ANNALES DE L'I. H. P., SECTION B

MICHEL TALAGRAND

The missing factor in Hoeffding's inequalities

Annales de l'I. H. P., section B, tome 31, nº 4 (1995), p. 689-702

<http://www.numdam.org/item?id=AIHPB_1995__31_4_689_0>

Ann. Inst. Henri Poincaré,

Vol. 31, n° 4, 1995, p. 689-702 *Probabilités et Statistiques*

The Missing factor in Hoeffding's inequalities

by

Michel TALAGRAND

University of Paris VI and The Ohio State University.

ABSTRACT. – A celebrated paper of Hoeffding establishes sharp bounds for the tails of sums of bounded independent random variables. An elementary observation allows to improve these bounds to optimal order under mild conditions. (The method we present also allows to improve many other exponential bounds.)

RÉSUMÉ. – Un article célèbre d'Hoeffding établit des bornes pour la déviation d'une somme de variables aléatoires indépendantes par rapport à sa moyenne. Combinant la méthode d'Hoeffding avec la transformation d'Esscher, on montre comment sous des hypothèses supplémentaires minimes ces bornes peuvent être améliorées d'une façon essentiellement optimale.

1. INTRODUCTION

Consider independent centered random variables (r.v.) $X_1, \ldots, X_n$. (Throughout the paper, X will always denote a centered r.v.) The study

Équipe d'Analyse, Tour 48, U.A. au C.N.R.S. n° 754, Université Paris VI, 4, place Jussieu, 75230 Paris Cedex 05 and Department of Mathematics, 231 West 18th Avenue, Columbus, Ohio 43210, U.S.A.

of the tail

$$P(\sum_{i\leq n} X_i \geq t)$$

certainly has a long history. In the case where the r.v. $X_1, \dots, X_n$ are identically distributed (and $E \exp \alpha |X_1| < \infty$ for some $\alpha > 0$), H. Cramér proved in 1938 that, if $\rho^2 = EX_1^2$, then, if $x = o(n^{1/6})$, we have

$$P(\sum_{i\leq n} X_i \geq x\rho\sqrt{n}) \sim \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-u^2/2} du. \tag{1.1}$$

Asymptotic expansions, however precise, do not diminish the need for inequalities valid for all n, t. Such inequalities have been obtained in particular by Yu, V. Prokhorov, G. Bennett, W. Hoeffding, S. V. Nagaev. For the purpose of simplicity, let us discuss Hoeffding's bounds. The thrust of his paper is (in the i.i.d. case) to obtain bounds that depend only on $\rho^2 = EX_1^2$ and on $b = \sup X_1$. He obtains a bound

$$P(\sum_{i\leq n} X_i \geq t) \leq \exp -nH\left(\rho^2, b, \frac{t}{n}\right) \tag{1.2}$$

for a certain function H (that will be described in (1.5) below). The sharpness of this bound can be tested in the case of the binomial law, *i.e.* when $P(X_i = 1-p) = p$, $P(X_i = -p) = 1-p$, for some $0 \leq p \leq 1$. In that case $\rho^2 = p(1-p)$, $b = 1-p$. For the binomial law, straight computation using Sterling's formulae allows to evaluate $P(\sum_{i\leq n} X_i \geq t)$. In particular, one then finds that, for $t \leq \frac{n}{2}\min(p, 1-p)$, and some constant K,

$$P(\sum_{i\leq n} X_i \geq t) \exp nH\left(\rho^2, b, \frac{t}{n}\right) \leq \frac{K\rho\sqrt{n}}{t} << 1. \tag{1.3}$$

for $t >> \rho\sqrt{n}$. (On the other hand, (1.2) is optimal for $t = b$).

The bounds of [H], [N], are all obtained from the inequality

$$P(\sum_{i\leq n} X_i \geq t) \leq \inf_{s\geq 0} e^{-st} E \exp s \sum_{i\leq n} X_i. \tag{1.4}$$

As pointed out by Hoeffding, (1.2) is the best that can be obtained from (1.4). The purpose of the present paper is to show how the technique used to obtain (1.1) can be combined with the calculations involving the use of (1.4) to obtain under very mild extra conditions the missing factor exemplified by (1.3) in inequalities like (1.2), yielding bounds of truly optimal order. The method we present is very general. We feel however that we serve better the reader by concentrating on two simple results rather than by trying to be encyclopedic. (Adaptation of the method to other cases is routine.)

