



Poisson

Chris Piech

CS109, Stanford University



HURRICANE MICHAEL

11:00 AM CDT

LAT: 29.6°N LON: 85.8°W
35 MI SW OF MEXICO BEACH FLORIDA

WINDS: 150 MPH
PRESSURE: 923 mb
MOVING: NNE at 14 MP

Birmingham

Macon

Panama City

Jacksonville

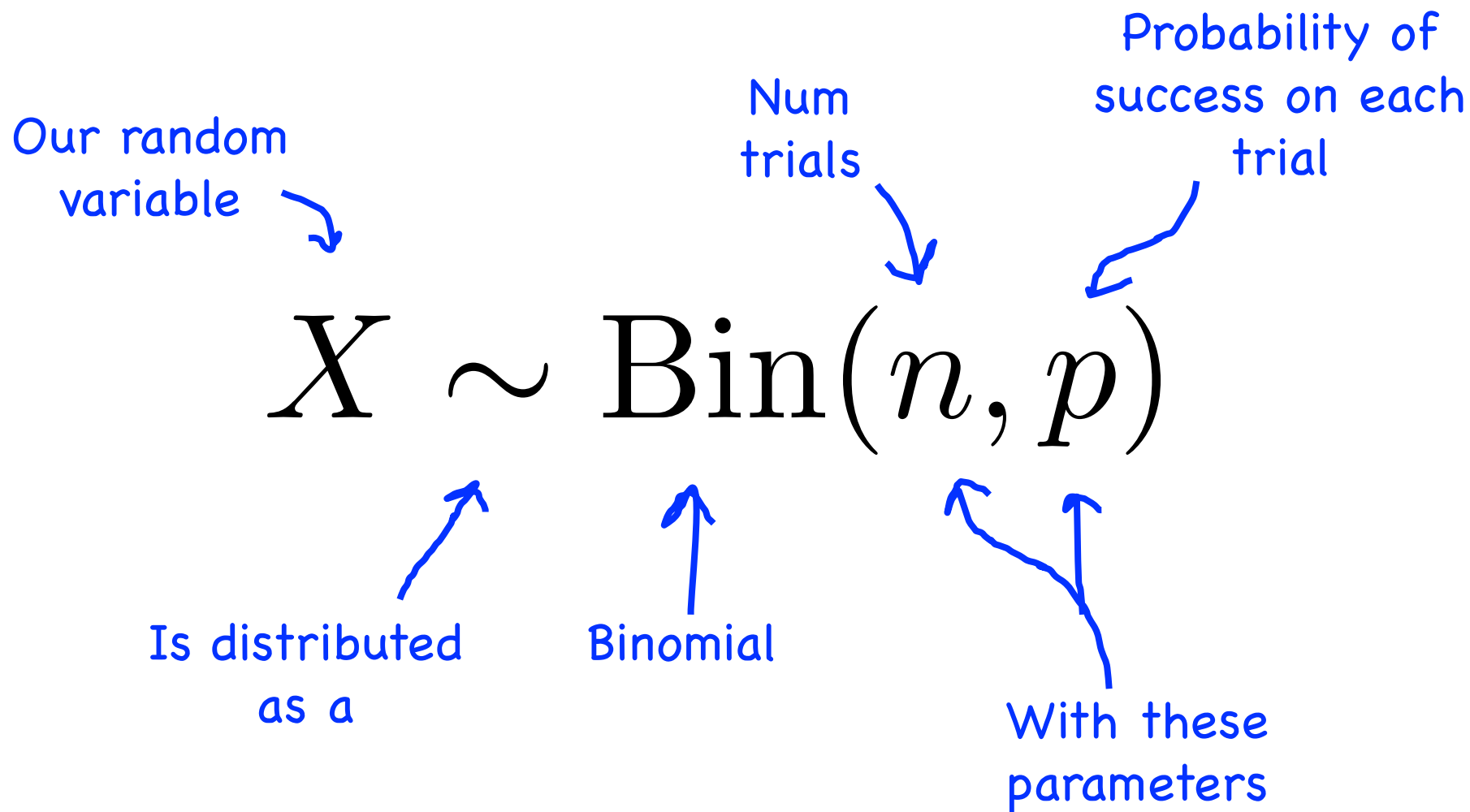
Probability for Extreme Weather?



Review


Binomial Random Variable

- Consider **n independent trials** of an experiment with success probability p .
 - X is **number of successes** in n trials
 - X is a **Binomial** Random Variable:
- 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



If X is a binomial with parameters n and p

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

Bernoulli vs Binomial

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

$$X \in \{0, 1\}$$

Bernoulli is a type of RV that can take on two values, 1 (for success) with probability p and 0 (for failure) with probability $(1 - p)$

