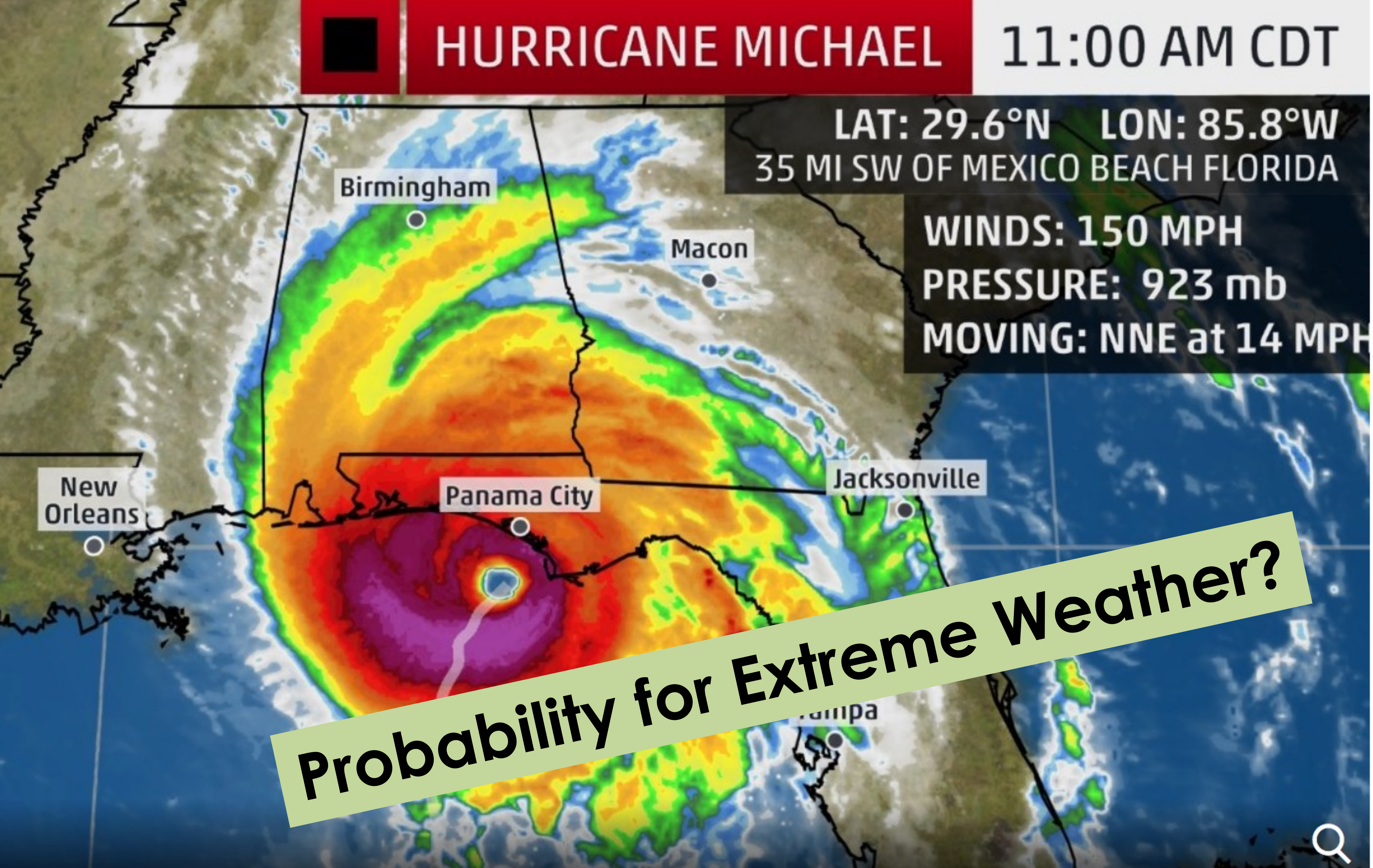
Poisson

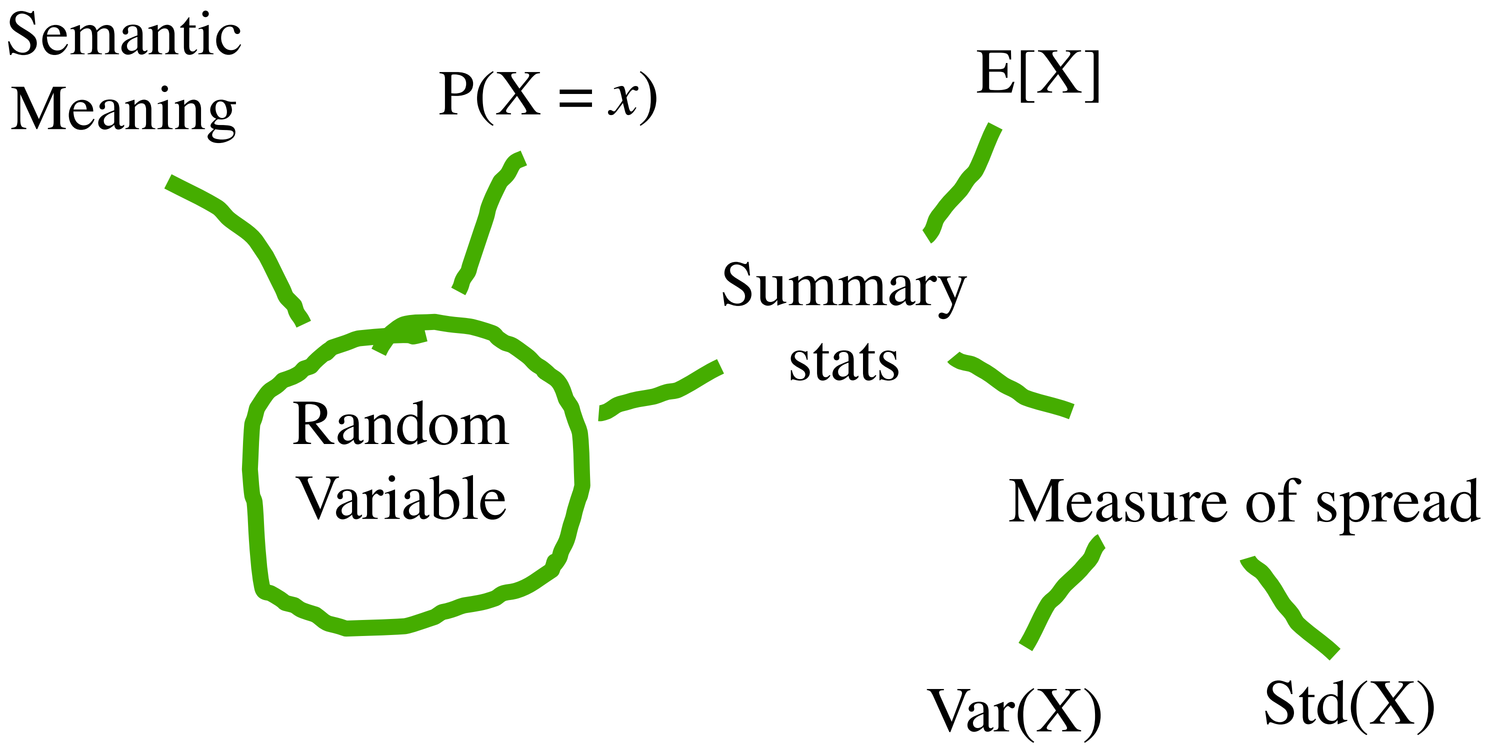
Chris Piech
CS109, Stanford University



Probability for Extreme Weather?

Review

Fundamental Properties of Random Variables



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**, in ***n* coin flips**, where each flip is heads with probability p:

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

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:

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

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

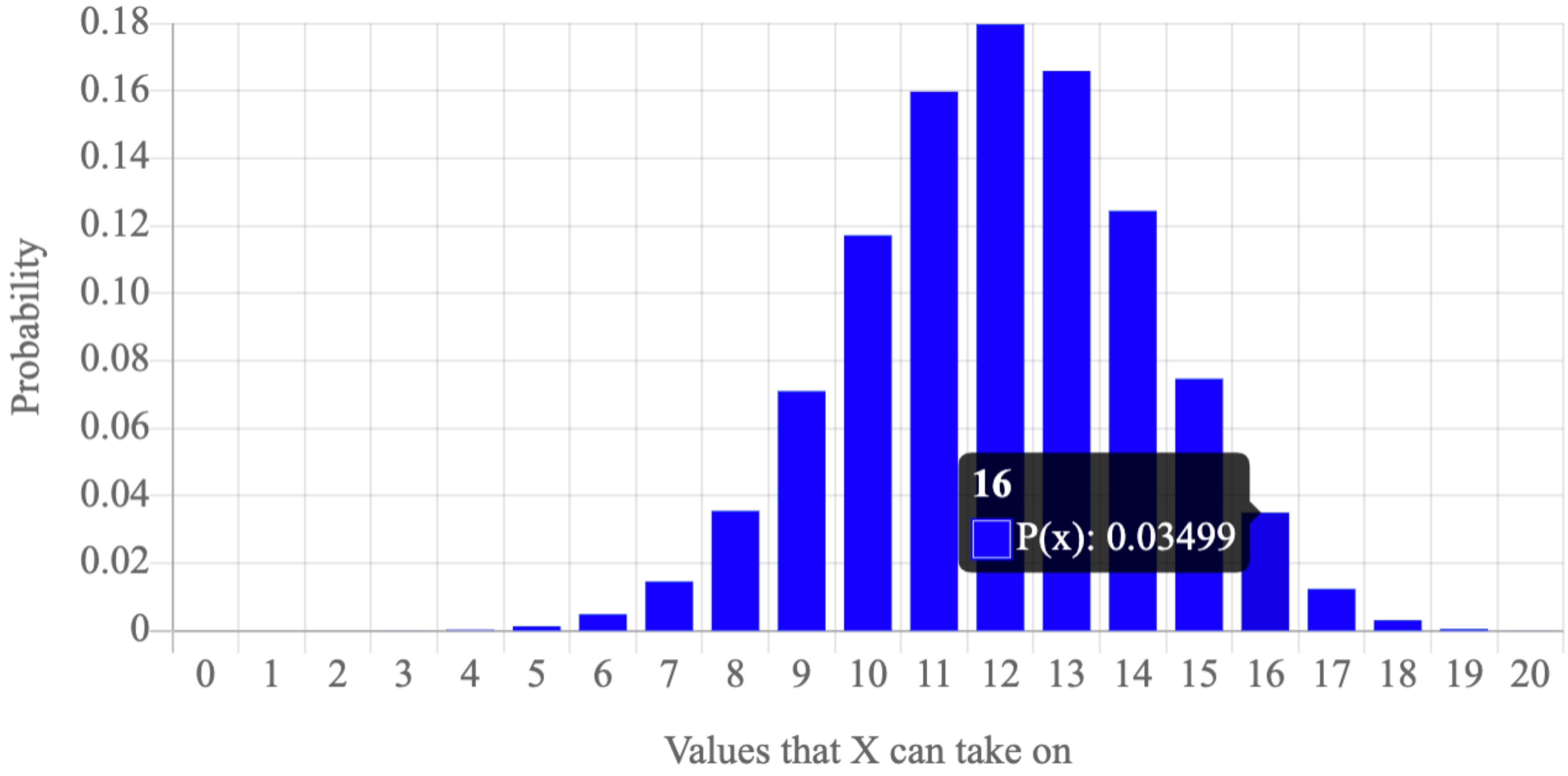
$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Probability that our variable takes on the value k

