Poisson

CS109, Stanford University

Apply to be a section leader for CS106A/B – add in the slides

Review

Natural Exponent Definition

Natural Exponent def:

$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n = e^{-\lambda}$$

Jacob
Bernoulli

https://en.wikipedia.org/wiki/E_(mathematical_constant)

Binomial Random Variable

```
(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)
```

The number of **successes**, in *n* independent **trials**, where each **trial** is a **success** with probability p:

Declare a Random Variable to be Binomial

Number of successes

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

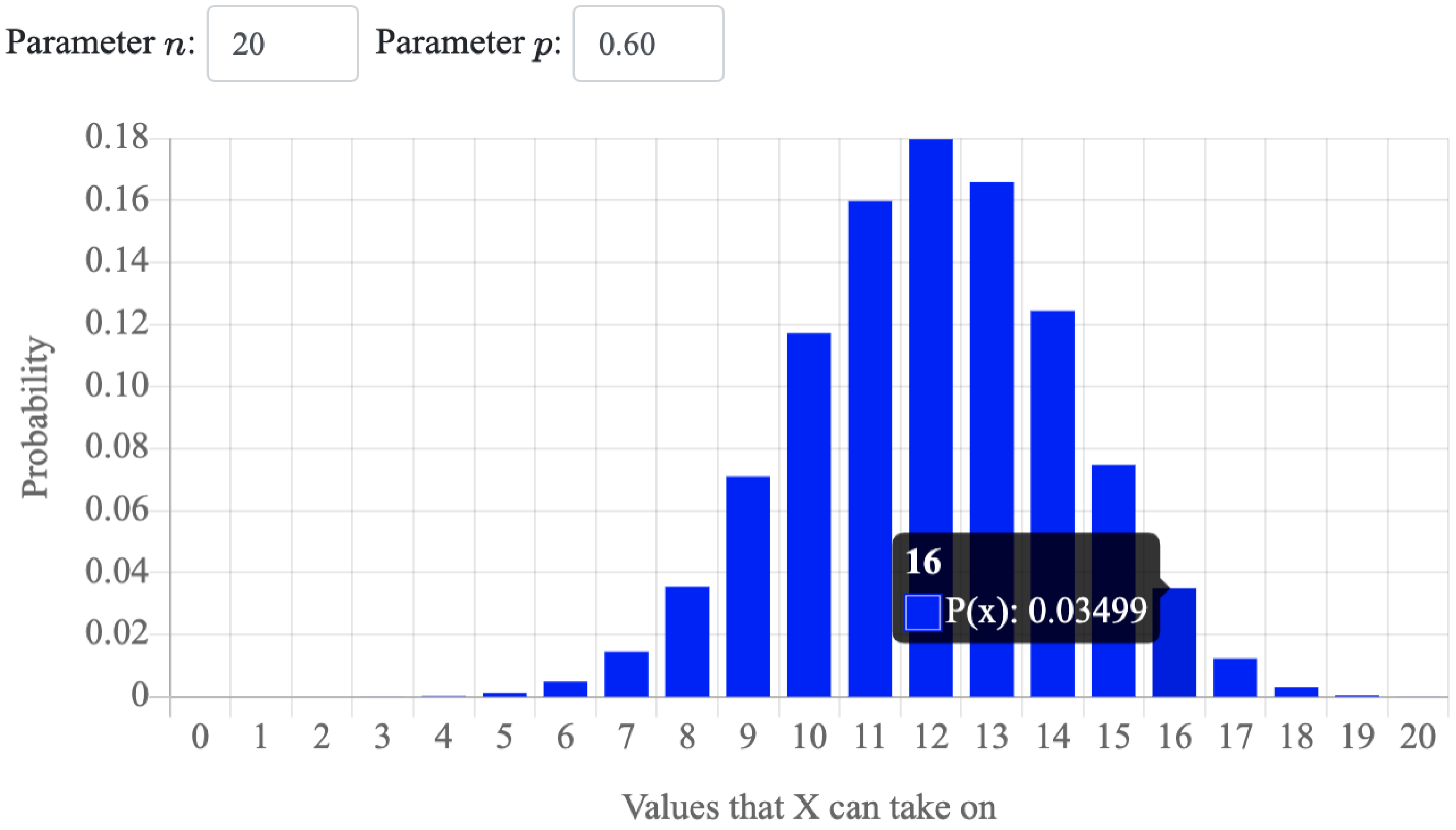
Probability Mass Function for a Binomial

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

Probability that there are k successes

* This is also called the binomial term

The PMF as a Graph: X ~ Bin(n = 20, p = 0.6)



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

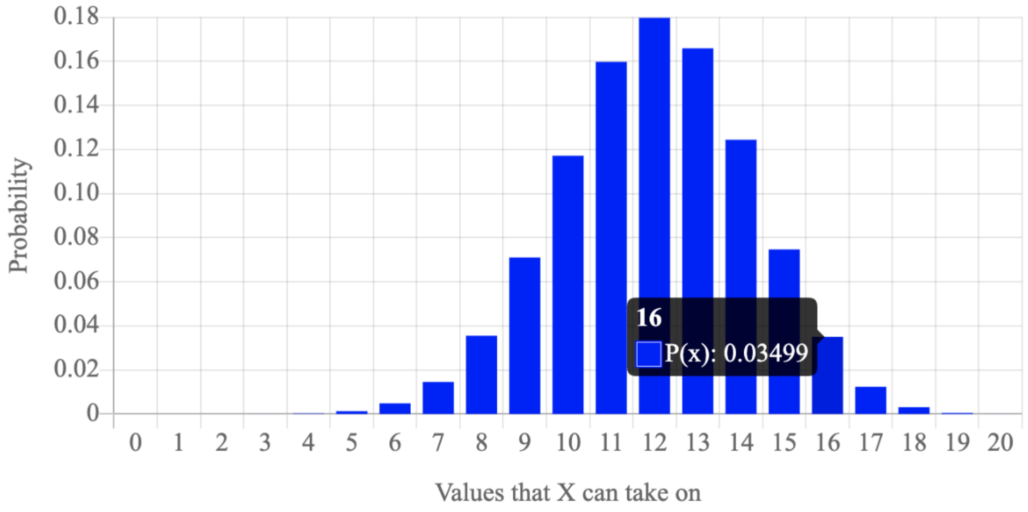
PMF equation: $\Pr(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$

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

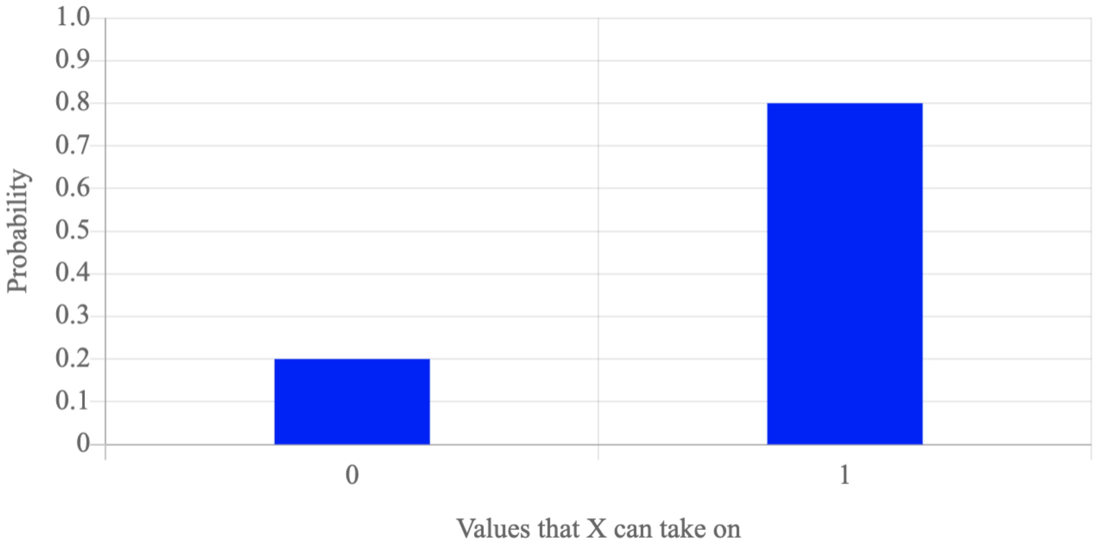
PMF equation: $\Pr(X = x) = \begin{cases} p & \text{if } x = 1 \\ 1 - p & \text{if } x = 0 \end{cases}$

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

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

PMF graph:

Parameter p: 0.80



What if we could summarize the whole beautiful PMF into a single number?

Expected Value

The probability of
that value

The value

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

Loop over all values x that X can take on

St. Petersburg Paradox

The Game:

- We have a fair coin (lands on heads with $p = 0.5$)
- Let n = number of coin flips to get the first heads
- You will win: $\$2^n$

How much would you pay to play?

Let X be your winnings.

$$E[X] = 2^1\left(\frac{1}{2}\right)^1 + 2^2\left(\frac{1}{2}\right)^2 + 2^3\left(\frac{1}{2}\right)^3 + \cdots = \sum_{i=1}^{\infty} 1 = \infty$$

What if you could play this game for only $1000…but just once?

St. Petersburg Paradox

The Game:

- We have a fair coin (lands on heads with $p = 0.5$)
- Let n = number of coin flips to get the first heads
- You will win: $\$2^n$
- If you win over $65,536 **I leave the country**.

How much would you pay to play?

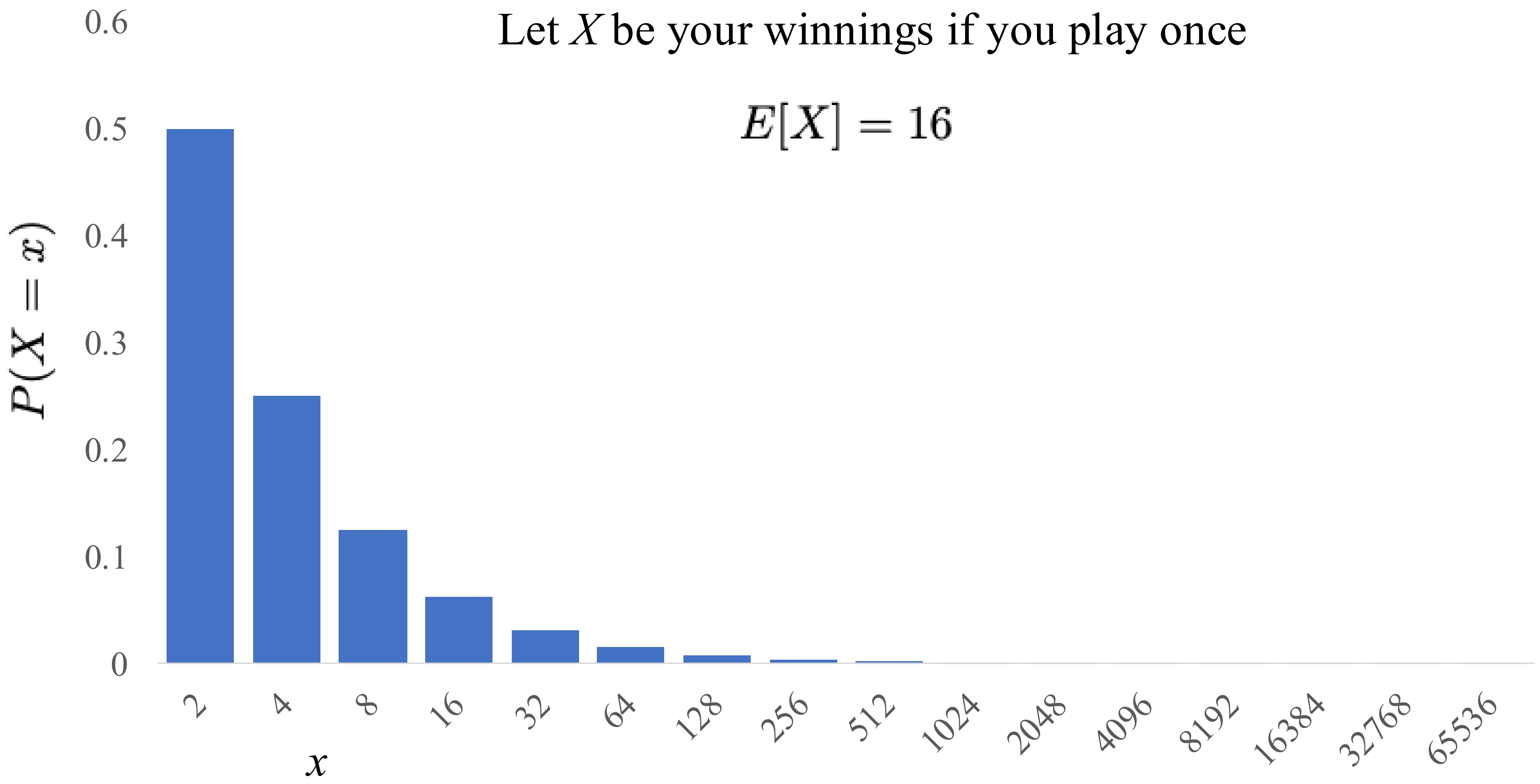
Let X be your winnings.

$$E[X] = 2^1\left(\frac{1}{2}\right)^1 + 2^2\left(\frac{1}{2}\right)^2 + \cdots + 2^{16}\left(\frac{1}{2}\right)^{16} = \sum_{i=1}^{16} 1 = 16$$

St Petersburg Probability Mass Function

Let X be your winnings if you play once

$$E[X] = 16$$



Properties of Expectation (more on this later)

Linearity:

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

Expectation of a sum is the sum of expectations

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

Unconscious statistician:

$$E[g(X)] = \sum_{x \in X} g(x) P(X = x)$$

LOUTS Examples

Unconscious statistician:

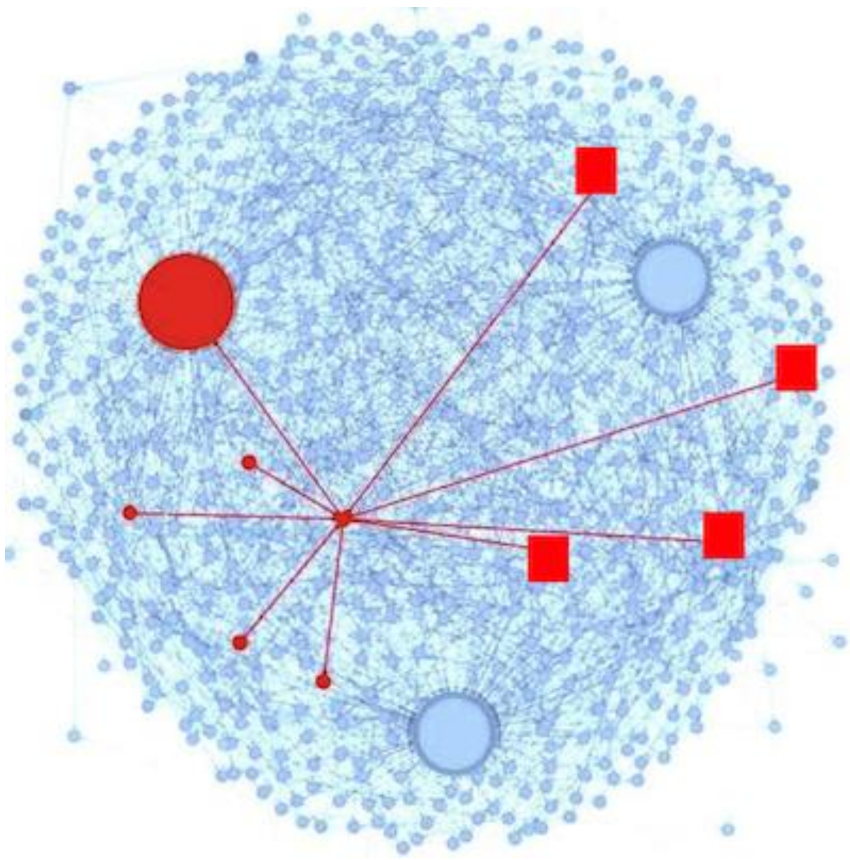
$$E[g(X)] = \sum_{x \in X} g(x) P(X = x)$$

$$g(x) = x^2 \qquad E[X^2] = \sum_{x \in X} x^2 P(X = x)$$

$$g(x) = (x - 4)^2 \qquad E[(X-4)^2] = \sum_{x \in X} (x - 4)^2 P(X = x)$$

End Review

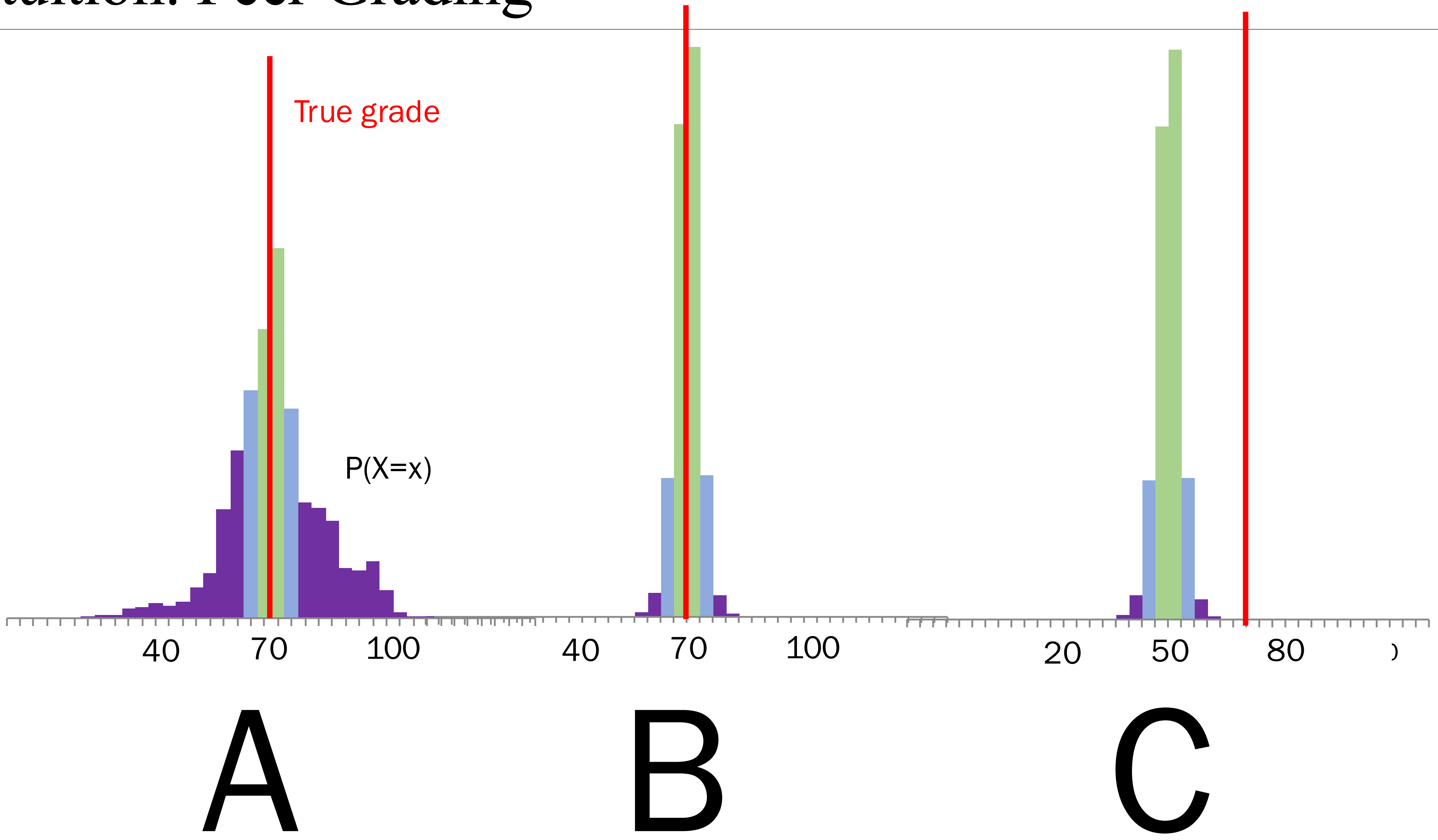
Intuition: Peer Grading



Peer Grading on Coursera HCI.

