An improved Hoeffding's inequality

Songfeng Zheng

Department of Mathematics
Missouri State University
Springfield, MO 65897
SongfengZheng@MissouriState.edu

Abstract. Using a refined arithmetic-geometric mean inequality, this paper gives an improved version of Hoeffding's inequality which has a closed form and is easy to evaluate. Numerical simulations comparing the performance of the proposed inequality and the original Hoeffding's inequality are also presented.

1 Introduction

In probability theory, Hoeffding's inequality [5] provides an upper bound on the probability that the sum of random variables deviates from its expected value. It has wide applications in random algorithm analysis [4], statistical learning theory [2, 6], information theory [7], and random matrix theory [1], to name a few. For convenience of reference, we give the full statement of Hoeffding's inequality and the complete proof below.

Theorem 1. (Hoeffding's Inequality) *Let $X_1, X_2, \cdots, X_n$ be independent (not necessarily identically distributed) random variables with $P(0 \le X_i \le 1) = 1$ for $i = 1, 2, \cdots, n$. Let $p_i = \mathrm{E}(X_i)$ for $i = 1, 2, \cdots, n$, $p = \frac{1}{n}\sum_{i=1}^n p_i$, and $q = 1 - p$. Then for any $0 < t < q$*

$$P(\bar{X} > p + t) \le \exp\left(-n\left[(p+t)\log\frac{p+t}{p} + (q-t)\log\frac{q-t}{q}\right]\right) \equiv U_H, \tag{1}$$

where $\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$.

Proof. Let $X = \sum_{i=1}^n X_i$. For any $\lambda > 0$, we have

$$P(\bar{X} > p + t) = P(X > n(p+t)) = P(e^{\lambda X} > e^{n\lambda(p+t)}) \le \frac{\mathrm{E}[e^{\lambda X}]}{e^{n\lambda(p+t)}} \tag{2}$$

$$= \frac{\mathrm{E}[e^{\lambda \sum_{i=1}^n X_i}]}{e^{n\lambda(p+t)}} = \frac{\mathrm{E}[\prod_{i=1}^n e^{\lambda X_i}]}{e^{n\lambda(p+t)}} = \frac{\prod_{i=1}^n \mathrm{E}[e^{\lambda X_i}]}{e^{n\lambda(p+t)}}, \tag{3}$$

where the inequality in Eq. (2) follows from Markov's inequality, and Eq. (3) uses the independence assumption.

Using the convexity of function $e^{\lambda x}$ on the interval $[0, 1]$, there is

$$e^{\lambda x} \le 1 - x + xe^{\lambda}. \tag{4}$$

Applying Eq. (4) to Eq. (3), yields

$$
\begin{aligned}
P(\bar{X} > p+t) &\leq e^{-n\lambda(p+t)} \prod_{i=1}^{n} \mathrm{E}[1 - X_i + X_i e^{\lambda}] \\
&= \exp\left(-n\lambda(p+t)\right) \prod_{i=1}^{n} (1 - p_i + p_i e^{\lambda}) \\
&\leq \exp\left(-n\lambda(p+t)\right) \left(\frac{\sum_{i=1}^{n}(1 - p_i + p_i e^{\lambda})}{n} \right)^{n} && (5) \\
&= \exp\left(-n\lambda(p+t)\right) (q + pe^{\lambda})^{n} && (6) \\
&= \exp\left(-n\left[\lambda(p+t) - \log(q + pe^{\lambda})\right]\right), && (7)
\end{aligned}
$$

where in Eq. (5), we used the arithmetic-geometric mean inequality which states that for positive $x_1, x_2, \cdots, x_n$,

$$
\prod_{i=1}^{n} x_i \leq \left(\frac{1}{n} \sum_{i=1}^{n} x_i \right)^{n}. \tag{8}
$$

Choosing λ to minimize Eq. (7), we obtain

$$
\lambda^* = \log\left[\frac{(p+t)q}{p(q-t)}\right]. \tag{9}
$$

Substituting λ^* into Eq. (7), we can obtain the desired result in Eq. (1). □

2 The improved Hoeffding's inequality

In the proof of Theorem 1, the arithmetic-geometric mean inequality in Eq. (8) was used as a major ingredient. There is a refined version of the arithmetic-geometric mean inequality which makes it possible to further improve the Hoeffding's inequality in Eq. (1).