* This is also called the binomial term

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

Parameter n: 20 Parameter p: 0.60



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

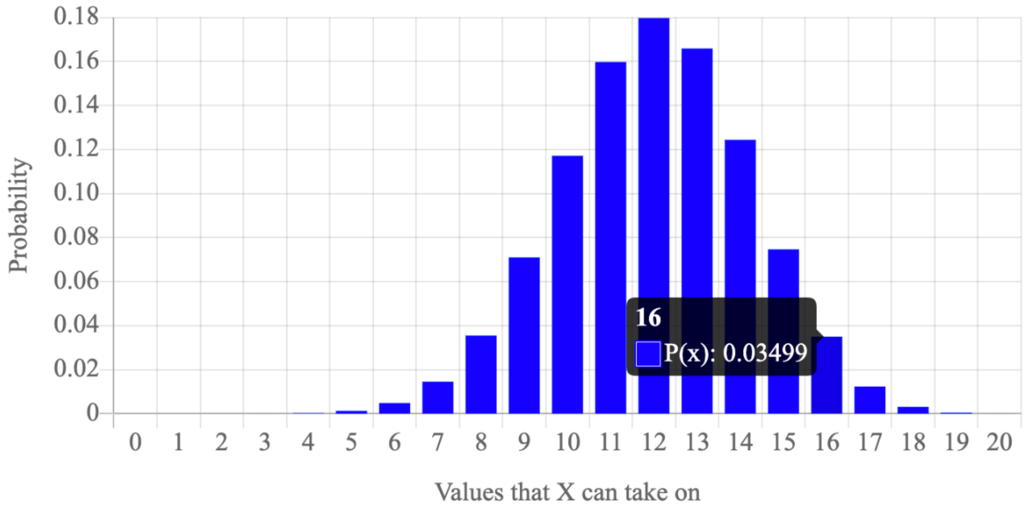
PMF equation: $\Pr(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$

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

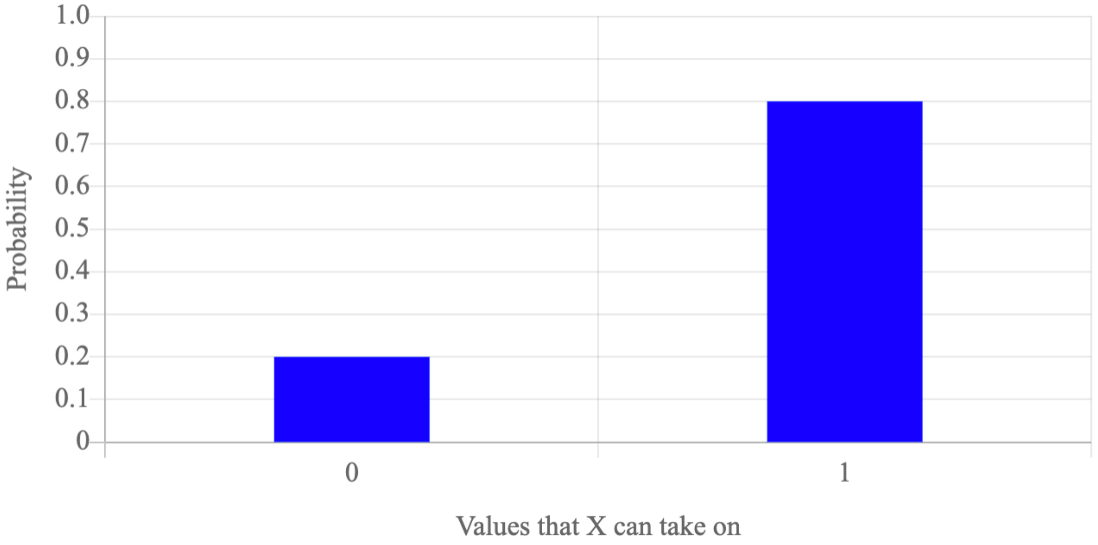
PMF equation: $\Pr(X = x) = \begin{cases} p & \text{if } x = 1 \\ 1 - p & \text{if } x = 0 \end{cases}$

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

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

PMF graph:

Parameter p: 0.80

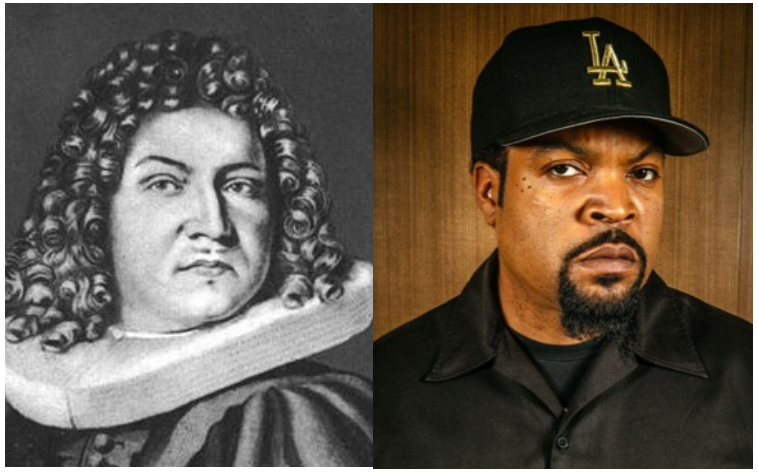


Natural Exponent Definition

Natural Exponent def:

