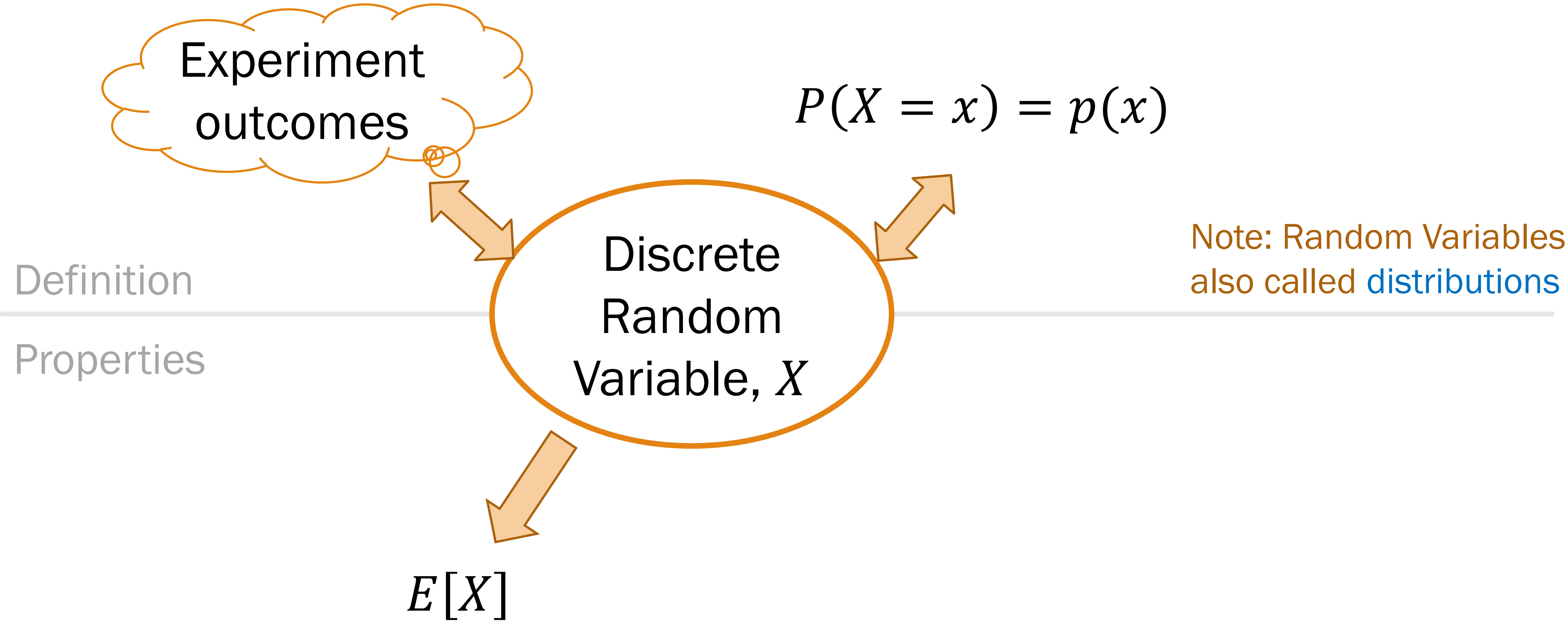
07: Variance, Bernoulli, Binomial

David Varodayan

January 22, 2020

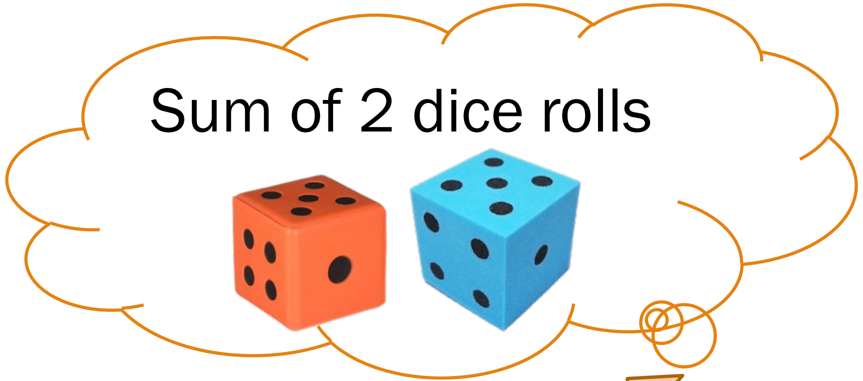
Adapted from slides by Lisa Yan

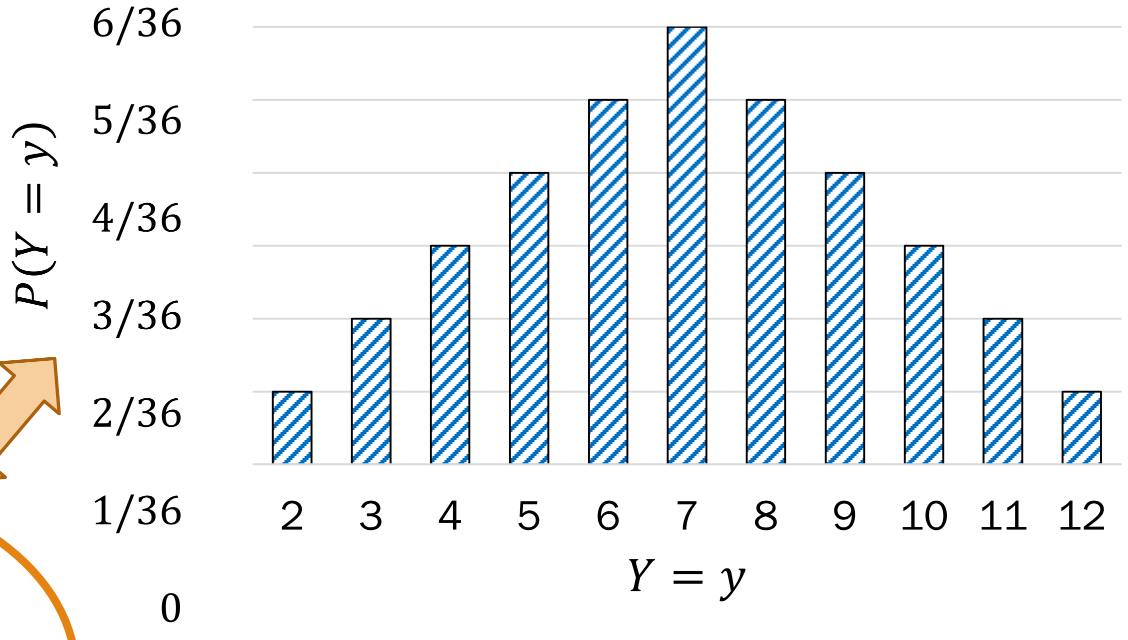
Discrete random variables



Note: Random Variables also called distributions

Sum of 2 dice rolls





Definition

Properties

Discrete Random Variable, X

$$E[X] = \sum_{x=2}^{12} x\, p(x) = 7$$

Important properties of expectation

1. Linearity:

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

- Let X = 6-sided dice roll, $Y = 2X - 1$.
- $E[X] = 3.5$
- $E[Y] = 6$

2. Expectation of a sum = sum of expectation:

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

Sum of two dice rolls:

- Let X = roll of die 1, Y = roll of die 2
- $E[X + Y] = 3.5 + 3.5 = 7$

3. Law of the unconscious statistician (LOTUS):

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

These properties let you avoid defining difficult PMFs.