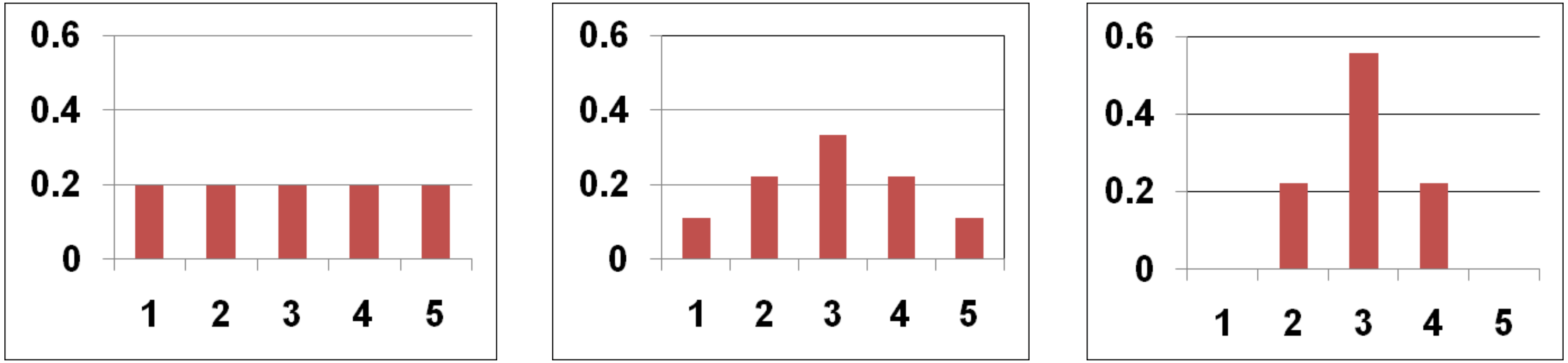
31,067 peer grades for 3,607 students.

Intuition: Peer Grading



Intuition: Measure of Spread

Consider the following 3 distributions (PMFs)



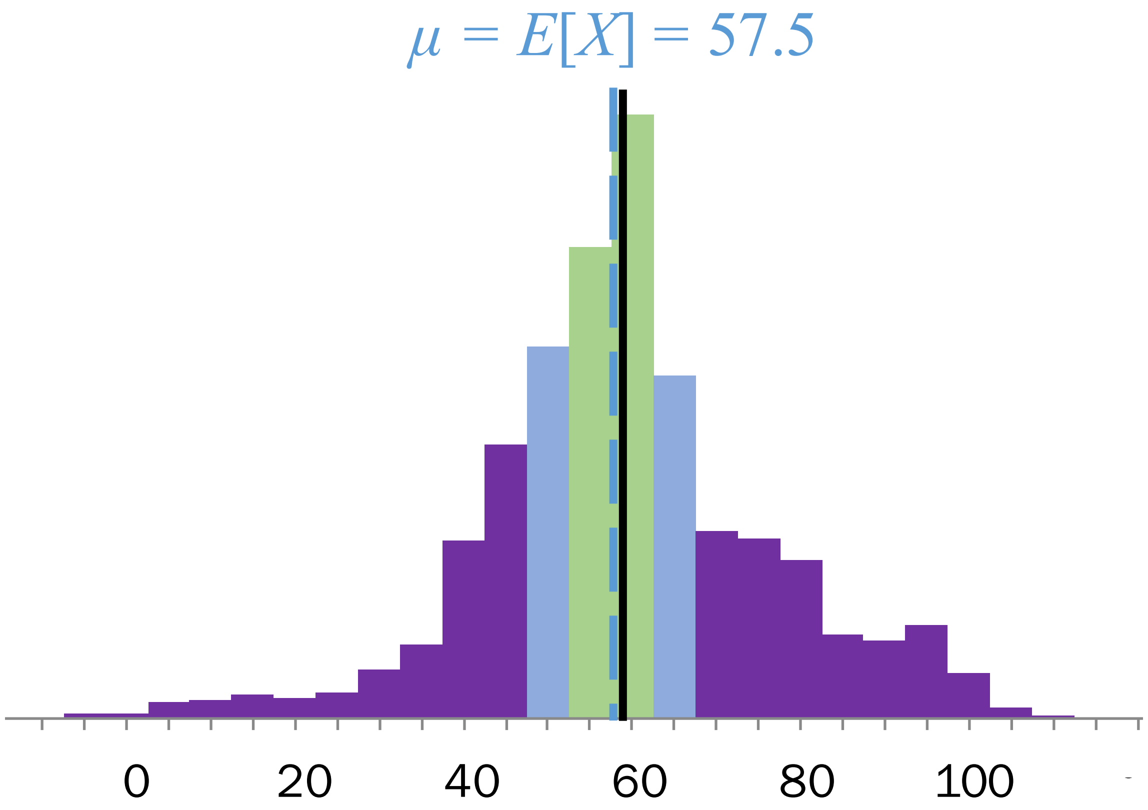
All have the same expected value, E[X] = 3

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

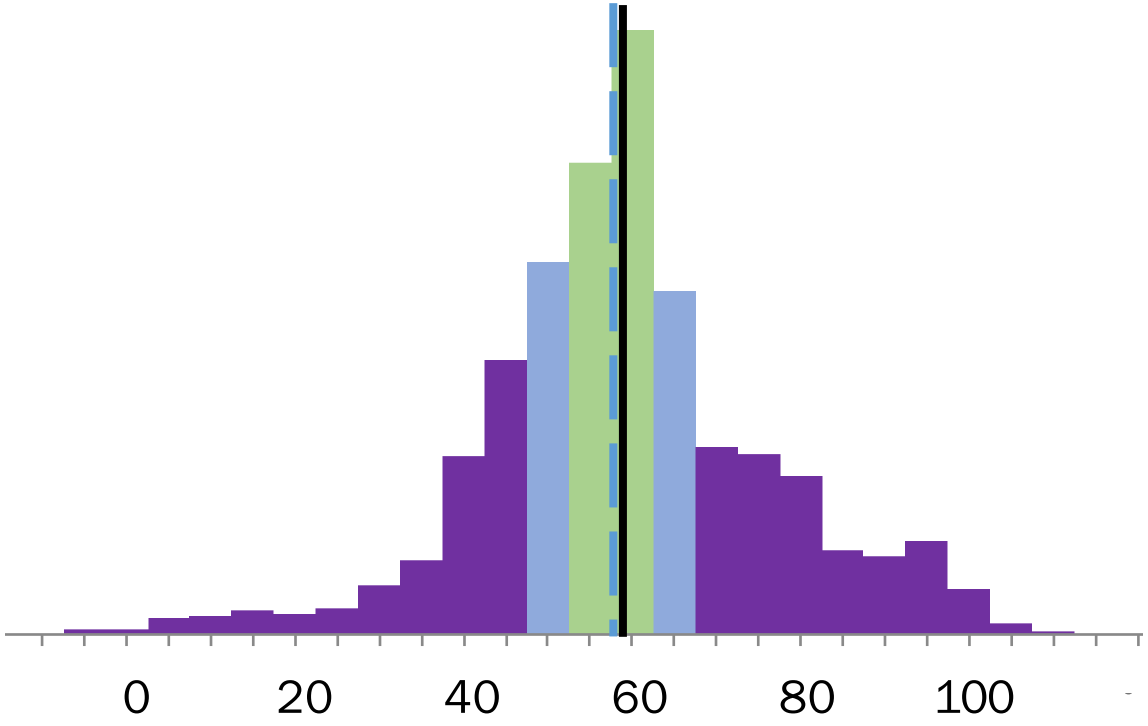


Peer grading in Coursera HCI

Let X be a random variable that represents a peer grade

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

$\mu = E[X] = 57.5$

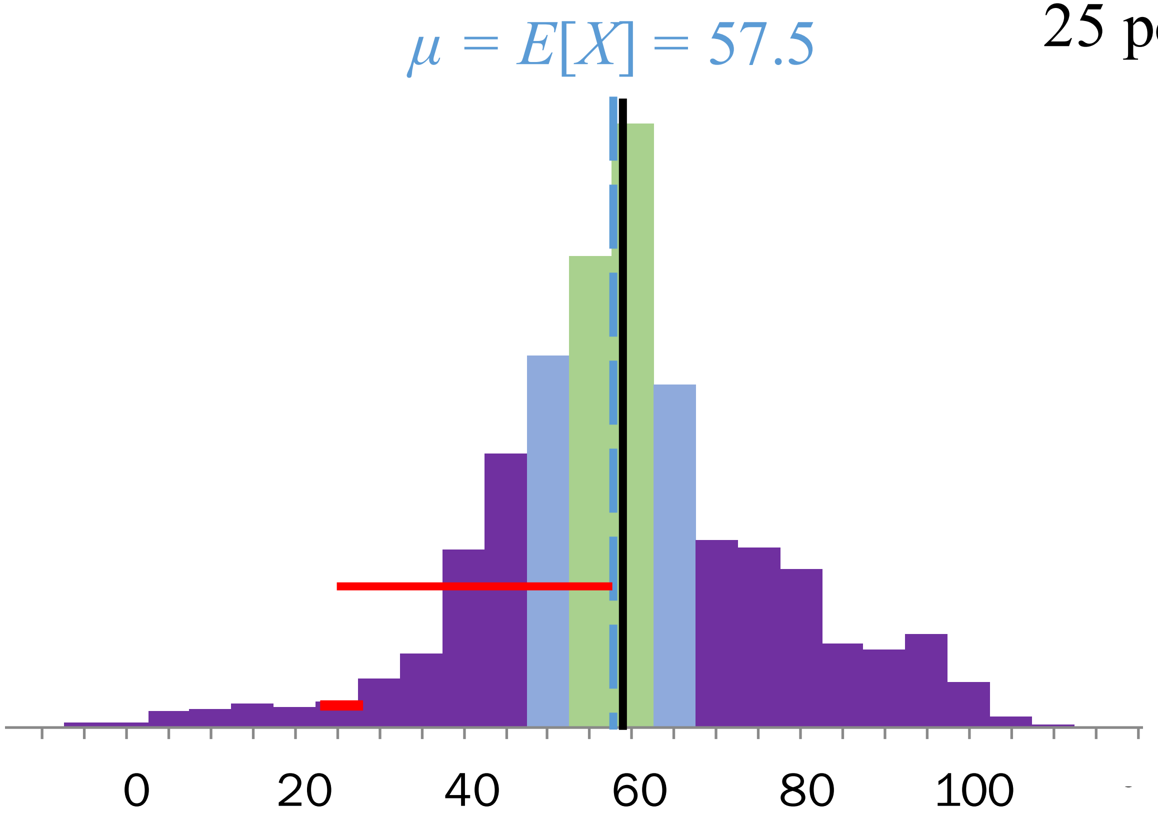


Peer grading in Coursera HCI

Let X be a random variable that represents a peer grade

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

x	$(x-\mu)^2$	$P(X=x)$
25 points	1056 points^2	0.02

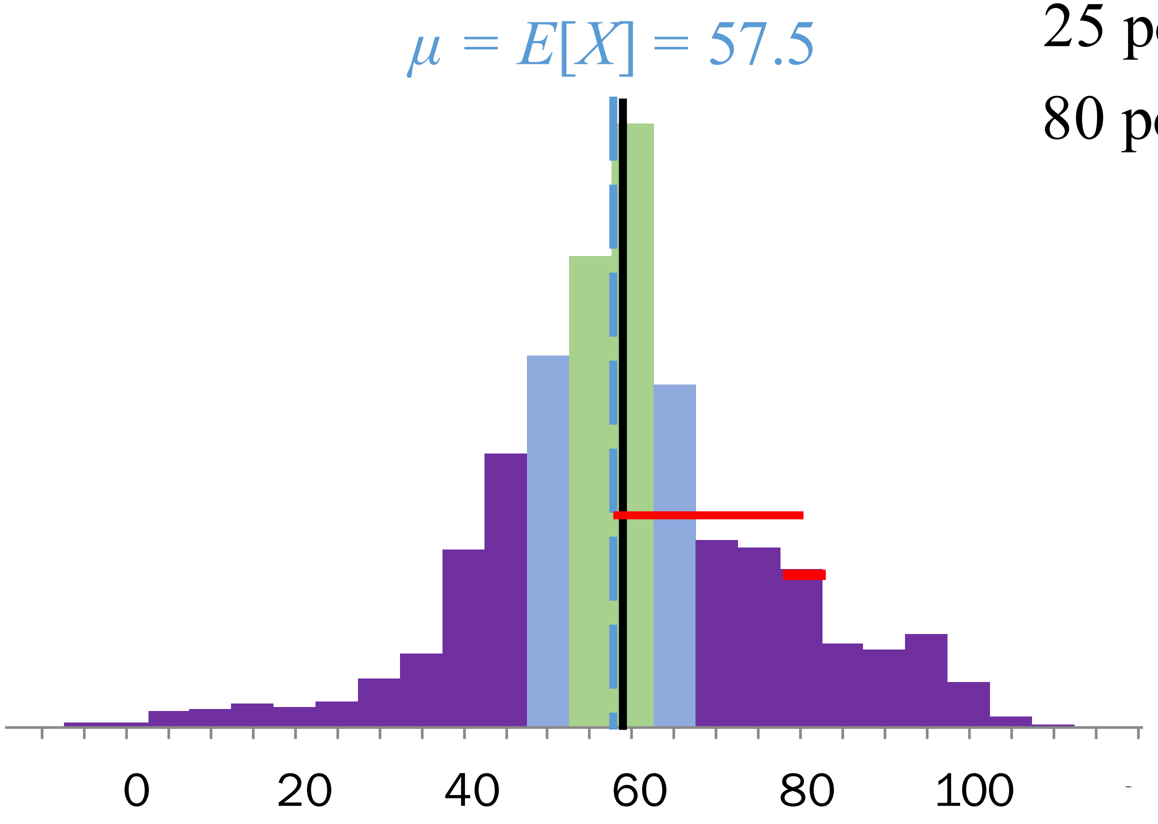


Peer grading in Coursera HCI

Let X be a random variable that represents a peer grade

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

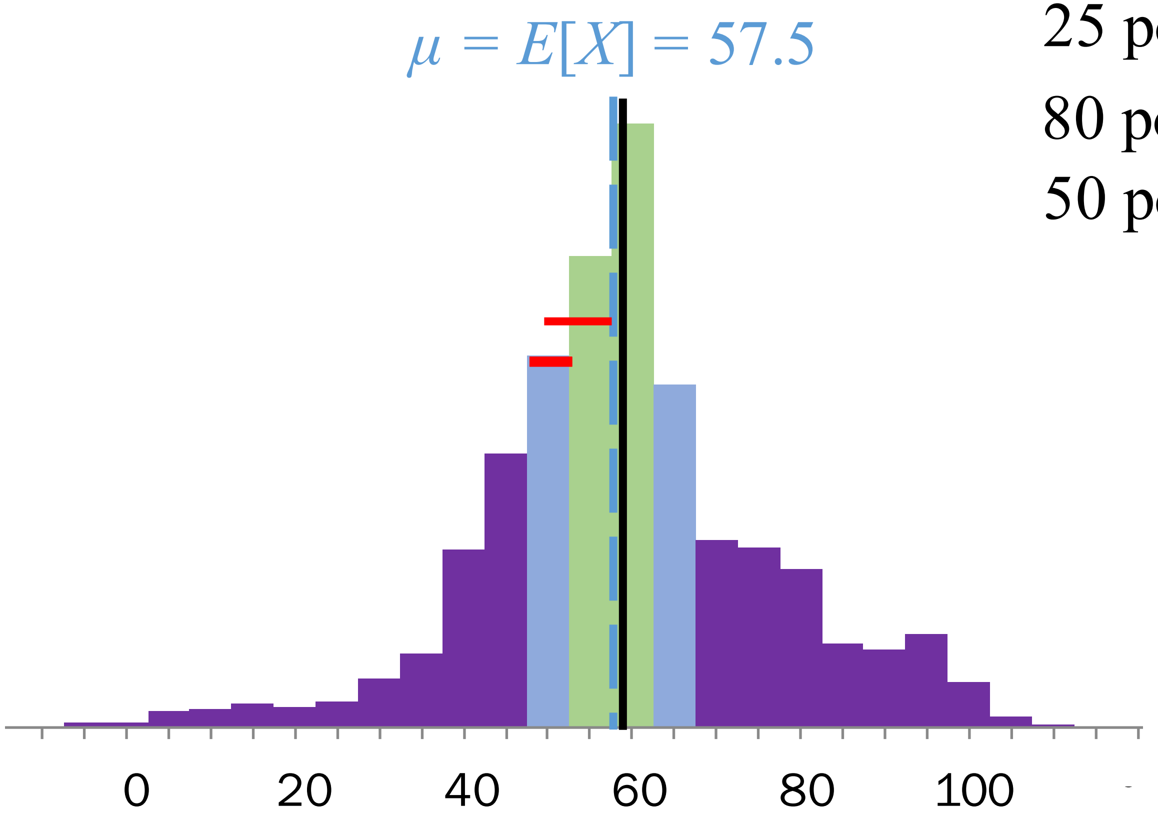
x	$(x-\mu)^2$	$P(X=x)$
25 points	1056 points^2	0.02
80 points	506 points^2	0.09



