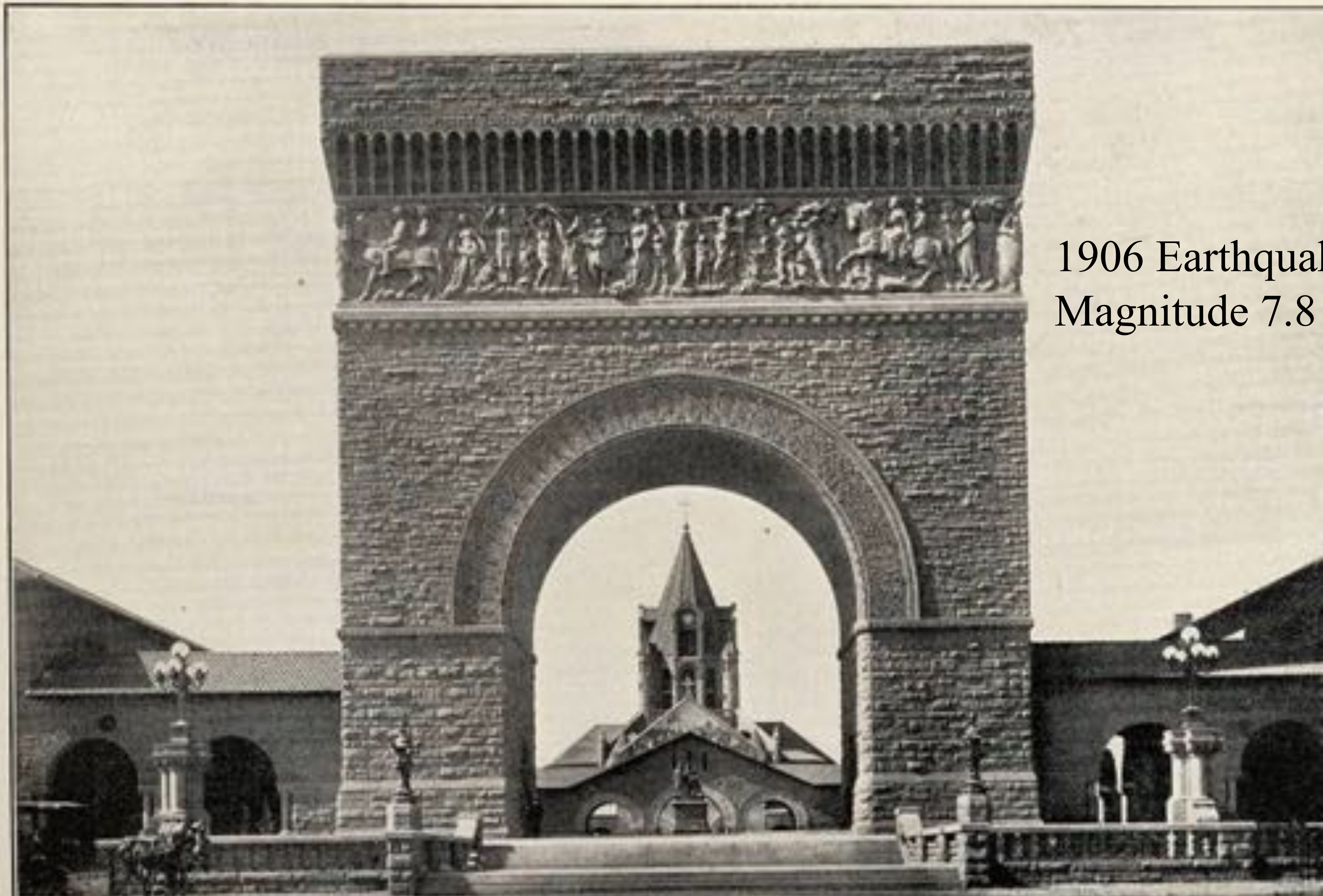
# Continuous Variables

Chris Piech
CS109, Stanford University




ILL. No. 65. MEMORIAL ARCH, WITH CHURCH IN BACKGROUND, STANFORD UNIVERSITY, SHOWING TYPES OF CARVED WO
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 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** | **time** to get successes | |
|---|---|---|---|
| One trial | $X \sim \text{Ber}(p)$ | $X \sim \text{Geo}(p)$ | One success |
| | ↑ ***n* = 1** | ↑ ***r* = 1** | |
| Several trials | $X \sim \text{Bin}(n, p)$ | $X \sim \text{NegBin}(r, p)$ | Several successes |
| Interval of time | $X \sim \text{Poi}(\lambda)$ | $X \sim \text{Exp}(\lambda)$ | 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:

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

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

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

# 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, $p$)

# 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<br>$p$: 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: | $\lambda$ : the rate of events in one interval |
| Range($X$): | $\{0, 1, \ldots, \infty\}$ |
| pmf: | $\Pr(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$ |
| E[$X$]: | $\lambda$ |
| Var($X$): | $\lambda$ |

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 ($\lambda$).

