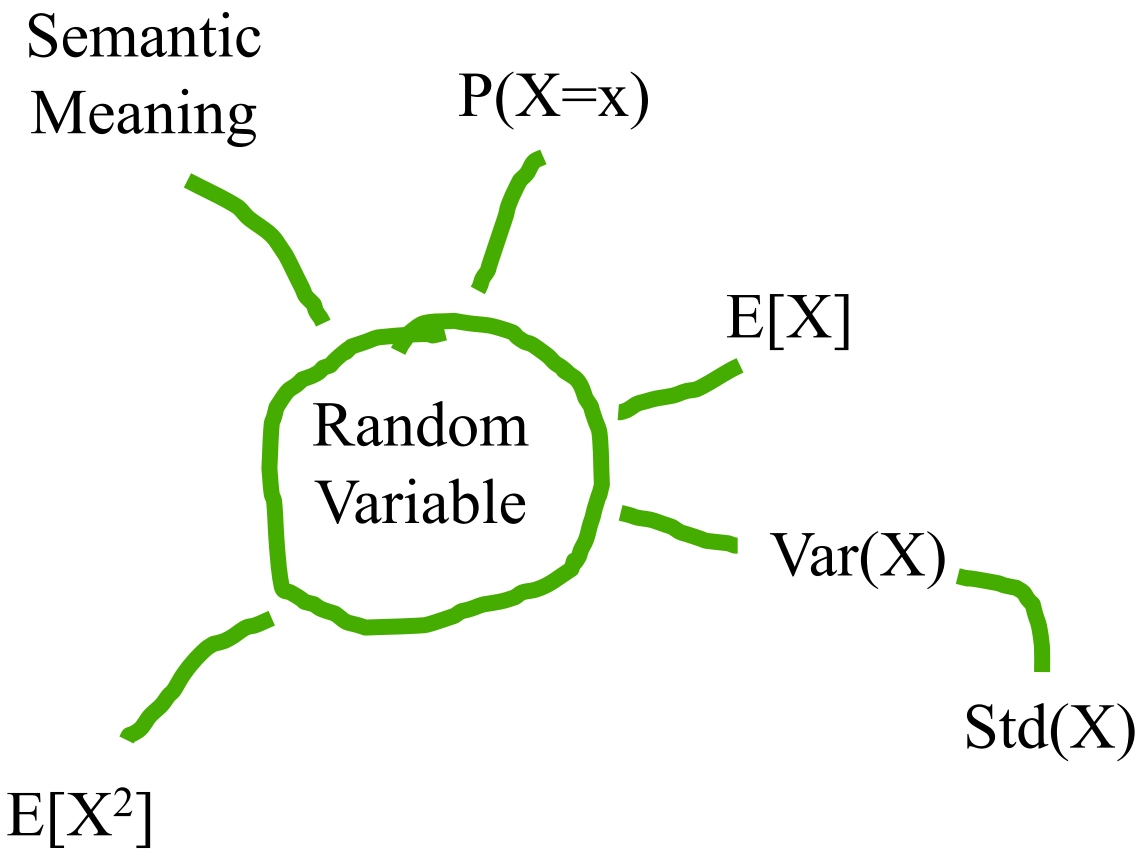
Poisson

Noah Arthurs
CS109, Stanford University

Review

Fundamental Properties



Bernoulli Random Variable

- Experiment results in “Success” or “Failure”
  - X is random **indicator** variable (1 = success, 0 = failure)
  - $P(X = 1) = p(1) = p$ $P(X = 0) = p(0) = 1 – p$
  - X is a **Bernoulli** Random Variable: $X \sim Ber(p)$
  - $E[X] = p$
  - $Var(X) = p(1 – p)$

Bernoulli vs Binomial

Bernoulli is an indicator RV

Binomial is the sum of *n* Bernoullis

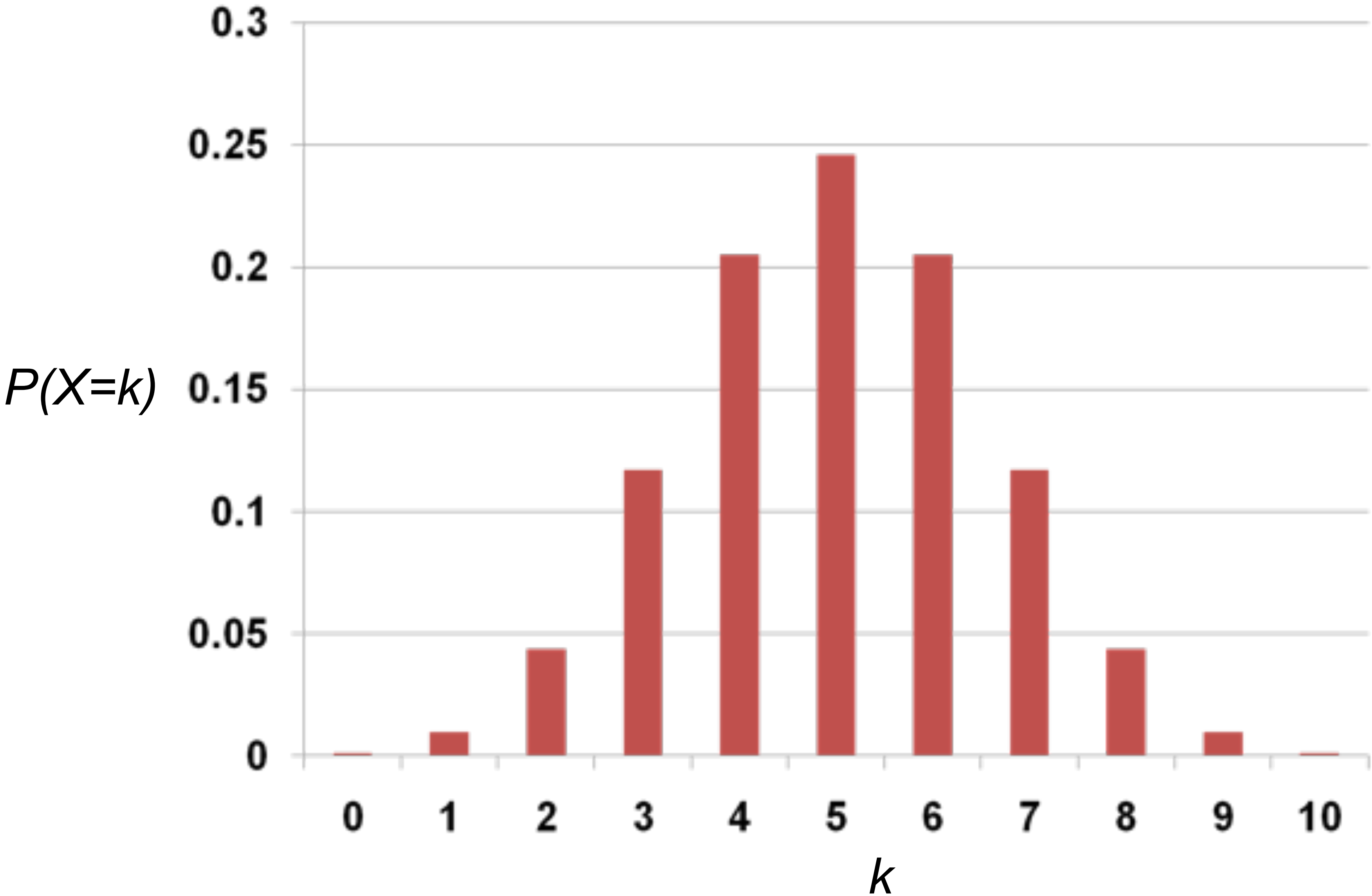
Binomial is for when you have n **independent** trials each with the **same probability of success** and you want to know **how many are successes**.

Properties of Bin(n, p)

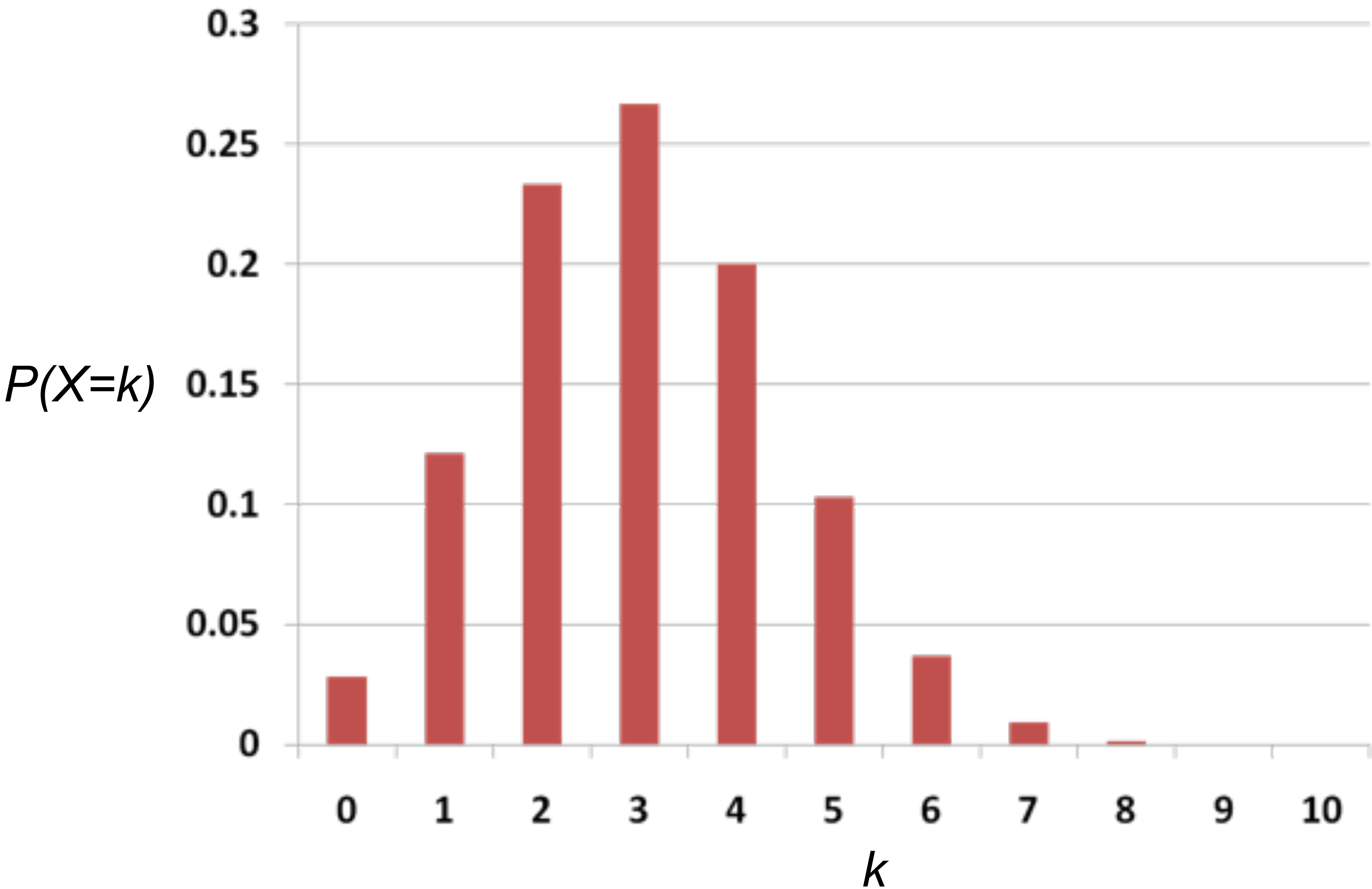
Consider: $X \sim \text{Bin}(n, p)$

- $P(X = i) = p(i) = \binom{n}{i} p^i (1-p)^{n-i} \quad i = 0,1,\ldots,n$
- $E[X] = np$
- $\text{Var}(X) = np(1-p)$
- Note: $\text{Ber}(p) = \text{Bin}(1, p)$

PMF for X ~ Bin(10, 0.5)



PMF for X ~ Bin(10, 0.3)



Probability you win a series?