We consider the function

$$\theta(x) = \frac{1}{\sqrt{2\pi}} e^{x^2/2} \int_x^\infty e^{-u^2/2} du.$$

Thus $\theta(0) = 1/2$, and it is easy to see that θ decreases for $x \geq 0$. It is shown in [I-K] p. 17 and [K-J], p. 505 that

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

We consider the function

$$H(\rho^2, b, x) = \left(1 + \frac{bx}{\rho^2}\right) \frac{\rho^2}{b^2+\rho^2} \log\left(1 + \frac{bx}{\rho^2}\right) + \left(1 - \frac{x}{b}\right) \frac{b^2}{b^2+\rho^2} \log\left(1 - \frac{x}{b}\right). \tag{1.5}$$

We denote by K a universal constant, that may vary at each occurrence.

THEOREM 1.1. – *Consider independent centered r.v.* $(X_i)_{i\leq n}$ *and let* $\sigma^2 = E(\sum_{i\leq n} X_i)^2$, $\rho^2 = \sigma^2/n$. *Assume that* $X_i \leq b$ *for all* $i \leq n$, *and that* $|X_i| \leq B$ *for all* $i \leq n$. *Then, for* $0 \leq t \leq \frac{\sigma^2}{KB}$, *we have*

$$P(\sum_{i\leq n} X_i \geq t) \leq \left(\theta\left(\frac{t}{\sigma}\right) + K\frac{B}{\sigma}\right) e^{-nH(\rho^2, b, \frac{t}{n})}. \tag{1.6}$$

Comments. – 1) Since $\theta(t/\sigma)$ is of order σ/t, it dominates B/σ for the range of t considered. In particular (1.6) yields

$$P(\sum_{i\leq n} X_i \geq t) \leq \frac{K\sigma}{t} e^{-nH(\rho^2, b, \frac{t}{n})}.$$

2) For $t = o(\sigma^{4/3} b^{-1/4})$, we have $nH(\rho^2, b, t/n) = t^2/2n\rho^2 + o(1)$. This is the range where (1.1) holds. By contrast, Theorem 1.1 holds for

much larger values of t. It implies in particular, that given $\varepsilon > 0$, there is a constant $K(\varepsilon)$, such that if $0 \leq t \leq \sigma^2/K(\varepsilon)B$, then

$$P(\sum_{i\leq n} X_i \geq t) \leq (1+\varepsilon)\theta\left(\frac{t}{\sigma}\right)e^{-nH(\rho^2,b,\frac{t}{n})}.$$

3) For simplicity we have made the blanket assumption $|X_i| \leq B$, although quite less is required. However, we do not know how to improve Theorem 3 of [H] without further assumption.

4) Certainly one can obtain a rather small numerical value for K, but numerical computations are better left to others with the talent for it.

THEOREM 1.2. – *Consider independent r.v. $(Y_i)_{i\leq n}$ such that $0 \leq Y_i \leq 1$. Set $\mu = n^{-1}\sum_{i\leq n} EY_i$. Then for $K \leq t \leq n\mu(1-\mu)/K$, we have*

$$(1.7)\quad P(\sum_{i\leq n} Y_i \geq n\mu + t) \leq \left(\theta\left(\frac{t}{\sqrt{n\mu(1-\mu)}}\right) + \frac{K}{\sqrt{n\mu(1-\mu)}}\right) \times e^{-nH(\mu(1-\mu),1-\mu,\frac{t}{n})}.$$

Comments. – 1) The quantity $\exp\left(-nH\left(\mu(1-\mu),1-\mu,\frac{t}{n}\right)\right)$ is exactly the bound (2.1) of [H]. When the r.v. Y_i have the common mean μ, Theorem 1.2 is closely related to Theorem 1.1 as is seen by considering $X_i = Y_i - \mu$, and using the pessimistic bound $\sigma \leq \sqrt{n\mu(1-\mu)}$. The point of Theorem 1.2 is that the variables Y_i need not have a common mean.

