Advanced Algorithms V (Fall 2020)

Instructor: Chihao Zhang
Scribed by: Jiaxin Lu, Yi Wang

Last modified on Oct 16, 2020

Starting from this lecture, we shall introduce the Chernoff-typed inequalities and their applications. Today we will talk about the vanilla Chernoff bound and the Hoeffding's inequality.

1 Chernoff-typed Bounds

1.1 Concentration inequalities

A concentration inequality is an upper bound on

$$\mathbf{Pr}\left[|X - \mathbf{E}[X]| \ge t\right].$$

On way to obtain a sharper bound is to choose certain non-decreasing function f and apply it on both sides of the inequality:

$$\mathbf{Pr}\left[|X - \mathbf{E}[X]| \ge t\right] = \mathbf{Pr}\left[f(|X - \mathbf{E}[X]|) \ge f(t)\right].$$

Then by Markov's inequality,

$$\mathbf{Pr}\left[|X - \mathbf{E}[X]| \ge t\right] = \mathbf{Pr}\left[f(|X - \mathbf{E}[X]|) \ge f(t)\right] \le \frac{\mathbf{E}\left[f(|X - \mathbf{E}[X]|)\right]}{f(t)}.$$

For example, if we choose $f(x) = x^2$, the inequality becomes to

$$\mathbf{Pr}\left[|X - \mathbf{E}[X]| \ge t\right] \le \frac{\mathbf{E}\left[(X - \mathbf{E}[X])^2\right]}{t^2} = \frac{\mathbf{Var}[X]}{t^2},$$

which is exactly the Chebyshev's inequality. It is natural to apply $f(x) = e^{\alpha x}$ for $\alpha > 0$ so that we can relate the upper bound to the moment generating function $\mathbf{E}\left[e^{\alpha X}\right]$ of X. In cases that $\mathbf{E}\left[e^{\alpha X}\right]$ is easy to estimate, we obtain sharper concentration.

1.2 Vanilla Chernoff Bound

When the random variable X can be written as the sum of independent Bernoulli variables, its moment generating function is easy to estimate.

Theorem 1. *Let $X_1, \ldots, X_n$ be independent random variables such that $X_i \sim \mathrm{Ber}(p_i)$ for each $i = 1, 2, \ldots, n$. Let $X = \sum_{i=1}^n X_i$ and denote $\mu \triangleq \mathbf{E}[X] = \sum_{i=1}^n p_i$, we have*

$$\mathbf{Pr}[X \geq (1+\delta)\mu] \leq \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mu} \tag{1}$$

If $0 < \delta < 1$, then we have

$$\mathbf{Pr}[X \leq (1-\delta)\mu] \leq \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{\mu} \tag{2}$$

Proof. We only prove (1) and the proof of (2) is similar. For every $\alpha > 0$, we have

$$\mathbf{Pr}[X \geq (1+\delta)\mu] = \mathbf{Pr}\left[e^{\alpha X} \geq e^{\alpha(1+\delta)\mu}\right] \leq \frac{\mathbf{E}\left[e^{\alpha X}\right]}{e^{\alpha(1+\delta)\mu}}. \tag{3}$$

Therefore, we need to estimate the moment generating function $\mathbf{E}\left[e^{\alpha X}\right]$. Since $X = \sum_{i=1}^n X_i$ is the sum of independent Bernoulli variables, we have

$$\mathbf{E}\left[e^{\alpha X}\right] = \mathbf{E}\left[e^{\alpha \sum_{i=1}^n X_i}\right] = \mathbf{E}\left[\prod_{i=1}^n e^{\alpha X_i}\right] = \prod_{i=1}^n \mathbf{E}\left[e^{\alpha X_i}\right].$$

Since $X_i \sim \mathrm{Ber}(p_i)$, we can compute $\mathbf{E}\left[e^{\alpha X_i}\right]$ directly:

$$\mathbf{E}\left[e^{\alpha X_i}\right] = p_i e^{\alpha} + (1-p_i) = 1 + (e^{\alpha}-1)p_i \leq \exp\left((e^{\alpha}-1)p_i\right).$$

