07: Variance, Bernoulli, Binomial

Jerry Cain

April 11, 2022

Table of Contents

Variance

Average temperatures

Stanford, CA

$E[\text{high}] = 68°\text{F}$

$E[\text{low}] = 52°\text{F}$

Washington, DC

$E[\text{high}] = 67°\text{F}$

$E[\text{low}] = 51°\text{F}$

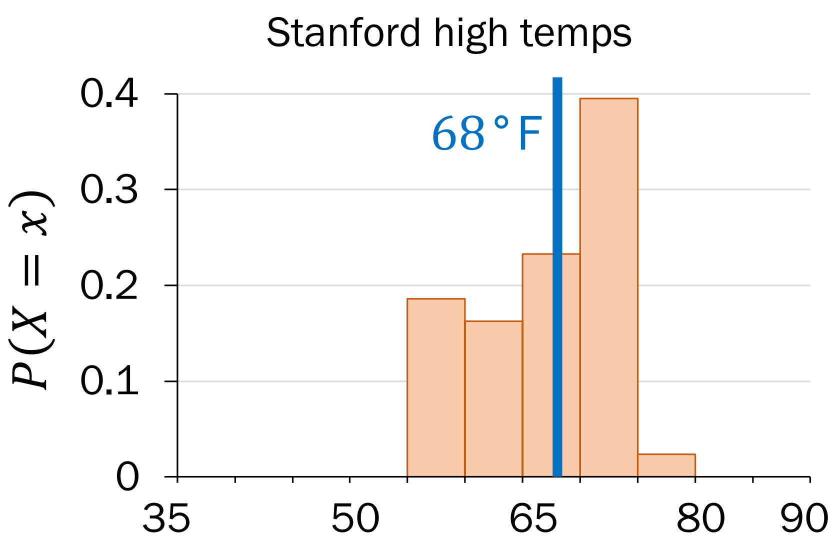
Is $E[X]$ enough?

Average temperatures

Stanford, CA

$E[\text{high}] = 68°\text{F}$

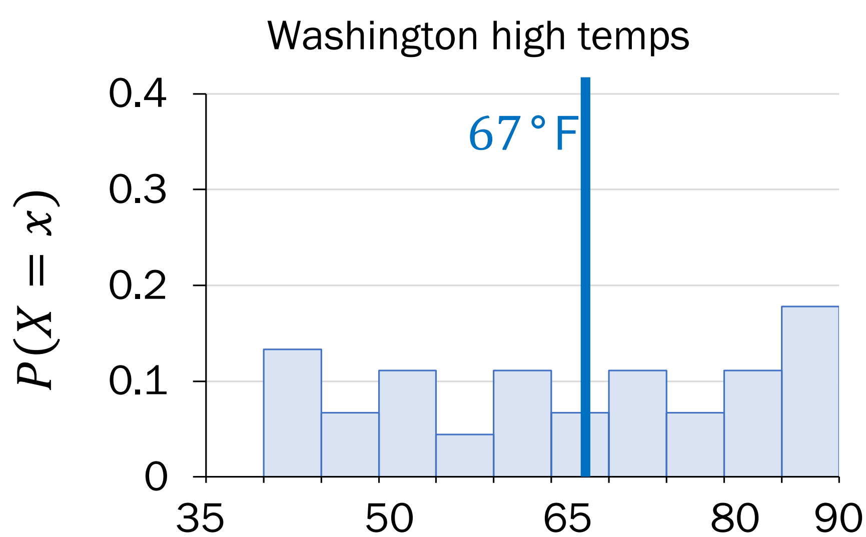
$E[\text{low}] = 52°\text{F}$



Washington, DC

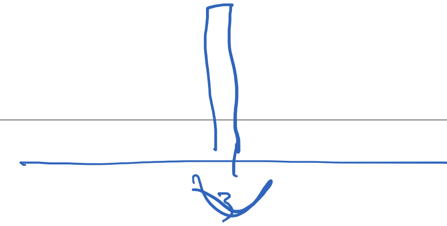
$E[\text{high}] = 67°\text{F}$

$E[\text{low}] = 51°\text{F}$

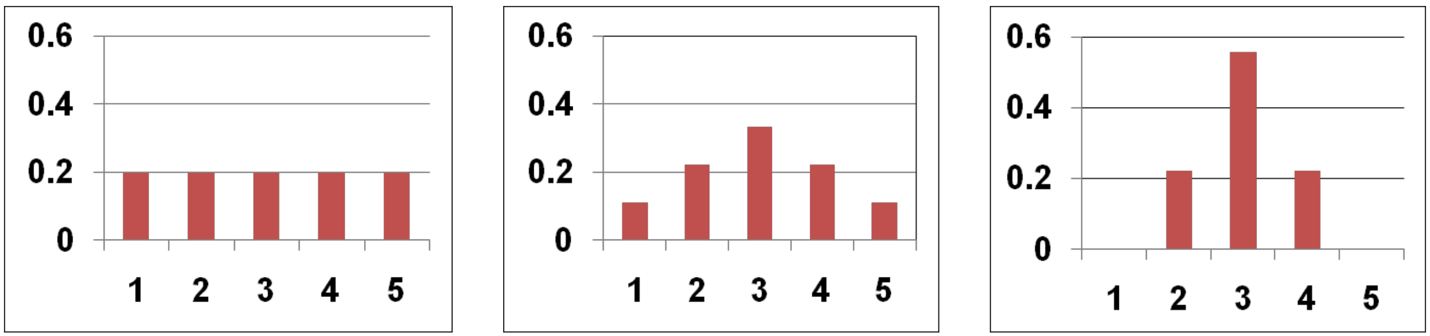


Normalized histograms are approximations of PMFs.

