Two Refinements of the Chernoff Bound for the Sum of Nonidentical Bernoulli Random Variables

Ye Xia

Department of Computer and Information Science and Engineering, University of Florida
301 CSE Building, P.O. Box 116120, Gainesville, Florida 32611-6120, USA

Abstract

We give two refinements to the Chernoff bound for the sum of nonidentical Bernoulli random variables with parameters p_i, where $0 < p_i < 1$ and $i = 1, ..., n$. Traditionally, the Chernoff bound is a function of the arithmetic mean of the p_i's, $n^{-1}\sum_{i=1}^n p_i$. The refined bounds contain the term $n^{-1}\sum_{i=1}^n p_i^2$, and hence, are able to capture the variations of p_i's.

Key words: Chernoff bound, Bernoulli Random Variable

---

1 Original Chernoff Bound

Let $X_1, X_2, ..., X_n$ be independent Bernoulli random variables, such that $P\{X_i = 1\} = p_i$, where $0 < p_i < 1$, for $1 \leq i \leq n$. Let $X = \sum_{i=1}^n X_i$, and let $p = n^{-1}\sum_{i=1}^n p_i$. Let $m = (1+\delta)np$ for $\delta > 0$. Then, the Chernoff bound for the sum of the Bernoulli random variables is (Motwani and Raghavan, 1995)

$$P\{X > m\} \leq \left(\frac{e^\delta}{(1+\delta)^{(1+\delta)}} \right)^{np}. \tag{1}$$

Note that the Chernoff bound in (1) does not capture the effect of the variation in the parameters, p_i, where $1 \leq i \leq n$. In this paper, we will find alternative bounds that capture the variation.

Let us define $\nu = n^{-1}\sum_{i=1}^n p_i^2$. Throughout the paper, we assume $n > m$, which is equivalent to $p < 1/(1+\delta)$. The cases where $n \leq m$ are trivial, with $P\{X > m\} = 0$.

Email address: `yx1@cise.ufl.edu` (Ye Xia).

In all subsequent refinements for the Chernoff bound, the tail probabilities have the form $P\{X > m\} = P\{X > (1+\delta)np\} \leq \theta^n$, where $0 < \theta < 1$ is dependent of n. For the Chernoff bound in (1), $\theta_{CHERN} = e^{\delta p}/(1+\delta)^{(1+\delta)p}$. One related work is from a special case of Lemma 1 in Chapter X of Petrov (1975). When applied to the Bernoulli random variables, it can be shown that, if $\nu \leq (1-\delta)p$, one can choose $\theta = \exp(-\frac{\delta^2 p^2}{2(p-\nu)}(1 - \frac{\delta p}{2(p-\nu)}))$. Due to the restricting condition for this result to hold, we will not explore this result further.

2 Refinement Using the Inequality $1 + x \leq e^{x-ax^2}$

Lemma 2.1 *Suppose* $0 \leq x \leq u$ *for some positive* u. *Then,* $1 + x \leq e^{x-ax^2}$, *where* $a = (u - \log(1+u))/u^2 > 0$.

Proof: First, $a > 0$ follows from $u > \log(1+u)$ for $u > 0$. Next, consider the function $f(x) \triangleq e^{x-ax^2} - (1+x)$. Note that $f(0) = f(u) = 0$. We claim that $f(x)$ first increases and then decreases on $[0, u]$. Hence, $f(x) \geq 0$ on $[0, u]$.

The first and second derivatives of $f(x)$ are $f'(x) = e^{x-ax^2}(1-2ax) - 1$, $f''(x) = e^{x-ax^2}((1-2ax)^2 - 2a)$. Let us define $g(x) = (1-2ax)^2 - 2a$. We will show that, for $u > 0$, $g(0) = 1 - 2a > 0$, and hence, $f''(0) > 0$. Since $g(0) = u^{-2}(u^2 - 2u + 2\log(1+u))$, it suffices to show that, for $u > 0$, $h(u) \triangleq u^2 - 2u + 2\log(1+u) > 0$. Note that $h(0) = 0$ and that, for $u > 0$, $h'(u) = 2u^2/(1+u) > 0$. Then, it must be true that $h(u) > 0$ for $u > 0$. Otherwise, suppose $h(u_o) \leq 0$ for some $u_o > 0$. By the Mean-Value theorem, there must exist some $u_1 \in (0, u_o)$, such that $h'(u_1) = (h(u_o) - h(0))/u_o \leq 0$. With a similar argument, we can show that $f'(u) < 0$ for $u > 0$