Being a statistician unconsciously

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

Let X be a discrete random variable.

- $P(X = x) = \frac{1}{3}$ for $x \in \{-1, 0, 1\}$

Let $Y = |X|$. What is $E[Y]$?

Today's plan

Variance

Bernoulli (Indicator) RVs

Binomial RVs

Average annual weather

Stanford, CA

$E[\text{high}] = 68°\text{F}$

$E[\text{low}] = 52°\text{F}$

Washington, DC

$E[\text{high}] = 67°\text{F}$

$E[\text{low}] = 51°\text{F}$

Is $E[X]$ enough?

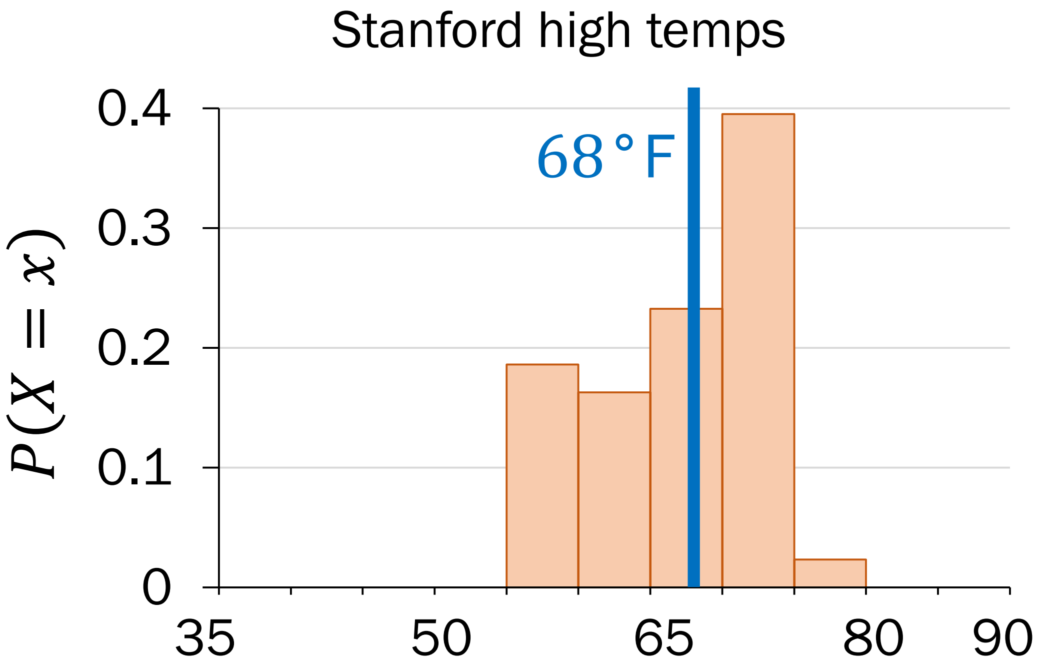
Average annual weather

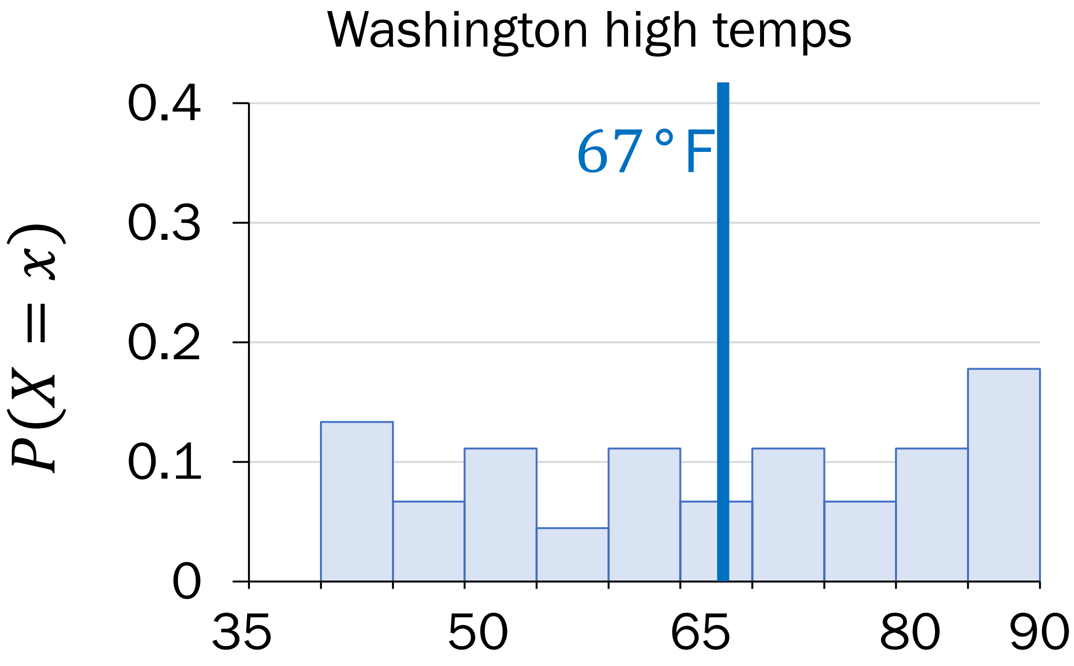
Stanford, CA

$E[\text{high}] = 68°\text{F}$

Washington, DC

$E[\text{high}] = 67°\text{F}$

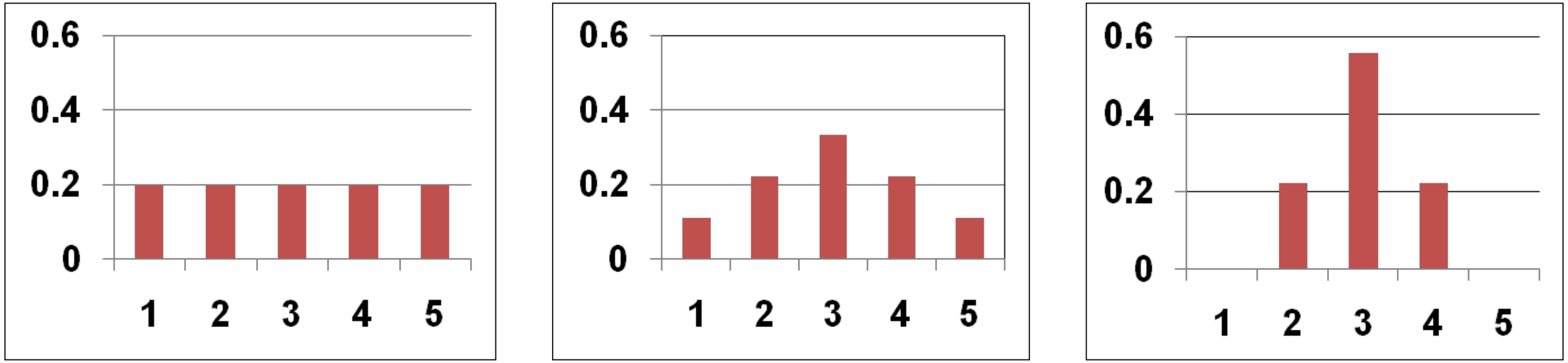




Normalized histograms are approximations of PMFs.

