PROBABILITY
AND
MATHEMATICAL STATISTICS
Vol. 30, Fasc. 2 (2010), pp. 185–205

IMPROVED BOUNDS ON BELL NUMBERS AND ON MOMENTS OF SUMS OF RANDOM VARIABLES

BY

DANIEL BEREND (BEER SHEVA, ISRAEL) AND TAMIR TASSA (RA'ANANA, ISRAEL)

Abstract. We provide bounds for moments of sums of sequences of independent random variables. Concentrating on uniformly bounded non-negative random variables, we are able to improve upon previous results due to Johnson et al. [10] and Latała [12]. Our basic results provide bounds involving Stirling numbers of the second kind and Bell numbers. By deriving novel effective bounds on Bell numbers and the related Bell function, we are able to translate our moment bounds to explicit ones, which are tighter than previous bounds. The study was motivated by a problem in operation research, in which it was required to estimate the L_p-moments of sums of uniformly bounded non-negative random variables (representing the processing times of jobs that were assigned to some machine) in terms of the expectation of their sum.

2000 AMS Mathematics Subject Classification: Primary: 60E15; Secondary: 05A18, 11B73, 26D15.

Key words and phrases: Sums of random variables, moments, bounds on moments, binomial distribution, Poisson distribution, Stirling numbers, Bell numbers.

1. INTRODUCTION

Numerous results in probability theory relate to sums of random variables. The law of large numbers and the central limit theorem are such instances. In this paper we deal, more specifically, with bounds for moments of sums of independent random variables. A classical result in this direction, due to Khintchine [11], is the following:

Let $X_1, X_2, \ldots, X_n$ be i.i.d. symmetric random variables assuming only the values 1 and -1. Then for every $p > 0$ there exist positive constants C_1, C_2, depending only on p, such that, for every choice of $a_1, a_2, \ldots, a_n \in \mathbf{R}$,

$$C_1\Big(\sum_{k=1}^{n} a_k^2\Big)^{1/2} \leqslant \Big\|\sum_{k=1}^{n} a_k X_k\Big\|_p \leqslant C_2\Big(\sum_{k=1}^{n} a_k^2\Big)^{1/2},$$

where $\|X\|_p = [E(|X|^p)]^{1/p}$ is the L_p-norm of X.

For more on results of this type we refer, for example, to [6].

Hereinafter we consider (possibly infinite) sequences of independent non-negative random variables X_i, $1 \leqslant i \leqslant t$, and aim at estimating the p-moments, $E(X^p)$, $p \geqslant 1$, of their sum, $X = \sum_{i=1}^t X_i$. The work of Latała [12] provides the last word in this subject. One of the corollaries of his work is the following one:

THEOREM A ([12], Theorem 1 and Lemma 8). *Let* X_i, $1 \leqslant i \leqslant t < \infty$, *be independent non-negative random variables*, $X = \sum_{i=1}^t X_i$, *and* $p \geqslant 1$. *Then, for every* $c > 0$,

$$\|X\|_p \leqslant 2e \cdot \max\left\{ \frac{(1+c)^p}{cp}\, \mathcal{E}_1, \left(1 + \frac{1}{c}\right) p^{-1/p}\, \mathcal{E}_p \right\}, \tag{1.1}$$

where

$$\mathcal{E}_k = \big(\sum_{i=1}^t E(X_i^k) \big)^{1/k}, \qquad k = 1, p. \tag{1.2}$$

In his Corollary 3, Latała [12] suggests to take $c = (\ln p)/p$ in (1.1). This yields the following uniform bound:

$$\|X\|_p \leqslant 2e \cdot \left(1 + \frac{p}{\ln p}\right) \cdot \max\{\mathcal{E}_1, \mathcal{E}_p\}, \qquad p > 1. \tag{1.3}$$

In fact, taking $c = p^{1/p} - 1$, the coefficients of both $\mathcal{E}_1$ and $\mathcal{E}_p$ in (1.1) coincide and equal $(p^{1/p} - 1)^{-1}$. Since the latter expression is bounded from above by $p/\ln p$, we may improve (1.3) and arrive at the following:

THEOREM A$'$. *Under the assumptions of Theorem* A,

$$\|X\|_p \leqslant 2e \cdot \frac{p}{\ln p} \cdot \max\{\mathcal{E}_1, \mathcal{E}_p\}, \qquad p > 1. \tag{1.4}$$

Finally, by finding the value of c that minimizes the upper bound in (1.1), we arrive at the following explicit version of Latała's theorem (the proof of which is postponed to the Appendix).

THEOREM A$''$. *Under the assumptions of Theorem* A,

$$\|X\|_p \leqslant \begin{cases} 2e \cdot \left(\dfrac{p}{p-1}\right)^{p-1} \cdot \mathcal{E}_1, & \mathcal{E}_1 \geqslant \dfrac{(p-1)^{p-1}}{p^{(p-1)^2/p}} \mathcal{E}_p, \\ 2e \cdot \dfrac{\mathcal{E}_p^{p/(p-1)}}{p^{1/p} \mathcal{E}_p^{1/(p-1)} - \mathcal{E}_1^{1/(p-1)}}, & \textit{otherwise}. \end{cases} \tag{1.5}$$

An estimate similar to that of Theorem A$'$, but with a better constant, was established already in [10]. It was shown there (Theorem 2.5) that

$$\|X\|_p \leqslant K \cdot \frac{p}{\ln p} \cdot \max\{\mathcal{E}_1, \mathcal{E}_p\}, \qquad K = 2,\ p > 1. \tag{1.6}$$

In addition, it was shown that estimate (1.6) cannot hold with $K < e^{-1}$ (see Proposition 2.9 in [10]). We note that Theorem A$''$ may offer better estimates than (1.6), even with $K < e^{-1}$. For example, if

$$\mathcal{E}_1 \geqslant \frac{(p-1)^{p-1}}{p^{(p-1)^2/p}} \mathcal{E}_p,$$

then the upper bound in Theorem A$''$ is less than $2e^2\mathcal{E}_1$, while the upper bound in (1.6) is at least $\big(2p/(\ln p)\big)\mathcal{E}_1$. In other cases, however, (1.6) may be sharper than Theorem A$''$.

In this paper we deal with the same setup, but restrict our attention to the case of uniformly bounded random variables, where $0 \leqslant X_i \leqslant 1$ for all $1 \leqslant i \leqslant t$. In this case, $\mathcal{E}_p \leqslant \mathcal{E}_1^{1/p}$, whence the existing bounds yield bounds depending only on $\mathcal{E}_1$. We derive here improved bounds in that case that depend only on $\mathcal{E}_1$. Our bounds offer sharper constants and are almost tight as $p \to 1^+$, where the previous explicit bounds, (1.3) and (1.6), explode. Our results apply also to possibly infinite sequences.

Some of our bounds involve Stirling numbers of the second kind and Bell numbers. A key ingredient in our estimates are new and effective bounds on Bell numbers. While previous bounds on Bell numbers were only asymptotic, we provide here an effective bound on Bell numbers that applies to all natural numbers. That result is an interesting result on its own right and may have applications for other problems of discrete mathematics.

