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

Lecture 5 — August 30, 2023

Prof. Samson Zhou *Scribe: Wenjing Chen*

1 Concentration Inequalities and Moments

Definition. For $p > 0$, the p-th **moment** of a random variable X over Ω is:

$$\mathbb{E}[X^p] = \sum_{x \in \Omega} Pr[X = x] \cdot x^p.$$

Examples of definitions related to moments include:

- expectation $\mathbb{E}[X]$ (1-st moment),
- variance $\mathbb{E}[X - \mathbb{E}[X]]^2 = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$ (related to 2-nd moment).

By using different moments, we can obtain different concentration inequalities. For instance, Chebyshev's inequality is proved by applying Markov's inequality on the variance of the random variable $X - \mathbb{E}[X]$. In particular, Chebyshev's inequality implies the Law of Large Numbers. Let $X_1, ..., X_n$ be random variables that are independent identically distributed (i.i.d.) with mean μ and variance σ^2. Consider the sample average $X = \frac{1}{n}\sum_i X_i$, we have that

$$Var[X] = \frac{1}{n^2}\sum_i Var[X_i] = \frac{\sigma^2}{n}.$$

For any fixed parameter ε, if we apply the Chebyshev's inequality on X,

$$Pr(|X - \mu| > \varepsilon) \leq \frac{Var[X]}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2}.$$

When n goes to infinity, we have that the probability of the difference between the sample average and the mean being larger than ε becomes arbitrarily small. This implies

Theorem 1 (Law of Large Numbers, informal)**.** *The sample average will always concentrate to the mean, given enough samples.*

2 Practical Examples and Accuracy Boosting

Suppose we design a randomized algorithm A to estimate a hidden statistic Z of a dataset and we know $0 < Z \leq 1000$. Each time we use the algorithm A, it outputs a number X such that $\mathbb{E}[X] = Z$ and $Var[X] = 100Z^2$. Then by applying Chebyshev's inequalities, we have

$$Pr(|X - Z| > 30Z) \leq \frac{Var[X]}{900Z^2} = \frac{1}{9}.$$

Since $Z \leq 1000$, it holds that

$$\begin{aligned} Pr(|X - Z| < 30000) &= 1 - Pr(|X - Z| \geq 30000) \\ &\geq 1 - Pr(|X - Z| > 30Z) = \frac{8}{9}. \end{aligned}$$

Then we get the additive error is at most 30000 with high probability. However, in many real-world applications, we would like to propose algorithms with small parameterizable additive error ε. How can we achieve that?

2.1 Accuracy Boosting

If we repeat A a total of $\frac{10^{12}}{\varepsilon^2}$ times and take the average, then the variance of the average is $\frac{\varepsilon^2}{10^{10}} Z^2$ and by Chebyshev's inequalities, we have

$$Pr[|X - Z| \geq \varepsilon] \leq \frac{Z^2}{10^{10}}.$$

Since $Z \leq 1000$, we prove that

$$Pr[|X - Z| \geq \varepsilon] \leq 0.0001.$$

This instance implies the **accuracy boosting** method. To improve the accuracy of your algorithm, we can run it many times independently and take the average.

3 More Powerful Concentration Inequalities.

However, the concentration inequalities discussed before has limitations on the tightness of the bound. For example, suppose we flip a fair coin $n = 100$ times and let H be the total number of heads. Then it holds that

$$\mathbb{E}[H] = 50 \qquad \text{and} \qquad Var[H] = 25.$$

If we apply Markov's inequality, then we have

$$Pr[H \geq 60] \leq 0.833.$$

If we apply Chebyshev's inequality, then we have

$$Pr[H \geq 60] \leq 0.25.$$

Recall that Chebyshev's inequality is proved by applying Markov to the second moment of the random variable $X - \mathbb{E}[X]$, i.e.,

$$Pr[|X - \mathbb{E}[X]| \geq t] = Pr[|X - \mathbb{E}[X]|^2 \geq t^2] \leq \frac{Var[X]}{t^2}.$$

By using the similar tricks on the 4-th moment, we obtain that

$$Pr[H \geq 60] \leq 0.186,$$

which is tighter than the previous bound. However, the true value of the probability is $Pr[H \geq 60] \approx 0.0284$. By looking at the k-th moment for sufficiently high k gives a number of very strong (and useful!) concentration inequalities with exponential tail bounds. Examples of exponential tail bounds include Chernoff bounds, Bernstein's inequality, Hoeffding's inequality, etc. Below we introduce Bernstein's inequality.

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

$$Pr[|X - \mu| \geq t] \leq 2e^{-\frac{t^2}{2\sigma^2 + \frac{4}{3}Mt}}.$$

Suppose $M = 1$ and let $t = k\sigma$, then

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

However, Chebyshev's inequality gives

$$Pr[|X - \mu| \geq k\sigma] \leq \frac{1}{k^2}.$$

By comparing Bernstein's inequality and Chebyshev's inequality, we can see an exponential improvement by Bernstein's inequality. If we depict the tail bound of $\exp(-\frac{k^2}{4})$ across different values of k, we would get a plot that is similar to the normal random variable, which is consistent with the result of the central limit theorem.

Theorem 3 (Stronger Central Limit Theorem, informal)**.** *The distribution of the sum of n bounded independent random variables converges to a Gaussian (normal) distribution as n goes to infinity.*

The theorem is very important since many random variables can be approximated as the sum of a large number of small and roughly independent random effects. Thus, their distribution looks Gaussian by CLT.