**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)

# 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($\lambda$)

**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?*

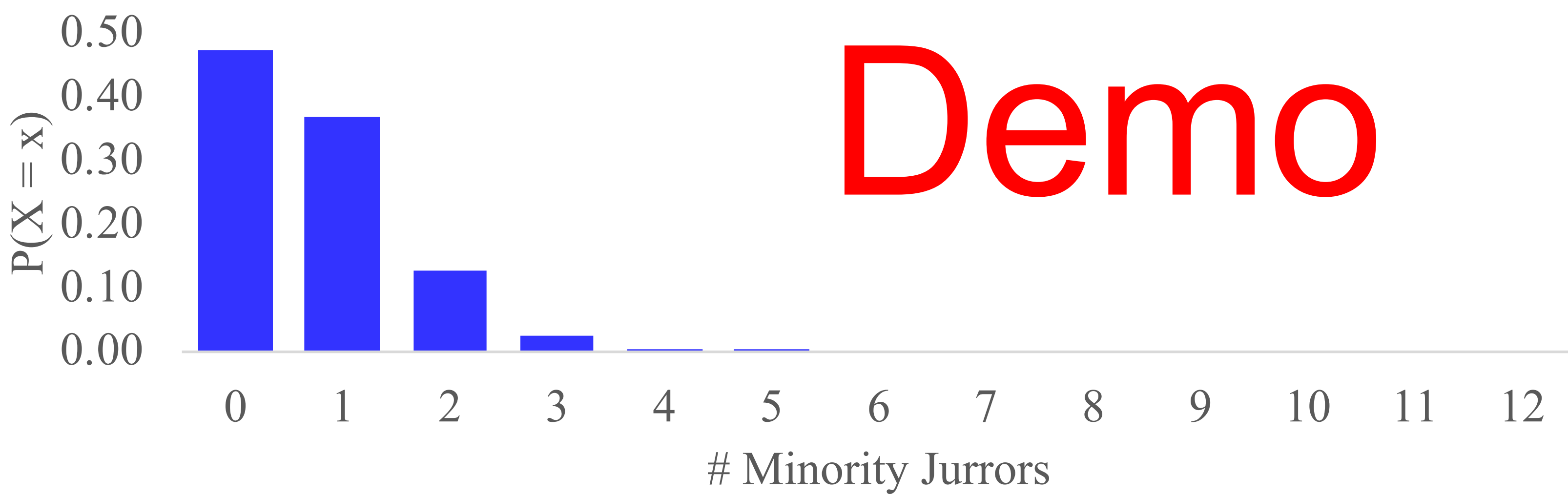
- Article by Erin Miller –January 22, 2010
- Thanks to (former CS109er) Josh Falk for this article

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 ball) constant for every draw = 60/1000
  - X = # 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 <u>over half</u> of juries to have at least one non-white person on them*



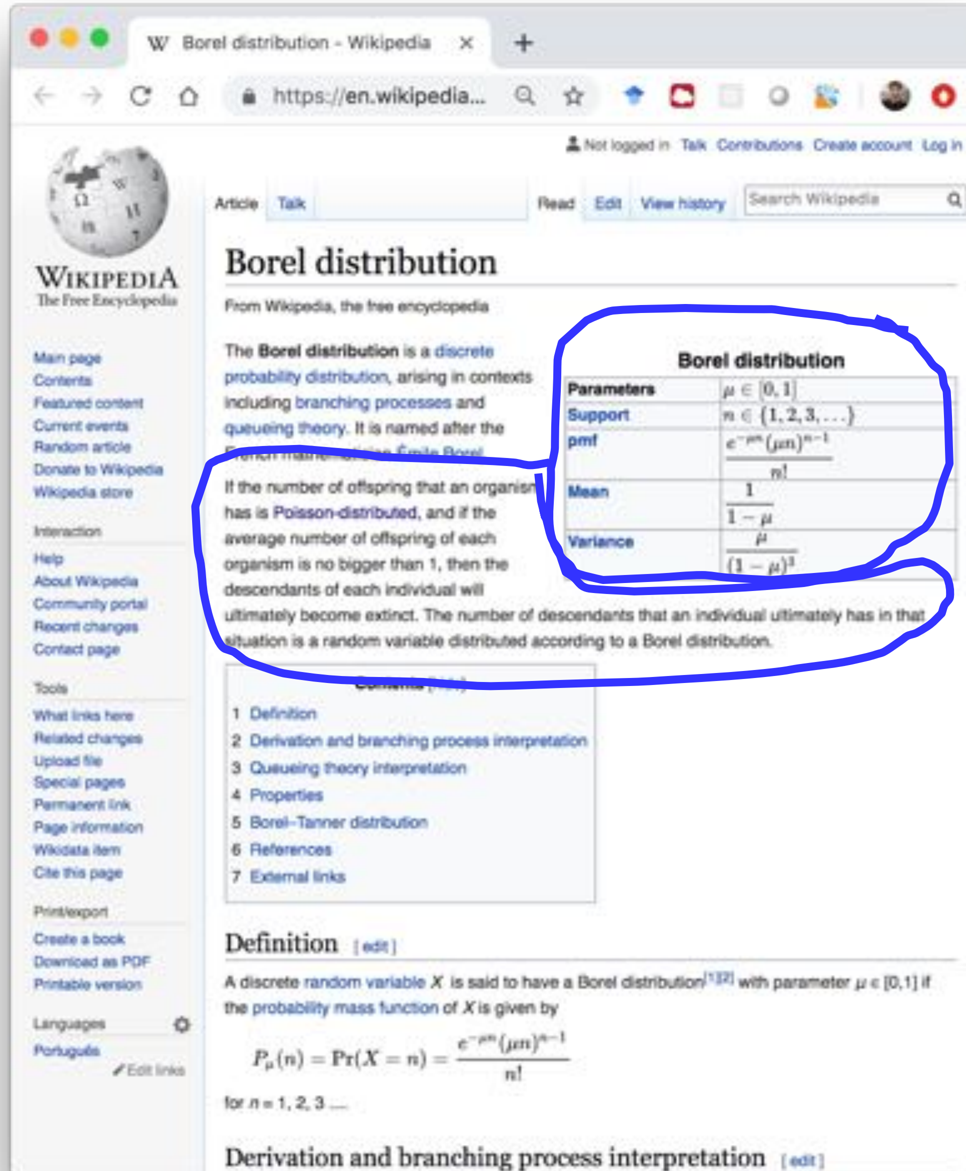

Demo

# 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 μ = 0.2 …



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)?