Section 2 includes our main results. In Subsection 2.1 we provide a short discussion of Stirling and Bell numbers, and a statement of our bounds on Bell numbers and the related Bell function. In Subsection 2.2 we state our main results concerning bounds for sums of random variables. The proofs of our results are given in Section 3. In Section 4 we discuss our results and compare them with the best previously known ones (namely, with (1.6), due to Johnson et al. [10]). Finally, the Appendix includes proofs of some of our statements (Subsections 5.1, 5.2 and 5.3) and a description of the operation research problem that motivated this study (Subsection 5.4).

2. MAIN RESULTS

2.1. Effective bounds on Bell numbers. The Stirling numbers of the second kind and Bell numbers are related to counting the number of partitions of sets. The first counts, for integers $n \geqslant k \geqslant 1$, the number of partitions of a set of size n into k non-empty sets. This number is denoted by $S(n, k)$ (or sometimes by $\left\{ n \atop k \right\}$). The second counts the number of all partitions of a set of size n, and is denoted by B_n. (For more information we refer to Riordan [16].)

Let $A(n, k)$ denote the number of colorings of n elements using exactly k colors (namely, each color must be used at least once in the coloring). $A(n, k)$ is given by

(2.1)
$$A(n,k) = \sum \left\{ \binom{n}{\mathbf{r}} : \mathbf{r} = (r_1, \ldots, r_k) \in \mathbf{N}^k,\ |\mathbf{r}| = n \text{ and } r_i \geqslant 1,\ 1 \leqslant i \leqslant k \right\},$$

where $|\mathbf{r}| = \sum_{i=1}^{k} r_i$ and $\binom{n}{\mathbf{r}} = n!/(r_1! \ldots r_k!)$. With this notation, the Stirling number of the second kind may be expressed as

$$S(n,k) = \frac{A(n,k)}{k!}. \tag{2.2}$$

The Bell number B_n may be written in terms of the Stirling numbers:

$$B_n = \sum_{k=1}^{n} S(n,k). \tag{2.3}$$

In [2], de Bruijn derives the following asymptotic estimate for the Bell number B_n:

$$\frac{\ln B_n}{n} = \ln n - \ln\ln n - 1 + \frac{\ln\ln n}{\ln n} + \frac{1}{\ln n} + \frac{1}{2}\left(\frac{\ln\ln n}{\ln n}\right)^2 + O\left[\frac{\ln\ln n}{(\ln n)^2}\right].$$

In particular, we conclude that for every $\varepsilon > 0$ there exists $n_0 = n_0(\varepsilon)$ such that, for all $n > n_0$,

$$\left(\frac{n}{e\ln n}\right)^n < B_n < \left(\frac{n}{e^{1-\varepsilon}\ln n}\right)^n. \tag{2.4}$$

The problem with estimate (2.4) is that it is *ineffective* in the sense that the value of $n_0 = n_0(\varepsilon)$ is implicit in the asymptotic analysis in [2]. We prove here an upper bound that is less tight than (2.4), but applies for all n.

THEOREM 2.1. *The Bell numbers satisfy*

$$B_n < \left(\frac{0.792n}{\ln(n+1)}\right)^n, \qquad n \in \mathbf{N}. \tag{2.5}$$

By Dobinski's formula (see [5] and [15]),

$$B_n = \frac{1}{e}\sum_{k=1}^{\infty} \frac{k^n}{k!}. \tag{2.6}$$

This suggests a natural extension of the Bell numbers for any real index p,

$$B_p = \frac{1}{e}\sum_{k=1}^{\infty} \frac{k^p}{k!}. \tag{2.7}$$

We refer to B_p as the *Bell function*. Note that, for $p \geqslant 0$, we have $B_p = E(Y^p)$, where $Y \sim P(1)$ is a Poisson random variable with mean $E(Y) = 1$. For the sake of proving Theorem 2.1, we establish *effective* asymptotic bounds for the Bell function:

THEOREM 2.2. *The Bell function B_p, given by* (2.7), *satisfies for all $\varepsilon > 0$*

$$B_p < \left(\frac{e^{-0.6+\varepsilon}p}{\ln(p+1)}\right)^p, \quad p > p_0(\varepsilon), \tag{2.8}$$

where

$$p_0(\varepsilon) = \max\{e^4, d^{-1}(\varepsilon)\} \tag{2.9}$$

and $d(\cdot)$ is given by

$$d(p) = \ln\ln(p+1) - \ln\ln p + \frac{1+e^{-1}}{\ln p}. \tag{2.10}$$

2.2. Bounding moments of sums of random variables. Throughout this paper we let X_i, $1 \leqslant i \leqslant t \leqslant \infty$, be a (possibly infinite) sequence of independent random variables for which $P(0 \leqslant X_i \leqslant 1) = 1$, $X = \sum_{i=1}^{t} X_i$, and $\mu = E(X)$.

Our first result is an estimate for moments of integral order.

THEOREM 2.3. *Letting $S(p,k)$ denote the Stirling number of the second kind, the following estimate holds*:

$$E(X^p) \leqslant \sum_{k=1}^{\min(t,p)} S(p,k) \cdot E(X)^k \cdot e^{-(k-1)k/2t}, \quad p \in \mathbf{N}. \tag{2.11}$$

The above estimate implies that $E(X^p) \leqslant \sum_{k=1}^{p} S(p,k) \cdot E(X)^k$ for all t. Hence, using (2.3) we infer the following bound:

COROLLARY 2.1. *Letting B_p denote the p-th Bell number*,

$$E(X^p) \leqslant B_p \cdot \max\{E(X), E(X)^p\}, \quad p \in \mathbf{N}. \tag{2.12}$$

(In Section 5.3 we give an alternative proof of Corollary 2.1 that relies on the results of de la Peña et al. [14].)

Relying on Corollary 2.1 and an interpolation argument, we arrive at the following explicit bound for all real $p \geqslant 1$:

THEOREM 2.4. *For all $p \geqslant 1$*,

$$\|X\|_p \leqslant 0.792 \cdot \nu(p) \cdot \frac{p}{\ln(p+1)} \cdot \max\{E(X)^{1/p}, E(X)\}, \tag{2.13}$$

where

$$\nu(p) = \left(1 + \frac{1}{\lfloor p \rfloor}\right)^{(\{p\}\cdot(1-\{p\}))/p} \tag{2.14}$$

and $\lfloor p \rfloor$ and $\{p\}$ denote the integer and fractional parts of p, respectively.

We note that for all $p \geqslant 1$

$$\nu(p) \leqslant \left(1 + \frac{1}{\lfloor p \rfloor}\right)^{1/(4p)} \leqslant 2^{1/4},$$

so that inequality (2.13) may be simplified into the weaker form:

$$\|X\|_p \leqslant 0.942 \cdot \frac{p}{\ln(p+1)} \cdot \max\{E(X)^{1/p}, E(X)\}, \quad p \geqslant 1. \tag{2.15}$$

Finally, using the asymptotic upper bounds for the Bell numbers provided by (2.4) and by Theorem 2.2, we may obtain better estimates for large moment orders:

THEOREM 2.5. *For all $p \geqslant 1$ and $\varepsilon > 0$, let $\nu(p)$ be as given in* (2.14) *and $p_0(\varepsilon)$ be as given in* (2.9) *and* (2.10). *Then, for all $\varepsilon > 0$ and $p > p_0(\varepsilon)$,*

$$\|X\|_p \leqslant e^{-0.6+\varepsilon} \cdot \nu(p) \cdot \frac{p}{\ln(p+1)} \cdot \max\{E(X)^{1/p}, E(X)\}. \tag{2.16}$$

In addition, for every $\varepsilon > 0$ there exists $\hat{p}_0(\varepsilon) > 1$ such that, for all $p > \hat{p}_0(\varepsilon)$,

$$\|X\|_p \leqslant e^{-1+\varepsilon} \cdot \nu(p) \cdot \frac{p}{\ln(p+1)} \cdot \max\{E(X)^{1/p}, E(X)\}. \tag{2.17}$$

Note that the constant $e^{-1+\varepsilon}$ in (2.17) cannot be replaced by any constant smaller than e^{-1}, as implied by the lower bound from (2.4) on Bell numbers, and by the previously mentioned result of Johnson et al. ([10], Proposition 2.9).