Lemma 1. (Refined Arithmetic-Geometric Mean Inequality) *Suppose that $x_k \in [a,b]$ and $p_k \geq 0$ for $k = 1, 2, \cdots, n$, where $a > 0$, and further suppose that $\sum_{k=1}^{n} p_k = 1$. Let $\bar{x} = \sum_{k=1}^{n} p_k x_k$. We have*

$$
\frac{1}{2b} \sum_{k=1}^{n} p_k (x_k - \bar{x})^2 \leq \bar{x} - \prod_{k=1}^{n} x_k^{p_k} \leq \frac{1}{2a} \sum_{k=1}^{n} p_k (x_k - \bar{x})^2. \tag{10}
$$

The proof of Lemma 1 is given in [3]. We are interested in a special case in which each $p_k = 1/n$, then

$$
\sum_{k=1}^{n} p_k (x_k - \bar{x})^2 = \frac{1}{n} \sum_{k=1}^{n} (x_k - \bar{x})^2 = \mathrm{var}(\mathbf{x}),
$$

where $\mathbf{x} = (x_1, x_2, \cdots, x_n)$, and $\text{var}(\mathbf{x})$ denotes the variance[1] of the array $\mathbf{x} = (x_1, x_2, \cdots, x_n)$. Thus, for this special case, there is

$$\frac{1}{2b}\text{var}(\mathbf{x}) \le \frac{x_1 + x_2 + \cdots + x_n}{n} - \left(\prod_{k=1}^{n} x_k\right)^{1/n} \le \frac{1}{2a}\text{var}(\mathbf{x}),$$

and consequently

$$\prod_{k=1}^{n} x_k \le \left(\frac{x_1 + x_2 + \cdots + x_n}{n} - \frac{1}{2b}\text{var}(\mathbf{x})\right)^n. \tag{11}$$

By comparing Eq. (8) to Eq. (11), clearly, Eq. (11) provides a tighter bound of the product of the given numbers by incorporating the variance information.

Theorem 2. (Improved Hoeffding's Inequality) *With the same notations and settings as in Theorem 1, let*

$$A = (q - t)(p - \frac{\sigma^2}{2}), \quad B = -(p + t)(q + \sigma^2), \quad C = \frac{\sigma^2}{2}(1 + p + t),$$

where $\sigma^2 = \sum_{i=1}^{n}(p_i - p)^2/n$. *Denote*

$$w = \frac{-B + \sqrt{B^2 - 4AC}}{2A}$$

Then

$$P(\bar{X} > p + t) \le w^{-n(p+t)}\left(q + pw - \frac{\sigma^2(w-1)^2}{2w}\right)^n \equiv U_N. \tag{12}$$

Before we give the proof of Theorem 2, we summarize some useful properties of the coefficients A, B and C as

Lemma 2. *With the same notations as Theorem 2, there is* $A > 0$, $A + B + C < 0$, *and* $B^2 - 4AC > 0$.

Proof. Since

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(p_i - p)^2 = \frac{1}{n}\sum_{i=1}^{n}p_i^2 - p^2,$$

there is

$$p - \sigma^2 = p - \frac{1}{n}\sum_{i=1}^{n}p_i^2 + p^2 = \frac{1}{n}\sum_{i=1}^{n}(p_i - p_i^2) + p^2 > 0.$$

Consequently,

$$A = (q - t)\left(p - \frac{\sigma^2}{2}\right) > 0, \tag{13}$$

[1] Although it is not exactly the variance, calling it as variance simplifies our notation greatly.

