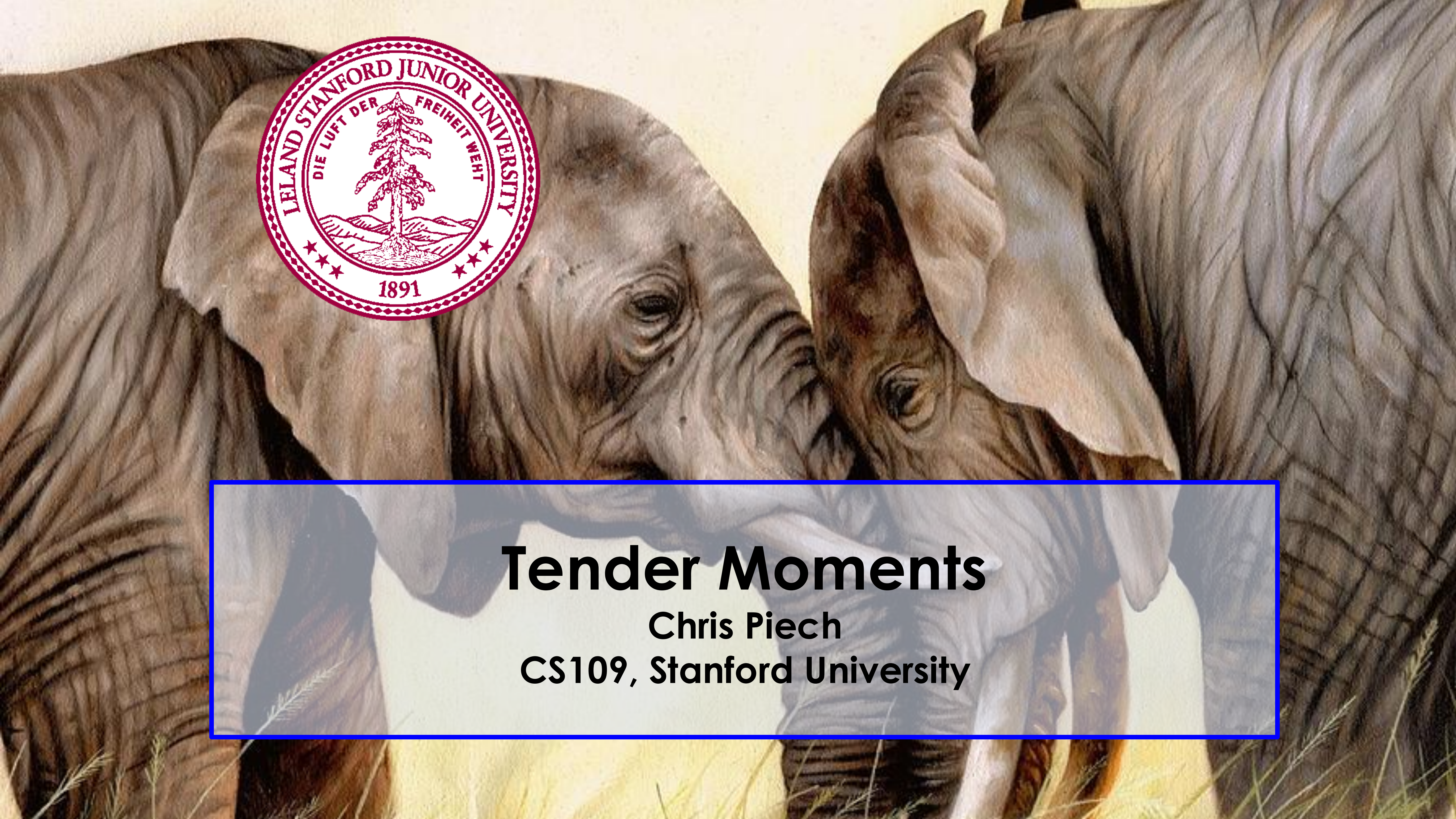
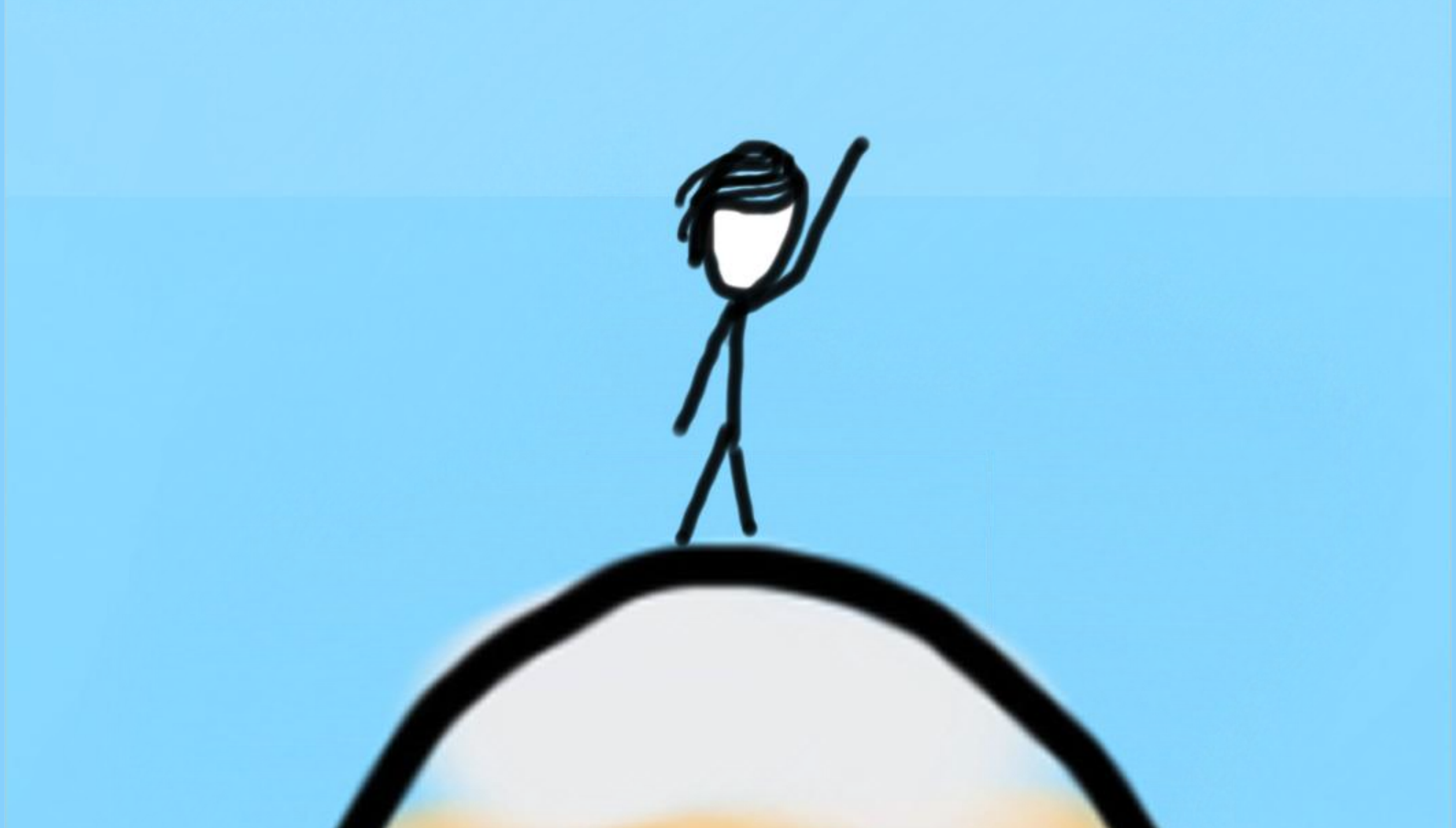


Tender Moments

Chris Piech
CS109, Stanford University

Learning Goals

1. Use New Random Variables!
2. Calculate Expectation of a Random Variable
3. Calculate Variance of a Random Variable (time permitting)



Review

- A **random variable** is a number which takes on values probabilistically.

- A discrete random variable is fully described by a **probability mass function**.

Let Y be a random variable

$$Y$$

For example Y is the number of heads in 5 coin flips

Let Y be a random variable

$$Y = 2$$

*note: here equals means == in coding

It is an event when
Y takes on a value

For example Y is the number of heads in 5 coin flips

Let Y be a random variable

$$Y < 3$$

It is an event when
you ask any comparison question

For example Y is the number of heads in 5 coin flips

If this is a number

$$P(Y = 2)$$

Then this is a probability
(between 0 and 1)

For example Y is the number of heads in 5 coin flips

If this is a variable

$$P(Y = k)$$

Then this is a function

For example Y is the number of heads in 5 coin flips

This is a function

$$P(Y = k)$$

$$k = 5$$

$$0.03125$$

For example Y is the number of heads in 5 coin flips

Random Variables are a big deal, because they allow other people to give you a PMF (and other helpful equations)

Classics

Declare a Random Variable to be Binomial

Our random variable

Num trials

Probability of success on each trial

$$X \sim \text{Bin}(n, p)$$

Is distributed as a

Binomial

With these parameters

Exactly k heads in n coin flips. Probability of exactly k heads:

```
(H, H, H, H, T, T, T, T, T, T)
(H, H, H, T, H, T, T, T, T, T)
(H, H, H, T, T, H, T, T, T, T)
(H, H, H, T, T, T, H, T, T, T)
(H, H, H, T, T, T, T, H, T, T)
(H, H, H, T, T, T, T, T, H, T)
(H, H, H, T, T, T, T, T, T, H)
(H, H, T, H, H, T, T, T, T, T)
(H, H, T, H, T, H, T, T, T, T)
(H, H, T, H, T, T, H, T, T, T)
(H, H, T, H, T, T, T, H, T, T)
(H, H, T, H, T, T, T, T, H, T)
(H, H, T, H, T, T, T, T, T, H)
(H, H, T, T, H, H, T, T, T, T)
(H, H, T, T, H, T, H, T, T, T)
(H, H, T, T, H, T, T, H, T, T)
(H, H, T, T, H, T, T, T, H, T)
(H, H, T, T, H, T, T, T, T, H)
```

$$\binom{n}{k} p^k (1-p)^{n-k}$$

Automatically Know the PMF

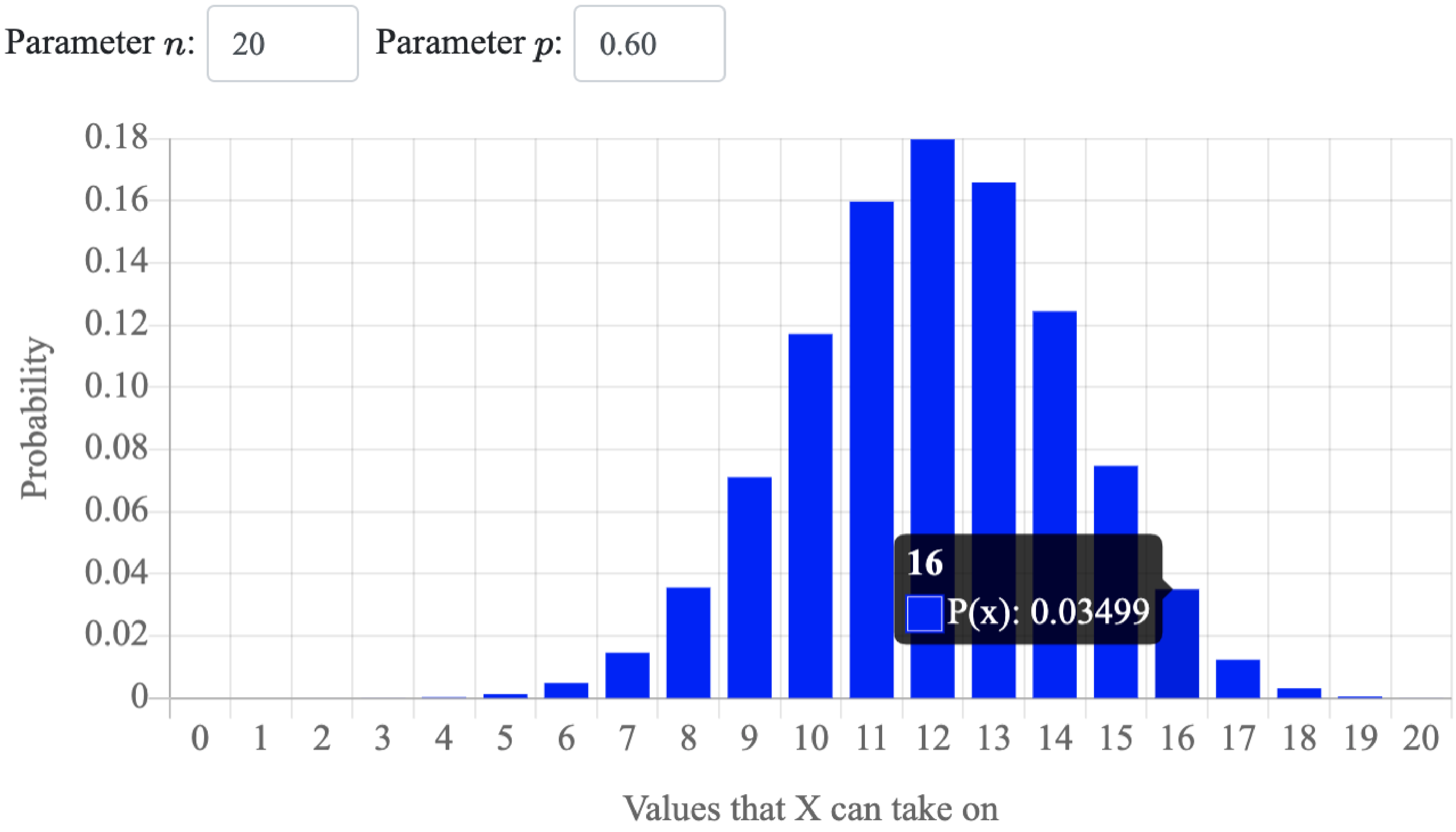
Probability Mass Function for a Binomial

$$P(X = k) = \underbrace{\binom{n}{k} p^k (1-p)^{n-k}}$$

Probability that our variable takes on the value k

* This is also called the binomial term

The PMF as a Graph: $X \sim \text{Bin}(n = 20, p = 0.6)$



End Review

Where are We in CS109?