Variance = “spread”

Consider the following three distributions (PMFs):



- Expectation: $E[X] = 3$ for all distributions
- But the “spread” in the distributions is different!
- Variance, $\text{Var}(X)$: a formal quantification of “spread”

Variance

The variance of a random variable X with mean $E[X] = \mu$ is

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

- Also written as: $E[(X - E[X])^2]$
- Note: Var(X) ≥ 0
- Other names: **2nd central moment**, or square of the standard deviation

- An easier way to compute variance: $\text{Var}(X) = E[X^2] - (E[X])^2$

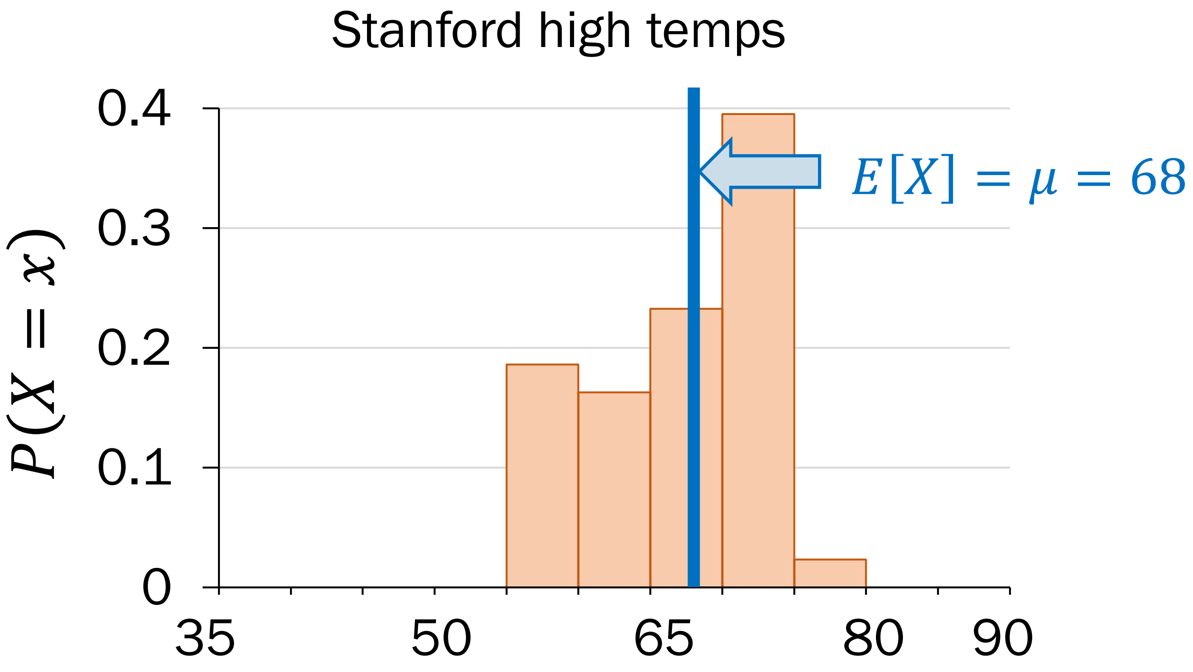
we'll come back to this

Variance of Stanford weather

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

Stanford, CA

$E[\text{high}] = 68°\text{F}$



X	$(X - \mu)^2$
57°F	124 (°F)^2
71°F	9 (°F)^2
75°F	49 (°F)^2
69°F	1 (°F)^2
…	…

Variance $E[(X - \mu)^2] = 39$ (°F)^2

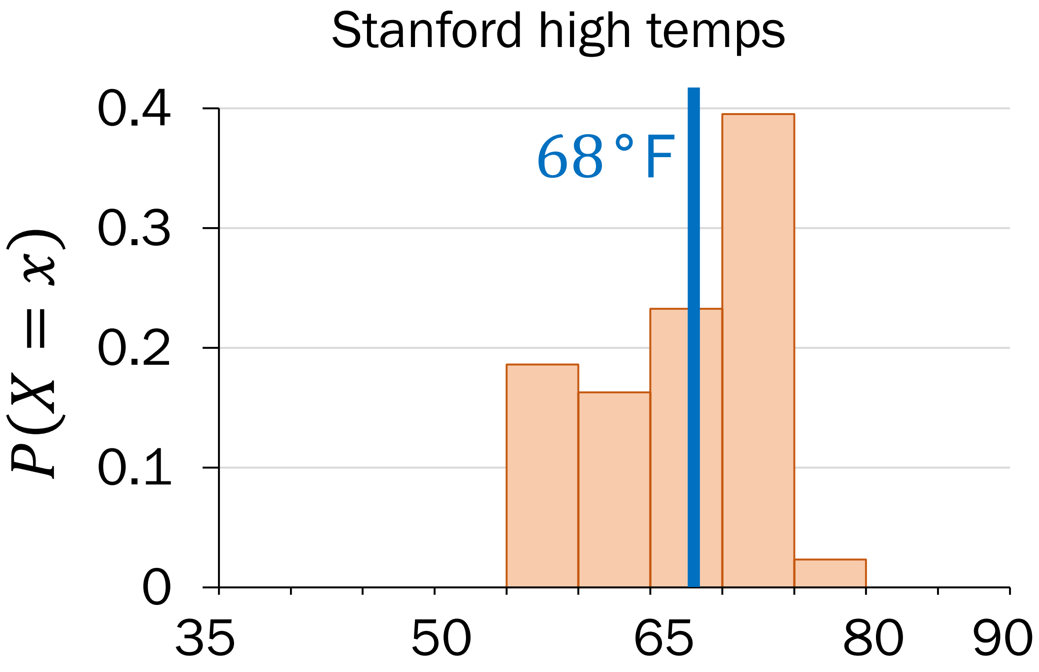
Standard deviation $= 6.2$°F

Comparing variance

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

Stanford, CA

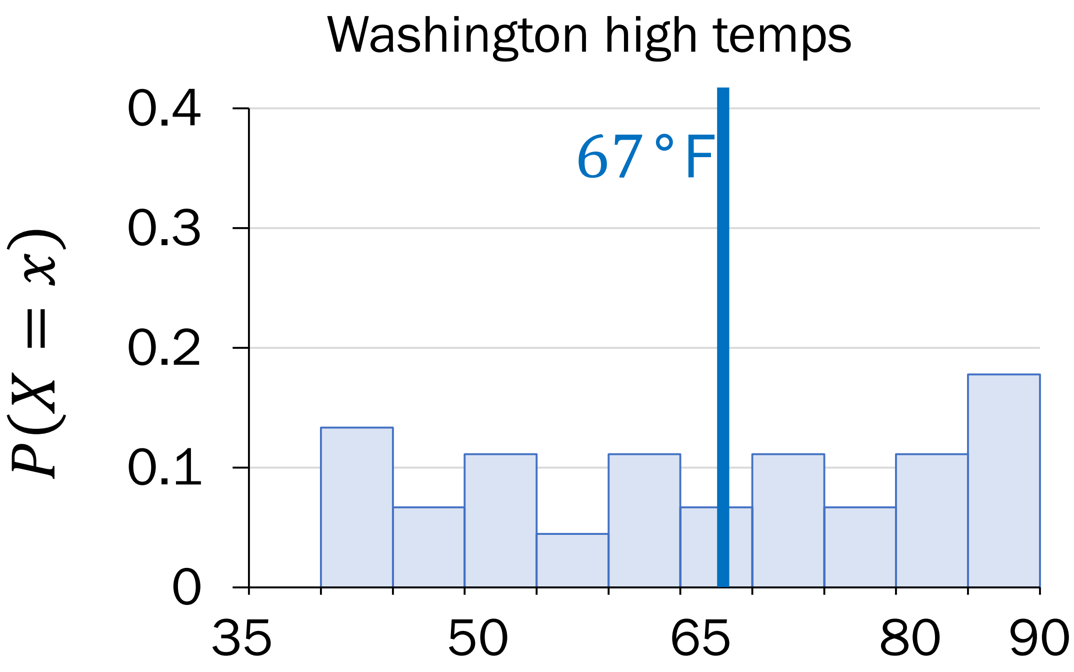
$E[\text{high}] = 68°\text{F}$



$\text{Var}(X) = 39\ (°\text{F})^2$

Washington, DC

$E[\text{high}] = 67°\text{F}$



$\text{Var}(X) = 248\ (°\text{F})^2$

Computing variance, a proof

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

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

Let $E[X] = \mu$

$$= \sum_x (x - \mu)^2 p(x)$$

$$= \sum_x (x^2 - 2\mu x + \mu^2) p(x)$$

$$= \sum_x x^2 p(x) - 2\mu \sum_x x p(x) + \mu^2 \sum_x p(x)$$

$$= E[X^2] - 2\mu E[X] + \mu^2 \cdot 1$$

Everyone, please welcome the second moment!

$$= E[X^2] - 2\mu^2 + \mu^2$$