3. PROOFS OF THE MAIN RESULTS

3.1. Proofs of Theorems 2.1 and 2.2.

LEMMA 3.1. *Let $x_0 = x_0(p)$ be defined for all $p > 0$ by*

$$x_0 \ln x_0 = p. \tag{3.1}$$

Then

$$x_0 < \frac{\alpha p}{\ln p}, \quad p > e^{e+1}, \tag{3.2}$$

where $\alpha = 1 + 1/e$.

Proof. As the function $u(x) = x \ln x$ is increasing for all $x \geqslant 1$, it suffices to show that

$$u\left(\frac{\alpha p}{\ln p}\right) = \frac{\alpha p}{\ln p} \cdot \ln\left(\frac{\alpha p}{\ln p}\right) > p, \quad p > e^{e+1}.$$

This is easily seen to be equivalent to

$$(\alpha-1)p+\frac{\alpha p\ln\alpha}{\ln p}-\frac{\alpha p\ln\ln p}{\ln p}>0,\qquad p>e^{e+1},$$

or

$$v(p):=(\alpha-1)\ln p+\alpha\ln\alpha-\alpha\ln\ln p>0,\qquad p>e^{e+1}. \tag{3.3}$$

First, we observe that

$$v(e^{e+1})=v(e^{\alpha/(\alpha-1)})=(\alpha-1)\frac{\alpha}{\alpha-1}+\alpha\ln\alpha-\alpha\ln\frac{\alpha}{\alpha-1}=0. \tag{3.4}$$

As

$$v'(p)=\frac{\alpha-1}{p}-\frac{\alpha}{p\ln p}>0,\qquad p>e^{\alpha/(\alpha-1)}=e^{e+1}, \tag{3.5}$$

inequality (3.3) follows from (3.4) and (3.5). ■

P r o o f o f T h e o r e m 2.2. Employing (a version of) Stirling's formula (cf. [17]), we infer by (2.7) that

$$B_p\leqslant\frac{1}{e}\sum_{n=1}^{\infty}\frac{n^p}{\sqrt{2\pi n}(n/e)^n}<\sum_{n=1}^{\infty}\frac{e^n n^p}{n^n}.$$

Define

$$h(x)=h_p(x)=\ln\frac{e^x x^p}{x^x}=x+p\ln x-x\ln x,\qquad x\geqslant 1.$$

As $h'(x)=p/x-\ln x$, the function h increases on $[1,x_0]$ and decreases on $[x_0,\infty)$, where $x_0=x_0(p)>1$ is determined by (3.1). Hence,

$$h(x)\leqslant h(x_0)=x_0+p\ln x_0-x_0\ln x_0=x_0+p\ln x_0-p\,,\qquad x\geqslant 1.$$

Invoking the bound (3.2) on x_0, we infer that

$$h(x)<\frac{\alpha p}{\ln p}+p\ln p-p\ln\ln p+p(\ln\alpha-1),\qquad x\geqslant 1,$$

for all $p>e^{e+1}$. Using the definition of $d(p)$ in (2.10), we conclude that

$$h(x)<p\cdot\big(d(p)+\ln p-\ln\ln(p+1)+\ln\alpha-1\big).$$

It is easy to check that $d(p)$ is decreasing for $p>1$ and maps the interval $(1,\infty)$ onto the interval $(0,\infty)$. Hence, for every $\varepsilon>0$ there exists a single $p_0=p_0(\varepsilon)>1$

such that $d(p_0) = \varepsilon$ and $d(p) < \varepsilon$ for all $p > p_0$. We conclude that, for every $\varepsilon > 0$ and $p > \max\{e^{e+1}, d^{-1}(\varepsilon)\}$,

$$h(x) < p \cdot \big(\ln p - \ln\ln(p+1) + \varepsilon + \ln\alpha - 1\big). \tag{3.6}$$

For larger x we can do much better by observing that

$$h(x) = x - (x-p)\ln x < x - \frac{x}{2}\ln x < -x, \qquad x > 2p, \;\; p > e^4. \tag{3.7}$$

Combining (3.6) and (3.7), we conclude that for $p > \max\{e^4, d^{-1}(\varepsilon)\}$:

$$\begin{aligned} B_p &< \sum_{n=1}^{\infty} e^{h(n)} = \sum_{n=1}^{\lfloor 2p\rfloor} e^{h(n)} + \sum_{n=\lfloor 2p\rfloor+1}^{\infty} e^{h(n)} \\ &< \lfloor 2p\rfloor \exp\big(p\cdot(\ln p - \ln\ln(p+1) + \varepsilon + \ln\alpha - 1)\big) + \sum_{n=\lfloor 2p\rfloor+1}^{\infty} e^{-n}. \end{aligned} \tag{3.8}$$

The first term on the right-hand side of (3.8) may be bounded as follows for all $p > e^4$:

$$\begin{aligned} &\lfloor 2p\rfloor \cdot \exp\big(p\cdot(\ln p - \ln\ln(p+1) + \varepsilon + \ln\alpha - 1)\big) \\ &\qquad = \lfloor 2p\rfloor \cdot \left(\frac{e^{\varepsilon+\ln\alpha-1}p}{\ln(p+1)}\right)^p < \left(\frac{e^{-0.6007+\varepsilon}p}{\ln(p+1)}\right)^p. \end{aligned} \tag{3.9}$$

The second term on the right-hand side of (3.8) may be bounded by evaluating the sum of the geometric series,

$$\sum_{n=\lfloor 2p\rfloor+1}^{\infty} e^{-n} = \frac{e^{-(\lfloor 2p\rfloor+1)}}{1-e^{-1}} < e^{-\lfloor 2p\rfloor}. \tag{3.10}$$

Using (3.9) and (3.10) in (3.8) we arrive at

$$B_p < \left(\frac{e^{-0.6007+\varepsilon}p}{\ln(p+1)}\right)^p + e^{-\lfloor 2p\rfloor}, \qquad p > p_0(\varepsilon). \tag{3.11}$$

Finally, as it is easy to verify that

$$\left(\frac{e^{-0.6007+\varepsilon}p}{\ln(p+1)}\right)^p + e^{-\lfloor 2p\rfloor} \leqslant \left(\frac{e^{-0.6+\varepsilon}p}{\ln(p+1)}\right)^p, \qquad p > e^4, \tag{3.12}$$

the required estimate (2.8) follows from (3.11) and (3.12). ∎

P r o o f o f T h e o r e m 2.1. Using $\varepsilon = d(e^4) \approx 0.346$ in Theorem 2.2, we obtain

$$B_p < \left(\frac{0.776p}{\ln(p+1)}\right)^p, \quad p > e^4. \tag{3.13}$$

Since it may be verified that

$$B_p < \left(\frac{0.792p}{\ln(p+1)}\right)^p, \quad p = 1, 2, \ldots, 54 = \lfloor e^4 \rfloor, \tag{3.14}$$

the theorem follows from (3.13) and (3.14). ■

3.2. Proof of Theorem 2.3. The main tool in proving Theorem 2.3 is the following general-purpose proposition:

PROPOSITION 3.1. *For every convex function f,*

$$E\big(f(X)\big) \leqslant E\big(f(Y)\big), \tag{3.15}$$

where Y is a binomial random variable with distribution $Y \sim B\,(t, \mu/t)$ in case $t < \infty$, and a Poisson random variable with distribution $Y \sim P(\mu)$ otherwise.

