Information Processing Letters 33 (1989/90) 305–308
North-Holland

A GUIDED TOUR OF CHERNOFF BOUNDS

Torben HAGERUP and Christine RÜB

Fachbereich Informatik, Universität des Saarlandes, D-6600 Saarbrücken, FRG

Communicated by R. Wilhelm
Received 14 April 1989
Revised 4 September 1989

We give elementary derivations of the various inequalities collectively known as Chernoff bounds. Chernoff bounds are strong upper bounds on the probability of obtaining very few or very many heads in series of independent coin tossings. This note aims at making known results and their proofs accessible to a wider audience; it contains little or no new material.

Keywords: Chernoff bounds, coin tossing, Bernoulli trials, probabilistic analysis

The following notation is used throughout: $\Pr(A)$ denotes the probability of an event A, $E(X)$ the expected value of a random variable X, ln the natural logarithm function and exp its inverse. We write $\exp(x)$ and e^x interchangeably. We need the inequalities stated in the following lemma.

Lemma 1.

$$1 + a \leqslant e^a, \quad a \in \mathbb{R}. \tag{1}$$

$$\left(1 + \frac{b}{x}\right)^x \leqslant e^b, \quad b \in \mathbb{R}, \ x > 0. \tag{2}$$

$$-\tfrac{1}{2}\epsilon^2 \leqslant \epsilon - (1+\epsilon)\ln(1+\epsilon) \leqslant -\tfrac{1}{3}\epsilon^2, \quad 0 \leqslant \epsilon \leqslant 1. \tag{3}$$

$$-\tfrac{1}{2}\epsilon^2 \geqslant \epsilon - (1+\epsilon)\ln(1+\epsilon), \quad -1 < \epsilon \leqslant 0. \tag{4}$$

Proof. Let $g(a) = e^a - (1+a)$, $a \in \mathbb{R}$. Then $g''(a) = e^a > 0$ for all $a \in \mathbb{R}$, while $g'(0) = 0$. Hence $g(a) \geqslant g(0) = 0$ for all $a \in \mathbb{R}$, which proves (1). Putting $a = b/x$ and raising both sides of (1) to the xth power gives (2). Now for $q \in \{0, 1\}$, let

$$f_q(\epsilon) = \epsilon - (1+\epsilon)\ln(1+\epsilon) + \tfrac{1}{2}\epsilon^2 - \tfrac{1}{6}q\epsilon^3, \quad -1 < \epsilon \leqslant 1.$$

Then

$$f_q'(\epsilon) = -\ln(1+\epsilon) + \epsilon - \tfrac{1}{2}q\epsilon^2,$$

$$f_q''(\epsilon) = -\frac{1}{1+\epsilon} + 1 - q\epsilon,$$

and

$$f_q'''(\epsilon) = \frac{1}{(1+\epsilon)^2} - q.$$

We can successively deduce the signs of the derivatives f_q''', f_q'', f_q' and of f_q, as given in Table 1.

Table 1

q	0			1		
	$\epsilon<0$	$\epsilon=0$	$\epsilon>0$	$\epsilon<0$	$\epsilon=0$	$\epsilon>0$
$f_q'''(\epsilon)$	+	+	+	+	0	−
$f_q''(\epsilon)$	−	0	+	−	0	−
$f_q'(\epsilon)$	+	0	+	+	0	−
$f_q(\epsilon)$	−	0	+	−	0	−

The sign variation of f_0 immediately implies (4) and the left half of (3). From the sign variation of f_1 we get for $0 \leqslant \epsilon \leqslant 1$,

$$\epsilon - (1+\epsilon)\ln(1+\epsilon) \leqslant -\epsilon^2\left(\tfrac{1}{2} - \tfrac{1}{6}\epsilon\right) \leqslant \tfrac{1}{3}\epsilon^2. \qquad \square$$

Let $n \in \mathbb{N}$ and let $p_1, \ldots, p_n \in \mathbb{R}$ with $0 \leqslant p_i \leqslant 1$, $i = 1, \ldots, n$. Put $p = (p_1 + \cdots + p_n)/n$ and $m = np$ and let $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$ be independent 0-1 random variables with

$$\Pr(X_i = 1) = p_i \quad \text{for } i = 1, \ldots, n, \qquad \Pr(Y_i = 1) = p \quad \text{for } i = 1, \ldots, n.$$

We are interested in the behaviour of the random variables $S = X_1 + \cdots + X_n$ and $S' = Y_1 + \cdots + Y_n$. We first quickly derive the most useful results.

