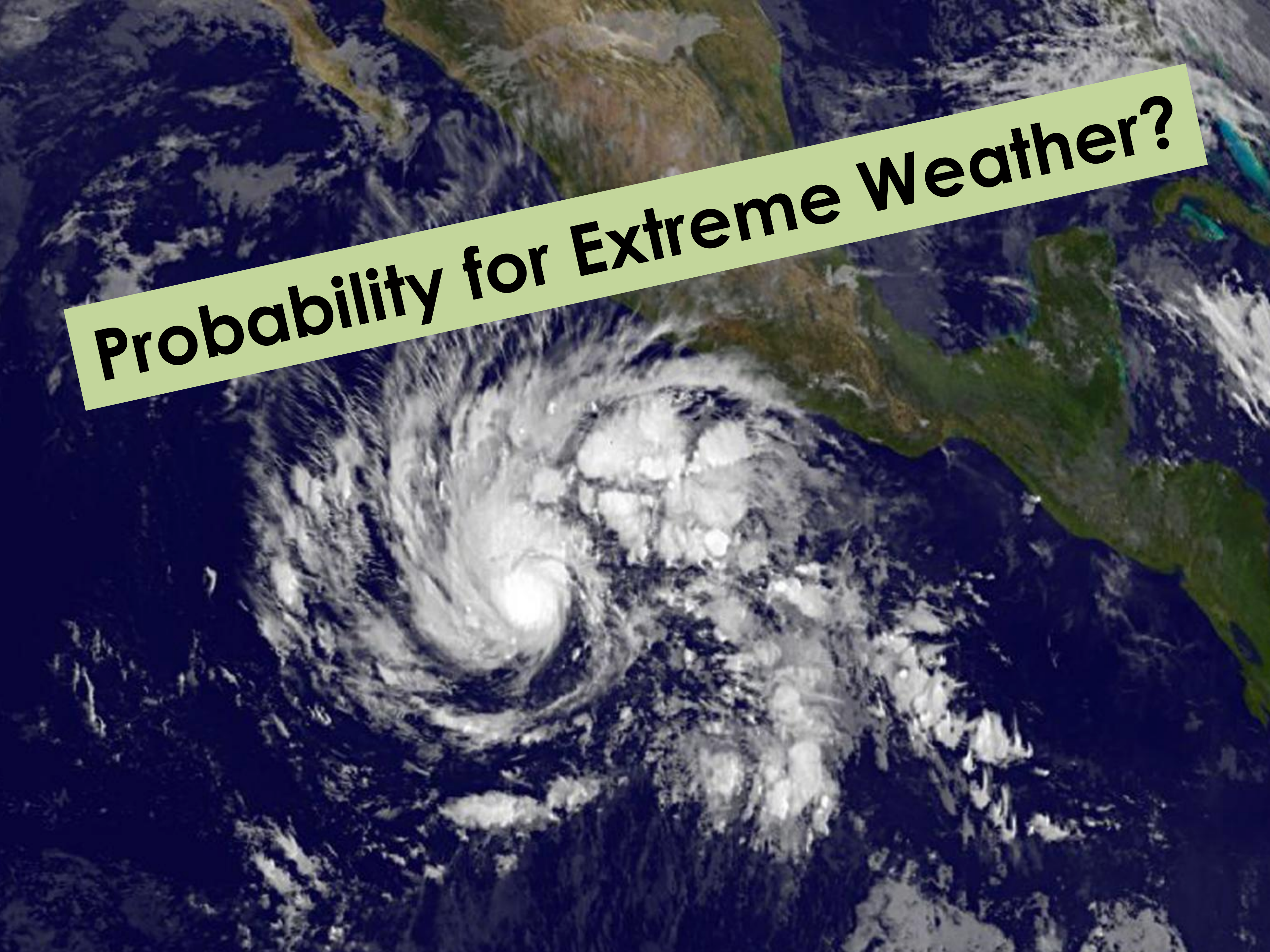
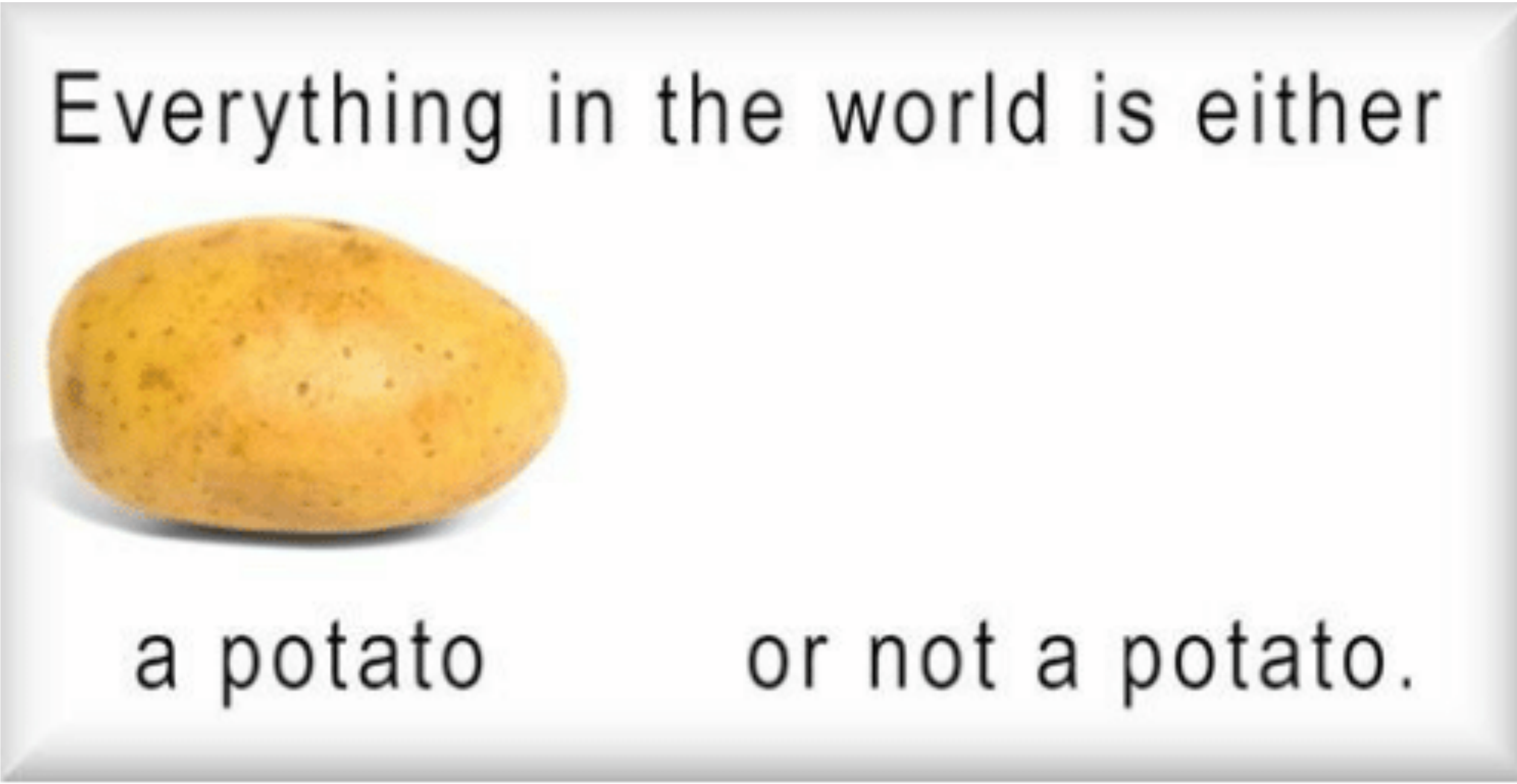
More Discrete Distributions

Chris Piech
CS109, Stanford University


Probability for Extreme Weather?