because $q > t$. We also have the following:

$$\begin{aligned} A+B+C &= (q-t)\left(p-\frac{\sigma^2}{2}\right) - (p+t)(q+\sigma^2) + \frac{\sigma^2}{2}(1+p+t) \\ &= (pq - pt - q\sigma^2/2 + t\sigma^2/2) - (pq + tq + p\sigma^2 + t\sigma^2) \\ &\quad + (\sigma^2/2 + p\sigma^2/2 + t\sigma^2/2) \\ &= -pt - qt - q\sigma^2/2 - p\sigma^2 + \sigma^2/2 + p\sigma^2/2 \\ &= -t < 0, \end{aligned} \tag{14}$$

where the last equality follows since $p + q = 1$.

Finally, let's calculate

$$\begin{aligned} B^2 - 4AC &= (p+t)^2(q+\sigma^2)^2 - 4(q-t)\left(p-\frac{\sigma^2}{2}\right)\frac{\sigma^2}{2}(1+p+t) \\ &= (p^2+t^2+2pt)(q^2+\sigma^4+2q\sigma^2) - (\sigma^2+\sigma^2 p+\sigma^2 t)(2pq - 2pt - q\sigma^2 + t\sigma^2) \\ &= p^2q^2 + q^2t^2 + 2pq^2t + p^2\sigma^4 + t^2\sigma^4 + 2pt\sigma^4 + 2p^2q\sigma^2 + 2q\sigma^2t^2 + 4pqt\sigma^2 \\ &\quad - (2pq\sigma^2 - 2pt\sigma^2 - q\sigma^4 + t\sigma^4 + 2p^2q\sigma^2 - 2p^2t\sigma^2 - pq\sigma^4 + pt\sigma^4 \\ &\qquad + 2pqt\sigma^2 - 2pt^2\sigma^2 - qt\sigma^4 + t^2\sigma^4) \\ &= (p^2q^2 + p^2\sigma^4 - 2pq\sigma^2 + q\sigma^4 + pq\sigma^4) + (2pt\sigma^4 - t\sigma^4 - pt\sigma^4 + qt\sigma^4) \\ &\quad + (4pqt\sigma^2 - 2pqt\sigma^2) + (q^2t^2 + 2pq^2t + 2q\sigma^2t^2 + 2p\sigma^2t + 2p^2\sigma^2t + 2p\sigma^2t^2) \end{aligned} \tag{15}$$

In Eq. (15),

$$2pt\sigma^4 - t\sigma^4 - pt\sigma^4 + qt\sigma^4 = (p+q)t\sigma^4 - t\sigma^4 = 0,$$

since $p + q = 1$; and

$$\begin{aligned} p^2q^2 + p^2\sigma^4 - 2pq\sigma^2 + q\sigma^4 + pq\sigma^4 &= (pq)^2 - 2pq\sigma^2 + \sigma^4(p^2 + pq + q) \\ &= (pq)^2 - 2pq\sigma^2 + \sigma^4 = (pq - \sigma^2)^2 \geq 0 \end{aligned}$$

and

$$\begin{aligned} r(t) &= (4pqt\sigma^2 - 2pqt\sigma^2) + (q^2t^2 + 2pq^2t + 2q\sigma^2t^2 + 2p\sigma^2t + 2p^2\sigma^2t + 2p\sigma^2t^2) \\ &= q^2t^2 + 2pq^2t + 2(p+q)t^2\sigma^2 + 2t\sigma^2(pq + p + p^2) \\ &= q^2t^2 + 2pq^2t + 2t^2\sigma^2 + 4pt\sigma^2 > 0 \end{aligned}$$

since every term in $r(t)$ is positive. Hence,

$$B^2 - 4AC = (pq - \sigma^2)^2 + r(t) > 0. \tag{16}$$

□

We now proceed to

Proof. **[Proof of Theorem 2]** For each i, since $p_i \in [0, 1]$, we have $1 + p_i(e^\lambda - 1) \in [1, e^\lambda]$. Thus, according to Eq. (11),

$$\prod_{i=1}^{n}(1 - p_i + p_i e^\lambda) \leq \left(\frac{\sum_{i=1}^{n}(1 - p_i + p_i e^\lambda)}{n} - \frac{\sigma^2(e^\lambda - 1)^2}{2e^\lambda}\right)^n = \left(q + pe^\lambda - \frac{\sigma^2(e^\lambda - 1)^2}{2e^\lambda}\right)^n. \tag{17}$$

