COMMUNICATIONS IN STATISTICS—THEORY AND METHODS
2017, VOL. 0, NO. 0, 1–8
https://doi.org/10.1080/03610926.2017.1367818

An improved Bennett's inequality

Songfeng Zheng

Department of Mathematics, Missouri State University, Springfield, MO, USA

ABSTRACT
This paper gives an improvement to Bennett's inequality for tail probability of sum of independent random variables, without imposing any additional condition. The improved version has a closed form expression. Using a refined arithmetic-geometric mean inequality, we further improve the obtained inequality. Numerical comparisons show that the proposed inequalities often improve the upper bound significantly in the far tail area, and these improvements get more prominent for larger sample size.

ARTICLE HISTORY
Received 19 June 2017
Accepted 10 August 2017

KEYWORDS
Arithmetic-geometric mean inequality; Bennett's inequality; Lambert *W* function; tail probability.

MATHEMATICS SUBJECT CLASSIFICATION
60E15; 62E17

1. Introduction

Bennett's inequality (Bennett 1962) provides an upper bound on the probability that the sum of independent random variables deviates from its expected value. For convenience of reference, we give the full statement of Bennett's inequality and the sketched proof below.

Theorem 1 (Bennett). *Assume* $X_1, \dots, X_n$ *are independent random variables, and* $\mathrm{E}(X_i) = 0$, $\mathrm{E}(X_i^2) = \sigma_i^2$, $|X_i| < M$ *almost surely. Then, for any* $0 \le t < nM$,

$$P\left(\sum_{i=1}^{n} X_i \ge t\right) \le \exp\left(-\frac{n\sigma^2}{M^2} h\left(\frac{tM}{n\sigma^2}\right)\right) \triangleq B_O(t) \tag{1}$$

where $h(x) = (1+x)\log(1+x) - x$ *and* $n\sigma^2 = \sum_{i=1}^{n} \sigma_i^2$.

Proof. Since $X_1, \dots, X_n$ are independent, by exponential Markov's inequality, for any $\lambda > 0$, there is

$$P\left(\sum_{i=1}^{n} X_i \ge t\right) \le \frac{\mathrm{E}\left(\exp\left(\sum_{i=1}^{n} \lambda X_i\right)\right)}{e^{\lambda t}} = \frac{\mathrm{E}\left(\prod_{i=1}^{n} \exp(\lambda X_i)\right)}{e^{\lambda t}} = \frac{\prod_{i=1}^{n} \mathrm{E}(\exp(\lambda X_i))}{e^{\lambda t}}$$

where

$$\mathrm{E}(\exp(\lambda X_i)) = \mathrm{E}\left(\sum_{k=0}^{\infty} \frac{\lambda^k X_i^k}{k!}\right) = 1 + \sum_{k=2}^{\infty} \frac{\lambda^k \mathrm{E}(X_i^2 X_i^{k-2})}{k!} \le 1 + \sum_{k=2}^{\infty} \frac{\lambda^k \sigma_i^2 M^{k-2}}{k!}$$
$$= 1 + \frac{\sigma_i^2}{M^2} \sum_{k=2}^{\infty} \frac{\lambda^k M^k}{k!} = 1 + \frac{\sigma_i^2}{M^2}(e^{\lambda M} - 1 - \lambda M) \tag{2}$$

CONTACT Songfeng Zheng SongfengZheng@MissouriState.edu Department of Mathematics, Missouri State University, Springfield MO 65897, USA.

$$\leq \exp\left(\frac{\sigma_i^2}{M^2}(e^{\lambda M} - 1 - \lambda M)\right) \tag{3}$$

Then, we have

$$P\left(\sum_{i=1}^{n} X_i \geq t\right) \leq e^{-\lambda t}\prod_{i=1}^{n} \mathrm{E}(\exp(\lambda X_i)) \tag{4}$$

$$\leq e^{-\lambda t}\prod_{i=1}^{n} \exp\left(\frac{\sigma_i^2}{M^2}(e^{\lambda M} - 1 - \lambda M)\right) \tag{5}$$

$$= e^{-\lambda t}\exp\left(\frac{n\sigma^2}{M^2}(e^{\lambda M} - 1 - \lambda M)\right) \tag{6}$$

The proof then proceeds by minimizing bound (6) with respect to all $\lambda > 0$. □

There are some efforts trying to refine the Bennett's inequality. In (Fan, Grama, and Liu 2015a), a missing factor of order $1/t$ is added to Bennett's inequality under the Bernstein's condition; in (Pinelis 2014), under the condition imposed to the third order moments, the author developed a sharp improvement to Bennett's inequality. However, they did not consider the differences among the variances of the random variables.

