Table of Contents

07: Variance, Bernoulli, Binomial

Jerry Cain

January 24, 2024

Lecture Discussion on Ed

Variance

Average temperatures

Stanford, CA

$E[\text{high}] = 68°\text{F}$

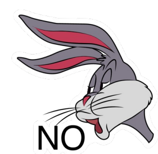
$E[\text{low}] = 52°\text{F}$

Washington, DC

$E[\text{high}] = 67°\text{F}$

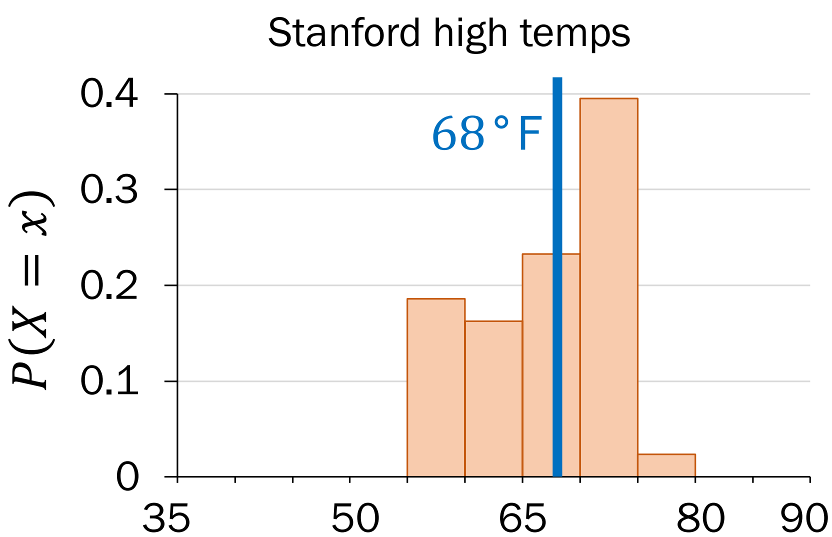
$E[\text{low}] = 51°\text{F}$

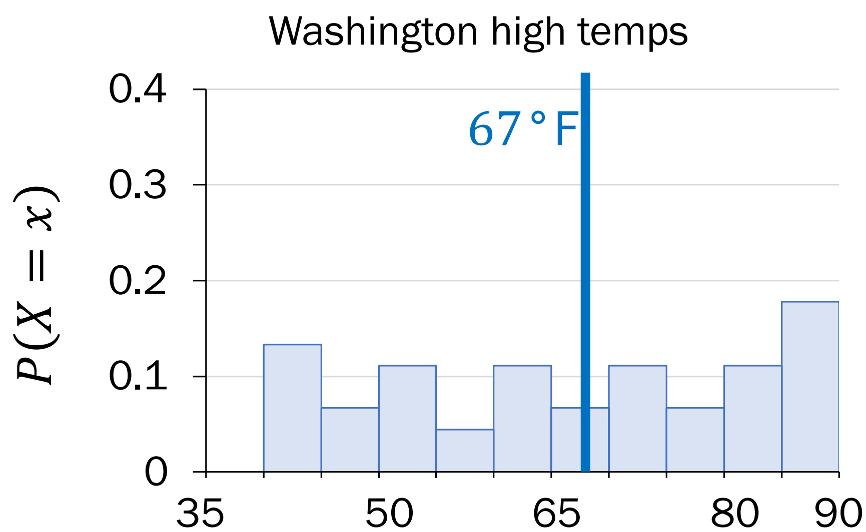
Is $E[X]$ enough? Does is capture everything?



Average temperatures

Stanford, CA	Washington, DC
$E[\text{high}] = 68°\text{F}$	$E[\text{high}] = 67°\text{F}$
$E[\text{low}] = 52°\text{F}$	$E[\text{low}] = 51°\text{F}$

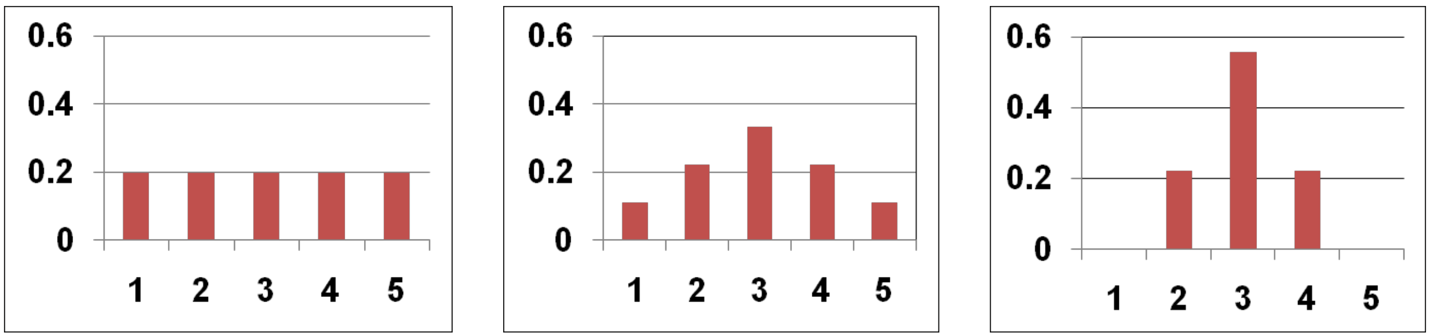




Normalized histograms are approximations of probability mass functions, i.e., PMFs.

