Debugging Intuition

- How to calculate the probability of at least *k* successes in *n* trials?
  - X is number of successes in *n* trials each with probability *p*
  - $P(X \geq k) =$

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

Don't care about the 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: $P(X \geq k) = \sum_{i=k}^{n} \binom{n}{i} p^i (i-p)^{n-i}$

Variance

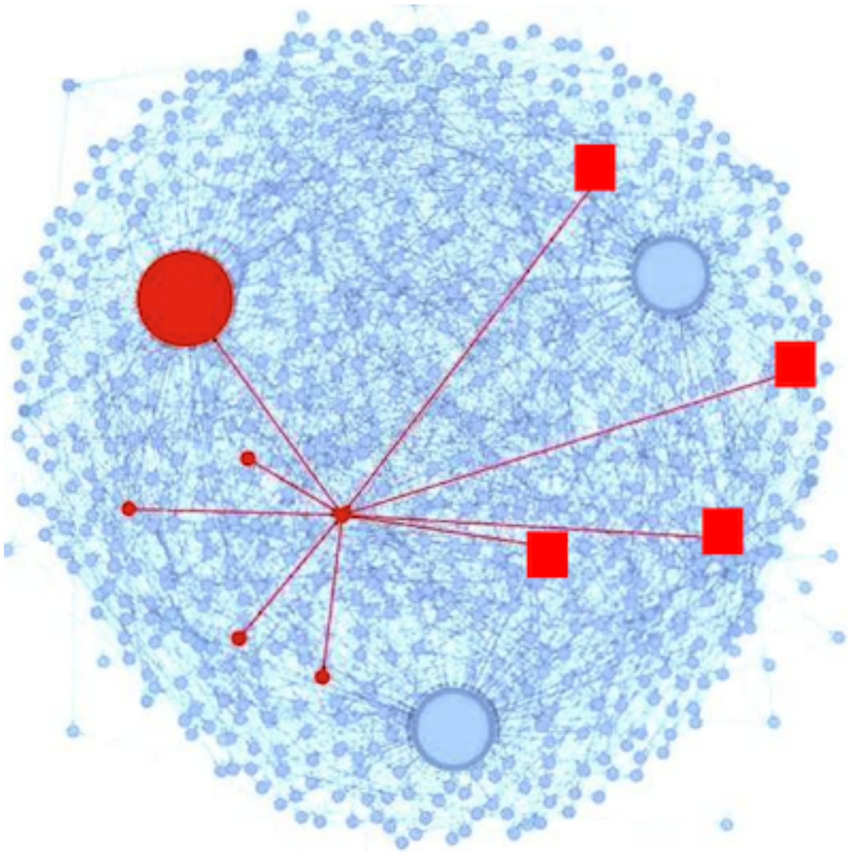
Chris Piech

CS109, Stanford University

Learning Goals

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

Is Peer Grading Accurate Enough?



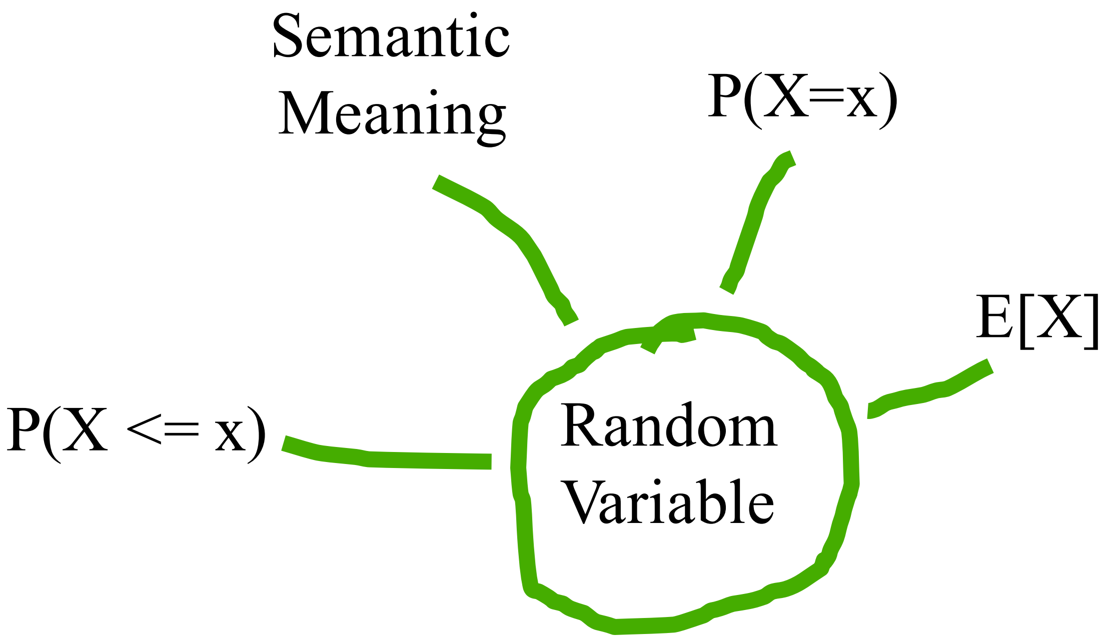
Peer Grading on Coursera HCI.

31,067 peer grades for 3,607 students.

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

Random variables are vars that probabilistically take on values.

Fundamental Properties



Breaking Vegas

Lets say we have an algorithm to break Vegas, but it has a probabilistic outcome.

Let Z be a random variable which is the amount of money we have when we finish our Vegas breaking algorithm.

Does there exist an algorithm with $E[Z] > 0$?

Recall, Geometric Series

$$a^0 + a^1 + a^2 + \ldots$$

$$= \sum_{i=0}^{\infty} a^i$$

$$= \frac{1}{1-a}$$

$$\text{where } 0 < a < 1$$

Breaking Vegas

- Consider even money bet (e.g., bet "Red" in roulette)
  - p = 18/38 you win \$Y, otherwise (1 – p) you lose \$Y
  - Consider this algorithm for one series of bets:
    1. Y = \$1
    2. Bet Y
    3. If Win then stop
    4. If Loss then Y = 2 * Y, goto 2
  - Let Z = winnings upon stopping
  - E[Z] $= \left(\frac{18}{38}\right)1 + \left(\frac{20}{38}\right)\left(\frac{18}{38}\right)(2-1) + \left(\frac{20}{38}\right)^2\left(\frac{18}{38}\right)(4-2-1) + \ldots$

$$= \sum_{i=0}^{\infty}\left(\frac{20}{38}\right)^i\left(\frac{18}{38}\right)\left(2^i - \sum_{j=0}^{i-1} 2^j\right) = \left(\frac{18}{38}\right)\sum_{i=0}^{\infty}\left(\frac{20}{38}\right)^i = \left(\frac{18}{38}\right)\frac{1}{1-\frac{20}{38}} = 1$$

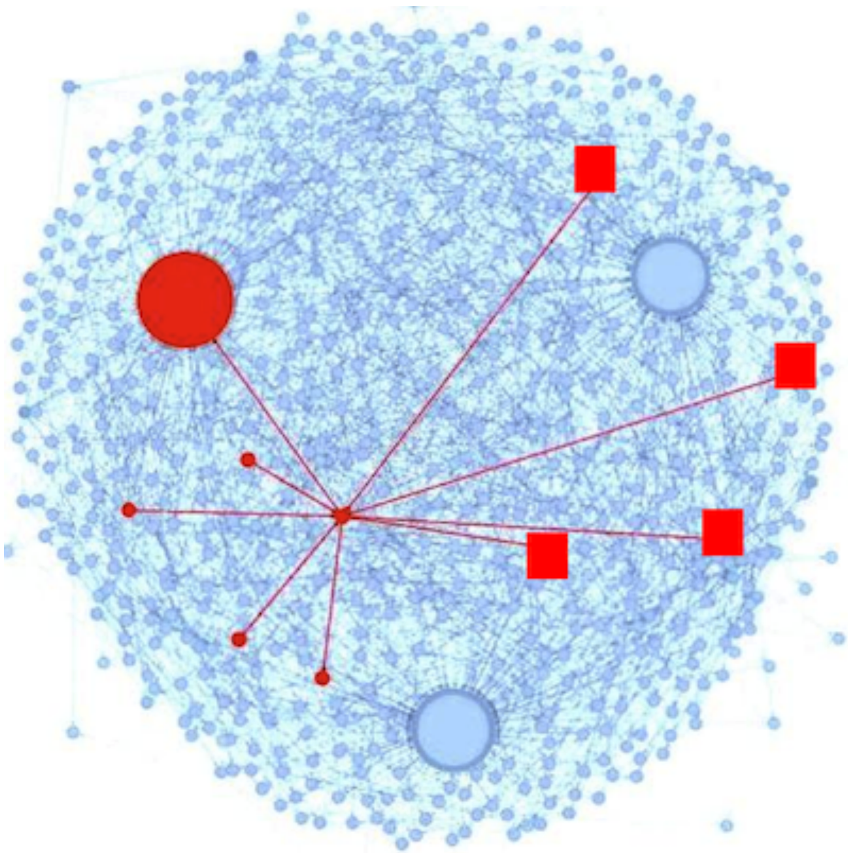
  - Expected winnings ≥ 0. Use algorithm infinitely often!

Vegas Breaks You

- Why doesn't everyone do this?
  - Real games have maximum bet amounts
  - You have finite money
    - Not able to keep doubling bet beyond certain point
  - Casinos can kick you out
- But, if you had:
  - No betting limits, and
  - Infinite money, and
  - Could play as often as you want...
- Then, go for it!
  - And tell me which planet you are living on

Is E[X] enough?

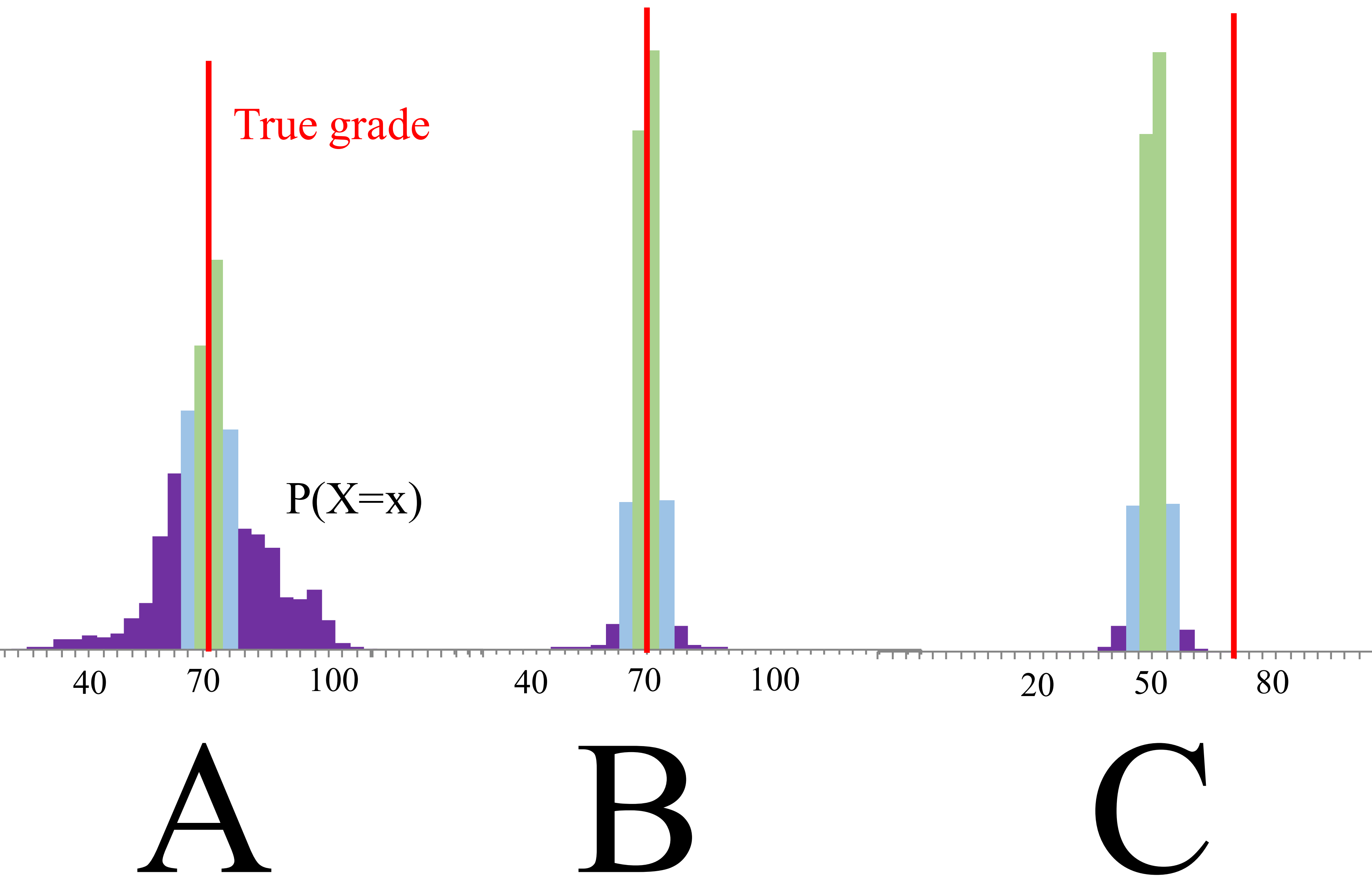
Intuition



Peer Grading on Coursera HCI.

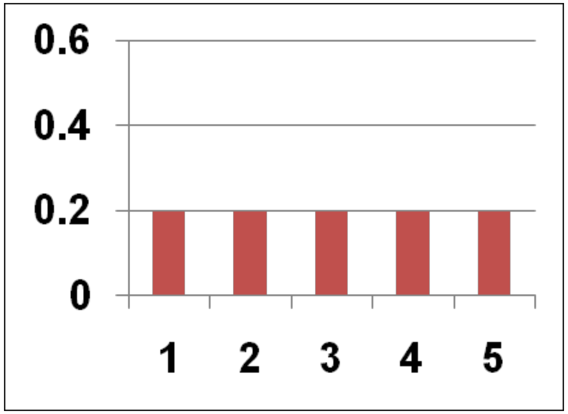
31,067 peer grades for 3,607 students.

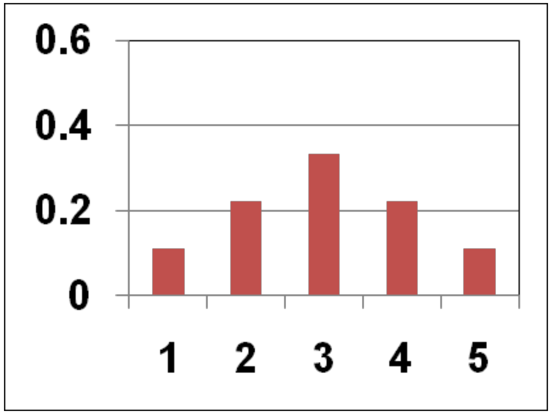
X is the score a peer grader gives to an assignment submission

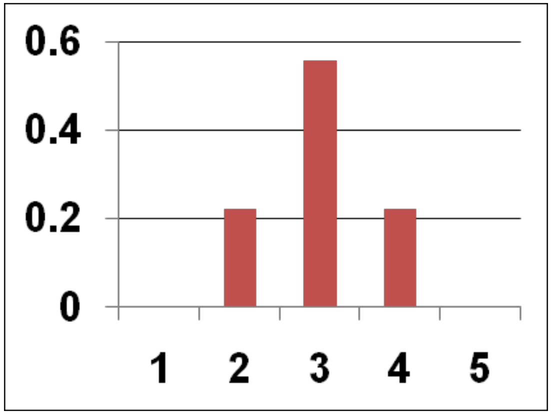


Variance

- Consider the following 3 distributions (PMFs)



