Tail Bounds

CSE 312 24Su

Lecture 17

Announcements

> Reminder about concept checks 14, 15, 16 late due date tonight

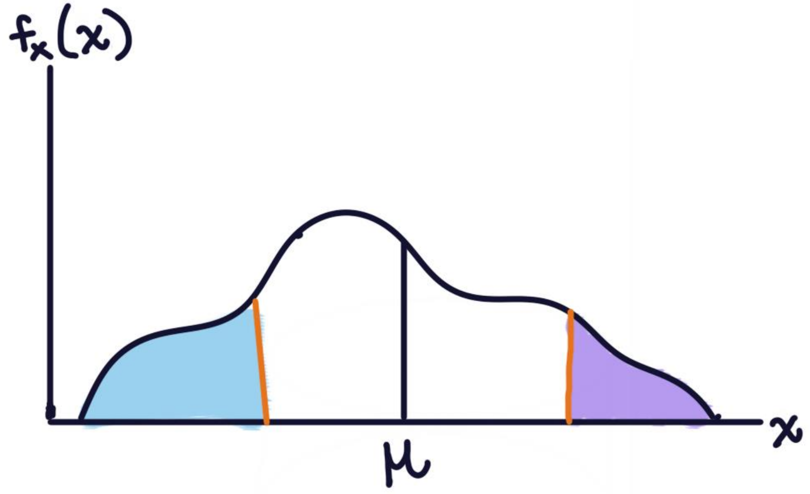
> Review/breather lecture on Wednesday, fill out form for requested topics/questions

What's a Tail Bound?

A **tail bound** (or concentration inequality) bounds the probability in the "tails" of the distribution. e.g., statements like $\mathbb{P}(X \geq 4) \leq 0.8$, $\mathbb{P}(X \geq 4) \leq 0.8$

We've seen this before! We can:

- Compute these probabilities exactly in some cases
- Approximate X as normal using CLT if X is the sum of a bunch of i.i.d random variables



But what if we barely know anything about X and it doesn't fit into the frameworks we've learned about? Can we still make some tail bound guarantees?

Upper vs. Lower Bound

If we find something like $\mathbb{P}(A) \leq b$, we found an upper bound

This *highest/"uppermost"* value the probability of A could be is b

If we find something like $\mathbb{P}(A) \geq b$, we found a lower bound

This *lowest/smallest* value the probability of A could be is b

Tail Bounds

We're going to learn about 3 tail bounds that we can use when all we know about X is it's expected value and/or variance:

- Markov's Inequality
- Chebyshev's Inequality
- Chernoff Bound

And....

- The **union bound** is not a *tail* bound, but we'll still talk about it :)