Variance = "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

-97 103

Variance

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

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

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

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

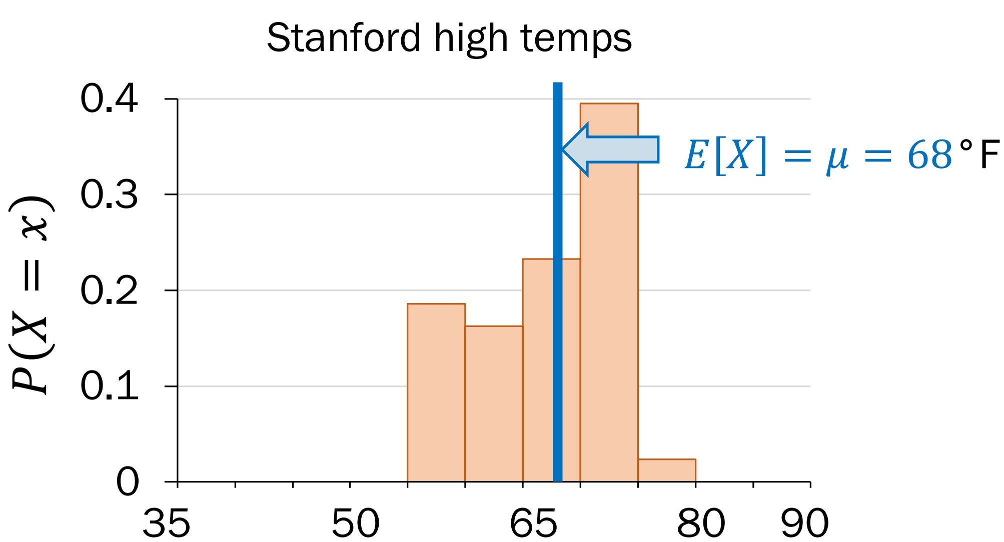
Variance of Stanford weather

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

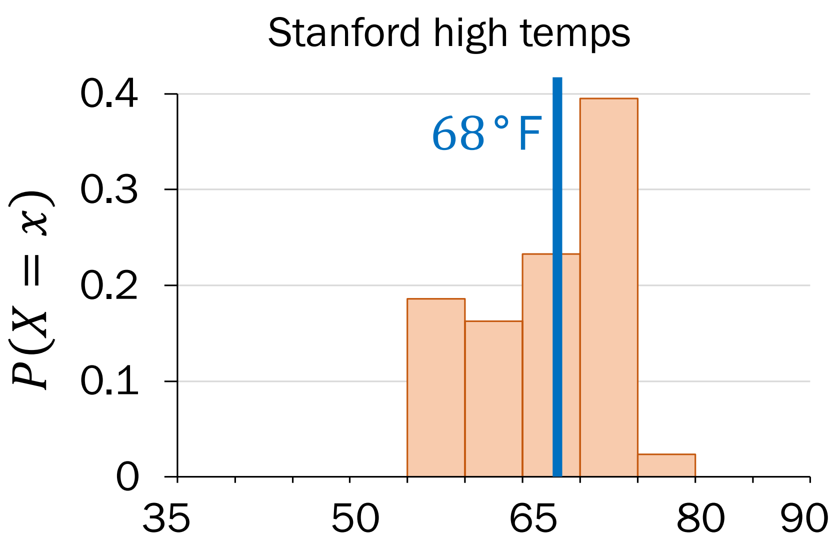
$\text{Var}(X) = E[(X - E[X])^2]$ Variance of X