$$= E[X^2] - \mu^2$$

$$= E[X^2] - (E[X])^2$$

Variance of a 6-sided die

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

Let Y = outcome of a single die roll. Recall $E[Y] = 7/2$.

Calculate the variance of Y.

1. Approach #1: Definition

$$\text{Var}(Y) = \frac{1}{6}\left(1 - \frac{7}{2}\right)^2 + \frac{1}{6}\left(2 - \frac{7}{2}\right)^2 + \frac{1}{6}\left(3 - \frac{7}{2}\right)^2 + \frac{1}{6}\left(4 - \frac{7}{2}\right)^2 + \frac{1}{6}\left(5 - \frac{7}{2}\right)^2 + \frac{1}{6}\left(6 - \frac{7}{2}\right)^2 = 35/12$$

2. Approach #2: A property

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

$$\text{Var}(Y) = 91/6 - (7/2)^2 = 35/12$$

Properties of variance

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

def standard deviation $\text{SD}(X) = \sqrt{\text{Var}(X)}$ — Units of X

Property 1 $\text{Var}(X) = E[X^2] - (E[X])^2$

Often easier to compute than definition.

Property 2 $\text{Var}(aX + b) = a^2\text{Var}(X)$

Unlike expectation, variance is NOT linear!!

Properties of variance

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

def standard deviation $\quad \text{SD}(X) = \sqrt{\text{Var}(X)}$ $\quad$ Units of X

Property 1 $\quad \text{Var}(X) = E[X^2] - (E[X])^2$

Property 2 $\quad \text{Var}(aX + b) = a^2\text{Var}(X)$

Unlike expectation, variance is NOT linear!!

Proof: $\text{Var}(aX + b)$

$$= E[(aX + b)^2] - (E[aX + b])^2 \quad \text{Property 1}$$

$$= E[a^2X^2 + 2abX + b^2] - (aE[X] + b)^2$$

$$= a^2E[X^2] + 2abE[X] + b^2 - (a^2(E[X])^2 + 2abE[X] + b^2)$$

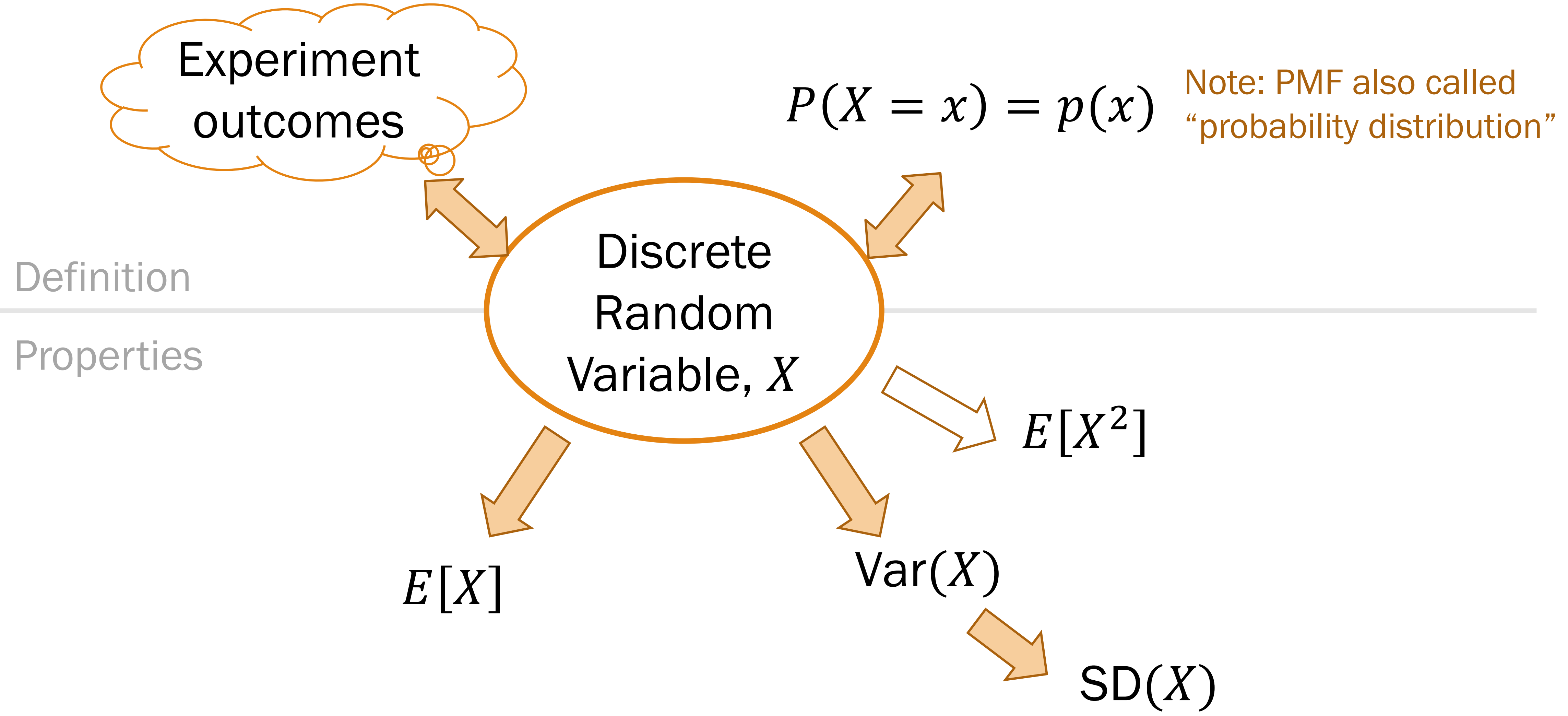
$$= a^2E[X^2] - a^2(E[X])^2$$

(Factoring/ Linearity of Expectation)