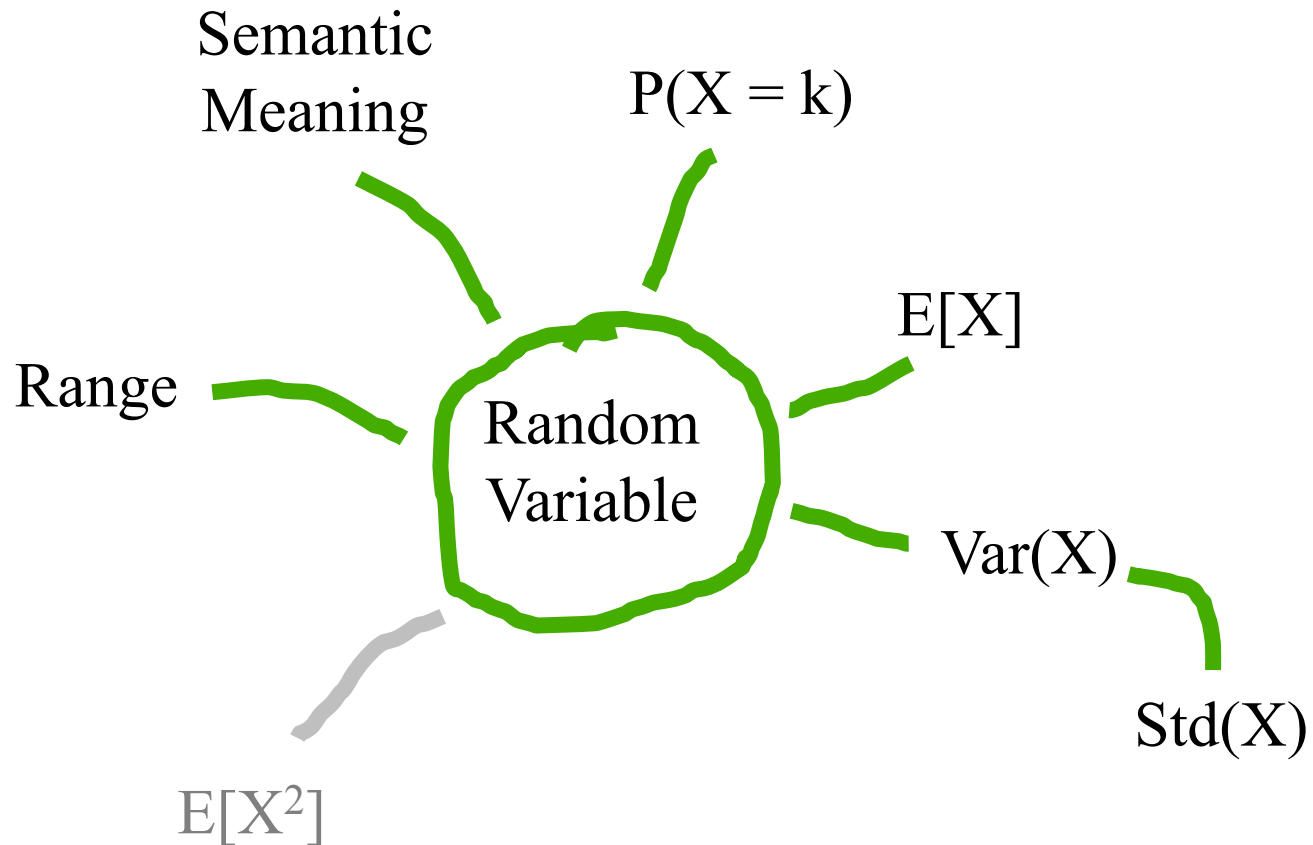
$$Y \sim \text{Bin}(n, p)$$

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

$$\text{s.t. } X_i \sim \text{Bern}(p)$$

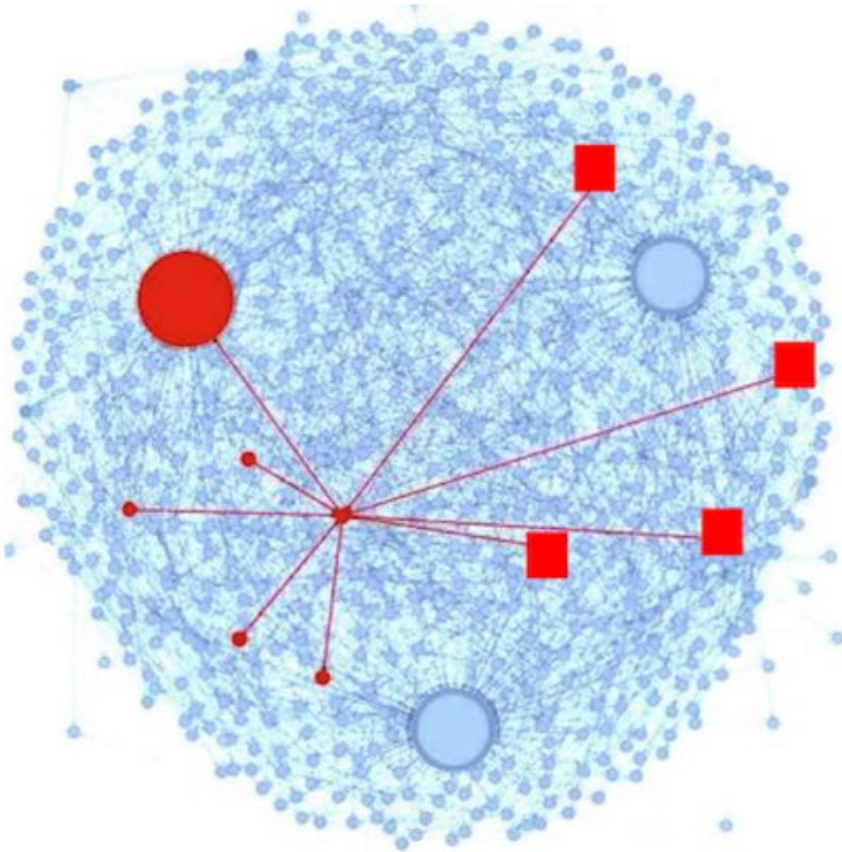
Binomial is the sum of n Bernoullis

Fundamental Properties



Is Peer Grading Accurate Enough?

Looking ahead

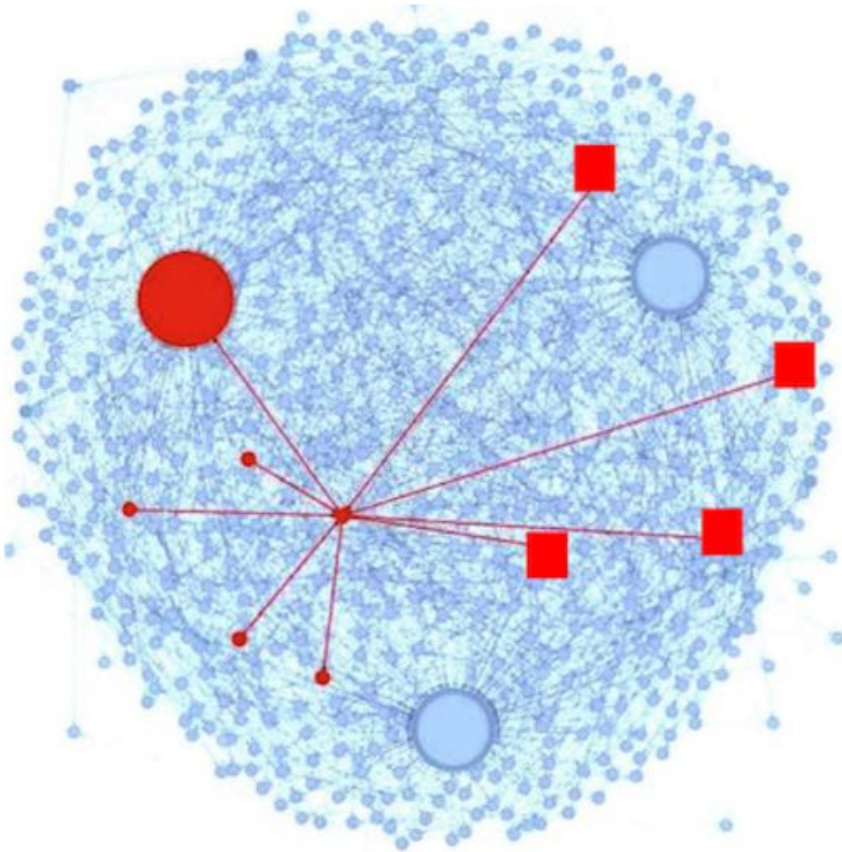


Peer Grading on Coursera
HCI.

31,067 peer grades for
3,607 students.

Is Peer Grading Accurate Enough?

Looking ahead



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

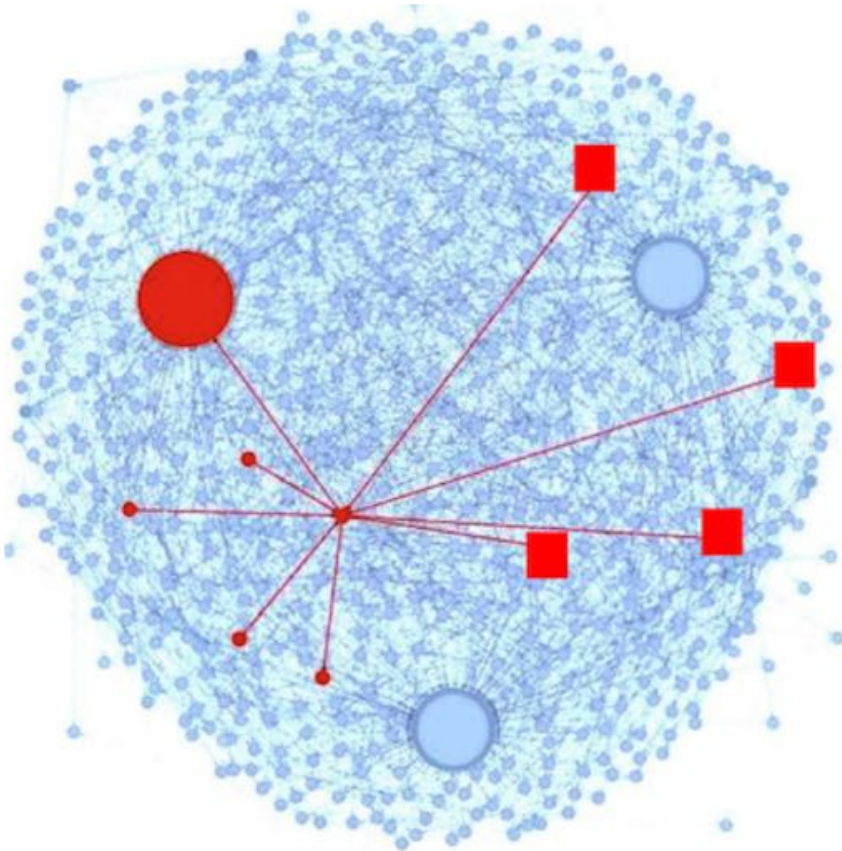
$$s_i \sim \text{Bin}(\text{points}, \theta)$$

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

Problem param
↙

Is Peer Grading Accurate Enough?

Looking ahead

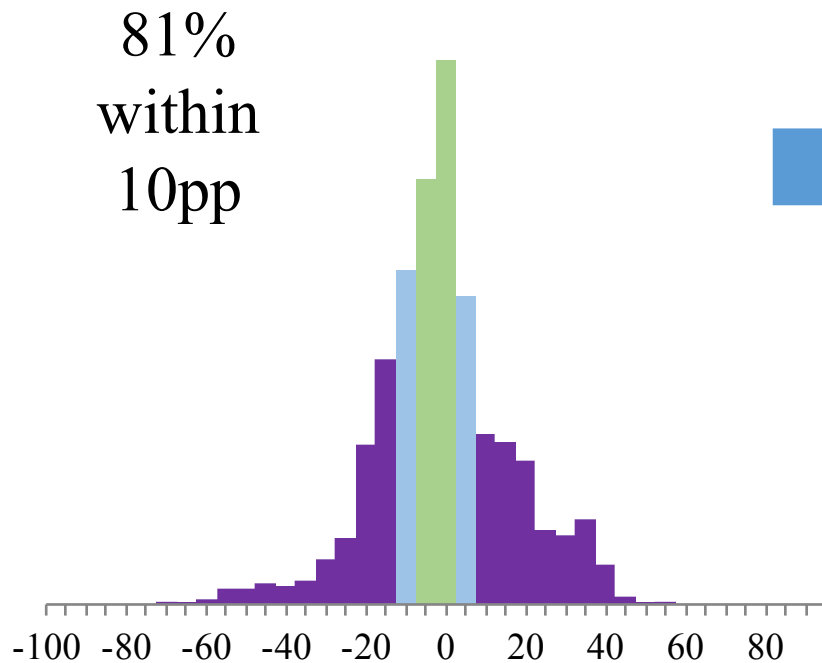


1. Defined random variables for:
 - True grade (s_i) for assignment i
 - Observed (z_i^j) score for assign i
 - Bias (b_j) for each grader j
 - Variance (r_j) for each grader j
2. Designed a probabilistic model that defined the distributions for all random variables
3. Found the variable assignments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

