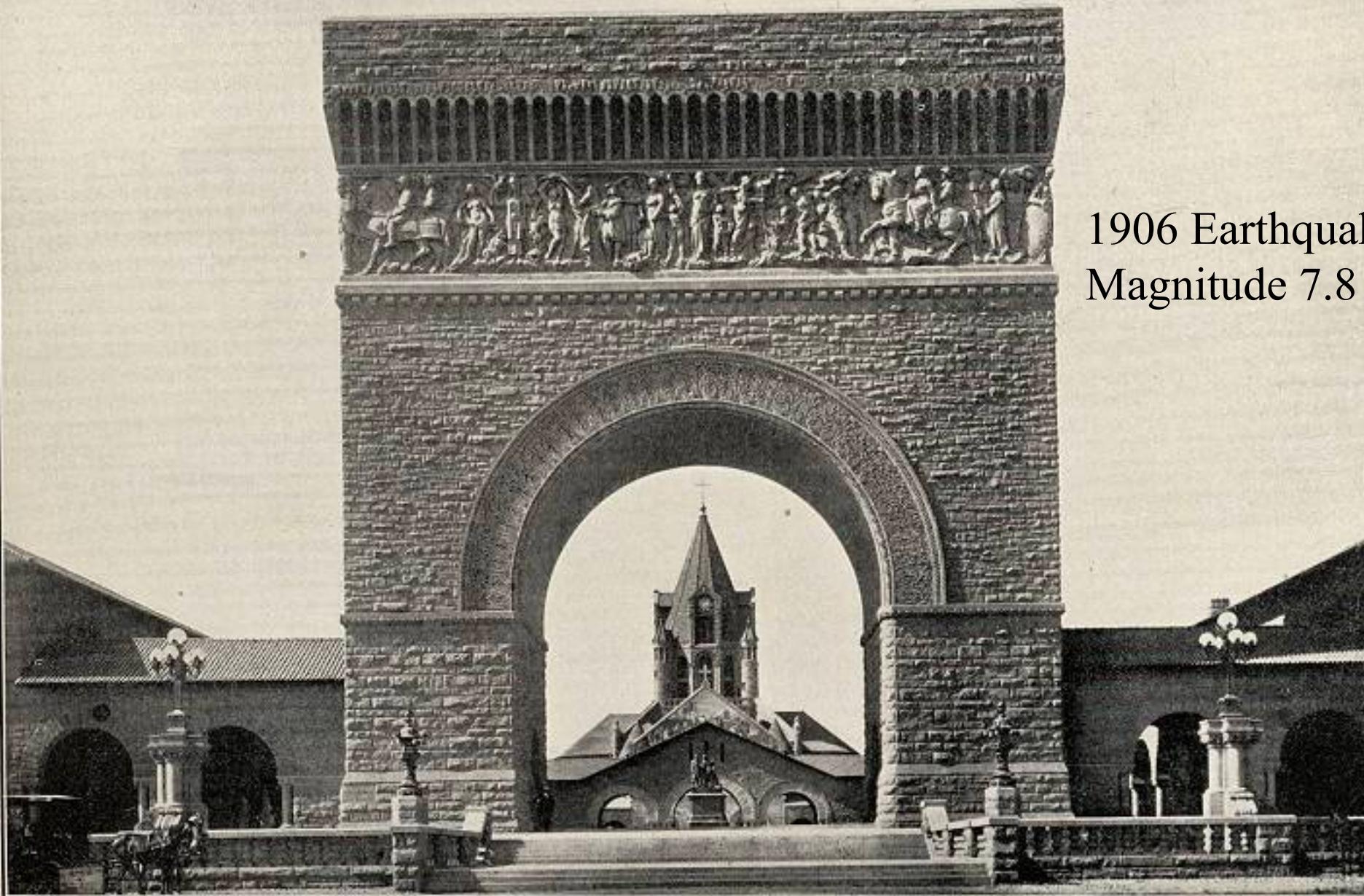




Continuous Variables

Chris Piech
CS109, Stanford University

1906 Earthquake
Magnitude 7.8



ILL. No. 65. MEMORIAL ARCH, WITH CHURCH IN BACKGROUND, STANFORD UNIVERSITY, SHOWING TYPES OF CARVED WORK WITH THE SANDSTONE.

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



Discrete Distributions

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

Grid of Random Variables

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

Geometric Random Variable

- X is Geometric Random Variable: $X \sim \text{Geo}(p)$
 - 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:

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

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



Negative Binomial Random Variable

- X is Negative Binomial RV: $X \sim \text{NegBin}(r, p)$
 - X is number of independent trials until r successes
 - p is probability of success on each trial
 - X takes on values $r, r + 1, r + 2\dots$, with probability:

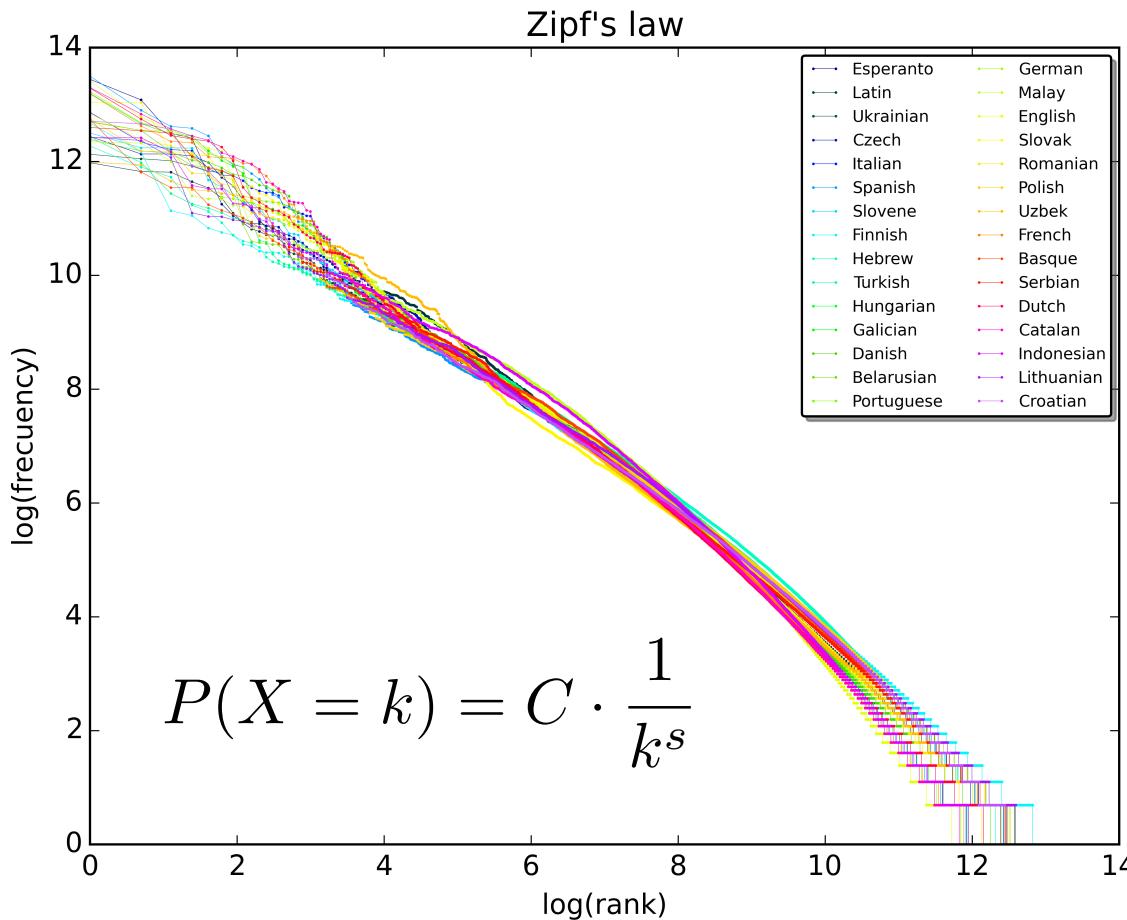
$$P(X = n) = \binom{n-1}{r-1} p^r (1-p)^{n-r}, \text{ where } n = r, r+1, \dots$$

- Note: $\text{Geo}(p) \sim \text{NegBin}(1, p)$



Zipf Random Variable

- X is Zipf RV: $X \sim \text{Zipf}(s)$
 - X is the rank index of a chosen word



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 $X \sim \text{NegBin}(r, p)$

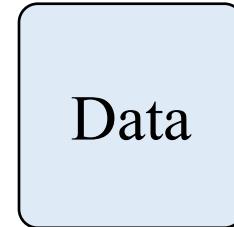
Zipf:

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

Bit Coin Mining



SHA-256 Hash(



)



Number that looks like random bits



You “mine a bitcoin” if, for given data D , you find a number N such that $\text{Hash}(D, N)$ 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 $\text{Hash}(D, N)$ 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

Supreme Court case: Berghuis v. Smith

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

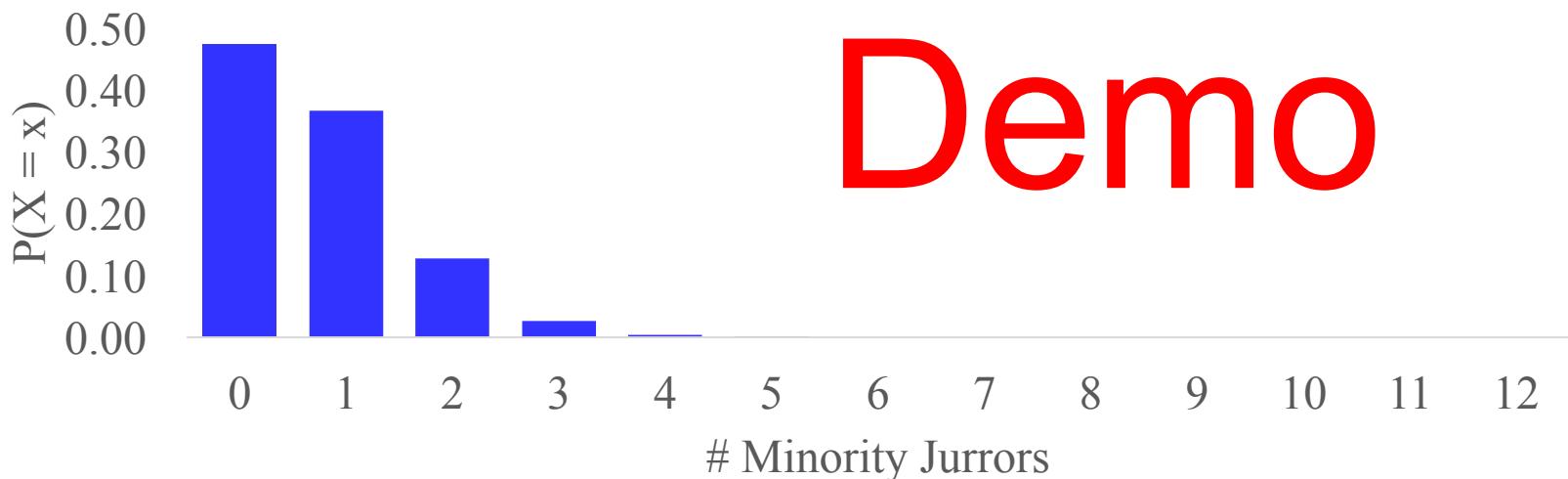
- Article by Erin Miller –January 22, 2010
- Thanks to (former CS109er) Josh Falk for this article

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 a third to a half of juries would have at least one minority person**” on them.