We have established that $f'(0) = 0$, $f''(0) > 0$, and for any $u > 0$, $f'(u) < 0$. In order to satisfy this, $f'(x)$ must crosses the x-axis an odd number of times on $(0, u]$. We claim that it crosses the x-axis only once on $(0, u]$. Suppose it crosses the x-axis at least three times. Then, there will be at least three critical points on $[0, u]$. This violates the fact that $f''(x)$ has only two zeros on $\mathbb{R}$. As a result, starting at $f'(0) = 0$, $f'(x)$ first strictly increases above 0 and then strictly decreases below 0 on $[0, u]$. Combining with the fact that $f(u) = 0$, $f(x)$ first increases above 0 and then decreases back to 0 at $x = u$. Hence, $f(x) \geq 0$ for all $x \in [0, u]$. ∎

Theorem 2.2

$$P\{X > m\} \leq (\theta_{EXP\_REFINE})^n, \tag{2}$$

where

$$\theta_{EXP\_REFINE} = \frac{e^{p\delta}}{\left(\frac{(1+\delta)p-\nu}{p-\nu}\right)^{(1+\delta)p-\nu}}. \tag{3}$$

Proof: For any $t > 0$, $P\{X > m\} \leq e^{-tm}\mathbf{E}e^{tX} = e^{-tm}\prod_{i=1}^{n}\mathbf{E}e^{tX_i}$, by the Markov inequality and the independence assumption of the X_i's. The moment generating function for X_i is $\mathbf{E}e^{tX_i} = 1 + p_i(e^t - 1)$. Hence, $P\{X > m\} \leq \prod_{i=1}^{n}(1+p_i(e^t-1))e^{-tm}$. Applying Lemma 2.1 with $x = p_i(e^t-1)$ and $u = e^t - 1$, we get $a = (e^t - 1 - t)/(e^t - 1)^2$, and

$$\prod_{i=1}^{n}(1 + p_i(e^t - 1))e^{-tm} \leq \exp\Big(np(e^t - 1) - n\nu(e^t - 1 - t) - tm\Big). \qquad (4)$$

For $p \neq \nu$, the minimum of (4) is achieved at $t^* = \log\Big(((1+\delta)p - \nu)/(p - \nu)\Big)$, and its value is as the bound in the theorem. ■

Recall that, throughout the paper, we assume $0 < p_i < 1$ for all i. Under this assumption, it is always true that $\nu < p$. It can be shown that the derivative of θ_{EXP_REFINE} with respect to ν is $\log(1 + p\delta/(p-\nu)) - p\delta/(p-\nu)$, which is less than zero. Hence, as ν increases, θ_{EXP_REFINE} decreases. Finally, it can be shown easily that $\theta_{EXP\_REFINE} \leq \theta_{CHERN}$. The strength of this refinement over the original Chernoff bound lies in the fact that e^{x-ax^2} is a much tighter bound for $1 + x$ than e^x on $[0, u]$.

3 Refinement Using a Refined Arithmetic Mean-Geometric Mean Inequality

Out of various refinements for the arithmetic mean-geometric mean inequality, the following is suitable for our problem.