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 $t > 0$

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

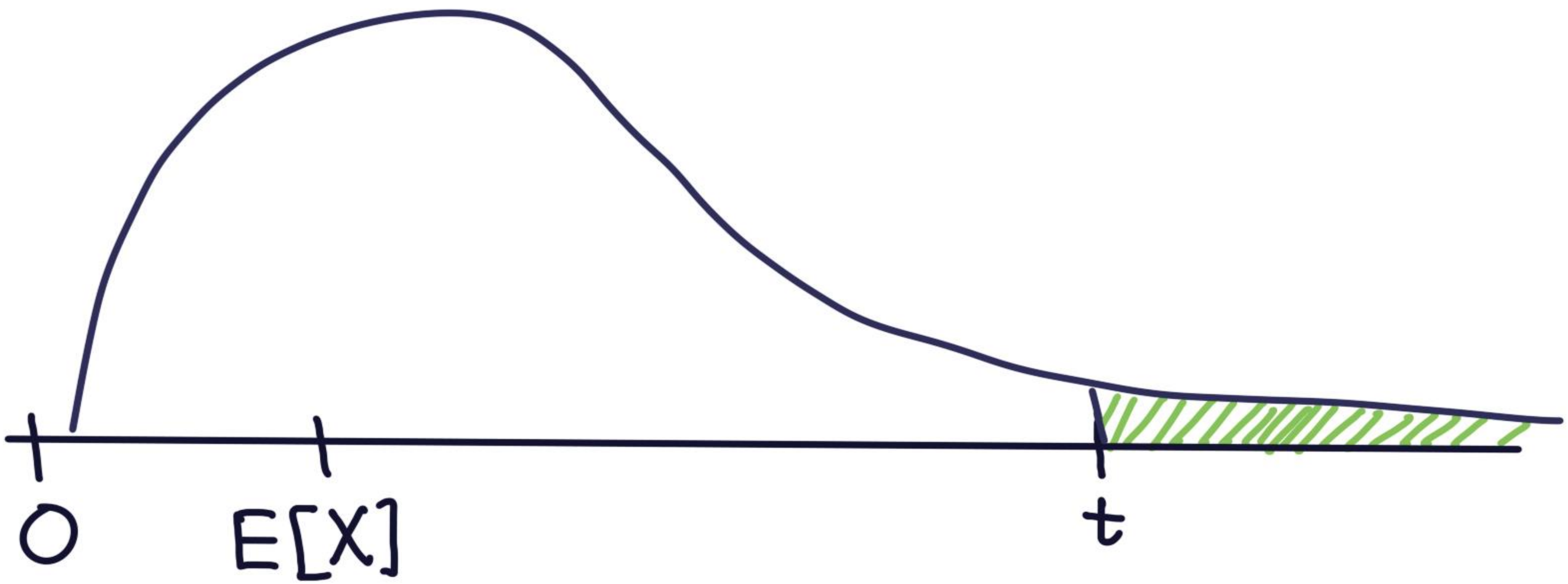
Markov's Inequality

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

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

Requirements:

1. X must be non-negative
2. We know the expectation of X


0
E[X]
t

Proof

$$\mathbb{E}[X] = \mathbb{E}[X|X < t]\mathbb{P}(X < t) + \mathbb{E}[X|X \geq t]\mathbb{P}(X \geq t)$$

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

$\mathbb{E}[X|X \geq t]\mathbb{P}(X \geq t) \geq 0$ if X is non-negative

$$\geq t \cdot \mathbb{P}(X \geq t)$$

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

Doing some algebra…we get exactly what's in Markov's inequality! →

Markov's Inequality

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

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

Example: Let's see how good this bound is…

Suppose you roll a fair (6-sided) die until you see a 6. Let X be the number of rolls. Bound the probability that $X \geq 12$.

$X \sim \text{Geo}\left(\frac{1}{6}\right)$, so $\mathbb{E}[X] = 1/(\frac{1}{6}) = 6$

Applying Markov's Inequality…

$\mathbb{P}(X \geq 12)$

Markov's Inequality

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

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

Example: Let's see how good this bound is…

Suppose you roll a fair (6-sided) die until you see a 6. Let X be the number of rolls. Bound the probability that $X \geq 12$.

$X \sim \text{Geo}\left(\frac{1}{6}\right)$, so $\mathbb{E}[X] = 1/(\frac{1}{6}) = 6$

Applying Markov's Inequality…

$$\mathbb{P}(X \geq 12) \leq \frac{\mathbb{E}[X]}{12} = \frac{6}{12} = \frac{1}{2}$$

Markov's Inequality

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

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

Example: Let's see how good this bound is…

Suppose you roll a fair (6-sided) die until you see a 6. Let X be the number of rolls. Bound the probability that $X \geq 12$.

$X \sim \text{Geo}\left(\frac{1}{6}\right)$, so $\mathbb{E}[X] = 1/(\frac{1}{6}) = 6$

Applying Markov's Inequality…

$$\mathbb{P}(X \geq 12) \leq \frac{\mathbb{E}[X]}{12} = \frac{6}{12} = \frac{1}{2}$$

Exact probability?

$$1 - \mathbb{P}(X < 12) \approx 1 - 0.865 = 0.135$$

Markov's Inequality

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

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

Example: Ads

Suppose the average number of ads you see on a website is 25. Give an upper bound on the probability of seeing a website with 75 or more ads.

Markov's Inequality

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

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

Example: Ads

Suppose the average number of ads you see on a website is 25. Give an upper bound on the probability of seeing a website with 75 or more ads.

$$\mathbb{P}(X \geq 75) \leq \frac{\mathbb{E}[X]}{75} = \frac{25}{75} = \frac{1}{3}$$

Markov's Inequality

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

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

Example: More Ads

Suppose the average number of ads you see on a website is 25. Give an upper bound on the probability of seeing a website with 20 or more ads.

Markov's Inequality

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

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

Example: More Ads

Suppose the average number of ads you see on a website is 25. Give an upper bound on the probability of seeing a website with 20 or more ads.

$$\mathbb{P}(X \geq 20) \leq \frac{\mathbb{E}[X]}{20} = \frac{25}{20} = 1.25$$

Well, that's…true. Technically.

But without more information we couldn't hope to do much better. What if every page gives exactly 25 ads? Then the probability really is 1.

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?