$$= a^2(E[X^2] - (E[X])^2)$$

$$= a^2\text{Var}(X) \quad \text{Property 1}$$

Discrete random variables



Today's plan

Variance

Bernoulli (Indicator) RVs

Binomial RVs

Bernoulli Random Variable

Consider an experiment with two outcomes: "success" and "failure."

def A **Bernoulli** random variable X maps "success" to 1 and "failure" to 0.
Other names: **indicator** random variable, boolean random variable

$X \sim \text{Ber}(p)$	PMF	$P(X=1) = p(1) = p$ $P(X=0) = p(0) = 1-p$
	Expectation	$E[X] = p$
Range: $\{0,1\}$	Variance	$\text{Var}(X) = p(1-p)$

Examples:

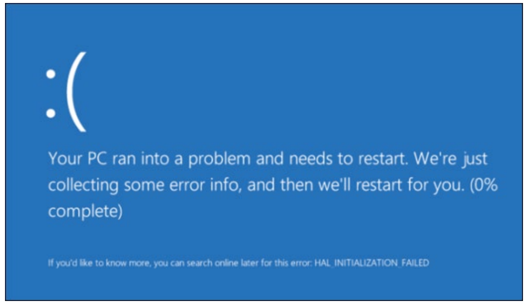
- Coin flip
- Random binary digit
- Whether a disk drive crashed

Bernoulli/indicator RVs are often used for this nice property of expectation.

Defining Bernoulli RVs

$X \sim \text{Ber}(p)$ $\quad p_X(1) = p$

$E[X] = p$ $\quad p_X(0) = 1 - p$



Run a program

- Crashes w.p. p
- Works w.p. $1 - p$

Let X: 1 if crash

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

$$P(X = 1) = p$$

$$P(X = 0) = 1 - p$$

Serve an ad.

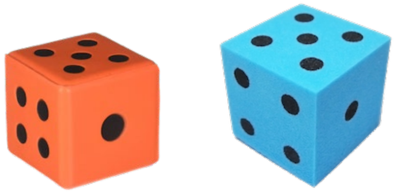
- Clicked w.p. p
- Ignored w.p. $1 - p$

Let X: 1 if clicked

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

$$P(X = 1) = p$$

$$P(X = 0) = 1 - p$$



Roll two dice.

- Success: roll two 6's
- Failure: anything else

Let X : 1 if success

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

$$P(X = 1) =$$

Announcements

Problem Set 2

Out: last Friday
Due: Monday 1/27
Covers: through last Friday

Section Resources

Handout (and python notebook):
posted Monday/Tuesday
Solutions:
posted Friday evening

Today's plan

Variance

Bernoulli (Indicator) RVs

Binomial RVs

Binomial Random Variable

Consider an experiment: n independent trials of Ber(p) random variables.

def A **Binomial** random variable X is the number of successes in n trials.

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

Range: $\{0, 1, \ldots, n\}$

PMF: $k = 0, 1, \ldots, n$:

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

Expectation: $E[X] = np$

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

Examples:

- # heads in n coin flips
- # of 1's in randomly generated length n bit string
- # of disk drives crashed in 1000 computer cluster (assuming disks crash independently)

By Binomial Theorem, we can prove $\sum_{k=0}^{n} P(X = k) = 1$

Reiterating notation

1. The random variable

2. is distributed as a

3. Binomial

4. with parameters

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

The parameters of a Binomial random variable:

- n: number of independent trials
- p: probability of success on each trial

Reiterating notation

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

If X is a binomial with parameters n and p, the PMF of X is

$$\underbrace{P(X = k)}_{\text{Probability that } X \text{ takes on the value } k} = \underbrace{\binom{n}{k} p^k (1-p)^{n-k}}_{\text{Probability Mass Function for a Binomial}}$$

Probability that X takes on the value k

Probability Mass Function for a Binomial

Three coin flips

$X \sim \text{Bin}(n, p) \quad p(k) = \binom{n}{k} p^k (1-p)^{n-k}$

Three fair ("heads" with $p = 0.5$) coins are flipped.

- X is number of heads
- $X \sim \text{Bin}(3, 0.5)$

Compute the following event probabilities:

$$P(X = 0)$$

$$P(X = 1)$$

$$P(X = 2)$$

$$P(X = 3)$$

$$P(X = 7)$$

P(event) PMF

Three coin flips

$X \sim \text{Bin}(n, p) \quad p(k) = \binom{n}{k} p^k (1-p)^{n-k}$

Three fair ("heads" with $p = 0.5$) coins are flipped.

- X is number of heads
- $X \sim \text{Bin}(3, 0.5)$

Compute the following event probabilities:

$$P(X = 0) \quad = p(0) \quad = \binom{3}{0} p^0 (1-p)^3 = \frac{1}{8}$$

$$P(X = 1) \quad = p(1) \quad = \binom{3}{1} p^1 (1-p)^2 = \frac{3}{8}$$

$$P(X = 2) \quad = p(2) \quad = \binom{3}{2} p^2 (1-p)^1 = \frac{3}{8}$$

$$P(X = 3) \quad = p(3) \quad = \binom{3}{3} p^3 (1-p)^0 = \frac{1}{8}$$

$$P(X = 7) \quad = p(7) \quad = 0$$

P(event) PMF

Binomial Random Variable

Consider an experiment: n independent trials of Ber(p) random variables.

def A **Binomial** random variable X is the number of successes in n trials.

$X \sim \text{Bin}(n, p)$	PMF	$k = 0, 1, \ldots, n$: $P(X = k) = p(k) = \binom{n}{k} p^k (1-p)^{n-k}$
	Expectation	$E[X] = np$
Range: $\{0,1,\ldots,n\}$	Variance	$\text{Var}(X) = np(1-p)$

Examples:

- # heads in n coin flips
- # of 1's in randomly generated length n bit string
- # of disk drives crashed in 1000 computer cluster (assuming disks crash independently)

By Binomial Theorem, we can prove $\sum_{k=0}^{n} P(X = k) = 1$

Binomial RV is sum of Bernoulli RVs

Bernoulli

- $X \sim \text{Ber}(p)$

Binomial

- $Y \sim \text{Bin}(n, p)$
- The sum of n independent Bernoulli RVs

$$Y = \sum_{i=1}^{n} X_i, \qquad X_i \sim \text{Ber}(p)$$

$\text{Ber}(p) = \text{Bin}(1, p)$

Binomial Random Variable

Consider an experiment: n independent trials of Ber(p) random variables.

def A **Binomial** random variable X is the number of successes in n trials.

