Exponentially Small Probabilities

Lu Yuxuan

Shandong University

November 30, 2025

1. Chernoff Bound

A common feature: With large probability, $X \approx \mathbb{E}X$
We will estimate $\mathbb{P}(X \geq \mathbb{E}X + t)$ that decrease expoentially as $t \to \infty$.

Theorem 1 (Markov's inequlaity)

Let $X \geq 0$ be a random vairable, for any $a \geq 0$, we have

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}X}{a}. \tag{1}$$

Bernstein's Method (1924)

For any $u \geq 0$, we have

$$\mathbb{P}(X \geq \mathbb{E}X + t) = \mathbb{P}(e^{ux} \geq e^{u(\mathbb{E}X+t)}) \leq \frac{\mathbb{E}e^{uX}}{e^{u(\mathbb{E}X+t)}}. \tag{2}$$

If we consider $X = X_1 + X_2 + \cdots + X_n$ where $X_i \sim \mathrm{Be}(p_i)$, we have

$$\mathbb{P}(X \geq \mathbb{E}X + t) \leq \exp\left\{-u\left(\sum_{i=1}^{n} p_i + t\right)\right\} \cdot \prod_{i=1}^{n}(1 - p_i + e^u \cdot p_i). \tag{3}$$

Moreover, if $X \sim \mathrm{Bi}(n, p)$, i.e. $p_i = p(i \in \{1, 2, \cdots, n\})$, we have

$$\mathbb{P}(X \geq \mathbb{E}X + t) \leq \exp\{-u(np + t)\} \cdot (1 - p + e^u \cdot p)^n. \tag{4}$$

If $t > (1-p)n = n - np$, i.e. $np + t > n$, the probability is 0 as $n \to \infty$.

Chernorff Bound

If $t \le (1-p)n$, choose u such that

$$e^u = \frac{(1-p)(np+t)}{(n-np-t)p}. \tag{5}$$

Therefore, we have the Chernoff Bound(1952).

Lemma 2 (Chernoff Bound, initial version)

Let $X \sim \mathrm{Bi}(n,p)$, for any $t \ge 0$, we have

$$\mathbb{P}(X \ge \mathbb{E}X + t) \le \begin{cases} \left(\frac{np}{np+t}\right)^{np+t} \cdot \left(\frac{n-np}{n-np-t}\right)^{n-np-t}, & 0 \le t \le n-np \\ 0, & t > n-np \end{cases}. \tag{6}$$

A large but simpler bound?

Let $\varphi(x) = (1+x)\log x - x$, for $x \geq -1$, we have

- $\varphi(x) \geq 0$ for any x; $\varphi(0) = 0$.
- $\varphi'(x) = \log(1+x)$, thus

$$\varphi(x) \geq \frac{x^2}{2}, \quad \forall -1 \leq x \leq 0. \tag{7}$$

- Since $\varphi(0) = \varphi'(0) = 0$ and

$$\varphi''(x) = \frac{1}{1+x} \geq \frac{1}{(1+x/3)^3} = \left(\frac{x^2}{2(1+x/3)}\right)''. \tag{8}$$

thus

$$\varphi(x) \geq \frac{x^2}{2(1+x/3)}. \tag{9}$$

A Large But Simpler Bound

Let $\lambda = np$, so we can rewrite

$$\mathbb{P}(X \geq \mathbb{E}X + t) \leq \left(\frac{\lambda}{\lambda + t}\right)^{\lambda+t} \cdot \left(\frac{n-\lambda}{n-\lambda-t}\right)^{n-\lambda-t}, \quad 0 \leq t \leq n-\lambda \tag{10}$$

as

$$\mathbb{P}(X \geq \mathbb{E}X + t) \leq \exp\left\{-\lambda\varphi\left(\frac{t}{\lambda}\right) - (n-\lambda)\varphi\left(\frac{-t}{n-\lambda}\right)\right\}, \quad 0 \leq t \leq n-\lambda. \tag{11}$$

Repacing X by $n - X$ and we obtain also

$$\mathbb{P}(X \leq \mathbb{E}X - t) \leq \exp\left\{-\lambda\varphi\left(\frac{-t}{\lambda}\right) - (n-\lambda)\varphi\left(\frac{t}{n-\lambda}\right)\right\}, \quad 0 \leq t \leq \lambda. \tag{12}$$

Thus we have

$$\mathbb{P}(X \geq \mathbb{E}X + t) \leq \exp\left\{-\lambda\varphi\left(\frac{t}{\lambda}\right)\right\}, \quad \mathbb{P}(X \leq \mathbb{E}X - t) \leq \exp\left\{-\lambda\varphi\left(\frac{-t}{\lambda}\right)\right\}, t \geq 0. \tag{13}$$

Chernoff Bound

Using $\varphi(x) \geq x^2/(2(1+x/3))$, we have

$$\exp\left\{-\lambda\varphi\left(\frac{t}{\lambda}\right)\right\} \leq \exp\left\{-\frac{t^2}{2(\lambda+t/3)}\right\}, \exp\left\{-\lambda\varphi\left(\frac{-t}{\lambda}\right)\right\} \leq \exp\left\{-\frac{t^2}{2\lambda}\right\}. \tag{14}$$

Therefore, we have the Chernoff Bound.

