b\le 1$ and $1-a\le t\le 1+b$;
Case 20: $a>0$ and $0<b\le 1$ and $a+b>1$ and $t\le -a$;
Case 21: $a>0$ and $0<b\le 1$ and $a+b>1$ and $-a\le t\le 1-a$;
Case 22: $a>0$ and $0<b\le 1$ and $a+b>1$ and $1-a\le t\le b$;
Case 23: $a>0$ and $0<b\le 1$ and $a+b>1$ and $b\le t\le 1+b$

(actually, the condition $a>0$ is redundant in Cases 20 through 23).

Clearly, $\tilde G(a,b,t)$ is piecewise polynomial in a,b,t. More specifically, for each pair $(\mathrm{i},\mathrm{j})\in\{1,2\}\times\{0,1,2,3\}$ there exists a polynomial $G_{\mathrm{ij}}=G_{\mathrm{ij}}(t)=G_{\mathrm{ij}}(a,b,t)$ such that $\tilde G(a,b,t)=G_{\mathrm{ij}}$ for all (a,b,t) satisfying the conditions of Case ij. It is also clear that the function $\tilde G$ is continuous on $\mathbb{R}^3$.

Note that the expressions $G_{10} = G_{20} = a^2(5a^2 + 8ab + 3b^2)$, $G_{12} = b^2(3a + b)(a(1 + b - t)^3 + b(1 - a - t)^3)$, and $G_{13} = G_{23} = ab^2(3a + b)(1 + b - t)^3$ are all manifestly nonnegative, in the respective cases. So, of the 8 cases, there remain only three cases to consider: Cases 11, 21, and 22.

One can check that

$$\begin{aligned}4G_{11} = {} & 10p_{00231} + 10p_{00321} + 2p_{01210} + 10p_{01211} + 20p_{01300} + 10p_{01311} + 10p_{03210} \\ & + 2p_{10120} + 6p_{10122} + 10p_{10220} + 4p_{10222} + 7p_{11110} + 5p_{11111} + 6p_{11121} + 16p_{11200} \\ & + 6p_{11210} + 34p_{11220} + p_{11221} + 17p_{11310} + 12p_{12011} + 4p_{12100} + 11p_{12110} + 3p_{12211} \\ & + 3p_{12310} + 4p_{20023} + 2p_{20130} + 4p_{20131} + 8p_{21012} + 4p_{21022} + 15p_{21111} + 7p_{21121} + 11p_{21130},\end{aligned}$$

where

$$p_{\mathrm{ijklm}} := v_1^{\mathrm{i}} \cdots v_5^{\mathrm{m}} \quad \text{and} \quad (v_1, \dots, v_5) := (t + a, b - t, a, b, 1 - a - b). \tag{3.16}$$

Since all the terms p_{ijklm} in the above representation of the polynomial $4G_{11}$ are nonnegative for (a, b, t) as in Case 11, this representation immediately "certifies" the nonnegativity of G_{11}; the existence of such a "certificate" follows by certain results in real algebraic geometry; see e.g. [11,22,23,29,30].

Since the verification of the above, purely algebraic representation of $4G_{11}$ is quite tedious (and better done with the aid of a computer algebra software package), let us present an alternative proof of the nonnegativity of G_{11}, which involves some calculus. Toward this end, first note that G_{11} is a third-degree polynomial (in t) and hence on any interval may change in its direction of convexity at most once. At that,

$$\begin{aligned}&G_{11}(-a) = a^2(a + b)(5a + 3b) \geq 0, \\ &G_{11}'(-a) = 3b(3a + b)\big(a + b - b^2\big) \geq 0, \\ &G_{11}(b) = b^2(a + b)(3a + b)\big(a + a(1 - a - b) + a(1 - a - b)^2 + (1 - a - b)^3\big) \geq 0, \\ &G_{11}'(b) = -3(3a + b)\big(a + ab + b^2\big)\big(b(1 - b)^2 + a\big(1 - b + b^2\big)\big) \leq 0\end{aligned}$$

for (a, b) as in Case 11, so that indeed $G_{11}(t) \geq 0$ for $t \in [-a, b]$.

