BULLETIN OF THE
HELLENIC MATHEMATICAL SOCIETY
Volume 62, 2018 (31–43)

ON THE BERNSTEIN-HOEFFDING METHOD

CHRISTOS PELEKIS, JAN RAMON, AND YUYI WANG

Abstract. We consider extensions of Hoeffding's "exponential method" approach for obtaining upper estimates on the probability that a sum of independent and bounded random variables is significantly larger than its mean. We show that the exponential function in Hoeffding's approach can be replaced with any function which is non-negative, increasing and convex. As a result we generalize and improve upon Hoeffding's inequality. Our approach allows to obtain "missing factors" in Hoeffding's inequality. The later result is a rather weaker version of a theorem that is due to Michel Talagrand. Moreover, we characterize the class of functions with respect to which our method yields optimal concentration bounds. Finally, using ideas from the theory of Bernstein polynomials, we show that similar ideas apply under information on higher moments of the random variables.

1. Prologue, related work and main results

Given a real number $p \in (0,1)$ let $\mathcal{B}(p)$ denote the set consisting of all $[0,1]$-valued random variables whose mean is equal to p. This work is motivated by the problem of obtaining sharp upper bounds on the "tail probability"

$$\mathbb{P}\left[\sum_{i=1}^{n} X_i \geq t\right],$$

where t is a fixed real number t such that $\sum_{i=1}^{n} p_i < t < n$ and $X_1, \ldots, X_n$ are independent random variables such that $X_i \in \mathcal{B}(p_i)$, for $i \in \{1, \ldots, n\}$.

If $t \leq \sum_i p_i$, then the problem is trivial; just choose X_i to be equal to p_i with probability 1. Throughout the text, whenever a random variable $X \in \mathcal{B}(p)$ is under consideration, it will be tacitly assumed that $p \in (0,1)$, thus excluding the uninteresting cases where X is either identically equal to zero, or identically equal to one.

There is a vast amount of literature that is related to the aforementioned problem and the interested reader is invited to take a look at the works of Bentkus [1], [2], [3], Fan et al. [5], From [7], From et al. [8], Györfi et al. [9], Hoeffding [10], Kha et al. [14], Krafft et al. [11], McDiarmid [12], Pinelis [17],[18], Schmidt et al. [19], Siegel [21], Talagrand [22], Xia [23], Zheng [24], among many others.

Probably the first systematic approach that allows one to obtain upper bounds on the probability that a sum of independent $[0,1]$-valued random variables is larger than its mean, was devised by Hoeffding in [10]. Hoeffding's approach is based on a method of Bernstein (see [10, page 14]) and from now on will be referred to as the *Bernstein-Hoeffding* method. The Bernstein-Hoeffding method is, briefly, the following.

2010 *Mathematics Subject Classification.* 60E15.
Key words and phrases. Hoeffding's inequality; convex order; Bernstein polynomials.

Markov's inequality and the assumption that the random variables are independent imply that

$$\mathbb{P}\left[\sum_{i=1}^{n} X_i \geq t\right] \leq e^{-ht}\prod_{i=1}^{n}\mathbb{E}[e^{hX_i}] \leq e^{-ht}\left\{\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[e^{hX_i}]\right\}^n, \text{ for all } h>0,$$

where the last inequality comes from the arithmetic-geometric means inequality. Using the fact that the function $f(t)=e^{ht}$ is *convex* one can show (see Lemma 2.2 below) that

$$\mathbb{E}[e^{hX_i}] \leq \mathbb{E}[e^{hB_i}],$$

where B_i is a Bernoulli random variable of mean p_i. Therefore, it follows that

$$\mathbb{P}\left[\sum_{i=1}^{n} X_i \geq t\right] \leq e^{-ht}\{(1-p)+pe^h\}^n, \text{ for all } h>0,$$

where $p=\frac{1}{n}\sum_{i=1}^{n} p_i$. Minimizing the expression in the right hand side of the last inequality with respect to h, we find $e^h=\frac{t(1-p)}{p(n-t)}$ and hence we obtain the following celebrated result of Hoeffding (see [10, Theorem 1]).

Theorem 1.1 (Hoeffding, 1963)**.** *Let the random variables $X_1,\dots,X_n$ be independent and such that $0\leq X_i\leq 1$, for each $i=1,\dots,n$. Set $p=\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i]$. Then for any t such that $np<t<n$ we have*

$$\mathbb{P}\left[\sum_{i=1}^{n} X_i \geq t\right] \leq \inf_{h>0}\left\{e^{-ht}\left(1-p+pe^h\right)^n\right\}.$$

Furthermore,

$$\inf_{h>0}\left\{e^{-ht}\left(1-p+pe^h\right)^n\right\} = \left(\frac{p(n-t)}{t(1-p)}\right)^t\left(\frac{(1-p)n}{n-t}\right)^n := H(n,p,t).$$

The function $H(n,p,t)$ in the last expression is the so-called *Hoeffding bound* (or *Hoeffding's function*) on tail probabilities for sums of independent, bounded random variables. Here and later, we denote by $\mathrm{Ber}(q)$ a Bernoulli random variable with mean q and by $\mathrm{Bin}(n,q)$ a binomial random variable of parameters n and q. If two random variables W,Z have the same distribution we will write $W\sim Z$. Let us remark that the Hoeffding bound is sharp, in the sense that the Bernoulli random variables $\mathrm{Ber}(p_i)$ attain the bound, i.e.,

$$\inf_{h>0}\ e^{-ht}\left\{\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[e^{hB_i}]\right\}^n = H(n,p,t),$$

where B_i is a $\mathrm{Ber}(p_i)$ random variable. The main ideas behind this work are hidden in the fact that

$$\prod_i \mathbb{E}\left[e^{hB_i}\right] \leq \mathbb{E}\left[e^{hB}\right],$$

where $B\sim\mathrm{Bin}(n,p)$ with $p=\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[B_i]$, and the fact that the function $f(x)=e^{hx}, h>0$, is non-negative, increasing and convex. In a subsequent section we will show that, while applying the Bernstein-Hoeffding method, one can replace the exponential function $f(x)=e^{hx}, h>0$, with any function $f(\cdot)$ having the aforementioned properties. Let us also remark that a slightly looser but more widely used version of Hoeffding's bound is the bound $\exp\left(-2n(t/n-p)^2\right)$, which follows from the fact that $H(n,p,t)\leq\exp\left(-2n\left(\frac{t}{n}-p\right)^2\right)$ (see [10, formula (2.3)]).