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

Range: $\{0,1,\ldots,n\}$

PMF	$k = 0, 1, \ldots, n$: $P(X = k) = p(k) = \binom{n}{k} p^k (1-p)^{n-k}$
Expectation	$E[X] = np$
Variance	$\text{Var}(X) = np(1-p)$

Examples:

- # heads in n coin flips
- # of 1's in randomly generated length n bit string
- # of disk drives crashed in 1000 computer cluster (assuming disks crash independently)

Binomial Random Variable

Consider an experiment: n independent trials of Ber(p) random variables.

<u>def</u> A **Binomial** random variable X is the number of successes in n trials.

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

Range: $\{0,1,\ldots,n\}$

PMF $k = 0, 1, \ldots, n$:

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

Expectation $E[X] = np$

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

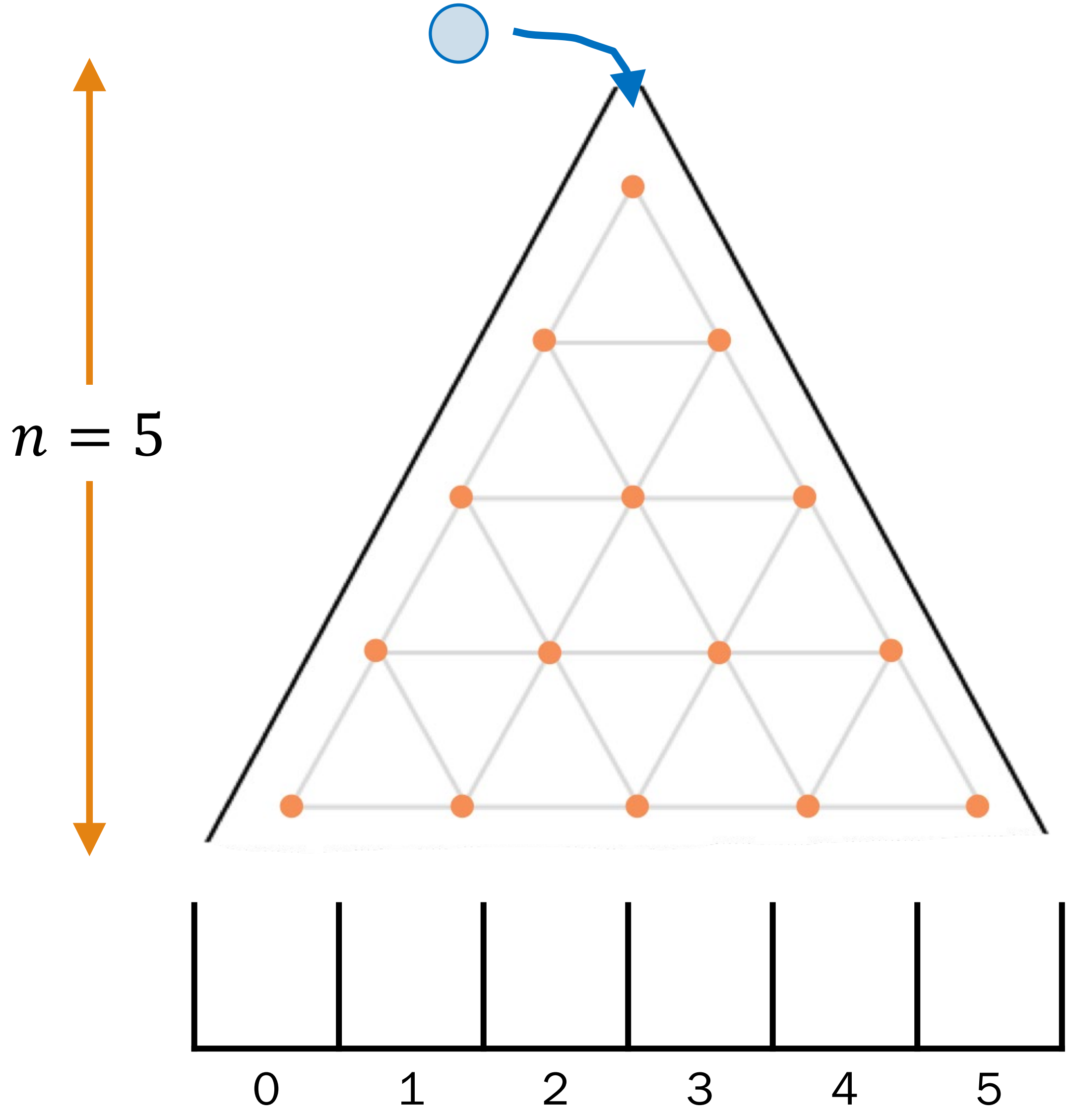
We'll prove this later in the course

Examples:

- # heads in n coin flips
- # of 1's in randomly generated length n bit string
- # of disk drives crashed in 1000 computer cluster (assuming disks crash independently)