This paper derives an improved version to Equation (1), under the same conditions. The derived improvement has a closed form expression. Using a refined arithmetic-geometric mean inequality, we further improve the derived bound. We will compare the derived new bounds to the original Bennett's inequality by graphical plots.

2. Improved Bennett's inequality

We first present a lemma which will be used in this paper

Lemma 2. *For $a < 0$ and $b > 1$, there is a unique positive solution to the equation $e^t = at + b$, which is given by*

$$t = -W\left(-\frac{1}{a}\exp(-b/a)\right) - b/a$$

where $W(\cdot)$ is the Lambert W function (Corless et al. 1996).

Proof. Let $f(t) = e^t - at - b$, then $f(0) = 1 - b < 0$ since $b > 1$, and as $t \to \infty$, $f(t) \to \infty > 0$. The derivative $f'(t) = e^t - a > 0$ because $a < 0$. Thus, $f(t)$ is a monotonically increasing function. By the intermediate value theorem, there is a unique positive solution to the equation $f(t) = e^t - at - b = 0$.

The given equation is

$$e^t = at + b = a\left(t + \frac{b}{a}\right)$$

that is

$$\exp\left(t + \frac{b}{a}\right) = a\exp\left(\frac{b}{a}\right)\left(t + \frac{b}{a}\right)$$

or equivalently,

$$-\left[t+\frac{b}{a}\right]\exp\left(-\left[t+\frac{b}{a}\right]\right)=-\frac{1}{a}\exp\left(-\frac{b}{a}\right)$$

By the definition of Lambert W function (Corless et al. 1996)

$$-\left[t+\frac{b}{a}\right]=W\left(-\frac{1}{a}\exp\left(-\frac{b}{a}\right)\right)$$

Consequently, the solution is

$$t=-W\left(-\frac{1}{a}\exp\left(-\frac{b}{a}\right)\right)-\frac{b}{a}$$

We have to show that the solution is positive, that is

$$-W\left(-\frac{1}{a}\exp\left(-\frac{b}{a}\right)\right)>\frac{b}{a} \tag{7}$$

Since the Lambert W function $W(x)$ is increasing when $x>0$, $a<0$ and $b>1$, there is

$$W\left(-\frac{1}{a}\exp\left(-\frac{b}{a}\right)\right)<W\left(-\frac{b}{a}\exp\left(-\frac{b}{a}\right)\right)=-\frac{b}{a} \tag{8}$$

where the equality follows from the identity $W(xe^x)=x$. Clearly, Equations (8) and (7) are equivalent. □

In the proof of Bennett's inequality sketched in Sec. 1, we used the inequality $1+x\le e^x$ to get Equation (3), which is often a loose relaxation. Instead, we can proceed directly with Equation (2), that is,

$$\mathrm{E}(\exp(\lambda X_i))\le 1+\frac{\sigma_i^2}{M^2}(e^{\lambda M}-1-\lambda M) \tag{9}$$

and this will give us a tighter bound. We have the following theorem.

Theorem 3. *With the same assumptions as in Theorem 1, let*

$$A=\frac{M^2}{\sigma^2}+\frac{nM}{t}-1 \quad \text{and} \quad B=\frac{nM}{t}-1$$

and $\Lambda=A-W(Be^A)$, where $W(\cdot)$ is the Lambert W function. Then, for any $0\le t<nM$,

$$P\left(\sum_{i=1}^n X_i\ge t\right)\le\exp\left(-\Lambda t/M+n\log\left[1+\frac{\sigma^2}{M^2}(e^\Lambda-1-\Lambda)\right]\right)\triangleq B_I(t) \tag{10}$$