In this article we shall be interested in improvements upon Hoeffding's theorem. This is a topic that has attracted the attention of several authors (see, for example [2, 18, 21, 22]). Let us bring to the reader's attention the following result which is due to Talagrand (see [22, Theorem 1.2]). Talagrand's work focuses on obtaining a "missing factor" in Hoeffding's inequality. The missing factor is obtained by combining the Bernstein-Hoeffding method together with a technique (i.e., suitable change of probability measure) that is used in the proof of Cramér's theorem on large deviations, yielding the following.

Theorem 1.2 (Talagrand, 1995)**.** *Let the random variables $X_1, \dots, X_n$ be independent and such that $0 \leq X_i \leq 1$, for each $i = 1, \dots, n$. Set $p = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[X_i]$. Then, for some absolute constant K, and every real number t such that $np + K \leq t \leq np + np(1-p)/K$, we have*

$$\mathbb{P}\left[\sum_{i=1}^{n} X_i \geq t\right] \leq \left\{\theta\left(\frac{t-np}{\sqrt{np(1-p)}}\right) + \frac{K}{\sqrt{np(1-p)}}\right\} \cdot H(n,p,t),$$

where $H(n,p,t)$ is the Hoeffding bound and $\theta(\cdot)$ is a non-negative function such that

$$\frac{1}{\sqrt{2\pi}(1+x)} \leq \frac{2}{\sqrt{2\pi}(x+\sqrt{x^2+4})} \leq \theta(x) \leq \frac{4}{\sqrt{2\pi}(3x+\sqrt{x^2+8})}, \text{ for } x \geq 0.$$

See [22] for the proof of this theorem and the precise definition of the function $\theta(\cdot)$. In other words, Talagrand's result improves upon Hoeffding's by inserting a "missing" factor of order $\approx \frac{\sqrt{np(1-p)}}{\sqrt{np(1-p)}+t-np} < 1$ in the Hoeffding bound. Notice that Talagrand's result holds true for moderate values of t, i.e., for $t \in [np + K, np + np(1-p)/K]$, for some absolute constant K whose value does *not* seem to be known. Talagrand (see [22, page 692]) mentions that one can obtain a rather small numerical value for K, but numerical computations are left to others with the talent for it. One of the purposes of this article is to improve upon Hoeffding's inequality by obtaining "missing" factors with explicit numerical value for the constant.

Another improvement upon Hoeffding's theorem is due to Bentkus. In the proof of [2, Theorem 1.2] (see [2, page 1659]) Bentkus, implicitly, obtains the following result.

Theorem 1.3 (Bentkus, 2004)**.** *Let the random variables $X_1, \dots, X_n$ be independent and such that $0 \leq X_i \leq 1$, for each $i = 1, \dots, n$. Set $p = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[X_i]$. Then, for any positive real t such that $np < t < n$, we have*

$$\mathbb{P}\left[\sum_{i=1}^{n} X_i \geq t\right] \leq \inf_{a<t} \frac{1}{t-a}\mathbb{E}\left[\max\{0, B-a\}\right],$$

where $B \sim Bin(n,p)$.

The quantity $\inf_{a<t} \frac{1}{t-a}\mathbb{E}\left[\max\{0, B-a\}\right]$ is estimated in [2, Lemma 4.2]. We will see in the forthcoming sections that Bentkus' result is optimal in a slightly broader sense, i.e., it is the best bound that can be obtained from the inequality

$$\mathbb{P}\left[\sum_{i=1}^{n} X_i \geq t\right] \leq \frac{1}{f(t)}\mathbb{E}[f(B)],$$

where f is a non-negative, convex and increasing function.

We shall be interested in employing the Bernstein-Hoeffding method to a larger class of generalized moments. Such approaches have been already considered by Bentkus [2], Eaton [4], Pinelis [16],[18]. Nevertheless, we were not able to find a systematic study of the classes

of functions that are considered in our paper. We now proceed by defining a class of functions that pertains to the Berstein-Hoeffding method. For fixed $t > 0$, we denote by $\mathcal{F}_{ic}(t)$ the class of functions

$$\mathcal{F}_{ic}(t) := \{f : [0, \infty) \to [0, \infty) : f \text{ is convex, increasing on } [t, +\infty) \text{ and } f(t) = 1\}.$$

Examples of functions belonging to the class $\mathcal{F}_{ic}(t)$ are: $f(x) = \frac{|x-\epsilon|}{t-\epsilon}$ for fixed $\epsilon < t$, $f(x) = \frac{1}{t-\epsilon}\max(0, x-\epsilon)$ for fixed $\epsilon < t$, $f(x) = e^{h(x-t)}$ for $h > 0$ and so on. Our first result shows that the Bernstein-Hoeffding method can be adapted to the class $\mathcal{F}_{ic}(t)$.

Theorem 1.4. *Let the random variables $X_1, \dots, X_n$ be independent and such that $0 \leq X_i \leq 1$, for each $i = 1, \dots, n$. Set $p = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i]$. Then the following hold true.*

a) *For any fixed real number t such that $np < t < n$, we have*

$$\mathbb{P}\left[\sum_{i=1}^{n} X_i \geq t\right] \leq \inf_{f \in \mathcal{F}_{ic}(t)} \mathbb{E}[f(B)],$$

where $B \sim Bin(n, p)$ is a binomial random variable and $\mathcal{F}_{ic}(t)$ is the class of functions defined above.

b) *Moreover, we have*

$$\inf_{f \in \mathcal{F}_{ic}(t)} \mathbb{E}[f(B)] = \inf_{a<t} \frac{1}{t-a}\mathbb{E}\left[\max\{0, B-a\}\right].$$

By employing Theorem 1.4 to a particular function from the class $\mathcal{F}_{ic}(t)$, we deduce the following improvement upon Hoeffding's inequality.

