CSCE 689: Special Topics in Modern Algorithms for Data Science

Lecture 5

Samson Zhou

Present and Future

- Today: Discuss potential project groups
- Friday: Email me the members/group name

- Future: Set up meetings to discuss proposed projects

Recall: Concentration Inequalities

- Concentration inequalities bound the probability that a random variable is “far away” from its expectation

- Often used in understanding the performance of statistical tests, the behavior of data sampled from various distributions, and for our purposes, the guarantees of randomized algorithms

Last Time: Moments

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

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

Last Time: Variance

- The variance of a random variable X over Ω is:

$$\mathrm{Var}[X] = \mathrm{E}[X^2] - (\mathrm{E}[X])^2$$

- Can rewrite $\mathrm{Var}[X] = \mathrm{E}[(X - \mathrm{E}[X])^2]$ since $\mathrm{E}\big[\mathrm{E}[X]\big] = \mathrm{E}[X]$

- "How far numbers are from the average"

Last Time: Chebyshev's Inequality

- Let X be a random variable with expected value $\mu := \mathrm{E}[X]$ and variance $\sigma^2 := \mathrm{Var}[X]$

- $\Pr[|X - \mathrm{E}[X]| \geq t] \leq \frac{\mathrm{Var}[X]}{t^2}$ becomes $\Pr[|X - \mathrm{E}[X]| \geq t] \leq \frac{\sigma^2}{t^2}$

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

- "Bounding the deviation of a random variable in terms of its variance"

Last Time: Law of Large Numbers

- Let $X_1, \dots, X_n$ be random variables that are independent identically distributed (i.i.d.) with mean μ and variance σ^2
- Consider the sample average $X = \frac{1}{n}\sum_i X_i$. How does it compare to μ?
- $\mathrm{Var}[X] = \frac{1}{n^2}\sum_i \mathrm{Var}[X_i] = \frac{\sigma^2}{n}$
- Law of Large Numbers: The sample average will always concentrate to the mean, given enough samples

Use Case

- Suppose we design a randomized algorithm A to estimate a hidden statistic Z 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 $\mathrm{E}[X] = Z$ and $\mathrm{Var}[X] = 100Z^2$

- What can we say about A?

- $\Pr[|X - Z| \geq 30Z] \leq \frac{1}{9}$ and $Z \leq 1000$ so $\Pr[|X - Z| < 30{,}000] > \frac{8}{9}$

Accuracy Boosting

- How can we use A to get additive error ε?

Accuracy Boosting

- How can we use A to get additive error ε?
- Repeat A a total of $\frac{10^{12}}{\varepsilon^2}$ times and take the average
- The variance of the average is $\frac{\varepsilon^2}{10^{10}} Z$ and $\Pr[|X-\mu| \geq k] \leq \frac{\sigma^2}{k^2}$
- $\Pr[|X-Z| \geq \varepsilon] \leq \frac{Z}{10^{10}}$ and $Z \leq 1000$ so $\Pr[|X-Z| < \varepsilon] > 0.999$

Accuracy Boosting

- Algorithmic consequence of Law of Large Numbers

- To improve the accuracy of your algorithm, run it many times independently and take the average

Limitations

- Suppose we flip a fair coin $n = 100$ times and let H be the total number of heads
- $\mathrm{E}[H] = 50$ and $\mathrm{Var}[H] = 25$

- Markov's inequality: $\Pr[H \geq 60] \leq 0.833$
- Chebyshev's inequality: $\Pr[H \geq 60] \leq 0.25$
- Truth: $\Pr[H \geq 60] \approx 0.0284$

Intuition for Previous Inequalities

- Recall: We proved Markov's inequality by looking at the first moment of the random variable X

$$\Pr[X \geq t \cdot \mathrm{E}[X]] \leq \frac{1}{t}$$

- Recall: We proved Chebyshev's inequality by applying Markov to the second moment of the random variable $X - \mathrm{E}[X]$