Review



$$P(X) + P(X^C) = 1$$

Inclusion Exclusion

$$P(A \cup B) = P(A) + P(B) - P(AB)$$

$$P(A \cup B \cup C \cup D) =$$

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

Our random variable

Num trials

Probability of success on each trial

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

Is distributed as a

Binomial

With these parameters

If X is a binomial with parameters n and p

$$P(X = k) = \overbrace{\binom{n}{k} p^k (1-p)^{n-k}}^{\text{Probability Mass Function for a Binomial}}$$

Probability that our variable takes on the value k

Bernoulli vs Binomial

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

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

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*)

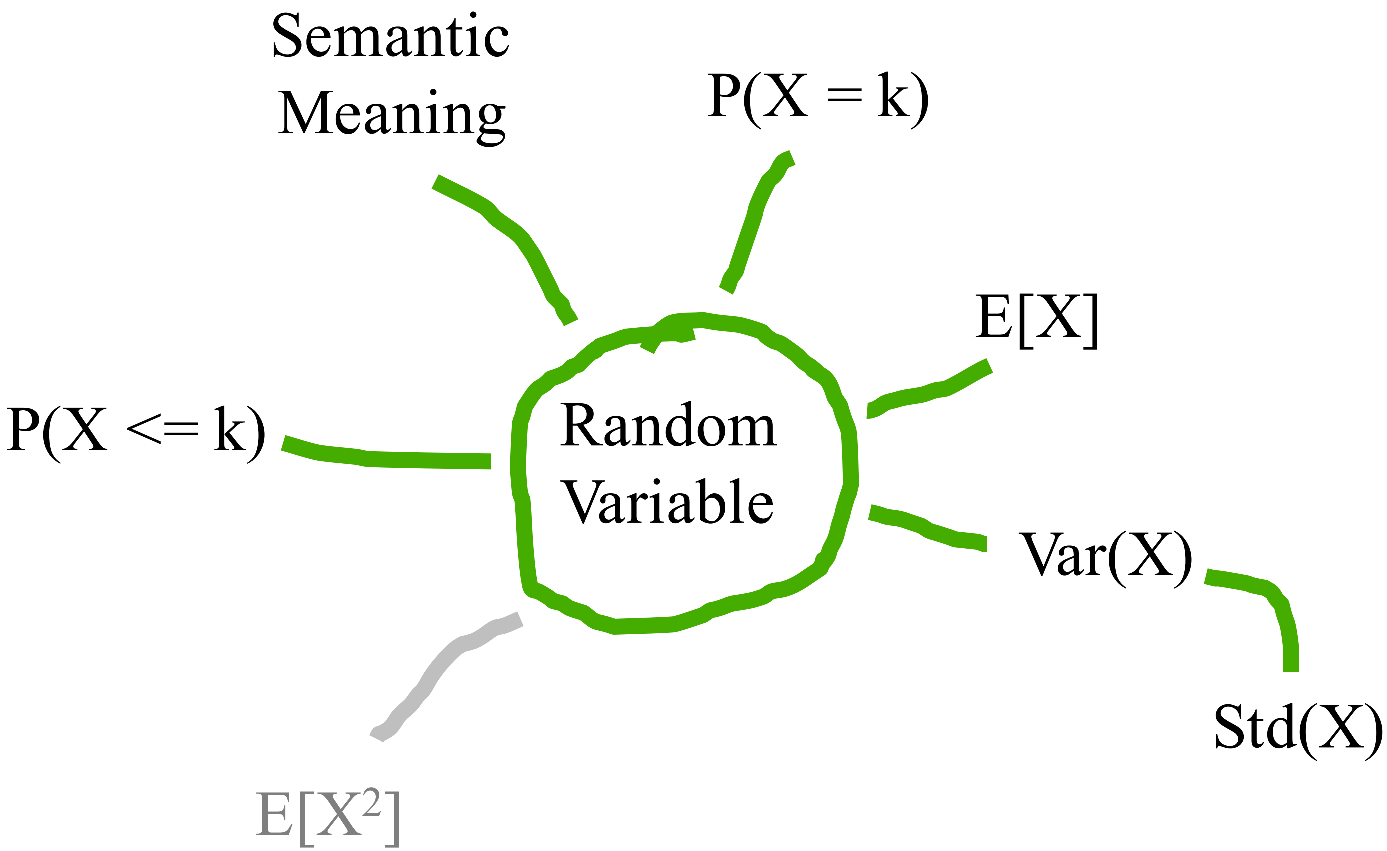
$$Y \sim \mathrm{Bin}(n, p)$$

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

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

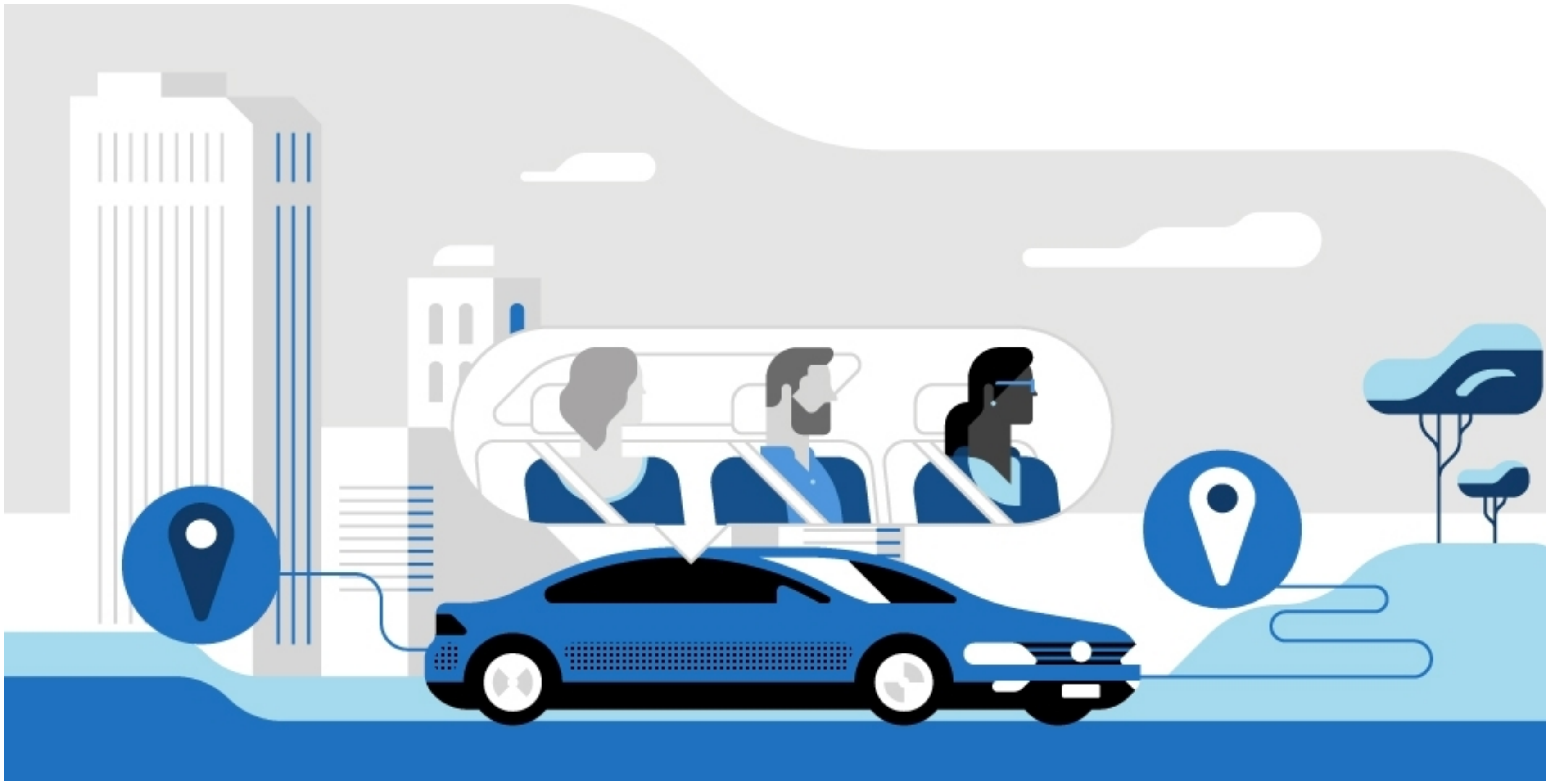
Binomial is the sum of *n* Bernoullis

Fundamental Properties



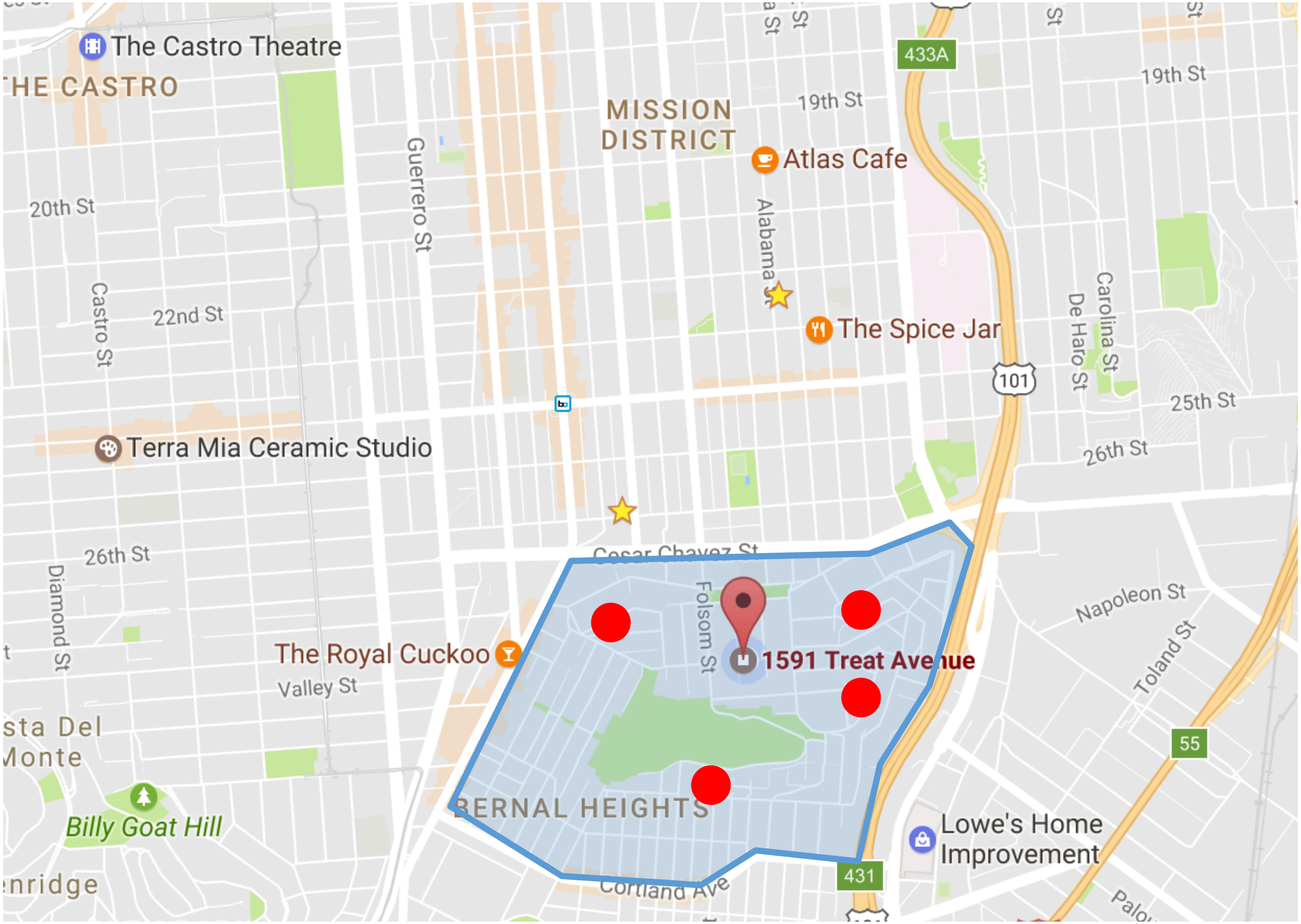
End Review

Algorithmic Ride Sharing



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

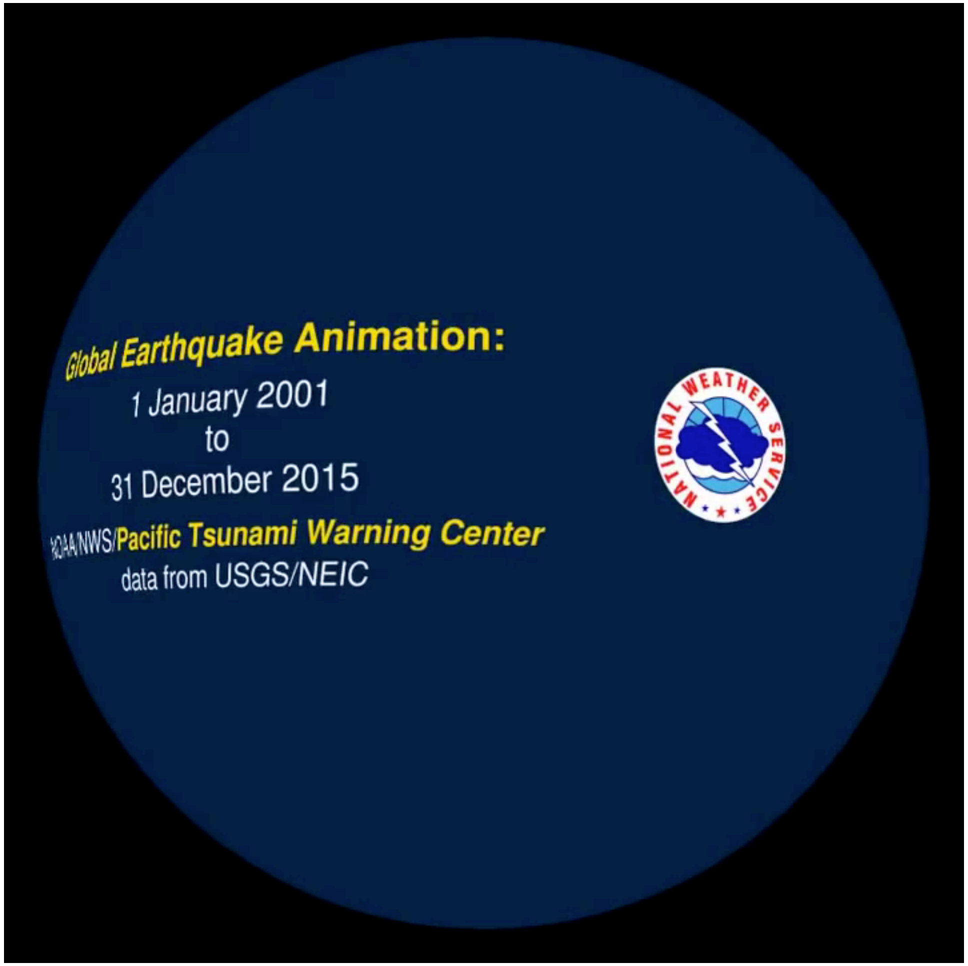
Probability of *k* requests from this area in the next 1 min