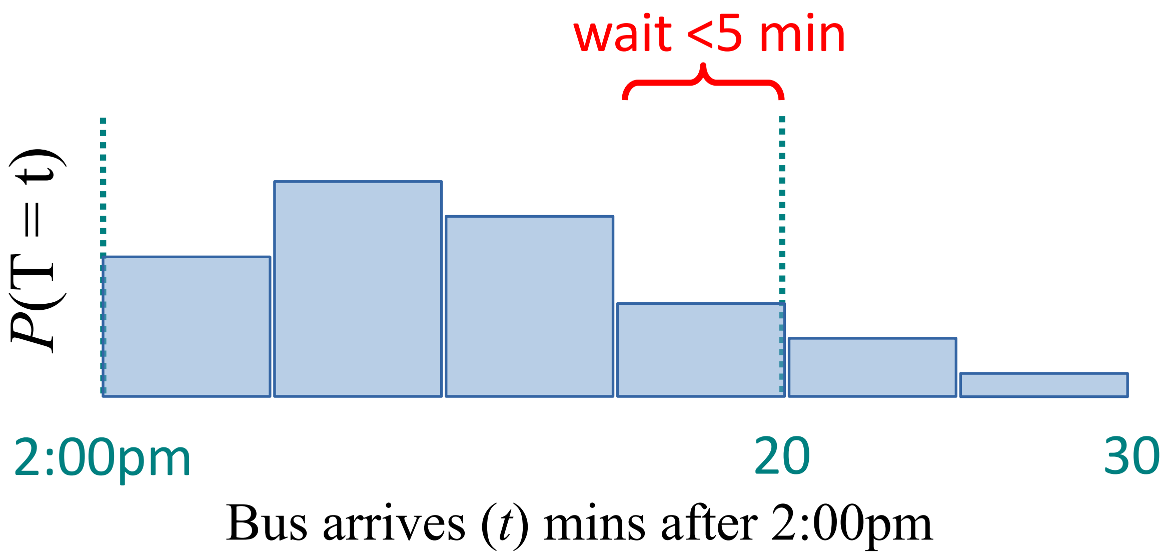
What is the probability that the bus arrives at: 2:17pm and 12.12333911102389234 seconds?