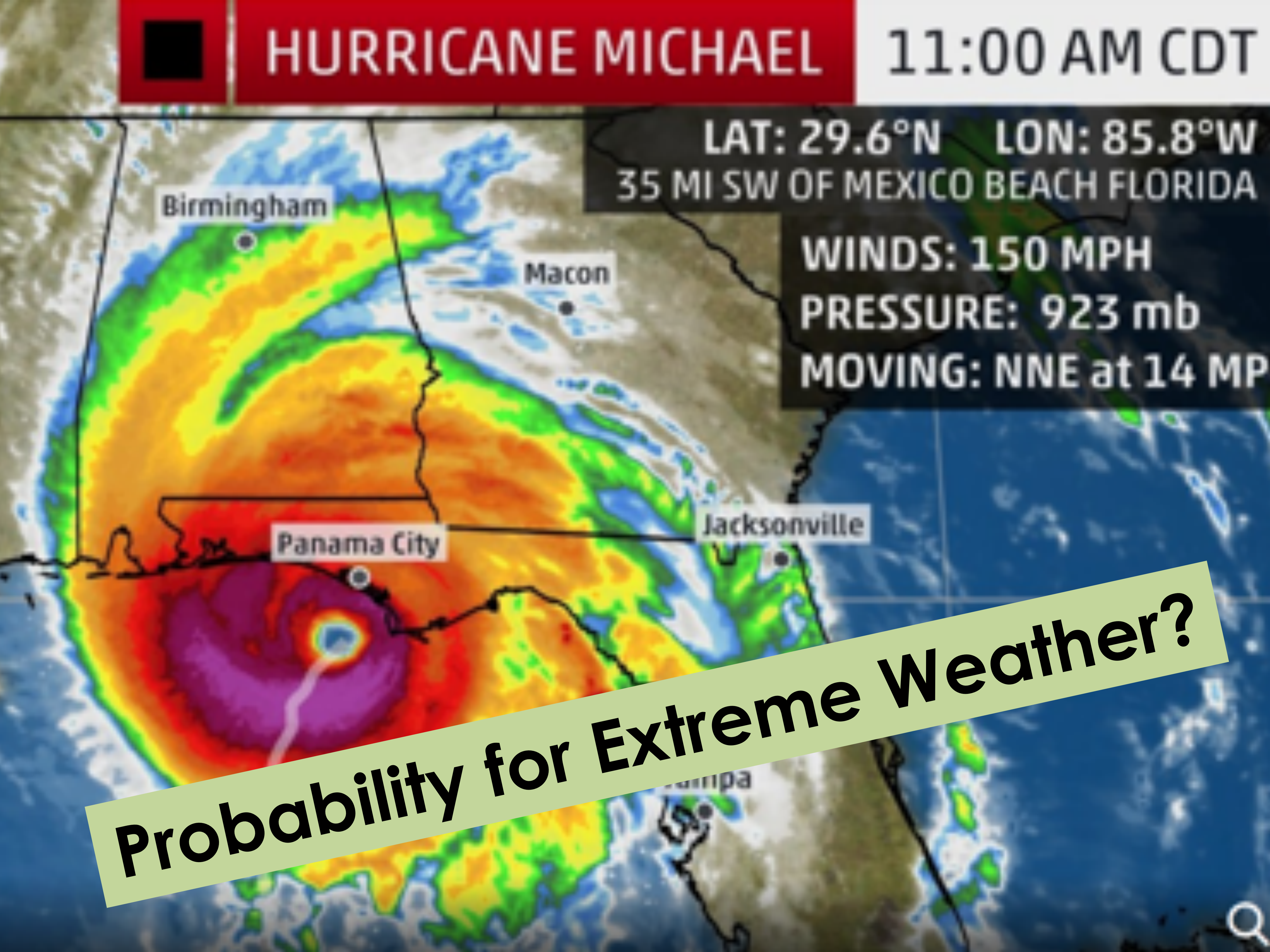
Warriors played the Raptors in a best of 7 series during the 2019 NBA finals. What was the probability (going in) that the warriors would win the series? Each game is **independent**. Each game, the warriors have a 0.55 probability of winning? Win series if you win at least 4 games.

Let X be the number of games won. X ~ Bin(n= 7, p = 0.55).
P(X > 3)?

$$P(X \geq 4) = \sum_{i=4}^{7} P(X = i)$$

$$= \sum_{i=4}^{7} \binom{7}{i} p^i (1-p)^{7-i}$$

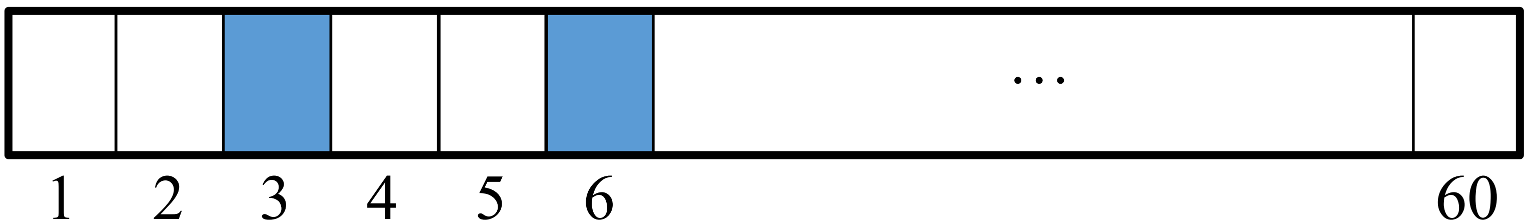
$$= \sum_{i=4}^{7} \binom{7}{i} 0.55^i (0.45)^{7-i} \quad \approx 0.61$$


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?

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

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

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

Can we do any better than milli-seconds?

Binomial in the Limit

On average 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

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

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

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):

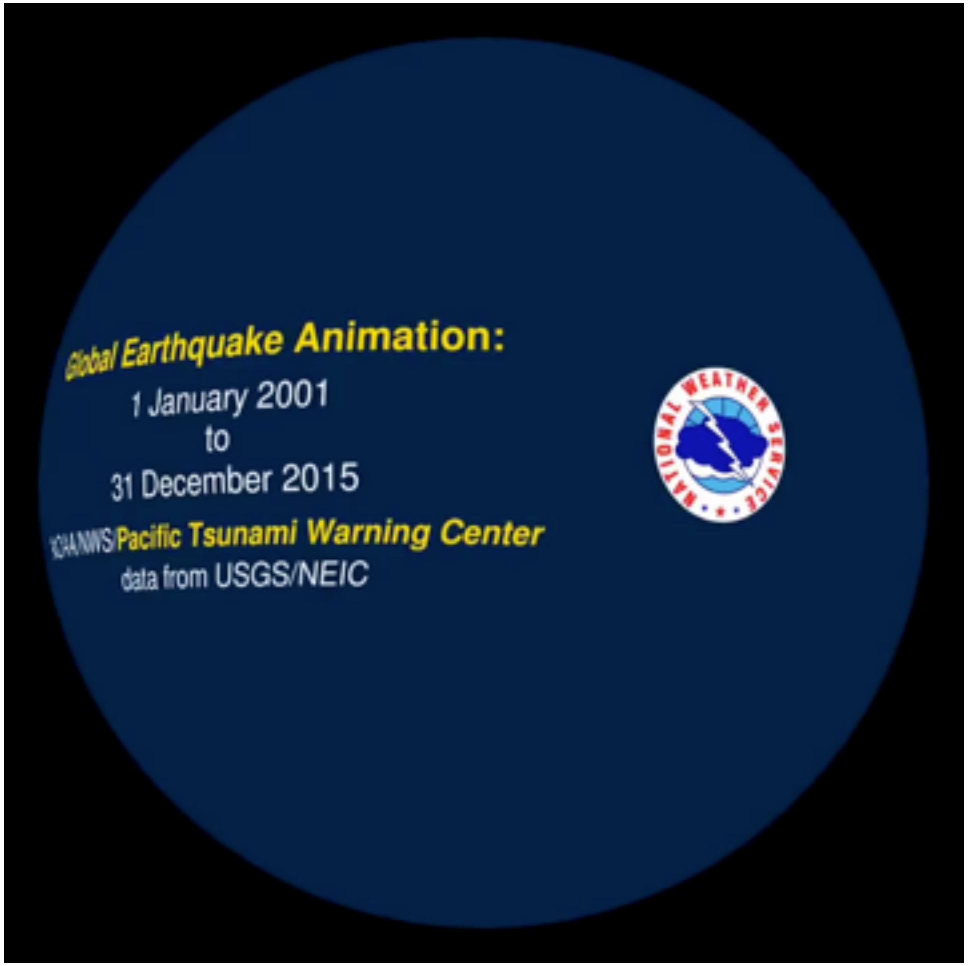
$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

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

Make sure that the rate is the expected number of occurrences within the timeframe you care about.

End Review

Earthquakes

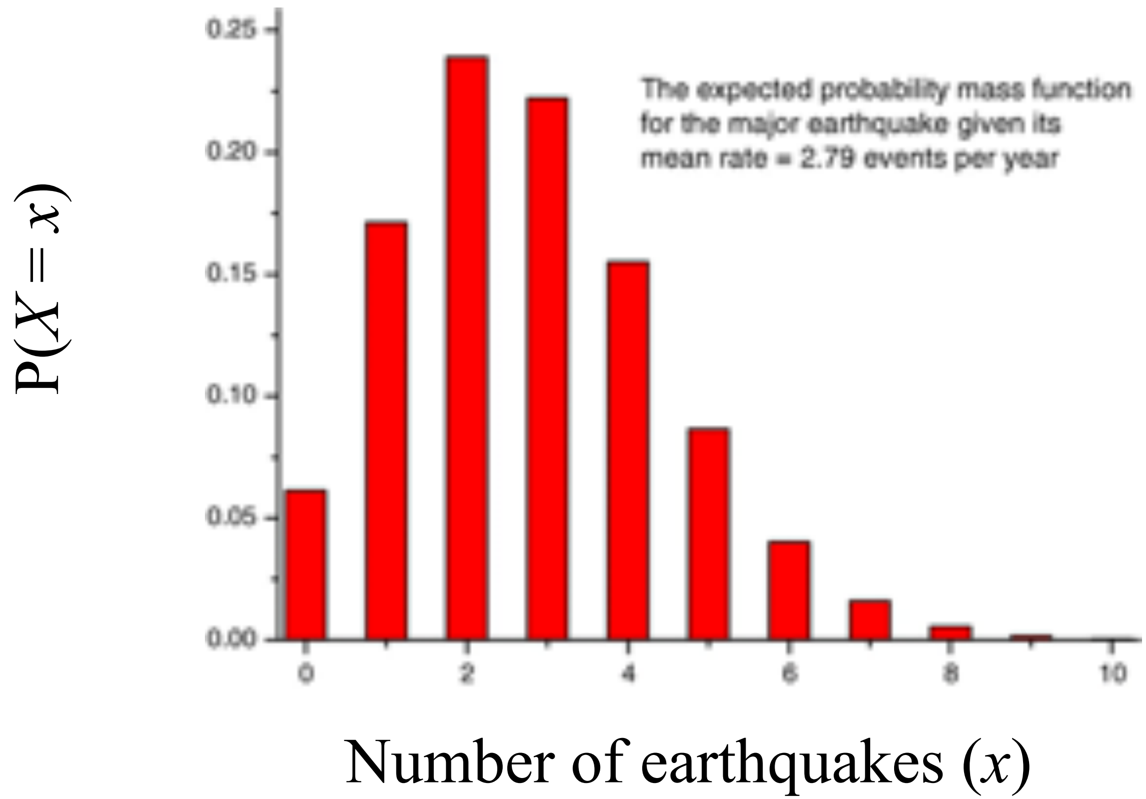


Average of 2.79 major earthquakes per year.
What is the probability of 3 major earthquakes next year?

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?