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

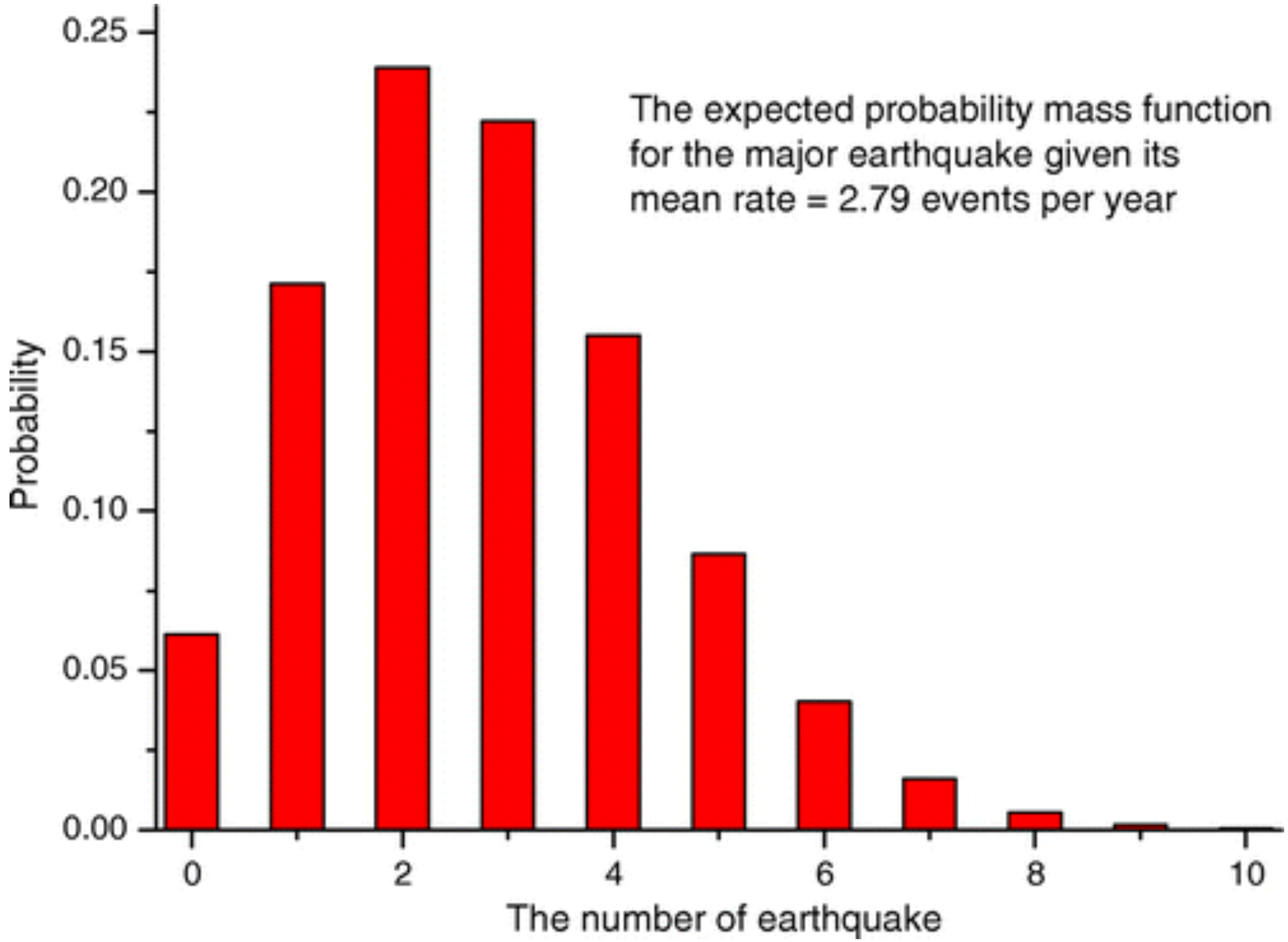
On average $\lambda = 5$ requests per minute

Earthquakes



Average of 2.79 major earthquakes per year.
What is the probability of more than 1 earthquake next year?

Earthquake Probability Mass Function



Binomial in the Limit

- Recall the Binomial distribution

$$P(X=i)=\frac{n!}{i!(n-i)!}p^i(1-p)^{n-i}$$

- Let $\lambda = np$ (equivalently: $p = \lambda/n$)

$$P(X=i)=\frac{n!}{i!(n-i)!}\left(\frac{\lambda}{n}\right)^i\left(1-\frac{\lambda}{n}\right)^{n-i} = \frac{n(n-1)\ldots(n-i+1)}{n^i}\frac{\lambda^i}{i!}\frac{(1-\lambda/n)^n}{(1-\lambda/n)^i}$$

- When n is large, p is small, and λ is "moderate":

$$\frac{n(n-1)\ldots(n-i+1)}{n^i}\approx 1 \qquad (1-\lambda/n)^n \approx e^{-\lambda} \qquad (1-\lambda/n)^i \approx 1$$

- Yielding:

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

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.

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

  - 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

- Consider "rare" 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 ~ $\text{Poi}(\lambda)$

Poisson is great when you have a rate!

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

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

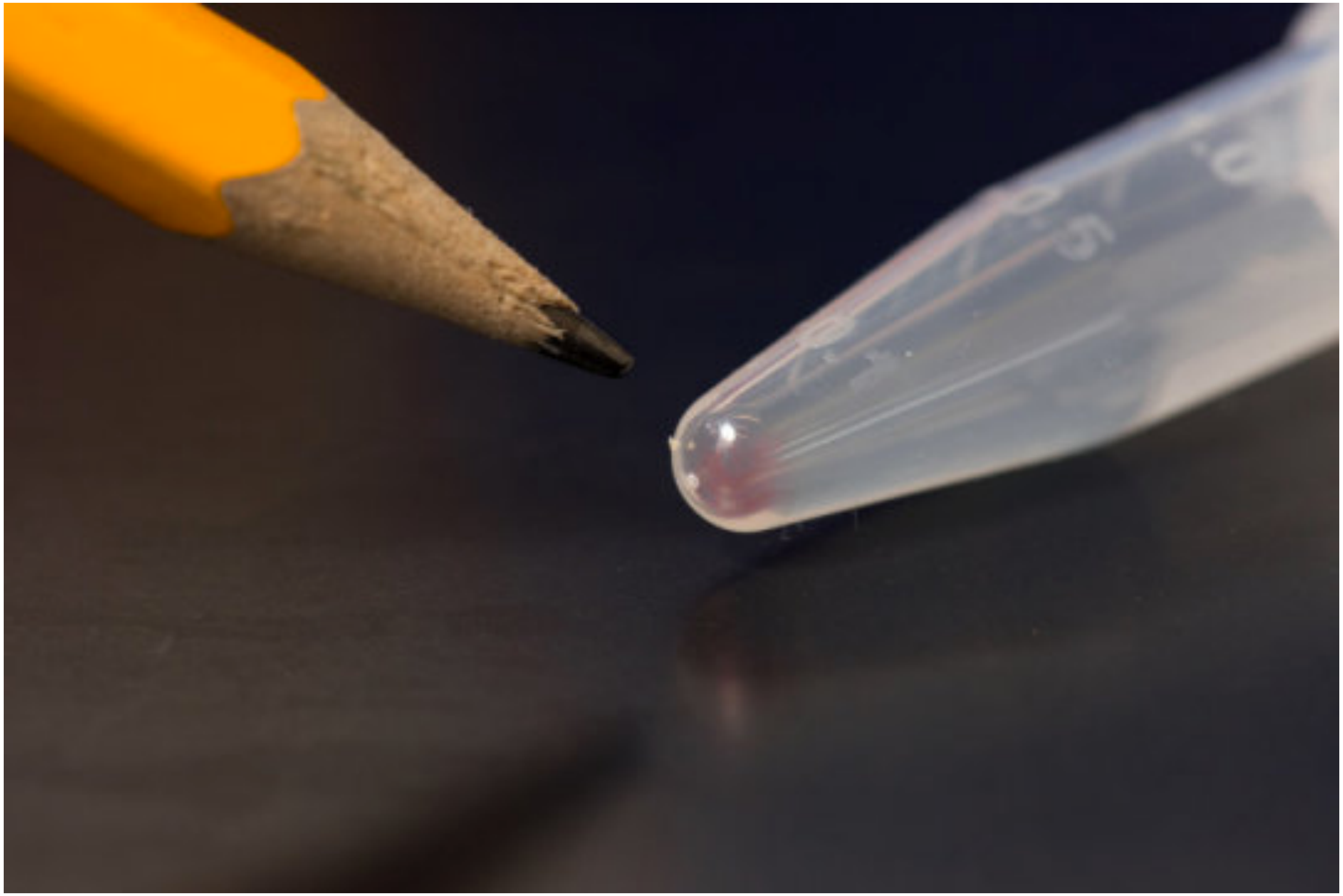
ABSTRACT

Yes.

Sending Bits

- We want to know the probability of 4 bits being corrupted?
  - Probability of each base pair being corrupted is 0.1
  - n = 4 bits
  - X = number of bits corrupted. X ~ Bin(4, 0.1)

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.

