More Tail Bounds

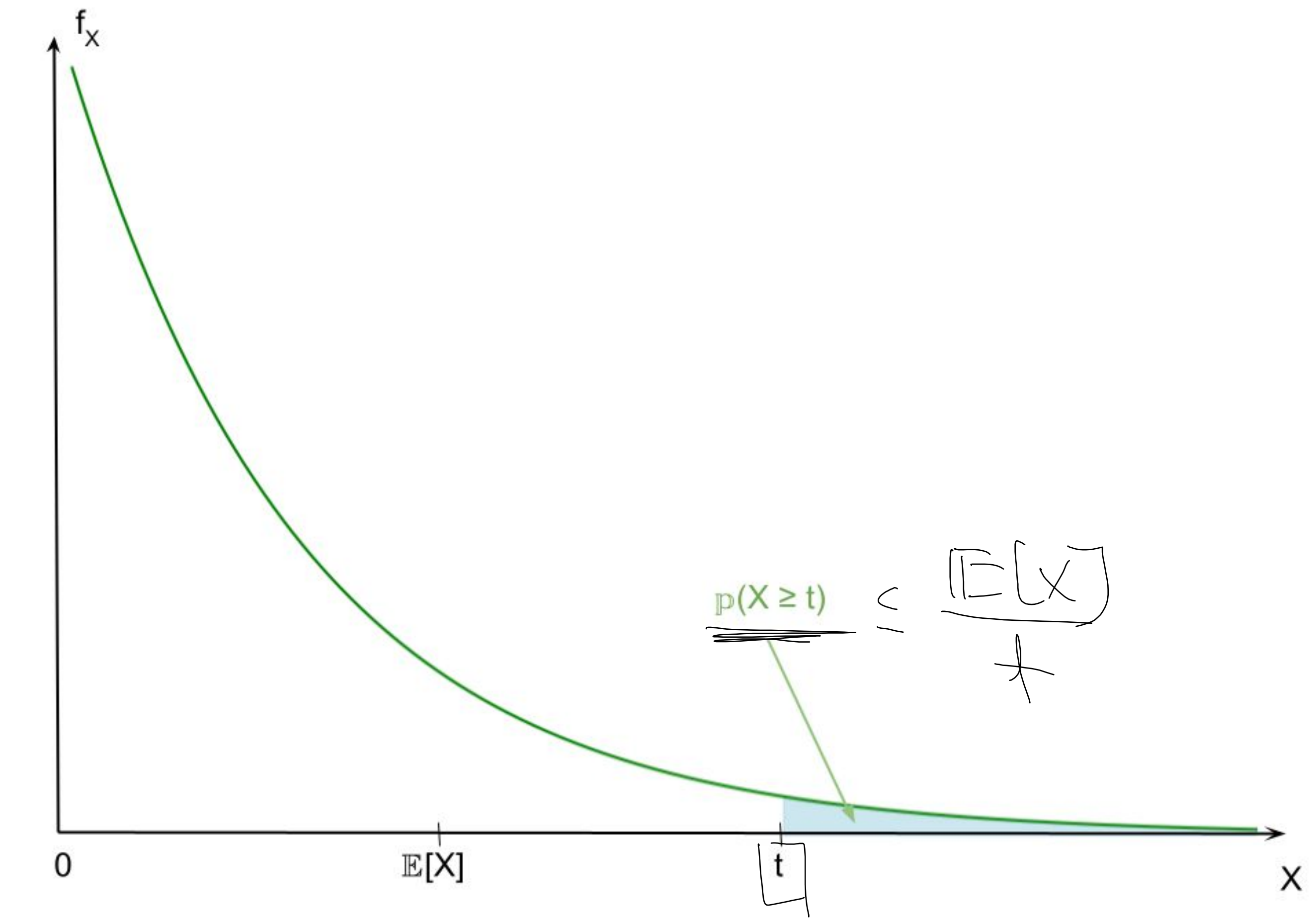
CSE 312 Spring 25
Lecture 22

Recap(Markov's Inequality)