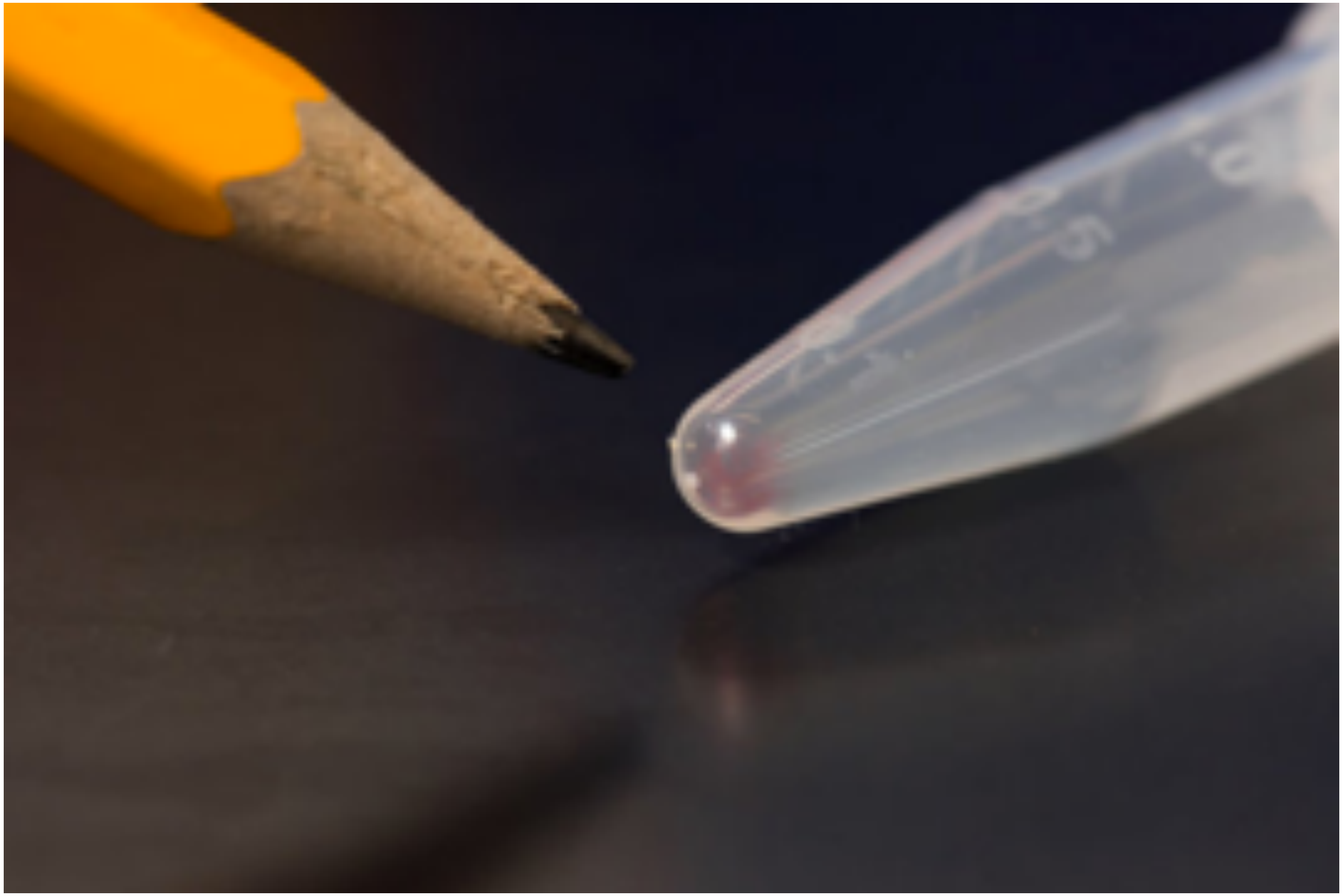
By J. K. Gardner and L. Knopoff

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 steam 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 ~ Poi($\lambda = 10^4 * 10^{-6} = 0.01$)   Hi MoM!

Approximate

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

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

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

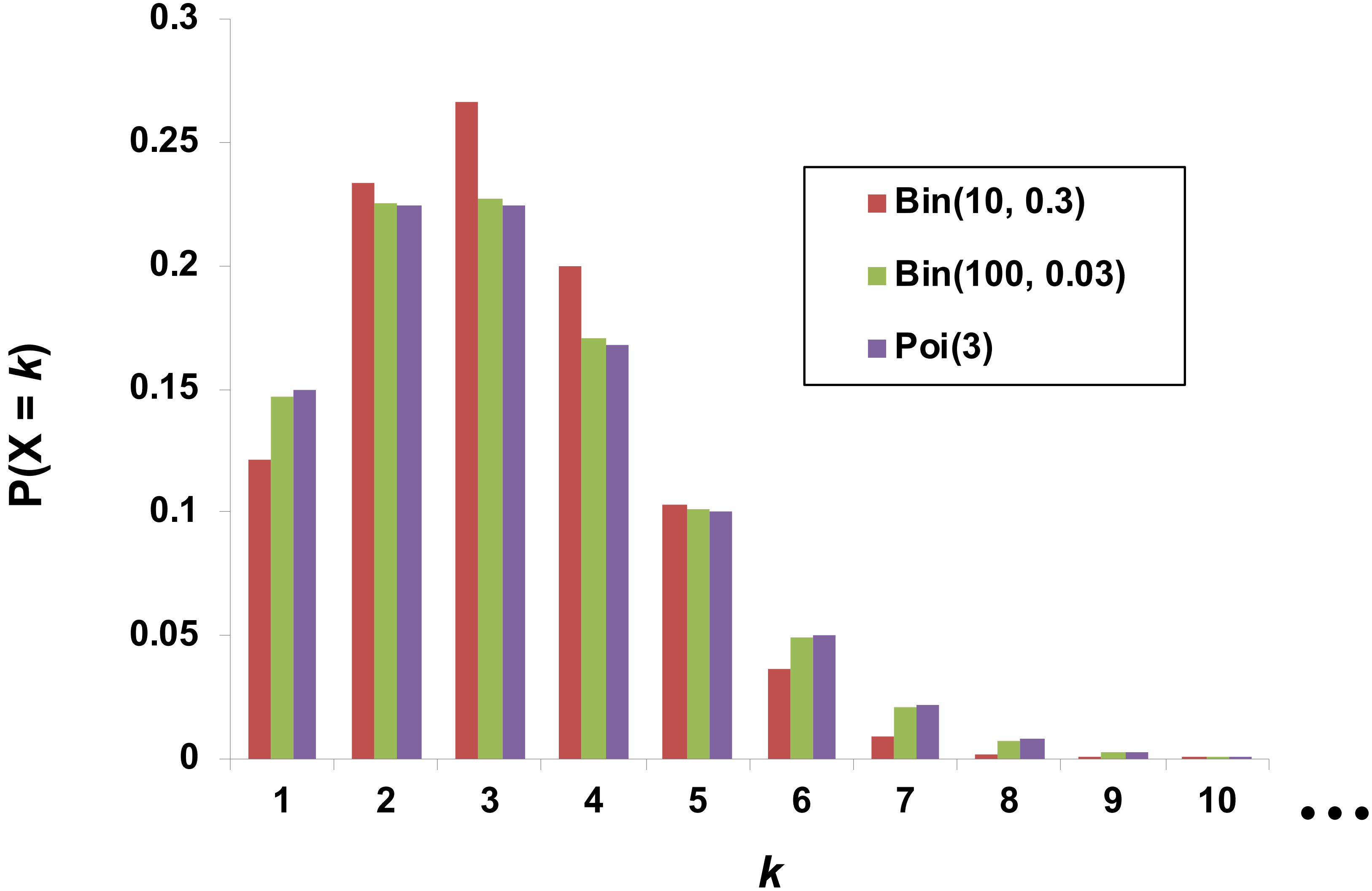
Poisson is Binomial in the Limit

- 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

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

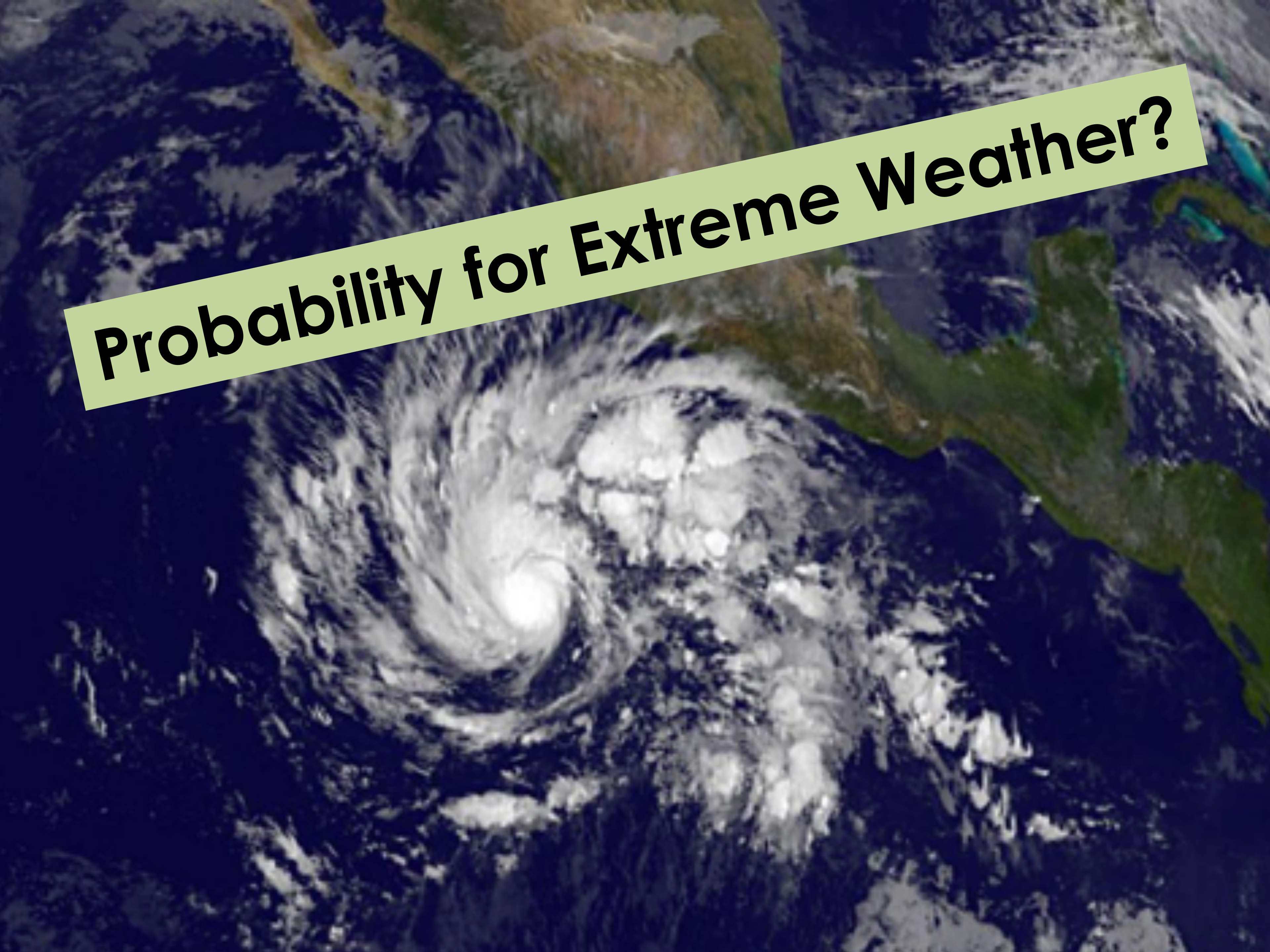
- Consider requests to a web server
  - In past, server load averages 120 hits/minute
  - X = # hits server receives in a second
  - What is $P(X < 5)$?
- Solution