A simplified version of this proposition, for the case where X is a sum of a finite sequence of Bernoulli random variables, appears in [9], Theorem 3. For the sake of completeness, we include a proof of this proposition in the Appendix. We now proceed to prove our first estimate that is stated in Theorem 2.3.

Let us put $\mu = E(X)$. Consider first the case of infinite sequences, $t = \infty$. In that case, according to Proposition 3.1, $E(X^p) \leqslant E(Y^p)$, where $Y \sim P(\mu)$ is a Poisson random variable. Defining $m_p(\mu) = E(Y^p)$, it may be shown that

$$m_{p+1}(\mu) = \mu \cdot \left(m_p(\mu) + \frac{dm_p(\mu)}{d\mu}\right)$$

(see [13]). This recursive relation implies, as shown in [13], that

$$E(Y^p) = \sum_{k=1}^{p} S(p,k) \cdot \mu^k.$$

Hence

$$E(X^p) \leqslant \sum_{k=1}^{p} S(p,k) \cdot \mu^k, \tag{3.16}$$

in accord with inequality (2.11) for $t = \infty$.

As for finite sequences, $t < \infty$, the bounds offered by Proposition 3.1 involve moments of the binomial distribution. The moment generating function of a binomial random variable $Y \sim B(t,q)$ is

(3.17)
$$M_Y(z) = E(z^Y) = \sum_{\ell=0}^{t} P(Y=\ell)z^\ell = \sum_{\ell=0}^{t} \binom{t}{\ell} q^\ell (1-q)^{t-\ell} z^\ell = (qz+1-q)^t.$$

This function enables the computation of the factorial moments of Y through

$$M_Y^{(k)}(1) = E\big(\prod_{i=0}^{k-1}(Y-i)\big).$$

Since

$$Y^p = \sum_{k=1}^{p} S(p,k) \prod_{i=0}^{k-1}(Y-i)$$

(see [7], p. 72), we have

$$E(Y^p) = \sum_{k=1}^{p} S(p,k) \cdot M_Y^{(k)}(1). \tag{3.18}$$

As (3.17) implies

$$M_Y^{(k)}(1) = q^k \cdot \prod_{i=0}^{k-1}(t-i) \tag{3.19}$$

(note that when $k \geqslant t+1$, we get $M_Y^{(k)}(1) = 0$), we infer by (3.18) and (3.19) that

$$E(Y^p) = \sum_{k=1}^{\min(t,p)} S(p,k) \cdot q^k \cdot \prod_{i=0}^{k-1}(t-i). \tag{3.20}$$

Using (3.20) in Proposition 3.1, where $Y \sim B(t, \mu/t)$, we conclude that, when $t < \infty$ and $p \in \mathbf{N}$,

$$E(X^p) \leqslant \sum_{k=1}^{\min(t,p)} S(p,k) \cdot \mu^k \cdot \prod_{i=0}^{k-1}\left(1 - \frac{i}{t}\right).$$

Finally, as

$$\prod_{i=0}^{k-1}\left(1 - \frac{i}{t}\right) \leqslant \prod_{i=0}^{k-1} e^{-i/t} = e^{-(k-1)k/2t},$$

we infer that

$$E(X^p) \leqslant \sum_{k=1}^{\min(t,p)} S(p,k) \cdot \mu^k \cdot e^{-(k-1)k/2t}.$$

This concludes the proof. ∎

Estimate (2.11) of Theorem 2.3 may be relaxed as follows:

$$E(X^p) \leqslant \sum_{k=1}^{\min(t,p)} S(p,k) \cdot E(X)^k, \quad p \in \mathbf{N}. \tag{3.21}$$

We proceed to describe a straightforward proof of (3.21) that does not depend on Proposition 3.1 and is much shorter than the proof given above.

P r o o f o f (3.21). As $X = \sum_{i=1}^{t} X_i$, we conclude that

$$X^p = \sum_{\{\mathbf{r}:|\mathbf{r}|=p\}} \binom{p}{\mathbf{r}} \mathbf{X}^{\mathbf{r}},$$

where $\mathbf{r} = (r_1, \ldots, r_t) \in \mathbf{N}^t$ is a multi-index of non-negative values, and

$$\mathbf{X} = (X_1, \ldots, X_t), \quad \mathbf{X}^{\mathbf{r}} = \prod_{i=1}^{t} X_i^{r_i}.$$

By the independence of the X_i's and their being bounded between zero and one we get

$$E(X^p) = \sum_{\{\mathbf{r}:|\mathbf{r}|=p\}} \binom{p}{\mathbf{r}} \prod_{1 \leqslant i \leqslant t} E(X_i^{r_i}) \leqslant \sum_{\{\mathbf{r}:|\mathbf{r}|=p\}} \binom{p}{\mathbf{r}} \prod_{\substack{1 \leqslant i \leqslant t \\ r_i > 0}} E(X_i). \tag{3.22}$$

The sum on the right-hand side of (3.22) may be split into partial sums, $\Sigma = \Sigma_1 + \Sigma_2 + \ldots + \Sigma_{\min(t,p)}$, where Σ_k denotes the sum of terms in which the product consists of k multiplicands. Thus, Σ_1 is the partial sum consisting of all terms of the form $E(X_i)$, $1 \leqslant i \leqslant t$, the partial sum Σ_2 includes all terms of the form $E(X_i)E(X_j)$, $1 \leqslant i < j \leqslant t$, and so forth. With this in mind, we may rewrite (3.22) as

$$E(X^p) \leqslant \sum_{k=1}^{\min(t,p)} A(p,k) \cdot \Big(\sum_{1 \leqslant i_1 < i_2 < \ldots < i_k \leqslant t} \prod_{j=1}^{k} E(X_{i_j}) \Big). \tag{3.23}$$

Next, we observe that for all k, $1 \leqslant k \leqslant \min(t,p)$,

$$E(X)^k = \Big(\sum_{i=1}^{t} E(X_i) \Big)^k \geqslant k! \cdot \Big(\sum_{1 \leqslant i_1 < i_2 < \ldots < i_k \leqslant t} \prod_{j=1}^{k} E(X_{i_j}) \Big). \tag{3.24}$$

Hence, (3.21) follows from (3.23), (3.24) and (2.2). ∎

3.3. Proofs of Theorems 2.4 and 2.5. The proofs of Theorems 2.4 and 2.5 rely upon an interpolation argument, given by the next lemma.