Peer grading in Coursera HCI

Let X be a random variable that represents a peer grade

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

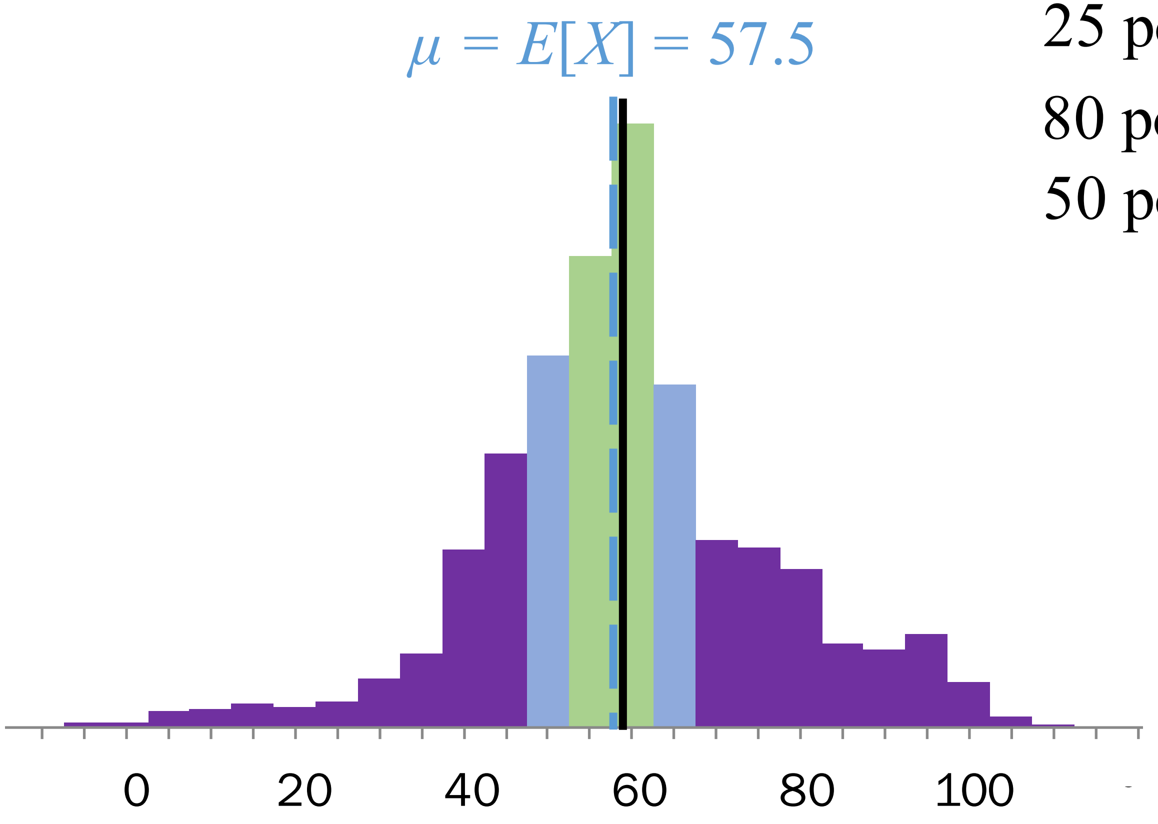


x	$(x - \mu)^2$	$P(X = x)$
25 points	1056 points^2	0.02
80 points	506 points^2	0.09
50 points	56 points^2	0.12

Peer grading in Coursera HCI

Let X be a random variable that represents a peer grade

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



x	$(x - \mu)^2$	$P(X = x)$
25 points	1056 points^2	0.02
80 points	506 points^2	0.09
50 points	56 points^2	0.12
	…	

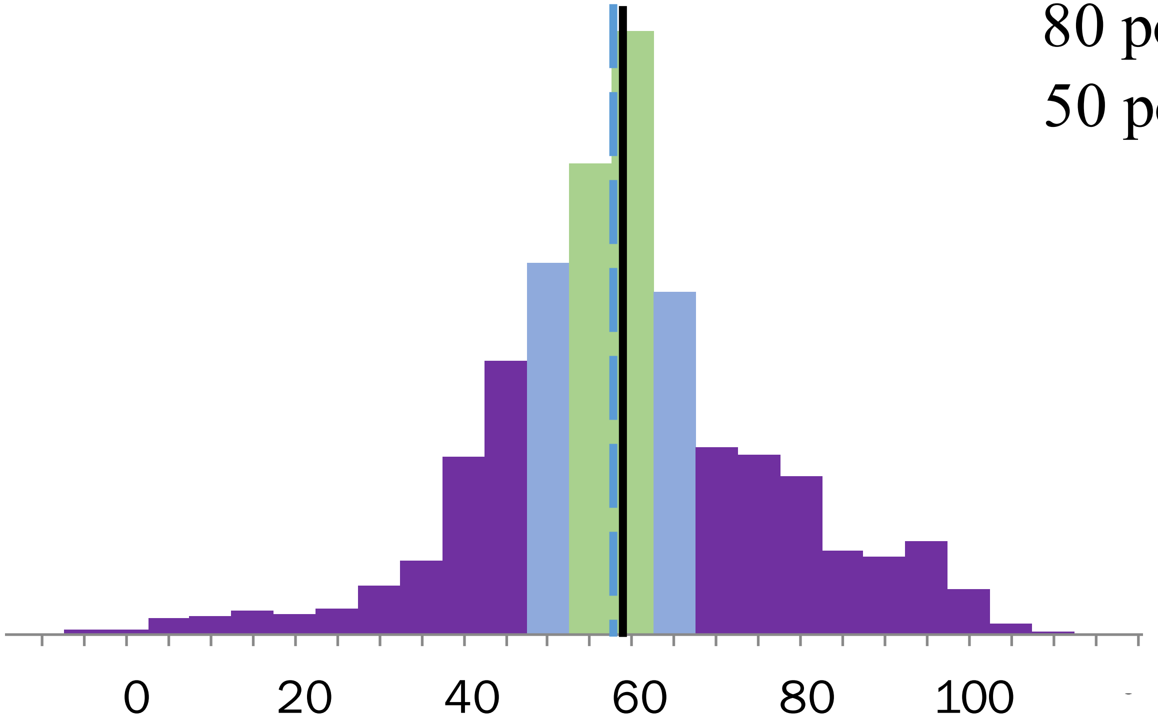
$$E\,[(X - \mu)^2] = 52 \text{ points}^2$$

Peer grading in Coursera HCI

Let X be a random variable that represents a peer grade

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

$\mu = E[X] = 57.5$

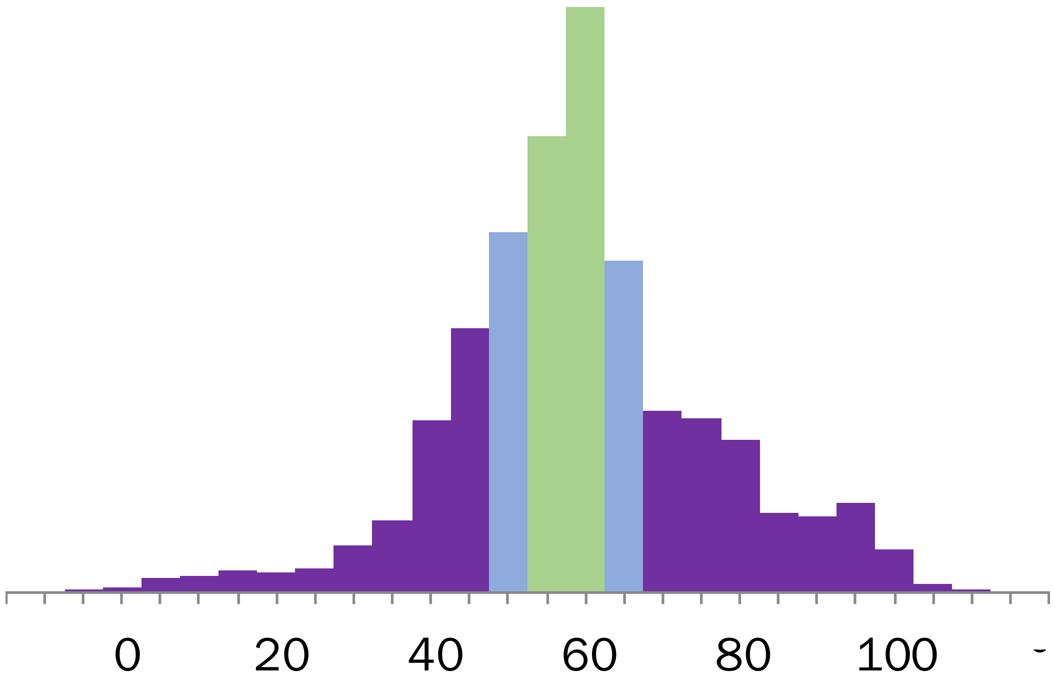


x	$(x-\mu)^2$	$P(X=x)$
25 points	1056 points^2	0.02
80 points	506 points^2	0.09
50 points	56 points^2	0.12
	…	

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

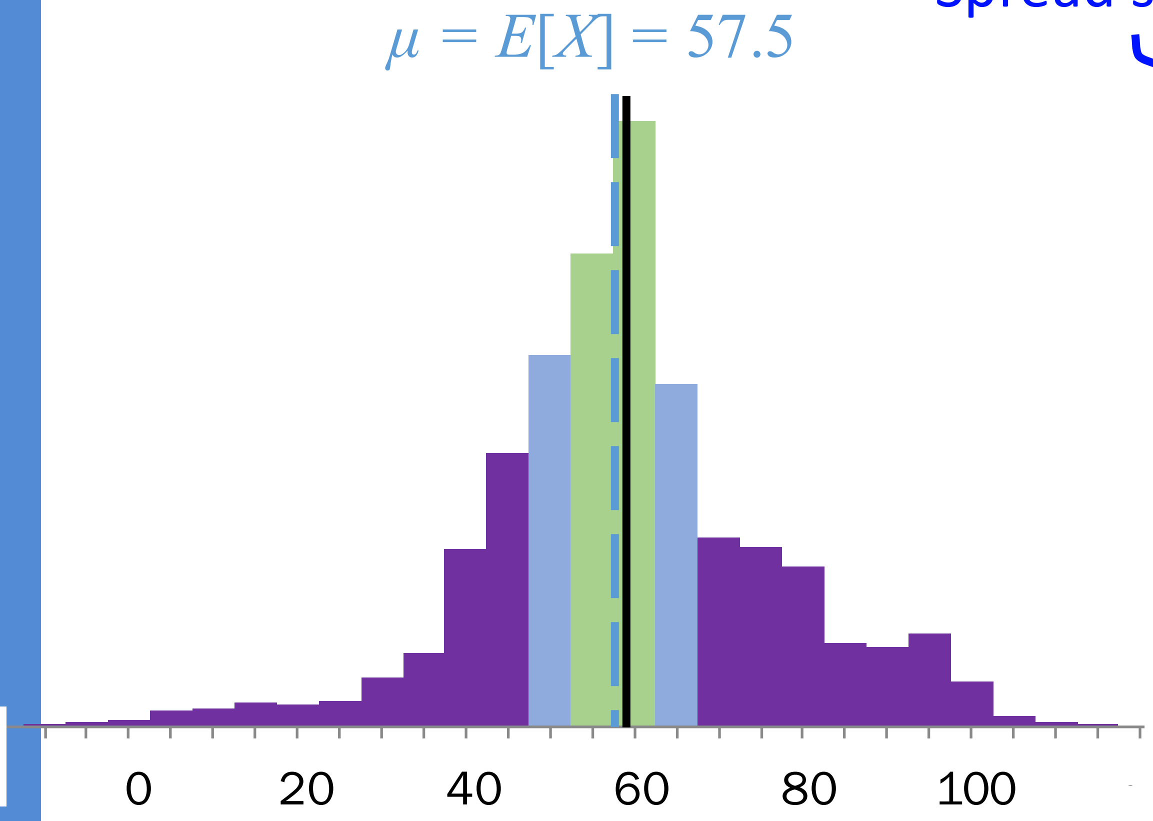
$$\text{Std}(X) = 7.2 \text{ points}$$

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



How Should We Measure Spread?

Let X be a random variable



Spread stat.

On average..

distance

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

The random variable X

The mean of X

Different Possibility: $E[|X - E[X]|]$?

Variance

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

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

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

Computing Variance

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

Law of unconscious statistician

Notation

$\mu = E[X]$

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

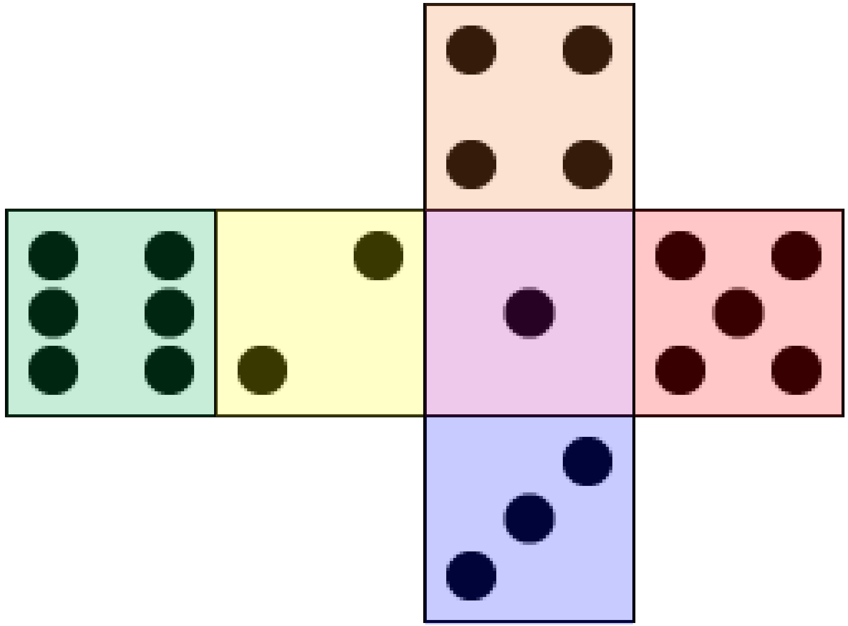
Units are in points squared

Example: Variance of a Dice Roll

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

Let X be the result of rolling a 6 sided dice.

What is Var(X)?

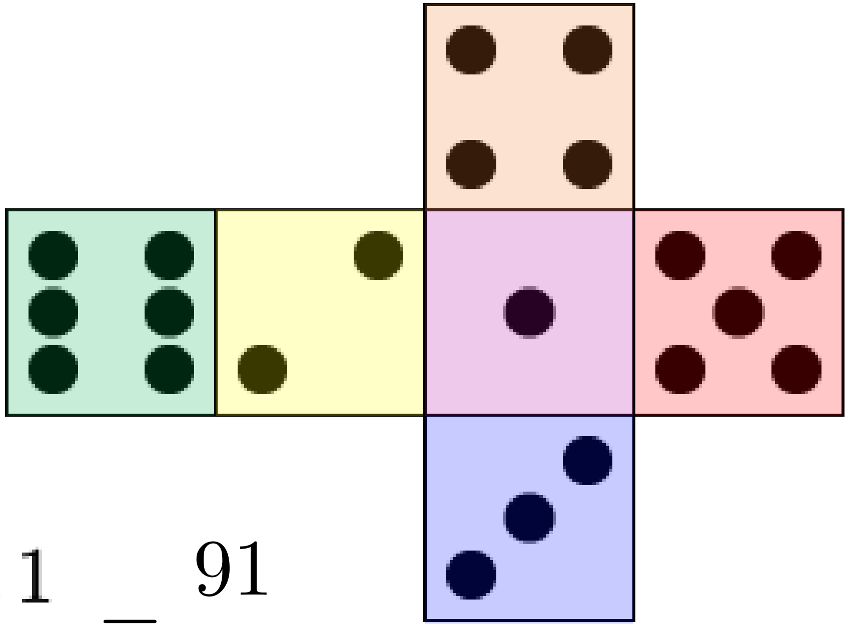


Example: Variance of a Dice Roll

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

Let X be the result of rolling a 6 sided dice.

What is Var(X)?



$$E[X] = 3.5$$

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

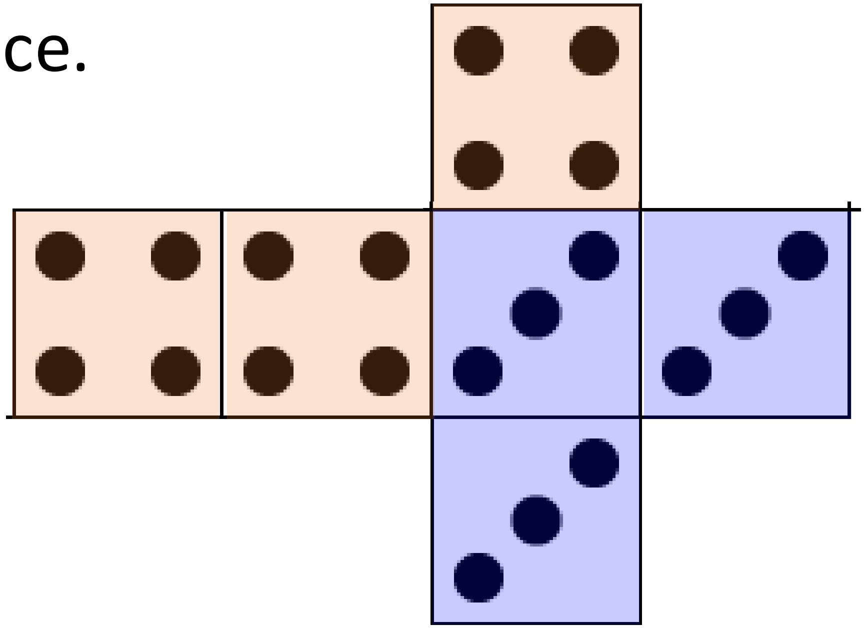
$$\begin{aligned}\text{Var}(X) &= E[X^2] - E[X]^2 \\ &= \frac{91}{6} - (3.5)^2 = 2.91\end{aligned}$$

Example: Variance of a Dice Roll

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

Let X be the result of rolling **this weird** 6 sided dice.

What is Var(X)?