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

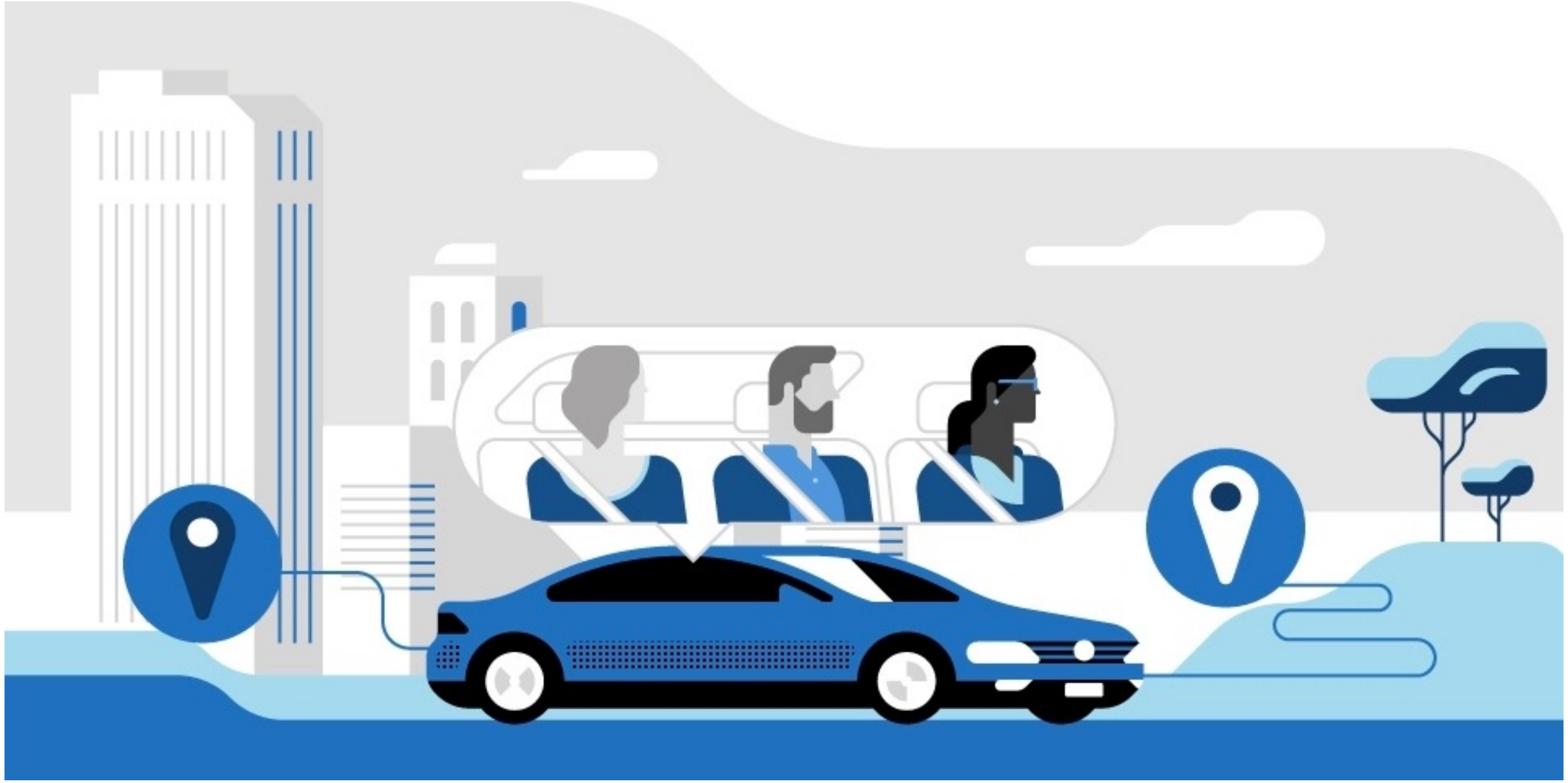
Jacob
Bernoulli



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

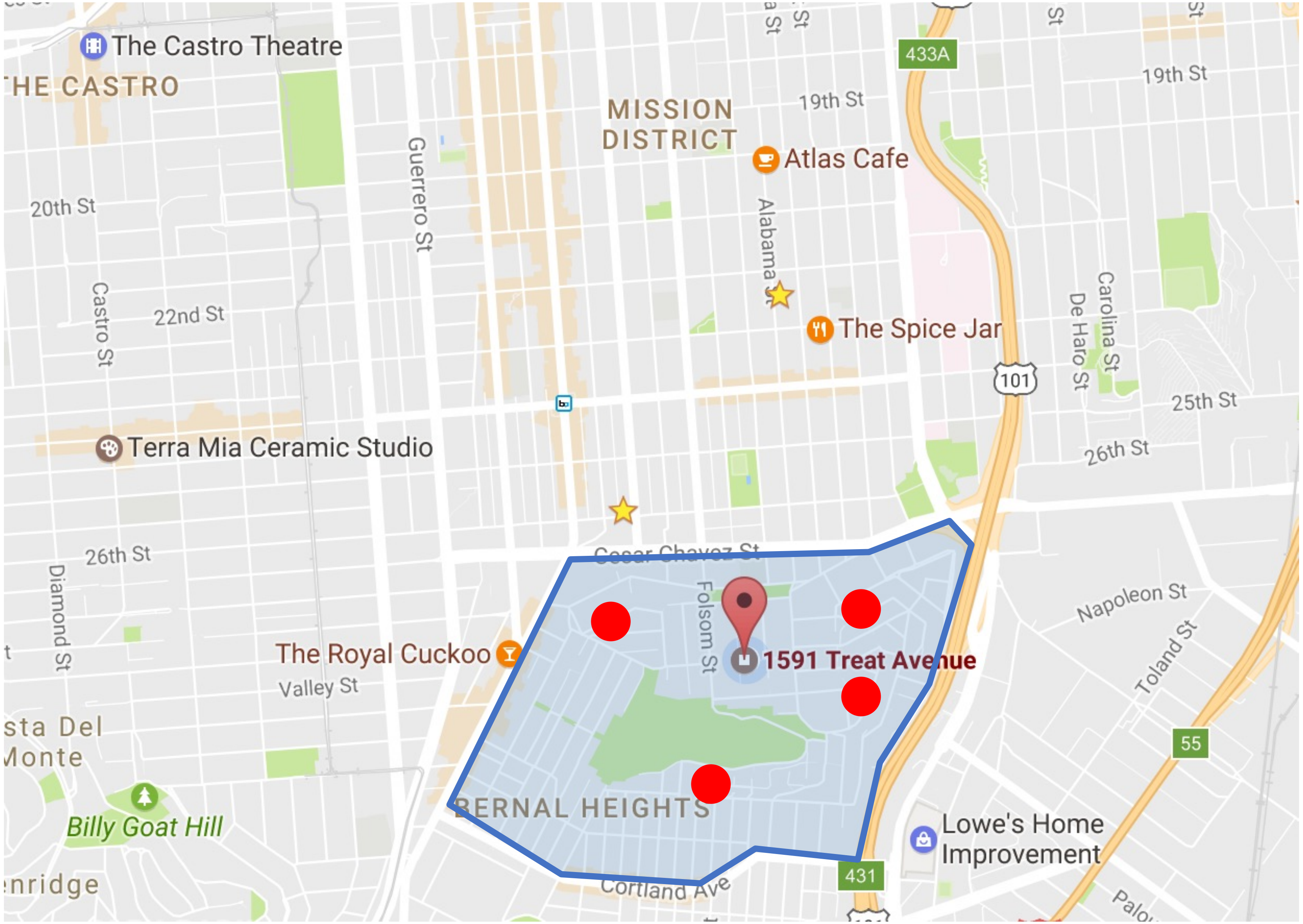
End Review

Algorithmic Ride Sharing

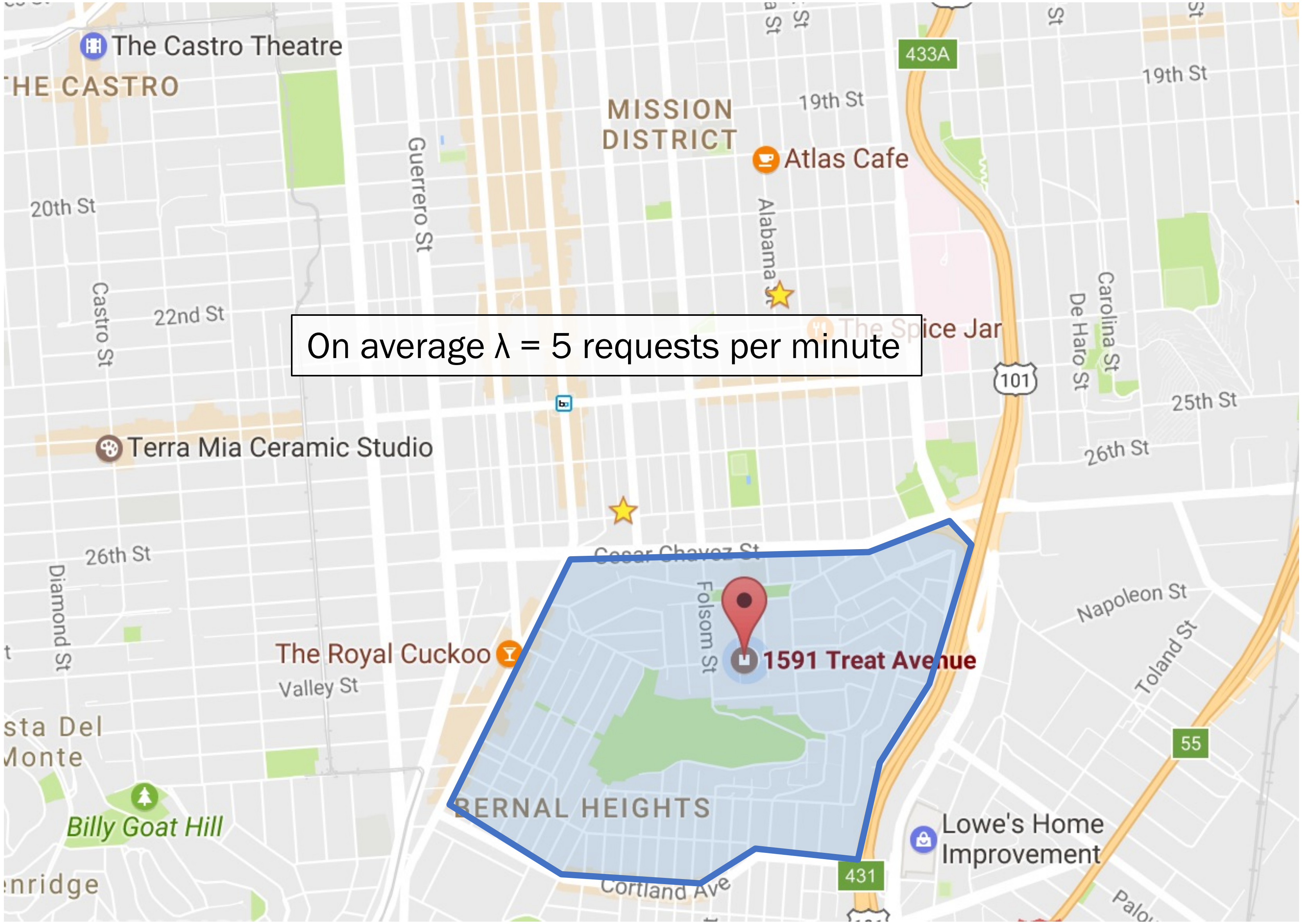


Probability of *k* **requests** from this area in the next 1 min

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



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

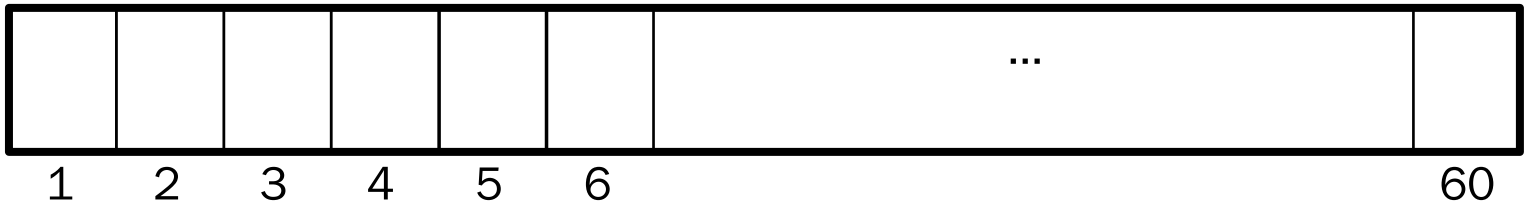


On average λ = 5 requests per minute

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



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

On average λ = 5 requests per minute

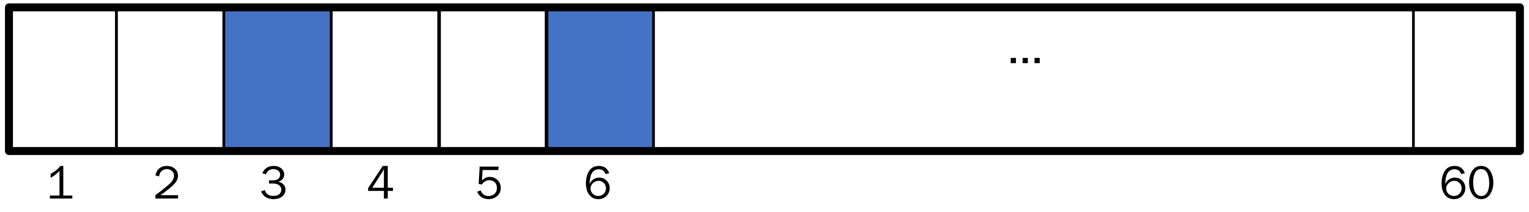
We can break the next minute down into seconds