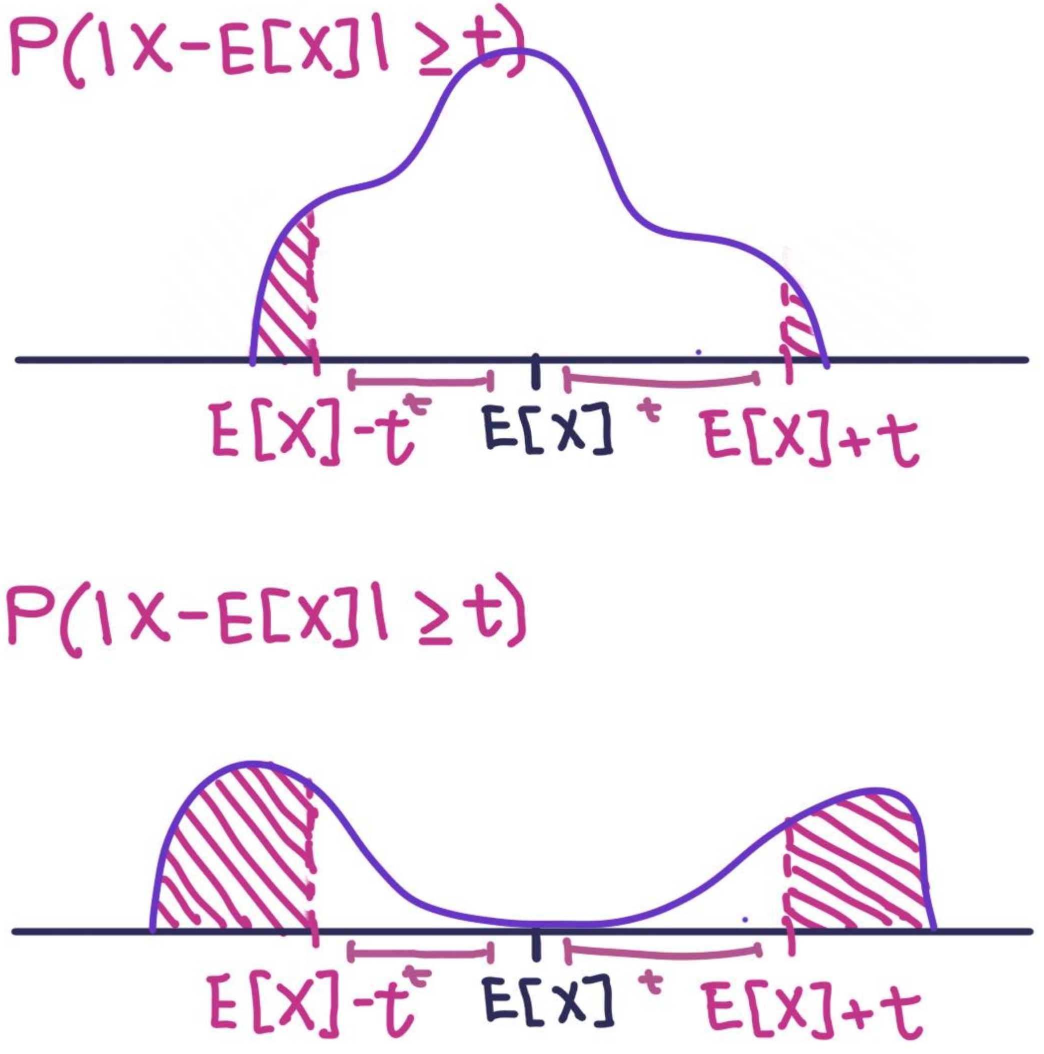
The variance!

Chebyshev's Inequality

Chebyshev's Inequality

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

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



Requirements:

1. We know the *expectation* of X
2. We know the *variance* of X

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 $k > 0$

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

Requirements:

1. We know the *expectation* of X
2. We know the *variance* of Xc

"probability we're at least k standard deviations away from the mean is $\leq \frac{1}{k^2}$"

Markov's inequality gave: $\mathbb{P}(X \geq 75) \leq 1/3$

Example: Ads (but better!)

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 75 or more ads.

Chebyshev's Inequality

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

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

Markov's inequality gave: $\mathbb{P}(X \geq 75) \leq 1/3$

Example: Ads (but better!)

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 75 or more ads.

$$\mathbb{P}(X \geq 75) = \mathbb{P}(X - 25 \geq 75 - 25) \leq \mathbb{P}(|X - 25| \geq 50) \leq \frac{16}{50^2}$$

Chebyshev's Inequality

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

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

Markov's inequality gave: $\mathbb{P}(X \geq 12) \leq 0.5$

Example: 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 $X \geq 12$

$$X \sim \text{Geo}\left(\frac{1}{6}\right), \mathbb{E}[X] = 6,\ \text{Var}(X) = \frac{1-\left(\frac{1}{6}\right)}{\left(\frac{1}{6}\right)^2} = \frac{\frac{5}{6}}{\frac{1}{36}} = 30$$