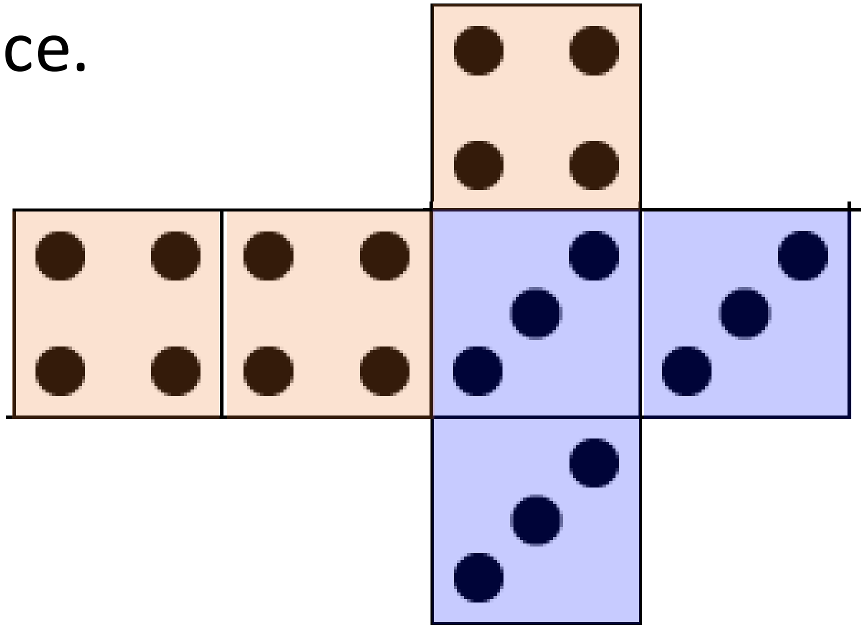


Example: Variance of a Dice Roll

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

Let X be the result of rolling **this weird** 6 sided dice.

What is Var(X)?

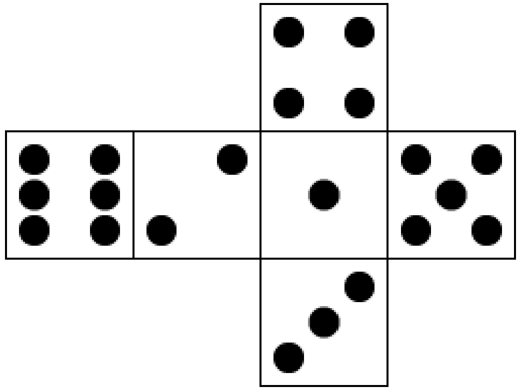


$$E[X] = 3.5$$

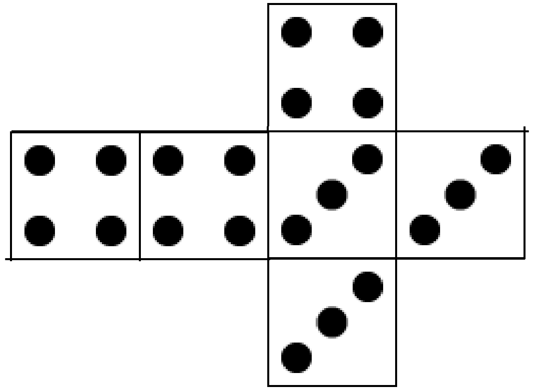
$$E[X^2] = 3^2 \cdot \frac{3}{6} + 4^2 \cdot \frac{3}{6} = 12.5$$

$$\begin{aligned}\text{Var}(X) &= E[X^2] - E[X]^2 \\ &= 12.5 - (3.5)^2 = 0.25\end{aligned}$$

Variance of a 6 Sided Dice

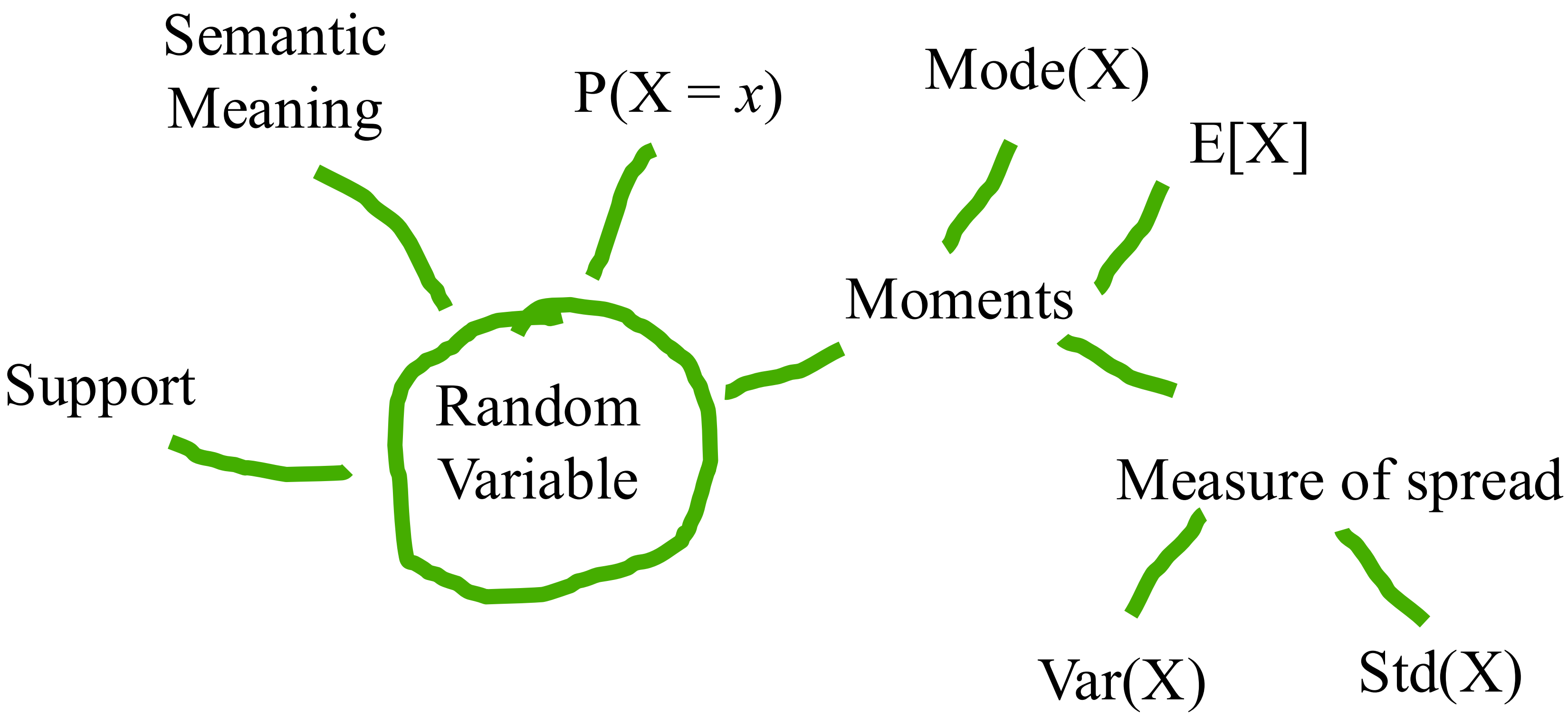


$$\text{Var}(X) = 2.91$$
$$\text{Std}(X) = 1.7$$



$$\text{Var}(X) = 0.25$$
$$\text{Std}(X) = 0.5$$

Fundamental Properties of Random Variables



You Get So Much For Free!

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

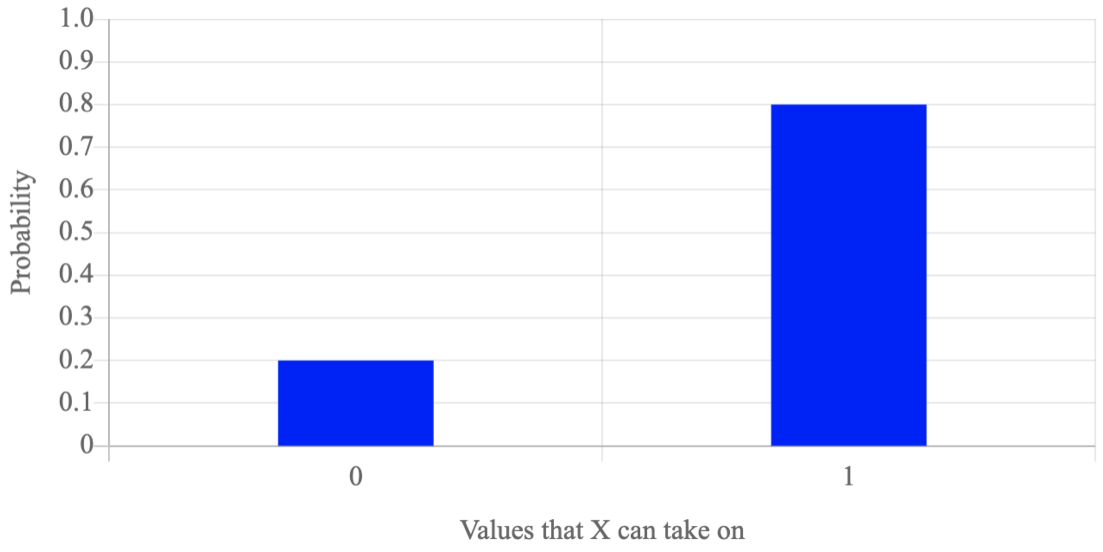
PMF equation: $$\Pr(X = x) = \begin{cases} p & \text{if } x = 1 \\ 1 - p & \text{if } x = 0 \end{cases}$$

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

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

PMF graph:

Parameter p: 0.80



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

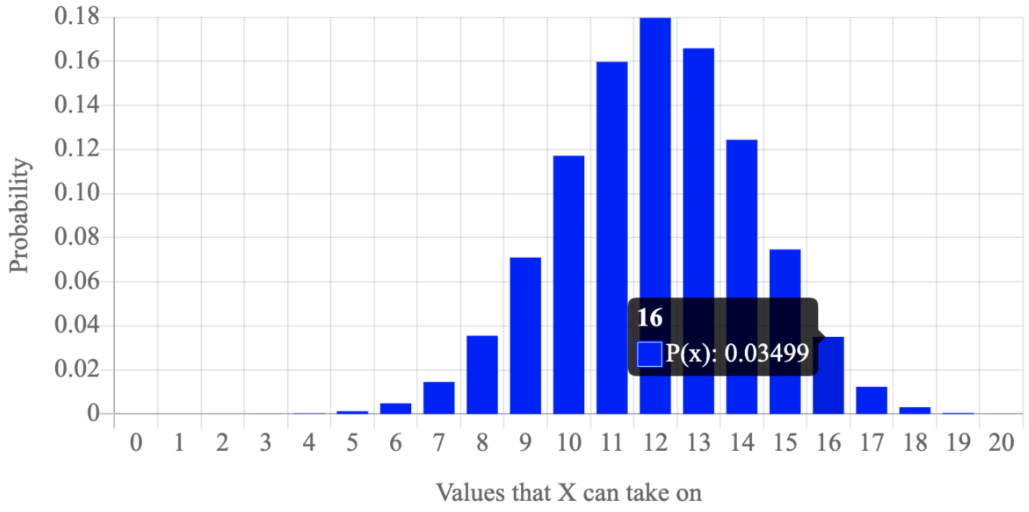
PMF equation: $$\Pr(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}$$

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



Curious? Proof of Variance for a Binomial (Hard Way)

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

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

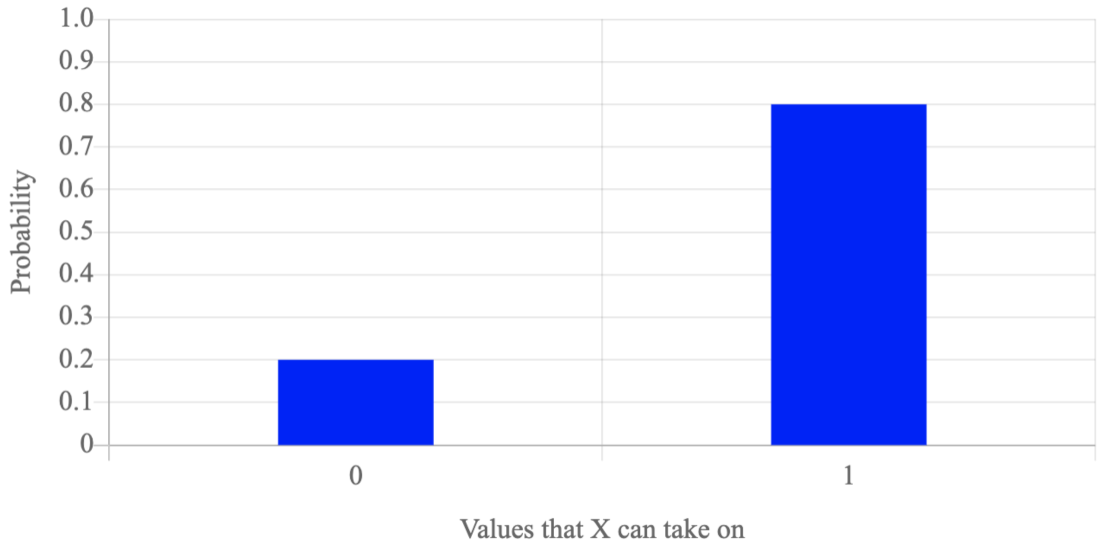
Now the easy way….

Variance of a Bernoulli

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



$$E[X^2] = \sum_{x \in \{0,1\}} x^2 P(X = x)$$
$$= 0^2 \cdot (1 - p) + 1^2 \cdot p$$
$$= p$$

$$\text{Var}(X) = E[X^2] - E[X]^2$$
$$= p - p^2$$
$$= p \cdot (1 - p)$$

Variance of a Binomial?

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

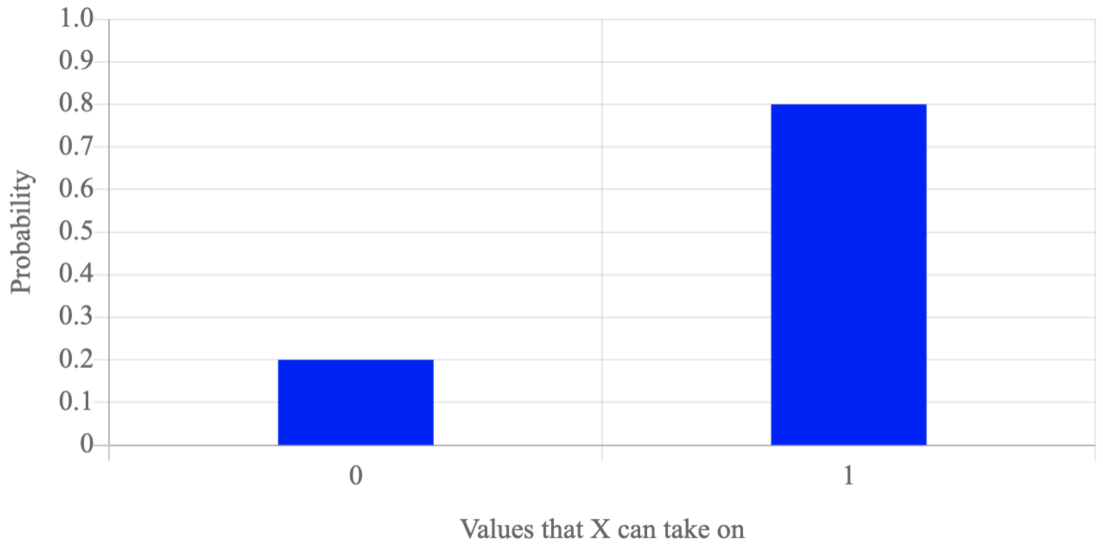
PMF equation: $\Pr(X = x) = \begin{cases} p & \text{if } x = 1 \\ 1 - p & \text{if } x = 0 \end{cases}$

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

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

PMF graph:

Parameter p: 0.80



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

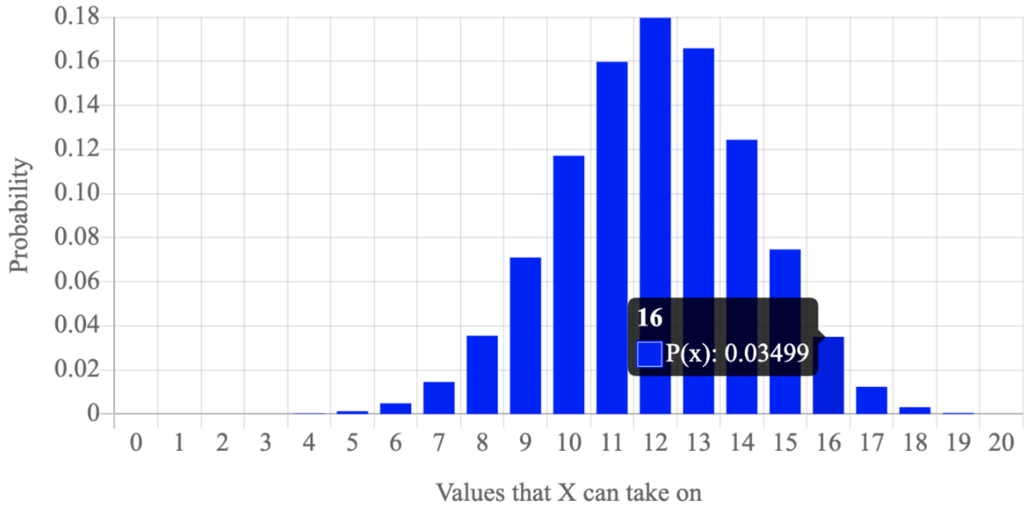
PMF equation: $\Pr(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}$

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



Variance of a Binomial (Easy Way)

Definitions

$$X_i \sim \mathrm{Bern}(p)$$

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

$$X = \sum_{i=1}^{n} X_i$$

Proved

$$\mathrm{Var}(X_i) = p \cdot (1 - p)$$

Want to Show

$$\mathrm{Var}(X) = n \cdot p \cdot (1 - p)$$

Proof

$$\mathrm{Var}(X) = \mathrm{Var}\Big(\sum_{i=1}^{n} X_i\Big)$$

$$= \sum_{i=1}^{n} \mathrm{Var}(X_i)$$

$$= \sum_{i=1}^{n} p \cdot (1 - p)$$

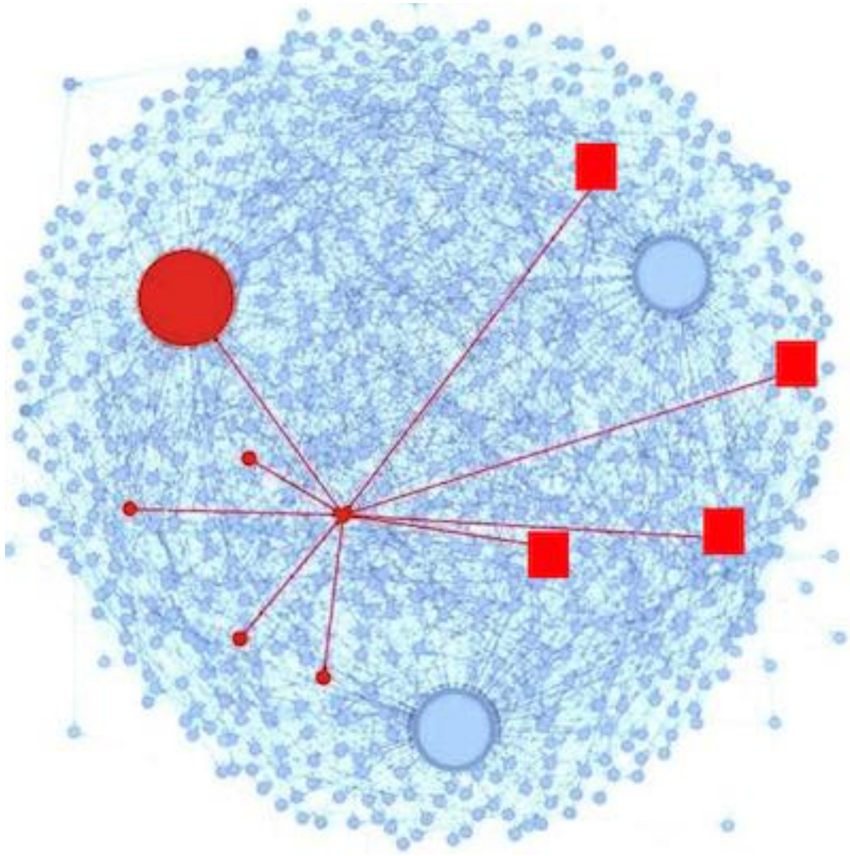
$$= n \cdot p \cdot (1 - p)$$

Is this true? Is the variance of the sum the sum of variance?