2) When σ^2 is of order $n\mu(1-\mu)$, the Berry-Essen theorem allows to get rid of the restriction $t \geq K$.

The point of Theorems 1.1, 1.2 is that for values of t up to order σ^2, we get both the optimal exponential term, and the essentially correct factor in front of it. In the situation of Theorem 1.2, the worst case is $\mu = 1/2$. In that case, the exponent of the exponential term is very close to $-2t^2/n$ for values of t of order up to $n^{3/4}$. This is essentially the range where asymptotic expressions like (1.1) hold, so much of the strength of Theorem 1.2 is lost when one replaces the last term by $\exp(-2t^2/n)$. The corresponding corollary is nonetheless worthy to state.

COROLLARY 1.3. – *In the situation of Theorem* 1.2 *we have, for $t \geq K$, that*

$$(1.8)\quad P(\sum_{i\leq n} Y_i \geq n\mu + t) \leq \left(\theta\left(\frac{2t}{\sqrt{n}}\right) + \frac{K}{\sqrt{n}}\right)e^{-2t^2/n}$$

$$= \frac{K}{\sqrt{n}}e^{-2t^2/n} + \frac{1}{\sqrt{2\pi}}\int_{2t/\sqrt{n}}^{\infty} e^{-u^2/2}du.$$

The paper is organized as follows. The basic observation is presented in Section 2. The elementary estimates needed to take advantage of it are presented in Section 3; and the Theorems are proved in Section 4.

2. THE APPROACH

For $i \leq n$, consider the function ψ_i given by

$$\psi_i(s) = \log Ee^{sX_i}.$$

Since, in this paper, we assume that X_i is bounded, ψ_i is infinitely differentiable. We set $\psi(s) = \sum_{i \leq n} \psi_i(s)$.

The basic inequality is as follows.

PROPOSITION 2.1. – *Consider $s \geq 0$, and $t = \psi'(s)$. Then*

$$P\Big(\sum_{i \leq n} X_i \geq t\Big) \leq e^{\psi(s)-st}[\theta(s\rho) + 2h(s)] \tag{2.1}$$

where $\rho^2 = \psi''(s)$, and where $h(s)$ is given by (2.4) *below.*

The term $h(s)$ will be shown to be of lower order. The approach is now as follows. Given $t \geq 0$, to obtain a bound for $P(\sum_{i \leq n} X_i \geq t)$, we apply (2.1) where s is chosen so that $t = \psi'(s)$. We then find a lower bound for s and $\psi''(s)$ (and thus an upper bound for $\theta(s\rho)$). We then observe that

$$\psi(s) - st = \inf_{u \geq 0} \psi(u) - ut$$

and we note that the quantity $\inf_{u \geq 0} e^{\psi(u)-ut}$ is bounded by the Hoeffding bounds, since it is how these bounds are derived.

To prove Proposition 2.1, we follow Feller [F], p. 554. We set $S = \sum_{i \leq n} X_i$. We consider the r.v.

$$R = e^{sS-\psi(s)}$$

so that $ER = 1$. (The Esscher transformation). Thus, we can consider the probability Q such that $dQ = RdP$. Thus $dP = R^{-1}dQ$.

We have

$$(2.2)\quad P(S \geq t) = \int 1_{\{S\geq t\}} dP = \int 1_{\{S\geq t\}} e^{-sS+\psi(s)} dQ$$
$$= e^{\psi(s)}(e^{-st}Q(S \geq t) - s\int_t^\infty e^{-su}Q(S \geq u)du)$$

where the last equality follows by integration by parts. We now observe that the r.v. X_i are still independent when Q is the basic probability. We can thus appeal to the Berry-Essen theorem as in Feller, p. 542, to get

$$(2.3)\qquad |Q(S \geq u) - \frac{1}{\sqrt{2\pi}}\int_{(u-a)/\rho} e^{-v^2/2} dv| \leq h(s)$$

where $a = \int SdQ$, $h(s) = 6r(s)\rho^{-3}(s)$, and where

$$(2.4)\, r(s) = \sum_{i\leq n}\int |X_i|^3 dQ,\quad \rho^2 = \rho^2(s) = \int S^2 dQ - \left(\int SdQ\right)^2$$