You are here

Counting Theory

Core Probability

Random Variables

Probabilistic Models

Uncertainty Theory

Machine Learning

Classic Random Variables (with PMFs)

$X \sim \text{Bern}(p)$

Successes in one trial

$X \sim \text{Geo}(p)$

Trials until one success

$Y \sim \text{Bin}(n, p)$

Successes in n trials

$Y \sim \text{NegBin}(r, p)$

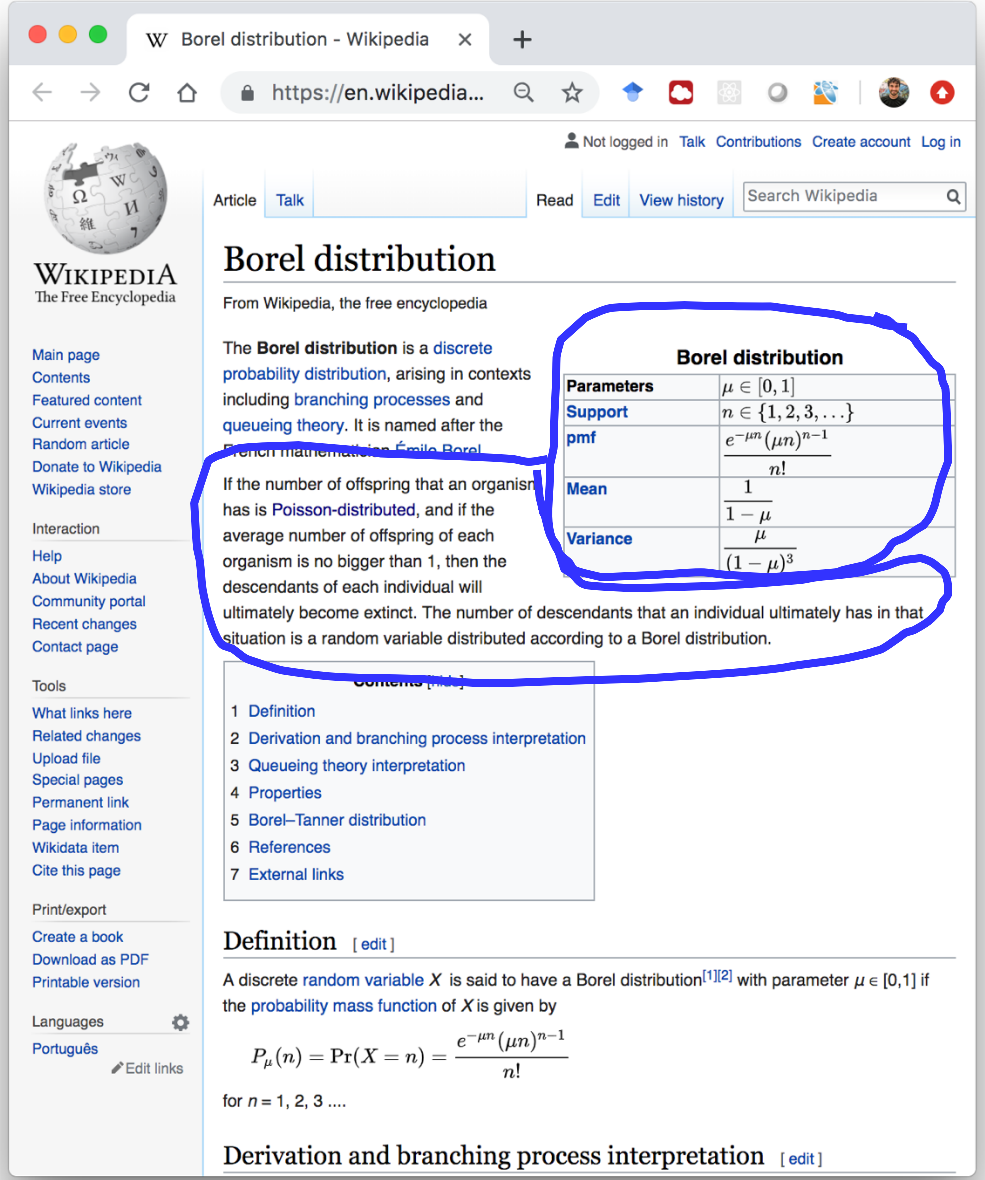
Trials until r success

Goal: Be Able to Use a New Random Variable

You are learning about servers…

You read about the MD1 queue…

You find a paper that says the length of a server "busy period" is distributed as a Borel with parameter μ = 0.2 …



Geometric Random Variable

X is **Geometric** Random Variable: X ~ Geo(*p*)

- X is number of independent trials until first success
- *p* is probability of success on each trial
- Assumes p does not change
- X takes on values 1, 2, 3, …, with probability:

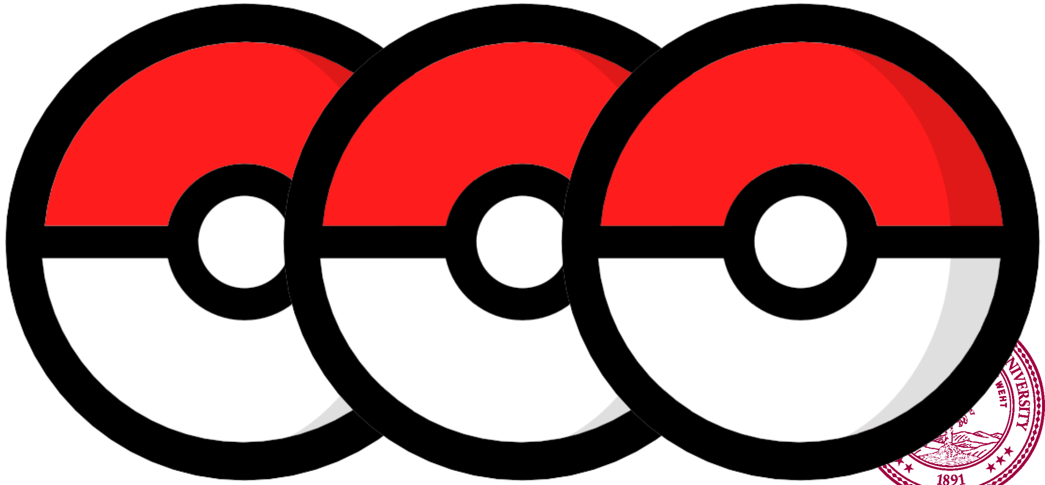
$$P(X = n) = (1 - p)^{n-1} p$$

Negative Binomial Random Variable

X is **Negative Binomial** RV: X ~ NegBin(*r*, *p*)

- X is number of independent trials until *r* successes
- *p* is probability of success on each trial
- Assumes *p* does not change.
- X takes on value *n* with probability:

$$P(X=n)=\binom{n-1}{r-1}p^r(1-p)^{n-r}, \text{ where } n=r,r+1,\ldots$$



Classic Random Variables (with PMFs)

$X \sim \text{Bern}(p)$

$X \sim \text{Geo}(p)$

Trials until one success

$$P(X = x) = (1-p)^{x-1}p$$

$Y \sim \text{Bin}(n, p)$

Successes in *n* trials

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$$

$Y \sim \text{NegBin}(r, p)$

Trials until *r* success

$$P(X = x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}$$

Recipe For Solving Problems:

1. Recognize a classic random variable type

2. Define a random variable to be that type, with parameters

$$X \sim \mathrm{Bin}(n, p)$$

3. Profit off the PMF

Dating at Stanford

Each person you date has a 0.2 probability of being someone you spend your life with. What is the probability you need to date more than 5 people? **Your meta goal: what steps would you take to answer this question?**

Equity in the Courts

Berghuis v. Smith

If a group is underrepresented in a jury pool, how do you tell?

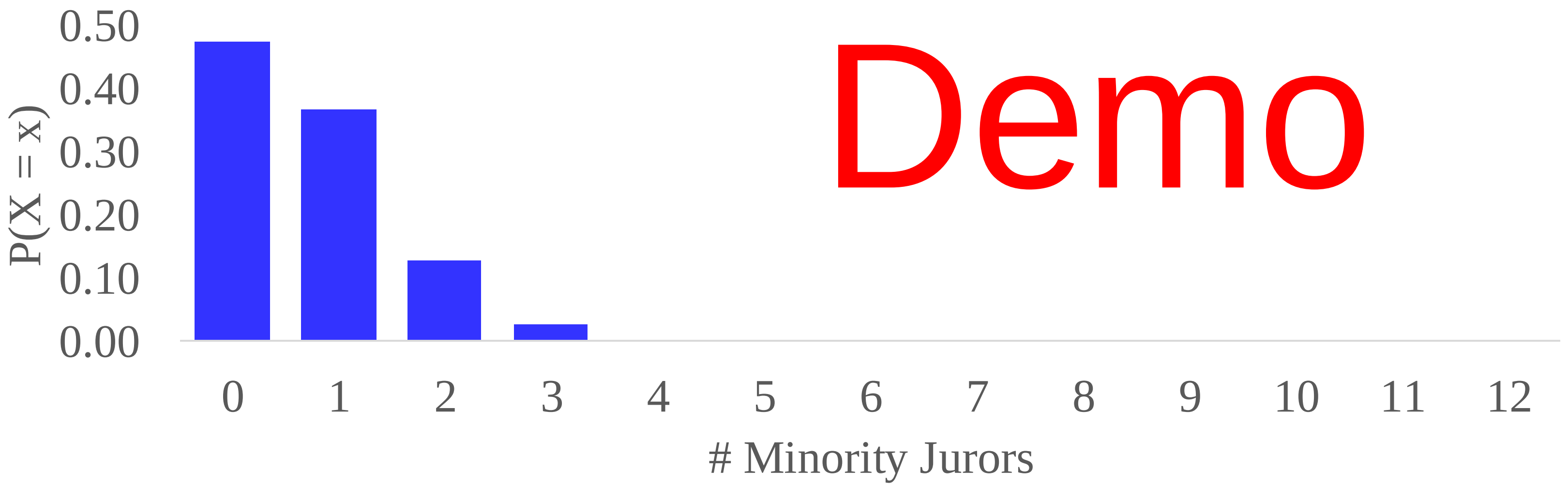
Justice Breyer [Stanford Alum] opened the questioning by invoking the binomial theorem. He hypothesized a scenario involving **"an urn with a thousand balls, and sixty are yellow, and nine hundred forty are navy-blue, and then you select them at random… twelve at a time."** According to Justice Breyer and the binomial theorem, if the purple balls were under represented jurors then **"you would expect… something like a third to a half of juries would have at least one yellow ball"** on them.

Equity in the Courts

Approximation using Binomial distribution

- Assume P(blue ball) constant for every draw = 60/1000
- X = # blue balls drawn. $X \sim \text{Bin}(12, 60/1000 = 0.06)$
- $P(X \geq 1) = 1 - P(X = 0) \approx 1 - 0.4759 = 0.5240$

In Breyer's description, should actually expect just <u>over half</u> of juries to have at least one non-white person on them



Demo

Bitcoin Mining

SHA-256 Hash(,)

Data
Fixed

Salt
Choice

Number that looks like random bits

You "mine a bitcoin" if, for given data *D*, you find a salt number *N* such that Hash(*D*, *N*) produces a string that starts with *g* zeroes.

You "mine a bitcoin" if, for given data D, you find a number N such that Hash(D, N) produces a string that starts with g zeroes.

(a) What is the probability that Hash outputs a bit string which starts with g zeroes (in other words you mine a bitcoin)?

Let X be the number of zeros in the first g bits. $X \sim \text{Bin}(n = g, p = 0.5)$

$$P(X = g) = \binom{g}{g} \frac{1}{2}^{g} = \frac{1}{2}^{g}$$

Call this answer p_a

(b) What is the probability that you will need under 100 attempts to mine 2 bit coins?

Let Y be the number of tries until you mine 2 bitcoins. $Y \sim \text{NegBin}(r = 2, p = p_a)$

$$P(Y < 100) = \sum_{x=2}^{99} P(Y = y)$$

$$= \sum_{x=2}^{99} \binom{x-1}{r-1} p^r (1-p)^{x-r}$$

Classic Random Variables (with PMFs)

$X \sim \text{Bern}(p)$

$X \sim \text{Geo}(p)$

Trials until one success

$$P(X = x) = (1-p)^{x-1}p$$

$Y \sim \text{Bin}(n, p)$

Successes in *n* trials

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$$

$Y \sim \text{NegBin}(r, p)$

Trials until *r* success

$$P(X = x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}$$

Can Jacob Bernoulli Have a Variable Named After Him?

Here yee. I want to have a random variable named after myself. Huzzah.

Can Jacob Bernoulli Have a Variable Named After Him?

Here yee. I want to have a random variable named after myself. Huzzah.

Yes - the Bernoulli random variable: $X \sim \text{Bern}(p)$

Can Jacob Bernoulli Have a Variable Named After Him?

Here yee. I want to have a random variable named after myself. Huzzah.