LEMMA 3.2. *Let $p \geqslant 1$ be real and let $\lfloor p \rfloor \in \mathbf{N}$ and $\gamma = p - \lfloor p \rfloor \in [0, 1)$ denote its integer and fractional parts, respectively. Then for any random variable X we have*

$$\|X\|_p \leqslant \|X\|_{\lfloor p \rfloor}^{\theta} \cdot \|X\|_{\lfloor p \rfloor + 1}^{1-\theta}, \quad \textit{where } \theta = \frac{\lfloor p \rfloor (1-\gamma)}{p} \in (0, 1].$$

Proof. Using Hölder's inequality,

$$E(X \cdot Y) \leqslant \|X\|_s \cdot \|Y\|_{s/(s-1)}, \quad s \in (1, \infty),$$

we may bound the p-th moment of a random variable X in the following manner:

$$\|X\|_p^p = E(X^p) = E(X^{\theta p} \cdot X^{(1-\theta)p}) \leqslant \|X^{\theta p}\|_s \cdot \|X^{(1-\theta)p}\|_{s/(s-1)}, \quad s \in (1, \infty).$$

(For the sake of simplicity, and without restricting generality, we assume herein that X is non-negative.) Since

$$\|X^\alpha\|_\sigma = E(X^{\alpha\sigma})^{1/\sigma} = \|X\|_{\alpha\sigma}^{\alpha}, \quad \sigma \geqslant 1, \ \alpha \geqslant 1/\sigma,$$

we conclude that

$$\|X\|_p^p \leqslant \|X\|_{\theta p s}^{\theta p} \cdot \|X\|_{(1-\theta)ps/(s-1)}^{(1-\theta)p}, \quad s \in (1, \infty). \tag{3.25}$$

Now set $s = \lfloor p \rfloor / (\theta p) = 1/(1-\gamma)$. Since $\theta p s = \lfloor p \rfloor$, while

$$\frac{(1-\theta)ps}{s-1} = \frac{\big(1 - \big(\lfloor p \rfloor (1-\gamma)\big)/p\big) \cdot p/(1-\gamma)}{(1-\gamma)^{-1} - 1} = \lfloor p \rfloor + 1, \tag{3.26}$$

the desired inequality follows from (3.25) and (3.26). ■

Proof of Theorem 2.4. In view of Corollary 2.1 and Theorem 2.1, we conclude that

$$\|X\|_p \leqslant \frac{cp}{\ln(p+1)} \cdot \max\{E(X)^{1/p}, E(X)\}, \quad p \in \mathbf{N}, \tag{3.27}$$

where we put $c = 0.792$. Fixing $p \in \mathbf{N}$ and $\gamma \in [0, 1)$, we proceed to estimate $\|X\|_{p+\gamma}$. By Lemma 3.2,

$$\|X\|_{p+\gamma} \leqslant \|X\|_p^{\theta} \cdot \|X\|_{p+1}^{1-\theta}, \quad \theta = \frac{p \cdot (1-\gamma)}{p+\gamma}. \tag{3.28}$$

Employing (3.27), we obtain

$$\|X\|_{p+\gamma} \leqslant c \cdot \left(\frac{p}{\ln(p+1)}\right)^{\theta} \cdot \left(\frac{p+1}{\ln(p+2)}\right)^{1-\theta} \cdot E(X) \tag{3.29}$$

in case $E(X) \geqslant 1$, and

$$\|X\|_{p+\gamma} \leqslant c \cdot \left(\frac{p}{\ln(p+1)}\right)^{\theta} \cdot \left(\frac{p+1}{\ln(p+2)}\right)^{1-\theta} \cdot E(X)^{\theta/p+(1-\theta)/(p+1)} \tag{3.30}$$

if $E(X) < 1$. Since

$$\frac{\theta}{p} + \frac{1-\theta}{p+1} = \frac{1}{p+\gamma},$$

we may combine (3.29) and (3.30) as follows:

$$\|X\|_{p+\gamma} \leqslant c \cdot \left(\frac{p}{\ln(p+1)}\right)^{\theta} \cdot \left(\frac{p+1}{\ln(p+2)}\right)^{1-\theta} \cdot \max\{E(X)^{1/(p+\gamma)}, E(X)\}. \tag{3.31}$$

It is easy to check that

$$\theta = (1-\gamma) - \eta \quad \text{and} \quad 1-\theta = \gamma + \eta, \tag{3.32}$$

where

$$\eta = \frac{\gamma(1-\gamma)}{p+\gamma}. \tag{3.33}$$

Hence,

$$p^{\theta}(p+1)^{1-\theta} = p^{1-\gamma}(p+1)^{\gamma} \cdot \left(1 + \frac{1}{p}\right)^{\eta}. \tag{3.34}$$

As, by Jensen's inequality,

$$p^{1-\gamma}(p+1)^{\gamma} \leqslant (1-\gamma) \cdot p + \gamma \cdot (p+1) = p + \gamma, \tag{3.35}$$

we infer by (3.31), (3.34) and (3.35) that

$$\|X\|_{p+\gamma} \leqslant c \cdot \frac{(p+\gamma) \cdot (1+1/p)^{\eta}}{\big(\ln(p+1)\big)^{\theta}\big(\ln(p+2)\big)^{1-\theta}} \cdot \max\{E(X)^{1/(p+\gamma)}, E(X)\}. \tag{3.36}$$

We claim, and prove later, that

$$\frac{1}{\big(\ln(p+1)\big)^{\theta}\big(\ln(p+2)\big)^{1-\theta}} \leqslant \frac{1}{\ln(p+\gamma+1)}. \tag{3.37}$$

Therefore, by (3.36), (3.37) and (3.33),

$$\|X\|_{p+\gamma} \leqslant c\left(1+\frac{1}{p}\right)^{(\gamma(1-\gamma))/(p+\gamma)} \frac{p+\gamma}{\ln(p+\gamma+1)} \cdot \max\{E(X)^{1/(p+\gamma)}, E(X)\}$$

in accord with (2.13).

Hence, all that remains is to prove (3.37). To that end, we use the definition (3.28) of θ, in order to express γ in terms of θ, and rewrite (3.37) as follows:

(3.38)
$$\ln\left(p+1+\frac{p(1-\theta)}{p+\theta}\right) \leqslant \big(\ln(p+1)\big)^{\theta}\big(\ln(p+2)\big)^{1-\theta}, \quad p \in \mathbf{N},\ 0<\theta\leqslant 1.$$

By extracting the logarithm of both sides of (3.38), our new goal is proving that

$$f\left(p+1+\frac{p(1-\theta)}{p+\theta}\right) \leqslant \theta f(p+1)+(1-\theta)f(p+2), \quad p \in \mathbf{N},\ 0<\theta\leqslant 1,$$

where $f(x)=\ln\ln x$. Writing

$$g(\theta)=f\left(p+1+\frac{p(1-\theta)}{p+\theta}\right),$$

we aim at showing that

$$g(\theta) \leqslant (1-\theta)g(0)+\theta g(1), \quad 0\leqslant\theta\leqslant 1,\ p\in\mathbf{N}.$$

To this end, it suffices to show that g is convex in $[0,1]$, namely, that $g''(\theta)\geqslant 0$ for $\theta\in[0,1]$. As the latter claim may be proved by standard techniques, we omit the further details. The proof is thus complete. ■