Chebyshev's Inequality

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

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

Fill out the poll everywhere:
pollev.com/cse312

Markov's inequality gave: $\mathbb{P}(X \geq 12) \leq 0.5$

Example: 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 $X \geq 12$

$$X \sim \text{Geo}\left(\frac{1}{6}\right), \mathbb{E}[X] = 6,\ \text{Var}(X) = \frac{1-\left(\frac{1}{6}\right)}{\left(\frac{1}{6}\right)^2} = \frac{\frac{5}{6}}{\frac{1}{36}} = 30$$

Chebyshev's Inequality

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

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

Markov's inequality gave: $\mathbb{P}(X \geq 12) \leq 0.5$

Example: 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 $X \geq 12$

$$\mathbb{P}(X \geq 12) = \mathbb{P}(X - 6 \geq 12 - 6) \leq \mathbb{P}(|X - 6| \geq 6) \leq \frac{30}{6^2} = \frac{5}{6}$$

Not any better than Markov ☹

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: Geometric RV (generalized)

Let X be a geometric rv with parameter p

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

With Chebyshev's…

$$\mathbb{E}[X] = \frac{1}{p}, \mathbb{P}\left(X \geq \frac{2}{p}\right) = \mathbb{P}\left(X - \frac{1}{p} \geq \frac{1}{p}\right) \leq \mathbb{P}\left(\left|X - \frac{1}{p}\right| \geq \frac{1}{p}\right) \leq \frac{\frac{1-p}{p^2}}{1/p^2} = 1 - p$$

With Markov's…

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

For large p, Chebyshev is better.

Chebyshev's Inequality

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

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

Example: 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?

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

$X_i \sim \mathrm{Ber}(0.6)$, so $\mathbb{E}[X_i] = 0.6$, $\mathrm{Var}(X_i) = 0.6 \cdot 0.4$

$\mathbb{E}[\bar{X}] = \frac{1}{1000}(1000 \cdot 0.6) = 0.6$

$\mathrm{Var}(\bar{X}) = \frac{1}{1000^2}(1000 \cdot (0.6 \cdot 0.4))$

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: 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?

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

$X_i \sim \text{Ber}(0.6)$, so $\mathbb{E}[X_i] = 0.6$, $\text{Var}(X_i) = 0.6 \cdot 0.4$

$\mathbb{E}[\bar{X}] = \frac{1}{1000}(1000 \cdot 0.6) = 0.6$

$\text{Var}(\bar{X}) = \frac{1}{1000^2}(1000 \cdot (0.6 \cdot 0.4))$

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

Chebyshev's Inequality

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

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

Example: Near the mean

Suppose you run a poll of $\boldsymbol{n}$ 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?

$\bar{X} = \sum X_i / \boldsymbol{n}$.

$X_i \sim \text{Ber}(0.6)$, so $\mathbb{E}[X_i] = 0.6$, $\text{Var}(X_i) = 0.6 \cdot 0.4$

$$\mathbb{E}[\bar{X}] = \frac{1}{\boldsymbol{n}}(\boldsymbol{n} \cdot 0.6) = 0.6$$

$$\text{Var}(\bar{X}) = \frac{1}{\boldsymbol{n}^2}\big(\boldsymbol{n} \cdot (0.6 \cdot 0.4)\big) = \frac{0.24}{\boldsymbol{n}}$$

$$\mathbb{P}(|\bar{X} - \mathbb{E}[\bar{X}]| \geq .1) \leq \frac{0.24/n}{.1^2} = \frac{0.24}{\boldsymbol{n} \cdot .1^2}$$

Chebyshev's Inequality

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

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

When is Chebyshev’s better than Markov’s?

> Chebyshev gets more powerful as the variance shrinks

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

What do we mean by a “better” bound?

When is Chebyshev's better than Markov's?

> Chebyshev gets more powerful as the variance shrinks

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

What do we mean by a "better" bound?

> it gives us *more* information

> a tighter bound is one that restricts the possible probabilities more

> e.g., $P(X \geq 4) \leq 0.2$ is a tighter bound than $P(X \geq 4) \leq 0.4$ and $P(X \leq 4) \geq 0.8$ is a tighter bound than $P(X \leq 4) \geq 0.6$

Chernoff Bound

(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 $0 \leq \delta \leq 1$

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

LEFT TAIL *RIGHT TAIL*

Requirements:

1. X is a sum of independent Bernoulli random variables.
2. We know $\mathbb{E}[X]$

Chernoff Bound

(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 $0 \leq \delta \leq 1$

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

LEFT TAIL *RIGHT TAIL*

Wait a second…why do we need a bound for this??

Is X binomial?

Chernoff Bound

(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 $0 \le \delta \le 1$

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

LEFT TAIL *RIGHT TAIL*

Wait a second…why do we need a bound for this??

Is X binomial? No! Because the X_i's might not be identically distributed. But even if X is binomial, with REALLY large values of n, computing exact probabilities is computationally expensive.