Proof. As in the proof of Theorem 1, we can get Equation (4), and applying Equation (9), there is

$$\begin{aligned}P\left(\sum_{i=1}^n X_i\ge t\right)&\le e^{-\lambda t}\prod_{i=1}^n \mathrm{E}(\exp(\lambda X_i))\\&\le e^{-\lambda t}\prod_{i=1}^n\left[1+\frac{\sigma_i^2}{M^2}(e^{\lambda M}-1-\lambda M)\right] &(11)\\&\le e^{-\lambda t}\left[1+\frac{\sigma^2}{M^2}(e^{\lambda M}-1-\lambda M)\right]^n &(12)\end{aligned}$$

$$= \exp\left(-\lambda t + n\log\left[1 + \frac{\sigma^2}{M^2}(e^{\lambda M} - 1 - \lambda M)\right]\right) \tag{13}$$

where Equation (12) follows from the arithmetic-geometric mean inequality which states that for positive $x_1, x_2, \ldots, x_n$,

$$\frac{1}{n}\sum_{i=1}^{n} x_i \geq \left(\prod_{i=1}^{n} x_i\right)^{1/n} \tag{14}$$

By the elementary inequality $1 + x \leq e^x$, there is

$$\left[1 + \frac{\sigma^2}{M^2}(e^{\lambda M} - 1 - \lambda M)\right]^n \leq \exp\left(\frac{n\sigma^2}{M^2}(e^{\lambda M} - 1 - \lambda M)\right)$$

Thus, comparing Equations (12) to (6), (12) provides a better bound. Consequently, we could get a tighter upper bound for the tail probability by minimizing the bound (12) or (13) with respect to $\lambda > 0$.

Letting

$$g(\lambda) = -\lambda t + n\log\left[1 + \frac{\sigma^2}{M^2}(e^{\lambda M} - 1 - \lambda M)\right]$$

Equation (13) becomes

$$P\left(\sum_{i=1}^{n} X_i \geq t\right) \leq \exp(g(\lambda))$$

To get a tight bound, we need to minimize the function $g(\lambda)$ with respect to $\lambda > 0$. For this purpose, calculating

$$g'(\lambda) = -t + \frac{n\sigma^2}{M^2}\frac{Me^{\lambda M} - M}{1 + \frac{\sigma^2}{M^2}(e^{\lambda M} - 1 - \lambda M)}$$

and setting $g'(\lambda) = 0$, there is

$$\frac{n\sigma^2}{M^2}(Me^{\lambda M} - M) = t + \frac{t\sigma^2}{M^2}(e^{\lambda M} - 1 - \lambda M)$$

or

$$\frac{\sigma^2}{M^2}(nM - t)e^{\lambda M} = -\frac{t\sigma^2}{M^2}\lambda M + \frac{\sigma^2}{M^2}(nM - t) + t \tag{15}$$

Since $t < nM$, we let

$$a = -\frac{\frac{t\sigma^2}{M^2}}{\frac{\sigma^2}{M^2}(nM - t)} = -\frac{t}{nM - t} < 0$$

and

$$b = \frac{\frac{\sigma^2}{M^2}(nM - t) + t}{\frac{\sigma^2}{M^2}(nM - t)} = 1 + \frac{M^2}{\sigma^2}\frac{t}{nM - t} > 1$$

Using the defined a and b, Equation (15) could be written as

$$e^{\lambda M} = a\lambda M + b$$

By Lemma 2, we have

$$\begin{aligned}\lambda M &= -W\left(-\frac{1}{a}\exp\left(-\frac{b}{a}\right)\right) - \frac{b}{a} \\ &= \frac{M^2}{\sigma^2} + \frac{nM}{t} - 1 - W\left(\left(\frac{nM}{t} - 1\right)\exp\left(\frac{M^2}{\sigma^2} + \frac{nM}{t} - 1\right)\right) \\ &= A - W(Be^A) \end{aligned} \tag{16}$$

where

$$A = \frac{M^2}{\sigma^2} + \frac{nM}{t} - 1 \quad \text{and} \quad B = \frac{nM}{t} - 1$$

We let $\Lambda = A - W(Be^A)$. Finally, substituting λM (i.e., Λ) to Equation (13), we obtain the upper bound for the tail probability as

$$P\left(\sum_{i=1}^{n} X_i \geq t\right) \leq \exp\left(-\Lambda t/M + n\log\left[1 + \frac{\sigma^2}{M^2}(e^{\Lambda} - 1 - \Lambda)\right]\right)$$

which is what to be proved. □

3. Further improvement

In the proof of Theorem 3, the arithmetic-geometric mean inequality in Equation (14) was used as a major ingredient. There is a refined version of the arithmetic-geometric mean inequality (Cartwright and Field 1978) which makes it possible to further improve the result in Theorem 3.