Yes - the Bernoulli random variable: $X \sim \text{Bern}(p)$

- The Bernoulli is an **indicator** random variable (value is either 0 or 1).
- $P(X = 1) = p$
- $P(X = 0) = 1 - p$

(this is the whole PMF)

- Examples: a single coin flip, one ad click, any binary event

Random Variable Sums

The Binomial

Heads Tails Heads Tails Heads Heads Tails Tails Heads

Random Variable Sums

The Binomial ...is a sum of Bernoulli random variables

Heads Tails Heads Tails Heads Heads Tails Tails Heads

Random Variable Sums

The Binomial

…is a sum of Bernoulli random variables

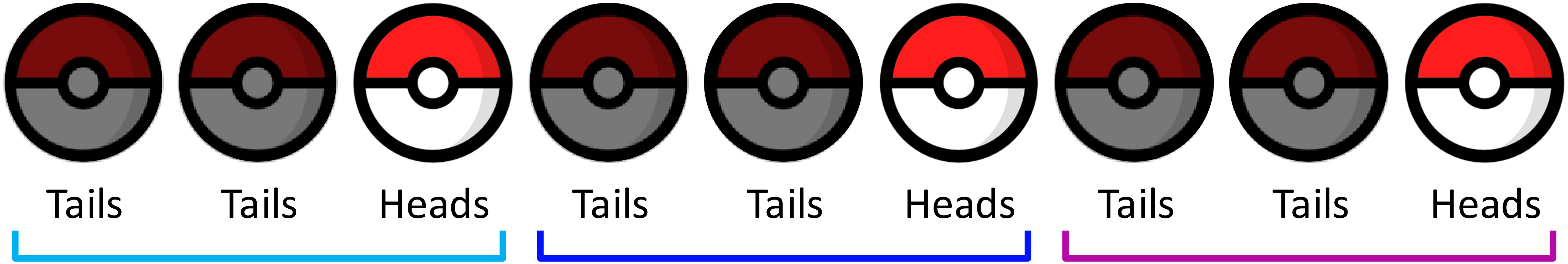
Let $X_1 \sim \text{Bern}(p = 1/2)$ and $X_2 \sim \text{Bern}(p = 1/2)$.

$$Y \sim \text{Bin}(n = 2, p = 1/2)$$

$$Y = X_1 + X_2$$

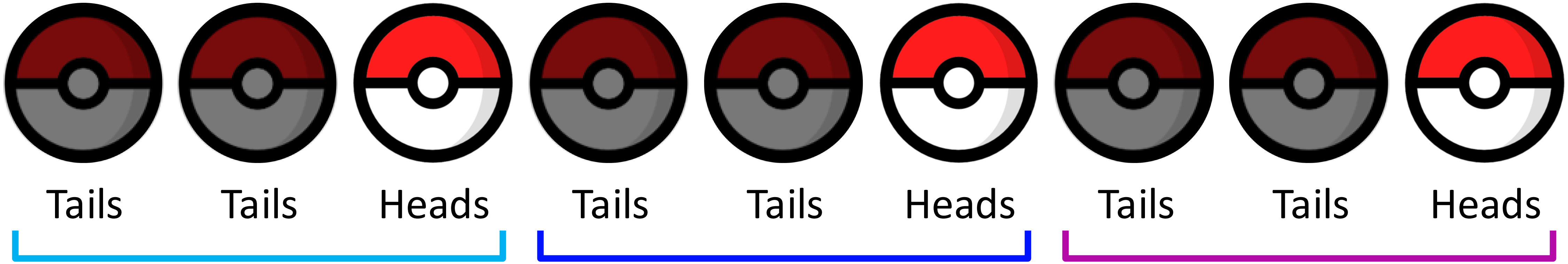
Random Variable Sums

The Negative Binomial



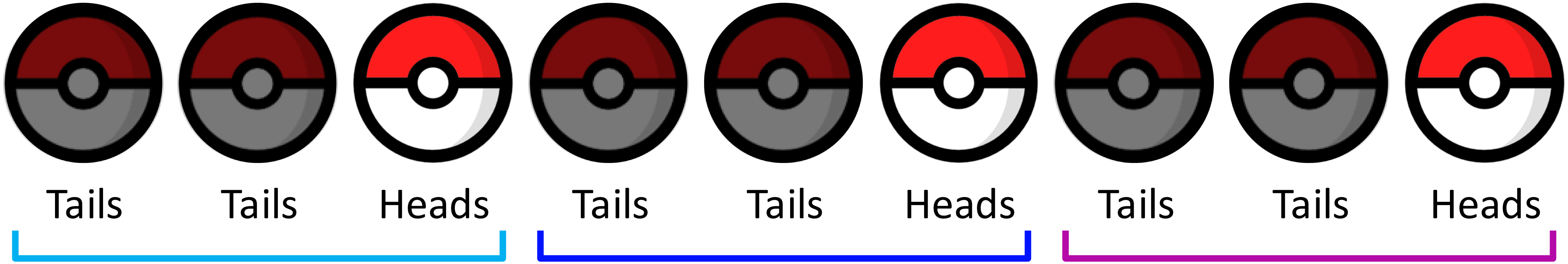
Random Variable Sums

The Negative Binomial ...is a sum of Geometric random variables



Random Variable Sums

The Negative Binomial ...is a sum of Geometric random variables



Let $X_1 \sim \text{Geo}(p = 1/3)$, $X_2 \sim \text{Geo}(p = 1/3)$, and $X_3 \sim \text{Geo}(p = 1/3)$.

$$Y \sim \text{NegBin}(r = 3, p = 1/3)$$

$$Y = X_1 + X_2 + X_3$$

Classic Random Variables (with PMFs)

$X \sim \text{Bern}(p)$

Successes in one trial

$$P(X = x) = \begin{cases} p & \text{if } x = 1 \\ 1 - p & \text{if } x = 0 \end{cases}$$

$X \sim \text{Geo}(p)$

Trials until one success

$$P(X = x) = (1 - p)^{x-1} p$$

$Y \sim \text{Bin}(n, p)$

Successes in *n* trials

$$P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}$$

$Y \sim \text{NegBin}(r, p)$

Trials until *r* success

$$P(X = x) = \binom{x - 1}{r - 1} p^r (1 - p)^{x-r}$$

[Short Pedagogical Pause]

Time for some tender moments*

*Moments: numbers that summarize different aspects of a random variable

Expectation

Expected Value

The probability of
that value

The value

$$E[X] = \sum_x x \cdot \mathrm{P}(X = x)$$

Loop over all values x that X can take on

Expected Value

Expected value answers the question:

What is the average value we could expect some random variable to be?

Also called: ***Mean***, *Expectation*, ***Weighted Average***, ***Center of Mass***, *1st Moment*

$$E[X] = \sum_x x \cdot \mathrm{P}(X = x)$$

Example: Expected Value of Dice Roll

Let X be the result of rolling a 6-sided dice.

$$\boxed{P(X = x) = \frac{1}{6}} \text{ for } x \in \{1, 2, 3, 4, 5, 6\}$$

What is the expectation of X?

Example: Expected Value of Dice Roll

$$E[X] = \sum_x x \cdot \mathrm{P}(X = x)$$

Let X be the result of rolling a 6-sided dice.

$$P(X = x) = \frac{1}{6} \text{ for } x \in \{1, 2, 3, 4, 5, 6\}$$

What is the expectation of X?

$$E[X] = \sum_{x=1}^{6} x \cdot P(X = x)$$

Example: Expected Value of Dice Roll

$$E[X] = \sum_x x \cdot \mathrm{P}(X = x)$$

Let X be the result of rolling a 6-sided dice.

$$P(X = x) = \frac{1}{6} \text{ for } x \in \{1, 2, 3, 4, 5, 6\}$$

What is the expectation of X?

$$\begin{aligned} E[X] &= \sum_{x=1}^{6} x \cdot P(X = x) \\ &= 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} \\ &= 3.5 \end{aligned}$$

Example: Expected Value of Dice Roll

$$E[X] = \sum_x x \cdot \mathrm{P}(X = x)$$

Let X be the result of rolling a 6-sided dice.

$$P(X = x) = \frac{1}{6} \text{ for } x \in \{1, 2, 3, 4, 5, 6\}$$

What is the expectation of X?

$$E[X] = \sum_{x=1}^{6} x \cdot P(X = x)$$

$$= 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6}$$

$$= 3.5$$

E[X] is not always an actual possible outcome for X

Lying With Statistics

Imagine a university has 3 classes, with 5, 10, and 150 students in each class.

We randomly choose a **class** with equal probability.

Let X be the chosen class's size. What is $\mathrm{E}[X]$?

Lying With Statistics

Imagine a university has 3 classes, with 5, 10, and 150 students in each class.

We randomly choose a **class** with equal probability.

Let X be the chosen class's size. What is $\mathrm{E}[X]$?

$$P(X = 5) = 1/3$$
$$P(X = 10) = 1/3$$
$$P(X = 150) = 1/3$$

$$E[X] = \sum_{x \in \{5,10,150\}} x \cdot P(X = x)$$
$$= 5 \cdot \frac{1}{3} + 10 \cdot \frac{1}{3} + 150 \cdot \frac{1}{3}$$
$$= 55$$

Lying With Statistics

Imagine a university has 3 classes, with 5, 10, and $\boxed{150}$ students in each class.

We randomly choose a **student** with equal probability.

Let X be the chosen student's class size. What is $\mathrm{E}[X]$?

Lying With Statistics

Imagine a university has 3 classes, with 5, 10, and 150 students in each class.

We randomly choose a **student** with equal probability.

Let X be the chosen student's class size. What is $\mathrm{E}[X]$?

$$P(X = 5) = 5/165$$
$$P(X = 10) = 10/165$$
$$P(X = 150) = 150/165$$

$$E[X] = \sum_{x \in \{5,10,150\}} x \cdot P(X = x)$$
$$= 5 \cdot \frac{5}{165} + 10 \cdot \frac{10}{165} + 150 \cdot \frac{150}{165}$$
$$= 137$$

Expectation from Data

List called data

X
3
2
6
10
1
1
5
4
...

$$E[X] = \sum_x x \cdot \mathrm{P}(X = x)$$

$$\approx \sum_x x \cdot \frac{\mathrm{count}(X = x)}{N}$$

Length of data

$$\approx \frac{1}{N} \sum_x x \cdot \mathrm{count}(X = x)$$

$$\approx \frac{1}{N} \sum_{v \in \mathrm{data}} v$$

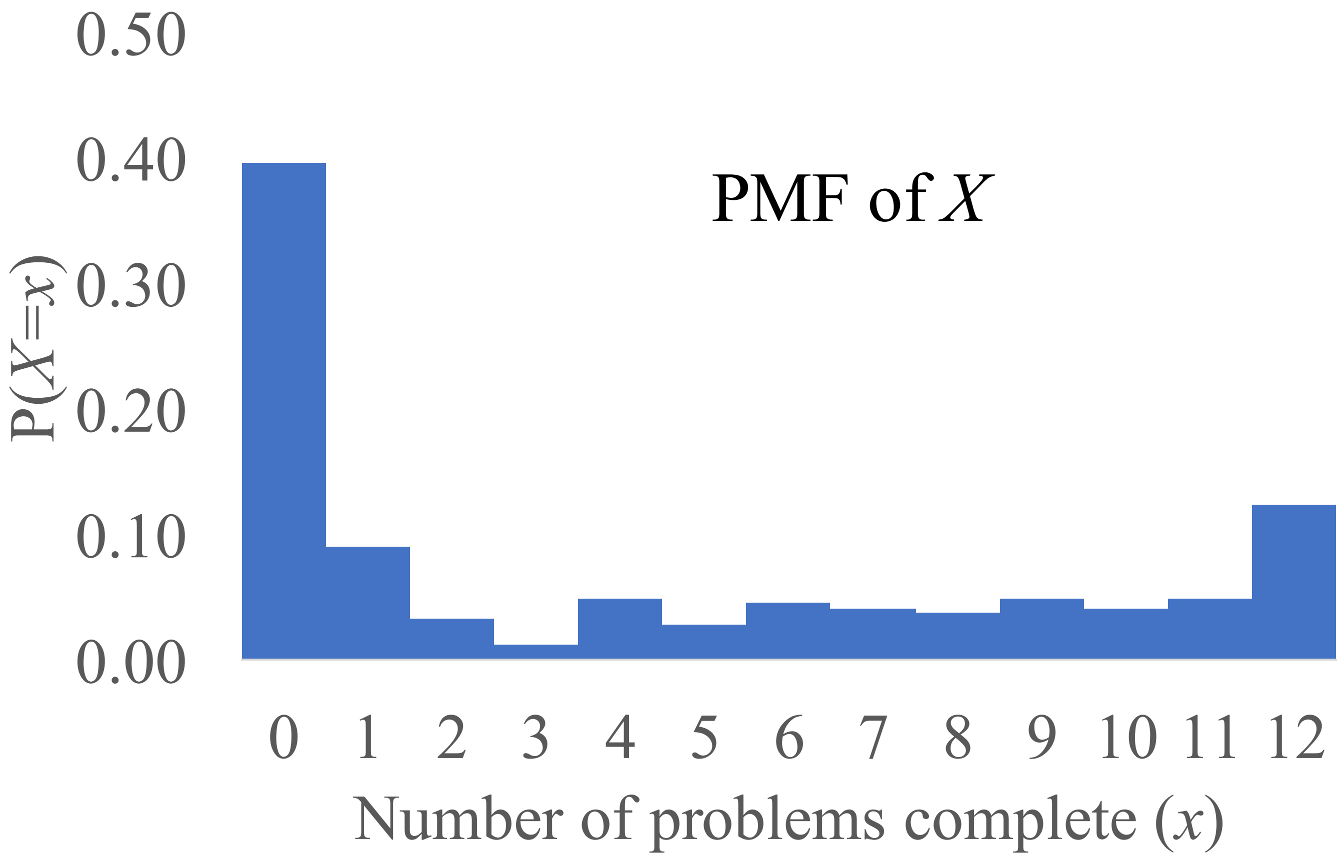
Expectation is a single number summary…

Expectation leaves much to be desired…

Expectation vs PMF

Let X be the number of problems that a randomly selected student has completed, as of 11a today.

X takes on values, with uncertainty. X is a random variable.



$$P(X = 12) \approx \frac{\text{Count}(X = 12)}{N}$$

$$E[X] = 6$$

Why People Care?

Properties of Expectation (proof later)

Linearity:

$$E[aX + b] = aE[X] + b$$

- Consider X = 6-sided die roll, Winnings = 2X – 1.
- E[X] = 3.5 E[2X-1] = 6

Expectation of a sum is the sum of expectations

$$E[X + Y] = E[X] + E[Y]$$

Unconscious statistician:

$$E[g(x)] = \sum_{x \in X} g(x) P(X = x)$$

Law of the Unconscious Statistician (LOTUS)

$$E[g(X)] = \sum_x g(x)P(X = x)$$

This lets you get the expectation of **any** function of a random variable.