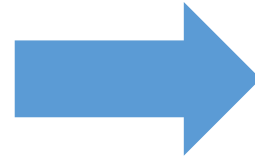
↑
Inference or Machine Learning

Yes, With Probabilistic Modelling

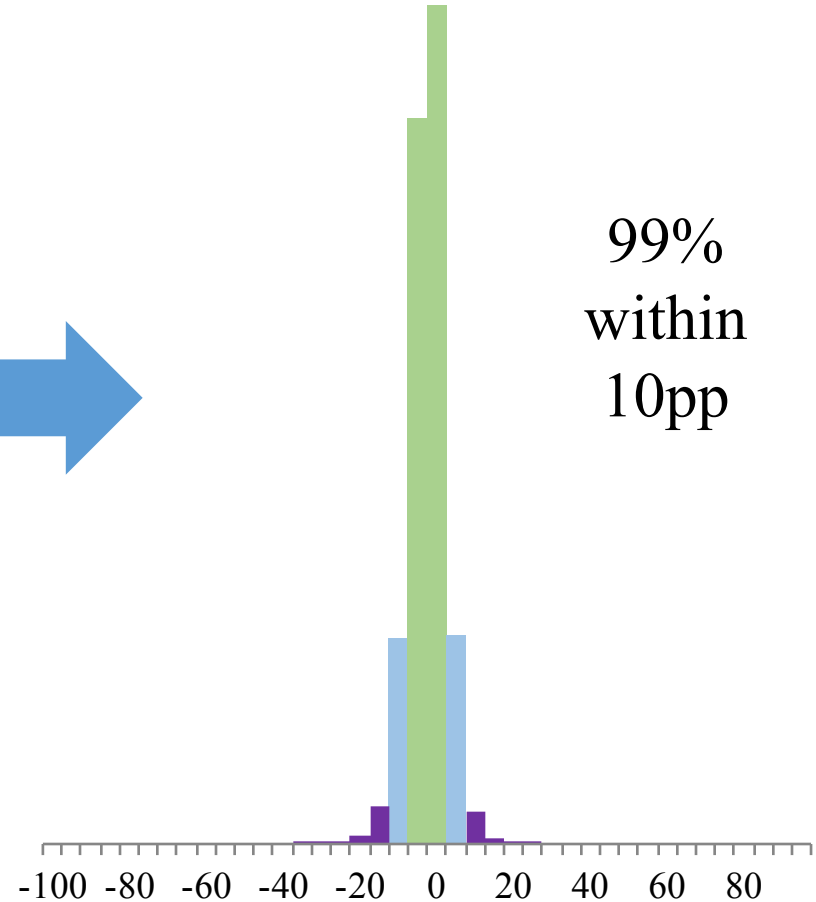
Before:



After:



99%
within
10pp



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

End Review

Good to know

Natural Exponent def:

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

Jacob
Bernoulli



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

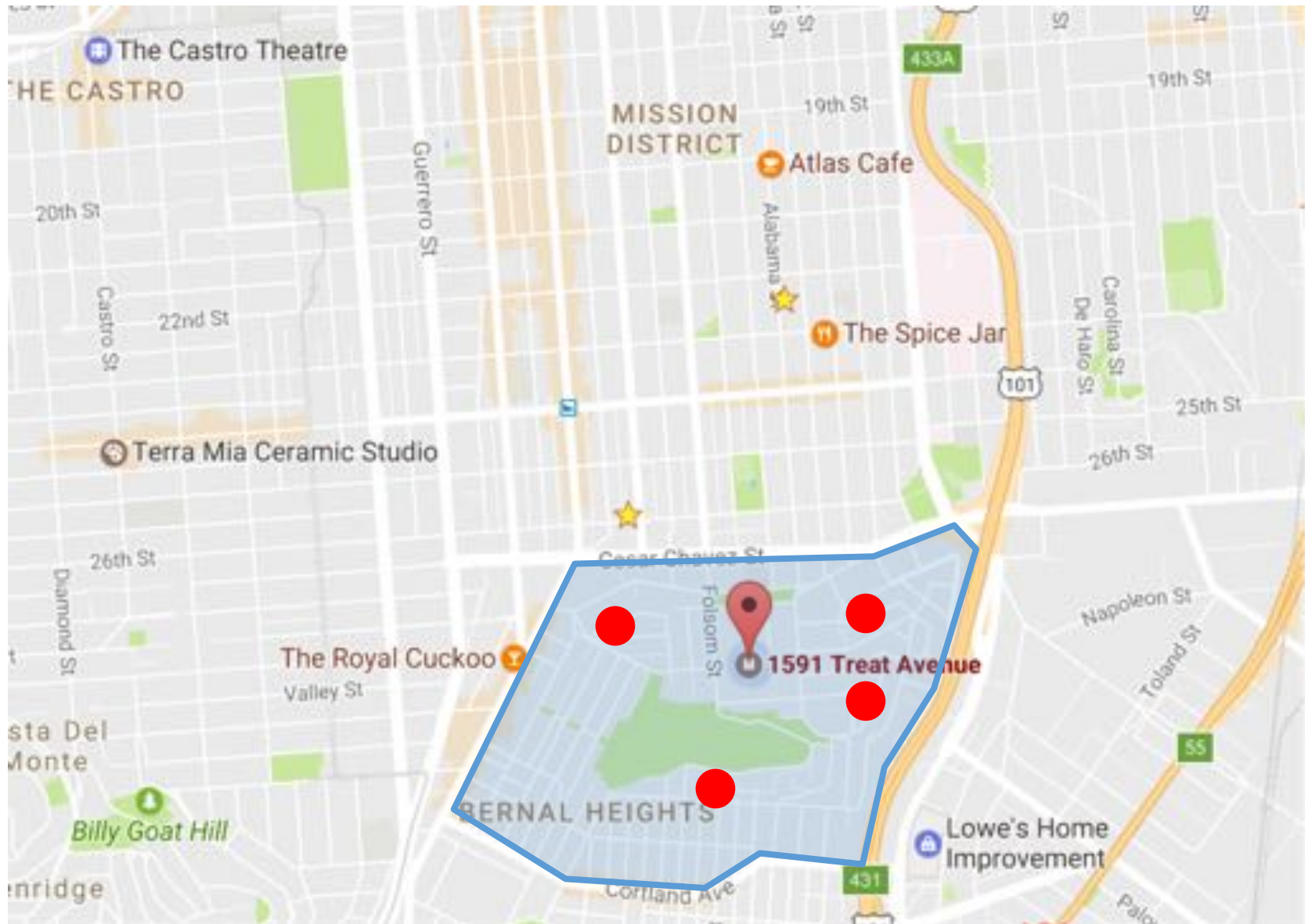
Algorithmic Ride Sharing



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



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



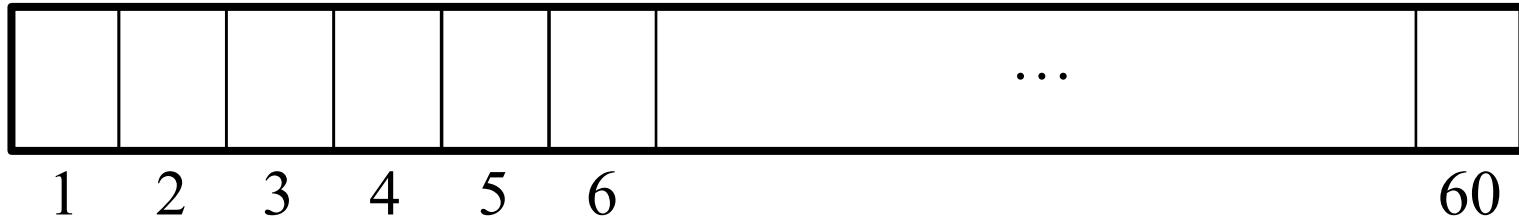
Probability of k requests from this area in the next 1 min



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

On average $\lambda = 5$ requests per minute

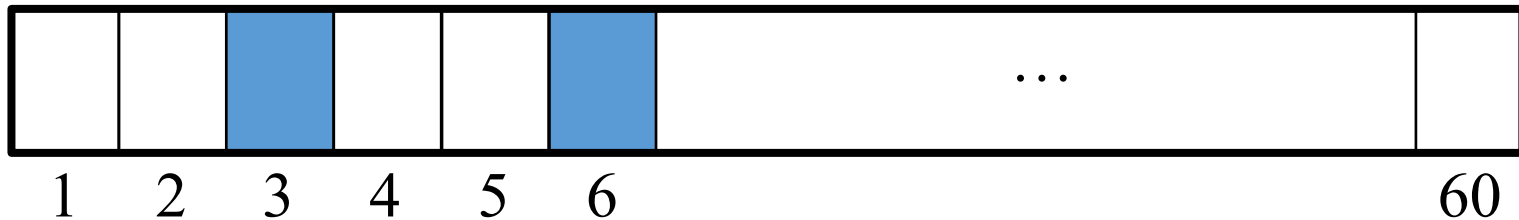
We can break the next minute down into seconds



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

On average $\lambda = 5$ requests per minute

We can break the next minute down into seconds

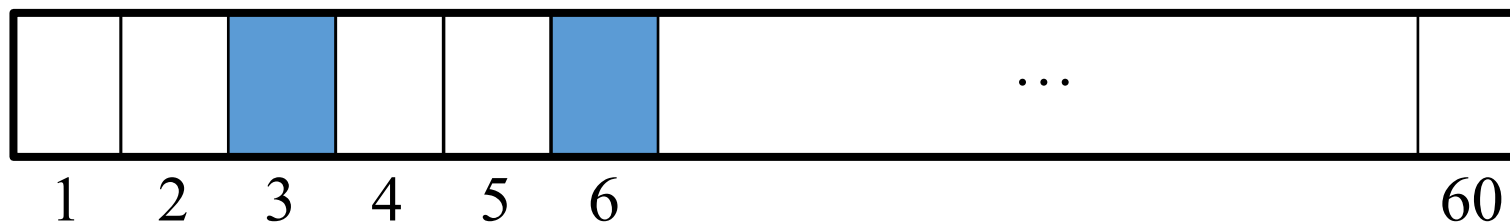


At each second either get a request or you don't.

