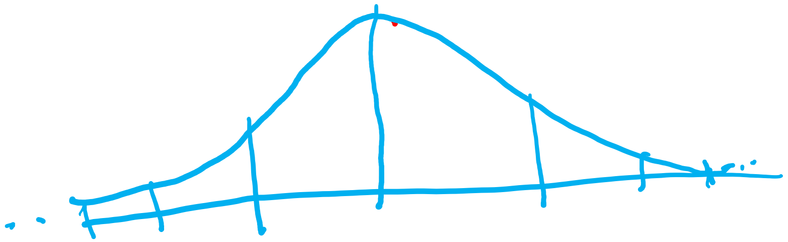
No activity slide today



Tail Bounds 2

CSE 312 Spring 21
Lecture 22

Announcements

Final logistical information coming to Ed in the next two days.

Pset 6 grades back. Trying an experiment – regrade requests will open tomorrow.

Real World ~~2~~ 3 is out, due Tuesday June 1.

Pset 8 out tonight (due in one week)

Our First bound

Two statements are equivalent.
Left form is often easier to use.
Right form is more intuitive.

Markov's Inequality

Let X be a random variable supported (only) on non-negative numbers. For any $t > 0$

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}[X]}{t}$$

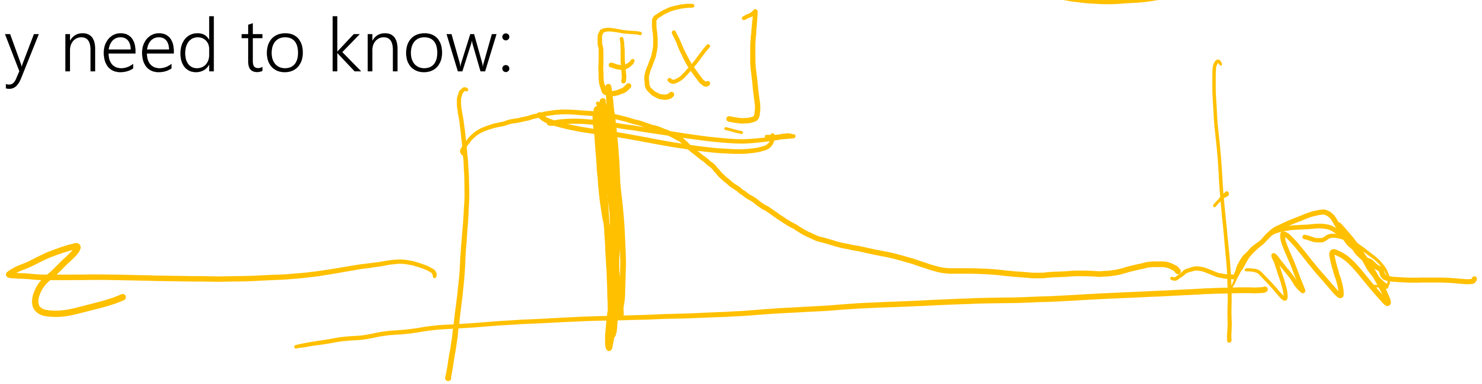
Markov's Inequality

Let X be a random variable supported (only) on non-negative numbers. For any $\mathrm{k} > 0$

$$\mathbb{P}(X \geq k\mathbb{E}[X]) \leq \frac{1}{k}$$

To apply this bound you only need to know:

1. it's non-negative
2. Its expectation.



So…what do we do?

A better inequality!

We're trying to bound the tails of the distribution.

What parameter of a random variable describes the tails?

The variance!

Chebyshev's Inequality

Two statements are equivalent.
Left form is often easier to use.
Right form is more intuitive.