Markov's Inequality

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

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

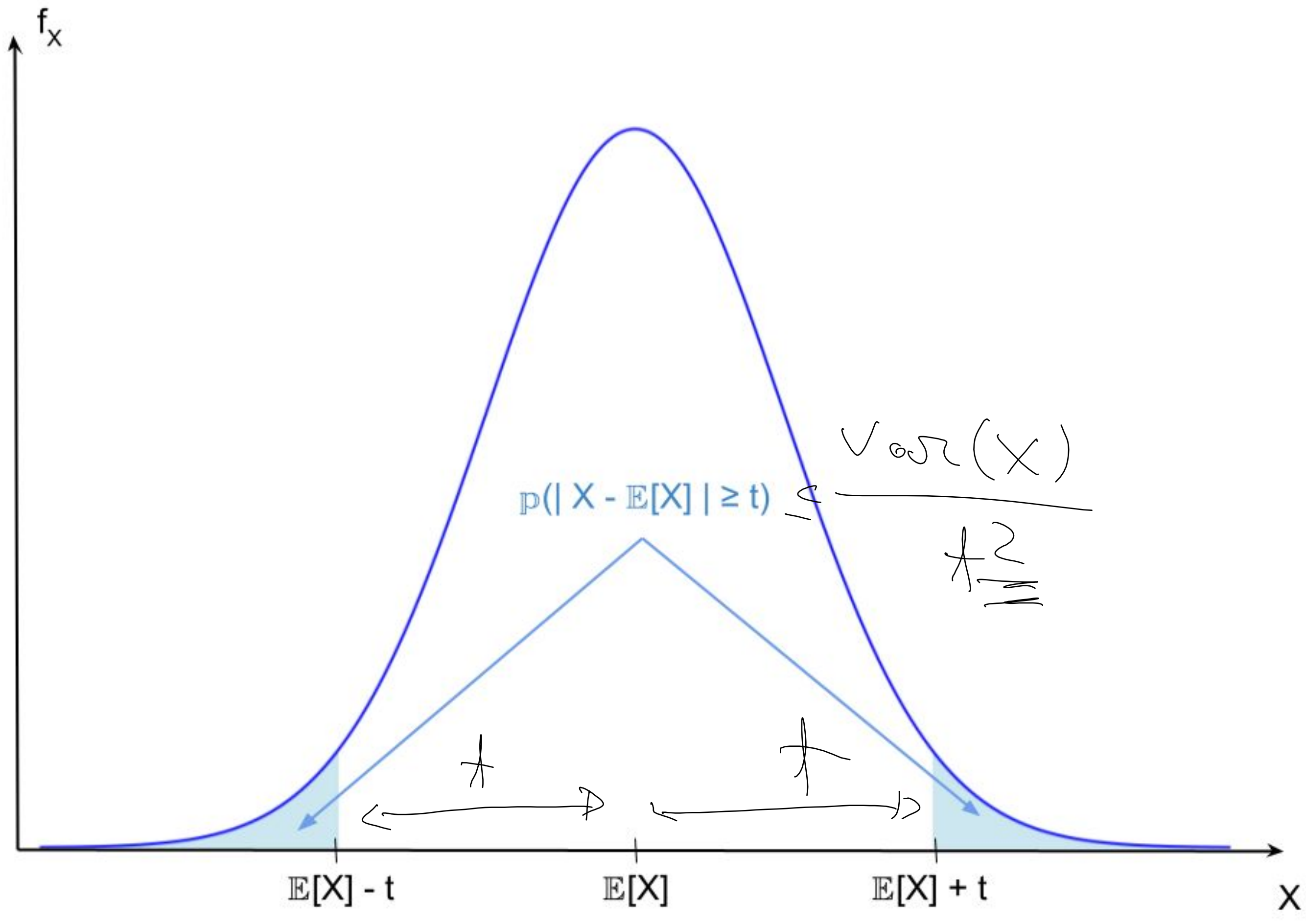


Recap(Chebyshev's Inequality)

Chebyshev's Inequality

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

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



Recap(Markov's and Chebyshev's)

Suppose you run a poll of $\mathbf{1000}$ 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?