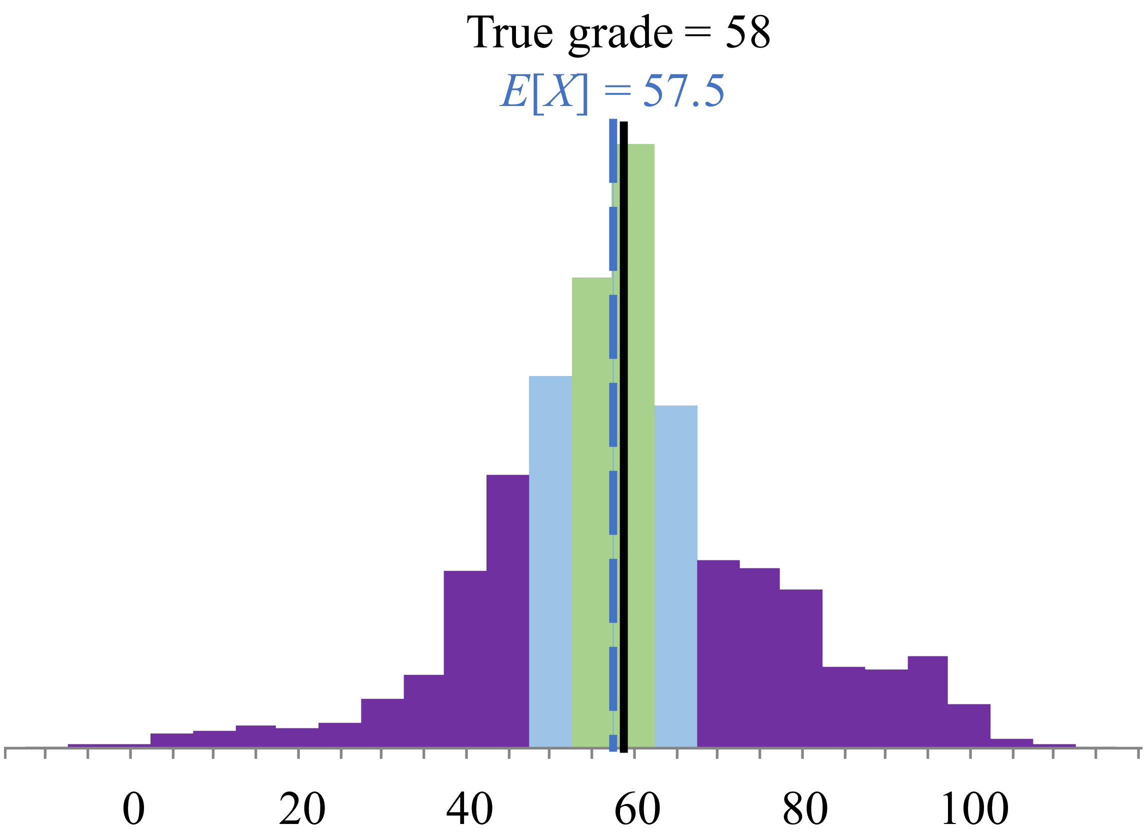




- All have the same expected value, E[X] = 3
- But "spread" in distributions is different
- Variance = a formal quantification of "spread"

Peer Grades in Coursera HCI

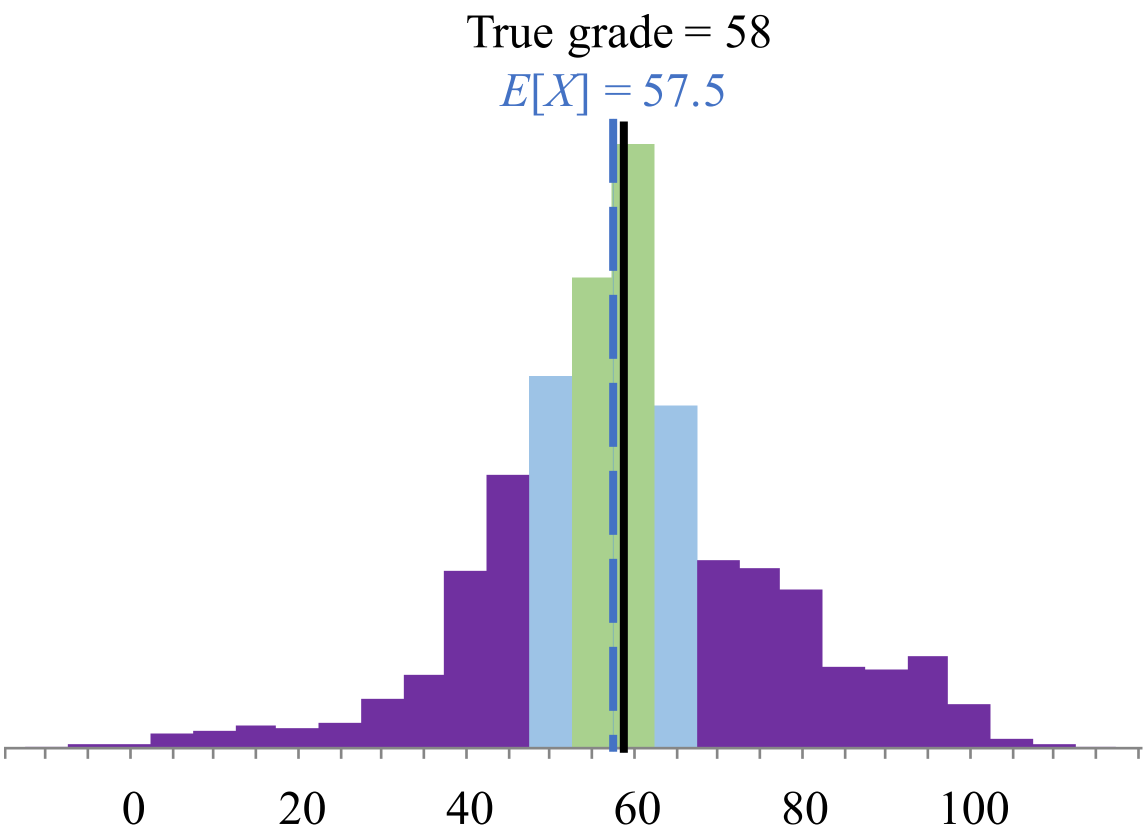
Let X be a random variable that represents a peer grade



Peer Grades in Coursera HCI

Let X be a random variable that represents a peer grade

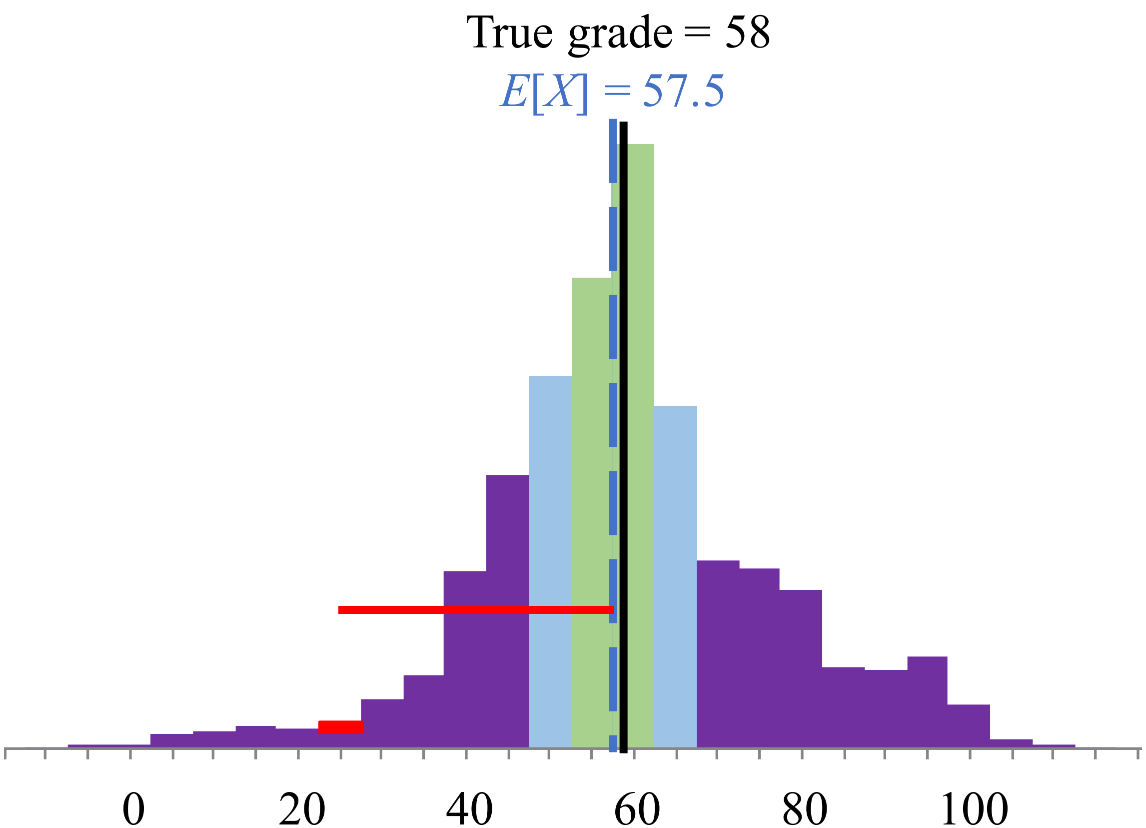
$$\mathrm{Var}(X) = E[(X - \mu)^2]$$



Peer Grades in Coursera HCI

Let X be a random variable that represents a peer grade

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

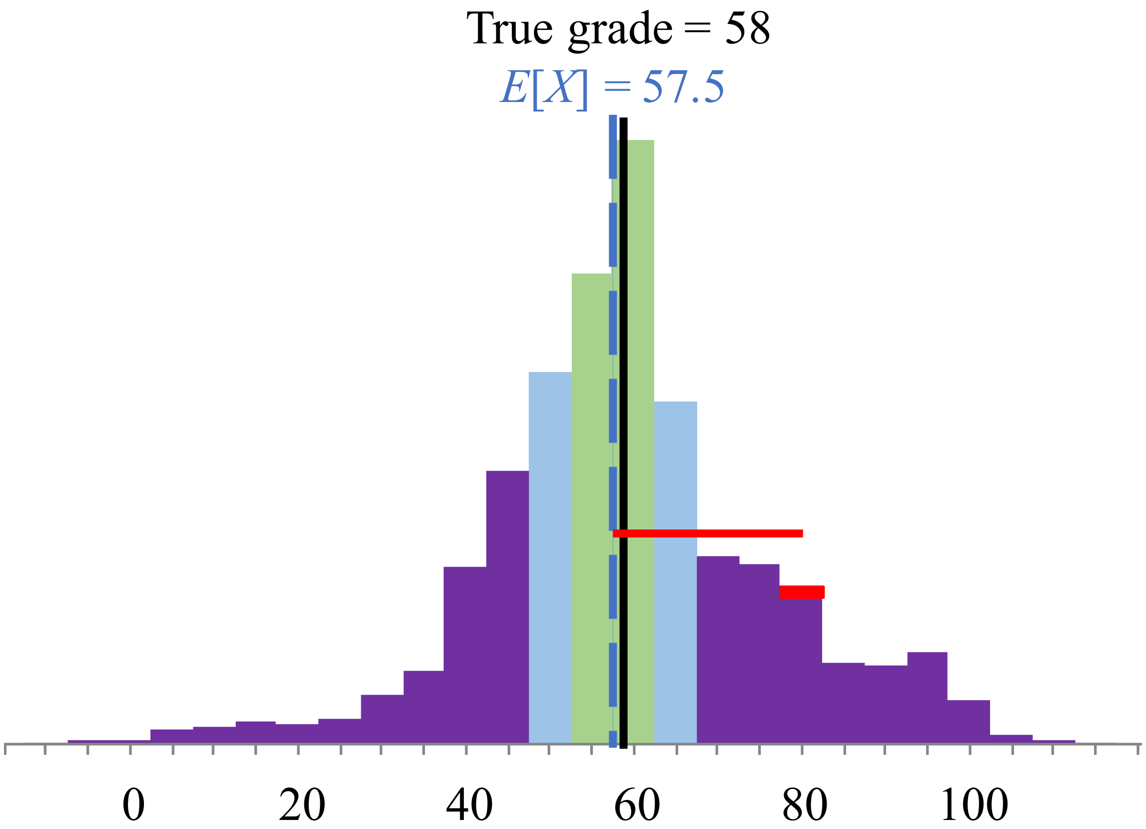


X	$(X - \mu)^2$
25 points	1056 points^2

Peer Grades in Coursera HCI

Let X be a random variable that represents a peer grade

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

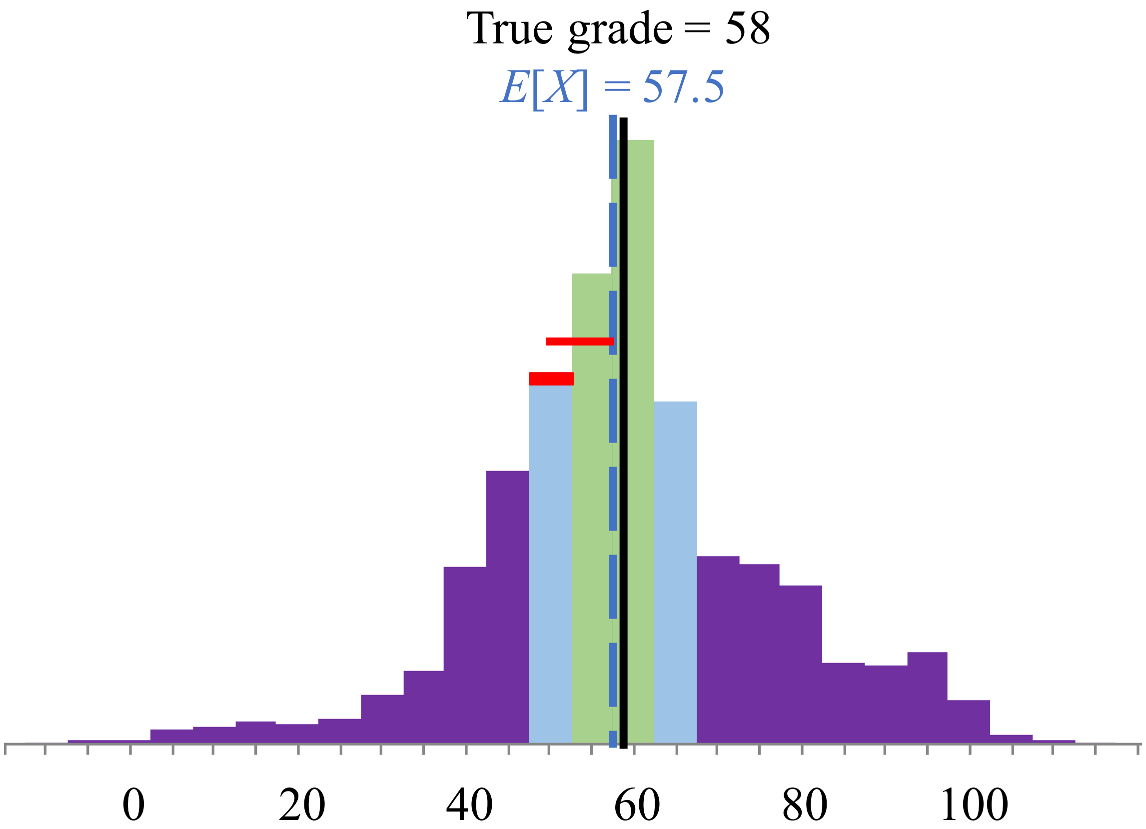


X	$(X - \mu)^2$
25 points	1056 points^2
80 points	506 points^2

Peer Grades in Coursera HCI

Let X be a random variable that represents a peer grade

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

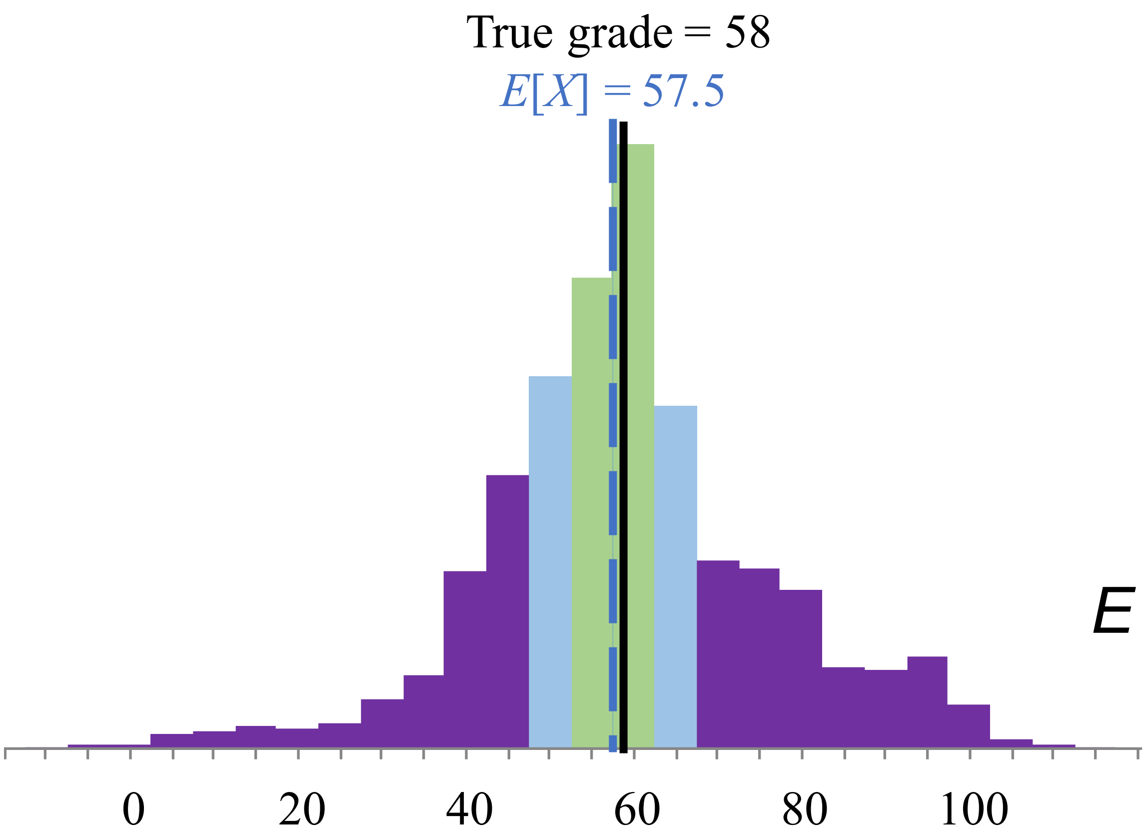


X	$(X - \mu)^2$
25 points	1056 points^2
80 points	506 points^2
50 points	56 points^2