# 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)?

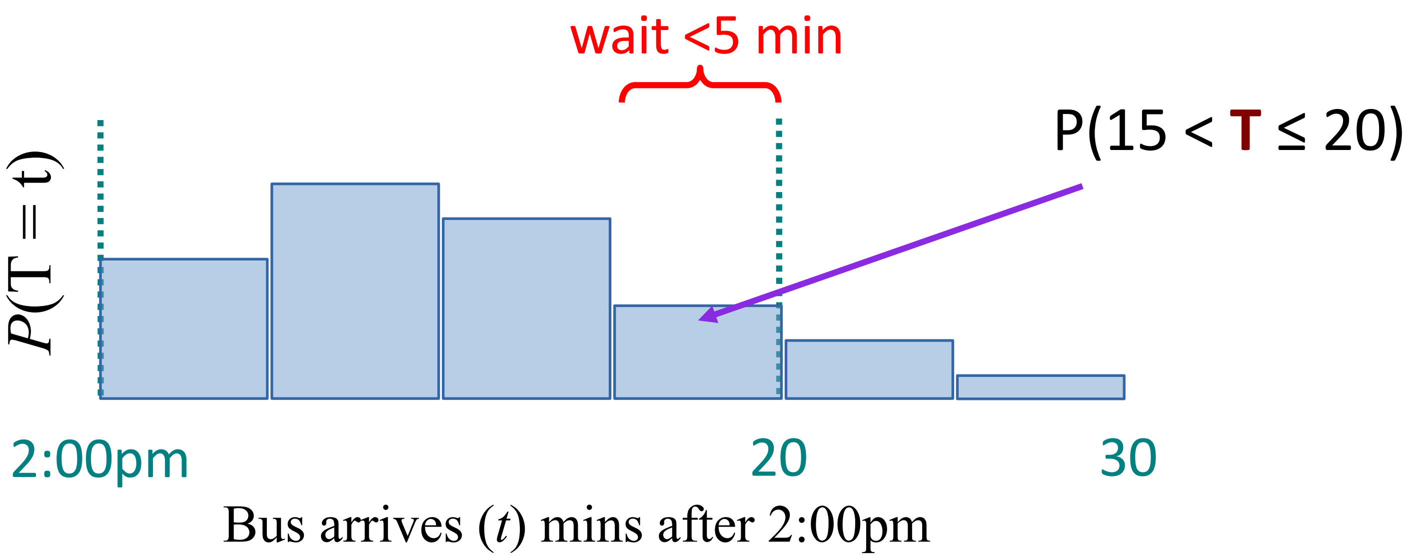



image: Haha169

# 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)?

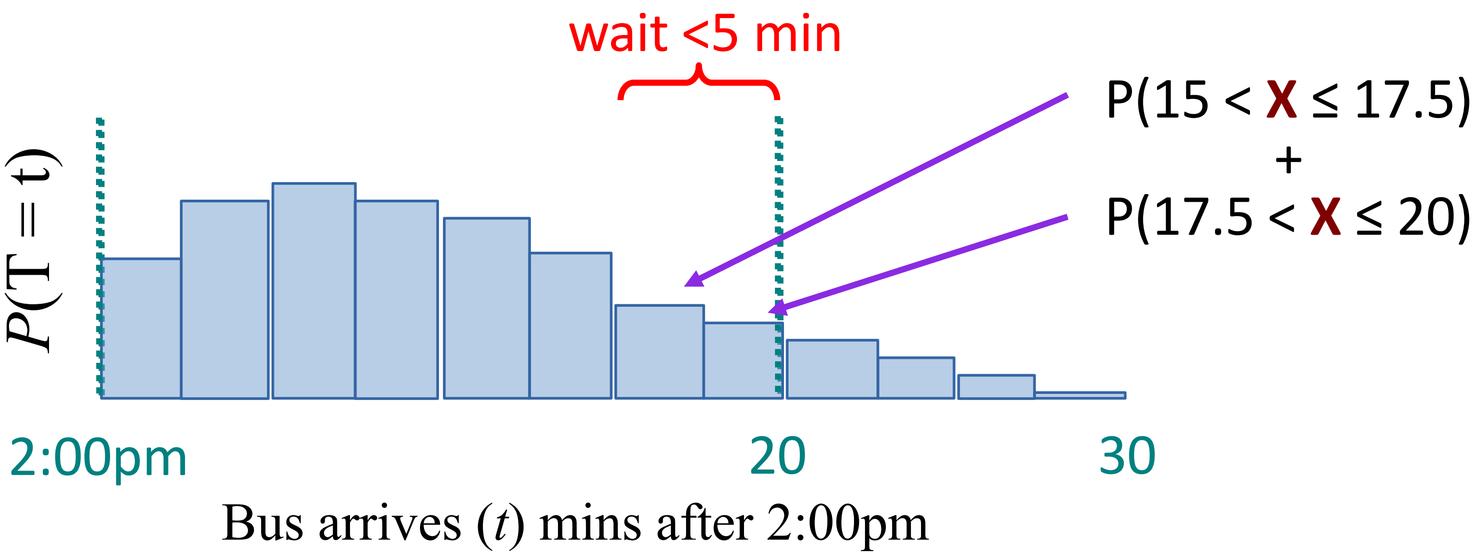



image: Haha169

# 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)?

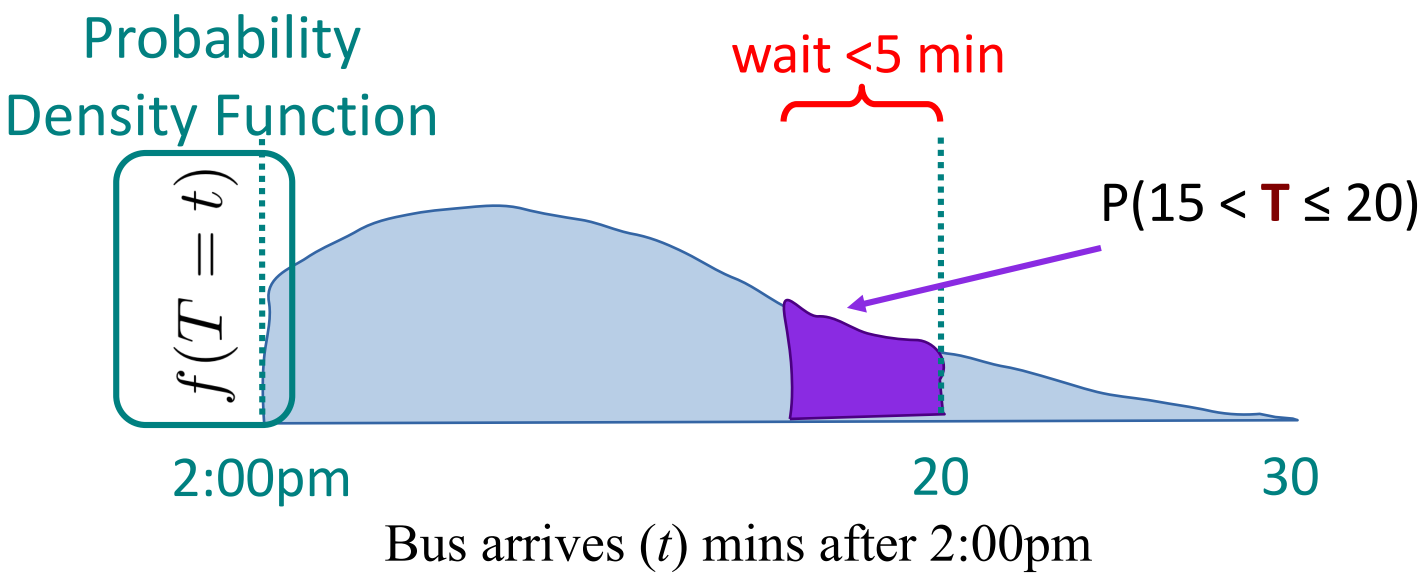



image: [Haha169]