Theorem 1.5. *Let the random variables $X_1, \dots, X_n$ be independent and such that $0 \leq X_i \leq 1$, for each $i = 1, \dots, n$. Set $p = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i]$ and let t be a fixed* positive integer *such that $\frac{enp}{ep-p+1} \leq t < n$. Then*

$$\mathbb{P}\left[\sum_{i=1}^{n} X_i \geq t\right] \leq \frac{1+h}{e^h} \cdot (H(n,p,t) - T(n,p,t;h)) + \left(1 - \frac{1+h}{e^h}\right)\mathbb{P}\left[B_{n,p} = t\right],$$

where $H(n,p,t)$ is the Hoeffding function, $B_{n,p}$ is a binomial random variable of parameters n and p,

$$T(n,p,t;h) = \sum_{i=0}^{t-1} e^{h(i-t)}\mathbb{P}\left[B_{n,p} = i\right],$$

and h is such that $e^h = \frac{t(1-p)}{p(n-t)}$, i.e., it is the optimal real such that

$$\frac{1}{e^{ht}}\mathbb{E}[e^{hB}] = \inf_{s>0} \frac{1}{e^{st}}\mathbb{E}[e^{sB}],$$

with $B \sim Bin(n, p)$.

Let us illustrate that the bound of the previous result is indeed an improvement upon Hoeffding's inequality. To see this, notice that the bound provided by Theorem 1.5 is

$$\leq \frac{1+h}{e^h} \cdot H(n,p,t) + \left(1 - \frac{1+h}{e^h}\right)\mathbb{P}\left[B_{n,p} = t\right]$$

and the later quantity is a convex combination of $H(n,p,t)$ and $\mathbb{P}\left[B_{n,p} = t\right]$. Now Hoeffding's Theorem 1.1 implies that $\mathbb{P}\left[B_{n,p} = t\right] \leq \mathbb{P}\left[B_{n,p} \geq t\right] \leq H(n,p,t)$ and therefore the bound of the

previous result is *smaller* than Hoeffding's. Let us also mention that, when t is *not* an integer, one may use the bound of Theorem 1.5 with t replaced by $\lfloor t \rfloor := \max\{k \in \mathbb{N} : k \leq t\}$, since

$$\mathbb{P}\left[\sum_{i=1}^{n} X_i \geq t\right] \leq \mathbb{P}\left[\sum_{i=1}^{n} X_i \geq \lfloor t \rfloor\right].$$

In other words, Theorem 1.5 improves upon Hoeffding's by adjusting "missing factors" in the Hoeffding bound. Notice that the bound provided by Theorem 1.5 holds true for large t, i.e., for t in the interval $\left[\frac{enp}{ep-p+1}, n\right)$, in contrast to Talagrand's result which holds true for moderate values of t, and so it may be seen as complementary to Theorem 1.2. It is unclear how to see whether the intervals $[np + K, np + np(1-p)/K]$ and $\left[\frac{enp}{ep-p+1}, n\right)$ overlap without knowing the constant K. In case the intervals do overlap, it may be informative to include a few words on comparison between the two factors. Since $e^h = \frac{t(1-p)}{p(n-t)}$ it follows that the "missing" factor of Theorem 1.5 can be written as

$$\frac{1+h}{e^h} = \frac{p}{1-p}\left(\frac{n}{t} - 1\right)\left(1 + \ln\frac{1-p}{p(n/t-1)}\right).$$

On the other hand, Talagrand's result provides a factor that is approximately

$$\frac{\sqrt{np(1-p)}}{\sqrt{2\pi}(\sqrt{np(1-p)} + t - np)} + \frac{K}{\sqrt{np(1-p)}}.$$

If we assume that $K = 0$ then elementary, though quite tedious, calculations show that Talagrand's bound is sharper than the bound of Theorem 1.5. Our bound has the advantage that it does *not* involve unknown constants and that it is obtained using a rather simple argument (see also Fan et al. [5] for refinements of Talagrand's inequality having explicit values for the constants).

Our last result may be seen as an extension of Theorem 1.1 for sums of bounded, independent random variables whose first m moments are known. Before being more precise, let us first fix some notation. Given real numbers $\mu_1, \ldots, \mu_m \in (0,1)$, we denote by $\mathcal{B}(\mu_1, \ldots, \mu_m)$ the set of all $[0,1]$-valued random variables whose i-th moment equals $\mu_i, i = 1, \ldots, m$. Formally,

$$\mathcal{B}(\mu_1, \ldots, \mu_m) := \{X : 0 \leq X \leq 1, \mathbb{E}[X] = \mu_1, \mathbb{E}[X^2] = \mu_2 \ldots, \mathbb{E}[X^m] = \mu_m\}.$$

Notice that the set may be empty. Note also that if $\mathcal{B}(\mu_1, \ldots, \mu_m)$ is non-empty then we have $\mu_1 \geq \mu_2, \geq \cdots \geq \mu_m$. Recall the definition of the class $\mathcal{F}_{ic}(t)$, defined above.

Theorem 1.6. *Fix positive integers, $n, m \geq 2$ and for $i = 1, \ldots, n$ let $\{\mu_{ij}\}_{j=1}^{m}$ be a finite sequence of real numbers such that the class $\mathcal{B}(\mu_{i1} \ldots, \mu_{im})$ is non-empty. Let $X_1, \ldots, X_n$ be independent random variables such that $X_i \in \mathcal{B}(\mu_{i1}, \ldots, \mu_{im})$, for $i = 1, \ldots, n$, and fix $t \in (\mu, n)$, where $\mu = \sum_i \mu_{i1}$. Then*

$$\mathbb{P}\left[\sum_{i=1}^{n} X_i \geq t\right] \leq \inf_{f \in \mathcal{F}_{ic}(t)} \mathbb{E}\left[f(Z_{nm})\right],$$

where $Z_{nm} = \sum_{i=1}^{n} Z_i$ is an independent sum of random variables Z_i such that

$$\mathbb{P}\left[Z_i = j/m\right] = \binom{m}{j} \cdot \mathbb{E}\left[X_i^j (1 - X_i)^{m-j}\right], \text{ for } j = 0, 1, \ldots, m.$$