Chebyshev's Inequality

Let X be a random variable. For any $t > 0$

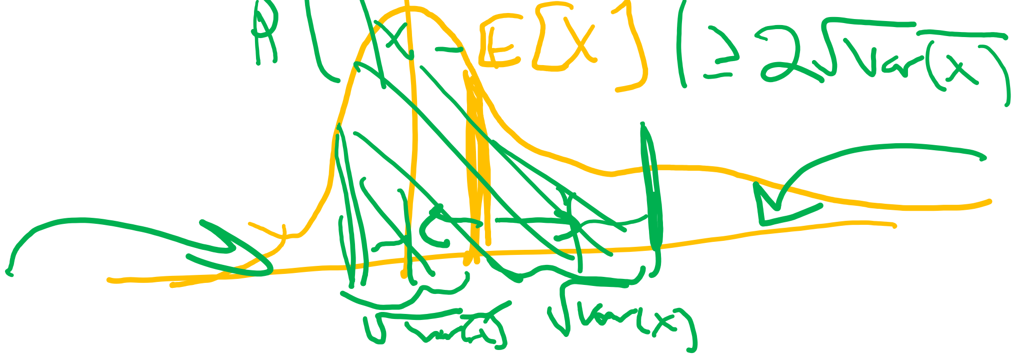
$$\mathbb{P}(|X - \mathbb{E}[X]| \geq t) \leq \frac{\operatorname{Var}(X)}{t^2}$$

Chebyshev's Inequality

Let X be a random variable. For any $\mathbf{k} > 0$

$$\mathbb{P}\left(|X - \mathbb{E}[X]| \geq k\sqrt{\operatorname{Var}(X)}\right) \leq \frac{1}{k^2}$$

$$\mathbb{P}\left(|X - \mathbb{E}[X]| \geq 2\sqrt{Var(X)}\right) \leq \frac{1}{k^2}$$



Proof of Chebyshev

Chebyshev's Inequality

Let X be a random variable. For any $t > 0$

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq t) \leq \frac{\mathrm{Var}(X)}{t^2}$$

Let $Z = X - \mathbb{E}[X]$

Markov's Inequality

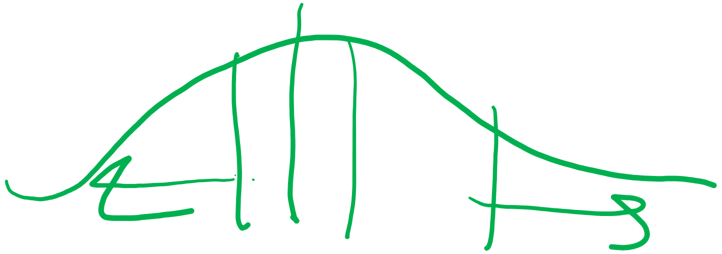
$\mathbb{E}[Z] = 0$

$$\mathbb{P}(|Z| \geq t) = \mathbb{P}(Z^2 \geq t^2) \leq \frac{\mathbb{E}[Z^2]}{t^2} = \frac{\mathbb{E}[Z^2] - (\mathbb{E}[Z])^2}{t^2} = \frac{\mathrm{Var}(Z)}{t^2} = \frac{\mathrm{Var}(X)}{t^2}$$

Inequalities are equivalent (square each side).

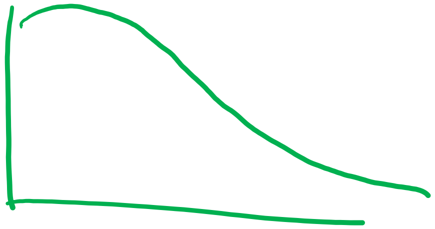
Z is just X shifted. Variance is unchanged.

Example with geometric RV



Suppose you roll a fair (6-sided) die until you see a 6. Let X be the number of rolls. $X \sim geo\left(\frac{1}{6}\right)$. $\mathbb{E}[X] = \frac{1}{1/6} = 6$

$\mathrm{Var}(X) = \frac{1-p}{p^2} = \frac{5/6}{1/36}$



Bound the probability that $X \geq 12$

$$\mathbb{P}(X \geq 12) = \mathbb{P}(\underbrace{X - 6 \geq 6}_{A})$$

$$\leq \mathbb{P}(\underbrace{|X - 6| \geq 6}_{B})$$

$$\leq \frac{\frac{5/6}{1/36}}{6^2}$$

A B

A/B

Chebyshev's Inequality

Let X be a random variable. For any $t > 0$

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq t) \leq \frac{\mathrm{Var}(X)}{t^2}$$

Example with geometric RV

Suppose you roll a fair (6-sided) die until you see a 6. Let X be the number of rolls.