Let $\epsilon \geqslant 0$ and $t \geqslant 0$. Then

$$\Pr(S \geqslant (1+\epsilon)m) \leqslant e^{-t(1+\epsilon)m}\, e^{t(1+\epsilon)m}\, \Pr(e^{tS} \geqslant e^{t(1+\epsilon)m}) \leqslant e^{-t(1+\epsilon)m} E(e^{tS}).$$

Since $X_1, \ldots, X_n$ are independent, we further get (also using (1))

$$E(e^{tS}) = E(e^{t(X_1 + \cdots + X_n)}) = E(e^{tX_1} \cdots e^{tX_n}) = \prod_{i=1}^{n} E(e^{tX_i}) = \prod_{i=1}^{n} \left(p_i\, e^t + (1-p_i)\right)$$

$$= \prod_{i=1}^{n} \left(1 + p_i(e^t - 1)\right) \leqslant \prod_{i=1}^{n} e^{p_i(e^t-1)} = \exp\left(\sum_{i=1}^{n} p_i(e^t - 1)\right) = e^{m(e^t-1)}.$$

Putting $t = \ln(1+\epsilon)$ yields

$$\Pr(S \geqslant (1+\epsilon)m) \leqslant (1+\epsilon)^{-(1+\epsilon)m} e^{m\epsilon}$$

and hence

$$\Pr(S \geqslant (1+\epsilon)m) \leqslant \left(\frac{e^{\epsilon}}{(1+\epsilon)^{1+\epsilon}}\right)^{m}. \tag{5}$$

By the right half of (3),

$$\Pr(S \geqslant (1+\epsilon)m) \leqslant e^{-\epsilon^2 m/3}, \quad 0 \leqslant \epsilon \leqslant 1. \tag{6}$$

Correspondingly, for $0 \leqslant \epsilon \leqslant 1$ and $t \geqslant 0$,

$$\Pr(S \leqslant (1-\epsilon)m) = \Pr(m - S \geqslant \epsilon m) \leqslant \Pr(e^{t(m-S)} \geqslant e^{t\epsilon m})$$

$$\leqslant e^{-t\epsilon m} E(e^{t(m-S)}) = e^{tm(1-\epsilon)} E(e^{-tS})$$

and

$$E(e^{-tS}) = \prod_{i=1}^{n} E(e^{-tX_i}) = \prod_{i=1}^{n} (p_i e^{-t} + (1-p_i)) = \prod_{i=1}^{n} (1 - p_i(1-e^{-t}))$$

$$\leqslant \prod_{i=1}^{n} e^{-p_i(1-e^{-t})} = \exp\left(-(1-e^{-t})\sum_{i=1}^{n} p_i\right) = e^{-m(1-e^{-t})}.$$

Putting $t = -\ln(1-\epsilon)$ yields

$$\Pr(S \leqslant (1-\epsilon)m) \leqslant \left[\left(\frac{1}{1-\epsilon}\right)^{1-\epsilon} e^{-\epsilon}\right]^m,$$

from which, by (4), the left half of (3), and continuity at $\epsilon = 1$, we get

$$\Pr(S \leqslant (1-\epsilon)m) \leqslant e^{-\epsilon^2 m/2} \leqslant \left(\frac{e^{\epsilon}}{(1+\epsilon)^{1+\epsilon}}\right)^m, \quad 0 \leqslant \epsilon \leqslant 1. \tag{7}$$

As a consequence of (5),

$$\Pr(S \geqslant (1+\epsilon)m) \leqslant \left(\frac{e}{1+\epsilon}\right)^{(1+\epsilon)m}.$$

