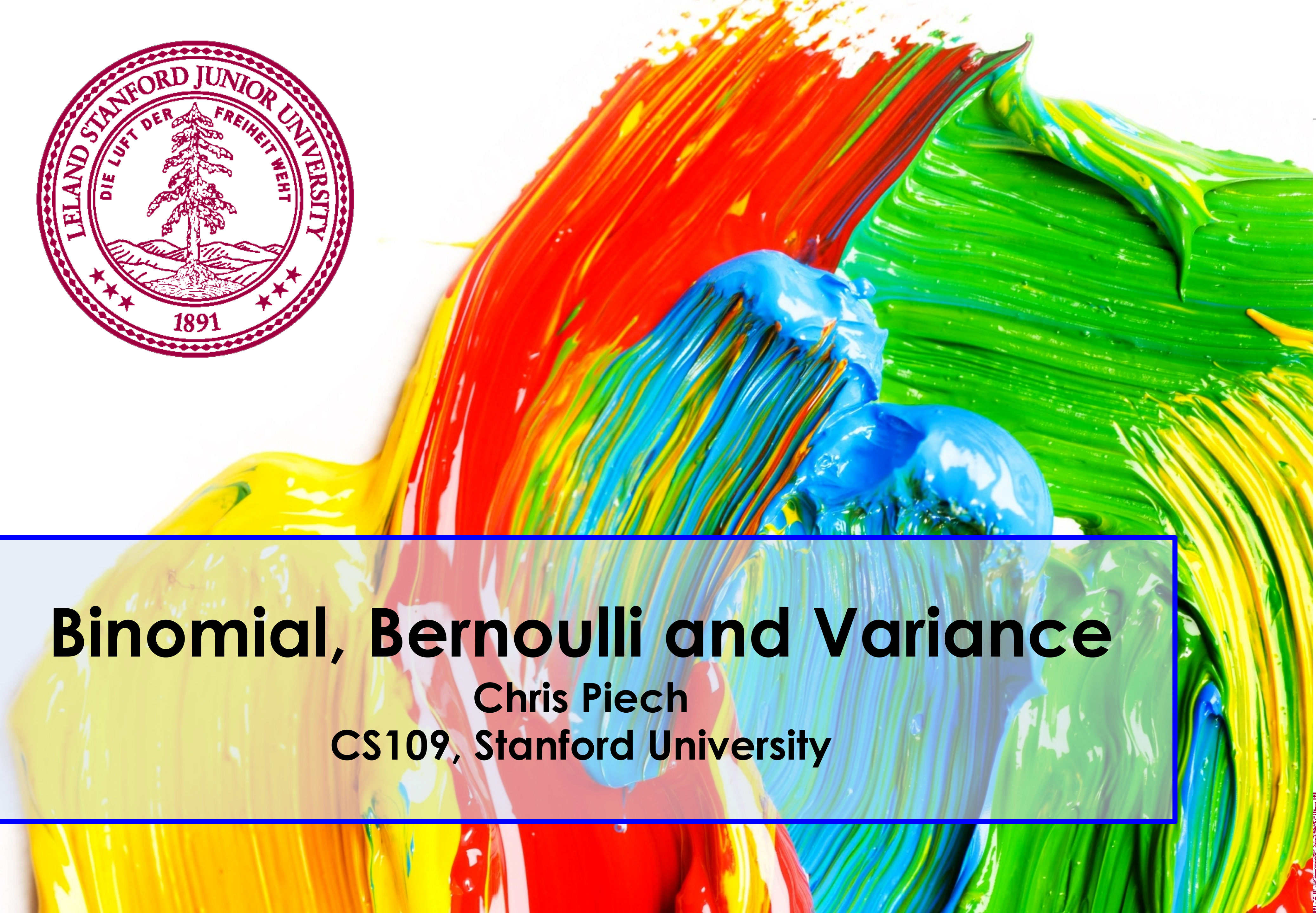


Binomial, Bernoulli and Variance

Chris Piech
CS109, Stanford University

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).

High-Resolution Course Feedback (Things we are working on)

Can there be more virtual OH options.

> Added 8 hours zoom only office hours (4 are brand new)

I do wish that the secret code thing was put into the recordings

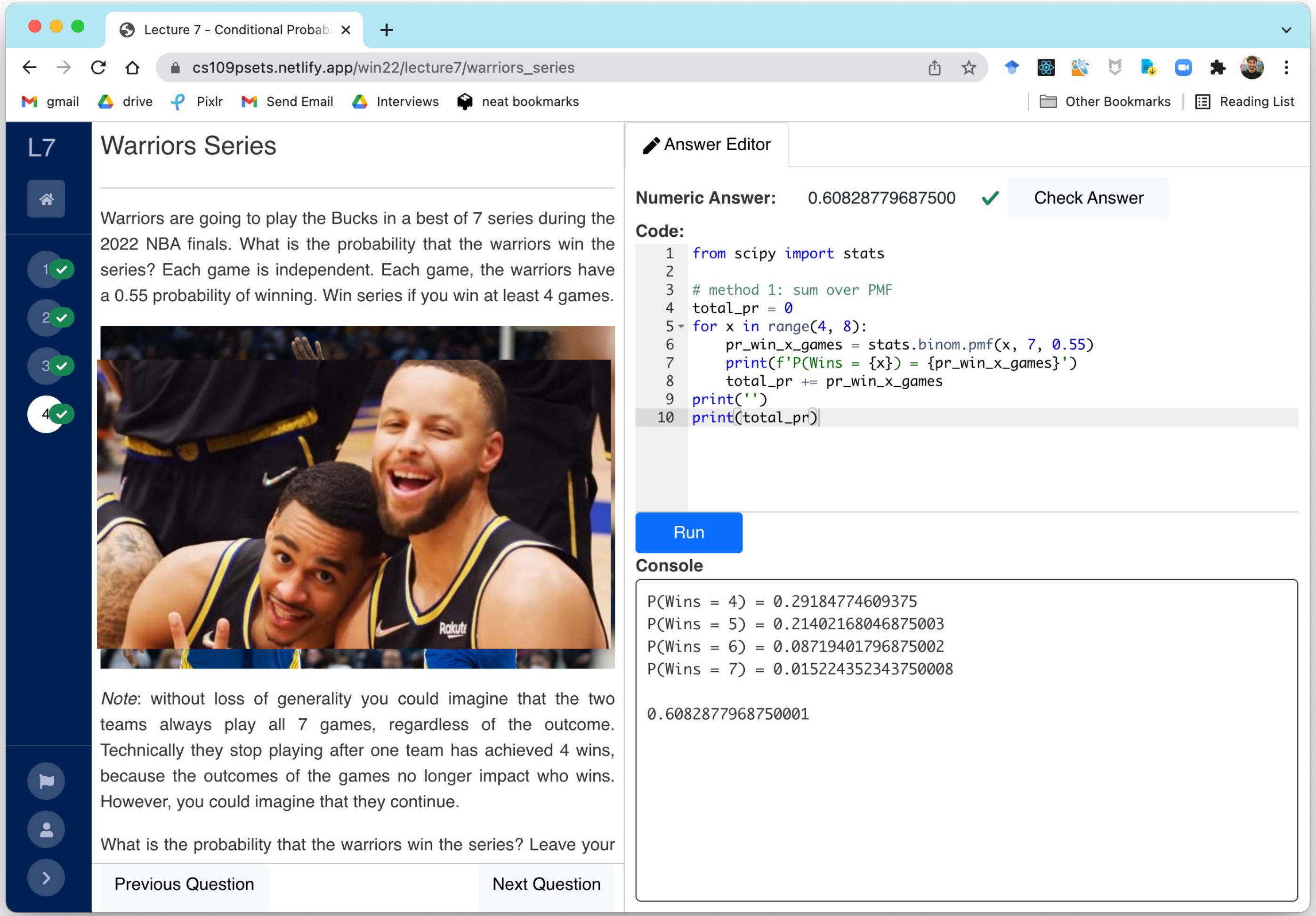
> Done. Enter it before the next class

The internet is so bad in that room, I usually can't even load simple webpages. That's a problem for me because I like to reference the course reader or look up concepts during class and I can't.

> Raised the issue! Let's get this fixed!

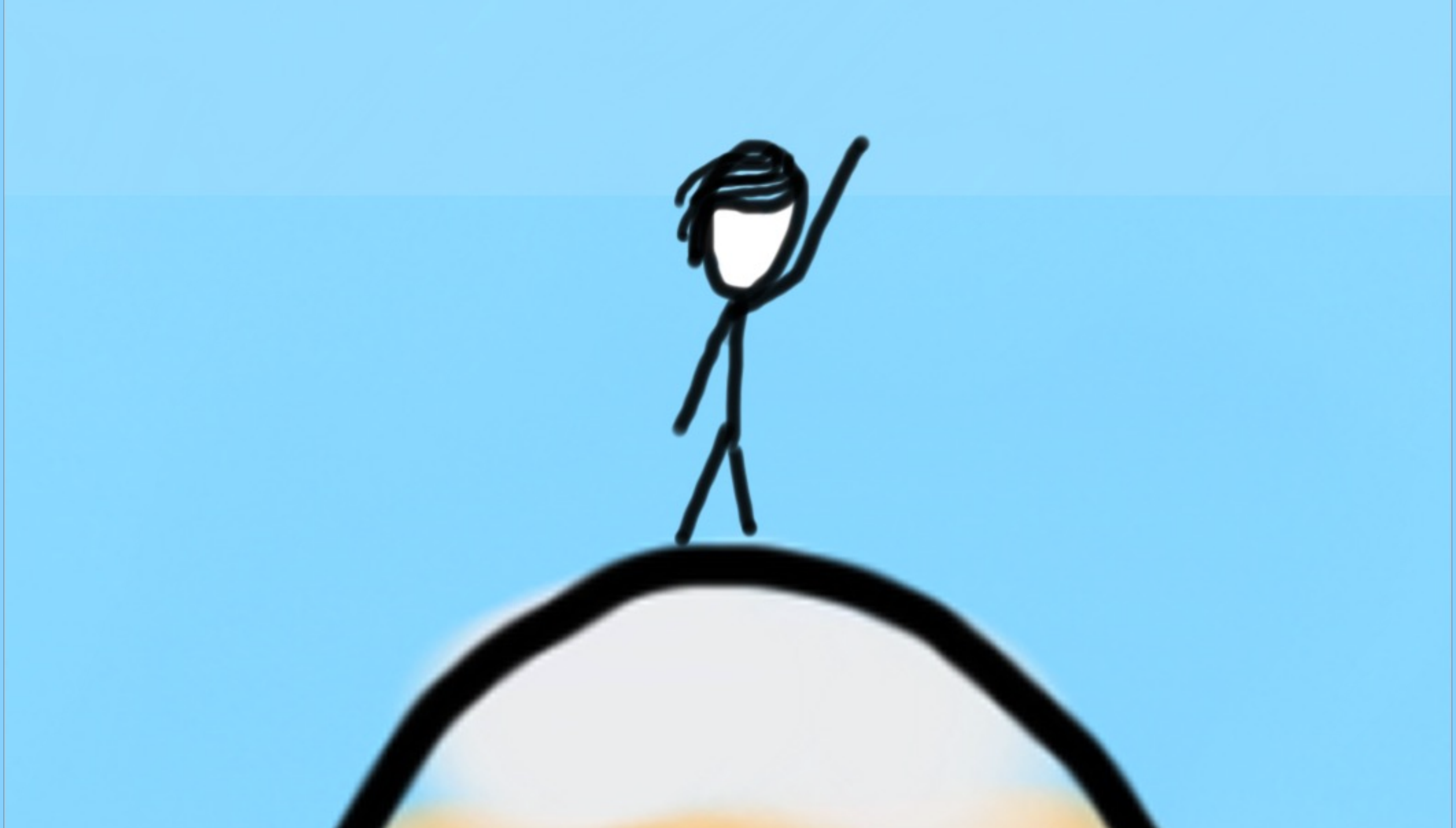
Follow Along!

https://cs109psets.netlify.app/fall22/lecture7/

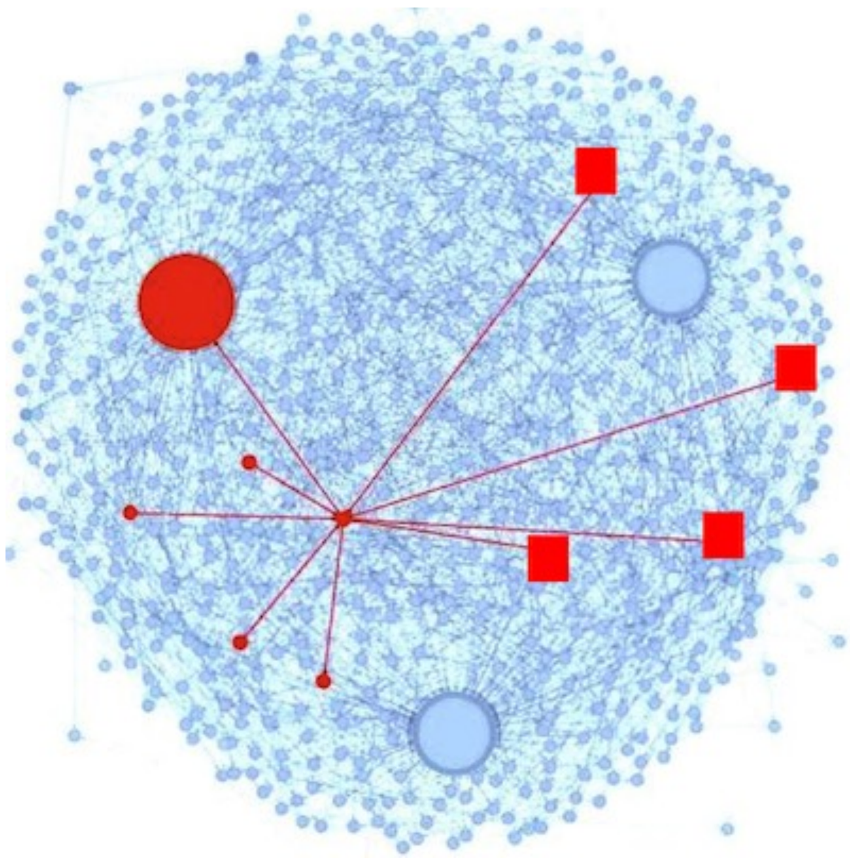


Learning Goals

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.

Tuned Models of Peer Assessment. C Piech, J Huang, A Ng, D Koller

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

$$P(Y = k)$$

$$k = 5$$

$$0.03125$$

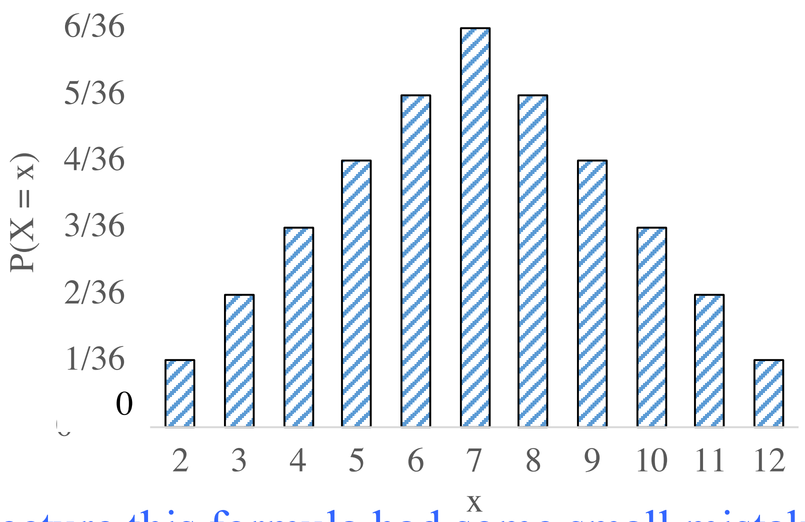
For example Y is the number of heads in 5 coin flips

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

PMF as an Equation

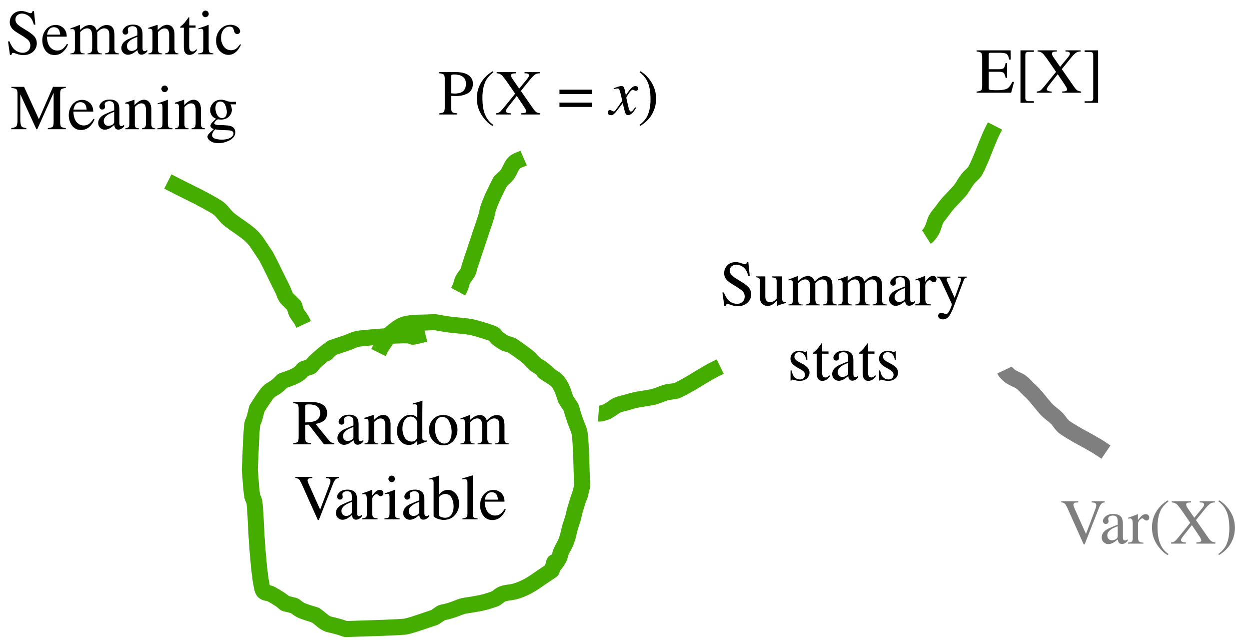
$$p(X = x) = \begin{cases} \frac{x-1}{36} & \text{if } x \in \mathbb{Z}\ ,\ 1 \leq x \leq 6 \\ \frac{13-x}{36} & \text{if } x \in \mathbb{Z}\ ,\ 7 \leq x \leq 12 \\ 0 & \text{else} \end{cases}$$

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



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

Fundamental Properties



Expected Value

The value

The probability of that value

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

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
...

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

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

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

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
...	...	...

