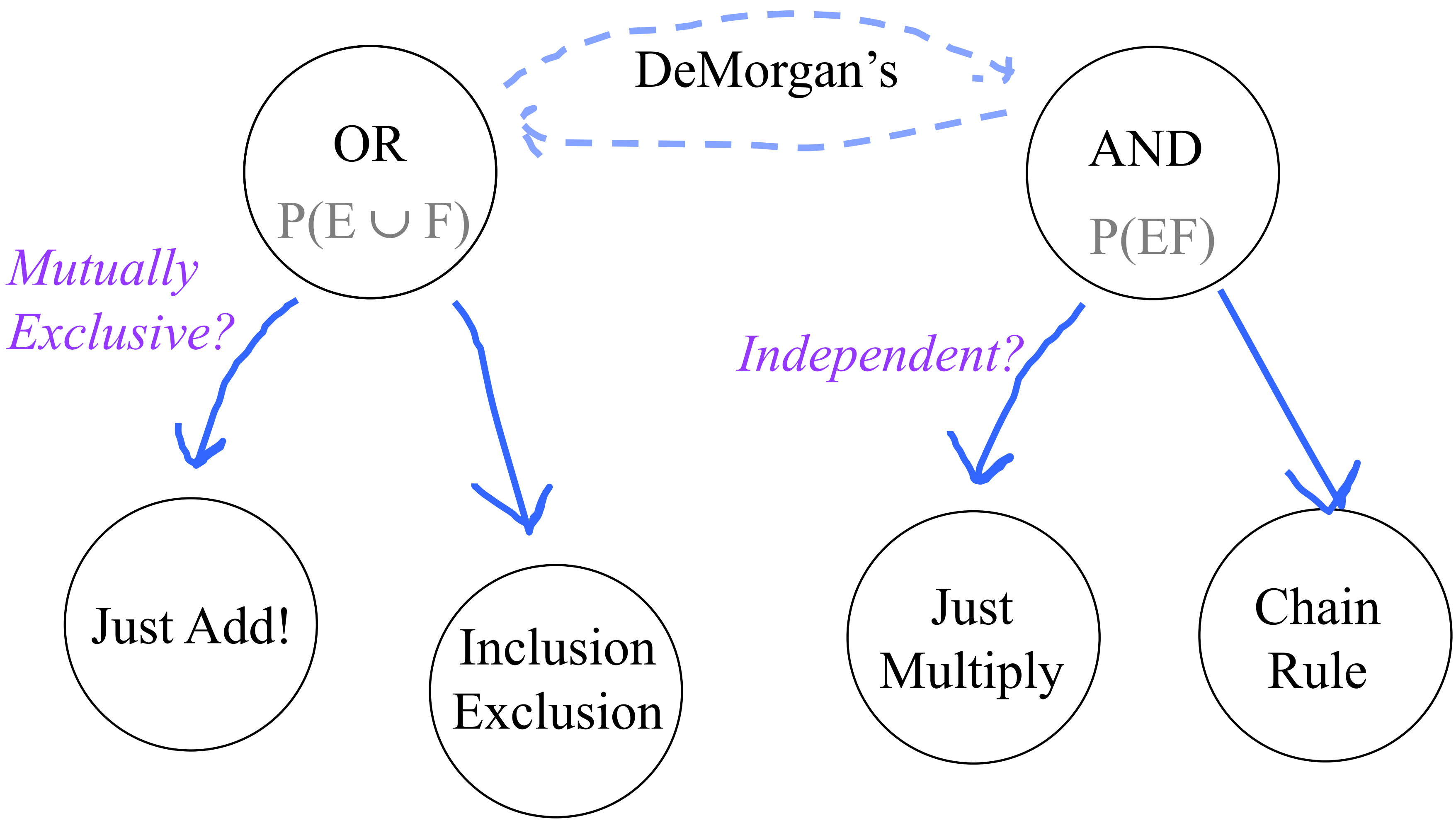
Random Variables

Noah Arthurs
CS109, Stanford University

Review

Earlier this Week



Independence

For events A and B, the following are equivalent:

A and B are independent

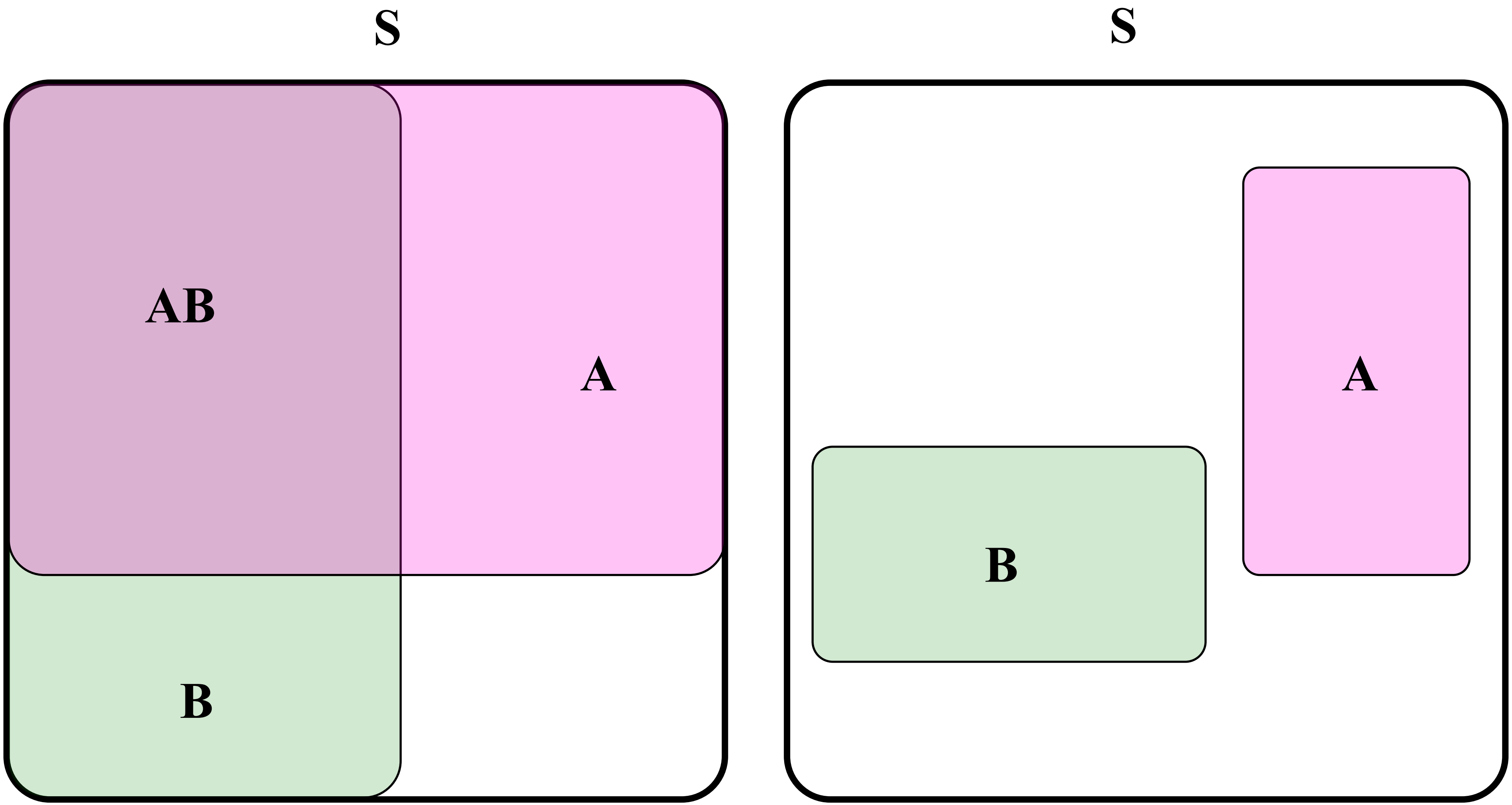
A and B are **not** dependent

$$P(AB) = P(A)P(B)$$

$$P(A|B) = P(A)$$

$$P(B|A) = P(B)$$

Independence vs. Mutual Exclusion



$$P(AB) = P(A)P(B)$$

$$P(AB) = 0$$

Generalized Independence

- General definition of Independence:

  Events E_1 , E_2 , ..., E_n are independent if **for every subset** with r elements (where $r \leq n$) it holds that:

$$P(E_{1'}E_{2'}E_{3'}\ldots E_{r'}) = P(E_{1'})P(E_{2'})P(E_{3'})\ldots P(E_{r'})$$

Two Dice

- Roll two 6-sided dice, yielding values D_1 and D_2
  - Let E be event: $D_1 = 1$
  - Let F be event: $D_2 = 6$
  - Are E and F independent? Yes!
- Let G be event: $D_1 + D_2 = 7$
  - Are E and G independent? Yes!
  - $P(E) = 1/6$, $P(G) = 1/6$, $P(E\ G) = 1/36$ [roll (1, 6)]
  - Are F and G independent? Yes!
  - $P(F) = 1/6$, $P(G) = 1/6$, $P(F\ G) = 1/36$ [roll (1, 6)]
  - Are E, F and G independent? No!
  - $P(EFG) = 1/36 \neq 1/216 = (1/6)(1/6)(1/6)$

No Transitive Property

Roll two 6-sided dice, yielding values D_1 and D_2

- Let E be event: $D_1 = 1$
- Let F be event: $D_2 = 6$
- Let G be event: $D_1 = 1$

We then have:

- E and F are independent
- F and G are independent
- E and G are dependent (they are the same exact event!)