Only if X_is are independent!

Is Peer Grading Accurate Enough?

Looking ahead



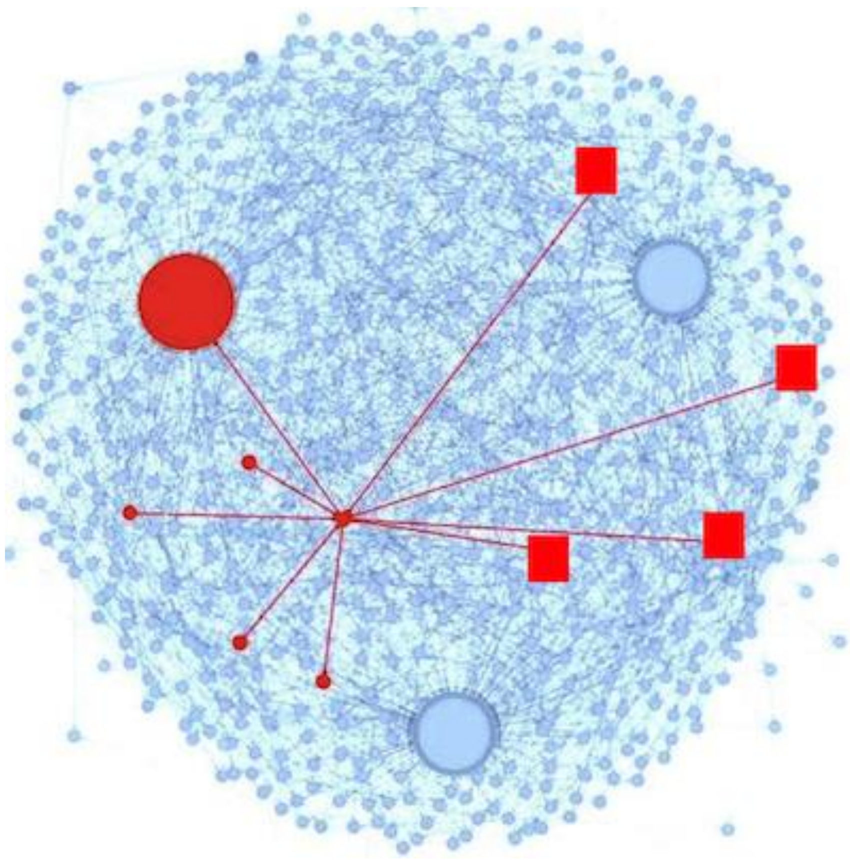
Peer Grading on Coursera HCI.

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

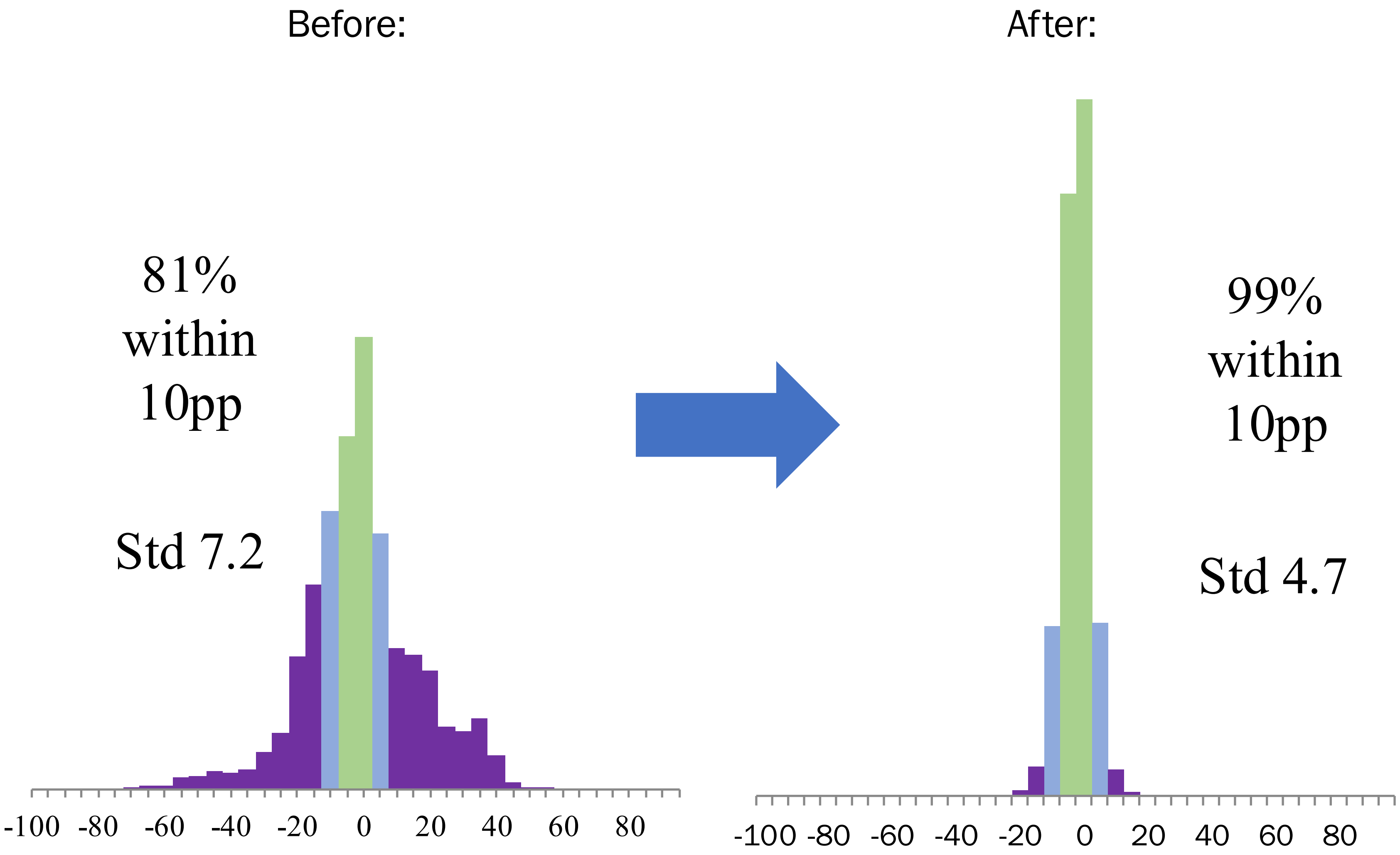
Problem param

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

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

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

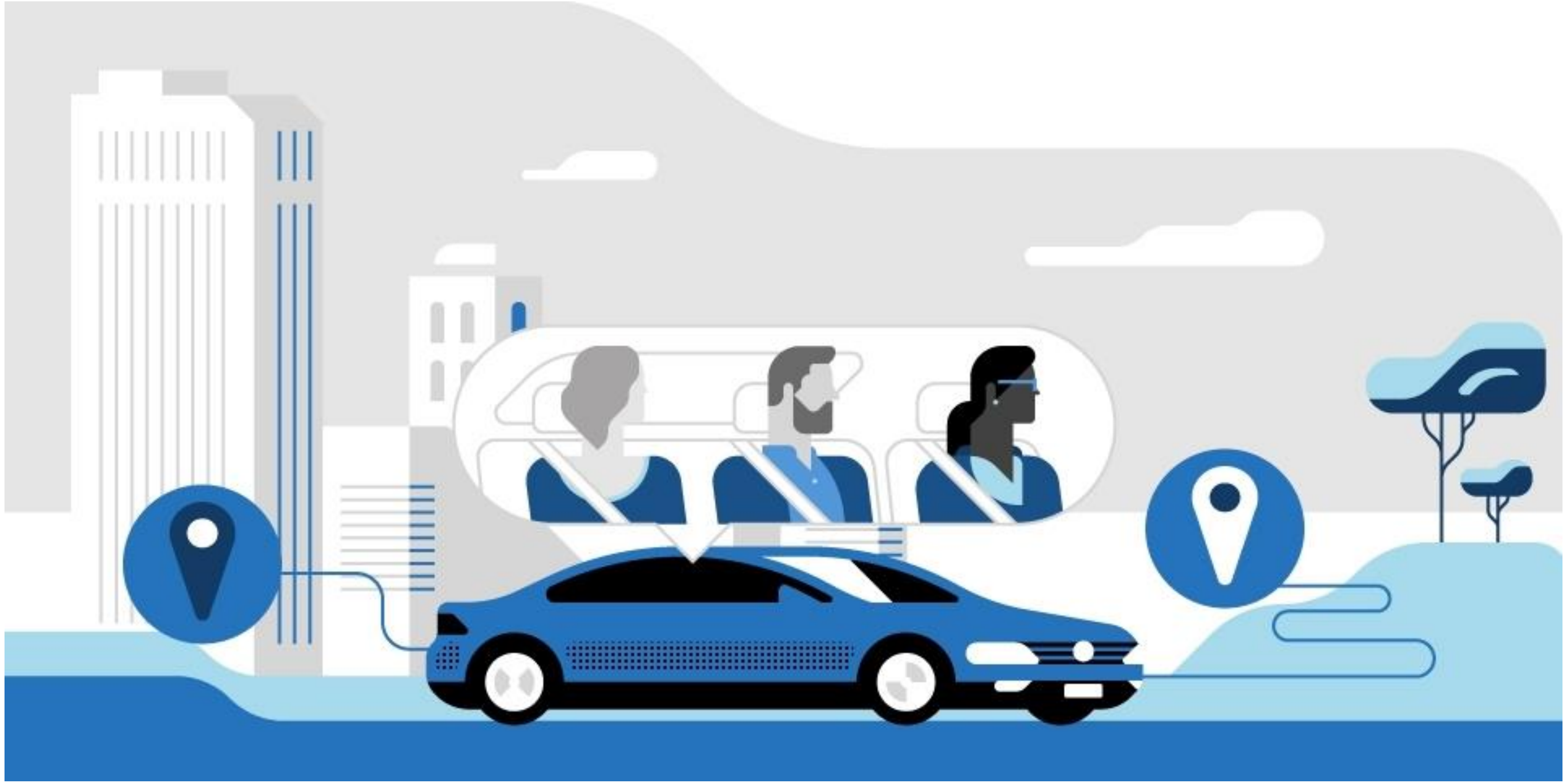
Yes, With Probabilistic Modelling



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

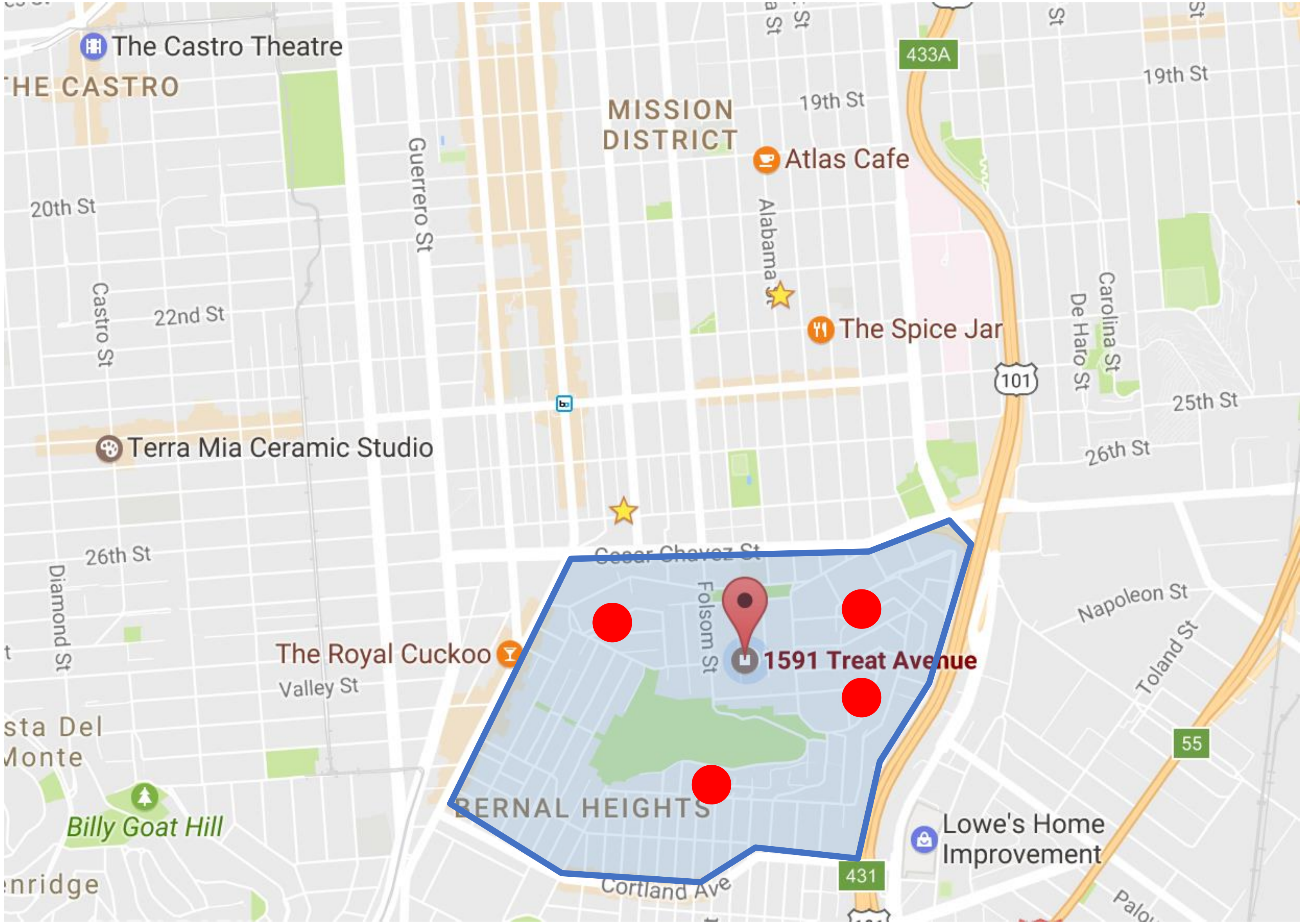
Ready..

Algorithmic Ride Sharing



Probability of *k* **requests** from this area in the next 1 min

Probability of ***k* requests** from this area in the next 1 min



Probability of ***k* requests** from this area in the next 1 min

On average $\lambda = 5$ requests per minute

Probability of ***k* requests** from this area in the next 1 min

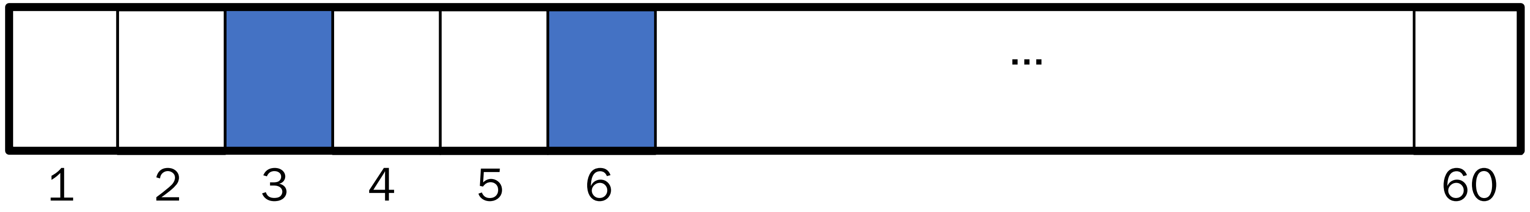
On average $\lambda = 5$ requests per minute

We can break the next minute down into seconds

Probability of ***k* requests** from this area in the next 1 min

On average λ = 5 requests per minute

We can break the next minute down into seconds

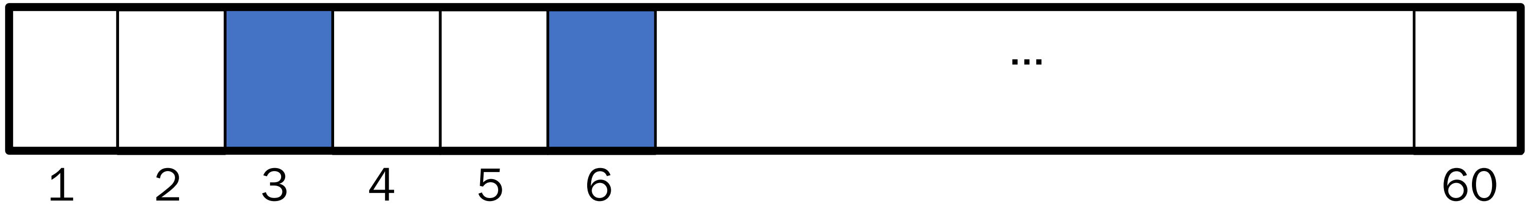


At each second either get a request or you don’t.

Probability of ***k* requests** from this area in the next 1 min

On average λ = 5 requests per minute

We can break the next minute down into seconds



At each second either get a request or you don't.