Chernoff Bound

(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 $0 \le \delta \le 1$

$$\mathbb{P}(X \le (1-\delta)\mu) \le e^{\left(-\frac{\delta^2 \mu}{2}\right)} \quad \textit{and} \quad \mathbb{P}(X \ge (1+\delta)\mu) \le e^{\left(-\frac{\delta^2 \mu}{3}\right)}$$

LEFT TAIL *RIGHT TAIL*

Wait a second…why do we need a bound for this??

Can we use CLT? Yes, but remember that CLT is still an approximation. This bound will give us a definite upper bound for the probability.

Chernoff Bound

(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 $0 \leq \delta \leq 1$

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

LEFT TAIL — *RIGHT TAIL*

Requirements:

1. X is a sum of independent Bernoulli random variables.
2. We know $\mathbb{E}[X]$

Example: Polling (again, but better!)

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?

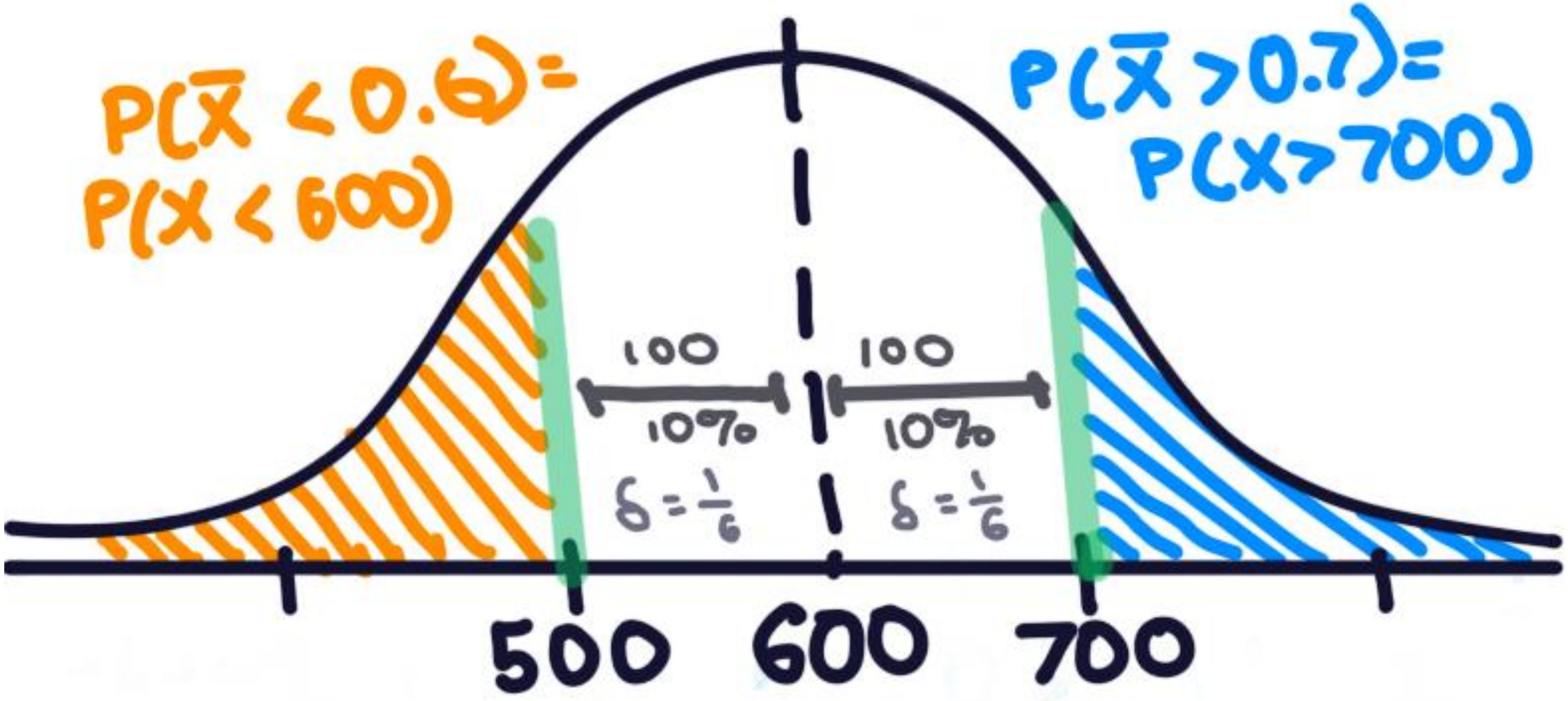
Goal: bound $\mathbb{P}(|\bar{X} - 0.6| \geq 0.1) = \mathbb{P}(\bar{X} \leq 0.5) + \mathbb{P}(\bar{X} \geq 0.7)$