Moreover,

$$\inf_{f\in\mathcal{F}_{ic}(t)} \mathbb{E}\left[f(Z_{nm})\right] = \inf_{a<t} \frac{1}{t-a}\mathbb{E}\left[\max\{0, Z_{nm}-a\}\right].$$

To the best of our knowledge, this is the first result that considers the performance of the Bernstein-Hoeffding method under additional information on higher moments. Notice that the probability distribution of the random variable Z_{nm} depends solely on the given sequence of moments $\{\mu_{ij}\}_{i,j}$. Indeed, using the binomial formula, it is easy to see that

$$\mathbb{E}\left[X_i^j(1-X_i)^{m-j}\right] = \sum_{k=0}^{m-j}\binom{m-j}{k}(-1)^{m-j-k}\mu_{i,m-k}$$

and so each Z_i is uniquely determined by the given sequence of moments $\{\mu_{ij}\}_{i,j}$. Let us also mention that the random variables Z_i arise in the study of the so-called *Hausdorff moment problem* (see Feller [6]).

The remaining part of our article is organized as follows. In Section 2 we prove Theorem 1.4, by employing ideas from the theory of convex orders. Moreover, we show that the functions from $\mathcal{F}_{ic}(t)$ that minimize $\frac{1}{f(t)}\mathbb{E}[f(B)]$ are those that occur in the aforementioned result of Bentkus, i.e., Theorem 1.3. In Section 3 we prove Theorem 1.5 by employing Theorem 1.4 to a suitable class of functions. In Section 4, we prove Theorem 1.6 using ideas from the theory of Bernstein polynomials. Finally, in Section 5, we provide some pictorial comparisons between the bound given in Theorem 1.5 and a refinement of Hoeffding's that is due to Zheng [24].

2. Proof of Theorem 1.4

In this section we prove Theorem 1.4, which allows to improve upon Hoeffding' bound by suitably choosing function from the class $\mathcal{F}_{ic}(t)$. Notice that Theorem 1.4 implies that there may be some space for improvement upon Hoeffding's bound. We will employ this result and en route find a function $\phi \in \mathcal{F}_{ic}(t)$ such that

$$\mathbb{E}[\phi(B)] < \inf_{h>0} e^{-ht}\mathbb{E}[e^{hB}],$$

where $B \sim \mathrm{Bin}(n,p)$. Hence there is indeed space for improvement upon Hoeffding's bound.

The proof of Theorem 1.4 will require some well-known results and the following notion of ordering between random variables (see [20]).

Definition 2.1. Let X and Y be two random variables such that

$$\mathbb{E}[f(X)] \leq \mathbb{E}[f(Y)], \text{ for all convex functions } f:\mathbb{R}\to\mathbb{R},$$

provided the expectations exist. Then X is said to be smaller than Y in the *convex order*, denoted $X \leq_{cx} Y$.

We begin with the following, well-known, result whose proof is included for the sake of completeness.

Lemma 2.2. *Fix real numbers a, b such that $a < b$. Let X be a random variable that takes values on the interval $[a,b]$ and is such that $\mathbb{E}[X] = p$. Let B be the random variable that takes on the values a and b with probabilities $\frac{b-\mathbb{E}[X]}{b-a}$ and $\frac{\mathbb{E}[X]-a}{b-a}$, respectively. Then for any convex function, $f:[a,b]\to\mathbb{R}$, we have*

$$\mathbb{E}[f(X)] \leq \mathbb{E}[f(B)].$$

Proof. Given X, we couple the random variables by setting B_X to be either equal to a with probability $\frac{b-X}{b-a}$, or equal to b with probability $\frac{X-a}{b-a}$. It is easy to see that $\mathbb{E}[B_X|X] = X$ and so

$$\mathbb{E}[B_X] = \mathbb{E}[\mathbb{E}[B_X|X]] = \mathbb{E}[X] = p.$$

Jensen's inequality now implies that

$$\mathbb{E}[f(X)] = \mathbb{E}[f(\mathbb{E}[B_X|X])] \leq \mathbb{E}\left[\mathbb{E}[f(B_X|X)]\right] = \mathbb{E}[f(B_X)],$$

as required. □

The following two results are well-known (see Theorems 3.*A*.12 and 3.*A*.37 in [20] and Theorem 4 in [10]). The first one shows that convex order is closed under convolutions.

Lemma 2.3. *Let $X_1, \ldots, X_n$ be a set of independent random variables and let $Y_1, \ldots, Y_n$ be another set of independent random variables. If $X_i \leq_{cx} Y_i$, for $i = 1, \ldots, n$, then*

$$\sum_{i=1}^{n} X_i \leq_{cx} \sum_{i=1}^{n} Y_i.$$

The next lemma shows that a sum of independent Bernoulli random variables is dominated, in the sense of convex order, by a certain binomial random variable.

Lemma 2.4. *Fix n real numbers $p_1, \ldots, p_n$ from $(0, 1)$. Let $B_1, \ldots, B_n$ be independent Bernoulli random variables with $B_i \sim Ber(p_i)$. Then*

$$\sum_{i=1}^{n} B_i \leq_{cx} B,$$

where $B \sim Bin(n, p)$ is a binomial random variable of parameters n and $p := \frac{1}{n}\sum_i p_i$.

We are now ready to prove the first statement of Theorem 1.4.

Proof of Theorem 1.4, a). Fix $f \in \mathcal{F}_{ic}(t)$. Since $f(\cdot)$ is non-negative, increasing in $[t, \infty)$ and $f(t) = 1$, Markov's inequality implies that

$$\mathbb{P}\left[\sum_{i=1}^{n} X_i \geq t\right] \leq \mathbb{E}\left[f\left(\sum_{i=1}^{n} X_i\right)\right].$$

Since $f(\cdot)$ is convex, Lemmata 2.2 and 2.3 imply that

$$\mathbb{E}\left[f\left(\sum_{i=1}^{n} X_i\right)\right] \leq \mathbb{E}\left[f\left(\sum_{i=1}^{n} B_i\right)\right],$$

where $B_i \sim \mathrm{Ber}(\mathbb{E}[X_i]), i = 1, \ldots, n$. Now Lemma 2.4 yields

$$\mathbb{E}\left[f\left(\sum_{i=1}^{n} B_i\right)\right] \leq \mathbb{E}\left[f\left(B\right)\right]$$

and the result follows. □

Similar ideas as above have been employed to sums of independent Bernoulli random variables by León and Perron in [13].