Let X = Number of requests in the minute

$$X \sim \text{Bin}(n = 60, p = 5/60)$$

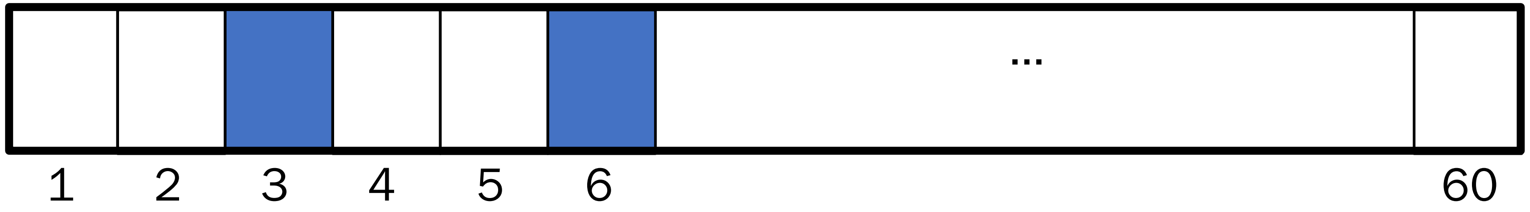
$$P(X = k) = \binom{n}{k}(p)^k(1-p)^{n-k}$$

$$P(X = 3) = \binom{60}{3}(5/60)^3(1-5/60)^{57}$$

Probability of ***k* requests** from this area in the next 1 min

On average λ = 5 requests per minute

We can break the next minute down into seconds



At each second either get a request or you don't.
Let X = Number of requests in the minute

$$X \sim \text{Bin}(n = 60, p = 5/60)$$

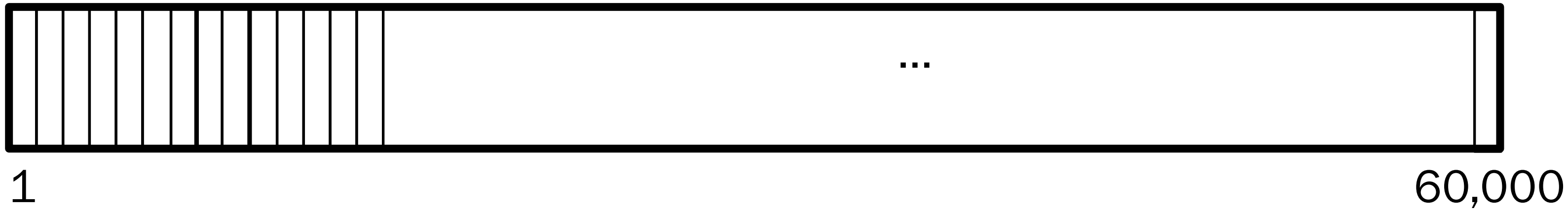
$$P(X = k) = \binom{n}{k} (p)^k (1 - p)^{n-k}$$

But what if there are two requests in the same second?

Probability of ***k* requests** from this area in the next 1 min

On average $\lambda = 5$ requests per minute

We can break that next minute down into ***milli***-seconds



At each ***milli***-second either get a request or you don't.

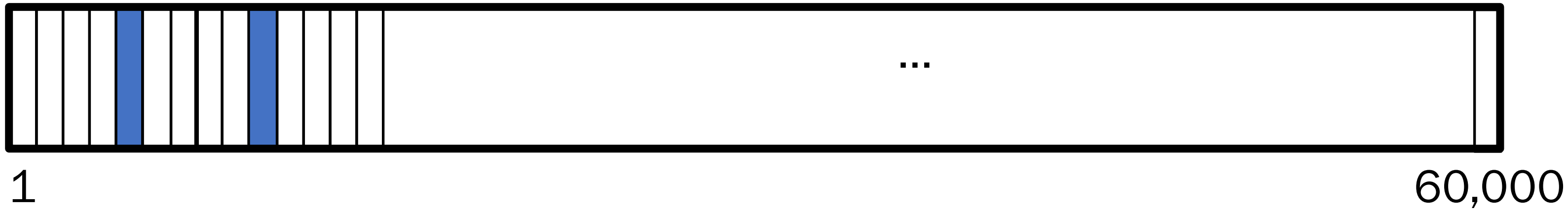
Let X = Number of requests in the minute

But what if there are two requests in the same second?

Probability of ***k* requests** from this area in the next 1 min

On average λ = 5 requests per minute

We can break that next minute down into ***milli***-seconds



At each ***milli***-second either get a request or you don't.
Let X = Number of requests in the minute

$$X \sim \text{Bin}(n = 60000, p = \lambda/n)$$

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

Can we do any better than milli-seconds?

Probability of ***k* requests** from this area in the next 1 min

On average λ = 5 requests per minute

We can break that minute down into ***infinitely small*** buckets

OMG so small

1 ∞

Let X = Number of requests in the minute

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

$$P(X = k) = \lim_{n \to \infty} \binom{n}{k} (\lambda/n)^k (1 - \lambda/n)^{n-k}$$

Who wants to see some cool math?

Probability of ***k* requests** from this area in the next 1 min

$$P(X = k) = \lim_{n\to\infty} \binom{n}{k} (\lambda/n)^k (1 - \lambda/n)^{n-k}$$

$$= \lim_{n\to\infty} \frac{n!}{(n-k)!k!} \cdot \frac{\lambda^k}{n^k} \cdot \frac{(1-\lambda/n)^n}{(1-\lambda/n)^k}$$ By expanding each term

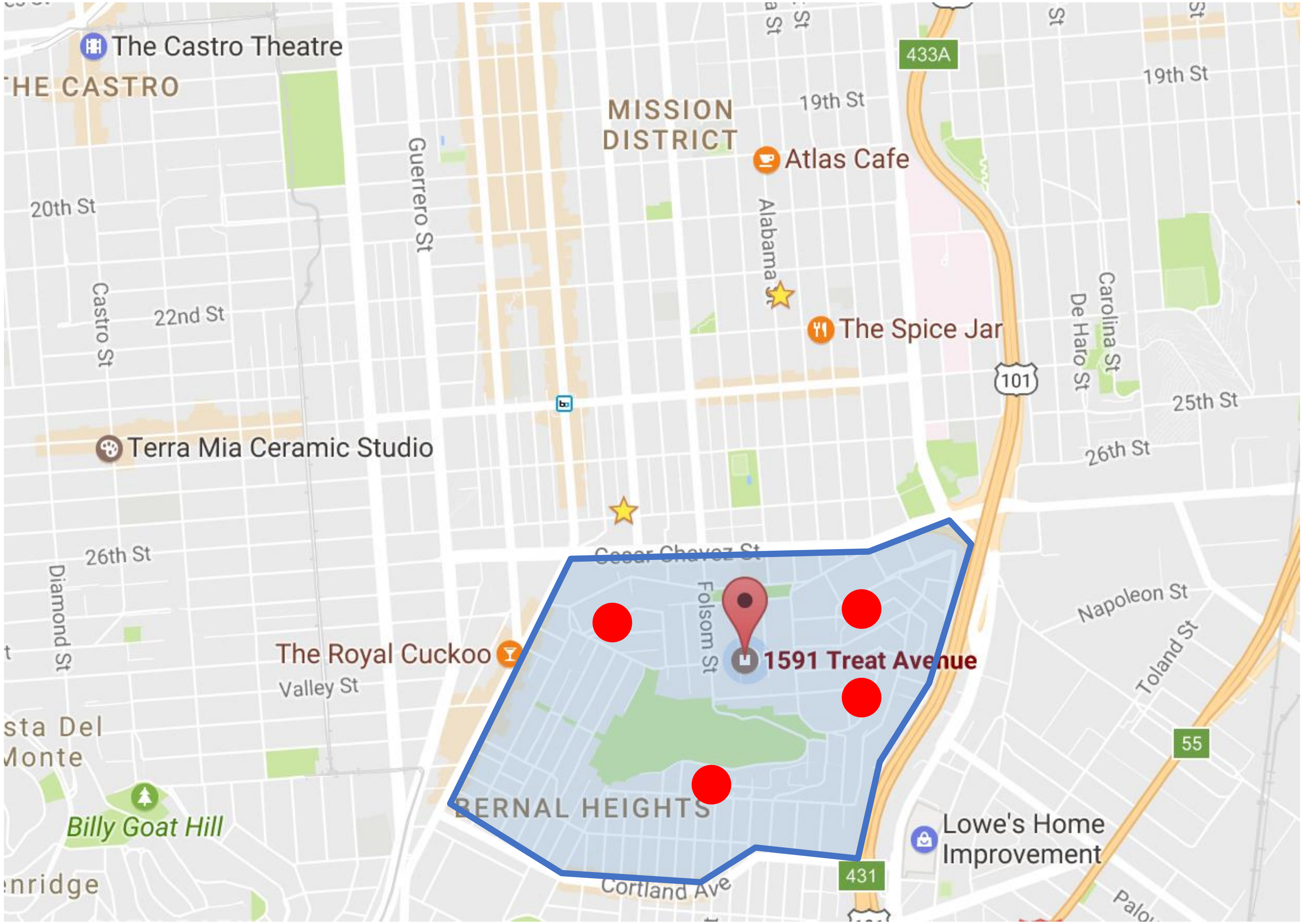
$$= \lim_{n\to\infty} \frac{n!}{(n-k)!k!} \cdot \frac{\lambda^k}{n^k} \cdot \frac{e^{-\lambda}}{1}$$ By definition of natural exp

$$= \lim_{n\to\infty} \frac{n!}{(n-k)!n^k} \cdot \frac{\lambda^k}{k!} \cdot \frac{e^{-\lambda}}{1}$$ Rearranging terms

$$= \lim_{n\to\infty} \frac{n^k}{n^k} \cdot \frac{\lambda^k}{k!} \cdot \frac{e^{-\lambda}}{1}$$ Limit analysis

$$= \frac{\lambda^k e^{-\lambda}}{k!}$$ Simplifying

Probability of ***k* requests** from this area in the next 1 min



Simeon-Denis Poisson

Simeon-Denis Poisson (1781-1840) was a prolific French mathematician

Published his first paper at 18, became professor at 21, and published over 300 papers in his life

- He reportedly said *"Life is good for only two things, discovering mathematics and teaching mathematics."*

Going with French Martin Freeman

Poisson Random Variable

X is a **Poisson** Random Variable: the number of occurrences in a fixed interval of time.

$$X \sim \mathrm{Poi}(\lambda)$$

- λ is the "rate"
- X takes on values 0, 1, 2…
- has distribution (PMF):

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

Poisson Process

1. Consider events that occur over time
   - Earthquakes, radioactive decay, hits to web server, etc.
   - Have time interval for events (1 year, 1 sec, whatever...)
   - Events arrive at rate: λ events per interval of time

2. Split time interval into $n \rightarrow \infty$ sub-intervals
   - Assume at most one event per sub-interval
   - Event occurrences in sub-intervals are independent
   - With many sub-intervals, probability of event occurring in any given sub-interval is small

3. # events in original time interval ~ Poi(λ)

To the reader!

Poisson Random Variable

Notation: $X \sim \text{Poi}(\lambda)$

Description: Number of events in a fixed time frame if (a) the events occur with a constant mean rate and (b) they occur independently of time since last event.

Parameters: $\lambda \in \{0, 1, \ldots\}$, the constant average rate.

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

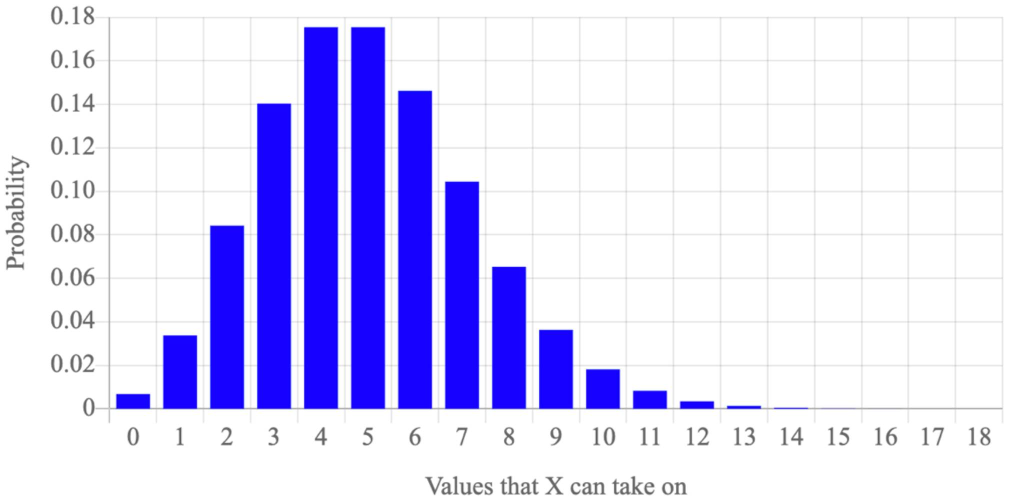
PMF equation: $\Pr(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$

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

Variance: $\text{Var}(X) = \lambda$

PMF graph:

Parameter λ: 5



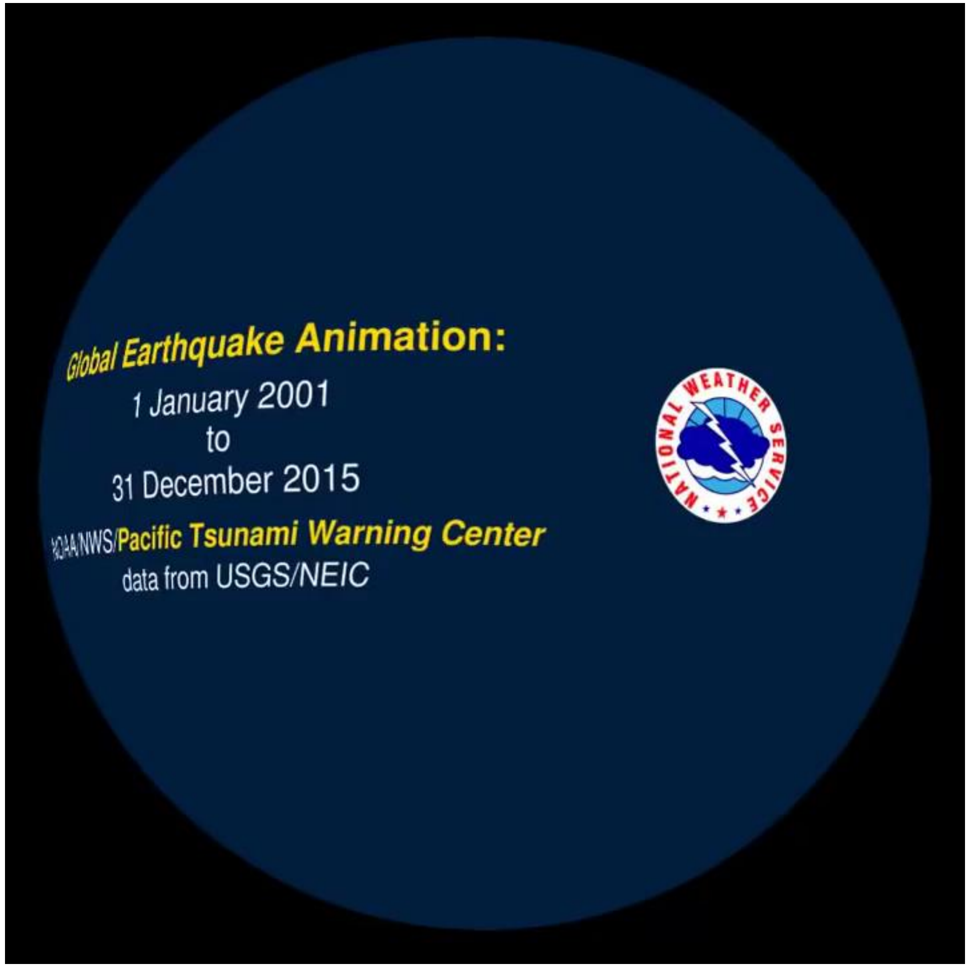
Poisson is great when you have a rate!

Poisson is great when you have a rate and you care about # of occurrences!

Make sure that the time unit for “rate” and match the probability question

Two quick examples!

Earthquakes



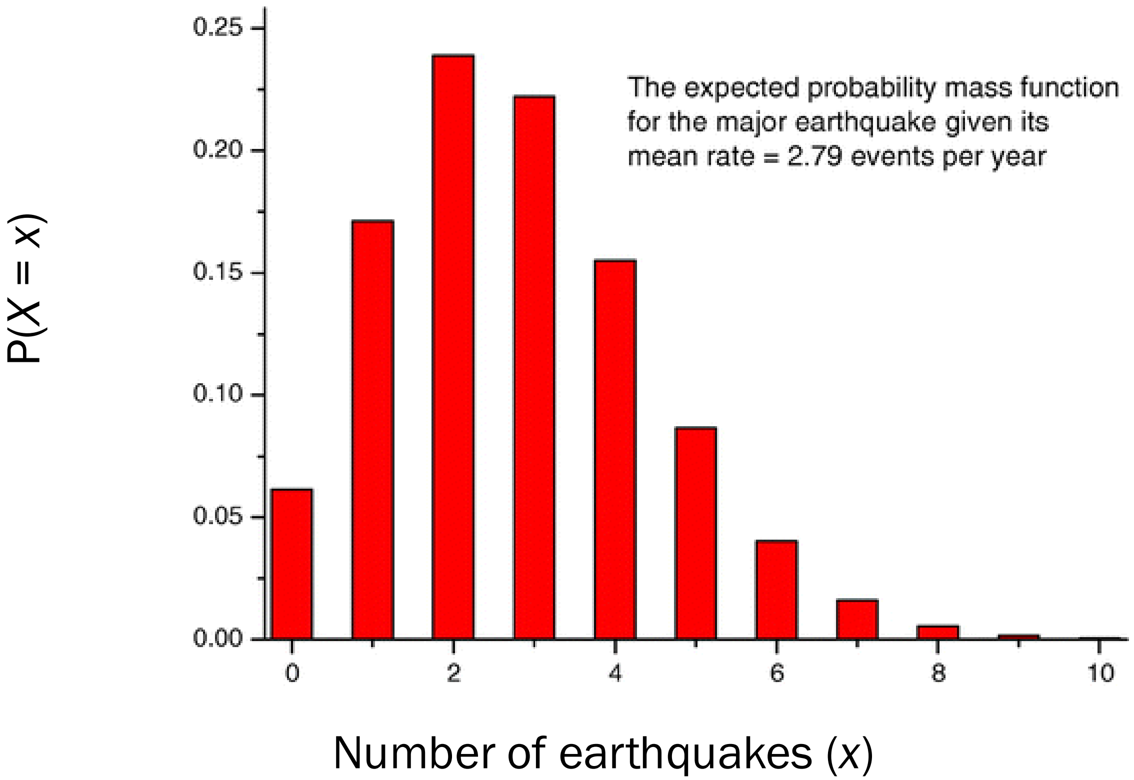
Average of 2.79 major earthquakes per year.

What is the probability of 3 major earthquakes next year?

Earthquake Probability Mass Function

Let X = number of earthquakes next year

$$X \sim \text{Poi}(2.79)$$



$$P(X=3) = \frac{\lambda^k e^{-\lambda}}{k!} = \frac{2.79^3 e^{-2.79}}{3!} \approx 0.23$$

Earthquakes

Bulletin of the
Seismological Society of America

Vol. 64 | October 1974 | No. 5

IS THE SEQUENCE OF EARTHQUAKES IN SOUTHERN CALIFORNIA, WITH AFTERSHOCKS REMOVED, POISSONIAN?