Peer Grades 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

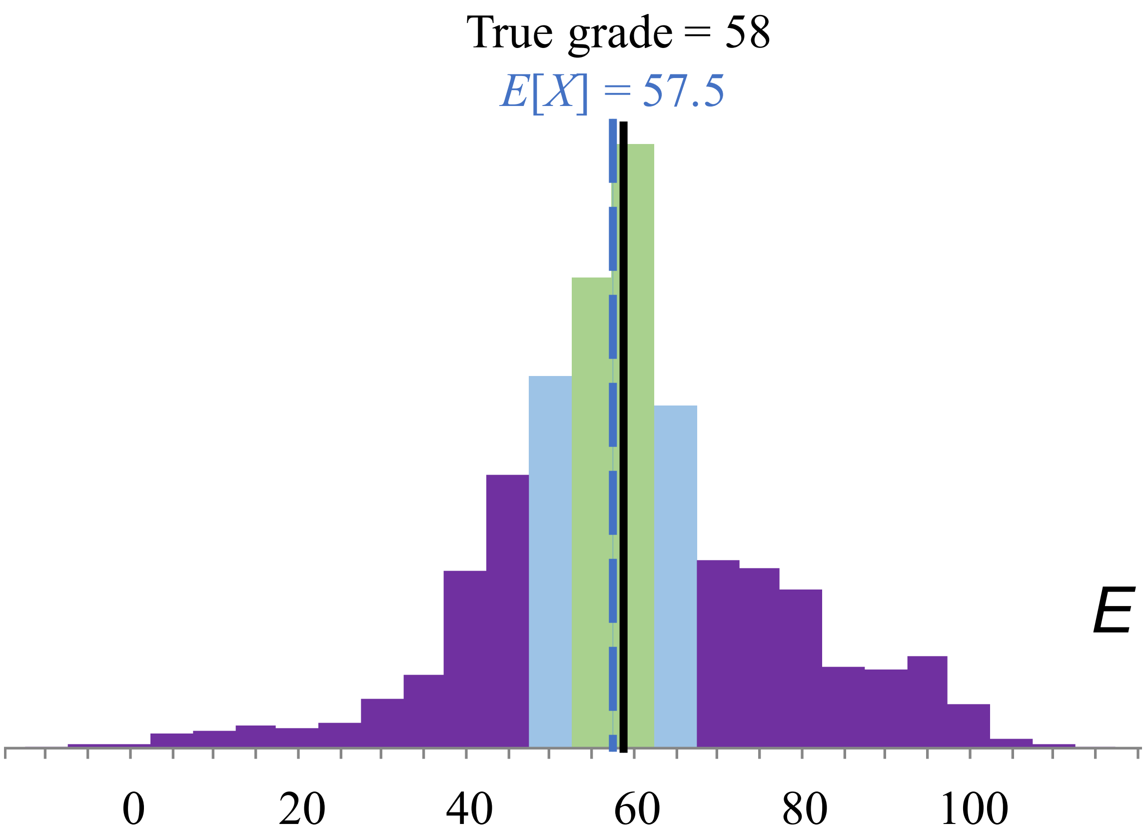
…

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

Peer Grades 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

…

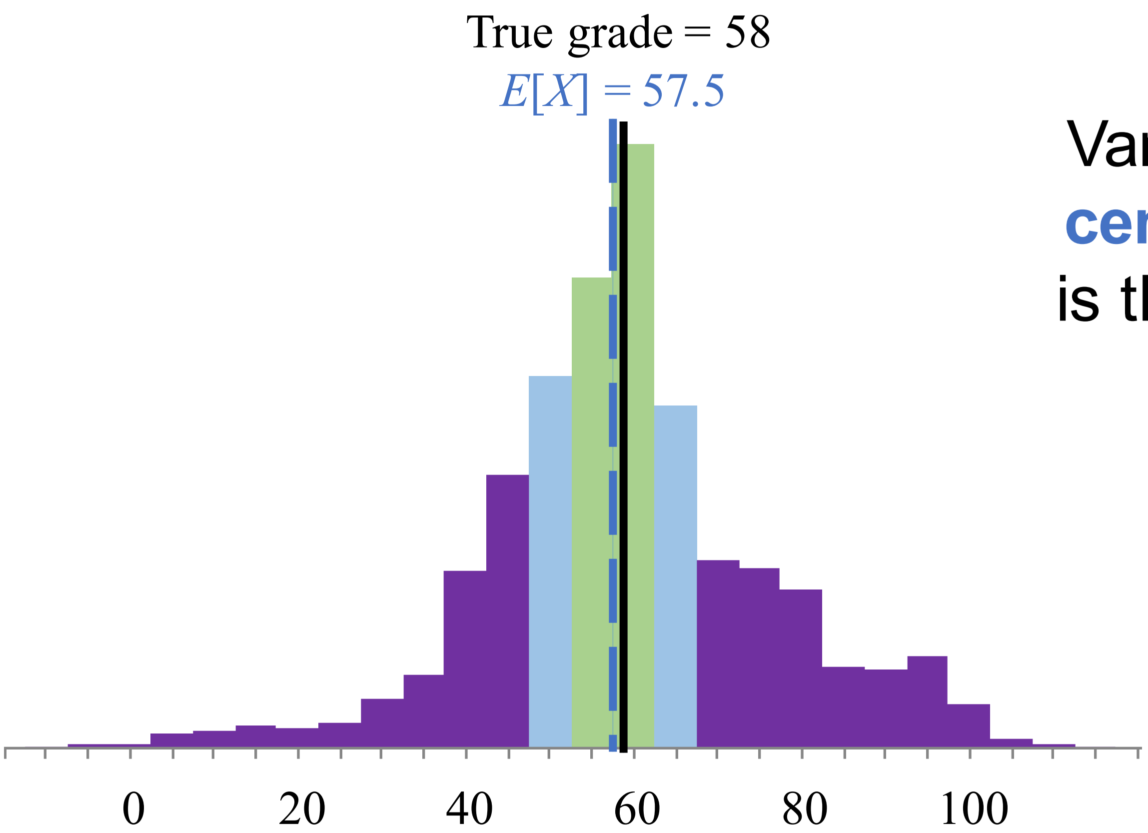
$$E[(X-\mu)^2] = 52 \text{ points}^2$$

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

Second Moment

Let X be a random variable that represents a peer grade

$$E[X^2]$$

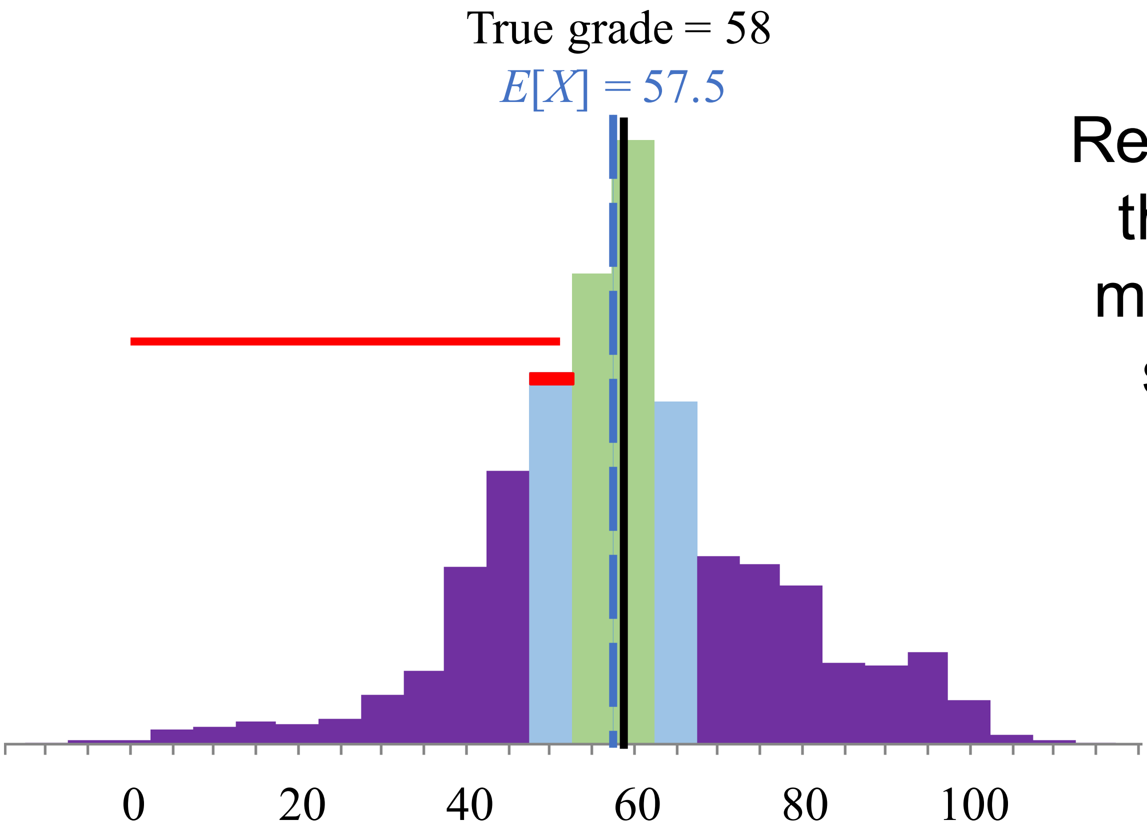


Variance is the second **central** moment. What is the second moment?

Second Moment

Let X be a random variable that represents a peer grade

$$E[X^2]$$

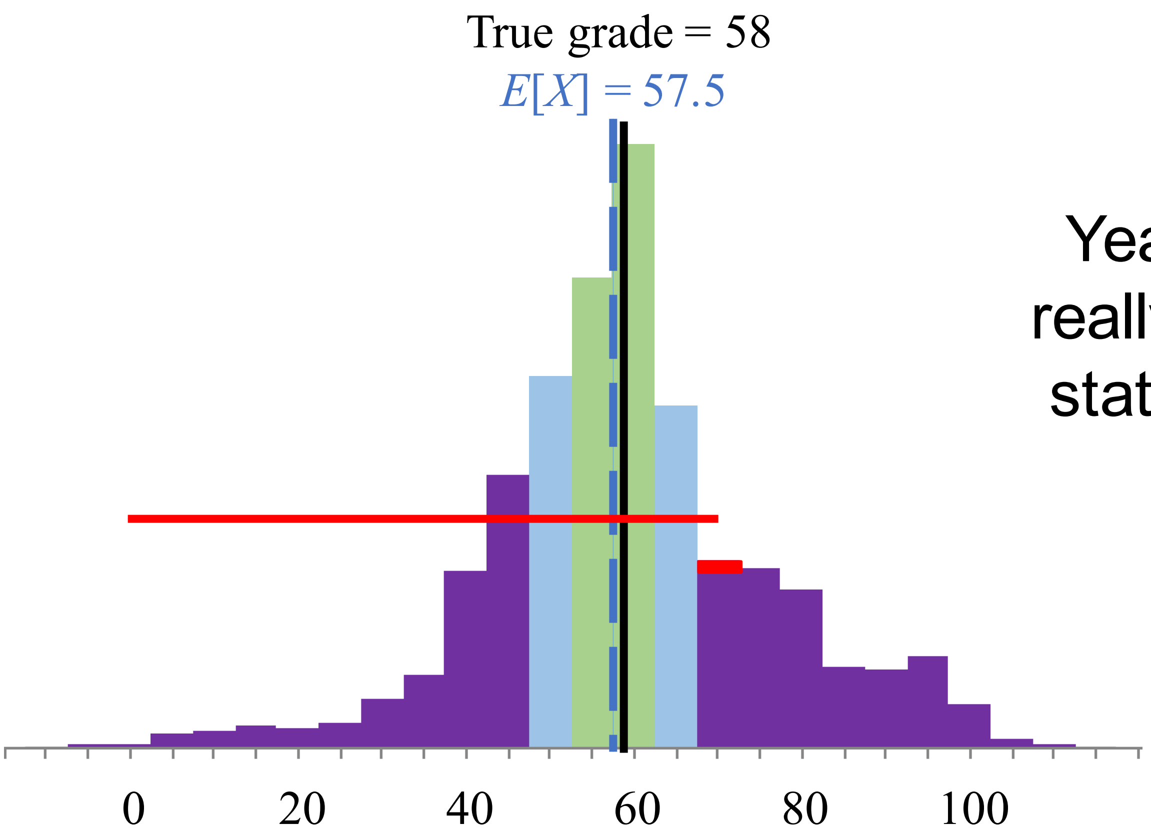


Recall that Variance is the second **central** moment. What is the second moment?

Second Moment

Let X be a random variable that represents a peer grade

$$E[X^2]$$



Yea second moment really isn't that useful a statistic in this case…

Variance

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

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

- Note: Var(X) ≥ 0

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

Computing Variance

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

Note: $\mu = E[X]$

Ladies and gentlemen, please welcome the 2[nd] moment!

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]

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

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

Properties of Variance

- $\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$
  - Proof:

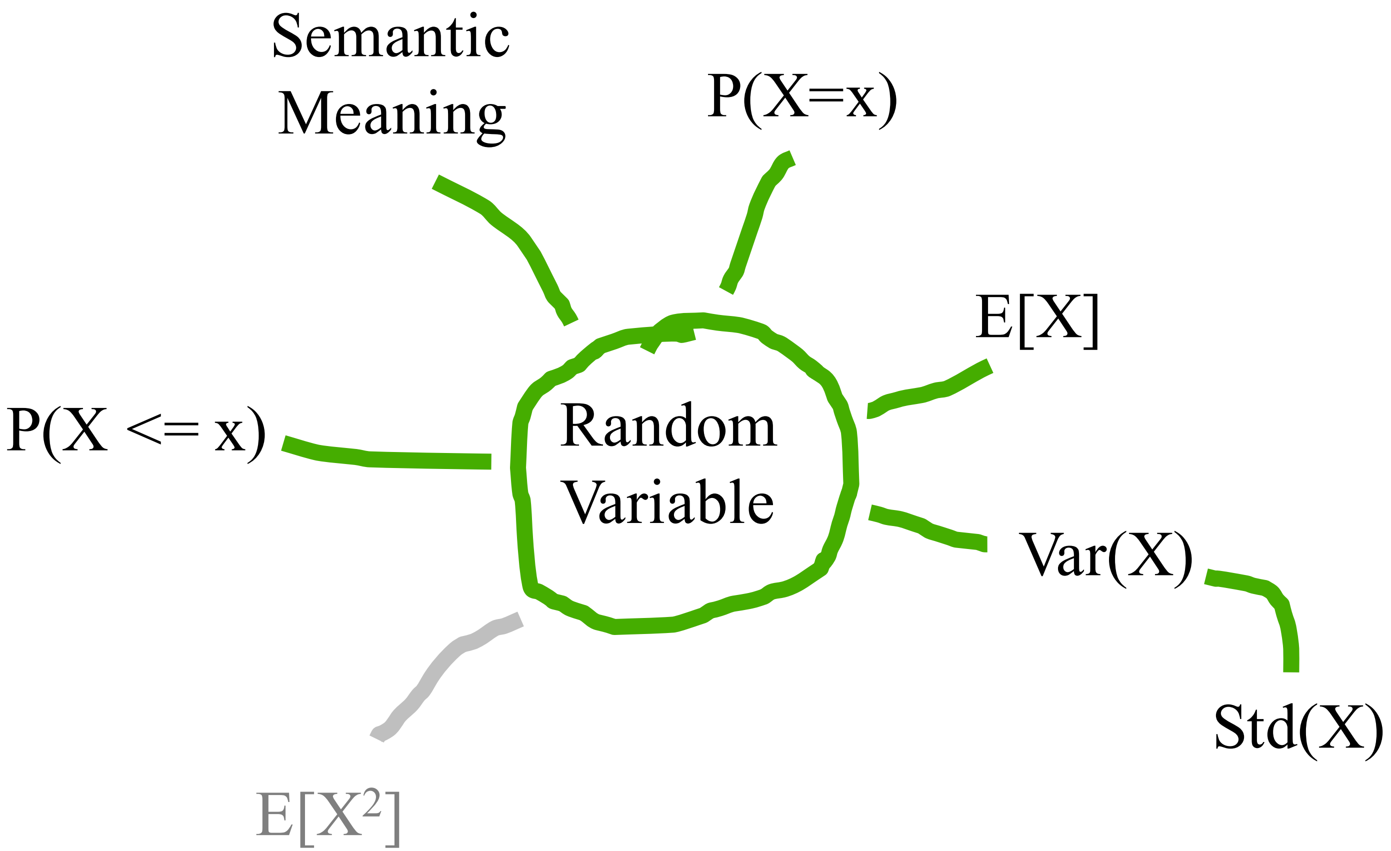
$$
\begin{aligned}
\mathrm{Var}(aX + b) &= E[(aX + b)^2] - (E[aX + b])^2 \\
&= E[a^2X^2 + 2abX + b^2] - (aE[X] + b)^2 \\
&= a^2E[X^2] + 2abE[X] + b^2 - (a^2(E[X])^2 + 2abE[X] + b^2) \\
&= a^2E[X^2] - a^2(E[X])^2 = a^2(E[X^2] - (E[X])^2) \\
&= a^2\mathrm{Var}(X)
\end{aligned}
$$

- Standard Deviation of X, denoted SD(X), is:

$$\mathrm{SD}(X) = \sqrt{\mathrm{Var}(X)}$$

  - $\mathrm{Var}(X)$ is in units of X^2
  - $\mathrm{SD}(X)$ is in same units as X

Fundamental Properties



Lots of fun with Random Variables

Classics

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
- Same eyes as Ice Cube

Bernoulli Random Variable

- Experiment results in "Success" or "Failure"
  - X is random **indicator** variable (1 = success, 0 = failure)
  - $P(X = 1) = p(1) = p$ $\qquad P(X = 0) = p(0) = 1 – p$
  - X is a <u>**Bernoulli**</u> Random Variable: $X \sim Ber(p)$
  - $E[X] = p$
  - $Var(X) = p(1 – p)$
- Examples
  - coin flip
  - random binary digit
  - whether a disk drive crashed
  - whether someone likes a netflix movie

Feel the Bern!