$X_i \sim Ber(0.6)$

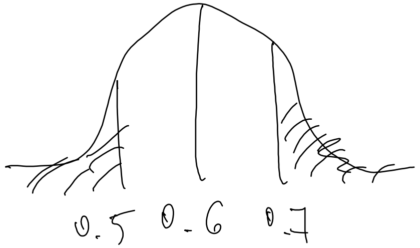
$\bar{X} = \sum X_i / 1000$

$\mathbb{E}[\bar{X}] = \mathbb{E}\left[\frac{1}{1000}\sum_{i=1}^{1000} X_i\right] = \frac{1}{1000}\sum_{i=1}^{1000}\mathbb{E}[X_i] = \frac{1}{\cancel{1000}}\cancel{1000}\cdot 0.6 = 0.6$

$\mathrm{Var}(\bar{X}) = Var\left[\frac{1}{1000}\sum_{i=1}^{1000} X_i\right] = \frac{1}{1000^2} Var\left[\sum_{i=1}^{1000} X_i\right]$

$\overset{indep}{=} \frac{1}{1000^2}\sum_{i=1}^{1000} Var(X_i) = \frac{1}{1000^{\cancel{2}}}\cdot \cancel{1000}\cdot 0.4 \cdot 0.6 =$

Recap(Markov's and Chebyshev's)



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?

$\bar{Y} = 1 - \bar{X} \iff \bar{X} = 1 - \bar{Y}$ $\qquad \mathbb{E}[\bar{Y}] = 1 - \mathbb{E}[\bar{X}] = 0.4$

$\bar{Y}$ = pct. of voters not voting for us

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

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

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

$$\mathbb{P}(\bar{X} \geq .7) \leq \frac{.6}{.7} = \frac{6}{7}$$

$$\mathbb{P}(\bar{X} \leq .5) = \mathbb{P}(1 - \bar{Y} \leq .5) = \mathbb{P}(1 - .5 \leq \bar{Y})$$

$$= \mathbb{P}(0.5 \leq \bar{Y}) = \mathbb{P}(\bar{Y} \geq 0.5) \leq \frac{0.4}{0.5} = \frac{4}{5}$$

Markov's Inequality

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

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

$$\mathbb{P}(X \leq .5 \cup X \geq .7)$$
$$= \mathbb{P}(X \leq .5) + \mathbb{P}(X \geq .7) \leq \frac{6}{7} + \frac{4}{5}$$

Recap(Markov's and Chebyshev's)

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?

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

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

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

$\mathrm{Var}[X]$

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

t

Chebyshev's Inequality

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

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

Takeaway

Chebyshev gets more powerful as the variance shrinks.

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

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \qquad X_i \text{ indep}$$

$$\mathrm{Var}[\bar{X}] = \frac{1}{\cancel{n^2}} \cancel{n} \cdot \mathrm{Var}(X_i) = \frac{\mathrm{Var}(X_i)}{n}$$

Chernoff Bound

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 $0 \leq \delta \leq 1$

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

Chernoff Bound

$\mathbb{P}(X \geq t) = \mathbb{P}(e^X \geq e^t)$

Chernoff Bound (left tail)

