Please download the activity slide! ☺

# Tail Bounds Continued

CSE 312 Summer 21
Lecture 20

# Announcements

Problem Set 7 is due on Monday, Aug 16 and not Thursday, Aug 12

Deadlines to be aware of:

- Real World 2 – Wednesday, Aug 11
- Review Summary 3 – Friday, Aug 13
- Problem Set 7 – Monday, Aug 16

# Final Logistics

- Will be released on Friday evening, Aug 13
- Timed 2 hours – NO PROCTORS except yourself!
- The key will be released on Tuesday, Aug 17, at midnight
- Interviews conducted Wednesday – Friday of the final week
- Submit your attempt on Gradescope before the interview

# Markov's Inequality

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 $t > 0$

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

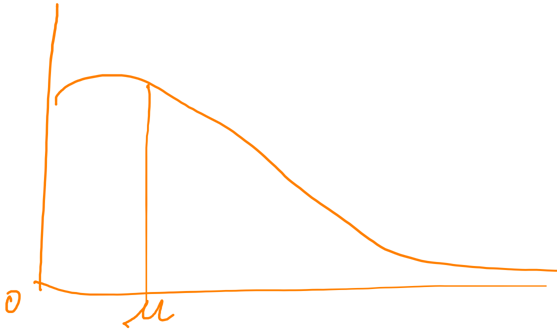
## Markov's Inequality

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 $t > 0$

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

## Chebyshev's Inequality

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

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

# 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?

$X_i \rightarrow$ indicator of support for you.

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

$$\mathbb{E}[\bar{X}] = 1000 \cdot \frac{0.6}{1000} = \frac{3}{5}$$

$$\mathrm{Var}(\bar{X}) = 1000 \cdot \frac{0.6 \cdot 0.4}{1000^2} = \frac{3}{12500}$$

$$\mathbb{P}(|\bar{X} - \mathbb{E}[\bar{X}]| \geq 0.1) \leq \frac{3/12500}{0.1^2} = 0.024$$

## 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

$Y \sim Geo(p)$ $E[Y] = \frac{1}{p}$

$E[X] = \frac{n}{p}$

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

NegBio$(n, p)$

$Var(X) = \frac{n(1-p)}{p^2}$

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

Markov

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

Chebyshev

$$\mathbb{P}\left(X \geq \frac{2n}{p}\right) = \mathbb{P}\left(X - \frac{n}{p} \geq \frac{n}{p}\right) \leq \mathbb{P}\left(\left|X - \frac{n}{p}\right| \geq \frac{n}{p}\right)$$

$$\leq \frac{Var(X)}{(n/p)^2} = \frac{n(1-p)/p^2}{n^2/p^2} = \frac{1-p}{n}$$

Decreases as $n$ increases

# 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 $X \geq \frac{2n}{p}$?

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.

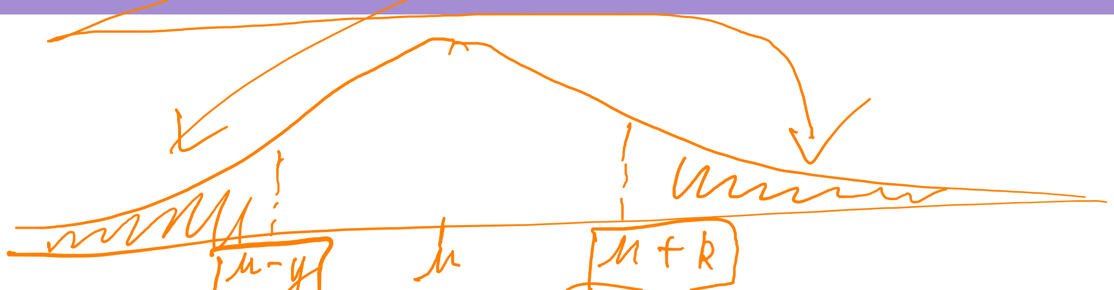
# More Assumptions → Better Guarantee

$$e^{\left(-\frac{\delta^2 \mu}{3}\right)} = \exp\left(-\frac{\delta^2 \mu}{3}\right)$$

## (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 $0 \le \delta \le 1$

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



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

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

# Right Tail

$0.6 \times 1000$

$1 + \frac{0.1}{0.6} = \frac{0.7}{\cancel{0.6}} \times \cancel{0.6} \times 1000$

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 $\mathbb{P}\left(\frac{X}{1000} \geq 0.7\right) = \mathbb{P}(X \geq 0.7 \cdot 1000)$

$= \mathbb{P}\left(X \geq \left(1 + \boxed{\frac{0.1}{0.6}}\right) \cdot (0.6 \cdot 1000)\right)$

So $\delta = \frac{1}{6}$ and $\mu = 0.6 \cdot 1000$

$\mathbb{P}(X \geq 700) \leq \exp\left(-\frac{\frac{1}{6^2} \cdot 0.6 \cdot 1000}{3}\right)$