(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** $0 \leq \delta \leq 1$

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

LEFT TAIL | *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?*

Goal: bound $\mathbb{P}(|\bar{X} - 0.6| \geq 0.1) = \mathbb{P}(\bar{X} \leq 0.5) + \mathbb{P}(\bar{X} \geq 0.7)$

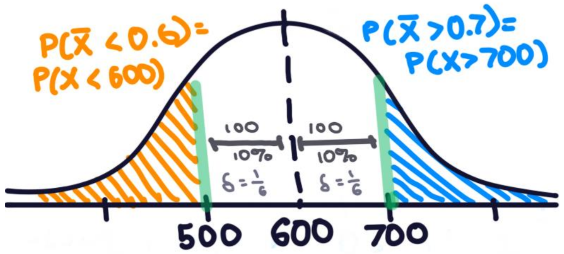
Goal: bound $\mathbb{P}(|\bar{X} - 0.6| \geq 0.1) = \mathbb{P}(\bar{X} \leq 0.5) + \mathbb{P}(\bar{X} \geq 0.7)$

Example: Polling (1. bound the 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?

$X = \sum X_i$, where $X_i \sim \text{Ber}(0.6), \mu = \mathbb{E}[X] = 1000 \cdot 0.6 = 600$

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



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 $0 \leq \delta \leq 1$

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

Goal: bound $\mathbb{P}(|\bar{X} - 0.6| \geq 0.1) = \mathbb{P}(\bar{X} \leq 0.5) + \mathbb{P}(\bar{X} \geq 0.7)$

Example: Polling (1. bound the 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?

$X = \sum X_i$, where $X_i \sim \mathrm{Ber}(0.6), \mu = \mathbb{E}[X] = 1000 \cdot 0.6 = 600$

$$\mathbb{P}\left(\frac{X}{1000} \leq 0.5\right) = \mathbb{P}(X \leq 500)$$

$$500 = (1 - \delta)600 \text{ -> } \delta = \frac{1}{6}$$

$$\ldots = \mathbb{P}\left(X \leq \left(1 - \frac{1}{6}\right)\mu\right) \leq e^{-\frac{\frac{1}{6^2} \cdot 600}{2}}$$

≈ 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 $0 \leq \delta \leq 1$

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

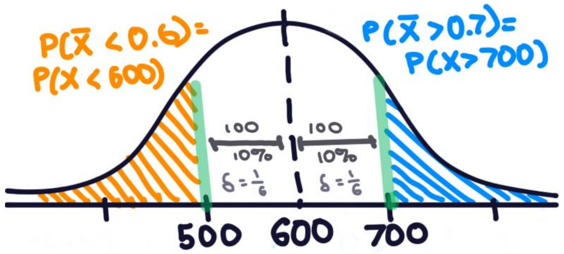
Goal: bound $\mathbb{P}(|\bar{X} - 0.6| \geq 0.1) = \mathbb{P}(\bar{X} \leq 0.5) + \mathbb{P}(\bar{X} \geq 0.7)$

Example: Polling (2. bound the 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?

$X = \sum X_i$, where $X_i \sim \text{Ber}(0.6), \mu = \mathbb{E}[X] = 1000 \cdot 0.6 = 600$

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



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 $0 \leq \delta \leq 1$

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

Goal: bound $\mathbb{P}(|\bar{X} - 0.6| \geq 0.1) = \mathbb{P}(\bar{X} \leq 0.5) + \mathbb{P}(\bar{X} \geq 0.7)$

Example: Polling (2. bound the 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?

$X = \sum X_i$, where $X_i \sim \text{Ber}(0.6), \mu = \mathbb{E}[X] = 1000 \cdot 0.6 = 600$

$$\mathbb{P}\left(\frac{X}{1000} \geq 0.7\right) = \mathbb{P}(X \geq 700)$$

$$700 = (1 + \delta)600 \text{ -> } \delta = \frac{1}{6}$$

$$\dots = \mathbb{P}\left(X \geq \left(1 + \frac{1}{6}\right)\mu\right) \leq e^{-\frac{\frac{1}{6^2} \cdot 600}{3}}$$

$$\approx 0.0039$$

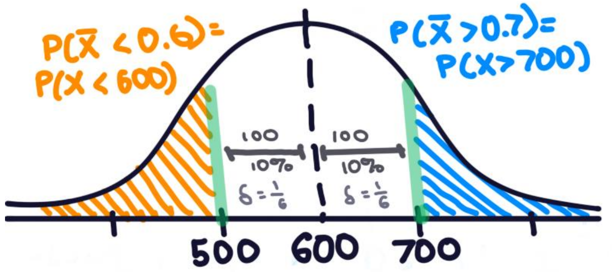
Chernoff Bound (right tail)

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 $0 \leq \delta \leq 1$

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

Example: Polling (3. Putting it all together)

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?