Binomial Random Variable

- Consider *n* independent trials of Ber(p) rand. var.
  - X is number of successes in *n* trials
  - X is a **Binomial** Random Variable: X ~ Bin(n, p)

$$P(X = i) = p(i) = \binom{n}{i} p^i (1-p)^{n-i} \qquad i = 0,1,\ldots,n$$

  - By Binomial Theorem, we know that $\sum_{i=0}^{\infty} P(X = i) = 1$
- Examples
  - # of 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

Bernoulli vs Binomial

Bernoulli is an indicator RV

Binomial is the sum of *n* Bernoullis

Three Coin Flips

- Three fair ("heads" with p = 0.5) coins are flipped
  - X is number of heads
  - X ~ Bin(3, 0.5)

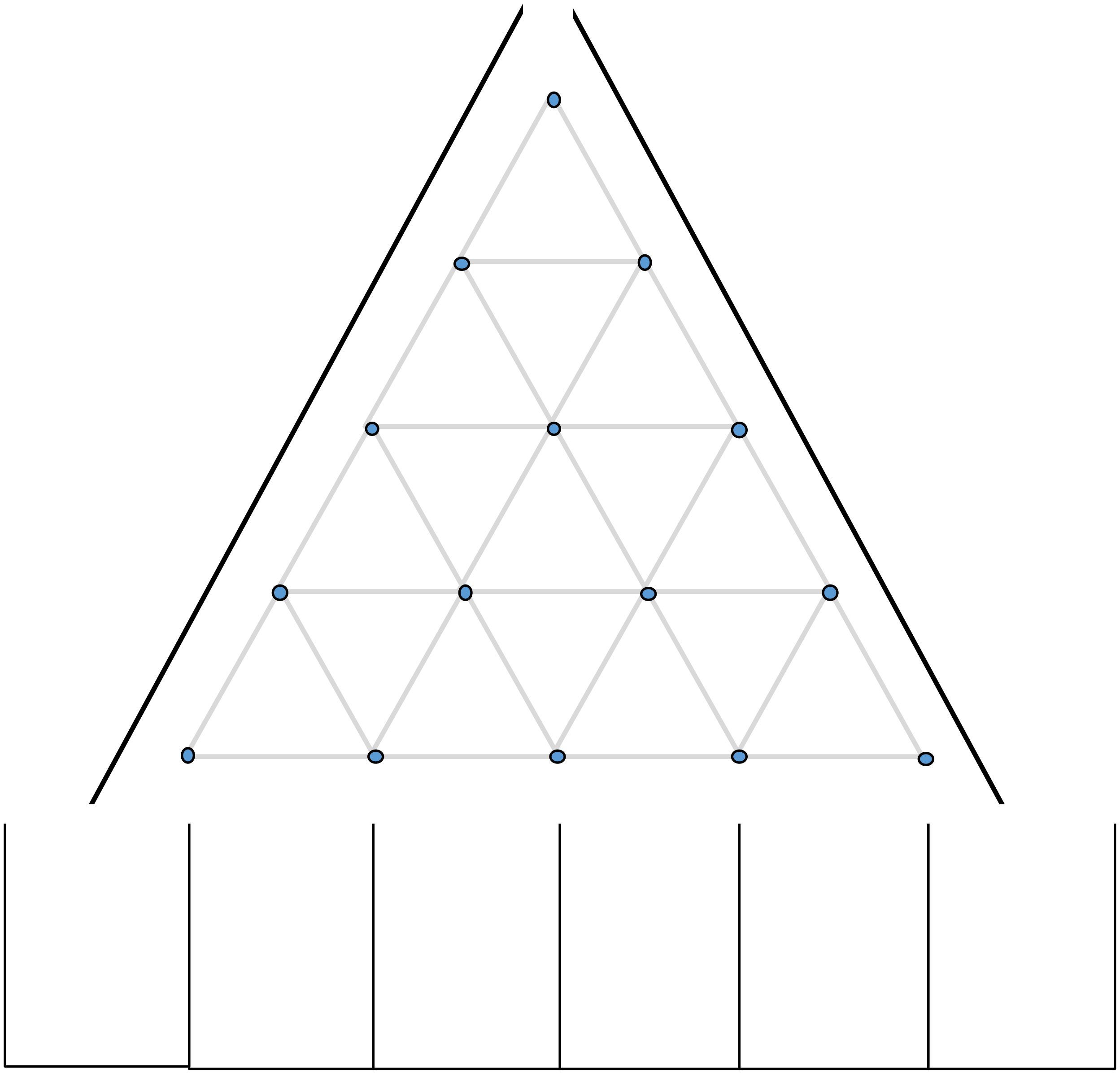
$$P(X=0) = \binom{3}{0} p^0 (1-p)^3 = \frac{1}{8}$$

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

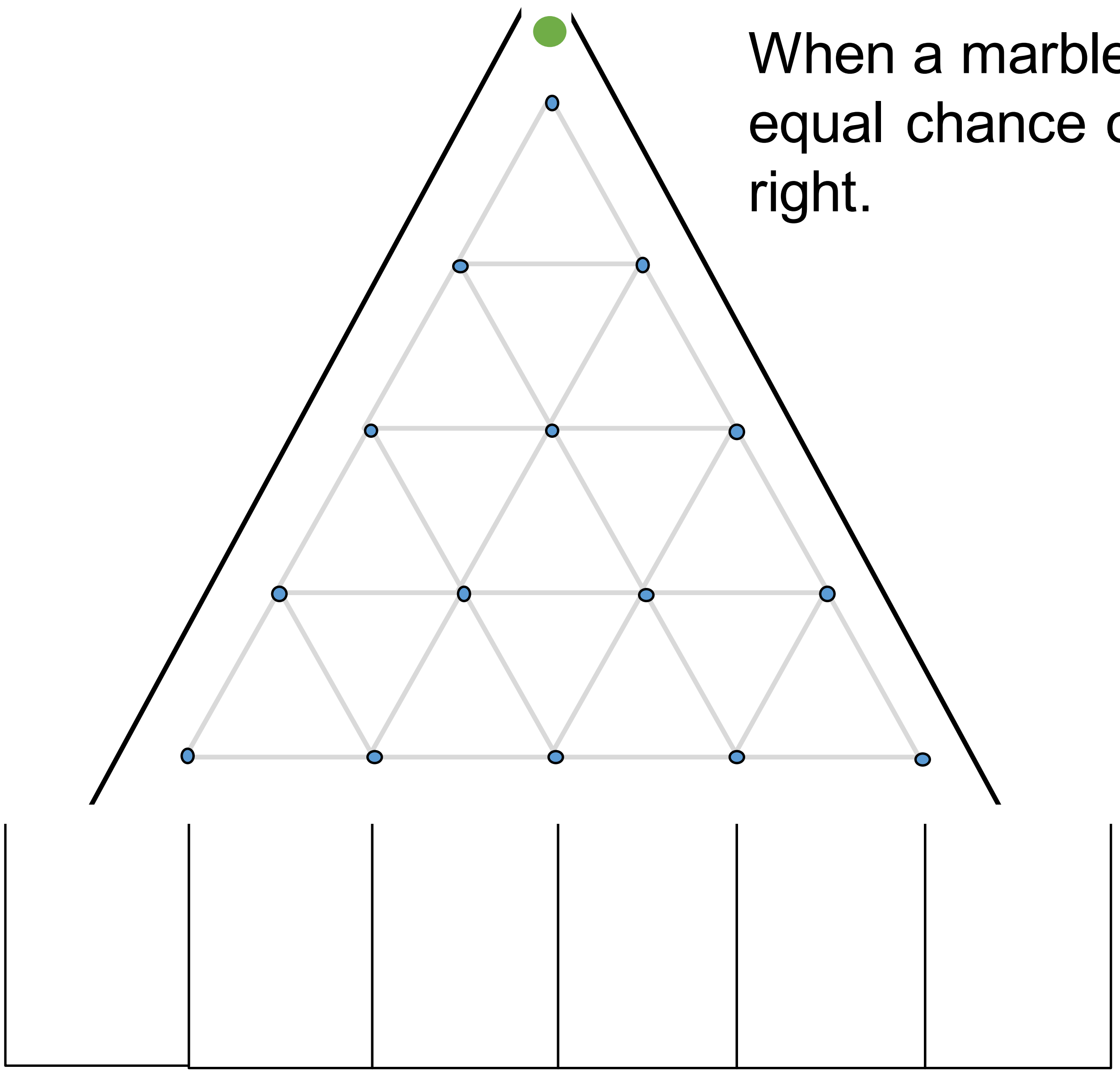
$$P(X=2) = \binom{3}{2} p^2 (1-p)^1 = \frac{3}{8}$$

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

Galton Board

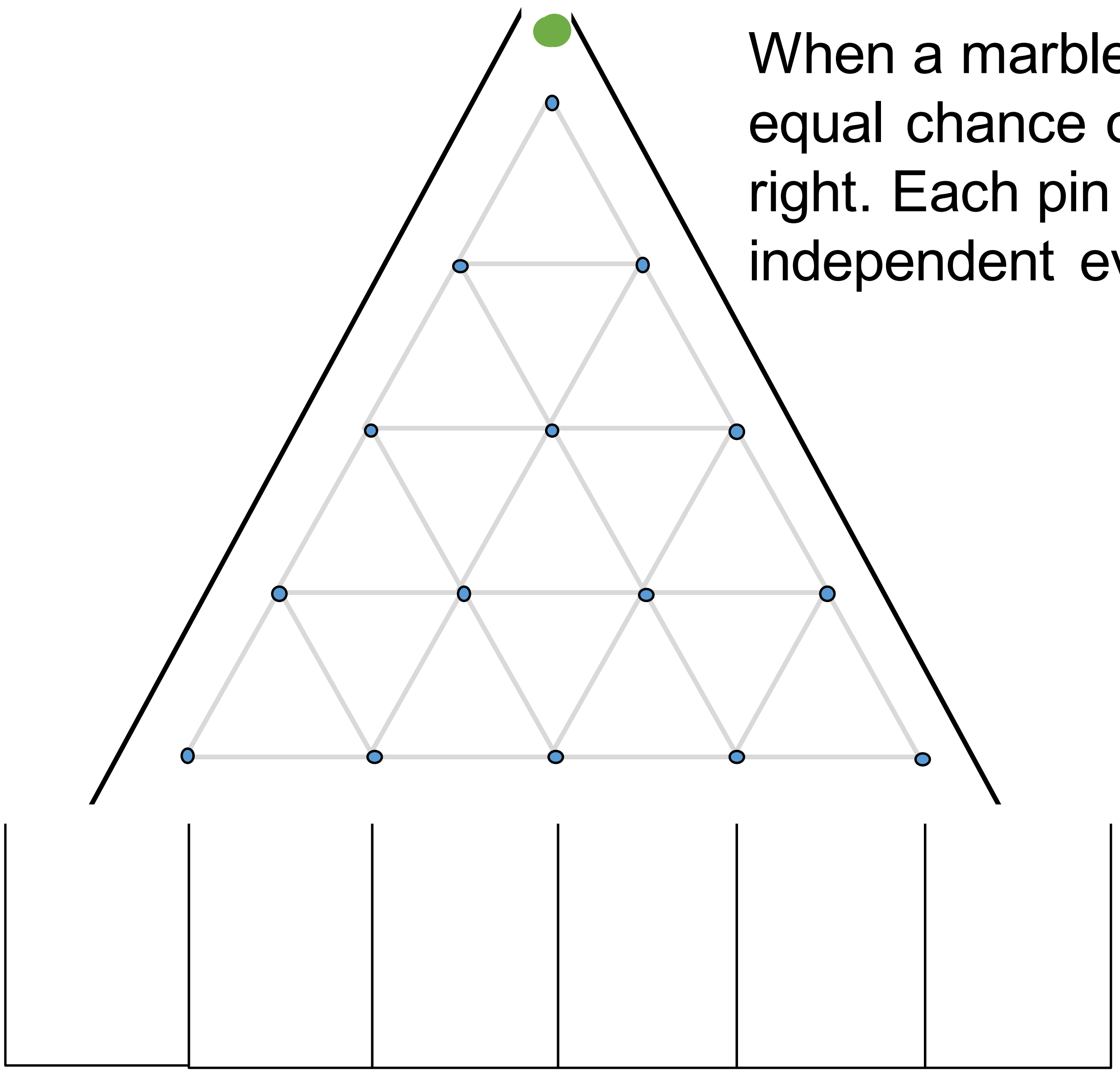


Galton Board



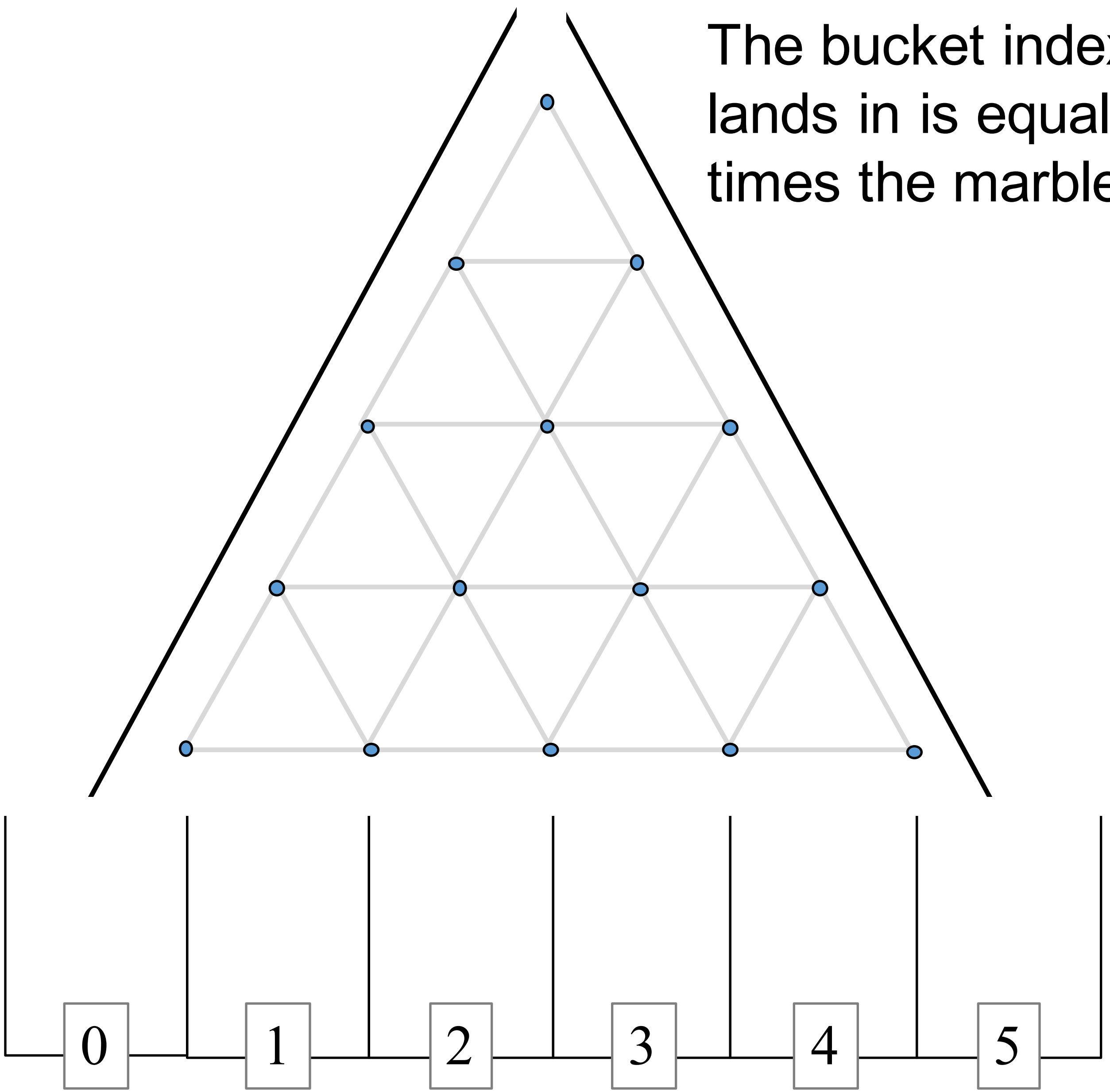
When a marble hits a pin, it has equal chance of going left or right.

