

Binomial, Bernoulli and Variance

Chris Piech

CS109, Stanford University



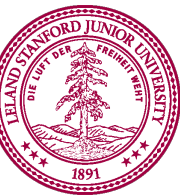
Announcements

High-Resolution Course Feedback

Subject:

[CS109] Please Submit Anonymous Feedback for Your Class

At the end of the course we see who completed the survey (but not your responses). I don't factor it into your grade, but it helps us a lot!



Follow Along!

<https://cs109psets.netlify.app/win22/lecture7/>

The screenshot shows a web browser window with the URL `cs109psets.netlify.app/win22/lecture7/warriors_series`. The page title is "L7 Warriors Series". The main content area contains a question: "Warriors are going to play the Bucks in a best of 7 series during the 2022 NBA finals. What is the probability that the warriors win the series? Each game is independent. Each game, the warriors have a 0.55 probability of winning. Win series if you win at least 4 games." Below the text is an image of two Golden State Warriors players, Draymond Green (number 23) and Stephen Curry (number 30), high-fiving on the court. A note below the image states: "Note: without loss of generality you could imagine that the two teams always play all 7 games, regardless of the outcome. Technically they stop playing after one team has achieved 4 wins, because the outcomes of the games no longer impact who wins. However, you could imagine that they continue." At the bottom of the question area, it asks: "What is the probability that the warriors win the series? Leave your" and provides "Previous Question" and "Next Question" buttons.

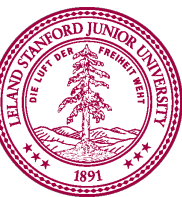
On the right side of the browser window, there is an "Answer Editor" section. It shows a "Numeric Answer" of `0.60828779687500` with a green checkmark and a "Check Answer" button. Below that is a "Code" section with a text area containing the following Python code:

```
1 from scipy import stats
2
3 # method 1: sum over PMF
4 total_pr = 0
5 for x in range(4, 8):
6     pr_win_x_games = stats.binom.pmf(x, 7, 0.55)
7     print(f'P(Wins = {x}) = {pr_win_x_games}')
8     total_pr += pr_win_x_games
9 print('')
10 print(total_pr)
```

Below the code is a blue "Run" button. Underneath is a "Console" section showing the output of the code:

```
P(Wins = 4) = 0.29184774609375
P(Wins = 5) = 0.21402168046875003
P(Wins = 6) = 0.08719401796875002
P(Wins = 7) = 0.015224352343750008

0.6082877968750001
```



Concept Checks

<https://edstem.org/us/courses/16856/discussion/999944>

Q1 Coin Flips

0 Points

You have one biased coin (where the probability of getting heads is 0.7 on any flip) and one unbiased coin. You randomly select one of these two coins (each has probability of 0.5 of being chosen) and proceed to flip it exactly 10 times.

Q1.1 Coin Flips: Seven Up

0 Points

What is the probability that you get 7 heads in all 10 flips? (to 3 decimal places, formatted as 0.abc)

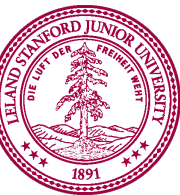
=0.192+-0.001

Answer:

Let F = you flip 7 heads. Let E = you choose the biased coin. Then $P(F|E) = \binom{10}{7} (0.7)^7 (0.3)^3$ and $P(F|E^C) = \binom{10}{7} (0.5)^7 (0.5)^3$. Using the Law of Total Probability, $P(F) = P(F|E)P(E) + P(F|E^C)P(E^C) \approx 0.192$

EXPLANATION

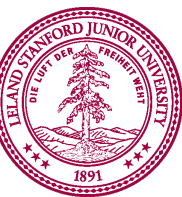
You got it correct!



Only Get Graded if Enrolled

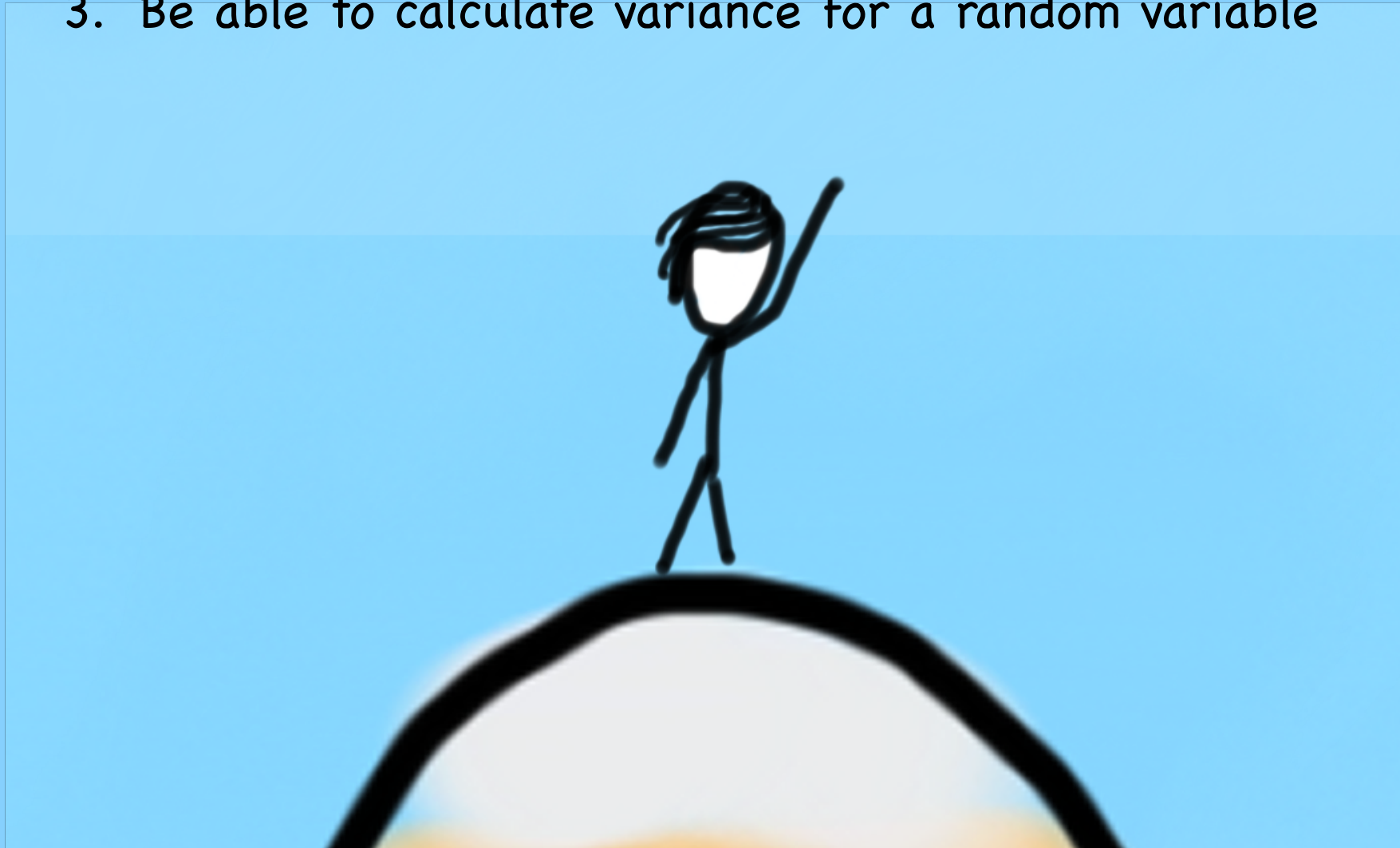
CS 109-01 LEC (8455)	M W F 01:30p-03:00p Huang Engineering Center, Rm NVIDIA AUD	View	N/A	✉ 242 Enrolled ✉ 0 Waitlisted
----------------------	--	----------------------	-----	----------------------------------

These 242 get graded

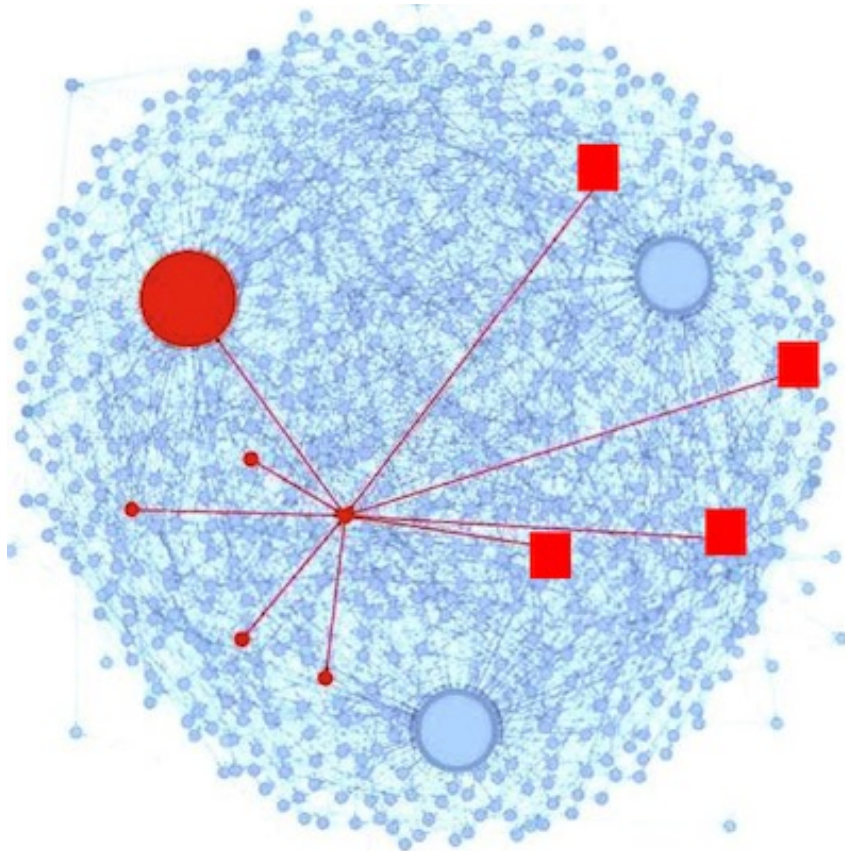


Learning Goals

1. Be able to recognize and use a Binomial Random Var
2. Be able to recognize and use a Bernoulli Random Var
3. Be able to calculate variance for a random variable

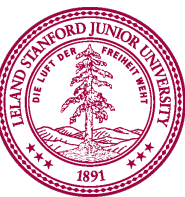


Is Peer Grading Accurate Enough?



Peer Grading on Coursera HCI.

31,067 peer grades for 3,607 students.



Review



A **random variable** is a number which takes on values probabilistically.



A discrete random variable is fully described by a **probability mass function**.



We can also calculate **summary statistics** such as expectation (and today, variance)



A **random variable** is a number which takes on values probabilistically.



A discrete random variable is fully described by a **probability mass function**.



We can also calculate **summary statistics** such as expectation (and today, variance)

Let Y be a random variable



Y

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

Let Y be a random variable



$$Y = 2$$

*note: here equals means `==` in coding

It is an event when
 Y takes on a value

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

Let Y be a random variable

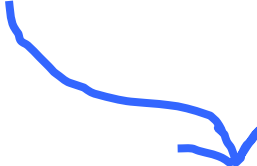
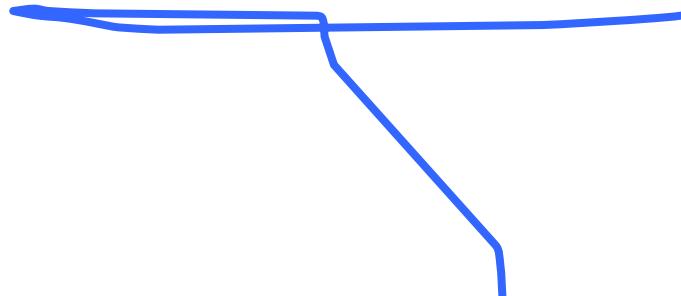


$$Y < 3$$

It is an event when
you ask any comparison question

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

If this is a number


$$P(Y = 2)$$


Then this is a probability
(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

This is a function

$$P(Y = k)$$

The diagram illustrates a function. A blue arrow points from the text $k = 5$ to the variable k in the expression $P(Y = k)$. A second blue arrow points from the expression $P(Y = k)$ down to the numerical value 0.03125 .

$$k = 5 \qquad 0.03125$$

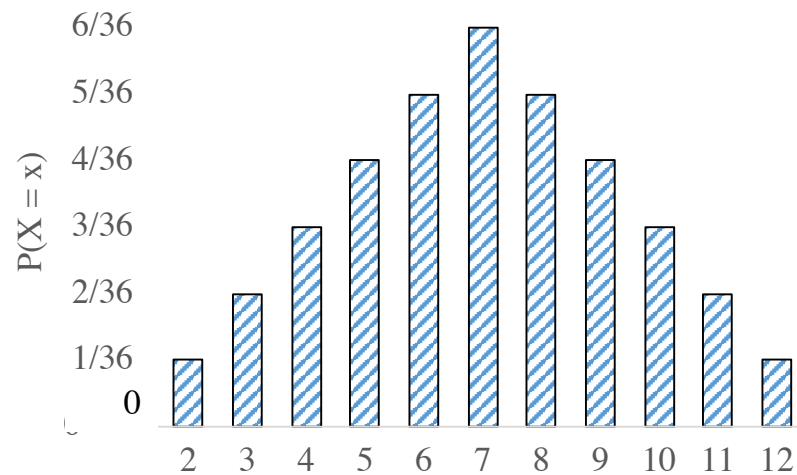
For example Y is the number of heads in 5 coin flips

Random Variables are a big deal, because they allow other people to give you a PMF (and other helpful equations)

PMF as an Equation

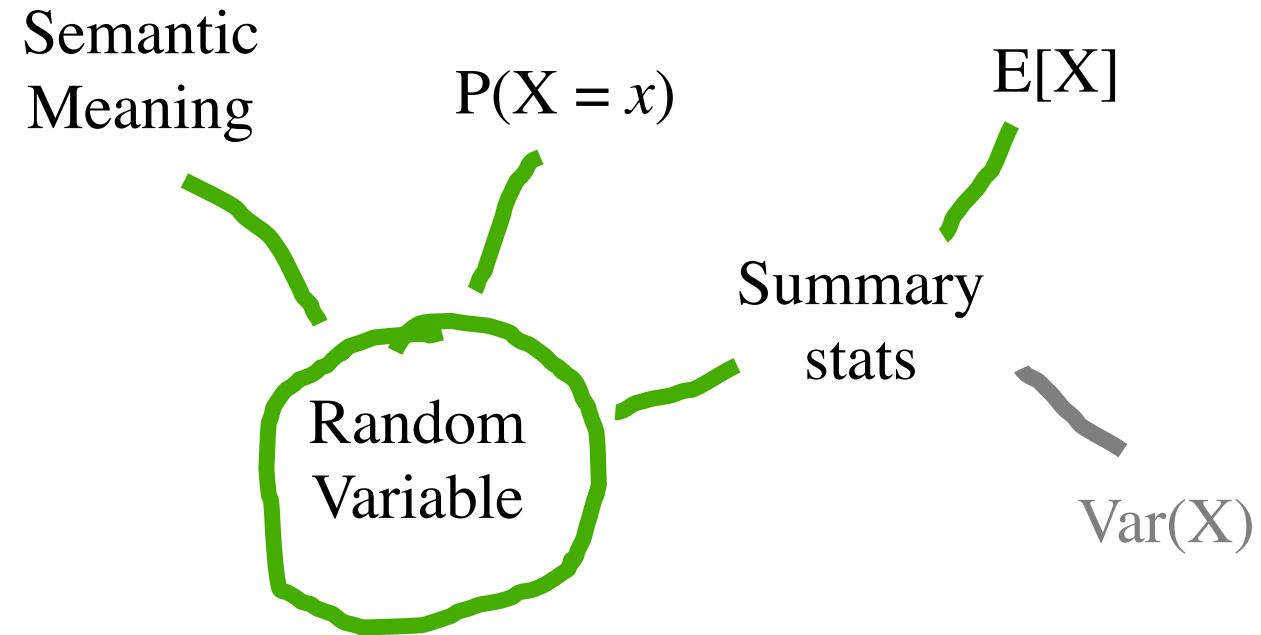
$$p(X = x) = \begin{cases} \frac{x-1}{36} & \text{if } x \in \mathbb{Z}, 1 \leq x \leq 6 \\ \frac{13-x}{36} & \text{if } x \in \mathbb{Z}, 7 \leq x \leq 12 \\ 0 & \text{else} \end{cases}$$

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



*errata: in lecture this formula had some small mistakes 😊

Fundamental Properties



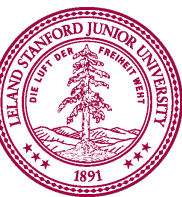
Expected Value

$$E[X] = \sum_x x \cdot P(X = x)$$

The value

The probability of that value

Loop over all values x that X can take on



Properties of Expectation

Linearity:

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

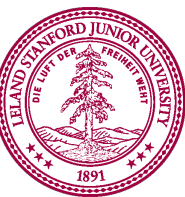
- Consider $X = 6$ -sided die roll, $Y = 2X - 1$.
- $E[X] = 3.5$ $E[Y] = 6$

Expectation of a sum is the sum of expectations

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

Unconscious statistician:

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



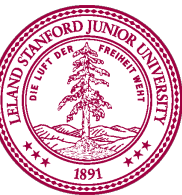
Expectation from Data

X
3
2
6
10
1
1
5
4
...

$$E[X] = \sum_x x \cdot P(X = x)$$

$$\approx \sum_x x \cdot \frac{\text{count}(X = x)}{N}$$

$$\approx \frac{1}{N} \sum_{\text{values } v} v$$



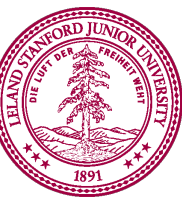
Expectation of Sum is Sum of Expectations

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

X	Y	$X+Y$
3	4	7
2	2	4
6	8	14
10	23	33
1	-3	-2
1	0	1
5	9	14
4	1	5
...

$$\frac{1}{n} \sum_{i=1}^n x_i + \frac{1}{n} \sum_{i=1}^n y_i = \frac{1}{n} \sum_{i=1}^n (x_i + y_i)$$

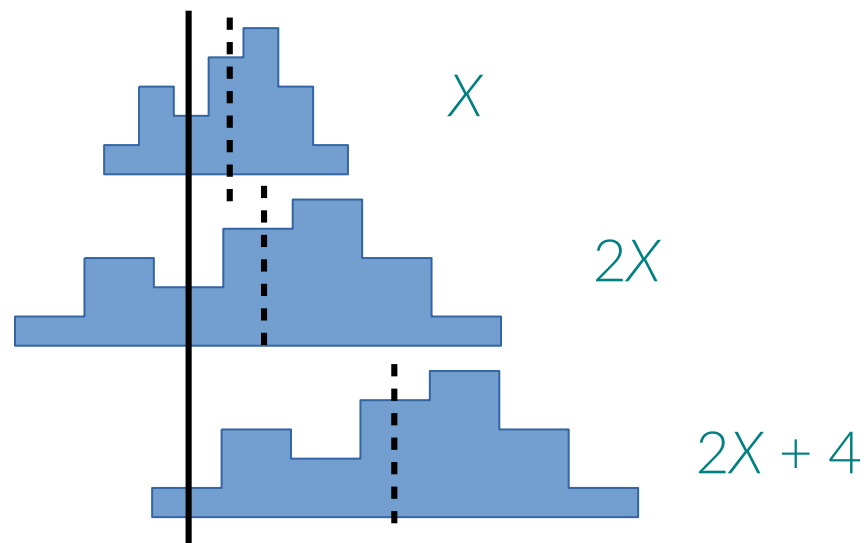
$$E(X) + E(Y) = E(X+Y)$$



Linearity of Expectation

Adding random variables or constants? **Add** the expectations. Multiplying by a constant? **Multiply** the expectation by the constant.

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



Is $E[X]$ enough?

No! PMF is complete!

End Review

Where are We in CS109?

You are here



Counting
Theory



Core
Probability



Random
Variables



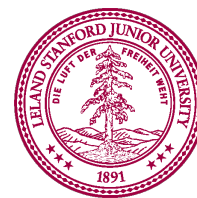
Probabilistic
Models



Uncertainty
Theory



Machine
Learning

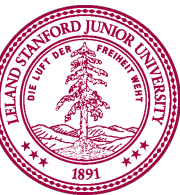


Classics



Coins are Everywhere...

1. **n independent trials** of the same experiment
(eg flipping a coin)
2. Each trial has a **probability of p** , of being a success (eg a heads)
3. What is the probability of **exactly k successes?**
(eg k heads)

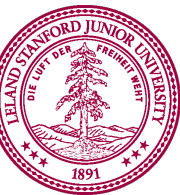


Many Random Variables Follow this Pattern

Examples

- # of heads in n coin flips
- # of 1's in randomly generated in length n bit string
- # of disk drives crashed in 1000 computer cluster
- # of people who vote for a candidate
- # of jury members selected from a demographic

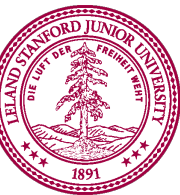
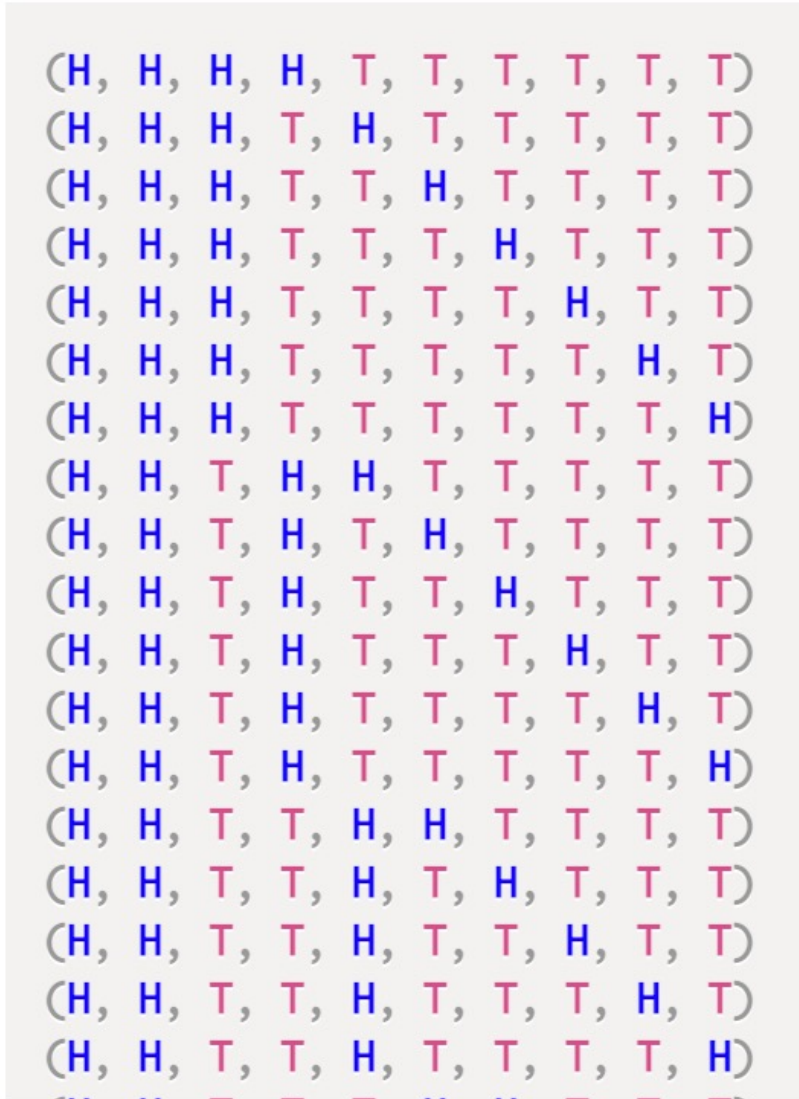
Note: All of these are random variables, and they have the same generative story



Exactly k heads in n coin flips

Probability of exactly k heads:

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



Let's Call it the Binomial



Here yee. This type of random variable is so common it needs a name so that I can talk about it generally.