Justin Breyer Meets CS109

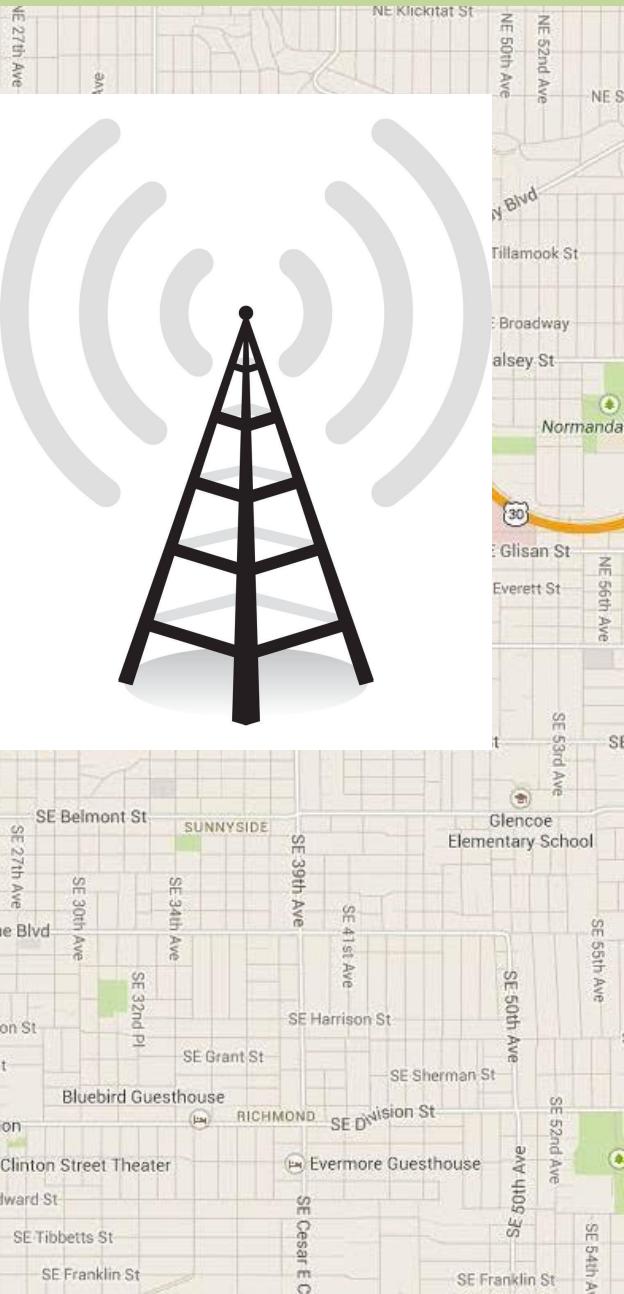
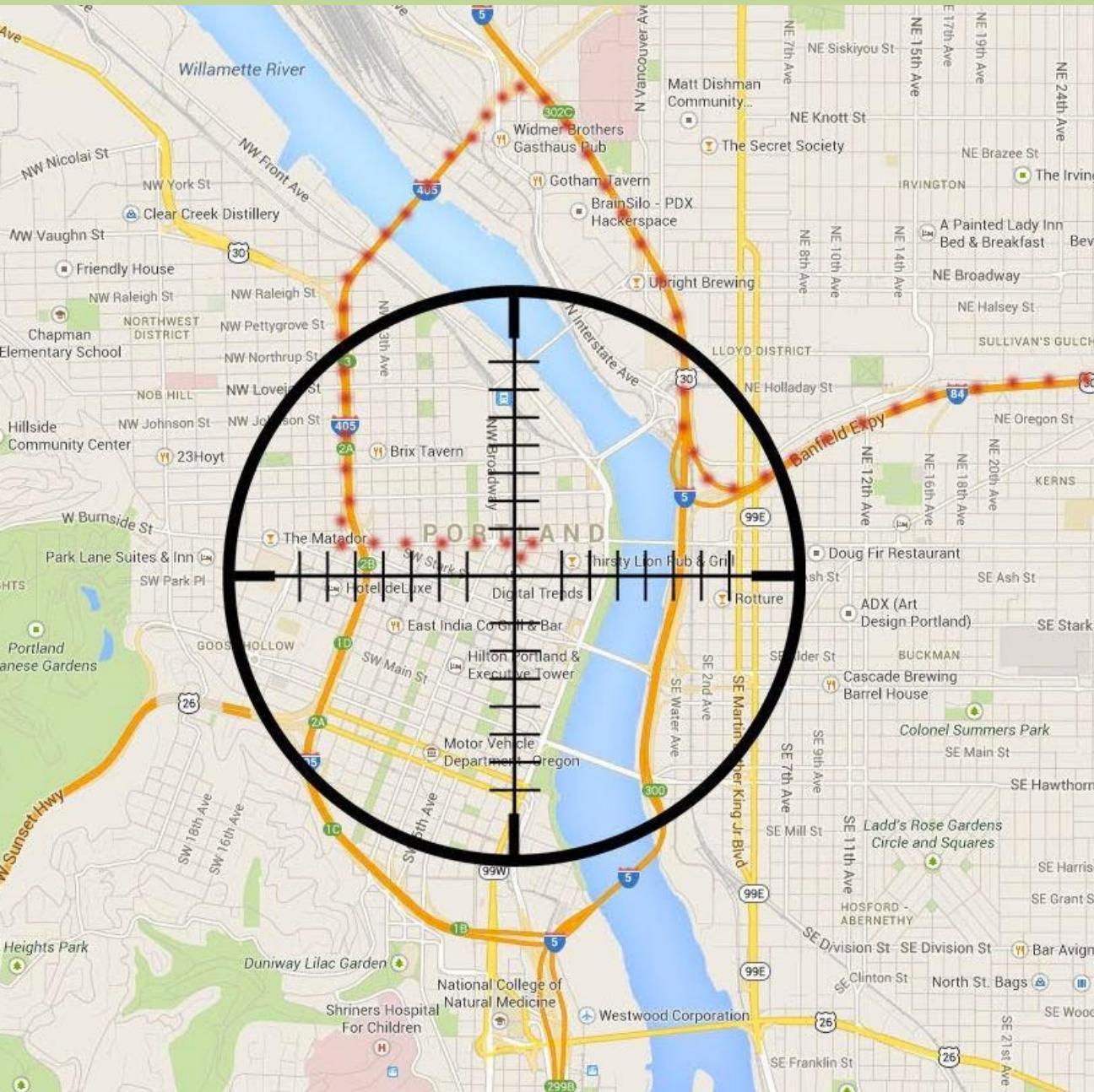
- Approximation using Binomial distribution
 - Assume $P(\text{blue ball})$ constant for every draw = 60/1000
 - $X = \# \text{ blue balls drawn. } X \sim \text{Bin}(12, 60/1000 = 0.06)$
 - $P(X \geq 1) = 1 - P(X = 0) \approx 1 - 0.4759 = 0.5240$

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



Big hole in our knowledge

Not all values are discrete



random() ?

Riding the Marguerite



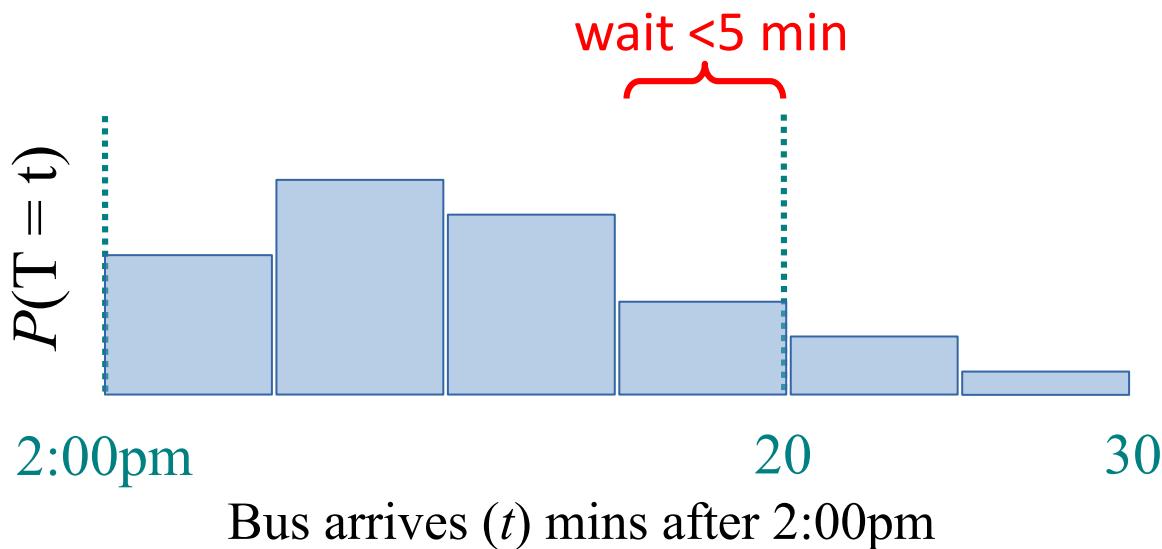
Riding the Marguerite



*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 $P(\text{wait} < 5 \text{ minutes})$?



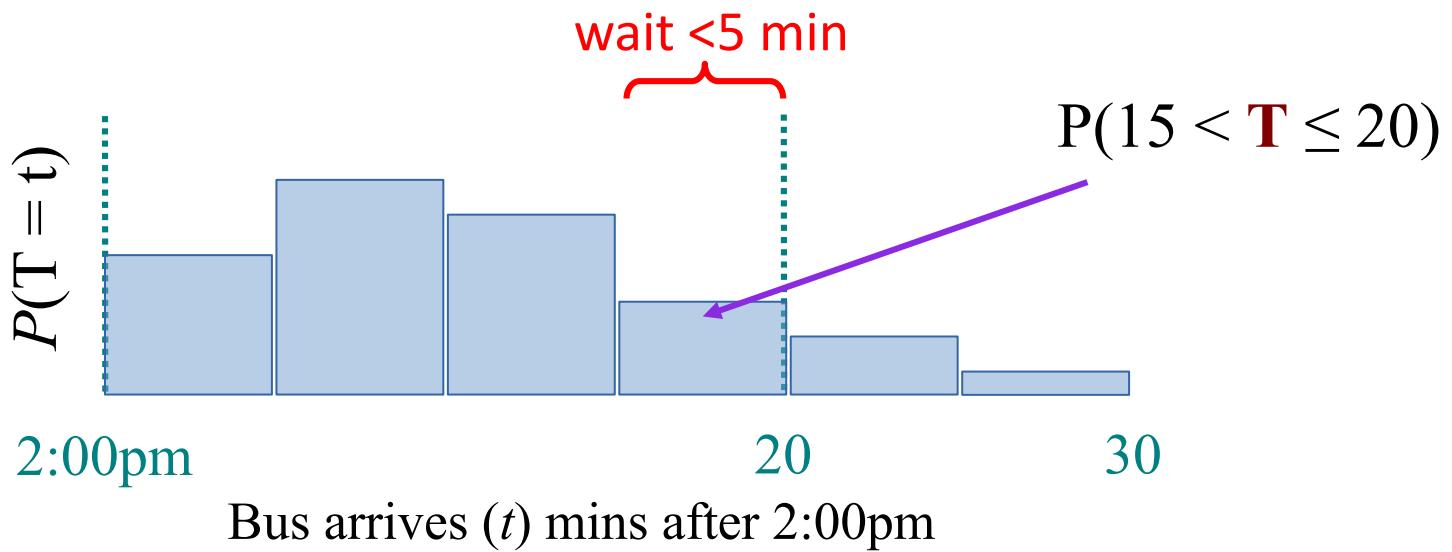
Riding the Marguerite



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 2:20pm.

What is $P(\text{wait} < 5 \text{ minutes})$?



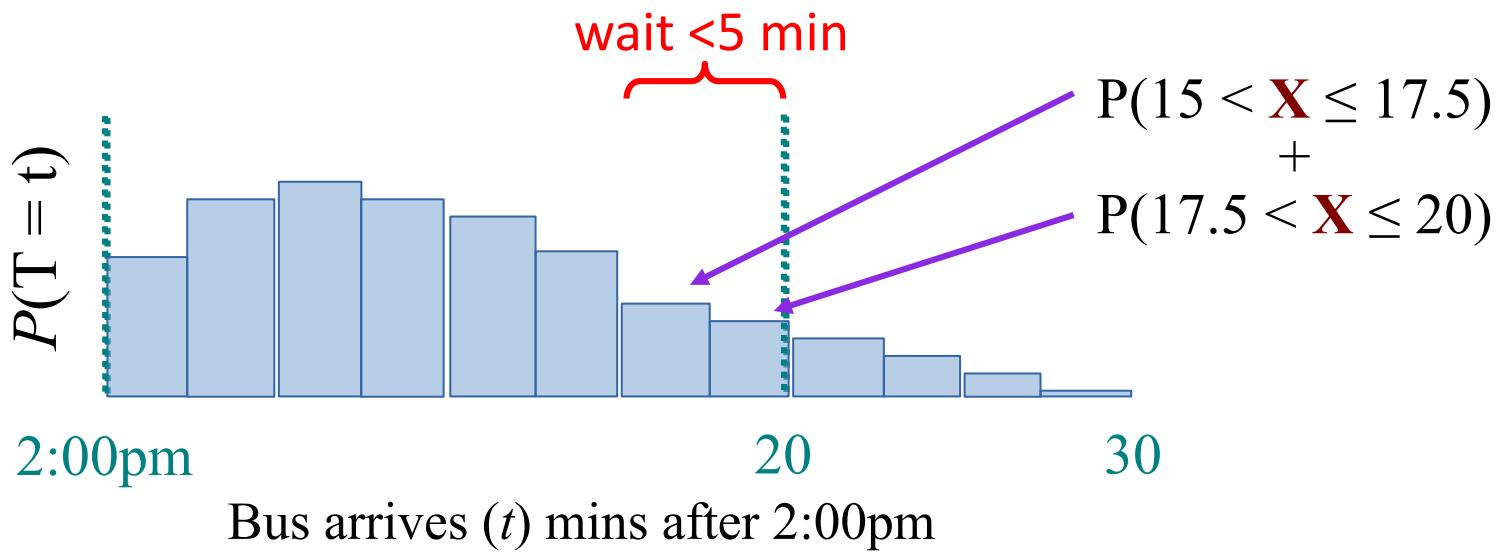
Riding the Marguerite



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 2:20pm.

