Hoeffding's Inequality

Lu Luyu

This note presents a self-contained proof of Hoeffding's inequality. We begin by establishing several foundational results that are instrumental to the main theorem.

Theorem 1 (Markov's inequality)**.** *Let X be a non-negative random variable and $t > 0$. Then*

$$\mathbb{P}[X \geq t] \leq \frac{\mathbb{E}(X)}{t}.$$

Proof. We denote the indicator function $\mathbb{1}_A(x) = 1$ if $x \in A$ and 0 otherwise. Then we have

$$t\mathbb{1}_{\{\omega|\omega \geq t\}}(x) \leq x.$$

Taking the expected value of both sides, the inequality follows. □

Corollary 2 (Chebyshev's inequality)**.** *Let X be a random variable with mean $\mathbb{E}[X] = \mu$ and variance $\mathbb{E}[X^2] = \sigma^2$. Then for all $t > 0$,*

$$\mathbb{P}[|X - \mu| \geq t] \leq \frac{\sigma^2}{t^2}.$$

Proof. This proof follows directly from Markov's inequality by considering the non-negative random variable $(X - \mu)^2$,

$$\mathbb{P}[|X - \mu| \geq t] = \mathbb{P}[|X - \mu|^2 \geq t^2] \leq \frac{\sigma^2}{t^2}.$$

□

Theorem 3 (Jensen's inequality)**.** *A function g is convex if for all x, y and all $\alpha \in [0, 1]$,*

$$g(\alpha x + (1 - \alpha)y) \leq \alpha g(x) + (1 - \alpha)g(y).$$

A function g is concave if $-g$ is convex. If g is convex, then

$$g(\mathbb{E}[X]) \leq \mathbb{E}[g(X)].$$

If g is concave, then

$$g(\mathbb{E}[X]) \geq \mathbb{E}[g(X)].$$

As a special case, the exponential function $g(x) = e^{sx}$ is convex.

Proof. Supoose g is convex, and let $L(x) = a + bx$ be a line, tangent to $g(x)$ at the point $E(X)$. Since g is convex, it lies above the line $L(x)$, so

$$\mathbb{E}(g(x)) \geq \mathbb{E}(L(x)) = a + b\mathbb{E}(x) = g(E(x)).$$

If g is concave, then $-g$ is convex, and the inequality reverses. The exponential function g(x)=e^{sx} is convex since its second derivative is positive. □

Lemma 4 (Hoeffding's Lemma)**.** *Let Y be a random variable satisfying $\mathbb{E}[Y] = 0$ and $a \leq Y \leq b$. Then for all $s \in \mathbb{R}$,*

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

Proof. For any $Y \in [a, b]$, we can rewrite Y as a convex combination of a and b, $Y = \alpha a + (1 - \alpha)b$, where $\alpha = (b - Y)/(b - a) \in [0, 1]$. Since e^x is convex, we have

$$e^{sY} \leq \frac{b - Y}{b - a}e^{sa} + \frac{Y - a}{b - a}e^{sb}.$$

Taking the expectation of both sides and using $\mathbb{E}(Y) = 0$, we get

$$\mathbb{E}(e^{sY}) \le \frac{b}{b-a} e^{sa} + \frac{-a}{b-a} e^{sb} = e^{g(u)}.$$

where $u = s(b-a)$, $g(u) = -\beta u + \log\{1 - \beta + \beta e^u\}$ and $\beta = -a/(b-a) \in (0,1)$.Since $g(0) = g'(0) = 0$, and the second derivative satisfies

$$g''(u) = \frac{\beta(1-\beta)e^u}{\{1-\beta+\beta e^u\}^2} = \left\{ \sqrt{\frac{1-\beta}{\beta e^u}} + \sqrt{\frac{\beta e^u}{1-\beta}} \right\}^{-2} \le \frac{1}{4},$$

an application of Taylor's theorem shows that for some $\xi \in (0, u)$,

$$g(u) = g(0) + g'(0)u + \frac{1}{2} g''(\xi) u^2 \le \frac{1}{8} u^2 = \frac{1}{8} s^2 (b-a)^2.$$

Hence,

$$\mathbb{E}(e^{sY}) \le e^{g(u)} \le \exp\left\{ \frac{1}{8} s^2 (b-a)^2 \right\},$$

which completes the proof. □

Theorem 5 (Hoeffding's inequality)**.** *Let $X_1, \cdots, X_n$ be independent random variables with $a_i \le X_i \le b_i$ for all i. Let $S_n = \sum_{i=1}^n X_i$. Then for all $t > 0$, it holds that*

$$\mathbb{P}[S_n - \mathbb{E}(S_n) \ge t] \le \exp\left\{ -\frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2} \right\},$$

and

$$\mathbb{P}\left[|S_n - \mathbb{E}(S_n)| \ge t \right] \le 2 \exp\left\{ -\frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2} \right\}.$$

Proof. We employ the Chernoff bound method. For any $s > 0$, applying Markov's inequality and leveraging the independence of $\{X_i\}$ yields

$$\begin{aligned} \mathbb{P}[S_n - \mathbb{E}(S_n) \ge t] &= \mathbb{P}\left[e^{s(S_n - \mathbb{E}(S_n))} \ge e^{st} \right] \\ &\le e^{-st} \mathbb{E}\left[e^{s(S_n - \mathbb{E}(S_n))} \right] \\ &= e^{-st} \prod_{i=1}^n \mathbb{E}\left[e^{s(X_i - \mathbb{E}(X_i))} \right]. \end{aligned}$$

By Hoeffding's Lemma, each term in the product can be bounded,

$$\mathbb{E}\left[e^{s(X_i - \mathbb{E}(X_i))} \right] \le \exp\left\{ \frac{1}{8} s^2 (b_i - a_i)^2 \right\}.$$

Substituting this into the main inequality gives

$$\mathbb{P}[S_n - \mathbb{E}(S_n) \ge t] \le \exp\left\{ -st + \frac{s^2}{8} \sum_{i=1}^n (b_i - a_i)^2 \right\}.$$

The bound is optimized by minimizing the quadratic exponent with respect to s. The minimum occurs at $s = 4t / \sum (b_i - a_i)^2$, which establishes the one-sided inequality,

$$\mathbb{P}[S_n - \mathbb{E}(S_n) \ge t] \le \exp\left\{ -\frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2} \right\}.$$

An identical bound for $\mathbb{P}[S_n - \mathbb{E}(S_n) \le -t]$ is obtained by applying the same argument to the variables $\{-X_i\}$. The two-sided inequality then follows from the union bound,

$$\mathbb{P}[|S_n - \mathbb{E}(S_n)| \ge t] \le 2 \exp\left\{ -\frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2} \right\}.$$ □