In particular,

$$\Pr(S \geqslant r) \leqslant 2^{-r}, \quad r \geqslant 6m. \tag{8}$$

We next derive some sharper but more complicated bounds. We henceforth consider only the case $p_1 = \cdots = p_n = p$. Let $0 < a < 1$, $a \geqslant p$ and $t \geqslant 0$. Choosing ϵ to make $(1+\epsilon)m = an$, we get from previous calculations

$$\Pr(S' \geqslant an) \leqslant e^{-tan}(p\,e^{t} + (1-p))^n.$$

For $t = \ln(a(1-p)/[p(1-a)])$ this becomes

$$\Pr(S' \geqslant an) \leqslant \left(\frac{p(1-a)}{a(1-p)}\right)^{an}\left(\frac{a(1-p)}{1-a} + (1-p)\right)^n = \left(\frac{p(1-a)}{a(1-p)}\right)^{an}\left(\frac{1-p}{1-a}\right)^n,$$

from which we obtain

$$\Pr(S' \geqslant an) \leqslant \left[\left(\frac{p}{a}\right)^a \left(\frac{1-p}{1-a}\right)^{1-a}\right]^n, \quad 0 < a < 1, \ a \geqslant p. \tag{9}$$

Let us introduce the abbreviation $S' \gtrsim k$ ("S' is at least as extreme as k") defined by

$$S' \gtrsim k \Leftrightarrow \begin{cases} S' \geqslant k & \text{if } k \geqslant pn, \\ S' \leqslant k & \text{if } k < pn. \end{cases}$$

Noting that the right-hand side of (9) is invariant under a simultaneous interchange of a with $1-a$ and p with $1-p$, we then get by considering the random variable $n - S'$,

$$\Pr(S' \gtrsim an) \leqslant \left[\left(\frac{p}{a}\right)^a \left(\frac{1-p}{1-a}\right)^{1-a}\right]^n, \quad 0 < a < 1. \tag{10}$$

Since by (2)

$$\left(\frac{1-p}{1-a}\right)^{1-a} = \left(1+\frac{a-p}{1-a}\right)^{1-a} \leqslant e^{a-p},$$

we also have

$$\Pr(S' \geqslant an) \leqslant \left[\left(\frac{p}{a}\right)^{a} e^{a-p}\right]^{n}, \quad 0 < a \leqslant 1. \tag{11}$$

Putting $na = k$, we may finally derive the inequality

$$\Pr(S' \geqslant k) \leqslant \left(\frac{np}{k}\right)^{k}\left(\frac{n-np}{n-k}\right)^{n-k} \leqslant \left(\frac{np}{k}\right)^{k} e^{k-np}, \quad 0 < k < n. \tag{12}$$

Bibliographic remarks

The material in this note was drawn from several sources. The fundamental technique goes back to Chernoff [2]. While we have not reproduced any of Chernoff's original formulas, putting $r = 1$ in his equation (5.11) and exponentiating it while combining it with equations (3.5) and (3.6) yields our equation (10). Our treatment is based mostly on the derivation in [6], and equations (5) and (7) were taken from there. Equation (6) and the left half of (7) apparently were first formulated in [1]. Equations (9) and (11) are from [5], while (12) appears in [7] and a close relative of the left half of (12) is the form preferred in [3]. More general results may be found in [4, p.104, Theorems 6 and 7].

References

[1] D. Angluin and L.G. Valiant, Fast probabilistic algorithms for Hamiltonian circuits and matchings, *J. Comput. System Sci.* **18** (1979) 155–193.

[2] H. Chernoff, A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations, *Ann. Math. Statist.* **23** (1952) 493–507.

[3] P. Erdös and J. Spencer, *Probabilistic Methods in Combinatorics* (Academic Press, New York, 1974).

[4] M. Hofri, *Probabilistic Analysis of Algorithms* (Springer, New York, 1987).

[5] R.M. Karp, Probabilistic analysis of algorithms, Class Notes, Univ. California, Berkeley, 1988.

[6] P. Raghavan, Probabilistic construction of deterministic algorithms: Approximating packing integer programs, *J. Comput. System Sci.* **37** (1988) 130–143.

[7] L.G. Valiant and G.J. Brebner, Universal schemes for parallel communication, In: *Proc. 13th Ann. ACM Symp. on Theory of Computing* (1979) 263–277.