Variance = measure of "spread"

Consider the following three distributions (PMFs):



- Expectation: $E[X] = 3$ for all distributions
- But the shape and spread across distributions are very 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

	$\text{Var}(X)$	Units of X^2
def standard deviation	$\text{SD}(X) = \sqrt{\text{Var}(X)}$	Units of X

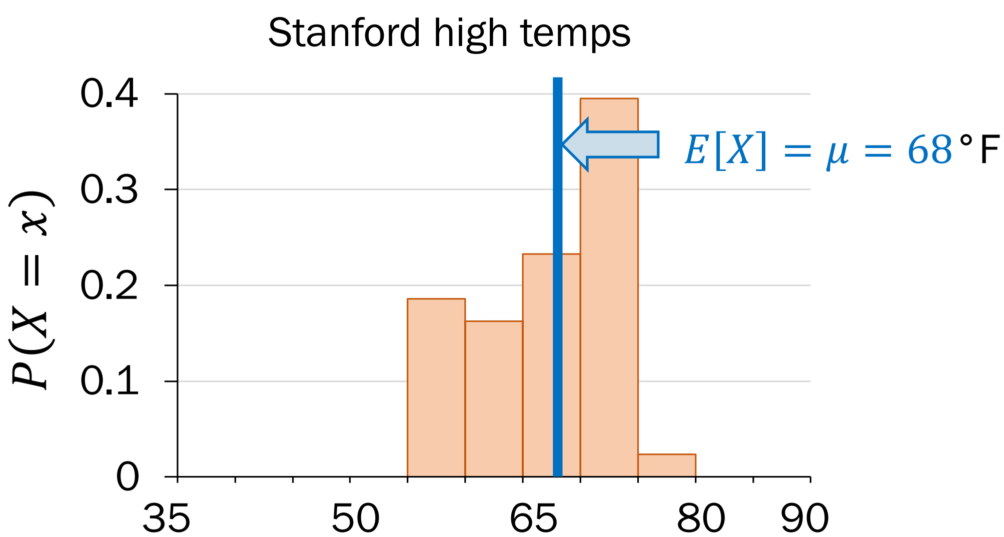
Variance of Stanford weather

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

Stanford, CA

$E[\text{high}] = 68°\text{F}$

$E[\text{low}] = 52°\text{F}$



X	$(X - \mu)^2$
57°F	121 (°F)^2
71°F	9 (°F)^2
75°F	49 (°F)^2
69°F	1 (°F)^2
…	…

Variance $E[(X - \mu)^2] = 39\ (°\text{F})^2$

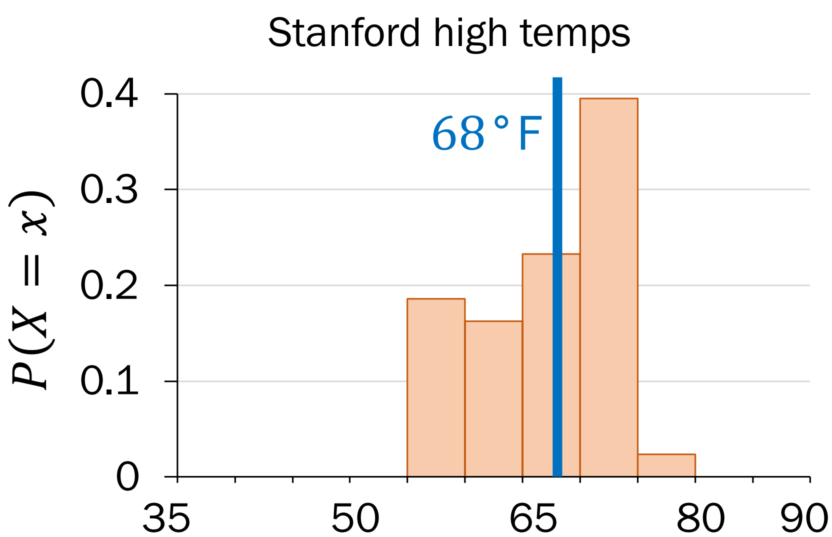
Standard deviation $= 6.2°\text{F}$

Comparing variance

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

Stanford, CA

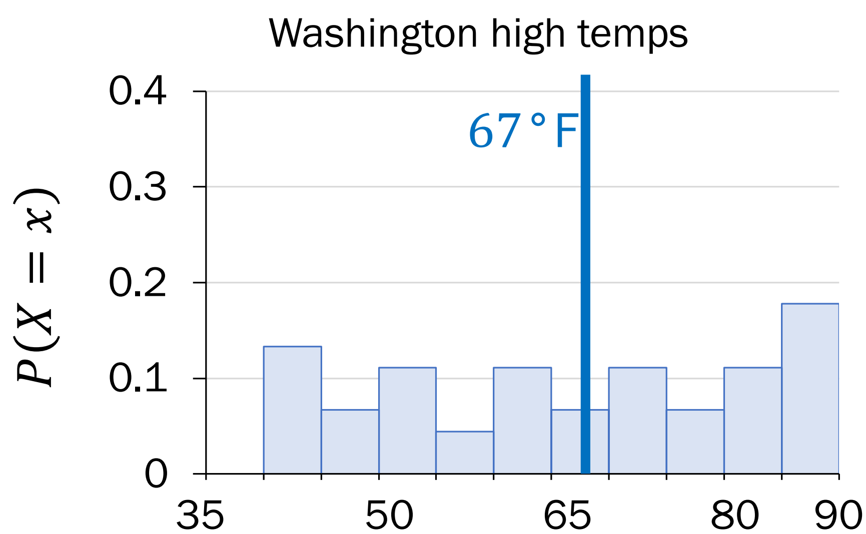
$E[\text{high}] = 68°\text{F}$



$\text{Var}(X) = 39\ (°\text{F})^2$

Washington, DC

$E[\text{high}] = 67°\text{F}$



$\text{Var}(X) = 248\ (°\text{F})^2$

Properties of Variance

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

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

- Property 1 is often easier to manipulate than the original definition
- Unlike expectation, variance is not linear

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

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

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 $E[X] = \mu$

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

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

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

Everyone, please welcome the second moment!

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

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

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

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

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

2nd moment

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

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

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

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

Property 2: A proof

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

<u>Proof</u>: $\text{Var}(aX+b)$

$$= E[(aX+b)^2] - (E[aX+b])^2$$ Property 1

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

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