Proof of Theorem 2.5. The proof of this theorem goes along the same lines of the proof of Theorem 2.4. We omit further details. ■

4. DISCUSSION

Here we compare our bounds to the previously known ones. According to Johnson et al. [10] (estimate (1.6)),

$$\|X\|_p \leqslant \frac{2p}{\ln p}\cdot\max\{\mathcal{E}_1,\mathcal{E}_p\}, \quad p\geqslant 1, \tag{4.1}$$

where $\mathcal{E}_k=\big(\sum_{i=1}^t E(X_i^k)\big)^{1/k}$. According to our Theorem 2.4, on the other hand, for all $p\geqslant 1$ we have

(4.2)
$$\|X\|_p \leqslant 0.792\cdot\left(1+\frac{1}{\lfloor p\rfloor}\right)^{(\{p\}\cdot(1-\{p\}))/p}\cdot\frac{p}{\ln(p+1)}\cdot\max\{E(X)^{1/p},E(X)\}.$$

First, we note that (4.2) relies only on $\mathcal{E}_1$ and does not involve $\mathcal{E}_p$, as does (4.1). However, estimate (4.2) applies only to the case of uniformly bounded variables, $P(0 \leqslant X_i \leqslant 1) = 1$, while (4.1) is not restricted to that case. It should be noted that, in the case where the variables X_i are not uniformly bounded by 1 (or some other arbitrary fixed constant), it is impossible to bound $E(X^p)$ in terms of $\mathcal{E}_1$ alone. As an example, let $X_i \sim a \cdot B\big(1, 1/(at)\big)$, $1 \leqslant i \leqslant t$, for some $a > 0$. Then $\mathcal{E}_1 = 1$ but $E(X^p)^{1/p} \geqslant \mathcal{E}_p = a^{(p-1)/p}$. As a may be arbitrarily large, $E(X^p)^{1/p}$ could not possibly be bounded in terms of $\mathcal{E}_1$ only in this case.

Estimate (4.2) is sharper than (4.1) in the sense that the bound in the latter tends to infinity when $p \to 1^+$, while that of (4.2) approaches

$$\frac{0.792}{\ln 2} E(X) \approx 1.14 E(X),$$

a small constant multiplicative factor above the limit of the left-hand side, $E(X) = \lim_{p\to 1^+} \|X\|_p$. When $\mathcal{E}_1 \geqslant 1$, estimate (4.2) is better than estimate (4.1) by a factor of at most

$$q(p) = \frac{0.792}{2} \cdot \left(1 + \frac{1}{\lfloor p \rfloor}\right)^{(\{p\}\cdot(1-\{p\}))/p} \cdot \frac{\ln p}{\ln(p+1)} \leqslant \frac{0.792}{2} \cdot 2^{1/4} \cdot 1 \approx 0.471.$$

For example,

$$q(2) = \frac{0.792}{2} \cdot \left(1 + \frac{1}{2}\right)^0 \cdot \frac{\ln 2}{\ln 3} \approx 0.25.$$

However, when $\mathcal{E}_1 \ll 1$, estimate (4.1) may become better than (4.2), as exemplified below.

EXAMPLE 4.1. Let $p = 2$, $\mathcal{E}_1 < 1$. Assume that all of the random variables are i.i.d. and that $(t^2/2) \cdot X_i \sim B\big(1, \frac{1}{2}\big)$. Hence, $\mathcal{E}_1 = 1/t$ and $\mathcal{E}_2 = \sqrt{2/t^3}$. The upper bound in (4.1) is

$$\frac{4}{\ln 2} \cdot \max(\mathcal{E}_1, \mathcal{E}_2) = \Theta(1/t).$$

However, the upper bound in (4.2) is of order of magnitude $\Theta(1/\sqrt{t})$. Hence, in this example (4.1) is sharper than (4.2).

Finally, we recall that Theorem 2.5 offers a further improvement of the estimate in Theorem 2.4 by a multiplicative factor of $e^{-0.6+\varepsilon}/0.792$ for all $\varepsilon > 0$ and $p > p_0(\varepsilon)$ for $p_0(\varepsilon)$ which is explicitly defined through (2.9) and (2.10), or by a multiplicative factor of $e^{-1+\varepsilon}/0.792$ for all $\varepsilon > 0$ and sufficiently large p.

Acknowledgment. The authors would like to express their gratitude to Michael Lin for his helpful comments on the paper.

5. APPENDIX

5.1. Proof of Theorem A$''$. Let us put

$$f(c) = \frac{(1+c)^p}{cp}\,\mathcal{E}_1, \qquad g(c) = \left(1+\frac{1}{c}\right)p^{-1/p}\,\mathcal{E}_p.$$

One checks easily that f is decreasing on the interval $\bigl(0, 1/(p-1)\bigr)$ and increasing on $\bigl(1/(p-1), \infty\bigr)$.

If $\mathcal{E}_1 > p^{1-1/p}\mathcal{E}_p$, then $f(c) > g(c)$ throughout the positive axis. Hence, the value of c that minimizes the right-hand side of (1.1) is that which minimizes $f(c)$, namely $c = 1/(p-1)$. This falls under the first case in (1.5) and gives the required result.

Next, if $\mathcal{E}_1 \leqslant p^{1-1/p}\mathcal{E}_p$, it may be easily verified that f and g intersect exactly once on the positive axis, at the point $c_0 = (p^{1-1/p}\mathcal{E}_p/\mathcal{E}_1)^{1/(p-1)} - 1$. There are two subcases to consider here: If f is non-increasing at c_0, which happens for

$$\mathcal{E}_1 \geqslant \frac{(p-1)^{p-1}}{p^{(p-1)^2/p}}\mathcal{E}_p,$$

then the value of c that minimizes the right-hand side of (1.1) is still the minimum point of f, which again falls under the first case in (1.5). If, on the other hand, f increases at c_0, namely

$$\mathcal{E}_1 < \frac{(p-1)^{p-1}}{p^{(p-1)^2/p}}\mathcal{E}_p,$$

then the point that minimizes the right-hand side of (1.1) is the intersection point of f and g. Plugging the common value of f and g in the right-hand side of (1.1), we obtain the second case in (1.5). ■

5.2. Proof of Proposition 3.1. We begin by stating and proving several lemmas. Hereinafter, $f(\cdot)$ is a convex function.