Galton Board



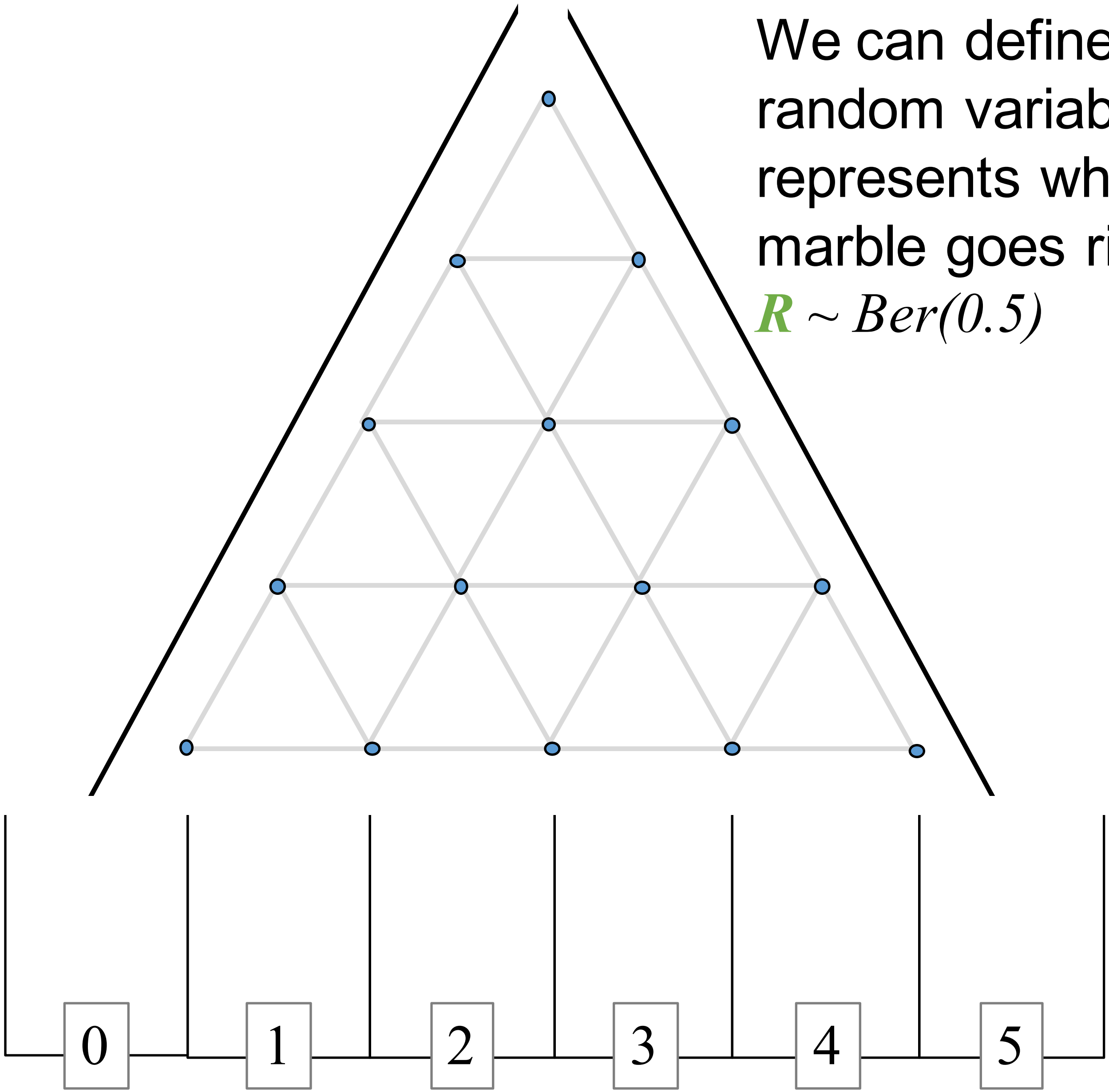
When a marble hits a pin, it has equal chance of going left or right. Each pin represents an independent event.

Galton Board



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

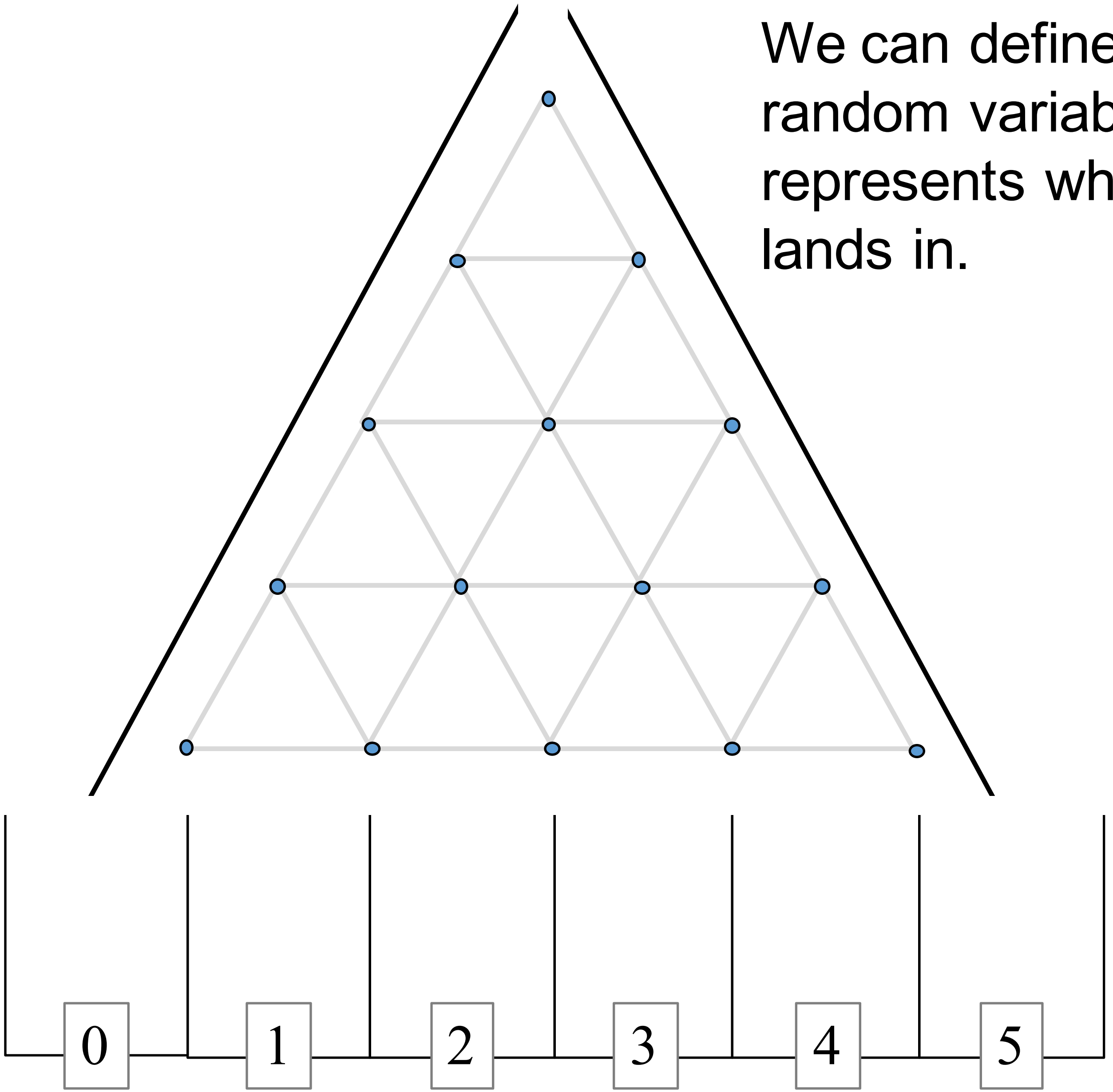
Galton Board



We can define an indicator random variable (***R***) which represents whether a particular marble goes right as a Bernoulli

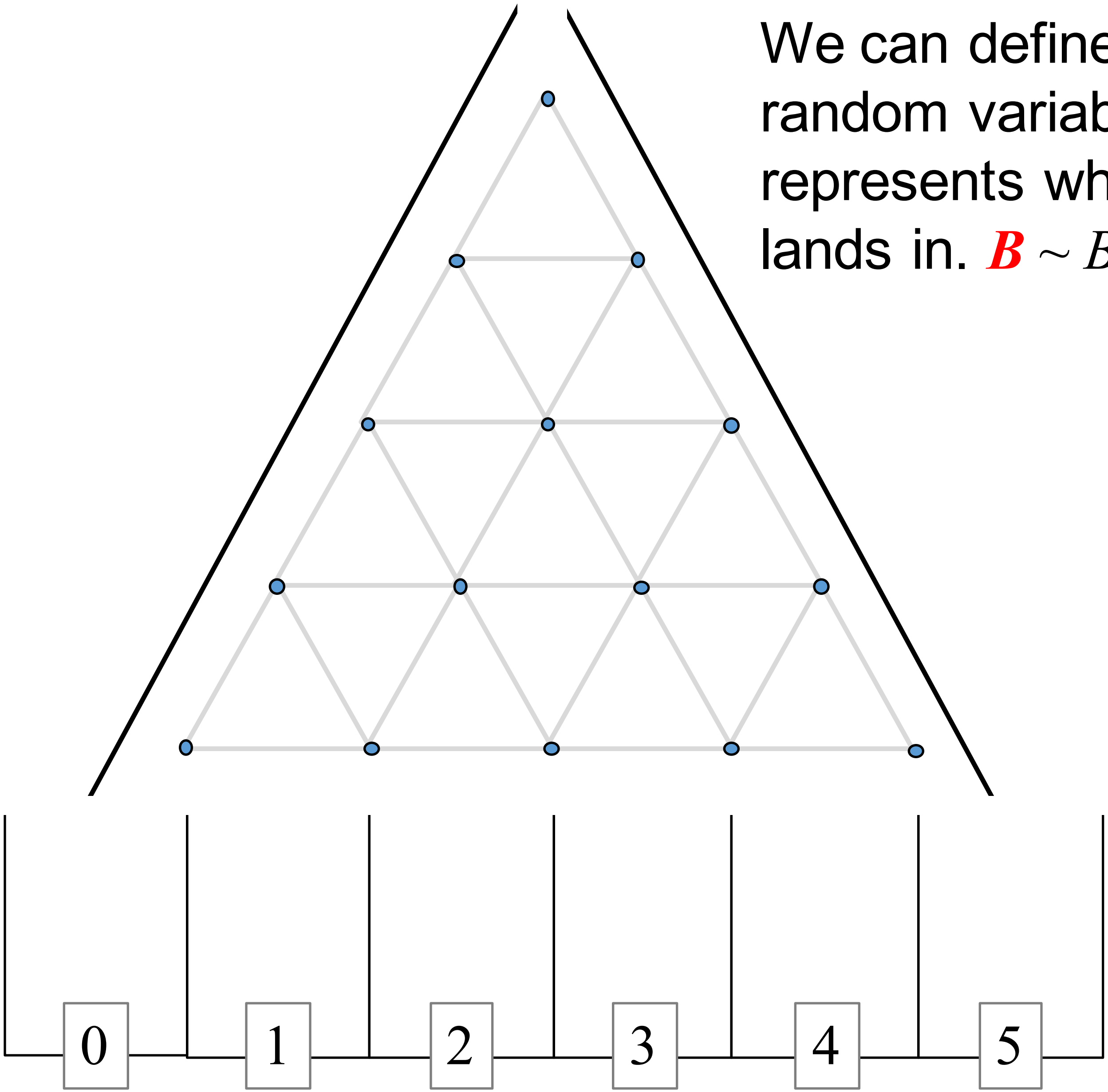
$R \sim Ber(0.5)$

Galton Board



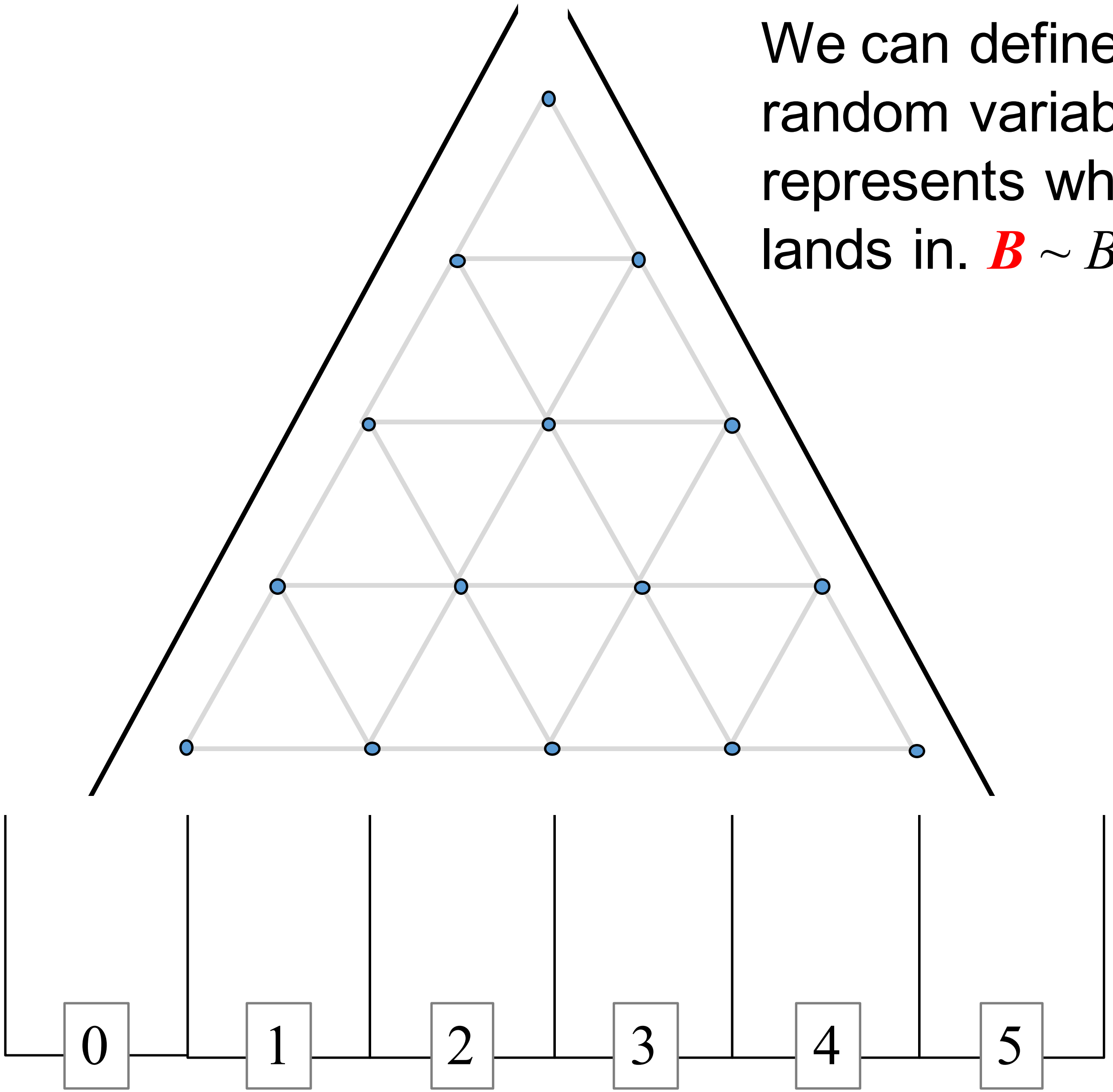
We can define an indicator random variable (***B***) which represents what bucket a marble lands in.