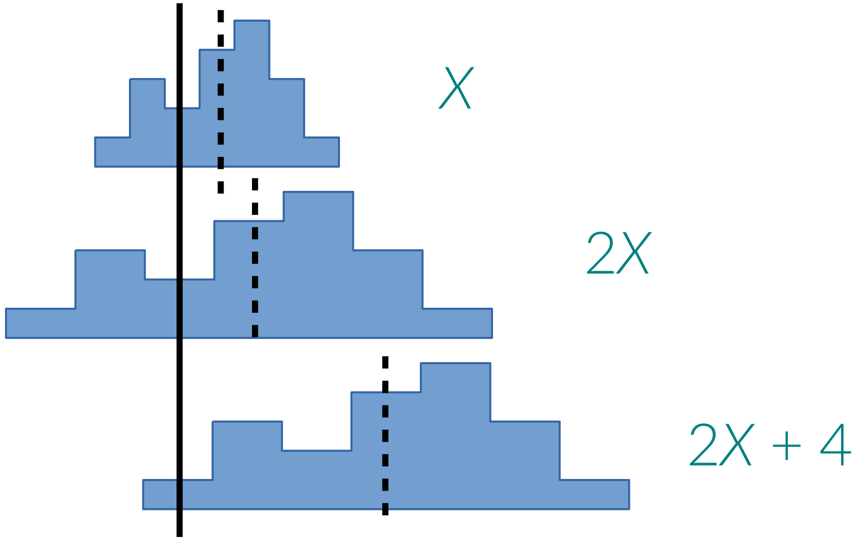
$$\frac{1}{n}\sum_{i=1}^{n} x_i \quad + \quad \frac{1}{n}\sum_{i=1}^{n} y_i \quad = \quad \frac{1}{n}\sum_{i=1}^{n} (x_i + y_i)$$

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

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

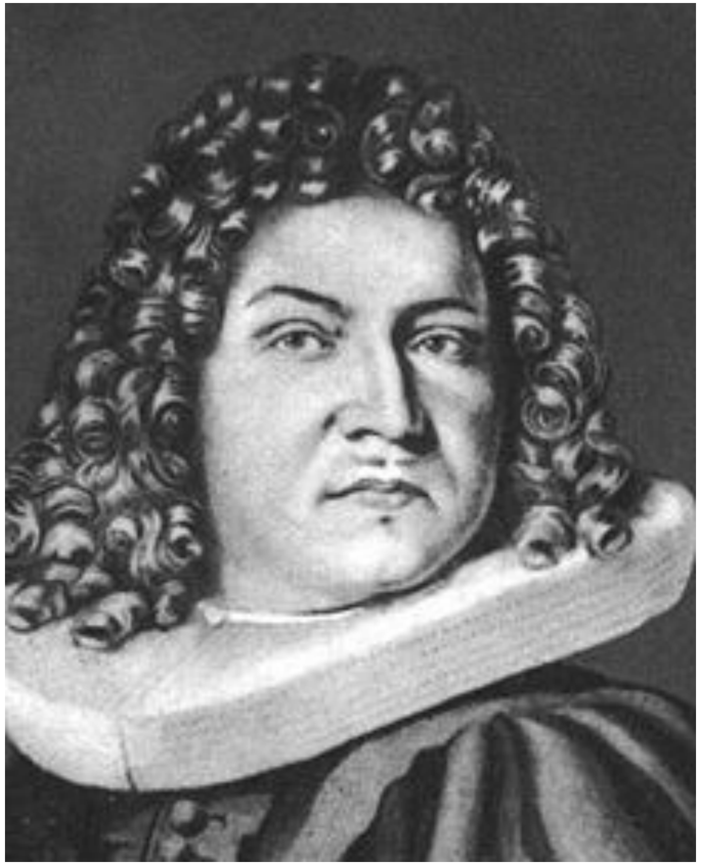
Exactly k heads in n coin flips

```
(H, H, H, H, T, T, T, T, T, T)
(H, H, H, T, H, T, T, T, T, T)
(H, H, H, T, T, H, T, T, T, T)
(H, H, H, T, T, T, H, T, T, T)
(H, H, H, T, T, T, T, H, T, T)
(H, H, H, T, T, T, T, T, H, T)
(H, H, H, T, T, T, T, T, T, H)
(H, H, T, H, H, T, T, T, T, T)
(H, H, T, H, T, H, T, T, T, T)
(H, H, T, H, T, T, H, T, T, T)
(H, H, T, H, T, T, T, H, T, T)
(H, H, T, H, T, T, T, T, H, T)
(H, H, T, H, T, T, T, T, T, H)
(H, H, T, T, H, H, T, T, T, T)
(H, H, T, T, H, T, H, T, T, T)
(H, H, T, T, H, T, T, H, T, T)
(H, H, T, T, H, T, T, T, H, T)
(H, H, T, T, H, T, T, T, T, H)
```

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* great^{12}-grandfather

Celebrity look alike: Ice Cube

Declare a Random Variable to be Binomial

Our random variable

Num trials

Probability of success on each trial

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

Is distributed as a

Binomial

With these parameters

Automatically Know the PMF

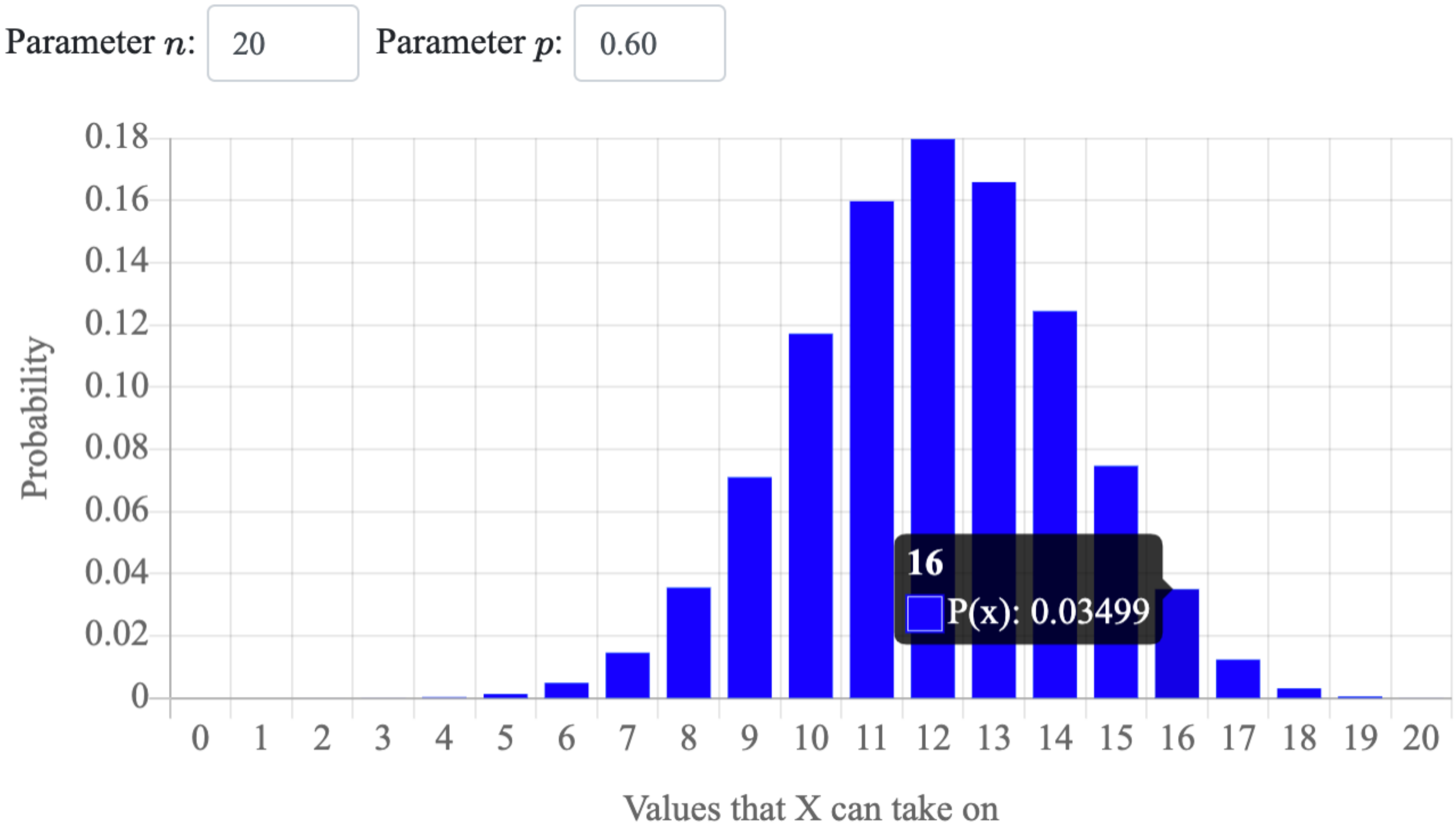
Probability Mass Function for a Binomial

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

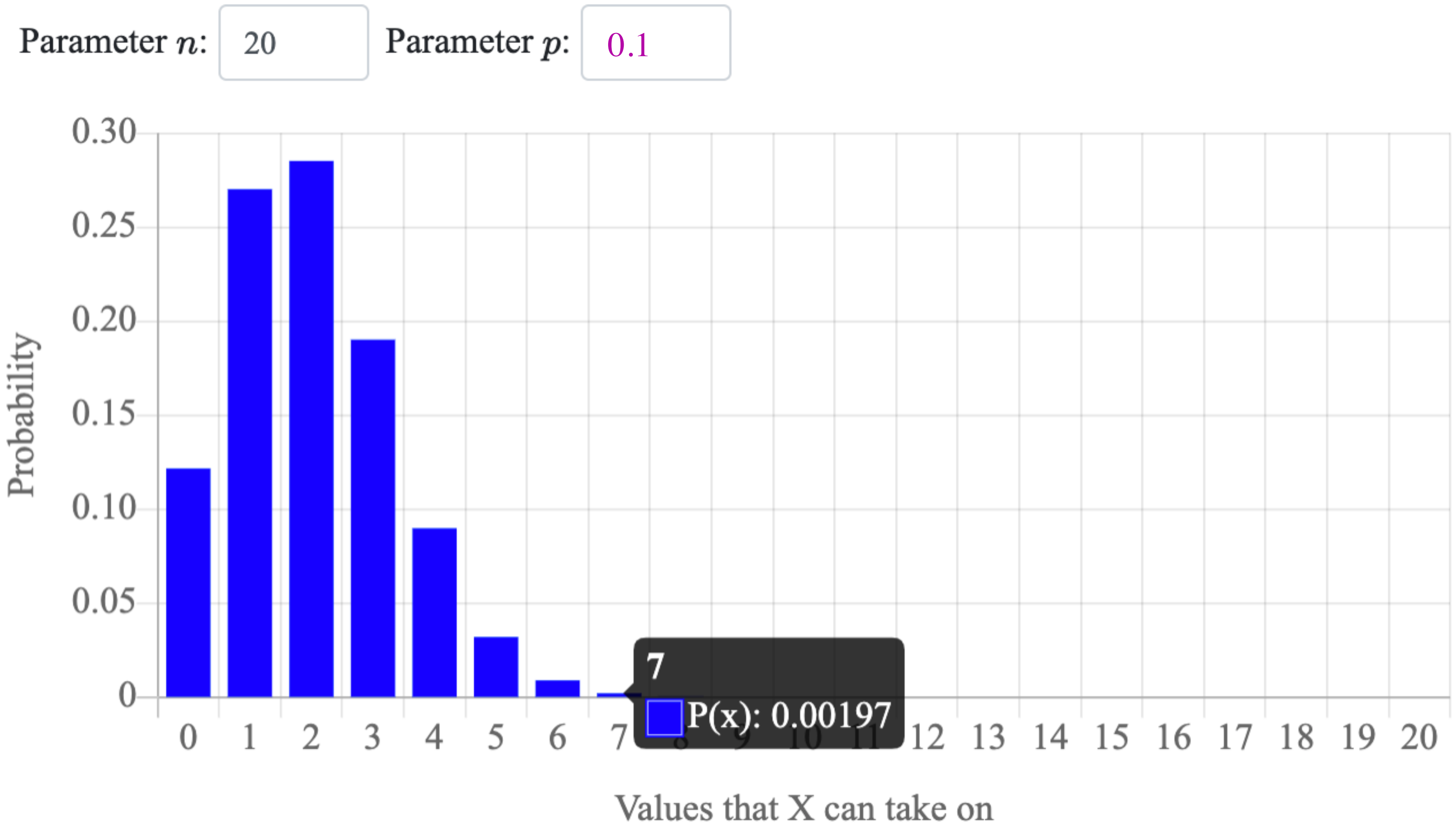
Probability that our variable takes on the value k

* This is also called the binomial term

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

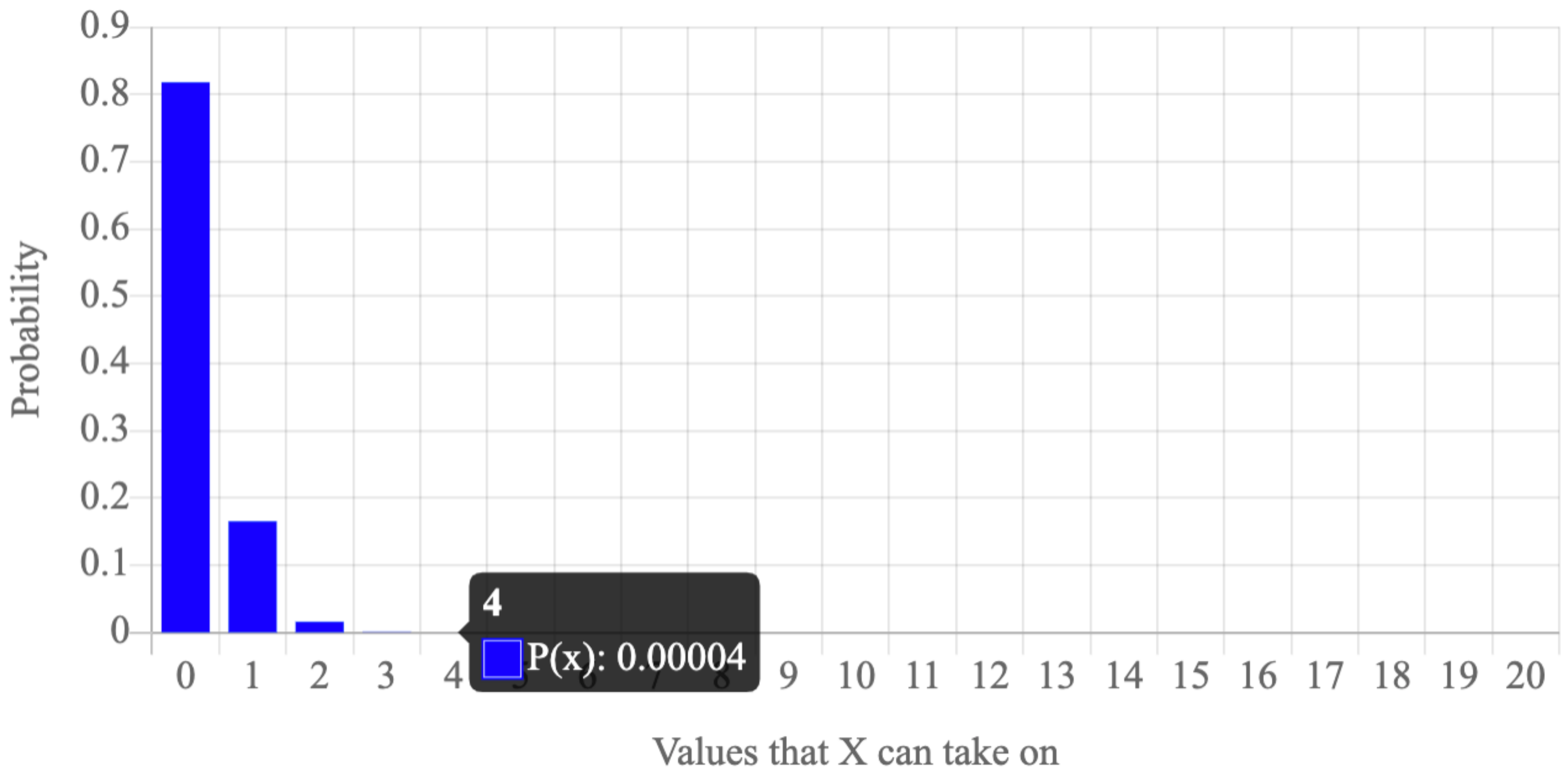


The PMF as a Graph: $X \sim \text{Bin}(n = 20, p = 0.1)$



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

Parameter n: 20 Parameter p: 0.01



Coins, now with Binomial.

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

- X is number of heads
- X ~ 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

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

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

$$\mathbf{P}(\mathbf{H} = 10) = \binom{1000}{10} (0.01)^{10} (0.99)^{998} \approx 0.125$$

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:

```
from scipy import stats

r = stats.binom.pmf(5, 1000, 0.01)
print(r)
```

Run

Console

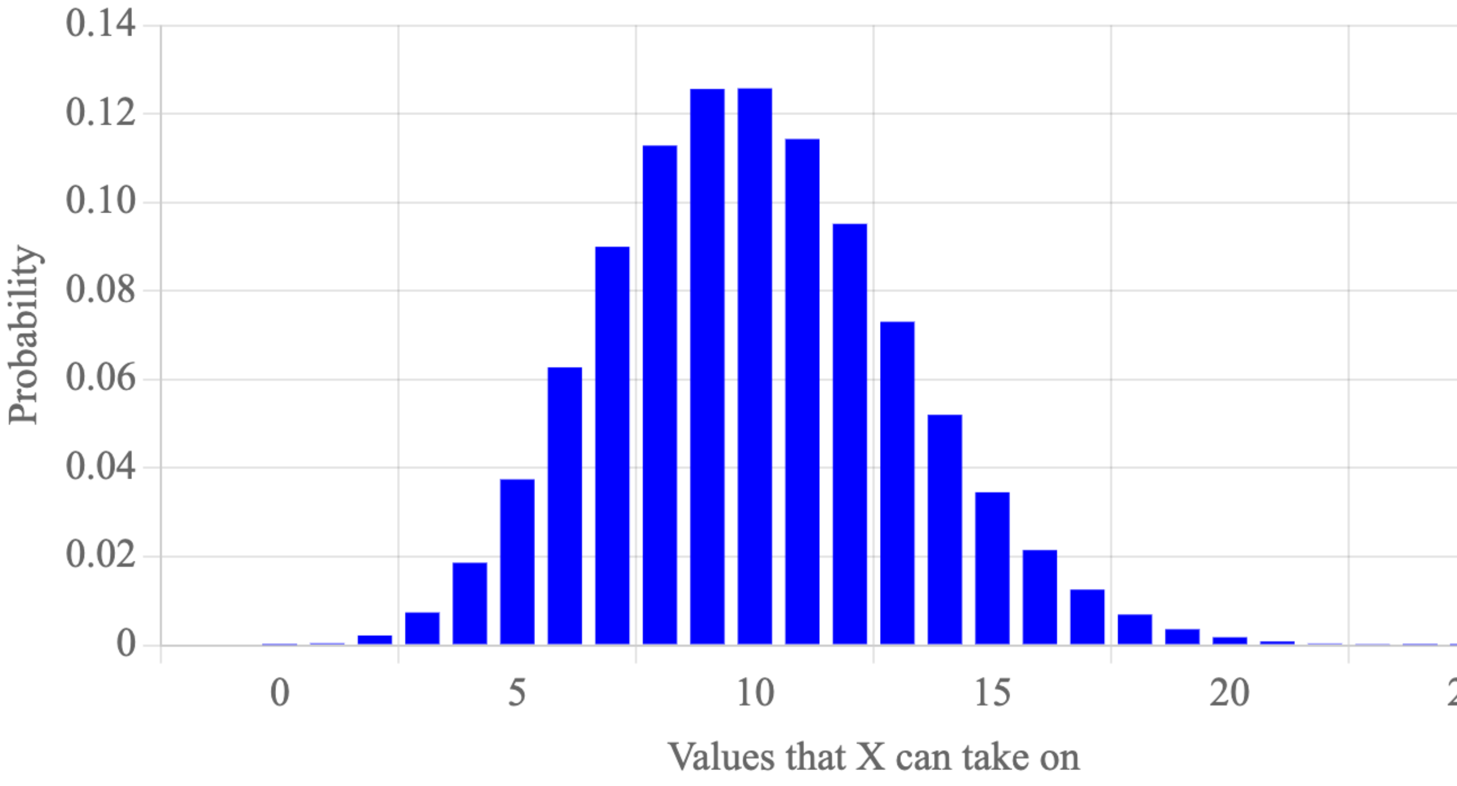
```
0.03745311160828357
```

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:

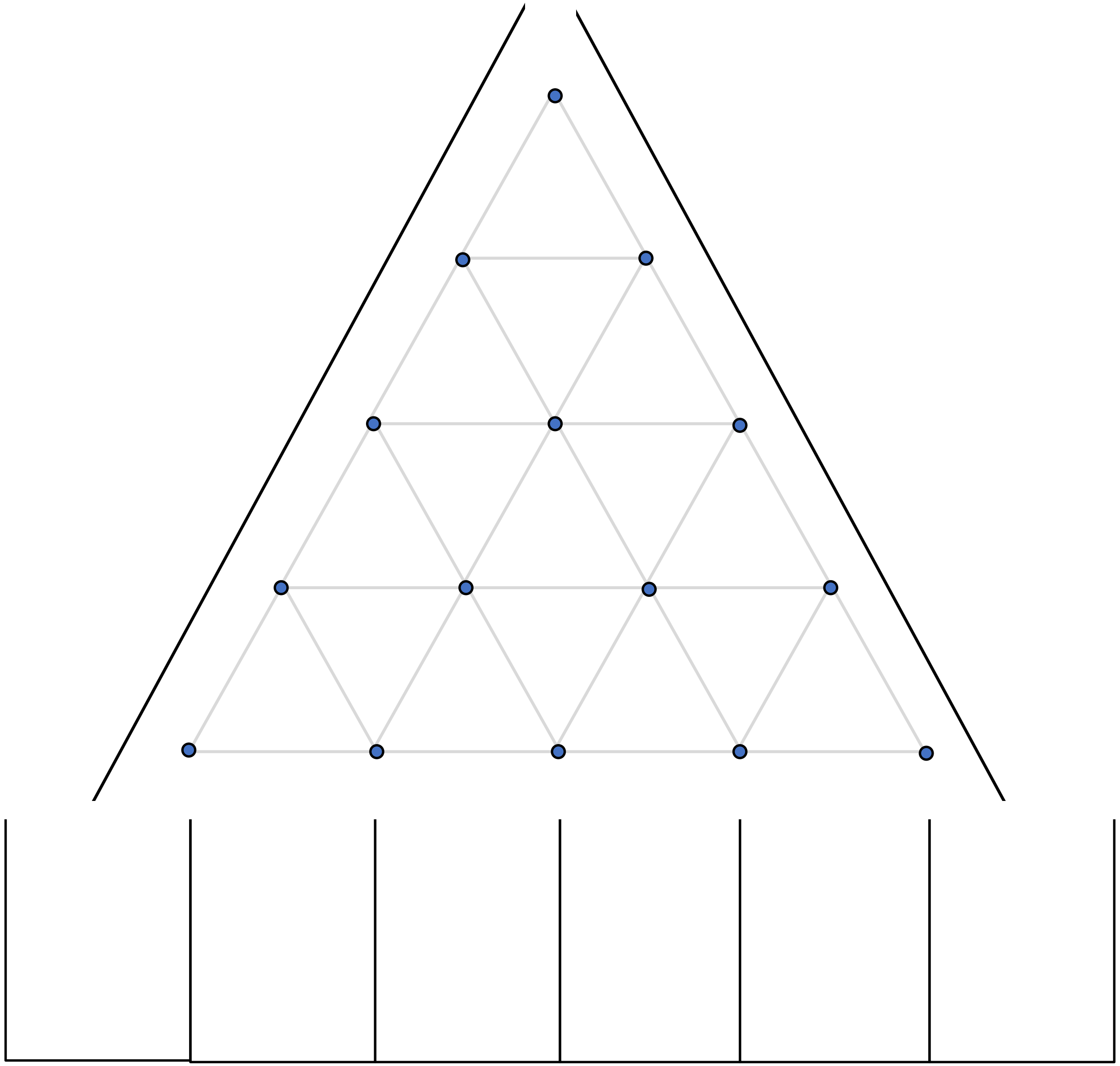
Parameter n: 1000 Parameter p: 0.01



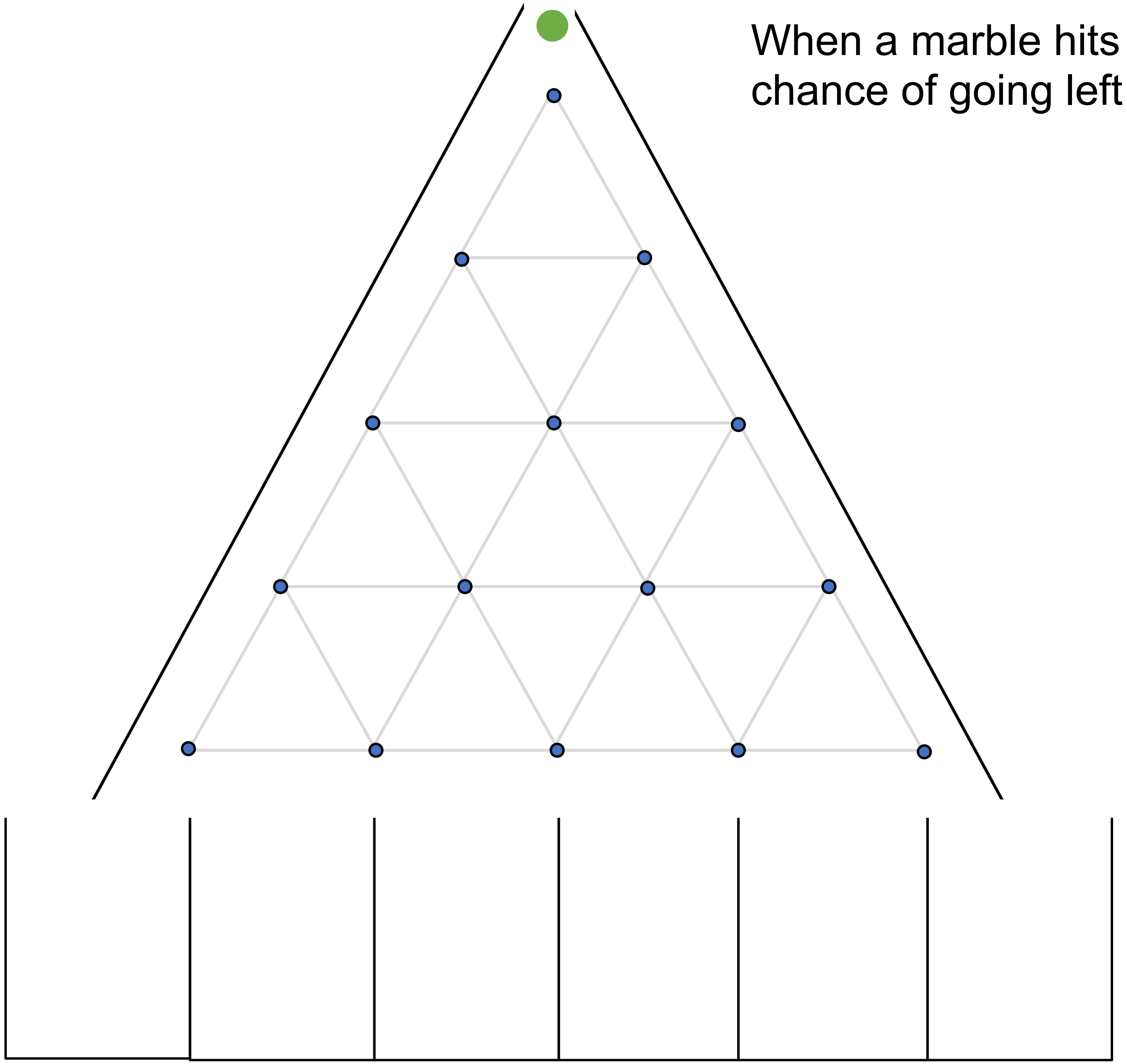
How Many Servers Crash?

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

Galton Board

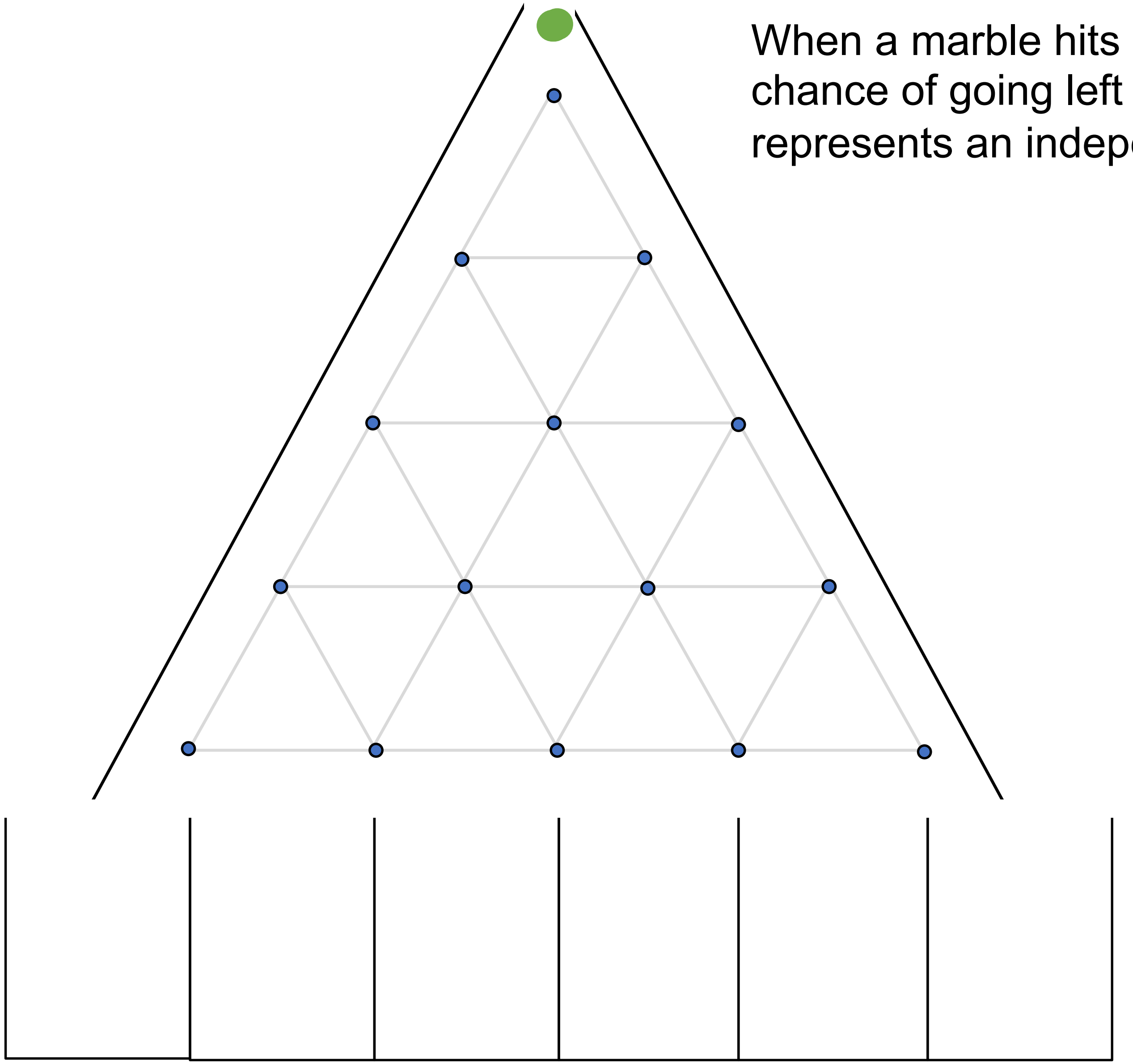


Galton Board



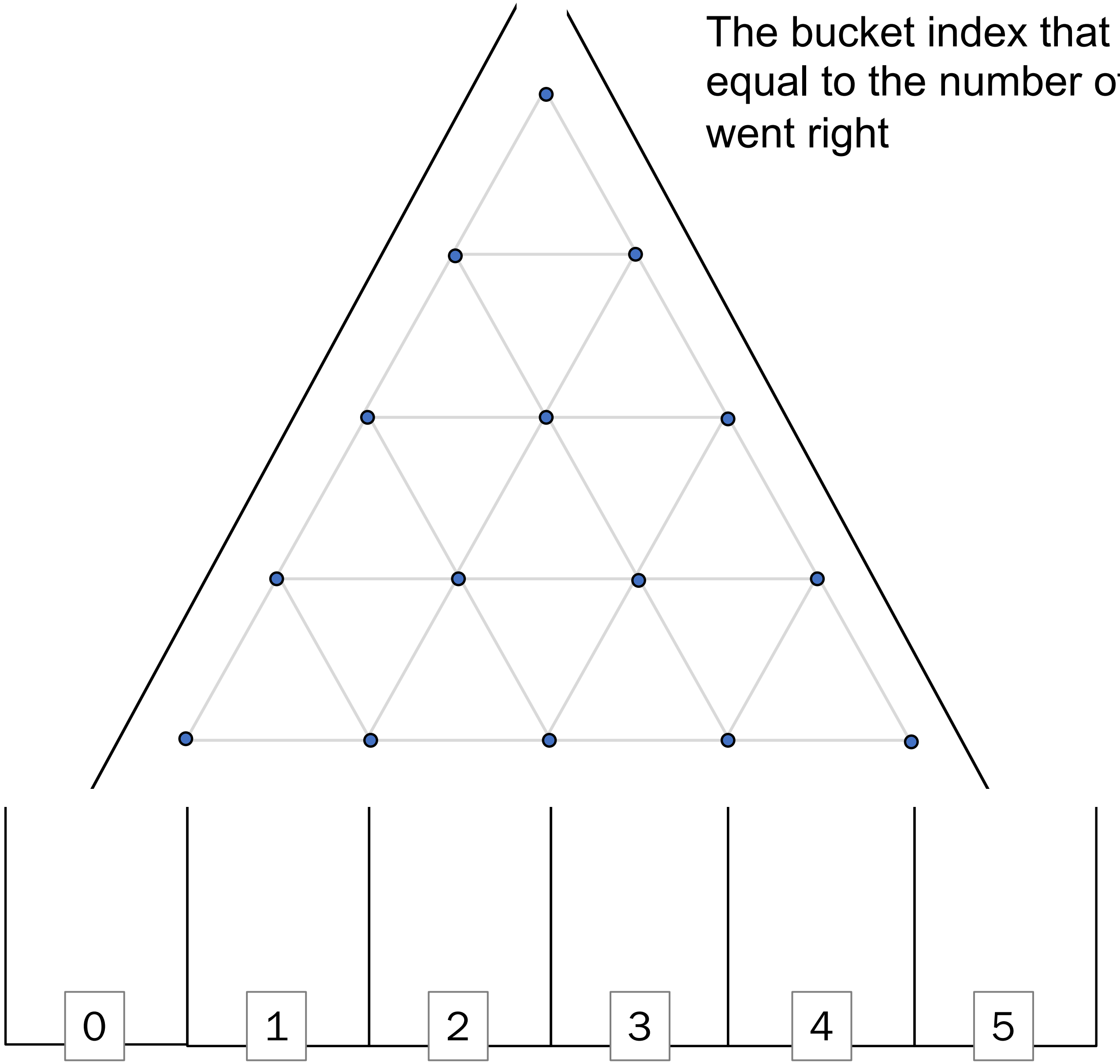
When a marble hits a pin, it has equal chance of going left or right.

Galton Board



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

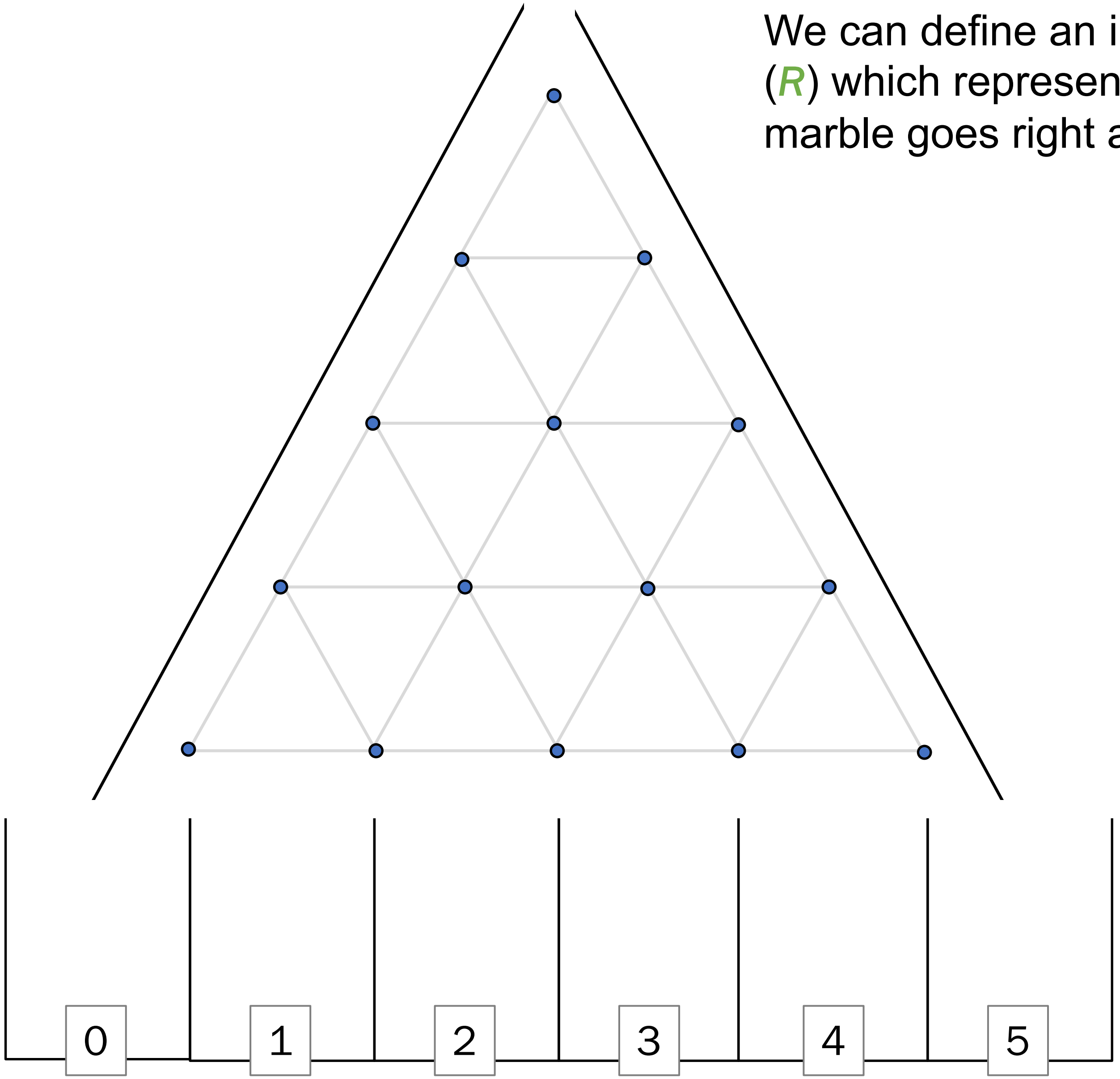
Galton Board



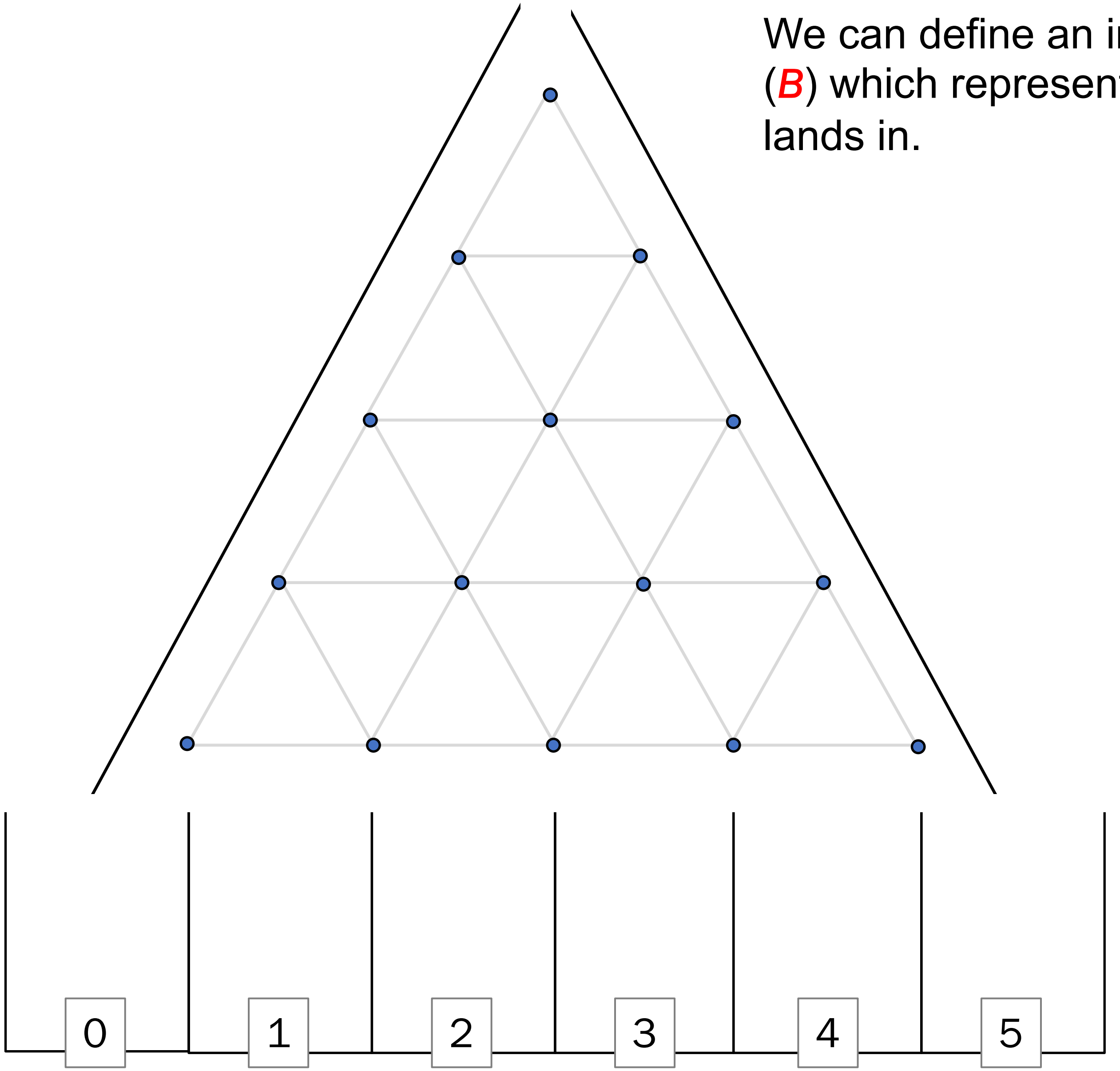
The bucket index that a marble lands in is equal to the number of times the marble went right