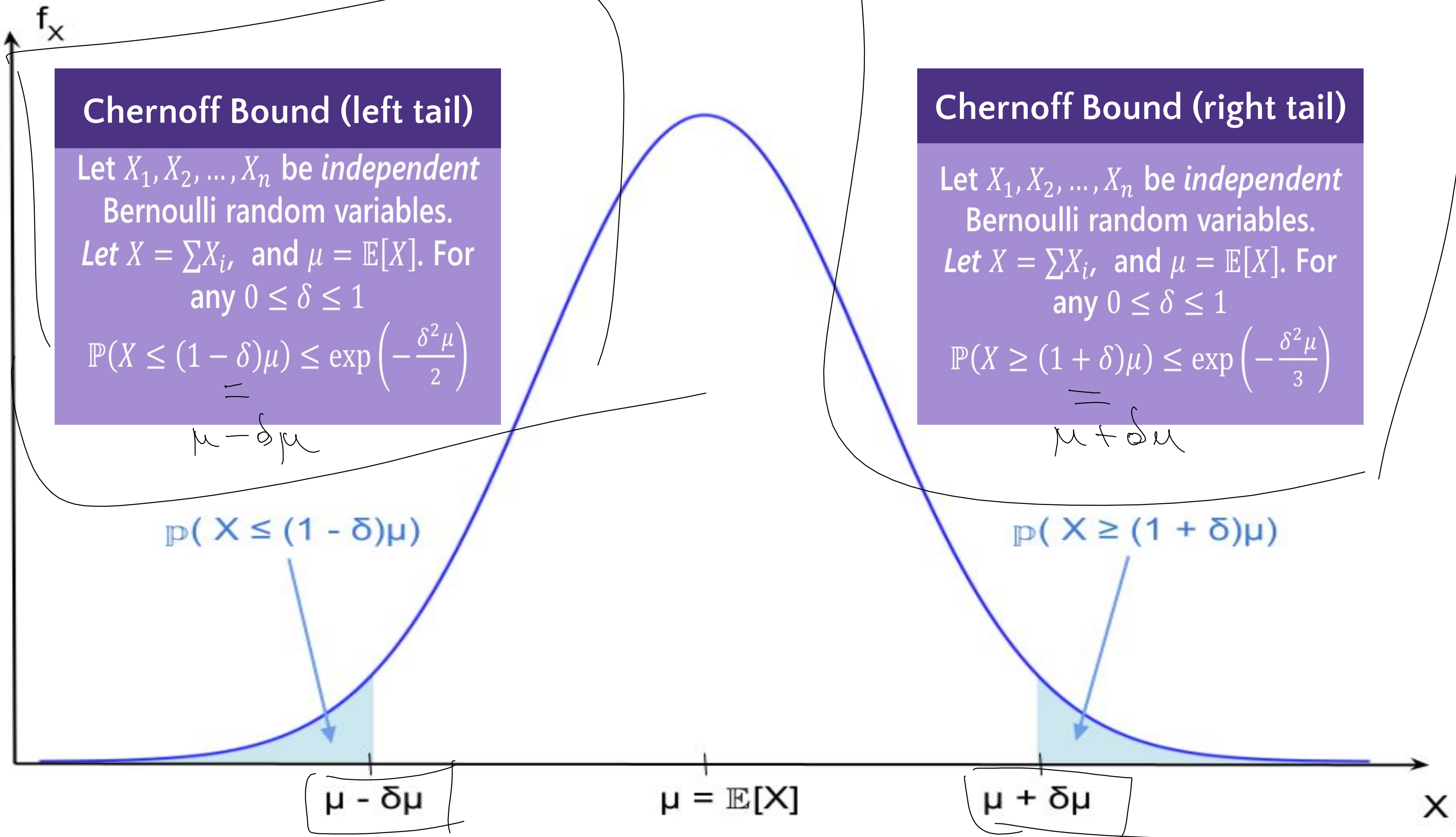
Let $X_1, X_2, \dots, X_n$ be *independent* Bernoulli random variables. *Let* $X = \sum X_i$, and $\mu = \mathbb{E}[X]$. For any $0 \leq \delta \leq 1$

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

Chernoff Bound (right 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 $0 \leq \delta \leq 1$

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



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?

Right 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?

$X = \sum_{i=1}^{1000} X_i$

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

$= \mathbb{P}(X \geq 700) =$

μ

$= \mathbb{P}\left(X \geq \left(1 + \frac{1}{6}\right) 600\right)$

δ

$\leq \exp\left(- \frac{(1/6)^2\, 600}{3}\right)$

$(1+\delta)\mu = 700$

$(1+\delta)600 = 700$

$(1+\delta) = \frac{700}{600}$

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

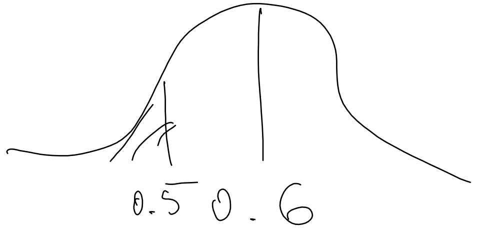
$\boxed{\delta = \frac{1}{6}}$

Chernoff Bound (right 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 $0 \leq \delta \leq 1$

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

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?

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

$= \mathbb{P}(X \leq 500)$

δ μ

$= \mathbb{P}\left(X \leq \left(1 - \frac{1}{6}\right) 600\right)$

$\leq \exp\left(-\frac{(1/6)^2 600}{2}\right)$

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 $0 \leq \delta \leq 1$

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

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!

Well if all the X_i have the same probability. It does work if they're independent but have different distributions. But there's bigger reasons to care…

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 $\boxed{\binom{20000}{10000}} \cdot \boxed{.51^{10000}} \cdot \boxed{.49^{10000}}$ 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

There's a whole book of these!

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.