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

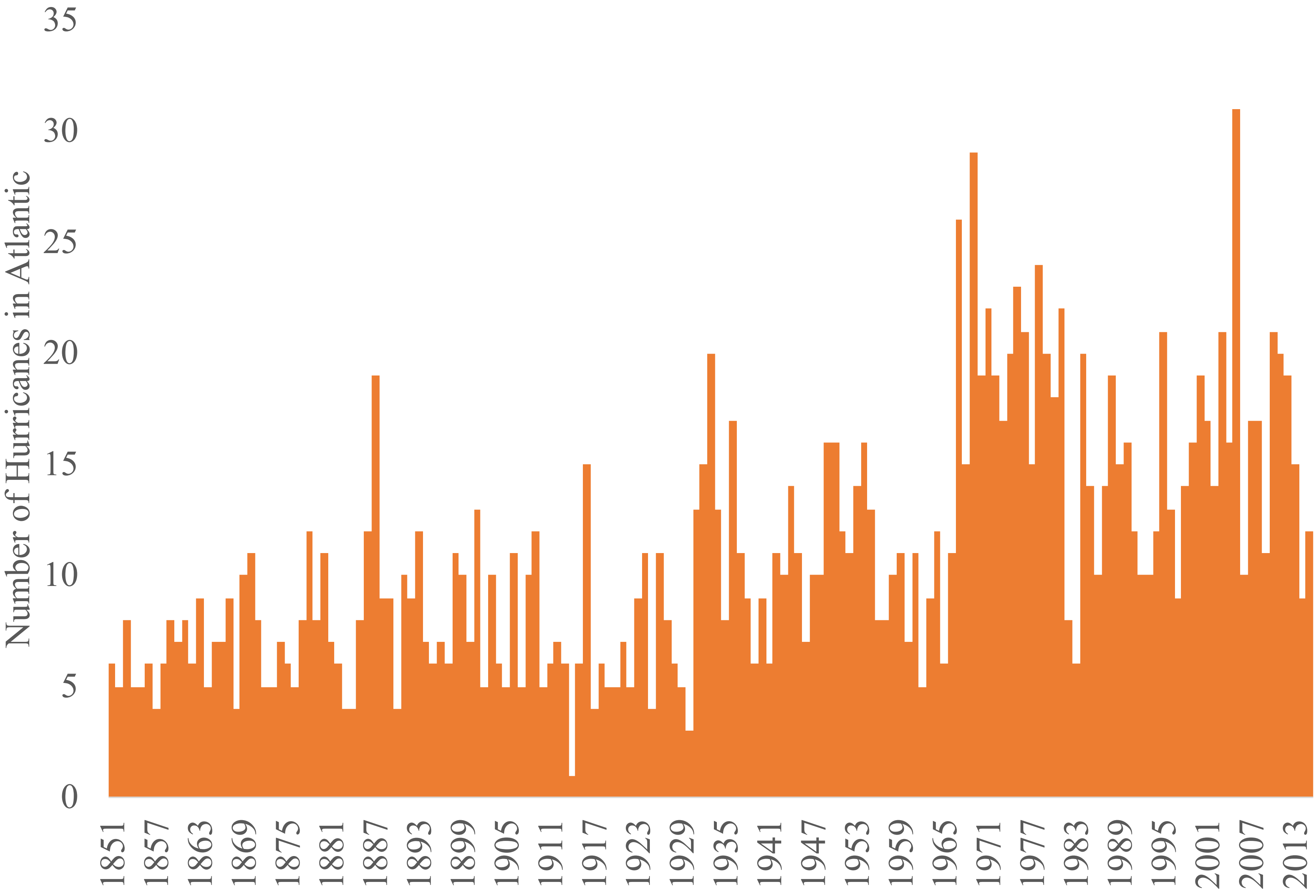
$$P(X < 5) = \sum_{i=0}^{4} P(X = i)$$

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

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

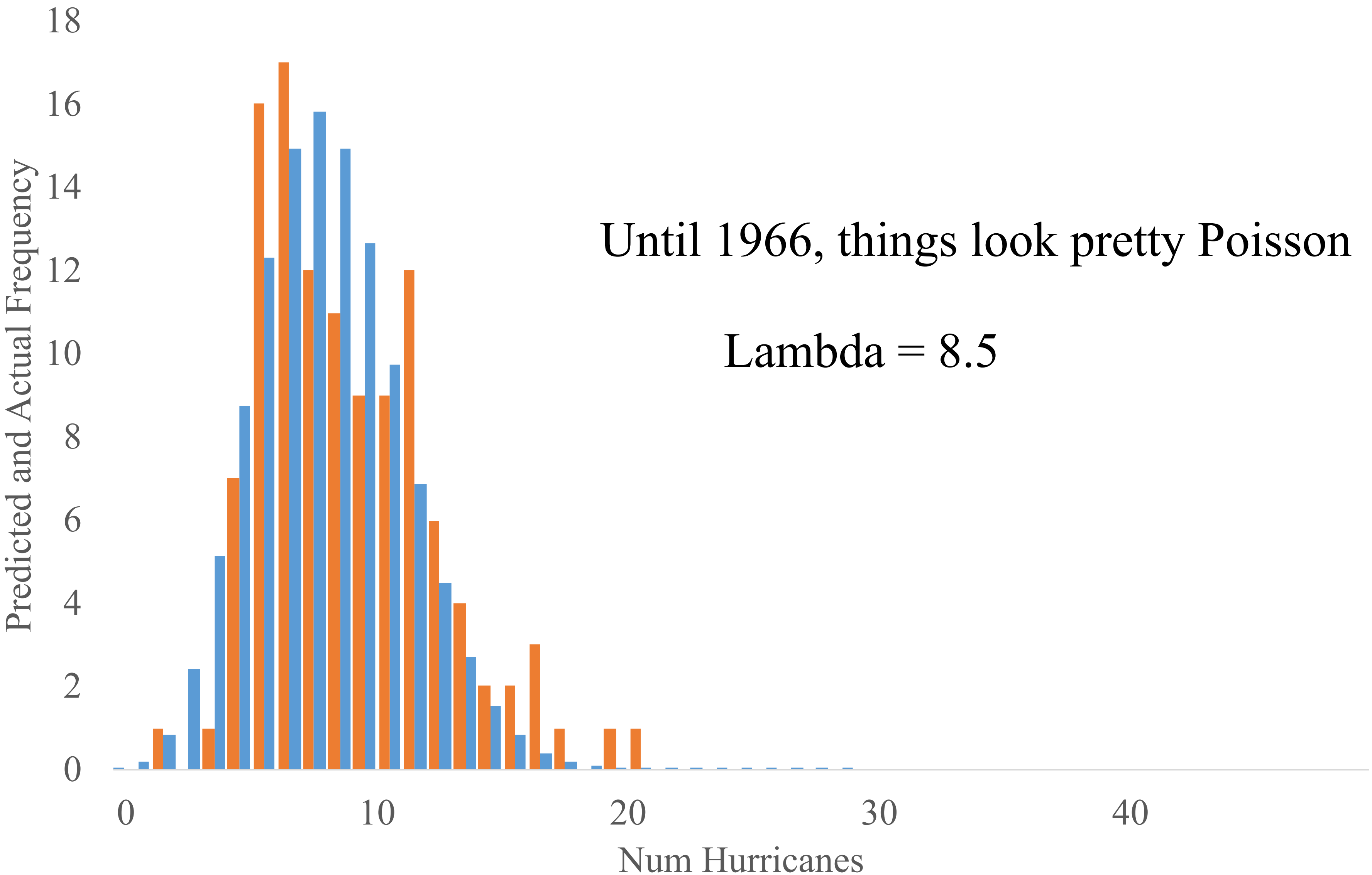

Probability for Extreme Weather?

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:

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

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

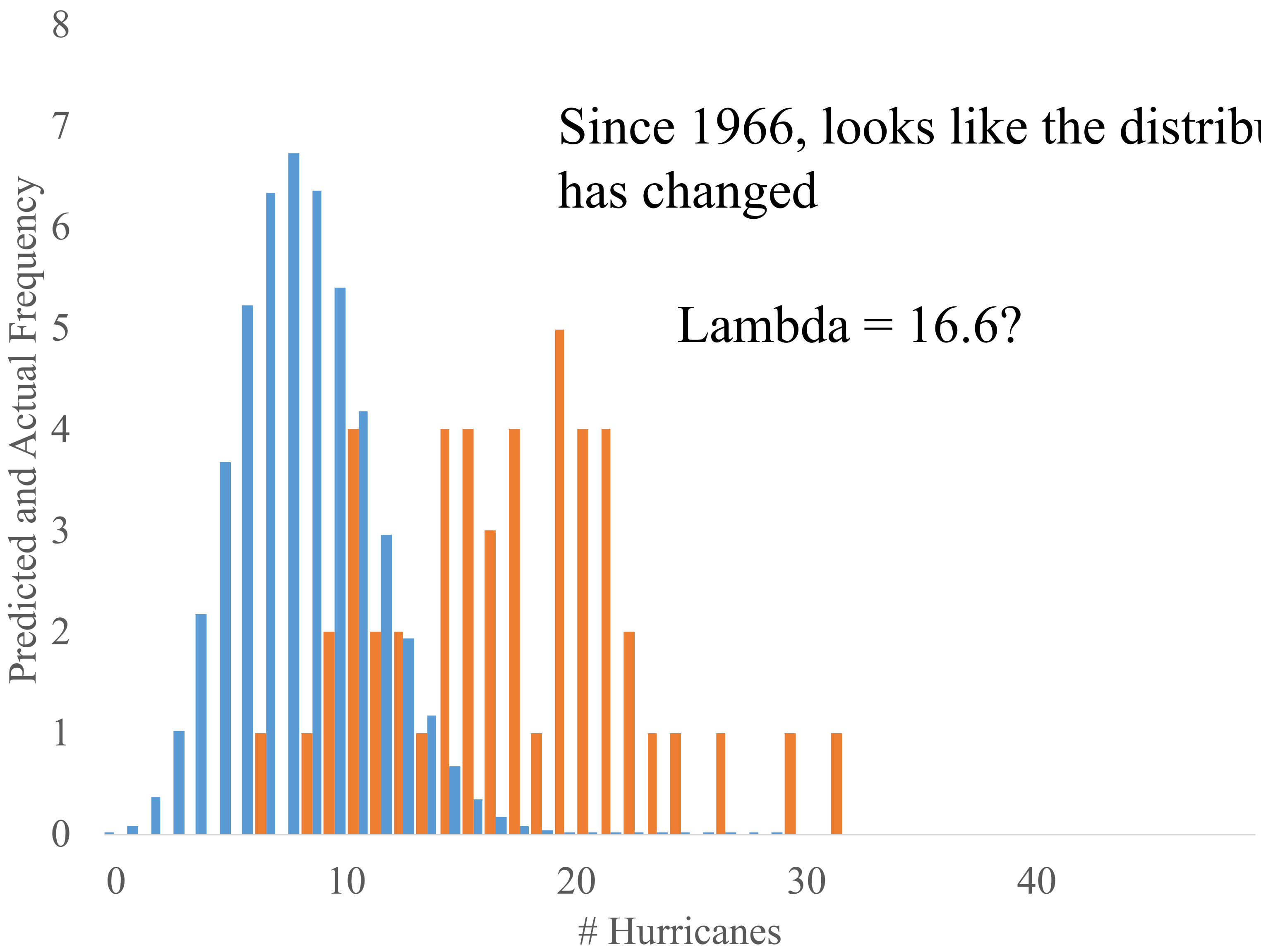
Improbability Drive

- 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 ~ 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