# 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

# Probability Density Function

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

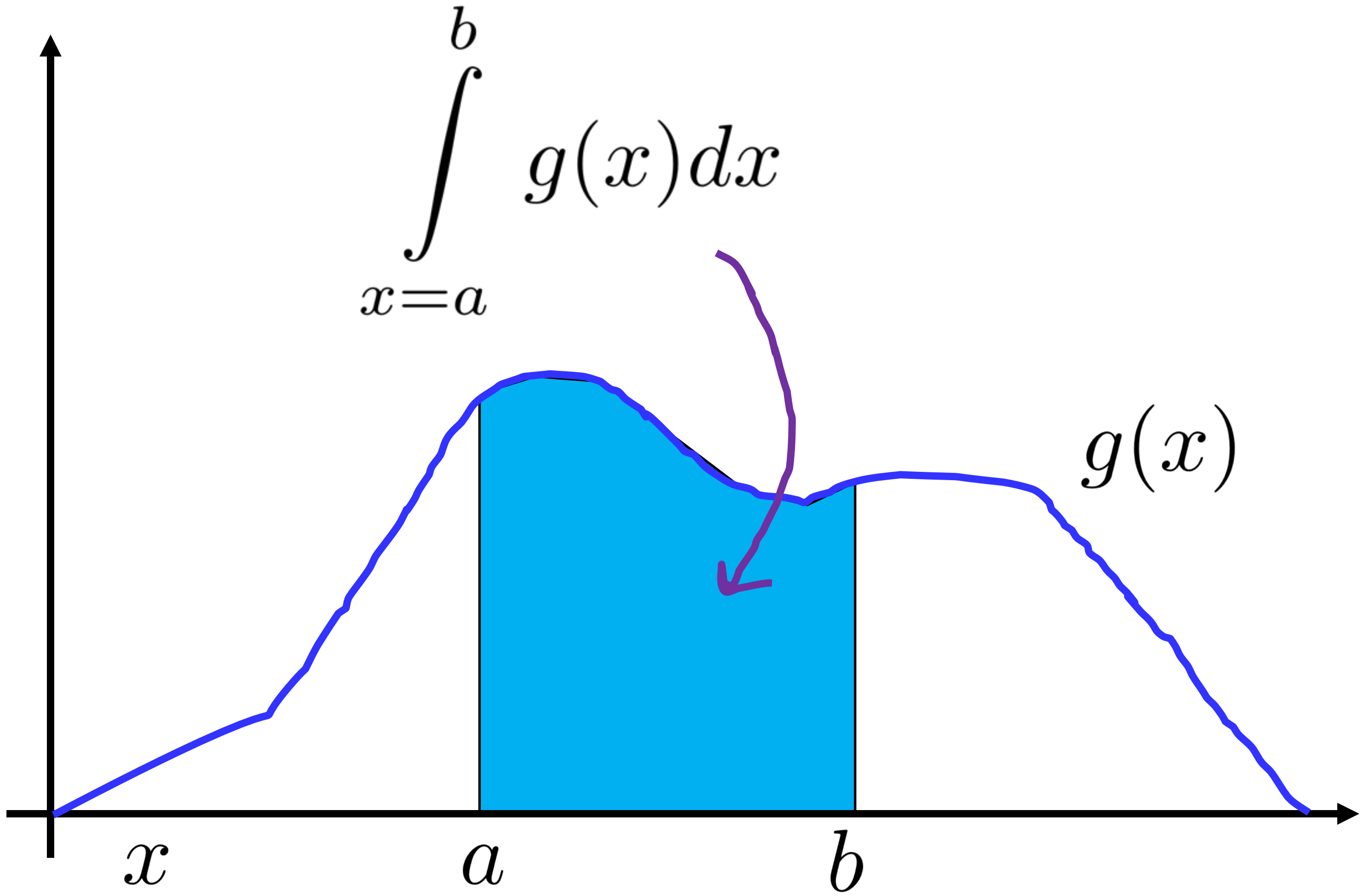
Units of probability *divided by units of X*.
**Integrate it** to get probabilities!