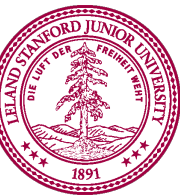
I call it: the Binomial Random Variable. Huzzah.

Jacob Bernoulli

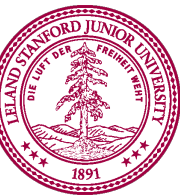
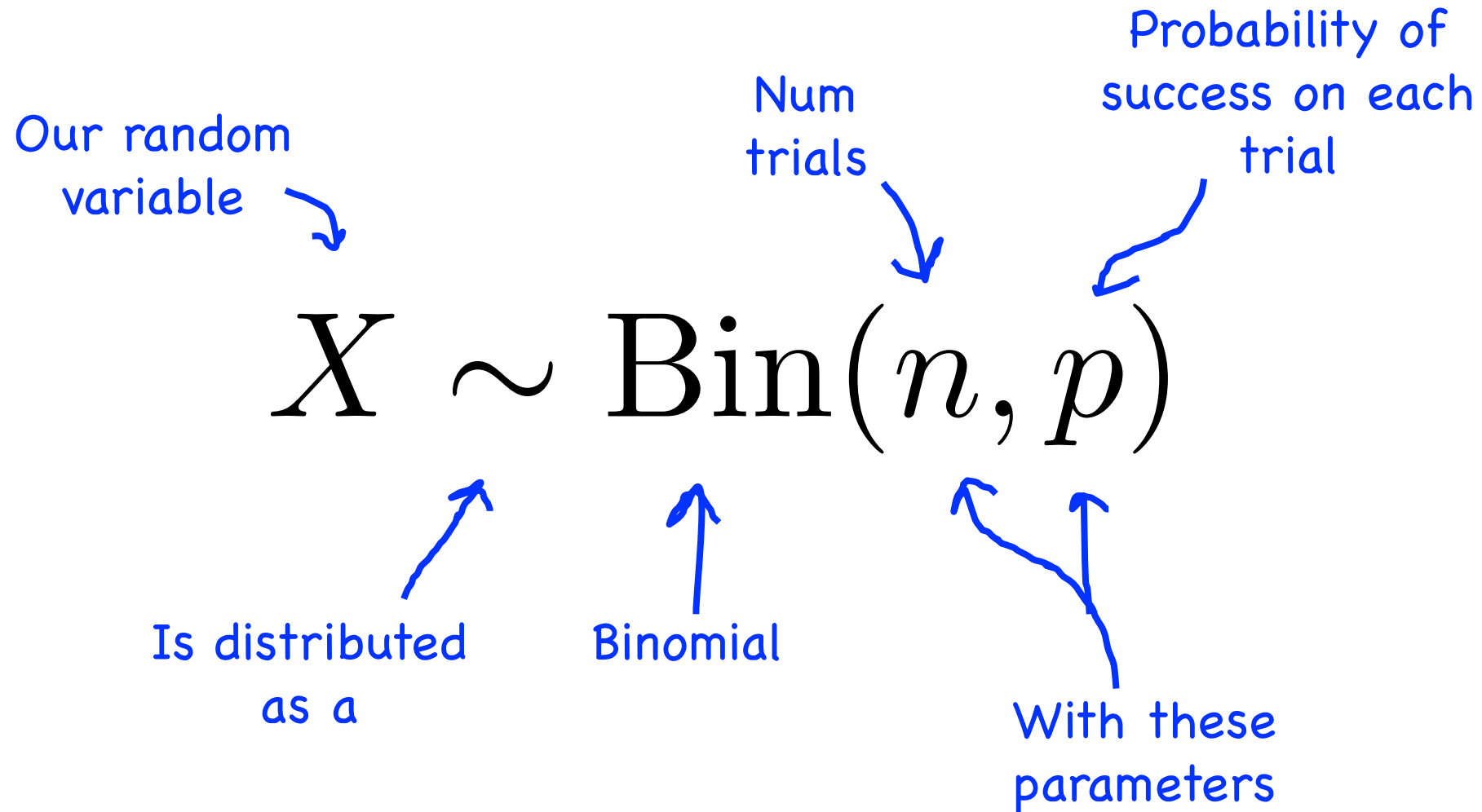
Jacob Bernoulli (1654-1705), also known as “James”, was a Swiss mathematician



One of many mathematicians in Bernoulli family
The Bernoulli Random Variable is named for him
He is my *academic* great¹²-grandfather
Celebrity look alike: Ice Cube



Declare a Random Variable to be Binomial



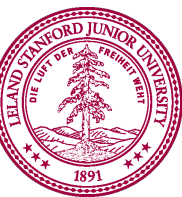
Automatically Know the PMF

Probability Mass Function
for a Binomial

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

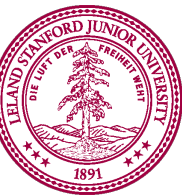
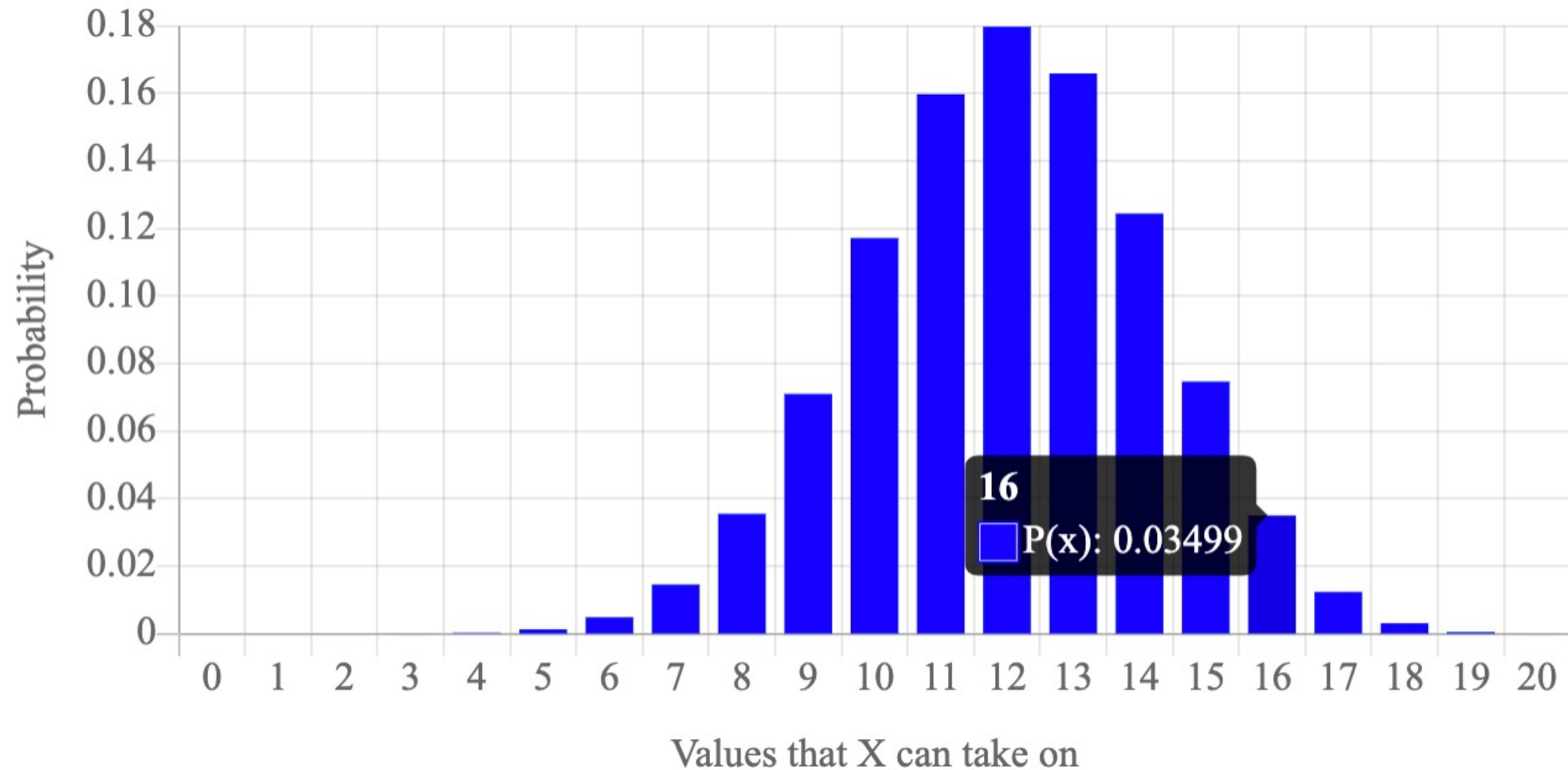
↑
Probability that our
variable takes on the
value k

↑
* This is also called
the binomial term



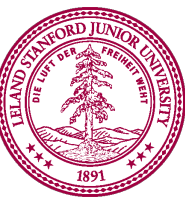
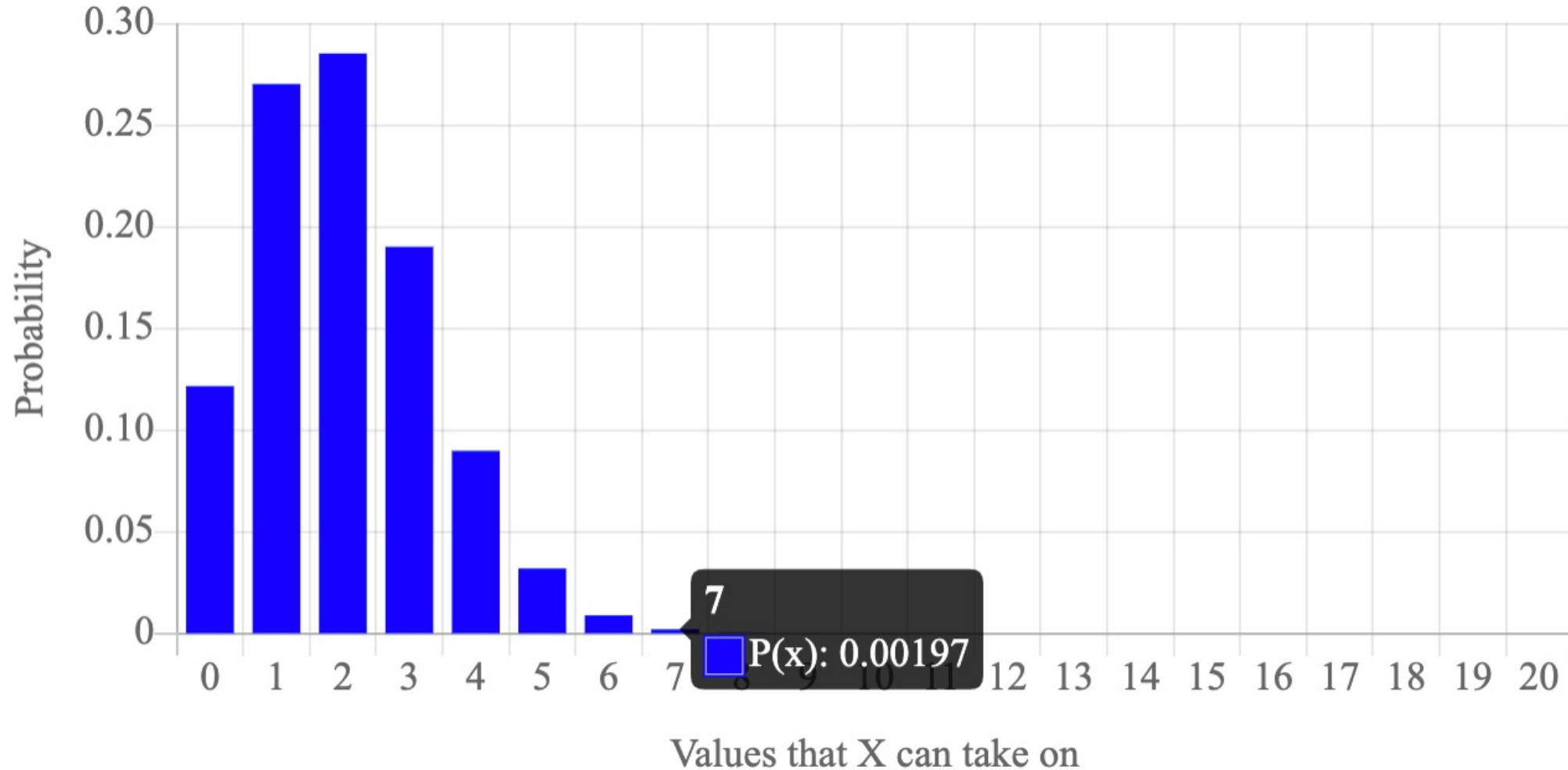
The PMF as a Graph: $X \sim \text{Bin}(n = 20, p = 0.6)$

Parameter n : Parameter p :



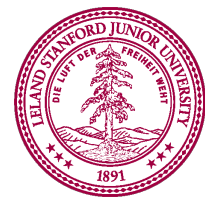
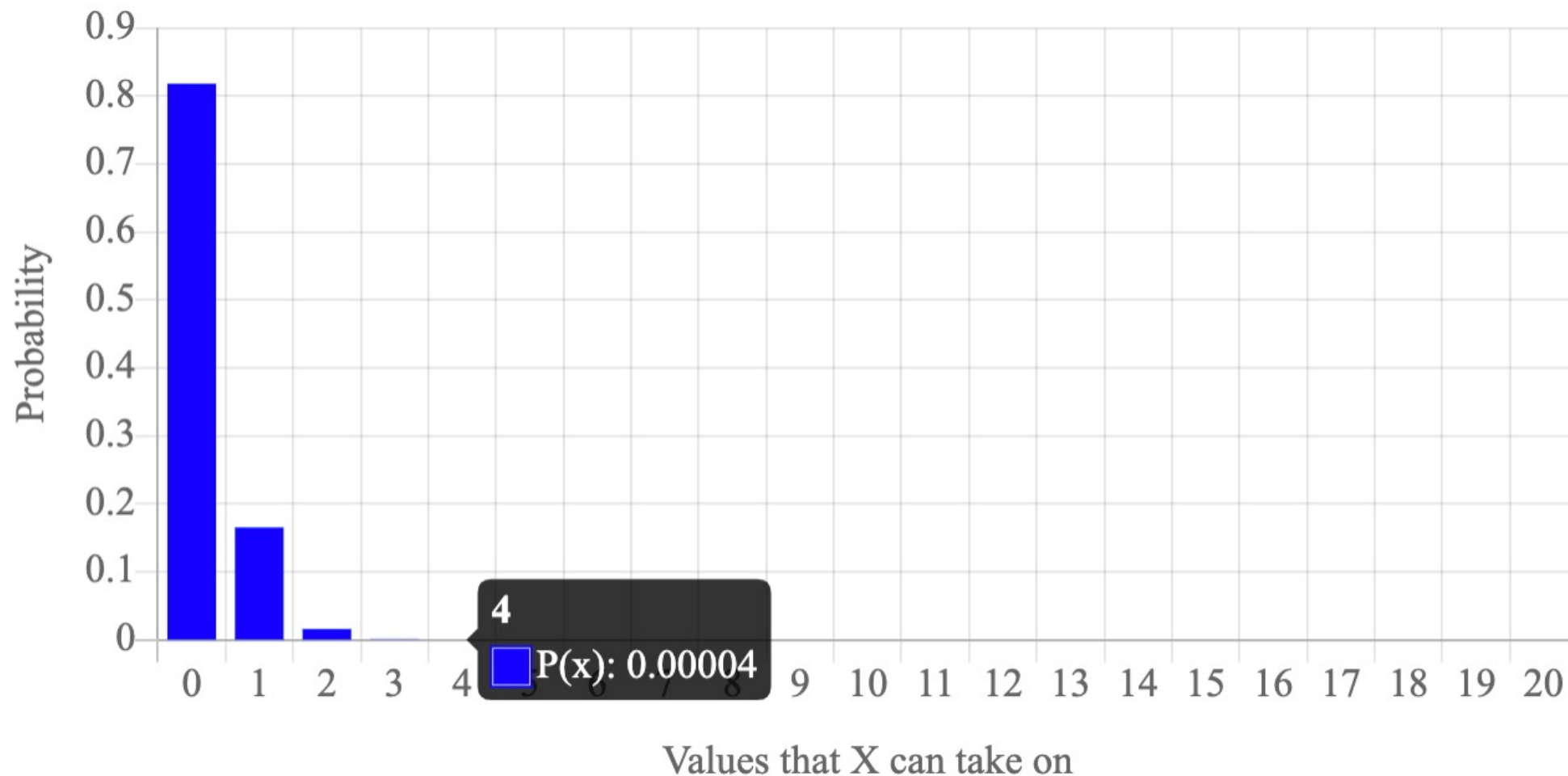
The PMF as a Graph: $X \sim \text{Bin}(n = 20, p = 0.1)$

Parameter n : Parameter p :



The PMF as a Graph: $X \sim \text{Bin}(n = 20, p = 0.01)$

Parameter n : Parameter p :



Coins, now with Binomial.

Three fair (“heads” with $p = 0.5$) coins are flipped

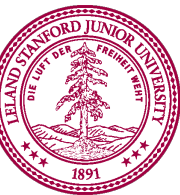
- X is number of heads
- $X \sim \text{Bin}(n = 3, p = 0.5)$

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

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

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

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



How Many Adds Clicked?



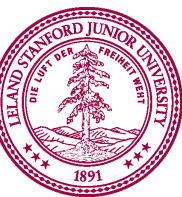
1000 ads served, each clicked with $p = 0.01$, otherwise ignored. What is the probability of 10 clicks?

H: number of clicks

$$\mathbf{H} \sim \text{Bin}(n = 1000, p = 0.01)$$

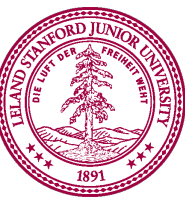
$$\mathbf{P}(\mathbf{H} = k) = \binom{1000}{k} (0.01)^k (0.99)^{1000-k}$$

$$\mathbf{P}(\mathbf{H} = 10) = \binom{1000}{10} (0.01)^{10} (0.99)^{990} \approx 0.125$$



How Many Adds Clicked? Redux

1000 ads served, each clicked with $p = 0.01$, otherwise ignored. What is the probability of 20 clicks?



How Many Adds Clicked? Redux

1000 ads served, each clicked with $p = 0.01$, otherwise ignored. What is the probability of 20 clicks?

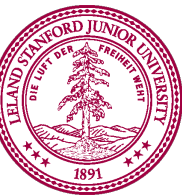
$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

```
[>>> from scipy import stats  
[>>> stats.binom.pmf(10, 1000, 0.01)  
0.1257402111262075  
[>>> stats.binom.pmf(20, 1000, 0.01)  
0.0017918782400182195
```

k

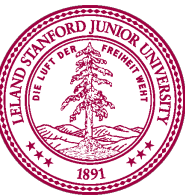
n

p




How Many Adds Clicked? Redux


1000 ads served, each clicked with $p = 0.01$, otherwise ignored. What is the probability of 5 clicks?



How Many Adds Clicked? Redux


1000 ads served, each clicked with $p = 0.01$, otherwise ignored. What is the probability of 5 clicks?

 Answer Editor

Numeric Answer: 0.03745311160828:  Check Answer

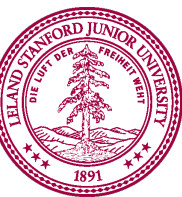
Code:

```
1 from scipy import stats
2
3 r = stats.binom.pmf(5, 1000, 0.01)
4 print(r)
```

 Run

Console

```
0.03745311160828357
```



How Many Adds Clicked? Redux

1000 ads served, each clicked with $p = 0.01$, otherwise ignored. What is the probability of x clicks?