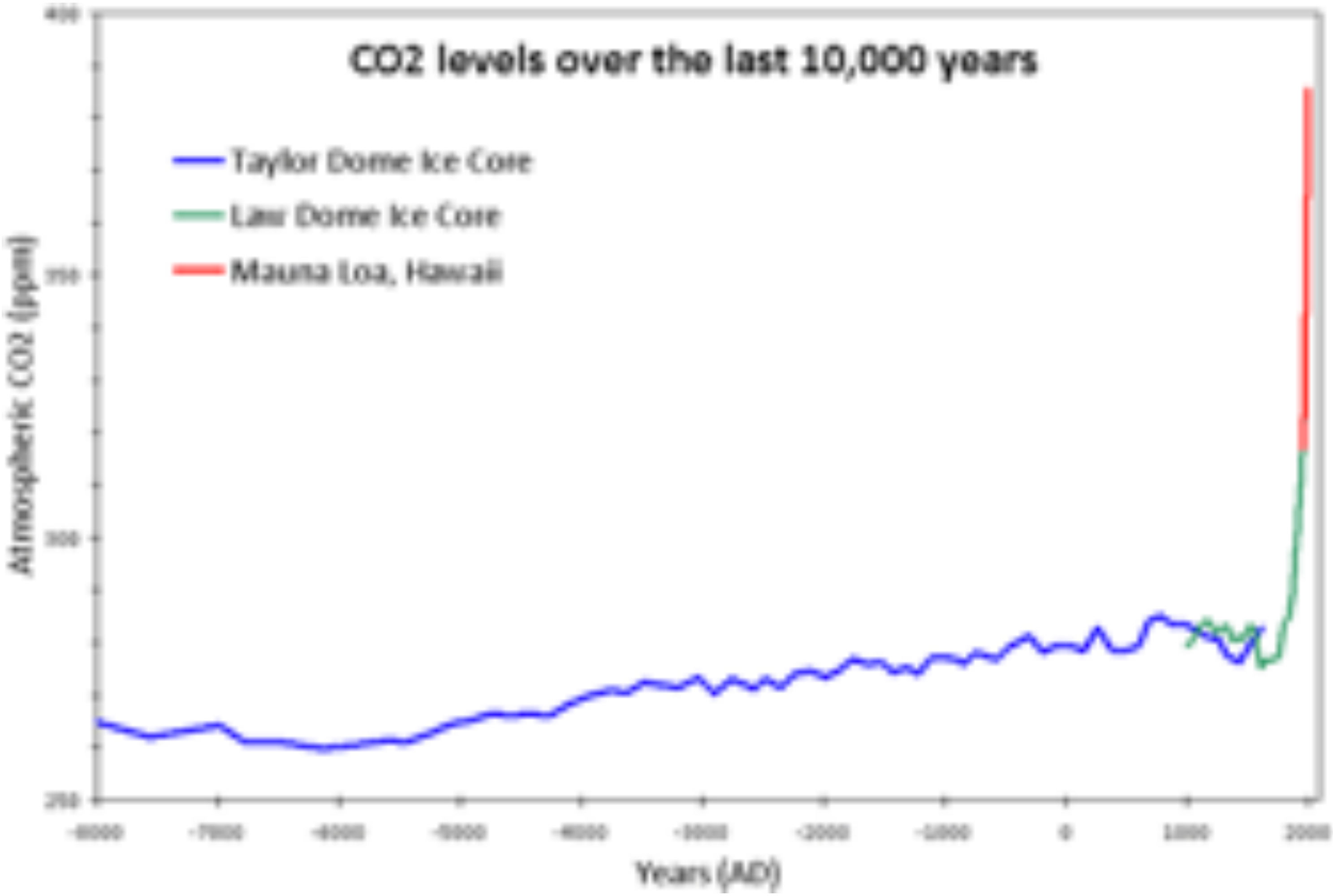
The Distribution has Changed



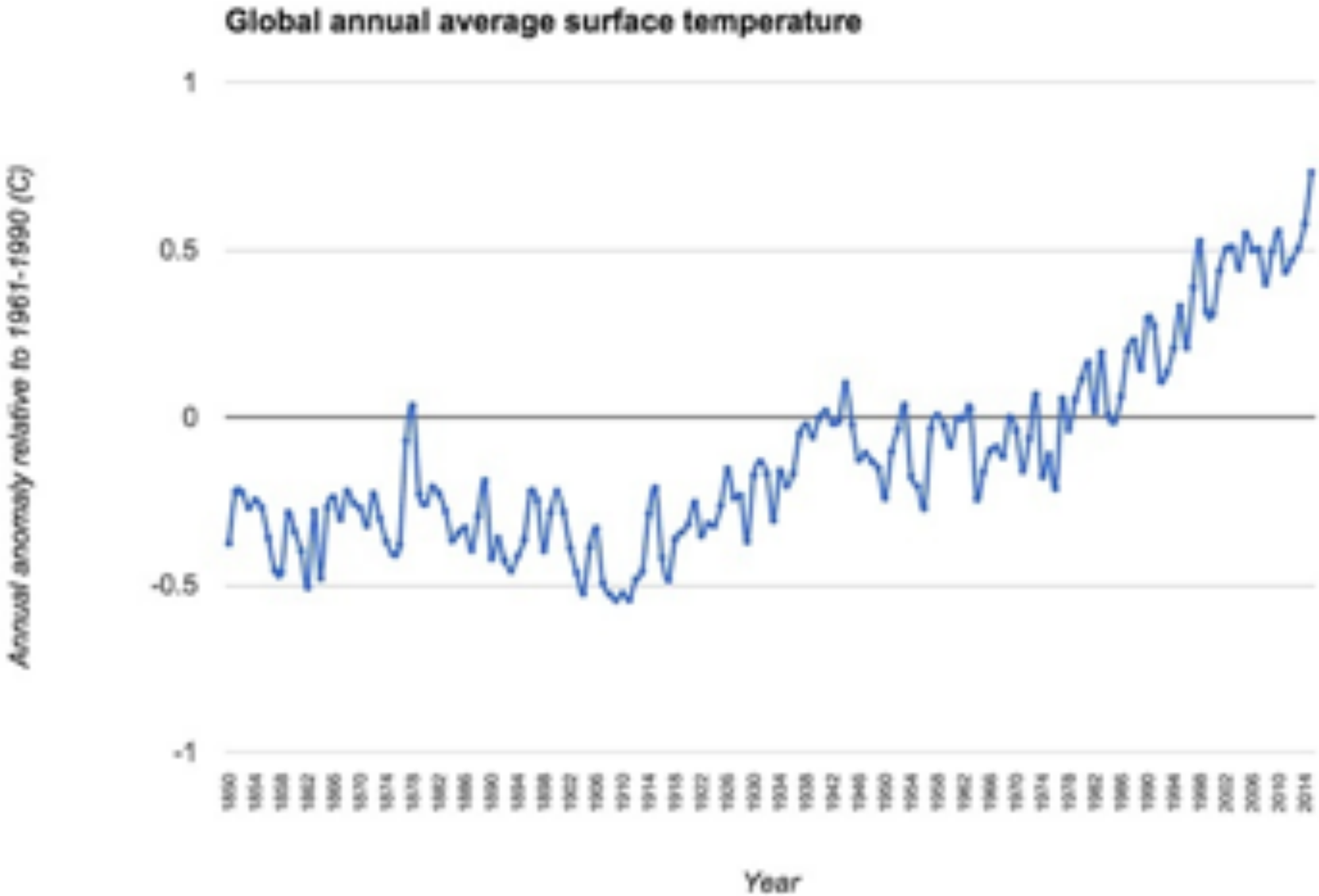
Since 1966, looks like the distribution has changed

Lambda = 16.6?

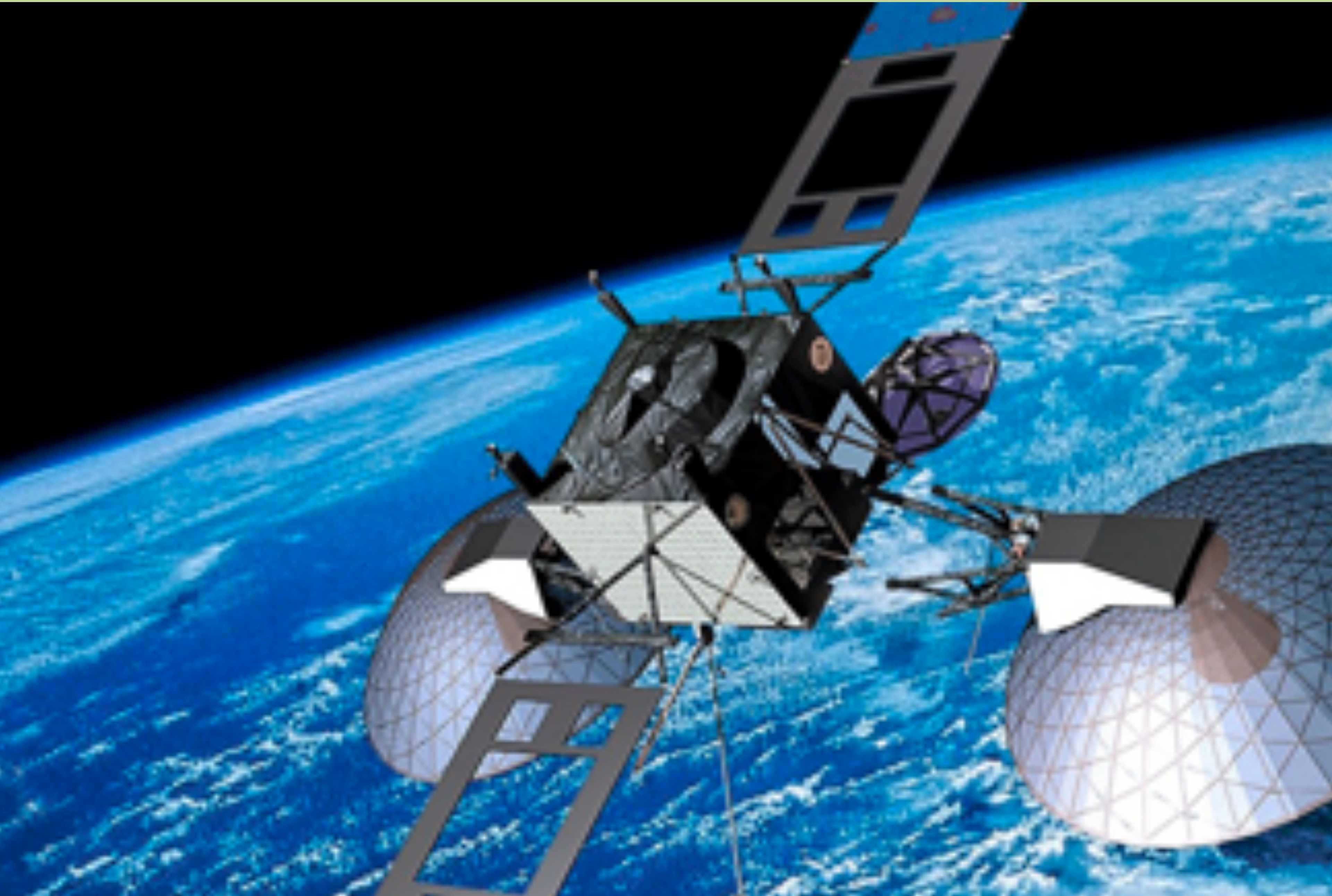
What's Up?


CO2 levels over the last 10,000 years
Taylor Dome Ice Core
Law Dome Ice Core
Mauna Loa, Hawaii
Atmospheric CO2 (ppm)
Years (AD)

What’s Up?


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

What’s Up?