Stanford, CA

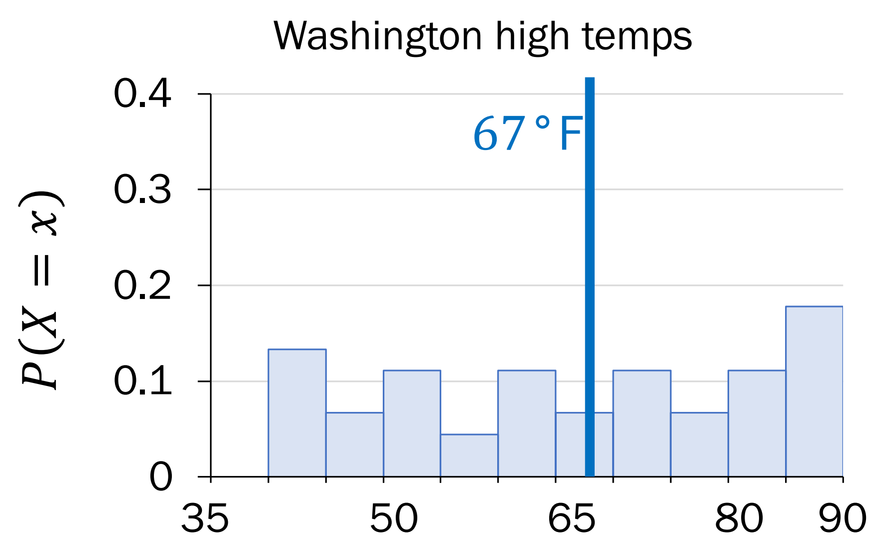
$E[\text{high}] = 68°\text{F}$

Washington, DC

$E[\text{high}] = 67°\text{F}$



$\text{Var}(X) = 39\ (°\text{F})^2$



$\text{Var}(X) = 248\ (°\text{F})^2$

Properties of Variance

Properties of variance

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

<u>def</u> 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 $\text{Var}(X) = E[X^2] - (E[X])^2$

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

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 $E[X] = \mu$

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

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

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

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

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

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

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

Everyone, please welcome the second moment!

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

<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

$1 \cdot p + 0 \cdot (1-p)$

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

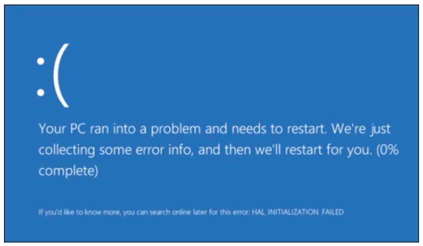
$X \sim \text{Ber}(p)$	PMF	$P(X=1) = p(1) = p$ $P(X=0) = p(0) = 1-p$
	Expectation	$E[X] = p$
Support: $\{0,1\}$	Variance	$\text{Var}(X) = p(1-p)$

Examples:
- Coin flip
- Random binary digit
- Whether Doris barks

Remember this nice property of expectation.

Defining Bernoulli RVs

$X \sim \text{Ber}(p)$	$p_X(1) = p$
$E[X] = p$	$p_X(0) = 1 - p$



Run a program
- Crashes w.p. p
- Works w.p. $1 - p$

Let X: 1 if crash

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

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

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

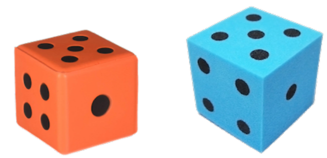
Serve an ad.
- User clicks w.p. 0.2
- Ignores otherwise

Let X: 1 if clicked

$$X \sim \text{Ber}(0.2)$$

$$P(X = 1) = 0.2$$

$$P(X = 0) = 0.8$$



Roll two dice.
- Success: roll two 6's
- Failure: anything else

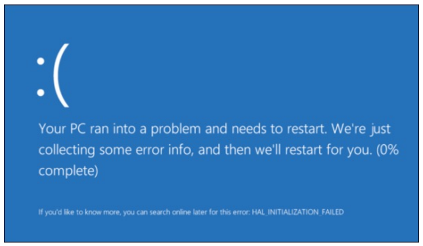
Let X : 1 if success

$$X \sim \text{Ber}(\tfrac{1}{36})$$

$$E[X] = \tfrac{1}{36}$$

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.

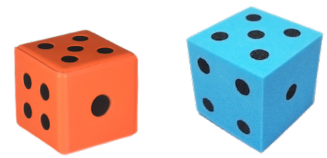
- User clicks w.p. 0.2
- Ignores otherwise

Let X: 1 if clicked

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

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

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



Roll two dice.

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

Let X : 1 if success

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

$$E[X] = \_\_$$

Binomial RV

Binomial Random Variable

$E[X] = \sum_{k=0}^{n} k \cdot \binom{n}{k} p^k (1-p)^{n-k}$

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

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

$X = B_1 + B_2 + B_3 + \cdots + B_n$ (each B_i with probability 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

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

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

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

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 (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) PMF

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

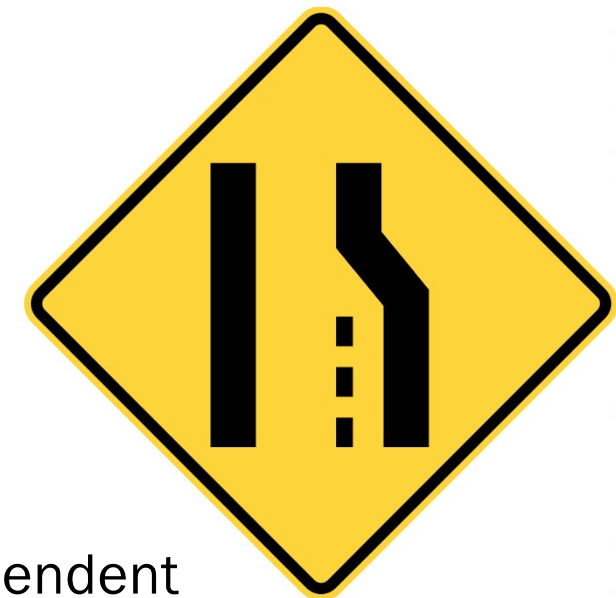
Examples:

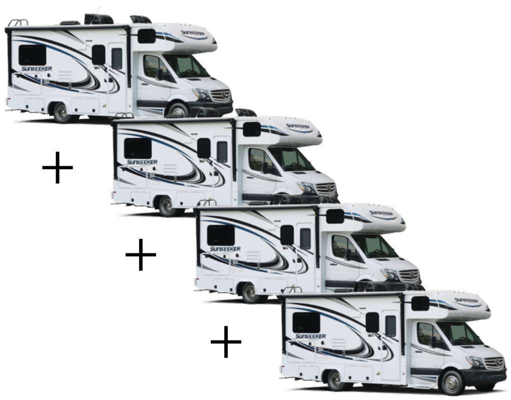
- # heads in n coin flips
- # of 1's in randomly generated length n bit string
- # of disk drives crashed in 1000 computer cluster (assuming disks crash independently)

Binomial 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)

Proof: before

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)

No, give me the variance proof right now

To simplify the algebra a bit, let $q = 1 - p$, so $p + q = 1$.

So:

$\mathrm{E}(X^2) = \sum_{k \geq 0}^{n} k^2 \binom{n}{k} p^k q^{n-k}$	Definition of Binomial Distribution: $p + q = 1$
$= \sum_{k=0}^{n} kn \binom{n-1}{k-1} p^k q^{n-k}$	Factors of Binomial Coefficient: $k\binom{n}{k} = n\binom{n-1}{k-1}$
$= np \sum_{k=1}^{n} k \binom{n-1}{k-1} p^{k-1} q^{(n-1)-(k-1)}$	Change of limit: term is zero when $k - 1 = 0$
$= np \sum_{j=0}^{m} (j+1) \binom{m}{j} p^j q^{m-j}$	putting $j = k - 1, m = n - 1$
$= np \left(\sum_{j=0}^{m} j \binom{m}{j} p^j q^{m-j} + \sum_{j=0}^{m} \binom{m}{j} p^j q^{m-j} \right)$	splitting sum up into two
$= np \left(\sum_{j=0}^{m} m \binom{m-1}{j-1} p^j q^{m-j} + \sum_{j=0}^{m} \binom{m}{j} p^j q^{m-j} \right)$	Factors of Binomial Coefficient: $j\binom{m}{j} = m\binom{m-1}{j-1}$
$= np \left((n-1)p \sum_{j=1}^{m} \binom{m-1}{j-1} p^{j-1} q^{(m-1)-(j-1)} + \sum_{j=0}^{m} \binom{m}{j} p^j q^{m-j} \right)$	Change of limit: term is zero when $j - 1 = 0$
$= np\left((n-1)p(p+q)^{m-1} + (p+q)^m\right)$	Binomial Theorem
$= np((n-1)p + 1)$	as $p + q = 1$
$= n^2p^2 + np(1-p)$	by algebra

Then:

$\mathrm{var}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2$	
$= np(1-p) + n^2p^2 - (np)^2$	Expectation of Binomial Distribution: $\mathrm{E}(X) = np$
$= np(1-p)$	

as required.

proofwiki.org

Exercises

Statistics: Expectation and variance

support $X = \{1,2,3,4\}$ $E[X] = 2.5$

1. a. Let X = the outcome of a fair 4-sided die roll. What is $E[X]$?

   b. Let Y = the sum of three rolls of a fair 4-sided die. What is $E[Y]$?

   7.5

2. Let Z = # of ***tails*** on 10 flips of a biased coin (w.p. 0.4 of heads). What is $E[Z]$?

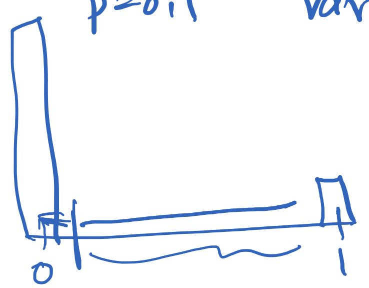
   $E(Z) = 6$

   $p(1-p)$

   $p = 0.5$ $Var(B_2) = 0.25$

   $p = 0.1$ $Var(B_1) = 0.09$

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



Statistics: Expectation and variance

1. a. Let X = the outcome of a fair 4-sided die roll. What is $E[X]$?
   b. Let Y = the sum of three rolls of a fair 4-sided die. What is $E[Y]$?