Union Bound

Union Bound

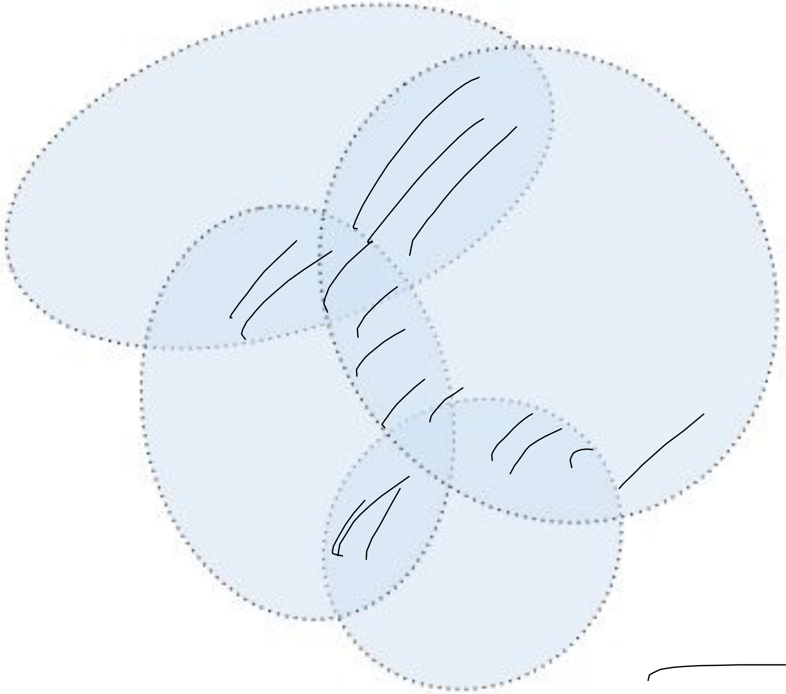
For any events E, F

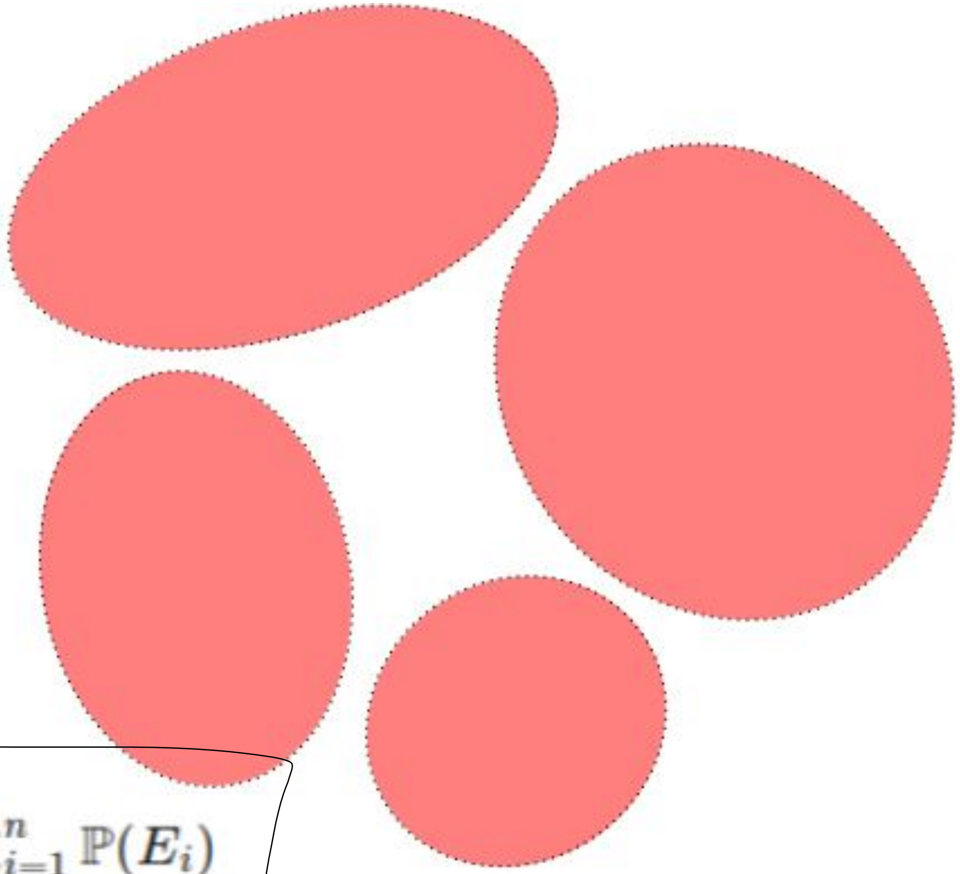
$$\mathbb{P}(E \cup F) \leq \mathbb{P}(E) + \mathbb{P}(F)$$