Factoring/ Linearity of Expectation

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

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

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

Bernoulli RV

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

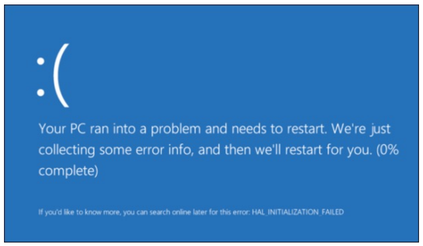
$X \sim \text{Ber}(p)$	PMF	$P(X=1) = p(1) = p$ $P(X=0) = p(0) = 1-p$
	Expectation	$E[X] = p$
Support: $\{0,1\}$	Variance	$\text{Var}(X) = p(1-p)$

Examples:

- Coin flip
- Random binary digit
- Whether Doris barks

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

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

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

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

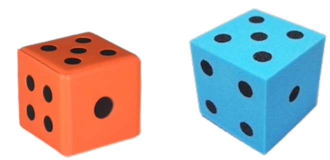
Serve an ad.
- User clicks w.p. 0.2
- Ignores otherwise

Let X: 1 if clicked

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

$$P(X = 1) = \_\_$$

$$P(X = 0) = \_\_$$



Roll two dice.
- Success: roll a 10
- Failure: anything else

Let X : 1 if success

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

$$E[X] = \_$$

Binomial RV

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.

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

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

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)

Reiterating notation

1. The random variable

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

2. is distributed as a

3. Binomial

4. with parameters

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

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

Three coin flips

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

Three fair (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)

Three coin flips

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

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

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

Compute the following event probabilities:

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

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

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

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

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

P(event)

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

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

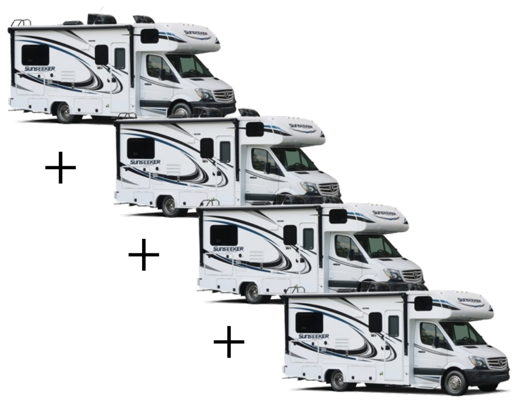
Examples:

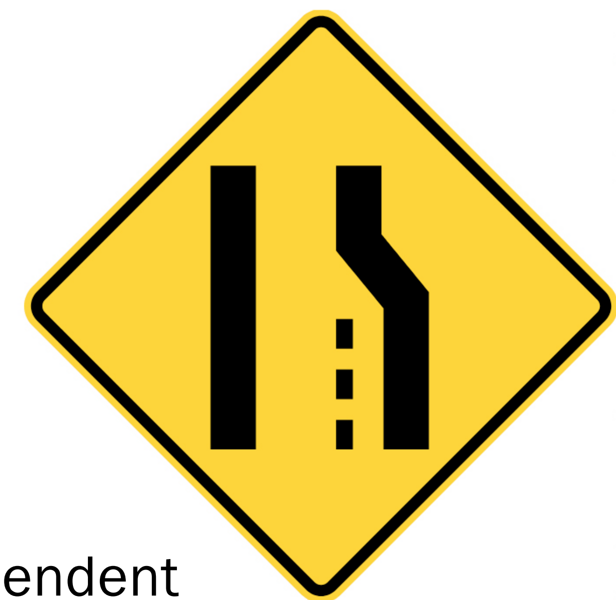
- # heads in n coin flips
- # of 1's in randomly generated length n bit string
- # of disk drives crashed in 1000 computer cluster (assuming disks crash independently)

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

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

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

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.

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

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

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)

Exercises

Statistics: Expectation and variance

1. a. Let X = the outcome of a fair 24-sided die roll. What is $E[X]$?
   b. Let Y = the sum of seven rolls of a fair 24-sided die. What is $E[Y]$?

2. Let Z = # of ***tails*** on 10 flips of a biased coin, with p = 0.71. What is $E[Z]$?

3. Compare the variances of $B_0 \sim \text{Ber}(0.0)$, $B_1 \sim \text{Ber}(0.1)$, $B_2 \sim \text{Ber}(0.5)$, and $B_3 \sim \text{Ber}(0.9)$.

Statistics: Expectation and variance

If you can identify common RVs, just look up statistics instead of rederiving from scratch.

1. a. Let X = the outcome of a fair 24-sided die roll. What is $E[X]$?
   b. Let Y = the sum of seven rolls of a fair 24-sided die. What is $E[Y]$?

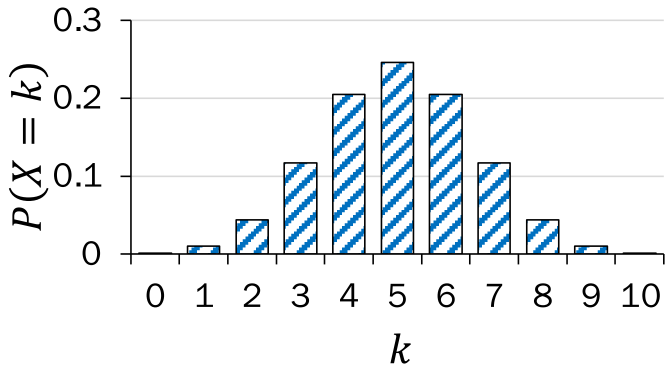
2. Let Z = # of ***tails*** on 10 flips of a biased coin, with p = 0.71. What is $E[Z]$?

3. Compare the variances of $B_0 \sim \text{Ber}(0.0)$, $B_1 \sim \text{Ber}(0.1)$, $B_2 \sim \text{Ber}(0.5)$, and $B_3 \sim \text{Ber}(0.9)$.

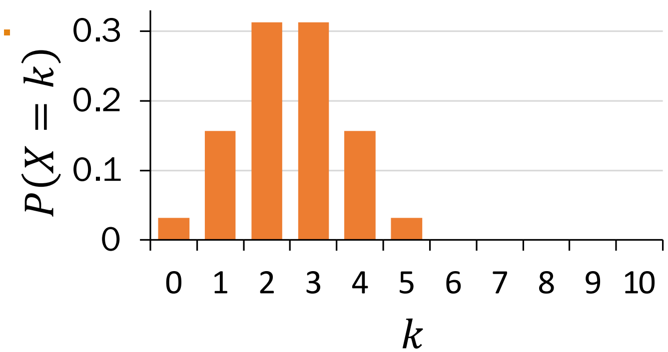
Visualizing Binomial PMFs

$$X \sim \text{Bin}(n, p) \qquad E[X] = np \qquad p(i) = \binom{n}{k} p^k (1-p)^{n-k}$$

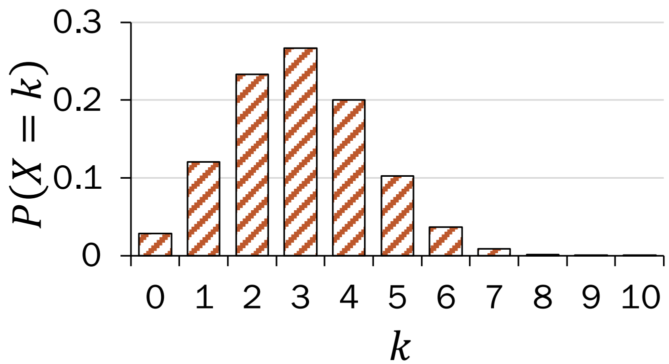
A.