What is $P(\text{wait} < 5 \text{ minutes})$?



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 2:20pm.

What is $P(\text{wait} < 5 \text{ minutes})$?

Probability
Density Function

$$f_T(t)$$

2:00pm

Bus arrives (t) mins after 2:00pm

wait <5 min

$$P(15 < T \leq 20)$$

20

30

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!

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

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

A probability!

Properties of PDFs

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

$$\int_{-\infty}^{\infty} f_X(x) \, dx = 1$$

$$P(a - \frac{\varepsilon}{2} \leq X \leq a + \frac{\varepsilon}{2}) = \int_{a-\varepsilon/2}^{a+\varepsilon/2} f(x) dx \approx \varepsilon f(a)$$

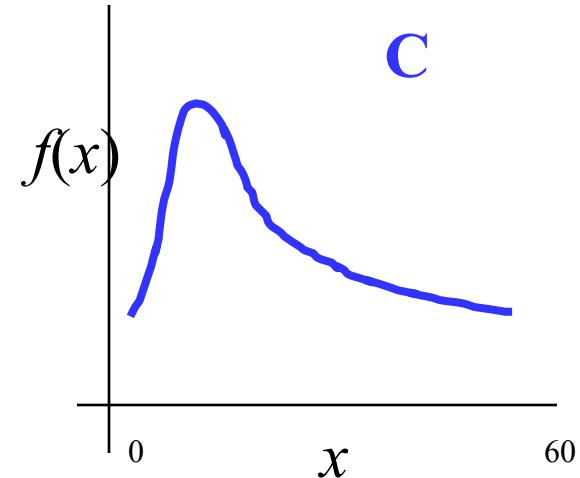
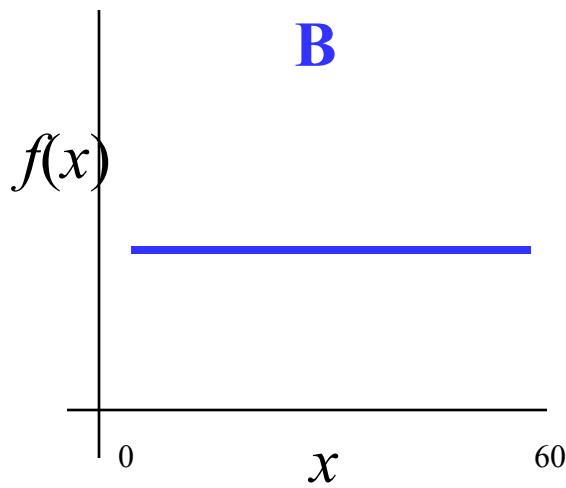
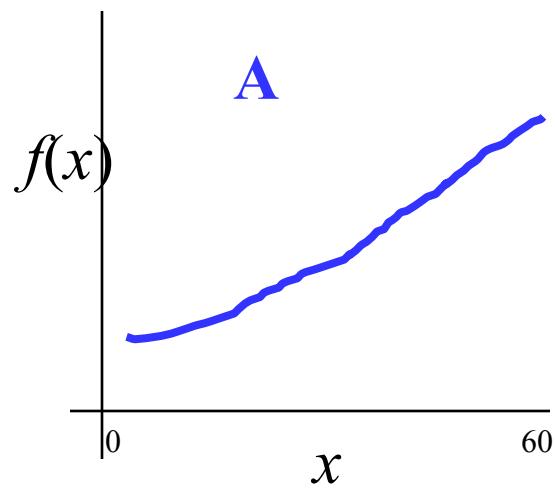
So, ratio of density can still be meaningful:

$$P(X = 1)/P(X = 2) \approx (\varepsilon f(1))/(\varepsilon f(2)) = f(1)/f(2)$$

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

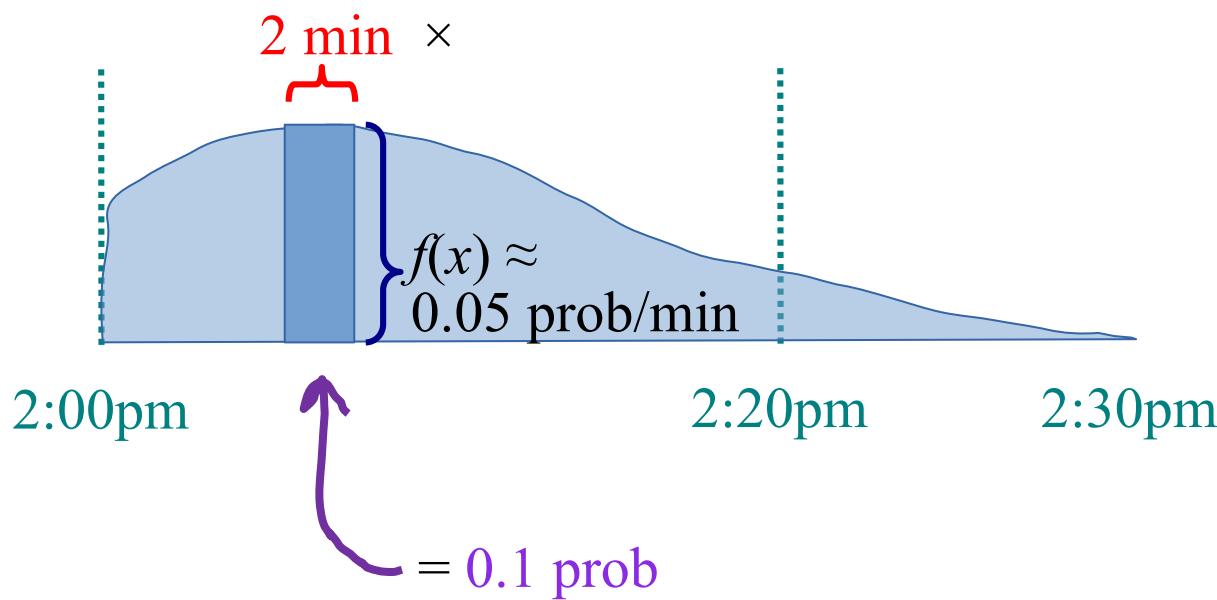
Integrals



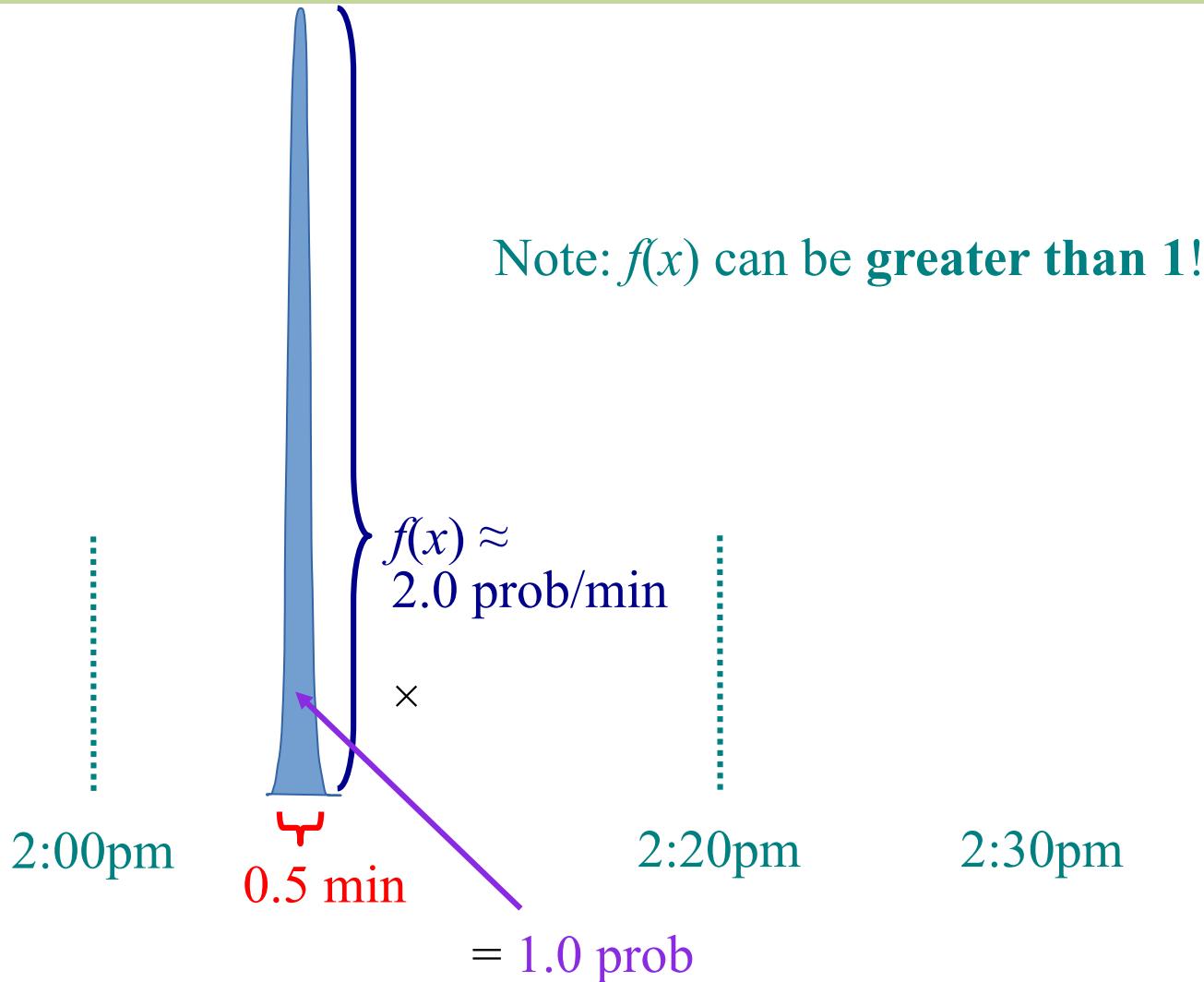
*loving, not scary

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

Rather, it has “units” of probability
divided by units of X .