LEMMA 5.1. *Let X be a random variable with $P(a \leqslant X \leqslant b) = 1$ for some $b > a \geqslant 0$, and $\mu = E(X)$. Let $\tilde{X}$ be a random variable, assuming only the values a and b, having the same expected value, i.e., $E(\tilde{X}) = \mu$. Then*

$$E\bigl(f(X)\bigr) \leqslant E\bigl(f(\tilde{X})\bigr) = f(a)\cdot\frac{b-\mu}{b-a} + f(b)\cdot\frac{\mu-a}{b-a}. \tag{5.1}$$

P r o o f. Let $Y \sim U(0,1)$ be independent of X. Define $\tilde{X}$ by

$$\tilde{X} = \begin{cases} a, & Y \leqslant (b-X)/(b-a), \\ b, & \text{otherwise.} \end{cases}$$

Then

$$E(\tilde{X}|X) = a \cdot \frac{b-X}{b-a} + b \cdot \frac{X-a}{b-a} = X, \tag{5.2}$$

whence

$$E(\tilde{X}) = E\big(E(\tilde{X}|X)\big) = E(X) = \mu.$$

Thus, $\tilde{X}$ indeed has the required distribution. Consequently, by Jensen's inequality and (5.2),

$$E\big(f(\tilde{X})\big) = E\Big(E\big(f(\tilde{X})|X\big)\Big) \geqslant E\Big(f\big(E(\tilde{X}|X)\big)\Big) = E\big(f(X)\big). \tag{5.3}$$

Next, we find that

$$\begin{aligned} P(\tilde{X} = a) &= \int_a^b P(\tilde{X} = a|X = r) \cdot P(X = r)dr = \int_a^b \frac{b-r}{b-a} \cdot P(X = r)dr \\ &= \frac{b-\mu}{b-a}. \end{aligned}$$

Hence $E\big(f(\tilde{X})\big)$ equals the value on the right-hand side of (5.1). This, combined with (5.3), completes the proof. ■

LEMMA 5.2. *Let X_i, $1 \leqslant i \leqslant t < \infty$, be independent random variables satisfying $P(0 \leqslant X_i \leqslant a_i) = 1$. For $1 \leqslant i \leqslant t$, let $\tilde{X}_i$ be a random variable assuming only the values 0 and a_i and having the same expectation as X_i, namely $\tilde{X}_i/a_i \sim B(1, \mu_i/a_i)$, where $\mu_i = E(X_i)$. Then*

$$E\Big(f\big(\sum_{i=1}^{t} X_i\big)\Big) \leqslant E\Big(f\big(\sum_{i=1}^{t} \tilde{X}_i\big)\Big). \tag{5.4}$$

P r o o f. We compute the expected value on the left-hand side of (5.4) by first fixing the values of $X_2, \ldots, X_t$, taking the average with respect to X_1, and then averaging with respect to $X_2, \ldots, X_t$:

$$E\Big(f\big(\sum_{i=1}^{t} X_i\big)\Big) = E\Big(E\Big(f\big(\sum_{i=1}^{t} X_i\big)|X_2, \ldots, X_t\Big)\Big). \tag{5.5}$$

For every fixed value of $X_2, \ldots, X_t$, the internal expected value on the right-hand side of (5.5) may be bounded using Lemma 5.1,

$$E\Big(f\big(\sum_{i=1}^{t} X_i\big)|X_2, \ldots, X_t\Big) \leqslant E\Big(f\big(\tilde{X}_1 + \sum_{i=2}^{t} X_i\big)|X_2, \ldots, X_t\Big). \tag{5.6}$$

Using (5.6) in (5.5) we conclude that

$$E\Big(f\big(\sum_{i=1}^{t} X_i\big)\Big) \leqslant E\Big(f\big(\tilde{X}_1 + \sum_{i=2}^{t} X_i\big)\Big).$$

By the same token, we may replace each of the other random variables $X_i, 2 \leqslant i \leqslant t$, with the corresponding $\tilde{X}_i$ without decreasing the expected value, thus proving the assertion (5.4). ∎

LEMMA 5.3. *Let $X_i, i = 1, 2, 3$, be independent random variables, with $X_i \sim B(1, p_i), i = 1, 2,$ $p_1 > p_2$, and X_3 be an arbitrary non-negative variable. Let $X_1' \sim B(1, p_1 - \varepsilon)$ and $X_2' \sim B(1, p_2 + \varepsilon)$, where $0 \leqslant \varepsilon \leqslant p_1 - p_2$, and assume that X_1', X_2', X_3 are also independent. Then*

$$E\big(f(X_1 + X_2 + X_3)\big) \leqslant E\big(f(X_1' + X_2' + X_3)\big).$$

Proof. First, we prove the inequality for the case where X_3 is constant, say $X_3 = a$. On the one hand, we have

$$\begin{aligned} &E\big(f(X_1 + X_2 + a)\big) \\ &\quad = p_1 p_2 f(2 + a) + (p_1 + p_2 - 2p_1 p_2) f(1 + a) + (1 - p_1)(1 - p_2) f(a). \end{aligned}$$

Setting $b = (p_1 - \varepsilon)(p_2 + \varepsilon) - p_1 p_2 \geqslant 0$, we find that

$$\begin{aligned} E\big(f(X_1' + X_2' + a)\big) &= (p_1 p_2 + b) f(2 + a) + (p_1 + p_2 - 2p_1 p_2 - 2b) f(1 + a) \\ &\quad + \big((1 - p_1)(1 - p_2) + b\big) f(a) \\ &= E\big(f(X_1 + X_2 + a)\big) + b\big(f(2 + a) - 2f(1 + a) + f(a)\big). \end{aligned}$$

Since the last term on the right-hand side is non-negative due to the convexity of f, this proves the inequality for constant X_3. The proof for general X_3 now follows:

$$\begin{aligned} E\big(f(X_1 + X_2 + X_3)\big) &= E\Big(E\big(f(X_1 + X_2 + X_3)|X_3\big)\Big) \\ &\leqslant E\Big(E\big(f(X_1' + X_2' + X_3)|X_3\big)\Big) \\ &= E\big(f(X_1' + X_2' + X_3)\big). \quad ∎ \end{aligned}$$

Proof of Proposition 3.1. Let us assume first that the sequence is finite. By Lemma 5.2, we may assume that all X_i's are Bernoulli distributed, say $X_i \sim B(1, p_i)$, $1 \leqslant i \leqslant t$. If not all p_i's are equal, we can find two of them, say p_1 and p_2, such that

$$p_1 > \frac{\mu}{t} = \frac{\sum_{i=1}^{t} p_i}{t} > p_2.$$

Employing Lemma 5.3, we can change p_1 and p_2, while keeping their sum constant and making one of them equal to μ/t, without reducing the p-th moment of the sum. Repeating this procedure over and over again until all p_i's coincide, we see that $E\big(f(X)\big) \leqslant E\big(f(Y)\big)$, where $Y \sim B(t, \mu/t)$.

For infinite sequences, we use the preceding part of the proof to conclude that for all finite values of t'

$$E\Big(f\big(\sum_{i=1}^{t'} X_i\big)\Big) \leqslant E\big(f(Y_{t'})\big),$$

where $Y_{t'} \sim B(t', \mu/t')$. Passing to the limit as $t' \to t = \infty$, we obtain the required result. ■

5.3. An alternative proof of Corollary 2.1. The second part of Corollary 3.1 in [14] states that, if $X_1, \ldots, X_n$ are non-negative random variables with

$$E(X_k^s | X_1, \ldots, X_{k-1}) \leqslant a_k^s$$

and

$$E(X_k^t | X_1, \ldots, X_{k-1}) \leqslant b_k$$

for all $1 \leqslant k \leqslant n$, then

$$E\big(\sum_{k=1}^n X_k\big)^t \leqslant E\big(\theta^t(1)\big) \cdot \max\Big(\sum_{k=1}^n b_k, \big(\sum_{k=1}^n a_k^s\big)^{t/s}\Big) \tag{5.7}$$

for all $t \geqslant 2$ and $0 < s \leqslant 1$, where $\theta(1)$ is a Poisson random variable with parameter one. When $X_1, \ldots, X_n$ are independent and bounded between zero and one, we may set $a_k^s = E(X_k^s)$ and $b_k = E(X_k)$. Taking $s = 1$, we infer from (5.7) that

$$E\big(\sum_{k=1}^n X_k\big)^t \leqslant E\big(\theta^t(1)\big) \cdot \max\Big(\sum_{k=1}^n E(X_k), \big(\sum_{k=1}^n E(X_k)\big)^t\Big). \tag{5.8}$$