Therefore,

$$\mathbf{E}\left[e^{\alpha X}\right] \leq \prod_{i=1}^n \exp\left((e^{\alpha}-1)p_i\right) = \exp\left((e^{\alpha}-1)\sum_{i=1}^n p_i\right) = \exp\left((e^{\alpha}-1)\mu\right). \tag{4}$$

Plugging into (3), we obtain

$$\mathbf{Pr}[X \leq (1+\delta)\mu] \leq \frac{\mathbf{E}\left[e^{\alpha x}\right]}{e^{\alpha(1+\delta)\mu}} \leq \left(\frac{\exp\left(e^{\alpha}-1\right)}{\exp\left(\alpha(1+\delta)\right)}\right)^{\mu} \tag{5}$$

Note that (5) holds for any $\alpha > 0$. Therefore, we would like to choose α so as to minimize $\frac{\exp(e^{\alpha}-1)}{\exp(\alpha(1+\delta))}$. To this end, we let

$$\left(\frac{\exp\left(e^{\alpha}-1\right)}{\exp\left(\alpha(1+\delta)\right)}\right)' = \exp\left(e^{\alpha}-1-\alpha-\alpha\delta\right)\cdot\left(e^{\alpha}-1-\delta\right) = 0.$$

This gives $\alpha = \log(1+\delta)$. Therefore

$$\mathbf{Pr}[X \leq (1+\delta)\mu] \leq \left(\frac{\exp\left(e^{\alpha}-1\right)}{\exp\left(\alpha(1+\delta)\right)}\right)^{\mu} = \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{\mu}.$$

□

The following form of Chernoff bound is more convenient to use (but weaker):

Corollary 2. *For any* $0 < \delta < 1$,

$$\mathbf{Pr}\left[X \geq (1+\delta)\mu\right] \leq \exp\left(-\frac{\delta^2}{3}\mu\right) \tag{6}$$

$$\mathbf{Pr}\left[X \leq (1-\delta)\mu\right] \leq \exp\left(-\frac{\delta^2}{2}\mu\right) \tag{7}$$

Proof. We only prove (6). It suffices to verify that for $0 < \delta < 1$, we have

$$\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}} \leq \exp\left(-\frac{\delta^2}{3}\right)$$

Taking logarithm of both sides, this is equivalent to

$$\delta - (1+\delta)\ln(1+\delta) \leq -\frac{\delta^2}{3}$$

Let $f(\delta) = \delta - (1+\delta)\ln(1+\delta) + \frac{\delta^2}{3}$ and note that

$$f'(\delta) = -\ln(1+\delta) + \frac{2}{3}\delta, \quad f''(\delta) = -\frac{1}{1+\delta} + \frac{2}{3}.$$

Then for $0 < \delta < 1/2$, $f''(\delta) < 0$, and for $1/2 < \delta < 1$, $f''(\delta) > 0$. Therefore, $f'(\delta)$ first decrease and then increase in $[0,1]$. Also note that $f'(0) = 0$, $f'(1) < 0$ and $f'(\delta) \leq 0$ when $0 \leq \delta \leq 1$. Therefore $f(\delta) \leq f(0) = 0$ and (6) holds. □

1.3 Application: Tossing Fair Coins

If we toss a fair coin n times, the average number of heads is $n/2$. We want to determine the value δ such that with high probability (say 99%), the total number of heads is in the interval of $[(1-\delta)\frac{n}{2}, (1+\delta)\frac{n}{2}]$. We use Chernoff bound to determine δ.

Let X denote the total number of heads, and $X_i \sim \text{Ber}\left(\frac{1}{2}\right)$ be the indicator of whether the i-th toss gives a head. Then by Chernoff bound, we have

$$\mathbf{Pr}\left[\left|X - \frac{n}{2}\right| \geq \delta \cdot \frac{n}{2}\right] \leq 2\exp\left(\frac{\delta^2}{3} \cdot \frac{n}{2}\right) \leq 0.01$$

So it suffices to choose

$$\delta = \Omega\left(\frac{1}{\sqrt{n}}\right)$$

1.4 Hoeffding's Inequality

One of annoying restrictions of Chernoff bound is that each X_i needs to be a Bernoulli random variable. Hoeffding's inequality generalizes Chernoff bound by allowing X_i to follow any distribution, provided its value is almost surely bounded.