Galton Board

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

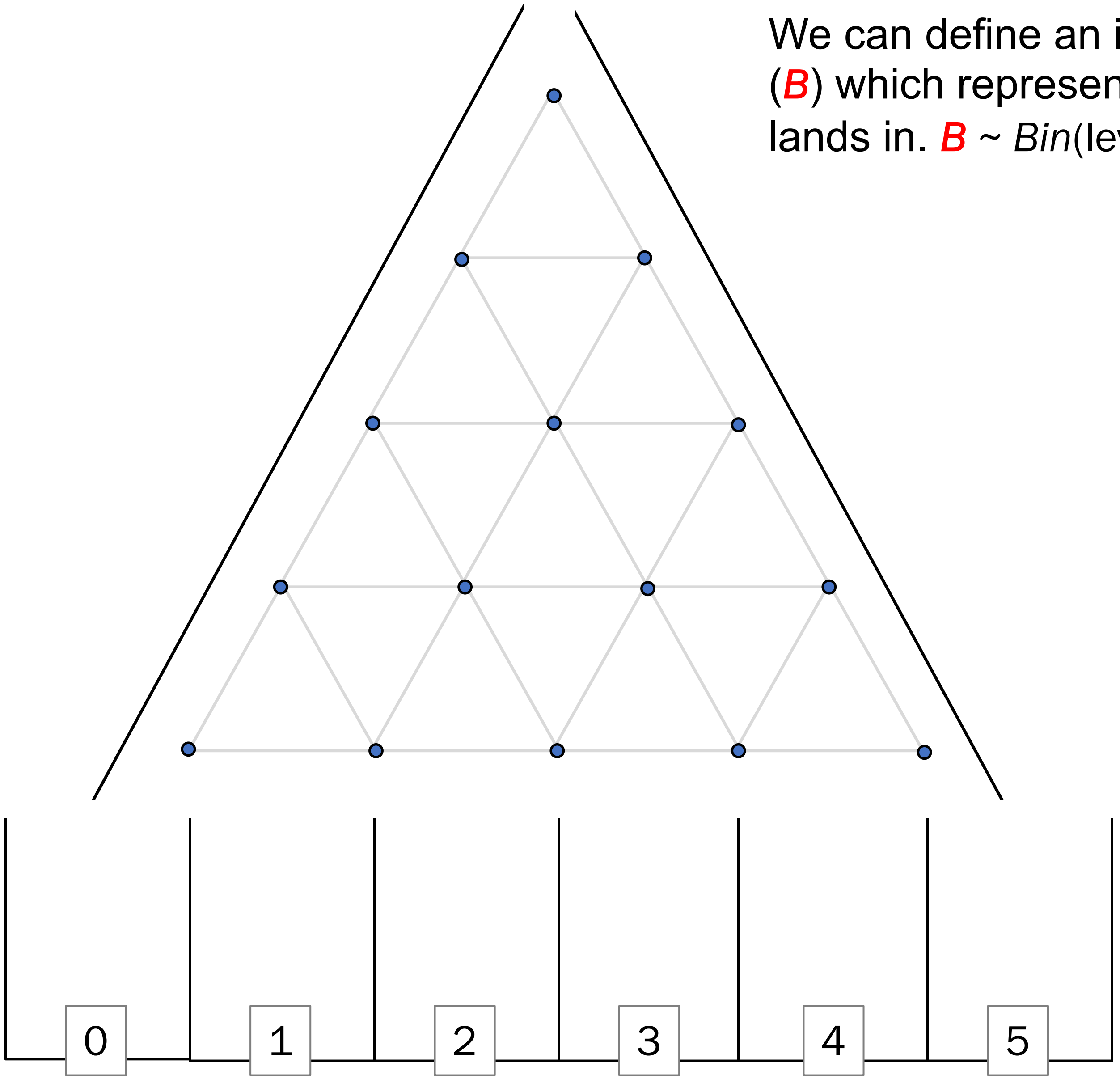


Galton Board



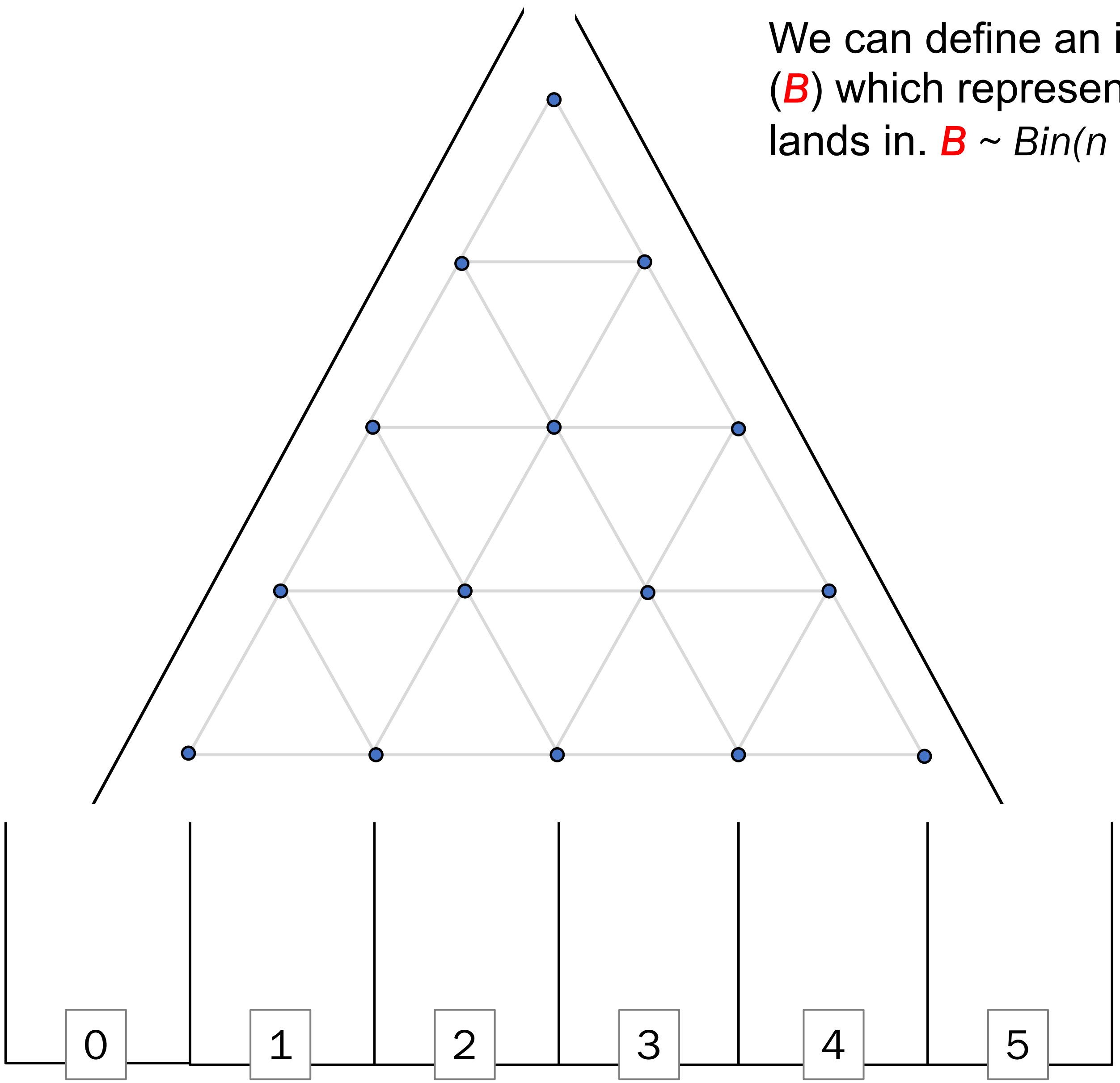
We can define an indicator random variable (*B*) which represents what bucket a marble lands in.

Galton Board



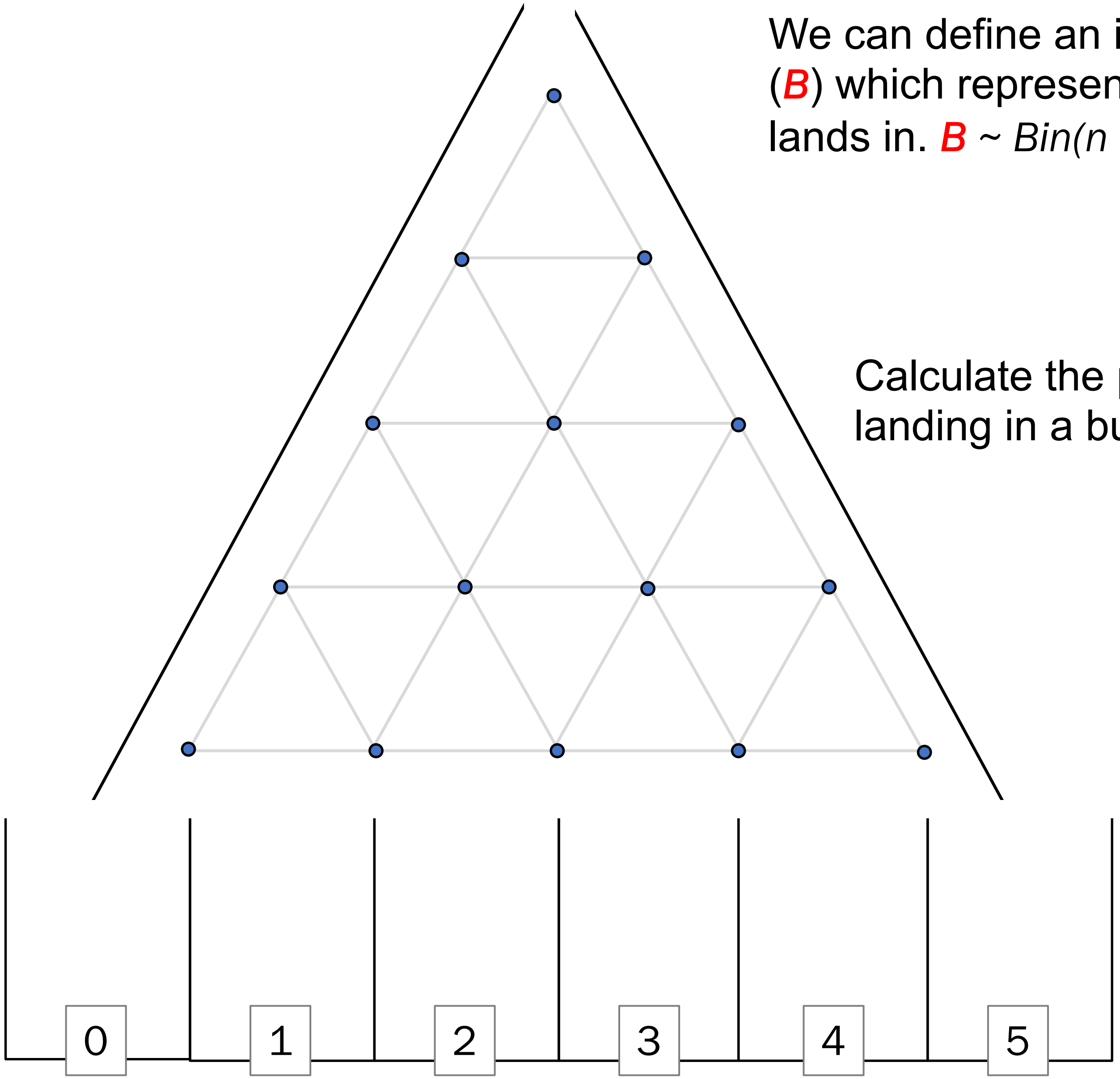
We can define an indicator random variable (*B*) which represents what bucket a marble lands in. *B* ~ *Bin*(levels, *0.5)*

Galton Board



We can define an indicator random variable (*B*) which represents what bucket a marble lands in. *B* ~ *Bin(n = 5, p = 0.5)*

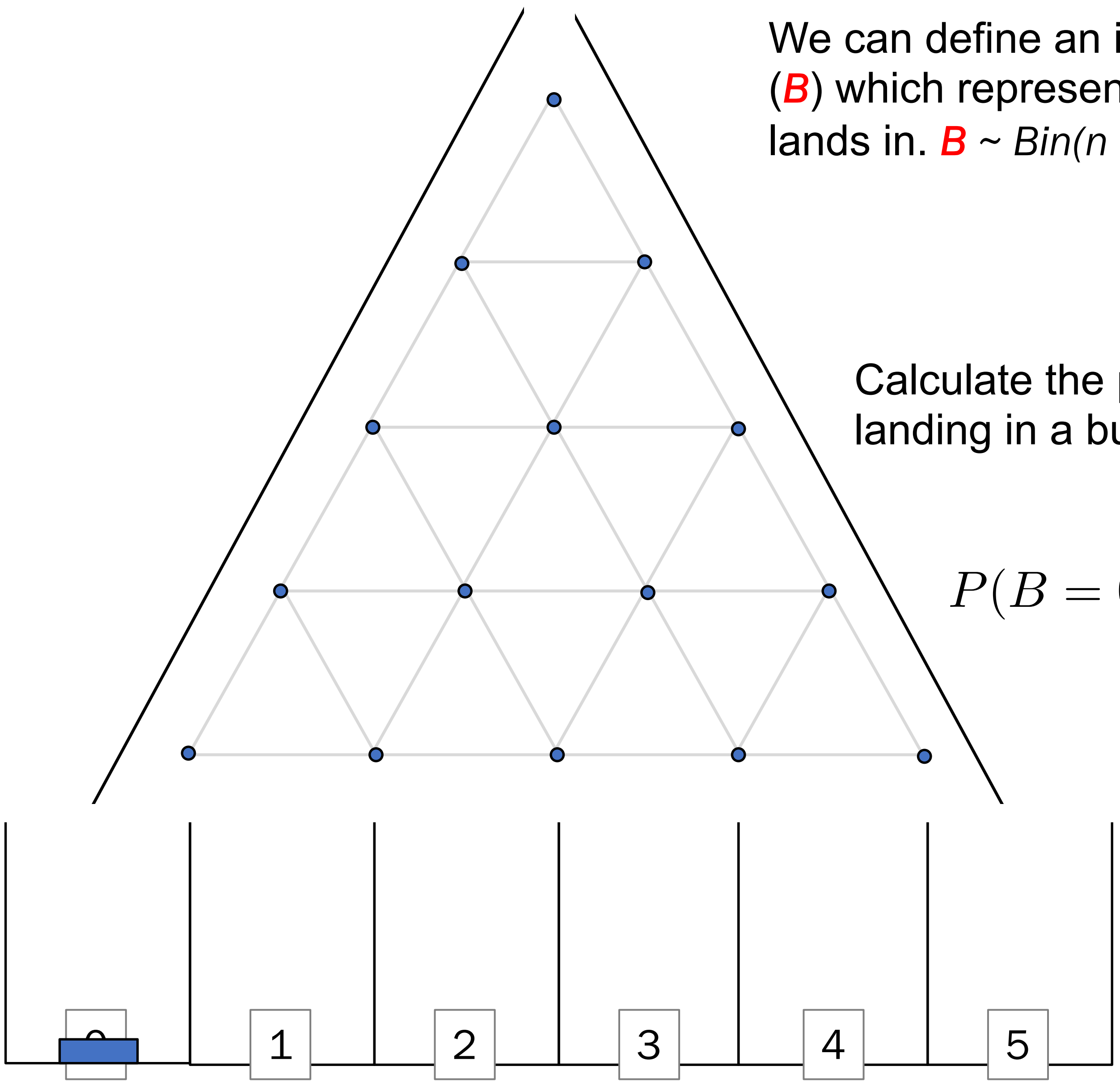
Galton Board



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

Calculate the probability of a marble landing in a bucket.

Galton Board

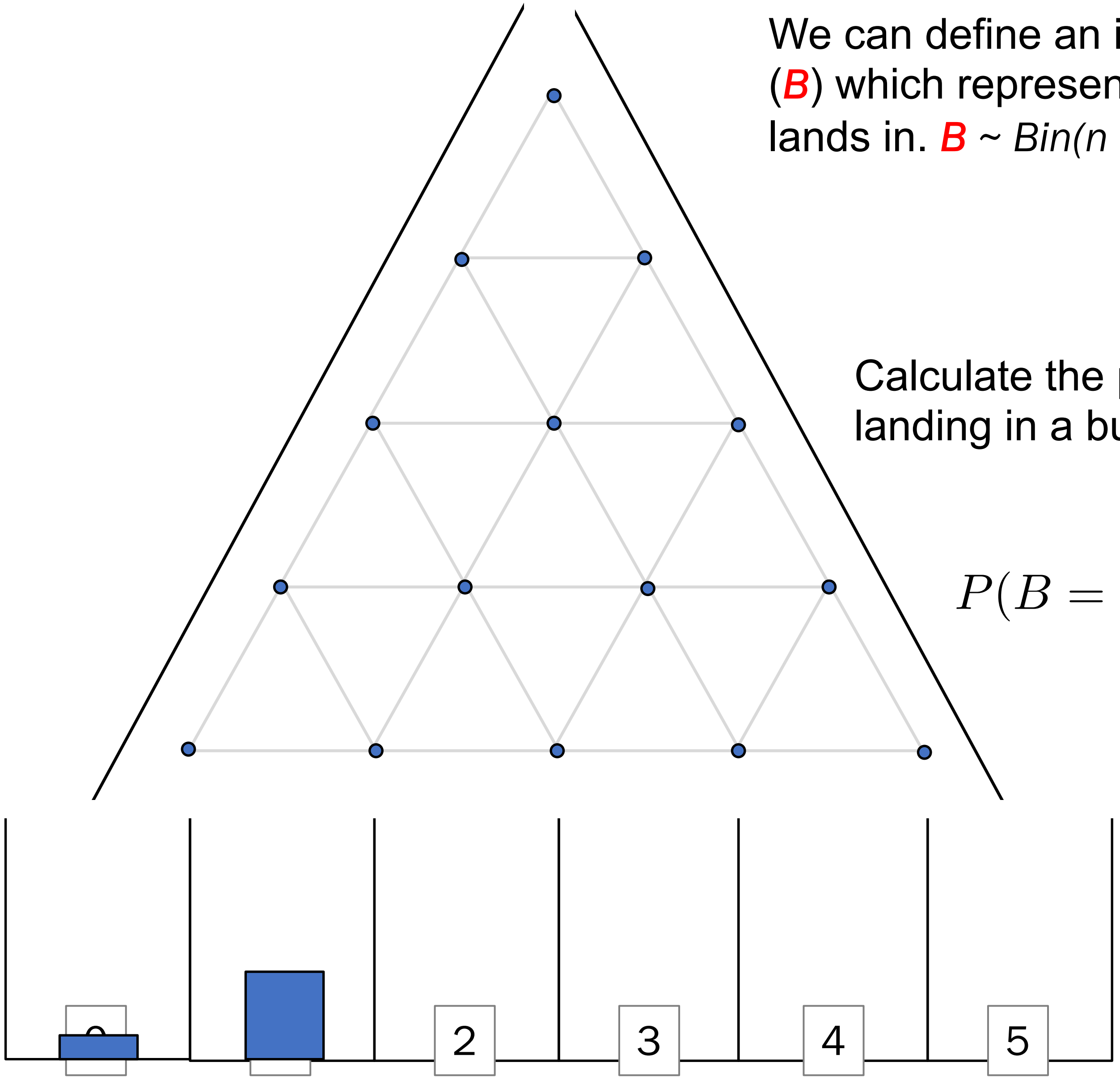


We can define an indicator random variable (*B*) which represents what bucket a marble lands in. *B* ~ *Bin(n = 5, p = 0.5)*