Examples:

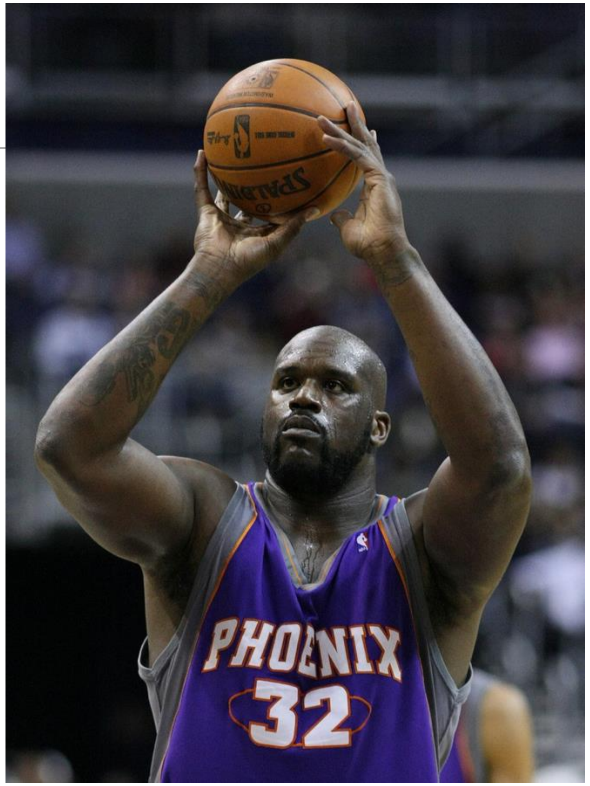
$$E[X^2] = \sum_x x^2 \cdot P(X = x)$$

$$E[\sin(X)] = \sum_x \sin(x) \cdot P(X = x)$$

$$E[\sqrt(X)] = \sum_x \sqrt{x} \cdot P(X = x)$$

Expectation of Classic Random Variables

Expected Value of Free Throws

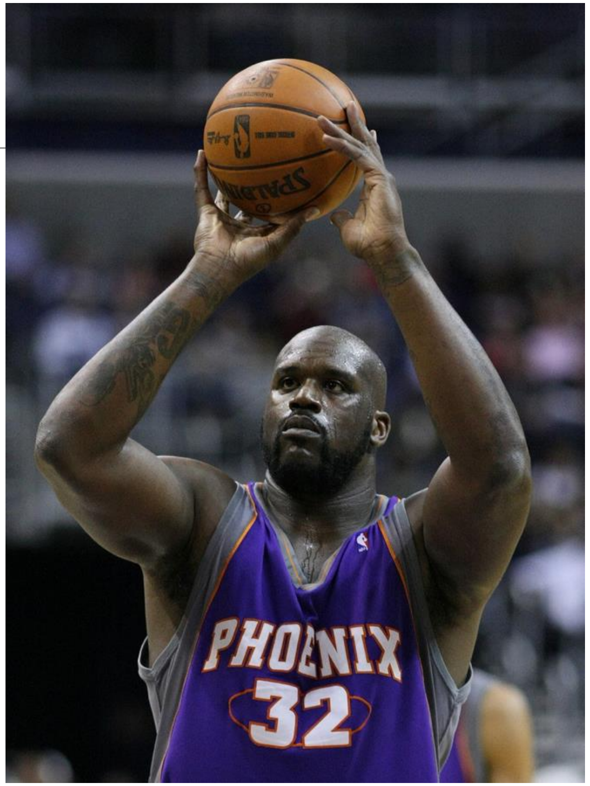


In basketball, players sometimes get a chance to shoot a free throw. If they make it, the team gets 1 point; otherwise they get no points.

Some players are not very good at free throws, such as Shaq. While in the NBA, Shaq made only 53% of his free throws.

Let X be the points gained from Shaq attempting a free throw.
What is $\mathrm{E}[X]$?

Expected Value of Free Throws



In basketball, players sometimes get a chance to shoot a free throw. If they make it, the team gets 1 point; otherwise they get no points.

Some players are not very good at free throws, such as Shaq. While in the NBA, Shaq made only 53% of his free throws.

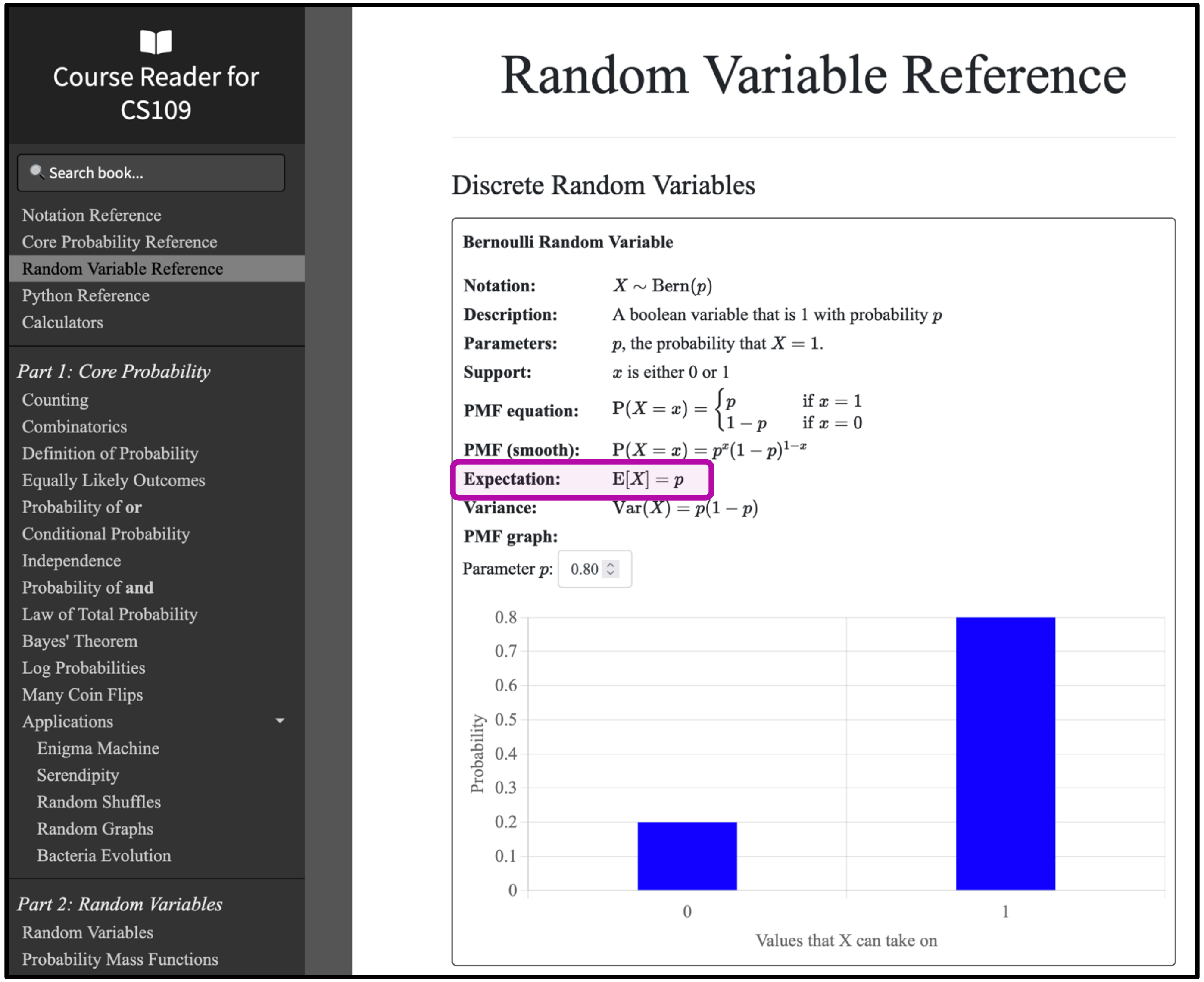
Let X be the points gained from Shaq attempting a free throw. What is $\mathrm{E}[X]$?

$$X \sim \mathrm{Bern}(p = 0.53)$$

$$\begin{aligned} E[X] &= 0 \cdot P(X=0) + 1 \cdot P(X=1) \\ &= 0 \cdot 0.47 + 1 \cdot 0.53 = \underline{0.53} \end{aligned}$$

For Bernoulli random variables, $\mathrm{E}[X] = p$ (always)

With Classic RVs, You Get Expectations For Free Too!



Course Reader for CS109

Search book...

Random Variable Reference

Discrete Random Variables

Bernoulli Random Variable

Notation: $X \sim \text{Bern}(p)$

Description: A boolean variable that is 1 with probability p

Parameters: p, the probability that $X = 1$.

Support: x is either 0 or 1

PMF equation: $P(X = x) = \begin{cases} p & \text{if } x = 1 \\ 1 - p & \text{if } x = 0 \end{cases}$

PMF (smooth): $P(X = x) = p^x(1 - p)^{1-x}$

Expectation: $\boxed{E[X] = p}$

Variance: $\text{Var}(X) = p(1 - p)$

PMF graph:

Parameter p: 0.80

We Can Now Calculate Expectation of Binomial

$$X \sim \mathrm{Bin}(n, p)$$

We Can Now Calculate Expectation of Binomial

$$X \sim \text{Bin}(n, p)$$

Let Y_i be 1 if trial i was a success, otherwise 0, with *i* from 1 to *n*. $Y_i \sim \text{Bern}(p)$.

The Binomial ...is a sum of Bernoulli random variables

Heads Tails Heads Tails Heads Heads Tails Tails Heads

We Can Now Calculate Expectation of Binomial

$$X \sim \mathrm{Bin}(n, p)$$

Let Y_i be 1 if trial i was a success, otherwise 0, with *i* from 1 to *n*. $Y_i \sim \mathrm{Bern}(p)$.

$$\mathrm{E}[X] = \mathrm{E}\left[\sum_{i=1}^{n} Y_i\right] \qquad \text{Since } X = \sum_{i=1}^{n} Y_i$$

We Can Now Calculate Expectation of Binomial

$$X \sim \text{Bin}(n, p)$$

Let Y_i be 1 if trial i was a success, otherwise 0, with i from 1 to n. $Y_i \sim \text{Bern}(p)$.

$$\begin{aligned} \mathrm{E}[X] &= \mathrm{E}\left[\sum_{i=1}^{n} Y_i\right] && \text{Since } X = \sum_{i=1}^{n} Y_i \\ &= \sum_{i=1}^{n} \mathrm{E}[Y_i] && \text{Expectation of sum} \end{aligned}$$

Expectation of a sum is the sum of expectations: $E[X+Y] = E[X] + E[Y]$

We Can Now Calculate Expectation of Binomial

$$X \sim \text{Bin}(n, p)$$

Let Y_i be 1 if trial i was a success, otherwise 0, with *i* from 1 to *n*. $Y_i \sim \text{Bern}(p)$.

$$\begin{aligned}
\mathrm{E}[X] &= \mathrm{E}\left[\sum_{i=1}^{n} Y_i\right] && \text{Since } X = \sum_{i=1}^{n} Y_i \\
&= \sum_{i=1}^{n} \mathrm{E}[Y_i] && \text{Expectation of sum} \\
&= \sum_{i=1}^{n} p && \text{Expectation of Bernoulli} \\
&= n \cdot p && \text{Sum } n \text{ times}
\end{aligned}$$

True for every binomial ever

You Get So Much For Free!

Binomial Random Variable

Notation: $X \sim \text{Bin}(n, p)$

Description: Number of "successes" in n identical, independent experiments each with probability of success p.

Parameters: $n \in \{0, 1, \ldots\}$, the number of experiments.
$p \in [0, 1]$, the probability that a single experiment gives a "success".

Support: $x \in \{0, 1, \ldots, n\}$

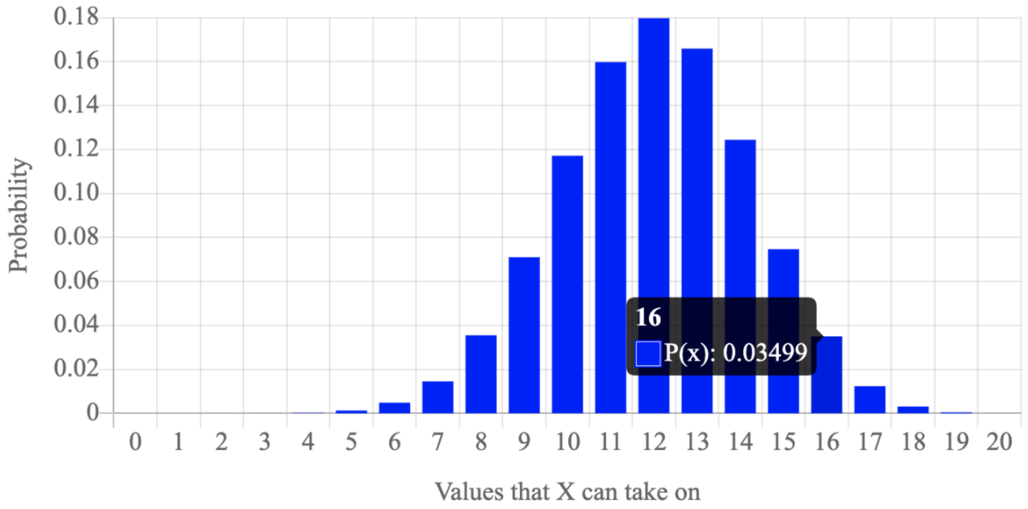
PMF equation: $\Pr(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$

Expectation: $\text{E}[X] = n \cdot p$

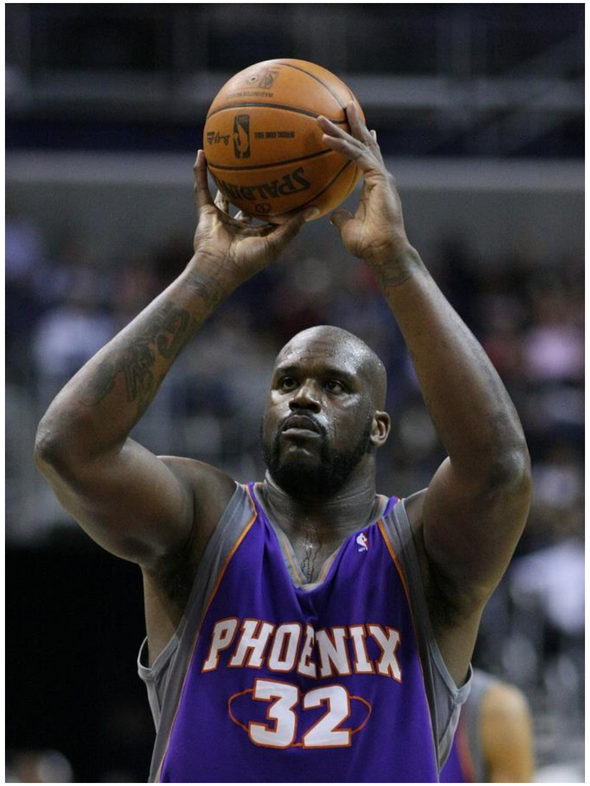
Variance: $\text{Var}(X) = n \cdot p \cdot (1-p)$