Since

$$e^{\psi(s)} = Ee^{sS}$$

by differentiation we have

$$(2.5)\qquad \psi'(s)e^{\psi(s)} = \int Se^{sS}dP,$$

so that $a = t$. Plugging (2.3) into (2.2) and integrating by part again yields

$$P(S \geq t) \leq 2e^{\psi(s)-st}h(s) + e^{\psi(s)}\int_t^\infty \frac{1}{\sqrt{2\pi}} e^{-su} e^{-\frac{(u-t)^2}{2\rho^2}} du$$

Making the change of variable $u = t + \rho v$, the last term is seen to be $e^{\psi(s)-st}\theta(s\rho)$ by a standard computation. To finish the proof, it suffices to see that $\rho^2 = \psi''(s)$. This is seen by differentiating again the relation (2.5). □

3. ESTIMATES

The basic lemma is completely elementary.

LEMMA 3.1. – *Consider a (bounded) r.v. X with $EX = 0$, $\tau^2 = EX^2$, $\gamma = EX^3$. Set $\varphi(s) = \log Ee^{sX}$. Then*

a) $$\varphi''(s) \geq \frac{\tau^2 + s\gamma}{(Ee^{sX})^2}$$

b) $$E(Xe^{sX}) \geq s\tau^2 + \frac{s^2}{2}\gamma$$

c) *If $X \leq b$, then*

$$E(Xe^{sX}) \leq \tau^2\left(\frac{e^{sb} - 1}{b}\right).$$

$$E(e^{sX}) \leq 1 + \tau^2\left(\frac{e^{sb} - sb - 1}{b^2}\right).$$

Proof. – a) By computation, we have $\varphi''(s) = g(s)(Ee^{sX})^{-2}$, where

$$g(s) = EX^2e^{sX}Ee^{sX} - (EXe^{sX})^2.$$

(This might be the point to observe that $g(s) \geq 0$ by Cauchy-Schwartz, so that $\varphi''(s) \geq 0$). Thus

$$g'(s) = EX^3e^{sX}Ee^{sX} - EX^2e^{sX}EXe^{sX}$$
$$g''(s) = EX^4e^{sX}Ee^{sX} - (EX^2e^{sX})^2.$$

Thus, by Cauchy-Schwartz, we have $g''(s) \geq 0$, so that $g'(s) \geq g'(0) = \gamma$, and $g(s) \geq g(0) + s\gamma = \tau^2 + s\gamma$.

b) Consider the function $g(s) = E(Xe^{sX})$. Then $g^{(3)}(s) = E(X^4e^{sX}) \geq 0$, so that $g''(s) \geq g''(0) = \gamma$, $g'(s) \geq g'(0) + \gamma s = \tau^2 + \gamma s$, and $g(s) \geq \tau^2 s + \gamma^2 s/2$.

c) Consider the function $f(s) = E(e^{sX})$. Then $f''(s) = E(X^2e^{sX}) \leq \tau^2e^{sb}$, so that $f'(s) \leq f'(0) + \tau^2(e^{sb} - 1)/b = \tau^2(e^{sb} - 1)/b$, and the result by integration. □

LEMMA 3.2. – *Assume that for some number $B > 0$, we have $X_i \leq B$ for all $i \leq n$ and $EX_i^2 \leq B^2$, $EX_i^3 \geq -BEX_i^2$. Then, if s satisfies $t = \psi'(s)$, we have*

$$s \geq \frac{t}{\sigma^2}\left(1 - \frac{KBt}{\sigma^2}\right) \tag{3.1}$$

$$t \le \sigma^2/6B \Rightarrow s \le 2t/\sigma^2 \tag{3.2}$$

$$s\sqrt{\psi''(s)} \ge \frac{t}{\sigma}\left(1 - \frac{KBt}{\sigma^2}\right) \tag{3.3}$$