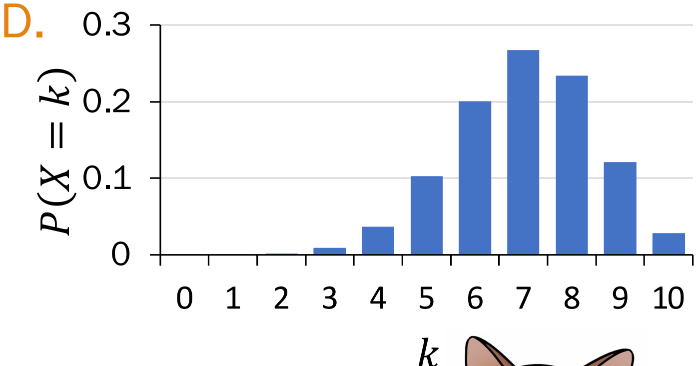
B.



C.



D.



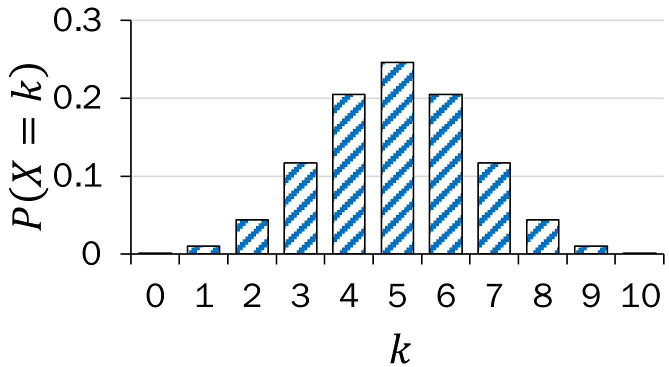
Match the distribution of X to the graph:

1. $\text{Bin}(10, 0.5)$
2. $\text{Bin}(10, 0.3)$
3. $\text{Bin}(10, 0.7)$
4. $\text{Bin}(5, 0.5)$

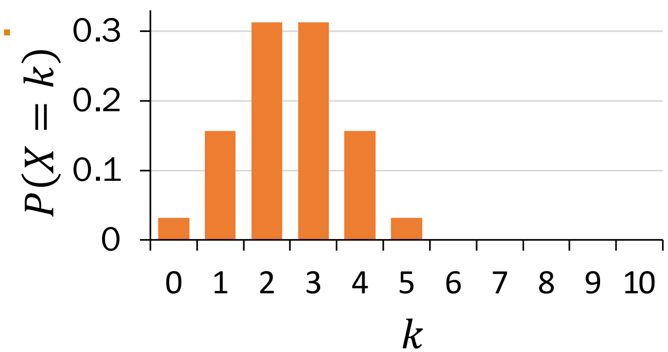
Visualizing Binomial PMFs

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

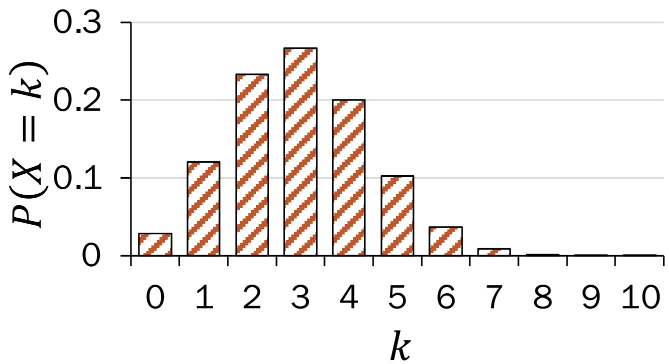
A.



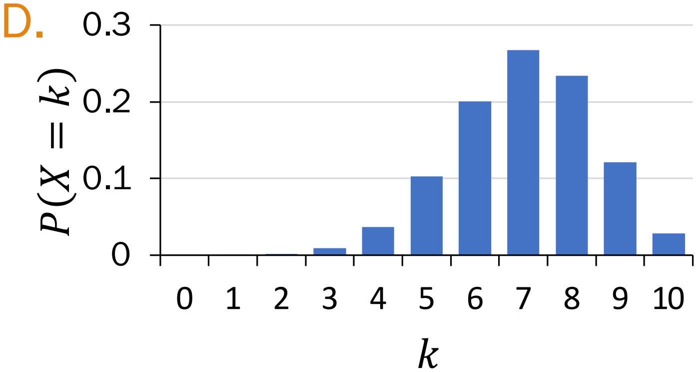
B.