Lemma 4 (Refined Arithmetic-Geometric mean inequality Cartwright and Field 1978). *Suppose that $x_k \in [a, b]$ with $a > 0$, $p_k \geq 0$ for $k = 1, 2, \ldots, n$, and further assume that $\sum_{k=1}^{n} p_k = 1$. Let $\bar{x} = \sum_{k=1}^{n} p_k x_k$. We have*

$$\frac{1}{2b}\sum_{k=1}^{n} p_k(x_k - \bar{x})^2 \leq \bar{x} - \prod_{k=1}^{n} x_k^{p_k} \leq \frac{1}{2a}\sum_{k=1}^{n} p_k(x_k - \bar{x})^2 \tag{17}$$

The proof of Lemma 4 is given in (Cartwright and Field 1978). We are interested in a special case in which each $p_k = 1/n$, then

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

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

$$\frac{1}{2b}\text{var}(\mathbf{x}) \leq \frac{1}{n}\sum_{i=1}^{n} x_i - \left(\prod_{k=1}^{n} x_k\right)^{1/n} \leq \frac{1}{2a}\text{var}(\mathbf{x})$$

and consequently

$$\prod_{k=1}^{n} x_k \leq \left(\frac{1}{n}\sum_{i=1}^{n} x_i - \frac{1}{2b}\text{var}(\mathbf{x})\right)^n \tag{18}$$

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

By comparing Equation (14) and Equation (18), clearly, Equation (18) provides a tighter bound for the product of the given numbers by incorporating the variance information.

We obtain Equations (12) from (11) by applying the arithmetic-geometric mean inequality in Equation (14). By applying the refined arithmetic-geometric mean inequality in Equation (18), it is possible to further improve the bound.

Let $\sigma_M^2 = \max(\sigma_1^2, \ldots, \sigma_n^2)$ and $V = \sum_{i=1}^n (\sigma_i^2 - \sigma^2)^2/n$. We have, for any $1 \le i \le n$,

$$1 + \frac{\sigma_i^2}{M^2}(e^{\lambda M} - 1 - \lambda M) \le 1 + \frac{\sigma_M^2}{M^2}(e^{\lambda M} - 1 - \lambda M) \triangleq M_\lambda$$

Applying Equation (18), there is

$$\prod_{i=1}^n \left[1 + \frac{\sigma_i^2}{M^2}(e^{\lambda M} - 1 - \lambda M)\right] \le \left[1 + \frac{\sigma^2}{M^2}(e^{\lambda M} - 1 - \lambda M) - \frac{(e^{\lambda M} - 1 - \lambda M)^2 V}{2M^4 M_\lambda}\right]^n \quad (19)$$

Substituting Equation (19) into Equation (11) in the proof of Theorem 3, for any $\lambda > 0$, there is

$$\begin{aligned} P\left(\sum_{i=1}^n X_i \ge t\right) &\le e^{-\lambda t} \prod_{i=1}^n \left[1 + \frac{\sigma_i^2}{M^2}(e^{\lambda M} - 1 - \lambda M)\right] \\ &\le e^{-\lambda t}\left[1 + \frac{\sigma^2}{M^2}(e^{\lambda M} - 1 - \lambda M) - \frac{(e^{\lambda M} - 1 - \lambda M)^2 V}{2M^4 M_\lambda}\right]^n \end{aligned} \quad (20)$$

Ideally, to find the best upper bound, we would like to choose a value for λ to minimize the right hand side of Equation (20), which would yield a high order equation that is not easy to solve. We note that Equation (20) holds for any $\lambda > 0$, thus, we can just choose the λ which minimizes Equation (12), that is, the λ which is defined in Equation (16). We summarize our discussion as in

Theorem 5. *With the same assumptions and notations as in Theorems 1 and 3, let* $\sigma_M^2 = \max(\sigma_1^2, \ldots, \sigma_n^2)$ *and* $V = \sum_{i=1}^n (\sigma_i^2 - \sigma^2)^2/n$, *and*

$$M_\Lambda = 1 + \frac{\sigma_M^2}{M^2}(e^\Lambda - 1 - \Lambda)$$

For any $0 \le t < nM$, *there is*

$$P\left(\sum_{i=1}^n X_i \ge t\right) \le e^{-\Lambda t/M}\left[1 + \frac{\sigma^2}{M^2}(e^\Lambda - 1 - \Lambda) - \frac{(e^\Lambda - 1 - \Lambda)^2 V}{2M^4 M_\Lambda}\right]^n \triangleq B_F(t) \quad (21)$$

It is straightforward to verify that the improved Bennett's inequality in Equation (10) can be obtained from Equation (21) by setting $V = 0$, that is, ignoring the variance of σ_i^2's. In this sense, we see that Equation (10) wastes part of the information contained in σ_i^2's. We also observe from Equation (21) that larger V could lead to tighter bound.