Theorem 3 (Hoeffding's inequality). *Let $X_1, \ldots, X_n$ be independent random variables where each $X_i \in [a_i, b_i]$*[1] *for certain $a_i \le b_i$. Assume $\mathbf{E}[X_i] = p_i$ for every $1 \le i \le n$. Let $X = \sum_{i=1}^n X_i$ and $\mu \triangleq \mathbf{E}[X] = \sum_{i=1}^n p_i$, then*

$$\mathbf{Pr}[|X - \mu| \ge t] \le 2\exp\left(-\frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right)$$

for all $t \ge 0$.

We learnt from the proof of the Chernoff bound that the key to establish concentration inequalities of this form is to obtain a nice upper bound on the moment generating function. Therefore, the following Hoeffding's lemma will be the main technical ingredient to prove Theorem 3.

Lemma 4 (Hoeffding's lemma). *Let X be a random variable with $\mathbf{E}[X] = 0$ and $X \in [a, b]$. Then it holds that*

$$\mathbf{E}\left[e^{\alpha X}\right] \le \exp\left(\frac{\alpha^2(b-a)^2}{8}\right) \text{ for all } \alpha \in \mathbb{R}$$

Proof. We first find a linear function to upper bound $e^{\alpha x}$ so that we could apply the linearity of expectation to bound $\mathbf{E}\left[e^{\alpha X}\right]$. By the convexity of the exponential function (Figure 1), we have

$$e^{\alpha x} \le \frac{e^{\alpha b} - e^{\alpha a}}{b - a}(x - a) + e^{\alpha a}, \quad \text{for all } a \le x \le b$$

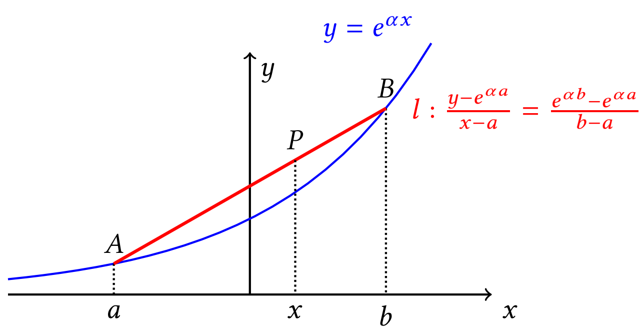

Figure 1: Bound $e^{\alpha x}$ by a linear function

Thus,

$$\begin{aligned}
\mathbf{E}[e^{\alpha x}] &\le \frac{e^{\alpha b} - e^{\alpha a}}{b - a}(-a) + e^{\alpha a} = \frac{-a}{b-a}e^{\alpha b} + \frac{b}{b-a}e^{\alpha a} \\
&= e^{\alpha a}\left(\frac{b}{b-a} - \frac{a}{b-a}e^{\alpha(b-a)}\right) \\
&= e^{-\theta t}(1 - \theta + \theta e^t) && \left(\theta = -\frac{a}{b-a},\ t = \alpha(b-a)\right) \\
&\triangleq e^{g(t)},
\end{aligned}$$

[1]In fact, $\mathbf{Pr}[X_i \in [a_i, b_i]] = 1$ suffices.

where

$$g(t) = -\theta t + \log(1 - \theta + \theta e^t)$$

By Taylor's theorem, for every real t there exists a δ between 0 and t such that,

$$g(t) = g(0) + tg'(0) + \frac{1}{2}g''(\delta)t^2$$

Note that,

$$\begin{aligned}
g(0) &= 0; \\
g'(0) &= -\theta + \left.\frac{\theta e^t}{1 - \theta + \theta e^t}\right|_{t=0} \\
&= 0; \\
g''(\delta) &= \frac{\theta e^t(1 - \theta + \theta e^t) - \theta e^t}{(1 - \theta + \theta e^t)^2} \\
&= \frac{(1-\theta)\theta e^t}{(1 - \theta + \theta e^t)^2} \\
&= \frac{(1-\theta)\theta}{\theta^2 z + 2(1-\theta)\theta + \frac{(1-\theta)^2}{z}} && (z = e^t) \\
&\leq \frac{(1-\theta)\theta}{2\theta(1-\theta) + 2(1-\theta)\theta} && (z > 0) \\
&= \frac{1}{4}.
\end{aligned}$$

