

07: Variance, Bernoulli, Binomial

Jerry Cain
April 11, 2022

Table of Contents

2	Variance
9	Properties of Variance
16	Bernoulli RV
20	Binomial RV
31	Exercises



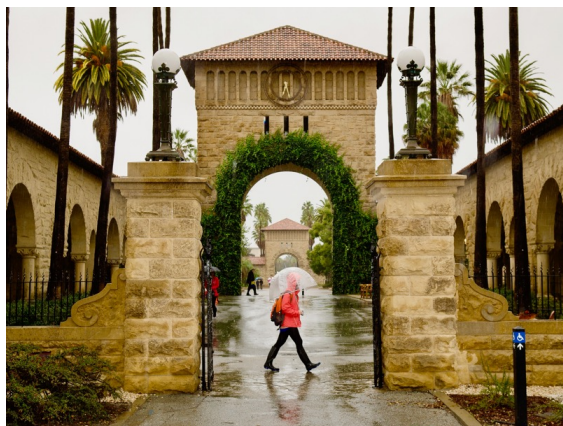
Variance

Average temperatures

Stanford, CA

$$E[\text{high}] = 68^\circ\text{F}$$

$$E[\text{low}] = 52^\circ\text{F}$$



Washington, DC

$$E[\text{high}] = 67^\circ\text{F}$$

$$E[\text{low}] = 51^\circ\text{F}$$



Is $E[X]$ enough?