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

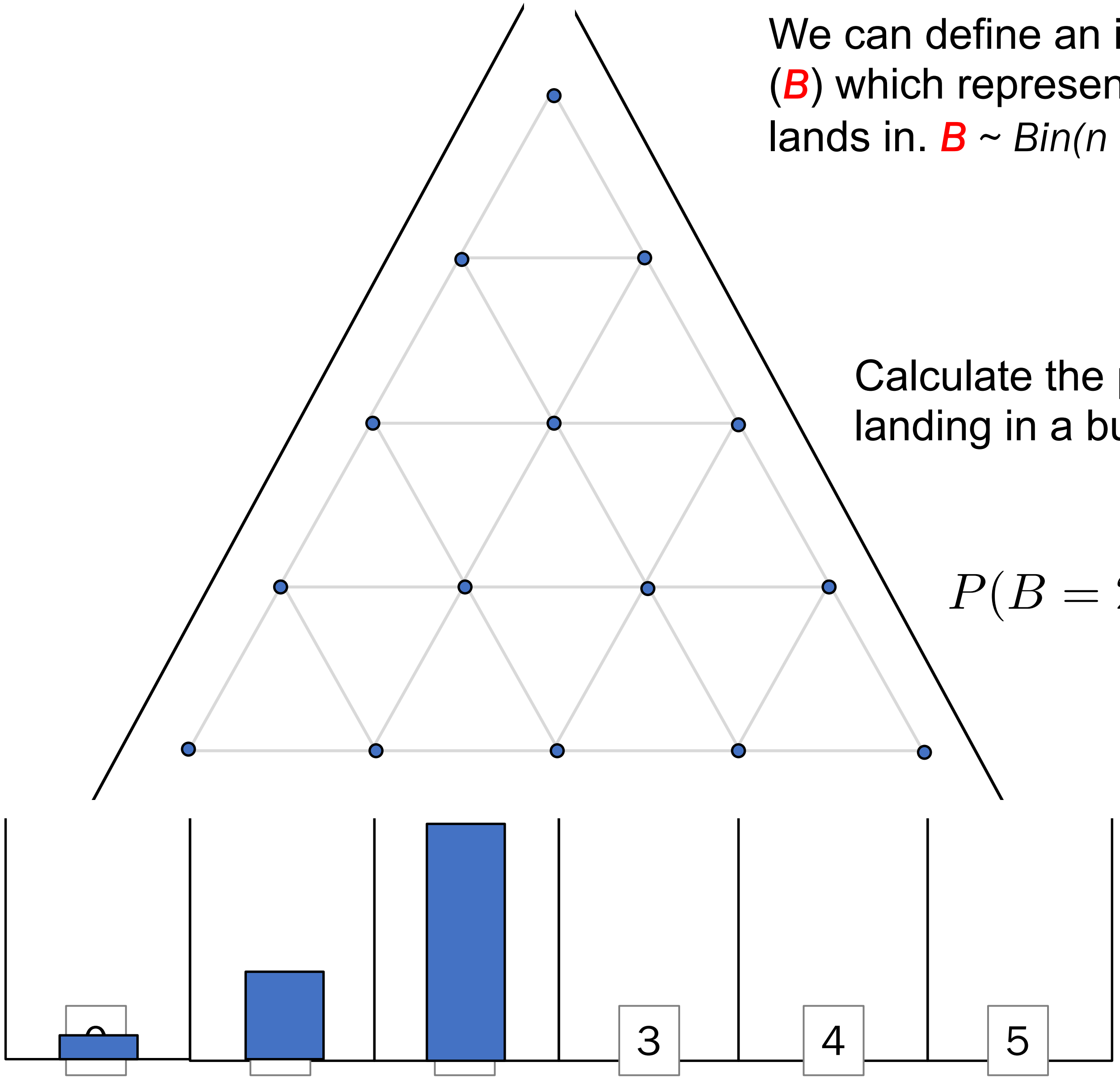


We can define an indicator random variable (*B*) which represents what bucket a marble lands in. *B ~ Bin(n = 5, p = 0.5)*

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

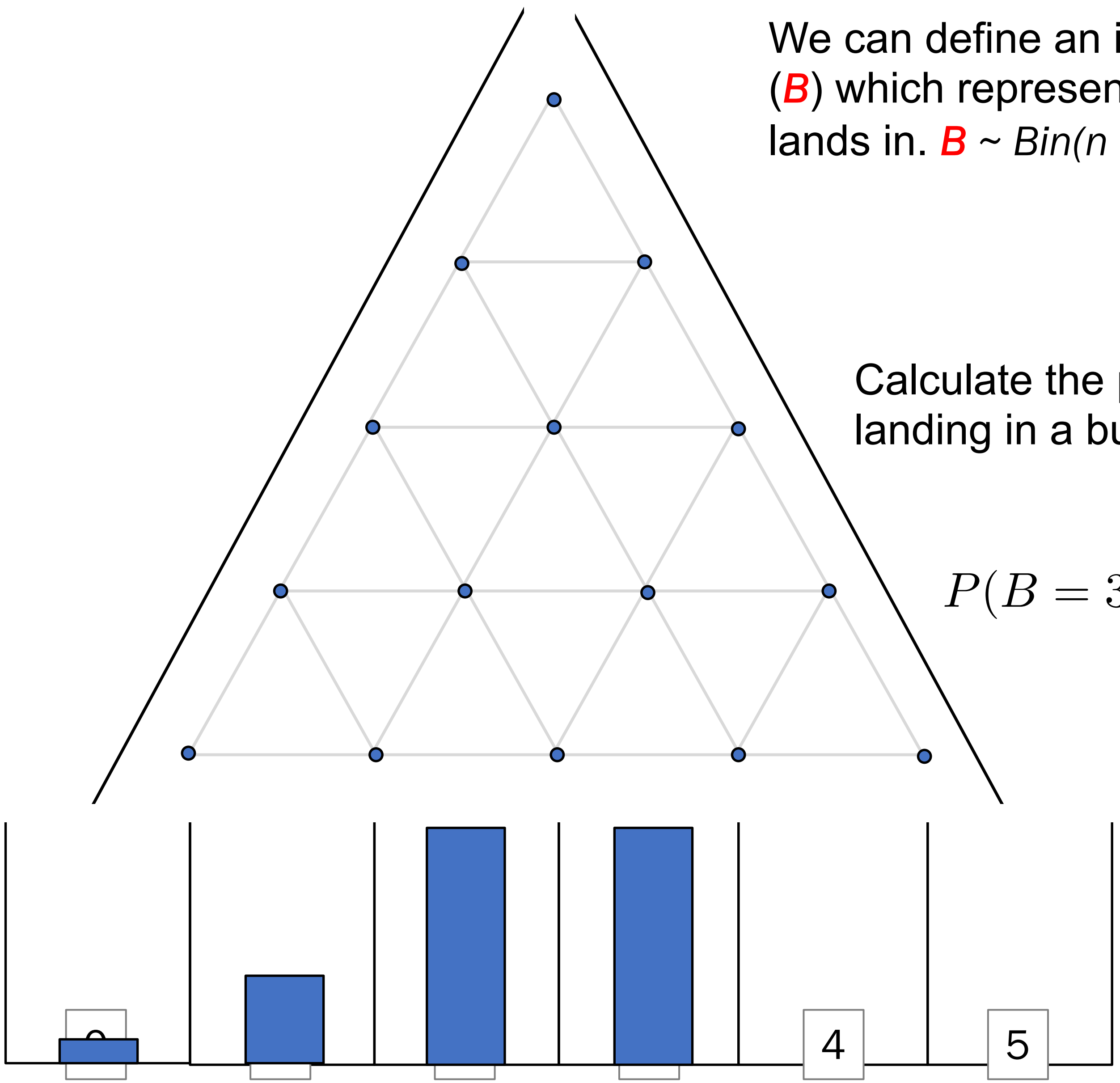


We can define an indicator random variable (*B*) which represents what bucket a marble lands in. *B* ~ *Bin(n = 5, p = 0.5)*

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

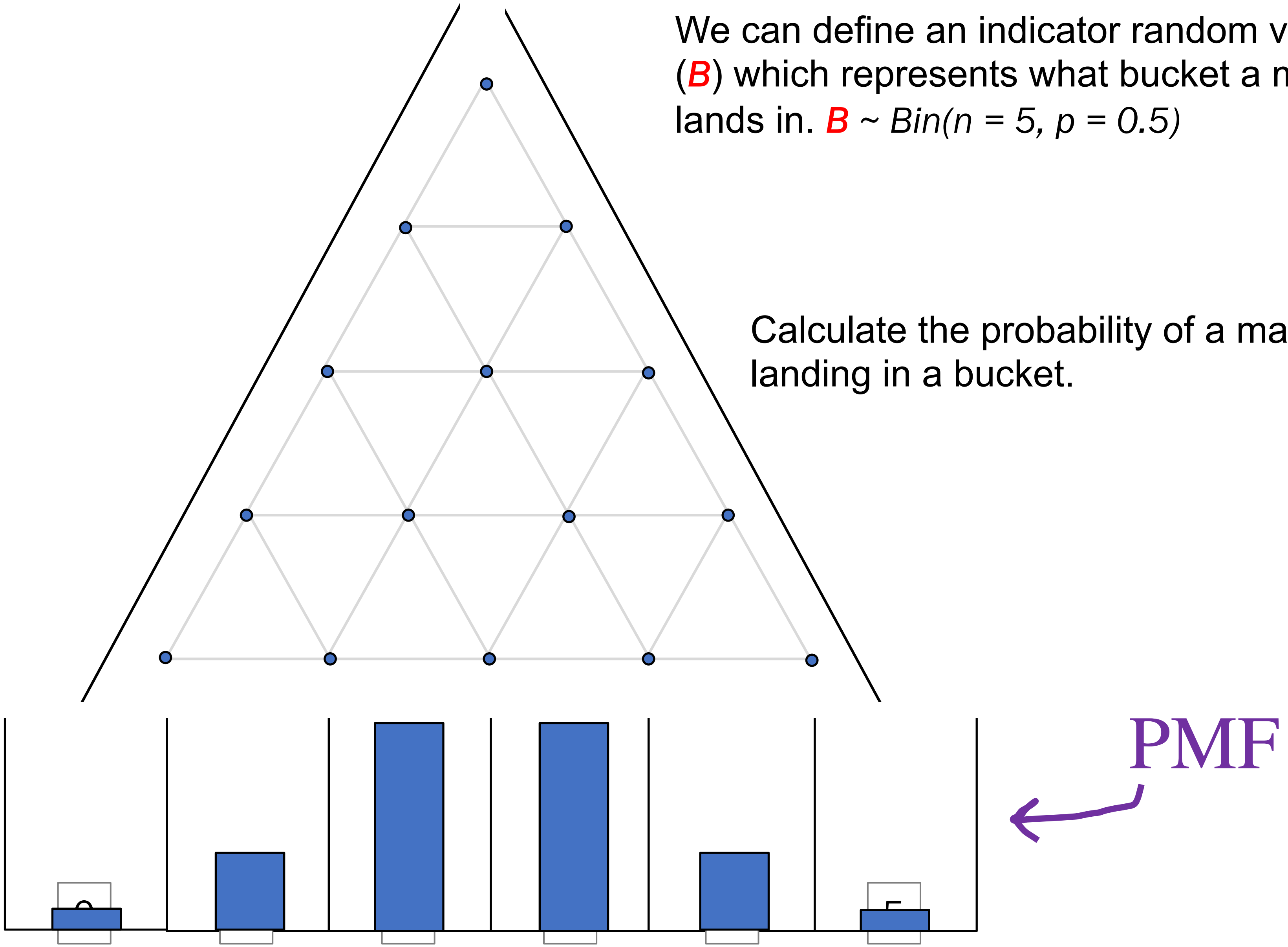


We can define an indicator random variable (*B*) which represents what bucket a marble lands in. *B* ~ *Bin(n = 5, p = 0.5)*

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* ~ $Bin(n = 5, p = 0.5)$

Calculate the probability of a marble landing in a bucket.

From Chaos to Order

What is the probability of winning a 7 game series?

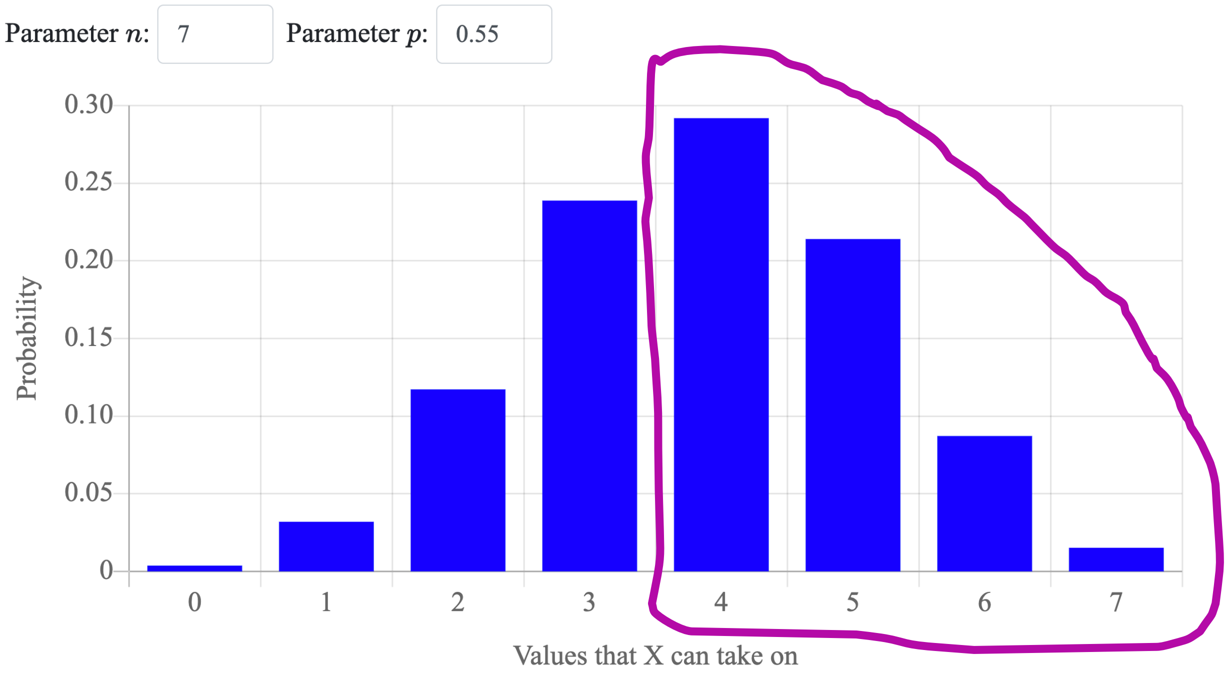
Warriors are going to play the Bucks in a best of 7 series during the 2022 NBA finals. What is the probability that the warriors win the series? Each game is **independent**. Each game, the warriors have a 0.55 probability of winning. Win series if you win at least 4 games.

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 ~ Bin(n= 7, p = 0.55). P(X > 3)?

$$P(X \geq 4) = \sum_{i=4}^{7} P(X = i)$$

$$= \sum_{i=4}^{7} \binom{7}{i} p^i (1-p)^{7-i}$$

$$= \sum_{i=4}^{7} \binom{7}{i} 0.55^i (0.45)^{7-i}$$

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*
- $P(X \geq k) =$

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

Chose slots for success, don't care about rest

ways to choose slots for success

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*
- $P(X \geq k) =$

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

Chose slots for success, don't care about rest

ways to choose slots for success

Probability that each is success

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

Not mutually exclusive…

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

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

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

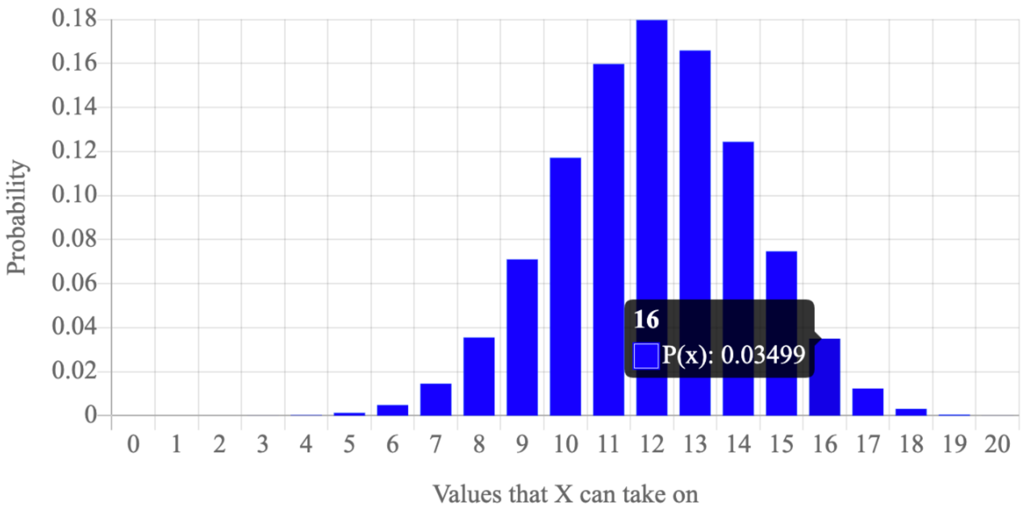
PMF equation: $\Pr(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$

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

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

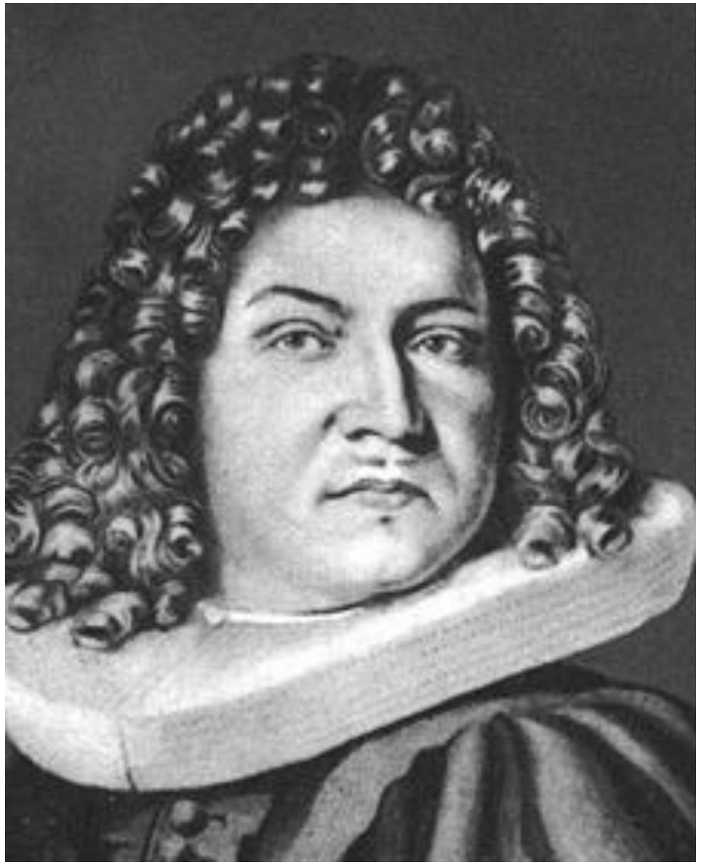
PMF graph:

Parameter n: 20 Parameter p: 0.60



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 ~ Bern(p)
- E[X] = p

Feel the Bern!

Examples

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

Does a Program Crash?

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

X: 1 if program crashes

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

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

Does a User Click an Ad?