Conditional Independence

- Two events E and F are called **conditionally independent given G**, if

$$P(EF|G) = P(E|G)P(F|G)$$

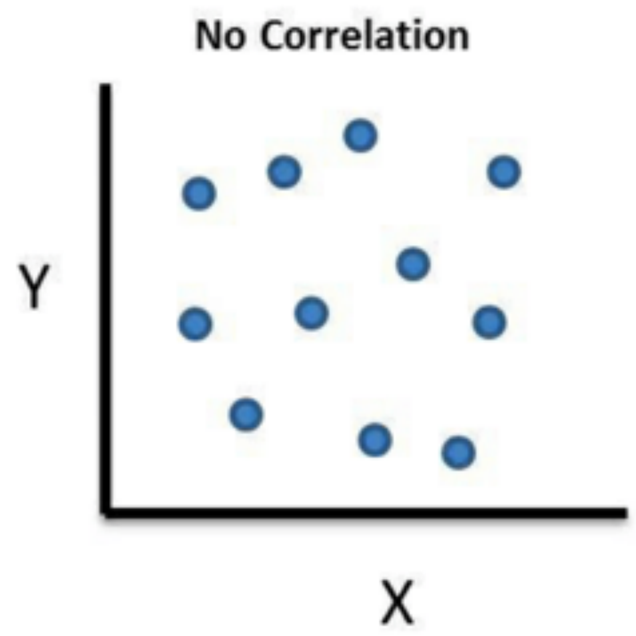
- Or, equivalently if:

$$P(E|FG) = P(E|G)$$

Conditional Independence

What does conditional independence tell you about independence and vice versa?

Nothing



End Review

Remember Learning to Code?

name

type

value

```
int a = 5;
double b = 4.2;
bit c = 1;
choice d = medium;
```

$$z \in \{\text{high}, \text{medium}, \text{low}\}$$

Random variables are like programming variables, with uncertainty

Pirates of the Random Variables

```
int a = 5;
```

A is the number of pirate ships in our *future* armada.

$$A \in \{1, 2, \ldots, 10\}$$

```
double b = 4.2;
```

B is the amount of money we get *after* we defeat Blackbeard.

$$B \in \mathbb{R}^+$$

```
bit c = 1;
```

C is 1 *if* we successfully defeat Blackbeard. 0 otherwise.

$$C \in \{0, 1\}$$

Random Variable

- A **Random Variable** is a variable will have a value. But there is uncertainty as to what value.
- Example:
  - 3 fair coins are flipped.
  - Y = number of “heads” on 3 coins
  - **Y is a random variable**
  - P(Y = 0) = 1/8 (T, T, T)
  - P(Y = 1) = 3/8 (H, T, T), (T, H, T), (T, T, H)
  - P(Y = 2) = 3/8 (H, H, T), (H, T, H), (T, H, H)
  - P(Y = 3) = 1/8 (H, H, H)
  - P(Y ≥ 4) = 0

It is confusing that both random variables and events use the same notation

Random variables and events are two *different* things

We can define an event to be a particular assignment to a random variables

Example Random Variable

- Consider 5 coin flips, each which independently come up heads with probability *p*
  - Recall:

$$P(2 \text{ heads}) = \binom{5}{2} p^2 (1-p)^3$$

$$P(3 \text{ heads}) = \binom{5}{3} p^3 (1-p)^2$$

  - Y = number of "heads" on 5 flips

$$Y \in \{0, 1, 2, 3, 4, 5\}$$

$$P(Y = k) = \binom{5}{k} p^k (1-p)^{5-k}$$

* Pro tip: no coin works like this… but many real world binary events do

Fun with Random Variables

- Probability Mass Function:

$$P(X = a)$$

- Expectation:

$$E[X]$$

- Variance:

$$\text{Var}(X)$$

Learning goals for today

1. Probability Mass Function

All the different assignments to a random variable make a function

Let Y be a random variable

$$Y$$

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

$$Y = 2$$

It is an event when
Y takes on a value

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

If this is a number

$$P(Y = 2)$$

Then this is a number (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

Random Variables -> Functions

$$P(Y = k)$$

$$k = 5$$

$$0.03125$$

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

Random Variables -> Functions

$$P(Y = k)$$

```java
private double eventProbability(int k) {
	int ways = choose(N, k);
	double a = Math.pow(P, k);
	double b = Math.pow(P, N-k);
	return ways * a * b;
}

private static final int N = 5;
private static final double P = 0.5;
```

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

If a random variable is discrete we call this function the **Probability Mass** ***Function***

Probability Mass Function

Let X be a random variable that represents the result of a **single dice roll**. X can take on the values {1, 2, 3, 4, 5, 6}