Probability of no corruptions

Storing Data on DNA

- Recall example of sending bit string over network
  - In DNA (and real networks) send 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
  - # visitors to a popular website
  - # of servers crashes in a day in giant data center
  - # Facebook login requests that go to particular server

Storing Data on DNA

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

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

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

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

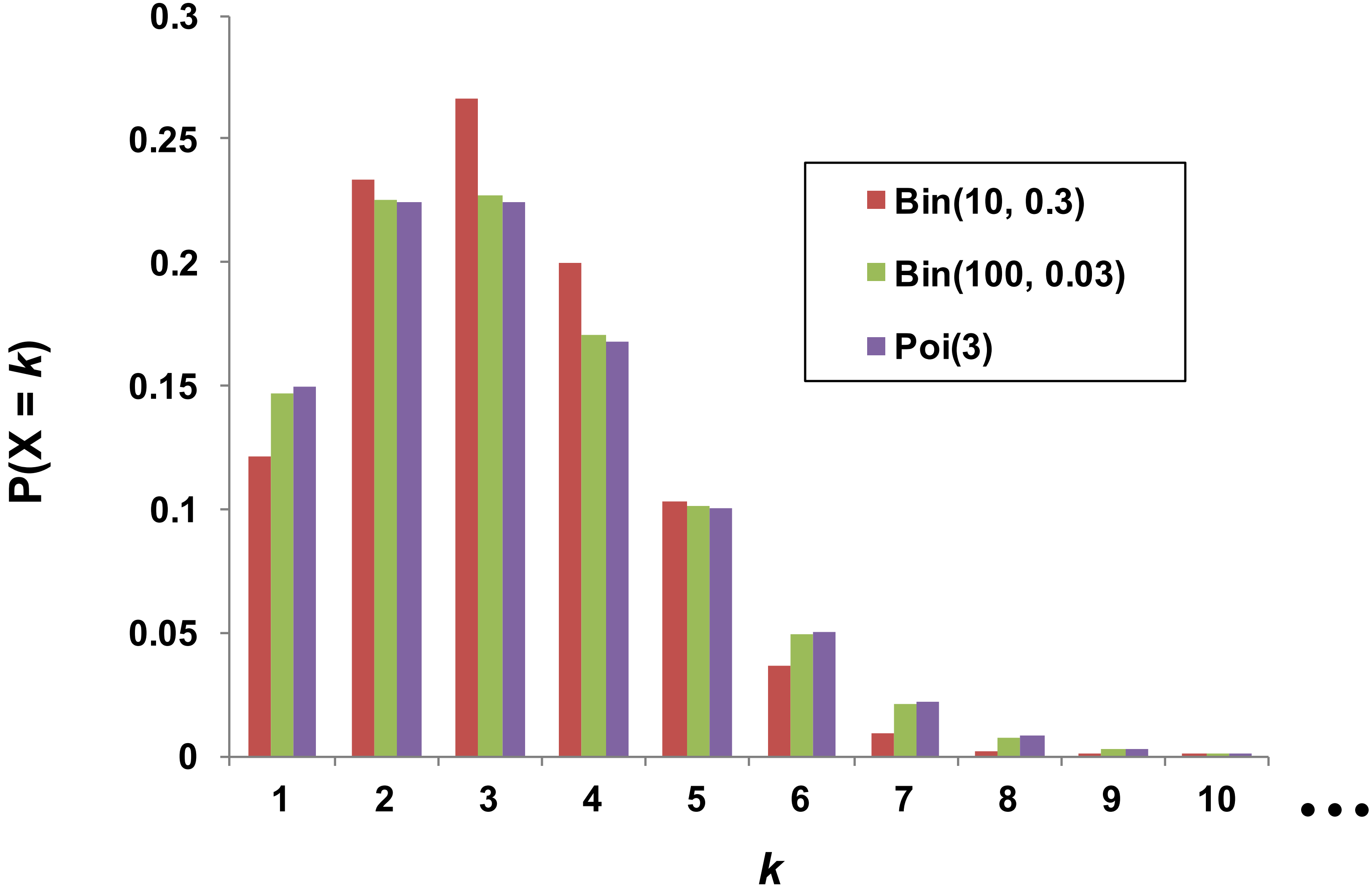
Poisson is Binomial in the Limit

- 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 $\lambda = np$ ($n \rightarrow \infty$ and $p \rightarrow 0$)
  - 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.

It's Really All About Raisin Cake

- Bake a cake using *many* raisins and *lots* of batter
- Cake is enormous (in fact, infinitely so…)
  - Cut slices of "moderate" size (w.r.t. # raisins/slice)
  - Probability p that a particular raisin is in a certain slice is very small (p = 1/# cake slices)
- Let X = number of raisins in a certain cake slice
- X ~ Poi(λ), where $\lambda = \dfrac{\text{total \# raisins}}{\text{\# cake slices}}$

CS = Baking Raisin Cake with Code

- Hash tables
  - strings = raisins
  - buckets = cake slices
- Server crashes in data center
  - servers = raisins
  - list of crashed machines = particular slice of cake
- Facebook login requests (i.e., web server requests)
  - requests = raisins
  - server receiving request = cake slice

