

Improved Hoeffding's Lemma and Hoeffding's Tail Bounds

David Hertz

Correspondence: Independent researcher. E-mail: hertz@013.net.il

Received: July 11, 2021 Accepted: August 11, 2021 Online Published: August 15, 2021

doi:10.5539/ijsp.v10n5p27 URL: <https://doi.org/10.5539/ijsp.v10n5p27>

Abstract

The purpose of this article is to improve Hoeffding's lemma and consequently Hoeffding's tail bounds. The improvement pertains to left skewed zero mean random variables $X \in [a, b]$, where $a < 0$ and $-a > b$. The proof of Hoeffding's improved lemma uses Taylor's expansion, the convexity of $\exp(sx)$, $s \in \mathbb{R}$, and an unnoticed observation since Hoeffding's publication in 1963 that for $-a > b$ the maximum of the intermediate function $\tau(1 - \tau)$ appearing in Hoeffding's proof is attained at an endpoint rather than at $\tau = 0.5$ as in the case $b > -a$. Using Hoeffding's improved lemma we obtain one sided and two sided tail bounds for $\mathbb{P}(S_n \geq t)$ and $\mathbb{P}(|S_n| \geq t)$, respectively, where $S_n = \sum_{i=1}^n X_i$ and the $X_i \in [a_i, b_i]$, $i = 1, \dots, n$ are independent zero mean random variables (not necessarily identically distributed). It is interesting to note that we could also improve Hoeffding's two sided bound for all $\{X_i : -a_i \neq b_i, i = 1, \dots, n\}$. This is so because here the one sided bound should be increased by $\mathbb{P}(-S_n \geq t)$, wherein the left skewed intervals become right skewed and vice versa.

Keywords: Hoeffding's lemma, Hoeffding's tail bounds, Chernoff's bound

1. Introduction

By googling Hoeffding bounds we obtain many applications and theoretical uses of Hoeffding's bounds to signal processing (e.g., to time series analysis, compressed sensing, sensor networks, financial signal processing, to name but a few) to machine learning, to information theory, communication, and coding (Raginsky and Sason, 2018), to randomized algorithms (Mitzenmacher and Upfal, 2017) to name but a few. Communication networks are treated in (Mitzenmacher and Upfal, 2017) where the authors rederive Hoeffding's results and present applications to packet routing in sparse networks. Our starting point is to present Hoeffding's lemma with a proof somewhat different than the original one and then present the improved Hoeffding's lemma and prove it. Most of the proof is not new and presented here for the sake of completeness. When the underlying distribution is skewed to the left we show how the one sided bound in Hoeffding's Lemma can be improved. The two sided Hoeffding's bound though can always be improved for skewed distributions either to the left or to the right.

We assume that X is a zero mean real valued random variable such that $X \in [a, b]$, $a < 0$, $b > 0$. Then, Hoeffding's lemma (Hoeffding, 1963) states that for all $s \in \mathbb{R}$, $s > 0$,

$$\mathbb{E}[e^{sX}] \leq \exp\left(\frac{s^2(b-a)^2}{8}\right). \quad (1)$$

Let A and G denote the arithmetic and geometric means of $|a|$ and b , respectively. Noting that $b - a = b + |a|$ we have

$$A := \frac{b + |a|}{2} \quad (2)$$

and

$$G := \sqrt{|a|b}. \quad (3)$$

Hence,

$$\exp\left(\frac{s^2(b-a)^2}{8}\right) \equiv \exp\left(\frac{s^2 A^2}{2}\right). \quad (4)$$

The improvement in Hoeffding's bound pertains to the case $|a| > b$. We have

$$\mathbb{E}[e^{sX}] \leq \begin{cases} \exp\left(\frac{s^2 A^2}{2}\right), & b \geq |a|, \\ \exp\left(\frac{s^2 G^2}{2}\right), & b \leq |a|, \end{cases} \quad (5)$$

where since $A \geq G$ gives for $-a > b$ a tighter bound than for $-a < b$. The proof will be given in the next Section. The organization of the remaining Sections is as follows. in Section 2 we present the proof of Hoeffding’s improved lemma. In Section 3 we present Hoeffding’s improved one sided tail bound and its proof. In Section 4 we present Hoeffding’s improved two sided tail bound and its proof. Finally, in Section 5 we give the conclusion.