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

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

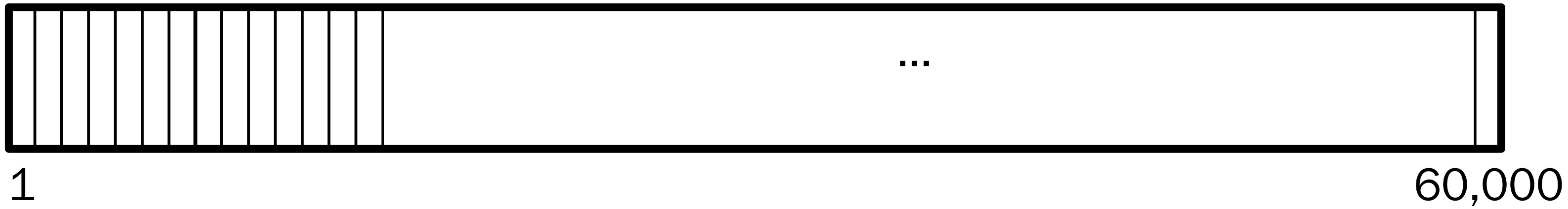
$$P(X = k) = \binom{n}{k} (p)^k (1-p)^{n-k}$$

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 λ = 5 requests per minute

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



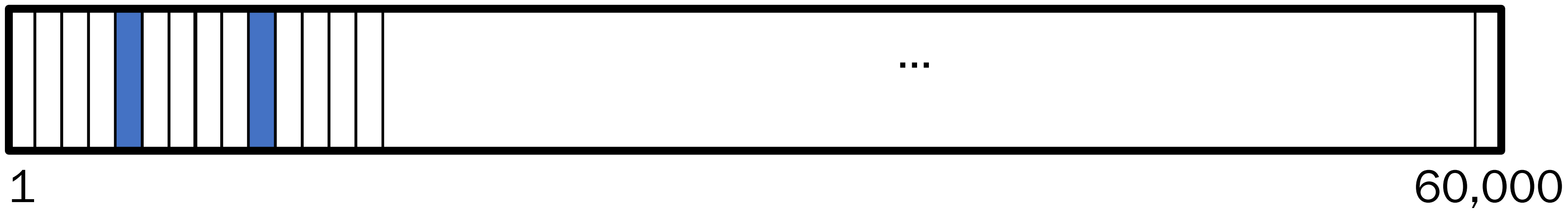
At each *milli*-second either get a request or you don't.
Let X = Number of requests in the minute

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

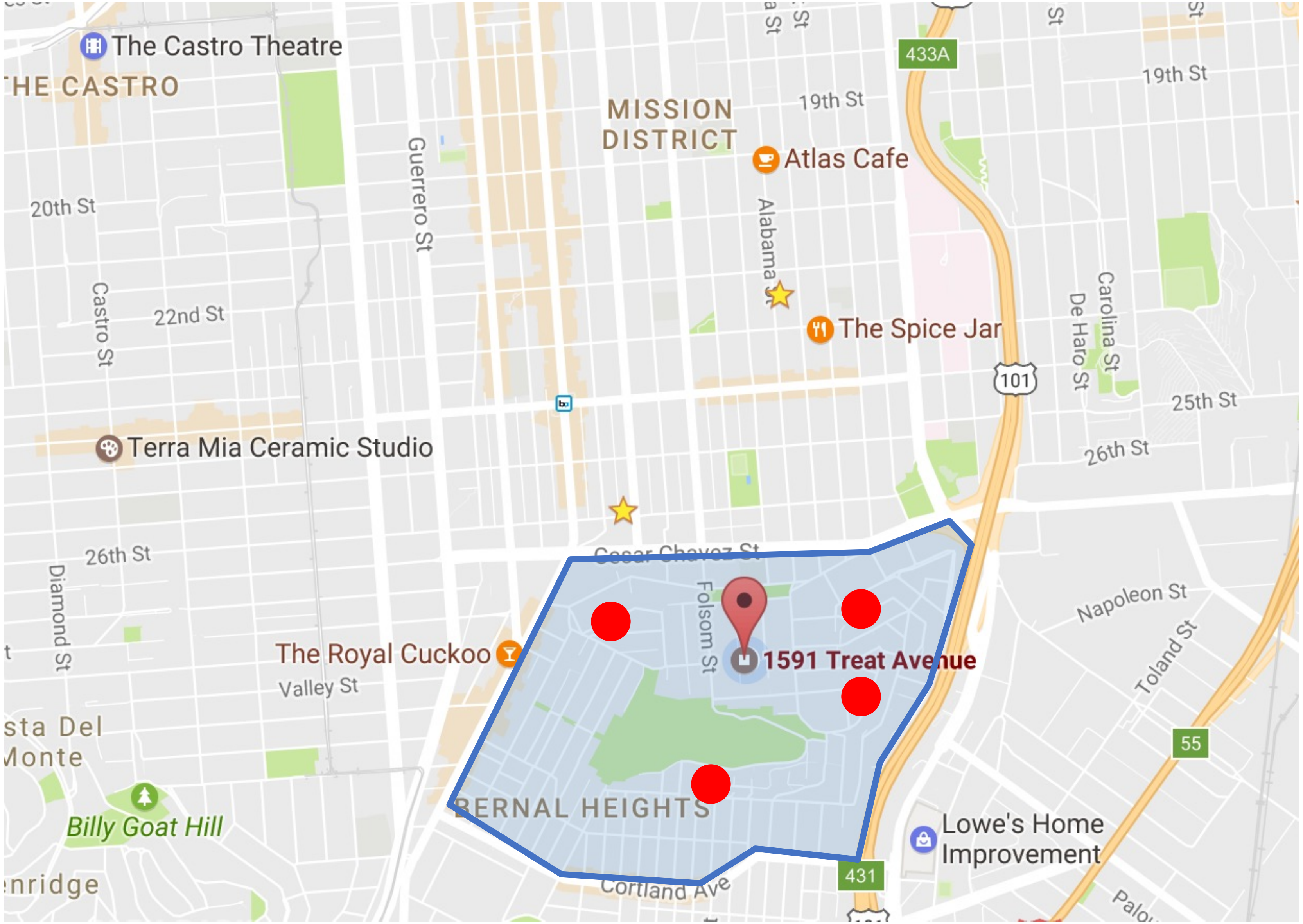
$$= \lim_{n\to\infty} \frac{n!}{(n-k)!k!} \cdot \frac{\lambda^k}{n^k} \cdot \frac{e^{-\lambda}}{1}$$ By definition of natural exp

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

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 $n \to \infty$ 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 ~ $\text{Poi}(\lambda)$

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

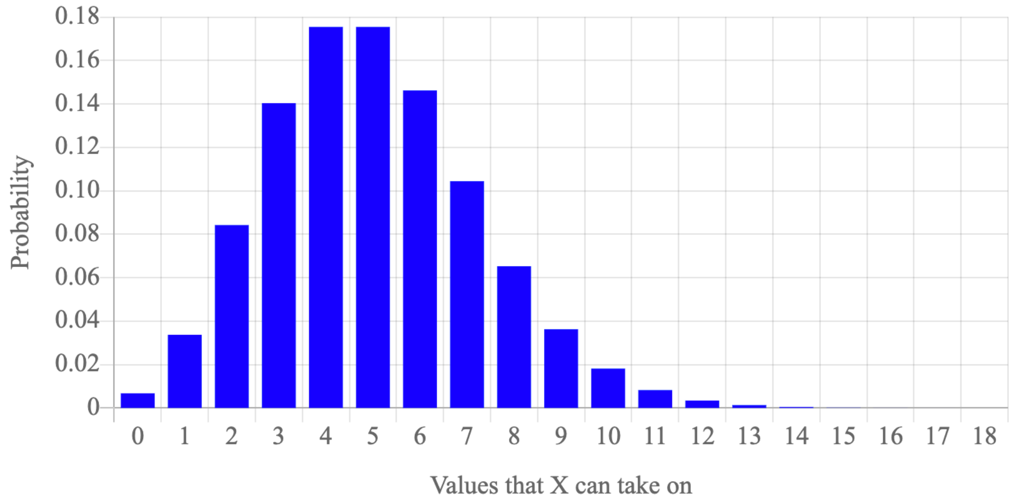
PMF equation: $\Pr(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$

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

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

PMF graph:

Parameter λ: 5

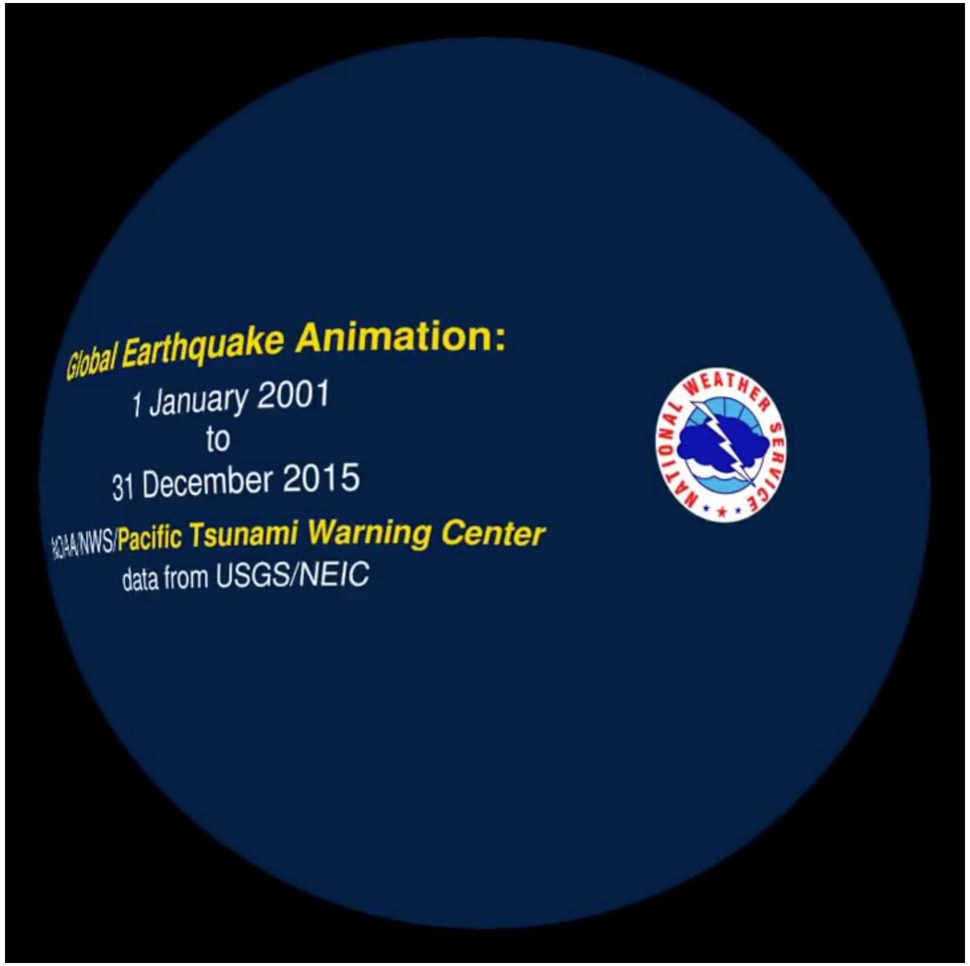


Poisson is great when you have a rate!

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

Earthquakes

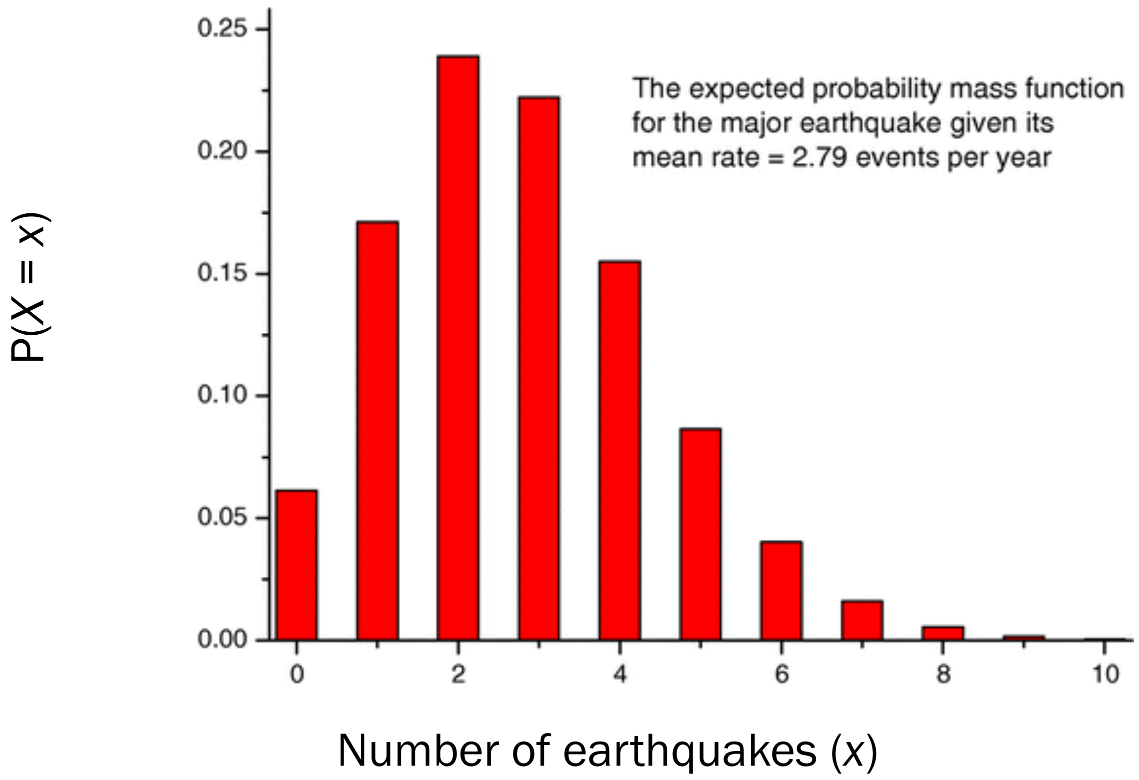


Average of 2.79 major earthquakes per year.
What is the probability of 3 major earthquakes next year?

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.

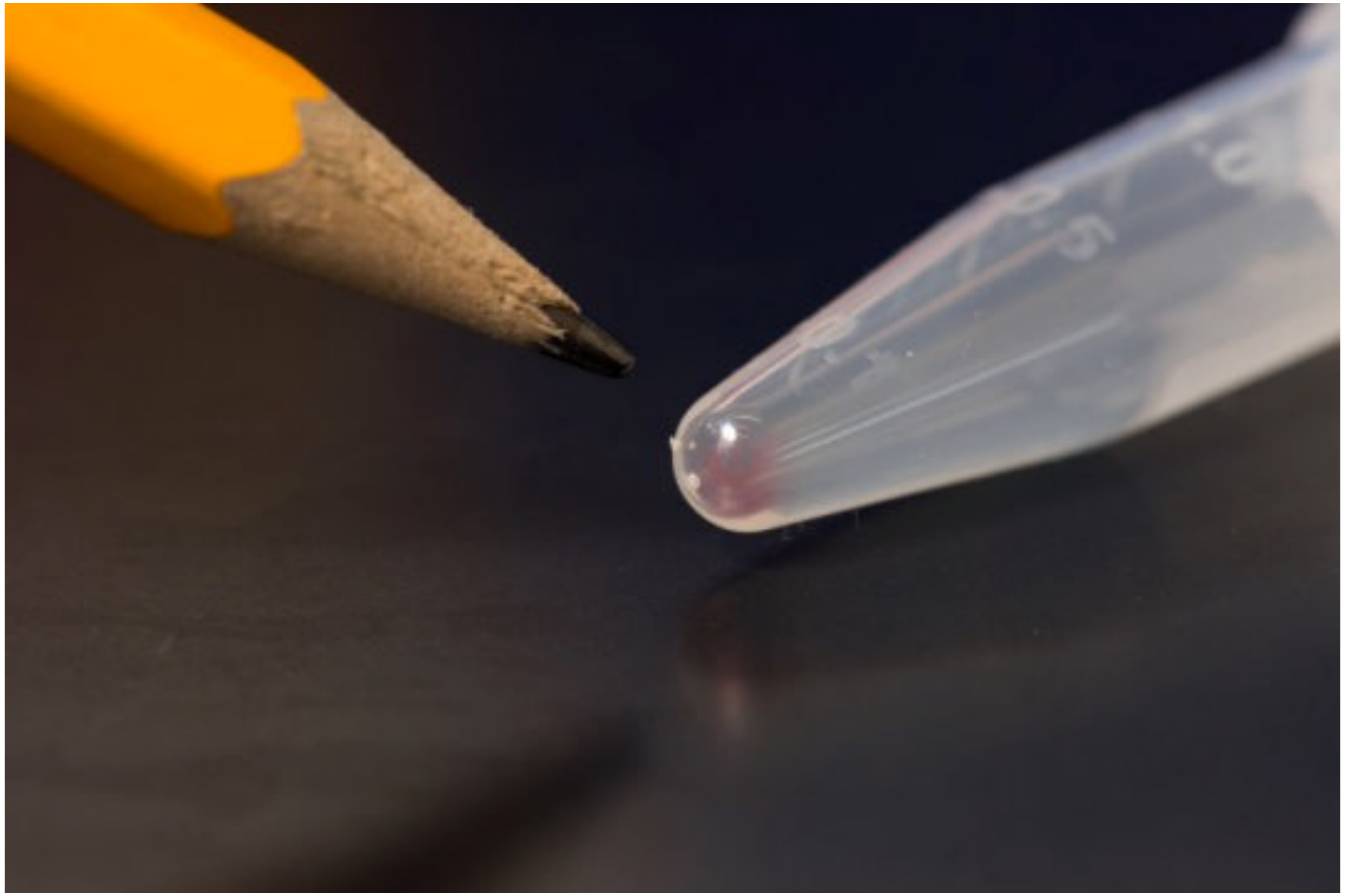
Poisson can approximate a Binomial!