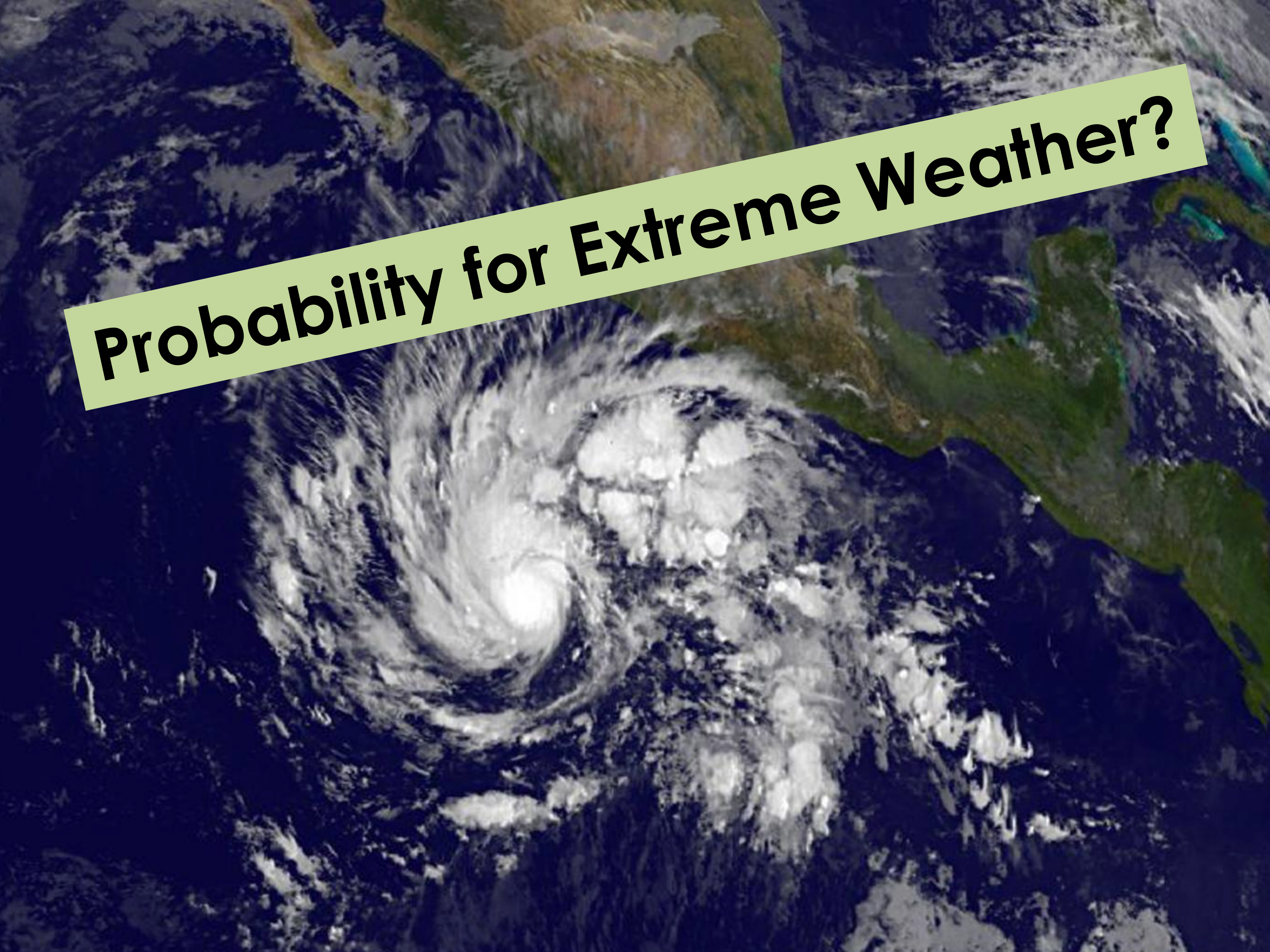
Poisson is Chill

- 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

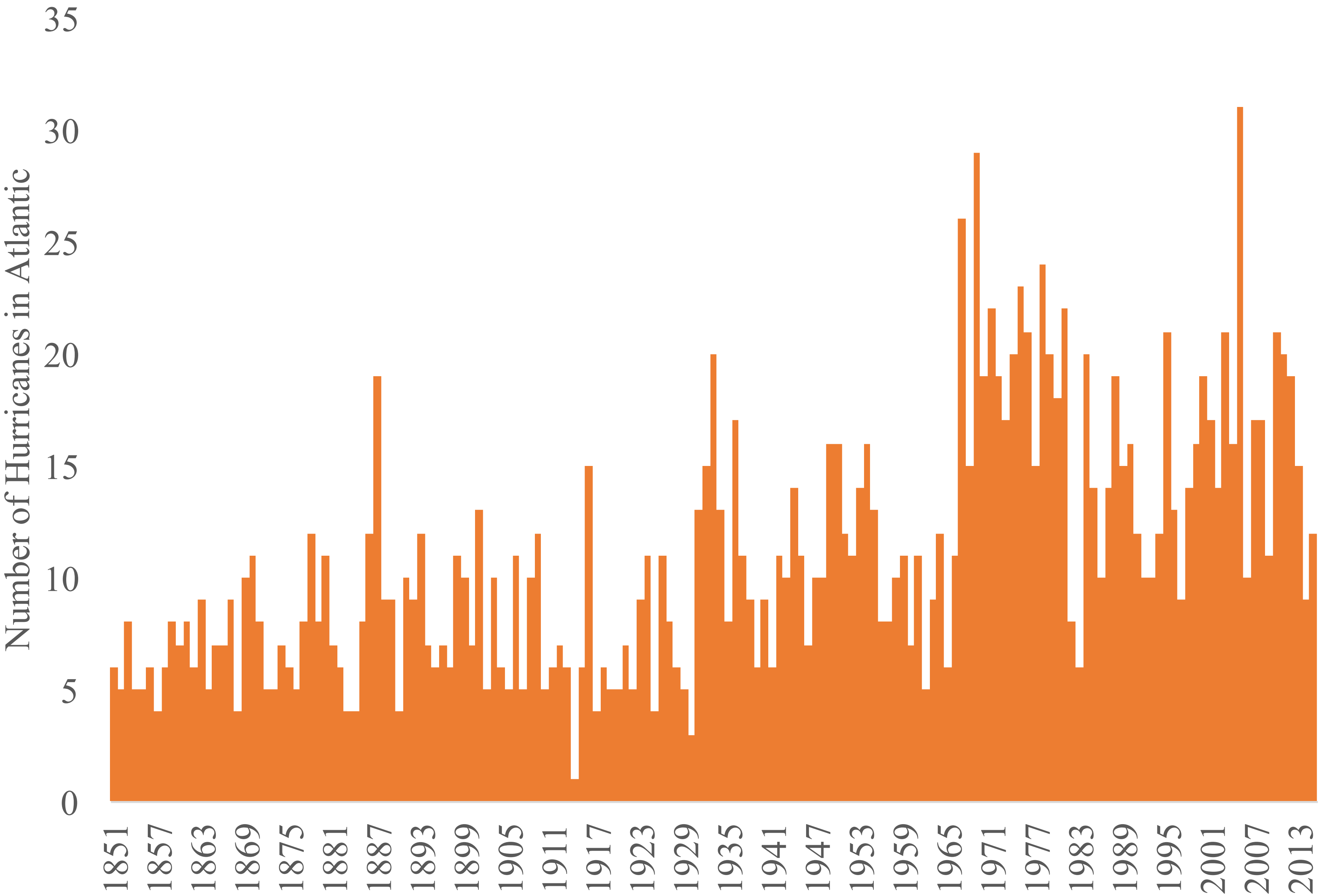
Web Server Load

- 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)?
- Model
  - Assume server cannot acknowledge > 1 hit/msec.
  - 1 sec = 1000 msec. (= large *n*)
  - P(hit server in 1 msec) = 2/1000 (= small *p*)
  - X ~ Poi(λ = 2)

$$P(X = 5) = e^{-2} \frac{2^5}{5!} \approx 0.0361$$

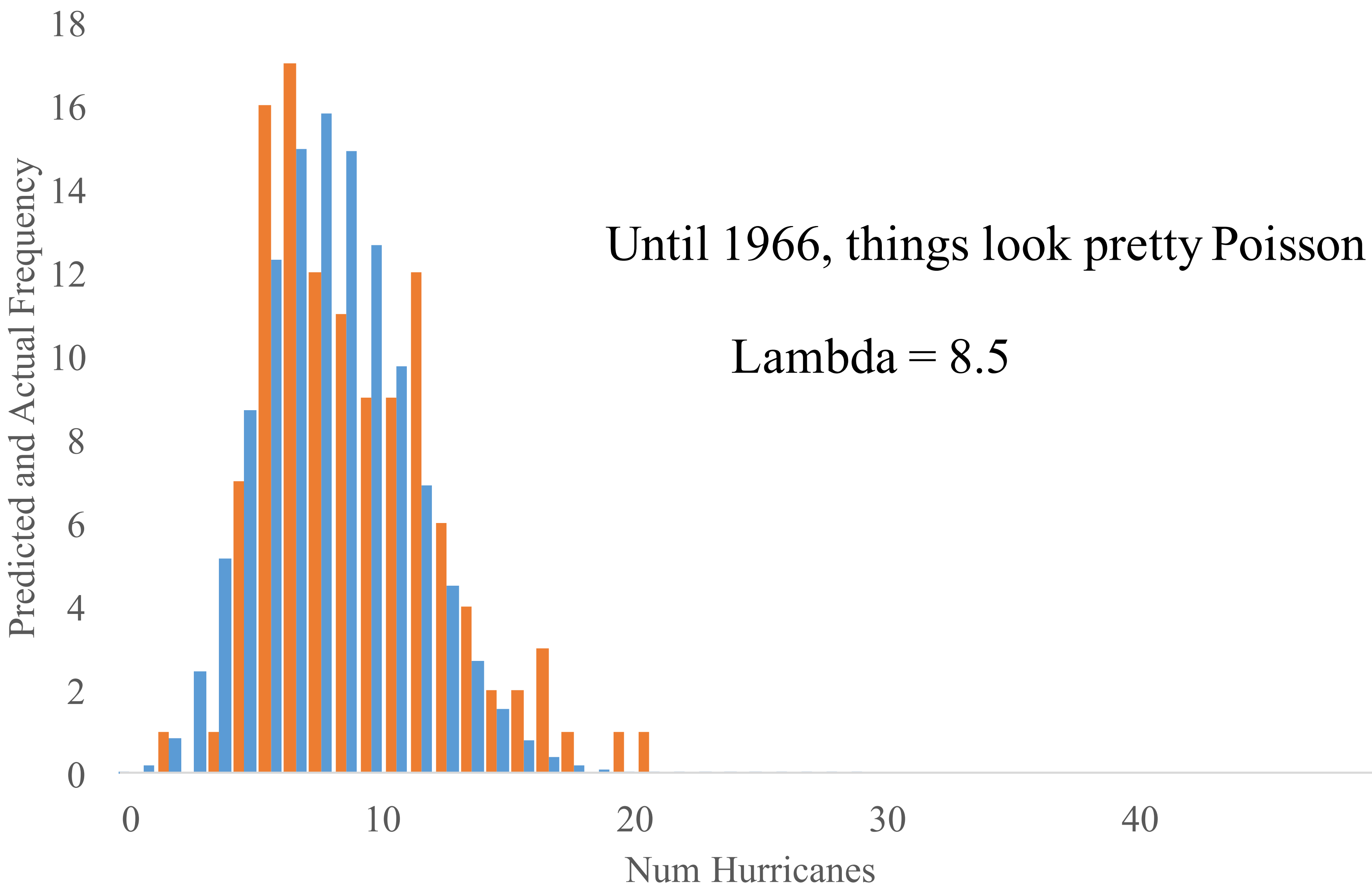

Probability for Extreme Weather?