2. Proof of Hoeffding’s Improved Lemma

Since e^{sx} is a convex function of x and $s \in \mathbb{R}, s > 0$ is a parameter we obtain

$$e^{sx} \leq \frac{b-x}{b-a} e^{sa} + \frac{x-a}{b-a} e^{sb} \tag{6}$$

Let

$$\lambda := \frac{-a}{b-a} \tag{7}$$

and

$$u := s(b-a) > 0. \tag{8}$$

Hence, since $\mathbb{E}X = 0$, using the convexity of $\exp(sx)$ and some algebra we obtain

$$\begin{aligned} \mathbb{E}e^{sX} &\leq \frac{b-\mathbb{E}X}{b-a} e^{sa} + \frac{\mathbb{E}X-a}{b-a} e^{sb} \\ &= (1-\lambda)e^{sa} + \lambda e^{sb} \\ &= (1-\lambda + \lambda e^u) e^{-\lambda u} \\ &= e^{\psi(u)}, \end{aligned} \tag{9}$$

where

$$\psi(u) := -\lambda u + \ln(1-\lambda + \lambda e^u). \tag{10}$$

Since $u > 0$ and $\psi(u)$ is well defined, by Taylor’s expansion we obtain

$$\psi(u) = \psi(0) + \psi'(0)u + 0.5\psi''(\mu)u^2, \text{ for some } \mu \in [0, u]. \tag{11}$$

We have

$$\psi'(u) = -\lambda + \frac{\lambda e^u}{1-\lambda + \lambda e^u} \tag{12}$$

and

$$\psi''(\mu) = \frac{\lambda e^\mu}{1-\lambda + \lambda e^\mu} \left(1 - \frac{\lambda e^\mu}{1-\lambda + \lambda e^\mu} \right). \tag{13}$$

Let

$$\tau(\mu) = \frac{\lambda}{(1-\lambda)e^{-\mu} + \lambda}, \mu \in [0, u]. \tag{14}$$

Since $\psi(0) = \psi'(0) = 0$ we obtain

$$\psi(u) = 0.5\tau(1-\tau)u^2. \tag{15}$$

Using (14) and the fact $u > 0$ we obtain that $\tau \in [\lambda, 1]$, where the endpoints λ and 1 correspond to $u = 0$ and $u = \infty$, respectively. Now, for $-a \leq b$ since $\lambda \leq 0.5$ the maximum of $\tau(\tau-1)$ is attained at $\tau = 0.5$ and is 0.25 giving Hoeffding’s original lemma. However, for $-a > b$, $\lambda > 0.5$ the maximum is attained at $\tau = \lambda$. Using (8) we obtain after some algebra

$$\psi(u) \leq 0.5\lambda(1-\lambda)u^2 = 0.5s^2G^2. \tag{16}$$

Hence, using (9) we obtain

$$\mathbb{E}[e^{sX}] \leq \exp\left(\frac{s^2G^2}{2}\right), -a \geq b. \tag{17}$$

This completes the proof ■

3. Hoeffding’s Improved One Sided Tail Bound

Let X_1, \dots, X_n be independent random variables such that $X_i \in [a_i, b_i], a_i < 0, b_i > 0$ and $\mathbb{E}X_i = 0$ for $i = 1, \dots, n$. Let $S_n := \sum_{i=1}^n X_i$ then $\mathbb{E}S_n = 0$. For all $s > 0$ we have

$$\begin{aligned} \mathbb{P}(S_n \geq t) &= \mathbb{P}\left(e^{sS_n} \geq e^{st}\right) && \text{Chernoff} \\ &\leq e^{-st} \mathbb{E}e^{sS_n} && \text{Markov's inequality} \\ &= e^{-st} \prod_{i=1}^n \mathbb{E}e^{sX_i} && X_i\text{'s are independent.} \end{aligned} \tag{18}$$

Now define

$$I := \{i : b_i \geq -a_i, i = 1 : n\} \tag{19}$$

and

$$J := \{i : b_i < -a_i, i = 1 : n\}. \tag{20}$$

Hence, using (18), we obtain

$$\begin{aligned} \mathbb{P}(S_n \geq t) &\leq e^{-st} \prod_{i=1}^n \mathbb{E}e^{sX_i} \\ &= \exp\left(-st + 0.5s^2\left(\sum_{i \in I} A_i^2 + \sum_{j \in J} G_j^2\right)\right) \\ &= \exp(-st + 0.5s^2 M_n^2), \end{aligned} \tag{21}$$

where the mixed sum M_n^2 is defined by

$$M_n^2 := \sum_{i \in I} A_i^2 + \sum_{j \in J} G_j^2. \tag{22}$$

Finally, minimizing the exponent in the last inequality we obtain

$$\mathbb{P}(S_n \geq t) \leq \exp(-0.5(t/M_n)^2). \tag{23}$$

Now let

$$\bar{M}_n^2 := \frac{M_n^2}{n}, \tag{24}$$

$$\bar{A}_n^2 := \frac{\sum_{i=1}^n A_i^2}{n}, \tag{25}$$

and

$$\bar{G}_n^2 := \frac{\sum_{i=1}^n G_i^2}{n}. \tag{26}$$

Since $G_i \leq A_i, i = 1, \dots, n$ we obtain

$$\bar{G}_n \leq \bar{M}_n \leq \bar{A}_n. \tag{27}$$

Consequently, we obtain the new bound

$$\begin{aligned} \mathbb{P}(S_n \geq t) &\leq \exp(-0.5(t/\bar{M}_n)^2 n) && \text{Improved bound} \\ &\leq \exp(-0.5(t/\bar{A}_n)^2 n) && \text{Original bound.} \end{aligned} \tag{28}$$