We now proceed with the second statement of Theorem 1.4. Let the random variables $X_1, \ldots, X_n$ be independent and such that $0 \leq X_i \leq 1$, for each $i = 1, \ldots, n$. Set $p =$

$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i]$ and fix a real number t such that $np < t < n$. We have already seen that, for every $f \in \mathcal{F}_{ic}(t)$, it holds

$$\mathbb{P}\left[\sum_{i=1}^{n} X_i \geq t\right] \leq \mathbb{E}\left[f\left(B\right)\right], \text{ where } B \sim \mathrm{Bin}(n,p).$$

Notice that

$$\mathbb{E}[f(B)] = \sum_{i=0}^{n} f(i)\cdot \mathbb{P}[B=i].$$

Notice also that a function $f \in \mathcal{F}_{ic}(t)$ that minimizes $\mathbb{E}[f(B)]$ has to be such that $\mathbb{E}[f(B)] \leq \frac{np}{t}$; indeed, since $\frac{np}{t}$ is the bound on $\mathbb{P}\left[\sum_i X_i \geq t\right]$ given by Markov's inequality (or by Theorem 1.4,(a) applied to the function $f(x) = x/t$) it follows that an optimal function has to provide a bound that is at least as good. Hence, for the purpose of characterising the function that minimize $\mathbb{E}[f(B)]$, we may assume that f belongs to the class $\mathcal{F}^*_{ic}(t)$, where $\mathcal{F}^*_{ic}(t)$ consists of all functions in $f \in \mathcal{F}_{ic}(t)$ that satisfy $\mathbb{E}[f(B)] \leq \frac{np}{t}$. The following result characterizes the functions $f \in \mathcal{F}^*_{ic}(t)$ that minimize $\mathbb{E}[f(B)]$.

Theorem 2.5. *Let $f \in \mathcal{F}^*_{ic}(t)$. Then there exists $\varepsilon \in [0,t)$ such that $\mathbb{E}\left[\phi_\varepsilon(B)\right] \leq \mathbb{E}\left[f(B)\right]$, where $\phi_\varepsilon(x) = \max\{0, \frac{1}{t-\varepsilon}\cdot(x-\varepsilon)\}$.*

Proof. Let $m_t := \min\{n \in \mathbb{N} : t < n\}$ be the smallest positive integer that is *strictly* larger than t. Note that, by definition, $0 < m_t - t \leq 1$. Let $\varepsilon = \frac{tf(m_t)-m_t}{f(m_t)-1}$ and for $x \geq 0$ define the function

$$\phi_\varepsilon(x) := \max\left\{0, \frac{1}{t-\varepsilon}(x-\varepsilon)\right\}.$$

In other words, $\phi_\varepsilon(\cdot)$ equals zero for $x < \varepsilon$ and for $x \geq \varepsilon$ it is a straight line starting from point $(\varepsilon, 0) \in \mathbb{R}^2$ and passing through the points $(t, f(t))$ and $(m_t, f(m_t))$. Since the function $f(\cdot)$ is convex it follows that for every integer i in the interval $[0,n]$ we have $\phi_\varepsilon(i) \leq f(i)$ and this, in turn, implies

$$\mathbb{E}\left[\phi_\varepsilon(B)\right] \leq \mathbb{E}\left[f(B)\right].$$

Clearly, we have $\varepsilon < t$ and it remains to show that $\varepsilon \geq 0$. Indeed, if $\varepsilon < 0$, then $\phi_\varepsilon(0) > 0$ and the function $f_1(x) = \frac{1-\phi_\varepsilon(0)}{t}x + \phi_\varepsilon(0)$ is such that

$$\frac{np}{t} + \phi_\varepsilon(0)\left(1 - \frac{np}{t}\right) = \mathbb{E}[f_1(B)] \leq \mathbb{E}[f(B)].$$

This implies that $\mathbb{E}[f(B)]$ is even worse than the bound obtained by Markov's inequality, and contradicts the assumption that $f \in \mathcal{F}^*_{ic}(t)$. The result follows. □

In other words, Theorem 2.5 implies that, in order to minimize $\mathbb{E}[f(B)]$ for $f \in \mathcal{F}^*_{ic}(t)$, it is enough to consider functions of the form $\max\{0, \frac{1}{t-\varepsilon}\cdot(x-\varepsilon)\}$, for $\varepsilon \in [0,t)$. The following result is an immediate consequence of Theorem 2.5 and finishes the proof of the second statement of Theorem 1.4.

Corollary 2.6. *Let the parameters n, p, t be as in Theorem 1.4. Then for any $t \in (np, n)$ we have*

$$\inf_f \ \mathbb{E}[f(B)] = \inf_{a<t} \ \frac{1}{t-a}\mathbb{E}\left[\max\{0, B-a\}\right],$$

where $B \sim Bin(n,p)$ and the infimum on the left hand side is taken over all functions $f \in \mathcal{F}_{ic}(t)$.

Notice that we can write the function $\rho_\varepsilon(x) := \max\{0, \frac{1}{t-\varepsilon} \cdot (x-\varepsilon)\}$, for $\varepsilon \in [0,t)$, in the form $g_h(x) := \max\{0, h \cdot (x-t)+1\}$, where $h = \frac{1}{t-\varepsilon}$, and that this correspondence is injective. Notice also that, since $\varepsilon \geq 0$, we have $h \geq 1/t$. The following question arises naturally from Corollary 2.6.