Proof of (3.1). – We have

$$t = \psi'(s) = \sum_{i \le n} \frac{E(X_i e^{sX_i})}{E(e^{sX_i})}. \tag{3.4}$$

By Lemma 3.1, c, we get, setting $\sigma_i^2 = EX_i^2$

$$t \le \sum_{i \le n} \frac{\sigma_i^2(e^{sB} - 1)}{BEe^{sX_i}} \le \sigma^2\left(\frac{e^{sB} - 1}{B}\right)$$

so that

$$s \ge \frac{1}{B} \log\left(1 + \frac{tB}{\sigma^2}\right).$$

In particular, since $\log(1 + x) \ge x - x^2/2$ for $x \ge 0$, we have

$$s \ge \frac{t}{\sigma^2}\left[1 - \frac{tB}{2\sigma^2}\right]. \tag{3.5}$$

Proof of (3.2). – By Lemma 3.1, b, we have, setting $\gamma_i = EX_i^3$

$$E(X_i e^{sX_i}) \ge s\sigma_i^2 + \frac{s^2}{2}\gamma_i \ge s\sigma_i^2 - s^2\sigma_i^2\frac{B}{2}. \tag{3.6}$$

By Lemma 3.1, c, setting $\zeta(s) = (e^{sB} - sB - 1)/B^2$, we have

$$\frac{1}{Ee^{sX_i}} \ge \frac{1}{1 + \sigma_i^2\zeta(s)} \ge 1 - \sigma_i^2\zeta(s). \tag{3.7}$$

Thus, combining (3.4), (3.6), (3.7), we have

$$\begin{aligned} t &\ge \sum_{i \le n}\left(s\sigma_i^2(1 - \sigma_i^2\zeta(s)) - s^2\sigma_i^2\frac{B}{2}\right) \\ &\ge s\sigma^2 - \frac{s^2\sigma^2 B}{2} - s\sum_{i \le n}\sigma_i^4\zeta(s) \\ t &\ge s\sigma^2\left[1 - \frac{sB}{2} - B^2\zeta(s)\right]. \end{aligned}$$

Denote by $f(s)$ the right hand side. Since $\zeta(s) \leq (e-2)s^2 \leq s^2$ for $s \leq 1/B$, $\zeta'(s) \leq (e-1)s \leq 2s$, one sees that if $0 \leq s \leq s_0 = 1/3B$, $f'(s) \geq 0$, so that $f(s)$ increases. Since ψ' increases (as noted in the proof of Lemma 3.1), it follows that if $t = \psi'(s) \leq f(s_0)$, we have $s \leq s_0$, so that $\zeta(s) \leq s^2$ and $s \leq 2t/\sigma^2$. Since $f(s_0) \geq \sigma^2/6B$, we have shown that for $t \leq \sigma^2/6B$, we have $s \leq 2t/\sigma^2$ (so that in particular $sB \leq 1/3$).

Proof of (3.3). – By Lemma 3.1, a, we have

$$\psi''(s) \geq \sum_{i \leq n} \frac{\sigma_i^2 + s\gamma_i}{(Ee^{sX_i})^2} \geq \sum_{i \leq n} \frac{\sigma_i^2(1 - Bs)}{(Ee^{sX_i})^2}.$$

By Lemma 3.1, c, and since $sB \leq 1$, we have

$$\begin{aligned} Ee^{sX_i} &\leq 1 + \sigma_i^2\left(\frac{e^{sB} - sB - 1}{B^2}\right) \\ &\leq 1 + s^2\sigma_i^2 \leq 1 + s^2B^2 \end{aligned}$$

so that

$$(Ee^{sX_i})^{-2} \geq 1 - 2s^2B^2$$

and

$$\psi''(s) \geq \sigma^2(1 - Bs)(1 - 2s^2B^2) \geq \sigma^2(1 - 3sB). \tag{3.8}$$

Combining with (3.5) and the fact that $s \leq 2t/\sigma^2$ yield the result. □

Remark. – Rather than assuming $X_i \leq B$, one could obtain a weaker, but sufficient result assuming only $Ee^{X_i/B} \leq 2$.