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 ~ 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 pdf of a Poisson. Your favorite programming language has a function for it

Twice since 1966 there have been years with over 30 hurricanes

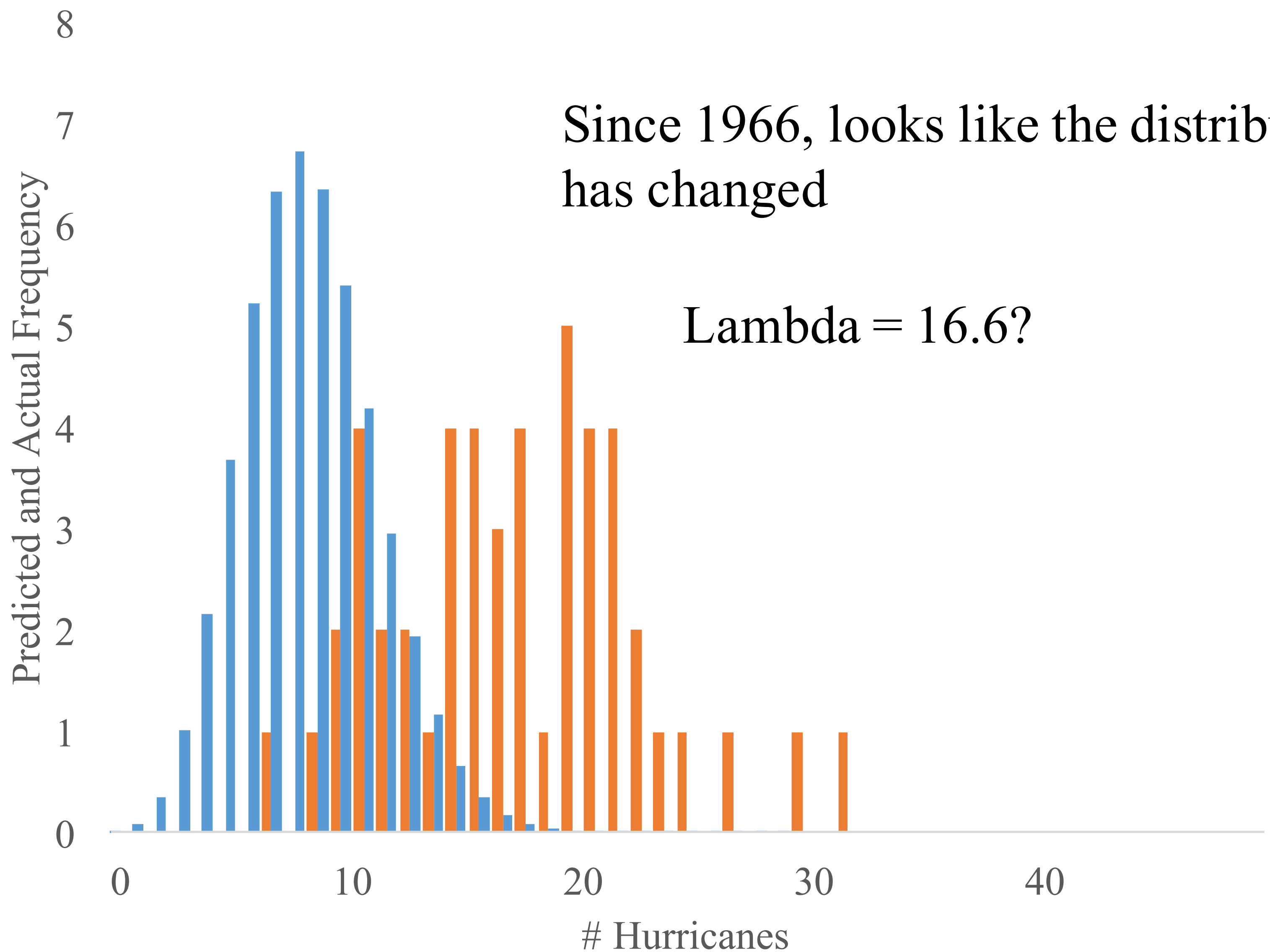
Improbability Drive

- 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 ~ Poi(8.5)
- Solution:

$$
\begin{aligned}
P(X > 30) &= 1 - P(X \leq 30) \\
&= 1 - \sum_{i=0}^{30} P(X = i) \\
&= 1 - 0.999999997823 \\
&= 2.2e - 09
\end{aligned}
$$

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

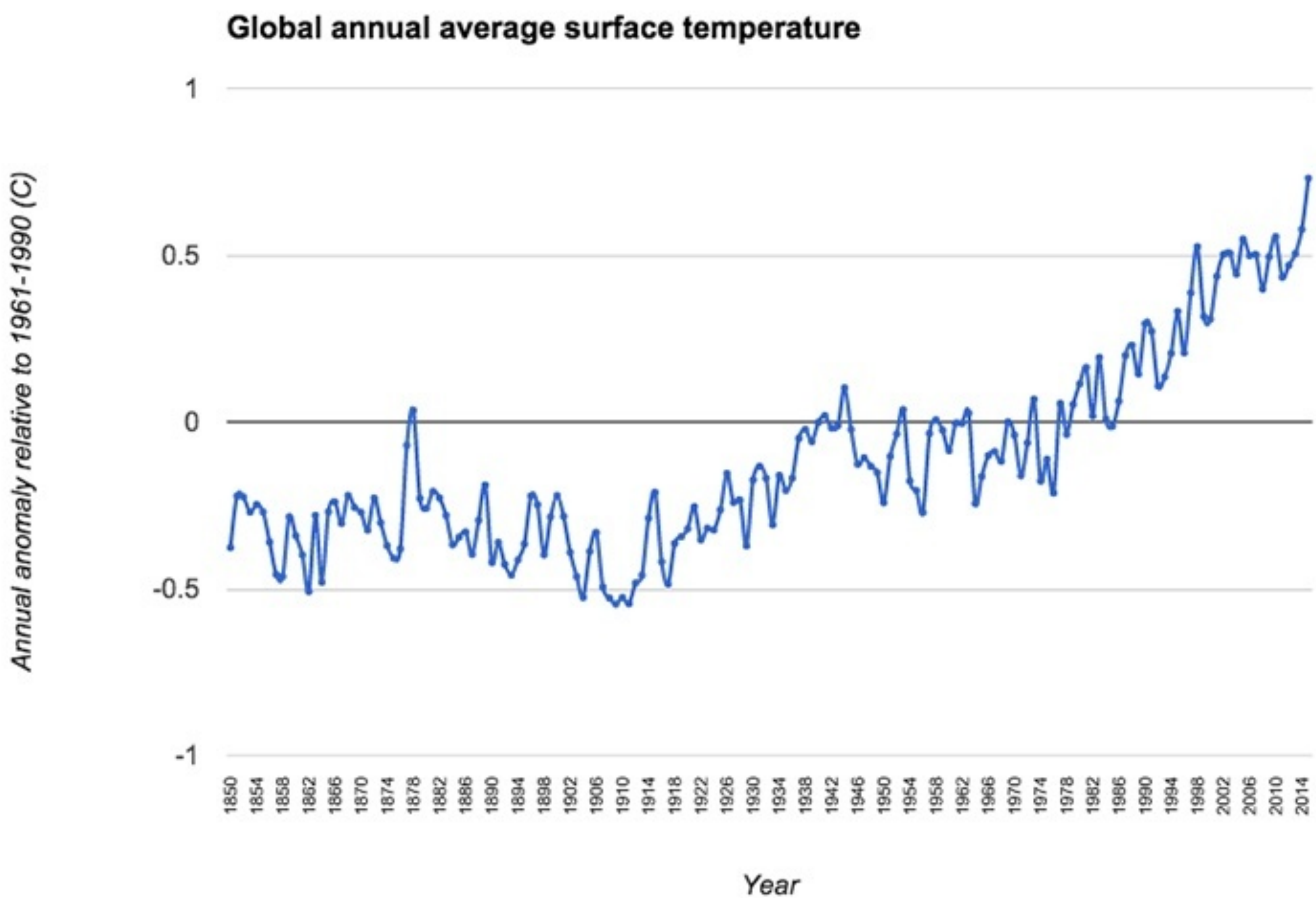
The Distribution has Changed



Since 1966, looks like the distribution has changed

Lambda = 16.6?

What's Up?



Global annual average surface temperature
Annual anomaly relative to 1961-1990 (C)
1
0.5
0
-0.5
-1
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

Python Scipy Poisson Methods

Function	Description
pmf(k)	Probability mass function.
cdf(k)	Cumulative distribution function.
entropy()	(Differential) entropy of the RV.
mean()	Mean of the distribution.
var()	Variance of the distribution.
std()	Standard deviation of the distribution.