Python Scipy RV Methods

```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)`	$\text{P}(X = k)$
`X.cdf(k)`	$\text{P}(X \leq k)$
`X.entropy()`	(Differential) entropy of X
`X.mean()`	$\text{E}[X]$
`X.var()`	$\text{Var}(X)$
`X,std()`	$\text{Std}(X)$

Discrete Distributions

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

Grid of Random Variables

	number of **successes**	**trials/time** to get successes	
One trial	$X \sim \mathrm{Ber}(p)$		One success
	***n* = 1**		
Several trials	$X \sim \mathrm{Bin}(n, p)$		Several successes
Interval of time	$X \sim \mathrm{Poi}(\lambda)$		One success

Geometric Random Variable

- X is **Geometric** 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:

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

$$E[X] = \frac{1}{p}$$

$$\mathrm{Var}(X) = \frac{1-p}{p^2}$$

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*)

Grid of Random Variables

	number of **successes**	**trials/time** to get successes	
One trial	$X \sim \mathrm{Ber}(p)$	$X \sim \mathrm{Geo}(p)$	One success
Several trials	↑ ***n* = 1** $X \sim \mathrm{Bin}(n, p)$	↑ ***r* = 1** $X \sim \mathrm{NegBin}(r, p)$	Several successes
Interval of time	$X \sim \mathrm{Poi}(\lambda)$	$X \sim \mathrm{Exp}(\lambda)$	One success

New Handout

Chris Piech
CS 109

Handout
Oct 12, 2018

All Discrete Distributions

Bernoulli

An indicator variable that takes on the value 1 or 0. Often the variable is defined to be 1 if an underlying event has occurred, 0 otherwise.

Notation	$X \sim \text{Bern}(p)$
Parameters:	p: The probability of the variable being 1
Range(X):	$\{0, 1\}$
pmf:	$\Pr(X = k) = \begin{cases} p & \text{if k = 1} \\ (1-p) & \text{if k = 0} \end{cases}$
E[X]:	p
Var(X):	$p(1-p)$

Note: Sometimes in machine learning algorithms a derivable version of the PMF is used: $f(X = k) = p^k(1-p)^{1-k}$

Binomial
A variable which represents the number of successes in a fixed number of independent trials. The probability of success must be the same for each trial.

Notation	$X \sim \text{Bin}(n, p)$
Parameters:	n : the number of trials p: the probability of success in each trial
Range(X):	$\{0, 1, \ldots, n\}$
pmf:	$\Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$
E[X]:	np
Var(X):	$np(1-p)$

Note: $\text{Bin}(1, p) = \text{Bern}(p)$

Poisson
The number of events occurring in a fixed interval of time or space if these events occur independently with a constant rate.

Notation	$X \sim \text{Poi}(\lambda)$
Parameters:	λ : the rate of events in one interval
Range(X):	$\{0, 1, \ldots, \infty\}$
pmf:	$\Pr(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$
E[X]:	λ
Var(X):	λ

Note: The Poisson is the number of events in an interval of time. The Exponential is a continuous distribution which is the time until the next event. They have the same parameter (λ).

Geometric
The number of independent Bernoulli trials until the first success.

Notation	$X \sim \text{Geo}(p)$
Parameters:	p: the probability of success of each trial
Range(X):	$\{1, 2, \ldots, \infty\}$