C.



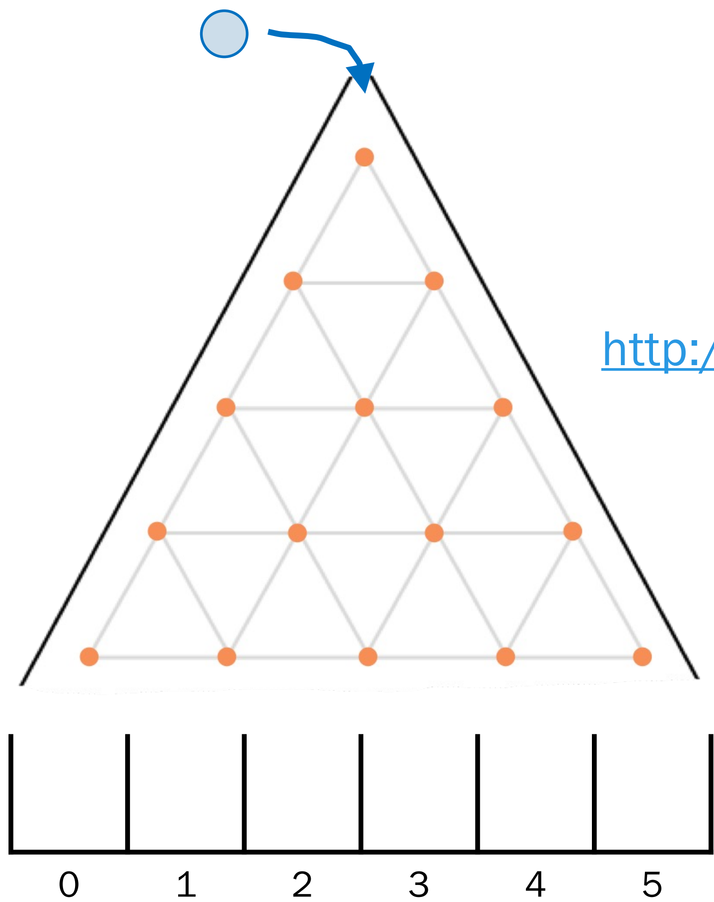
D.



Match the distribution of X to the graph:

1. $\text{Bin}(10, 0.5)$
2. $\text{Bin}(10, 0.3)$
3. $\text{Bin}(10, 0.7)$
4. $\text{Bin}(5, 0.5)$