PMF graph:

Parameter n: 20 Parameter p: 0.60



Expected Value of Free Throws

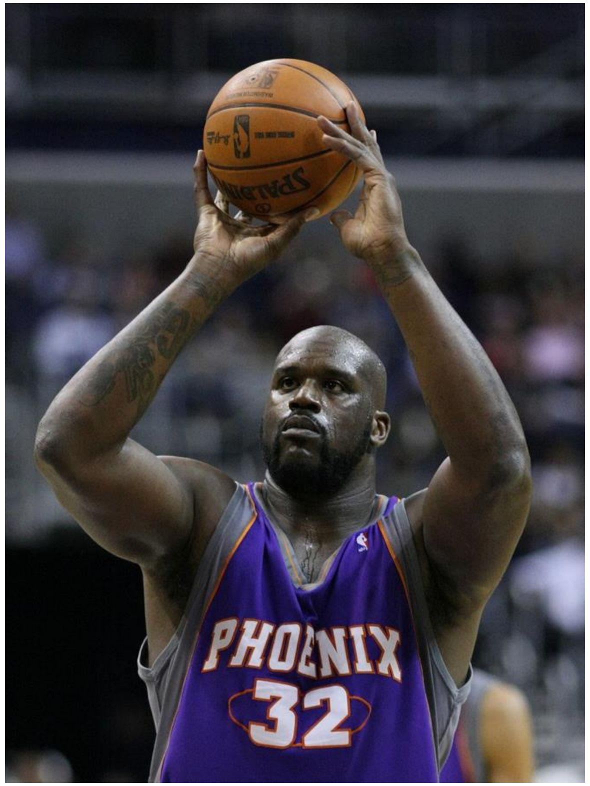


In basketball, players sometimes get a chance to shoot a free throw. If they make it, the team gets 1 point; otherwise they get no points.

Some players are not very good at free throws, such as Shaq. While in the NBA, Shaq made only 53% of his free throws.

Let Y be the points gained from Shaq attempting **500** free throws. What is $\mathrm{E}[Y]$?

Expected Value of Free Throws



In basketball, players sometimes get a chance to shoot a free throw. If they make it, the team gets 1 point; otherwise they get no points.

Some players are not very good at free throws, such as Shaq. While in the NBA, Shaq made only 53% of his free throws.

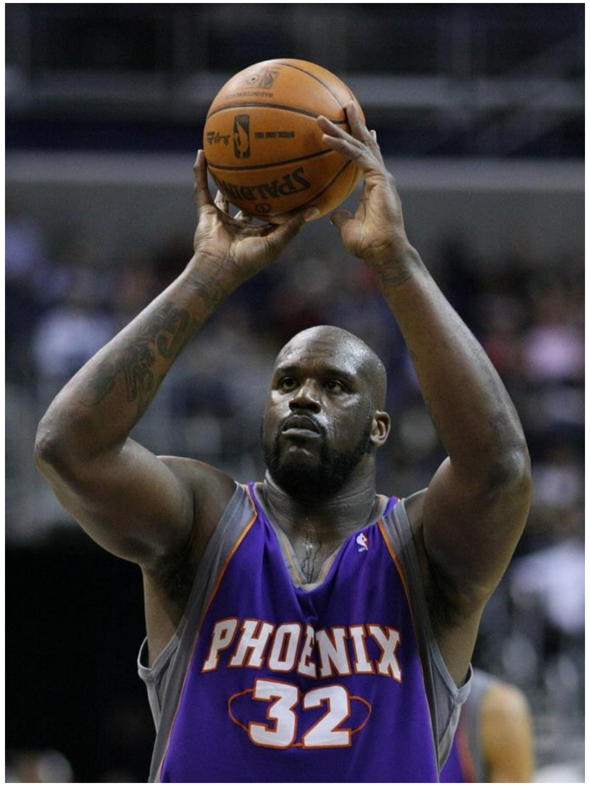
Let Y be the points gained from Shaq attempting **500** free throws. What is $\mathrm{E}[Y]$?

$$Y \sim Bin(n = 500, p = 0.53)$$

Expected Value of Free Throws

In basketball, players sometimes get a chance to shoot a free throw. If they make it, the team gets 1 point; otherwise they get no points.

Some players are not very good at free throws, such as Shaq. While in the NBA, Shaq made only 53% of his free throws.

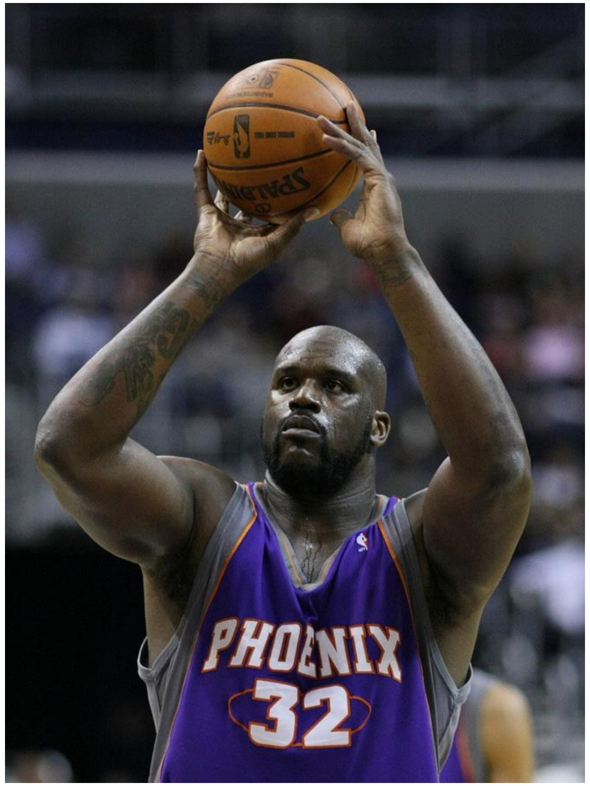


Let Y be the points gained from Shaq attempting **500** free throws. What is $\mathrm{E}[Y]$?

$$Y \sim Bin(n = 500, p = 0.53)$$

$$E[Y] = n \cdot p = 500 \cdot 0.53 = 265$$

Expected Value of Free Throws



In basketball, players sometimes get a chance to shoot a free throw. If they make it, the team gets 1 point; otherwise they get no points.

Some players are not very good at free throws, such as Shaq. While in the NBA, Shaq made only 53% of his free throws.

Let Y be the points gained from Shaq attempting **500** free throws. What is $\mathrm{E}[Y]$?

$$Y \sim Bin(n = 500, p = 0.53)$$

$$E[Y] = n \cdot p = 500 \cdot 0.53 = 265$$

Challenge: If Shaq was 10% better at shooting free throws, how many *more* free throws would you expect him to make, out of 500?

Expected Value of The Geometric

If $X \sim \text{Geo}(p)$, then $$E[X] = \frac{1}{p}$$

This definition has intuition built in:

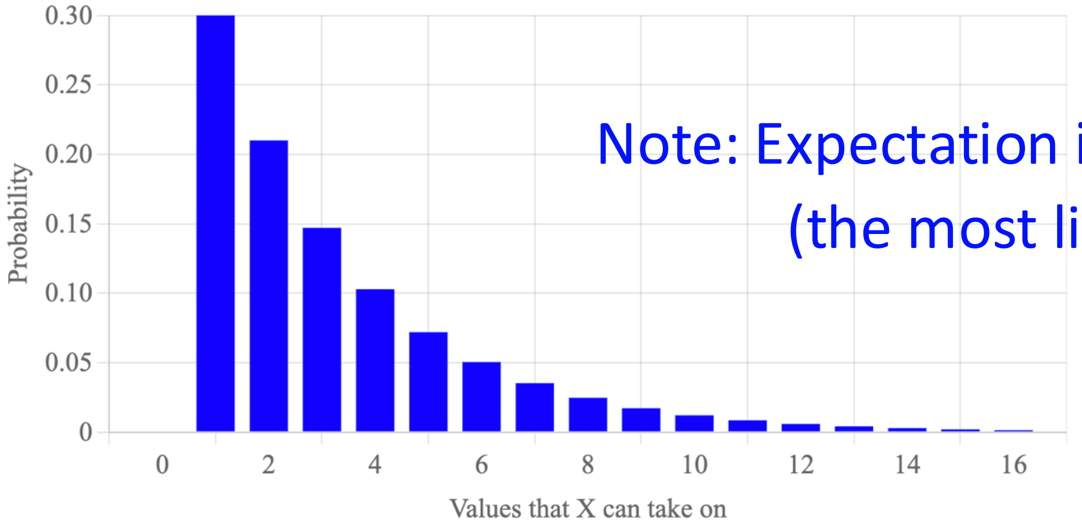
- If Shaq makes about half his free throws, then on average, it will take him two shots to make one free throw. $E[X] = (1/2)^{-1} = 2$.

Expected Value of The Geometric

If $X \sim \text{Geo}(p)$, then $$E[X] = \frac{1}{p}$$

This definition has intuition built in:

- If Shaq makes about half his free throws, then on average, it will take him two shots to make one free throw. $E[X] = (1/2)^{-1} = 2$.

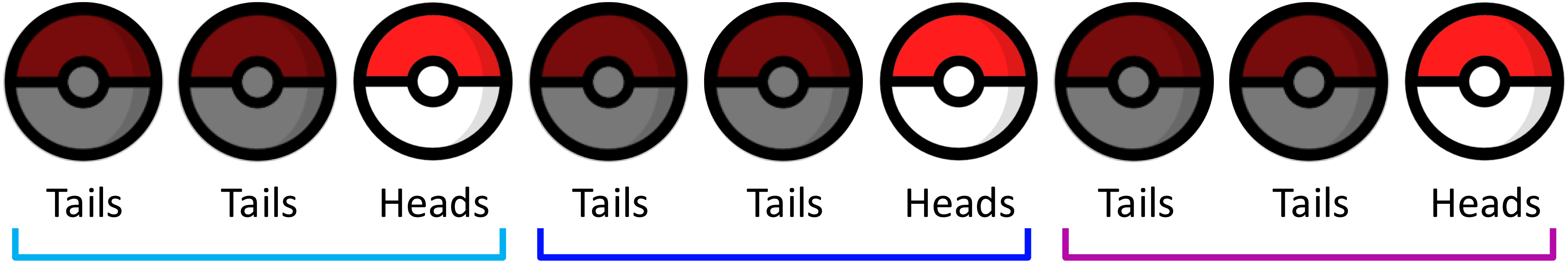


Note: Expectation is often **not** the mode (the most likely outcome)

Expected Value of The Negative Binomial

We can derive using the **sum of expectations** property, similar to binomials.

The Negative Binomial ...is a sum of Geometric random variables



Expected Value of The Negative Binomial

We can derive using the **sum of expectations** property, similar to binomials.

Let $X_i \sim \mathrm{Geo}(p)$, for each i from 1 to r.

$$E[X_i] = \frac{1}{p}$$

Let $Y \sim \mathrm{NegBin}(r, p)$.

Expected Value of The Negative Binomial

We can derive using the **sum of expectations** property, similar to binomials.

Let $X_i \sim \text{Geo}(p)$, for each i from 1 to r.

$$E[X_i] = \frac{1}{p}$$

Let $Y \sim \text{NegBin}(r, p)$.

$$E[Y] = E\left[\sum_{i=1}^{r} X_i\right]$$

Expected Value of The Negative Binomial

We can derive using the **sum of expectations** property, similar to binomials.

Let $X_i \sim \text{Geo}(p)$, for each i from 1 to r.

$$E[X_i] = \frac{1}{p}$$

Let $Y \sim \text{NegBin}(r, p)$.

$$E[Y] = E\left[\sum_{i=1}^{r} X_i\right]$$

$$= \sum_{i=1}^{r} E[X_i]$$

Expected Value of The Negative Binomial

We can derive using the **sum of expectations** property, similar to binomials.