Question 2.7. *What is the optimal ε such that*

$$\inf_{a<t} \frac{1}{t-a}\mathbb{E}\left[\max\{0, B-a\}\right] = \mathbb{E}\left[\rho_\varepsilon(B)\right] ?$$

We remark that such an ε will satisfy $\varepsilon \leq \lceil t \rceil - 1$, where $\lceil t \rceil := \min\{k \in \mathbb{N} : t \leq k\}$. To see this notice that if $\varepsilon > \lceil t \rceil - 1$, then $\rho_\varepsilon(\lceil t \rceil - 1) = 0$ and we may decrease ε, until it reaches the point $\lceil t \rceil - 1$, without increasing the value $\mathbb{E}\left[\rho_\varepsilon(B)\right]$. Since $\varepsilon \leq \lceil t \rceil - 1$ it follows that $h \leq \frac{1}{t-\lceil t \rceil + 1}$. Now, finding the optimal ε is equivalent to finding the optimal h. We are *not* able to find this h. Nevertheless, due to the following result, one can easily find h using, say, a binary search algorithm.

Proposition 2.8. *Let the parameters n, p, t be as in Theorem 1.4. Let $h > 0$ be such that*

$$\mathbb{E}\left[\max\{0, h \cdot (B-t)+1\}\right] = \inf_{s>0} \mathbb{E}\left[\max\{0, s \cdot (B-t)+1\}\right],$$

where $B \sim Bin(n,p)$. Then we may assume that $h = \frac{1}{t-j}$, for some positive integer $j \in \{0, 1, \ldots, \lceil t \rceil - 1\}$.

Proof. Recall that $h \in \left[\frac{1}{t}, \frac{1}{t+1-\lceil t \rceil}\right]$. We have

$$\mathbb{E}[g_h(B)] = \sum_{i=0}^{n} \binom{n}{i} p^i (1-p)^{n-i} \cdot g_h(i).$$

It is clear that the function $E(h) := \mathbb{E}[g_h(B)]$ is linear on the interval $\left[\frac{1}{t-j}, \frac{1}{t-j-1}\right]$, for every $j \in \{0, 1, \ldots, \lceil t \rceil - 1\}$. Hence the function $E(h)$ is continuous and piecewise linear on the interval $\left[\frac{1}{t}, \frac{1}{t-\lceil t \rceil + 1}\right]$ and this implies that it attains its minimum at the endpoints of $\left[\frac{1}{t-j}, \frac{1}{t-j-1}\right]$, for some $j \in \{0, 1, \ldots, \lceil t \rceil - 1\}$. The result follows. □

In the next section we obtain an improvement upon Hoeffding's bound.

3. Proof of Theorem 1.5

This section contains the proof of Theorem 1.5.

Proof of Theorem 1.5. Given $h > 0$ define the function $f(x) = \max\{0, h(x-t)+1\}$, for $x \geq 0$. Clearly, we have $f \in \mathcal{F}_{ic}(t)$. Let m_t be the largest positive integer for which $f(m_t) = 0$. Using Theorem 1.4 and the inequality $e^x > 1 + x$, for $x \in \mathbb{R}$, we estimate

$$\begin{aligned}
\mathbb{P}\left[\sum_{i=1}^{n} X_i \geq t\right] \leq \mathbb{E}[f(B)] &= \sum_{i=m_t+1}^{n} (h(i-t)+1)\mathbb{P}[B=i] \\
&< \sum_{i=m_t+1}^{n} e^{h(i-t)}\mathbb{P}[B=i] \\
&\leq H(n,p,t),
\end{aligned}$$

which shows that $\mathbb{E}[f(B)]$ is strictly smaller than Hoeffding's bound. Since we assume that $t \geq \frac{epn}{ep-p+1}$ it follows that $h \geq 1$ which in turn implies, since t is an integer, that $f(i) = 0$, for all $i \in \{0, 1, \ldots, t-1\}$. Hence we can write

$$\begin{aligned} H(n,p,t) - \mathbb{E}[f(B)] &= \sum_{i=0}^{n} e^{h(i-t)}\mathbb{P}[B=i] - \sum_{i=t+1}^{n} (h(i-t)+1)\mathbb{P}[B=i] \\ &= \sum_{i=0}^{t-1} e^{h(i-t)}\mathbb{P}[B=i] \\ &+ \sum_{i=t+1}^{n} \left(e^{h(i-t)} - (h(i-t)+1)\right)\mathbb{P}[B=i]. \end{aligned}$$

For $i \geq t+1$, we have

$$\begin{aligned} e^{h(i-t)} - (h(i-t)+1) &= \left(1 - \frac{1+h(i-t)}{e^{h(i-t)}}\right) e^{h(i-t)} \\ &\geq \left(1 - \frac{1+h}{e^{h}}\right) e^{h(i-t)} \end{aligned}$$

which implies that

$$\begin{aligned} H(n,p,t) - \mathbb{E}[f(B)] &\geq \left(1 - \frac{1+h}{e^{h}}\right) H(n,p,t) + \frac{1+h}{e^{h}} \cdot \sum_{i=0}^{t-1} e^{h(i-t)}\mathbb{P}[B=i] \\ &- \left(1 - \frac{1+h}{e^{h}}\right) \mathbb{P}\left[B_{n,p} = t\right]. \end{aligned}$$

The result follows. □

4. Proof of Theorem 1.6

In this section we prove Theorem 1.6. The proof borrows ideas from the theory of Bernstein polynomials (see Phillips [15, Chapter 7]). Recall that, for a function $f : [0,1] \to \mathbb{R}$, the *Bernstein polynomial* corresponding to f is defined as

$$B_m(f,x) = \sum_{j=0}^{m} \binom{m}{j} x^j (1-x)^{n-j} f\left(j/m\right),$$

for each positive integer m. The following is a folklore result regarding Bernstein polynomials.

Lemma 4.1. *If $f : [0,1] \to [0,\infty)$ is convex, then*

$$f(x) \leq B_m(f,x), \text{ for all } x \in [0,1].$$

If $f : [0,1] \to [0,\infty)$ is continuous, then

$$\sup_{x \in [0,1]} |f(x) - B_m(f,x)| \to 0, \text{ as } m \to \infty.$$

Proof. See [15] Theorems 7.1.5 and 7.1.8. We remark that the first statement is easy to prove and the second arose from Bernstein's search for a proof of Weierstrass' theorem. □

We can now provide a proof of Theorem 1.6.