$$P(X = x)$$

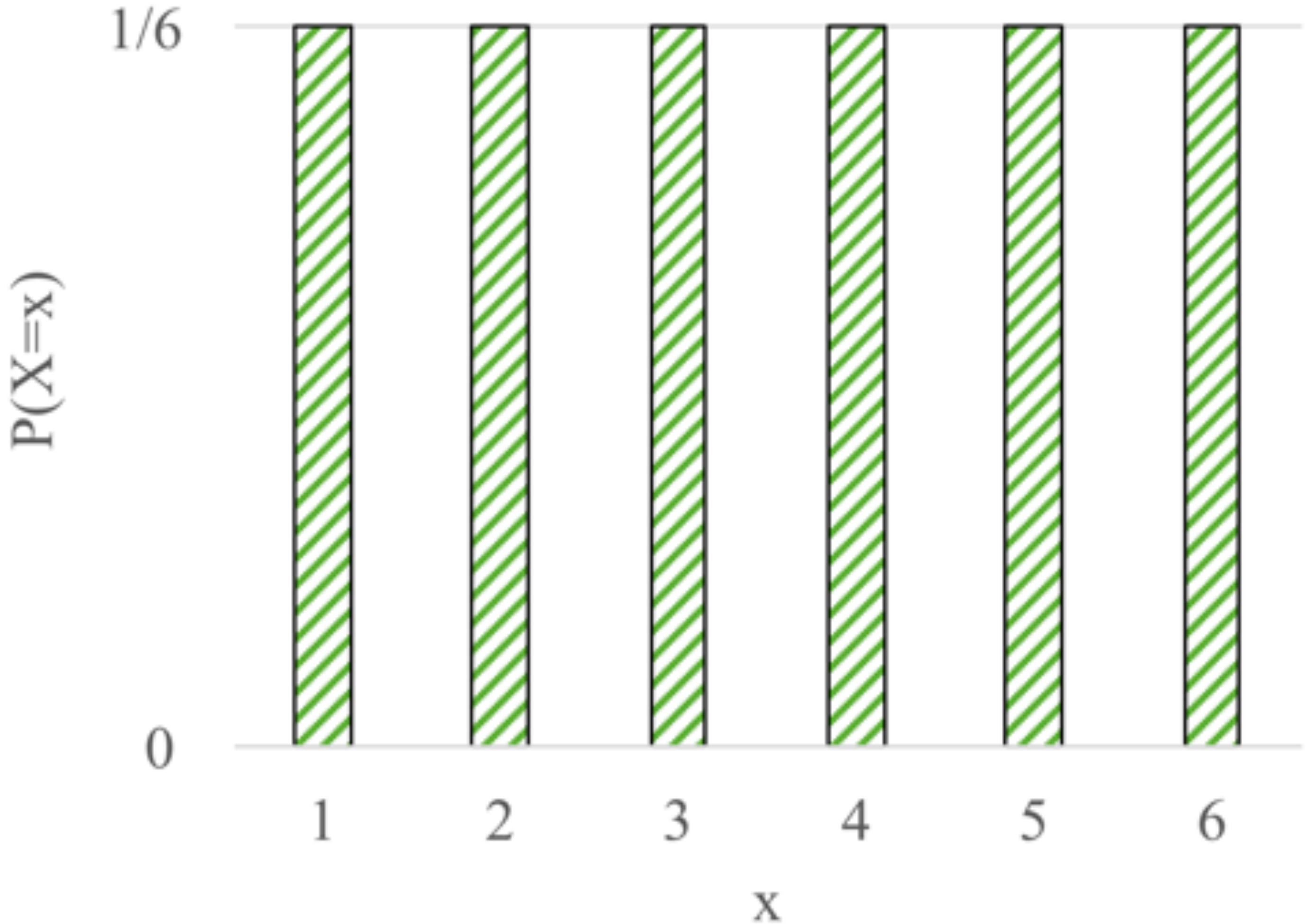
$$p(x)$$

This is shorthand notation for the PMF

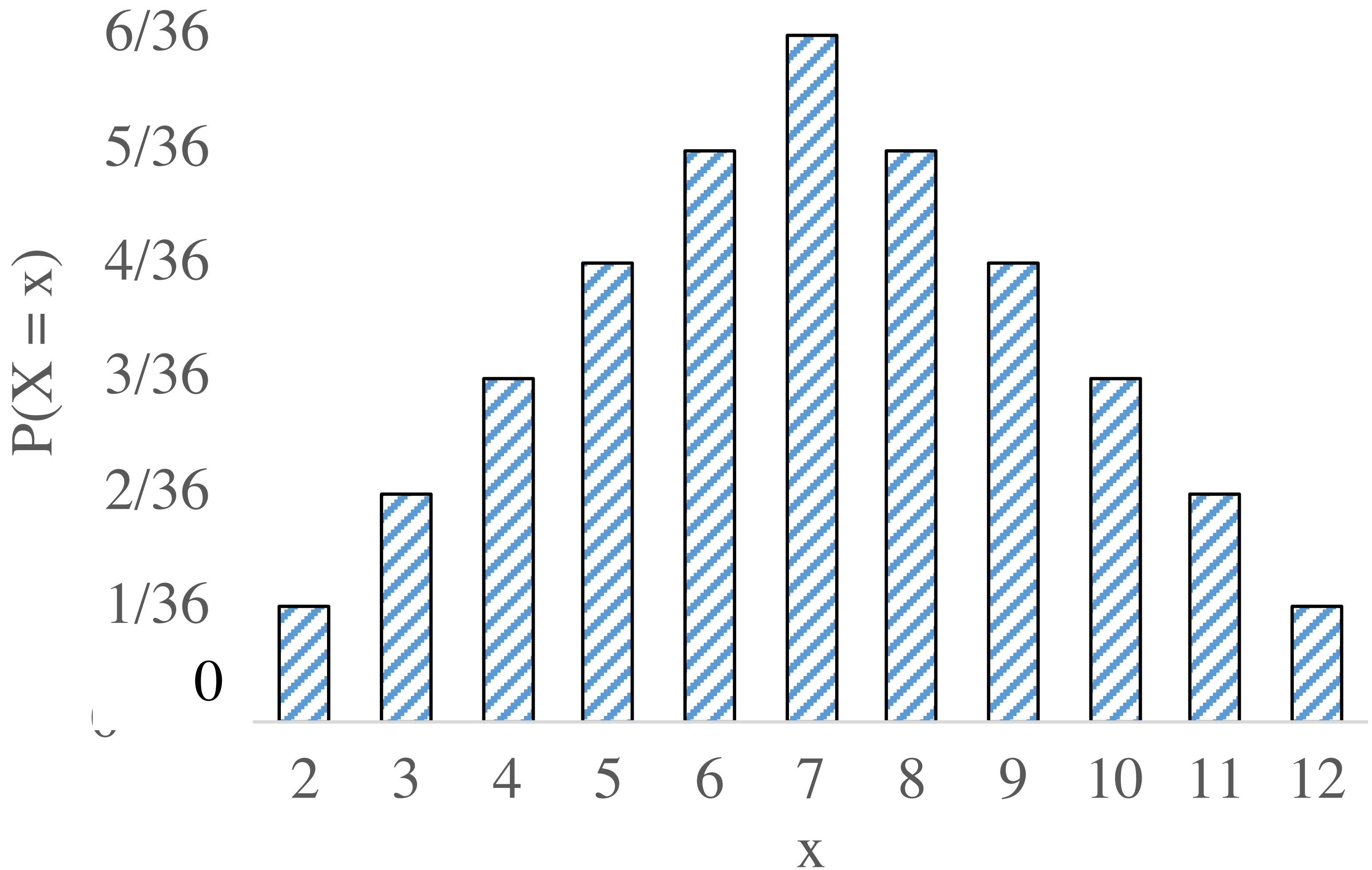
$$p_X(x)$$

This is also shorthand notation for the PMF

PMF For a Single 6 Sided Dice



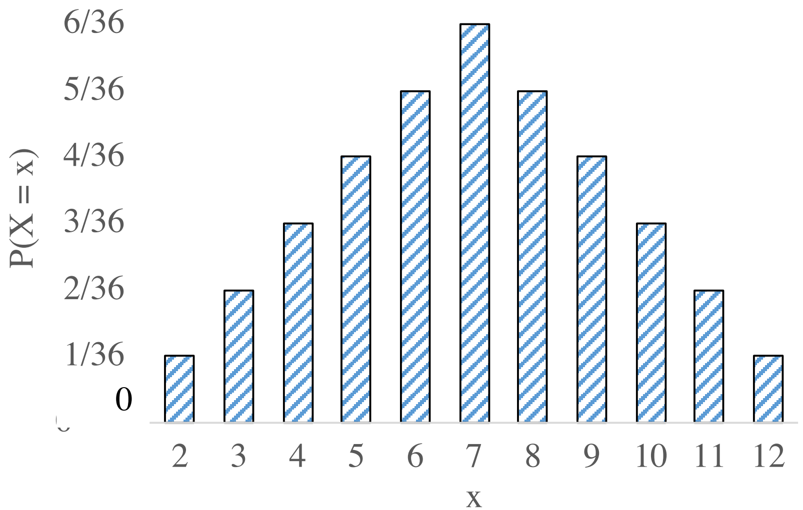
PMF for the sum of two dice



PMF as an Equation

$$p(X = x) = \begin{cases} \frac{x-1}{36} & \text{if } x \in \mathbb{Z}\ ,\ 1 \leq x \leq 6 \\ \frac{13-x}{36} & \text{if } x \in \mathbb{Z}\ ,\ 7 \leq x \leq 12 \\ 0 & \text{else} \end{cases}$$