2. Let Z = # of ***tails*** on 10 flips of a biased coin (w.p. 0.4 of heads). What is $E[Z]$?

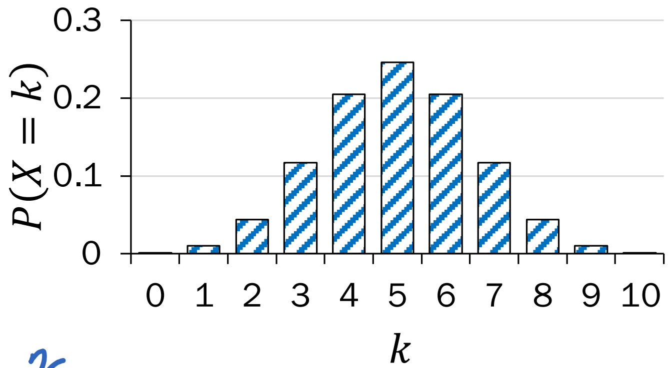
3. Compare the variances of $B_1 \sim \text{Ber}(0.1)$ and $B_2 \sim \text{Ber}(0.5)$.

If you can identify common RVs, just look up statistics instead of re-deriving from definitions.

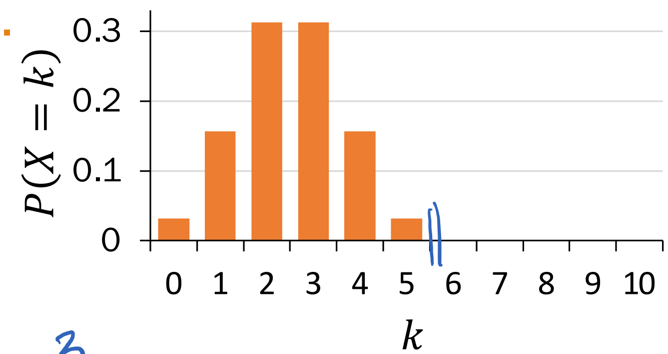
Visualizing Binomial PMFs

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

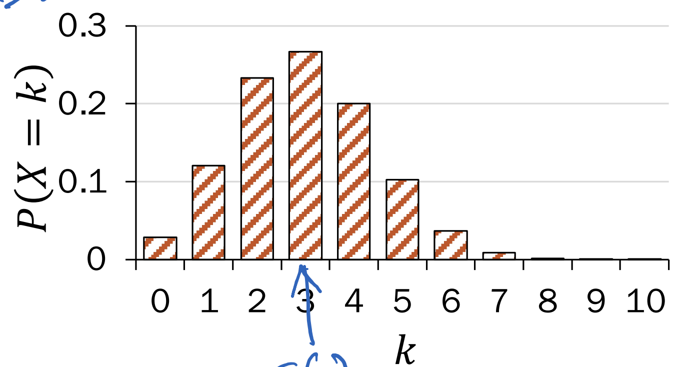
A.



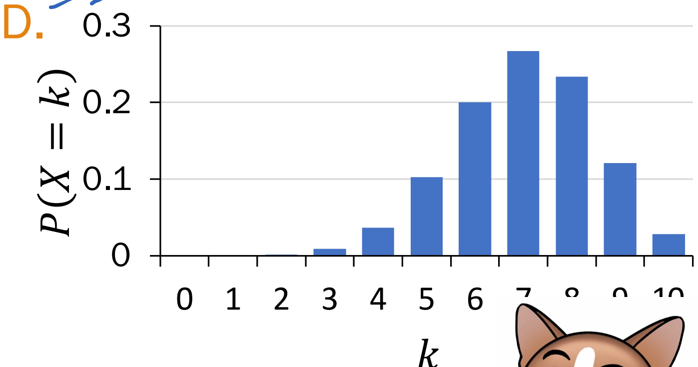
B.



C.



D.