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

C: 1 if ad is clicked

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

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

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. $Y_i \sim \text{Bern}(p)$

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

$$= \sum_{i=1}^{n} E[Y_i] \qquad \text{Expectation of sum}$$

$$= \sum_{i=1}^{n} p \qquad \text{Expectation of Bernoulli}$$

$$= n \cdot p \qquad \text{Sum } n \text{ times}$$

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

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

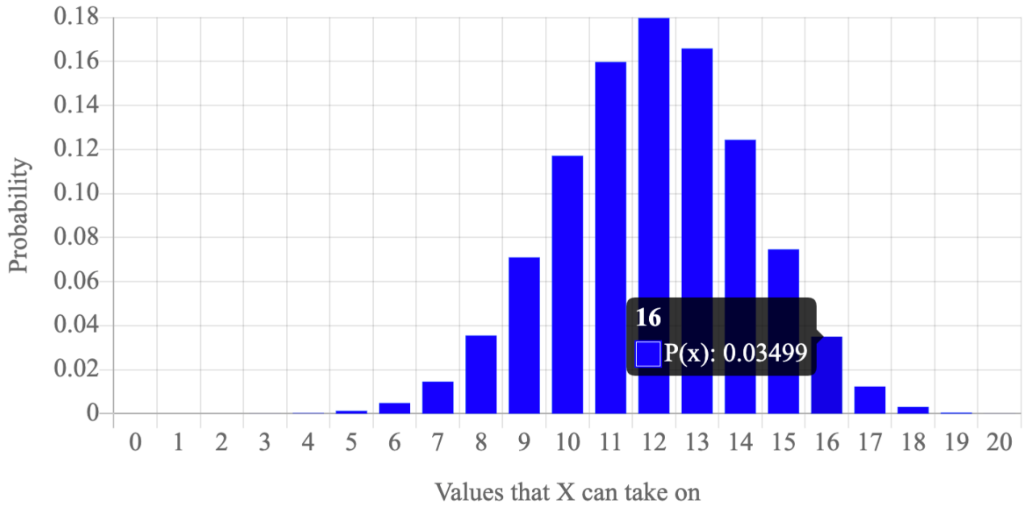
PMF equation: $\Pr(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$

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

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

PMF graph:

Parameter n: 20 Parameter p: 0.60



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

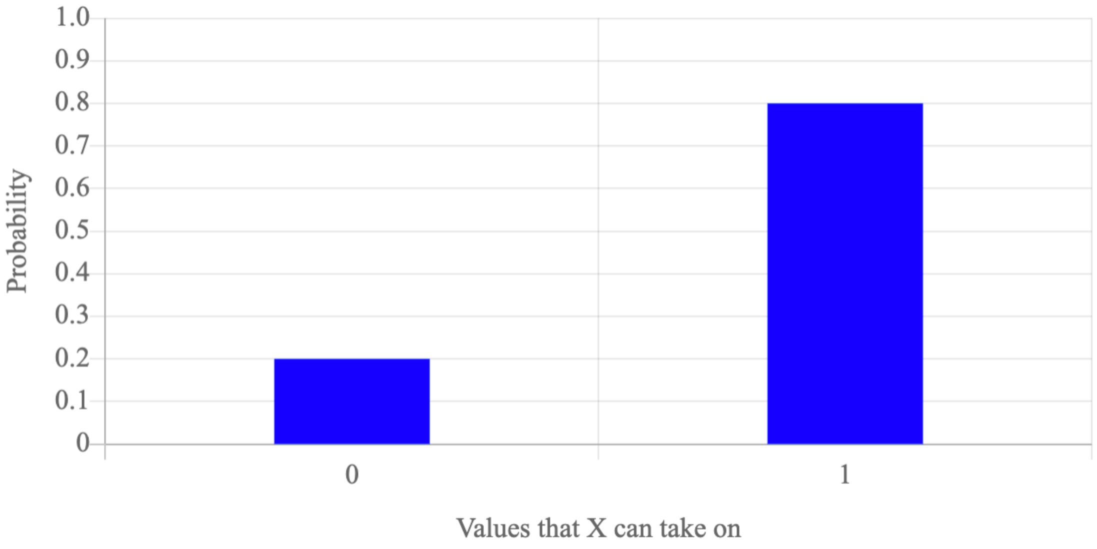
PMF equation: $\Pr(X = x) = \begin{cases} p & \text{if } x = 1 \\ 1-p & \text{if } x = 0 \end{cases}$

Expectation: $\text{E}[X] = p$

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

PMF graph:

Parameter p: 0.80

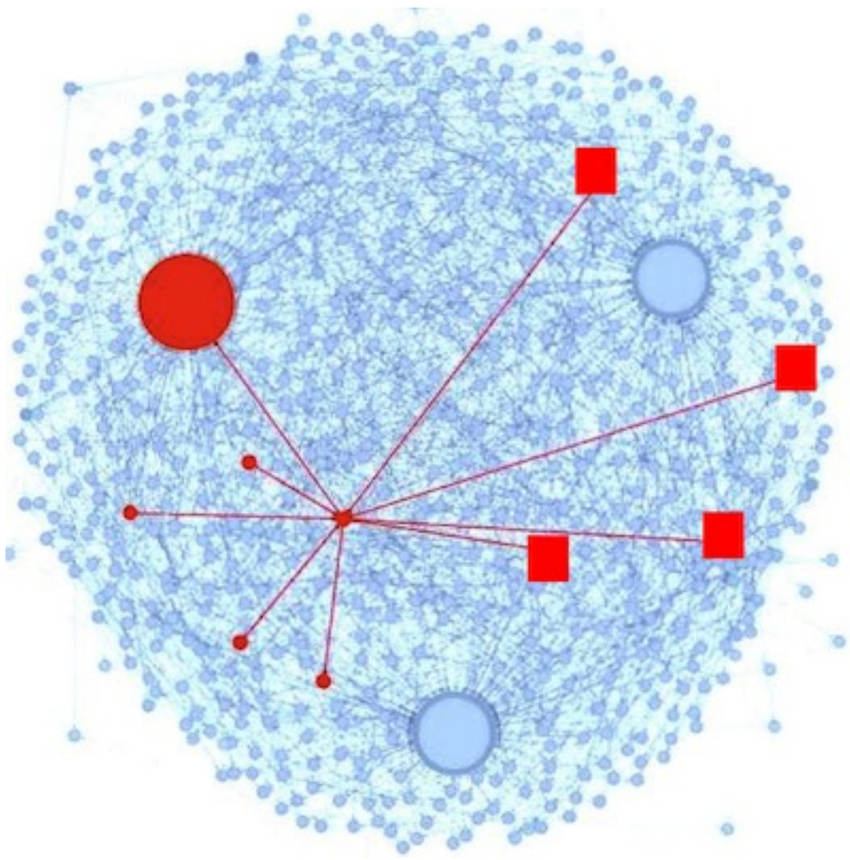


Expectation is a single number summary…

Expectation leaves much to be desired…

Can we invent *another* summary number?

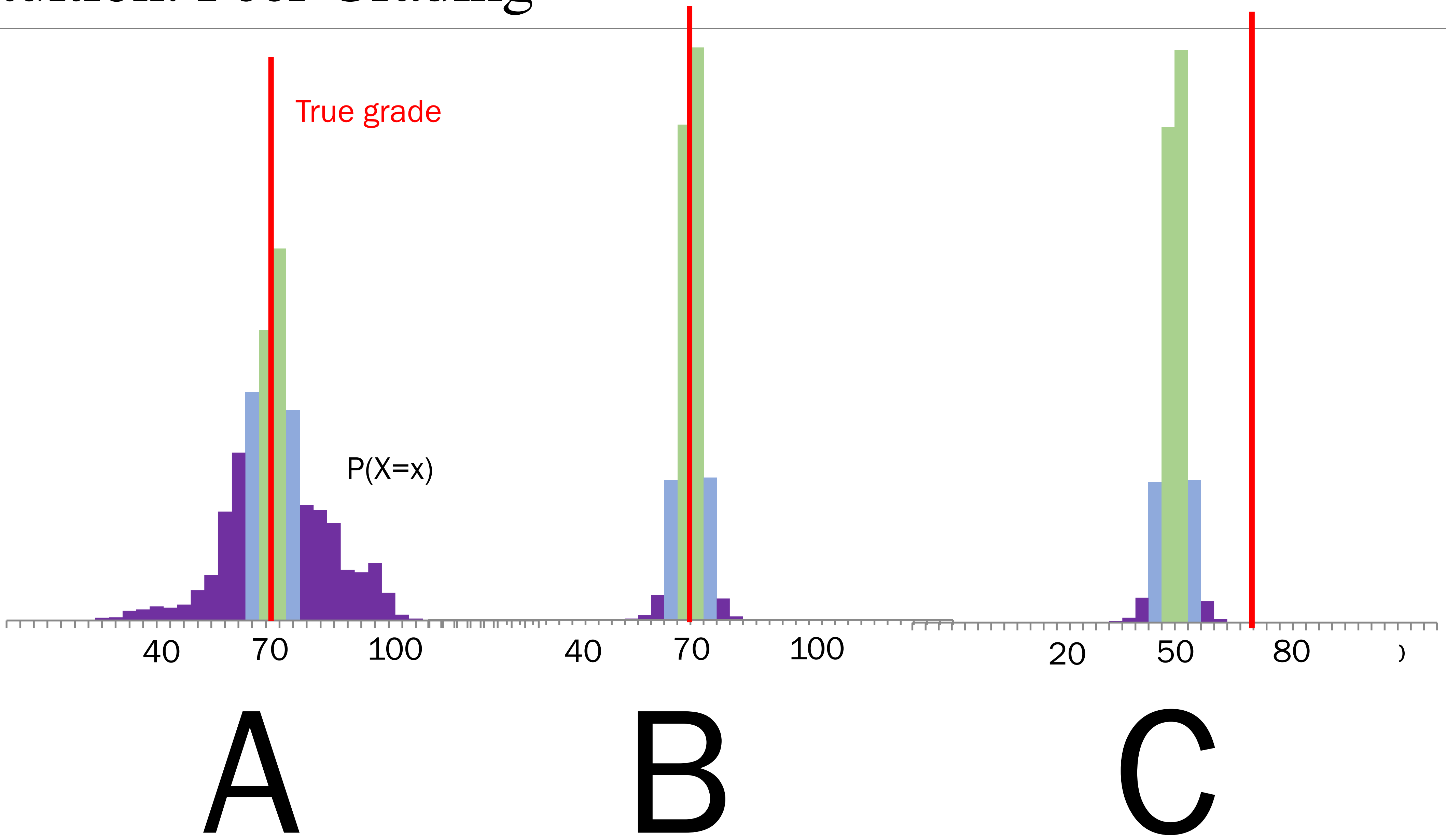
Intuition: Peer Grading



Peer Grading on Coursera HCI.

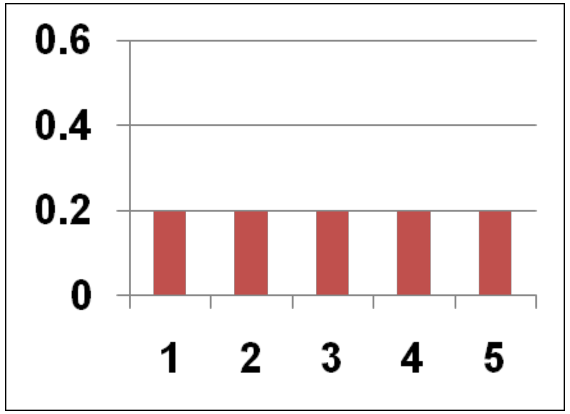
31,067 peer grades for 3,607 students.

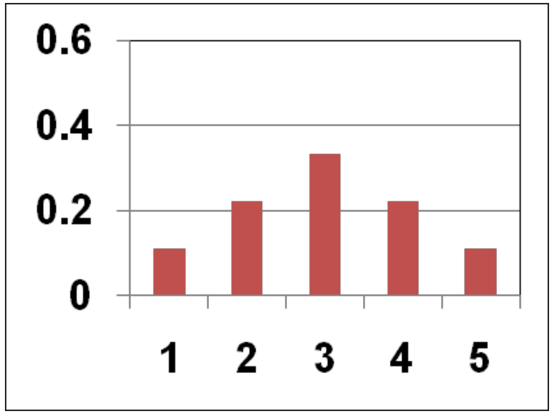
Intuition: Peer Grading

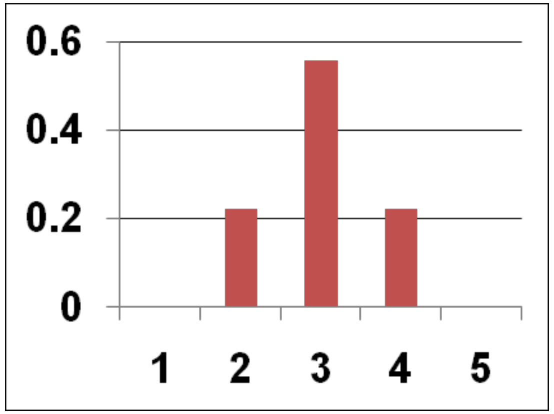


Intuition: Measure of Spread

Consider the following 3 distributions (PMFs)







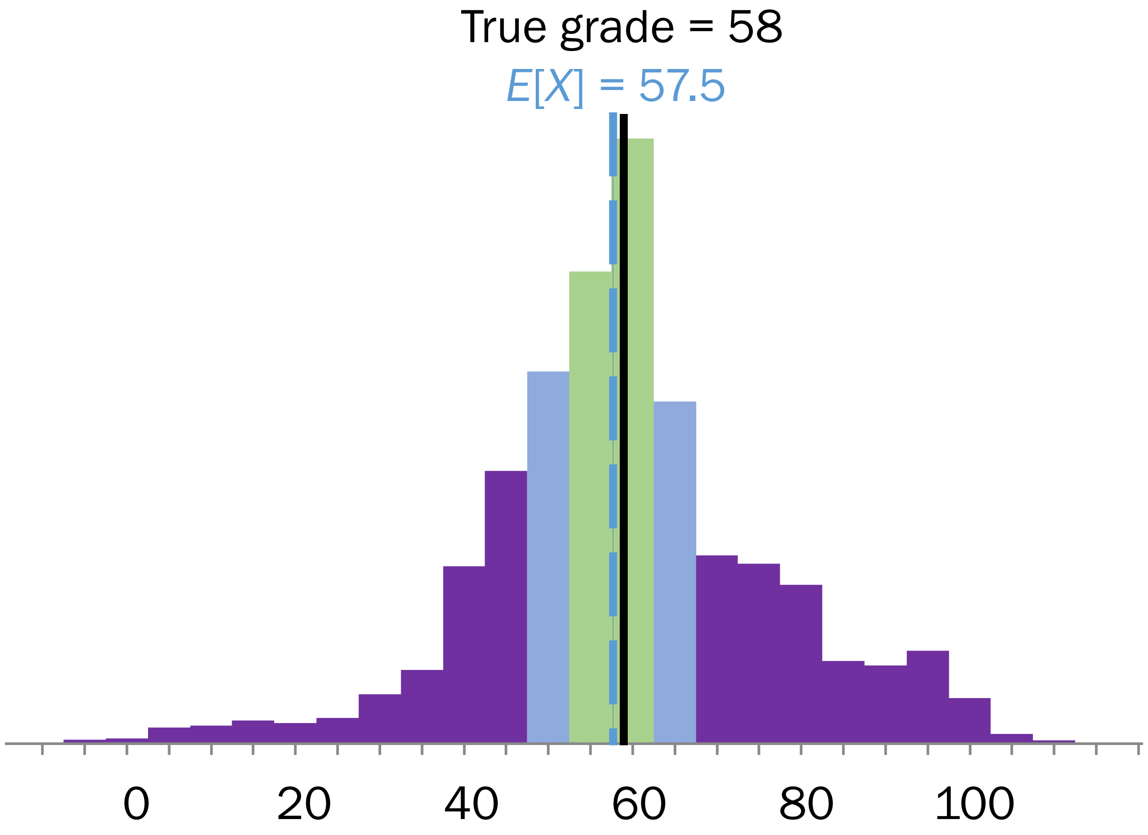
All have the same expected value, E[X] = 3

But “spread” in distributions is different

Invent a formal quantification of “spread”?

Peer grading in Coursera HCI

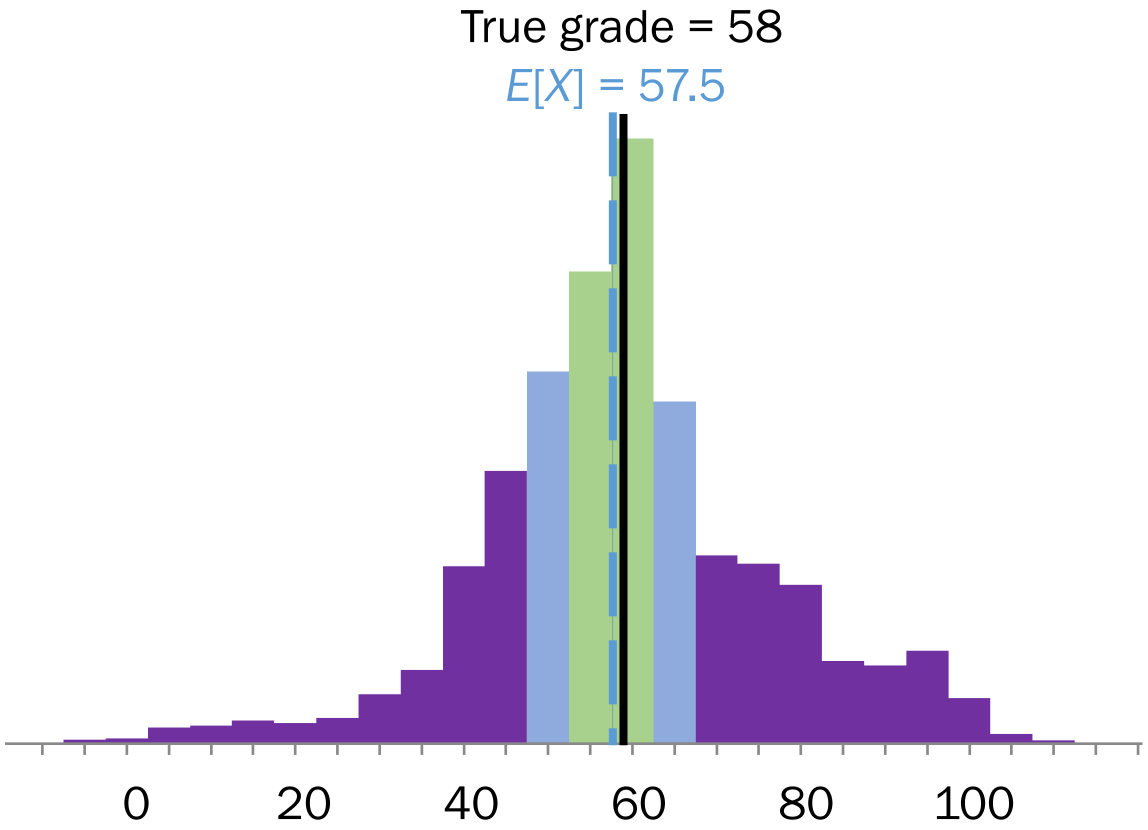
Let X be a random variable that represents a peer grade