Similar to the proof of Theorem 1, there is

$$\begin{aligned}
P(\bar{X} > p+t) &\leq \exp\left(-n\lambda(p+t)\right)\prod_{i=1}^{n}(1-p_i+p_ie^{\lambda}) \\
&\leq \exp\left(-n\lambda(p+t)\right)\left(q+pe^{\lambda}-\frac{\sigma^2(e^{\lambda}-1)^2}{2e^{\lambda}}\right)^n \qquad (18)\\
&= \exp\left(-n\left[\lambda(p+t)-\log\left(q+pe^{\lambda}-\frac{\sigma^2}{2}e^{\lambda}+\sigma^2-\frac{\sigma^2}{2}e^{-\lambda}\right)\right]\right), \qquad (19)
\end{aligned}$$

where Eq. (18) follows by applying Eq. (17).

To minimize Eq. (19), let

$$g(\lambda) = \lambda(p+t) - \log\left(q+pe^{\lambda}-\frac{\sigma^2}{2}e^{\lambda}+\sigma^2-\frac{\sigma^2}{2}e^{-\lambda}\right),$$

and we maximize $g(\lambda)$ with respect to $\lambda > 0$. For this purpose, set

$$g'(\lambda) = p+t-\frac{\left(p-\frac{\sigma^2}{2}\right)e^{\lambda}+\frac{\sigma^2}{2}e^{-\lambda}}{q+pe^{\lambda}-\frac{\sigma^2}{2}e^{\lambda}+\sigma^2-\frac{\sigma^2}{2}e^{-\lambda}} = 0,$$

which is equivalent to

$$p+t = \frac{\left(p-\frac{\sigma^2}{2}\right)e^{2\lambda}+\frac{\sigma^2}{2}}{\left(p-\frac{\sigma^2}{2}\right)e^{2\lambda}+(q+\sigma^2)e^{\lambda}-\frac{\sigma^2}{2}}$$

or

$$\left(p-\frac{\sigma^2}{2}\right)e^{2\lambda}+\frac{\sigma^2}{2} = (p+t)\left(p-\frac{\sigma^2}{2}\right)e^{2\lambda}+(p+t)(q+\sigma^2)e^{\lambda}-\frac{\sigma^2}{2}(p+t),$$

which is

$$(q-t)\left(p-\frac{\sigma^2}{2}\right)e^{2\lambda}-(p+t)(q+\sigma^2)e^{\lambda}+\frac{\sigma^2}{2}(1+p+t) = 0. \qquad (20)$$

Denote

$$A = (q-t)\left(p-\frac{\sigma^2}{2}\right), \quad B = -(p+t)(q+\sigma^2), \quad C = \frac{\sigma^2}{2}(1+p+t), \quad \text{and} \quad W = e^{\lambda},$$

then Eq. (20) is

$$AW^2+BW+C = 0. \qquad (21)$$

According to Eq. (16), $B^2-4AC > 0$. Thus, Eq. (21) has two real roots:

$$W = \frac{-B\pm\sqrt{B^2-4AC}}{2A}.$$

Since $\lambda > 0$, $W = e^{\lambda} > 1$, therefore we should take the root which is greater than 1. Let us check the two roots.