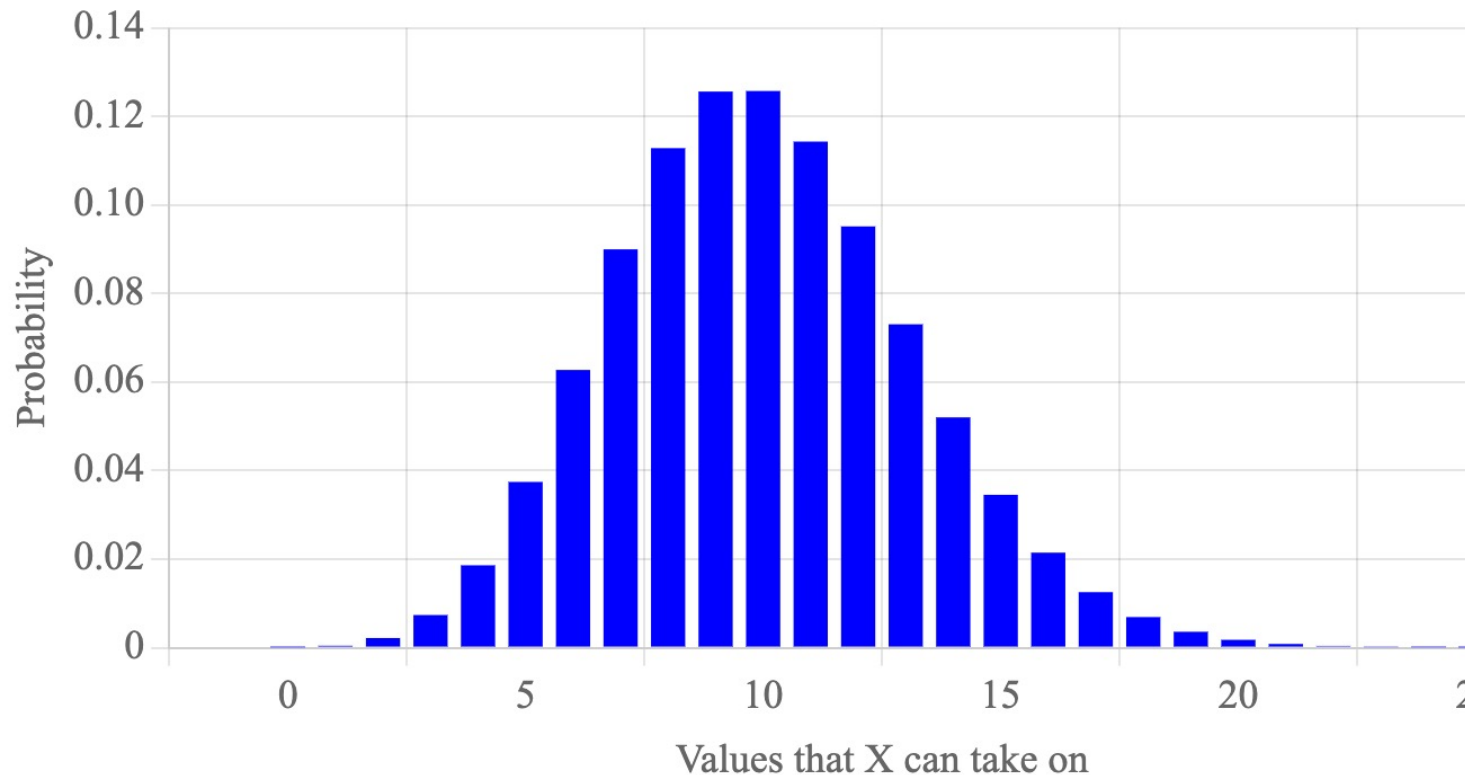
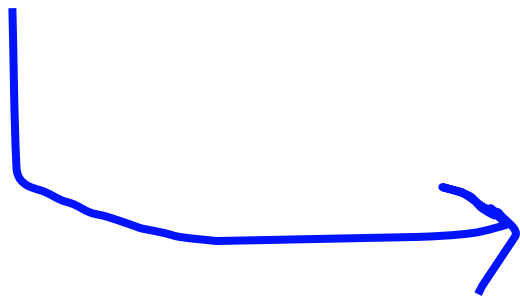
PMF graph:

Parameter n :

1000

Parameter p :

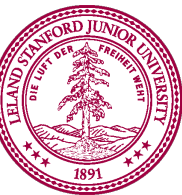
0.01



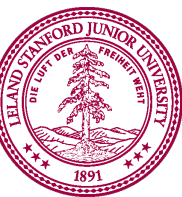
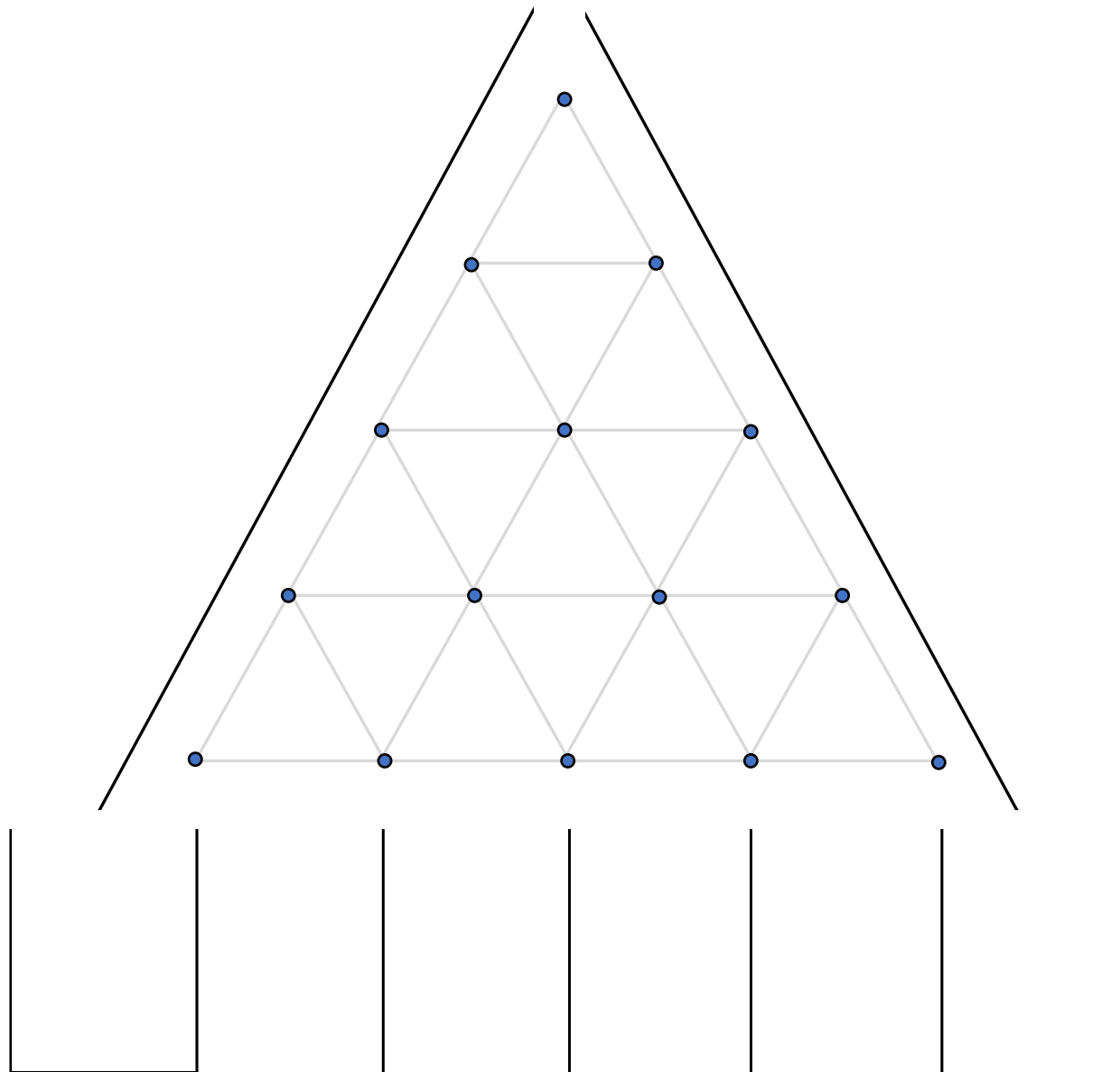
How Many Servers Crash?



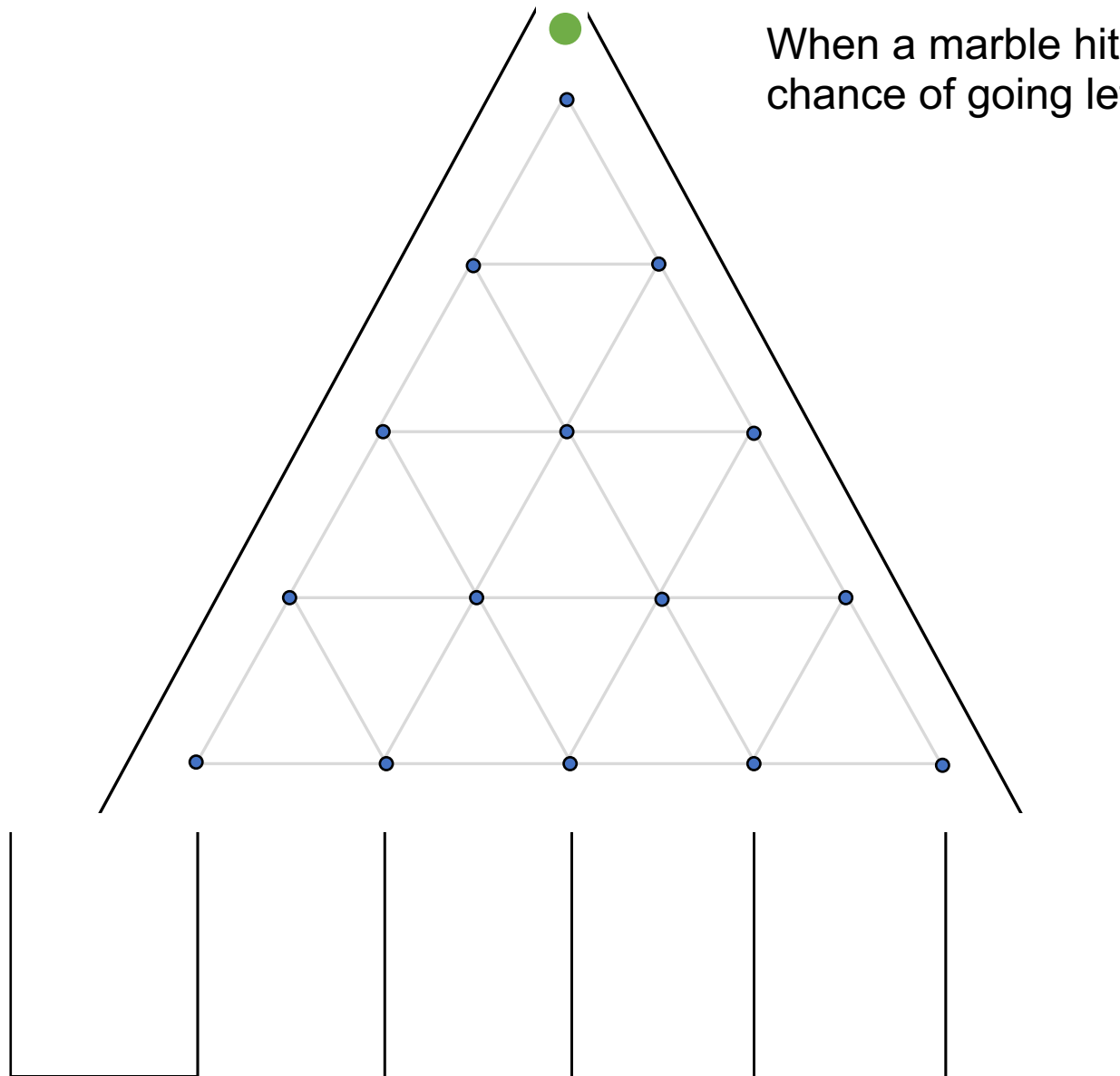
7 runs of program, each run crashes with probability 0.3.
What is the chance of exactly 3 crashes?



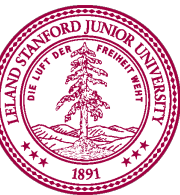
Galton Board



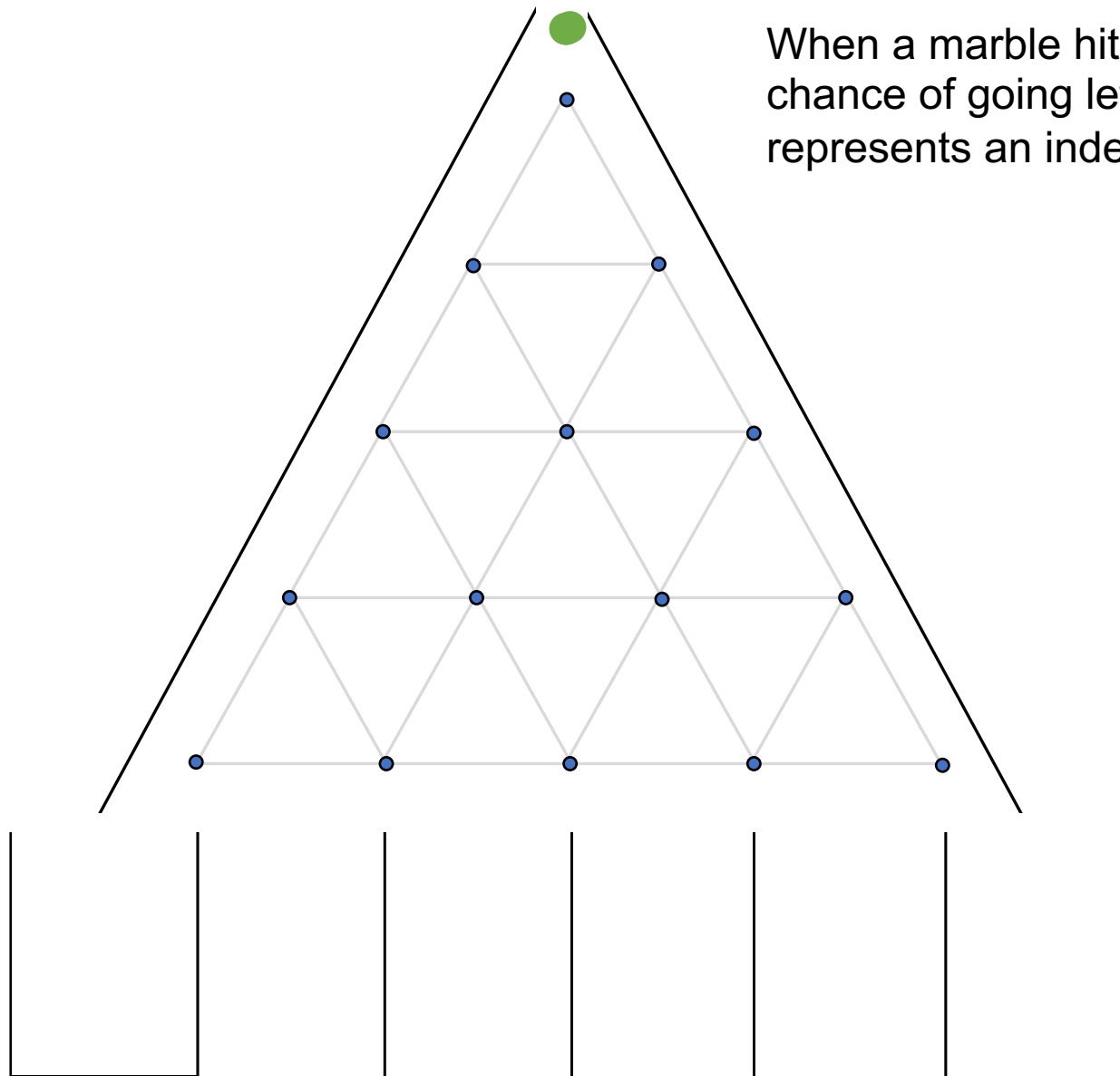
Galton Board



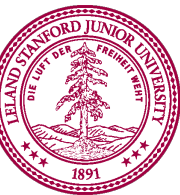
When a marble hits a pin, it has equal chance of going left or right.



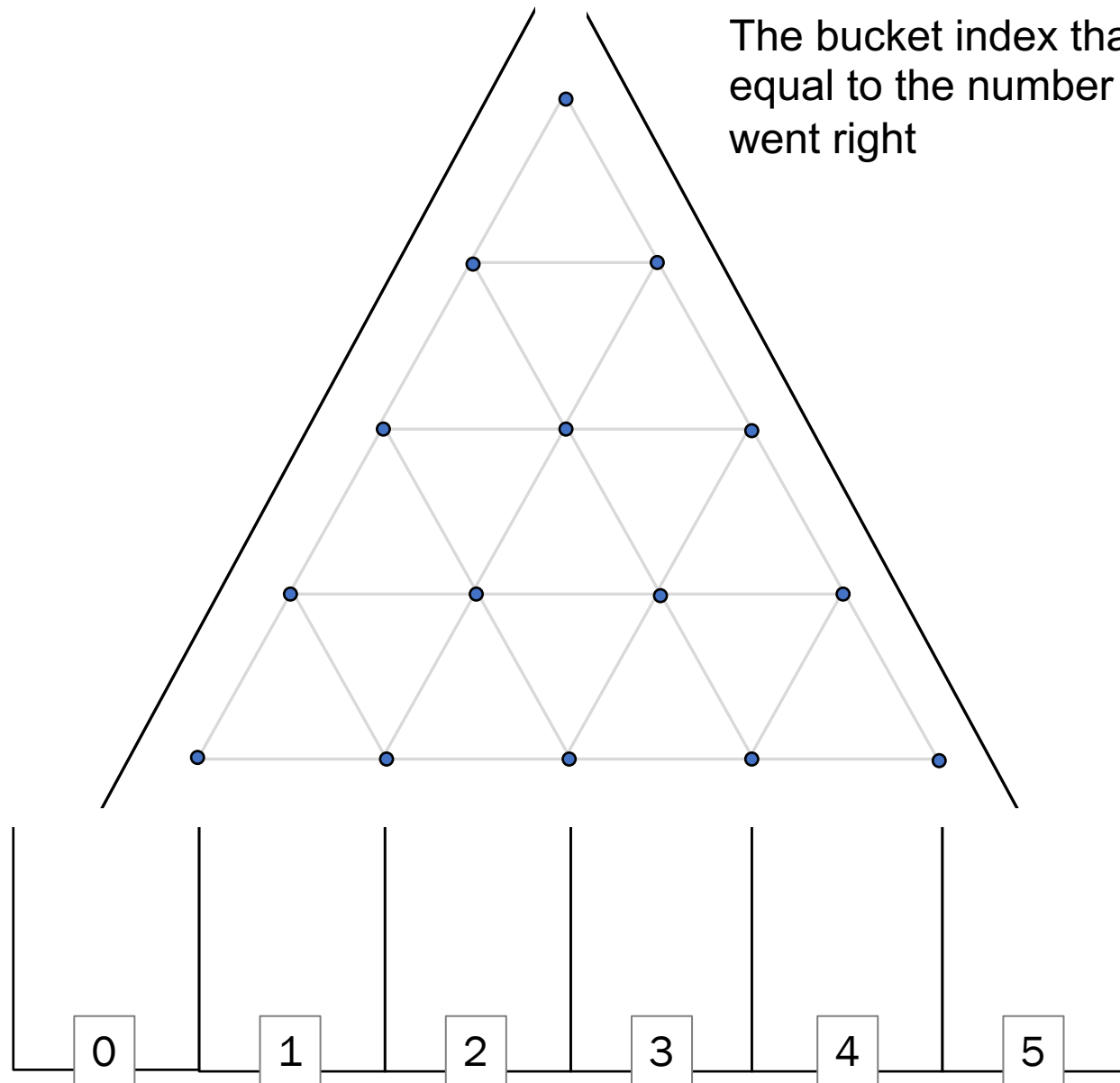
Galton Board



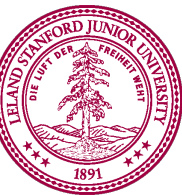
When a marble hits a pin, it has equal chance of going left or right. Each pin represents an independent event.



Galton Board

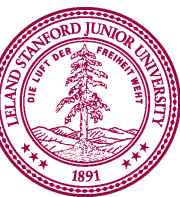
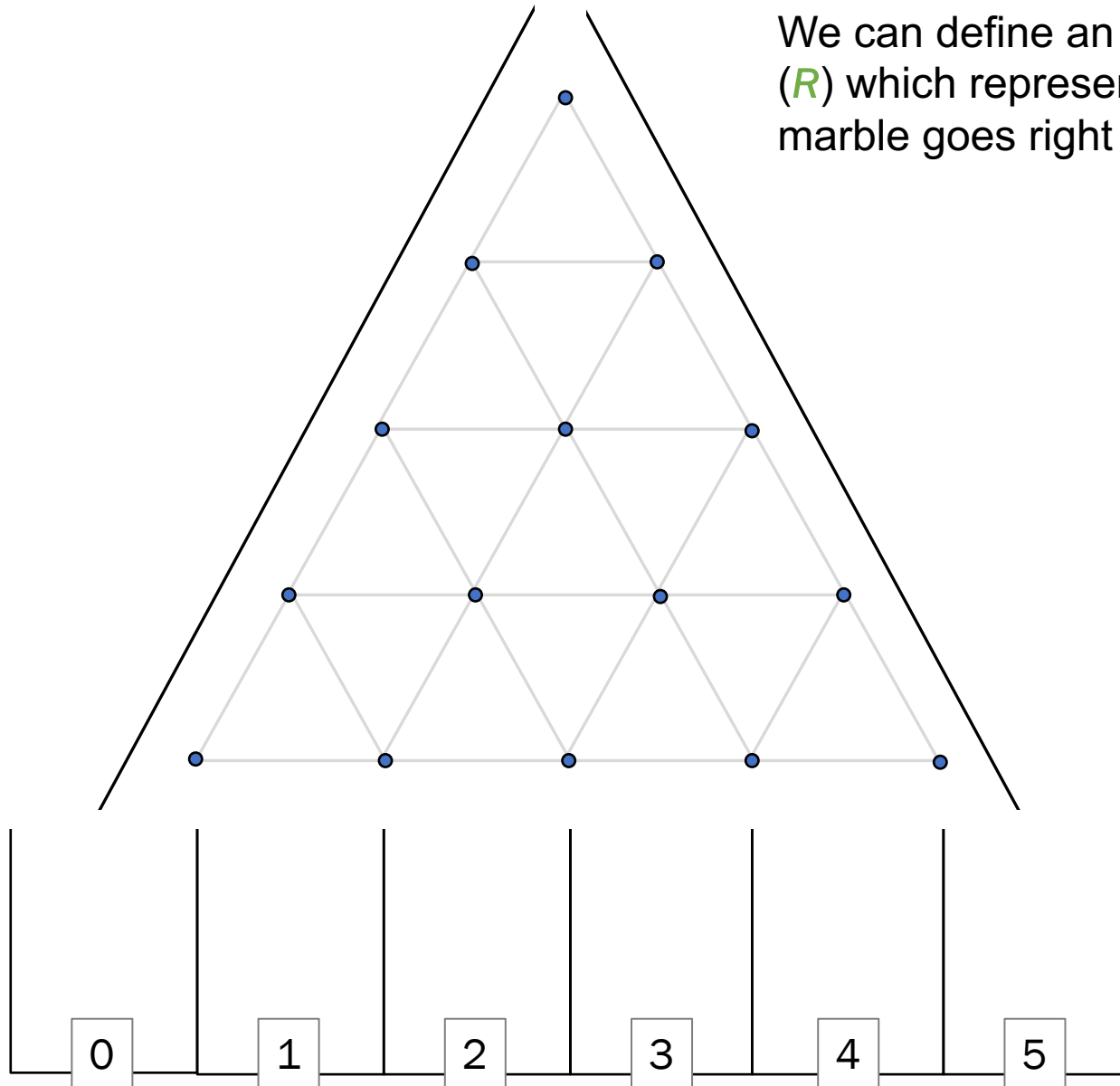


The bucket index that a marble lands in is equal to the number of times the marble went right