We are now ready for the main result.

THEOREM 3.3. – *There exists a universal constant K with the following property. Assume that for some number $B > 0$, the r.v. X_i satisfies*

$$X_i \leq B; \quad EX_i^2 \leq B^2; \quad E|X_i|^3 \leq BEX_i^2.$$

Then, if $t \leq \sigma^2/KB$, we have

$$P\Big(\sum_{i \leq n} X_i \geq t\Big) \leq \inf_{u \geq 0} e^{\psi(u) - ut}\left[\theta\left(\frac{t}{\sigma}\right) + \frac{KB}{\sigma}\right].$$

Proof. – Let $t \leq s\sigma^2/6B$, and let s satisfy $t = \psi'(s)$. Thus, by (3.2) we have $sB \leq 1/3$. We observe that

$$\begin{aligned}\int |X_i|^3 dQ &\leq \int |X_i|^3 e^{sX_i} dP \\ &\leq e^{sB} \int |X_i|^3 dP \leq 2B\sigma_i^2,\end{aligned}$$

Thus, by (3.8), we have $h(s) \leq KB/\sigma$.

We now show that, if $t \leq \sigma^2/2K'B$, we have

$$\theta\left(\frac{t}{\sigma}\left(1 - \frac{K'tB}{\sigma^2}\right)\right) \leq \theta\left(\frac{t}{\sigma}\right) + \frac{KB}{\sigma}.$$

Indeed,

$$\left|\theta\left(\frac{t}{\sigma}\right) - \theta\left(\frac{t}{\sigma}\left(1 - \frac{K'tB}{\sigma^2}\right)\right)\right| \leq \frac{K't^2B}{\sigma^3} \max\left\{|\theta'(u)| : \frac{t}{2\sigma} \leq u \leq \frac{t}{\sigma}\right\}.$$

Thus, it suffices to prove the following lemma.

LEMMA 3.4. – $\dfrac{-K}{\max(1,\lambda^2)} \leq \theta'(\lambda) \leq -\dfrac{1}{K}\dfrac{1}{\max(1,\lambda^2)}$.

Proof. – We observe that $-\theta'(\lambda) = \dfrac{1}{\sqrt{2\pi}} - \lambda\theta(\lambda)$, so that

$$\begin{aligned}-\theta'(\lambda) &= \frac{1}{\sqrt{2\pi}} \int_\lambda^\infty (t-\lambda) e^{-t^2/2+\lambda^2/2} dt \\ &= \frac{1}{\sqrt{2\pi}} \int_0^\infty u e^{-u^2/2 - u\lambda} du\end{aligned}$$

and the result follows by elementary estimates. □

Remark. – Actually $\lim_{\lambda\to 0} -\theta'(\lambda) = 1/\sqrt{2\pi}$ and $\lim_{\lambda\to\infty} -\lambda^2\theta'(\lambda) = 1/\sqrt{2\pi}$.

It remains to observe that

$$\psi(s) - st = \inf_{u\geq 0} \psi(u) - ut.$$

But we have noted that $\psi'' \geq 0$, so that ψ is convex. And the minimum of $\psi(u) - ut$ occurs at the point where the derivative is zero, *i.e.* for $\psi'(u) = t$. □

4. PROOFS OF THE THEOREMS

Theorem 1.1 is a direct consequence of Theorem 3.3, combined with Hoeffding's result that

$$\inf_{u\geq 0} e^{\psi(u)-ut} \leq e^{-nH(\rho^2,b,t/n)}.$$

We now turn to the proof of Theorem 1.2. For simplicity we set $h(\mu, x) = H(\mu(1-\mu), 1-\mu, x)$. First we recall that Hoeffding proves that

$$\inf_{u\geq 0} e^{\psi(u)-ut} \leq e^{-nh(\mu,t/n)} \tag{4.1}$$

Thus, recalling that $X_i = Y_i - EY_i$ by Theorem 3.3, for some universal constants K_0, K_1, we have that, for $t \leq \sigma^2/K_0$

$$P(\sum_{i\leq n} X_i \geq t) \leq \left(\theta\left(\frac{t}{\sigma}\right) + \frac{K_1}{\sigma}\right) e^{-nh(\mu,t/n)}. \tag{4.2}$$