Again, this is the probability for the sum of two dice

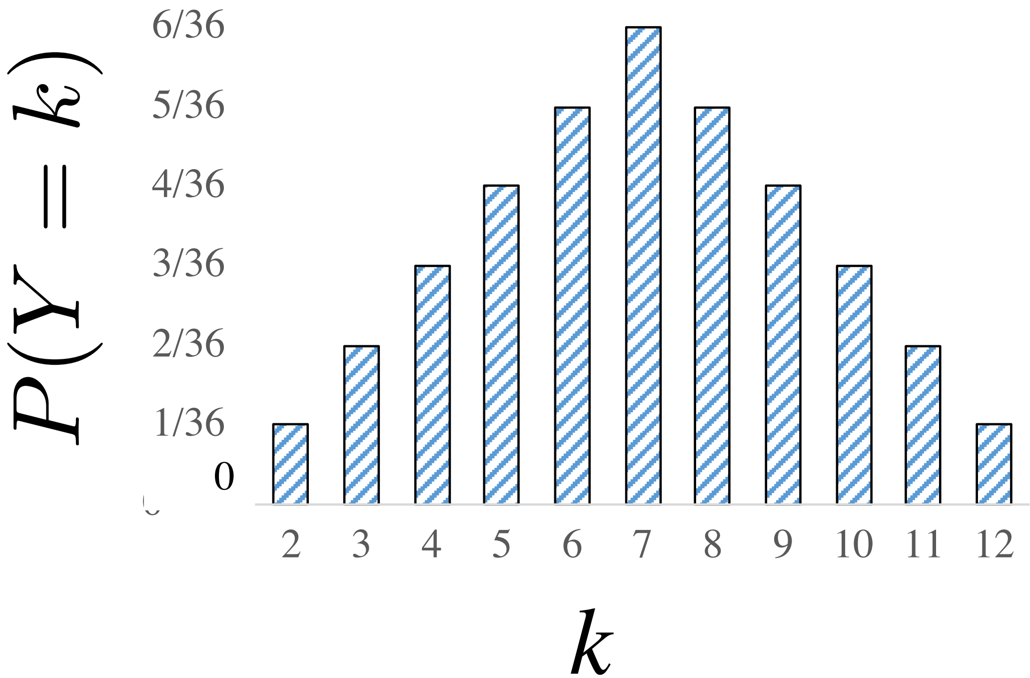


Sanity Check

$$\sum_{\text{all } k} P(Y = k) \stackrel{?}{=}$$

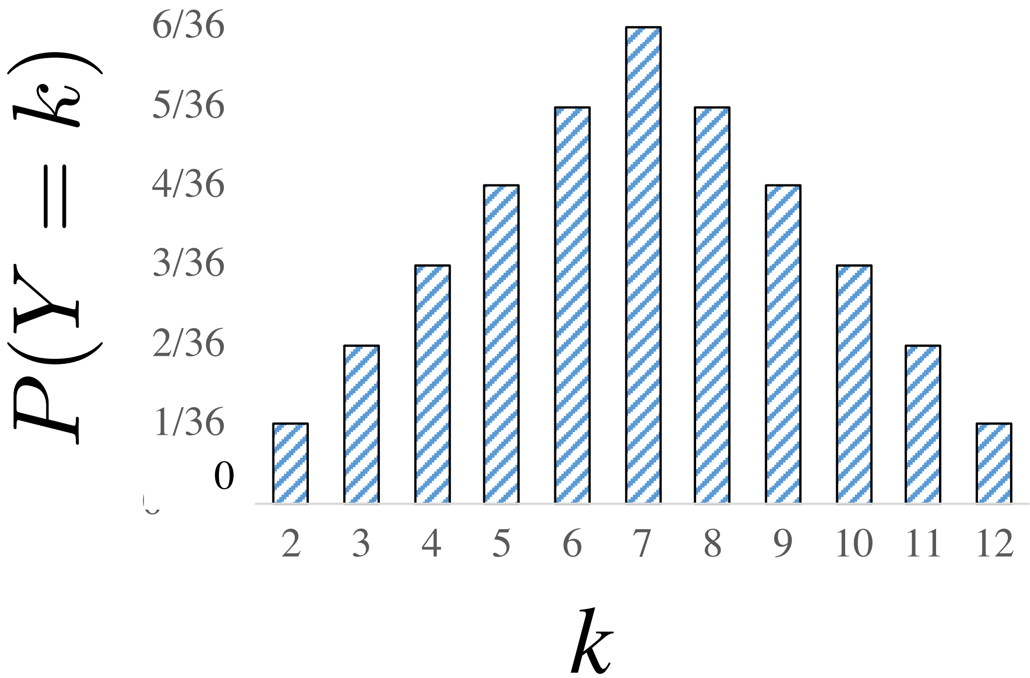
Sanity Check

$$\sum_{\text{all } k} P(Y = k) \stackrel{?}{=}$$



Sanity Check

$$\sum_k P(Y = k) = 1$$



2. Expectation

Expected Value

- The Expected Values for a discrete random variable X is defined as:

$$E[X] = \sum_{x:p(x)>0} x \cdot p(x)$$

- Note: sum over all values of x that have $p(x) > 0$.

- Expected value also called: ***Mean***, *Expectation*, ***Weighted Average***, ***Center of Mass***, *1st Moment*

Expected Value

- Roll a 6-Sided Die. X is outcome of roll
  - p(1) = p(2) = p(3) = p(4) = p(5) = p(6) = 1/6
- E[X] = $1\left(\frac{1}{6}\right)+2\left(\frac{1}{6}\right)+3\left(\frac{1}{6}\right)+4\left(\frac{1}{6}\right)+5\left(\frac{1}{6}\right)+6\left(\frac{1}{6}\right)=\frac{7}{2}$
- Y is random variable
  - P(Y = 1) = 1/3,   P(Y = 2) = 1/6,   P(Y = 3) = 1/2
- E[Y] = 1 (1/3) + 2 (1/6) + 3 (1/2) = 13/6

Lying with Statistics

"There are three kinds of lies:
lies, damned lies, and statistics"
– *Mark Twain*

- School has 3 classes with 5, 10 and 150 students
- Randomly choose a <u>class</u> with equal probability
- X = size of chosen class
- What is E[X]?
  - E[X] = 5 (1/3) + 10 (1/3) + 150 (1/3)
    = 165/3 = 55

Lying with Statistics

"There are three kinds of lies:
lies, damned lies, and statistics"
– *Mark Twain*

- School has 3 classes with 5, 10 and 150 students
- Randomly choose a <u>student</u> with equal probability
- Y = size of class that student is in
- What is E[Y]?
  - $E[Y] = 5\,(5/165) + 10\,(10/165) + 150\,(150/165)$
    $= 22635/165 \approx 137$
- Note: E[Y] is students' perception of class size
  - But E[X] is what is usually reported by schools!

Properties of Expectation

- **Linearity**:

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

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

- **Expectation of a sum** is the sum of expectations

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

- **Unconscious statistician**:

$$E[g(x)] = \sum_x g(x)p(x)$$

Wonderful

St Petersburg

- Game set-up
  - We have a fair coin (come up “heads” with p = 0.5)
  - Let n = number of coin flips (“heads”) before first “tails”
  - You win \$$2^n$

- How much would you pay to play?

St Petersburg

- Game set-up
  - We have a fair coin (come up "heads" with p = 0.5)
  - Let n = number of coin flips ("heads") before first "tails"
  - You win $\$2^n$
- How much would you pay to play?
- Solution
  - Let X = your winnings
  - E[X] = $\left(\frac{1}{2}\right)^1 2^0 + \left(\frac{1}{2}\right)^2 2^1 + \left(\frac{1}{2}\right)^3 2^2 + \left(\frac{1}{2}\right)^4 2^3 + \ldots = \sum_{i=0}^{\infty} \left(\frac{1}{2}\right)^{i+1} 2^i$

    $= \sum_{i=0}^{\infty} \frac{1}{2} = \infty$
  - I'll let you play for $1 thousand... but just once! Takers?

St Petersburg + Reality

- What if Noah has only \$65,536?
  - Same game
  - If you win over \$65,536 I leave the country.
- Solution
  - Let X = your winnings
  - E[X]

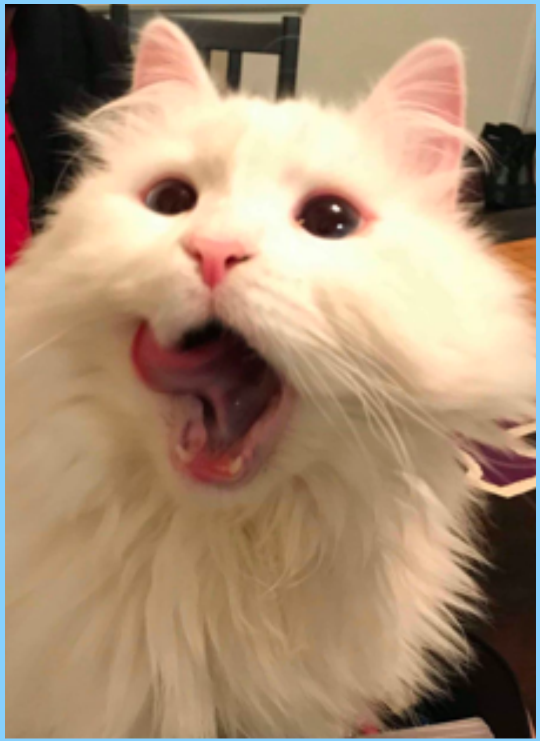
$$= \left(\frac{1}{2}\right)^1 2^0 + \left(\frac{1}{2}\right)^2 2^1 + \left(\frac{1}{2}\right)^3 2^2 + \left(\frac{1}{2}\right)^4 2^3 + \dots$$

$$= \sum_{i=0}^{k} \left(\frac{1}{2}\right)^{i+1} 2^i \text{ s.t. } k = \log_2(65,536)$$

$$= \sum_{i=0}^{16} \frac{1}{2} = 8.5$$

Learning Goals

1. Be able to use conditional independence
2. Be able to define a random variable (R.V.)
3. Be able to use + produce a PMF of a R.V.
4. Be able to calculate the expectation of the R.V.