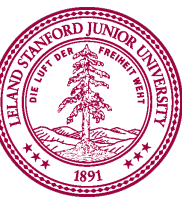
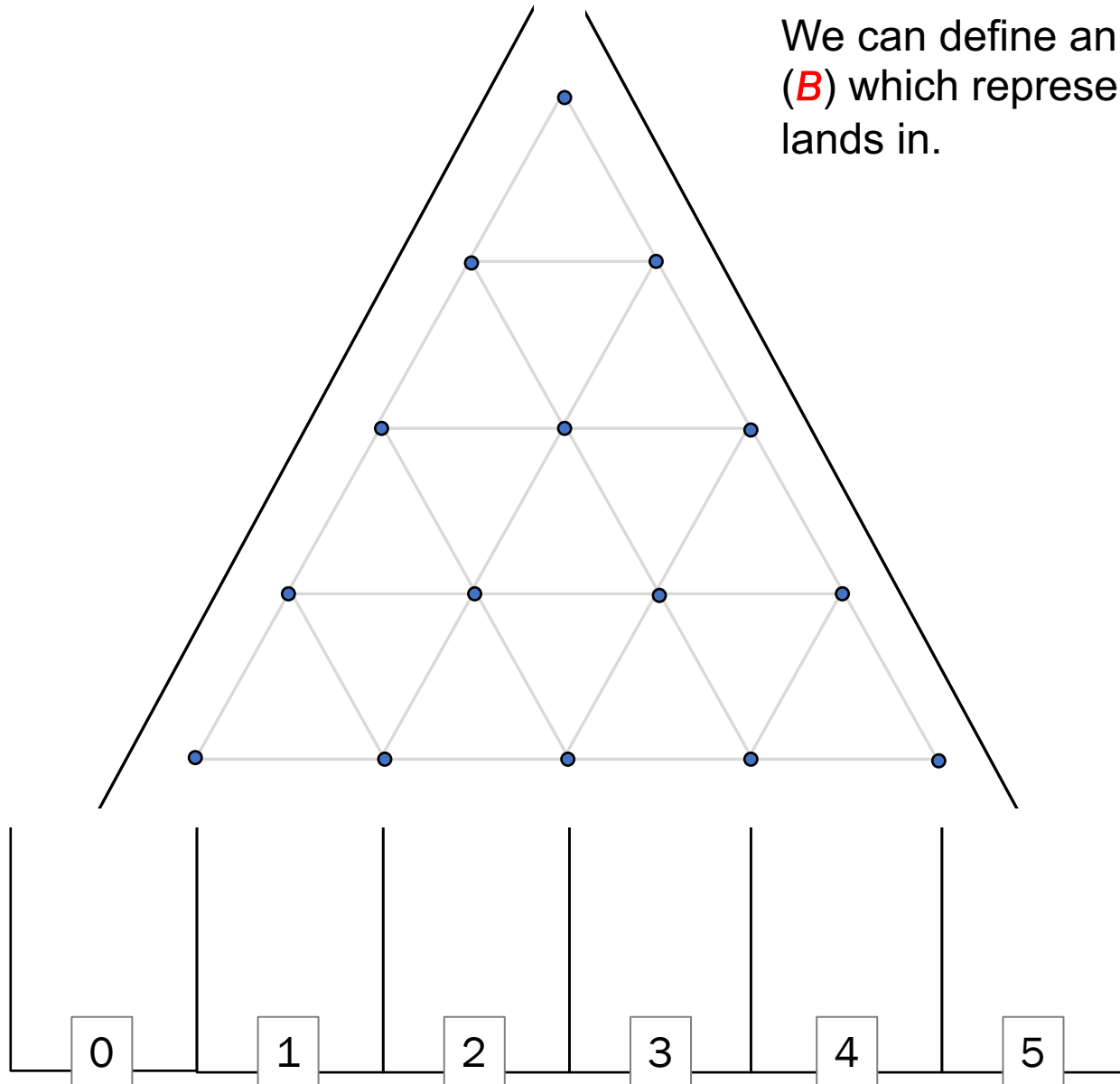
Galton Board

We can define an indicator random variable (R) which represents whether a particular marble goes right as a Bernoulli $R \sim \text{Ber}(0.5)$



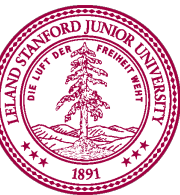
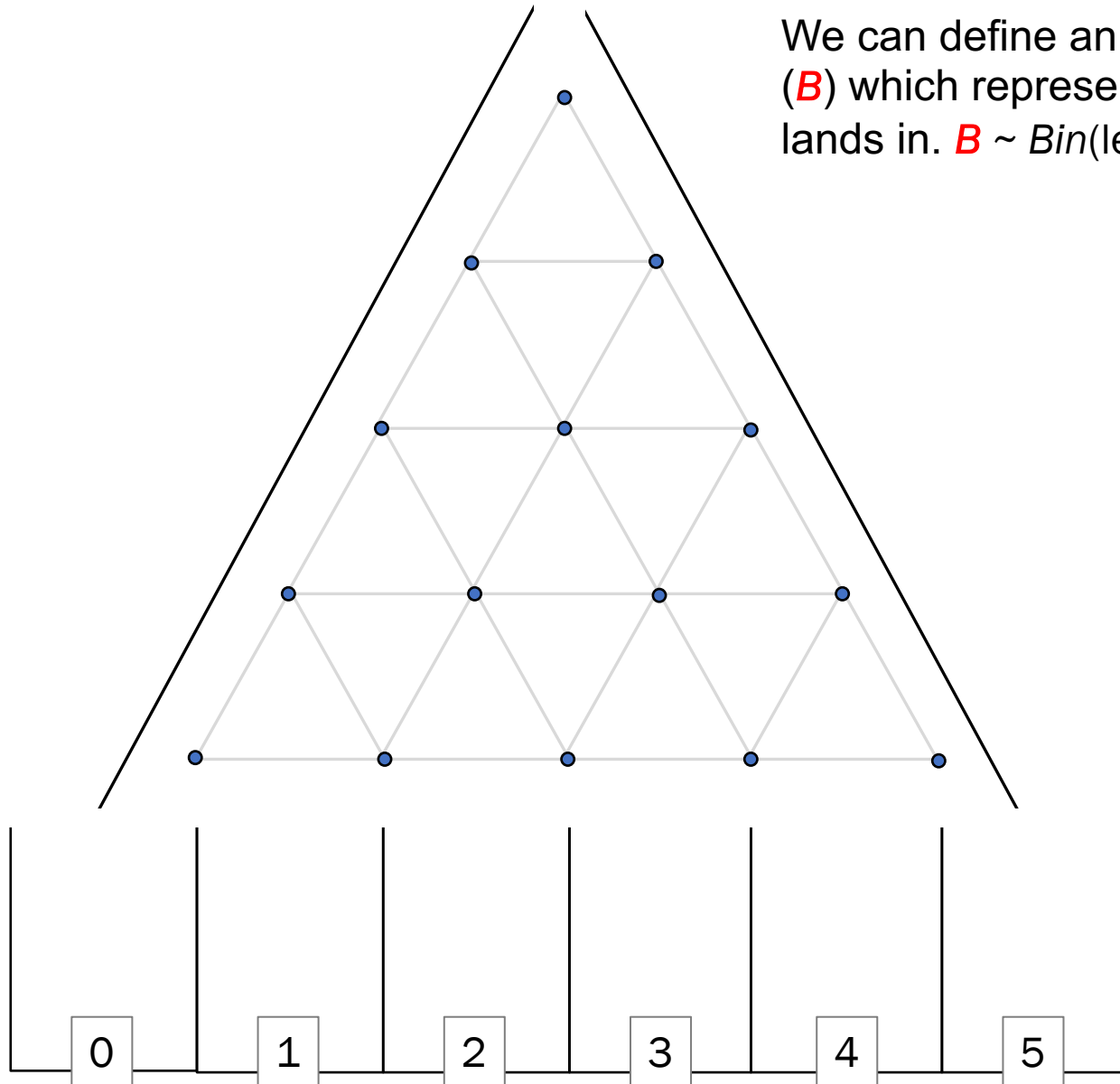
Galton Board

We can define an indicator random variable (B) which represents what bucket a marble lands in.



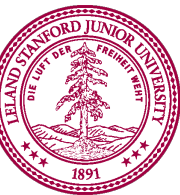
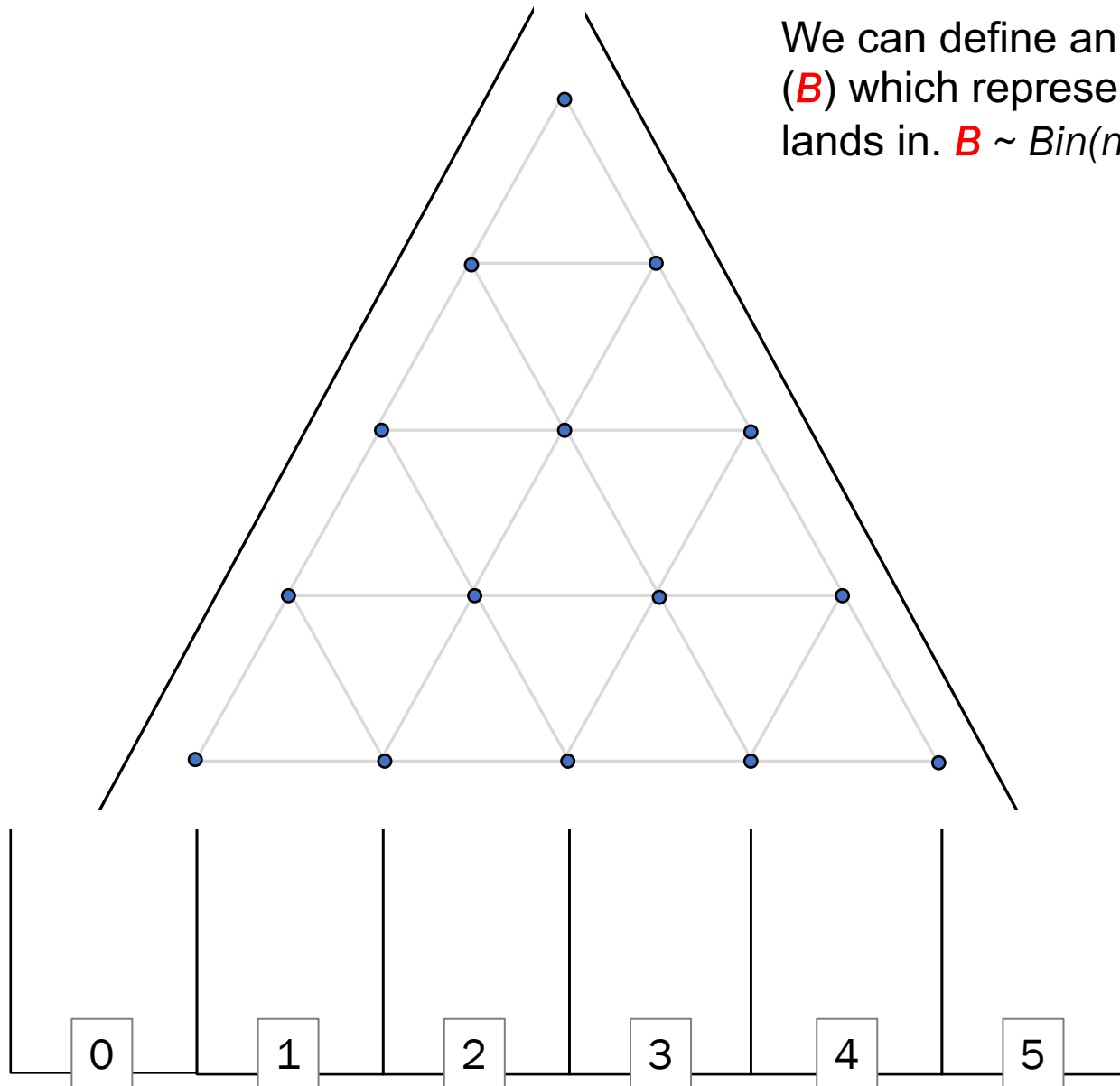
Galton Board

We can define an indicator random variable (B) which represents what bucket a marble lands in. $B \sim \text{Bin}(\text{levels}, 0.5)$

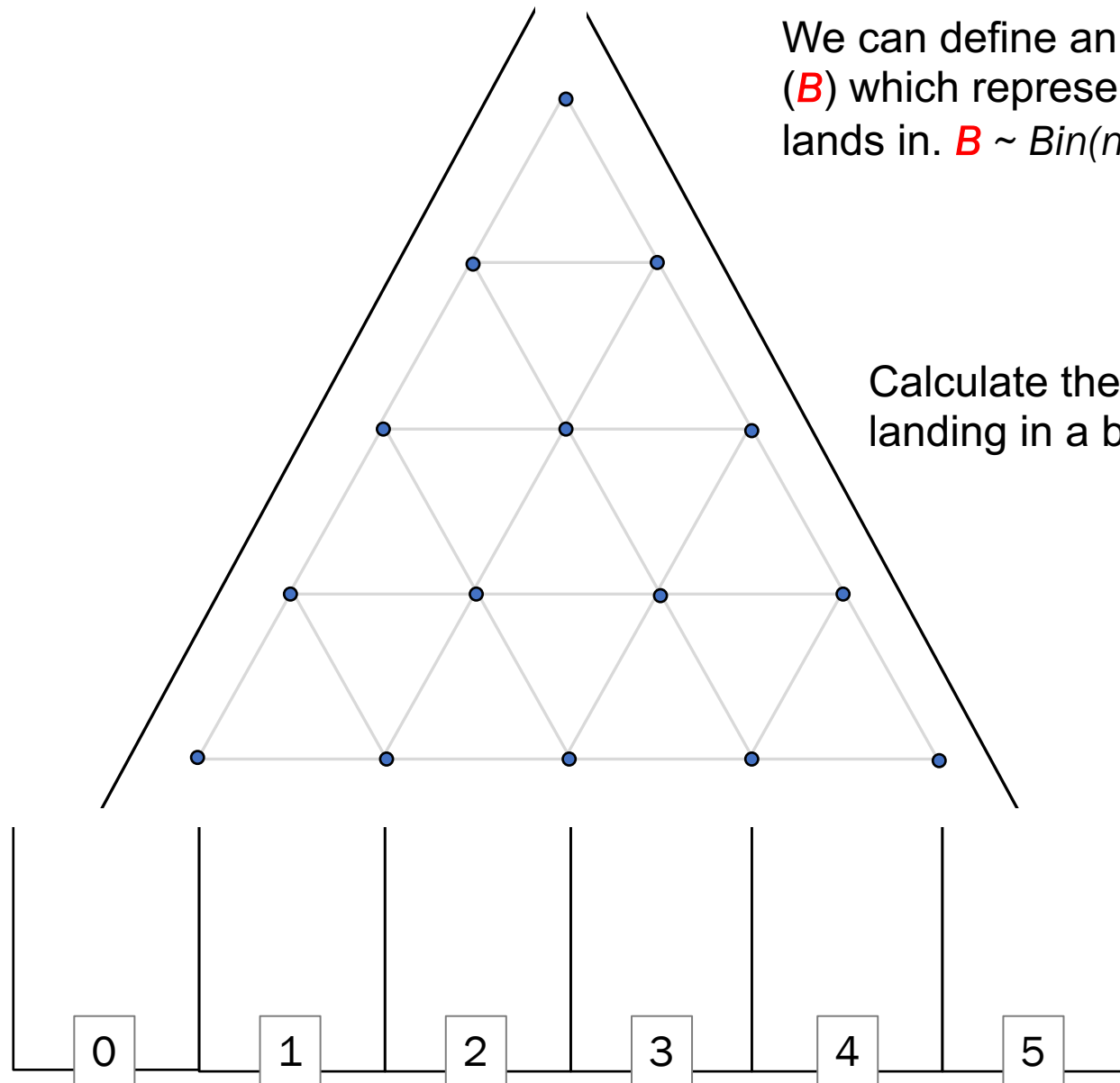


Galton Board

We can define an indicator random variable (B) which represents what bucket a marble lands in. $B \sim \text{Bin}(n = 5, p = 0.5)$

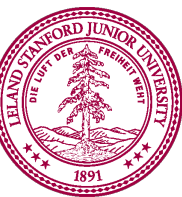


Galton Board

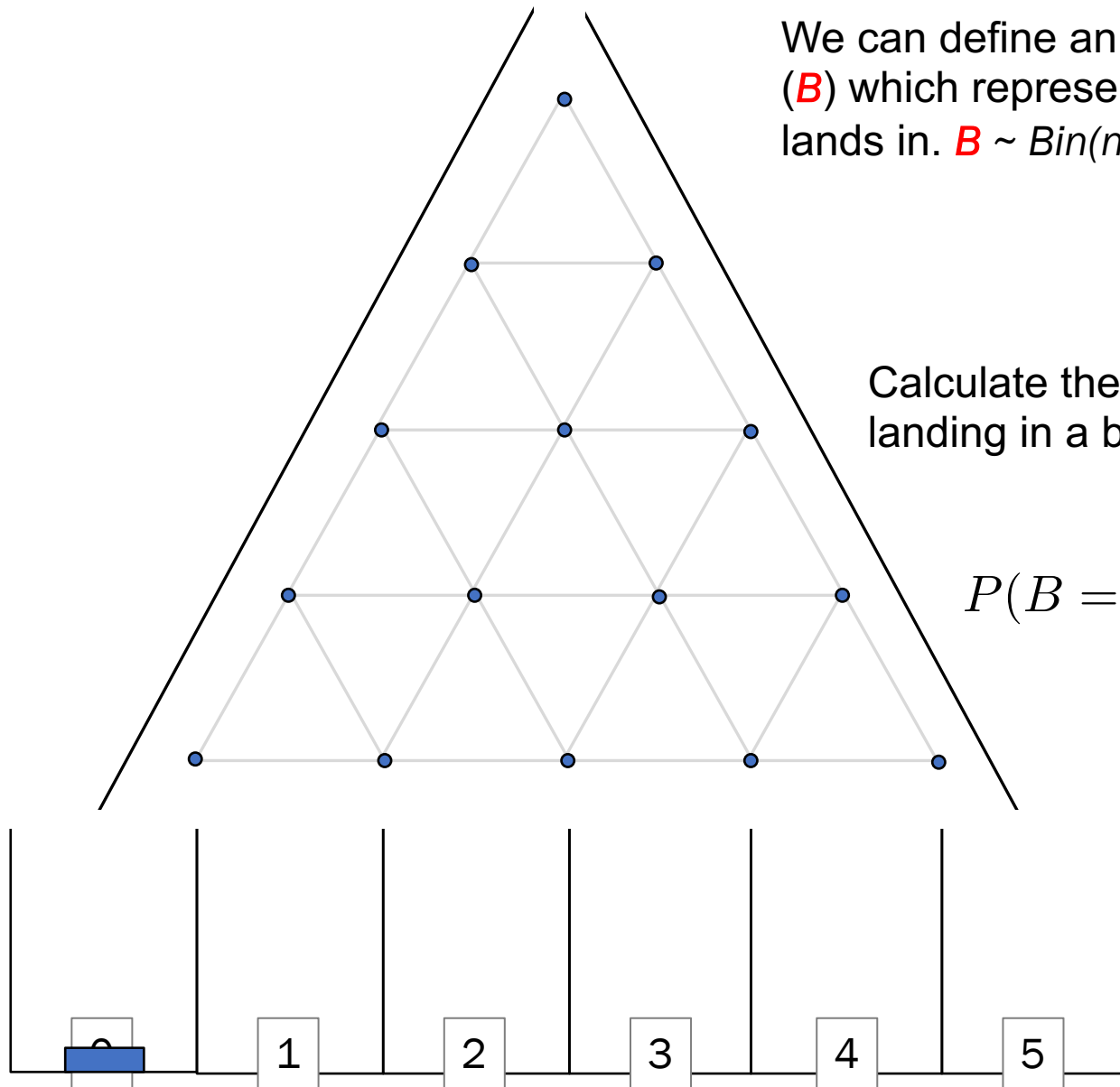


We can define an indicator random variable (B) which represents what bucket a marble lands in. $B \sim \text{Bin}(n = 5, p = 0.5)$

Calculate the probability of a marble landing in a bucket.