Let $X_i \sim \mathrm{Geo}(p)$, for each i from 1 to r.

$$E[X_i] = \frac{1}{p}$$

Let $Y \sim \mathrm{NegBin}(r, p)$.

$$E[Y] = E\left[\sum_{i=1}^{r} X_i\right]$$

$$= \sum_{i=1}^{r} E[X_i]$$

$$= \sum_{i=1}^{r} \frac{1}{p} = \frac{r}{p}$$

Expectations of Classic Random Variables

$X \sim \text{Geo}(p)$

$$E[X] = \frac{1}{p}$$

$X \sim \text{Bern}(p)$

$$E[X] = p$$

$Y \sim \text{NegBin}(r, p)$

$$E[Y] = \frac{r}{p}$$

$Y \sim \text{Bin}(n, p)$

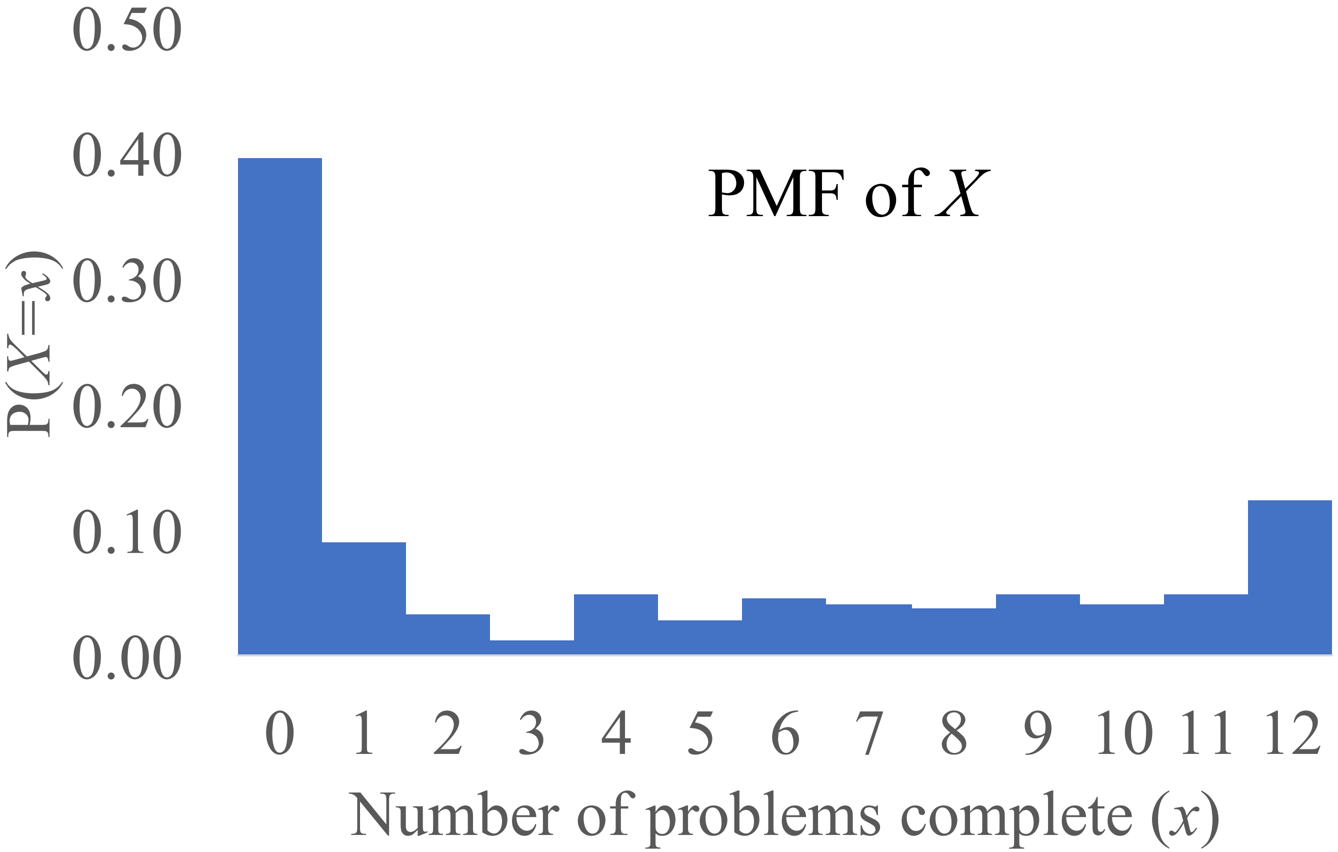
$$E[Y] = n \cdot p$$

Expectation is easy to work with, but still leaves much to be desired

Expectation vs PMF

Let X be the number of problems that a randomly selected student has completed, as of 11a today.

X takes on values, with uncertainty. X is a random variable.

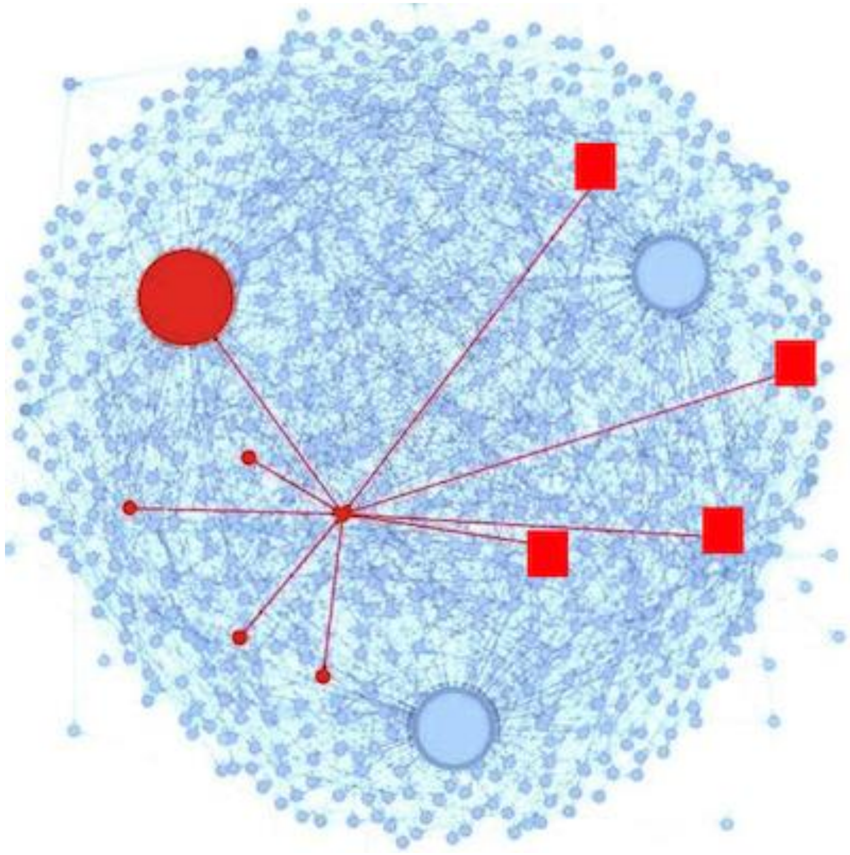


$$P(X = 12) \approx \frac{\text{Count}(X = 12)}{N}$$

$$E[X] = 6$$

Can we invent *another* summary number?

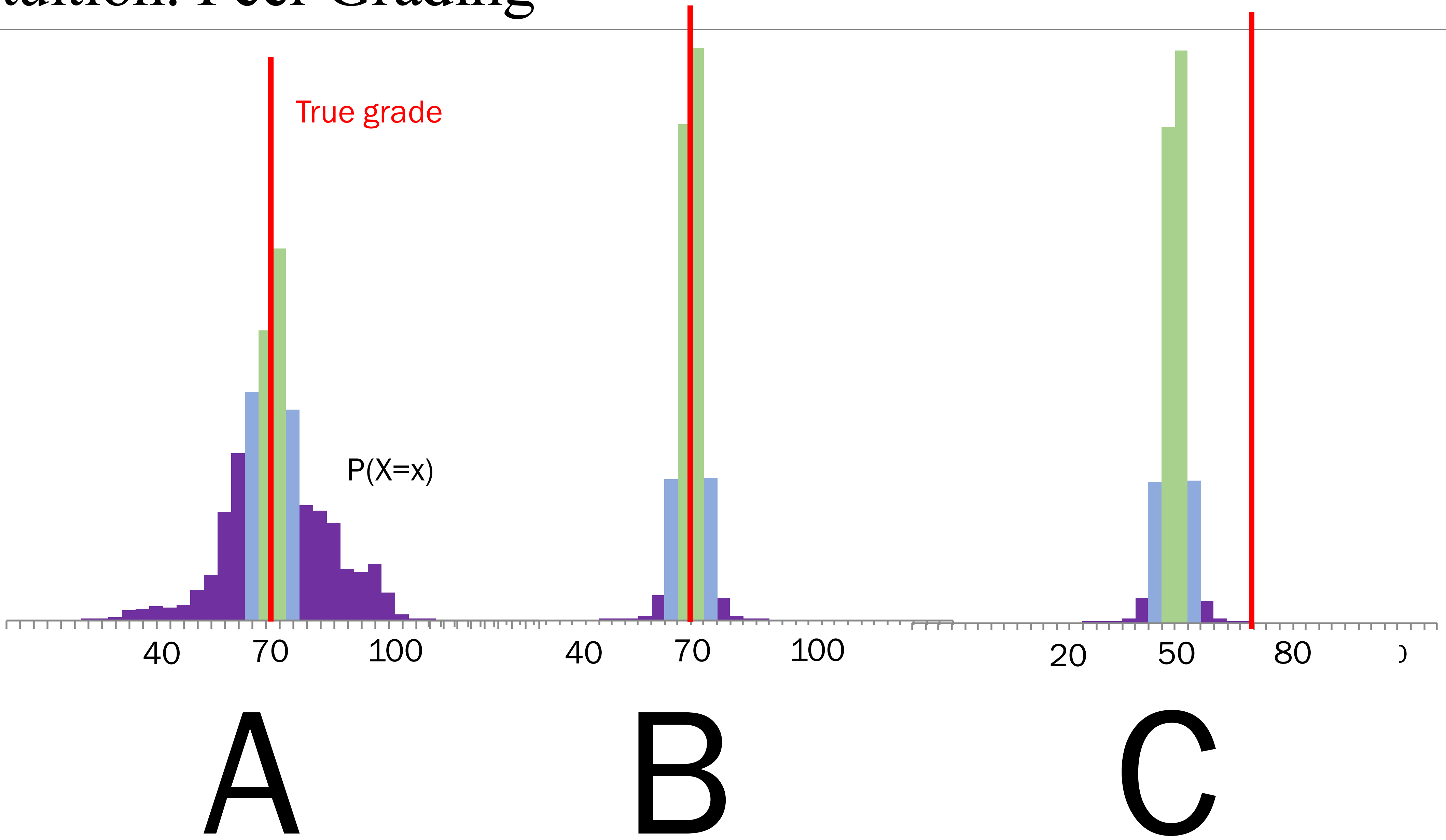
Intuition: Peer Grading



Peer Grading on Coursera HCI.

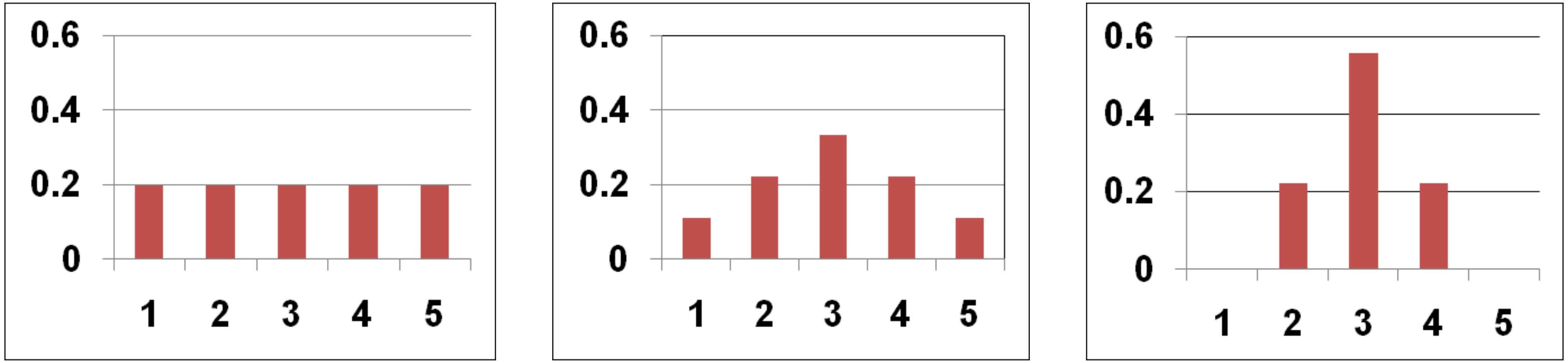
31,067 peer grades for 3,607 students.

Intuition: Peer Grading



Intuition: Measure of Spread

Consider the following 3 distributions (PMFs)



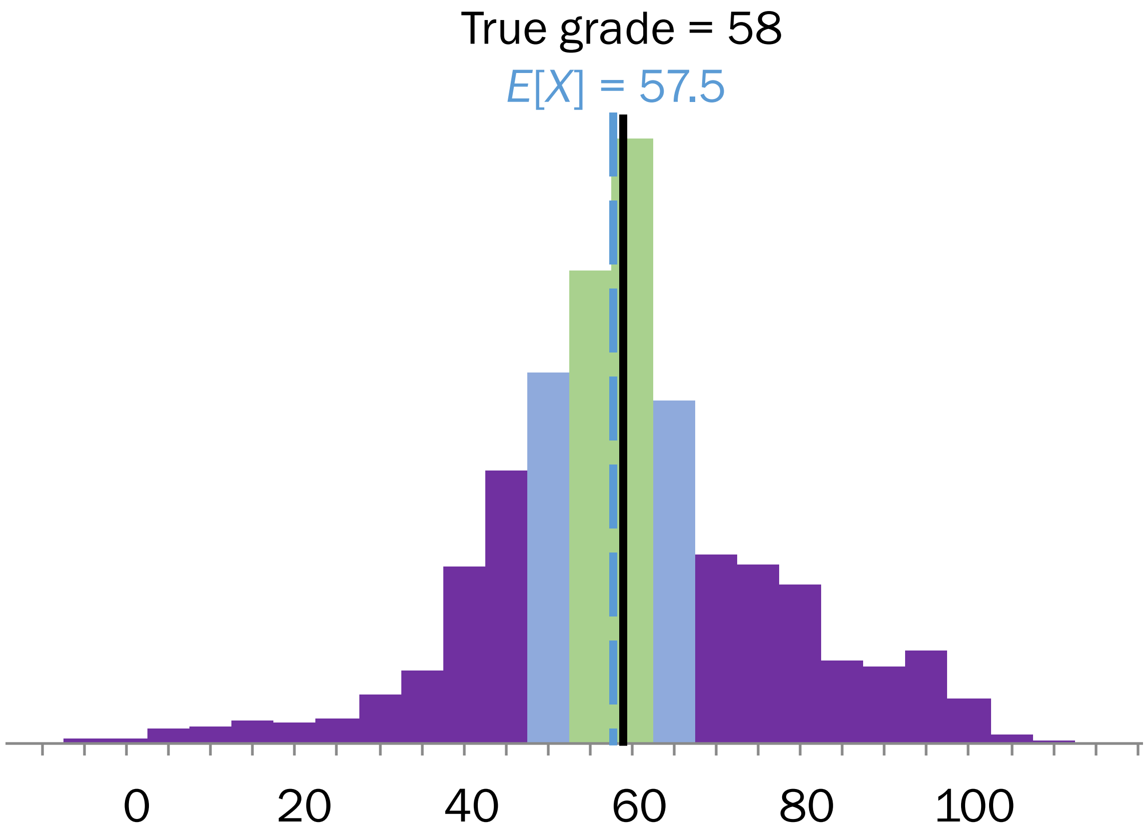
All have the same expected value, E[X] = 3

But “spread” in distributions is different

Invent a formal quantification of “spread”?