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



Simple Example



Consider a random 5000×5000 matrix, where each element in the matrix is $\text{Uniform}(0,1)$. What is the probability that a selected eigenvalue (λ) of the matrix is greater than 0?*

* With help from Wigner, Chris is going to rephrase this problem

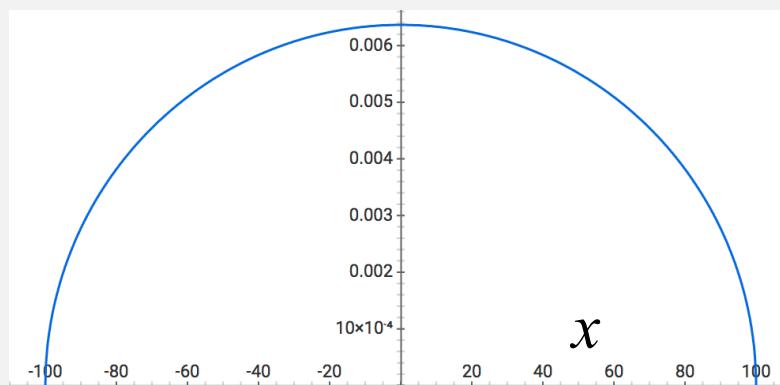
Simple Example from Quantum Physics

Let X be a continuous random variable:

Theory

$$f(x) = \frac{1}{15708} \sqrt{100^2 - x^2}$$

$f(x)$



Practice

From simulations

$p(x)$

0.006
0.005
0.004
0.003
0.002
0.001

x

-50 0 50 100

$$P(X > 0) = ?$$

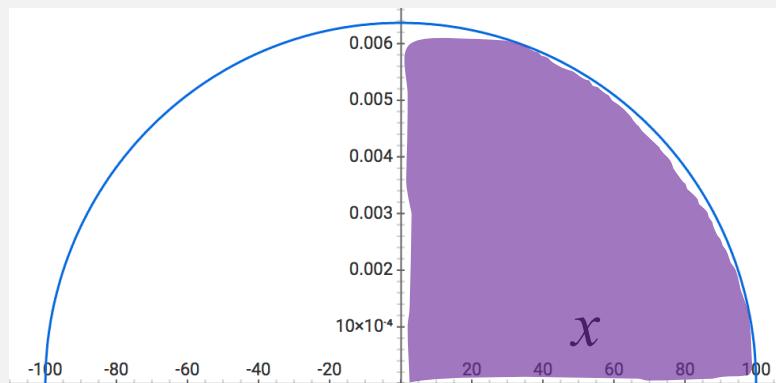
Simple Example from Quantum Physics

Let X be a continuous random variable:

Theory

$$f(x) = \frac{1}{15708} \sqrt{100^2 - x^2}$$

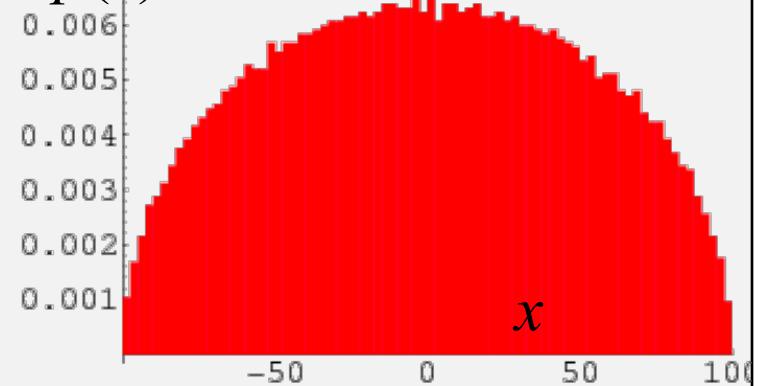
$f(x)$



Practice

From simulations

$p(x)$



Approach #1: Integrate over the PDF

$$P(X > 0) = \int_0^{100} f_X(x) dx$$

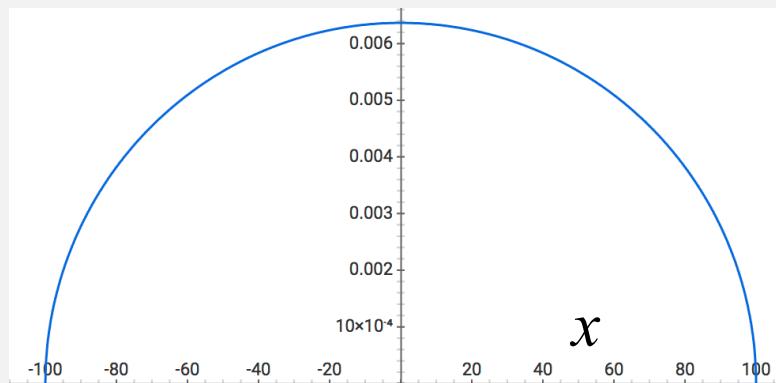
Simple Example from Quantum Physics

Let X be a continuous random variable:

Theory

$$f(x) = \frac{1}{15708} \sqrt{100^2 - x^2}$$

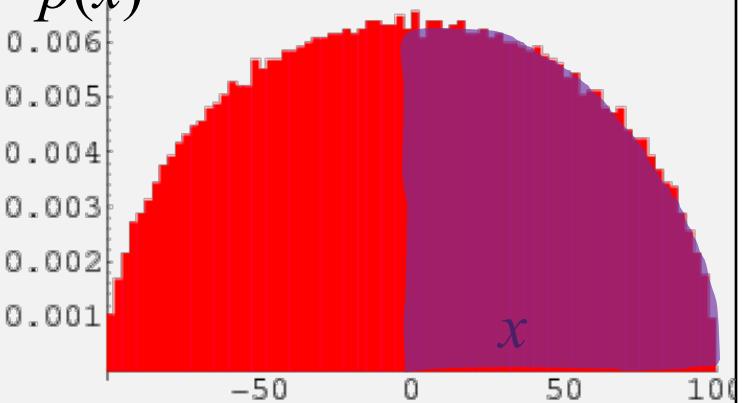
$f(x)$



Practice

From simulations

$p(x)$



Approach #2: Discrete Approximation

$$P(X > 0) \approx \sum_{i=0}^{100} P(X = i)$$

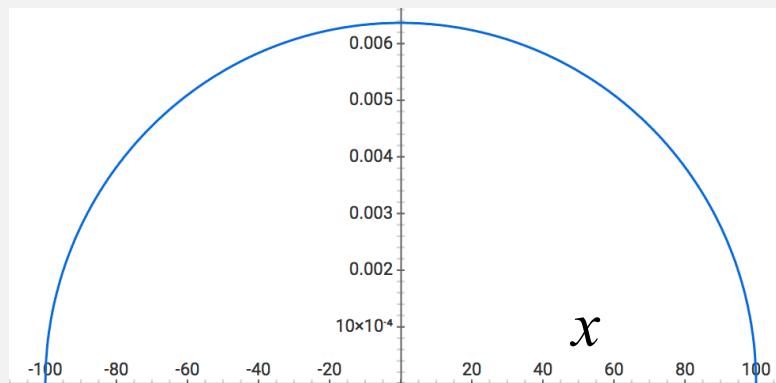
Simple Example from Quantum Physics

Let X be a continuous random variable:

Theory

$$f(x) = \frac{1}{15708} \sqrt{100^2 - x^2}$$

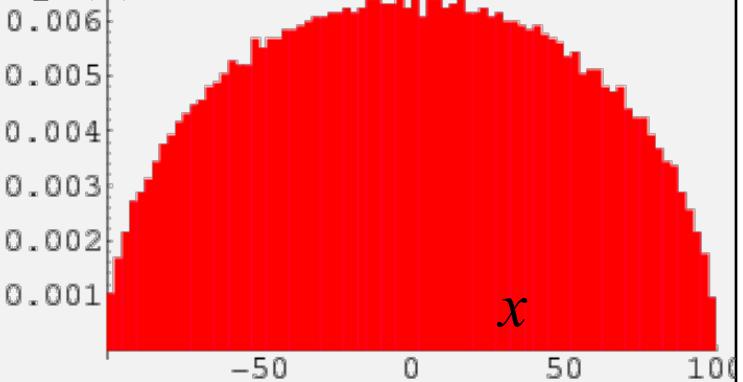
$f(x)$



Practice

From simulations

$p(x)$



Approach #3: Know Semi-Circles

$$P(X > 0) = \frac{1}{2}$$

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

A probability!

Uniform Random Variable

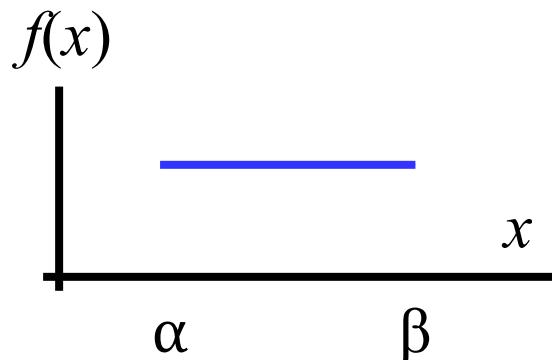
A **uniform** random variable is **equally likely** to be any value in an interval.

$$X \sim \text{Uni}(\alpha, \beta)$$



Probability Density

$$f(x) = \begin{cases} \frac{1}{\beta - \alpha} & \alpha \leq x \leq \beta \\ 0 & \text{otherwise} \end{cases}$$



Properties

$$E[X] = \frac{\beta - \alpha}{2}$$

$$\text{Var}(X) = \frac{(\beta - \alpha)^2}{12}$$

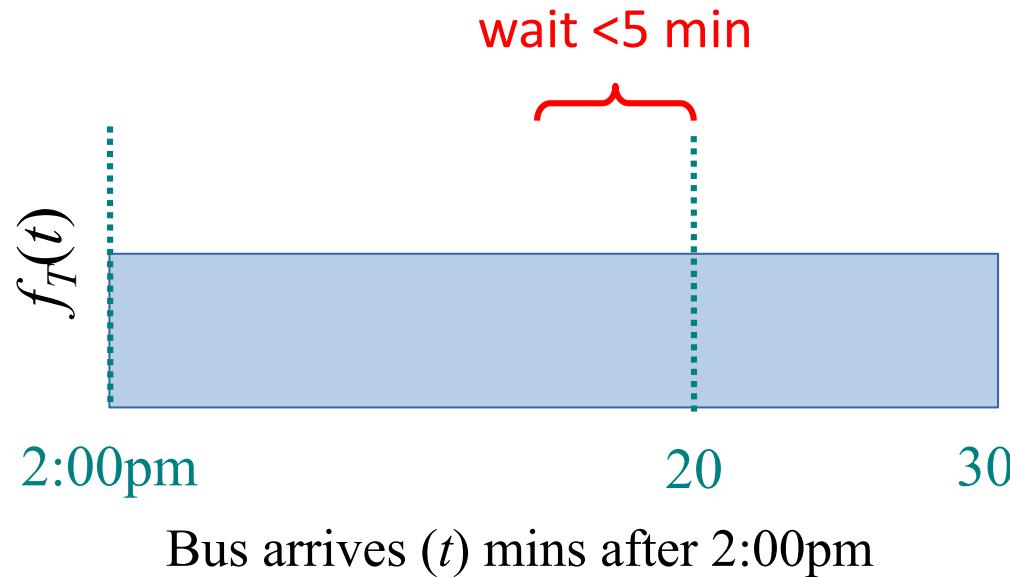
Uniform Bus



You are running to the bus stop. You don't know exactly when the bus arrives. **You believe all times between 2 and 2:30 are equally likely.**

You show up at 2:20pm. What is $P(\text{wait} < 5 \text{ minutes})$?

$$T \sim \text{Uni}(\alpha = 0, \beta = 30)$$



$$\begin{aligned} P(\text{Wait} < 5) &= \int_{15}^{20} \frac{1}{\beta - \alpha} dx \\ &= \left. \frac{x}{\beta - \alpha} \right|_{15}^{20} \\ &= \left. \frac{x}{30 - 0} \right|_{15}^{20} = \frac{5}{30} \end{aligned}$$

Expectation and Variance

For discrete RV X :

$$E[X] = \sum_x x p(x)$$

$$E[g(X)] = \sum_x g(x) p(x)$$

$$E[X^n] = \sum_x x^n p(x)$$

For continuous RV X :

$$E[X] = \int_{-\infty}^{\infty} x f(x) dx$$

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x) dx$$

$$E[X^n] = \int_{-\infty}^{\infty} x^n f(x) dx$$

For both discrete and continuous RVs:

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

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

$$\text{Var}(aX + b) = a^2 \text{Var}(X)$$

Expectation of Uniform

$$X \sim \text{Uni}(\alpha, \beta)$$

$$E[X] = \int_{-\infty}^{\infty} x \cdot f(x) \, dx$$

$$= \int_{\alpha}^{\beta} x \cdot \frac{1}{\beta - \alpha} \, dx$$

$$= \frac{1}{\beta - \alpha} \left[\frac{1}{2} x^2 \right]_{\alpha}^{\beta}$$

$$= \frac{1}{2} \frac{1}{\beta - \alpha} (\beta + \alpha)(\beta - \alpha)$$

just average
the start and
end!