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 steam 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"

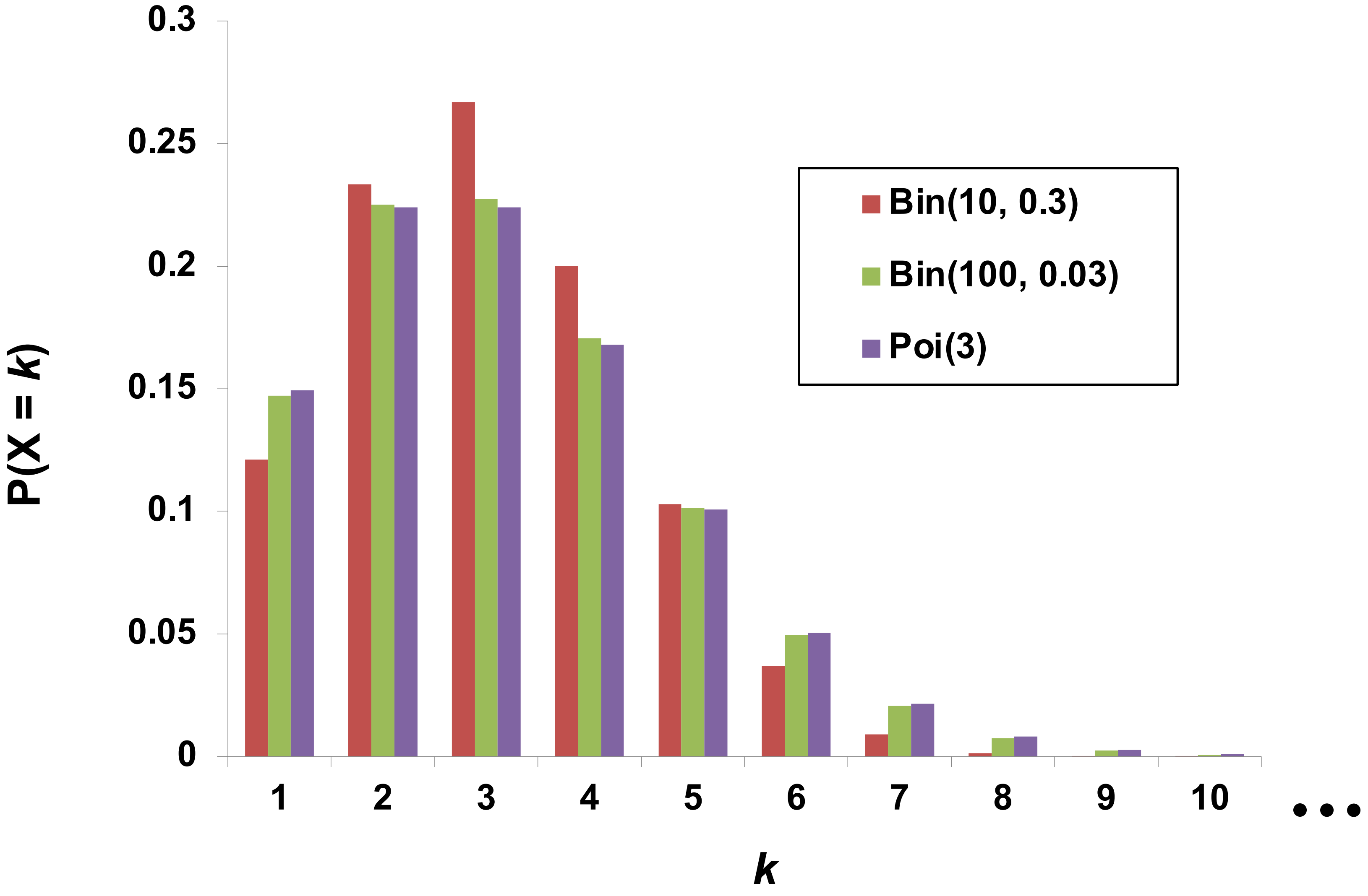
Different interpretations of "moderate"

- $n > 20$ 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)



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

Tender (Central) Moments with Poisson

Recall: Y ~ Bin(n, p)

- E[Y] = np
- Var(Y) = $np(1 - p)$

X ~ Poi(λ)   where $\lambda = np$ ($n \to \infty$ and $p \to 0$)

- E[X] = $np = \lambda$
- 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 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

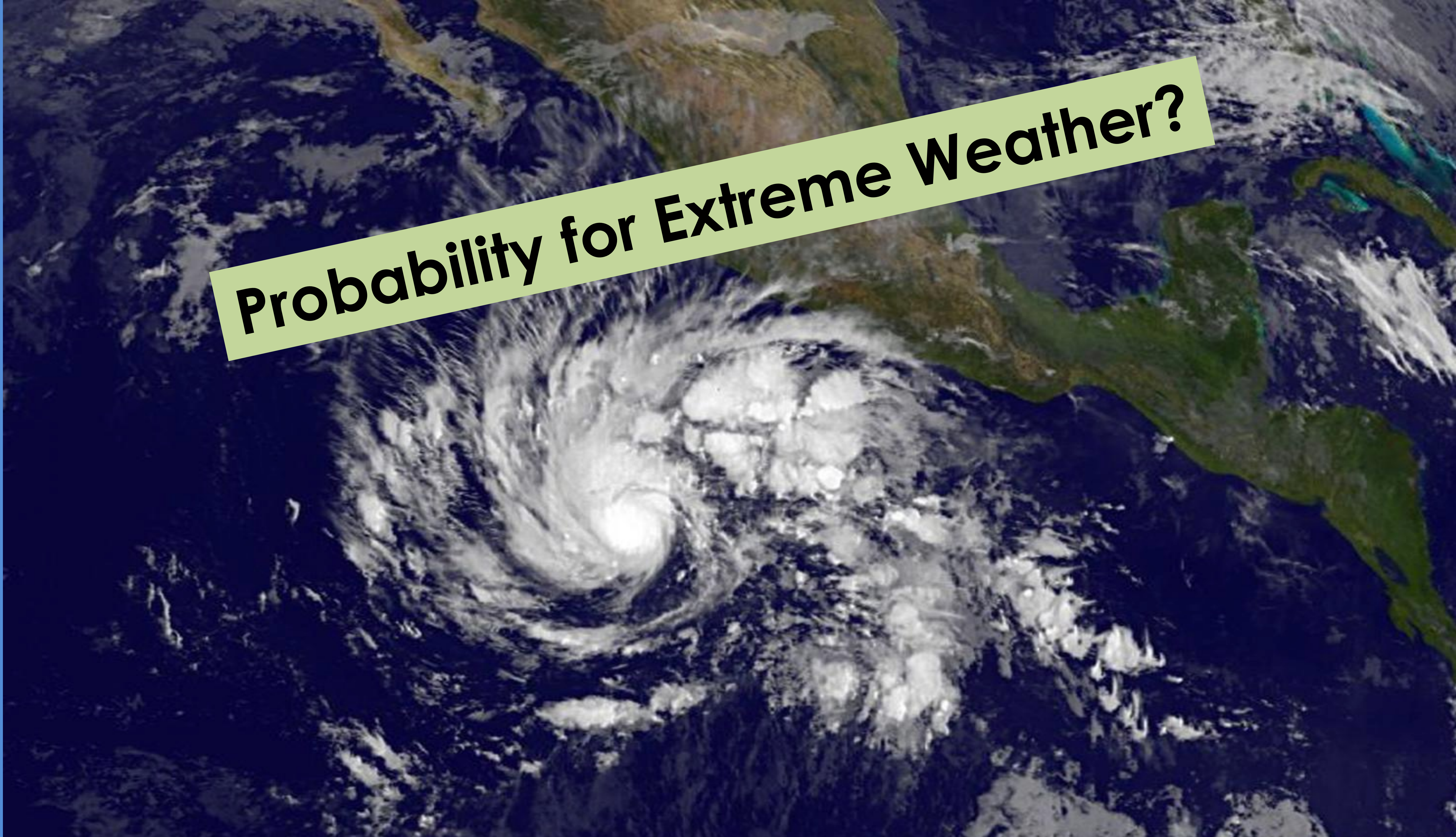
- In past, server load averages 2 hits/second
- X = # hits server receives in a second
- What is $P(X < 5)$?

Solution

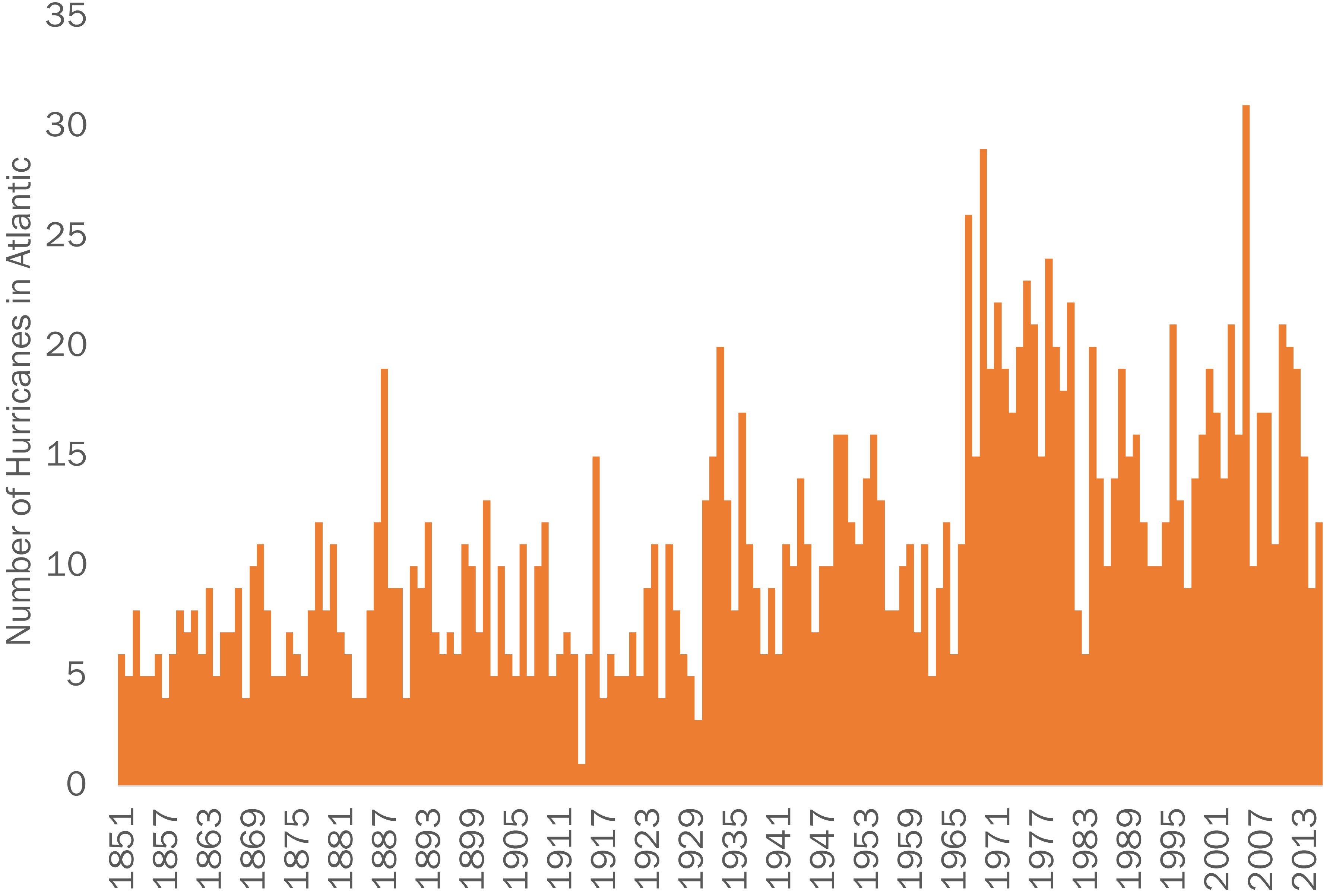
$$X \sim \text{Poi}(\lambda = 2)$$

$$
\begin{aligned}
P(X < 5) &= \sum_{i=0}^{4} P(X = i) \\
&= \sum_{i=0}^{4} e^{-\lambda} \frac{\lambda^i}{i!} && \text{Since } X \text{ is Poisson} \\
&= \sum_{i=0}^{4} e^{-2} \frac{2^i}{i!} \approx 0.95 && \text{Since } \lambda = 2
\end{aligned}
$$

Probability for Extreme Weather?