Stretch!
LEMON-LIME SODA
100% NATURAL FLAVORS
NO CAFFEINE
PERFECTLY CLEAR
Sprite
24
CANS

Debugging Intuition

- How to calculate the probability of at least *k* successes in *n* trials?
  - X is number of successes in *n* trials each with probability *p*
  - $P(X \geq k) =$

$$\binom{n}{k} p^k$$

Don't care about the rest

\# ways to choose slots for success

Probability that each is success

First clue that something is wrong. Think about *p* = 1

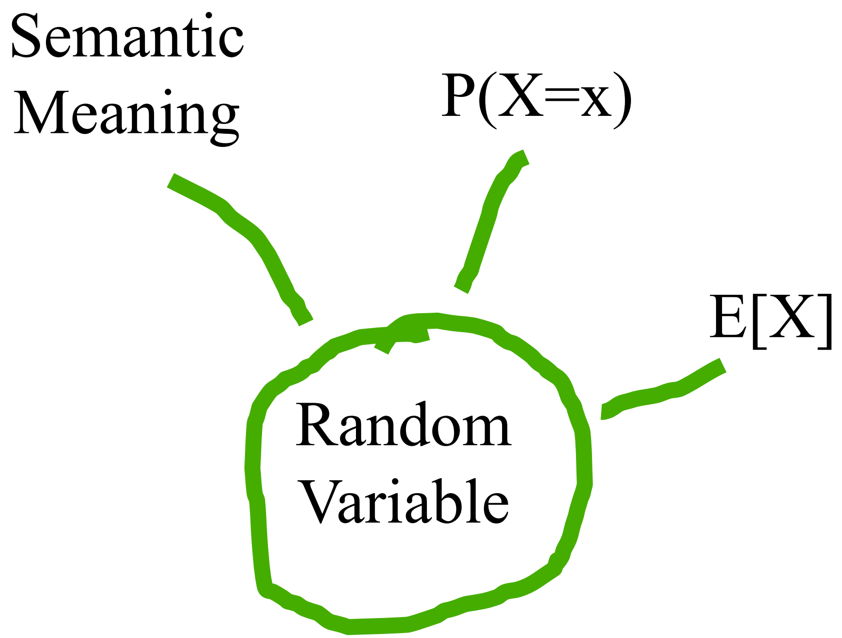
Not mutually exclusive…

Correct: $P(X \geq k) = \sum_{i=k}^{n} \binom{n}{i} p^i (\mathbf{1} - p)^{n-i}$

Variance

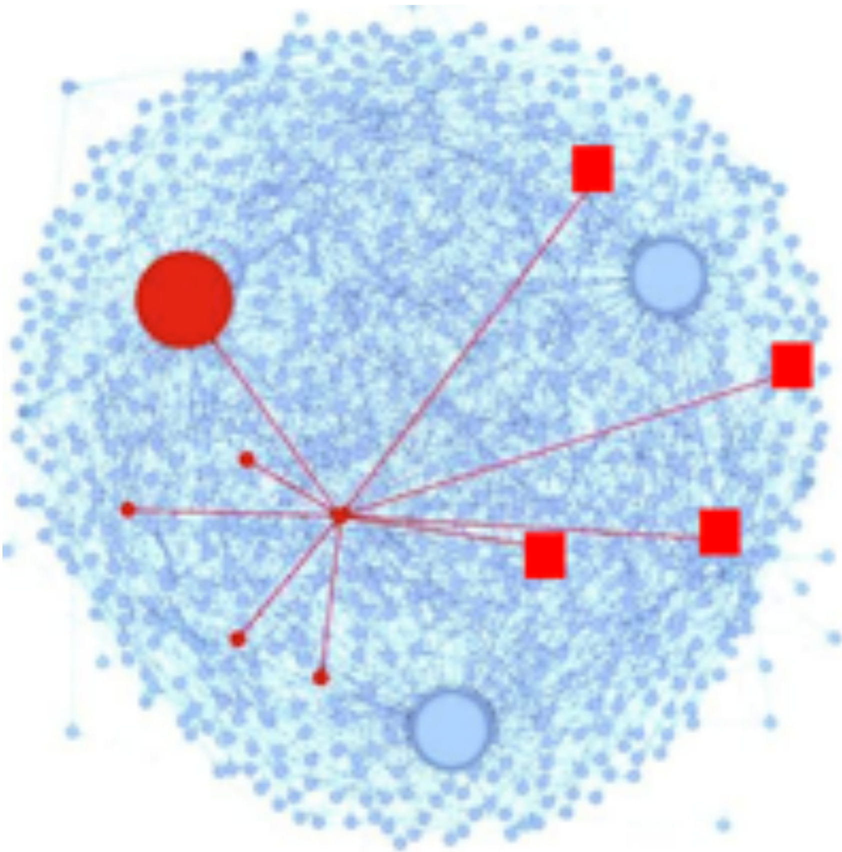
Noah Arthurs
CS109, Stanford University

Fundamental Properties



Is E[X] enough?

Is Peer Grading Accurate Enough?

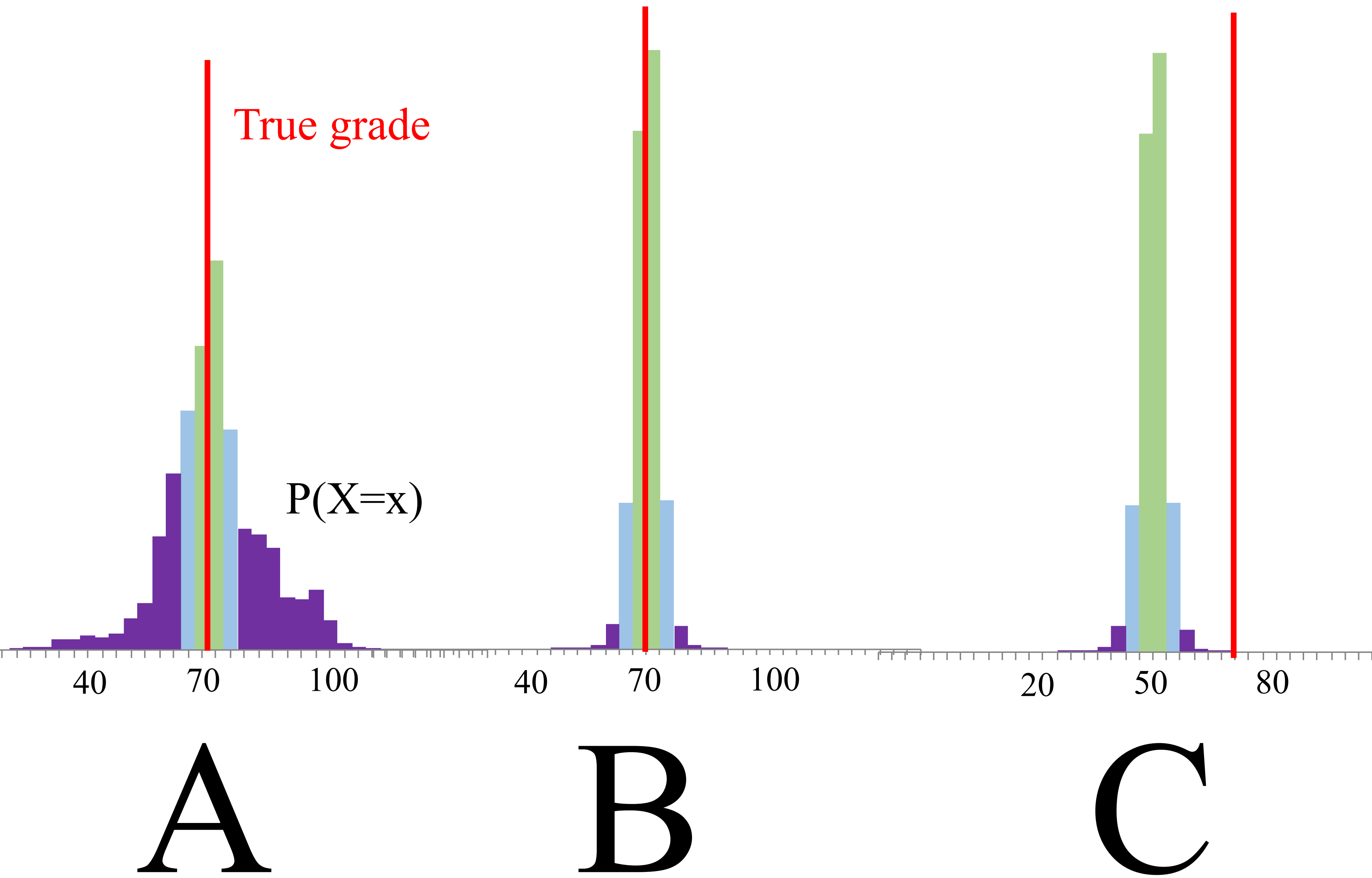


Peer Grading on Coursera HCI.

31,067 peer grades for 3,607 students.