Proof of Theorem 1.6. Let $f \in \mathcal{F}_{ic}(t)$. Since f is non-negative and increasing and $f(t) = 1$, Markov's inequality yields

$$\mathbb{P}\left[\sum_{i=1}^{n} X_i \geq t\right] \leq \mathbb{E}\left[f\left(\sum_{i=1}^{n} X_i\right)\right].$$

Since f is convex and $X_i \in [0,1]$, Lemma 4.1 implies that

$$\mathbb{E}\left[f\left(X_i\right)\right] \leq \mathbb{E}\left[B_m\left(f, X_i\right)\right].$$

Now note that

$$\mathbb{E}\left[B_m\left(f, X_i\right)\right] = \sum_{j=0}^{m} \binom{m}{j} \cdot \mathbb{E}\left[X_i^j (1 - X_i)^{m-j}\right] \cdot f(j/m).$$

For $j = 0, 1 \ldots, m$ let

$$\pi_j := \binom{m}{j} \cdot \mathbb{E}\left[X_i^j (1 - X_i)^{m-j}\right].$$

Notice also that

$$\mathbb{E}\left[X_i^j (1 - X_i)^{m-j}\right] = \sum_{k=0}^{m-j} \binom{m-j}{k} (-1)^{m-j-k} \mu_{i,m-k}$$

which implies that $\mathbb{E}\left[X_i^j (1 - X_i)^{m-j}\right]$ is the same for all random variables from the class $\mathcal{B}(\mu_{i,1}, \ldots, \mu_{i,m})$. It is easy to verify that $\sum_{j=0}^{m} \pi_j = 1$. Now, if we define the random variable Z_i that takes on the value $\frac{j}{m}$ with probability $\pi_j, j = 0, 1, \ldots, m$, we have $\mathbb{E}[f(X_i)] \leq \mathbb{E}[B_m(f, X_i)] = \mathbb{E}[f(Z_i)]$, for convex f. Since the convex order is closed under convolutions, by Lemma 2.3, the first statement follows. The proof of the second statement is almost identical to the proof of Theorem 2.5 and is left to the reader. □

5. Comparisons

In this section we perform pictorial comparisons between the bound given by Theorem 1.5 and a refinement of Hoeffding's bound which is due to Zheng [24]. In [24] it is shown, under the same assumptions as in Theorem 1.5, that certain refinements of the arithmetic-geometric means inequality yield the following estimate:

$$\mathbb{P}\left[\sum_{i=1}^{n} X_i \geq t\right] \leq w^{-t}\left(1 - p + pw - \frac{\sigma^2 (w-1)^2}{2w}\right)^n, \text{ for } np < t < n, \tag{1}$$

where $w = \frac{-B + \sqrt{B^2 - 4AC}}{2A}$ and $A = (1 - t/n)(p - \sigma^2/2)$, $B = -\frac{t}{n}(1 - p + \sigma^2)$, $C = \frac{\sigma^2}{2}(1 + t/n)$ and $\sigma^2 = \frac{1}{n}\sum_{i=1}^{n} (\mathbb{E}[X_i] - p)^2$.

Let us remark that comparisons between the bound provided in Theorem 1.5 and the bound given in (1) require quite tedious calculations. However, it is rather straightforward to put the computer to work and see that the two bounds are quite close to each other.

More precisely, let $Z(n, p, t)$ be the right hand side of (1) and let $Y(n, p, t)$ be the bound given in Theorem 1.5. In Figure 1 we fix the value of n, draw n numbers $p_1, \ldots, p_n$ from $[0, 1]$ (which serve as the expected values of the random variables) uniformly at random and then plot the function $f(t) = Y(n, p, \lfloor t \rfloor) - Z(n, p, t)$ for $t \in \left[\frac{enp}{ep - p + 1}, n\right)$. For moderate vales of t the bound given by (1) performs slightly better than our bound; while for larger values of t the two bounds are almost equal. In all cases, the two bounds are very close to each other. Notice that

as n gets larger the difference between the bounds appears to get closer to zero.