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

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

Union Bound





$$\mathbb{P}\left(\bigcup_{i=1}^{n} E_i\right) \le \sum_{i=1}^{n} \mathbb{P}(E_i)$$

Concentration Applications

A common pattern:

1. $E = \bigcup_{i=1}^{n} E_i$

2. $\mathbb{P}(E) = \mathbb{P}\left(\bigcup_{i=1}^{n} E_i\right) \leq \sum_{i=1}^{n} \mathbb{P}(E_i)$

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.

3. Bound $\mathbb{P}(E_i)$ by Markov
Cheby.
Chernoff.

Frogs

E = at least one square has ≥ 36 frogs

$E = \bigcup_{i=1}^{25} E_i$ ← square i has ≥ 36 frogs

25 squares

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).

1. $\mathbb{P}(E) = \mathbb{P}\left(\bigcup_{i=1}^{25} E_i\right) \leq \sum_{i=1}^{25} \mathbb{P}(E_i) \leq 25 \cdot b$

Union

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

F_i = # of frogs on square i

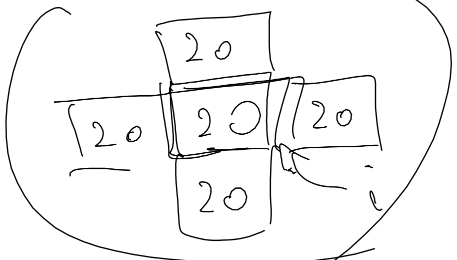
$F_i = \sum_{j=1}^{100} F_{ij}$ ← frog j ends up on square i

$\mathbb{P}(E_i) = \mathbb{P}(F_i \geq 36) =$

$= \mathbb{P}\left(F_i \geq \left(1 + \frac{16}{20}\right) 20\right)$

δ, μ

These events are dependent – adjacent squares affect each other!



$\mathbb{E}[F_i] = \sum_{j=1}^{100} \mathbb{E}[F_{ij}] = 100 \cdot \frac{1}{5} = 20$

$F_{ij} \sim \mathrm{Ber}(1/5)$

$\leq \exp\left(\frac{-\left(16/20\right)^2 20}{3}\right) = b$

Frogs

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 $.2$. Let X be the number that land at the location we're interested in.

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

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.

Tail Bounds – Summary

	Markov's Inequality	Chebyshev's Inequality	Chernoff Bound	Union Bound
Random Variable Type	Any, but positive	Any	$X = \sum_{i=1}^{n} X_i$ where X_i are independent Bernoulli	Any
Bound Type	Right tail	Both tails around the mean	Either left or right tail around the mean	Any union of events
Parameters Needed	$\mathbb{E}[X]$, t = start point of right tail	$\mathbb{E}[X], \mathrm{Var}(X)$, t = distance from mean where tails start	$\mu = \mathbb{E}[X]$, $\delta \in [0, 1]$ with $(1-\delta)\mu$ and $(1+\delta)\mu$ the starts of the left and right tails	Probabilities of E_i's
Bound	$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}[X]}{t}$	$\mathbb{P}(\lVert X - \mathbb{E}[X] \rVert \geq t) \leq \frac{Var(X)}{t^2}$	**Left**: $\mathbb{P}(X \leq (1-\delta)\mu) \leq \exp(-\frac{\delta^2 \mu}{2})$ **Right**: $\mathbb{P}(X \geq (1+\delta)\mu) \leq \exp(-\frac{\delta^2 \mu}{3})$	$\mathbb{P}\left(\bigcup_{i=1}^{n} E_i\right) \leq \sum_{i=1}^{n} \mathbb{P}(E_i)$
Comment	Sucks most of the time	Good when variance is small(e.g. mean of a large sample)	X_i **need not** have the same p. Great for large n	E_i are usually events we want to avoid with small $\mathbb{P}(E_i)$