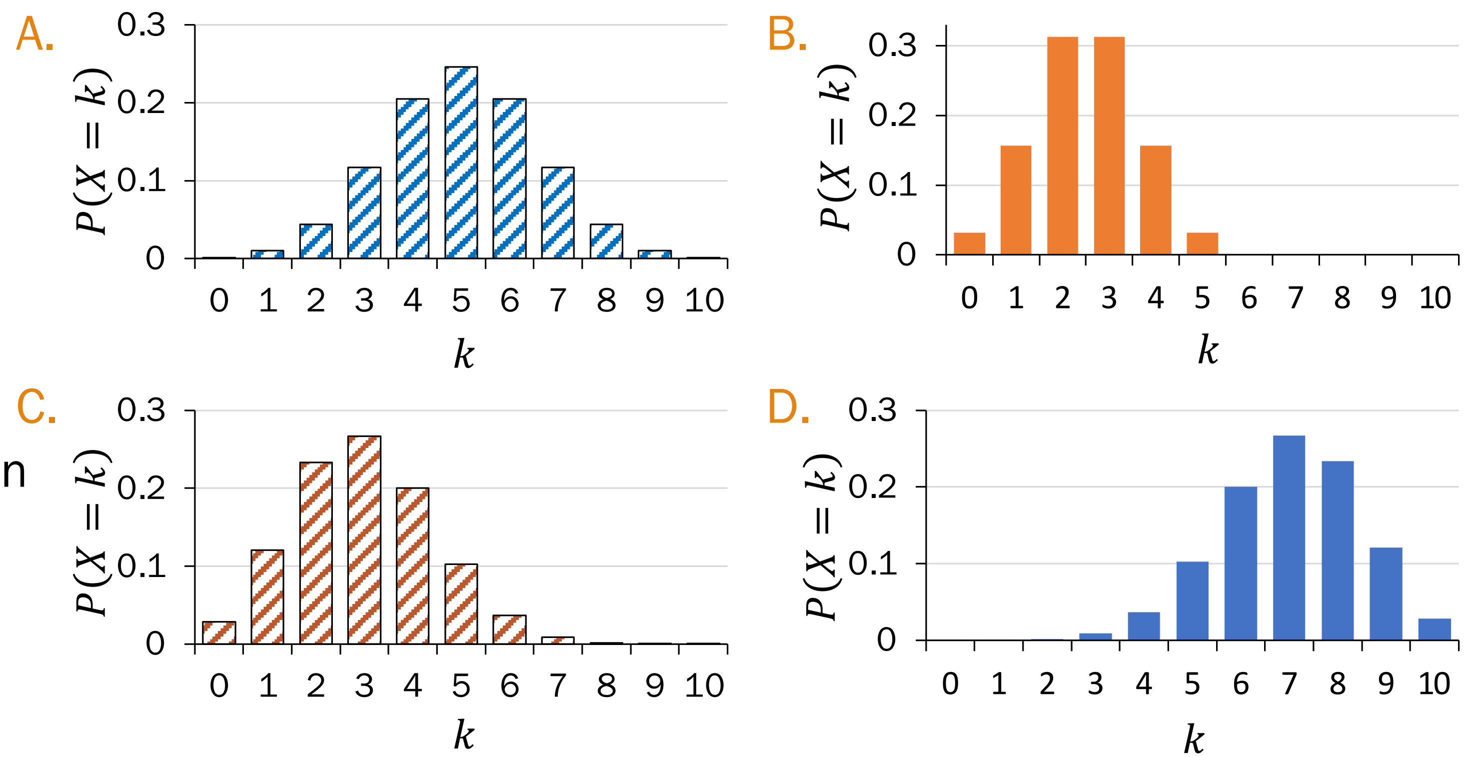


Match the distribution of X to the graph:

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

Visualizing Binomial PMFs

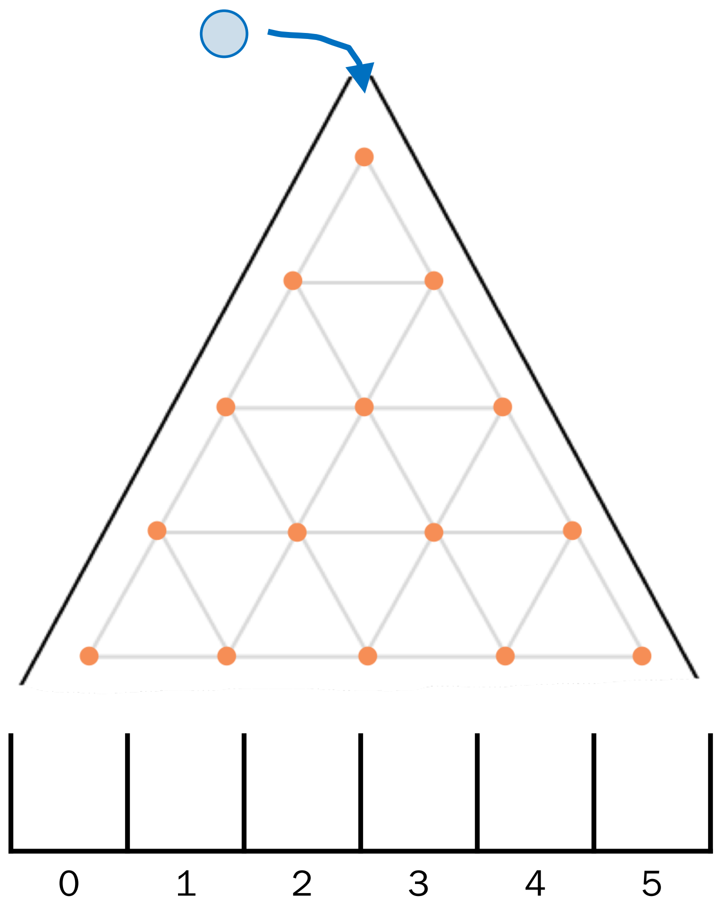
$X \sim \text{Bin}(n, p)$ $E[X] = np$ $p(i) = \binom{n}{k} p^k (1-p)^{n-k}$