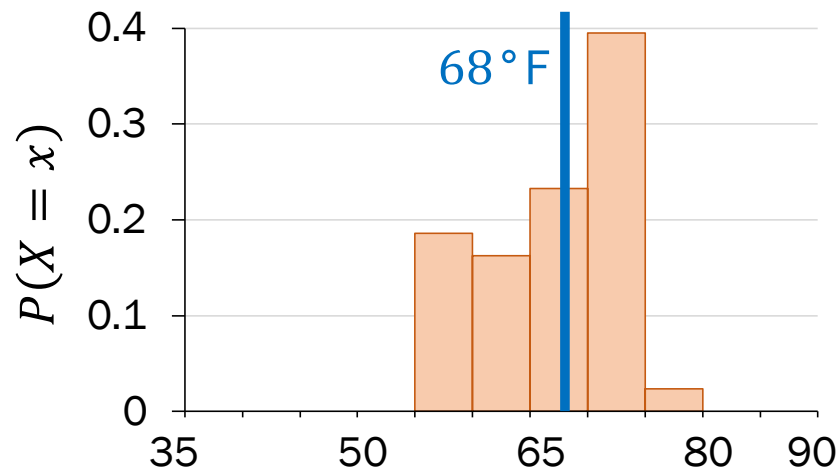
Average temperatures

Stanford, CA

$E[\text{high}] = 68^\circ\text{F}$

$E[\text{low}] = 52^\circ\text{F}$

Stanford high temps

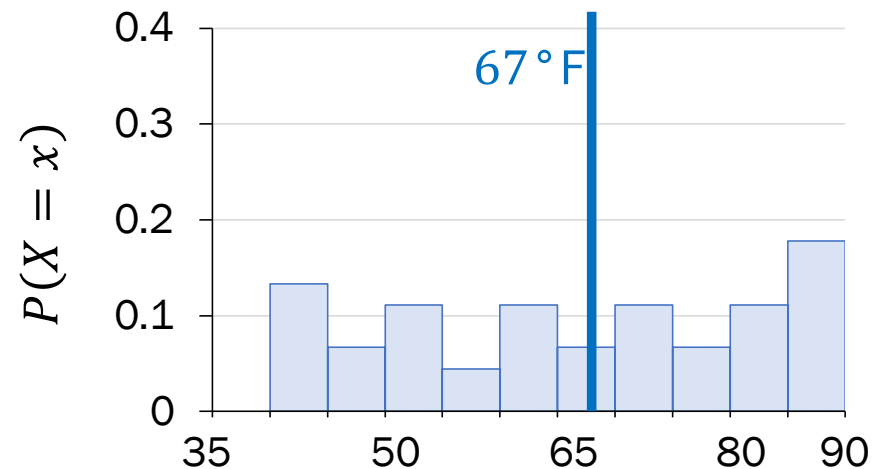


Washington, DC

$E[\text{high}] = 67^\circ\text{F}$

$E[\text{low}] = 51^\circ\text{F}$

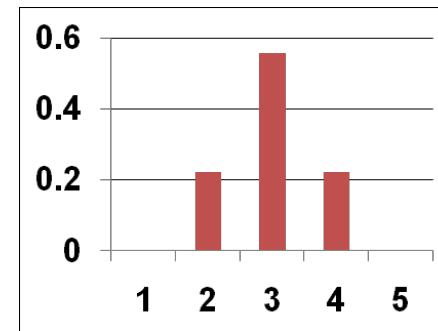
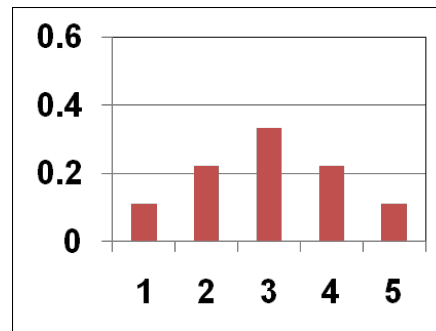
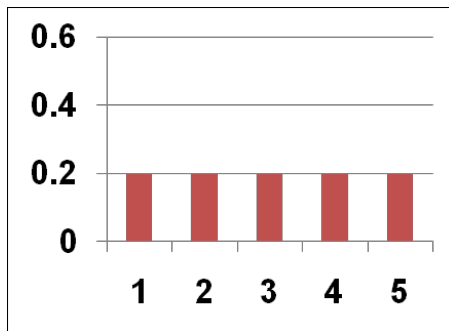
Washington high temps



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

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: $\text{Var}(X) \geq 0$
- Other names: **2nd central moment**, or square of the standard deviation

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

Variance of Stanford weather

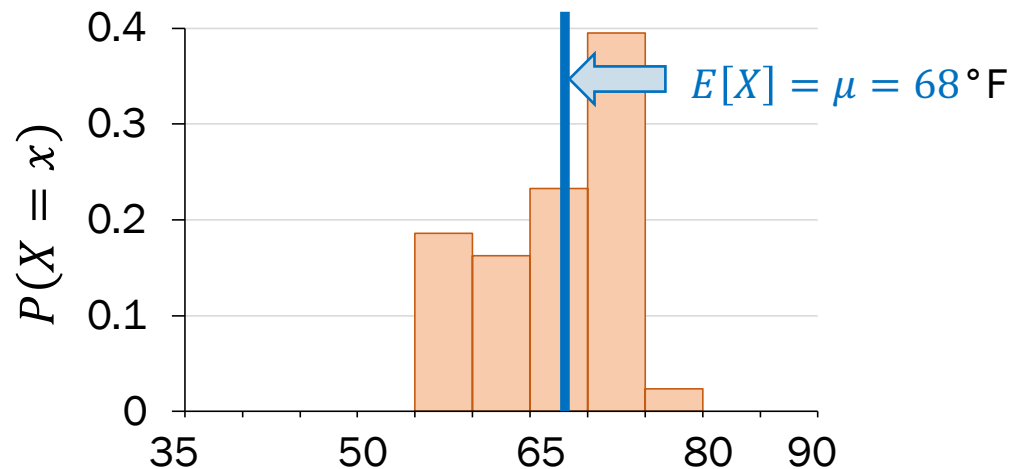
$$\text{Var}(X) = E[(X - E[X])^2] \quad \text{Variance of } X$$

Stanford, CA

$E[\text{high}] = 68^\circ\text{F}$

$E[\text{low}] = 52^\circ\text{F}$

Stanford high temps



X	$(X - \mu)^2$
57°F	$121 (\text{°F})^2$
71°F	$9 (\text{°F})^2$
75°F	$49 (\text{°F})^2$
69°F	$1 (\text{°F})^2$
...	...

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

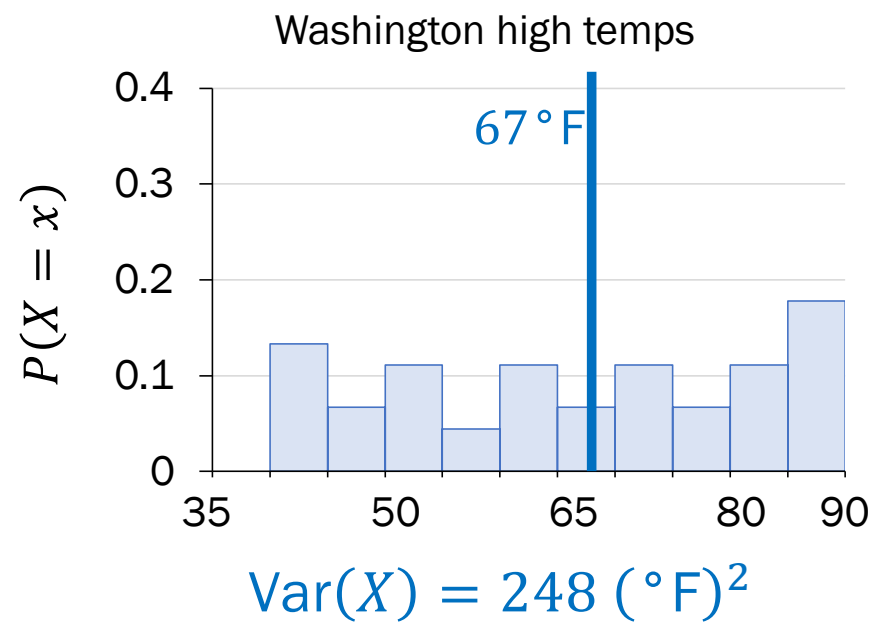
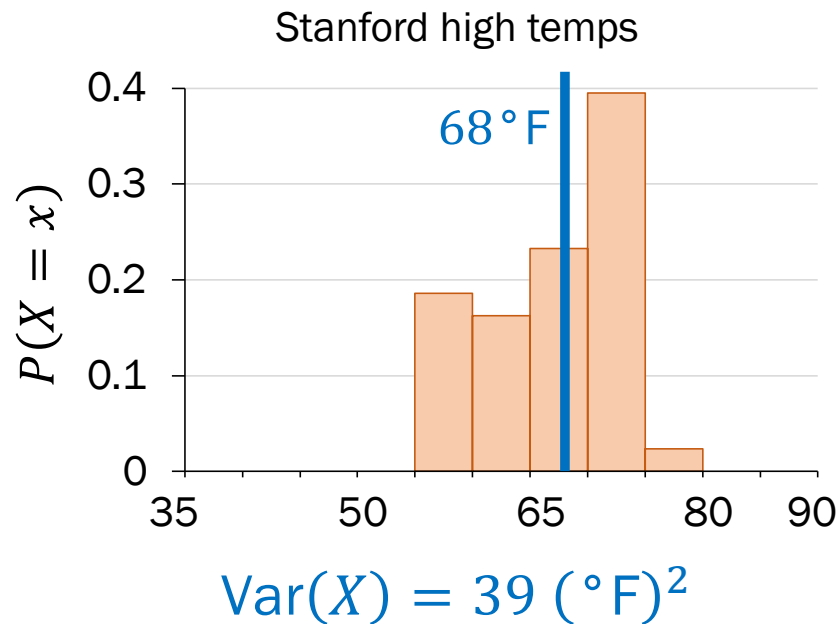
Standard deviation $= 6.2^\circ\text{F}$