Bound the probability that $X \geq 12$

$$\mathbb{P}(X \geq 12) \leq \mathbb{P}(|X - 6| \geq 6) \leq \frac{\frac{5/6}{1/36}}{6^2} = \frac{5}{6}$$

Not any better than Markov ☹

$\mathbb{P}(X \geq 12) \leq \frac{1}{2}$.

Chebyshev's Inequality

Let X be a random variable. For any $t > 0$

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq t) \leq \frac{\mathrm{Var}(X)}{t^2}$$

Example with geometric RV

Let X be a geometric rv with parameter p

Bound the probability that $X \geq \frac{2}{p}$

$$\mathbb{P}(X \geq 2/p) \leq \mathbb{P}(|X - 1/p| \geq 1/p) \leq \frac{\frac{1-p}{p^2}}{1/p^2} = 1 - p$$

Markov gives:

$$\mathbb{P}\left(X \geq \frac{2}{p}\right) = \frac{\mathbb{E}[X]}{2/p} = \frac{1}{p} \cdot \frac{p}{2} = \frac{1}{2}.$$

For large p, Chebyshev is better.

Chebyshev's Inequality

Let X be a random variable. For any $t > 0$

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq t) \leq \frac{\mathrm{Var}(X)}{t^2}$$

Better Example

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

Suppose the average number of ads you see on a website is 25. And the variance of the number of ads is 16. Give an upper bound on the probability of seeing a website with 30 or more ads.

Better Example

Suppose the average number of ads you see on a website is 25. And the variance of the number of ads is 16. Give an upper bound on the probability of seeing a website with 30 or more ads.

$$\underbrace{\mathbb{P}(X \geq 30)}_{\text{"}\mathbb{P}(X-25 \geq 5)\text{"}} \leq \mathbb{P}(|X - 25| \geq 5) \leq \frac{16}{25}$$

Near the mean

Suppose you run a poll of 1000 people where in the true population 60% of the population supports you. What is the probability that the poll is not within 10-percentage-points of the true value?

$\bar{X} = \sum X_i/1000$

$\mathbb{E}[\bar{X}] = 1000 \cdot \frac{.6}{1000} = \frac{3}{5}$

$\mathrm{Var}(\bar{X}) = 1000 \frac{.6 \cdot .4}{1000^2} = \frac{3}{12500}$

Chebyshev's Inequality

Let X be a random variable. For any $t > 0$

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq t) \leq \frac{\mathrm{Var}(X)}{t^2}$$

Near the mean

Suppose you run a poll of 1000 people where in the true population 60% of the population supports you. What is the probability that the poll is not within 10-percentage-points of the true value?

$$\bar{X} = \sum X_i / 1000$$

$$\mathbb{E}[\bar{X}] = 1000 \cdot \frac{.6}{1000} = \frac{3}{5}$$

$$\text{Var}(\bar{X}) = 1000 \frac{.6 \cdot .4}{1000^2} = \frac{3}{12500}$$

Chebyshev's Inequality

Let X be a random variable. For any $t > 0$

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq t) \leq \frac{\text{Var}(X)}{t^2}$$

Near the mean

Suppose you run a poll of 1000 people where in the true population 60% of the population supports you. What is the probability that the poll is not within 10-percentage-points of the true value?

$$\bar{X} = \sum X_i / 1000$$

$$\mathbb{E}[\bar{X}] = 1000 \cdot \frac{.6}{1000} = \frac{3}{5}$$

$$\mathrm{Var}(\bar{X}) = 1000 \frac{.6 \cdot .4}{1000^2} = \frac{3}{12500}$$

$$\mathbb{P}(|\bar{X} - \mathbb{E}[\bar{X}]| \geq .1) \leq \frac{3/12500}{.1^2} = .024$$

Chebyshev's Inequality

Let X be a random variable. For any $t > 0$

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq t) \leq \frac{\mathrm{Var}(X)}{t^2}$$

Chebyshev's – Repeated Experiments

How many coin flips (each head with probability p) are needed until you get n heads.

Let X be the number necessary. What is probability $X \geq 2n/p$?

Markov

Chebyshev

Chebyshev's – Repeated Experiments

How many coin flips (each head with probability p) are needed until you get n heads.

Let X be the number necessary. What is probability $X \geq 2n/p$?