1. Since $A > 0$, $\frac{-B+\sqrt{B^2-4AC}}{2A} > 1$ is equivalent to

$$\sqrt{B^2 - 4AC} > 2A + B. \tag{22}$$

   If $2A + B < 0$, Eq. (22) is true automatically. If $2A + B > 0$, Eq. (22) is equivalent to

$$B^2 - 4AC > 4A^2 + 4AB + B^2,$$

   which is further equivalent to

$$4A(A + B + C) < 0,$$

   which is true because $A > 0$ and $A + B + C < 0$ according to Eqs. (13) and (14).
2. Similarly, $\frac{-B-\sqrt{B^2-4AC}}{2A} > 1$ is equivalent to

$$\sqrt{B^2 - 4AC} < -(2A + B). \tag{23}$$

   If $2A + B > 0$, Eq. (23) is false automatically. If $2A + B < 0$, Eq. (23) is equivalent to

$$B^2 - 4AC < 4A^2 + 4AB + B^2,$$

   which is further equivalent to

$$4A(A + B + C) > 0,$$

   which is false because $A > 0$ and $A + B + C < 0$ according to Eqs. (13) and (14).

Hence, we should take

$$w = e^{\lambda^*} = \frac{-B + \sqrt{B^2 - 4AC}}{2A}$$

as the solution to Eq. (21). According to Eq. (18), we finally have

$$\begin{aligned} P(\bar{X} > p + t) &\leq \exp\left(-n\lambda^*(p + t)\right) \left(q + pe^{\lambda^*} - \frac{\sigma^2(e^{\lambda^*} - 1)^2}{2e^{\lambda^*}}\right)^n \\ &= w^{-n(p+t)} \left(q + pw - \frac{\sigma^2(w - 1)^2}{2w}\right)^n, \end{aligned}$$

which is what to be proved. □

Remark 1. Let $p_M = \max(p_1, p_2, \cdots, p_n) \leq 1$, there is $1 + p_i(e^\lambda - 1) \leq 1 + p_M(e^\lambda - 1) \leq e^\lambda$. Hence, Eq. (17) can be improved as

$$\prod_{i=1}^{n}(1 - p_i + p_i e^\lambda) \leq \left(q + pe^\lambda - \frac{\sigma^2(e^\lambda - 1)^2}{2(1 + p_M(e^\lambda - 1))}\right)^n. \tag{24}$$

Consequently, Eq. (18) becomes

$$P(\bar{X} > p + t) \leq \exp\left(-n\lambda(p + t)\right) \left(q + pe^\lambda - \frac{\sigma^2(e^\lambda - 1)^2}{2(1 + p_M(e^\lambda - 1))}\right)^n \tag{25}$$

However, to minimize Eq. (25), we need to solve a high order equation, which is not easy. We note that Eq. (25) holds for any $\lambda > 0$, we can just choose the λ which minimizes Eq. (18), that is,

$$w = e^{\lambda^*} = \frac{-B + \sqrt{B^2 - 4AC}}{2A}.$$

With this choice, the further improved bound is

$$\begin{aligned} P(\bar{X} > p + t) &\leq \exp\left(-n\lambda^*(p+t)\right)\left(q + pe^{\lambda^*} - \frac{\sigma^2(e^{\lambda^*}-1)^2}{2(1+p_M(e^{\lambda^*}-1))}\right)^n \\ &= w^{-n(p+t)}\left(q + pw - \frac{\sigma^2(w-1)^2}{2(1+p_M(w-1))}\right)^n. \end{aligned} \tag{26}$$

Clearly, compared to Eq. (12), Eq. (26) provides a tighter bound because $p_M \leq 1$.

Remark 2. It is straightforward to verify that the original Hoeffding's inequality in Eq. (1) can be obtained from the proposed result in Eq. (12) by setting $\sigma^2 = 0$, that is, ignoring the variance of p_i's. In this sense, we see that the original Hoeffding's inequality wastes the information contained in p_i's. We also observe from Eq. (18) that larger σ^2 could lead to tighter bound.

3 Simulations and Comparisons

In this section, we compare the performance of the original Hoeffding's inequality given in Eq. (1) to that of the improved version in Eq. (12). In Theorems 1 and 2, we assume that every $X_i \in [0,1]$ almost surely, thus $p_i = \mathrm{E}(X_i) \in [0,1]$ almost surely.