4. Comparisons

In this section, we numerically compare the original Bennett's inequality $B_O(t)$ given in Equation (1) and its enhanced versions $B_I(t)$ and $B_F(t)$ given in Equations (10) and (21), respectively. Without loss of generality, we assume $M = 1$, and let $t = pnM$ with $0 < p < 1$. We assume that the variance $\sigma_i^2 = i/n$, for $i = 1, 2, \ldots, n$.

For $n = 30$, and for p equally spaced between 0.1 and 0.8, we calculate the ratio $B_O(t)/B_I(t)$ and $B_O(t)/B_F(t)$. We show the curves of the ratios in Figure 1a, and Figure 1b demonstrates

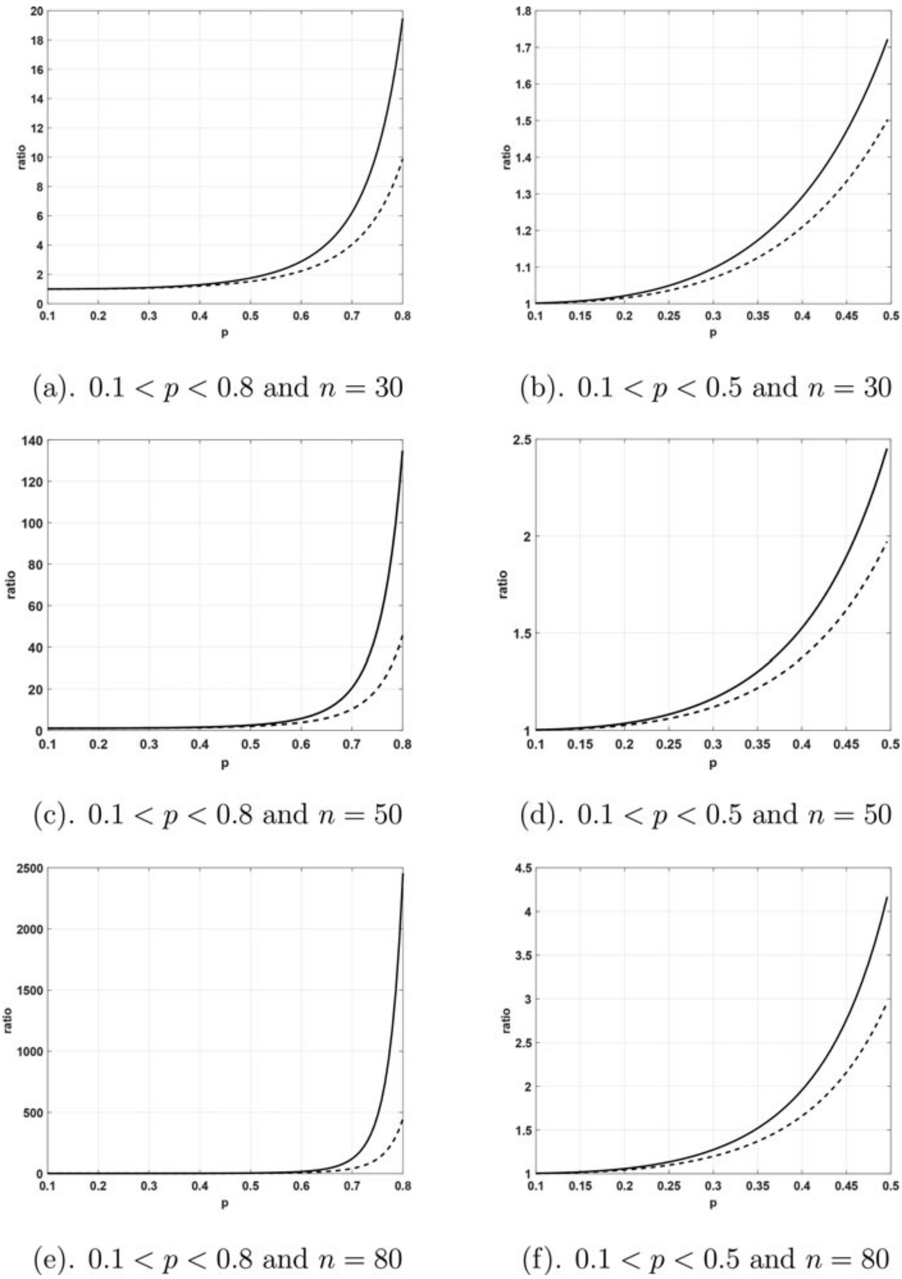

Figure 1. The ratio between the original Bennett bound $B_O(t)$ and the improved versions $B_I(t)$ and $B_F(t)$. Solid curve is the ratio $B_O(t)/B_F(t)$, while the dashed curve is $B_O(t)/B_I(t)$.

the ratios when $0.1 < p < 0.5$ for more details. We observe that at the far tail area (p large), the bound in Equation (21) improves the original Bennett bound by 3–20 times, while the bound in Equation (10) improves 2–10 times; for small p (that is, in the near tail area), the improvements are minor.

We then choose $n = 50$ and $n = 80$, and the corresponding curves are given in Figure 1c–f. We observe the same trend as in Figure 1a and 1b, that is, Equation (21) gives better bound than Equation (10), which is expected because Equation (21) incorporates more information. For both improved versions, the improvement is more significant at the far tail area (i.e., for

p large). Comparing the curves for different n, we see that for large n, the improvement is more prominent.

5. Conclusion and future work

In the proof of Bennett's inequality, we employ a tighter upper bound for the moment generating function, and this enables us to obtain a new bound for tail probability for sum of independent random variables, without imposing additional condition. This paper further improves the obtained upper bound by using a refined arithmetic-geometric mean inequality, which incorporates more variance information of the random variables. We compare the original Bennett's upper bound and the improved versions by plotting the curves of the ratios between different upper bounds, and the results show that the proposed upper bounds often improve the original Bennett's bound significantly in the far tail area, and this improvement becomes more significant for large sample size.