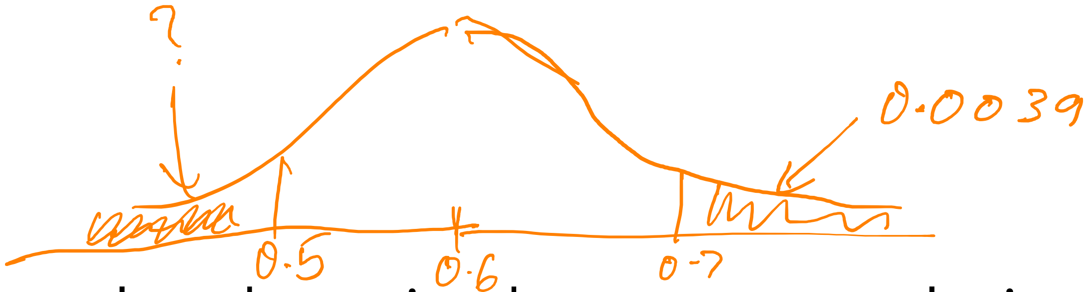
$\leq 0.0039$

## 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 $0 \leq \delta \leq 1$

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

# 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?

$\mathbb{P}(X \leq 0.5 \times 1000)$

$\mathbb{P}\left(X \leq \left(1 - \frac{0.1}{0.6}\right) \cdot 0.6 \cdot 1000\right)$

## Chernoff Bound (left 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 $0 \leq \delta \leq 1$

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

Fill out the poll everywhere so Kushal knows how long to explain
Go to pollev.com/cse312su21

# 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 $\mathbb{P}\left(\frac{X}{1000} \leq 0.5\right) = \mathbb{P}(X \leq 0.5 \cdot 1000)$

$= \mathbb{P}\left(X \leq \left(1 - \frac{0.1}{0.6}\right) \cdot (0.6 \cdot 1000)\right)$

So $\delta = \frac{1}{6}$ and $\mu = 0.6 \cdot 1000$

$\mathbb{P}(X \leq 500) \leq \exp\left(-\frac{\frac{1}{6^2} \cdot 0.6 \cdot 1000}{\boxed{2}}\right)$

$\leq 0.0003$

## Chernoff Bound (left 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 $0 \leq \delta \leq 1$

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

# 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)

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

$$\leq 0.0039 + 0.0003 = 0.0042$$

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

2.4%

# 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 $\delta$'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 $\binom{20000}{10000} \cdot 0.51^{10000} \cdot 0.49^{10000}$ is fraught with chances for floating point error and other issues. Chernoff is much better.

# But Wait! There's More

$P(X \geq 5)$

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.

$(1+\delta)\cdot\mu$ $(1-\delta)\mu$

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

## Markov's Inequality

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

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

## Chebyshev's Inequality

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

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

## (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 $0 \leq \delta \leq 1$

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

# One More Bound

**Union Bound**

For any events $E, F$

$\mathbb{P}(\boldsymbol{E} \cup \boldsymbol{F}) \leq \mathbb{P}(\boldsymbol{E}) + \mathbb{P}(\boldsymbol{F})$

Proof? $\boxed{\mathbb{P}(E \cup F)} = \mathbb{P}(E) + \mathbb{P}(F) \boxed{- \mathbb{P}(E \cap F)}$

And $\mathbb{P}(E \cap F) \geq 0$.

# Concentration Applications

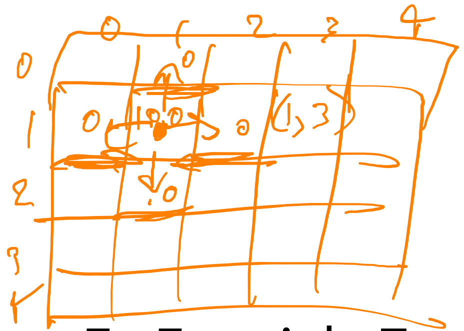
A common pattern:

Figure out "what could possibly go wrong" – often these are dependent.

Use a concentration inequality for each of the things that could go wrong.

Union bound over everything that could go wrong.

# Frogs

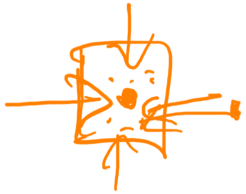



There are 20 frogs on each location in a 5x5 grid. Each frog will independently jump to the left, right, up, down, or stay where it is with equal probability. A frog at an edge of the grid magically warps to the corresponding edge (Pac-Man style).

Bound the probability that at least one square ends up with at least 36 frogs.

These events are dependent – adjacent squares affect each other!

# Frogs



$E[X]$

For an arbitrary location:

There are 100 frogs who could end up there (those above, below, left, right, and at that location). Each with probability $0.2$. Let $X$ be the number that land at the location we're interested in.

$0.2 \times 100 = 20$

$$\mathbb{P}(X \geq 36) = \mathbb{P}(X \geq (1+\delta)20) \leq \exp\left(-\frac{\left(\frac{4}{5}\right)^2 \cdot 20}{3}\right) \leq 0.015$$

There are 25 locations. Since all locations are symmetric, by the union bound the probability of at least one location having 36 or more frogs is at most $25 \cdot 0.015 \leq 0.375$.

# 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.