pmf:	$\Pr(X = k) = (1-p)^{k-1} p$
E[X]:	$1/p$
Var(X):	$\frac{1-p}{p^2}$

Etc…

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)

Bit Coin Mining

SHA-256 Hash(Data *Fixed* , Salt *Choice*)

Number that looks like random bits

You "mine a bitcoin" if, for given data *D*, you find a number *N* such that Hash(*D*, *N*) produces a string that starts with *g* zeroes.

Midterm Question: Bit Coin Mining

You "mine a bitcoin" if, for given data *D*, you find a number *N* such that Hash(*D*, *N*) produces a string that starts with *g* zeroes.

(a) What is the probability that the first number you try will produce a bit string which starts with *g* zeroes (in other words you mine a bitcoin)?

(b) How many different numbers do you expect to have to try before you mine five bitcoins?

Dating at Stanford

Each person you date has a 0.2 probability of being someone you spend your life with. What is the average number of people one will date? What is the standard deviation?

Equity in the Courts

Berghuis v. Smith

If a group is underrepresented in a jury pool, how do you tell?

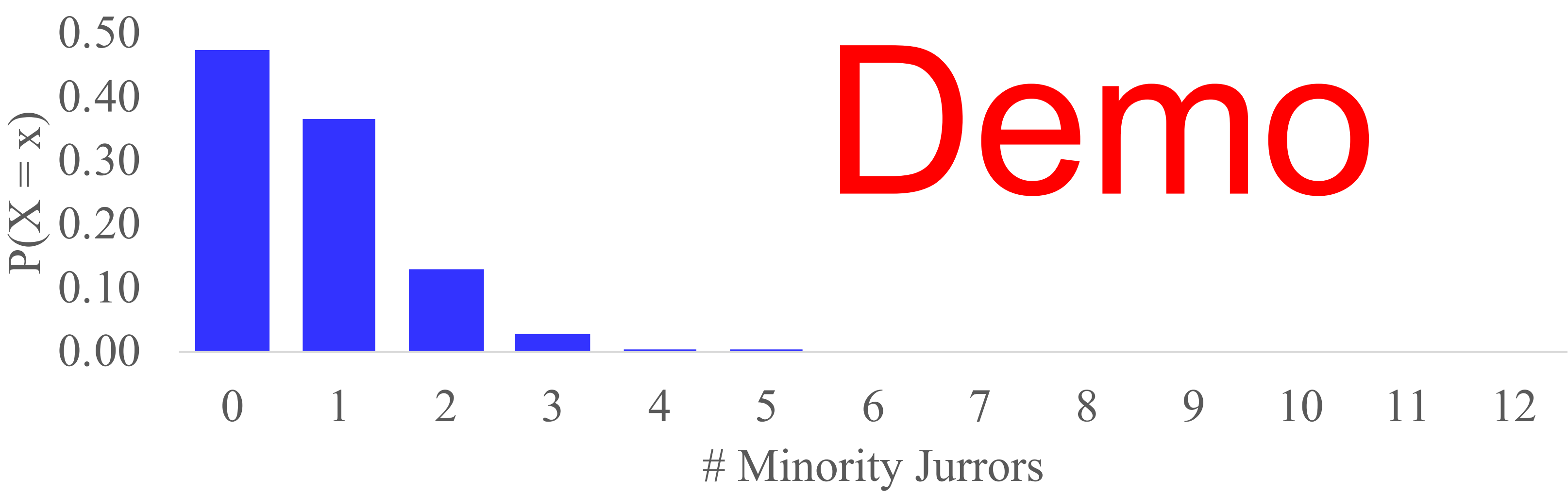
- Article by Erin Miller –January 22, 2010
- Thanks to (former CS109er) Josh Falk for this article

Justice Breyer [Stanford Alum] opened the questioning by invoking the binomial theorem. He hypothesized a scenario involving **"an urn with a thousand balls, and sixty are blue, and nine hundred forty are purple, and then you select them at random… twelve at a time."** According to Justice Breyer and the binomial theorem, if the purple balls were under represented jurors then **"you would expect… something like <u>a third to a half</u> of juries would have at least one minority person"** on them.

Justin Breyer Meets CS109

- Approximation using Binomial distribution
  - Assume P(blue) constant for every draw = 60/1000
  - X = # blue drawn. $X \sim \text{Bin}(12, 60/1000 = 0.06)$
  - $P(X \geq 1) = 1 - P(X = 0) \approx 1 - 0.4759 = 0.5240$

In Breyer's description, should actually expect just ***over half*** *of juries to have at least one non-white person on them*



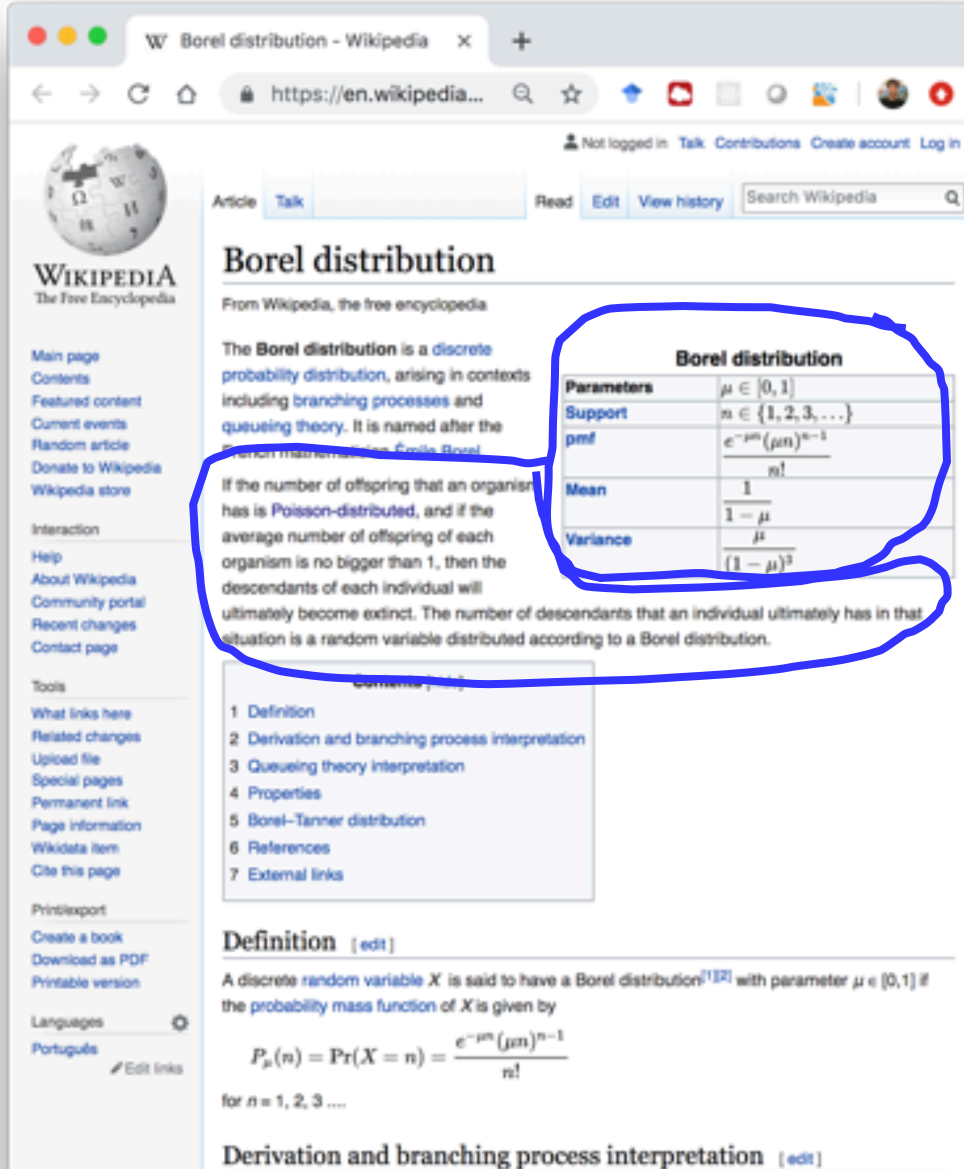
Demo

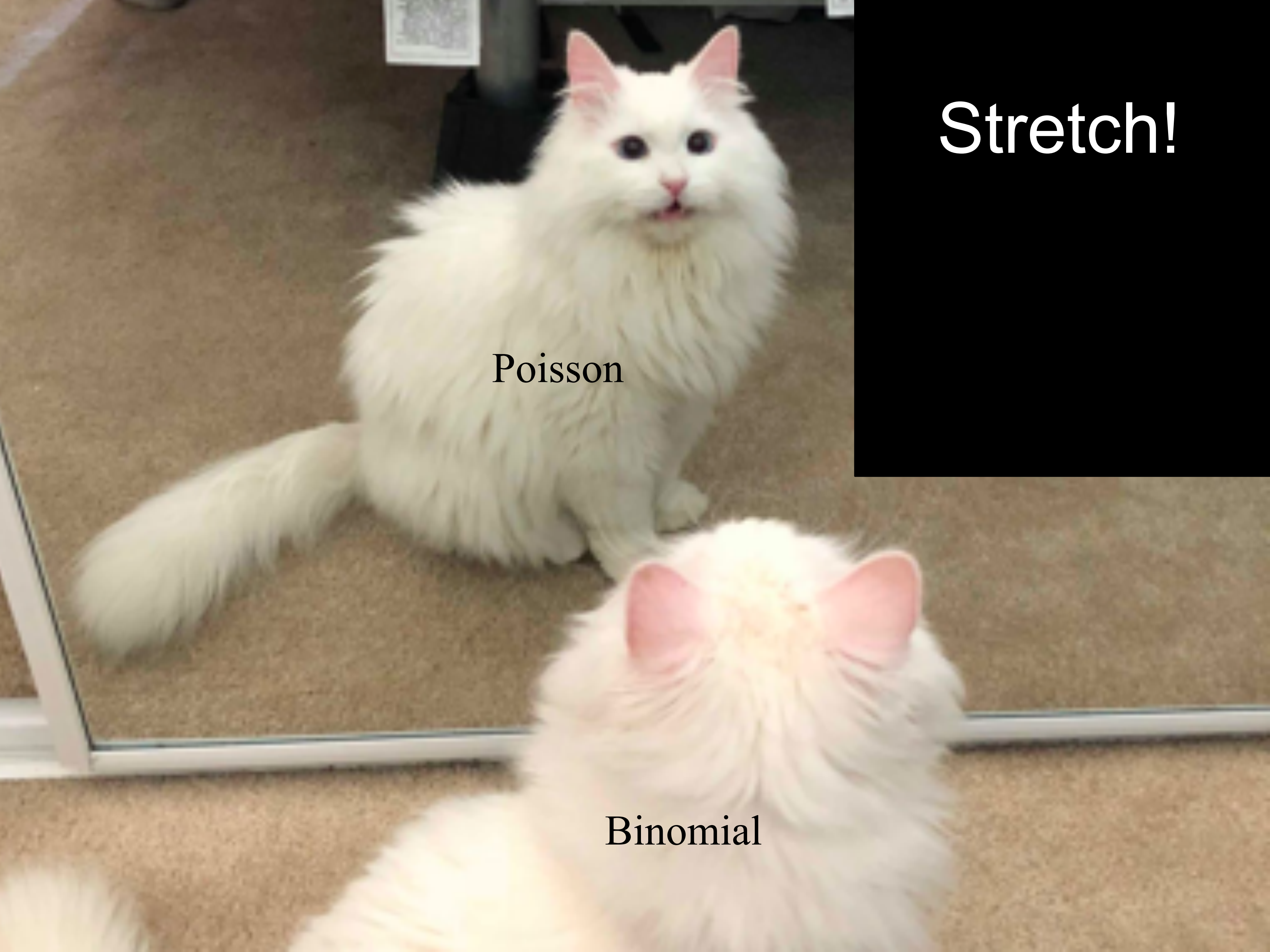
Learning Goal: Use new RVs

You are learning about servers…

You read about the MD1 queue…

You find a paper that says the server "busy period" is distributed as a Borel with parameter $\mu = 0.2$ …




Stretch!
Poisson
Binomial

Continuous Variables

Noah Arthurs
CS109, Stanford University

Learning Goals

1. Comfort using new discrete random variables
2. Integrate a density function (PDF) to get a probability
3. Use a cumulative function (CDF) to get a probability

Big hole in our knowledge

Not all values are discrete

`random()?`

Riding the Marguerite

Riding the Marguerite

You are running to the bus stop. You don't know exactly when the bus arrives. You have a distribution of probabilities.

You show up at 2:20pm.

What is P(wait < 5 minutes)?

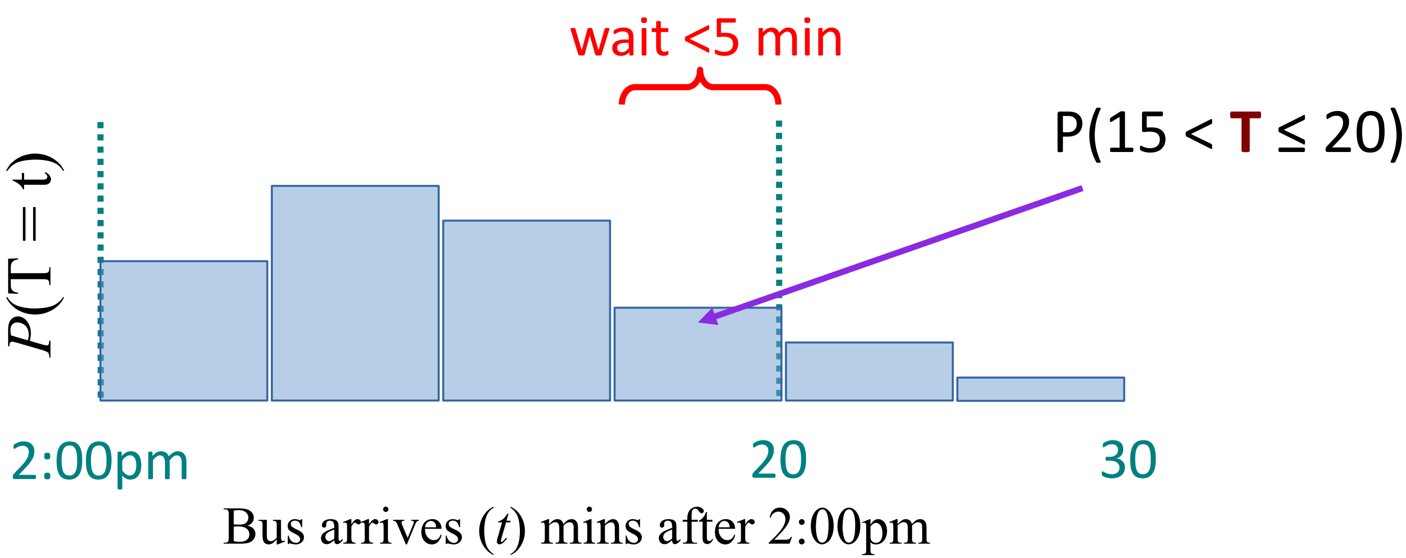