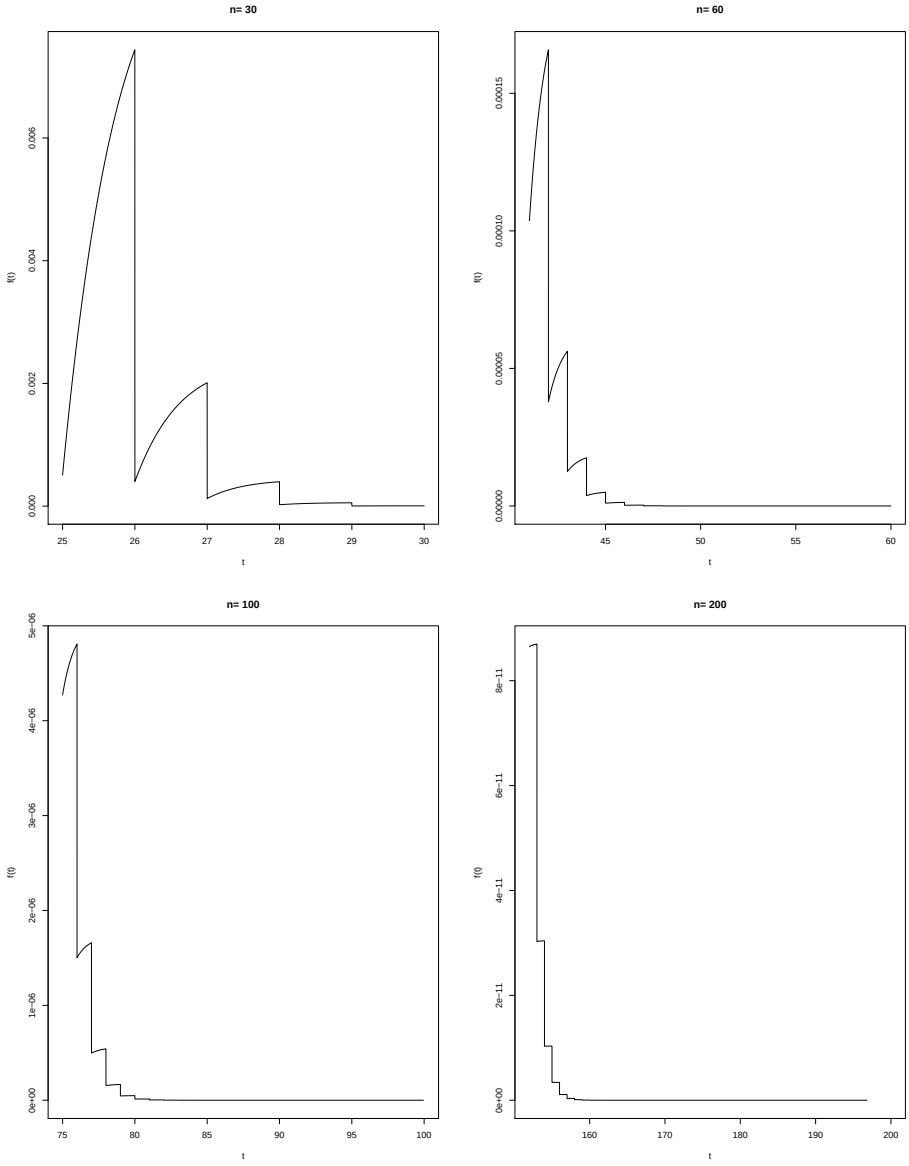

FIGURE 1. Pictorial comparisons between the bound of Theorem 1.5 and the bound given in Zheng [24].

Acknowledgements. The authors were supported by European Research Council Starting Grant 240186 "MiGraNT, Mining Graphs and Networks: a Theory-based approach". We are grateful to Dr. Xiequan Fan and to an anonymous referee for valuable suggestions and comments that improved the presentation of the paper.

References

[1] Bentkus, V.: A remark on Bernstein, Prokhorov, Bennett, Hoeffding and Talagrand inequalities, *Lithuanian Math. Journal* **42** (3) (2002) 262–269.

[2] Bentkus, V.: On Hoeffding's inequalities *Annals of Probability* **32**(2) (2004) 1650–1673.

[3] Bentkus, V., Geuze, G.D.C., Van Zuilen, M.C.A.: Optimal Hoeffding-like inequalities under a symmetry assumption, *Statistics* **40** (2) (2006) 159–164.

[4] Eaton, M.L.: A probability inequality for linear combinations of bounded random variables, *Annals of Statistics* **2** (3) (1974) 609–613.

[5] Fan, X., Grama, I., Liu, Q.: Sharp large deviation probabilities for sums of independent bounded random variables, *Sci China Math* **58** (9) (2015) 1939–1958.

[6] Feller, W.: An introduction to probability theory and its applications, Vol. 2, *Wiley* New York (1957).

[7] From, S.G.: An Improved Hoeffding's Inequality of Closed Form Using Refinemets of the Arithmetic Mean-Geometric Mean Inequality, *Communications in Statistics-Theory and Methods* (2013), doi: 10.1080/03610926.2012.756913

[8] From, S.G., Swift, A.W.: A refinement of Hoeffding's inequality, *J. of Stat. Computation and Simulation* **83** (5) (2013) 977–983.

[9] Györfi, L., Harremoës, P., Tusnády, G.: Some refinements of large deviation tail probabilities, arXiv:1205.1005

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

[11] Krafft, O., Schmitz, N.: A note on Hoeffding's inequality, *J. Amer. Statist. Assoc.* **64** (327) (1969) 907–912.

[12] McDiarmid, C.: On the method of bounded differences, *London Math. Soc. Lecture Note Ser.* **141** (1989) 148–188.

[13] León, C.A., Perron, F.: Extremal properties of sums of Bernoulli random variables, *Statistics & Probability Letters* **62** (2003) 345–354.

[14] Kha, F.D., Nagaev, S.V.: Probability inequalities for sums of independent random variables, *Theory of Probab. Appl.* **16** (4) (1971) 643–660.

[15] Phillips, G.M.: Interpolation and approximation by polynomials, *Springer* Verlag (2003).

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

[17] Pinelis, I.: On inequalities for sums of bounded random variables, *J. Math. Inequalities* **2** (1) (2008) 1–7.

[18] Pinelis, I.: On the Bennett-Hoeffding inequality, *Ann. Inst. Henri Poincaré Probab. Stat.* **50** (1) (2014) 15–27.

[19] Schmidt, J.P., Siegel, A., Srinivasan, A.: Chernoff-Hoeffding bounds for applications with limited independence, *SIAM J. Disc. Math*, **8** (2) (1995) 223–250.

[20] Shaked, M., Shanthikumar, G.J.: Stochastic Orders, *Springer*, New York (2007).

[21] Siegel, A.: Towards a usable theory of Chernoff-Hoeffding bounds for heterogeneous and partially dependent random variables, manuscript (1992)

[22] Talagrand, M.: The missing factor in Hoeffding's inequalities, *Ann. Inst. Henri Poincaré Probab. Stat.* **31**, 689–702 (1995).

[23] Xia, Y.: Two refinements of the Chernoff bound for the sum of nonidentical Bernoulli random variables, *Statistics & Probability Letters*, **78** (12), 1557–1559 (2008).

[24] Zheng, S.: A refined Hoeffding's upper tail probability bound for sums of independent random variables, *Statistics & Probability Letters* **131** (2017) 87–92.

Institute of Mathematics, Czech Academy of Sciences, Žitná 25, 115 67, Praha 1 Czech Republic.
E-mail address: pelekis.chr@gmail.com

INRIA Lille, 40 avenue Halley 59650 Villeneuve d'Ascq, France
E-mail address: Jan.Ramon@inria.fr

ETH Zurich, Distributed Computing Group, Gloriastrasse 35, 8092 Zurich, Switzerland
E-mail address: yuwang@ethz.ch