We first consider that p_i's are uniformly distributed in $(0,1)$. Table 1 displays the values of U_H and U_N for various combinations of n and λ. To get an intuitive idea about the improvement of the proposed inequality to the original inequality, we also list the ratio between these two values, i.e., $R = U_H/U_N$. From Table 1, it is observed that our method provides a modest improvement when λ is small, but for large λ values such that $p+\lambda$ is close to 1, the improvement is significant. Table 1 also demonstrates that for large n, the improvement is more prominent.

Table 2 shows the comparison results when p_i's are Bernoulli(0.5) distributed. We observe the similar pattern as in Table 1, that is, our method provides significant improvement when $p+\lambda$ is close to 1. However, if we compare Table 2 to Table 1, we see that for similar $p+\lambda$ values, the improvement of our method over the original inequality differs significantly. This is because, for Bernoulli(0.5), the variance of p_i's is larger than that for uniform distribution, and according to Eq. (18), larger σ^2 is more likely to yield a tighter bound.

To further verify our observation in Table 2, we generate p_i's according to Beta(2, 10) distribution, which is right skewed and has small variance (about 0.011). In this situation, we would expect that the improvement provided by our method is marginal, and the computation results in Table 3 demonstrates this.

n	p	σ^2	λ	U_H	U_N	$R = U_H/U_N$
10	0.3762	0.0485	0.1	0.8128	0.7807	1.0411
			0.2	0.4412	0.3884	1.1359
			0.4	0.0359	0.0250	1.4357
			0.45	0.0139	0.0090	1.5326
20	0.4612	0.0843	0.1	0.6697	0.5799	1.1548
			0.2	0.1979	0.1262	1.5680
			0.4	9.1973×10^{-4}	2.5867×10^{-4}	3.5556
			0.45	1.0027×10^{-4}	2.2438×10^{-5}	4.4687
50	0.4887	0.0868	0.1	0.3663	0.2526	1.4501
			0.2	0.0167	0.0052	3.2055
			0.4	1.3950×10^{-8}	5.1546×10^{-10}	27.0629
			0.45	3.2958×10^{-11}	6.7021×10^{-13}	49.1751
100	0.4822	0.0887	0.1	0.1345	0.0626	2.1473
			0.2	2.8703×10^{-4}	2.6379×10^{-5}	10.8807
			0.4	2.6983×10^{-16}	3.2065×10^{-19}	841.5330
			0.45	1.9909×10^{-21}	7.0177×10^{-25}	2.8369×10^{3}
150	0.5259	0.0970	0.1	0.0474	0.0129	3.6827
			0.2	3.4907×10^{-6}	6.0417×10^{-8}	57.7763
			0.4	6.9625×10^{-26}	7.5374×10^{-31}	9.2373×10^{4}
			0.45	2.3836×10^{-35}	3.2747×10^{-41}	7.2790×10^{5}

Table 1. Comparison of the original Hoeffding's inequality and the proposed inequality when p_i's are uniformly distributed in $(0, 1)$.

n	p	σ^2	λ	U_H	U_N	$R = U_H/U_N$
10	0.7000	0.2100	0.1	0.7731	0.5292	1.4609
			0.2	0.3125	0.1196	2.6126
20	0.6000	0.2400	0.1	0.6492	0.3128	2.0755
			0.2	1604	0.0254	6.3156
50	0.4800	0.2496	0.1	0.3668	0.0588	6.2430
			0.2	0.0170	1.7433×10^{-4}	97.5370
			0.4	1.7310×10^{-8}	3.0123×10^{-13}	5.7466×10^{4}
			0.45	4.9139×10^{-11}	1.5775×10^{-16}	3.1151×10^{5}
100	0.4600	0.2484	0.1	0.1347	0.0034	39.6811
			0.2	3.0424×10^{-4}	3.1165×10^{-8}	9.7623×10^{3}
			0.4	6.8869×10^{-16}	2.0737×10^{-25}	3.3211×10^{9}
			0.45	1.1004×10^{-20}	1.1004×10^{-31}	9.6548×10^{10}
150	0.5	0.25	0.1	0.0488	2.0459×10^{-4}	238.4513
			0.2	4.3626×10^{-6}	4.8180×10^{-12}	9.0549×10^{5}
			0.4	1.0538×10^{-24}	5.4437×10^{-39}	1.9358×10^{14}
			0.45	5.9926×10^{-33}	1.9059×10^{-49}	3.1441×10^{16}

Table 2. Comparison of the original Hoeffding's inequality and the proposed inequality when p_i's are Bernoulli(0.5) random variables.