By J. K. Gardner and L. Knopoff

Abstract

Yes.

Candy in Class!

Students ask on average **15 questions per class**.

Juliette only brought **10 pieces of candy**!

What is the probability she doesn't have enough candy?

* Assume: (a) question rate is constant (b) questions don't impact one another.

Let X be the number of questions asked in class. $X \sim \text{Poi}(\lambda = 15)$

$$P(X > 10) = \sum_{i=11}^{\infty} P(X = i) = 1 - P(X \leq 10)$$

$$P(X \leq 10) = \sum_{i=0}^{10} P(X = i)$$

$$= \sum_{i=0}^{10} \frac{\lambda^i e^{-\lambda}}{i!}$$

PMF of Poisson

$$= \sum_{i=0}^{10} \frac{15^i e^{-15}}{i!}$$

λ = 15

```
from scipy import stats
def main():
  lam = int(input("Questions per class: "))
  num_candy = int(input("Number of Candy: "))
  X = stats.poisson(lam)
  prob_enough_candy = 0
  for i in range(0, num_candy+1):
    pr_i_questions = X.pmf(i)
    prob_enough_candy += pr_i_questions
  print(prob_enough_candy)
```

Poisson in Python

```
from scipy import stats # great package
X = stats.poisson(2.5)  # X ~ Poi(λ = 2.5)
print(X.pmf(2))         # P(X = 2)
```

Function	Description
X.pmf(k)	$P(X = k)$
X.cdf(k)	$P(X \leq k)$
X.mean()	$E[X]$
X.var()	$Var(X)$
X.std()	$Std(X)$

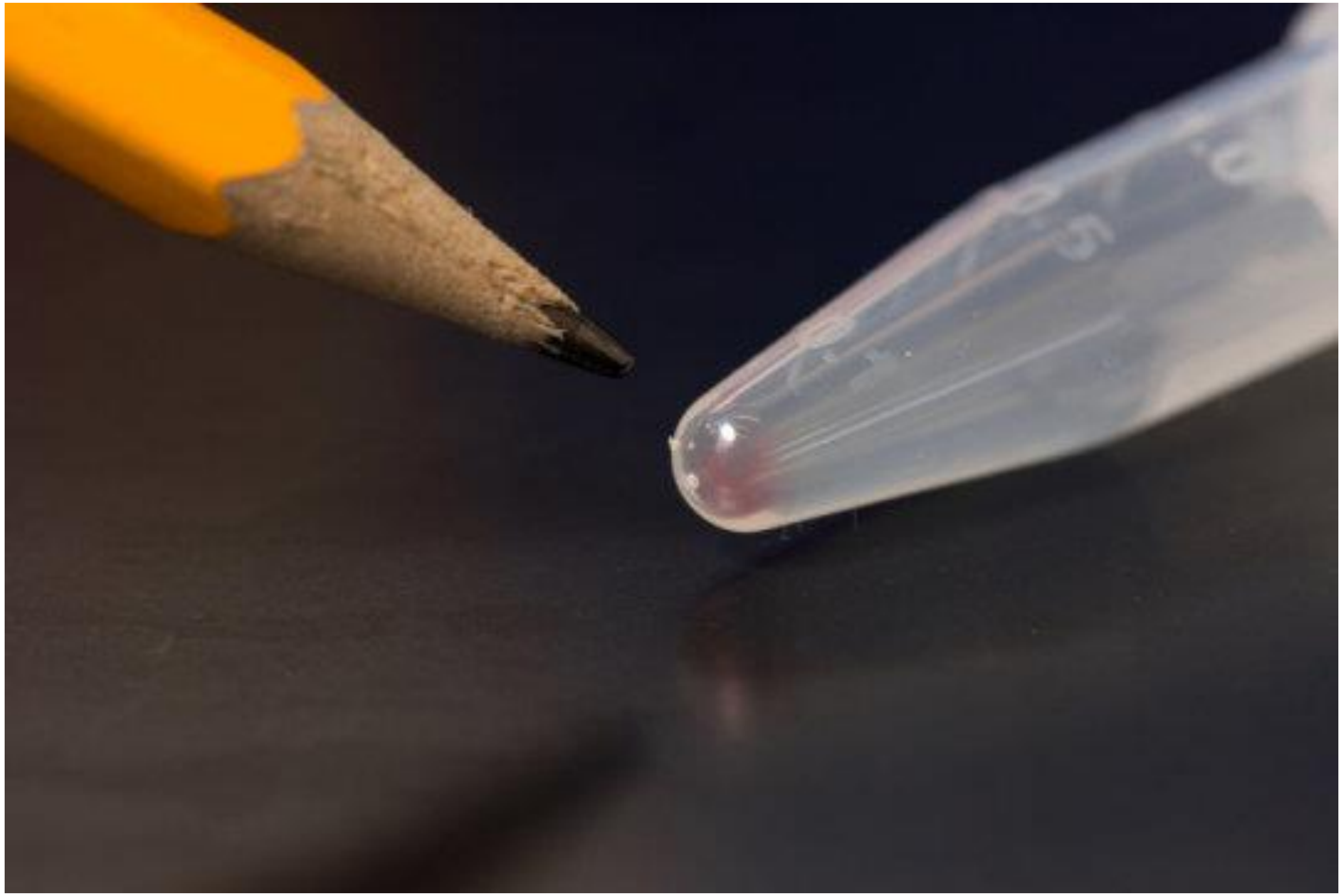
Poisson can approximate a Binomial!

Wait why would you want to do that?

1) Binomial can be expensive to compute.

2) Connections help build math intuition.

Storing Data in DNA



All the movies, images, emails and other digital data from more than 600 smartphones (10,000 gigabytes) can be stored in the faint pink smear of DNA at the end of this test tube.

Storing Data in DNA

Will more than 1% of DNA storage become corrupt?

- In DNA (and real networks) store large strings
- Length $n \approx 10^4$
- Probability of corruption of each base pair is very small $p \approx 10^{-6}$
- $X \sim \text{Bin}(10^4, 10^{-6})$ is unwieldy to compute

Extreme n and p values arise in many cases

- # bit errors in stream sent over a network
- # of servers crashes in a day in giant data center

Storing Data in DNA

Will the DNA storage become corrupt?

- In DNA (and real networks) store large strings
- Length $n \approx 10^4$
- Probability of corruption of each base pair is very small $p \approx 10^{-6}$
- X ~ Poi($\lambda = 10^4 * 10^{-6} = 0.01$)

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

$$P(X = 0) = e^{-\lambda} \frac{1}{0!}$$

$$= e^{-0.01} \approx 0.99$$

Poisson is a Binomial in the Limit

Poisson approximates Binomial where n is large, p is small, and $\lambda = np$ is "moderate"

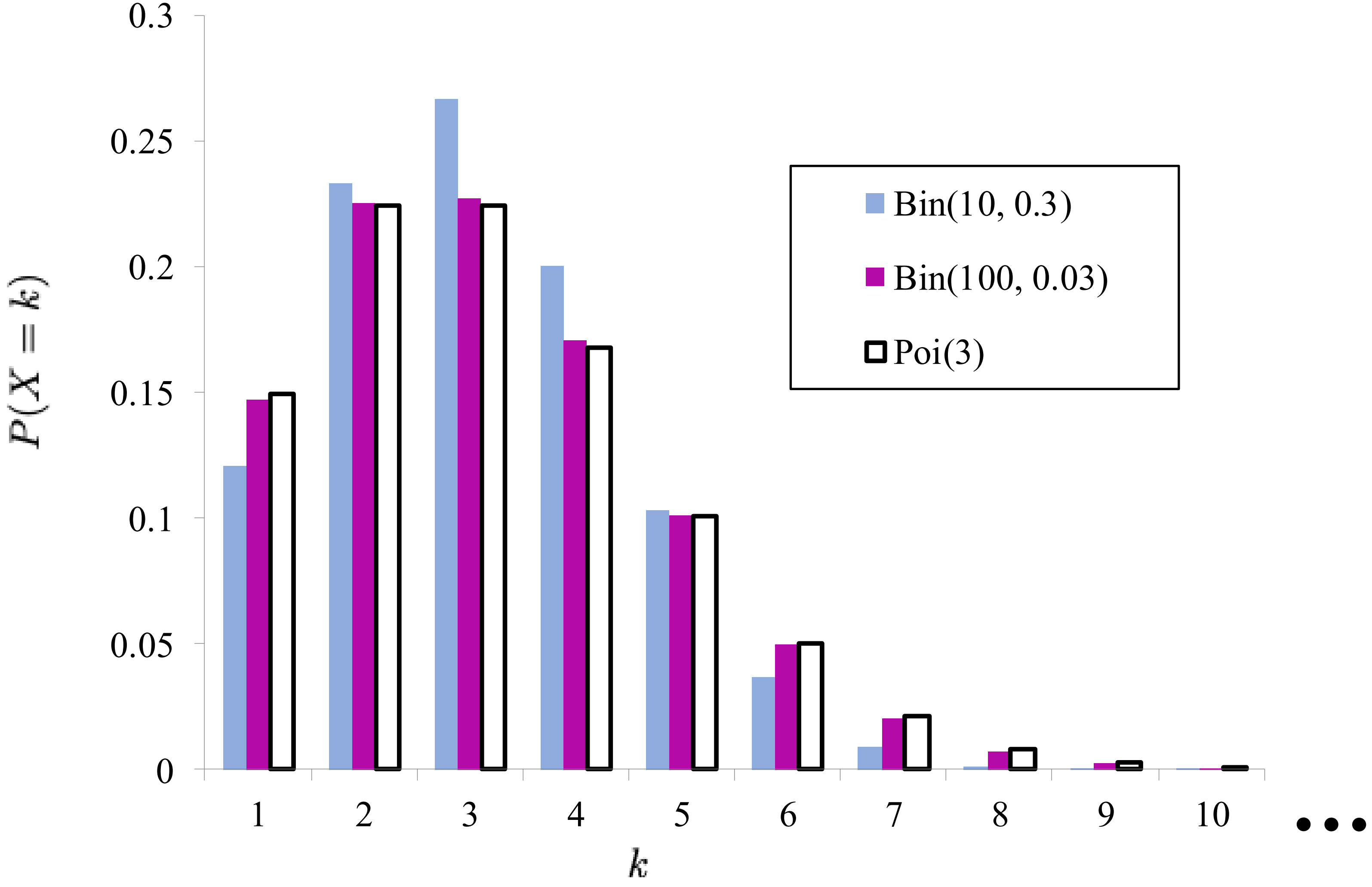
Different interpretations of "moderate"

- $n > 20$ and $p < 0.05$
- $n > 100$ and $p < 0.1$

Really, Poisson is Binomial as

$$n \rightarrow \infty \text{ and } p \rightarrow 0, \text{ where } np = \lambda$$

Bin(10,0.3) vs Bin(100,0.03) vs Poi(3)



A Real License Plate Seen at Stanford

No, it’s not mine…
but I kind of wish it was.

Poisson can be used to approximate a Binomial where *n* is large and *p* is small.

Tender (Central) Moments with Poisson

Recall: $Y \sim \text{Bin}(n, p)$

- $E[Y] = np$
- $\text{Var}(Y) = np(1 - p)$

$X \sim \text{Poi}(\lambda)$ where $\lambda = np$ ($n \to \infty$ and $p \to 0$)

- $E[X] = np = \lambda$
- $\text{Var}(X) = np(1 - p) = \lambda(1 - 0) = \lambda$
- Yes, expectation and variance of Poisson are same
- It brings a tear to my eye…

Poisson Paradigm

Poisson can still provide a good way to model an event, even when assumptions are "mildly" violated. Can apply Poisson approximation when...

1

"Successes" in trials are not entirely independent.

- Example: # entries in each bucket in large hash table

2

Probability of "Success" p in each trial varies slightly.

- Example: average # requests to web server/sec. may fluctuate slightly due to load on network

Web Server Load

Consider requests to a web server in 1 second

- In past, server load averages 2 hits/second
- X = # hits server receives in a second
- What is $P(X < 5)$?

Solution

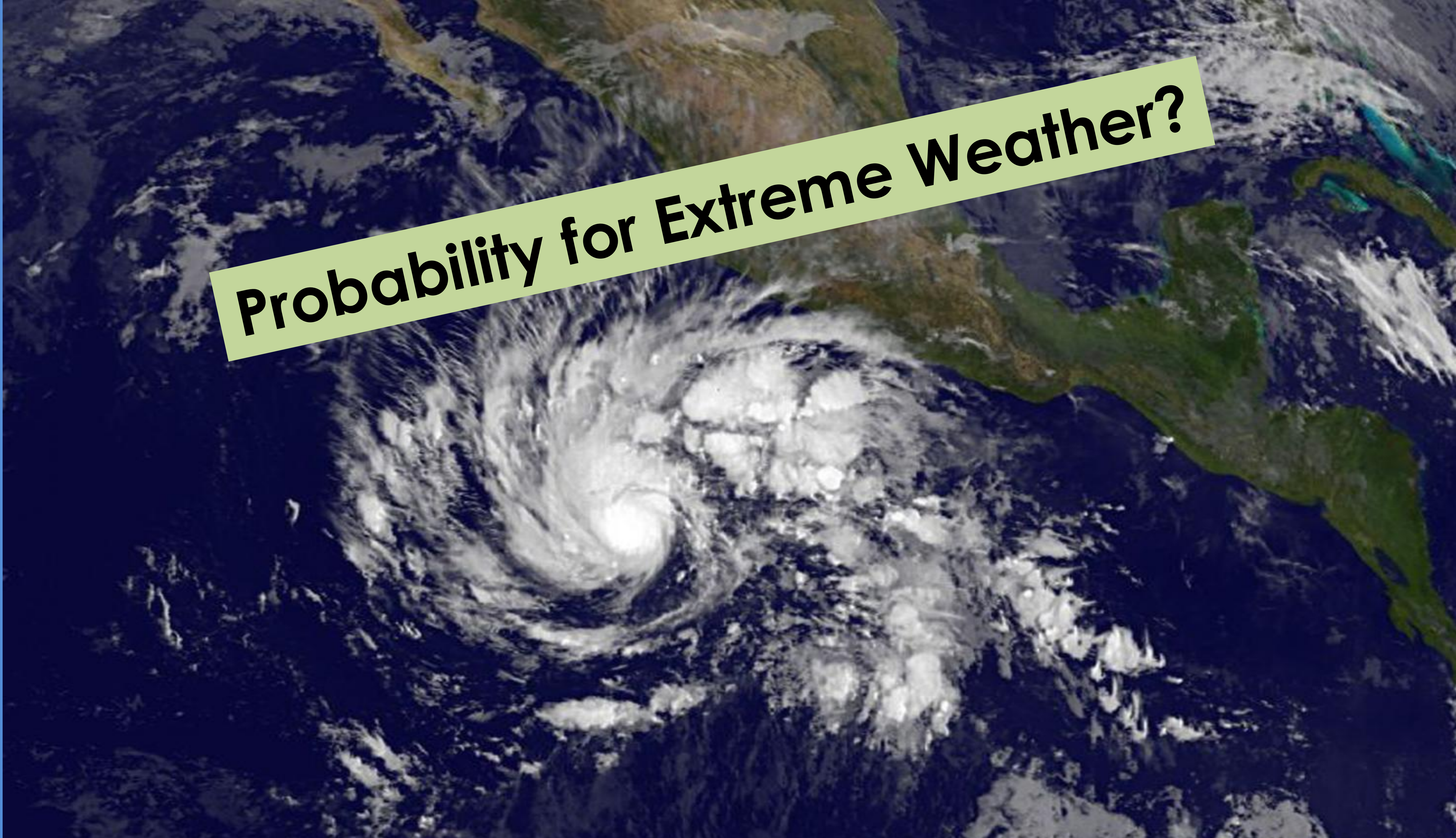
$$X \sim \text{Poi}(\lambda = 2)$$

$$P(X < 5) = \sum_{i=0}^{4} P(X = i)$$

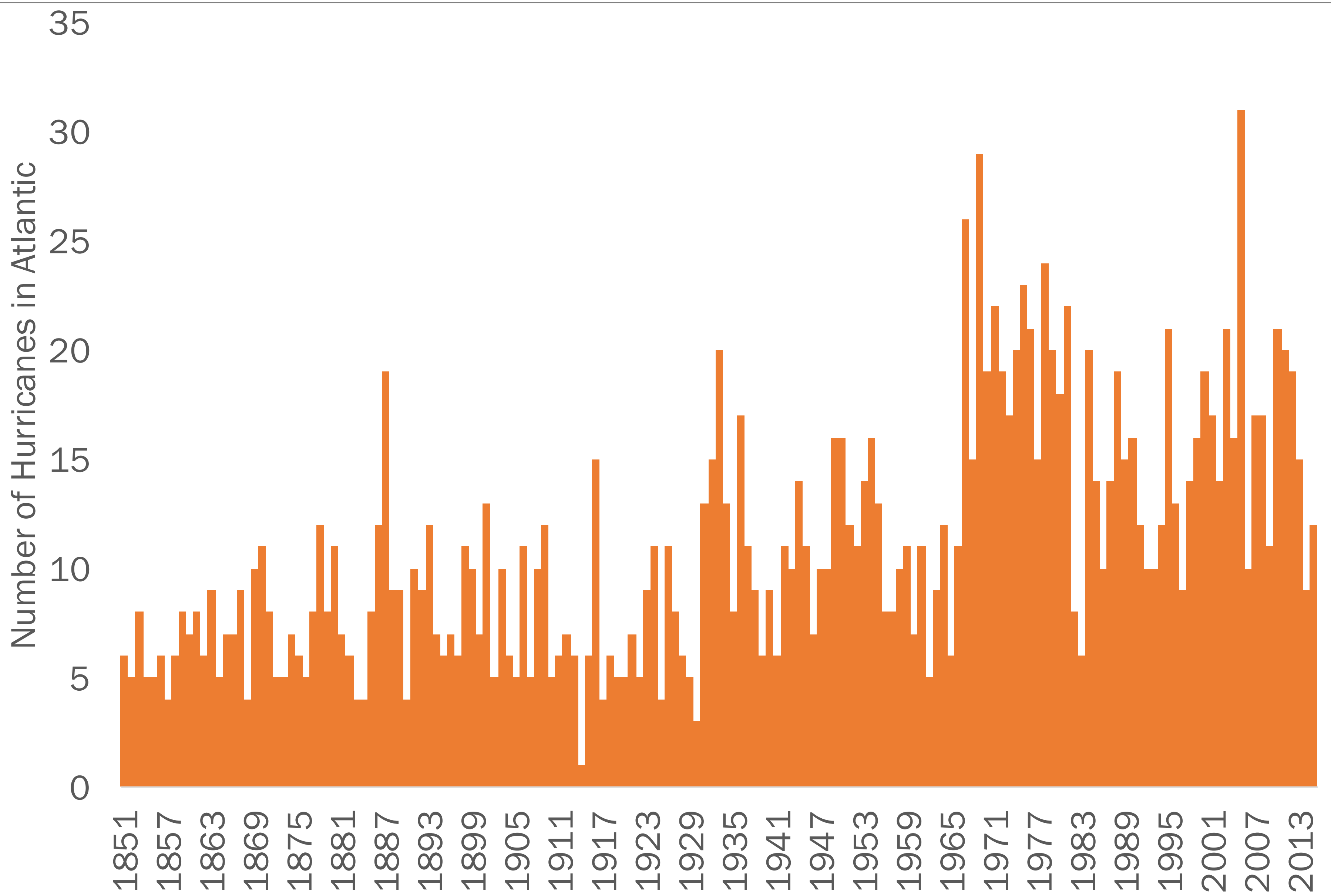
$$= \sum_{i=0}^{4} e^{-\lambda} \frac{\lambda^i}{i!} \qquad \text{Since } X \text{ is Poisson}$$

$$= \sum_{i=0}^{4} e^{-2} \frac{2^i}{i!} \approx 0.95 \qquad \text{Since } \lambda = 2$$

Probability for Extreme Weather?