We want $\mathbb{P}(|\bar{X} - 0.6| \geq 0.1) = \mathbb{P}(\bar{X} \leq 0.5) + \mathbb{P}(\bar{X} \geq 0.7)$

We know..

-> $\mathbb{P}(\bar{X} \leq 0.5) = \mathbb{P}(X \leq 500) \leq 0.0003$ (from Chernoff bound, left tail)

-> $\mathbb{P}(\bar{X} \geq 0.7) = \mathbb{P}(X \geq 700) \leq 0.0039$ (from Chernoff bound, right tail)

So, $\mathbb{P}(|\bar{X} - 0.6| \geq 0.1) \leq 0.0003 + 0.0039 = 0.0042$

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

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

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.

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).

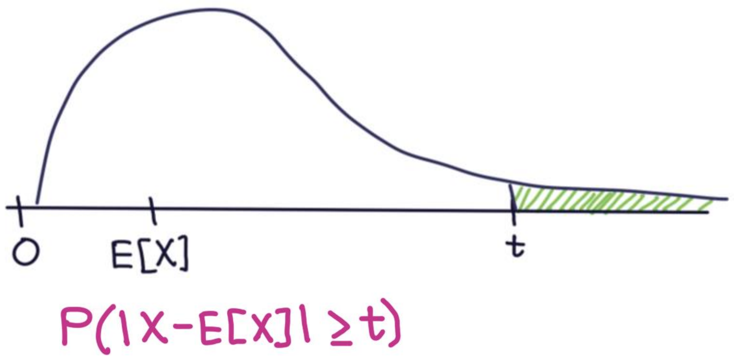
Choosing a minimum n for a poll – don’t need exact probability of failure, just to make sure it’s small.

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

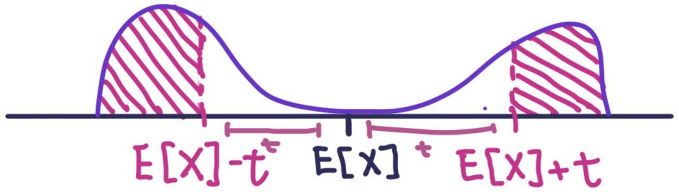
Learning more about the situation (e.g. learning variance instead of just mean, knowing bounds on the support of the starting variables) usually lets you get more accurate bounds.

Tail Bounds – Summary

- **Markov's inequality** - $\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}[X]}{t}$
  - Use if X is non-negative and we know the expectation
  - Useful when we don't know much about X

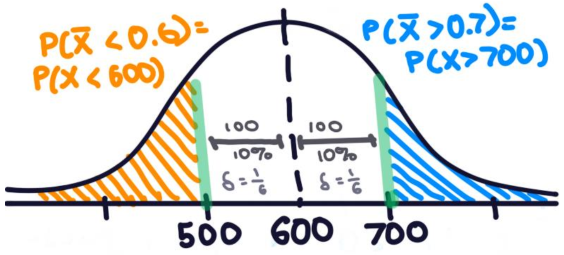


- **Chebyshev's inequality** - $\mathbb{P}(|X - \mathbb{E}[X]| \geq t) \leq \frac{\mathrm{Var(t)}}{t^2}$
  - Use if we know the expectation **and** variance of X
  - Gives better bounds with small variances



- **Chernoff Bound**

  $\mathbb{P}(X \leq (1-\delta)\mu) \leq e^{\left(-\frac{\delta^2 \mu}{2}\right)}$ *and* $\mathbb{P}(X \geq (1+\delta)\mu) \leq e^{\left(-\frac{\delta^2 \mu}{3}\right)}$
  - Use if X is a sum of independent Bernoulli random variables
  - Gives a very good bound usually, and is especially helpful when X is binomial and we can't easily computationally compute some summations/probability



One More Bound – Union Bound

Union Bound (not a *tail* bound, but still a bound)

Union Bound

For any events E, F

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

Sometimes we don't' have enough information to compute this probability exactly, so we use the union bound to bound that probability

Proof?

By *inclusion-exclusion*, $\mathbb{P}(E \cup F) = \mathbb{P}(E) + \mathbb{P}(F) - \mathbb{P}(E \cap F)$

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

Concentration Applications

A common pattern:

"What's the probability something goes wrong?"

> 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.

Example: 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 at least one square ends up with at least 36 frogs.

Example: 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 at least one square ends up with at least 36 frogs.

A_i is the event the i'th square has at least 36 frogs

$\mathbb{P}(A_1 \cup A_2 \cup A_3 \cup \cdots \cup A_{25})$
$\leq \mathbb{P}(A_1) + \mathbb{P}(A_2) + \mathbb{P}(A_3) + \cdots + \mathbb{P}(A_{25})$ *by the union bound*

How do we find $\mathbb{P}(\boldsymbol{A_i})$? Use another bound!