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

# 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)?

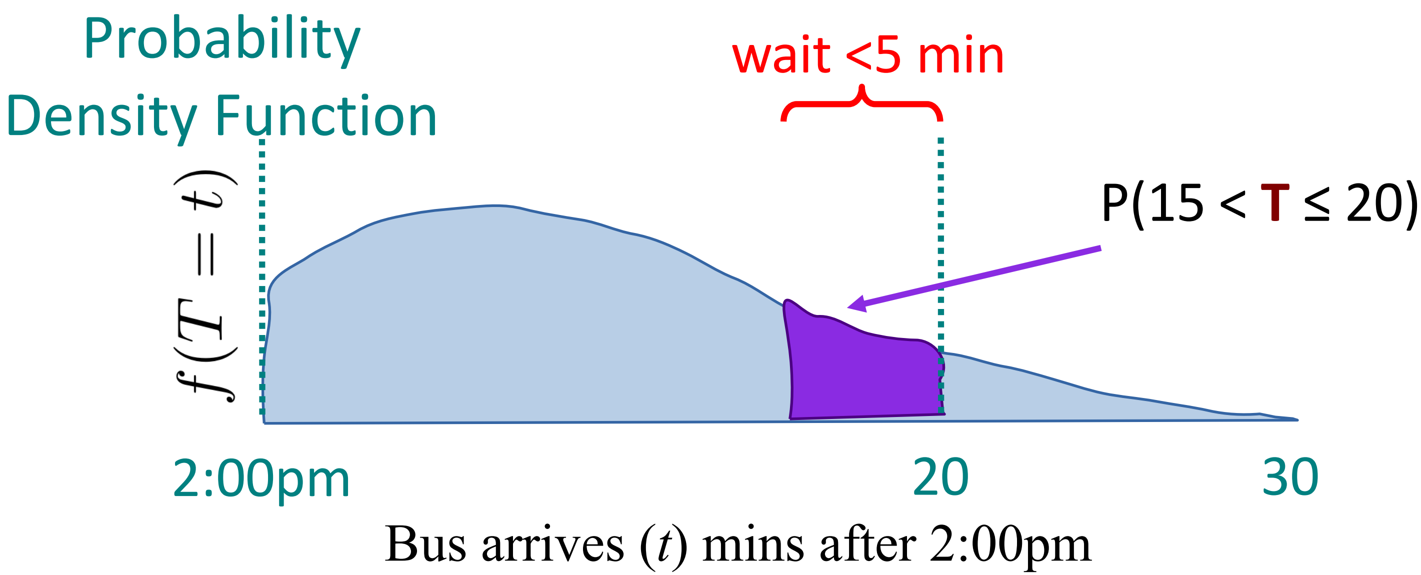



image: Haha169

# Properties of PDFs

The integral of a PDF gives a probability. Thus:

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

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

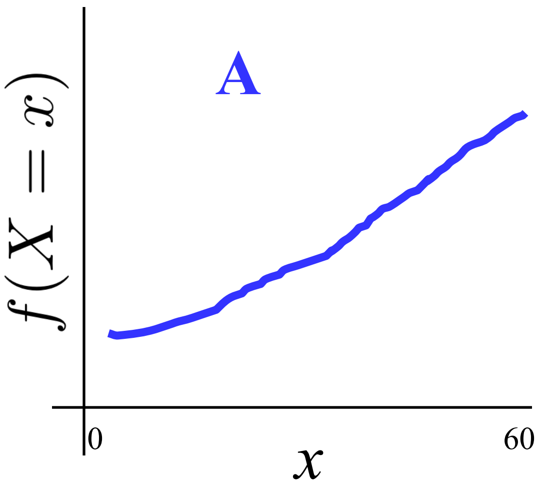
What do you get if you integrate over a probability *density* function?

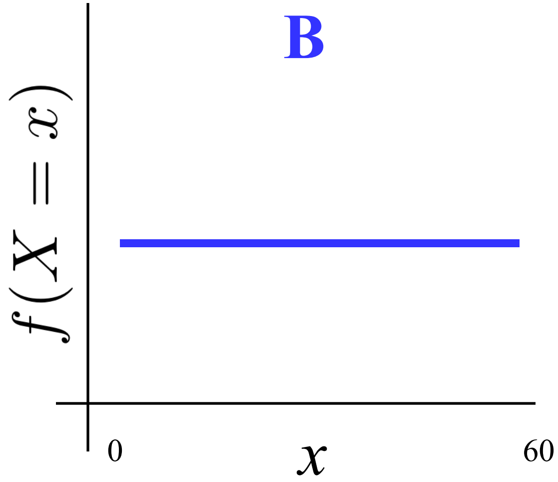
# A probability!

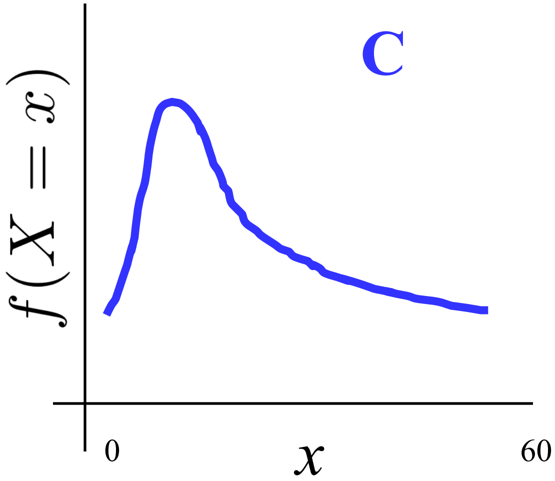
# Probability Density Function

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:



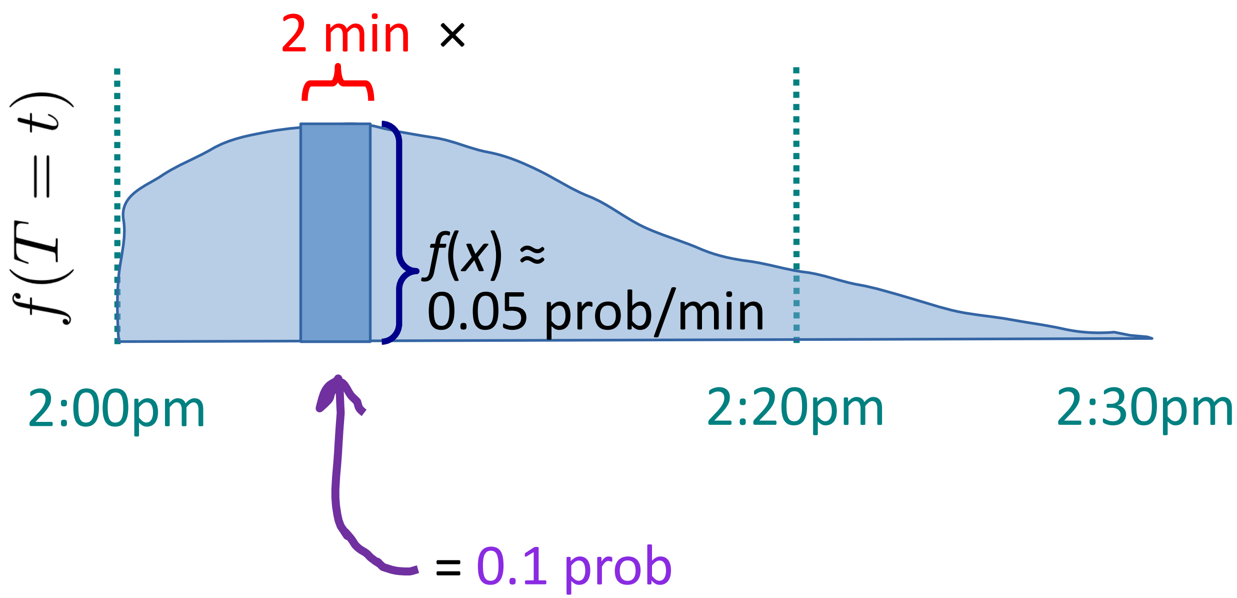




Which of these represent that you think the arrival is more likely to be close to 3:00pm

The ratio of probability densities is meaningful

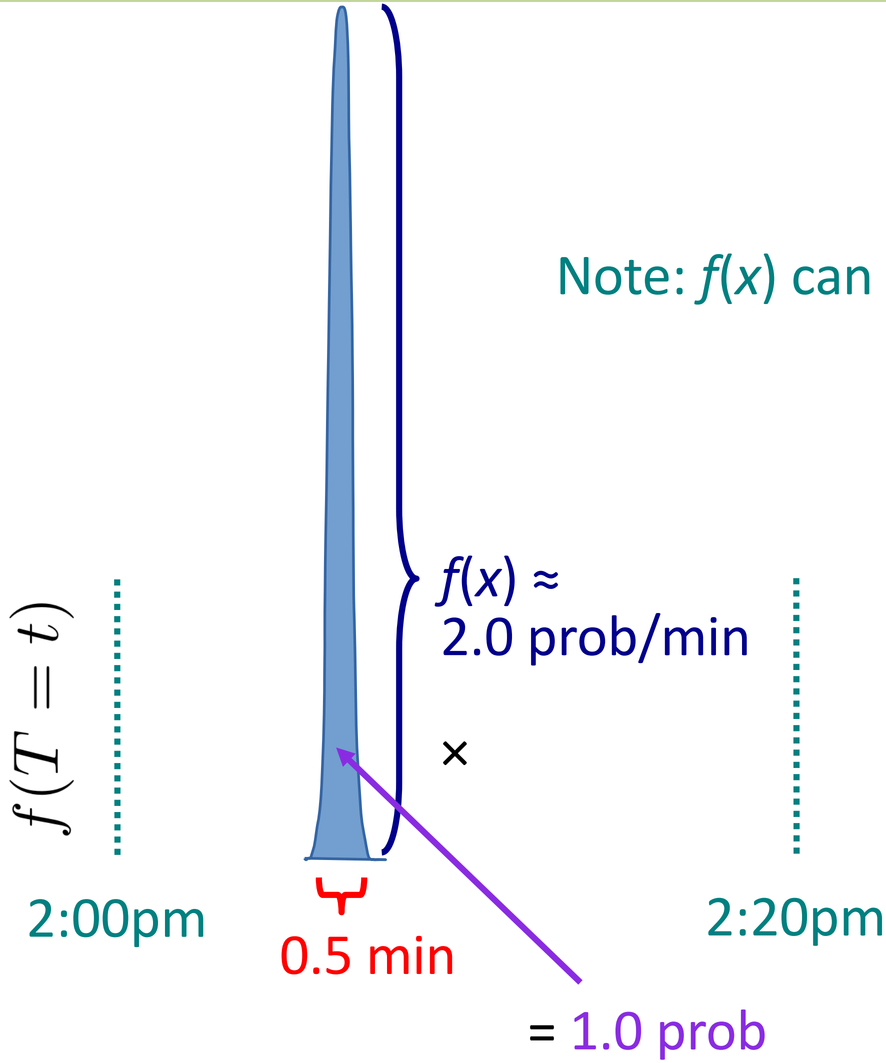
# $f(X = x)$ is *Not* a Probability

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.*

# Simple Example

Consider a random 5000x5000 matrix, where each element in the matrix is Uniform(0,1). What is the probability that a selected eigenvalue (λ) of the matrix is greater than 0?*

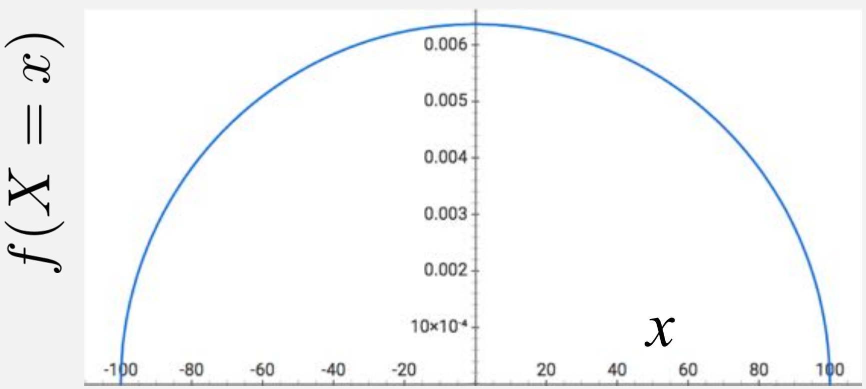
* With help from Wigner, Chris is going to rephrase this problem