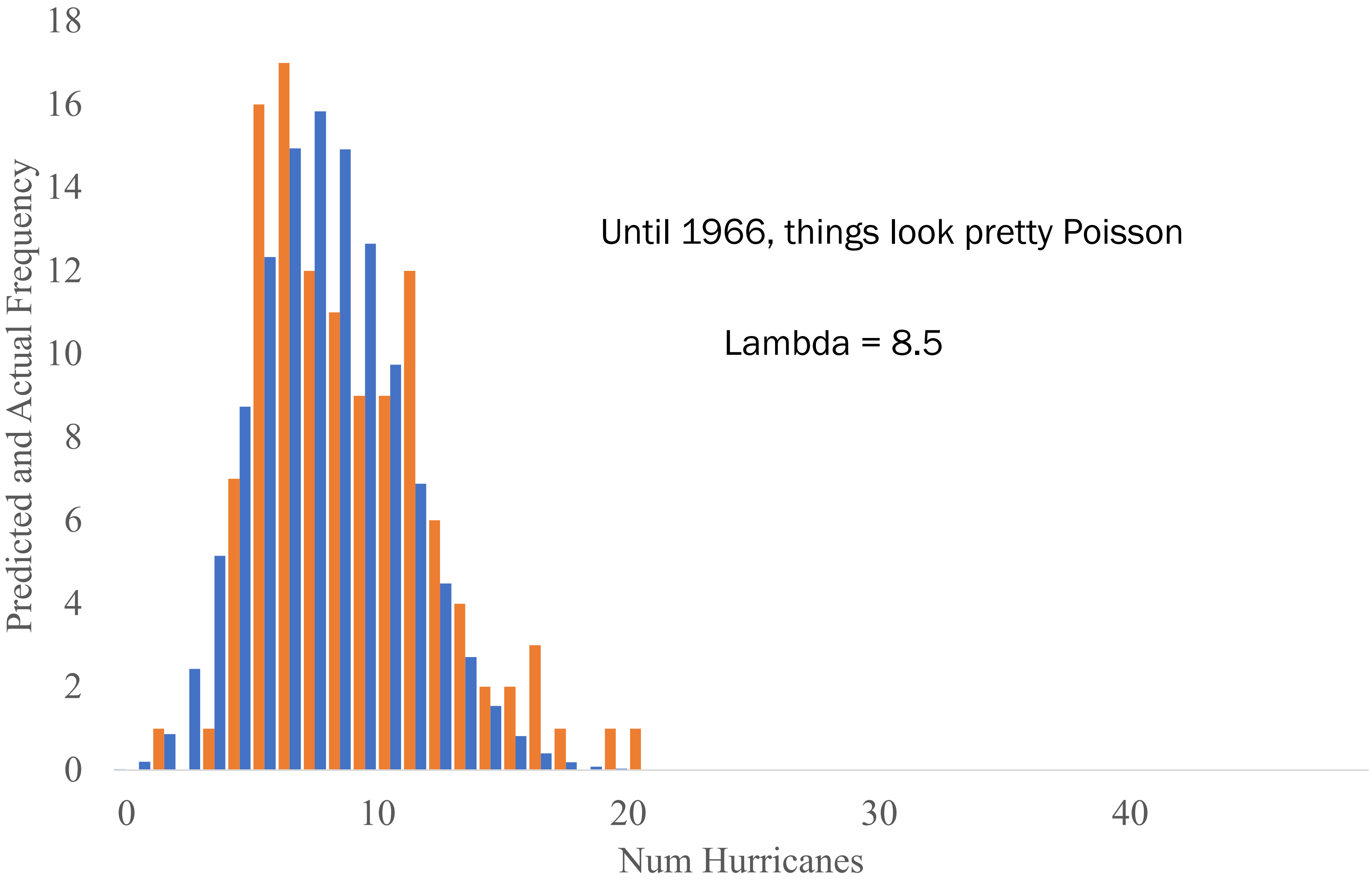


Hurricanes per Year since 1851



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)

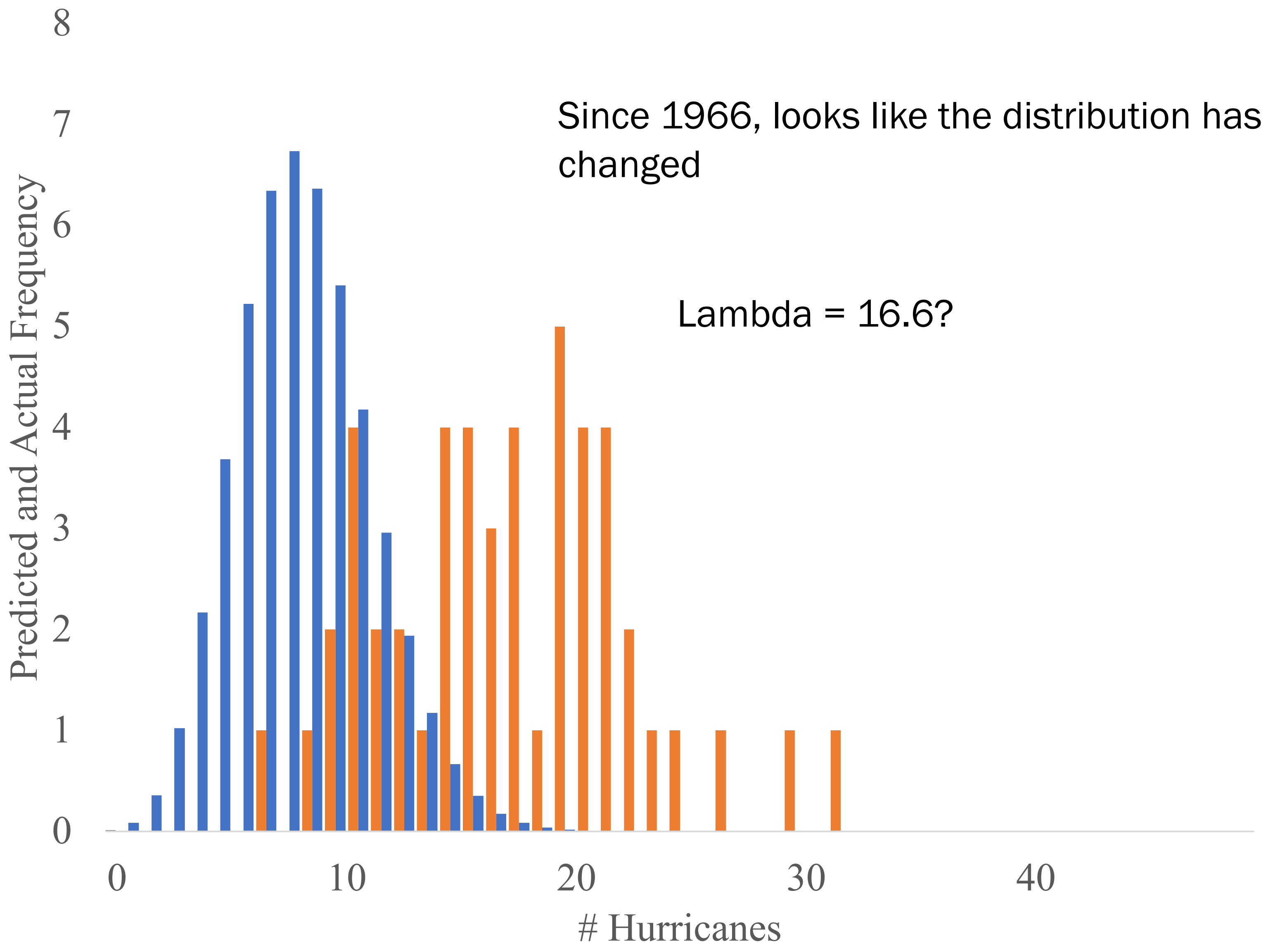
Solution:

$$P(X > 30) = 1 - P(X \leq 30)$$
$$= 1 - \sum_{i=0}^{30} P(X = i)$$
$$= 1 - 0.99999999997823$$
$$= 2.2e - 09$$