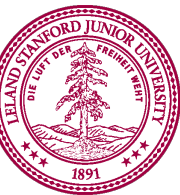
Galton Board



We can define an indicator random variable (B) which represents what bucket a marble lands in. $B \sim \text{Bin}(n = 5, p = 0.5)$

Calculate the probability of a marble landing in a bucket.

$$P(B = 0) = \binom{5}{0} \frac{1}{2}^5 \approx 0.03$$

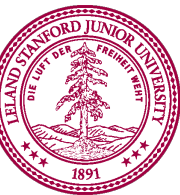
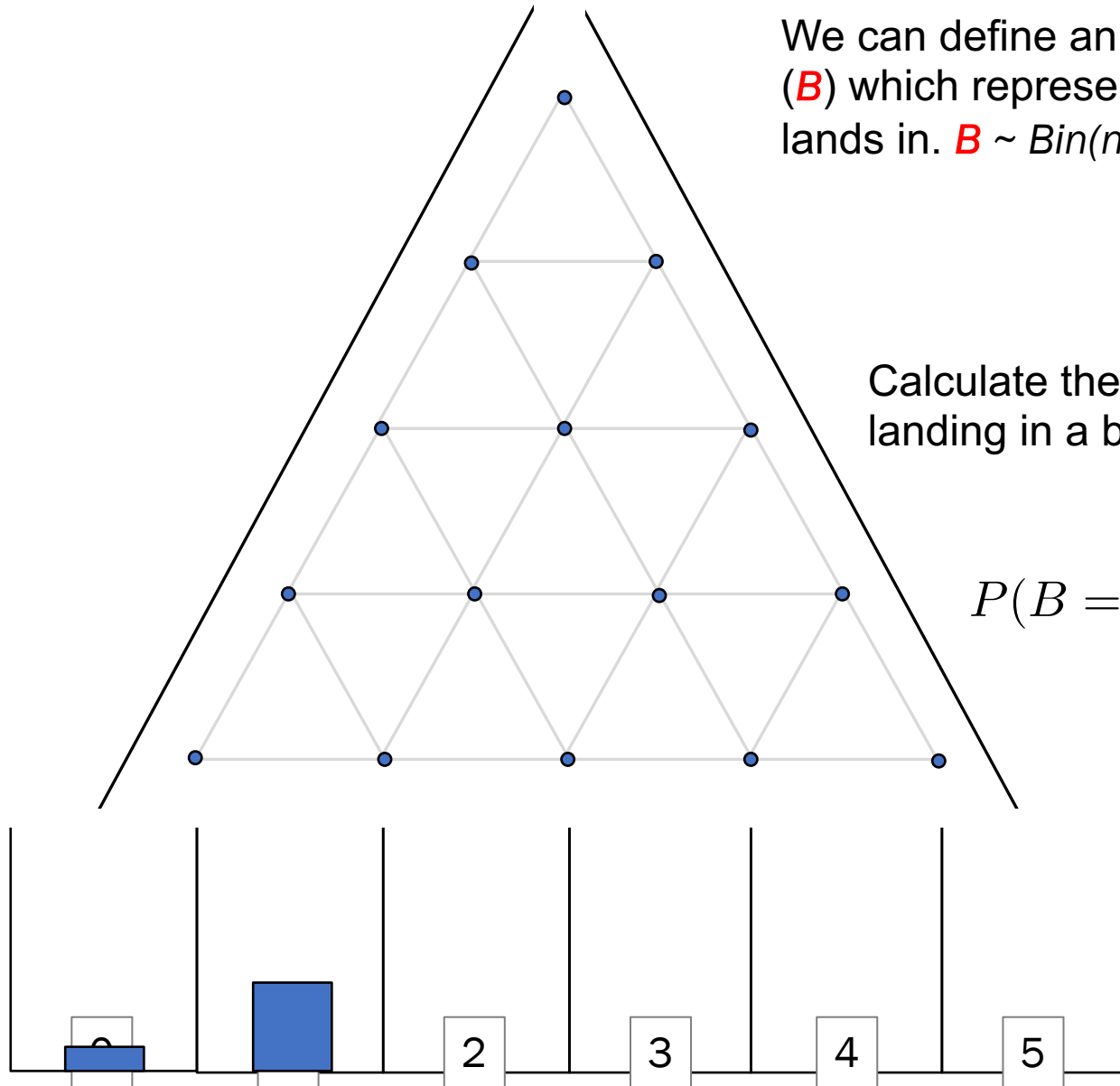


Galton Board

We can define an indicator random variable (B) which represents what bucket a marble lands in. $B \sim \text{Bin}(n = 5, p = 0.5)$

Calculate the probability of a marble landing in a bucket.

$$P(B = 1) = \binom{5}{1} \frac{1}{2}^5 \approx 0.16$$

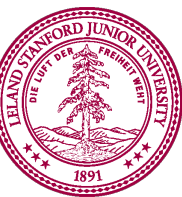
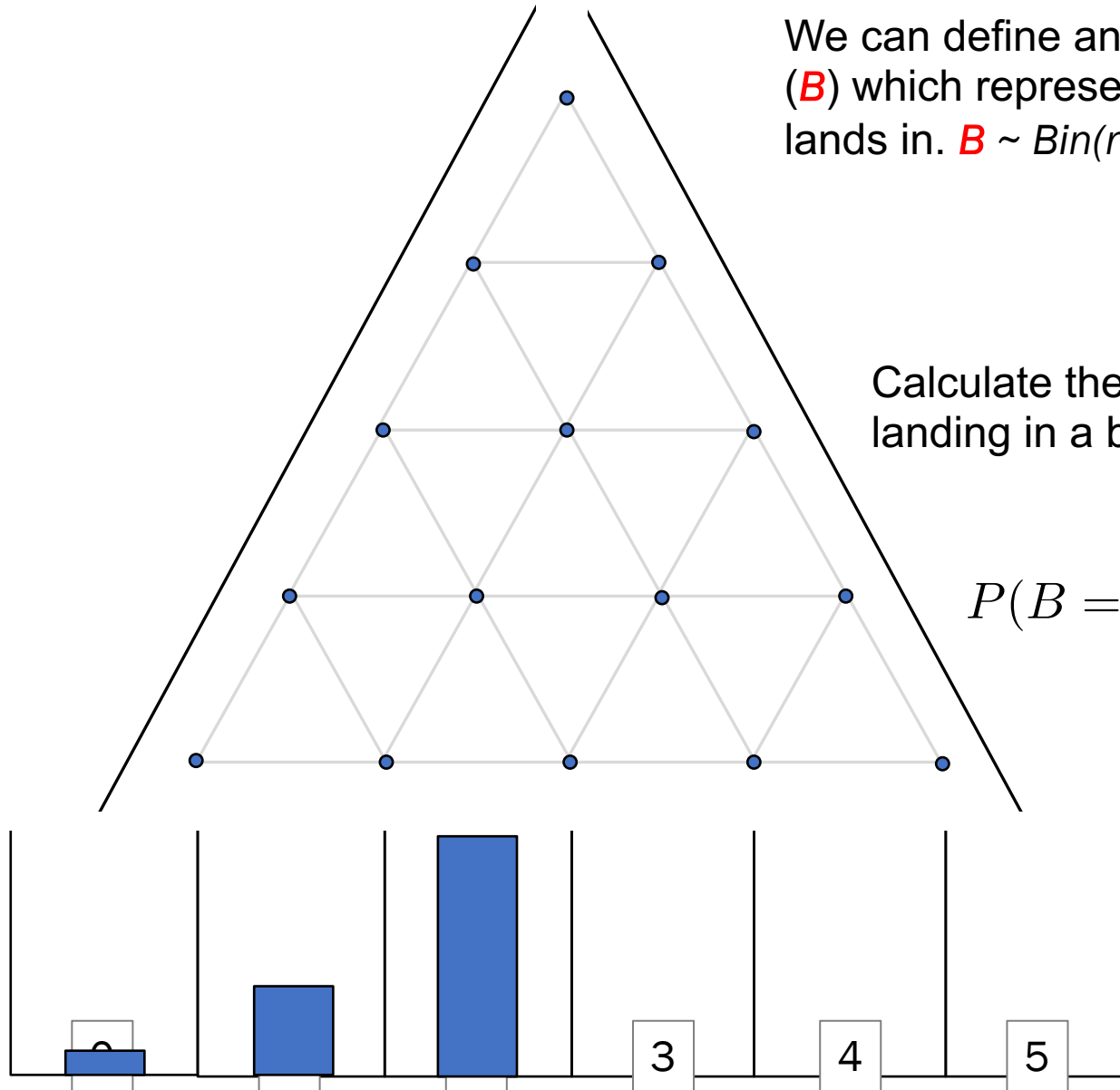


Galton Board

We can define an indicator random variable (B) which represents what bucket a marble lands in. $B \sim \text{Bin}(n = 5, p = 0.5)$

Calculate the probability of a marble landing in a bucket.

$$P(B = 2) = \binom{5}{2} \frac{1}{2^5} \approx 0.31$$

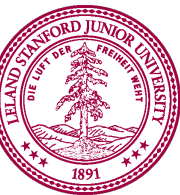
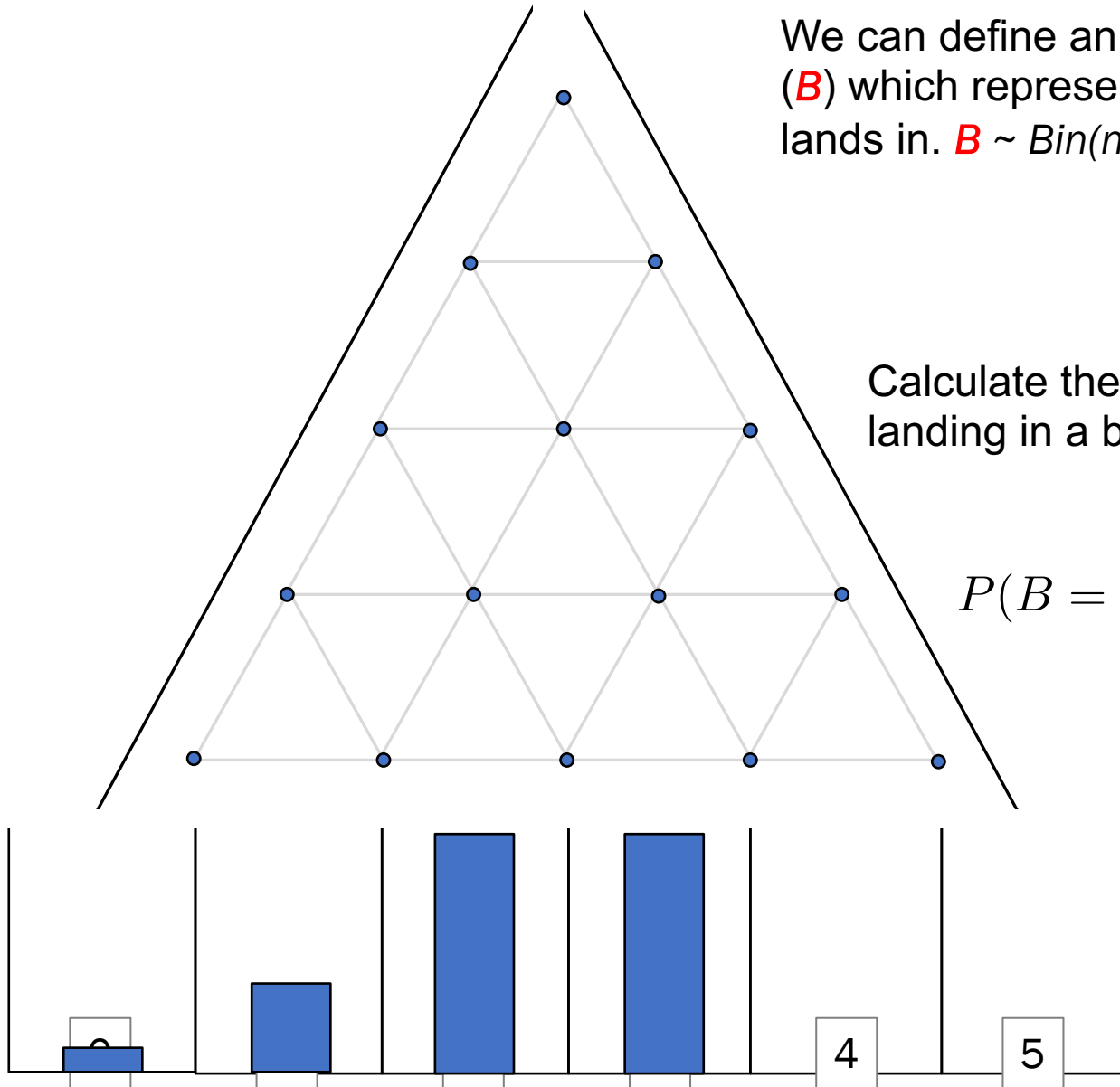


Galton Board

We can define an indicator random variable (B) which represents what bucket a marble lands in. $B \sim \text{Bin}(n = 5, p = 0.5)$

Calculate the probability of a marble landing in a bucket.

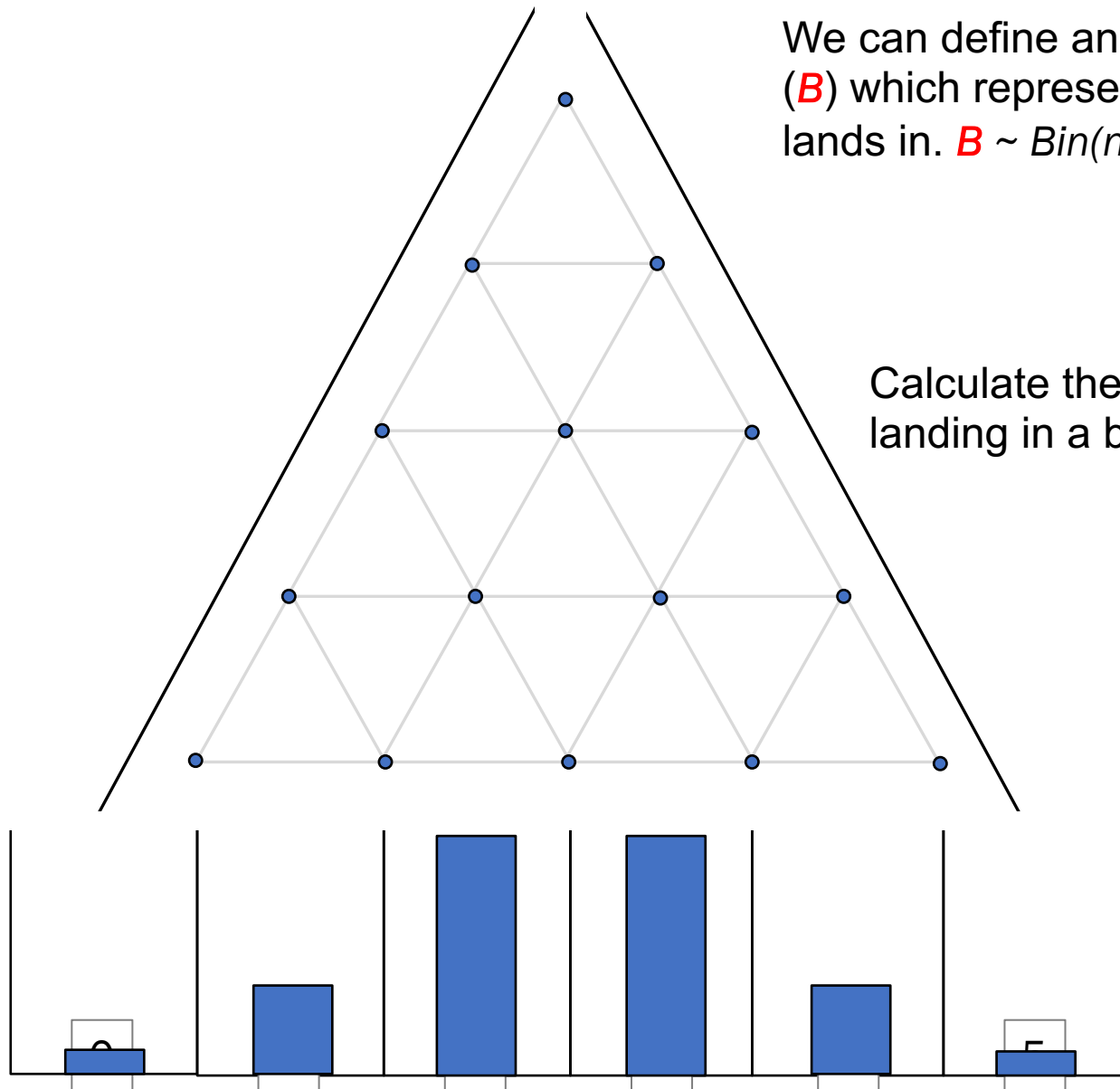
$$P(B = 3) = \binom{5}{2} \frac{1}{2^5} \approx 0.31$$



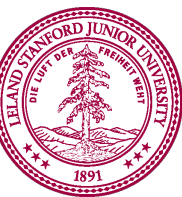
Galton Board

We can define an indicator random variable (B) which represents what bucket a marble lands in. $B \sim \text{Bin}(n = 5, p = 0.5)$

Calculate the probability of a marble landing in a bucket.



PMF





FROM CHAOS TO ORDER



GREEN

23

GOLDEN STATE
30
WARRIORS

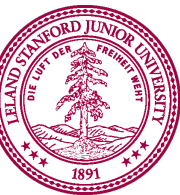
Rakuten

W

What is the probability of winning a 7 game series?

Warriors are going to play the Bucks in a best of 7 series during the 2022 NBA finals. What is the probability that the warriors win the series? Each game is **independent**. Each game, the warriors have a 0.55 probability of winning. Win series if you win at least 4 games.

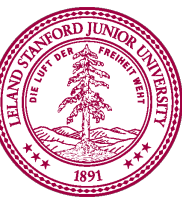
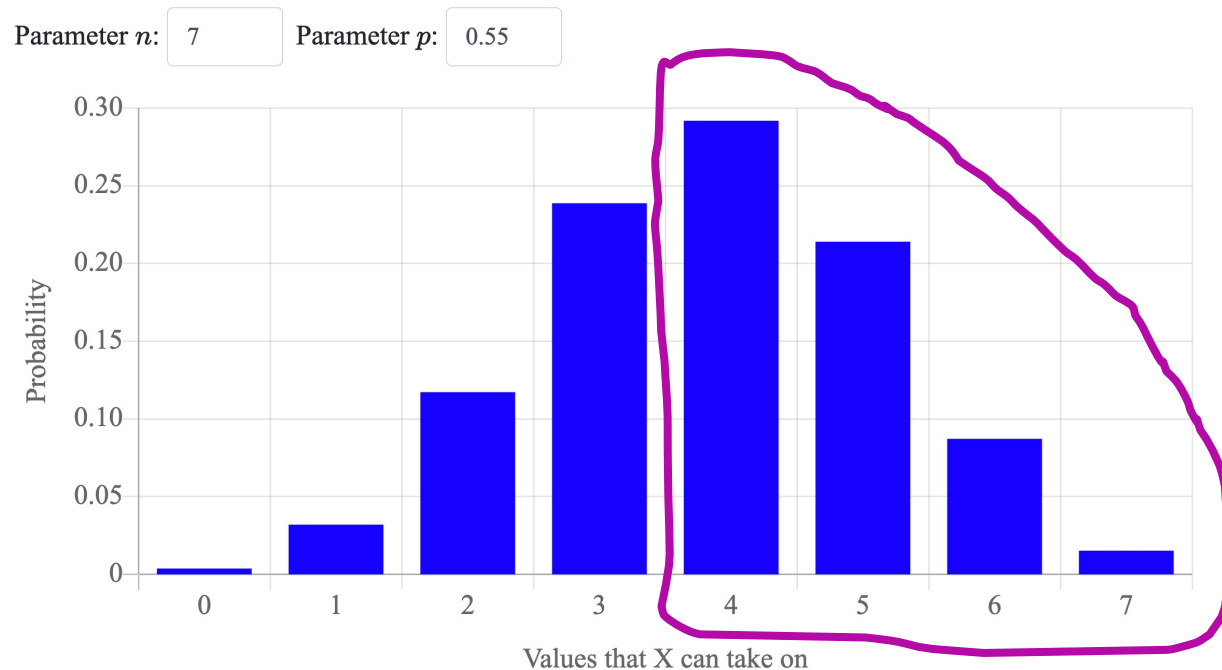
Let X be the number of games won. $X \sim \text{Bin}(n=7, p=0.55)$. $P(X > 3)$?



What is the probability of winning a 7 game series?

Warriors are going to play the Bucks in a best of 7 series during the 2022 NBA finals. What is the probability that the warriors win the series? Each game is **independent**. Each game, the warriors have a 0.55 probability of winning? Win series if you win at least 4 games.

Let X be the number of games won. $X \sim \text{Bin}(n=7, p=0.55)$. $P(X > 3)$?

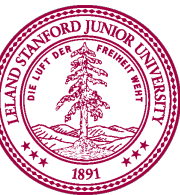


What is the probability of winning a 7 game series?

Warriors are going to play the Bucks in a best of 7 series during the 2022 NBA finals. What is the probability that the warriors win the series? Each game is **independent**. Each game, the warriors have a 0.55 probability of winning? Win series if you win at least 4 games.

Let X be the number of games won. $X \sim \text{Bin}(n=7, p=0.55)$. $P(X \geq 4)$?

$$\begin{aligned} P(X \geq 4) &= \sum_{i=4}^7 P(X = i) \\ &= \sum_{i=4}^7 \binom{7}{i} p^i (1-p)^{7-i} \\ &= \sum_{i=4}^7 \binom{7}{i} 0.55^i (0.45)^{7-i} \end{aligned}$$



Debugging Probability



Debugging Probability

How to calculate the probability of at least k successes in n independent trials?

- X is number of successes in n trials each with probability p

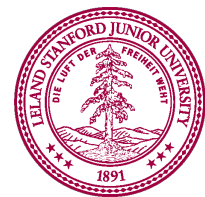
- $P(X \geq k) =$

Chose slots for success,
don't care about rest

ways to choose slots for
success

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

Probability that each is
success



Debugging Probability

How to calculate the probability of at least k successes in n independent trials?

- X is number of successes in n trials each with probability p
- $P(X \geq k) =$

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

Not mutually exclusive...

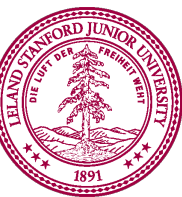
Chose slots for success, don't care about rest

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

ways to choose slots for success

Probability that each is success

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



You Get So Much For Free!

Binomial Random Variable

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

Description: Number of "successes" in n identical, independent experiments each with probability of success p .

Parameters: $n \in \{0, 1, \dots\}$, the number of experiments.
 $p \in [0, 1]$, the probability that a single experiment gives a "success".

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

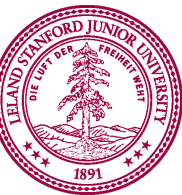
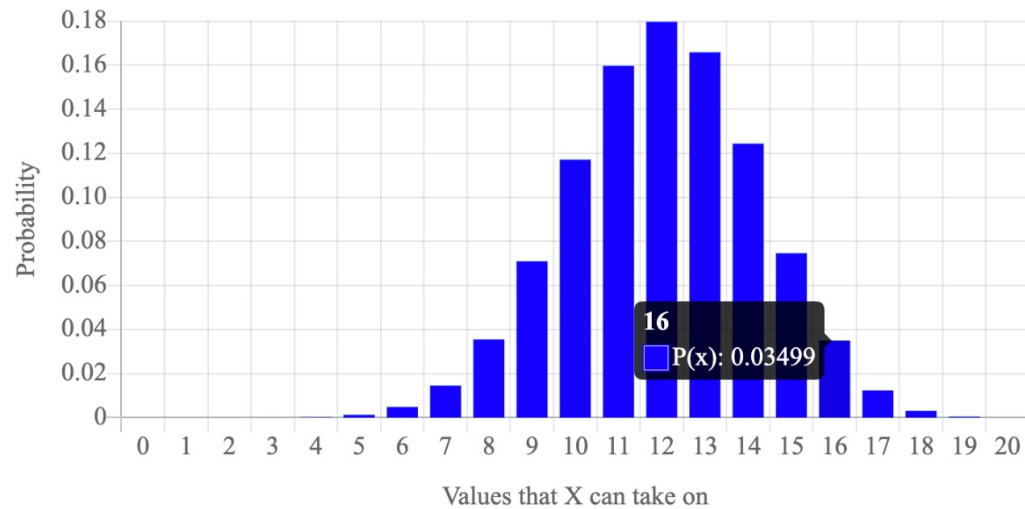
PMF equation: $\Pr(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}$

Expectation: $E[X] = n \cdot p$

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

PMF graph:

Parameter n : Parameter p :



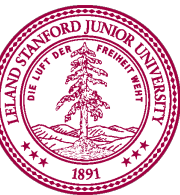
A solid blue vertical bar is positioned on the far left side of the image, extending from the top to the bottom.

Pedagogical Pause

Can Jacob Bernoulli Have a Variable Named After Him?



Here yee. I want to have a random variable named after myself. Huzzah.