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. $\text{Bin}(5, 0.5)$

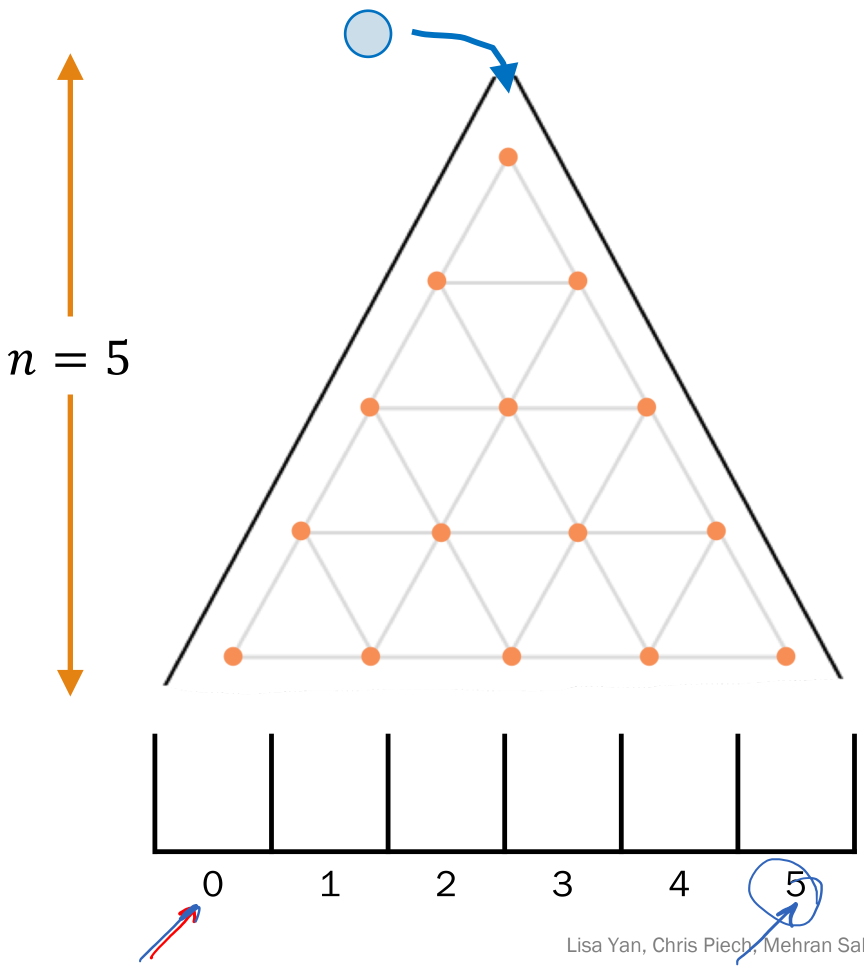
Galton Board



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

Galton Board

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



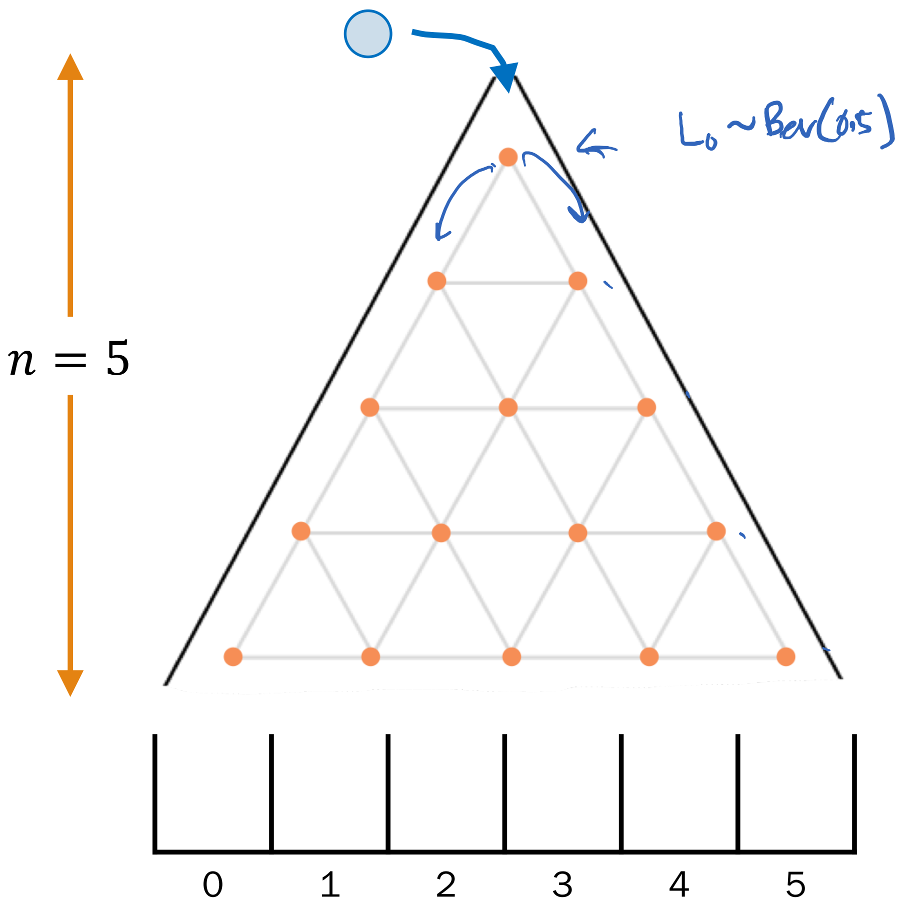
When a marble hits a pin, it has an equal chance of going left or right.
Let B = the bucket index a ball drops into.
What is the distribution of B?