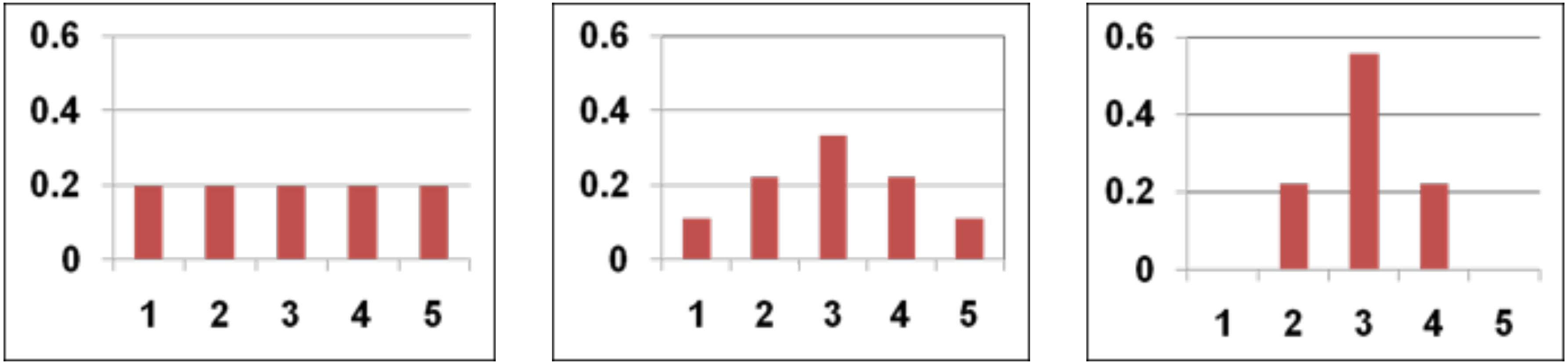
Tuned Models of Peer Assessment. C Piech, J Huang, A Ng, D Koller

X is the score a peer grader gives to an assignment submission



Variance

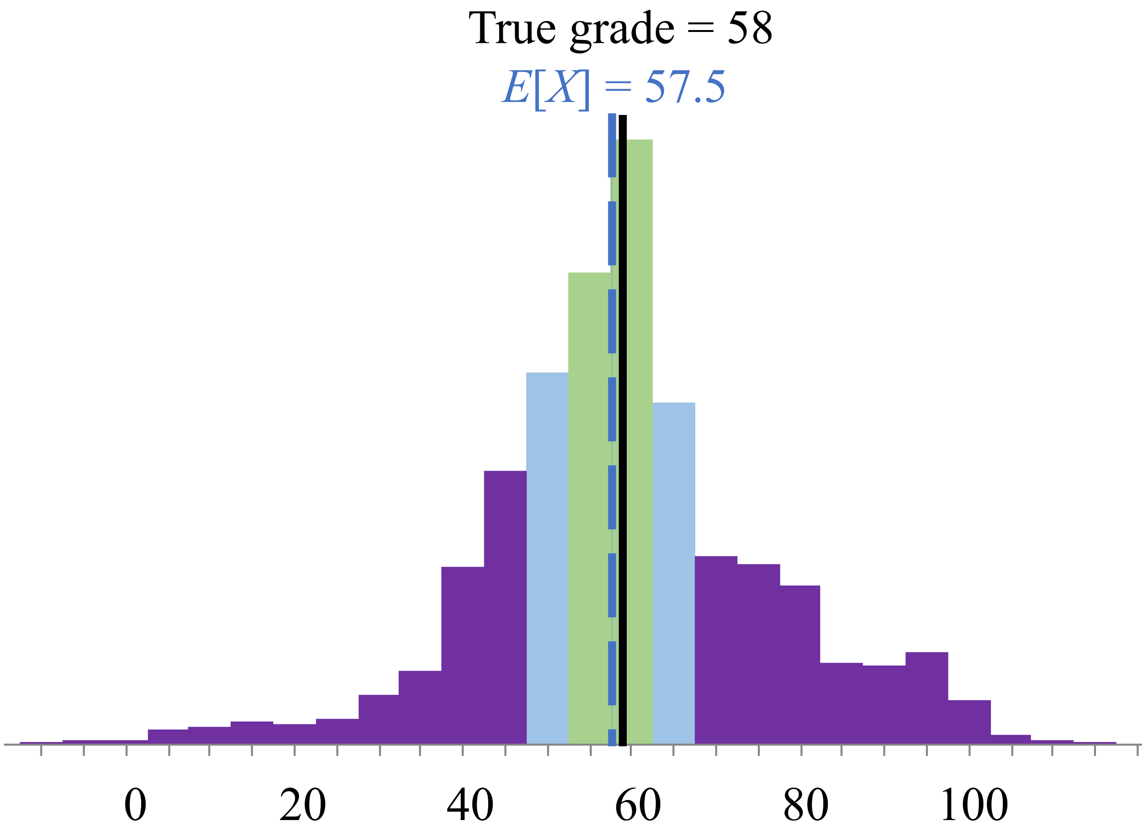
- Consider the following 3 distributions (PMFs)



- All have the same expected value, E[X] = 3
- But “spread” in distributions is different
- Variance = a formal quantification of “spread”

Peer Grades in Coursera HCI

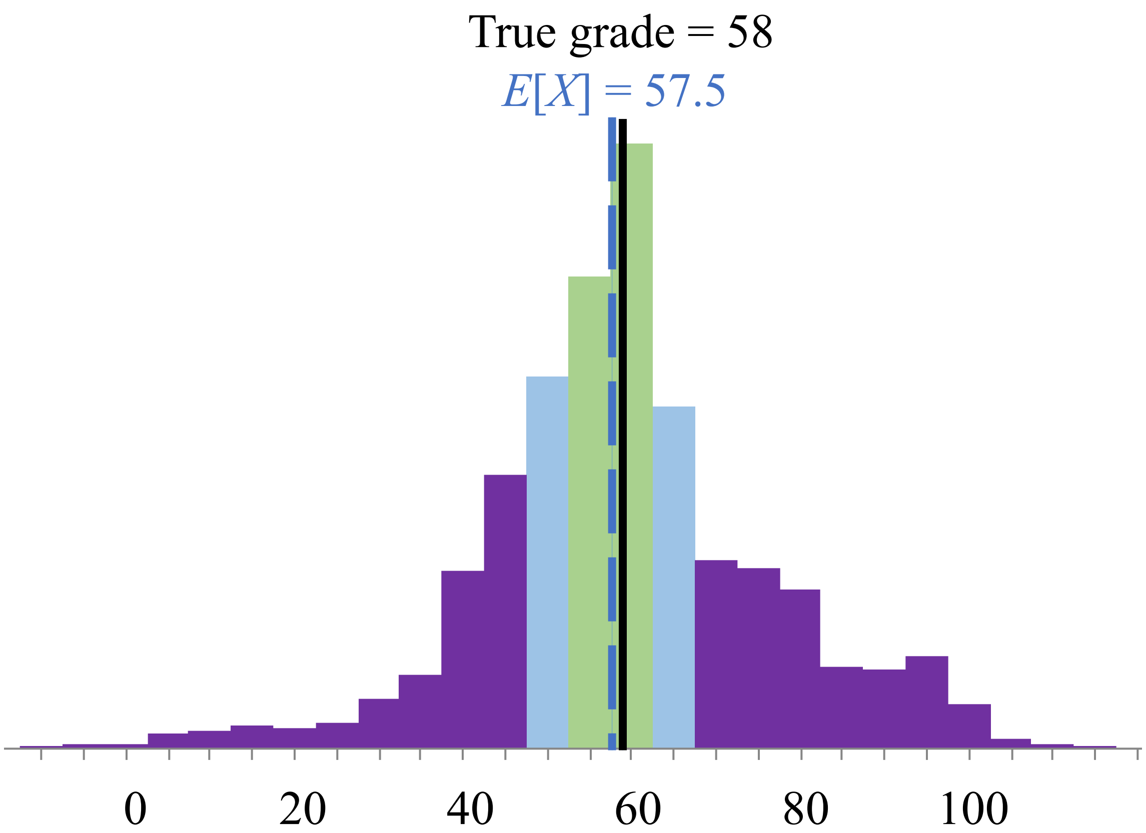
Let X be a random variable that represents a peer grade