Markov $\quad \mathbb{P}\left(X \geq \frac{2n}{p}\right) \leq \frac{n/p}{2n/p} = \frac{1}{2}$

Chebyshev $\quad \mathbb{P}\left(X \geq \frac{2n}{p}\right) \leq \mathbb{P}\left(\left|X - \frac{n}{p}\right| \geq \frac{n}{p}\right) \leq \frac{\mathrm{Var}(X)}{n^2/p^2} = \frac{n(1-p)/p^2}{n^2/p^2} = \frac{1-p}{n}$

Takeaway

Chebyshev gets more powerful as the variance shrinks.

Repeated experiments are a great way to cause that to happen.

More Assumptions → Better Guarantee

(Multiplicative) Chernoff Bound

Let $X_1, X_2, \ldots, X_n$ be *independent* Bernoulli random variables.
Let $X = \sum X_i$, and $\mu = \mathbb{E}[X]$. For any $0 \leq \delta \leq 1$

$$\boxed{\mathbb{P}(X \geq (1+\delta)\mu) \leq \exp\left(-\frac{\delta^2 \mu}{3}\right)} \text{ and } \boxed{\mathbb{P}(X \leq (1-\delta)\mu) \leq \exp\left(-\frac{\delta^2 \mu}{2}\right)}$$

$e^{-\delta^2\mu/3}$

$\frac{1}{t^2}$

Same Problem, New Solution

Suppose you run a poll of 1000 people where in the true population 60% of the population supports you. What is the probability that the poll is not within 10-percentage-points of the true value?

(Multiplicative) Chernoff Bound

Let $X_1, X_2, \dots, X_n$ be *independent* Bernoulli random variables.

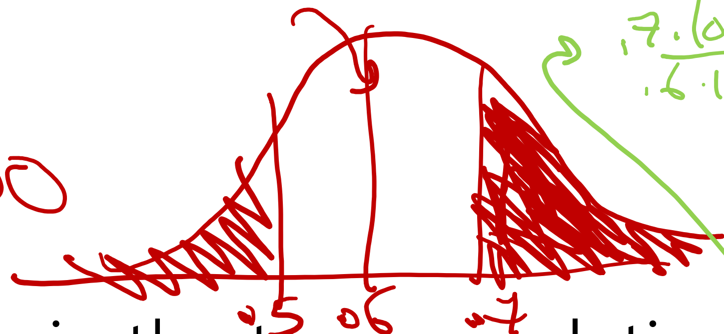
Let $X = \sum X_i$, and $\mu = \mathbb{E}[X]$. For any $0 \leq \delta \leq 1$

$$\mathbb{P}(X \geq (1+\delta)\mu) \leq \exp\left(-\frac{\delta^2 \mu}{3}\right) \text{ and } \mathbb{P}(X \leq (1-\delta)\mu) \leq \exp\left(-\frac{\delta^2 \mu}{2}\right)$$

Right Tail

$.7 \cdot 1000 = (1+\delta) \cdot .6 \cdot 1000$

$\frac{.7}{.6} = 1+\delta \quad \delta = \frac{1}{6}$



$\frac{.7 \cdot 1000}{.6 \cdot 1000} = 1+\delta$

$\frac{7}{6} = 1+\delta$

$\delta = \frac{1}{6}$

Suppose you run a poll of 1000 people where in the true population 60% of the population supports you. What is the probability that the poll is not within 10-percentage-points of the true value?

$.7 \cdot 1000 = (1+\delta)[.6 \cdot 1000]$

Want $\mathbb{P}\left(\frac{X}{1000} \geq .7\right) = \mathbb{P}(X \geq .7 \cdot 1000)$

$= \mathbb{P}(X \geq (1 + .1/.6) \cdot (.6 \cdot 1000))$

So $\delta = \frac{1}{6}$ and $\mu = .6 \cdot 1000$

$\mathbb{P}(X \geq 700) \leq \exp\left(-\frac{\frac{1}{6^2} \cdot .6 \cdot 1000}{3}\right)$

≤ 0.0039

Chernoff Bound (right tail)

Let $X_1, X_2, \ldots, X_n$ be *independent* Bernoulli random variables. *Let* $X = \sum X_i$, and $\mu = \mathbb{E}[X]$. For any $0 \leq \delta \leq 1$

$$\mathbb{P}(X \geq (1+\delta)\mu) \leq \exp\left(-\frac{\delta^2 \mu}{3}\right)$$

Left Tail

Suppose you run a poll of 1000 people where in the true population 60% of the population supports you. What is the probability that the poll is not within 10-percentage-points of the true value?