$$= \frac{1}{2} (\alpha + \beta)$$

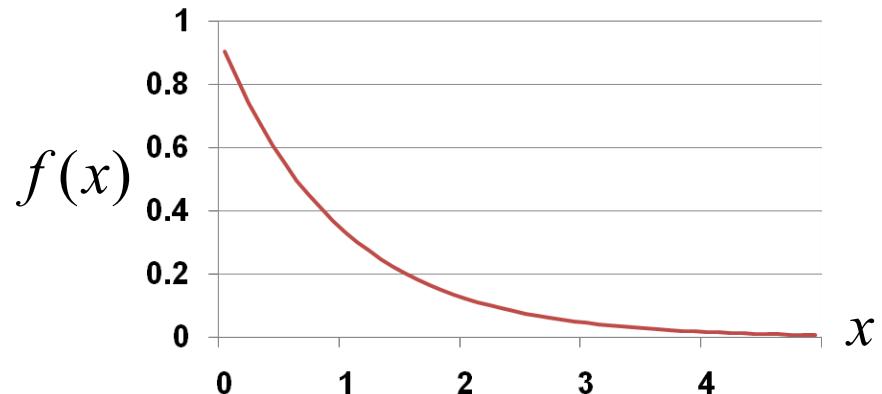
Exponential Random Variable

- X is an Exponential RV: $X \sim \text{Exp}(\lambda)$ Rate: $\lambda > 0$

- Probability Density Function (PDF):

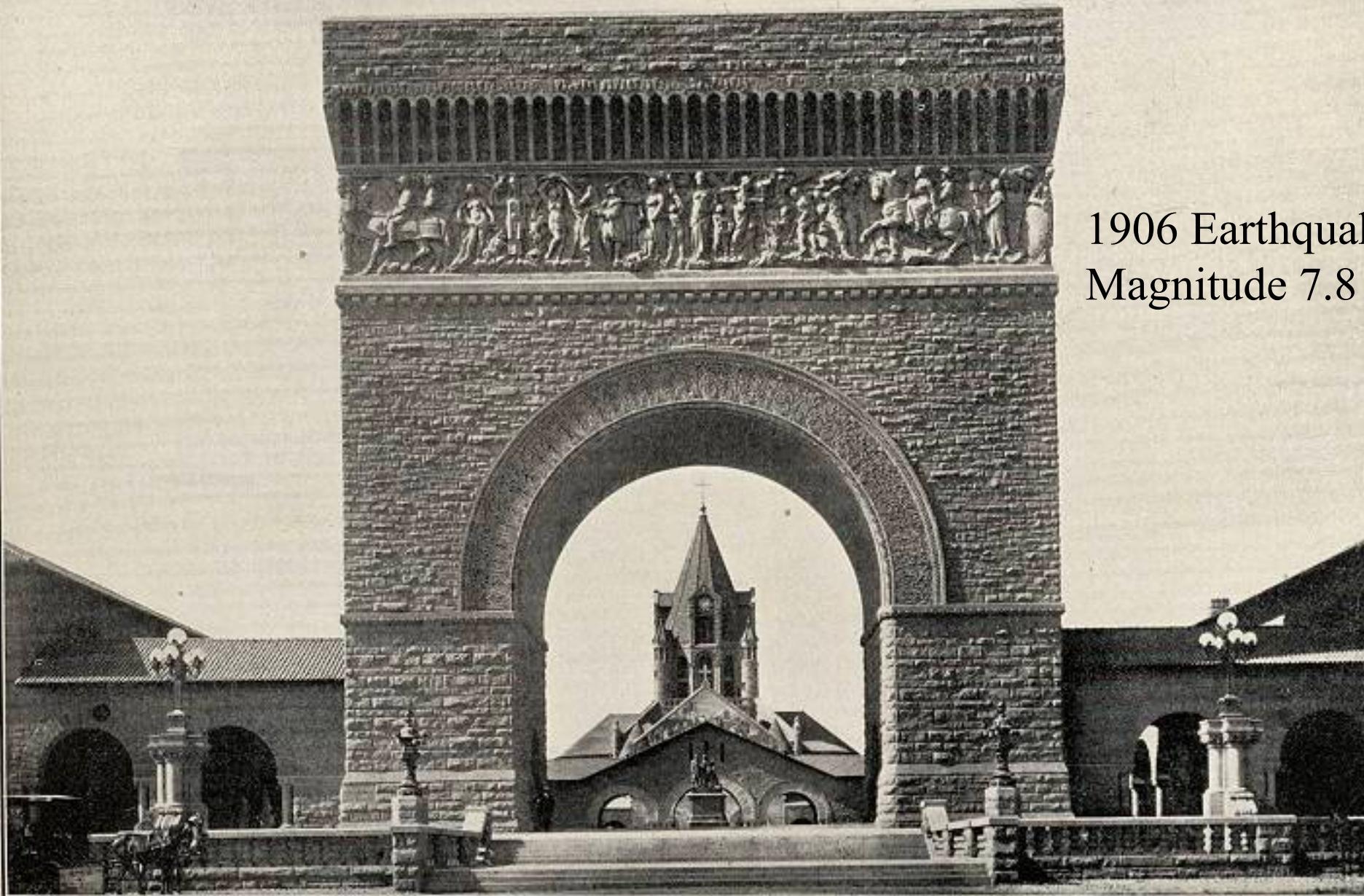
$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases} \quad \text{where } -\infty < x < \infty$$

- $E[X] = \frac{1}{\lambda}$
 - $Var(X) = \frac{1}{\lambda^2}$



- Represents time until some event
 - Earthquake, request to web server, end cell phone contract, etc.

1906 Earthquake
Magnitude 7.8



ILL. No. 65. MEMORIAL ARCH, WITH CHURCH IN BACKGROUND, STANFORD UNIVERSITY, SHOWING TYPES OF CARVED WORK WITH THE SANDSTONE.

Integral Review

$$\int e^{cx} dx = \frac{1}{c} e^{cx}$$

How Long Until Next Earthquake

Based on historical data, earthquakes of magnitude 8.0+ happen at a **rate of 0.002** per year*. What is the probability of an earthquake of magnitude 8+ in the next 30 years?

Y = Years until the next earthquake of magnitude 8.0+

$$\begin{aligned} Y &\sim \text{Exp}(\lambda = 0.002) & f_Y(y) &= \lambda e^{-\lambda y} \\ &&&= 0.002^{-0.002y} \\ P(Y < 30) &= \int_0^{30} 0.002e^{-0.002y} dy \\ &= 0.002 \left[-500e^{-0.002y} \right]_0^{30} \\ &= \frac{500}{500} (-e^{-0.06} + e^0) && \approx 0.06 \end{aligned}$$

*In California, according to the USGS, 2015

Is there a way to avoid
integrals?

Cumulative Density Function

A cumulative density function (CDF) is a “closed form” equation for the probability that a random variable is less than a given value

$$F(x) = P(X < x)$$



If you learn how to use a cumulative density function, you can avoid integrals!

$$F_X(x)$$

This is also shorthand notation for the PMF

Cumulative Density Function

$$F(x) = P(X < x)$$

$$x = 2$$

$$0.03125$$

CDF of an Exponential

$$F_X(x) = 1 - e^{-\lambda x}$$

$$P(X < x) = \int_{y=-\infty}^x f(y) \, dy$$

$$= \int_{y=0}^x \lambda e^{-\lambda y} \, dy$$

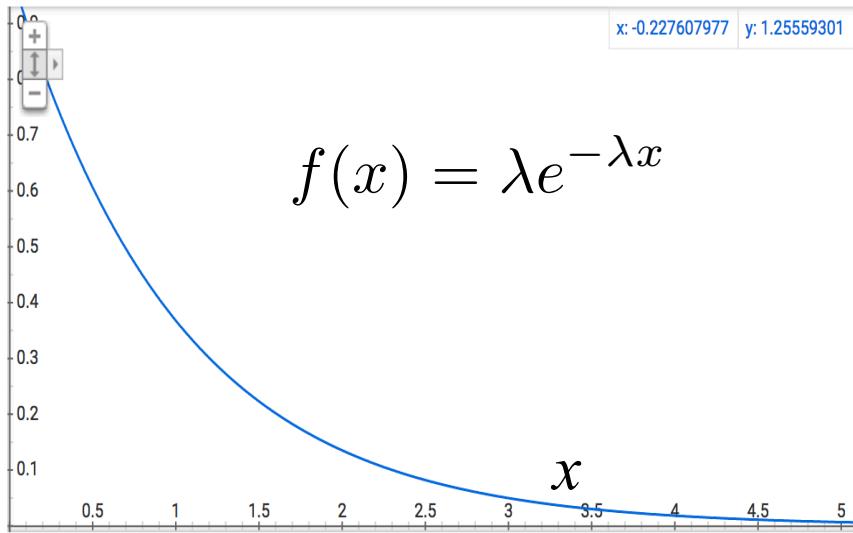
$$= \frac{\lambda}{\lambda} \left[-e^{-\lambda y} \right]_0^x$$