Next, concerning Case 21, one can verify the following certificate of the nonnegativity of G_{21}:

$$\begin{aligned}2G_{21} = {} & 10p_{00014} + 16p_{00021} + 24p_{00023} + 10p_{00104} + 24p_{00113} + 4p_{00211} + 4p_{00240} \\ & + 3p_{00302} + 4p_{00340} + 4p_{00430} + 8p_{01011} + 16p_{01021} + 60p_{01022} + 10p_{01040} + 3p_{01102} \\ & + 45p_{01112} + 6p_{01120} + 6p_{01130} + 4p_{01140} + 3p_{01212} + 16p_{02003} + 16p_{10003} + 45p_{10012} \\ & + 3p_{10032} + 2p_{10101} + 2p_{10111} + 9p_{10202} + 2p_{10311} + 4p_{10320} + 4p_{11001} + 12p_{11002} \\ & + 16p_{11003} + 8p_{11010} + 20p_{11011} + 2p_{11120} + 14p_{11211} + 8p_{11220} + 4p_{11221} + 2p_{12010} \\ & + 2p_{12110} + 12p_{12111} + 2p_{12311} + 6p_{20111} + 4p_{21101} + 4p_{21310} + 2p_{30041} + 2p_{30210},\end{aligned}$$

where (cf. (3.16))

$$p_{\mathrm{ijklm}} := v_1^{\mathrm{i}} \cdots v_5^{\mathrm{m}} \quad \text{and} \quad (v_1, \dots, v_5) := (t + a, 1 - a - t, b, 1 - b, a + b - 1).$$

In this case too, let us present an alternative proof of the nonnegativity of G_{21}, involving calculus. Here, similarly, first note that G_{21} is a third-degree polynomial (in t) and hence on any interval may change in its direction of convexity at most once. At that,

$$\begin{aligned}&G_{21}(-a) = a^2(a + b)(5a + 3b) \geq 0, \\ &G_{21}'(-a) = 3b(3a + b)\big(a + b - b^2\big) \geq 0,\end{aligned}$$

$$G_{21}(1-a) = p_{001} + 6p_{002} + 8p_{003} + 5p_{004} + p_{011} + 9p_{012} + 12p_{013} + 7p_{021} + 9p_{022} + 5p_{211} + 3p_{220} + p_{320} + p_{410} \geq 0,$$
$$G'_{21}(1-a) = -3(a+b)\big[(1-b) + 2(1-b)^2 + (a+b-1)\big(1+3(1-b)\big)\big] \leq 0$$

for (a,b) as in Case 21, where $p_{\mathrm{klm}} := b^{\mathrm{k}}(1-b)^{\mathrm{l}}(a+b-1)^{\mathrm{m}}$. So, indeed $G_{21}(t) \geq 0$ for $t \in [-a, 1-a]$.

It remains to consider Case 22. In this case, a certificate of nonnegativity is as follows:

$$G_{22} = p_{00201} + 3p_{00220} + p_{00310} + 5p_{01003} + 6p_{01010} + 20p_{01012} + 3p_{01030} + 2p_{01210} + 3p_{01302} + 24p_{02020} + 21p_{02110} + 8p_{02121} + 6p_{02200} + 5p_{03200} + 11p_{03210} + 5p_{10210} + 7p_{11001} + 4p_{11012} + 27p_{11030} + 2p_{11101} + 42p_{11110} + 9p_{11130} + 9p_{11400} + p_{12021} + p_{12030} + 6p_{12100} + 6p_{12111} + 5p_{12201} + 13p_{12210} + 3p_{20210} + 3p_{20300} + 4p_{21111} + 3p_{21120} + 4p_{21201} + 6p_{21210},$$

where

$$p_{\mathrm{ijklm}} := v_1^{\mathrm{i}} \cdots v_5^{\mathrm{m}} \quad \text{and} \quad (v_1, \dots, v_5) := (t+a-1, b-t, b, 1-b, a+b-1).$$

For a proof (for Case 22) involving calculus, first note that G_{22} is a third-degree polynomial (in t), and

$$\text{the coefficient of } t^3 \text{ in } G_{22} \text{ is } a+b > 0. \tag{3.17}$$

So, on any interval G_{22} may change in its direction of convexity at most once, and only from concavity to convexity (when moving left to right). At that, $G_{22}(b) = ab^2(3a+b) \geq 0$ and $G'_{22}(b) = -3a(a+b)(3a+b) \leq 0$ for (a,b) as in Case 22; moreover, $G_{22}(1-a) = G_{21}(1-a) \geq 0$. Thus, $G_{22}(t) \geq 0$ for $t \in [1-a, b]$. (In view of (3.17), the sign of $G'_{22}(1-a)$ is irrelevant for these considerations.) This concludes the proof of Lemma 3.6. (Yet another proof of the nonnegativity of $\tilde{G}(a,b,t)$, based in essence on a theory by Tarski [14,31,53], can be found in the mentioned paper [35].) □

References

[1] G. Bennett. Probability inequalities for the sum of independent random variables. *J. Amer. Statist. Assoc.* **57** (1962) 33–45.