Galton Board



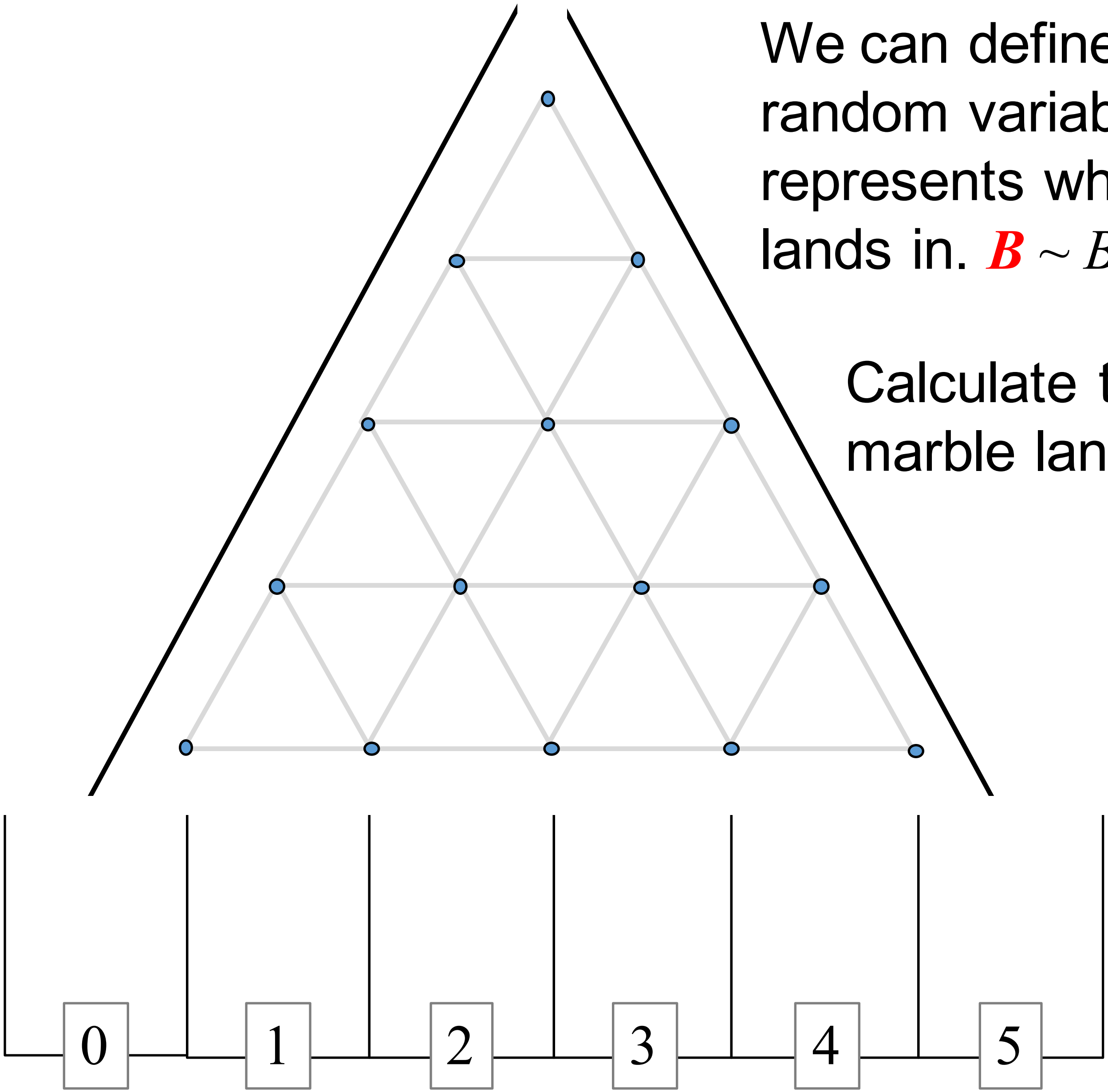
We can define an indicator random variable (***B***) which represents what bucket a marble lands in. $\boldsymbol{B} \sim Bin(\text{levels}, 0.5)$

Galton Board



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

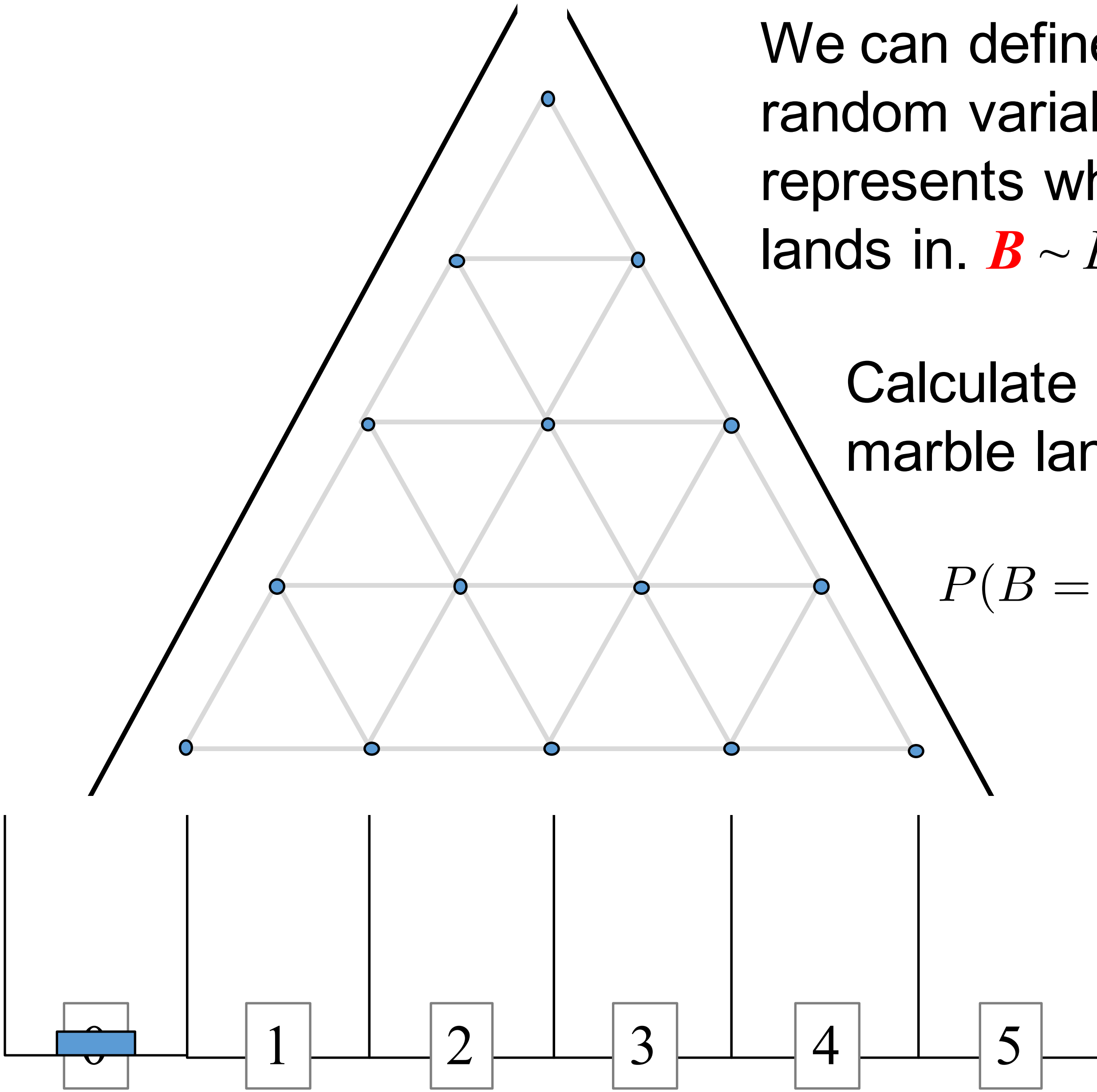
Galton Board



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

Calculate the probability of a marble landing in a bucket.

Galton Board

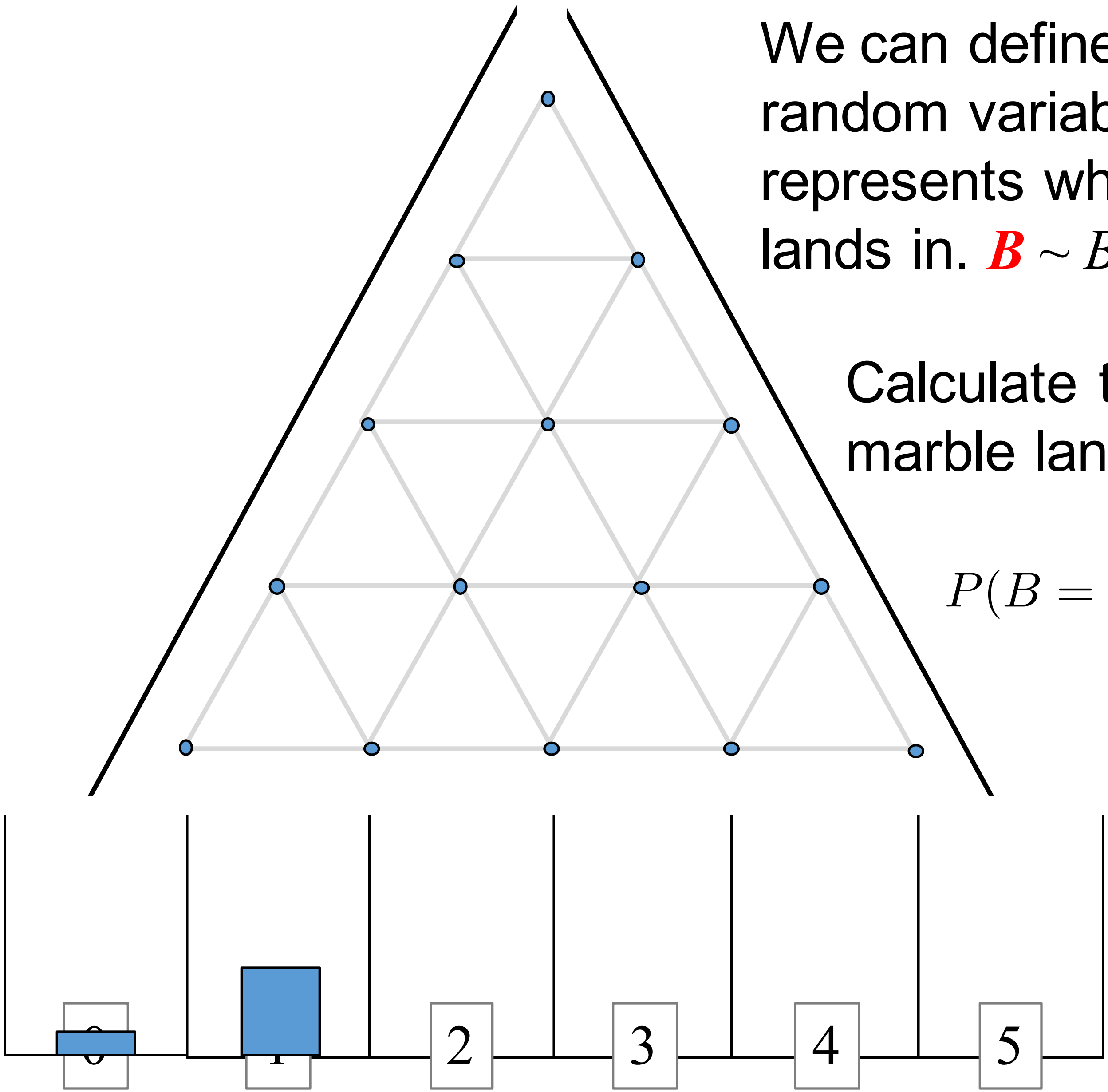


We can define an indicator random variable (***B***) which represents what bucket a marble lands in. $\boldsymbol{B} \sim Bin(5, 0.5)$

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

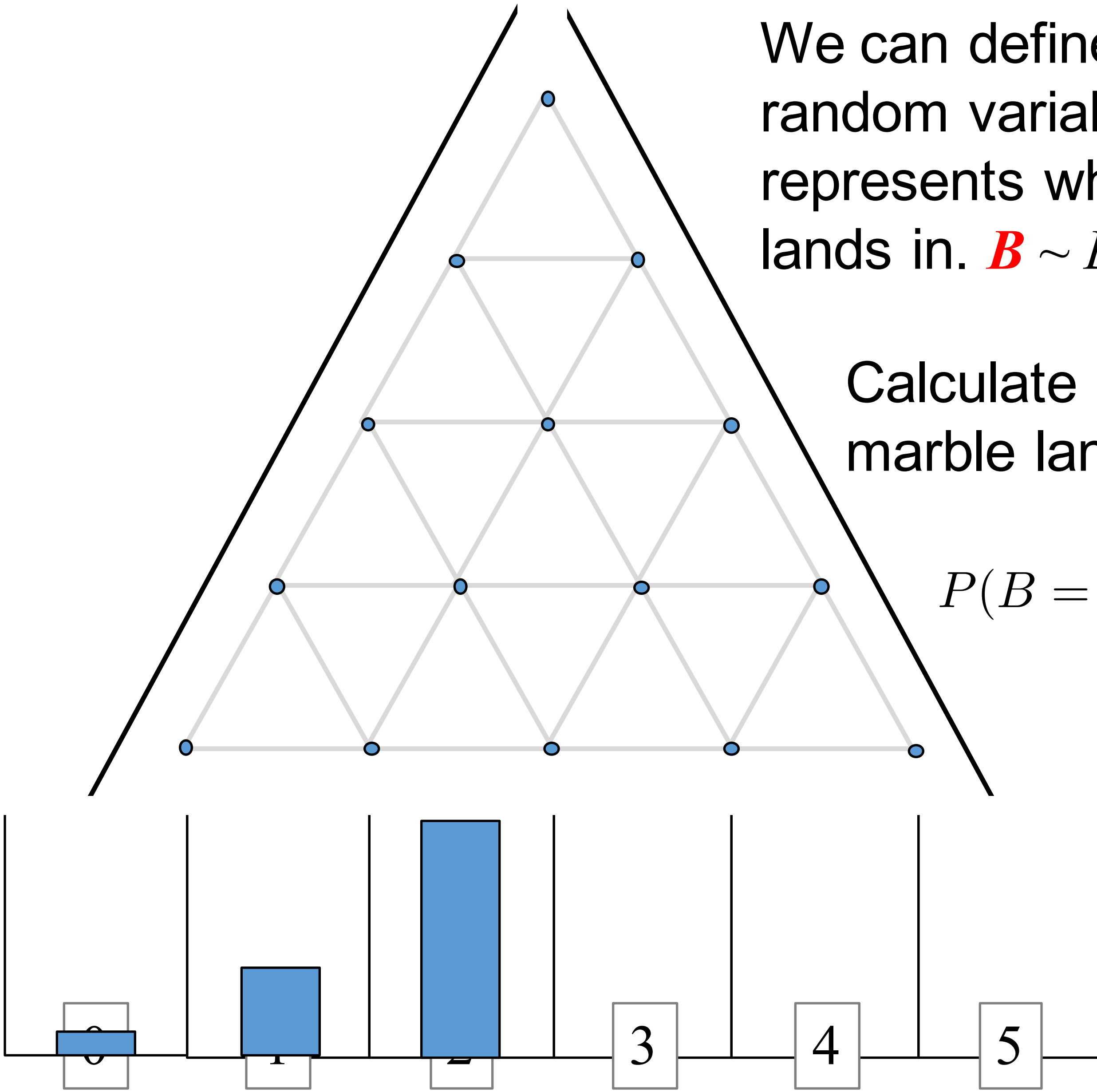


We can define an indicator random variable (***B***) which represents what bucket a marble lands in. ***B*** $\sim Bin(5, 0.5)$