We now show that

$$\sigma^2 \leq n\mu(1-\mu). \tag{4.3}$$

By concavity of the function $x(1-x)$, it suffices to show that $EX_i^2 \leq EY_i(1-EY_i)$. This follows from the fact that for $-p \leq x \leq 1-p$, we have $x^2 \leq p(1-p) + x(1-2p)$, so that, if $-p \leq X \leq 1-p$ and $EX = 0$, we have $EX^2 \leq p(1-p)$.

The main difficulty in the proof of Theorem 1.2 is that it can happen that $\sigma^2 \leq tK_0$, so that we cannot use (4.2). The way around this difficulty is to show in that case that, unless t is very small, using (1.2) with $b = 1$ yields a better bound than (1.7). (Thus, in that range (1.7) is indeed weaker than (1.2). The point, however, is that (1.7) involves only first moments.) By direct computation, we observe the precious fact that

$$\begin{gathered} H(\rho^2, b, 0) = 0; \quad \frac{d}{dx} H(\rho^2, b, 0) = 0 \\ \frac{d^2H}{dx^2}(\rho^2, b, x) = \frac{1}{\left(\frac{\rho^2}{b} + x\right)(b-x)} = \frac{1}{\rho^2 + x\left(b - \frac{\rho^2}{b}\right) - x^2}. \end{gathered} \tag{4.4}$$

In particular

$$\frac{d^2h(\mu,x)}{dx^2}=\frac{1}{\mu(1-\mu)-x(2\mu-1)-x^2}. \tag{4.5}$$

We fix t, and we assume $\sigma^2\le tK_0$, so that $\rho^2\le tK_0/n$. For $x\le t/n$, we have

$$\rho^2+x(1-\rho^2)-x^2\le\rho^2+x\le(1+K_0)t/n.$$

Thus, by (4.4) and integration we get

$$H(\rho^2,1,x)\ge\frac{n}{2t(1+K_0)}x^2$$

and thus

$$nH(\rho^2,1,t/n)\ge\frac{t}{2(1+K_0)}. \tag{4.6}$$

On the other hand, for $x\le\mu(1-\mu)/2$, we have

$$\begin{aligned}\mu(1-\mu)-x(2\mu-1)-x^2&\ge\mu(1-\mu)-\frac{\mu(1-\mu)}{2}-\frac{(\mu(1-\mu))^2}{4}\\&\ge\mu(1-\mu)\left(\frac{1}{2}-\frac{\mu(1-\mu)}{4}\right)\ge\mu(1-\mu)/3.\end{aligned}$$

Thus, by (4.5) and integration, we get for $t\le n\mu(1-\mu)/2$, that

$$nh(\mu,t/n)\le\frac{3t^2}{2n\mu(1-\mu)}. \tag{4.7}$$

Since $\theta(\lambda)\ge(\sqrt{2\pi}(1+\lambda))^{-1}$, the bound given by (1.7) is larger than

$$B_1=\frac{1}{\sqrt{2\pi}(1+t/\sqrt{n\mu(1-\mu)})}\exp\left(-\frac{3t^2}{2n\mu(1-\mu)}\right),$$

while, by (4.6), the bound given by (1.2) is less than

$$B_2=\exp\left(-\frac{t}{2(1+K_0)}\right).$$

It follows that, for some universal constant K, if $t \geq K$, $t \leq n\mu(1-\mu)/K$, then $B_2 \leq B_1$. $\Bigg($Observe that then

$$\frac{1}{1+t/\sqrt{n\mu(1-\mu)}} \geq \frac{1}{1+\sqrt{Kt}}\Bigg).$$

Thus, we have shown that provided $t \geq K$, we can always assume $t \leq \sigma^2/K_0$, so that (4.3) holds. We now observe that (provided K_0 is large enough) the function