$$= [-e^{-\lambda x}] - [-e^{\lambda 0}]$$

$$= 1 - e^{-\lambda x}$$

CDF: $X \sim \text{Exp}(\lambda = 1)$

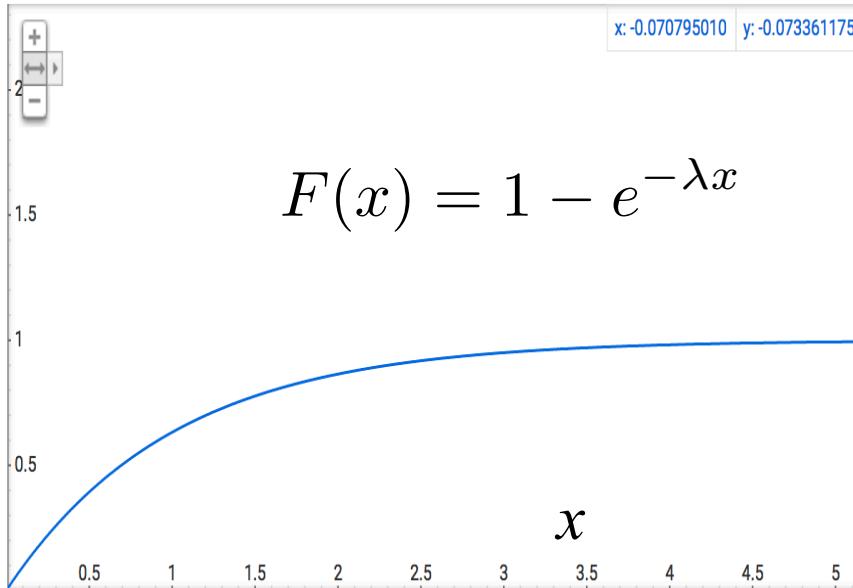
*Probability
density
function*



*Cumulative
density
function*

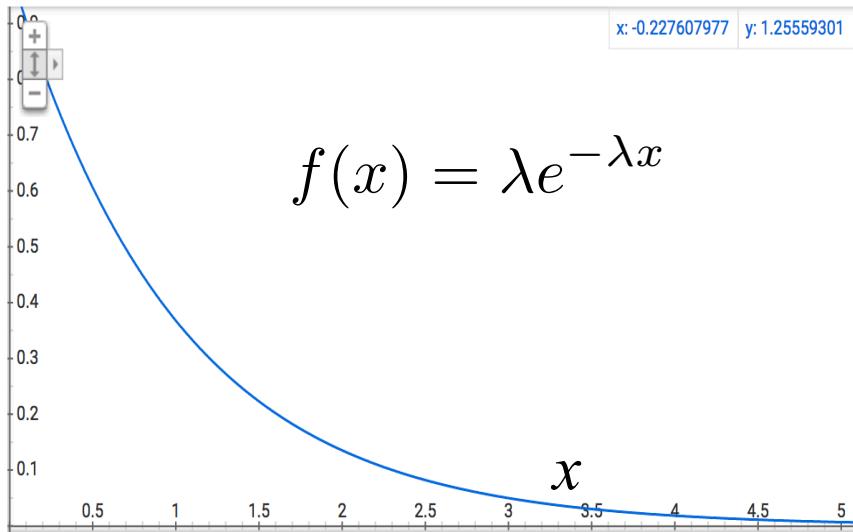
$$F_X(x) = P(X < x)$$

$$= \int_{y=-\infty}^x f(y) dy$$



CDF: $X \sim \text{Exp}(\lambda = 1)$

Probability
density
function

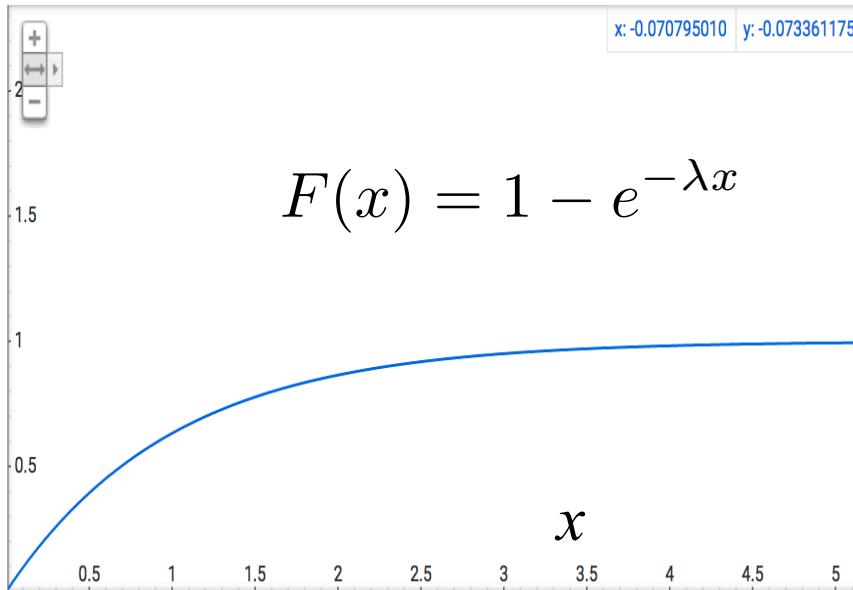


$$P(X < 2)$$

Cumulative
density
function

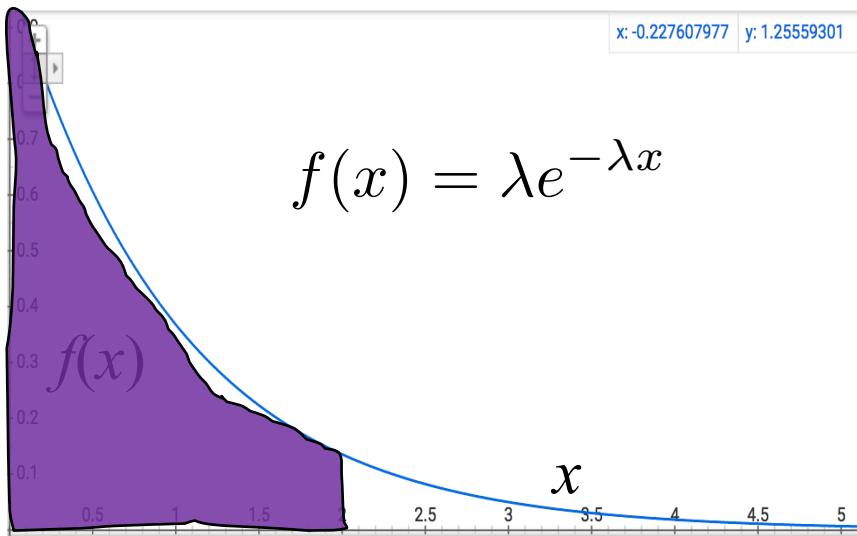
$$F_X(x) = P(X < x)$$

$$= \int_{y=-\infty}^x f(y) dy$$



CDF: $X \sim \text{Exp}(\lambda = 1)$

Probability
density
function



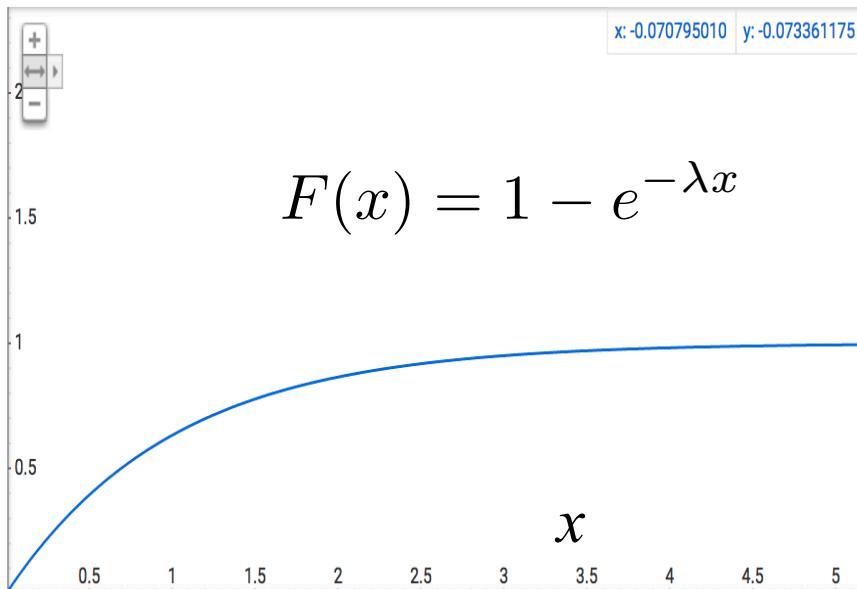
$$P(X < 2)$$

$$= \int_{x=-\infty}^2 f(x) dx$$

Cumulative
density
function

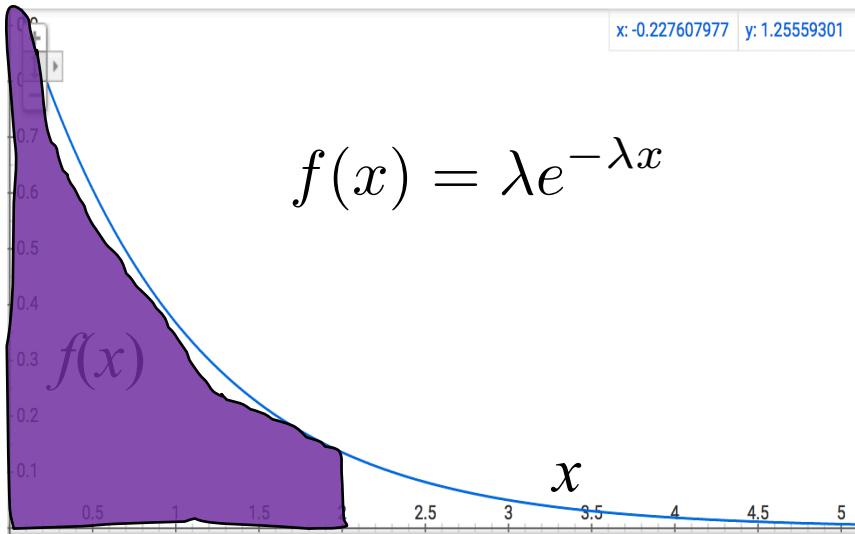
$$F_X(x) = P(X < x)$$

$$= \int_{y=-\infty}^x f(y) dy$$



CDF: $X \sim \text{Exp}(\lambda = 1)$

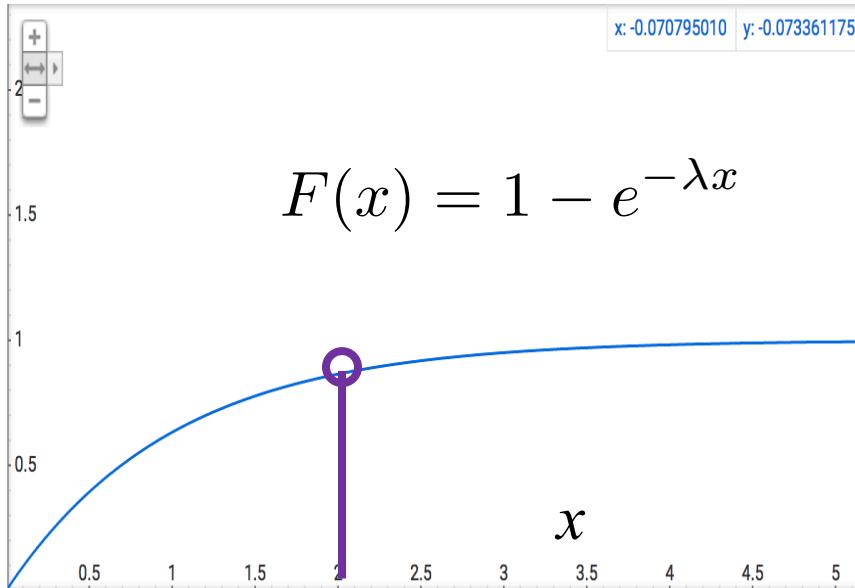
Probability
density
function



Cumulative
density
function

$$F_X(x) = P(X < x)$$

$$= \int_{y=-\infty}^x f(y) dy$$



$$P(X < 2)$$

$$= \int_{x=-\infty}^2 f(x) dx$$

or

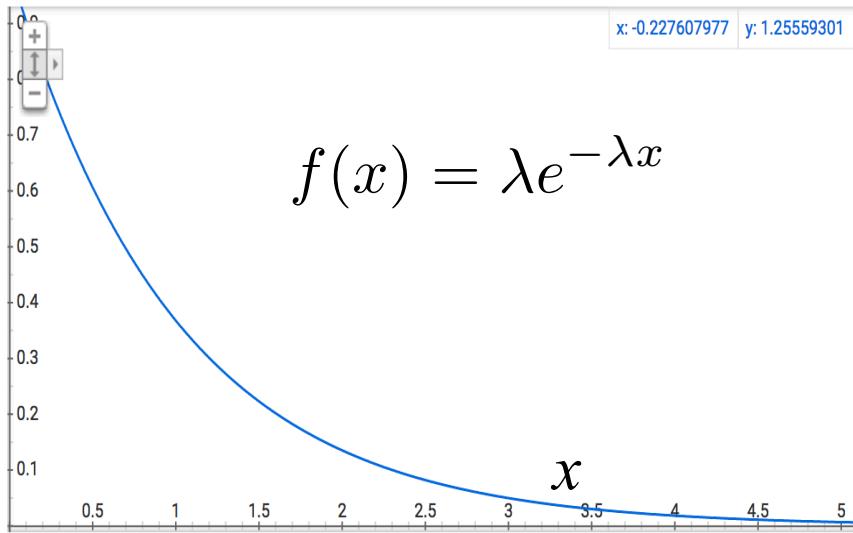
$$= F(2)$$

$$= 1 - e^{-2}$$

$$\approx 0.84$$

CDF: $X \sim \text{Exp}(\lambda = 1)$

Probability
density
function

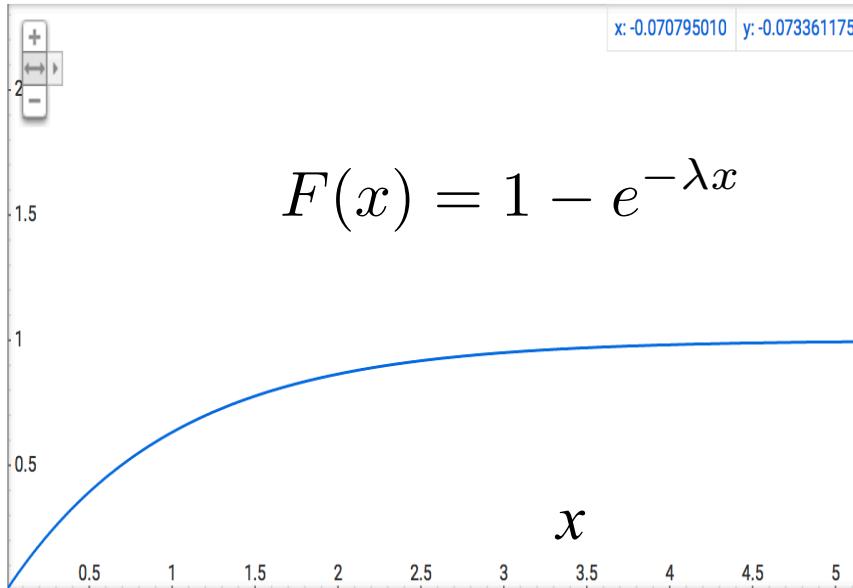


$$P(X > 1)$$

Cumulative
density
function

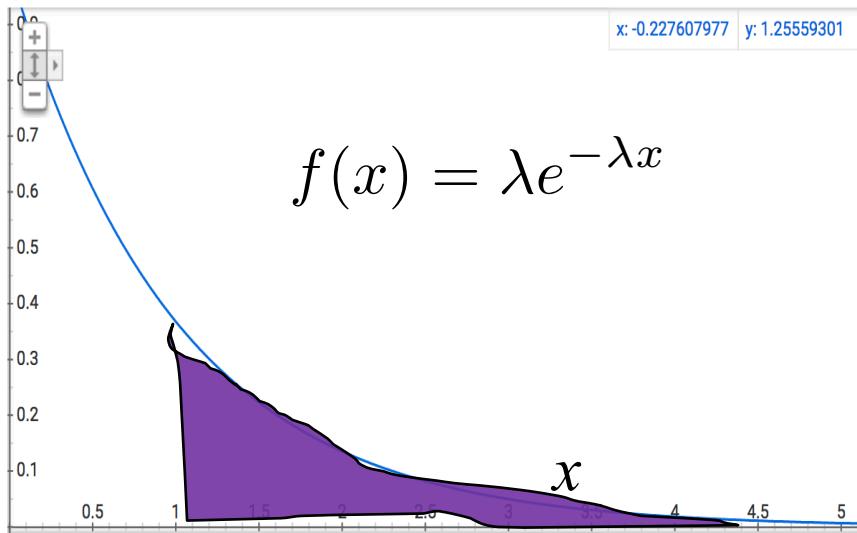
$$F_X(x) = P(X < x)$$

$$= \int_{y=-\infty}^x f(y) dy$$



CDF: $X \sim \text{Exp}(\lambda = 1)$

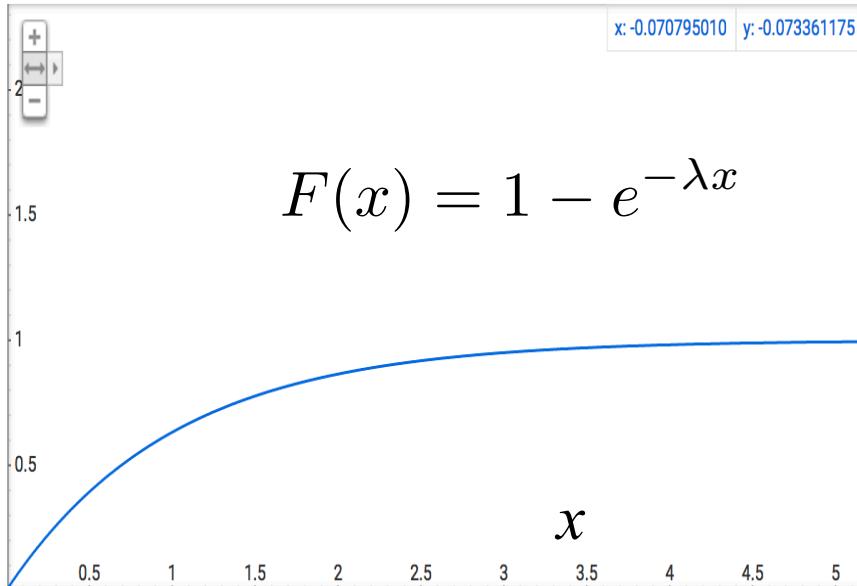
Probability
density
function



Cumulative
density
function

$$F_X(x) = P(X < x)$$

$$= \int_{y=-\infty}^x f(y) dy$$

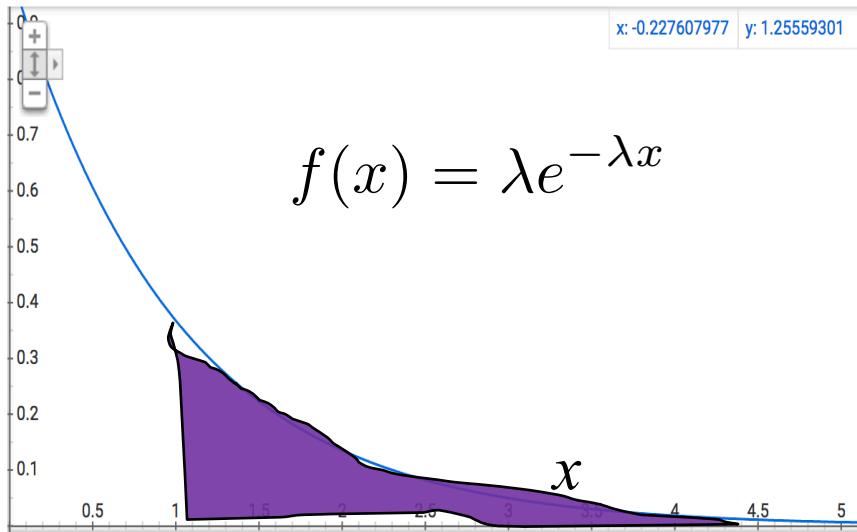


$$P(X > 1)$$

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

CDF: $X \sim \text{Exp}(\lambda = 1)$

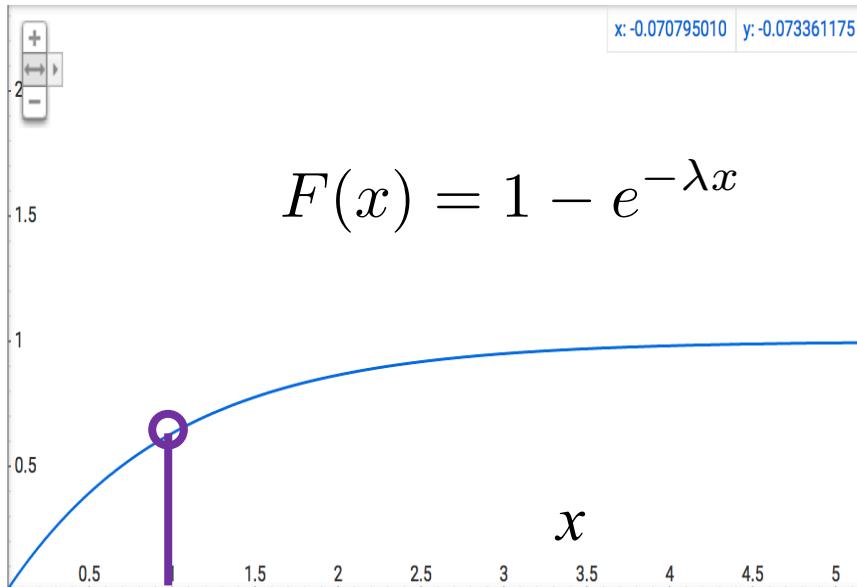
Probability
density
function



Cumulative
density
function

$$F_X(x) = P(X < x)$$

$$= \int_{y=-\infty}^x f(y) dy$$



$$P(X > 1)$$

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