Want $\mathbb{P}\left(\frac{X}{1000} \le .5\right) = \mathbb{P}(X \le .5 \cdot 1000)$

$= \mathbb{P}(X \le (1 - .1/.6) \cdot (.6 \cdot 1000))$

So $\delta = \frac{1}{6}$ and $\mu = .6 \cdot 1000$

$\mathbb{P}(X \le 500) \le \exp\left(-\frac{\frac{1}{6^2} \cdot .6 \cdot 1000}{2}\right)$

≤ 0.0003

Chernoff Bound (left tail)

Let $X_1, X_2, \dots, X_n$ be *independent* Bernoulli random variables. *Let* $X = \sum X_i$, and $\mu = \mathbb{E}[X]$. For any $0 \le \delta \le 1$

$$\mathbb{P}(X \le (1-\delta)\mu) \le \exp\left(-\frac{\delta^2 \mu}{2}\right)$$

Both Tails

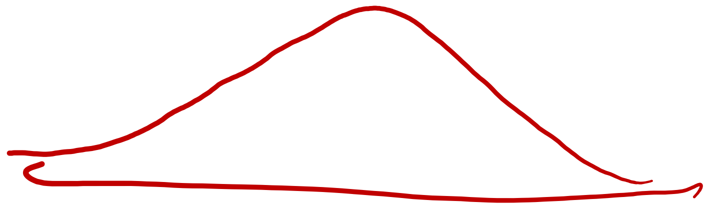
Let E be the event that X is not between 500 and 700 (i.e. we're not within 10 percentage points of the true value)

$$\mathbb{P}(E) = \mathbb{P}(X < 500) + \mathbb{P}(X > 700)$$

$$\leq .0039 + .0003 = .0042$$

Less than 1%. That's a better bound than Chebyshev gave!

Wait a Minute



I asked Wikipedia about the "Chernoff Bound" and I saw something different?

This is the "easiest to use" version of the bound. If you need something more precise, there are other versions.

Why are the tails different??

The strongest/original versions of "Chernoff bounds" are symmetric $(1+\delta$ and $1-\delta$ **correspond**), but those bounds are ugly and hard to use.

When computer scientists made the "easy to use versions", they needed to use some inequalities. The numerators now have plain old δ's, instead of $1+$ or $1-$. As part of the simplification to this version, there were different inequalities used so you don't get exactly the same expression.

Wait a Minute

This is just a binomial!

The concentration inequality will let you control n easily, even as a variable. That's not easy with the binomial.

What happens when n gets big?

Evaluating $\binom{20000}{10000}.51^{10000} \cdot .49^{10000}$ is fraught with chances for floating point error and other issues. Chernoff is much better.

But Wait! There's More

For this class, please limit yourself to:
Markov, Chebyshev, and Chernoff, as stated in these slides…

But for your information. There's more.

Trying to apply Chebyshev, but only want a "one-sided" bound (and tired of losing that almost-factor-of-two)Try Cantelli's Inequality

In a position to use Chernoff, but want additive distance to the mean instead of multiplicative? They got one of those.

Have a sum of independent random variables that aren't indicators, but are bounded, you better believe Wikipedia's got one

Have a sum of random **matrices** instead of a sum of random numbers. Not only is that a thing you can do, but the eigenvalue of the matrix concentrates

Tail Bounds – Takeaways

Useful when an experiment is complicated and you just need the probability to be small (you don't need the exact value).

Choosing a minimum n for a poll – don't need exact probability of failure, just to make sure it's small.

Designing probabilistic algorithms – just need a guarantee that they'll be extremely accurate

Learning more about the situation (e.g. learning variance instead of just mean, knowing bounds on the support of the starting variables) usually lets you get more accurate bounds.

Next Time

One more bound (the union bound)

Not a concentration bound -- one more tool for handling non-independence.

We'll see it in the context of some applications!