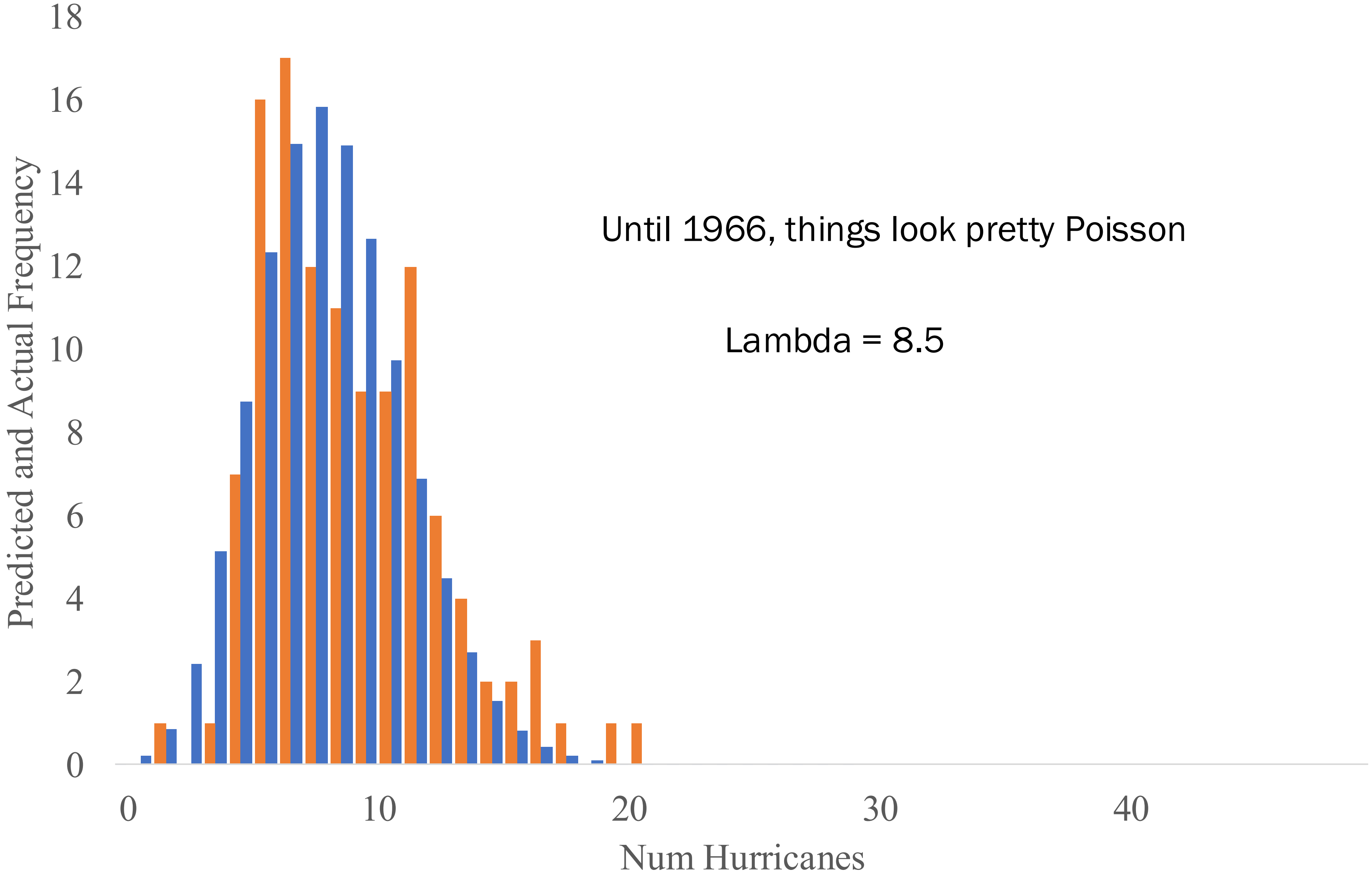


Hurricanes per Year since 1851



To the code!

Historically ~ Poisson(8.5)



Until 1966, things look pretty Poisson

Lambda = 8.5

Improbability Drive

What is the probability of over 15 hurricanes in a season given that the distribution doesn't change?

- Let X = # hurricanes in a year. X ~ Poi(8.5)

Solution:

$$P(X > 15) = 1 - P(X \leq 15)$$

$$= 1 - \sum_{i=0}^{15} P(X = i)$$

This is the pmf of a Poisson. Your favorite programming language has a function for it

$$= 0.0135$$

Twice since 1966 there have been two years with over 30 hurricanes

Improbability Drive

What is the probability of over 30 hurricanes in a season given that the distribution doesn't change?

- Let X = # hurricanes in a year. X ~ Poi(8.5)

Solution:

$$P(X > 30) = 1 - P(X \leq 30)$$

$$= 1 - \sum_{i=0}^{30} P(X = i)$$

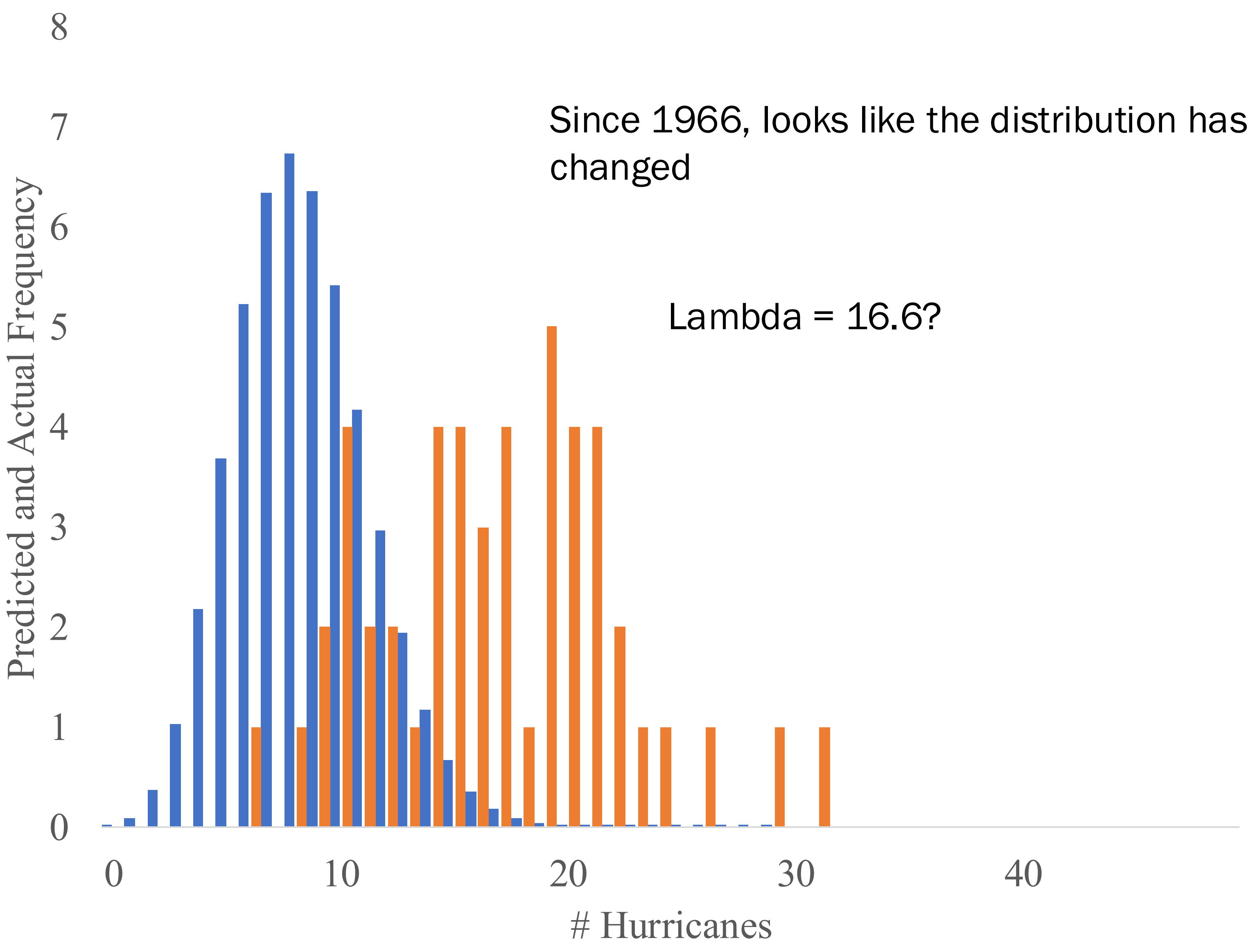
$$= 1 - 0.99999997823$$

$$= 2.2e - 09$$

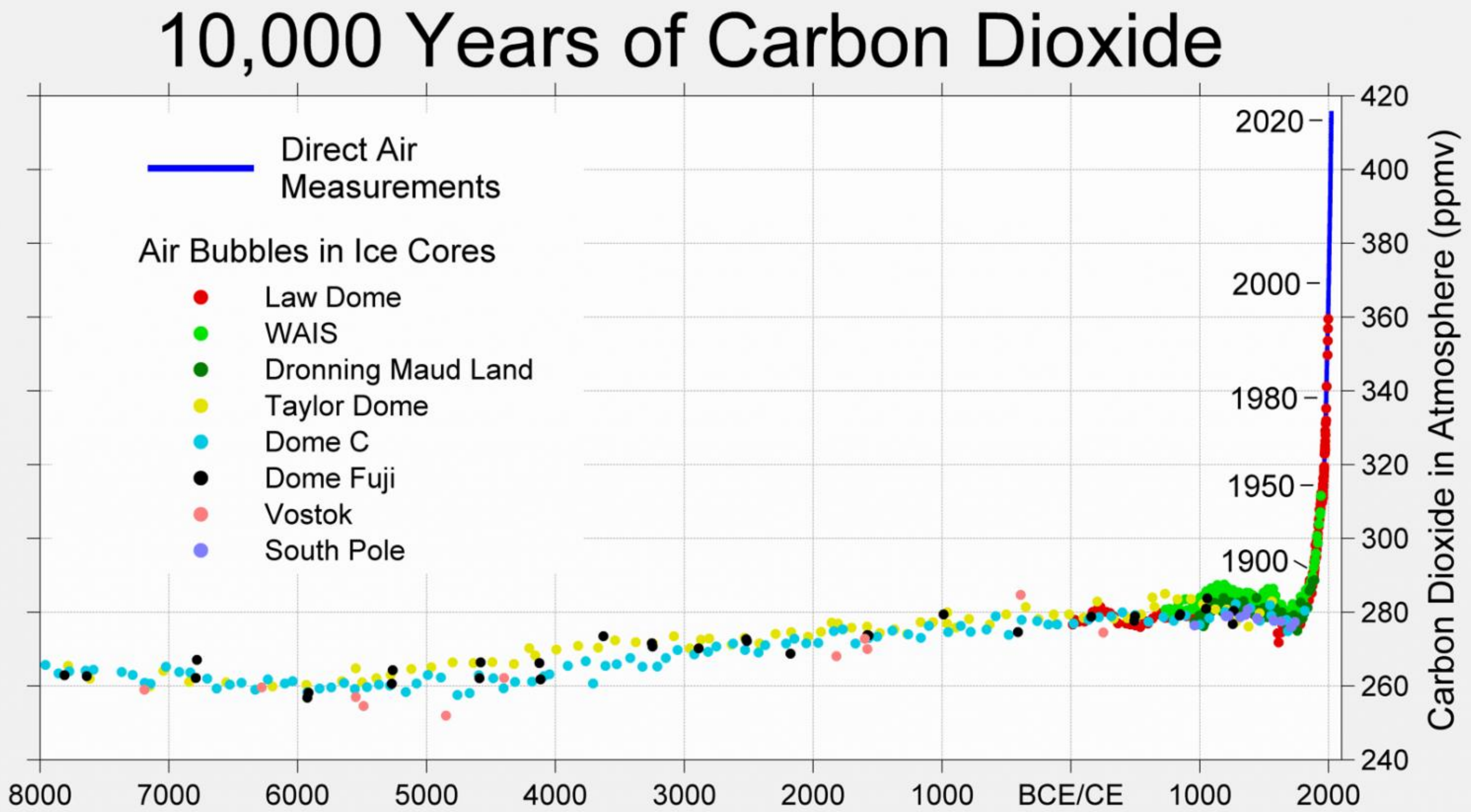
This is the pdf of a Poisson. Your favorite programming language has a function for it

* Challenge: Calculate the probability of two years with over 30 hurricanes

The Distribution has Changed

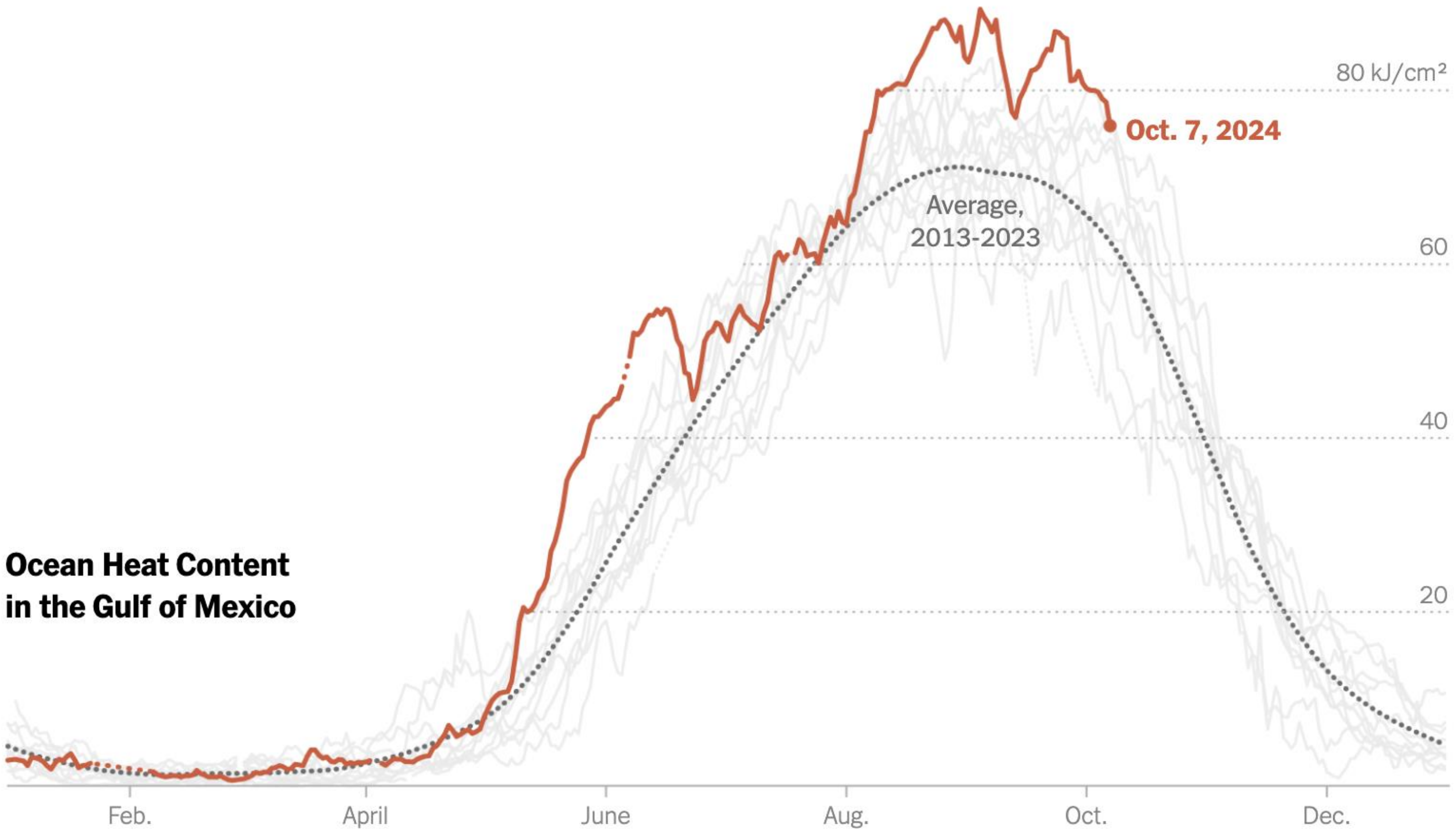


What's up?



10,000 Years of Carbon Dioxide
Direct Air Measurements
Air Bubbles in Ice Cores
Law Dome
WAIS
Dronning Maud Land
Taylor Dome
Dome C
Dome Fuji
Vostok
South Pole
2020
2000
1980
1950
1900
Carbon Dioxide in Atmosphere (ppmv)
420
400
380
360
340
320
300
280
260
240
8000
7000
6000
5000
4000
3000
2000
1000
BCE/CE
1000
2000

CO2 leads to Hotter Oceans



Source: Brian McNoldy; University of Miami Upper Ocean Dynamics Lab

What’s Up?

Next Time

Discrete Distributions

Bernoulli:
- indicator of coin flip X ~ Ber(*p*)

Binomial:
- # successes in *n* coin flips X ~ Bin(*n*, *p*)

Poisson:
- # successes in *n* coin flips X ~ Poi(λ)

Geometric:
- # coin flips until success X ~ Geo(*p*)

Negative Binomial:
- # trials until *r* successes X ~ NegBin(*r*, *p*)

Zipf:
- The popularity rank of a random word, from a natural language
- X ~ Zipf(s)