Pause

Discrete Distributions

- Don’t have to memorize all of the following distributions.
- We want you to get a sense of how random variables work

Geometric Random Variable

- X is **<u>Geometric</u>** Random Variable: X ~ Geo(*p*)
  - X is number of independent trials until first success
  - *p* is probability of success on each trial
  - X takes on values 1, 2, 3, …, with probability:

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

  - E[X] = 1/*p*   Var(X) = (1 – *p*)/*p*^{2}
- Examples:
  - Flipping a coin (P(heads) = *p*) until first heads appears
  - Urn with N black and M white balls. Draw balls (with replacement, *p* = N/(N + M)) until draw first black ball
  - Generate bits with P(bit = 1) = *p* until first 1 generated

Negative Binomial Random Variable

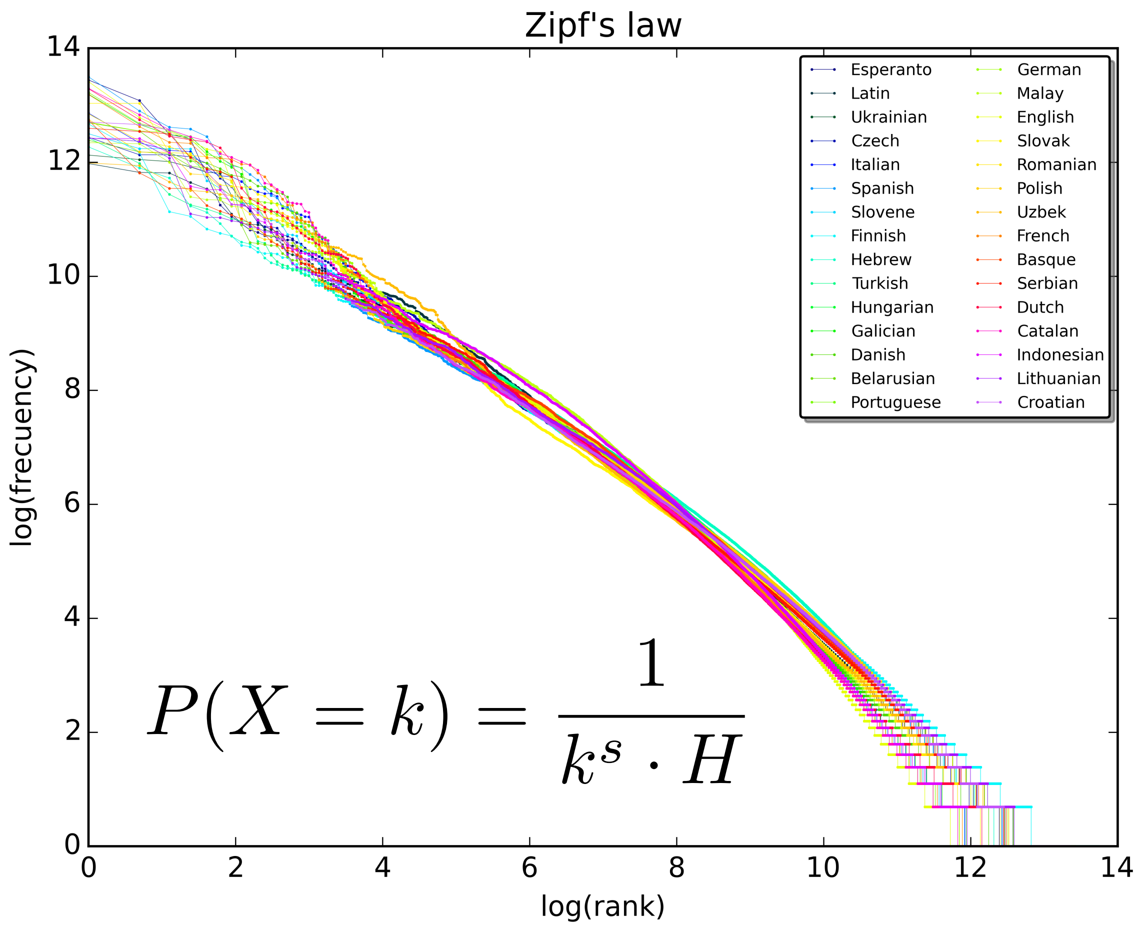
- X is **Negative Binomial** RV: X ~ NegBin(r, p)
  - X is number of independent trials until r successes
  - p is probability of success on each trial
  - X takes on values r, r + 1, r + 2…, with probability:

$$P(X = n) = \binom{n-1}{r-1} p^r (1-p)^{n-r}, \text{ where } n = r, r+1, \ldots$$

  - E[X] = r/p Var(X) = $r(1 - p)/p^2$
- Note: Geo(p) ~ NegBin(1, p)
- Examples:
  - # of coin flips until r-th "heads" appears
  - # of strings to hash into table until bucket 1 has r entries

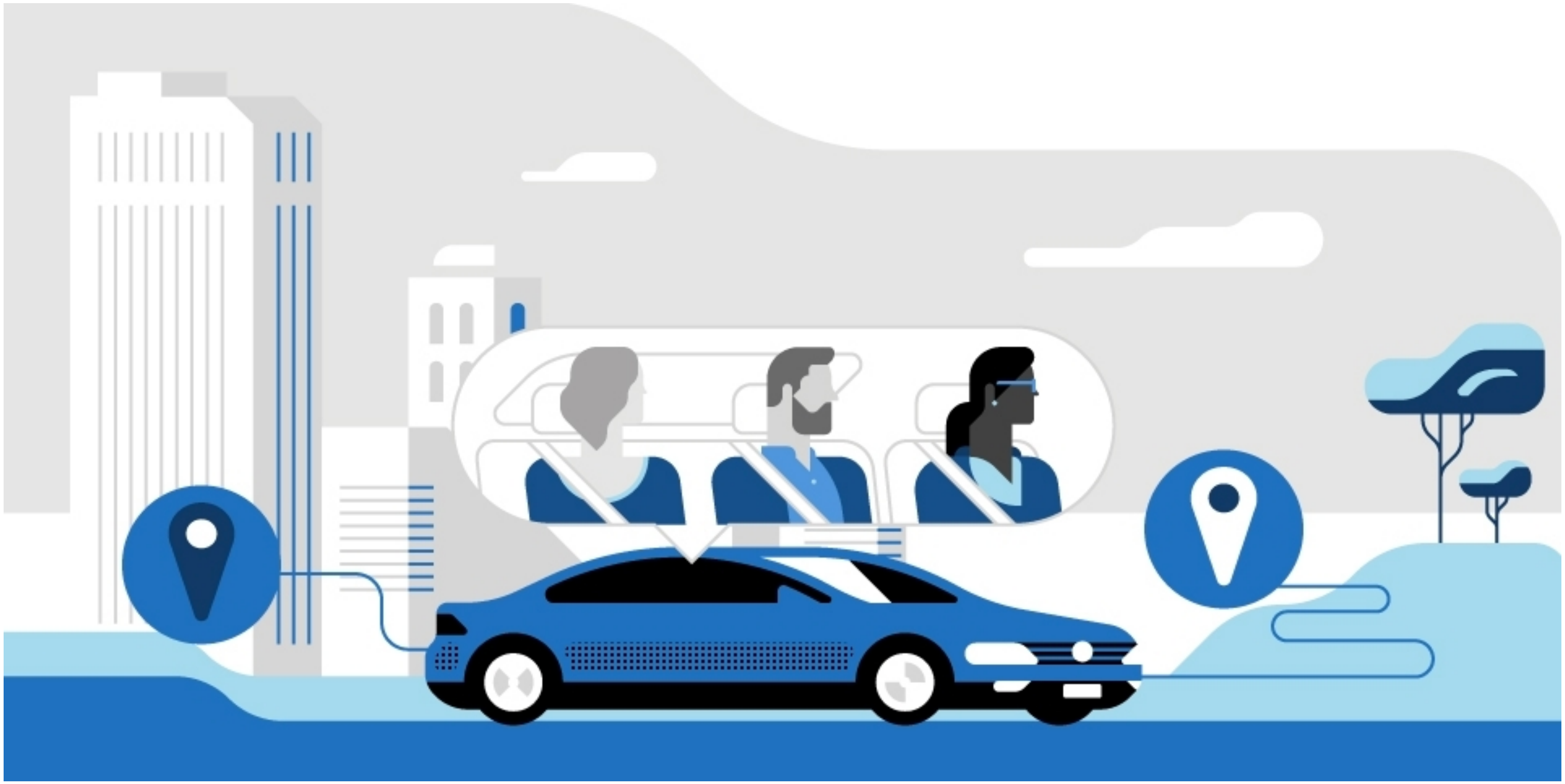
Zipf Random Variable

- X is **Zipf** RV: X ~ Zipf(*s*)
  - X is the rank index of a chosen word



Recap

Algorithmic Ride Sharing



The Poisson Common Path