Hoeffding's Inequality

John A. Gubner
Department of Electrical and Computer Engineering
University of Wisconsin–Madison

1. Preliminaries

1.1. Variance Bound

If X is a bounded random variable with $a \le X \le b$, where $-\infty < a \le b < \infty$, we derive the well-known result

$$\mathsf{var}(X) \le \frac{(b-a)^2}{4}$$

using the approach of [2]. Recall that

$$\mathsf{var}(X) = \min_{c\in\mathbb{R}} \mathsf{E}[(X-c)^2]$$

is achieved when $c = \mathsf{E}[X]$. So in particular, if we take c as the midpoint between a and b; i.e., $(a+b)/2$, we have

$$\mathsf{var}(X) \le \mathsf{E}\biggl[\biggl(X - \frac{a+b}{2}\biggr)^2\biggr].$$

The maximum possible distance between X and the midpoint occurs when X is at either end of the interval $[a,b]$. Hence,

$$\mathsf{var}(X) \le \mathsf{E}\biggl[\biggl(X - \frac{a+b}{2}\biggr)^2\biggr] \le \biggl(b - \frac{a+b}{2}\biggr)^2 = \biggl(\frac{b-a}{2}\biggr)^2 = \frac{(b-a)^2}{4}.$$

1.2. Moment Generating Functions

For real s, let $M(s) := \mathsf{E}[e^{sX}]$ denote the moment generating function of X. Recall that $M'(s) = \mathsf{E}[Xe^{sX}]$ and $M''(s) = \mathsf{E}[X^2e^{sX}]$ [1, pp. 278–279]. Clearly, $M(0) = 1$, and $M'(0) = \mathsf{E}[X]$. The cumulant generating function is $\psi(s) := \ln M(s)$, and it satisfies

$$\psi'(s) = \frac{M'(s)}{M(s)} \quad \text{and} \quad \psi''(s) = \frac{M(s)M''(s) - M'(s)^2}{M(s)^2} = \frac{M''(s)}{M(s)} - \biggl[\frac{M'(s)}{M(s)}\biggr]^2.$$

Note that $\psi(0) = 0$. Using the formulas for $M'(s)$ and $M''(s)$, we can write

$$\psi''(s) = \mathsf{E}\biggl[X^2\frac{e^{sX}}{M(s)}\biggr] - \biggl(\mathsf{E}\biggl[X\frac{e^{sX}}{M(s)}\biggr]\biggr)^2. \tag{1}$$

As in [3] we define a family of probability measures on subsets A of the sample space by

$$\mathsf{P}_s(A) := \mathsf{E}[\mathbf{1}_A L_s(X)],$$

where $L_s(X) := e^{sX}/M(s)$.[a] In other words, the expectation of any random variable Z under P_s is computed via

$$\mathsf{E}_s[Z] = \mathsf{E}[Z L_s(X)].$$

Then (1) says that

$$\psi''(s) = \mathsf{E}_s[X^2] - (\mathsf{E}_s[X])^2 = \mathsf{var}_s(X) \leq (b-a)^2/4. \tag{2}$$

2. Hoeffding's Lemma

Let X be a bounded random variable with $a \leq X \leq b$, where $-\infty < a \leq b < \infty$, and put $\mu := \mathsf{E}[X]$. Then Hoeffding's Lemma says that

$$\boxed{\mathsf{E}[e^{s(X-\mu)}] \leq e^{s^2(b-a)^2/8}.}$$

To derive it, rewrite it as $\mathsf{E}[e^{sX}] \leq e^{\mu s} e^{s^2(b-a)^2/8}$ and take the logarithm to get

$$\psi(s) \leq \mu s + s^2(b-a)^2/8. \tag{3}$$

By Taylor's Theorem with remainder [4, Th. 5.15],

$$\psi(s) = \psi(0) + \psi'(0)s + \psi''(\hat{s})s^2/2,$$

where $\hat{s}$ lies between 0 and s. By (2), we have $\psi''(\hat{s}) \leq (b-a)^2/4$, and since $\psi(0) = 0$ and $\psi'(0) = \mu$, the inequality (3) follows.