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

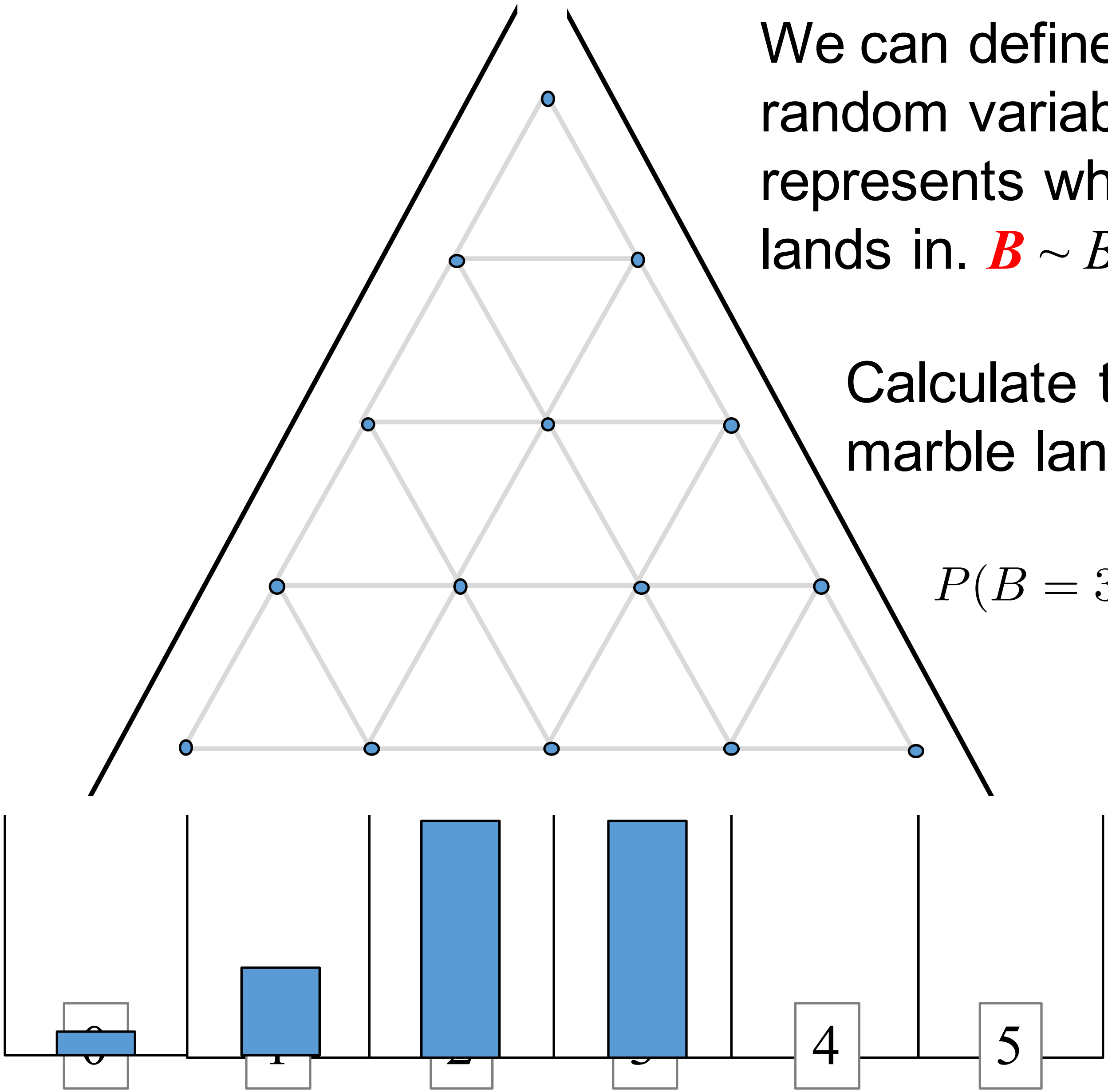


We can define an indicator random variable (***B***) which represents what bucket a marble lands in. ***B*** $\sim Bin(5, 0.5)$

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*** $\sim Bin(5, 0.5)$

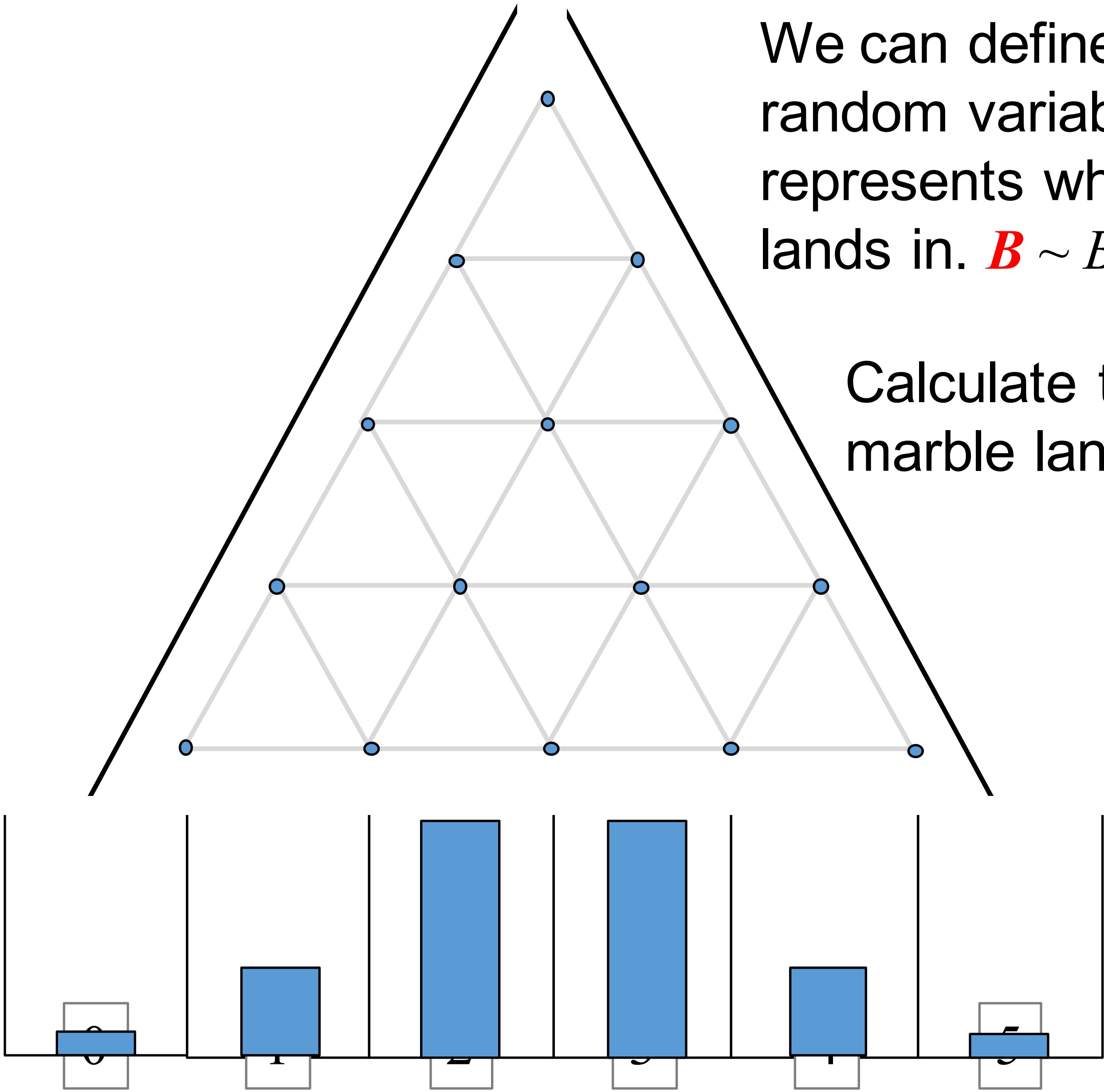
Calculate the probability of a marble landing in a bucket.

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

Galton Board

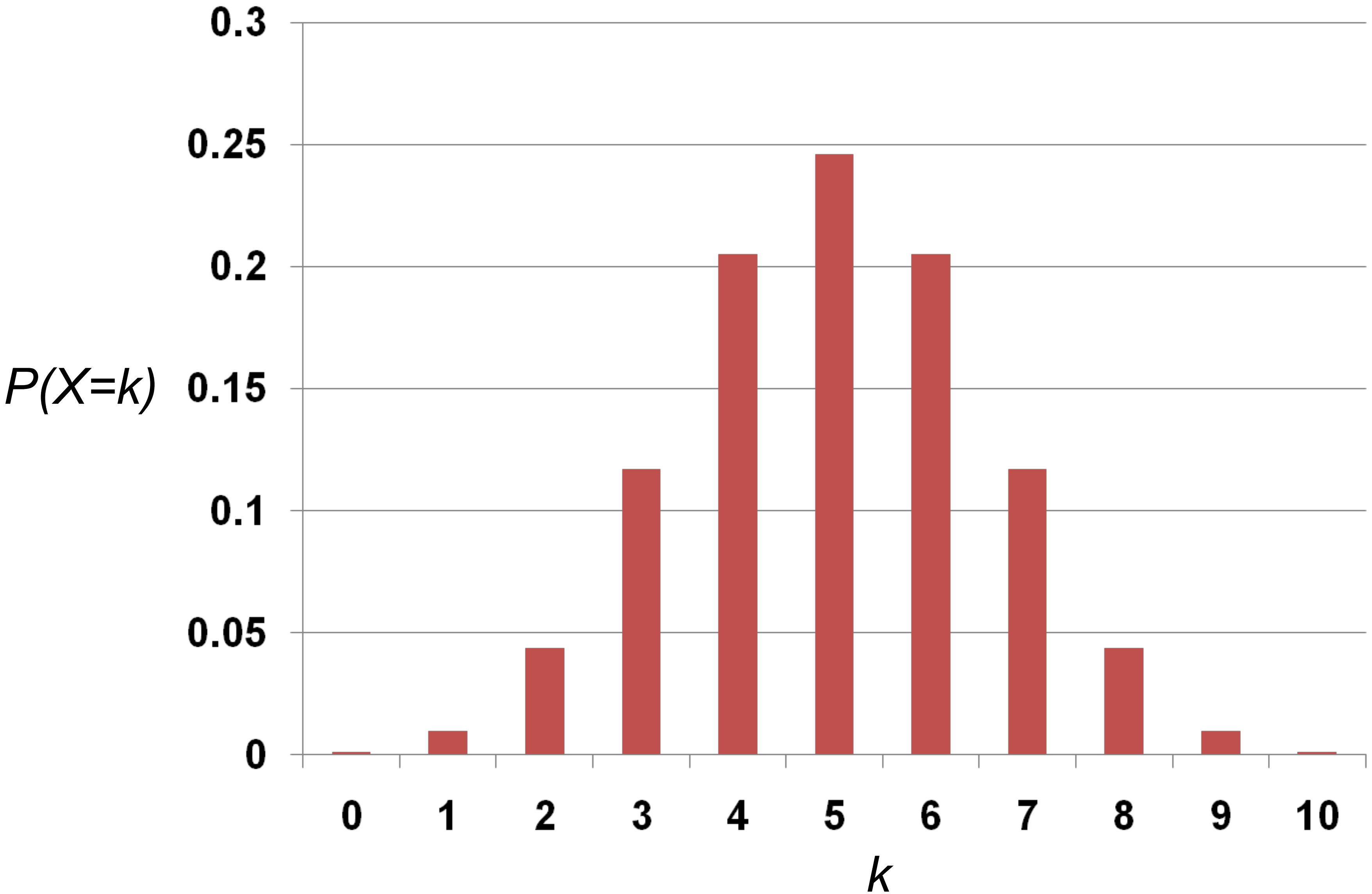
We can define an indicator random variable (***B***) which represents what bucket a marble lands in. ***B*** $\sim Bin(5, 0.5)$

Calculate the probability of a marble landing in a bucket.

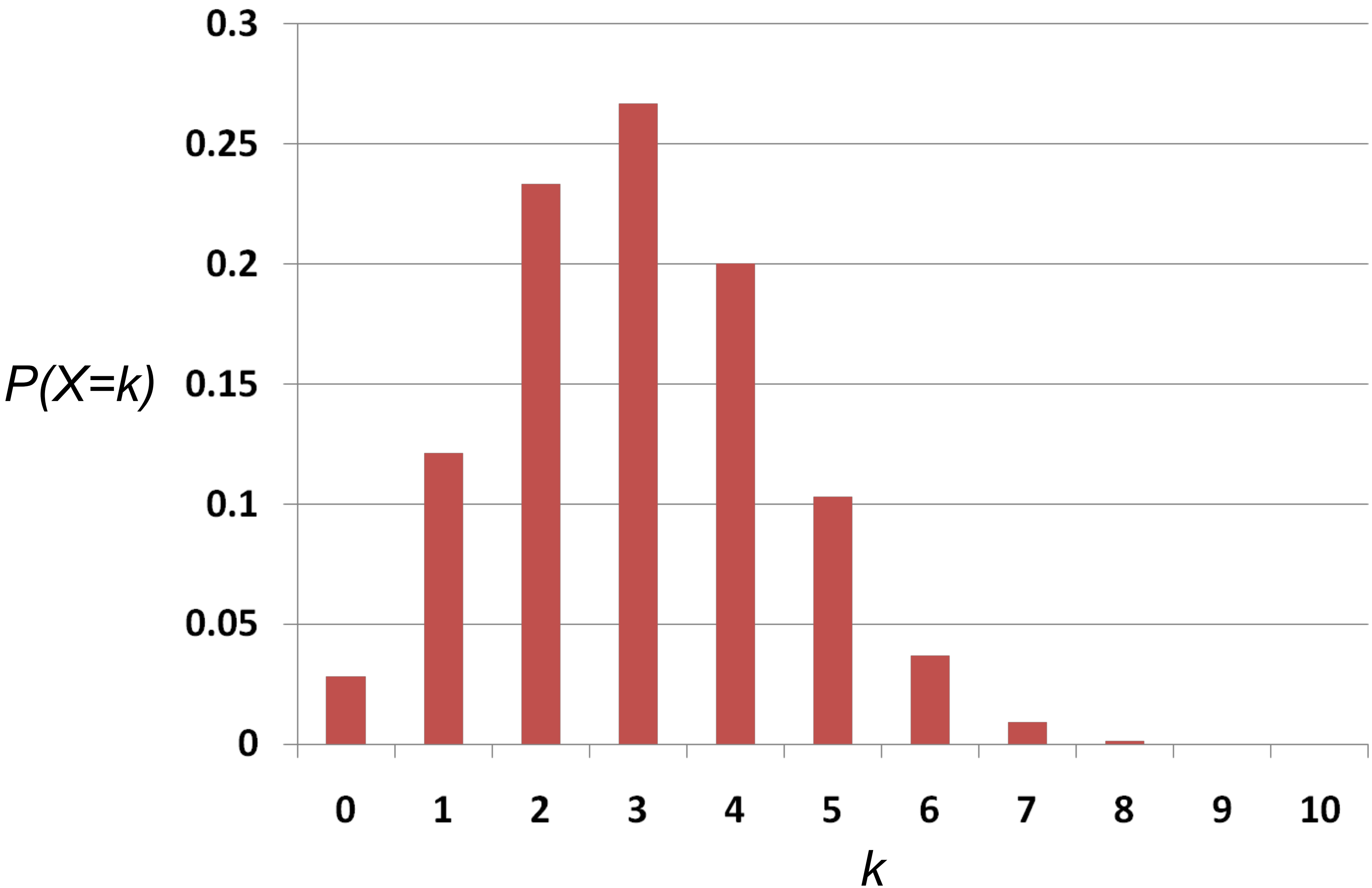


PDF

PMF for X ~ Bin(10, 0.5)



PMF for X ~ Bin(10, 0.3)



Properties of Bin(n, p)

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

- $P(X = i) = p(i) = \binom{n}{i} p^i (1 - p)^{n-i} \quad i = 0, 1, \ldots, n$
- $P(X \leq x) = F(x) = \sum_{i=0}^{x} p(i)$
- $\mathrm{E}[X] = np$
- $\mathrm{Var}(X) = np(1 - p)$
- $\mathrm{E}[X^2] = n^2 p^2 - np^2 + np$

- Note: $\mathrm{Ber}(p) = \mathrm{Bin}(1, p)$

I Really Want the Proof of Var :)

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


1001

Error Correcting Codes

- Error correcting codes
  - Have original 4 bit string to send over network
  - Add 3 “parity” bits, and send 7 bits total
  - Each bit independently corrupted (flipped) in transition with probability 0.1

Error Correcting Codes

Key for 7 bits

Send: 1110?

Receive: 1110000?

Receive: 1010100?

Flip set: $O_i \, E_j^C$
for all odd sets i
and even sets j

Error Correcting Codes

- Error correcting codes
  - Have original 4 bit string to send over network
  - Add 3 “parity” bits, and send 7 bits total
  - Each bit independently corrupted (flipped) in transition with probability 0.1
  - X = number of bits corrupted: $X \sim \text{Bin}(7, 0.1)$
  - But, parity bits allow us to correct at most 1 bit error
- P(a correctable message is received)?
  - $P(X = 0) + P(X = 1)$

Error Correcting Codes

- Using error correcting codes: X ~ Bin(7, 0.1)

$$P(X=0)=\binom{7}{0}(0.1)^0(0.9)^7 \approx 0.4783$$

$$P(X=1)=\binom{7}{1}(0.1)^1(0.9)^6 \approx 0.3720$$

  - P(X = 0) + P(X = 1) = 0.8503
- What if we didn't use error correcting codes?
  - X ~ Bin(4, 0.1)
  - P(correct message received) = P(X = 0)

$$P(X=0)=\binom{4}{0}(0.1)^0(0.9)^4 = 0.6561$$

- Using error correction improves reliability ~30%!