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

Generalizations

- Suppose we flip a fair coin $n = 100$ times and let H be the total number of heads

- What if we consider higher moments?
- Looking at the 4^{th} moment: $\Pr[H \geq 60] \leq 0.186$

- Markov's inequality: $\Pr[H \geq 60] \leq 0.833$
- Chebyshev's inequality: $\Pr[H \geq 60] \leq 0.25$
- Truth: $\Pr[H \geq 60] \approx 0.0284$

Concentration Inequalities

- Looking at the k^{th} moment for sufficiently high k gives a number of very strong (and useful!) concentration inequalities with exponential tail bounds

- Chernoff bounds, Bernstein's inequality, Hoeffding's inequality, etc.

Bernstein's Inequality

- Bernstein's inequality: Let $X_1, \ldots, X_n \in [-M, M]$ be independent random variables and let $X = X_1 + \cdots + X_n$ have mean μ and variance σ^2. Then for any $t \geq 0$:

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

Bernstein's Inequality

- Berstein's inequality: Let $X_1, \dots, X_n \in [-M, M]$ be independent random variables and let $X = X_1 + \cdots + X_n$ have mean μ and variance σ^2. Then for any $t \geq 0$:

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

- Example: Suppose $M = 1$ and let $t = k\sigma$. Then

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

Bernstein's Inequality

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

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

- Compare to Chebyshev's inequality:

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

- Exponential improvement!

Bernstein's Inequality

- Suppose we flip a fair coin $n = 100$ times and let H be the total number of heads

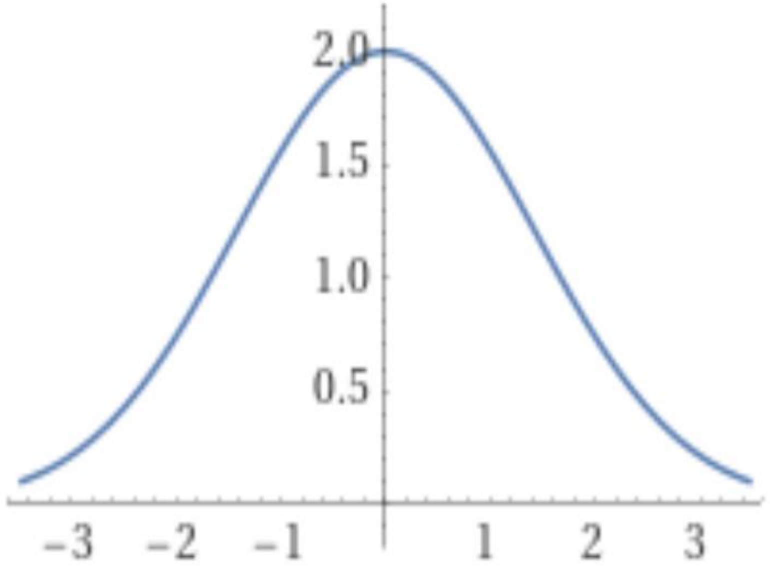
- Markov's inequality: $\Pr[H \geq 60] \leq 0.833$
- Chebyshev's inequality: $\Pr[H \geq 60] \leq 0.25$
- 4^{th} moment: $\Pr[H \geq 60] \leq 0.186$
- Bernstein's inequality: $\Pr[H \geq 60] \leq 0.15$
- Truth: $\Pr[H \geq 60] \approx 0.0284$

Bernstein's Inequality

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

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

- Plot across values of k looks like normal random variable



- PDF of Gaussian $\mathrm{N}(0, \sigma^2)$ is

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}$$

Central Limit Theorem

- Stronger Central Limit Theorem: The distribution of the sum of n bounded independent random variables converges to a Gaussian (normal) distribution as n goes to infinity

- Why is the Gaussian distribution is so important in statistics, data science, ML, etc.?

- 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.