or

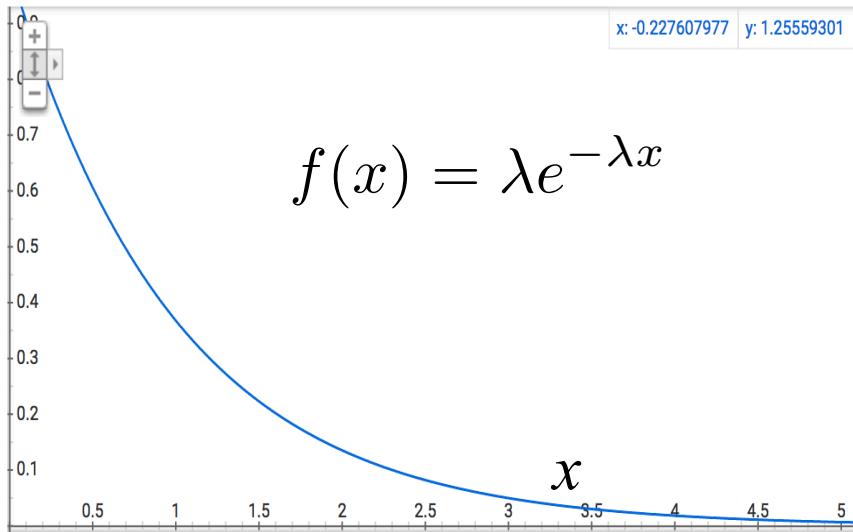
$$= 1 - F(1)$$

$$= e^{-1}$$

$$\approx 0.37$$

CDF: $X \sim \text{Exp}(\lambda = 1)$

Probability
density
function

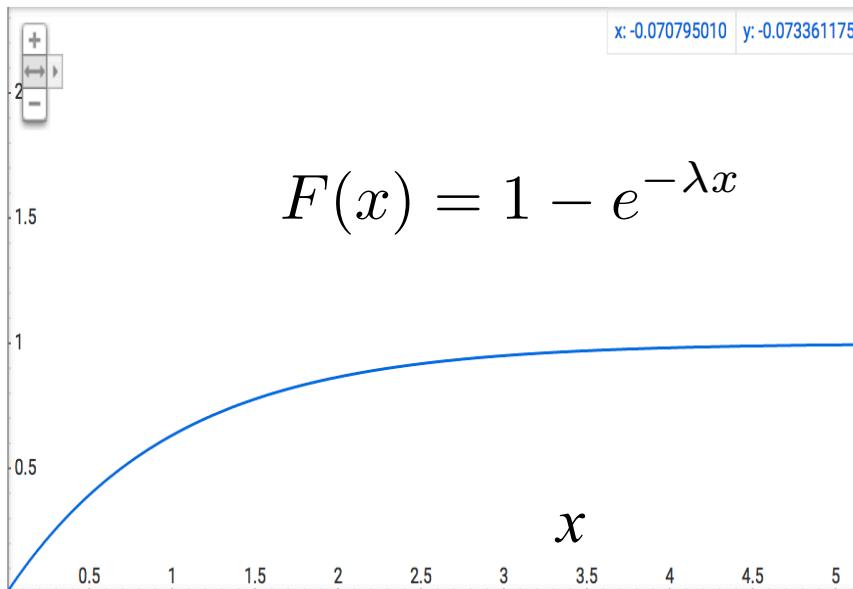


$$P(1 < X < 2)$$

Cumulative
density
function

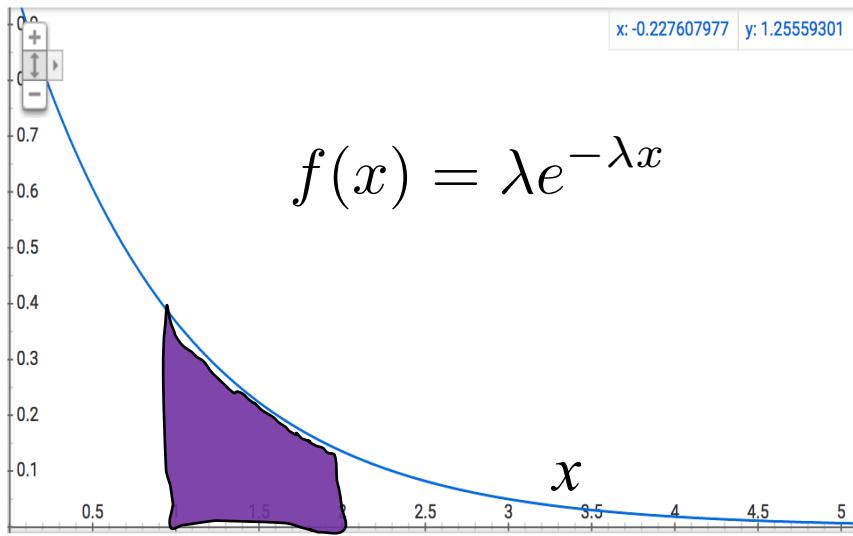
$$F_X(x) = P(X < x)$$

$$= \int_{y=-\infty}^x f(y) dy$$



CDF: $X \sim \text{Exp}(\lambda = 1)$

Probability
density
function



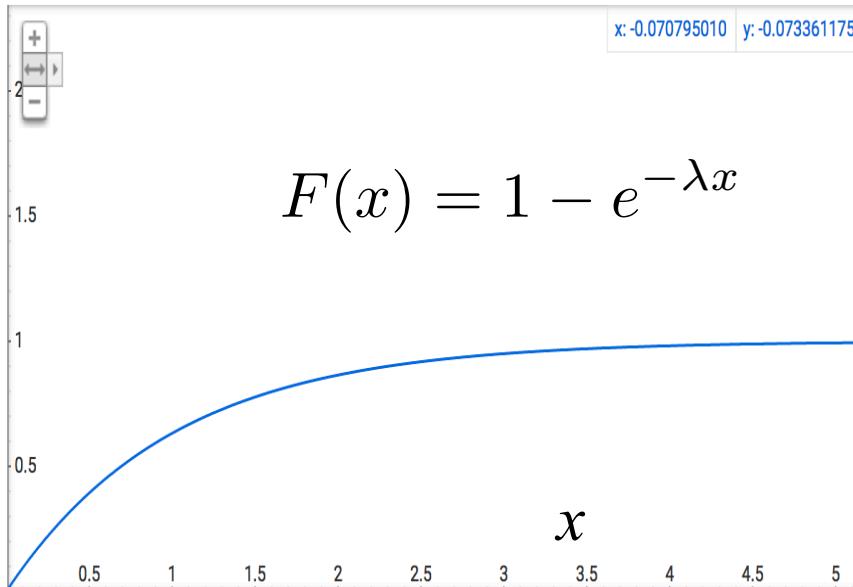
$$P(1 < X < 2)$$

$$= \int_{x=1}^2 f(x) \, dx$$

Cumulative
density
function

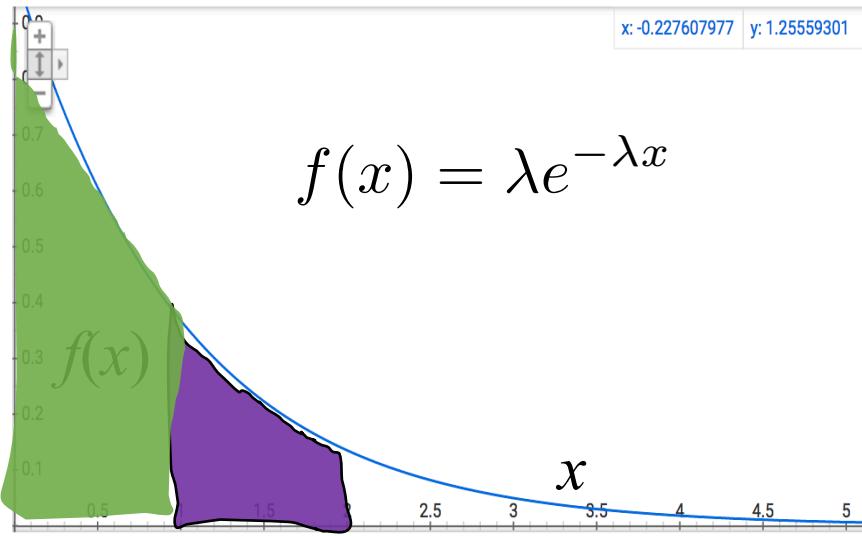
$$F_X(x) = P(X < x)$$

$$= \int_{y=-\infty}^x f(y) \, dy$$



CDF: $X \sim \text{Exp}(\lambda = 1)$

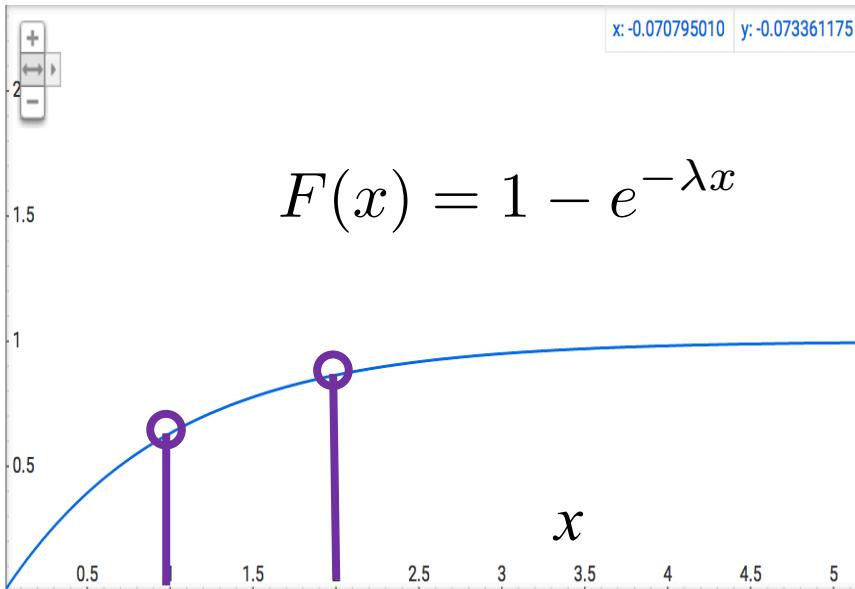
Probability
density
function



Cumulative
density
function

$$F_X(x) = P(X < x)$$

$$= \int_{y=-\infty}^x f(y) dy$$



$$P(1 < X < 2)$$

$$= \int_{x=1}^2 f(x) dx$$

or

$$= F(2) - F(1)$$

$$\begin{aligned} &= (1 - e^{-2}) \\ &\quad - (1 - e^{-1}) \\ &\approx 0.23 \end{aligned}$$

Probability of Earthquake in Next 4 Years?

Based on historical data, earthquakes of magnitude 8.0+ happen at a **rate of 0.002** per year*. What is the probability of an earthquake of magnitude 8+ in the next **4** years?

Y = Years until the next earthquake of magnitude 8.0+

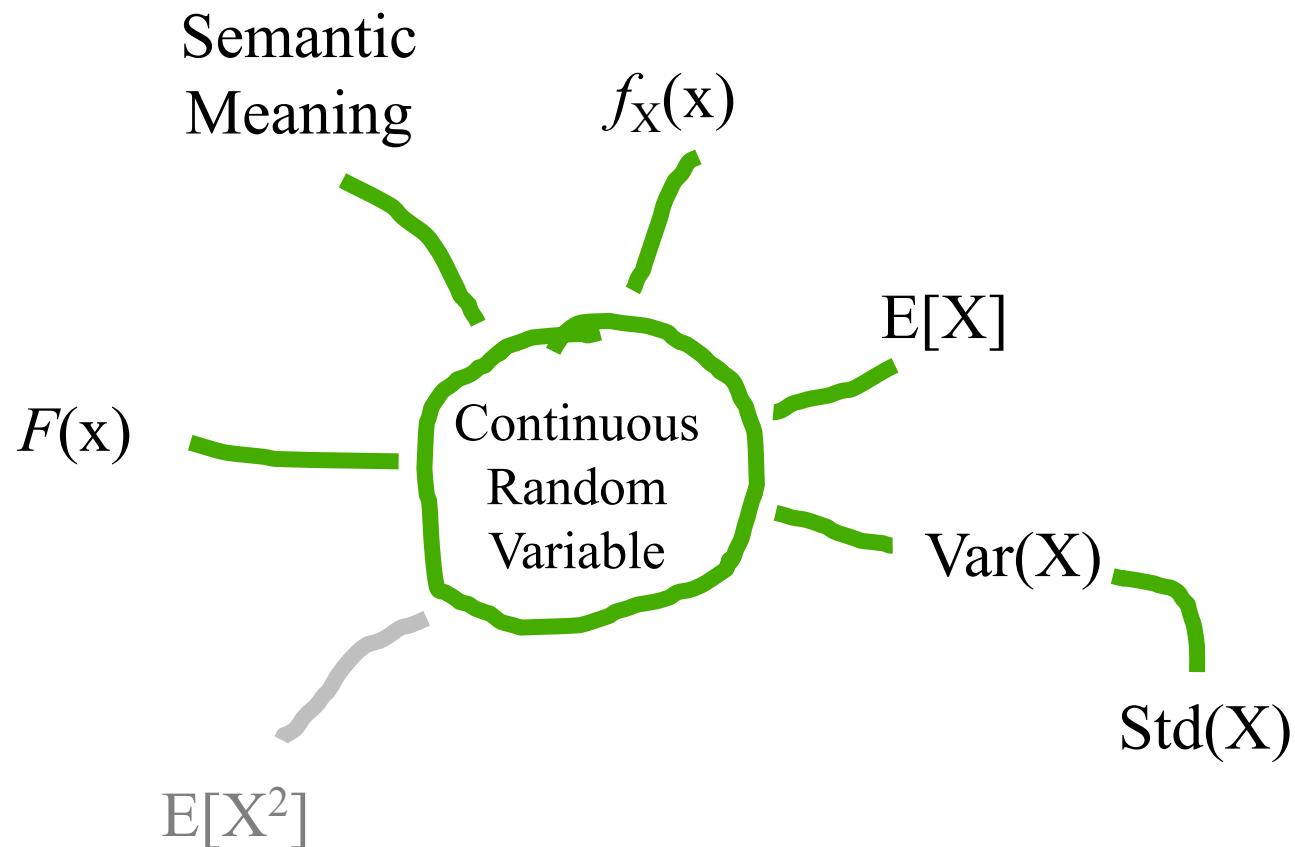
$$Y \sim \text{Exp}(\lambda = 0.002) \quad F(y) = 1 - e^{-0.002y}$$

$$\begin{aligned} P(Y < 4) &= F(4) \\ &= 1 - e^{-0.002 \cdot 4} \\ &\approx 0.008 \end{aligned}$$

Feeling lucky?

*According to USGS, 2015

Properties for Continuous Random Variables



Extra Problems

Visits to a Website

- Say visitor to your web site leaves after X minutes
 - On average, visitors leave site after 5 minutes
 - Assume length of stay is Exponentially distributed
 - $X \sim \text{Exp}(\lambda = 1/5)$, since $E[X] = 1/\lambda = 5$
 - What is $P(X > 10)$?

$$P(X > 10) = 1 - F(10) = 1 - (1 - e^{-\lambda 10}) = e^{-2} \approx 0.1353$$

- What is $P(10 < X < 20)$?

$$P(10 < X < 20) = F(20) - F(10) = (1 - e^{-4}) - (1 - e^{-2}) \approx 0.1170$$