Galton Board

$X \sim \text{Bin}(n, p) \quad p(k) = \binom{n}{k} p^k (1-p)^{n-k}$



Let B = the bucket index a ball drops into.
B is distributed as a Binomial RV,

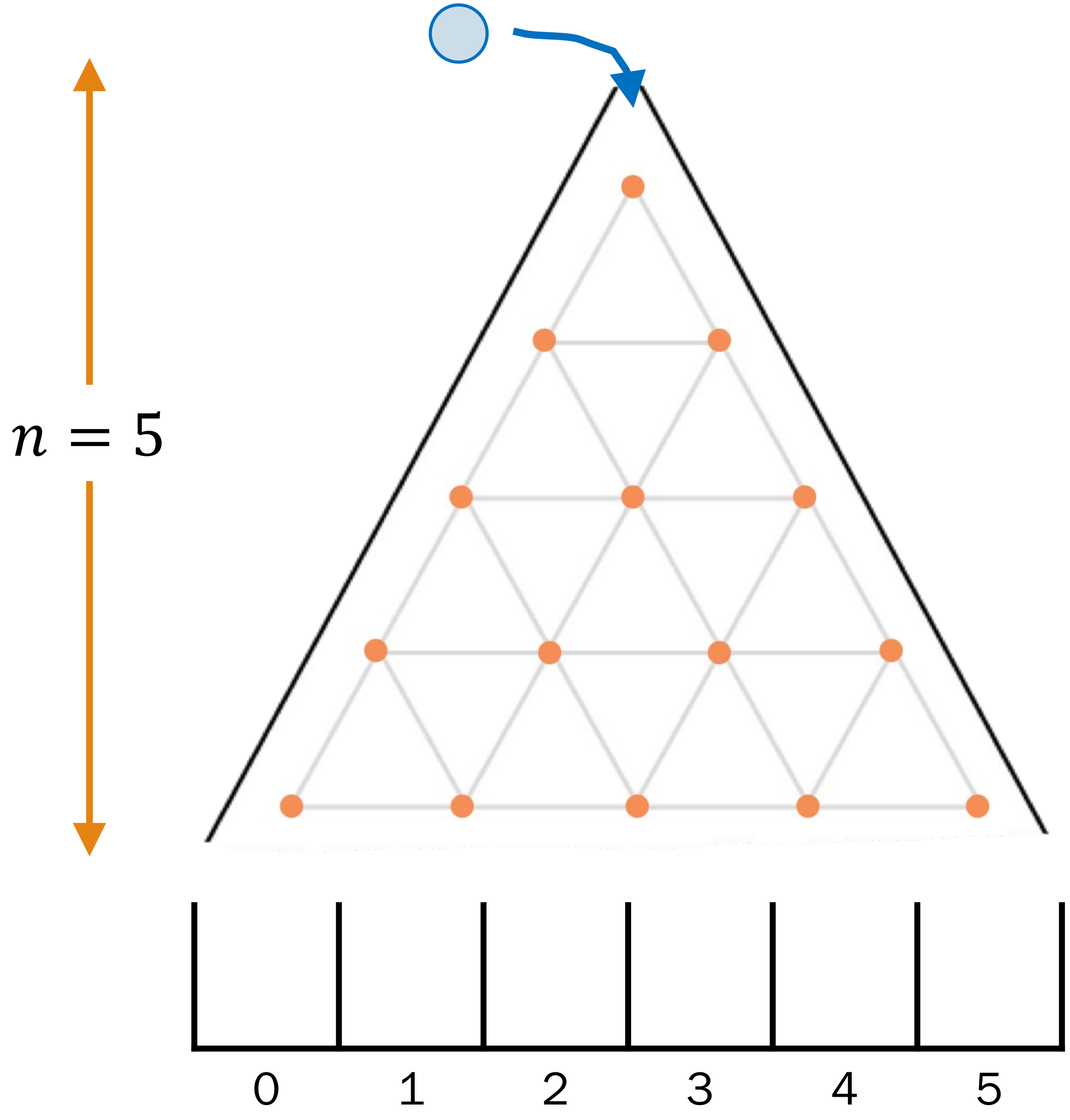
$$B \sim \text{Bin}(n = 5, p = 0.5)$$

If B is a sum of Bernoulli RVs,
what defines the ith trial, R_i?

http://web.stanford.edu/class/cs109/demos/galton.html

Galton Board

$X \sim \text{Bin}(n, p) \quad p(k) = \binom{n}{k} p^k (1-p)^{n-k}$



Let B = the bucket index a ball drops into.
B is distributed as a Binomial RV,

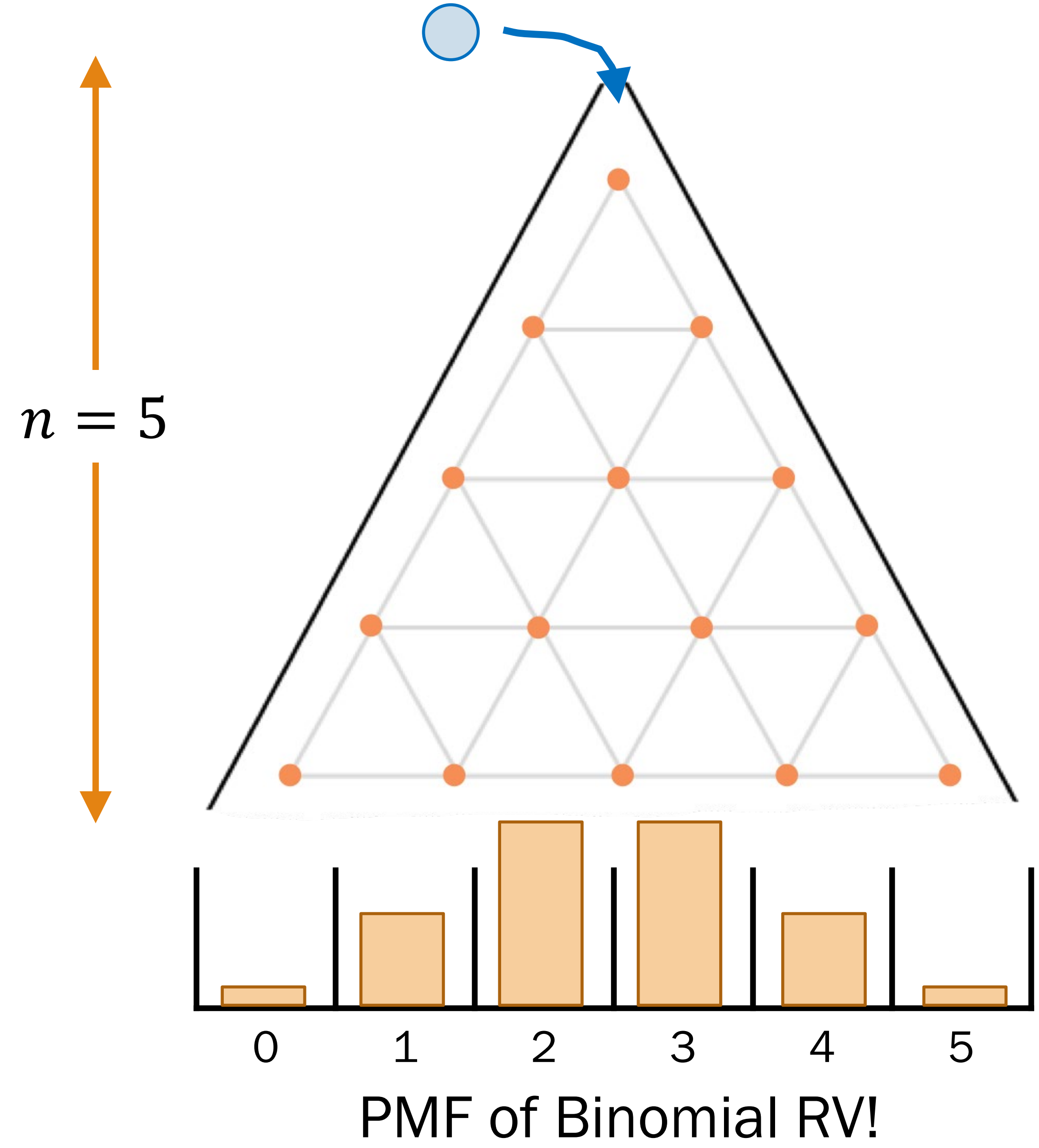
$$B \sim \text{Bin}(n = 5, p = 0.5)$$

If B is a sum of Bernoulli RVs,
what defines the ith trial, R_i?