Finally, since, by Dobinski's formula (2.6), $E\big(\theta^t(1)\big)$ is the Bell number B_t, inequality (5.8) coincides with inequality (2.12) in Corollary 2.1. ■

5.4. An application to stochastic scheduling. In the classical multiprocessor scheduling problem, one is given a set $\mathbf{J} = \{J_1, \ldots, J_n\}$ of n jobs, with known processing times $\tau(J_j) = \tau_j \in \mathbf{R}^+$, and a set $\mathbf{M} = \{M_1, \ldots, M_m\}$ of m machines. The goal is to find an assignment $A : \mathbf{J} \to \mathbf{M}$ of the jobs to the machines, such that the resulting loads on each machine minimize a given cost function,

$$T^f(A) = f(L_1, \ldots, L_m),$$

where $L_i = \sum_{\{j:\, A(J_j)=M_i\}} \tau_j$, $1 \leqslant i \leqslant m$. Typical choices for the cost function f are the maximum norm (in which case the problem is known as the *makespan* *problem*) or, more generally, the ℓ_p-norms, $1 \leqslant p \leqslant \infty$. The case $p = 2$ was studied in [3] and [4] and was motivated by storage allocation problems; the general case, $1 < p < \infty$, was studied in [1].

In stochastic scheduling, the processing times τ_j are random variables with known probability distribution functions. The random variables $\tau_j,\ 1 \leqslant j \leqslant n$, are non-negative. They are also typically independent and uniformly bounded. The goal is to find an assignment that minimizes the expected cost,

$$T^f(A) = E[f(L_1, \dots, L_m)].$$

As such problems are strongly NP-hard, even in the deterministic setting, the goal is to obtain a reasonable approximation algorithm, or to compare the performance of a given scheduling algorithm to that of an optimal scheduling.

We proceed to briefly sketch an application of our results for estimating the performance of a simple scheduling algorithm, called *List Scheduling*, as carried out in [18]. Assume that the target function is the ℓ_p-norm of the vector of loads, that is

$$T^f(A) = E\Big(\big(\sum_{i=1}^m L_i^p\big)^{1/p}\Big) \qquad (p > 1).$$

Then, by Jensen's inequality,

$$T^f(A) \leqslant \big(\sum_{i=1}^m E(L_i^p)\big)^{1/p}.$$

Since L_i is a sum of independent non-negative uniformly bounded random variables (where the uniform bound may be scaled to equal one), estimate (2.15), which is the simpler version of Theorem 2.4, implies that

$$E(L_i^p) \leqslant \left(\frac{0.942p}{\ln(p+1)}\right)^p \cdot \max\{E(L_i), E(L_i)^p\}.$$

Hence,

$$T^f(A) \leqslant \frac{0.942p}{\ln(p+1)} \cdot \big(\sum_{i=1}^m \max\{E(L_i), E(L_i)^p\}\big)^{1/p}. \tag{5.9}$$

The List Scheduling algorithm is an online algorithm that always schedules the next job to the machine that currently has the smallest (expected) load. As implied by the analysis of Graham in [8],

$$E(L_i) \leqslant 2\mu, \qquad 1 \leqslant i \leqslant m, \tag{5.10}$$

where

$$\mu := \max\left\{\frac{\sum_{j=1}^n E(\tau_j)}{m}, \max_{1\leqslant j\leqslant n} E(\tau_j)\right\}$$

is a quantity that depends only on the known expected processing times of the jobs and is independent of the assignment of jobs to the machines. Combining (5.10) with (5.9) we arrive at the following bound on the performance of the stochastic List Scheduling algorithm:

$$T^f(A) \leqslant \frac{0.942p}{\ln(p+1)} \cdot \big(m \cdot \max\{2\mu, (2\mu)^p\}\big)^{1/p}.$$

REFERENCES

[1] N. Alon, Y. Azar, G. Woeginger and T. Yadid, *Approximation schemes for scheduling*, in: *Proceedings of the 8th ACM-SIAM Symposium on Discrete Algorithms*, 1997, pp. 493–500.

[2] N. G. de Bruijn, *Asymptotic Methods in Analysis*, Dover, New York, NY, 1958.

[3] A. K. Chandra and C. K. Wong, *Worst-case analysis of a placement algorithm related to storage allocation*, SIAM J. Computing 4 (1975), pp. 249–263.

[4] R. A. Cody and E. G. Coffman, Jr., *Record allocation for minimizing expected retrieval costs on drum-like storage devices*, J. Assoc. Comput. Mach. 23 (1976), pp. 103–115.

[5] G. Dobinski, *Summierung der reihe* $\sum n^m/m!$ *für* $m = 1, 2, 3, 4, 5, ...$, Grunert Archiv (Arch. Math. Phys.) 61 (1877), pp. 333–336.

[6] T. Figiel, P. Hitczenko, W. B. Johnson, G. Schechtman and J. Zinn, *Extremal properties of Rademacher functions with applications to the Khintchine and Rosenthal inequalities*, Trans. Amer. Math. Soc. 349 (1997), pp. 997–1027.

[7] B. Fristedt and L. Gray, *A Modern Approach to Probability Theory*, Birkhäuser, Boston, MA, 1997.

[8] R. L. Graham, *Bounds on multiprocessing timing anomalies*, SIAM J. Appl. Math. 17 (1969), pp. 416–429.

[9] W. Hoeffding, *On the distribution of the number of successes in independent trials*, Ann. Math. Statist. 27 (1956), pp. 713–721

[10] W. B. Johnson, G. Schechtman and J. Zinn, *Best constants in moment inequalities for linear combinations of independent and exchangeable random variables*, Ann. Probab. 13 (1985), pp. 234–253.

[11] A. Khintchine, *Über dyadische Brüche*, Math. Z. 18 (1923), pp. 109–116.

[12] R. Latała, *Estimation of moments of sums of independent real random variables*, Ann. Probab. 25 (1997), pp. 1502–1513.

[13] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, New York, NY, 1984.

[14] V. H. de la Peña, R. Ibragimov and S. Sharakhmetov, *On extremal distribution and sharp* L_p*-bounds for sums of multilinear forms*, Ann. Probab. 31 (2003), pp. 630–675.

[15] J. Pitman, *Some probabilistic aspects of set partitions*, Amer. Math. Monthly 104 (1997), pp. 201–209.

[16] J. Riordan, *An Introduction to Combinatorial Analysis*, Wiley, New York, NY, 1980.

[17] H. Robbins, *A remark on Stirling's formula*, Amer. Math. Monthly 62 (1955), pp. 26–29.

[18] H. Shachnai and T. Tassa, *Approximation ratios for stochastic list scheduling*, preprint.

Departments of Mathematics
and of Computer Science
Ben-Gurion University of the Negev
Beer Sheva, Israel
E-mail: berend@math.bgu.ac.il

Department of Mathematics
and Computer Science
The Open University of Israel
Ra'anana, Israel
E-mail: tamirta@openu.ac.il

Received on 5.9.2008;
revised version on 8.11.2009