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

On average $\lambda = 5$ 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

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

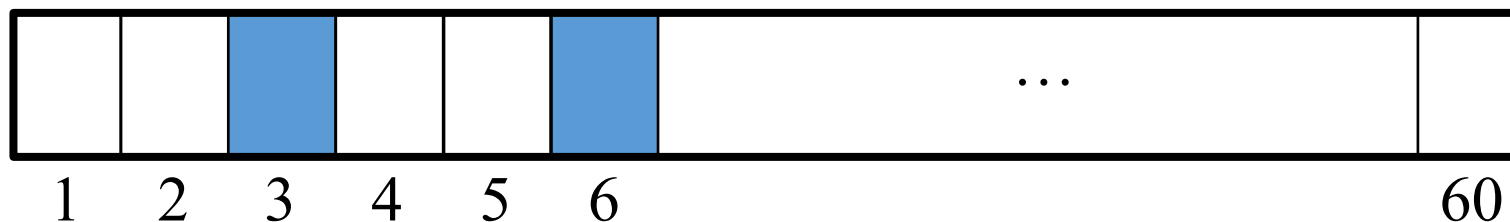
$$P(X = k) = \binom{n}{k} (p)^k (1 - p)^{n-k}$$

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

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

On average $\lambda = 5$ 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 $\lambda = 5$ 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

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

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

Can we do any better than milli-seconds?

Binomial in the Limit

On average $\lambda = 5$ 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

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

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

Who wants to see some cool math?

Binomial in the Limit

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

$$= \lim_{n \rightarrow \infty} \frac{n!}{(n-k)!k!} \cdot \frac{\lambda^k}{n^k} \cdot \frac{(1 - \lambda/n)^n}{(1 - \lambda/n)^k}$$

By expanding each term

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

By definition of natural exp

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

Rearranging terms

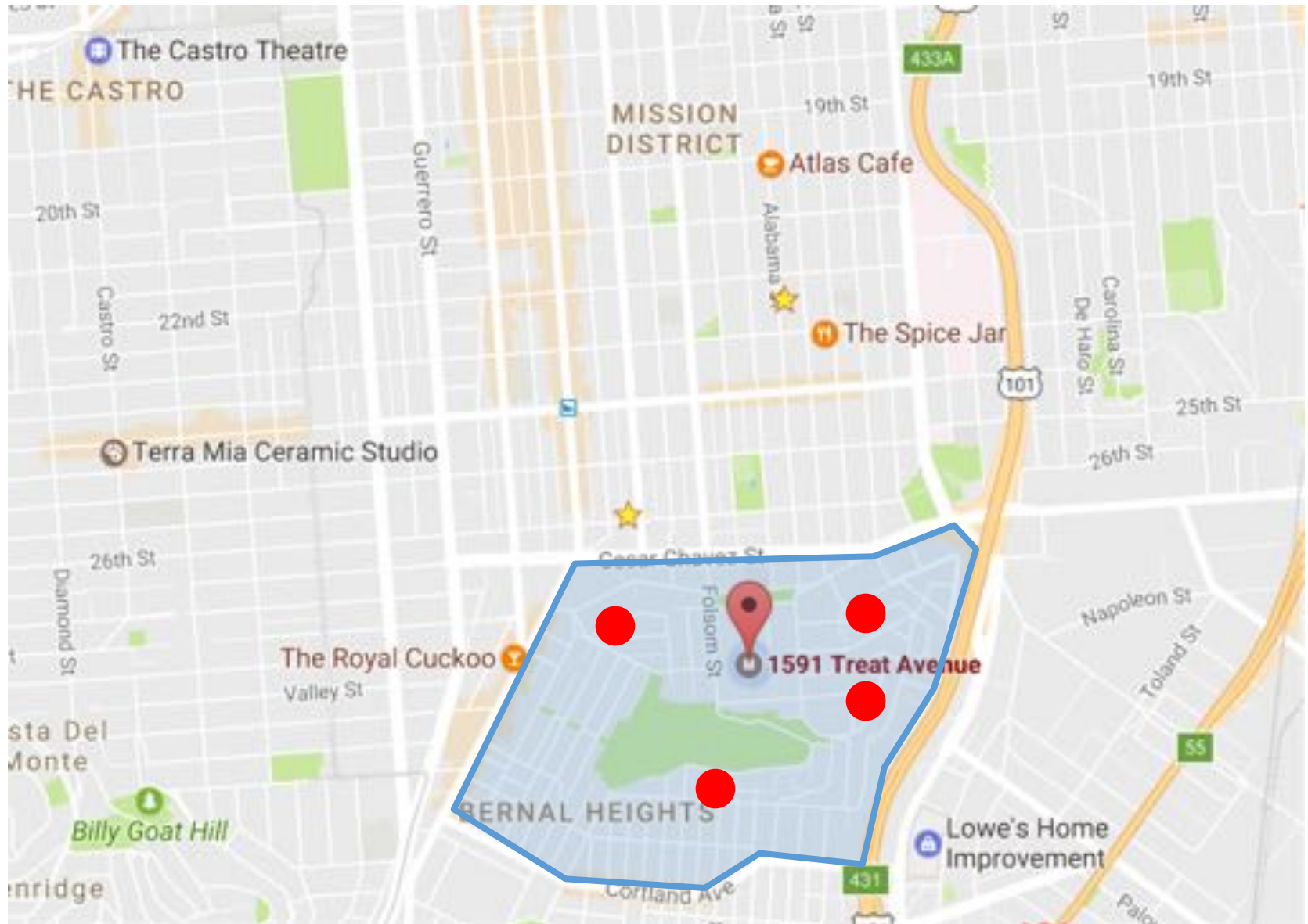
$$= \lim_{n \rightarrow \infty} \frac{n^k}{n^k} \cdot \frac{\lambda^k}{k!} \cdot \frac{e^{-\lambda}}{1}$$

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.”*
- I’m going with French Martin Freeman

Poisson Random Variable

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

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

- 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

- 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
- 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
- $N(t)$ = # events in original time interval $\sim \text{Poi}(\lambda)$



Poisson is great when you
have a rate!