Peer Grades in Coursera HCI

Let X be a random variable that represents a peer grade

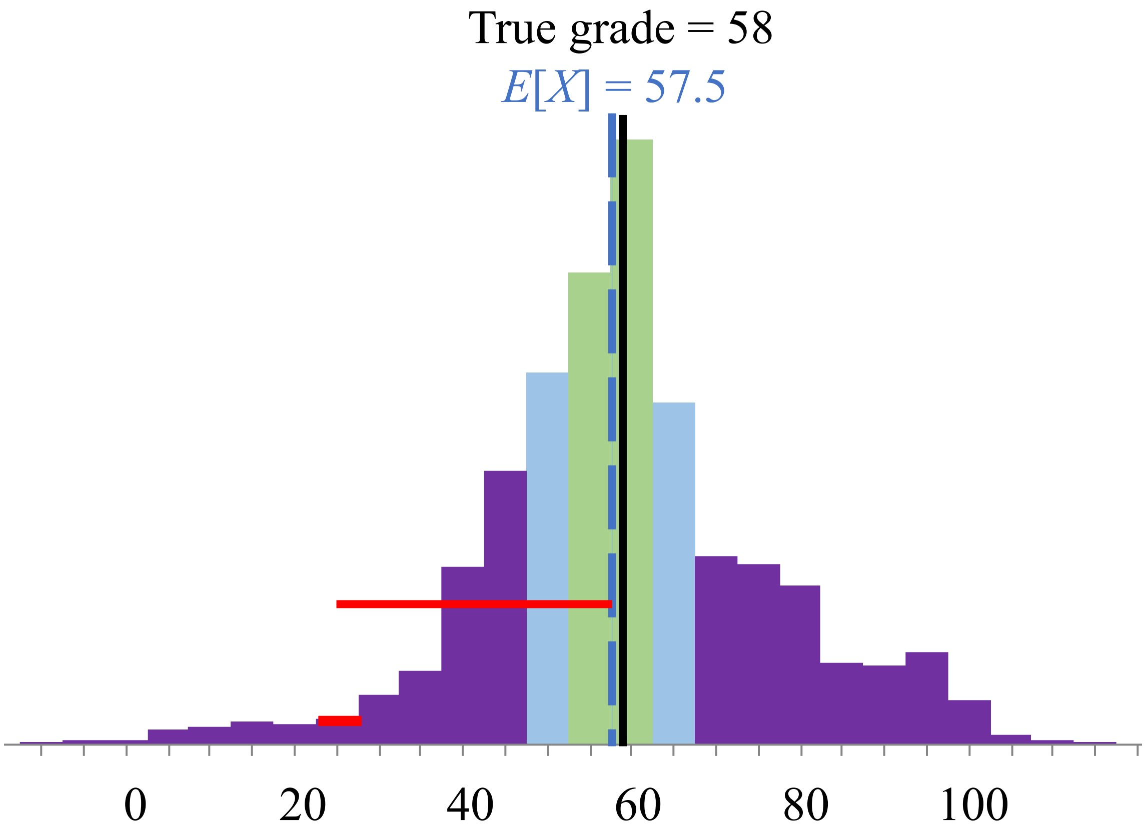
$$\mathrm{Var}(X) = E[(X - \mu)^2]$$



Peer Grades in Coursera HCI

Let X be a random variable that represents a peer grade

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

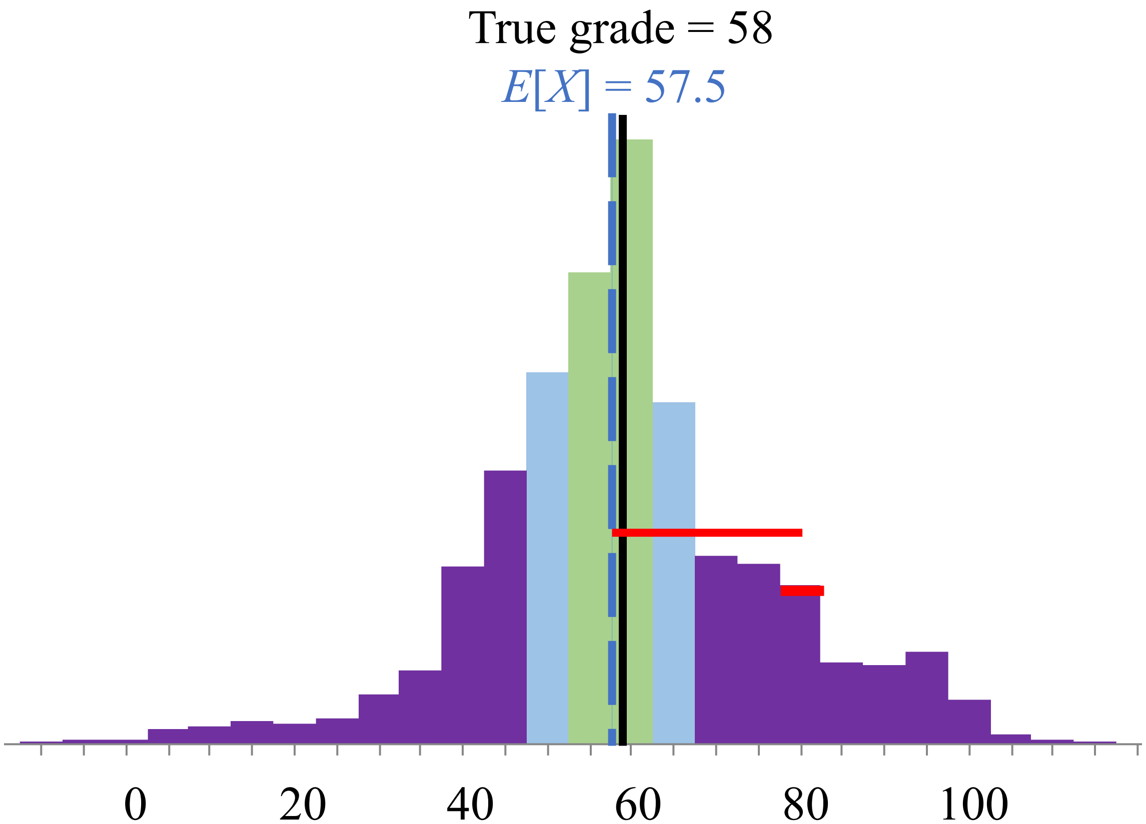


X	$(X - \mu)^2$
25 points	1056 points^2

Peer Grades in Coursera HCI

Let *X* be a random variable that represents a peer grade

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

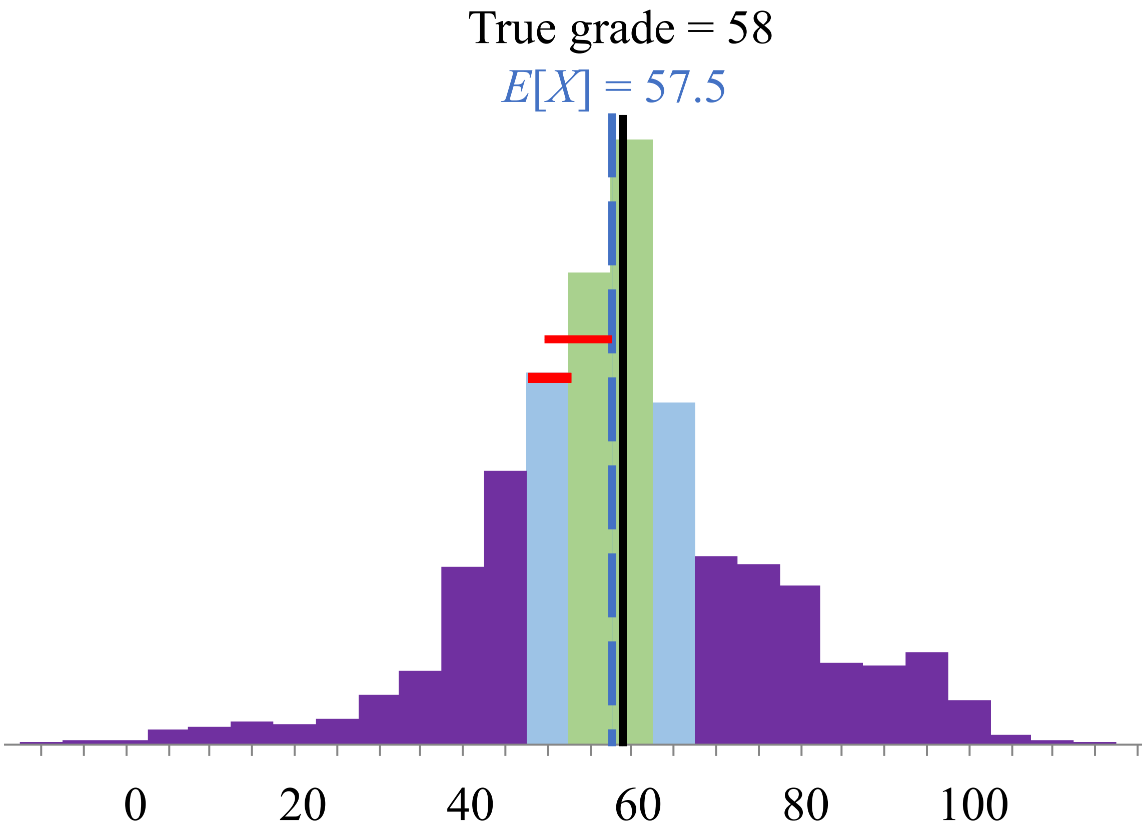


X	$(X - \mu)^2$
25 points	1056 points^2
80 points	506 points^2

Peer Grades in Coursera HCI

Let *X* be a random variable that represents a peer grade

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

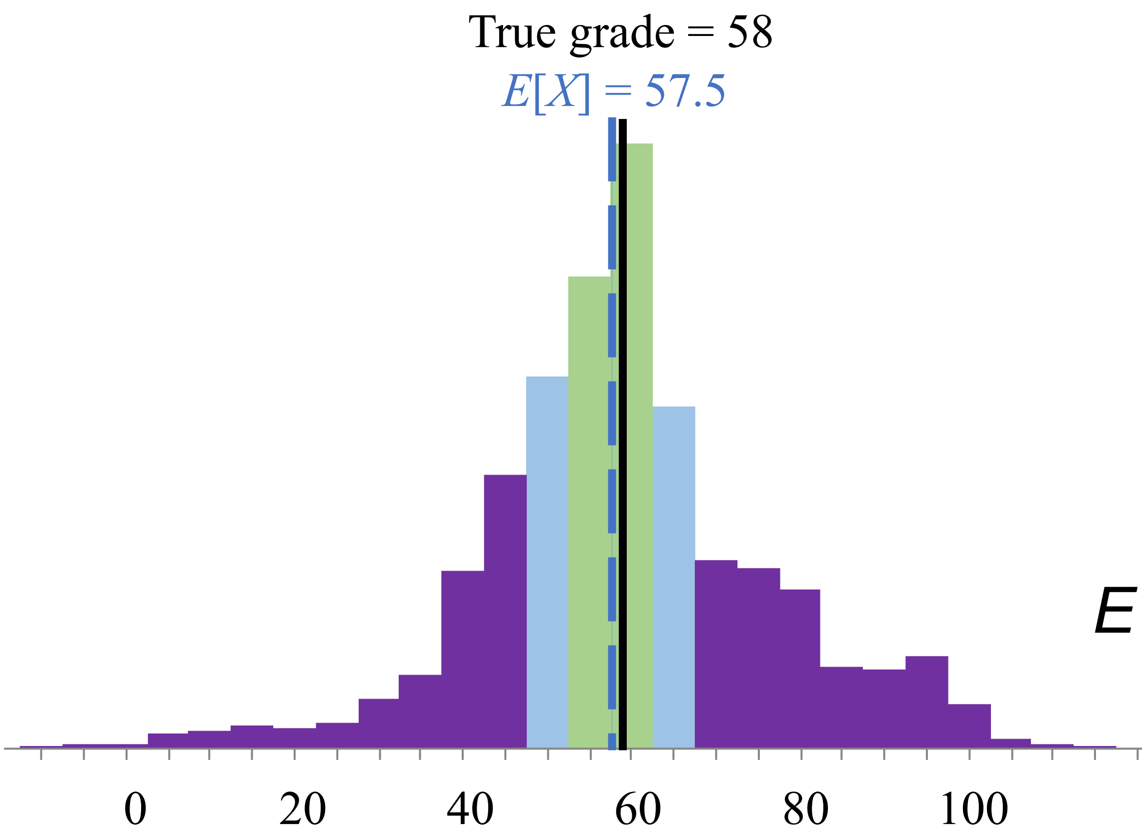


X	$(X - \mu)^2$
25 points	1056 points^2
80 points	506 points^2
50 points	56 points^2

Peer Grades 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

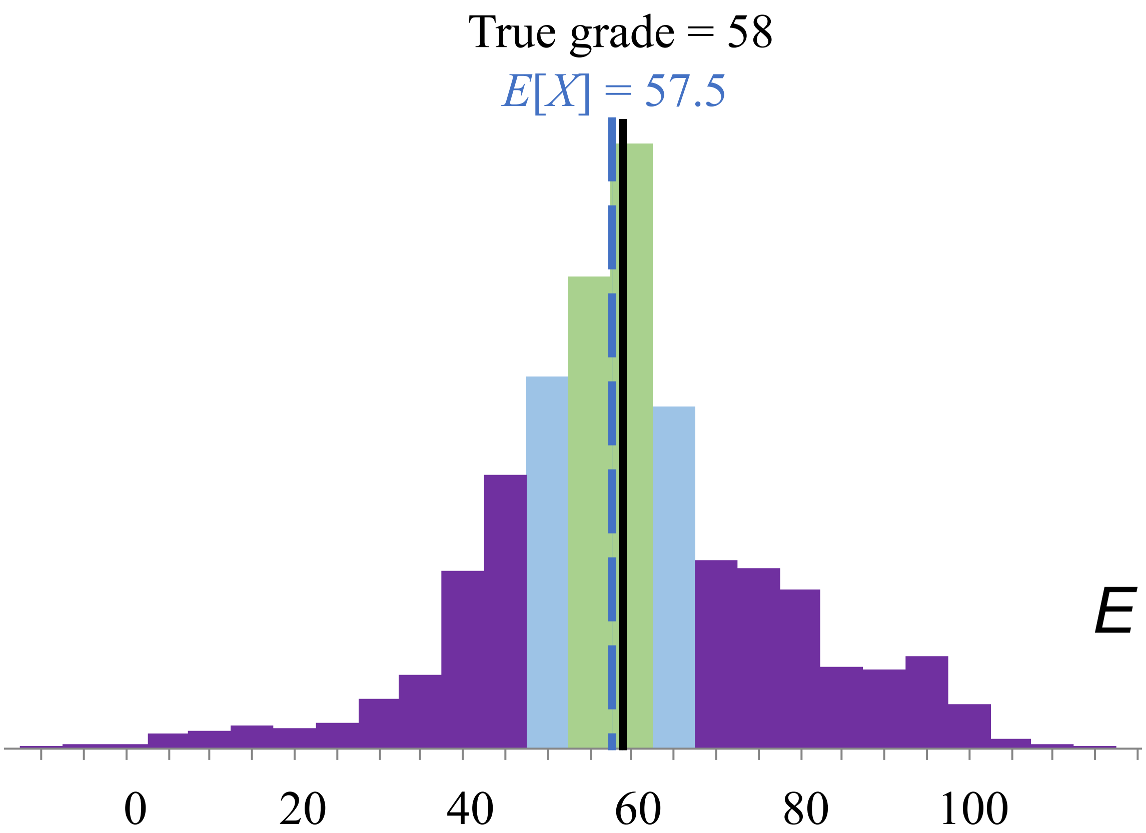
…