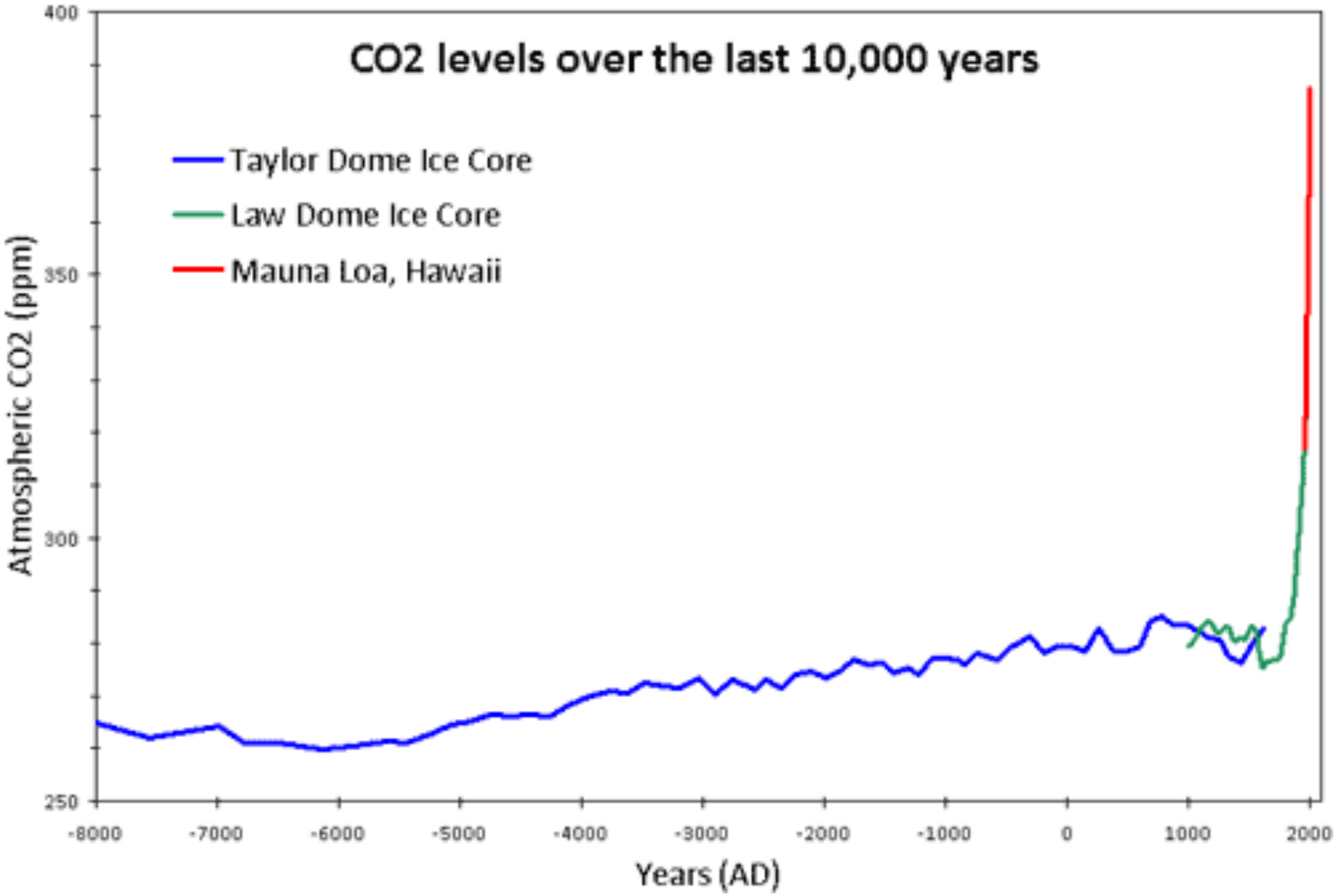
This is the pdf of a Poisson. Your favorite programming language has a function for it

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

The Distribution has Changed

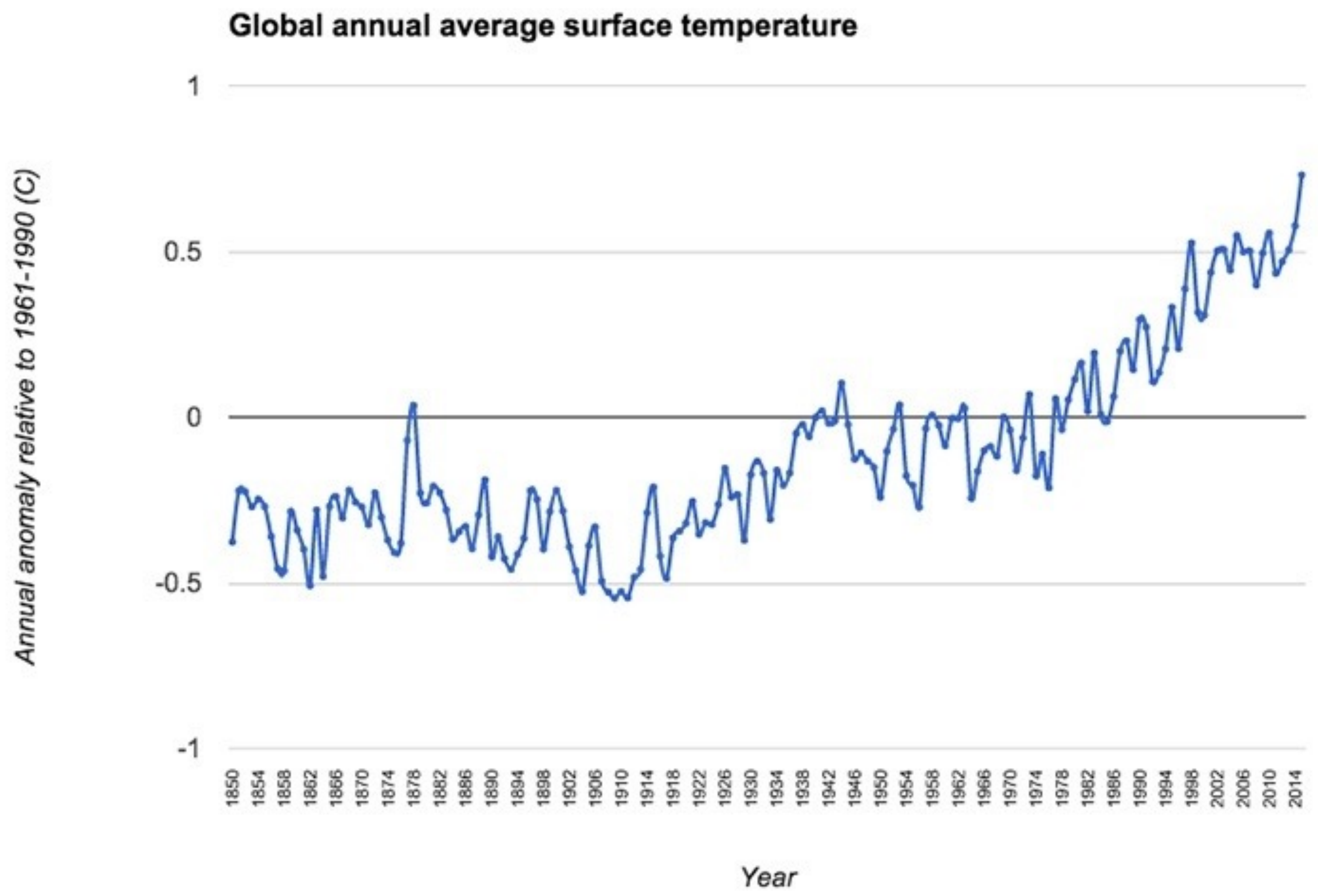


What's Up?



CO2 levels over the last 10,000 years
Taylor Dome Ice Core
Law Dome Ice Core
Mauna Loa, Hawaii
Atmospheric CO2 (ppm)
400
350
300
250
-8000
-7000
-6000
-5000
-4000
-3000
-2000
-1000
0
1000
2000
Years (AD)

What's Up?



Global annual average surface temperature
1
0.5
0
-0.5
-1
Annual anomaly relative to 1961-1990 (C)
1850
1854
1858
1862
1866
1870
1874
1878
1882
1886
1890
1894
1898
1902
1906
1910
1914
1918
1922
1926
1930
1934
1938
1942
1946
1950
1954
1958
1962
1966
1970
1974
1978
1982
1986
1990
1994
1998
2002
2006
2010
2014
Year

What's Up?

Python Scipy RV Methods

```
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.entropy()`	(Differential) entropy of X
`X.mean()`	E[X]
`X.var()`	Var(X)
`X,std()`	Std(X)

Next Time

Discrete Distributions

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

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

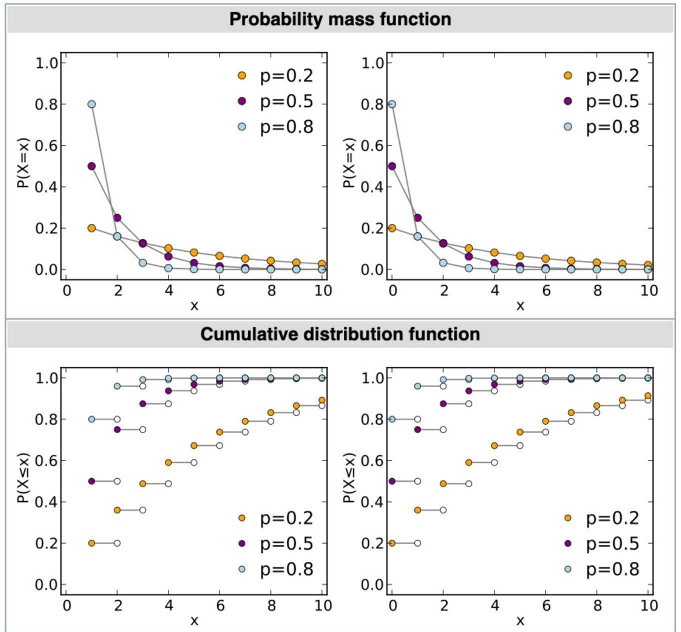
Poisson:
- # successes in *n* coin flips X ~ Poi(λ)

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)

Geometric



Parameters	$0 < p \leq 1$ success probability (real)	$0 < p \leq 1$ success probability (real)
Support	k trials where $k \in \{1, 2, 3, \dots\}$	k failures where $k \in \{0, 1, 2, 3, \dots\}$
PMF	$(1-p)^{k-1}p$	$(1-p)^k p$
CDF	$1-(1-p)^k$	$1-(1-p)^{k+1}$
Mean	$\frac{1}{p}$	$\frac{1-p}{p}$

The Poisson Common Path