Lemma 3 (Chernoff Bound)

If $X \in \mathrm{Bi}(n,p)$ and $\lambda = np$, we have

$$\begin{aligned} \mathbb{P}(X \geq \mathbb{E}X + t) \leq \exp\left\{-\frac{t^2}{2(\lambda+t/3)}\right\}, t \geq 0 \\ \mathbb{P}(X \leq \mathbb{E}X - t) \leq \exp\left\{-\frac{t^2}{2\lambda}\right\}, t \geq 0 \end{aligned}. \tag{15}$$

$$\mathbb{P}(X \geq \mathbb{E}X + t) \leq \exp\left\{-\frac{t^2}{2(\lambda + t/3)}\right\}, t \geq 0. \tag{16}$$

When t is small, i.e. $(t \leq \lambda)$, the exponent is $\Theta(t^2)$.
When t is large,

In particular, let $t = \varepsilon\mathbb{E}X$, we have

$$\mathbb{P}(X - \mathbb{E}X \geq \varepsilon\mathbb{E}X) \leq \exp\left\{-\mathbb{E}X \cdot \varphi\left(\frac{\varepsilon\mathbb{E}X}{\mathbb{E}X}\right)\right\} = \exp\left\{-\varphi(\varepsilon)\right\} \leq \exp\left\{-\frac{\varepsilon^2}{3}\mathbb{E}X\right\}. \tag{17}$$

since $\varphi(-\varepsilon) > \varphi(\varepsilon) \geq \varepsilon^2/3$ when $\varepsilon \leq 2/3$. Therefore, we have the Chernoff Bound.

Chernoff Bound

Theorem 4 (Chernoff Bound)

If $X \sim \mathrm{Bi}(n, p)$ *and for any* $0 < \varepsilon < 3/2$, *we have*

$$\mathbb{P}(|X - \mathbb{E}X| \geq \varepsilon \mathbb{E}X) \leq 2\exp\left\{-\frac{\varepsilon^2 \mathbb{E}X}{3}\right\}. \tag{18}$$

For $X \sim \mathrm{Be}(p_i)$ and $X \sim H(n, M, N)$, Chernoff Bound still works.
Single variable $\rightarrow$ More variables ?

McDiarmid's Inequality for n Independent Trials

Definition 5 (C-Lipschitz)

X is called C-Lipschitz if $|X(\omega) - X(\omega')| \leq C$ whenever ω and ω' differ in at most 1-coordinate.

Theorem 6 (McDiarmid's inequality)

Let X be C-Lipschitz random variable on a product probability space with n coordinates. Then, for any $t > 0$,

$$\mathbb{P}(|X - \mathbb{E}X| > \gamma \mathbb{E}X) \leq 2 \exp\left\{ -\frac{\gamma^2 \mathbb{E}X^2}{nC^2} \right\}. \tag{19}$$

If $\mathbb{E}X = o(\sqrt{n})$, McDiarmid's inequality doesn't work. **The dependency on n!**

Theorem 7 (Talagrand's Inequality)

Let X be a non-negative r.v. determined by n independent trials, and satisfying the following for some $r, c > 0$.

1. *X is C-Lipschitz.*
2. *For any s, and sample ω with $X(\omega) > s$, here are $\leq rs$ indices of ω, so that for all other samples ω' that agree with ω on those $\leq rs$ indices, $X(\omega' > s)$ also.*

Then, for all $t > \sqrt{\mathbb{E}X}$, if $\mathbb{E}X$ is sufficiently large and $\beta < 1/8c^2r$, we have:

$$\mathbb{P}(|X - \mathbb{E}X| \geq t) \leq 2\exp\left\{-\frac{\beta t^2}{\mathbb{E}X}\right\}. \tag{20}$$

When $t = \varepsilon\mathbb{E}X$, RHS is $2\exp\{-\beta\varepsilon^2\mathbb{E}X\}$, not depend on n. **How about dependent trials?**

Azuma's Inequality

Lemma 8 (Azuma's Inequality)

Let X *be a r.v. determined by not necessarily independent outcomes of* n *trials,* $T_1, T_2, \cdots, T_n$*, such that for any* $i \in [n]$*, and any two possible sequences of outcomes* $t_1, \cdots, t_{i-1}, t_i$ *and* $t_1, \cdots, t_{i-1}, t_i'$*, we have*

$$\left| \mathbb{E}\left(X \mid \bigcap_{j=1}^{i} T_j = t_j \right) - \mathbb{E}\left(X \mid \bigcap_{j=1}^{i-1} T_j = t_j, T_i = t_i \right) \right| \leq c_i. \tag{21}$$

then

$$\mathbb{P}(|X - \mathbb{E}X| > t) \leq 2 \exp \left\{ -\frac{t^2}{\sum_{i=1}^{n} c_i^2} \right\}. \tag{22}$$