Bernoulli Random Variable

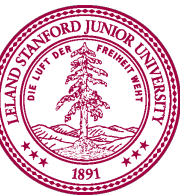
Experiment results in “Success” or “Failure”

- X is random **indicator** variable (1 = success, 0 = failure)
- $P(X = 1) = p(1) = p$ $P(X = 0) = p(0) = 1 - p$
- X is a **Bernoulli** Random Variable: $X \sim \text{Bern}(p)$
- $E[X] = p$

Examples

- coin flip
- random binary digit
- whether a disk drive crashed
- whether someone likes a netflix movie

Feel the Bern!



Does a Program Crash?



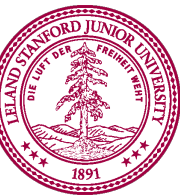
Run a program, crashes with prob. p , works with prob. $(1 - p)$

X : 1 if program crashes

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

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

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



Does a User Click an Ad?



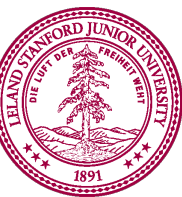
Serve an ad, clicked with prob. p , ignored with prob. $(1 - p)$

C : 1 if ad is clicked

$$P(C = 1) = p$$

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

$$\underline{C} \sim \text{Ber}(p)$$



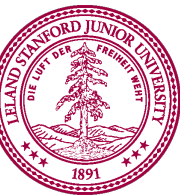
Bernoulli vs Binomial



Bernoulli is an indicator RV



Binomial is the sum of n
Bernoullis



We Can Now Calculate Expectation of Binomial

Let $X \sim \text{Bin}(n, p)$. Let Y_i be 1 if trial i was a success. $Y_i \sim \text{Bern}(p)$

$$\mathbf{E}[X] = \mathbf{E} \left[\sum_{i=1}^n Y_i \right]$$

$$= \sum_{i=1}^n \mathbf{E}[Y_i]$$

$$= \sum_{i=1}^n p$$

$$= n \cdot p$$

$$\text{Since } X = \sum_{i=1}^n Y_i$$

Expectation of sum

Expectation of Bernoulli

Sum n times



You Get So Much For Free!

Binomial Random Variable

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

Description: Number of "successes" in n identical, independent experiments each with probability of success p .

Parameters: $n \in \{0, 1, \dots\}$, the number of experiments.
 $p \in [0, 1]$, the probability that a single experiment gives a "success".

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

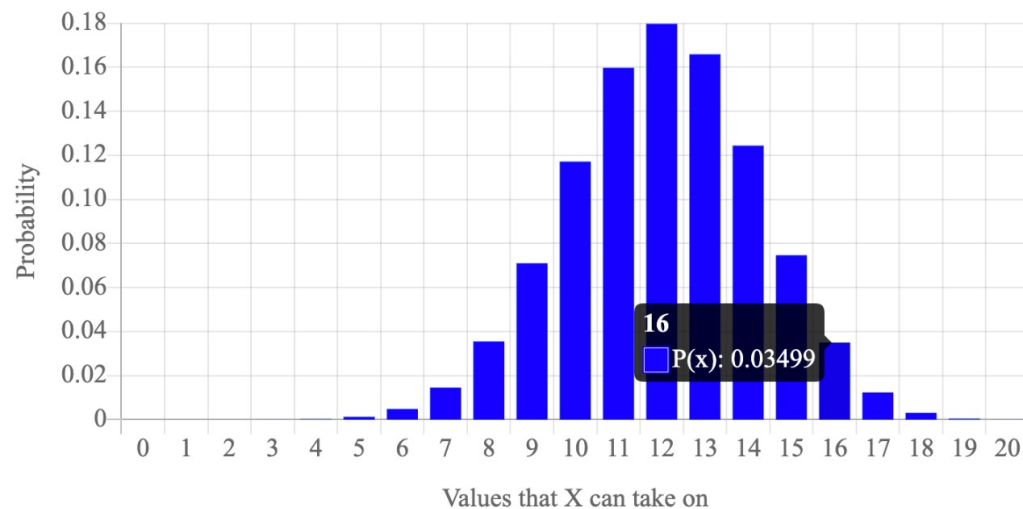
PMF equation: $\Pr(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$

Expectation: $E[X] = n \cdot p$

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

PMF graph:

Parameter n : Parameter p :



Bernoulli Random Variable

Notation: $X \sim \text{Bern}(p)$

Description: A boolean variable that is 1 with probability p

Parameters: p , the probability that $X = 1$.

Support: x is either 0 or 1

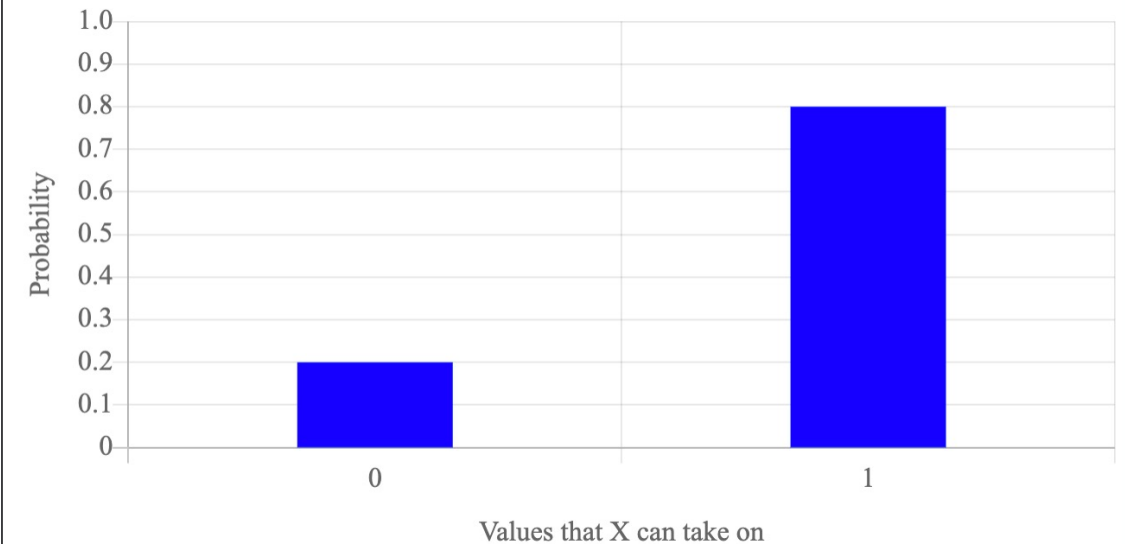
PMF equation: $\Pr(X = x) = \begin{cases} p & \text{if } x = 1 \\ 1 - p & \text{if } x = 0 \end{cases}$

Expectation: $E[X] = p$

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

PMF graph:

Parameter p :

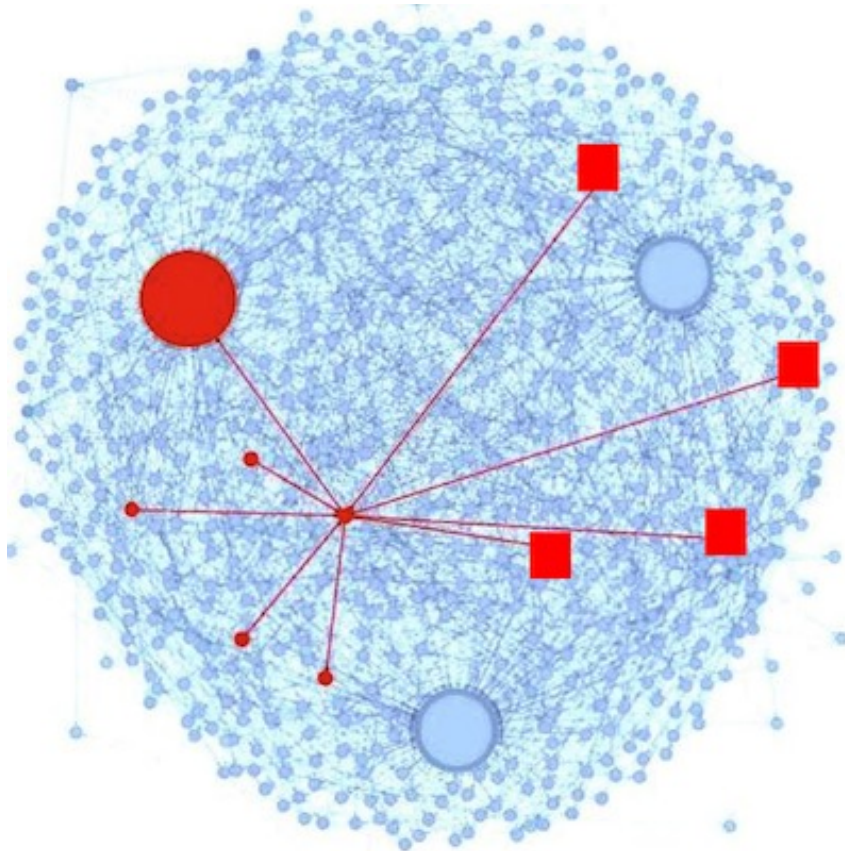


Expectation is a single number
summary...

Expectation leaves much to be
desired...

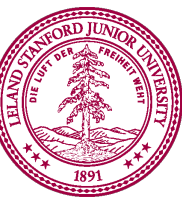
Can we invent *another* summary
number?

Intuition: Peer Grading

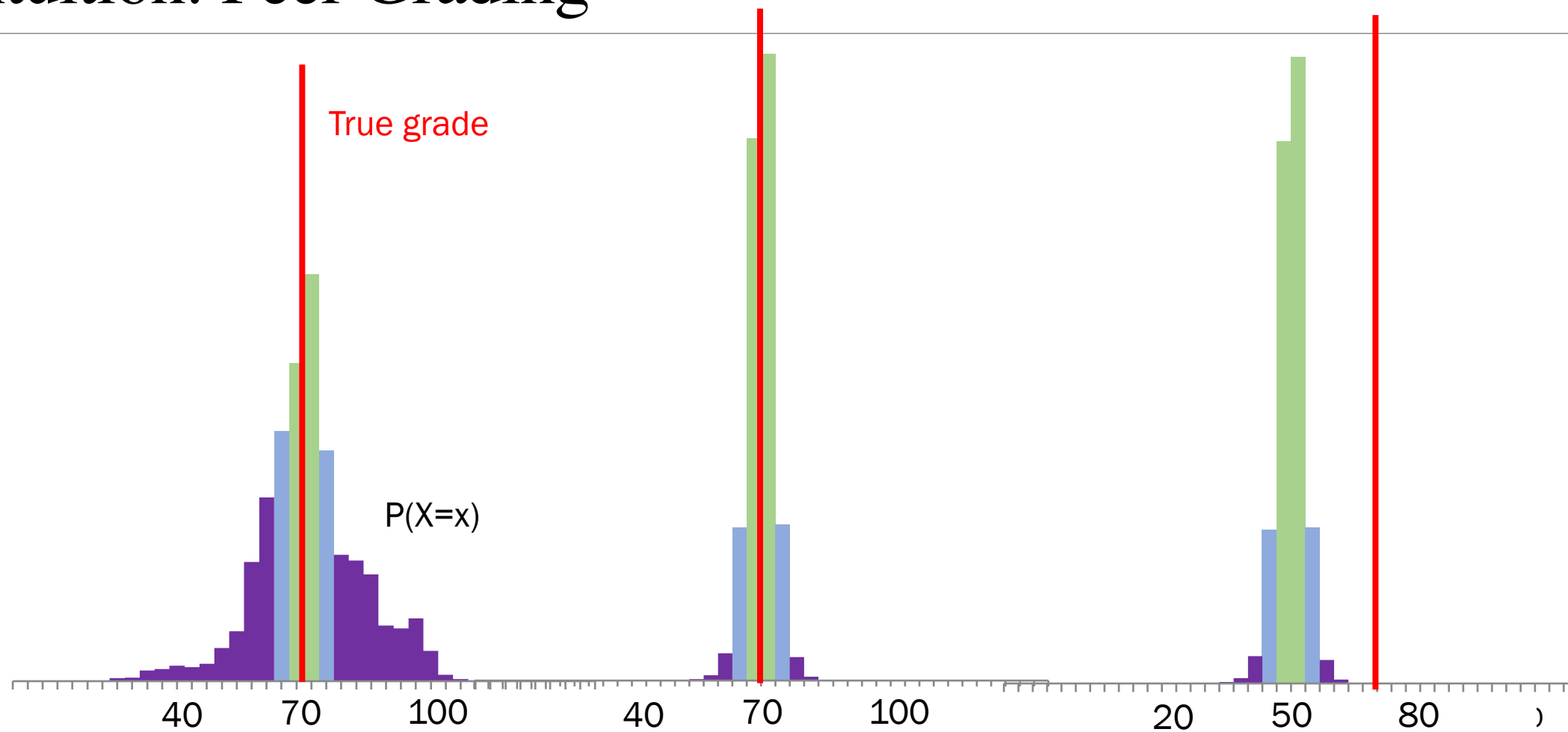


Peer Grading on Coursera HCI.

31,067 peer grades for 3,607 students.



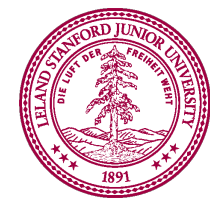
Intuition: Peer Grading



A

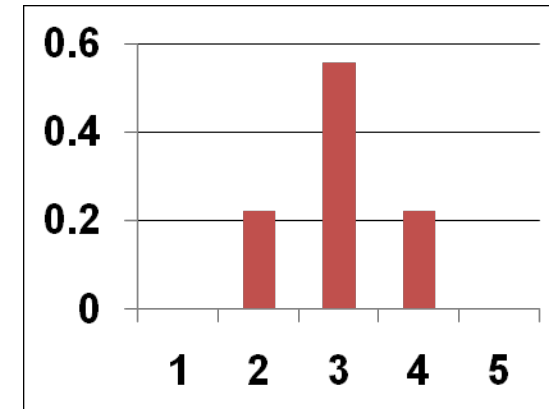
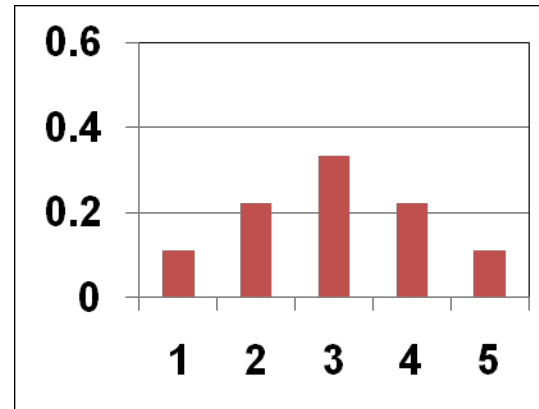
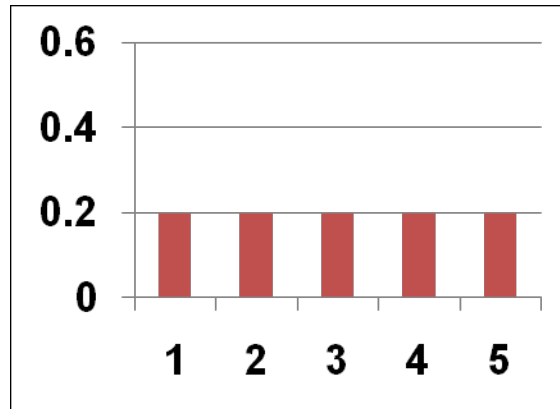
B

C



Intuition: Measure of Spread

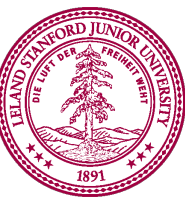
Consider the following 3 distributions (PMFs)



All have the same expected value, $E[X] = 3$

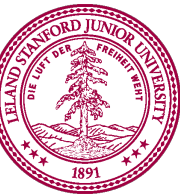
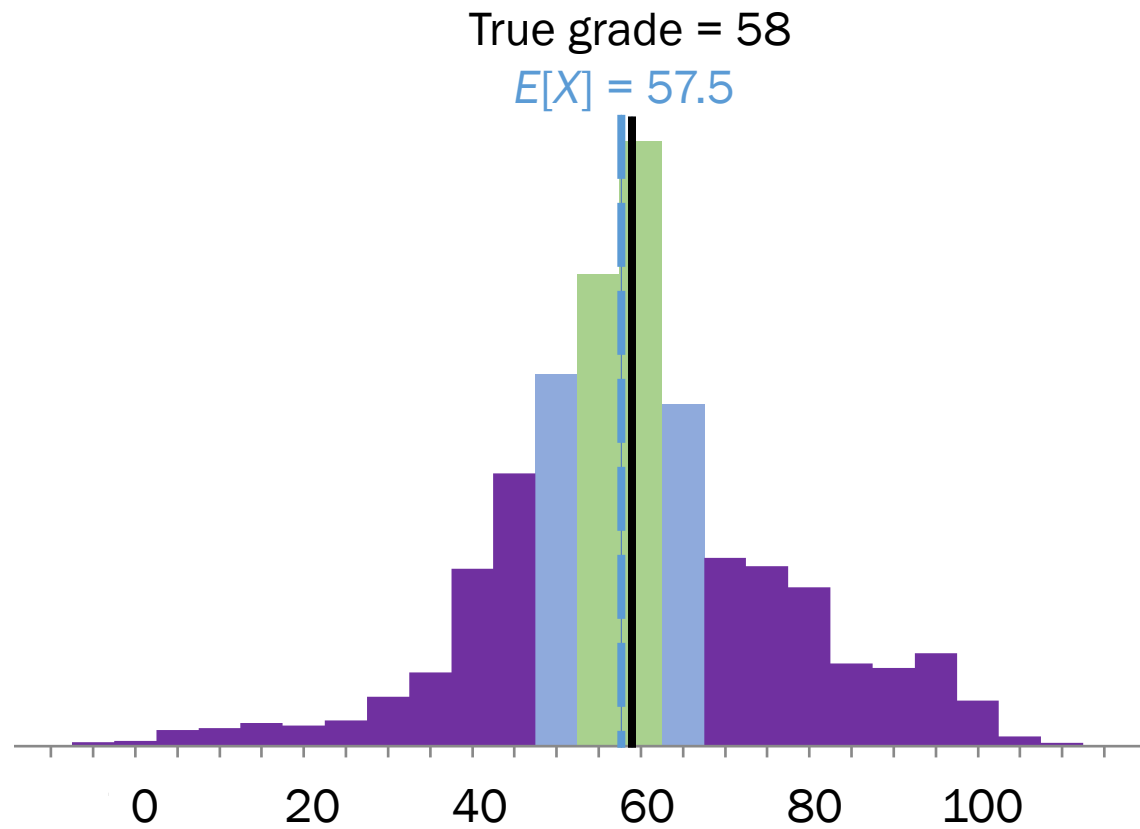
But “spread” in distributions is different

Invent a formal quantification of “spread”?



Peer grading in Coursera HCI

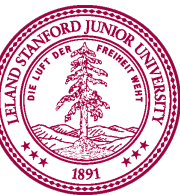
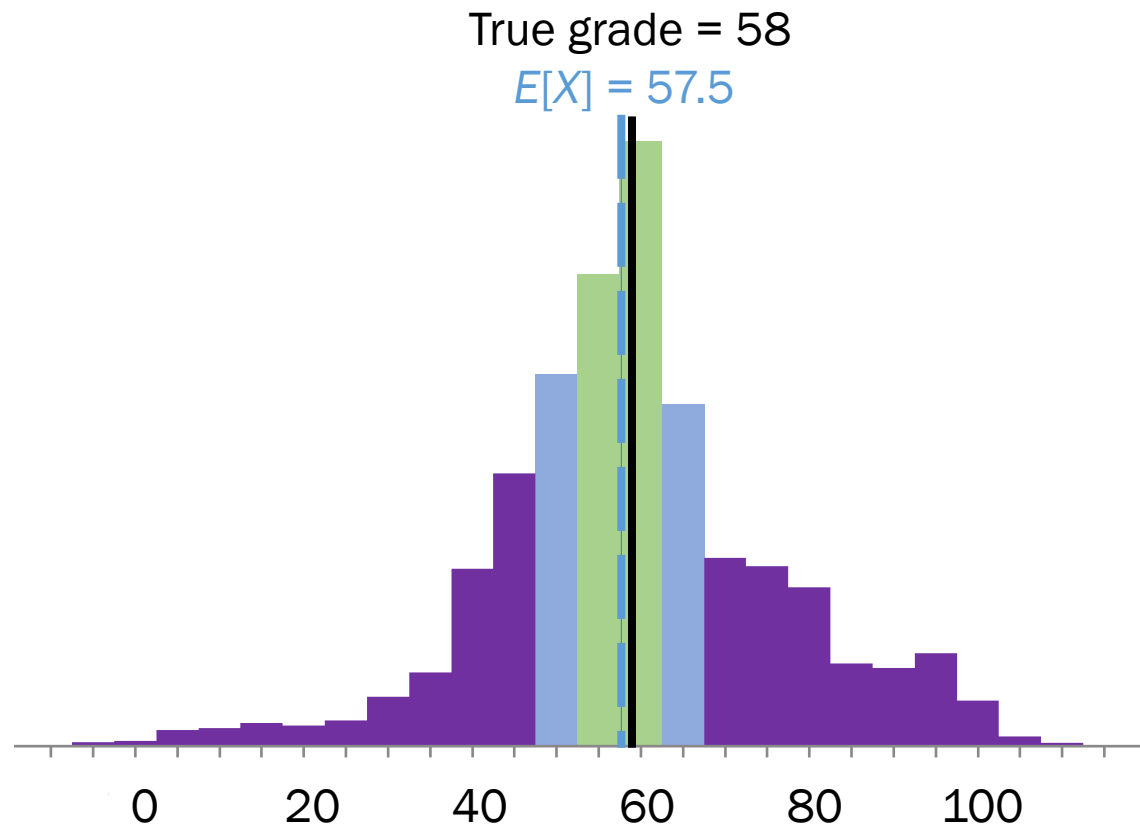
Let X be a random variable that represents a peer grade