3. Hoeffding's Inequalities

Let $X_1, \ldots, X_n$ be independent bounded random variables with $a_i \leq X_i \leq b_i$. Put

$$S_n := \sum_{i=1}^n X_i, \quad s_n := \sum_{i=1}^n \mathsf{E}[X_i], \quad \text{and} \quad \Delta_n := \sum_{i=1}^n (b_i - a_i)^2.$$

Then Hoeffding's inequalities are

$$\boxed{\mathsf{P}(S_n - s_n > t) \leq e^{-2t^2/\Delta_n} \quad \text{and} \quad \mathsf{P}(|S_n - s_n| > t) \leq 2e^{-2t^2/\Delta_n}.} \tag{4}$$

[a] The measure P_s is said to be exponentially tilted [5].

If the X_i are identically distributed with common bounds $a \le X_i \le b$ and common means μ, then Hoeffding's inequalities imply the simplified relations

$$\boxed{\mathsf{P}\left(\frac{S_n}{n} - \mu > t\right) \le e^{-2nt^2/(b-a)^2} \quad \text{and} \quad \mathsf{P}\left(\left|\frac{S_n}{n} - \mu\right| > t\right) \le 2e^{-2nt^2/(b-a)^2}.}$$

To derive the left-hand inequality in (4), observe that for any $\theta > 0$, we can write

$$\begin{aligned}
\mathsf{P}(S_n - s_n > t) &= \mathsf{P}(\theta(S_n - s_n) > \theta t) \\
&= \mathsf{P}(e^{\theta(S_n - s_n)} > e^{\theta t}) \\
&\le \frac{\mathsf{E}[e^{\theta(S_n - s_n)}]}{e^{\theta t}}, && \text{by Markov's inequality,} \\
&= e^{-\theta t} \prod_{i=1}^{n} \mathsf{E}[e^{\theta(X_i - \mathsf{E}[X_i])}], && \text{by independence,} \\
&\le e^{-\theta t} \prod_{i=1}^{n} e^{\theta^2 (b_i - a_i)^2/8}, && \text{by Hoeffding's Lemma,} \\
&= e^{-\theta t} e^{\theta^2 \Delta_n/8} \\
&= \exp[\theta^2 \Delta_n/8 - \theta t].
\end{aligned}$$

To minimize the bound as a function of θ, we minimize the argument of exp, which is a quadratic. Setting the derivative of the quadratic to zero shows that the optimal value of θ is $4t/\Delta_n$. Substituting this into the above inequality yields the desired result.

A similar analysis shows that $\mathsf{P}(S_n - s_n < -t) \le e^{-2t^2/\Delta_n}$, and then

$$\begin{aligned}
\mathsf{P}(|S_n - s_n| > t) &= \mathsf{P}(\{S_n - s_n > t\} \cup \{S_n - s_n < -t\}) \\
&\le \mathsf{P}(S_n - s_n > t) + \mathsf{P}(S_n - s_n < -t) \\
&\le 2e^{-2t^2/\Delta_n}.
\end{aligned}$$

References

[1] P. Billingsley, *Probability and Measure*, 3rd ed. New York: Wiley, 1995.

[2] JBD (`https://stats.stackexchange.com/users/339110/jbd`), "Understanding proof of a lemma used in Hoeffding inequality," [Online]. Available: `https://stats.stackexchange.com/q/550105`, accessed Nov. 23, 2023.

[3] pSrIoGcNeAsLs (`https://stats.stackexchange.com/users/302748/psriogcneasls`), "Understanding proof of a lemma used in Hoeffding inequality," [Online]. Available: `https://stats.stackexchange.com/q/585079`, accessed Nov. 23, 2023.

[4] W. Rudin, *Principles of Mathematical Analysis*, 3rd ed. New York: McGraw-Hill, 1976.

[5] Wikipedia Contributors, "Exponential tilting — Wikipedia, The Free Encyclopedia," [Online]. Available: `https://en.wikipedia.org/w/index.php?title=Exponential_tilting&oldid=1175268151`, accessed Nov. 24, 2023.