What is the probability that the bus arrives at:
2:17pm and 12.12333911102389234 seconds?

Riding the Marguerite

You are running to the bus stop. You don't know exactly when the bus arrives. You have a distribution of probabilities.

You show up at 2:15pm.

What is P(wait < 5 minutes)?

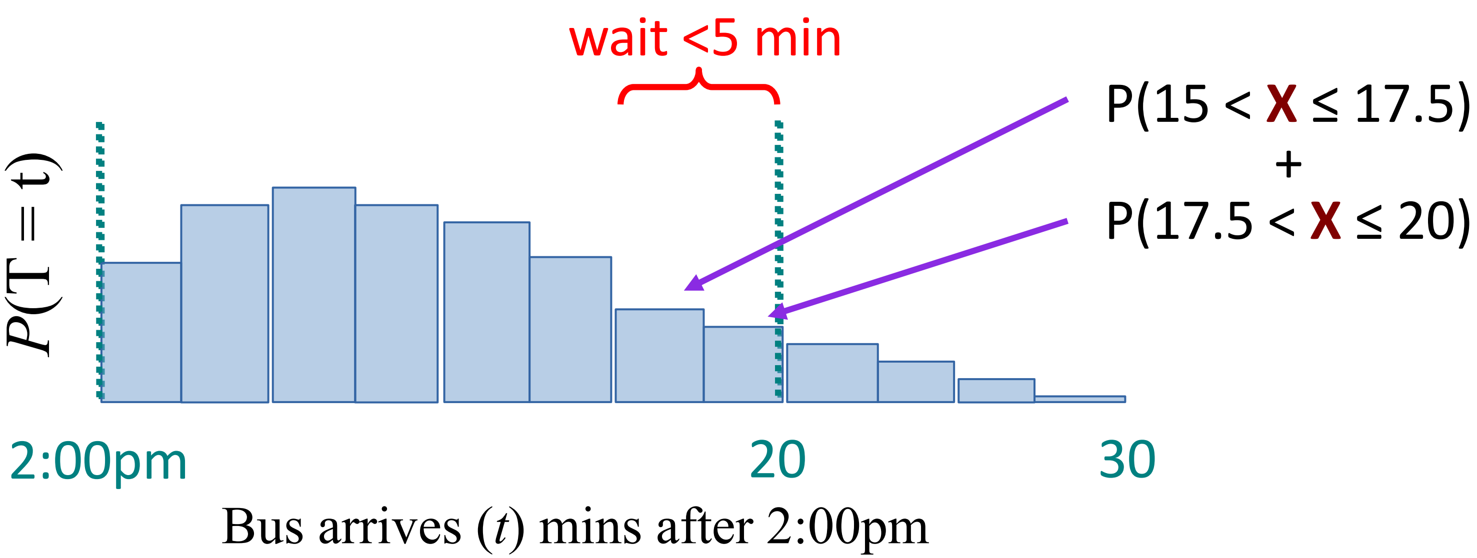


Riding the Marguerite

You are running to the bus stop. You don't know exactly when the bus arrives. You have a distribution of probabilities.

You show up at 2:15pm.

What is P(wait < 5 minutes)?

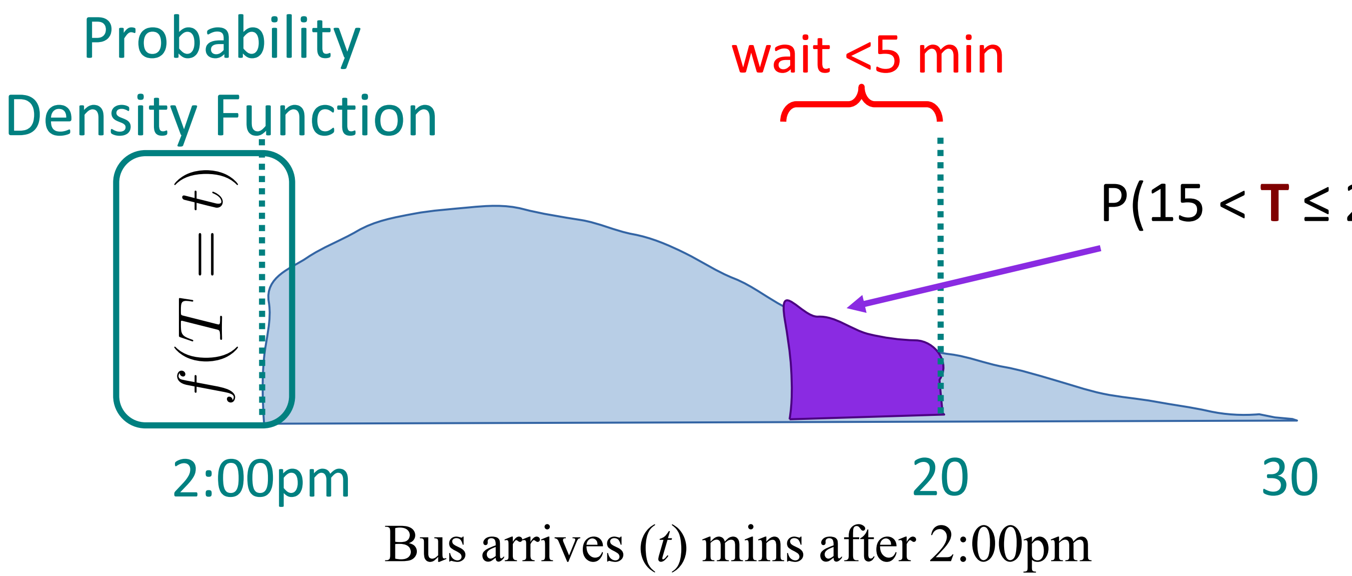


Riding the Marguerite

You are running to the bus stop. You don't know exactly when the bus arrives. You have a distribution of probabilities.

You show up at 2:15pm.

What is P(wait < 5 minutes)?



Probability Density Function

The **probability density function** (PDF) of a continuous random variable represents the relative likelihood of various values.

Units of probability *divided by units of X*.
Integrate it to get probabilities!

$$P(a < X < b) = \int_{x=a}^{b} \boxed{f(X = x)}\, dx$$

Probability Density Function

The **probability density function** (PDF) of a continuous random variable represents the relative likelihood of various values.

Units of probability *divided by units of X*.
Integrate it to get probabilities!

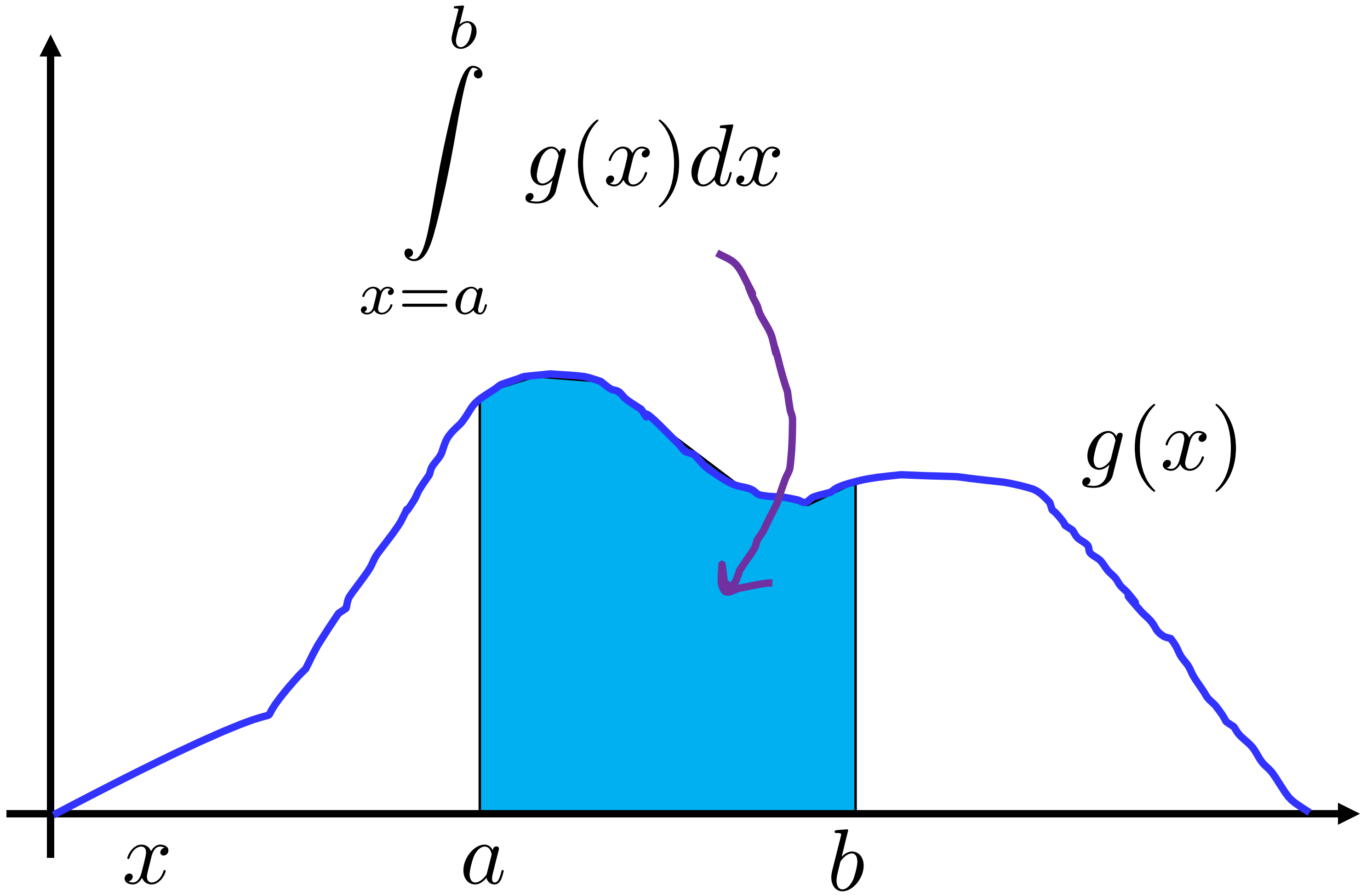
$$P(a < X < b) = \int_{x=a}^{b} \boxed{f_X(x)}\, dx$$

This is another way to write the PDF

Integrals

*loving, not scary

Integrals



Riding the Marguerite

You are running to the bus stop. You don't know exactly when the bus arrives. You have a distribution of probabilities.

You show up at 2:15pm.

What is P(wait < 5 minutes)?

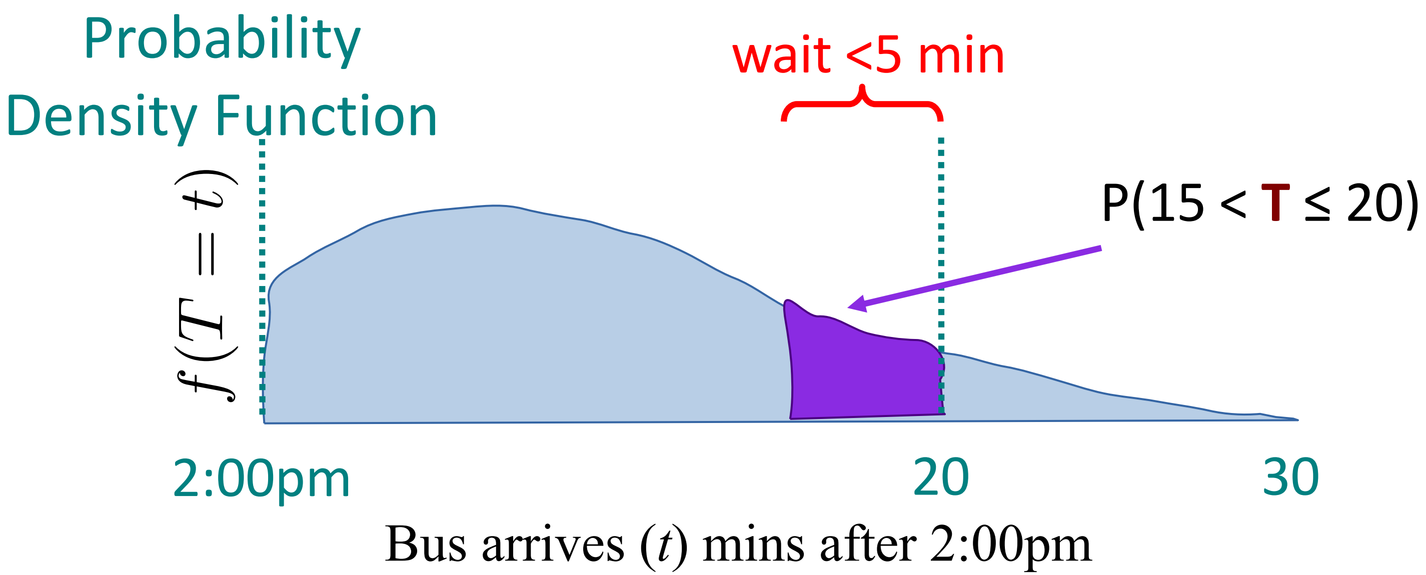


image: Haha169

Properties of PDFs

The integral of a PDF gives a probability. Thus:

$$0 \leq \int_{x=a}^{b} f(X = x)\ dx \leq 1$$

$$\int_{x=-\infty}^{\infty} f(X = x)\ dx = 1$$

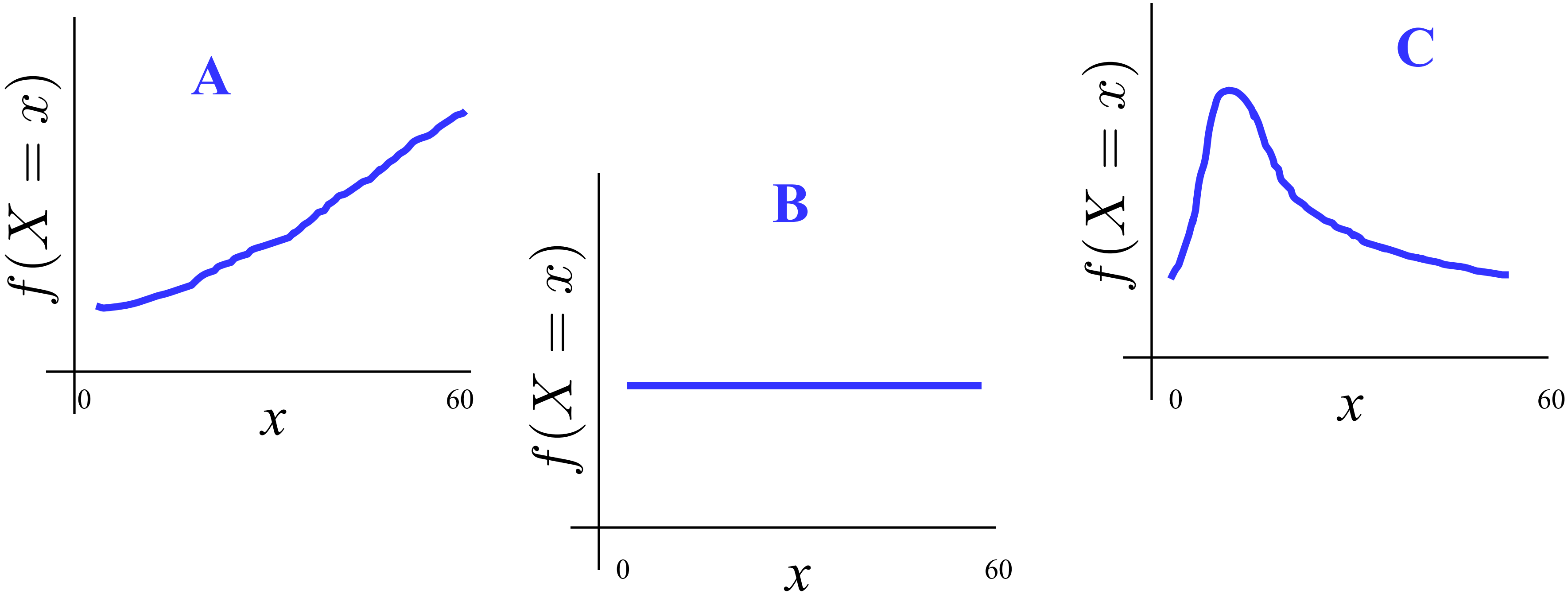
What do you get if you integrate over a probability *density* function?

A probability!

Probability Density Function

Probability density functions articulate ***relative*** belief.

Let X be a random variable which is the # of minutes after 2pm that the bus arrives at the stop:

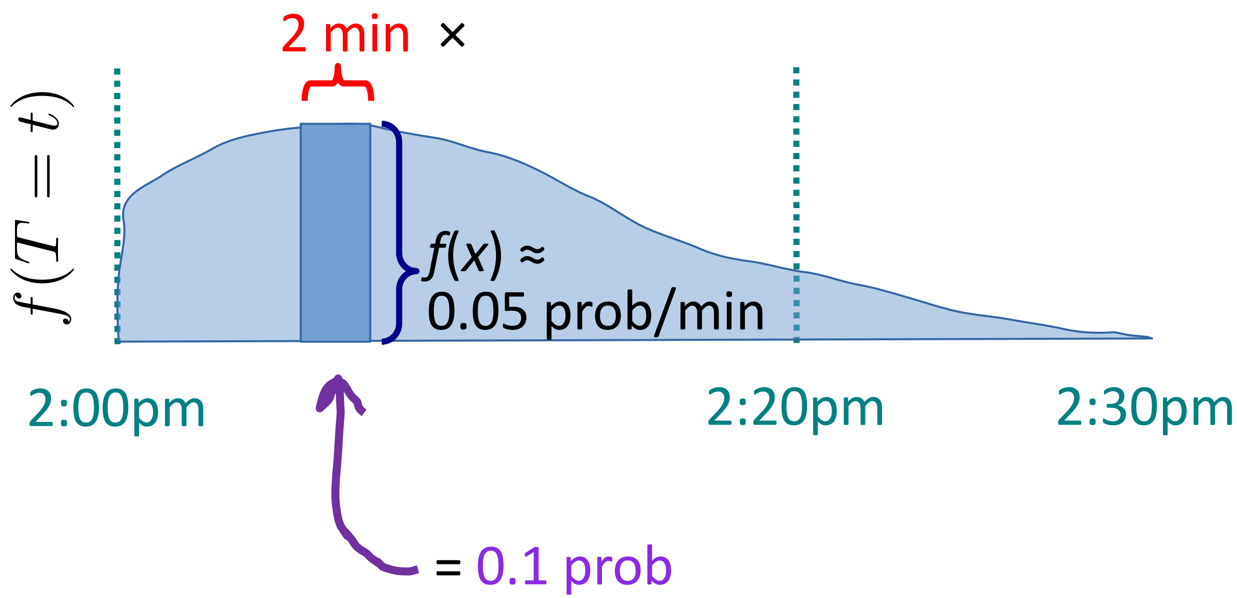


Which of these represent that you think the arrival is more likely to be close to 3:00pm

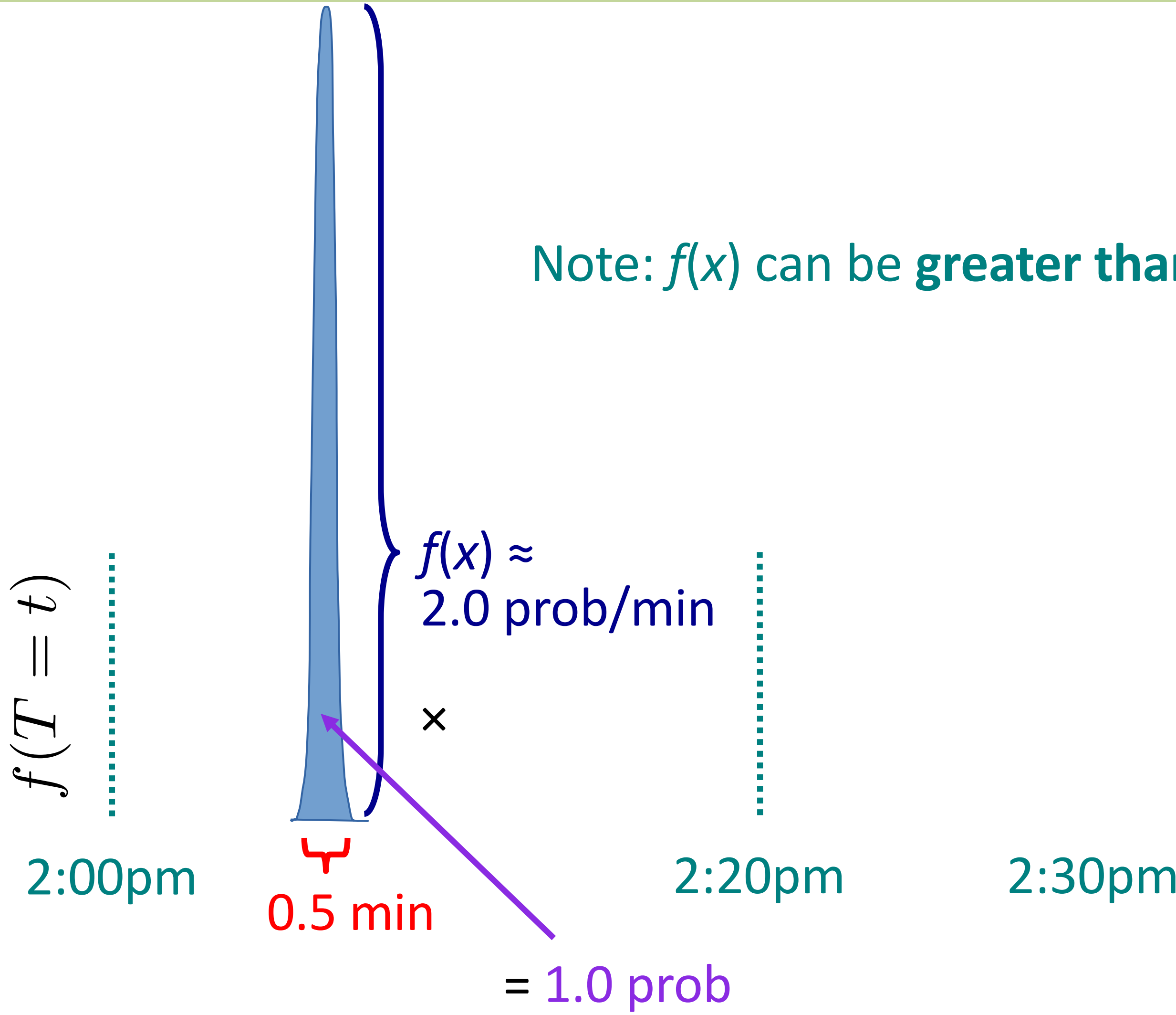
The ratio of probability densities is meaningful

$f(X = x)$ is *Not* a Probability

Rather, it has "units" of:
probability divided by units of *X*.



$f(X = x)$ is ***Not*** a Probability



Note: $f(x)$ can be **greater than 1**!

$f(X = x)$ vs $P(X = x)$

"The probability that a **discrete** random variable X takes on the value little *x*."

$$P(X = x)$$

Aka the PMF

"The ***derivative*** of the probability that a **continuous** random variable X takes on the value little *x*."

$$f(X = x)$$

Aka the PDF

*They are both measures of how **likely** X is to take on the value x.*