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

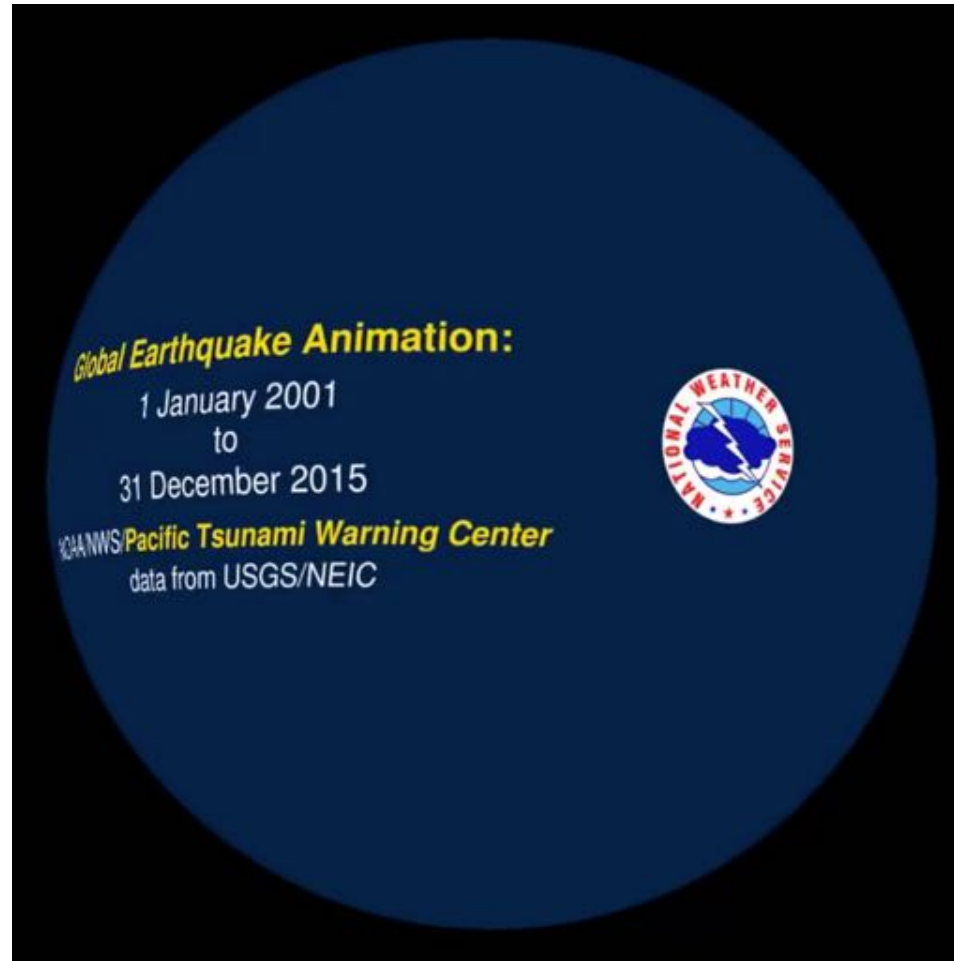




Make sure that the time unit for “rate” and match the probability question



Earthquakes



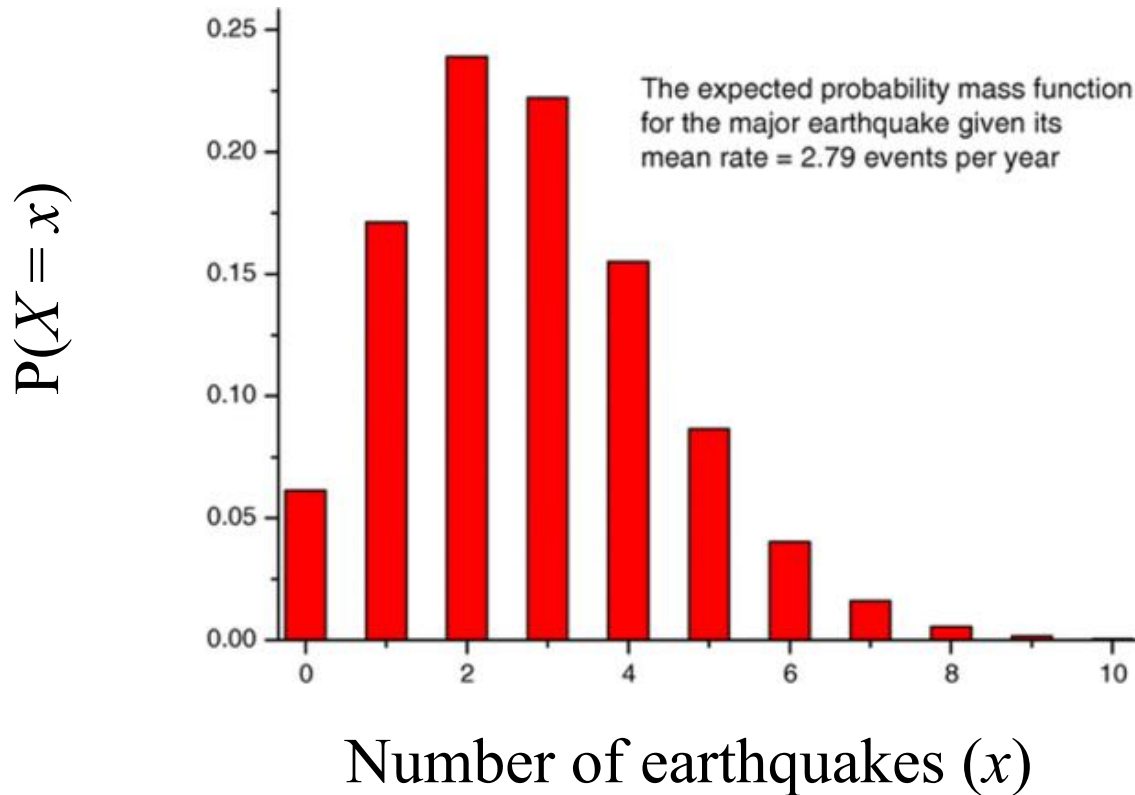
Average of 2.79 major earthquakes per year.
What is the probability of 3 major earthquakes next year?



Earthquake Probability Mass Function

Let X = number of earthquakes next year

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



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



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?

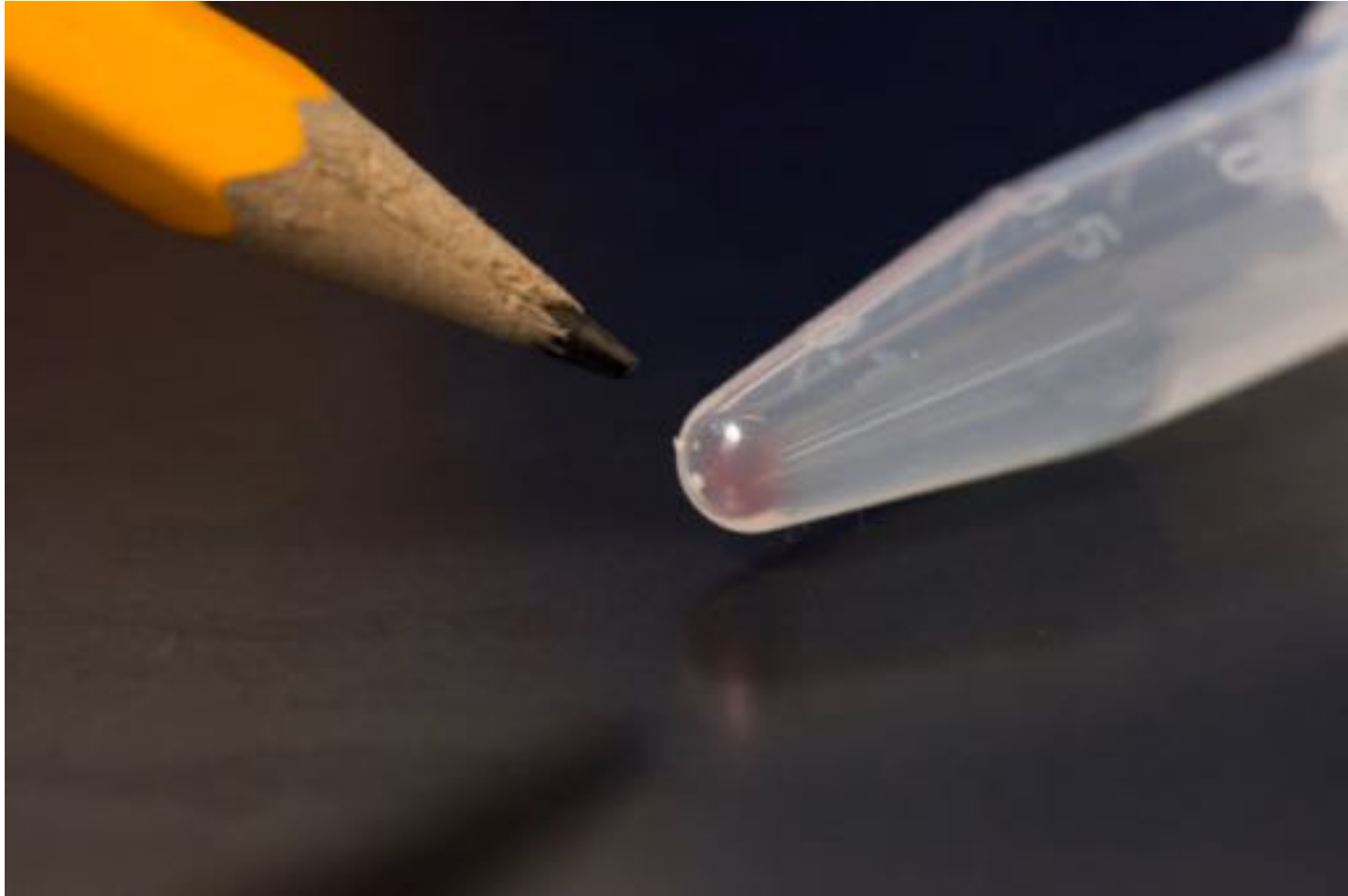
BY J. K. GARDNER and L. KNOPOFF

ABSTRACT

Yes.

Poisson can approximate a Binomial!

Storing Data on 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 on 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 $p \approx 10^{-6}$
 - $X \sim \text{Bin}(10^4, 10^{-6})$ 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 on DNA

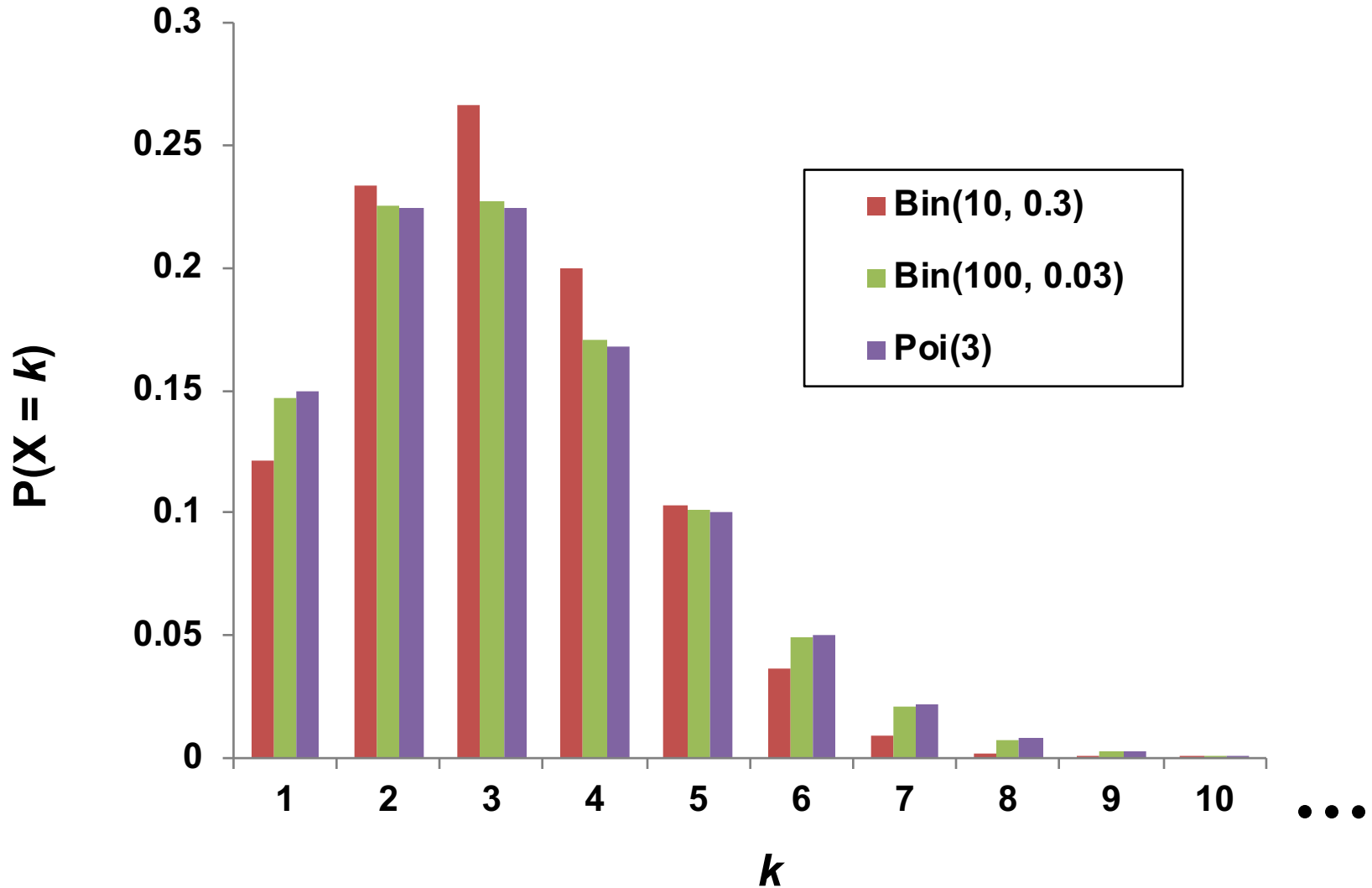
- Will the DNA storage become corrupt?
 - In DNA (and real networks) store large strings
 - Length $n \approx 10^4$
 - Probability of corruption of each base pair is very small $p \approx 10^{-6}$
 - $X \sim \text{Poi}(\lambda = 10^4 * 10^{-6} = 0.01)$