$$f(\sigma) = \theta\left(\frac{t}{\sigma}\right) + \frac{K_1}{\sigma}$$

increases for $t \geq K_0$ and $\sigma^2 \geq K_0 t$. Indeed, we have

$$f'(\sigma) = -\frac{1}{\sigma^2}\left(K_1 + t\theta'\left(\frac{t}{\sigma}\right)\right), \tag{4.8}$$

and the claim follows from Lemma 3.4. Theorem 1.2 is proved. □

We now prove Corollary 1.3. By (4.5), we have

$$\frac{d^2h}{dx^2}(\mu, x) = \frac{1}{\frac{1}{4} - \left(x + \left(\mu - \frac{1}{2}\right)\right)^2} \geq 4 + 16\left(x + \left(\mu - \frac{1}{2}\right)\right)^2$$

so that, by integration

$$\begin{aligned}\frac{dh}{dx}(\mu, x) &\geq 4x + \frac{16}{3}\left(\left(x + \left(\mu - \frac{1}{2}\right)\right)^3 - \left(\mu - \frac{1}{2}\right)^3\right) \\ &\geq 4x + \frac{4}{3}x^3\end{aligned}$$

since $(x+a)^3 - a^3$ is minimum at $a = -x/2$. Thus, by integration

$$h(\mu, x) \geq 2x^2 + \frac{x^4}{4}. \tag{4.9}$$

By (4.2), we have

$$P\Big(\sum_{i \leq n} X_i \geq n\mu + t\Big) \leq \exp\left(-\frac{2t^2}{n} - \frac{t^4}{4n^3}\right).$$

If $t \geq n^{7/8}$, this is better than the bound of (1.8) for n sufficiently large (which we can certainly assume). Thus it suffices to consider the case $t \leq n^{7/8}$. For $x \leq t/n$, we have

$$\mu(1-\mu) + x(2\mu - 1) - x^2 \leq \mu(1-\mu) + t/n.$$

Thus, by (4.4) again

$$nh(\mu, t/n) \geq \frac{t^2}{2n(\mu(1-\mu) + t/n)}.$$

Thus, by (4.5), and the argument of Theorem 1.2, it suffices to consider the case $t \leq n\mu(1-\mu)/K$. Then (1.7) holds. By (4.9), we have $h(\mu, t/n) \geq 2t^2/n$; and, as shown in the proof of Theorem 2.1, we have

$$\theta\left(\frac{t}{\sqrt{n\mu(1-\mu)}}\right) + \frac{K}{\sqrt{n\mu(1-\mu)}} \leq \theta\left(\frac{2t}{\sqrt{n}}\right) + \frac{K}{\sqrt{n}}. \qquad \square$$

REFERENCES

[B] G. BENNETT, Probability inequalities for the sum of independent random variables, *Amer. Stat. Assoc. J.*, **57**, 1962, pp. 33-45.

[F] W. FELLER, *An introduction to Probability Theory and its applications*, Vol. **II** (second edition), J. Wiley and Sons, 1971.

[H] W. HOEFFDING, Probability inequalities for sums of bounded random variables, *Amer. Stat. Assoc. J.*, **58**, 1963, pp. 13-30.

[K-J] S. KOTZ and N. L. JOHNSON, *Encyclopedia of Statistical Sciences*, Wiley, New York, 1982.

[I-K] K. ITO and H. P. MCKEAN, *Diffusion processes and their sample paths*, 1974, Springer Verlag.

[N] S. V. NAGAEV, Large deviations of sums of independent random variables, *Ann. Probab.*, **7**, 1977, pp. 745-789.

[P-U] I. S. PINELIS and S. A. UTEV, Exact exponential bounds for sums of independent random variables, *Theory Probab. Appl.*, 1990, pp. 340-345.

[P] YU V. PROKHOROV, An extremal problem in probability theory, *Theory of Probability and its applications*, **4**, 1959, pp. 201-3.

*(Manuscript received September 24, 1993;
revised June 30, 1994.)*