Genetic Inheritance

- Person has 2 genes for trait (eye color)
  - Child receives 1 gene (equally likely) from each parent
  - Child has brown eyes if either (or both) genes brown
  - Child only has blue eyes if both genes blue
  - Brown is “dominant” (d) , Blue is “recessive” (r)
  - Parents each have 1 brown and 1 blue gene
- 4 children, what is P(3 children with brown eyes)?
  - Child has blue eyes: p = (½) (½) = ¼ (2 blue genes)
  - P(child has brown eyes) = 1 – (¼) = 0.75
  - X = # of children with brown eyes. X ~ Bin(4, 0.75)

$$P(X = 3) = \binom{4}{3}(0.75)^3(0.25)^1 \approx 0.4219$$

Power of Your Vote

- Is it better to vote in small or large state?
  - Small: more likely your vote changes outcome
  - Large: larger outcome (electoral votes) if state swings
  - a (= 2n) voters equally likely to vote for either candidate
  - You are deciding (a + 1)st vote

$$P(2n \text{ voters tie}) = \binom{2n}{n}\left(\frac{1}{2}\right)^n\left(\frac{1}{2}\right)^n = \frac{(2n)!}{n!n!2^{2n}}$$

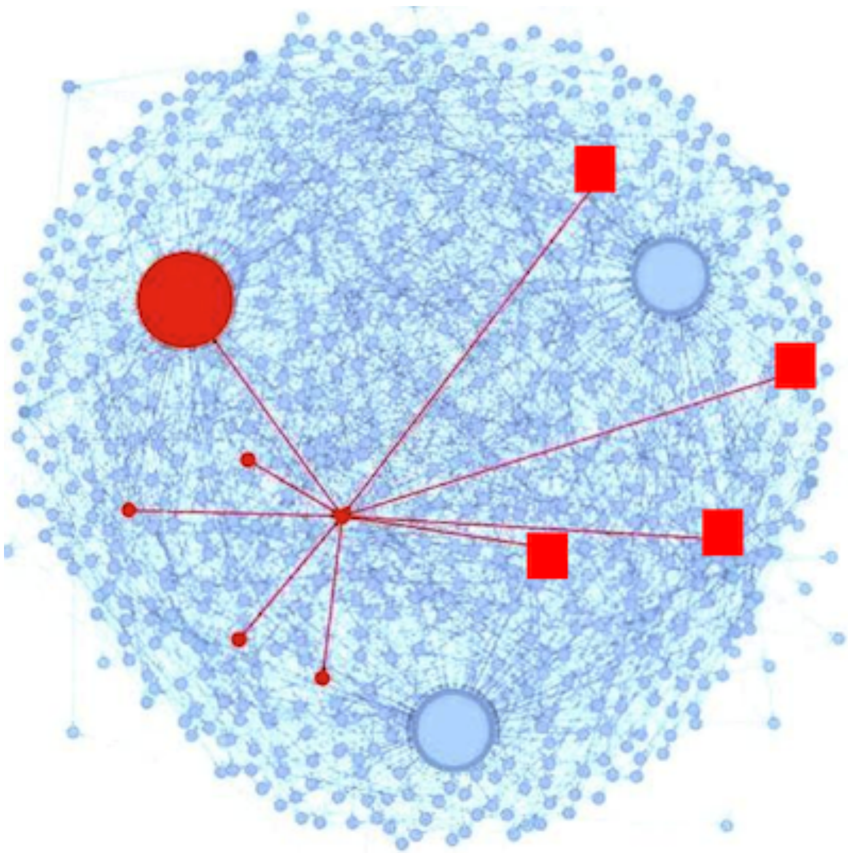
  - Use Stirling's Approximation: $n! \approx n^{n+1/2}e^{-n}\sqrt{2\pi}$

$$P(2n \text{ voters tie}) \approx \frac{(2n)^{2n+1/2}e^{-2n}\sqrt{2\pi}}{n^{2n+1}e^{-2n}2\pi 2^{2n}} = \frac{1}{\sqrt{n\pi}}$$

  - Power = P(tie) * Elec. Votes = $\frac{1}{\sqrt{(a/2)\pi}}(ac) = \frac{c\sqrt{2a}}{\sqrt{\pi}}$
  - Larger state = more power

Is Peer Grading Accurate Enough?

Looking ahead



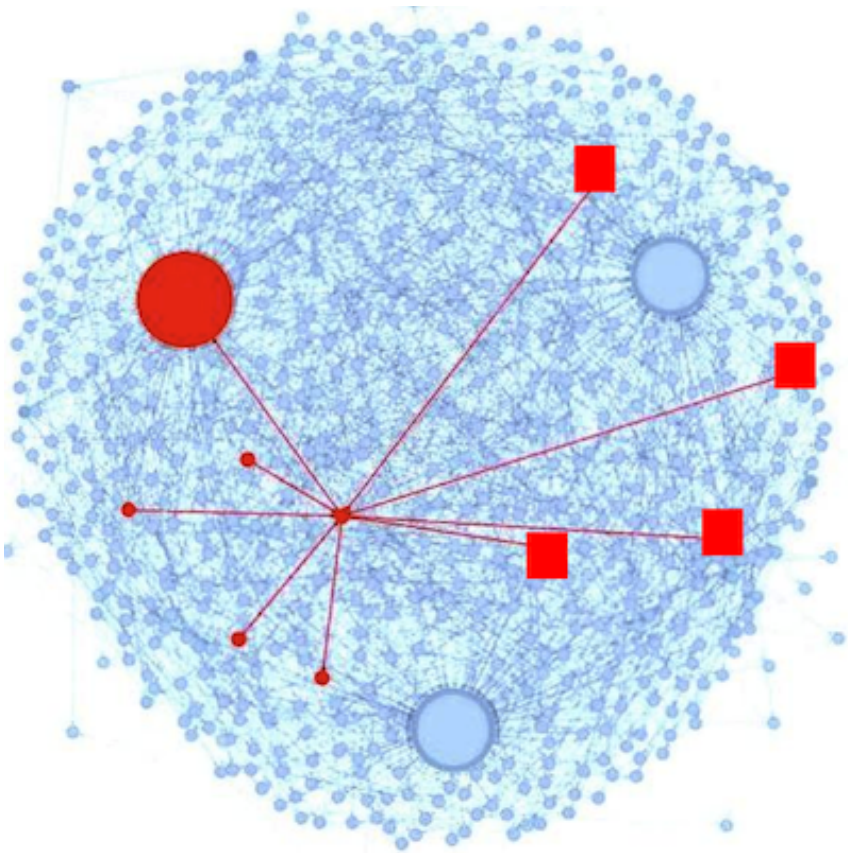
Peer Grading on Coursera HCI.

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

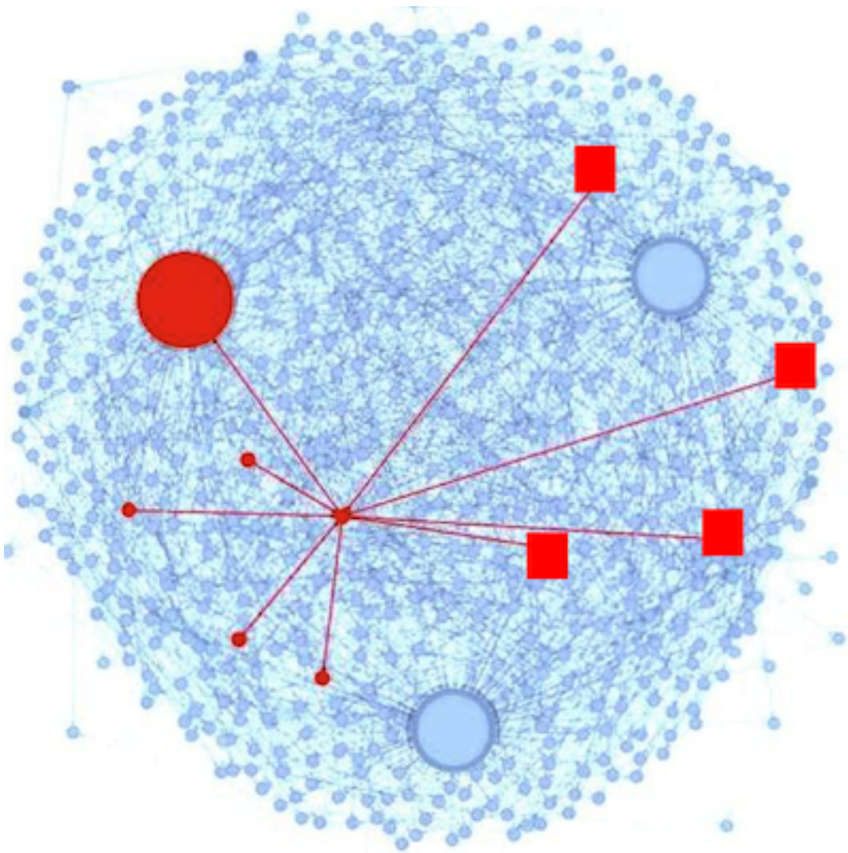
$$s_i \sim \text{Bin}(\text{points}, \theta)$$

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

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

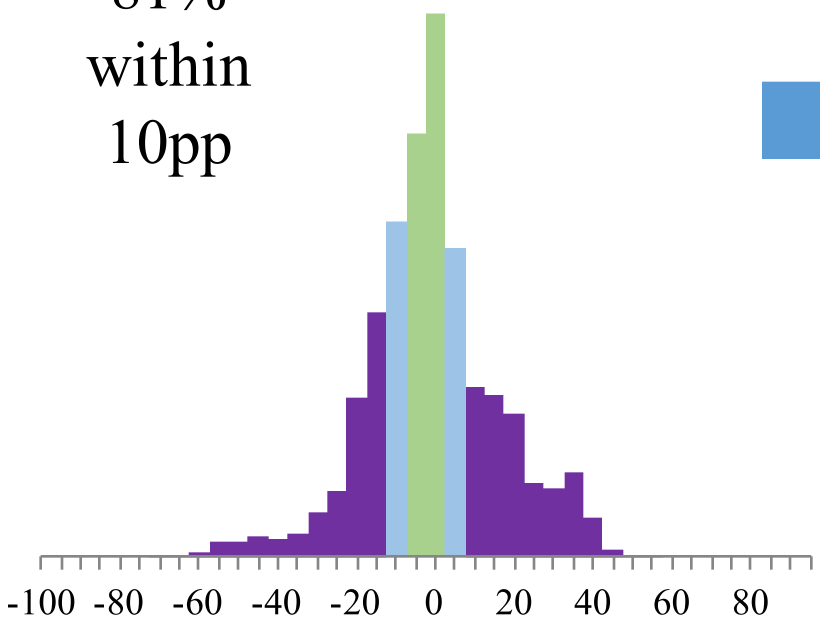
Inference or Machine Learning

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

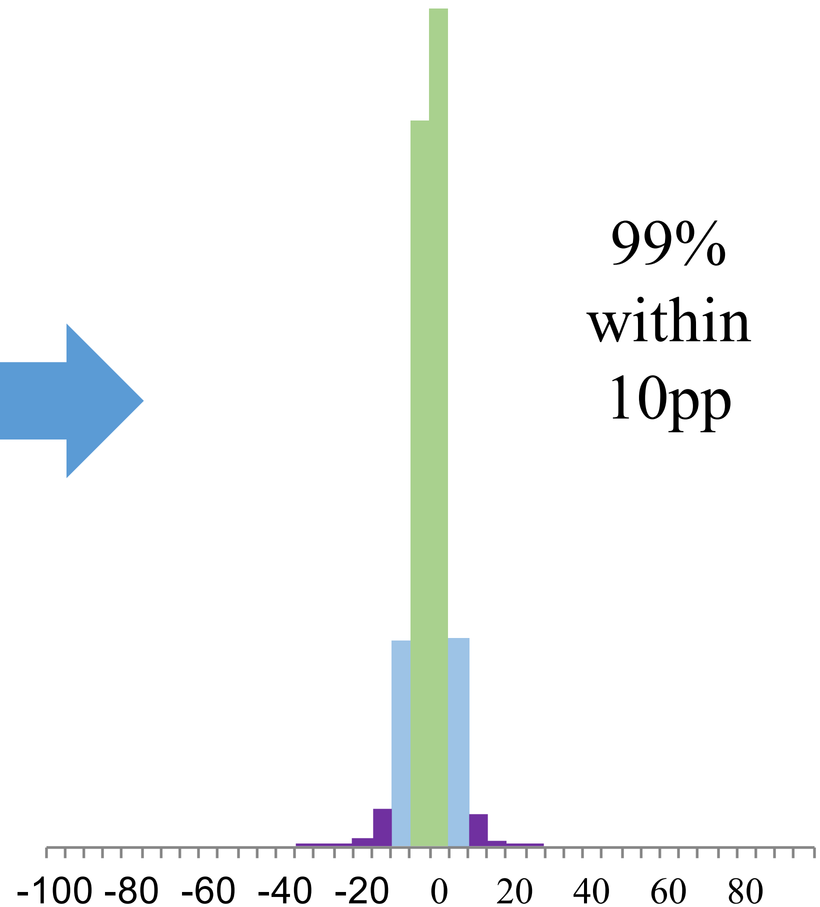
Yes, With Probabilistic Modelling

Before:

81%
within
10pp



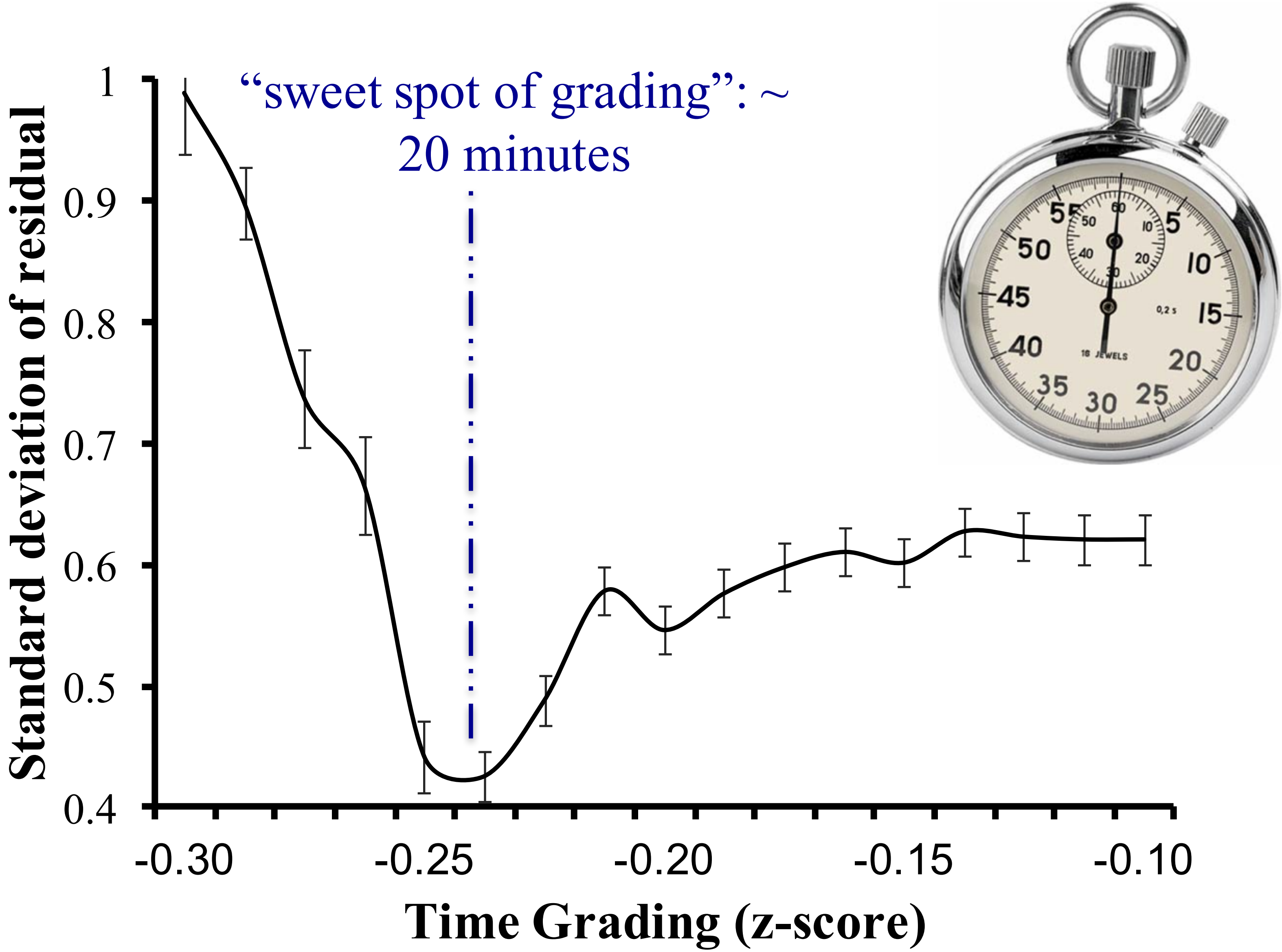
After:



99%
within
10pp

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

Grading Sweet Spot



Voilà, c'est tout