$$\begin{aligned} P(X = k) &= e^{-\lambda} \frac{\lambda^k}{k!} \\ P(X = 0) &= e^{-\lambda} \frac{1}{0!} \\ &= e^{-0.01} \approx 0.99 \end{aligned}$$

Poisson is 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)





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

A Real License Plate Seen at Stanford



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

Poisson is Chill

- Poisson can still provide a good approximation even when assumptions are “mildly” violated
- “Poisson Paradigm”
- Can apply Poisson approximation when...
 - “Successes” in trials are not entirely independent
 - Example: # entries in each bucket in large hash table
 - Probability of “Success” in each trial varies (slightly)
 - Small relative change in a very small p
 - 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

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

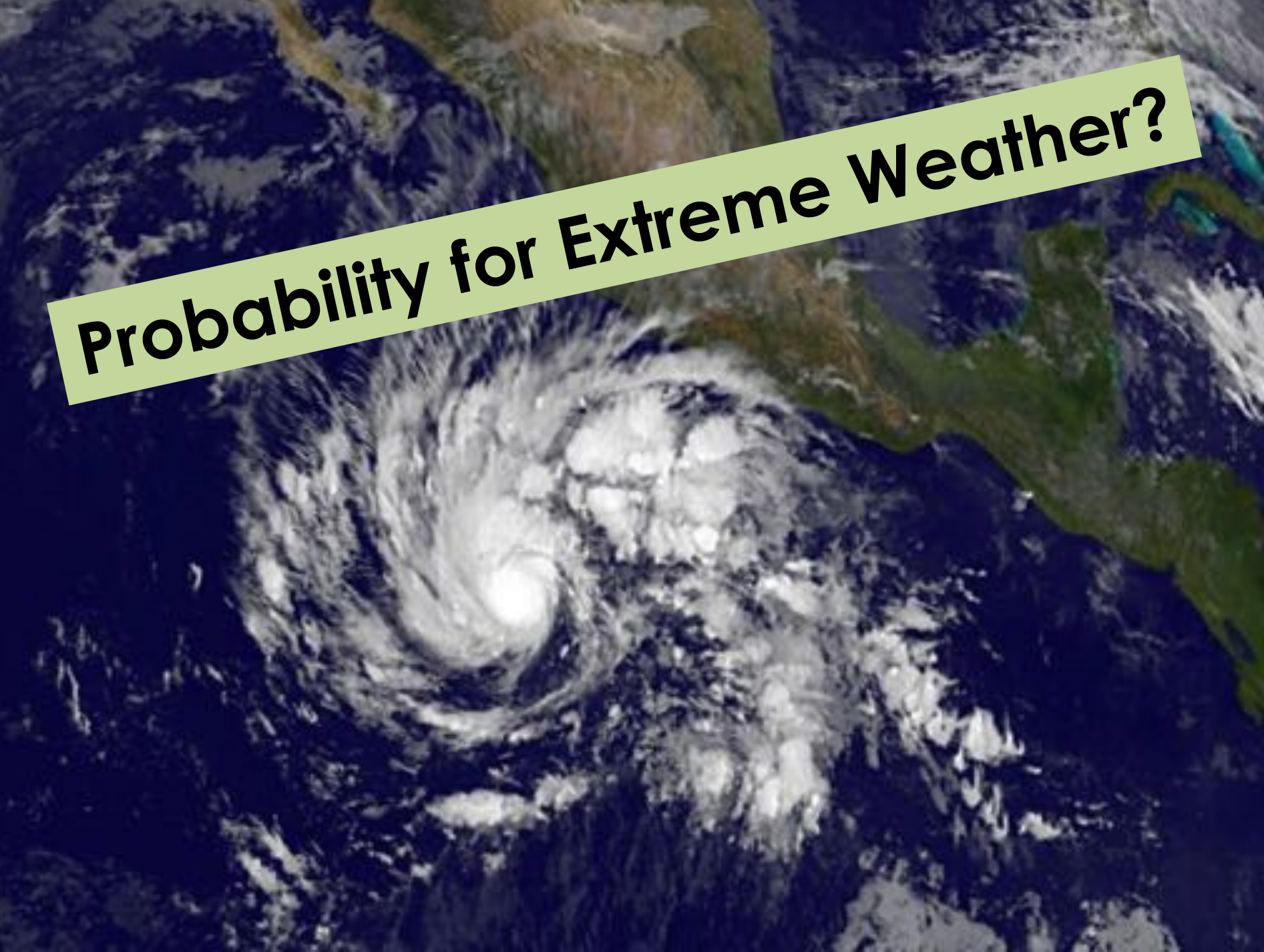
$$P(X < 5) = \sum_{i=0}^4 P(X = i)$$

$$= \sum_{i=0}^4 e^{-\lambda} \frac{\lambda^i}{i!}$$

Since X is Poisson

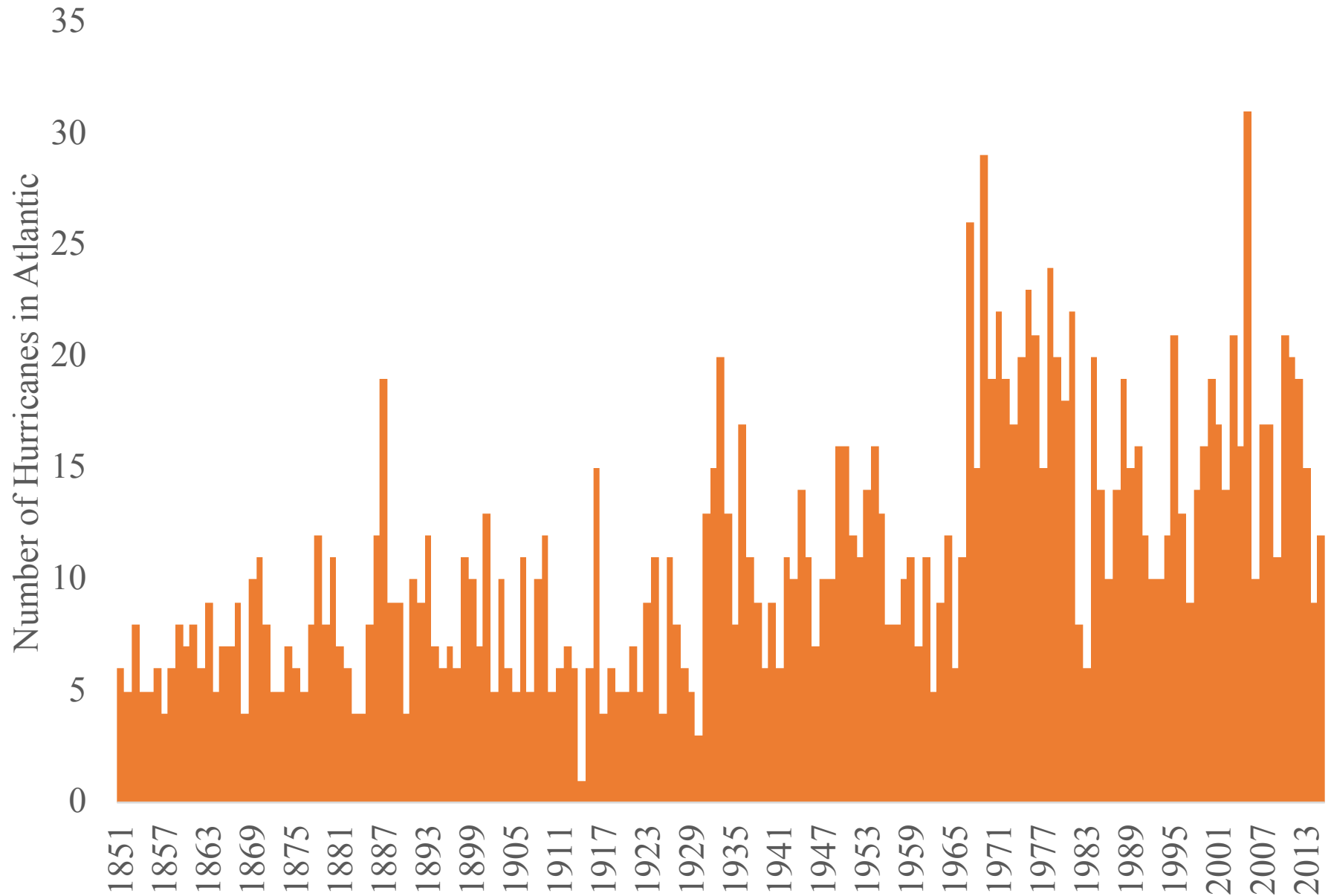
$$= \sum_{i=0}^4 e^{-2} \frac{2^i}{i!} \approx 0.95$$