Thus

$$g(t) \leq 0 + t \cdot 0 + \frac{1}{2}t^2 \cdot \frac{1}{4} = \frac{1}{8}t^2 = \frac{1}{8}\alpha^2(b-a)^2$$

Therefore, $\mathbf{E}\left[e^{\alpha x}\right] \leq \exp\left(\frac{\alpha^2(b-a)^2}{8}\right)$ holds. □

Armed with Hoeffding's lemma, it is routine to prove Hoeffding's inequality.

Proof of Theorem 3. First note that we can assume $\mathbf{E}[X_i] = 0$ and therefore $\mu = 0$ (if not so, replace X_i by $X_i - \mathbf{E}[X_i]$). By symmetry, we only need to prove that $\mathbf{Pr}[X \geq t] \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right)$. Since

$$\mathbf{Pr}[X \geq t] \overset{\alpha \geq 0}{=} \mathbf{Pr}\left[e^{\alpha X} \geq e^{\alpha t}\right] \leq \frac{\mathbf{E}\left[e^{\alpha X}\right]}{e^{\alpha t}}$$

and

$$\mathbf{E}\left[e^{\alpha X}\right] = \mathbf{E}\left[e^{\alpha \sum_{i=1}^n X_i}\right] = \prod_{i=1}^n \mathbf{E}\left[e^{\alpha X_i}\right].$$

Applying Hoeffding's lemma for each $\mathbf{E}\left[e^{\alpha X_i}\right]$ yields

$$\mathbf{E}\left[e^{\alpha X_i}\right] \leq \exp\left(-\frac{\alpha^2(b_i - a_i)^2}{8}\right).$$

Let $\alpha = \frac{4t}{\sum_{i=1}^n (b_i - a_i)^2}$, we have,

$$\mathbf{Pr}\left[X \geq t\right] \leq \frac{\prod_{i=1}^n \mathbf{E}\left[e^{\alpha X_i}\right]}{e^{\alpha t}} \leq \exp\left(-\alpha t + \frac{\alpha^2}{8}\sum_{i=1}^n (b_i - a_i)^2\right) = \exp\left(-\frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right)$$

□

1.5 Comparing Chernoff Bound and Hoeffding's Inequality

It is instructive to compare Hoeffding and Chernoff when X_i's are independent Bernoulli variables. Formally, let $X_1, \ldots, X_n$ be i.i.d. random variables where $X_i \sim \mathrm{Ber}(p)$ for all $i = 1, \ldots, n$. Set $X = \sum_{i=1}^n X_i$ and denote $\mathbf{E}\left[X\right] = np$ by μ. For $t = \delta\mu$, by Hoeffding's Inequality, we have

$$\mathbf{Pr}\left[|X - \mu| \geq t\right] \leq 2\exp\left(-2\delta^2 p^2 n\right).$$

By Chernoff Bound, we have

$$\mathbf{Pr}\left[|X - \mu| \geq t\right] \leq 2\exp\left(-\frac{1}{3}\delta^2 pn\right).$$

Comparing the exponent, it is easy to see that for some constant p like $p = 1/2$, Hoeffding's inequality is tighter up to certain constant factor. However, when p is close to 0, Chernoff bound is significantly better than Hoeffding's inequality, as its dependency to p is linear.

The following simple example demonstrates the difference. Suppose we have a box of N balls. Among them pN are red and $(1-p)N$ are blue. We draw a random ball from this box, record its color and put it back. The problem is in how many rounds we are sure about the value $\hat{p}$ (which is the percentage of red balls we record) we guess is within the range $(1 \pm 0.01)p$. The rounds required is $\Omega(1/p)$ if we apply Chernoff bound, and $\Omega(1/p^2)$ if we apply Hoeffding's inequality.

References