Peer grading in Coursera HCI

Let X be a random variable that represents a peer grade

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



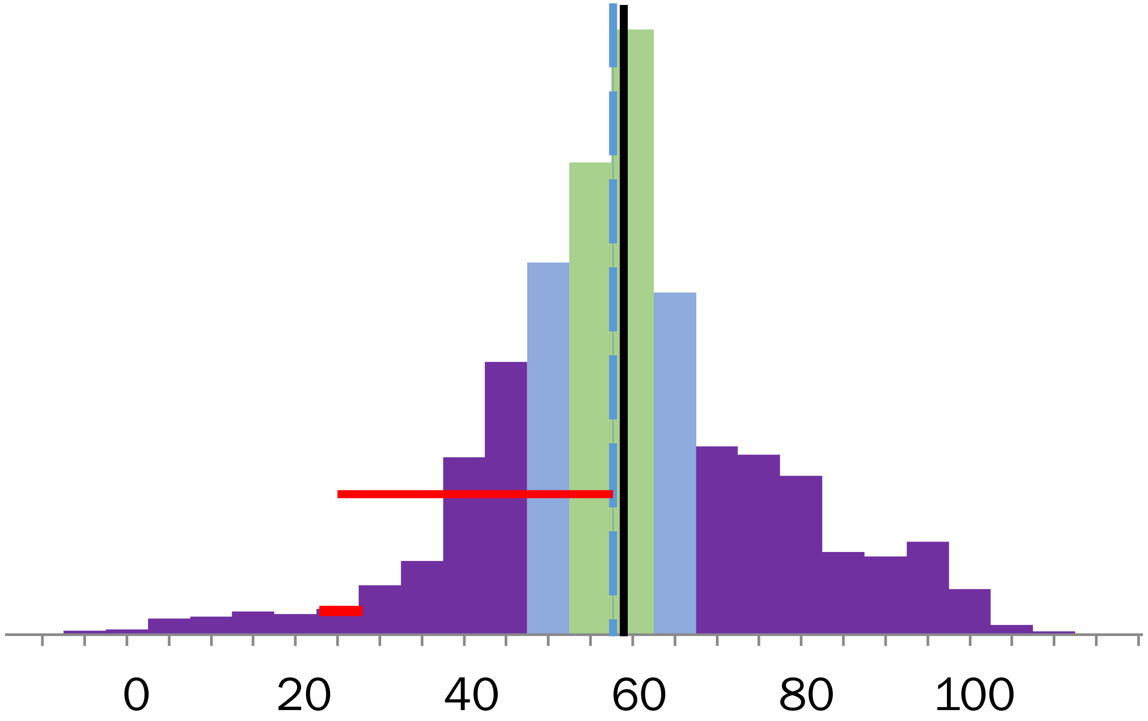
Peer grading in Coursera HCI

Let X be a random variable that represents a peer grade

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

True grade = 58

$E[X] = 57.5$

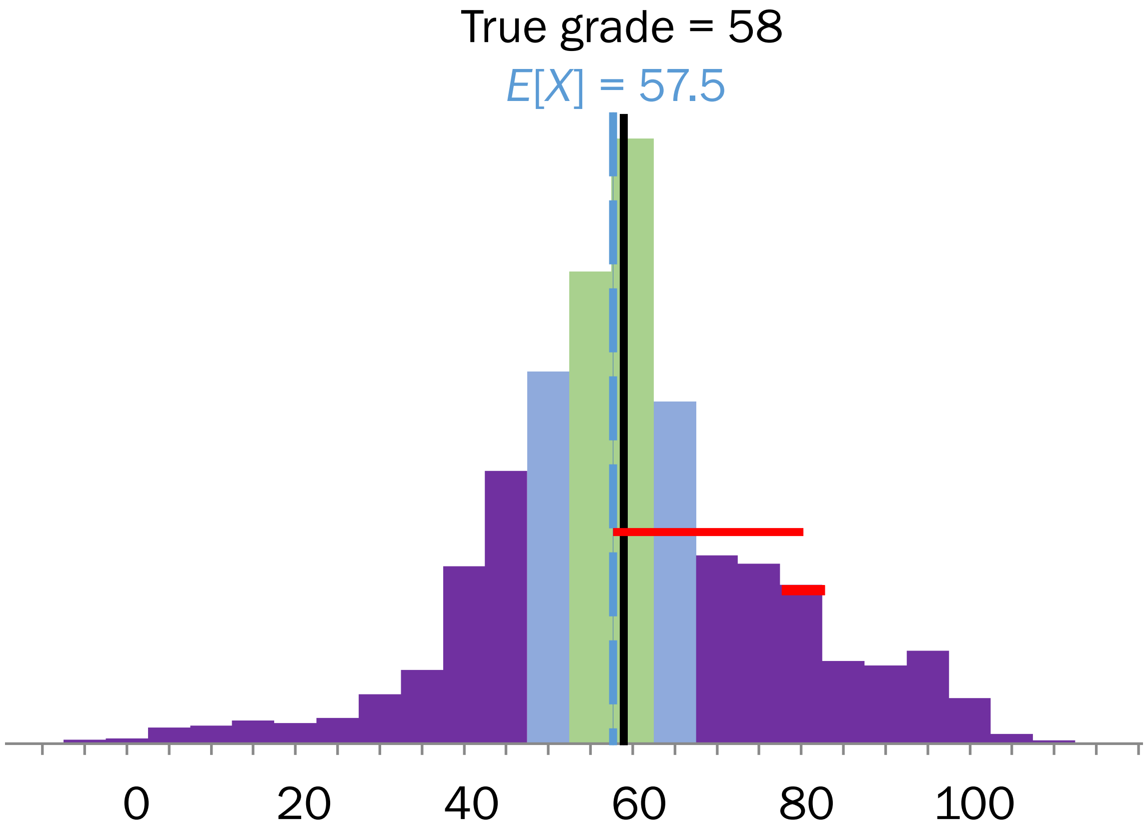


X	$(X - \mu)^2$
25 points	1056 points^2

Peer grading in Coursera HCI

Let X be a random variable that represents a peer grade

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

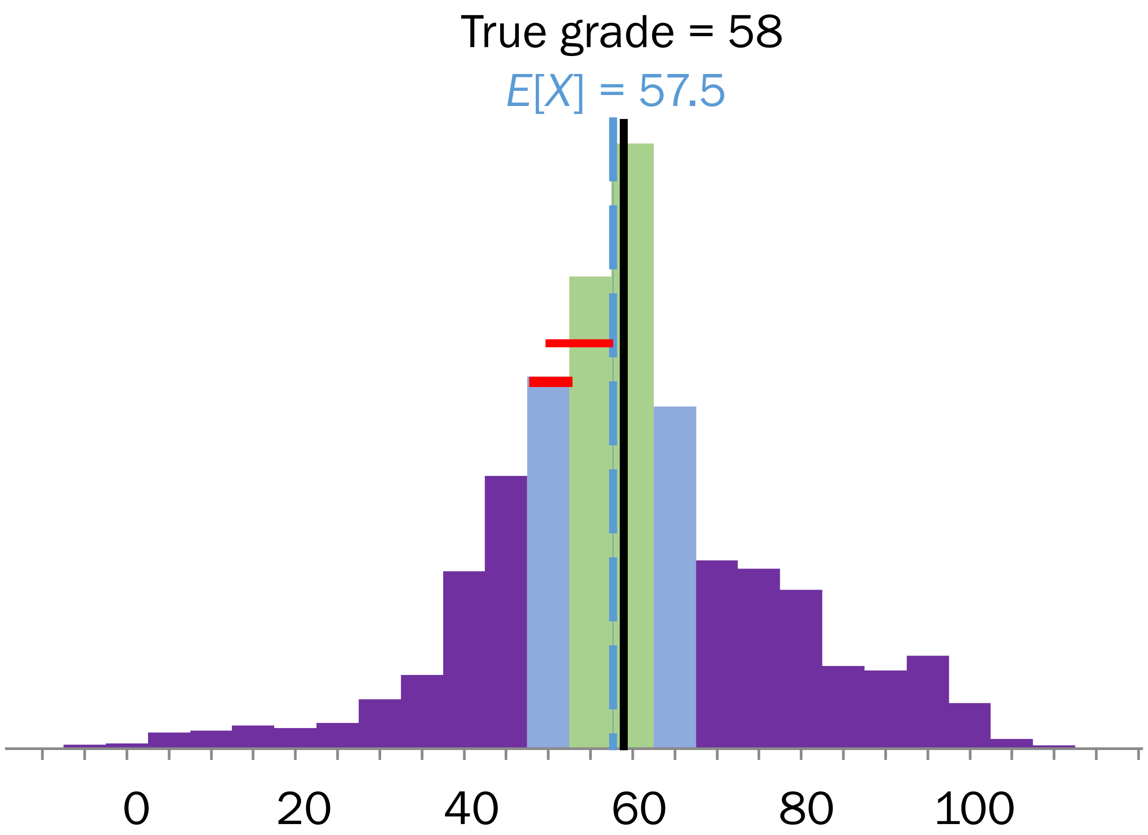


X	$(X - \mu)^2$
25 points	1056 points^2
80 points	506 points^2

Peer grading in Coursera HCI

Let *X* be a random variable that represents a peer grade

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

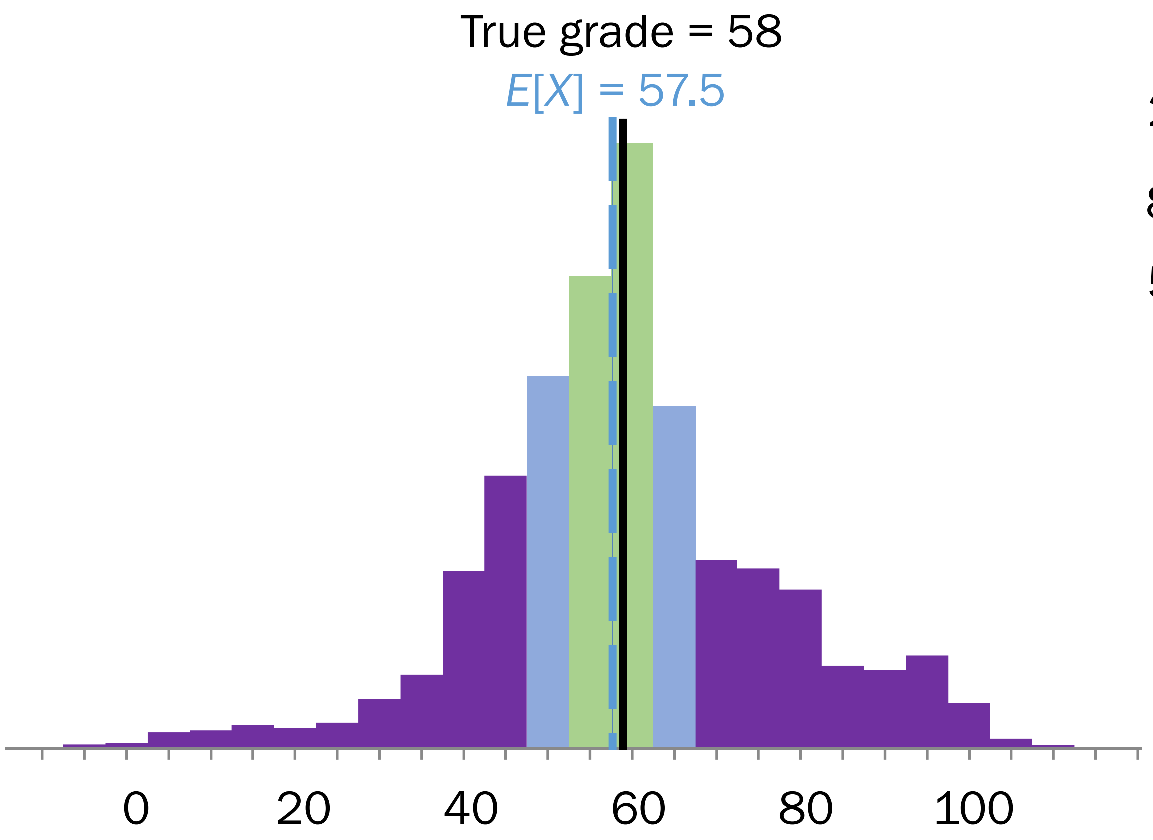


X	$(X - \mu)^2$
25 points	1056 points^2
80 points	506 points^2
50 points	56 points^2

Peer grading in Coursera HCI

Let X be a random variable that represents a peer grade

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



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

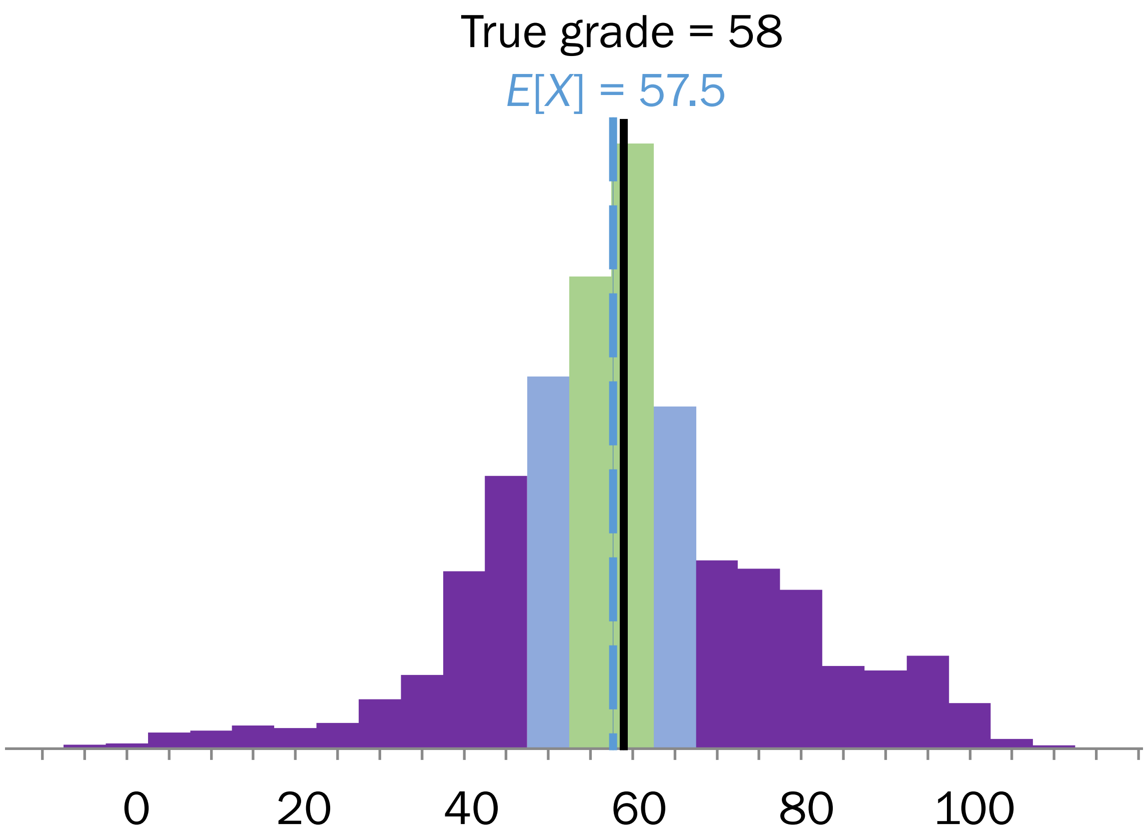
…

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

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

...

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

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

Variance

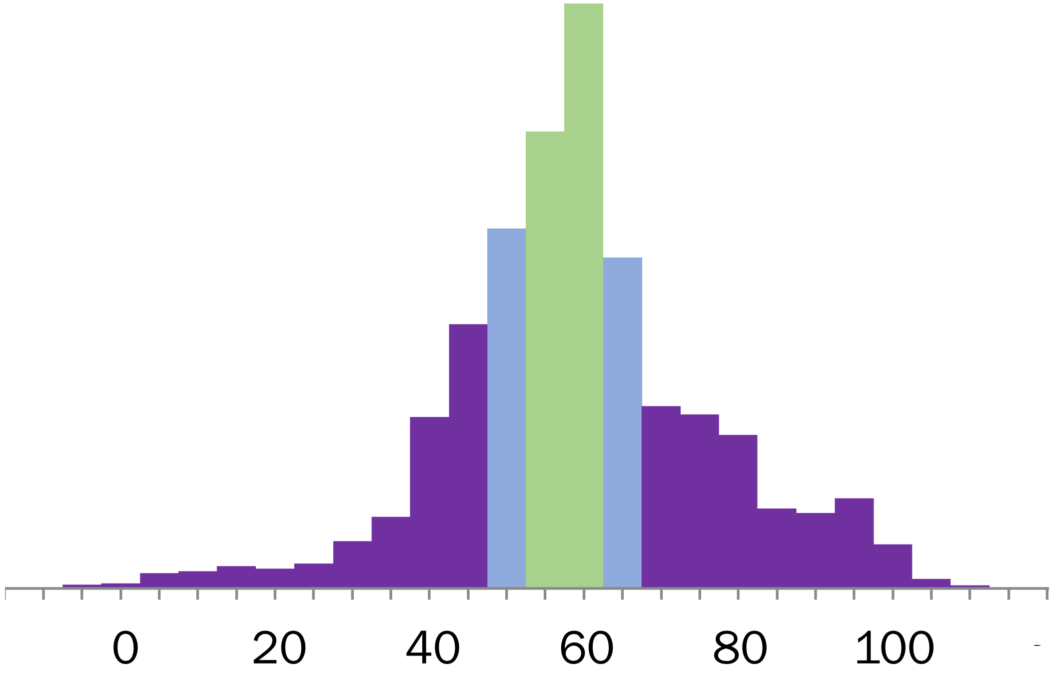
If X is a random variable with mean μ then the **variance** of X, denoted Var(X), is:

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

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

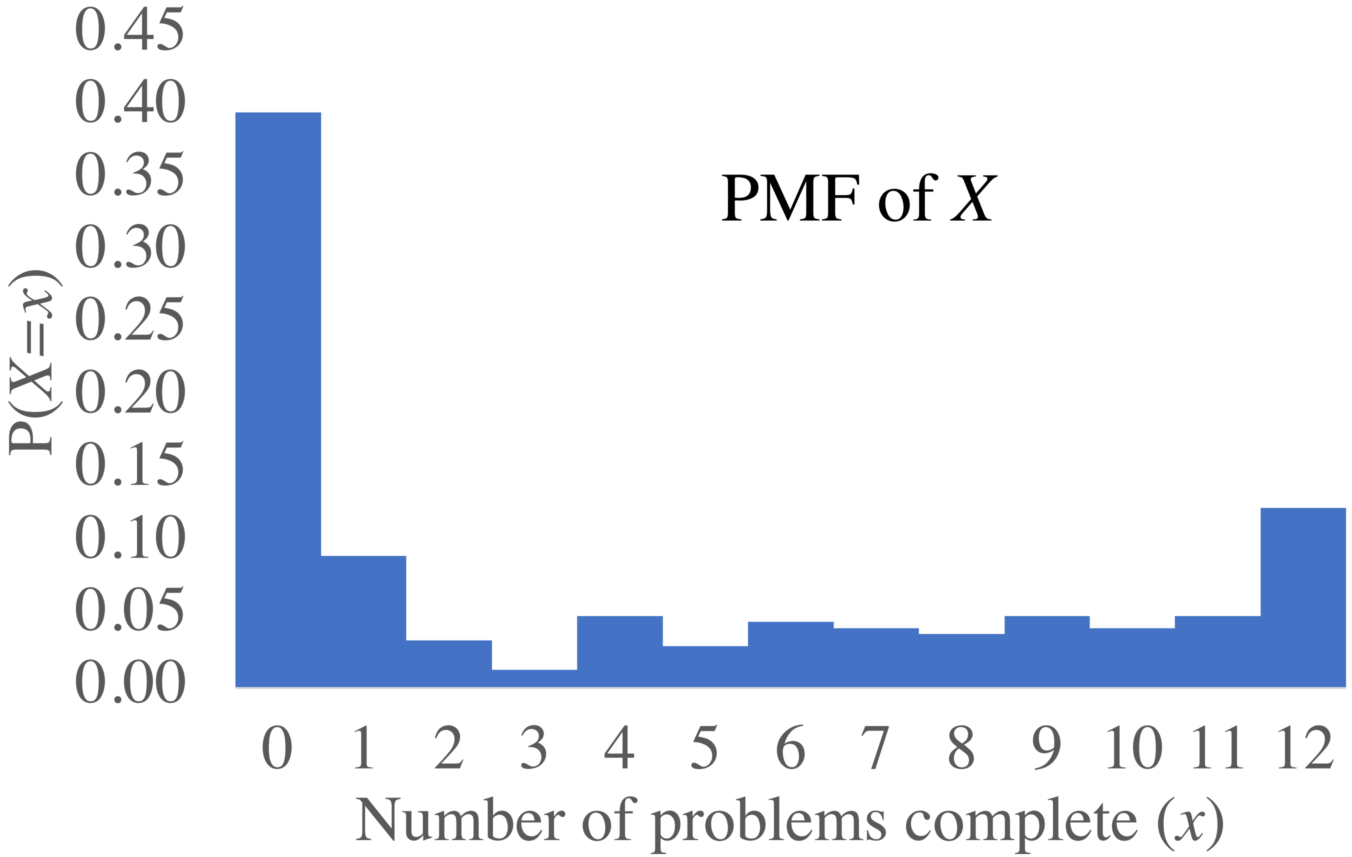
Normalized **histograms** are approximations of **probability mass functions**



Aside: Normalized 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

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

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 $E[X^2]$

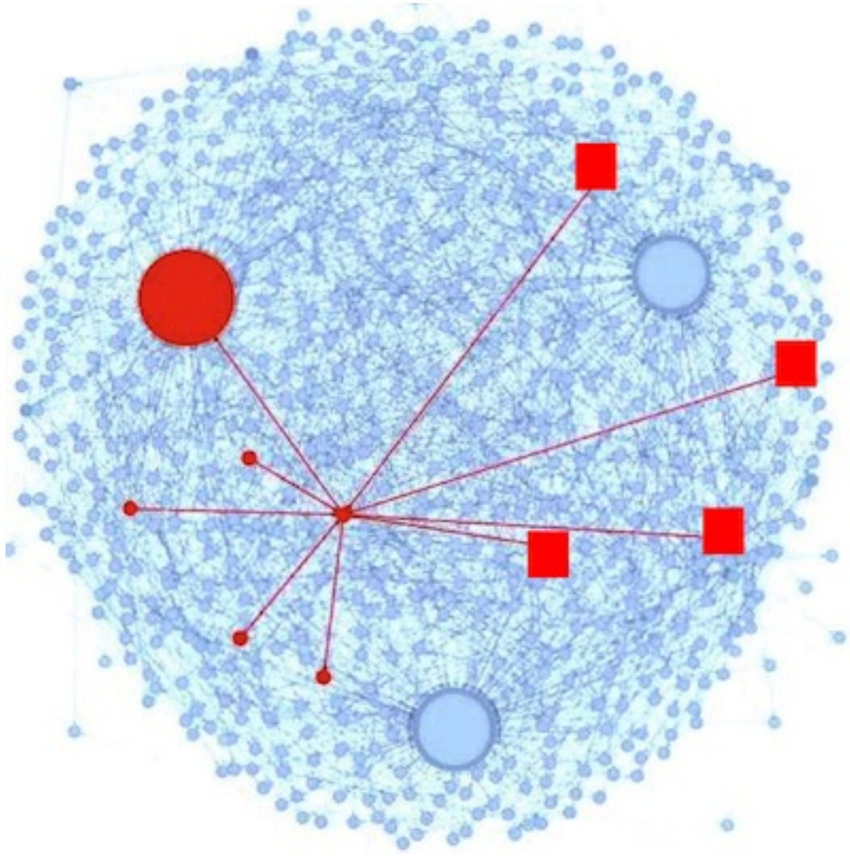
$$E[X^2] = (1^2)\frac{1}{6} + (2^2)\frac{1}{6} + (3^2)\frac{1}{6} + (4^2)\frac{1}{6} + (5^2)\frac{1}{6} + (6^2)\frac{1}{6} = \frac{91}{6}$$

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

$$= \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}$$

Is Peer Grading Accurate Enough?