Peer grading in Coursera HCI

Let X be a random variable that represents a peer grade

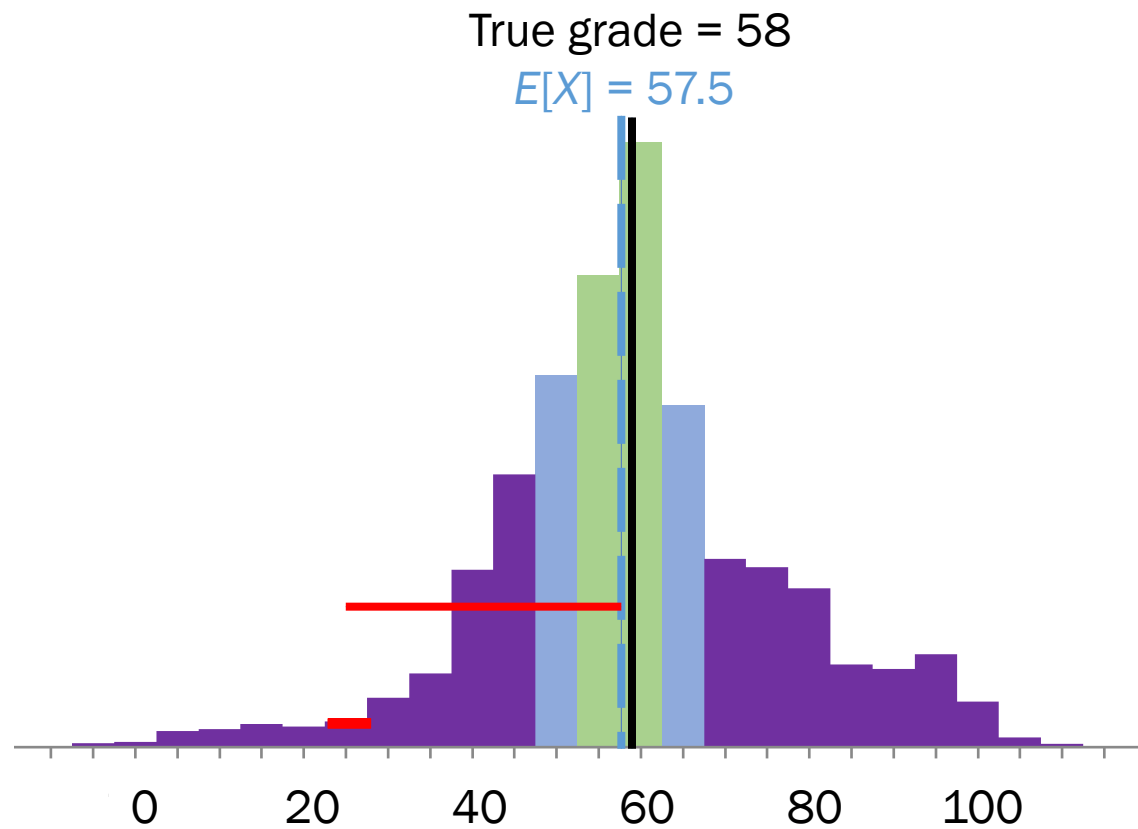
$$\text{Var}(X) = E[(X - \mu)^2]$$



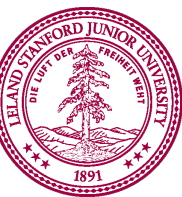
Peer grading in Coursera HCI

Let X be a random variable that represents a peer grade

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



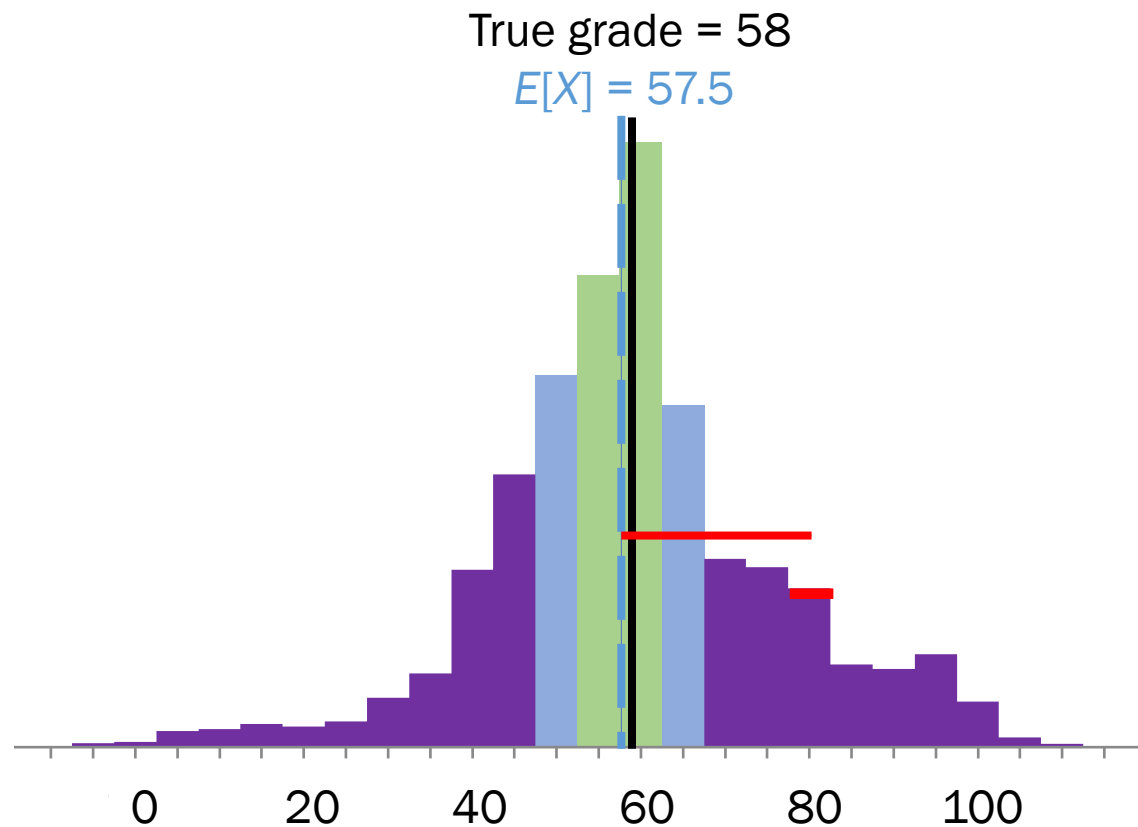
X	$(X - \mu)^2$
25 points	1056 points ²



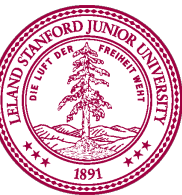
Peer grading in Coursera HCI

Let X be a random variable that represents a peer grade

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



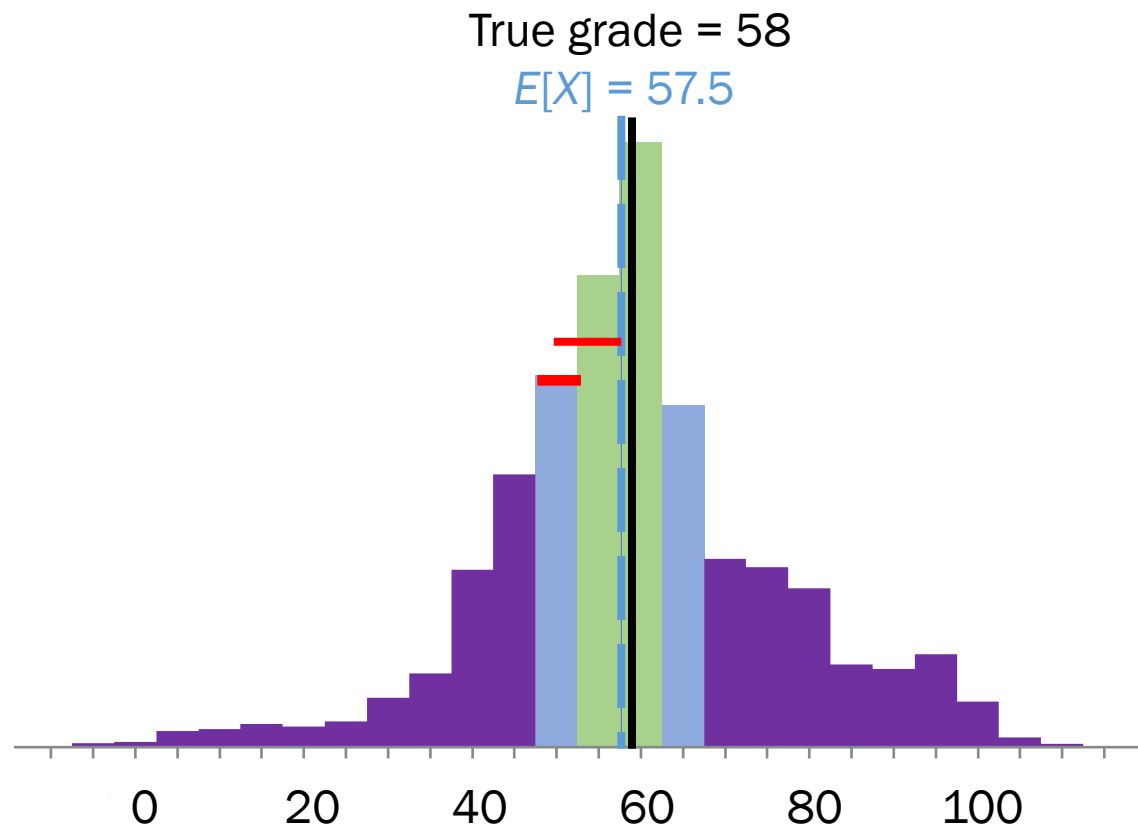
X	$(X - \mu)^2$
25 points	1056 points ²
80 points	506 points ²



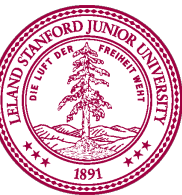
Peer grading in Coursera HCI

Let X be a random variable that represents a peer grade

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



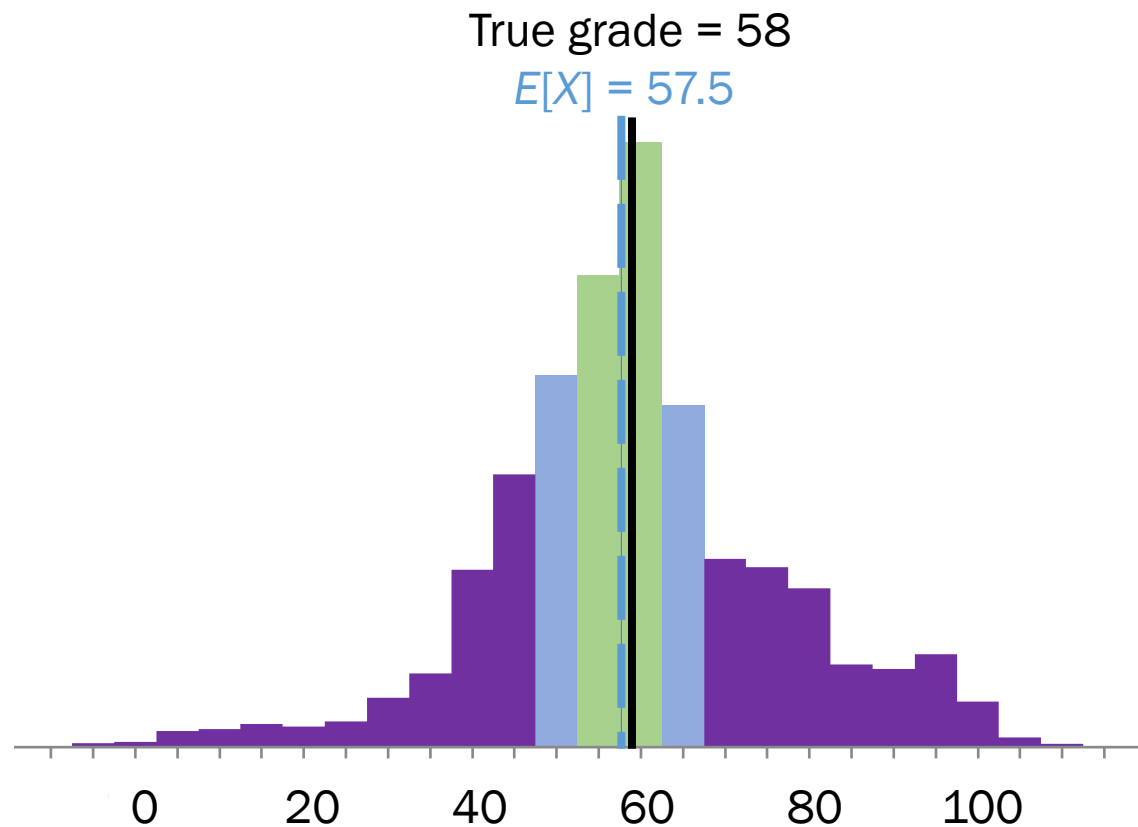
X	$(X - \mu)^2$
25 points	1056 points ²
80 points	506 points ²
50 points	56 points ²



Peer grading in Coursera HCI

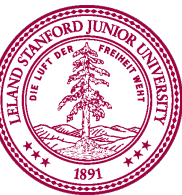
Let X be a random variable that represents a peer grade

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



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

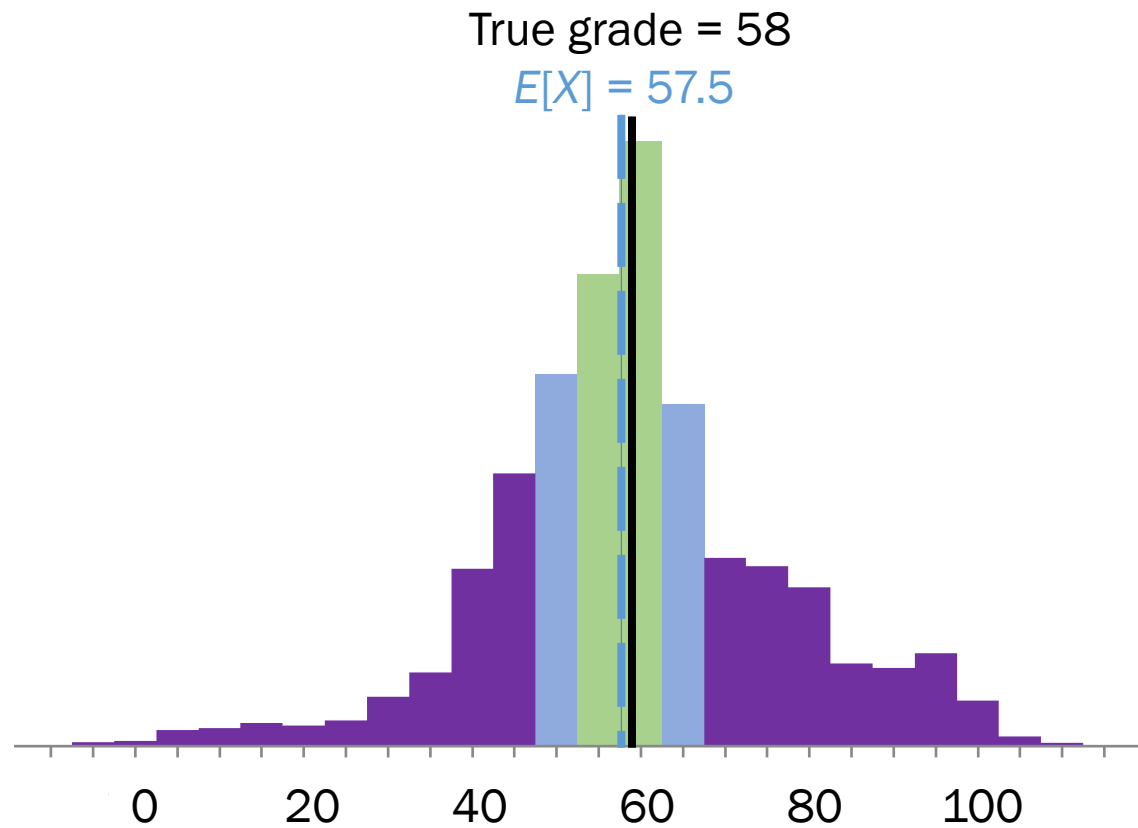
$$E[(X - \mu)^2] = 52 \text{ points}^2$$



Peer grading in Coursera HCI

Let X be a random variable that represents a peer grade

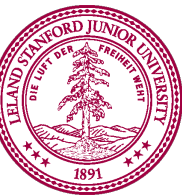
$$\text{Var}(X) = E[(X - \mu)^2]$$



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

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

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



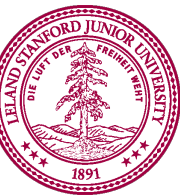
Variance

If X is a random variable with mean μ then the **variance** of X , denoted $\text{Var}(X)$, is:

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

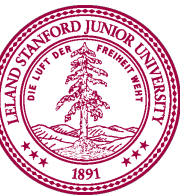
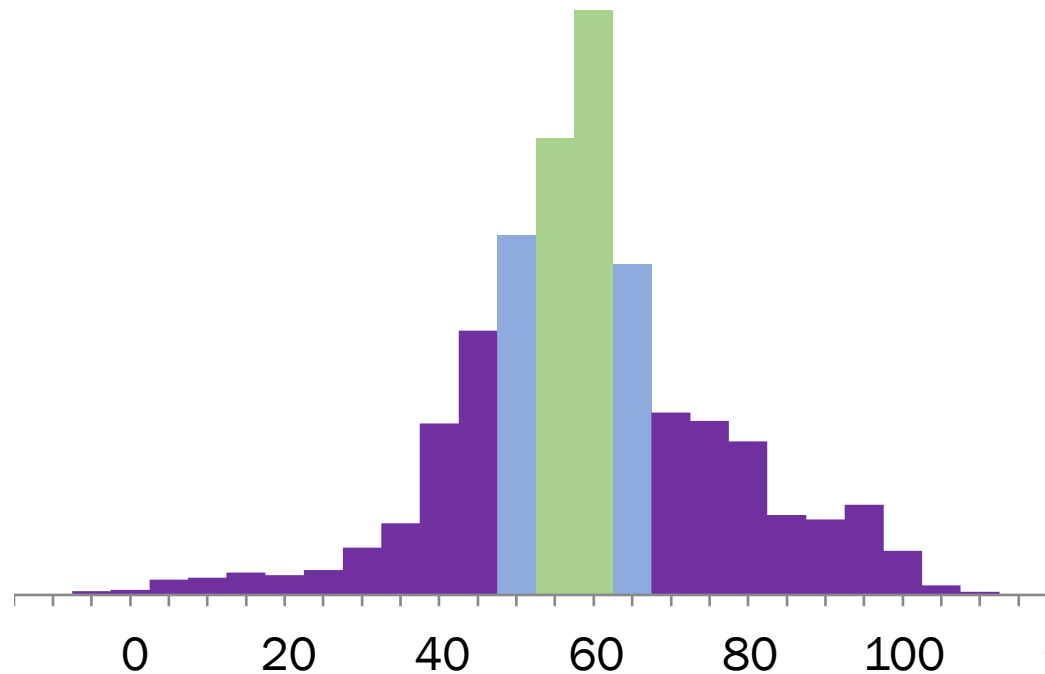
Variance is a formal definition of the **spread** of a random variable.

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





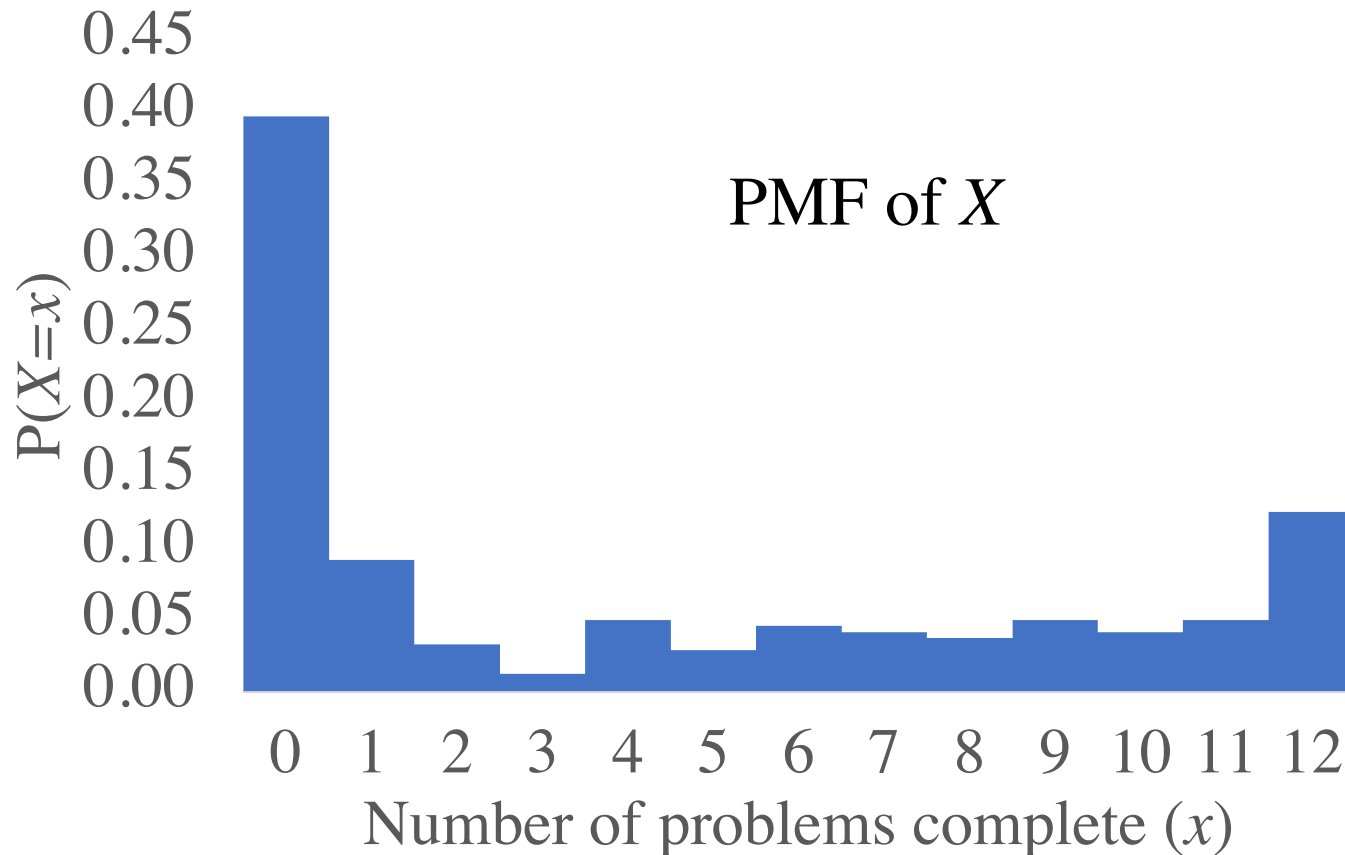
Normalized **histograms** are approximations of **probability mass functions**



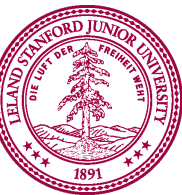
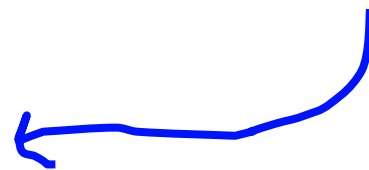
Normalize histograms are approximations of PMFs

Let X be the number of problems that a randomly selected student has completed, as of 11a today.