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

(by yourself)

Galton Board

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



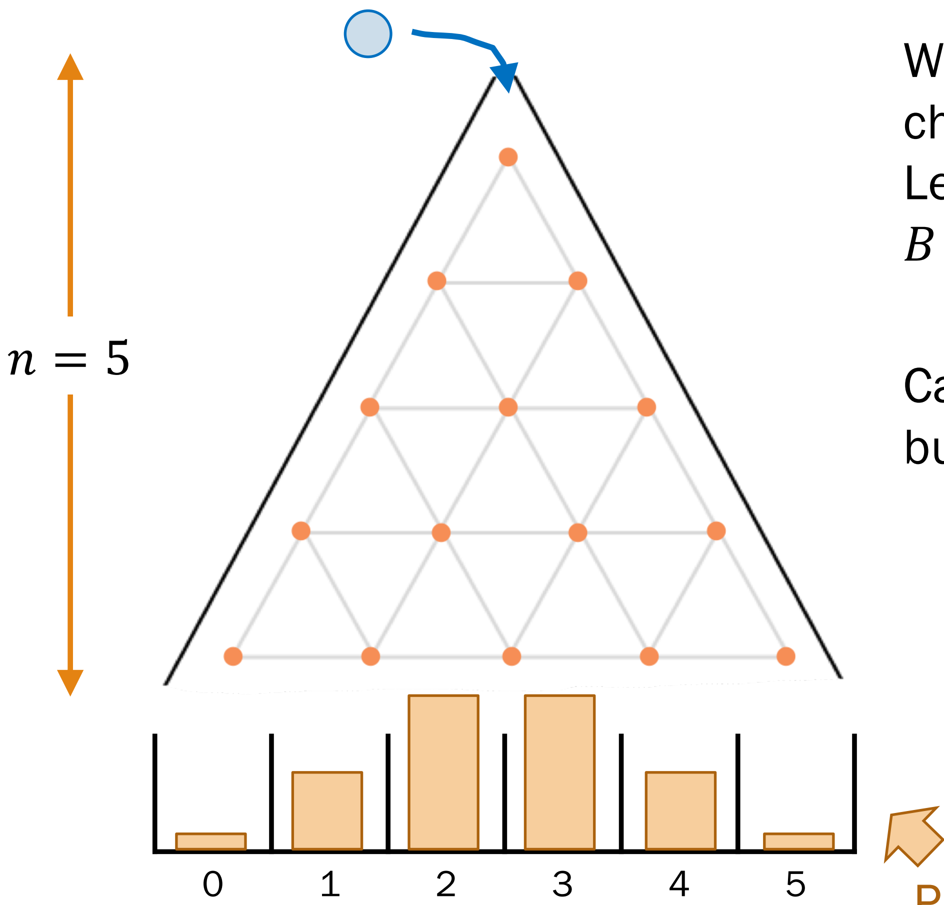
When a marble hits a pin, it has an equal chance of going left or right.
Let B = the bucket index a ball drops into.
What is the **distribution** of B?

- 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,

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

Calculate the probability of a ball landing in bucket k.

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

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

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

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 Utah Jazz will play the Golden State Warriors in a 7-game series during the 2022 NBA finals.
   - The Jazz 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(Utah Jazz 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?

$n=4$, $p_B = 0.75$

Random variable What differentiates our event from the rest of the sample space?

$X \in \{0,1,2,3,4\}$

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

X: # brown-eyed children,

$X \sim \text{Bin}(4, p)$] $p = 0.75$

p: $P(\text{brown}-\text{eyed child})$

Want: $P(X = 3)$

2. Identify known probabilities

$p = 0.75$

	R	L
R	.25	.25
L	.25	.25

3. Solve

$P(X=3) = \binom{4}{3}(0.75)^3(0.25)^1$

RRRL RLRR

RRLR LRRR

NBA Finals

Let's speculate: the Utah Jazz will play the Golden State Warriors in a 7-game series during the 2022 NBA finals.

- The Jazz have a probability of 58% of winning each game, independently. $p = 0.58$
- A team wins the series if they win at least 4 games (we play all 7 games).

What is P(Utah Jazz winning)?

1. Define events/ RVs & state goal

$X = \{0, 1, 2, \ldots, 7\}$
$X \sim \text{Bin}(7, 0.58)$

X: # games Utah Jazz win
$X \sim \text{Bin}(7, 0.58)$

Want:

Big Q: Fixed parameter or random variable?

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

NBA Finals

Let's speculate: the Utah Jazz will play the Golden State Warriors in a 7-game series during the 2022 NBA finals.

- The Jazz 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).

WWWW [???]
WWLWW [??]
WLWWW [??]

What is P(Utah Jazz winning)?

1. Define events/ RVs & state goal

X: # games Utah Jazz 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