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

Peer Grades 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
	…

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

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

Variance

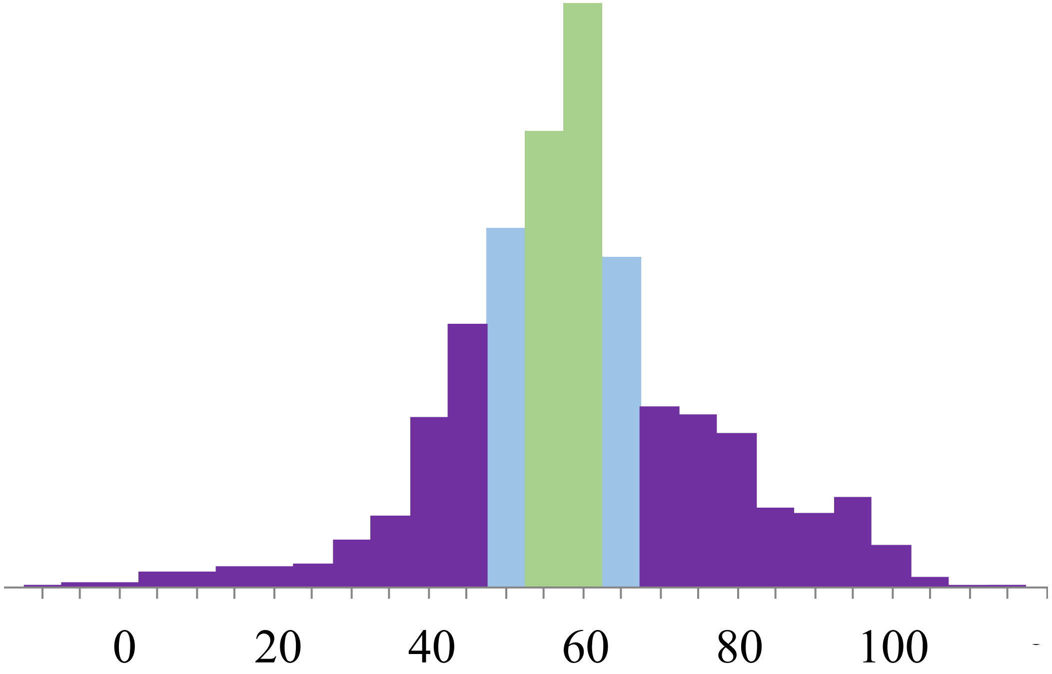
- If X is a random variable with mean μ then the **variance** of X, denoted Var(X), is:

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

- Note: Var(X) ≥ 0

- Also known as the 2nd **Central** Moment, or square of the Standard Deviation

Normalized **histograms** are approximations of **probability mass functions**



Computing Variance

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

Note: $\mu = E[X]$

Ladies and gentlemen, please welcome the 2nd moment!

Variance of a 6 sided dice

- Let X = value on roll of 6 sided die
- Recall that E[X] = 7/2
- Compute $E[X^2]$

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

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

$$= \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}$$