Comparing variance

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

Stanford, CA
 $E[\text{high}] = 68^\circ\text{F}$

Washington, DC
 $E[\text{high}] = 67^\circ\text{F}$





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

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

Computing variance, a proof

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

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

$$\text{Let } E[X] = \mu$$

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

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

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

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

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

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

Calculate the variance of Y .



1. Approach #1: Definition

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

2. Approach #2: A property

2nd moment

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

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

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

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

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

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

$$= a^2E[X^2] + 2abE[X] + b^2 - (a^2(E[X])^2 + 2abE[X] + b^2) \quad \left. \vphantom{E[X^2]} \right\} \begin{array}{l} \text{Factoring/} \\ \text{Linearity of} \\ \text{Expectation} \end{array}$$

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

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

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



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

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

Support: $\{0,1\}$

PMF

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

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

Expectation

$$E[X] = p$$

Variance

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

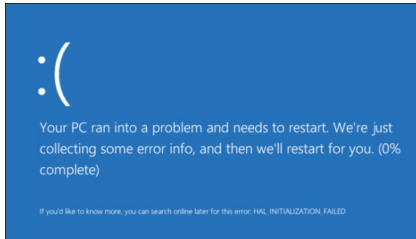
Examples:

- Coin flip
- Random binary digit
- Whether Doris barks

Remember this nice property of expectation.

Defining Bernoulli RVs

$$\begin{aligned} X \sim \text{Ber}(p) & \quad p_X(1) = p \\ E[X] = p & \quad p_X(0) = 1 - p \end{aligned}$$



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.

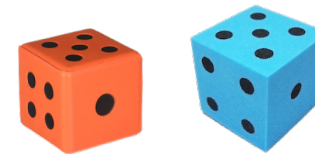
- User clicks w.p. 0.2
- Ignores otherwise

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

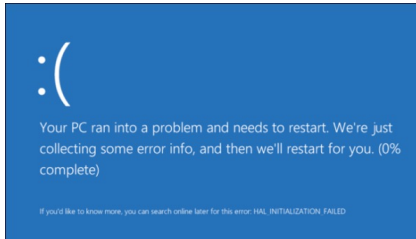
$$X \sim \text{Ber}(___)$$

$$E[X] = ___$$



Defining Bernoulli RVs

$$\begin{aligned} X \sim \text{Ber}(p) & \quad p_X(1) = p \\ E[X] = p & \quad p_X(0) = 1 - p \end{aligned}$$



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.

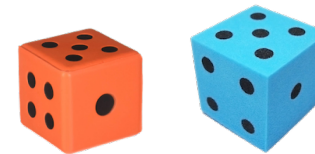
- User clicks w.p. 0.2
- Ignores otherwise