[2] V. Bentkus. A remark on the inequalities of Bernstein, Prokhorov, Bennett, Hoeffding, and Talagrand. *Liet. Mat. Rink.* **42** (2002) 332–342. MR1947624

[3] V. Bentkus, N. Kalosha and M. van Zuijlen. On domination of tail probabilities of (super)martingales: Explicit bounds. *Liet. Mat. Rink.* **46** (2006) 3–54. MR2251440

[4] V. Bentkus. On Hoeffding's inequalities. *Ann. Probab.* **32** (2004) 1650–1673. MR2060313

[5] E. Berger. Majorization, exponential inequalities and almost sure behavior of vector-valued random variables. *Ann. Probab.* **19** (1991) 1206–1226. MR1112413

[6] P. Billingsley. *Convergence of Probability Measures*. Wiley, New York, 1968. MR0233396

[7] E. Bolthausen and F. Götze. The rate of convergence for multivariate sampling statistics. *Ann. Statist.* **21** (1993) 1692–1710. MR1245764

[8] S. Boucheron, G. Lugosi and P. Massart. A sharp concentration inequality with applications. *Random Structures Algorithms* **16** (2000) 277–292. MR1749290

[9] O. Bousquet. A Bennett concentration inequality and its application to suprema of empirical processes. *C. R. Math. Acad. Sci. Paris* **334** (2002) 495–500. MR1890640

[10] O. Bousquet. Concentration inequalities for sub-additive functions using the entropy method. In *Stochastic Inequalities and Applications* 213–247. *Progr. Probab.* **56**. Birkhäuser, Basel, 2003. MR2073435

[11] G. Cassier. Problème des moments sur un compact de $\mathbf{R}^n$ et décomposition de polynômes à plusieurs variables. *J. Funct. Anal.* **58** (1984) 254–266. MR0759099

[12] L. H. Y. Chen and Q.-M. Shao. Normal approximation for nonlinear statistics using a concentration inequality approach. *Bernoulli* **13** (2007) 581–599. MR2331265

[13] A. Cohen, Y. Rabinovich, A. Schuster and H. Shachnai. Optimal bounds on tail probabilities: A study of an approach. In *Advances in Randomized Parallel Computing* 1–24. *Comb. Optim.* **5**. Kluwer Acad. Publ., Dordrecht, 1999. MR1782937

[14] G. E. Collins. Quantifier elimination for real closed fields by cylindrical algebraic decomposition. In *Quantifier Elimination and Cylindrical Algebraic Decomposition* (Linz, 1993) 85–121. *Texts Monogr. Symbol. Comput.* Springer, Vienna, 1998. MR1634190

[15] V. H. de la Peña. A general class of exponential inequalities for martingales and ratios. *Ann. Probab.* **27** (1999) 537–564. MR1681153

[16] J.-M. Dufour and M. Hallin. Improved Eaton bounds for linear combinations of bounded random variables, with statistical applications. *J. Amer. Statist. Assoc.* **88** (1993) 1026–1033. MR1242946

[17] K. Dzhaparidze and J. H. van Zanten. On Bernstein-type inequalities for martingales. *Stochastic Process. Appl.* **93** (2001) 109–117. MR1819486

[18] M. L. Eaton. A note on symmetric Bernoulli random variables. *Ann. Math. Statist.* **41** (1970) 1223–1226. MR0268930

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

[20] D. A. Freedman. On tail probabilities for martingales. *Ann. Probability* **3** (1975) 100–118. MR0380971

[21] D. H. Fuk and S. V. Nagaev. Probabilistic inequalities for sums of independent random variables. *Teor. Verojatnost. i Primenen.* **16** (1971) 660–675. MR0293695

[22] D. Handelman. Positive polynomials and product type actions of compact groups. *Mem. Amer. Math. Soc.* **54** (1985) xi+79. MR0783217

[23] D. Handelman. Representing polynomials by positive linear functions on compact convex polyhedra. *Pacific J. Math.* **132** (1988) 35–62. MR0929582

[24] W. Hoeffding. Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** (1963) 13–30. MR0144363

[25] S. Janson. Large deviations for sums of partly dependent random variables. *Random Structures Algorithms* **24** (2004) 234–248. MR2068873

[26] O. Kallenberg. *Foundations of Modern Probability*, 2nd edition. *Probability and Its Applications (New York)*. Springer, New York, 2002. MR1876169

[27] T. Klein and E. Rio. Concentration around the mean for maxima of empirical processes. *Ann. Probab.* **33** (2005) 1060–1077. MR2135312

[28] T. Klein, Y. Ma and N. Privault. Convex concentration inequalities and forward-backward stochastic calculus. *Electron. J. Probab.* **11** (2006) 486–512 (electronic). MR2242653