Peer grading in Coursera HCI

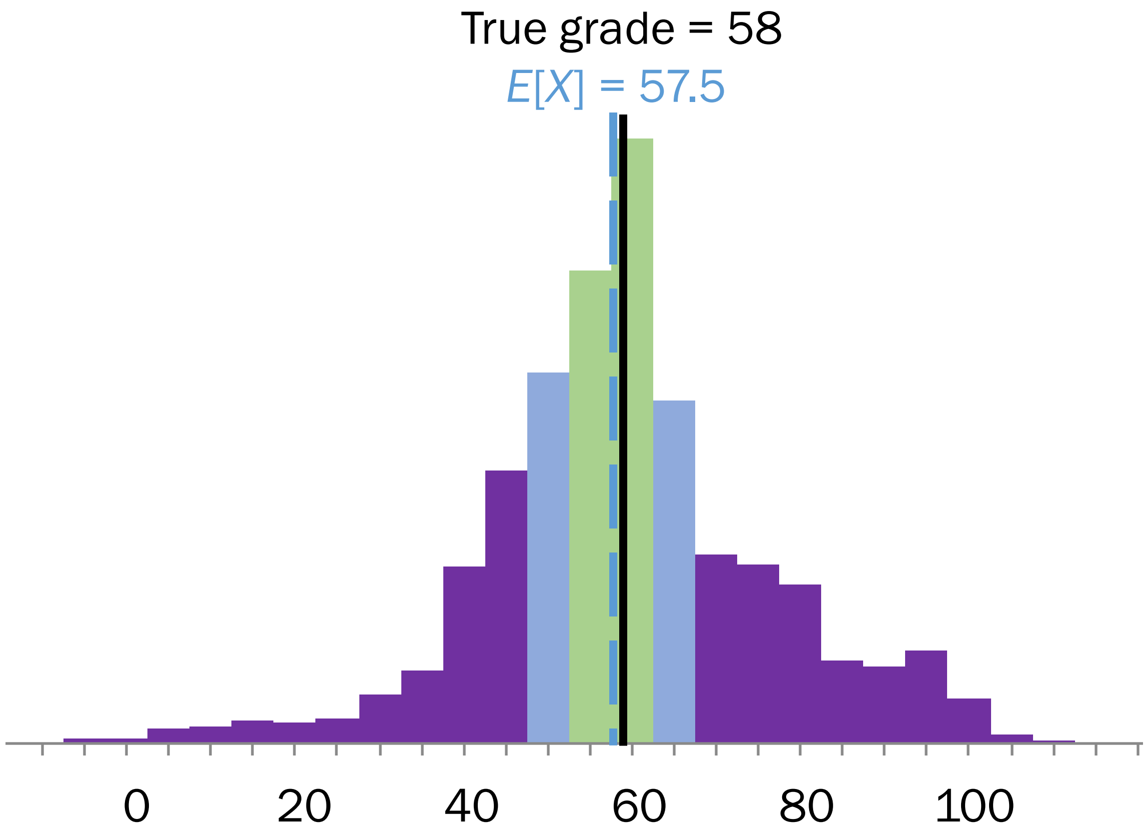
Let X be a random variable that represents a peer grade



Peer grading in Coursera HCI

Let X be a random variable that represents a peer grade

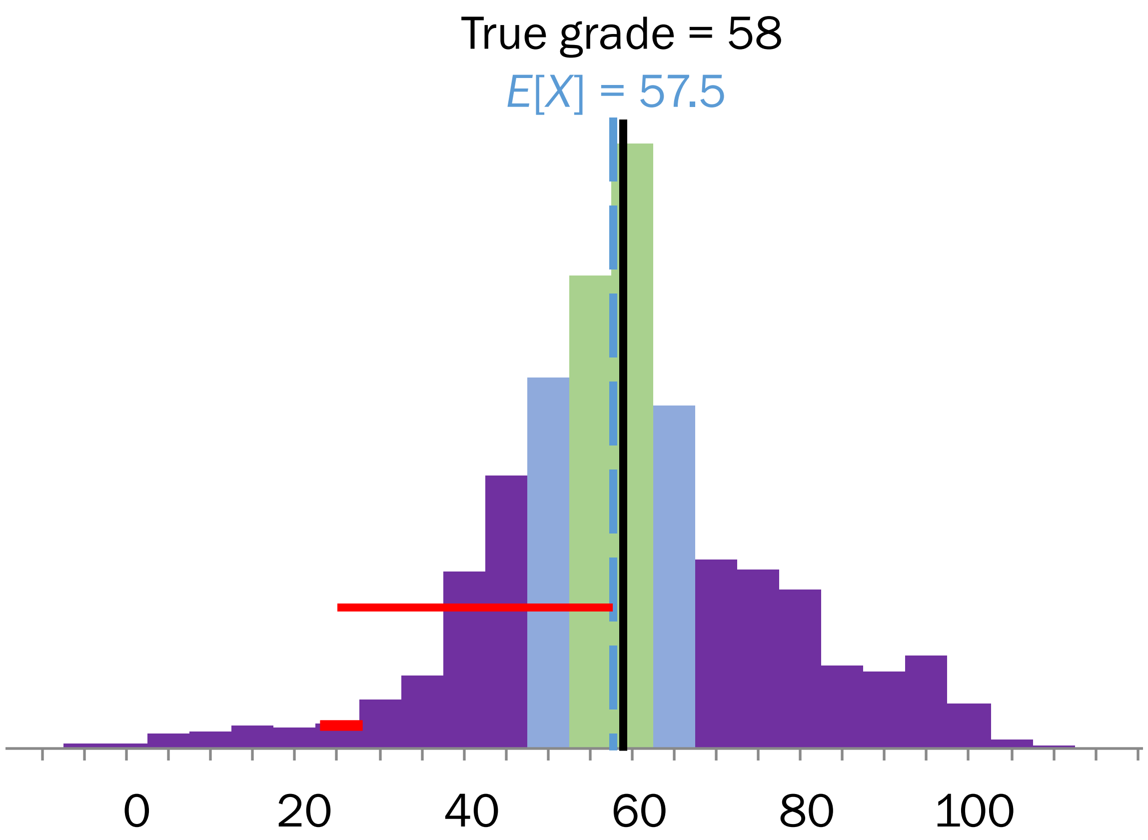
$$\mathrm{Var}(X) = E[(X - \mu)^2]$$



Peer grading in Coursera HCI

Let X be a random variable that represents a peer grade

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



X	$(X - \mu)^2$
25 points	1056 points^2

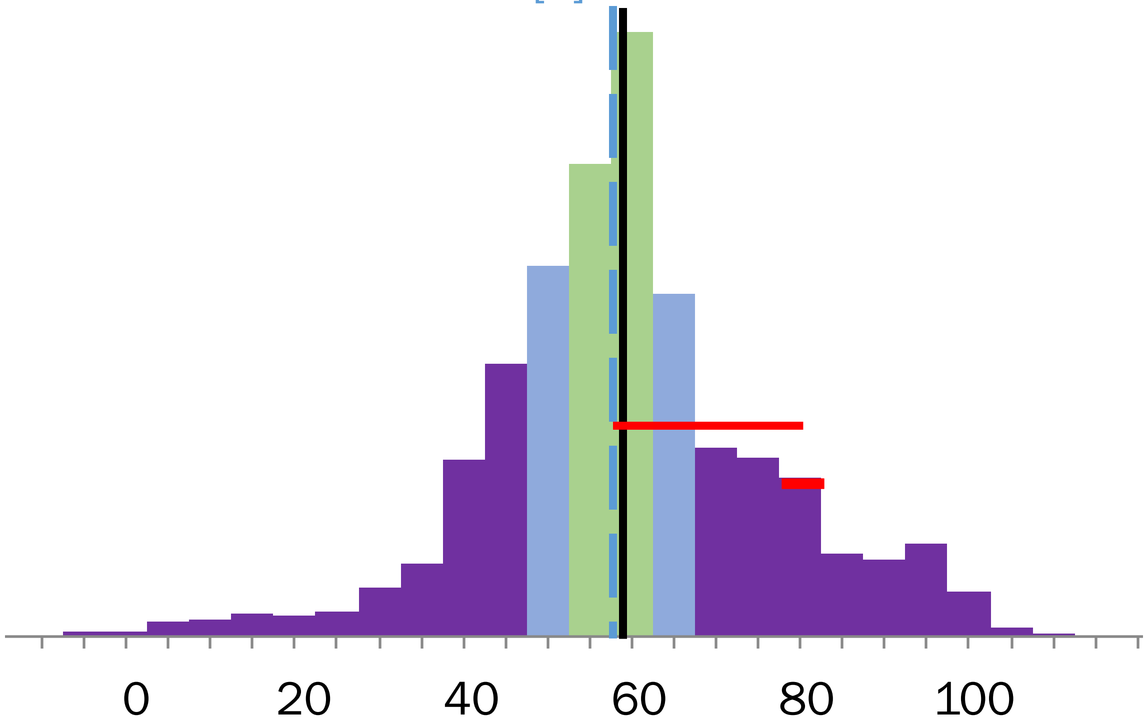
Peer grading in Coursera HCI

Let X be a random variable that represents a peer grade

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

True grade = 58

$E[X] = 57.5$

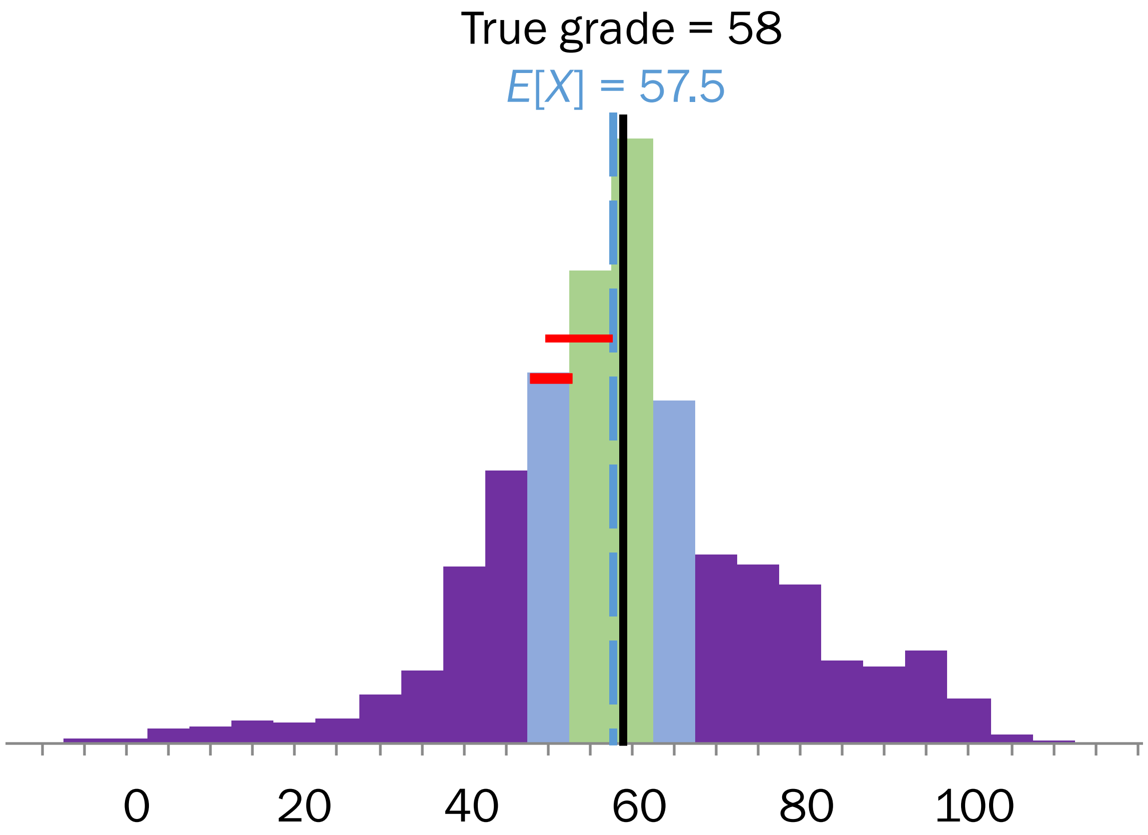


X	$(X - \mu)^2$
25 points	1056 points^2
80 points	506 points^2

Peer grading in Coursera HCI

Let X be a random variable that represents a peer grade

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

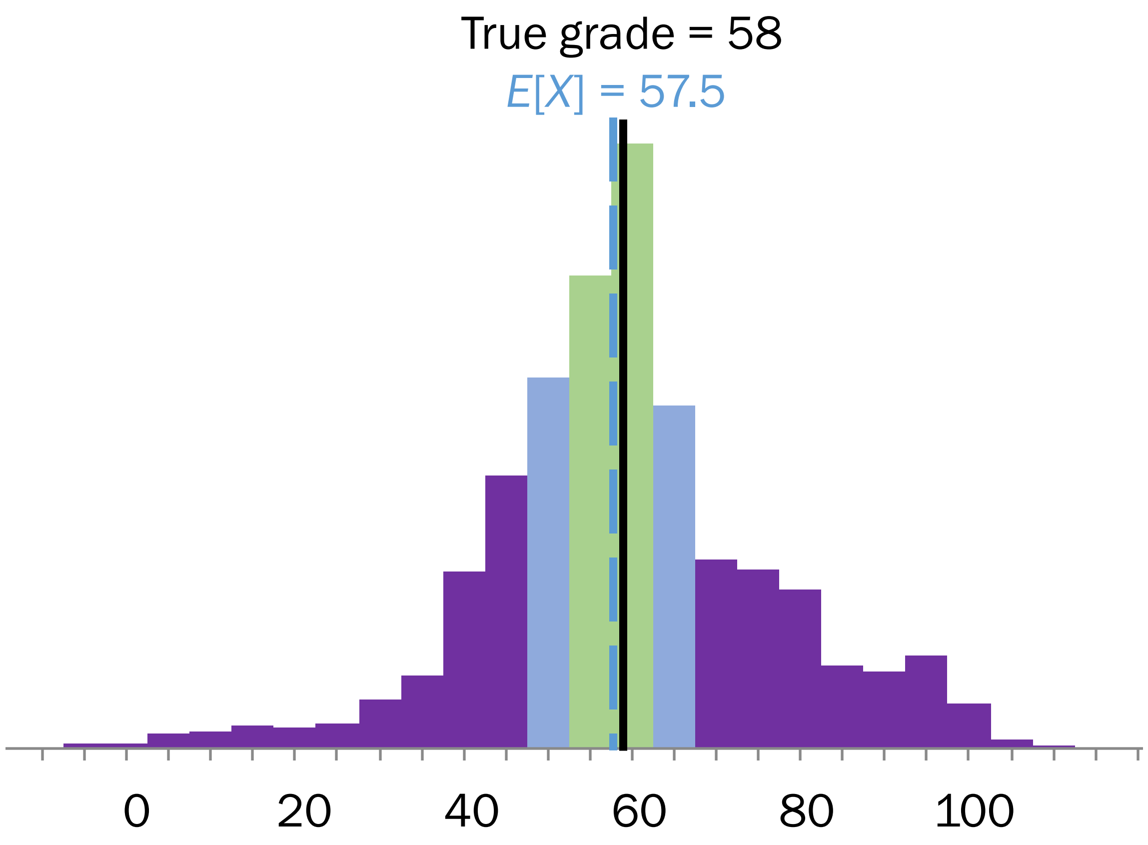


X	$(X-\mu)^2$
25 points	1056 points^2
80 points	506 points^2
50 points	56 points^2