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

Consider an experiment: n independent trials of $\text{Ber}(p)$ 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

$$E[X] = np$$

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

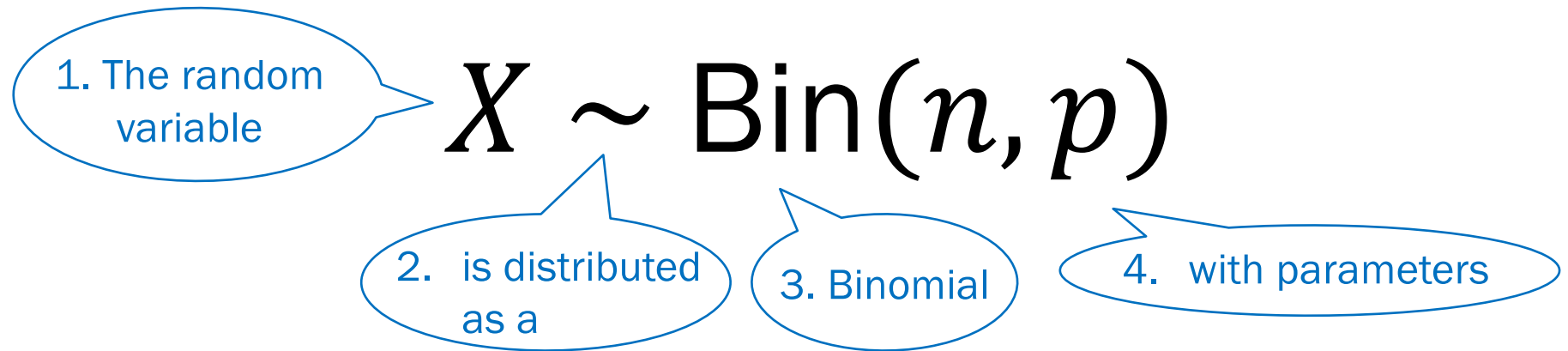
Variance

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

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)

Reiterating notation



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

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

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:

$$P(X = 0)$$

$$P(X = 1)$$

$$P(X = 2)$$

$$P(X = 3)$$

$$P(X = 7)$$

$P(\text{event})$



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

Lisa Yan, Chris Piech, Mehran Sahami, and Jerry Cain, CS109, Spring 2022

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 $\text{Ber}(p)$ random variables.

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

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

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

PMF

$k = 0, 1, \dots, n:$

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

Expectation

$$E[X] = np$$

Variance

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

Examples:

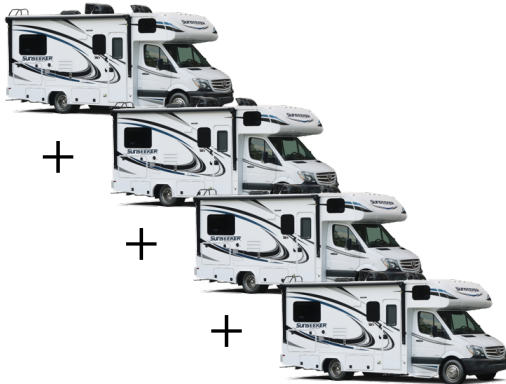
- # heads in n coin flips
- # of 1's in randomly generated length n bit string
- # of disk drives crashed in 1000 computer cluster (assuming disks crash independently)

Binomial RV is sum of Bernoulli RVs



Bernoulli

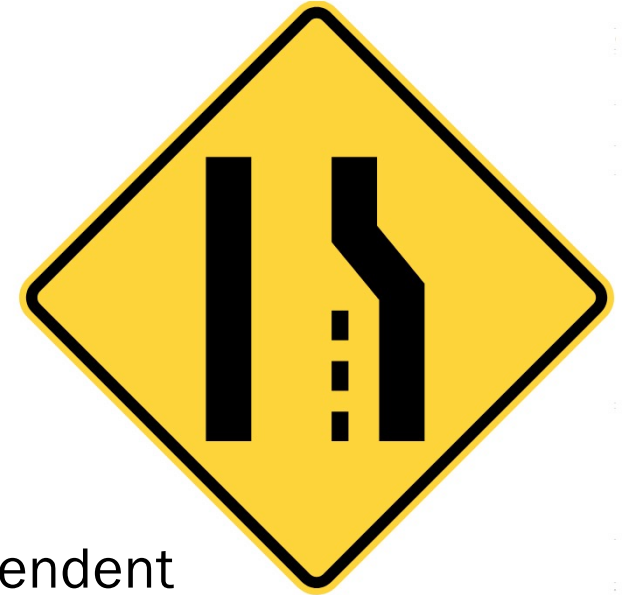
- $X \sim \text{Ber}(p)$



Binomial

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

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



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

Binomial Random Variable

Consider an experiment: n independent trials of $\text{Ber}(p)$ random variables.

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

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

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

PMF

$k = 0, 1, \dots, n:$

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

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:

Binomial Random Variable

Consider an experiment: n independent trials of $\text{Ber}(p)$ random variables.