Under the following Bernstein's condition: for a positive constant ϵ,

$$\mathrm{E}\left(X_i^k\right) \le \frac{1}{2} k! \epsilon^{k-2} \mathrm{E}\left(X_i^k\right) \qquad \text{for all } k \ge 2 \text{ and all } i \in [1, n]$$

it was proved in (Fan, Grama, and Liu 2015b) that for all $x > 0$,

$$P\left(\sum_{i=1}^{n} X_i \ge x\right) \le \exp\left\{-\bar{\lambda} x + n \log\left(1 + \frac{\bar{\lambda}^2 \sigma^2}{2(1-\bar{\lambda}\epsilon)}\right)\right\} \tag{22}$$

where

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^{n} \mathrm{E}\left(X_i^2\right) \qquad \text{and } \bar{\lambda} = \frac{2x/v^2}{2x\epsilon/v^2 + 1 + \sqrt{1 + 2x\epsilon/v^2}}$$

Equation (22) can be regarded as a counterpart of Theorem 3 under the Bernstein's condition. Similar to Equation (22), it would be interesting to develop a counterpart for Theorem 5 under the Bernstein's condition and this is our next step of work.

Acknowledgments

The author would like to extend his sincere gratitude to the anonymous reviewer for the suggestions, which helped improve the quality of this paper. This work was partially supported by a Summer Faculty Research Fellowship from Missouri State University.

References

Bennett, G. 1962. Probability inequalities for the sum of independent random variables. *Journal of the American Statistical Association* 57 (297):33–45.

Cartwright, D. I., and M. J. Field. 1978. A refinement of the arithmetic mean-geometric mean inequality. *Proceedings of the American Mathematical Society* 71:36–38.

Corless, R. M., G. H. Gonnet, D. E. G. Hare, D. J. Jeffery, and D. E. Knuth. 1996. On the Lambert W Function. *Advances in Computational Mathematics* 5:329–59.

Fan, X. Q., I. Grama, and Q. S. Liu. 2015a. Sharp large deviation results for sums of independent random variables. *Science China Mathematics* 58 (9):1939–58.

Fan, X. Q., I. Grama, and Q. S. Liu. 2015b. Exponential inequalities for martingales with applications. *Electronic Journal of Probability* 20 (1):1–22.

Pinelis, I. 2014. On the Bennett-Hoeffding inequality. *Annales de l'Institut Henri Poincaré Probability and Statistics* 50 (1):15–27.