Replacing Your Laptop

- $X = \# \text{ hours of use until your laptop dies}$
 - On average, laptops die after 5000 hours of use
 - $X \sim \text{Exp}(\lambda = 1/5000)$, since $E[X] = 1/\lambda = 5000$
 - You use your laptop 5 hours/day.
 - What is $P(\text{your laptop lasts 4 years})$?
 - That is: $P(X > (5)(365)(4)) = P(X > 7300)$

$$P(X > 7300) = 1 - F(7300) = 1 - (1 - e^{-7300/5000}) = e^{-1.46} \approx 0.2322$$

- Better plan ahead... especially if you are coterminating:

$$P(X > 9125) = 1 - F(9125) = e^{-1.825} \approx 0.1612 \quad (\text{5 year plan})$$

$$P(X > 10950) = 1 - F(10950) = e^{-2.19} \approx 0.1119 \quad (\text{6 year plan})$$

Exponential is Memoryless

- $X = \text{time until some event occurs}$
 - $X \sim \text{Exp}(\lambda)$
 - What is $P(X > s + t | X > s)$?

$$P(X > s + t | X > s) = \frac{P(X > s + t \text{ and } X > s)}{P(X > s)} = \frac{P(X > s + t)}{P(X > s)}$$

$$\frac{P(X > s + t)}{P(X > s)} = \frac{1 - F(s + t)}{1 - F(s)} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} = 1 - F(t) = P(X > t)$$

$$\text{So, } P(X > s + t | X > s) = P(X > t)$$

- After initial period of time s , $P(X > t | \bullet)$ for waiting another t units of time until event is same as at start
- “Memoryless” = no impact from preceding period s

Disk Crashes

- $X = \text{days of use before your disk crashes}$

$$f(x) = \begin{cases} \lambda e^{-x/100} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

- First, determine λ to have actual PDF

- Good integral to know: $\int e^u du = e^u$

$$1 = \int \lambda e^{-x/100} dx = -100\lambda \int \frac{-1}{100} e^{-x/100} dx = -100\lambda e^{-x/100} \Big|_0^\infty = 100\lambda \Rightarrow \lambda = \frac{1}{100}$$

- What is $P(50 < X < 150)$?

$$F(150) - F(50) = \int_{50}^{150} \frac{1}{100} e^{-x/100} dx = -e^{-x/100} \Big|_{50}^{150} = -e^{-3/2} + e^{-1/2} \approx 0.383$$

- What is $P(X < 10)$?

$$F(10) = \int_0^{10} \frac{1}{100} e^{-x/100} dx = -e^{-x/100} \Big|_0^{10} = -e^{-1/10} + 1 \approx 0.095$$