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

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

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

PMF

$k = 0, 1, \dots, n$:

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

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:

$$\begin{aligned}
 E(X^2) &= \sum_{k \geq 0}^n k^2 \binom{n}{k} p^k q^{n-k} \\
 &= \sum_{k=0}^n kn \binom{n-1}{k-1} p^k q^{n-k} \\
 &= np \sum_{k=1}^n k \binom{n-1}{k-1} p^{k-1} q^{(n-1)-(k-1)} \\
 &= np \sum_{j=0}^m (j+1) \binom{m}{j} p^j q^{m-j} \\
 &= np \left(\sum_{j=0}^m j \binom{m}{j} p^j q^{m-j} + \sum_{j=0}^m \binom{m}{j} p^j q^{m-j} \right) \\
 &= np \left(\sum_{j=0}^m m \binom{m-1}{j-1} p^j q^{m-j} + \sum_{j=0}^m \binom{m}{j} p^j q^{m-j} \right) \\
 &= np \left((n-1)p \sum_{j=1}^m \binom{m-1}{j-1} p^{j-1} q^{(m-1)-(j-1)} + \sum_{j=0}^m \binom{m}{j} p^j q^{m-j} \right) \\
 &= np((n-1)p(p+q)^{m-1} + (p+q)^m) \\
 &= np((n-1)p+1) \\
 &= n^2 p^2 + np(1-p)
 \end{aligned}$$

Definition of **Binomial Distribution**: $p + q = 1$

Factors of Binomial Coefficient: $k \binom{n}{k} = n \binom{n-1}{k-1}$

Change of limit: term is zero when $k - 1 = 0$

putting $j = k - 1$, $m = n - 1$

splitting sum up into two

Factors of Binomial Coefficient: $j \binom{m}{j} = m \binom{m-1}{j-1}$

Change of limit: term is zero when $j - 1 = 0$

Binomial Theorem

as $p + q = 1$

by algebra

Then:

$$\begin{aligned}
 \text{var}(X) &= E(X^2) - (E(X))^2 \\
 &= np(1-p) + n^2 p^2 - (np)^2 \quad \text{Expectation of Binomial Distribution: } E(X) = np \\
 &= np(1-p)
 \end{aligned}$$

as required.



Exercises

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 $B_2 \sim \text{Ber}(0.5)$.



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 $E[Z]$?

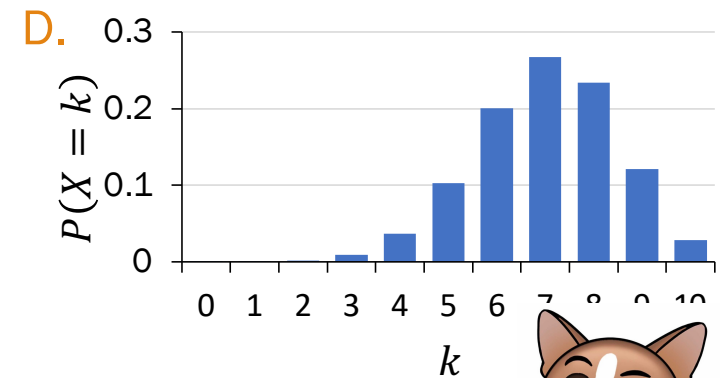
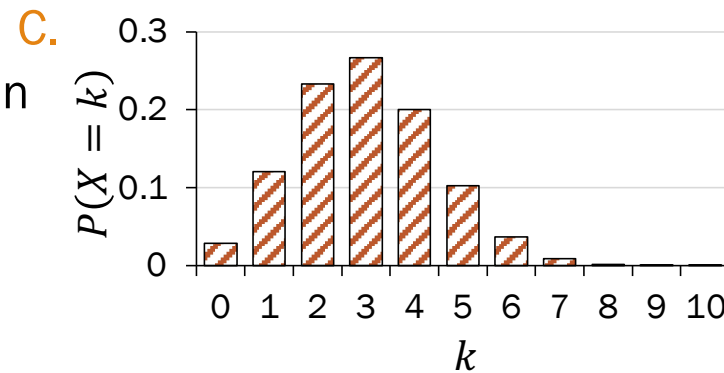
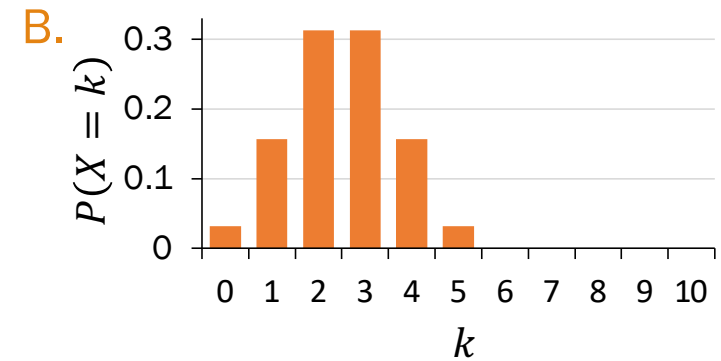
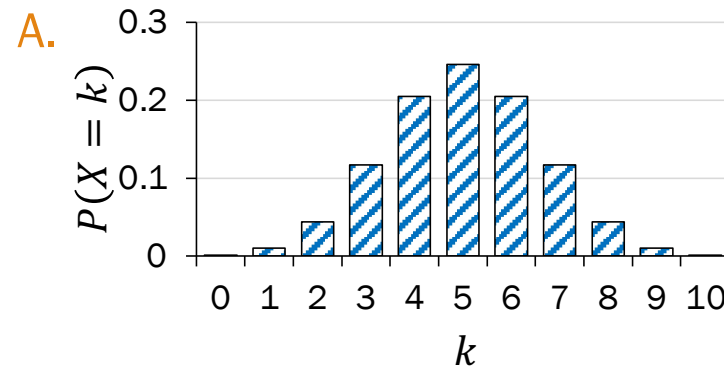
3. Compare the variances of $B_1 \sim \text{Ber}(0.1)$ and $B_2 \sim \text{Ber}(0.5)$.