Since $\lambda = 2$



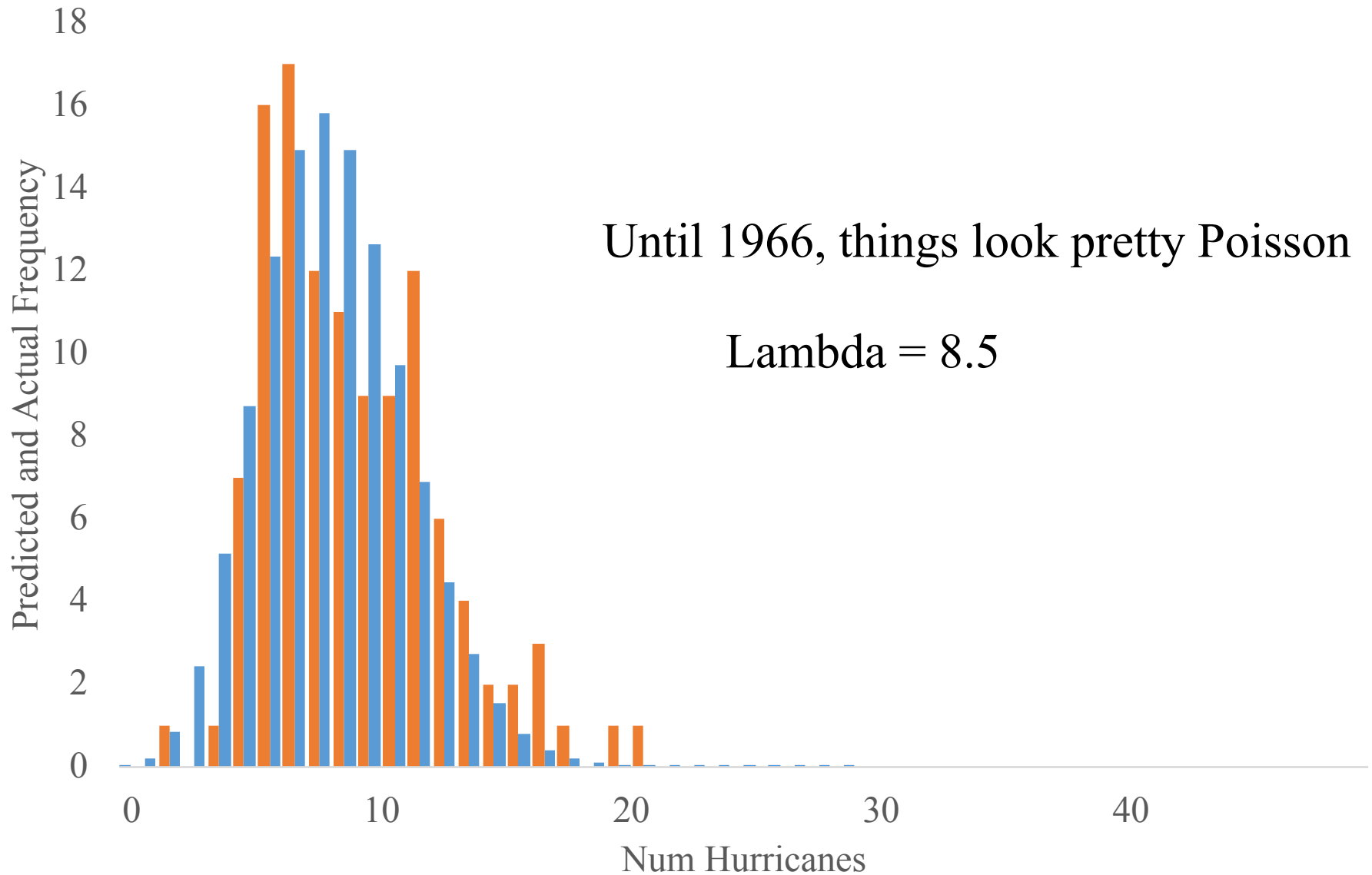
Probability for Extreme Weather?

Hurricanes per Year since 1851



To the code!

Historically ~ Poisson(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. $X \sim \text{Poi}(8.5)$

- Solution:

$$\begin{aligned}P(X > 15) &= 1 - P(X \leq 15) \\&= 1 - \sum_{i=0}^{15} P(X = i) \\&= 1 - 0.98 \\&= 0.02\end{aligned}$$

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

Twice since 1966 there have been
years with over 30 hurricanes

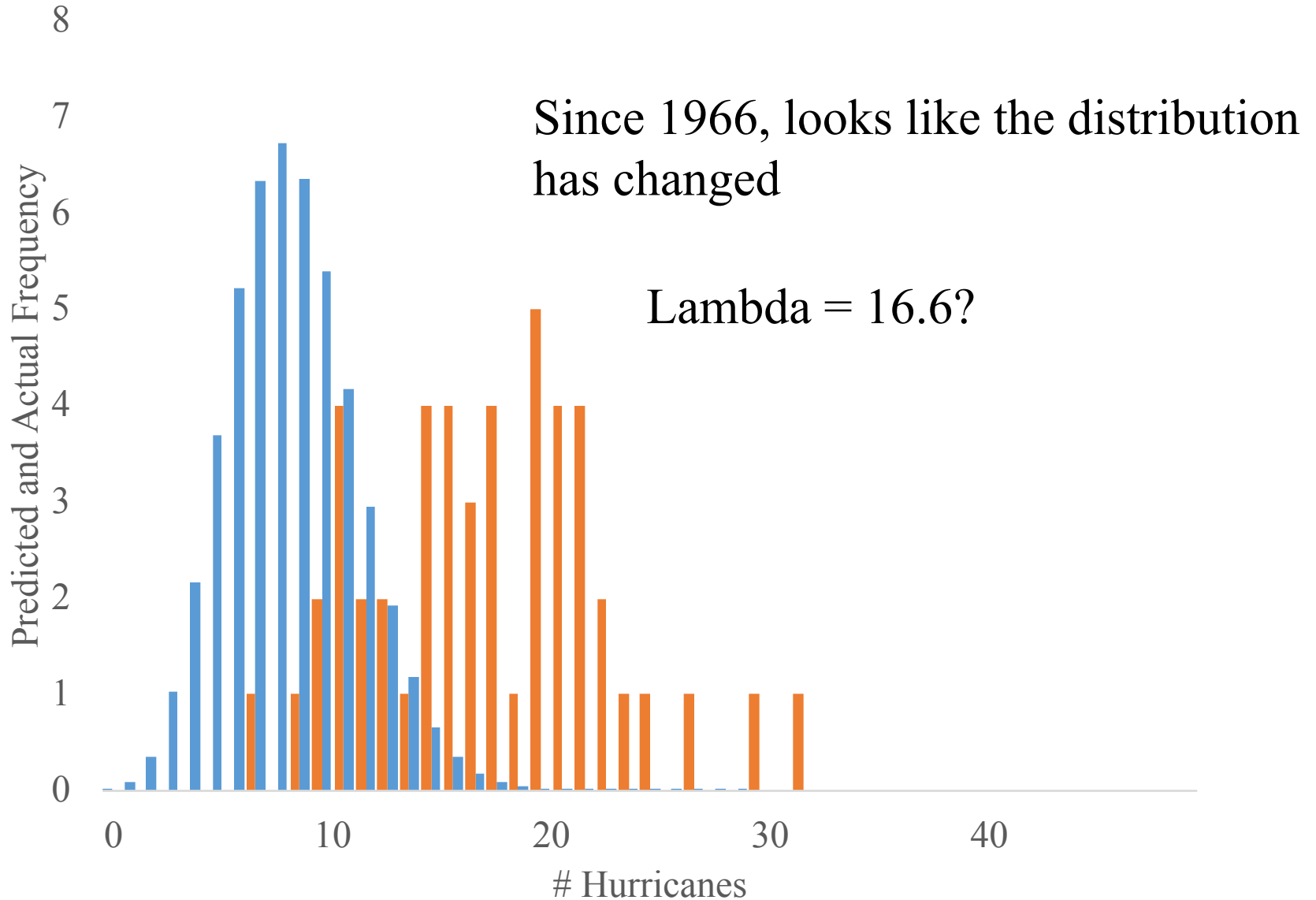
Improbability Drive

- What is the probability of over 30 hurricanes in a season given that the distribution doesn't change?
 - Let $X = \#$ hurricanes in a year. $X \sim \text{Poi}(8.5)$
- Solution:

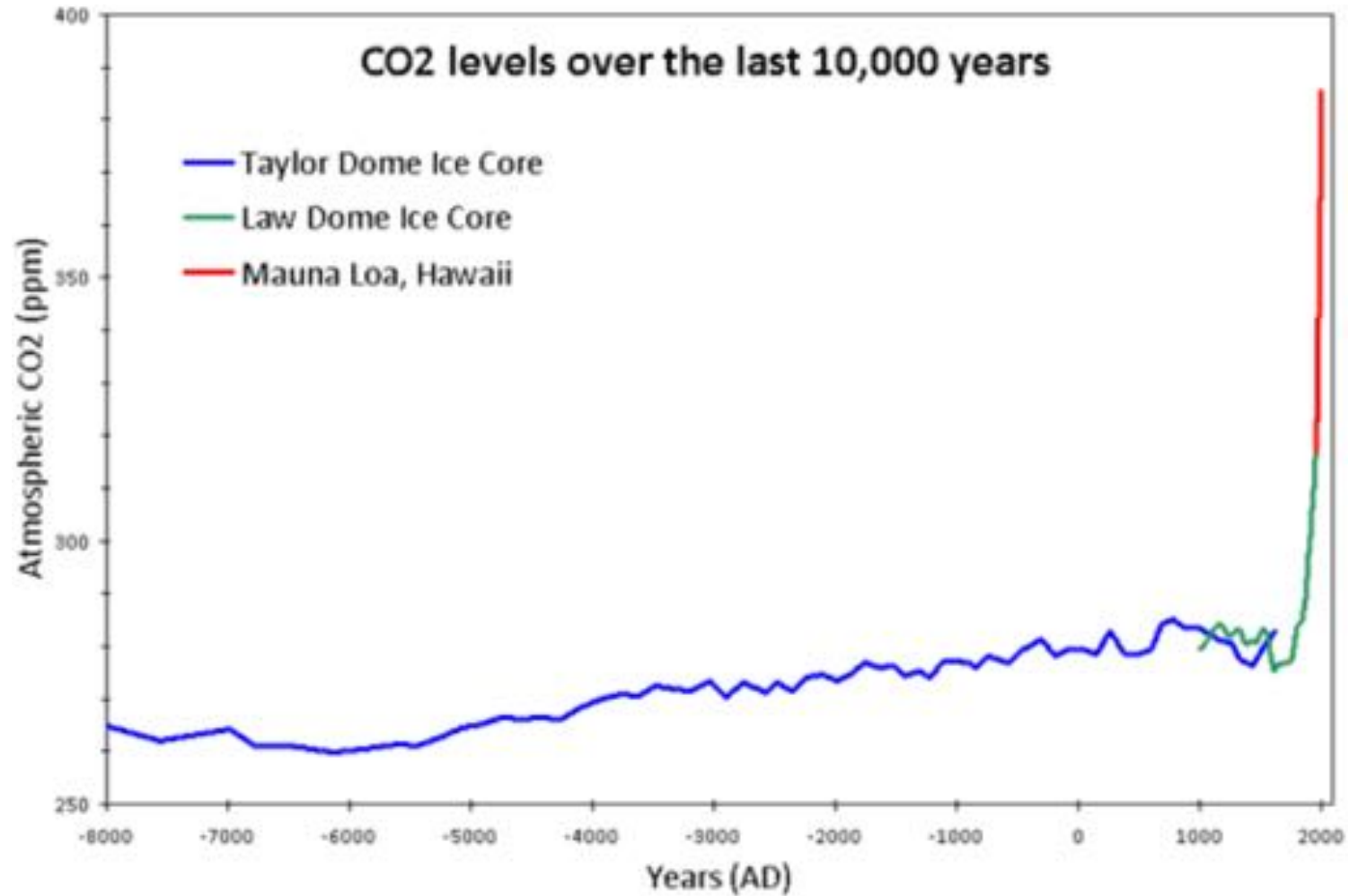
$$\begin{aligned}P(X > 30) &= 1 - P(X \leq 30) \\&= 1 - \sum_{i=0}^{30} P(X = i) \\&= 1 - 0.9999999997823 \\&= 2.2e - 09\end{aligned}$$

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

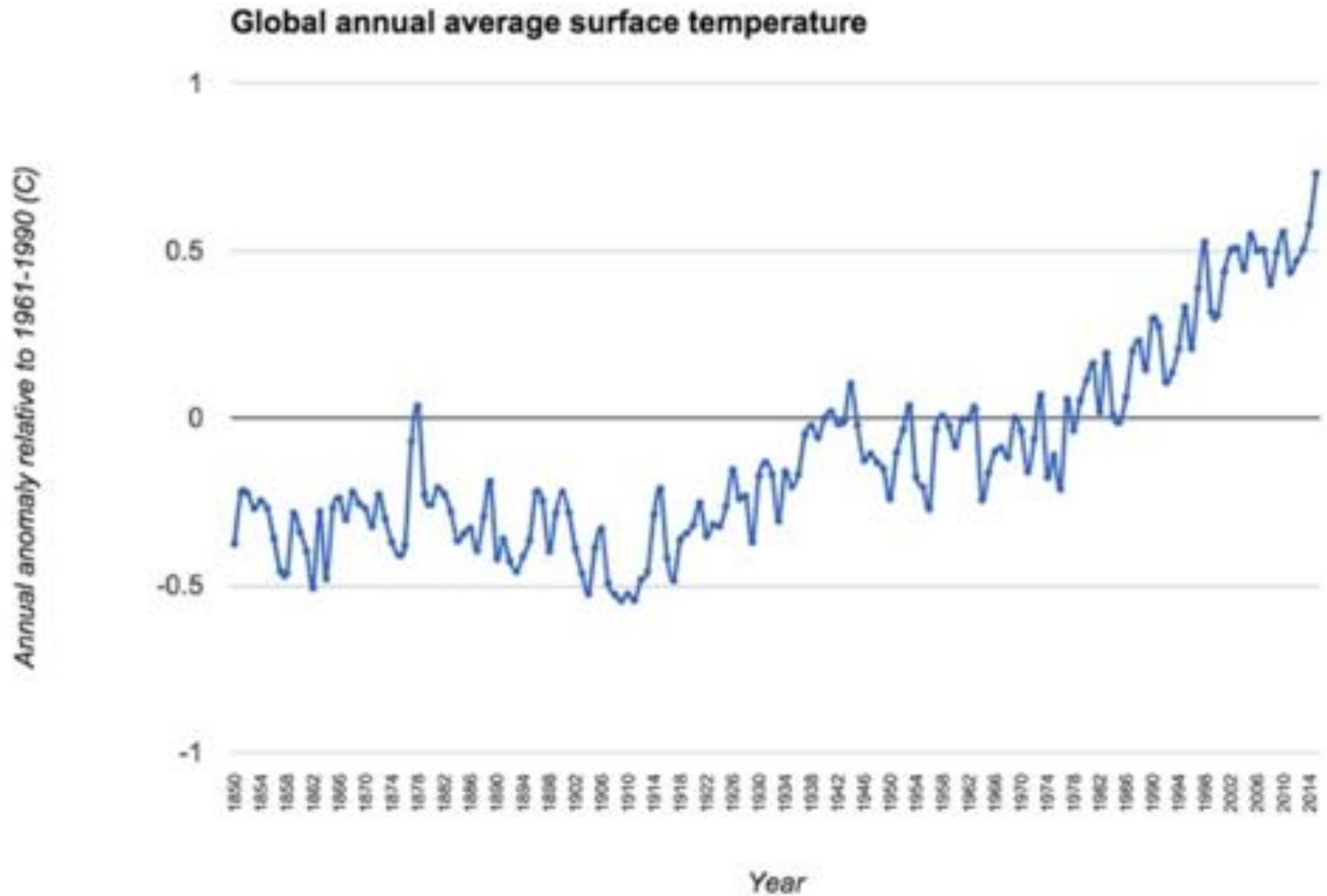
The Distribution has Changed



What's Up?



What's Up?



What's Up?



Python Scipy RV Methods

```
from scipy import stats # great package
X = stats.poisson(2.5)   #  $X \sim \text{Poi}(\lambda = 2.5)$ 
print(X.pmf(2))         #  $P(X = 2)$ 
```

Function	Description
<code>X.pmf(k)</code>	$P(X = k)$
<code>X.cdf(k)</code>	$P(X \leq k)$
<code>X.entropy()</code>	(Differential) entropy of X
<code>X.mean()</code>	$E[X]$
<code>X.var()</code>	$\text{Var}(X)$
<code>X.std()</code>	$\text{Std}(X)$

Next Time

Discrete Distributions

Bernoulli:

- indicator of coin flip $X \sim \text{Ber}(p)$

Binomial:

- # successes in n coin flips $X \sim \text{Bin}(n, p)$

Poisson:

- # successes in n coin flips $X \sim \text{Poi}(\lambda)$

Geometric:

- # coin flips until success $X \sim \text{Geo}(p)$

Negative Binomial:

- # trials until r successes $X \sim \text{NegBin}(r, p)$

Zipf:

- The popularity rank of a random word, from a natural language
- $X \sim \text{Zipf}(s)$

The Poisson Common Path