X takes on values, with uncertainty. X is a random variable.



$$P(X = 12) \approx \frac{\text{Count}(X = 12)}{N}$$



Computing Variance

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

Law of unconscious statistician

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

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

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

Ladies and gentlemen, please welcome the 2nd moment!

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

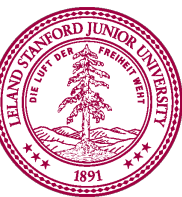
$$= E[X^2] - \mu^2$$

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

Notation

$$p(x) = P(X = x)$$

$$\mu = E[X]$$



How do you get $E[X^2]$?

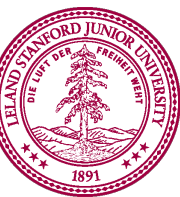
$$\text{Var}(X) = E[X^2] - E[X]^2$$

Unconscious statistician:

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

$E[X^2]$:

$$E[X^2] = \sum_x x^2 \cdot P(X = x)$$



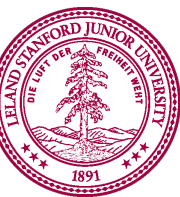
Standard Deviation?

$$\text{Std}(X) = \sqrt{\text{Var}(X)}$$

Units are in points



Units are in points squared



Variance of a 6 Sided Dice

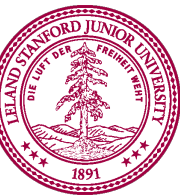
Let X = value on roll of 6 sided die

Recall that $E[X] = 7/2$

Compute $E[X^2]$

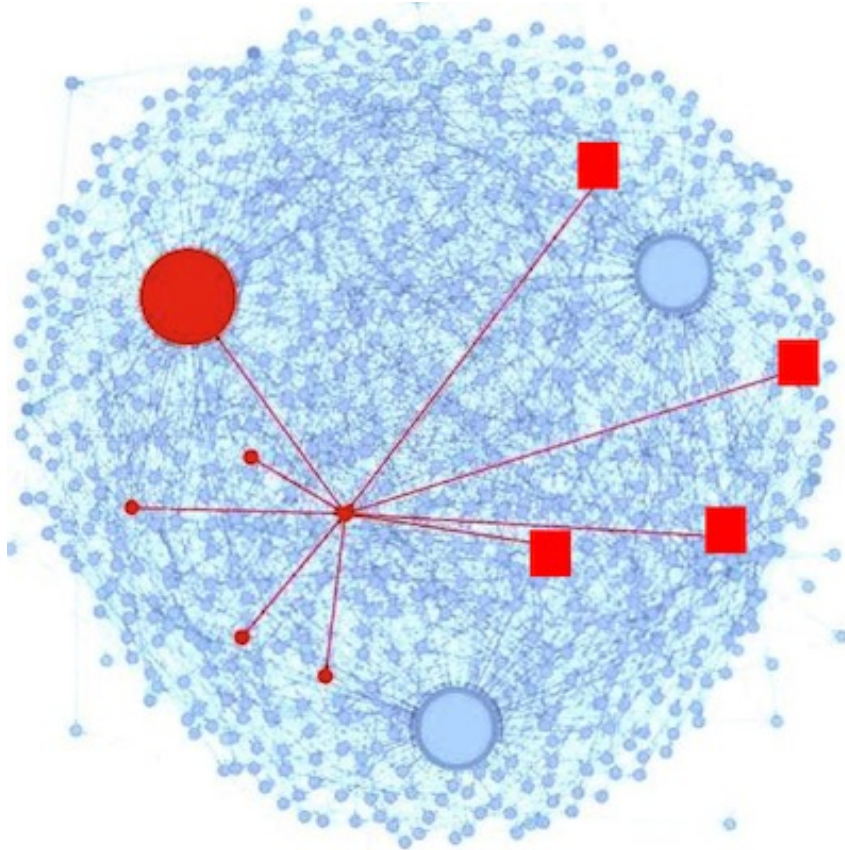
$$E[X^2] = (1^2)\frac{1}{6} + (2^2)\frac{1}{6} + (3^2)\frac{1}{6} + (4^2)\frac{1}{6} + (5^2)\frac{1}{6} + (6^2)\frac{1}{6} = \frac{91}{6}$$

$$\begin{aligned}\text{Var}(X) &= E[X^2] - (E[X])^2 \\ &= \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}\end{aligned}$$



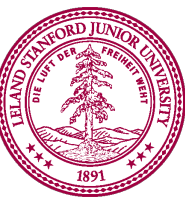
Is Peer Grading Accurate Enough?

Looking ahead



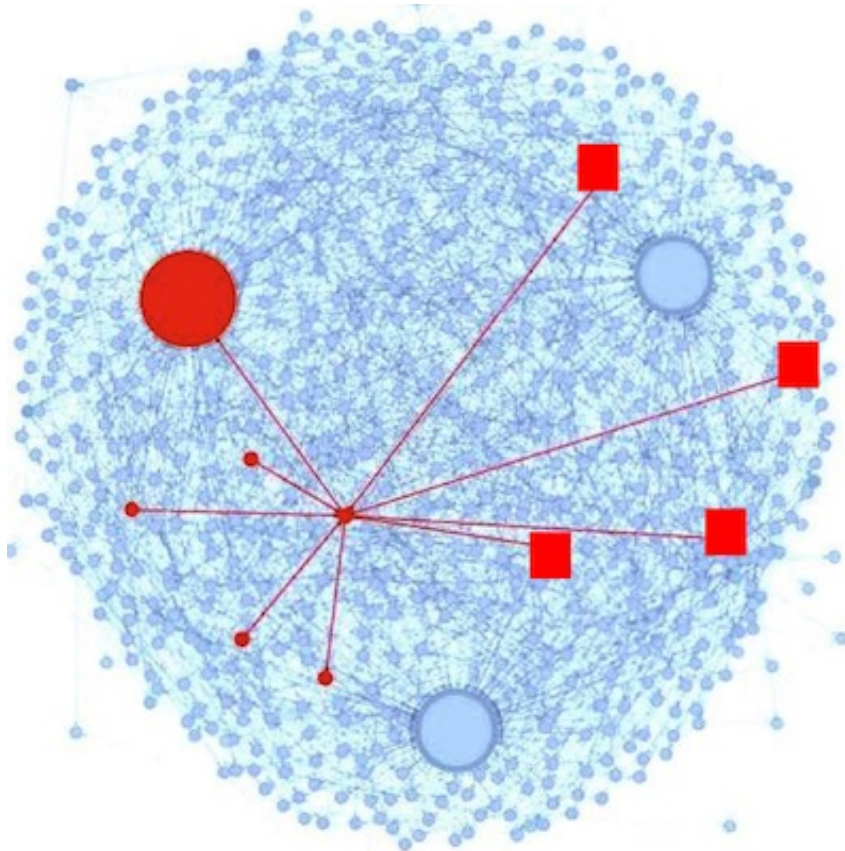
Peer Grading on Coursera HCI.

31,067 peer grades for 3,607 students.



Is Peer Grading Accurate Enough?

Looking ahead

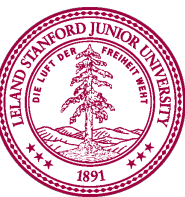


1. Defined random variables for:
 - True grade (s_i) for assignment i
 - Observed (z_i^j) score for assign i
 - Bias (b_j) for each grader j
 - Variance (r_j) for each grader j
2. Designed a probabilistic model that defined the distributions for all random variables

$$s_i \sim \text{Bin}(\text{points}, \theta)$$

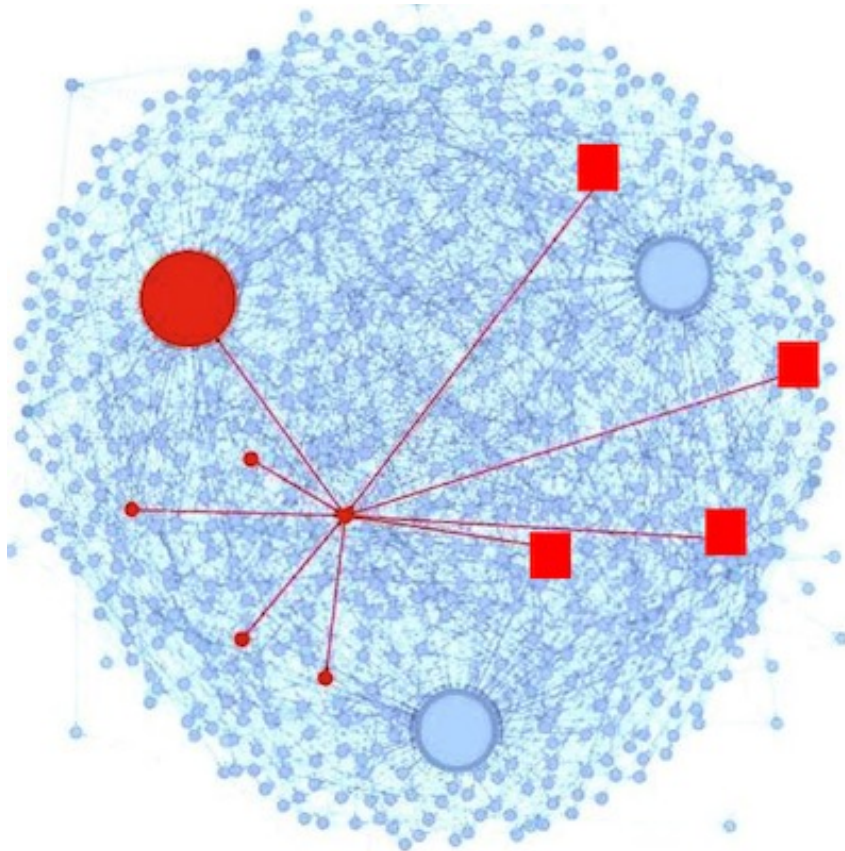
$$z_i^j \sim \mathcal{N}(\mu = s_i + b_j, \sigma = \sqrt{r_j})$$

Problem param
↙



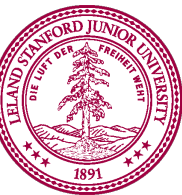
Is Peer Grading Accurate Enough?

Looking ahead

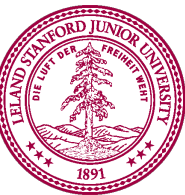
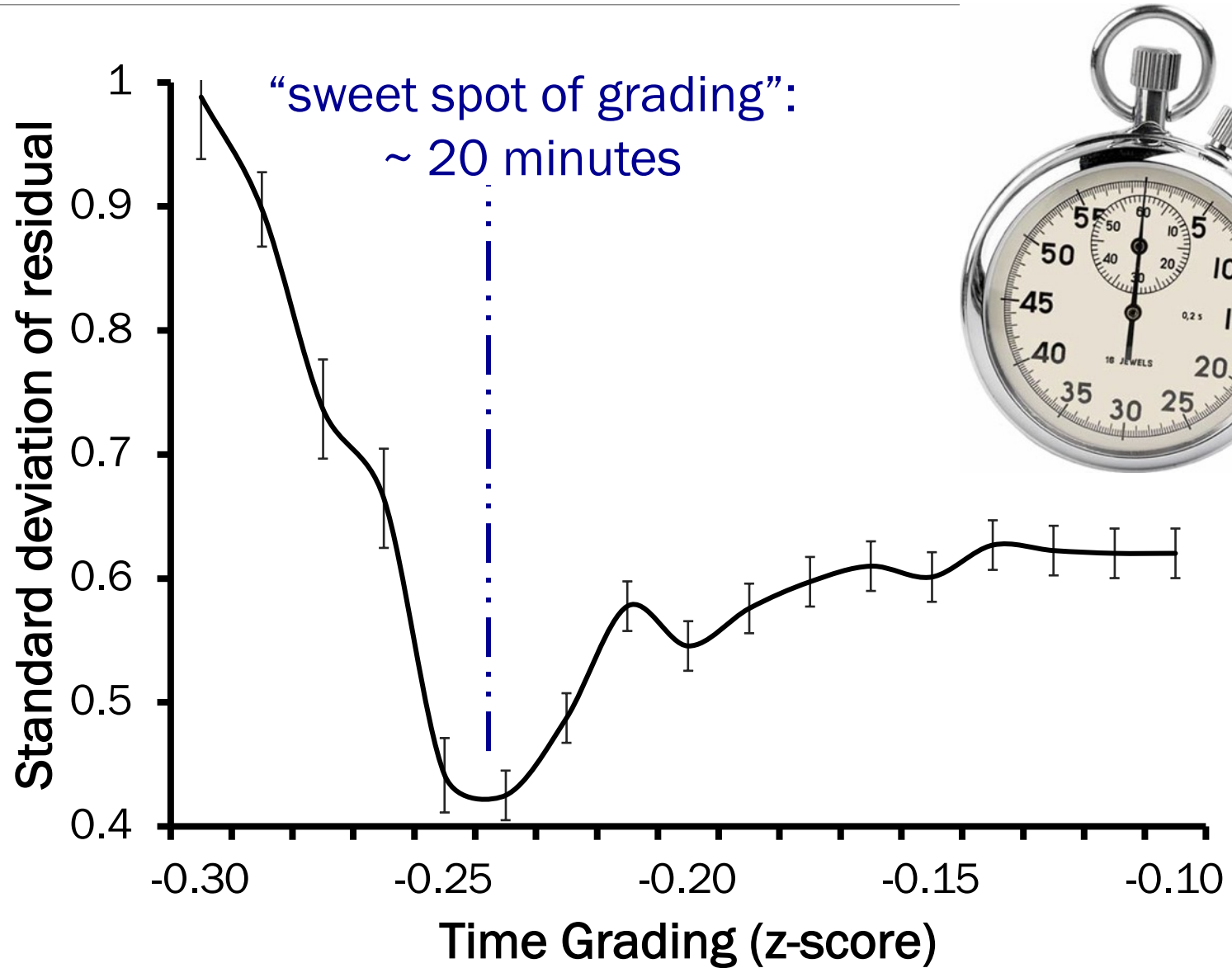


1. Defined random variables for:
 - True grade (s_i) for assignment i
 - Observed (z_i^j) score for assign i
 - Bias (b_j) for each grader j
 - Variance (r_j) for each grader j
2. Designed a probabilistic model that defined the distributions for all random variables
3. Found the variable assignments that maximized the probability of our observed data

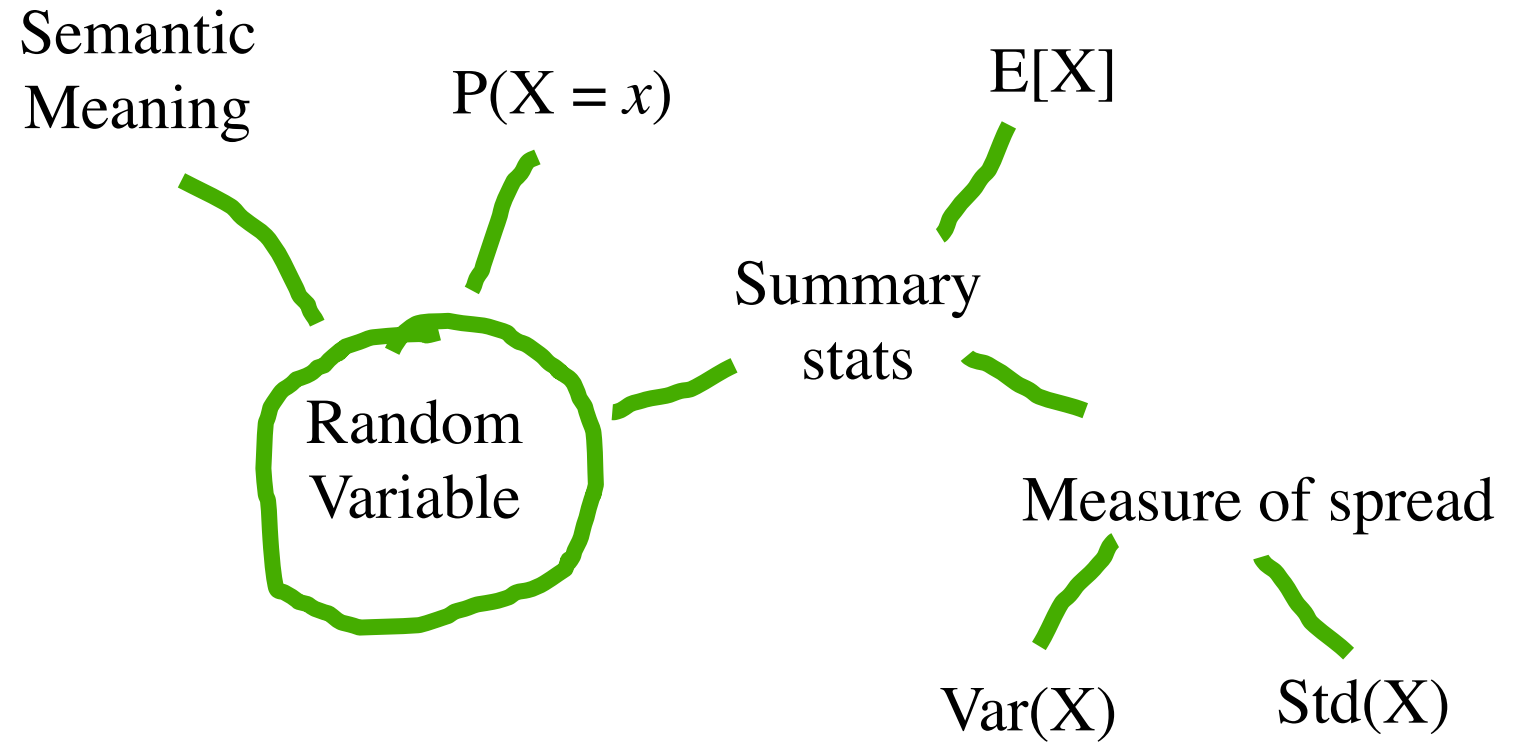
↑
Inference or Machine Learning



Grading Sweet Spot



Fundamental Properties of Random Variables



You Get So Much For Free!

Binomial Random Variable

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

Description: Number of "successes" in n identical, independent experiments each with probability of success p .

Parameters: $n \in \{0, 1, \dots\}$, the number of experiments.
 $p \in [0, 1]$, the probability that a single experiment gives a "success".

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

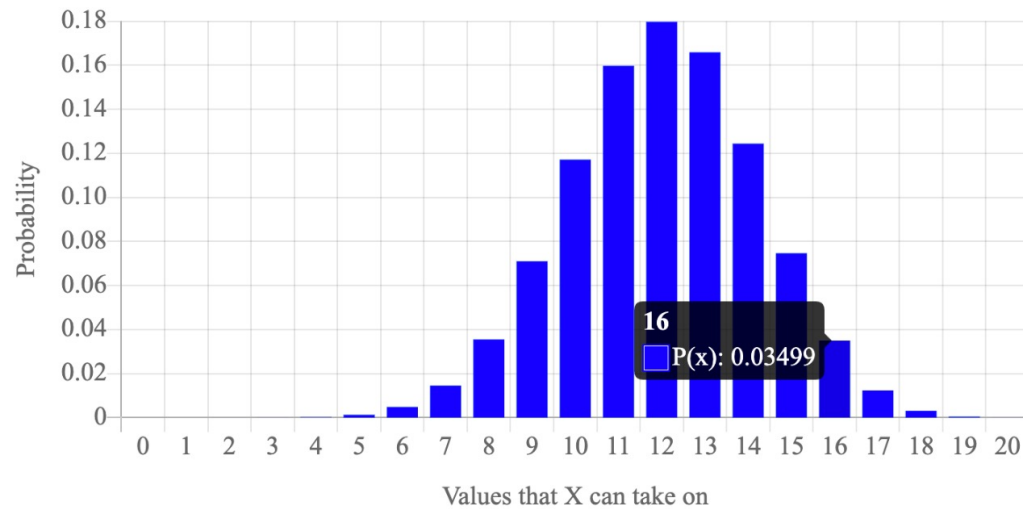
PMF equation: $\Pr(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$

Expectation: $E[X] = n \cdot p$

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

PMF graph:

Parameter n : Parameter p :



Bernoulli Random Variable

Notation: $X \sim \text{Bern}(p)$

Description: A boolean variable that is 1 with probability p

Parameters: p , the probability that $X = 1$.

Support: x is either 0 or 1

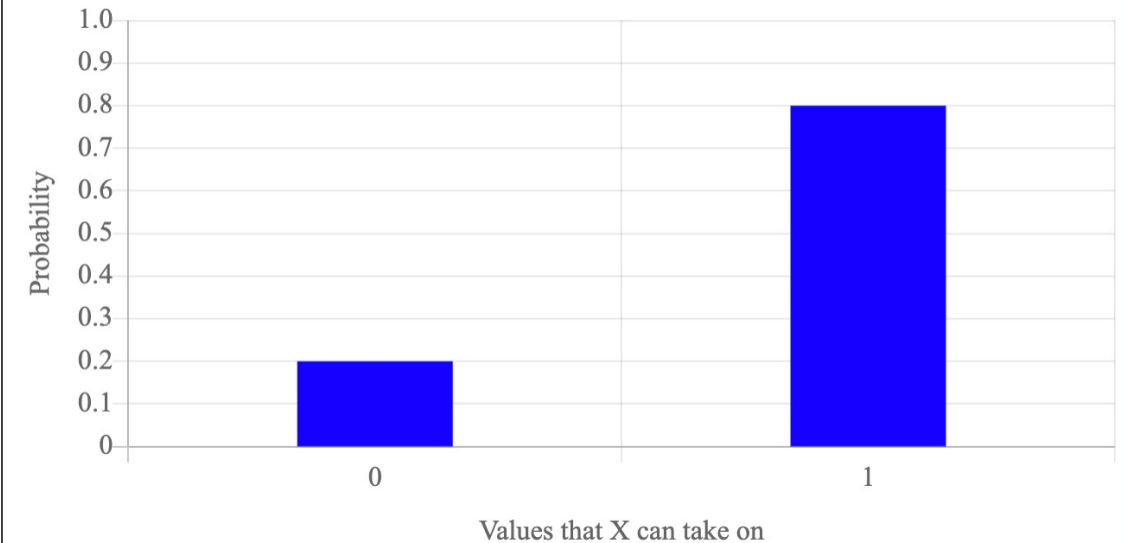
PMF equation: $\Pr(X = x) = \begin{cases} p & \text{if } x = 1 \\ 1 - p & \text{if } x = 0 \end{cases}$

Expectation: $E[X] = p$

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

PMF graph:

Parameter p :



Beyond CS109: Proof of Variance for a Binomial

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

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

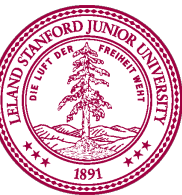
Factors of Binomial Coefficient: $j \binom{m}{j} = m \binom{m-1}{j-1}$

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

Binomial Theorem

as $p + q = 1$

by algebra



Voilà, c'est tout