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

Visualizing Binomial PMFs

$$X \sim \text{Bin}(n, p) \quad E[X] = np$$

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



Match the distribution of X to the graph:

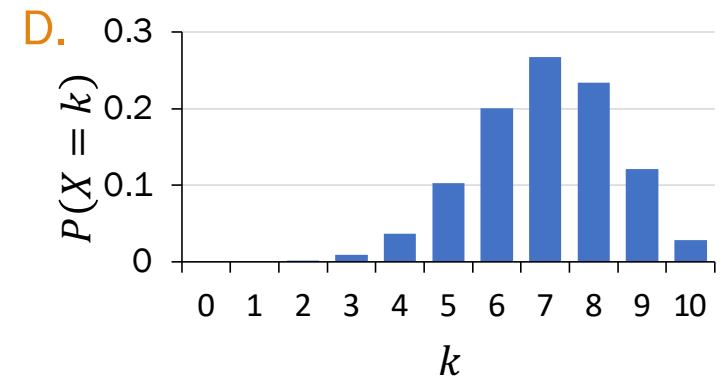
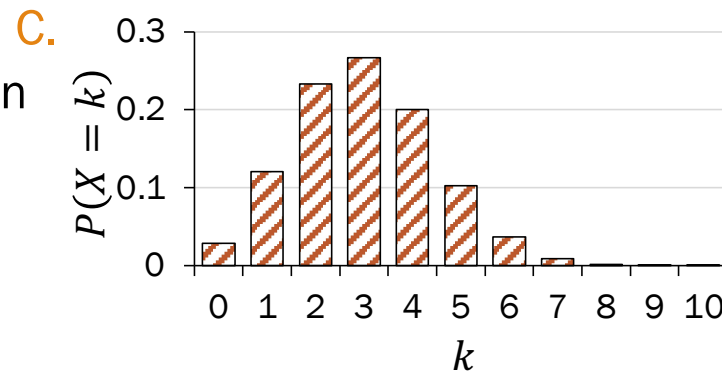
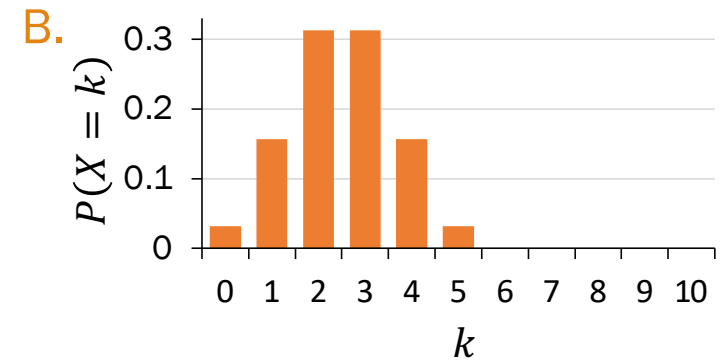
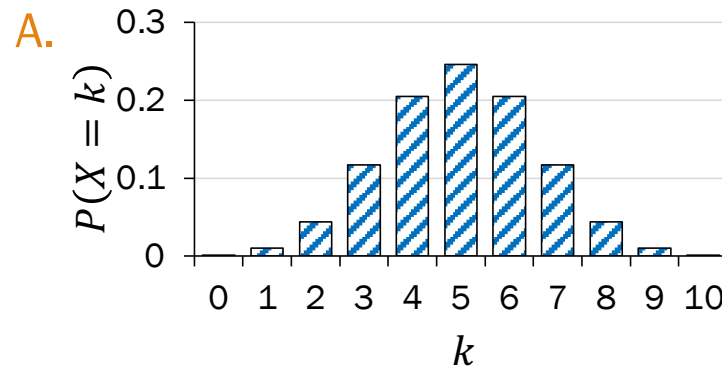
1. Bin(10,0.5)
2. Bin(10,0.3)
3. Bin(10,0.7)
4. Bin(5,0.5)



