CSCE 689: Special Topics in Modern Algorithms for Data Science Fall 2023

Lecture 6 — September 1, 2023

Prof. Samson Zhou *Scribe: Shuo Xing*

Chernoff bound. Recall Bernstein's Inequality (the following Theorem 1), introduced in the previous lecture.

Theorem 1. *Let $X_1, \cdots, X_n \in [-M, M]$ be independent random variables and let $X = X_1 + \cdots + X_n$ have mean μ and variance σ^2. Then for any $t \geq 0$, we have*

$$Pr[|X - \mu| \geq t] \leq 2 \exp\left(-\frac{t^2}{2\sigma^2 + \frac{4}{3}Mt}\right).$$

Suppose $M = 1$ and let $t = k\sigma$. Then we have

$$Pr[|X - \mu| \geq k\sigma] \leq 2 \exp\left(\frac{-k^2}{4}\right).$$

Furthermore, if we consider binary variables, we can obtain the following Corollary 1 (also known as *Chernoff Bounds*), based on the Bernstein's inequality.

Corollary 1 (Chernoff bounds)**.** *Let $X_1, \cdots, X_n \in \{0, 1\}$ be independent random variables and let $X = X_1 + \cdots + X_n$ have mean μ. Then for any $\delta \geq 0$, we have*

$$Pr[|X - \mu| \geq \delta\mu] \leq 2 \exp\left(\frac{-\delta^2\mu}{2 + \delta}\right),$$

Proof. Since $X_1, \cdots, X_n \in \{0, 1\}$, and by the definition of variance we have

$$\sigma^2 = E[X^2] - (E[X])^2 \leq E[X^2] \leq E[X] = \mu.$$

Therefore, by Bernstein's inequality, we have

$$\begin{aligned} Pr[|X - \mu| \geq \delta\mu] &\leq 2 \exp\left(\frac{-\delta^2\mu^2}{2\sigma^2 + \frac{2}{3}\delta\mu}\right) \\ &\leq 2 \exp\left(\frac{-\delta^2\mu^2}{2\mu + \delta\mu}\right) \\ &= 2 \exp\left(\frac{-\delta^2\mu}{2 + \delta}\right). \end{aligned}$$

Note that the $\frac{2}{3}$ in the denominator in the application of Bernstein's inequality is from resymmetrizing $\{0, 1\}$ to $\left\{-\frac{1}{2}, +\frac{1}{2}\right\}$. ■

Additionally, we can also obtain the following corollary 2 (*Multiplicative Error Chernoff bounds*).

Corollary 2. *Let $X_1, \cdots, X_n \in \{0,1\}$ be independent random variables and let $X = X_1 + \cdots + X_n$ have mean μ. Then for any $\delta \in (0,1)$, we have*

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

Median-of-means framework. Suppose there is a algorithm A that outputs a real number Z that is correct with probability $\frac{2}{3}$, and we want to be correct with probability 0.999 or $1 - 1/n^2$ or $1-\delta$. By Chernoff bounds, we can know that if we independently run the algorithm A a total of $O(\log 1/\delta)$ times and take the median, the output will be correct with probability $1-\delta$.

- Let A^* be the new algorithm (repeating the A algorithm $100 \log 1/\delta$ times). Let $X_i = 1$ if the i-th output of A is correct, and the total number of the correct output of algorithm A is $X = X_1 + \cdots + X_{\log 1/\delta}$. And we know that $E[X] \geq 200/3 \cdot \log 1/\delta$.
- We take the median of these outputs of A as the final output of A^*. Thus, when $X \geq 100/2 \log 1/\delta$, the algorithm A^* returns the correct answer.
- Therefore, by Chernoff bounds, we know that

$$\begin{aligned} Pr[X \leq 100/2 \log 1/\delta] &= Pr[X - 200/3 \cdot \log 1/\delta \leq -100/6 \cdot \log 1/\delta] \\ &\leq Pr[|X - E[X]| \geq 100/6 \cdot \log 1/\delta] \\ &= Pr[|X - E[X]| \geq 100/4 \cdot E[X]] \\ &\leq 2\exp\left(\frac{-10000/16 \cdot 2/3 \cdot \log 1/\delta}{3}\right) \\ &< \delta, \end{aligned}$$

  which means the algorithm A^* will return the correct answer with probability at least $1-\delta$.

We can illustrate the core principles of the *median-of-means framework* through an example:

- Suppose we design a randomized algorithm A to estimate a hidden statistic of a dataset and we know $0 < Z \leq 1000$.
- Suppose each time we use the algorithm A, it outputs a number X such that $E[X] = Z$ and $Var[X] = 100Z^2$.
- Suppose we want to estimate Z to accuracy ϵ with probability $1-\delta$.
- *Accuracy boosting*: Repeat A a total of $10^{12}/\epsilon^2$ time and take the mean (so that we have $Pr[|X - Z| < \epsilon] > 0.999$, i.e. the mean of the repeated algorithms outputs X that estimates Z to accuracy ϵ with probability 0.999).
- *Success boosting*: Find the mean a total of $O(\log 1/\delta)$ times and take the median to be correct with probability $1-\delta$.

Max load. Suppose that we have a n-sided die that we roll n times. On average, what is the largest number of times any outcome is rolled?

- First we fix a value $k \in [n]$.
- Let $X_i = 1$ if the i-th roll is k and $X_i = 0$ otherwise. Thus, $E[X_i] = 1 \times 1/n + 0 \times (n-1)/n = 1/n$.
- If we roll the die n times, the total number of rolls with value k is $X = X_1 + \cdots + X_n$ (such that $E[X] = 1$).
- By Chernoff bounds, we can know that

$$\begin{aligned} Pr[X \geq 3\log n] &\leq Pr[X \geq (1 + 2\log n)] \\ &\leq 2\exp\left(\frac{-(2\log n)^2}{2 + 2\log n}\right) \\ &\sim 2\exp\left(\frac{-(2\log n)^2}{2\log n}\right) \\ &= 2\exp\left(\frac{-(2\log n)^2}{2\log n}\right) \\ &\leq \frac{2}{n^2}. \end{aligned}$$

- The above inequality means that with probability at least $1 - 2/n^2$, we will get fewer than $3\log n$ rolls with value k.
- Thus, by union bound, we can know that no outcome will be rolled more than $3\log n$ times with probability at least $1 - 2/n$.