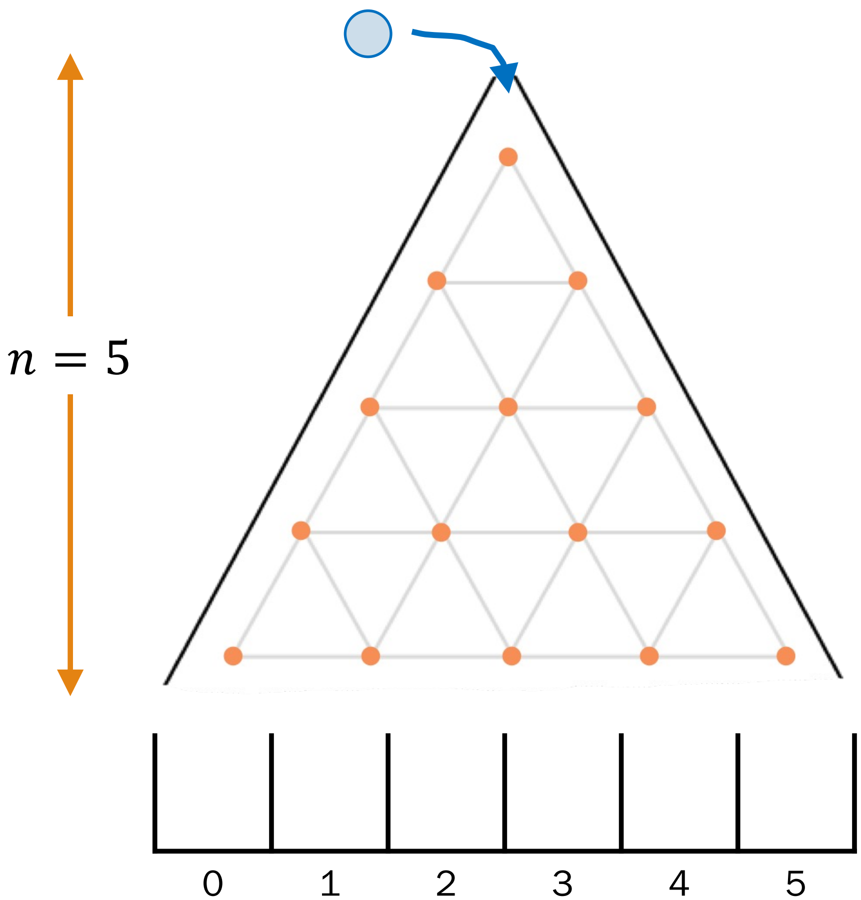
Galton Board



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

Galton Board

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



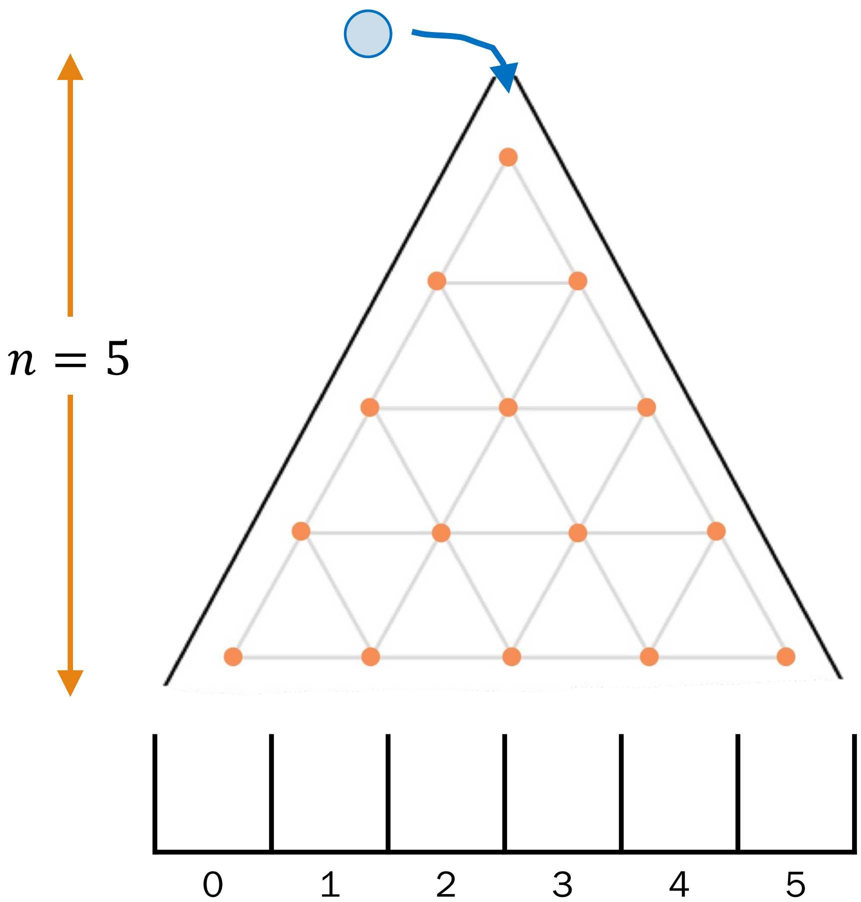
When a marble hits a pin, it has an equal chance of going left or right.

Let B = the bucket index a ball drops into. What is the distribution of B?

(Interpret: If B is a common random variable, report it, otherwise report PMF)

Galton Board

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



When a marble hits a pin, it has an equal chance of going left or right.