Visualizing Binomial PMFs

$$E[X] = np$$

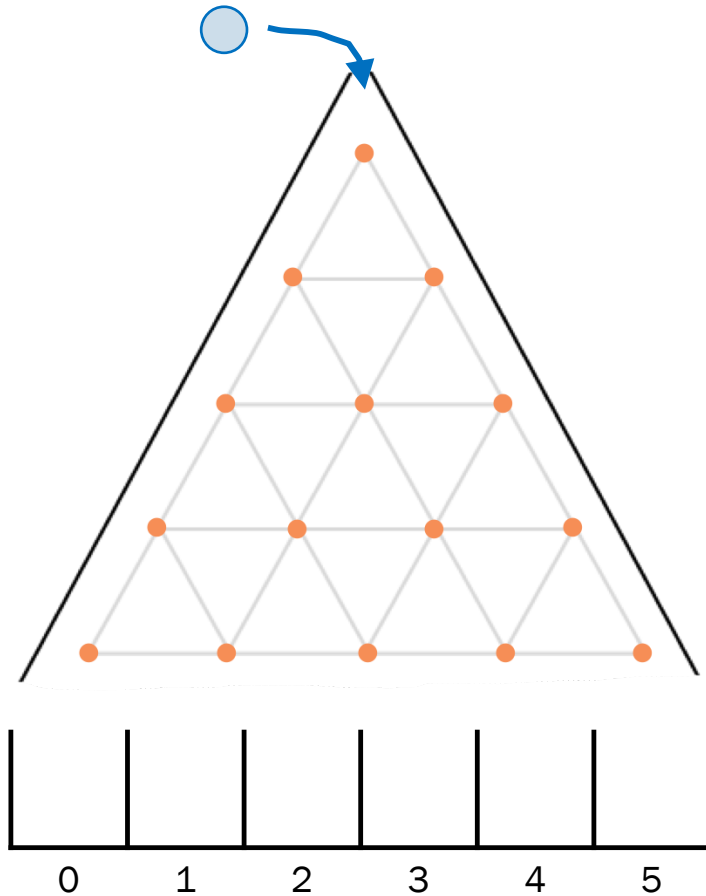
$$X \sim \text{Bin}(n, p) \quad p(i) = \binom{n}{k} p^k (1-p)^{n-k}$$



Match the distribution of X to the graph:

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

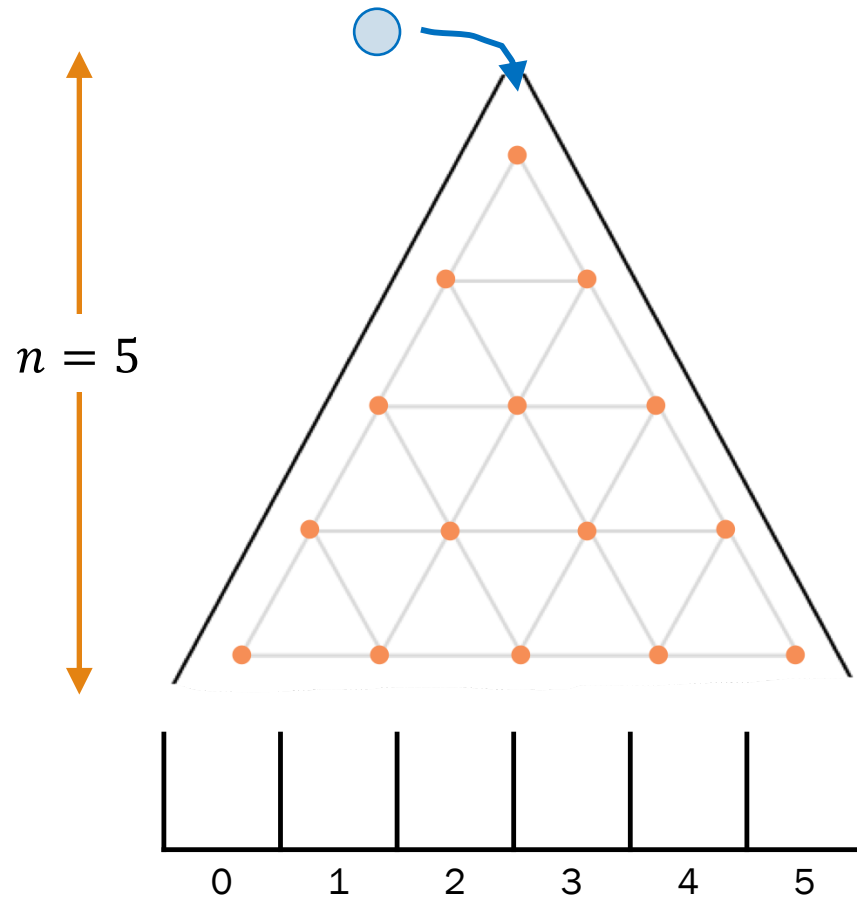
Galton Board



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

Galton Board

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



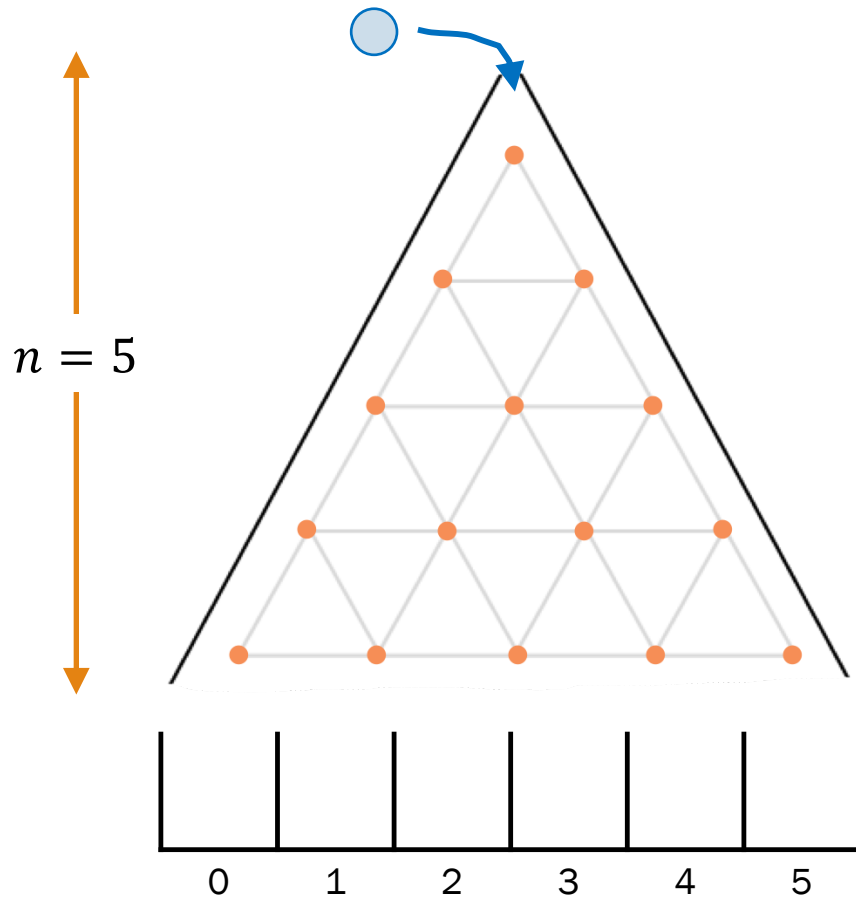
When a marble hits a pin, it has an equal chance of going left or right.
Let B = the **bucket index** a ball drops into.
What is the **distribution** of B ?

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



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.

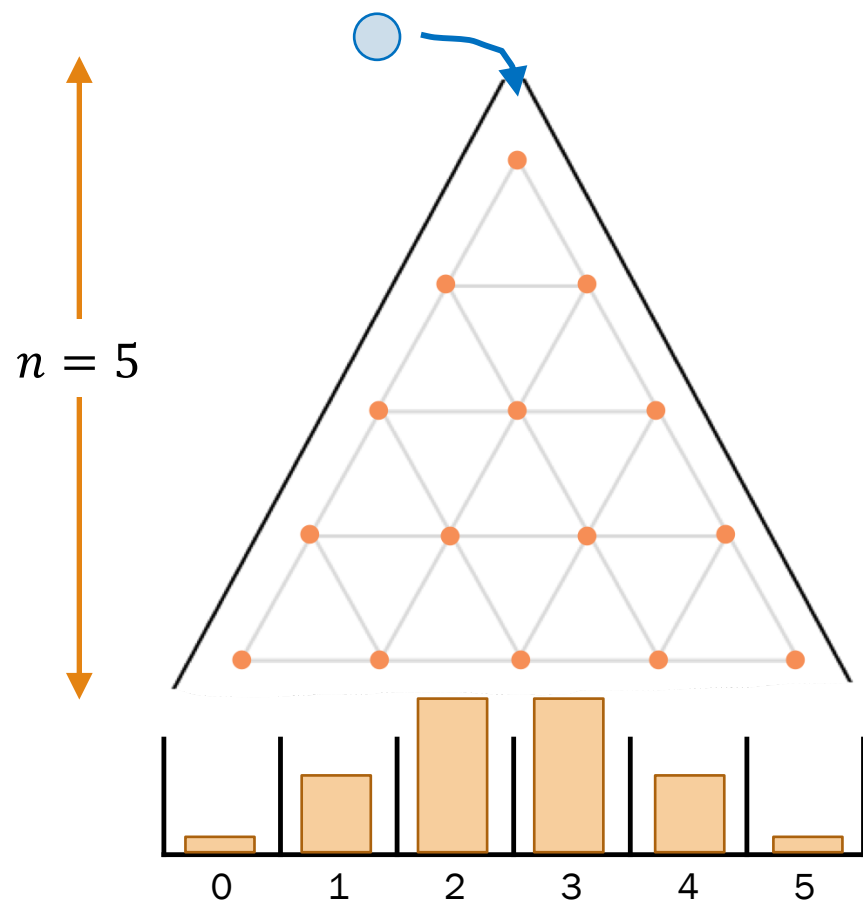
Let B = the **bucket index** a ball drops into.
What is the **distribution** of B ?

- Each pin is an independent trial
- One decision made for **level** $i = 1, 2, \dots, 5$
- 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 $P(\text{exactly 3 children have brown eyes})$?

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 $P(\text{Utah Jazz winning})$?



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 $P(\text{exactly 3 children have brown eyes})$?

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 $P(\text{exactly 3 children have brown eyes})$?

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

X : # brown-eyed children,

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

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

Want: $P(X = 3)$

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).

What is $P(\text{Utah Jazz winning})$?

1. Define events/ RVs & state goal

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).

What is $P(\text{Utah Jazz winning})$?

1. Define events/
RVs & state goal
2. Solve

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

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

Want: $P(X \geq 4)$

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