- When a marble hits a pin, it has an equal chance of going left or right
- Each pin is an independent trial
- One decision made for level $i = 1, 2, .., 5$
- $R_i = 1$ if ball went right on level i
- Bucket index $B =$ # times ball went right

Galton Board

$X{\sim}\text{Bin}(n,p) \quad p(k) = \binom{n}{k} p^k (1-p)^{n-k}$



Let B = the bucket index a ball drops into.
B is distributed as a Binomial RV,

$$B{\sim}\text{Bin}(n = 5, p = 0.5)$$

Calculate the probability of a ball landing in bucket k.

$$P(B = 0) = \binom{5}{0} 0.5^5 \approx 0.03$$

$$P(B = 1) = \binom{5}{1} 0.5^5 \approx 0.16$$

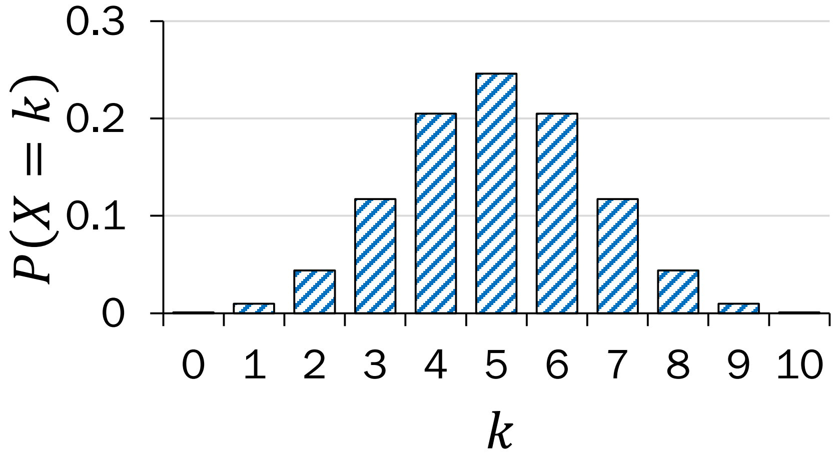
$$P(B = 2) = \binom{5}{2} 0.5^5 \approx 0.31$$

Visualizing Binomial PMFs

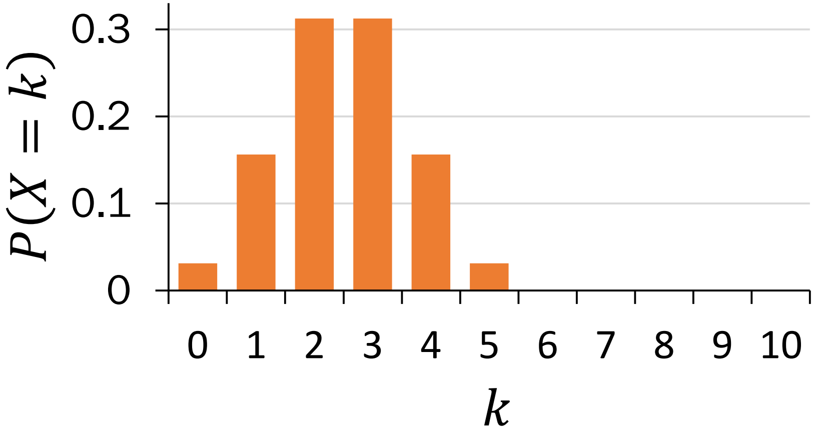
$E[X] = np$

$X \sim \text{Bin}(n, p) \quad p(i) = \binom{n}{k} p^k (1-p)^{n-k}$

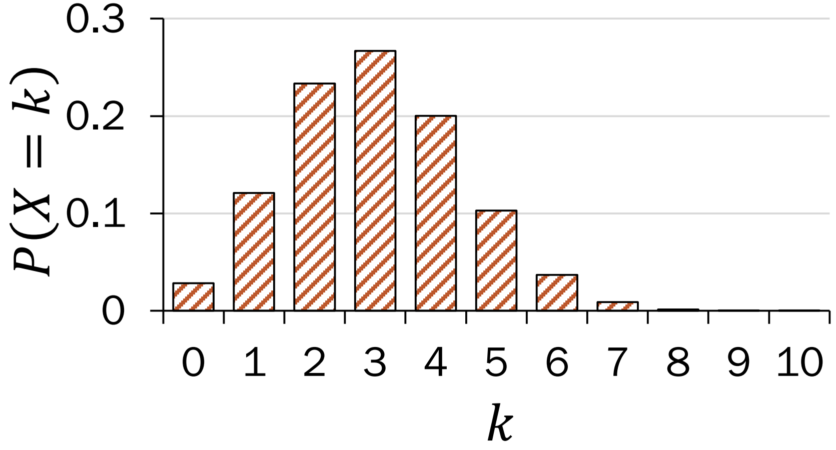
A.



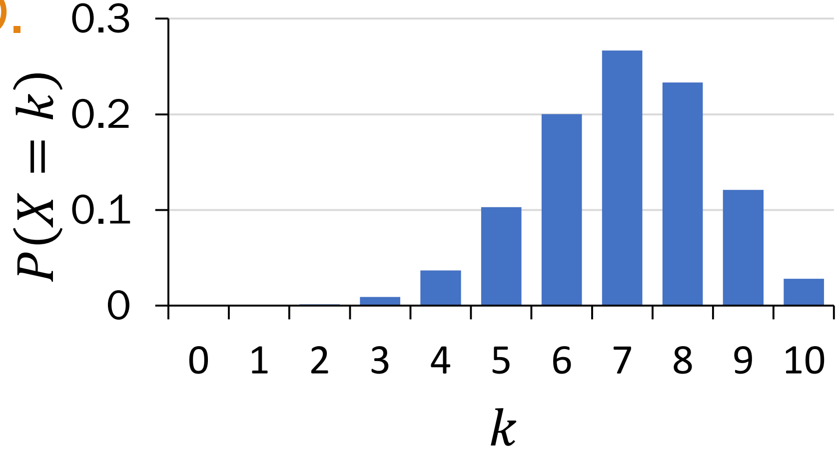
B.



C.



D.



Match the distribution to the graph:

1. Bin(10,0.5)
2. Bin(10,0.3)
3. Bin(10,0.7)
4. Bin(5,0.5)

NBA Finals

$X \sim \text{Bin}(n, p) \quad p(k) = \binom{n}{k} p^k (1-p)^{n-k}$

Let's say the Golden State Warriors are going to play the Toronto Raptors in a 7-game series during the 2020 NBA finals.

- The Warriors have a probability of 58% of winning each game, independently.
- A team wins the series if they win at least 4 games (we play all 7 games).

What is P(Warriors winning)?

1. Define events/ RVs & state goal

X: # games Warriors win

$X \sim \text{Bin}(7, 0.58)$

Want: → Desired probability? (select all that apply)

A. $P(X > 4)$
B. $P(X \geq 4)$
C. $P(X > 3)$
D. $1 - P(X \leq 3)$
E. $1 - P(X < 3)$

NBA Finals

$X \sim \text{Bin}(n, p) \quad p(k) = \binom{n}{k} p^k (1-p)^{n-k}$

Let's say the Golden State Warriors are going to play the Toronto Raptors in a 7-game series during the 2020 NBA finals.

- The Warriors have a probability of 58% of winning each game, independently.
- A team wins the series if they win at least 4 games (we play all 7 games).

What is P(Warriors winning)?

1. Define events/ RVs & state goal

X: # games Warriors win

$X \sim \text{Bin}(7, 0.58)$

Want: $P(X \geq 4)$

2. Solve

$$P(X \geq 4) = \sum_{k=4}^{7} P(X = k) = \sum_{k=4}^{7} \binom{7}{k} 0.58^k (0.42)^{7-k}$$

Cool Algebra/Probability Fact: this is identical to the probability of winning if we define winning = first to win 4 games