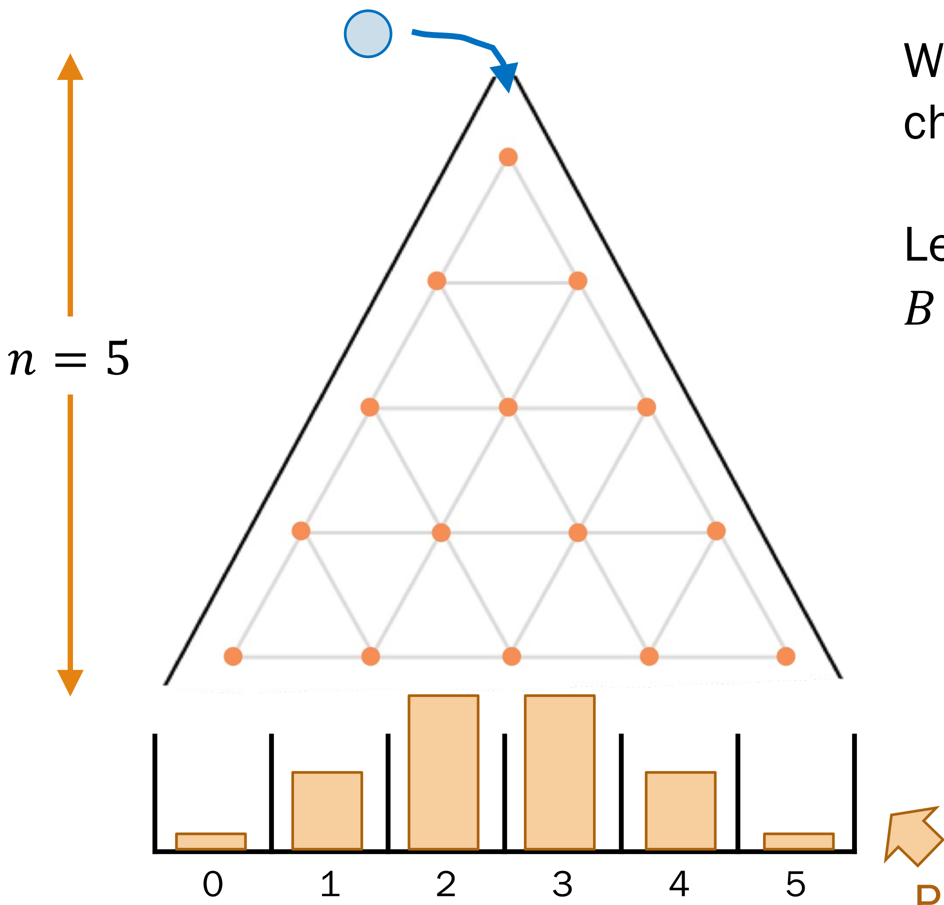
Let B = the bucket index a ball drops into.
What is the distribution of B?

- Each pin is an independent trial
- One decision made for level $i = 1, 2, .., 5$
- Consider a Bernoulli RV with success R_i if ball went right on level i
- Bucket index $B =$ # times ball went right

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

Galton Board

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



When a marble hits a pin, it has an equal chance of going left or right.

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

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

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

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

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

PMF of Binomial RV!

Genetics and NBA Finals

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

1. Each parent has 2 genes per trait (e.g., eye color).

- Child inherits 1 gene from each parent with equal likelihood.
- Brown eyes are "dominant", blue eyes are "recessive":
  - Child has brown eyes if either or both genes for brown eyes are inherited.
  - Child has blue eyes otherwise (i.e., child inherits two genes for blue eyes)
- Assume parents each have 1 gene for blue eyes and 1 gene for brown eyes.

Two parents have 4 children. What is P(exactly 3 children have brown eyes)?

2. Let's speculate that the Boston Celtics will play the Milwaukee Bucks in a 7-game series during the 2024 NBA finals.

- The Celtics 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(Celtics winning)?

Genetic inheritance

1. Each parent has 2 genes per trait (e.g., eye color).

- Child inherits 1 gene from each parent with equal likelihood.
- Brown eyes are "dominant", blue eyes are "recessive":
  - Child has brown eyes if either or both genes for brown eyes are inherited.
  - Child has blue eyes otherwise (i.e., child inherits two genes for blue eyes)
- Assume parents each have 1 gene for blue eyes and 1 gene for brown eyes.

Two parents have 4 children. What is P(exactly 3 children have brown eyes)?

Big Q: Fixed parameter or random variable?

Parameters	What is common among all outcomes of our experiment?
Random variable	What differentiates our event from the rest of the sample space?

Genetic inheritance

1. Each parent has 2 genes per trait (e.g., eye color).

- Child inherits 1 gene from each parent with equal likelihood.
- Brown eyes are "dominant", blue eyes are "recessive":
  - Child has brown eyes if either or both genes for brown eyes are inherited.
  - Child has blue eyes otherwise (i.e., child inherits two genes for blue eyes)
- Assume parents each have 1 gene for blue eyes and 1 gene for brown eyes.

Two parents have 4 children. What is P(exactly 3 children have brown eyes)?

1. Define events/ RVs & state goal
2. Identify known probabilities
3. Solve

X: # brown-eyed children,
$X \sim \text{Bin}(4, p)$
p: $P(\text{brown}-\text{eyed child})$

Want: $P(X = 3)$

NBA Finals

Let's speculate: the Boston Celtics will play the Milwaukee Bucks in a 7-game series during the 2024 NBA finals.

- The Celtics 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(Celtics winning)?

1. Define events/ RVs & state goal

X: # games Celtics win

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

Want:

Big Q: Fixed parameter or random variable?

Parameters	# of total games prob Celtics winning a game
Random variable	# of games Boston Celtics win
Event based on RV	

NBA Finals

Let's speculate: the Boston Celtics will play the Milwaukee Bucks in a 7-game series during the 2024 NBA finals.

- The Celtics 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(Celtics winning)?

1. Define events/ RVs & state goal

X: # games Celtics 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