If we define $k := t/\bar{M}_n$ then

$$\begin{aligned} \mathbb{P}(S_n/\bar{M}_n \geq k) &= \mathbb{P}\left(\sum_{i=1}^n (X_i/\bar{M}_n) \geq k\right) \\ &\leq \exp(-0.5k^2 n). \end{aligned} \tag{29}$$

Hence, \bar{M}_n is a natural unit for $X_i, i = 1, \dots, n$ and S_n that we denote by $[X_i] = [S_n] = \bar{M}_n$. Using this unit we obtain

$$\mathbb{P}(S_n \geq k) \leq \exp(-0.5k^2 n). \tag{30}$$

Further, note that if $X_i \in [a, b], i = 1, \dots, n$ are zero mean independent random variables (not necessarily identically distributed) then using (2) and (3) we obtain

$$\mathbb{P}(S_n \geq t) \leq \exp\left(-0.5(t/A)^2 n\right), \text{ if } -a \leq b \tag{31}$$

and

$$\mathbb{P}(S_n \geq t) \leq \exp\left(-0.5(t/G)^2 n\right), \text{ if } -a > b. \tag{32}$$

4. Hoeffding’s Improved Two Sided Bound

We will prove that

$$\mathbb{P}(|S_n| \geq t) \leq \exp(-0.5(t/\bar{M}_n)^2n) + \exp(-0.5(t/\bar{N}_n)^2n), \tag{33}$$

where \bar{M}_n^2 is as defined in (24) and \bar{N}_n^2 is a sort of its complement. I.e.,

$$\bar{N}_n^2 := N_n^2/n, \tag{34}$$

where

$$N_n^2 := \sum_{i \in I} A_i^2 + \sum_{j \in J} G_j^2, \tag{35}$$

and I, J are as defined in (19) and (20), respectively.

Proof of (33).

Since $S_n = \sum_{i=1}^n X_i, X_i \in [a_i, b_i]$ it follows that $-S_n = \sum_{i=1}^n (-X_i), (-X_i) \in [-b_i, -a_i], -b_i < 0, -a_i > 0$. Hence, in this case we should use \bar{N}_n instead of \bar{M}_n and we obtain

$$\begin{aligned} \mathbb{P}(|S_n| \geq t) &= \mathbb{P}(S_n \geq t \text{ or } -S_n \leq -t) \\ &= \mathbb{P}(S_n \geq t) + \mathbb{P}(-S_n \leq -t) \\ &\leq \exp(-0.5(t/\bar{M}_n)^2n) + \exp(-0.5(t/\bar{N}_n)^2n). \end{aligned}$$

This completes the proof ■

Note that unless $\bar{M}_n = \bar{N}_n$ we can not define k as in the one sided case. It is easily seen and interesting to note that if $[a_i, b_i] = [a, b], i = 1, \dots, n$, we obtain

$$\mathbb{P}(|S_n| \geq t) \leq \exp(-0.5(t/G)^2n) + \exp(-0.5(t/A)^2n). \tag{36}$$

Hence, unless $A = G$ we thus improved Hoeffding’s two sided tail bound.

5. Conclusion

In this article, we presented Hoeffding’s improved lemma and Hoeffding’s improved one and two sided tail bounds for bounded random variables. The improvement pertains only to intervals $[a, b]$, where $-a > b$ and are associated with the geometric mean of $-a$ and b . For $-a < b$ Hoeffding’s lemma and Hoeffding one sided tail bound remain intact and are associated only with the average of $-a$ and b . It is interesting to note that we could also improve Hoeffding’s two sided bound for all $\{X_i : -a_i \neq b_i, i = 1, \dots, n\}$. This is so because here the one sided bound should be increased by the bound for $\mathbb{P}(-S_n \geq t)$, wherein left skewed intervals become right skewed and vice versa. Perhaps, further research will focus on trying to improve other inequalities that use Hoeffding’s results.

References

Hoeffding, W. (1994). *Probability inequalities for sums of bounded random variables*. In The collected works of Wassily Hoeffding (pp. 409-426). Springer, New York, NY.

Mitzenmacher, M., & Upfal, E. (2017). *Probability and computing: Randomization and probabilistic techniques in algorithms and data analysis*. Cambridge university press.

Raginsky, M., & Sason, I. (2012). Concentration of measure inequalities in information theory, communications and coding. arXiv preprint arXiv:1212.4663. <https://doi.org/10.1561/9781601987259>

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).