Looking ahead



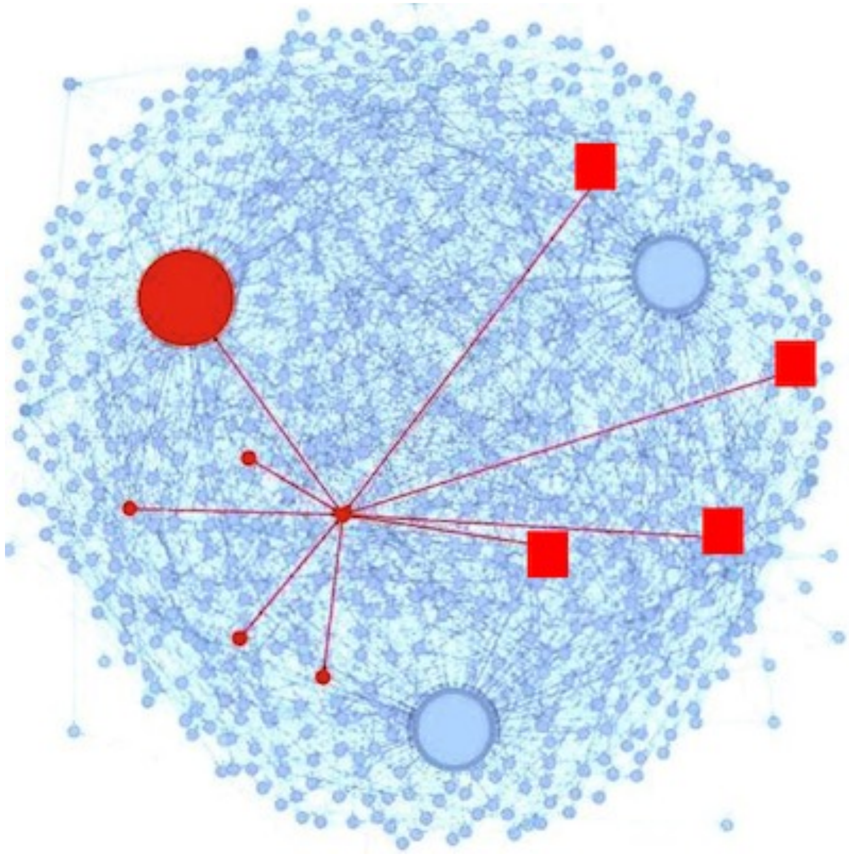
Peer Grading on Coursera HCI.

31,067 peer grades for 3,607 students.

Tuned Models of Peer Assessment. C Piech, J Huang, A Ng, D Koller

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

Problem param

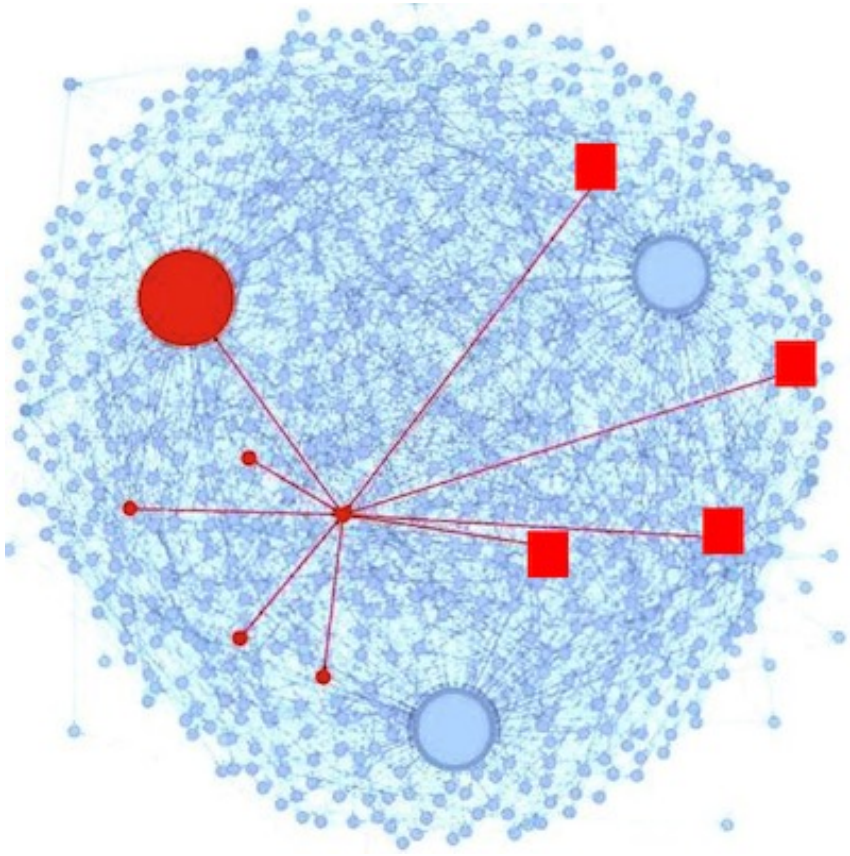
$$s_i \sim \mathrm{Bin}(\mathrm{points}, \theta)$$

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

Tuned Models of Peer Assessment. C Piech, J Huang, A Ng, D Koller

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*

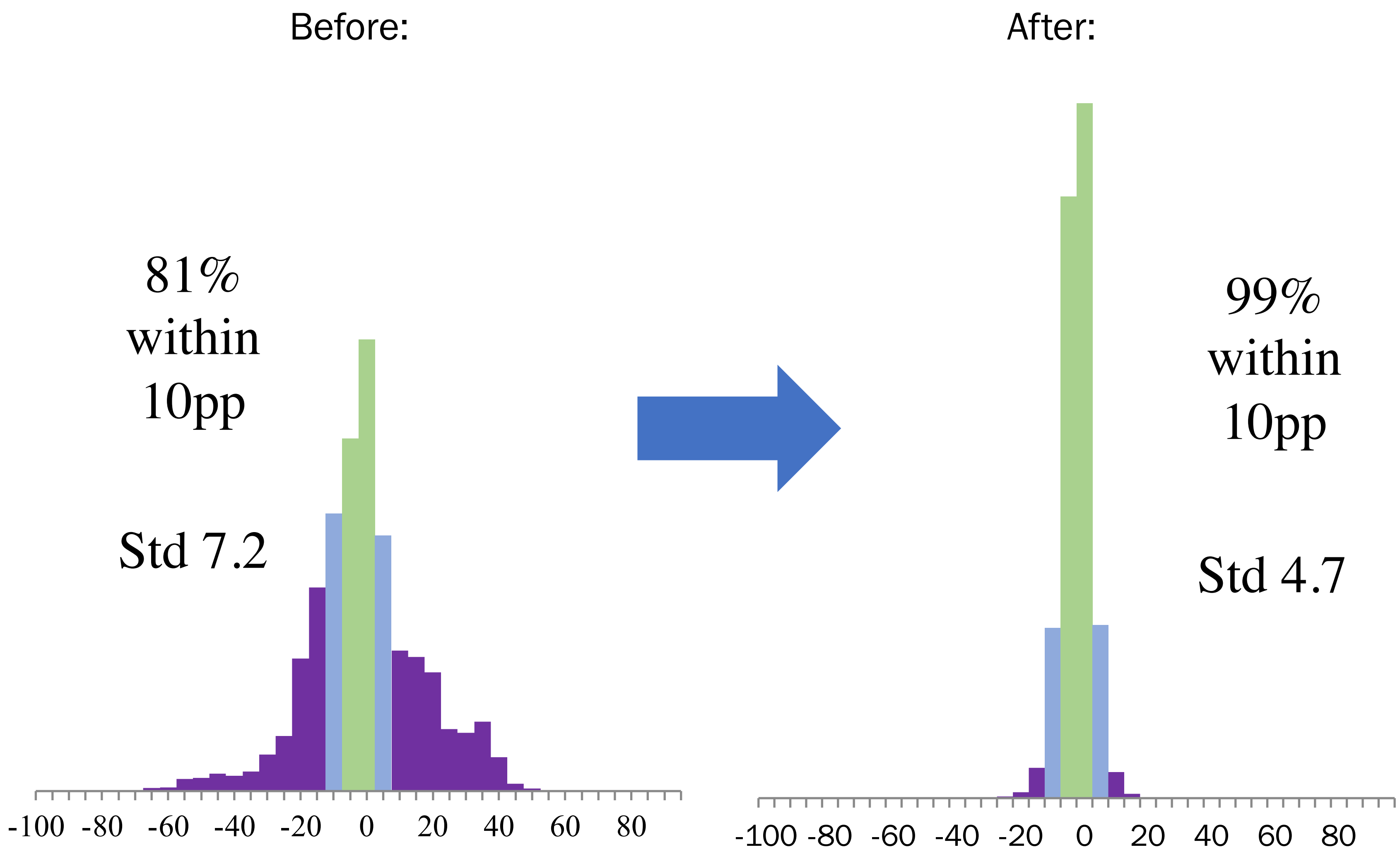
2. Designed a probabilistic model that defined the distributions for all random variables

3. Found the variable assignments that maximized the probability of our observed data

Inference or Machine Learning

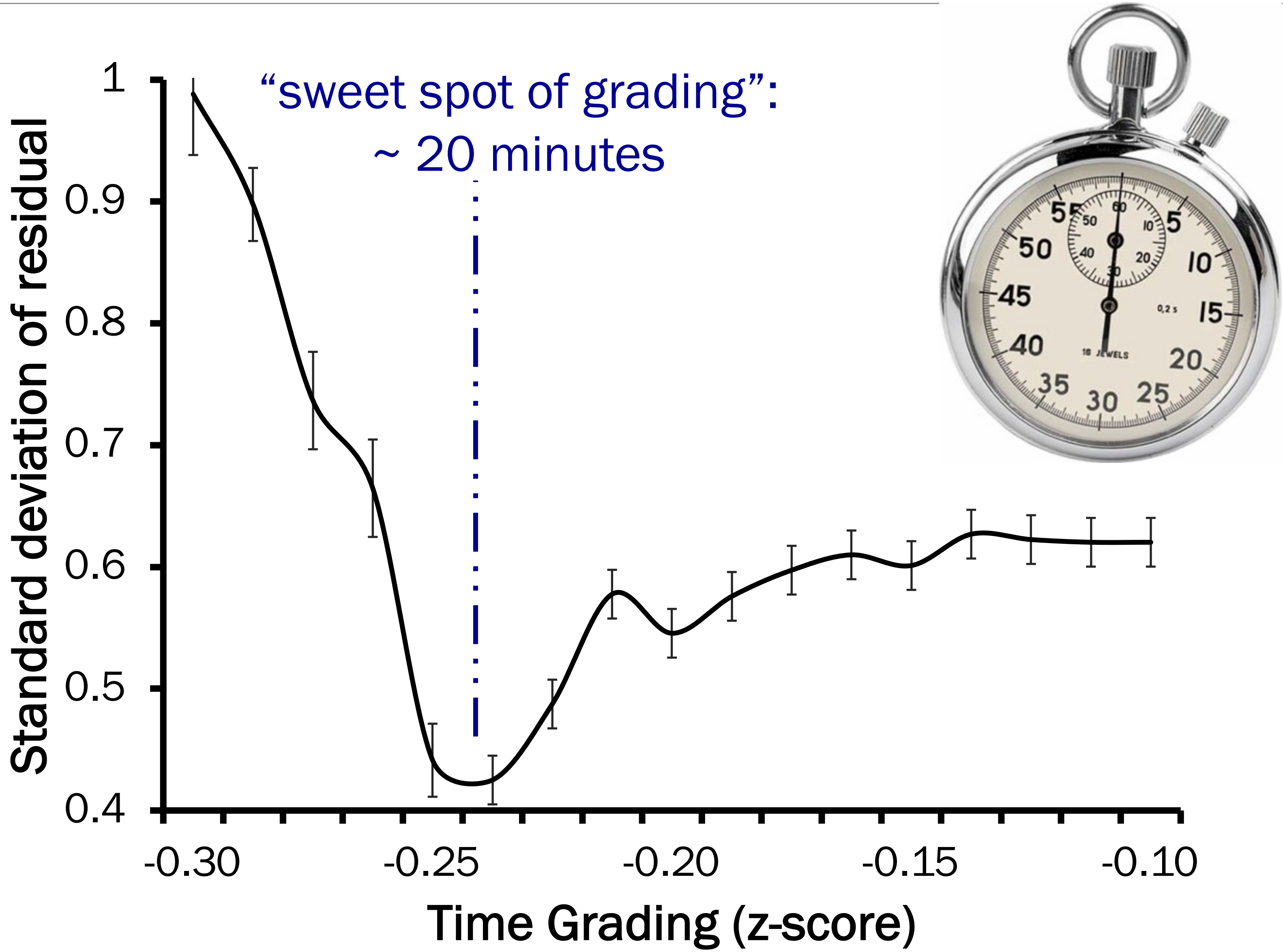
Tuned Models of Peer Assessment. C Piech, J Huang, A Ng, D Koller

Yes, With Probabilistic Modelling

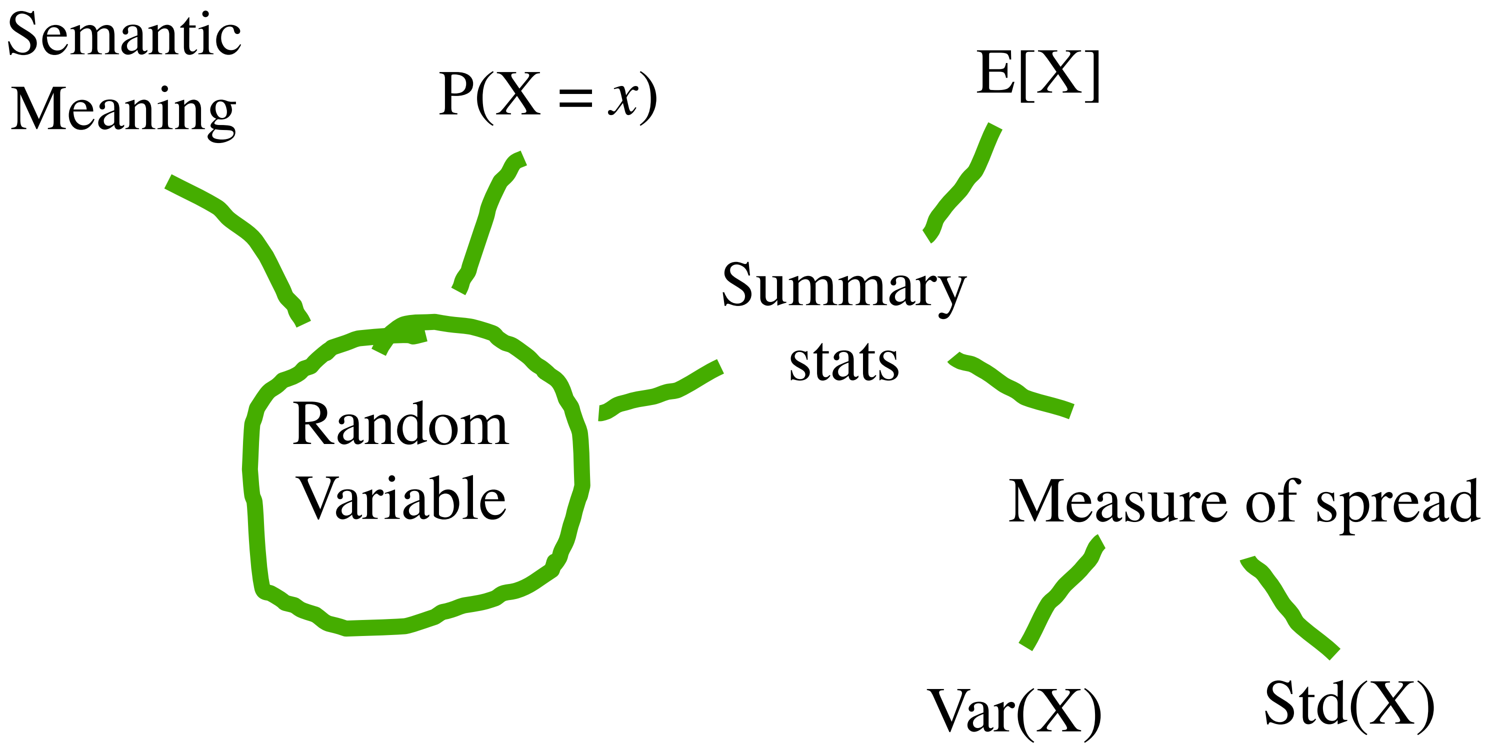


Tuned Models of Peer Assessment. C Piech, J Huang, A Ng, D Koller

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

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

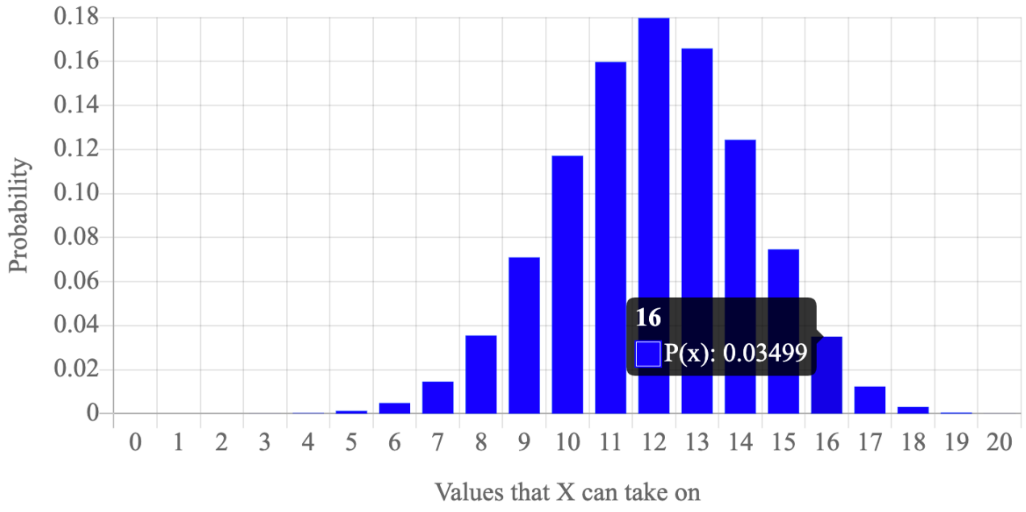
PMF equation: $\Pr(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$

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

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

PMF graph:

Parameter n: 20 Parameter p: 0.60



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

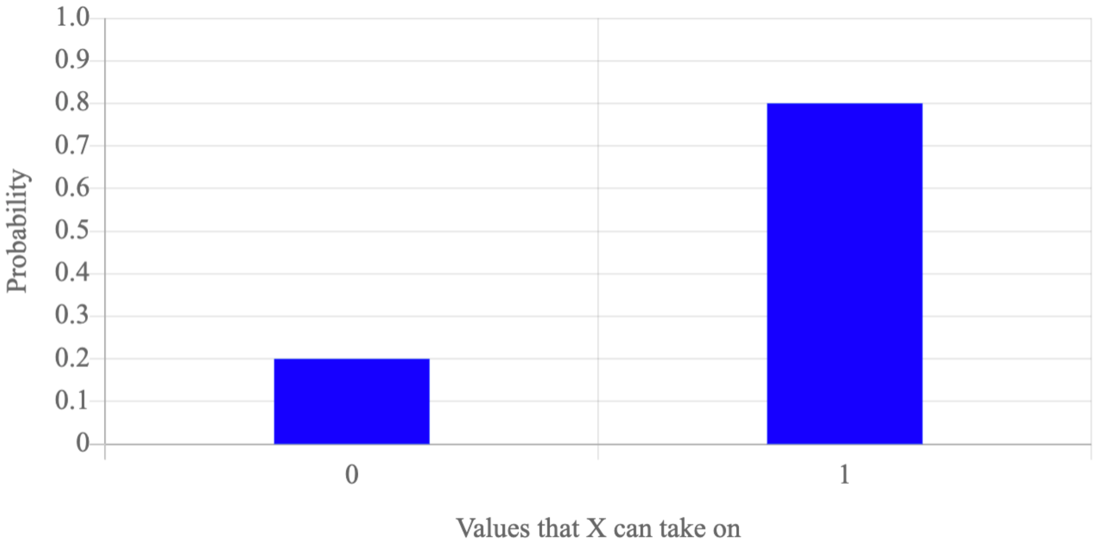
PMF equation: $\Pr(X = x) = \begin{cases} p & \text{if } x = 1 \\ 1-p & \text{if } x = 0 \end{cases}$

Expectation: $\text{E}[X] = p$

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

PMF graph:

Parameter p: 0.80



Beyond CS109: Proof of Variance for a Binomial

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

Voilà, c'est tout