These events are dependent – adjacent squares affect each other!

Example: 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 at least one square ends up with at least 36 frogs.

A_i is the event the i'th square has at least 36 frogs

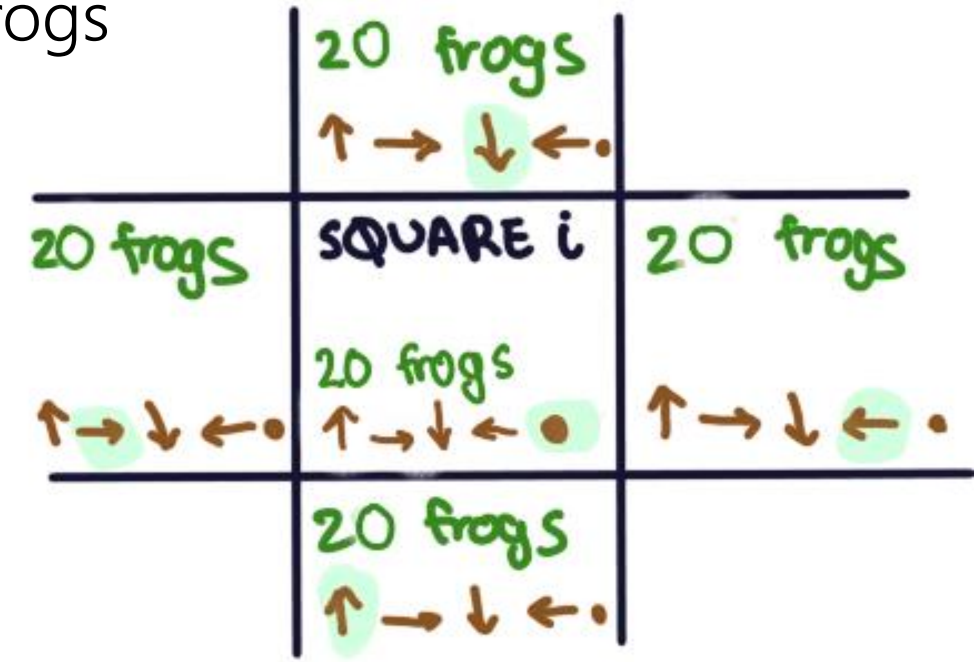
How do we find $\mathbb{P}(A_i)$? Use another bound!

Let Y be the number frogs in i'th square

$Y = \sum_{j=1}^{100} X_j$, $X_j \sim \text{Ber}(1/5)$, $E[Y] = \frac{100}{5} = 20$

$\mathbb{P}(A_i) = \mathbb{P}(Y \geq 36) = \mathbb{P}\left(Y \geq \left(1 + \frac{4}{5}\right) 20\right)$

$\leq e^{\left(-\frac{\left(\frac{4}{5}\right)^2 \cdot 20}{3}\right)} \leq 0.015$ *by the Chernoff bound*



Example: 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 at least one square ends up with at least 36 frogs.

A_i is the event the i'th square has at least 36 frogs

$$\mathbb{P}(A_1 \cup A_2 \cup A_3 \cup \cdots \cup A_{25})$$
$$\leq \mathbb{P}(A_1) + \mathbb{P}(A_2) + \mathbb{P}(A_3) + \cdots + \mathbb{P}(A_{25})$$ *by the union bound*
$$\leq 0.015 + 0.015 + 0.015 + \cdots + 0.015 = 25 \cdot 0.015$$ *by the Chernoff bound*
$$= 0.375$$

Example: 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 – Summary

- **Markov's inequality** - $\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}[X]}{t}$
  - Use if X is non-negative and we know the expectation
  - Useful when we don't know much about X
- **Chebyshev's inequality** - $\mathbb{P}(|X - \mathbb{E}[X]| \geq t) \leq \frac{\mathrm{Var(t)}}{t^2}$
  - Use if we know the expectation **and** variance of X
  - Gives better bounds with small variances
- **Chernoff Bound**
  $\mathbb{P}(X \leq (1-\delta)\mu) \leq e^{\left(-\frac{\delta^2\mu}{2}\right)}$ *and* $\mathbb{P}(X \geq (1+\delta)\mu) \leq e^{\left(-\frac{\delta^2\mu}{3}\right)}$
  - Use if X is a sum of independent Bernoulli random variables
  - Gives a very good bound usually, and is especially helpful when X is binomial and we can't easily computationally compute some summations/probability
- **Union Bound** - $\mathbb{P}(A \cup B) \leq \mathbb{P}(A) + \mathbb{P}(B)$
  - Use if we don't have enough information to find the union (e.g,. ways for at least of __ to occur, for A, or B, or C, or … to occur)