[29] J.-L. Krivine. Anneaux préordonnés. *J. Analyse Math.* **12** (1964) 307–326. MR0175937

[30] J.-L. Krivine. Quelques propriétés des préordres dans les anneaux commutatifs unitaires. *C. R. Acad. Sci. Paris* **258** (1964) 3417–3418. MR0169083

[31] S. Łojasiewicz. Sur les ensembles semi-analytiques. In *Actes du Congrès International des Mathématiciens (Nice, 1970)* **2** 237–241. Gauthier-Villars, Paris, 1971. MR0425152

[32] P. Massart. About the constants in Talagrand's concentration inequalities for empirical processes. *Ann. Probab.* **28** (2000) 863–884. MR1782276

[33] S. V. Nagaev. Some limit theorems for large deviations. *Theory Probab. Appl.* **10** (1965) 214–235. MR0185644

[34] S. V. Nagaev. Large deviations of sums of independent random variables. *Ann. Probab.* **7** (1979) 745–789. MR0542129

[35] I. Pinelis. On the Bennett–Hoeffding inequality. Preprint. Available at arXiv:0902.4058v1 [math.PR].

[36] 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. Available at arXiv:0906.0177v3 [math.ST].

[37] I. F. Pinelis and A. I. Sakhanenko. Remarks on inequalities for probabilities of large deviations. *Theory Probab. Appl.* **30** (1985) 143–148. MR0779438

[38] I. S. Pinelis and S. A. Utev. Sharp exponential estimates for sums of independent random variables. *Theory Probab. Appl.* **34** (1989) 340–346. MR1005745

[39] I. Pinelis. An approach to inequalities for the distributions of infinite-dimensional martingales. In *Probability in Banach Spaces, 8 (Brunswick, ME, 1991)* 128–134. *Progr. Probab.* **30**. Birkhäuser Boston, Boston, MA, 1992. MR1227615

[40] I. Pinelis. On a majorization inequality for sums of independent random vectors. *Statist. Probab. Lett.* **19** (1994) 97–99. MR1256696

[41] I. Pinelis. Optimum bounds for the distributions of martingales in Banach spaces. *Ann. Probab.* **22** (1994) 1679–1706. MR1331198

[42] I. Pinelis. Optimal tail comparison based on comparison of moments. In *High Dimensional Probability (Oberwolfach, 1996)* 297–314. *Progr. Probab.* **43**, Birkhäuser, Basel, 1998. MR1652335

[43] I. Pinelis. Fractional sums and integrals of r-concave tails and applications to comparison probability inequalities. In *Advances in Stochastic Inequalities (Atlanta, GA, 1997)* 149–168. *Contemp. Math.* **234**. Amer. Math. Soc., Providence, RI, 1999. MR1694770

[44] I. Pinelis. Dimensionality reduction in extremal problems for moments of linear combinations of vectors with random coefficients. In *Stochastic Inequalities and Applications* 169–185. *Progr. Probab.* **56**. Birkhäuser, Basel, 2003. MR2073433

[45] I. Pinelis. Binomial upper bounds on generalized moments and tail probabilities of (super)martingales with differences bounded from above. In *High Dimensional Probability* 33–52. *IMS Lecture Notes Monogr. Ser.* **51**. IMS, Beachwood, OH, 2006. MR2387759

[46] I. Pinelis. On normal domination of (super)martingales. *Electron. J. Probab.* **11** (2006) 1049–1070. MR2268536

[47] I. Pinelis. Exact inequalities for sums of asymmetric random variables, with applications. *Probab. Theory Related Fields* **139** (2007) 605–635. MR2322709

[48] I. Pinelis. Optimal two-value zero-mean disintegration of zero-mean random variables. *Electron. J. Probab.* **14** (2009) 663–727. MR2486818

[49] G. G. Roussas. Exponential probability inequalities with some applications. In *Statistics, Probability and Game Theory* 303–319. *IMS Lecture Notes Monogr. Ser.* **30**. Inst. Math. Statist., Hayward, CA, 1996. MR1481786

[50] G. R. Shorack and J. A. Wellner. *Empirical Processes with Applications to Statistics*. *Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics*. Wiley, New York, 1986. MR0838963

[51] M. Talagrand. The missing factor in Hoeffding's inequalities. *Ann. Inst. H. Poincaré Probab. Statist.* **31** (1995) 689–702. MR1355613

[52] M. Talagrand. New concentration inequalities in product spaces. *Invent. Math.* **126** (1996) 505–563. MR1419006

[53] A. Tarski. *A Decision Method for Elementary Algebra and Geometry*. RAND Corporation, Santa Monica, CA, 1948. MR0028796