Peer grading in Coursera HCI

Let X be a random variable that represents a peer grade

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



X	$(X - \mu)^2$
25 points	1056 points^2
80 points	506 points^2
50 points	56 points^2

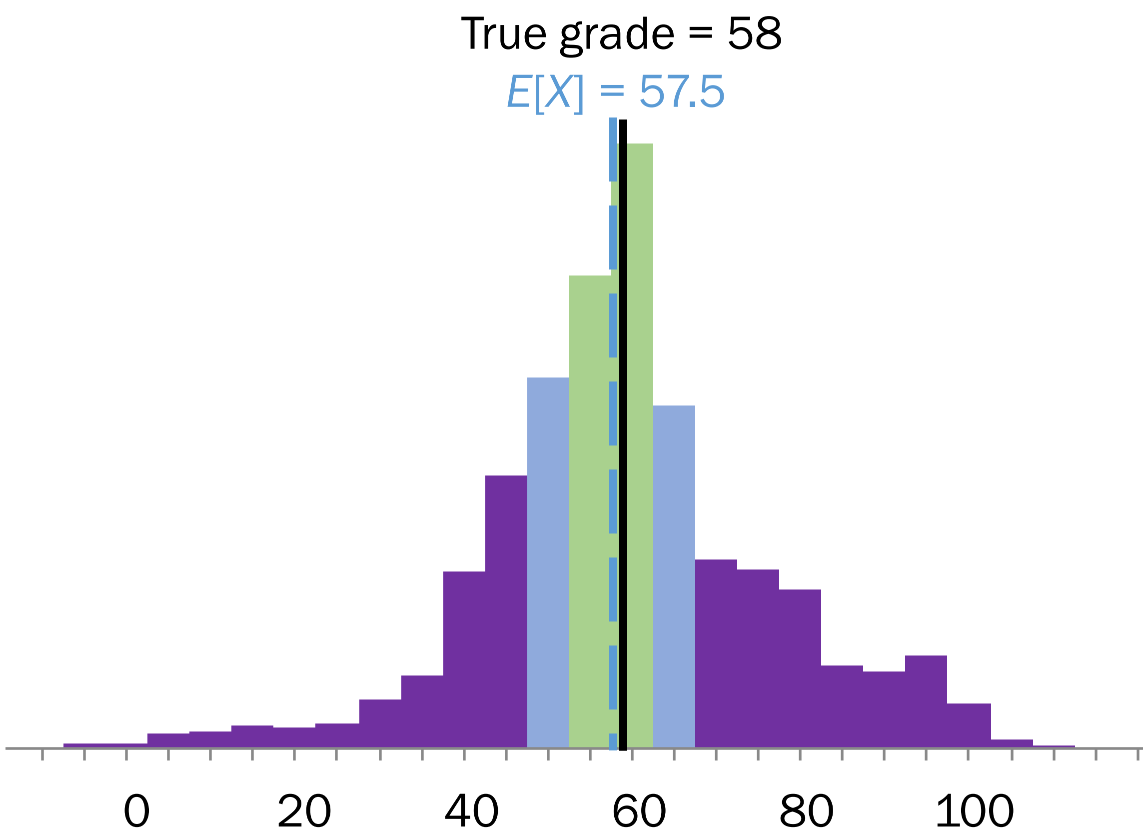
...

$$E[(X - \mu)^2] = 52 \text{ points}^2$$

Peer grading in Coursera HCI

Let X be a random variable that represents a peer grade

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



X	$(X - \mu)^2$
25 points	1056 points^2
80 points	506 points^2
50 points	56 points^2

...

$$E[(X - \mu)^2] = 52 \text{ points}^2$$

$$Std(X) = 7.2 \text{ points}$$

Computing Variance

$$\begin{aligned}
\mathrm{Var}(X) &= E[(X-\mu)^2] \\
&= \sum_x (x-\mu)^2 P(X=x) \\
&= \sum_x (x^2 - 2\mu x + \mu^2) P(X=x) \\
&= \sum_x x^2 P(X=x) - 2\mu \sum_x x P(X=x) + \mu^2 \sum_x P(X=x) \\
&= E[X^2] - 2\mu E[X] + \mu^2 \\
&= E[X^2] - 2\mu^2 + \mu^2 \\
&= E[X^2] - \mu^2 \\
&= \boxed{E[X^2] - (E[X])^2}
\end{aligned}$$

Law of unconscious statistician

Notation

$\mu = E[X]$

If extra time…

How do you get E[X^2]?

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

Unconscious statistician:

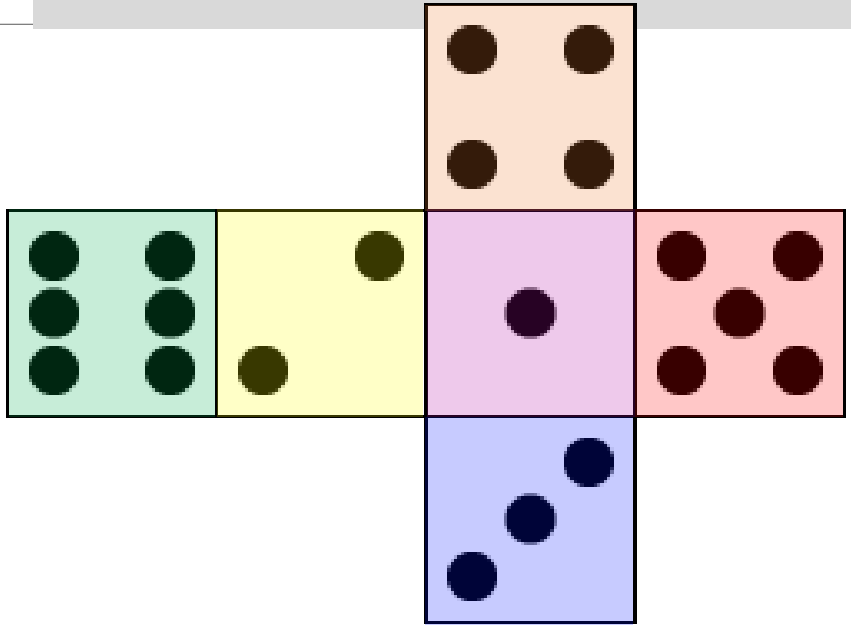
$$E[g(X)] = \sum_x g(x)P(X = x)$$

E[X^2]:

$$E[X^2] = \sum_x x^2 \cdot P(X = x)$$

Example: Variance of a Dice Roll

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

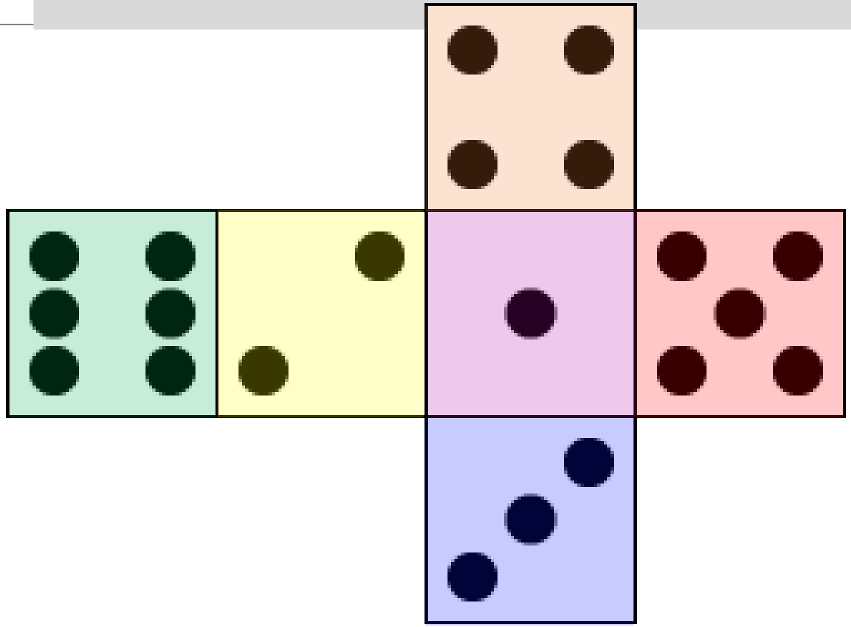


Let X be the result of rolling a 6 sided dice.

What is $\mathrm{Var}(X)$?

Example: Variance of a Dice Roll

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



Let X be the result of rolling a 6 sided dice.

What is Var(X)?

$$E[X] = 3.5$$

$$E[X^2] = 1^2 \cdot \frac{1}{6} + 2^2 \cdot \frac{1}{6} + 3^2 \cdot \frac{1}{6} + 4^2 \cdot \frac{1}{6} + 5^2 \cdot \frac{1}{6} + 6^2 \cdot \frac{1}{6} = \frac{91}{6}$$

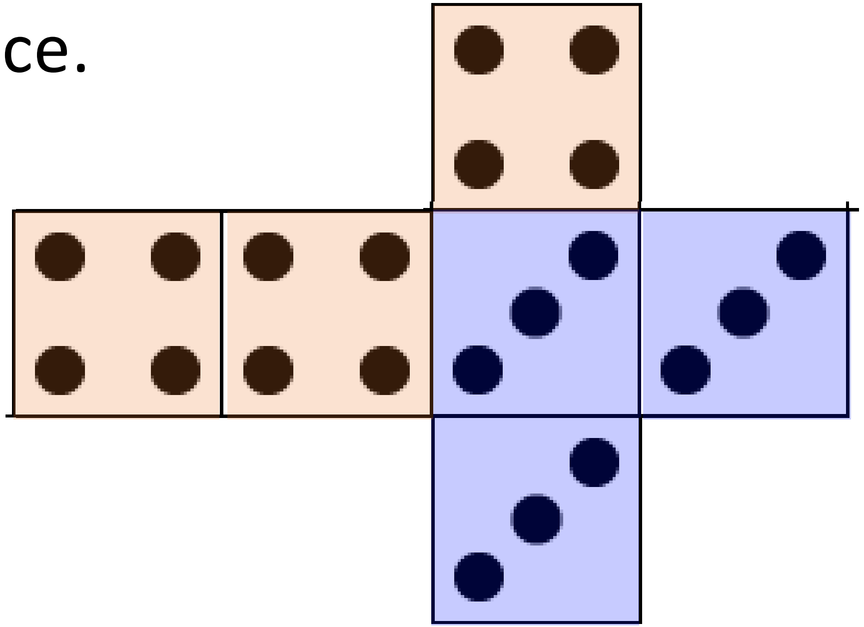
$$\begin{aligned}\text{Var}(X) &= E[X^2] - E[X]^2 \\ &= \frac{91}{6} - (3.5)^2 = 2.91\end{aligned}$$

Example: Variance of a Dice Roll

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

Let X be the result of rolling **this weird** 6 sided dice.

What is Var(X)?

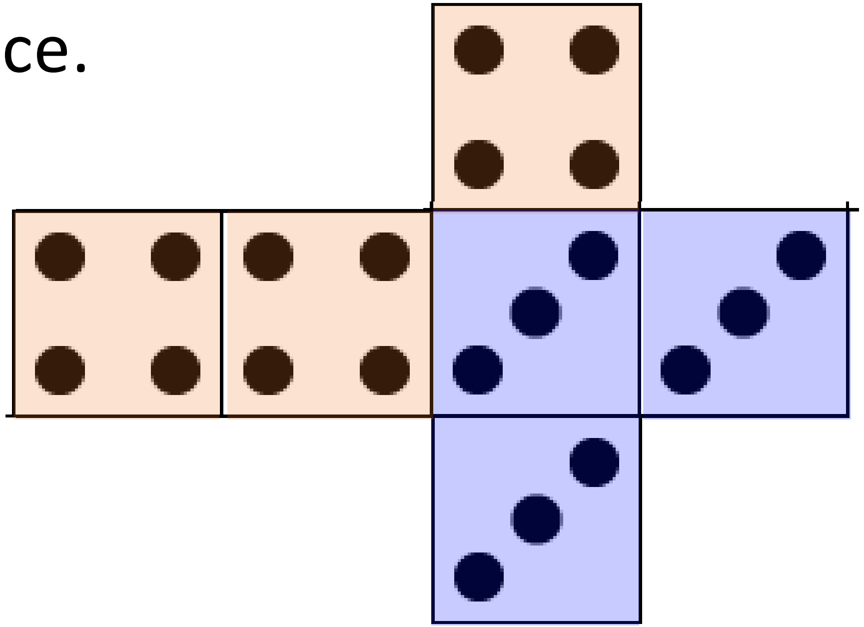


Example: Variance of a Dice Roll

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

Let X be the result of rolling **this weird** 6 sided dice.

What is Var(X)?



$$E[X] = 3.5$$

$$E[X^2] = 3^2 \cdot \frac{3}{6} + 4^2 \cdot \frac{3}{6} = 12.5$$

$$\begin{aligned}\text{Var}(X) &= E[X^2] - E[X]^2 \\ &= 12.5 - (3.5)^2 = 0.25\end{aligned}$$