Theorem 3.1 *(Cartwright and Field, 1978) Suppose that $x_k \in [\alpha, \beta]$ and $q_k \geq 0$ for $k = 1, ..., n$, where $\alpha > 0$, and suppose that $\sum_{k=1}^{n} q_k = 1$. Then, writing $\bar{x} = \sum_{k=1}^{n} q_k x_k$, we have*

$$\frac{1}{2\beta}\sum_{k=1}^{n} q_k(x_k - \bar{x})^2 \leq \bar{x} - \prod_{k=1}^{n} x_k^{q_k} \leq \frac{1}{2\alpha}\sum_{k=1}^{n} q_k(x_k - \bar{x})^2. \qquad (5)$$

Define $\sigma^2 = n^{-1}\sum_{i=1}^{n} p_i^2 - (n^{-1}\sum_{k=1}^{n} p_i)^2$. Note that $\sigma^2 \leq p - p^2$.

Theorem 3.2

$$P\{X > m\} \leq (\theta_{AG\_REFINE})^n, \qquad (6)$$

where

$$\theta_{AG\_REFINE} = (1 + py_1 - \frac{\sigma^2 y_1^2}{2(1 + y_1)})(\frac{1}{y_1 + 1})^{(1+\delta)p}. \qquad (7)$$

Here, y_1 *is given by* $y_1 = \left(-a_1 + \sqrt{a_1^2 - 4a_0a_2}\right)/(2a_2) > 0$*, and* a_0*,* a_1 *and* a_2 *are given by*

$$a_2 = (1-(1+\delta)p)(2p-\sigma^2)$$
$$a_1 = 2(p - p\delta - p^2 - p^2\delta - \sigma^2)$$
$$a_0 = -2\delta p.$$

Proof: Applying Theorem 3.1 with x_k replaced by $1 + p_k(e^t - 1)$, $\beta = e^t$ and $q_k = 1/n$ for all k, we get

$$\prod_{i=1}^{n}(1 + p_i(e^t - 1))e^{-tm} \le (1 + p(e^t - 1) - \frac{1}{2}e^{-t}\sigma^2(e^t-1)^2)^n e^{-tm}. \quad (8)$$

We will minimize the right hand side of (8). Define $y \triangleq e^t - 1$, and

$$\phi(y) \triangleq \log(1 + py - \frac{\sigma^2 y^2}{2(1+y)}) - (1+\delta)p\log(y+1).$$

The right hand side of (8) is equal to $\exp(n\phi(y))$. The derivative of $\phi(y)$ is

$$\phi'(y) = \frac{a_2y^2 + a_1y + a_0}{(y+1)((2p-\sigma^2)y^2 + 2(1+p)y + 2)}, \quad (9)$$

where a_0, a_1 and a_2 are given as in the statement of the theorem. Let the numerator and denominator of (9) be denoted by $g_n(y)$ and $g_d(y)$, respectively. It is easy to show that the roots of $g_d(y)$ are not greater than 0, and hence, $\phi'(y)$ is continuous on $y > 0$. Under the condition $1 - (1+\delta)p > 0$, $p > 0$ and $\delta > 0$, we have $a_2 > 0$ and $-4a_0a_2 > 0$. Hence, the roots of $g_n(y) = 0$ are real, and the only positive root is $y_1 = \left(-a_1 + \sqrt{a_1^2 - 4a_0a_2}\right)/(2a_2) > 0$. Since $h_d(0) = 2$ and $h_d(y)$ does not have a zero on $y > 0$, it must be true $h_d(y) > 0$ for $y > 0$. Therefore, at y_1, $\phi'(y)$ crosses the x-axis increasingly, and hence, $\phi''(y_1) > 0$. Thus, y_1 is the minimum of $\phi(y)$ for $y > 0$. The right hand side of (8) achieves the minimum on $t > 0$ at $t^* = \log(1 + y_1)$, and the minimum is as the bound in the theorem. ■

Note that it must be true that $\theta_{AG\_REFINE} \le \theta_{AG}$.

References

D.I. Cartwright and M. J. Field. A Refinement of the Arithmetic Mean-Geometric Mean Inequality. *Proceedings of the American Mathematical Society*, 71(1):36–38, August 1977.

Rajeev Motwani and Prabhakar Raghavan. *Randomized Algorithms*. Cambridge University Press, 1995.

V. V. Petrov. *Sums of Independent Random Variables*. Springer Verlag, 1975.