4 Conclusions

By applying a refined arithmetic-geometric inequality which incorporates variance information of the given numbers, this paper introduces an improved version to Hoeffd-

n	p	σ^2	λ	U_H	U_N	$R = U_H/U_N$
10	0.1509	0.0053	0.1	0.7139	0.7068	1.0099
			0.2	0.2958	0.2876	1.0285
			0.4	0.0139	0.0130	1.0737
			0.45	0.0050	0.0046	1.0860
20	0.1799	0.0060	0.1	0.5479	0.5381	1.0183
			0.2	0.1094	0.1038	1.0546
			0.4	3.5130×10^{-4}	3.0583×10^{-4}	1.1487
			0.45	5.0320×10^{-5}	4.2821×10^{-5}	1.1751
50	0.1598	0.0061	0.1	0.1973	0.1871	1.0545
			0.2	0.0027	0.0024	1.1645
			0.4	8.7132×10^{-10}	5.9035×10^{-10}	1.4759
			0.45	5.8351×10^{-12}	3.7135×10^{-12}	1.5713
100	0.1691	0.0109	0.1	0.0437	0.0365	1.1985
			0.2	1.0768×10^{-5}	6.3927×10^{-6}	1.6844
			0.4	1.9655×10^{-18}	5.1534×10^{-19}	3.8140
			0.45	1.0196×10^{-22}	2.1529×10^{-23}	4.7360
150	0.1587	0.0097	0.1	0.0075	0.0058	1.2986
			0.2	1.9330×10^{-8}	9.1753×10^{-9}	2.1067
			0.4	5.5459×10^{-28}	8.3303×10^{-29}	6.6575
			0.45	1.6210×10^{-34}	1.7972×10^{-35}	9.0196

Table 3. Comparison of the original Hoeffding's inequality and the proposed inequality when p_i's are Beta(2, 10) random variables.

ing's inequality. Simulation results show that the improvement is moderate for small λ but significant if $p + \lambda$ is close to 1. Both theoretical analysis and simulation results indicate that if the variance of p_i's are large, the improvement is prominent.

References

1. Z. D. Bai, *Methodologies in spectral analysis of large dimensional random matrices, a review*, Statistica Sinica, 9 (1999), pp. 611-677.
2. O. Bousquet, S. Boucheron, and G. Lugosi, *Introduction to Statistical Learning Theory*, In Advanced lectures on machine learning, pp. 169-207, Springer-Verlag, New York, NY, 2004.
3. D. I. Cartwright and M. J. Field, *A refinement of the arithmetic mean-geometric mean inequality*, Proc. Amer. Math. Soc., 71 (1978), pp. 36 – 38.
4. D. P. Dubhashi and A. Panconesi, *Concentration of Measure for the Analysis of Randomized Algorithms.* Cambridge University Press, 2012.
5. W. Hoeffding, *Probability Inequalities for Sums of Bounded Random Variables*, Journal of the American Statistical Association, Vol. 58, No. 301 (1963), pp. 13-30.
6. S. Mendelson, *A few notes on Statistical Learning Theory*, In Advanced lectures on machine learning, pp. 1-40, Springer-Verlag, New York, NY, 2003.
7. M. Raginsky and I. Sason, *Concentration of Measure Inequalities in Information Theory, Communications, and Coding*, Foundations and Trends in Communications and Information Theory, Vol. 10, No. 1-2 (2013), pp 1-246.