# Simple Example from Quantum Physics

Let X be a continuous random variable[1]

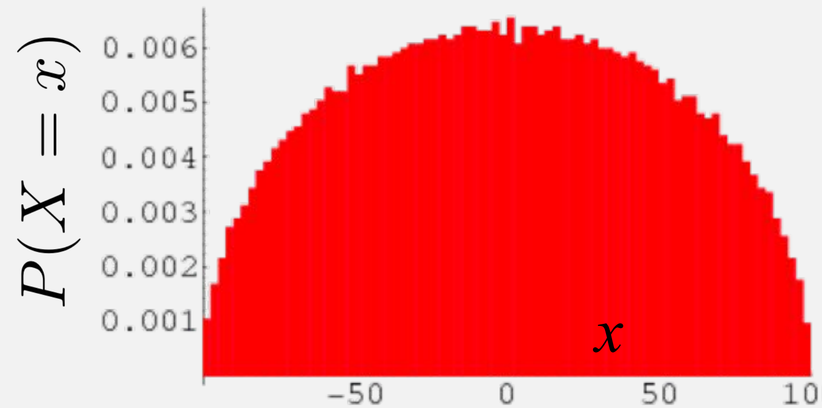
Theory

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




Practice

From simulations




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

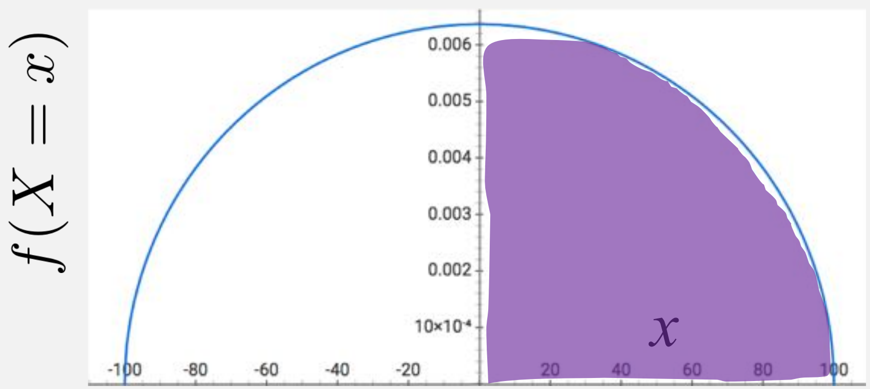
[1] X represents the eigenvalue of a 5000x5000 matrix of uniform random variables

# Simple Example from Quantum Physics

Let X be a continuous random variable:

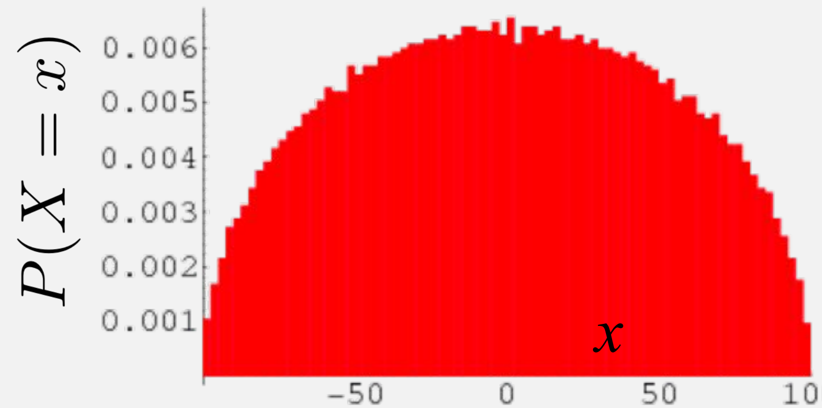
**Theory**

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




**Practice**

From simulations




Approach #1: Integrate over the PDF

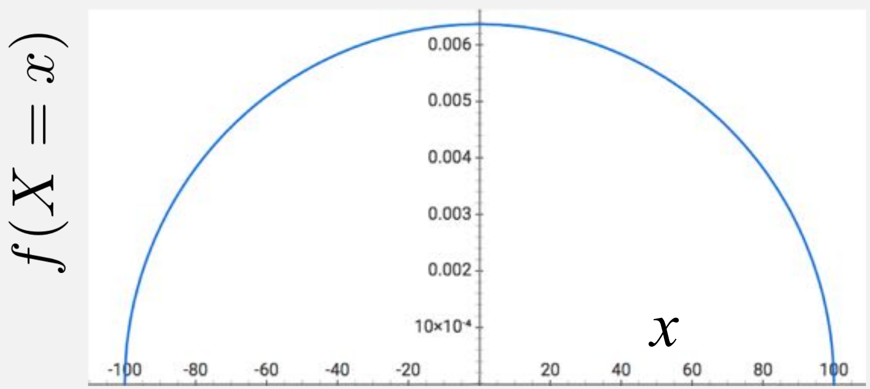
$$P(X > 0) = \int_{0}^{100} f(X = x)\, dx$$

# Simple Example from Quantum Physics

Let X be a continuous random variable:

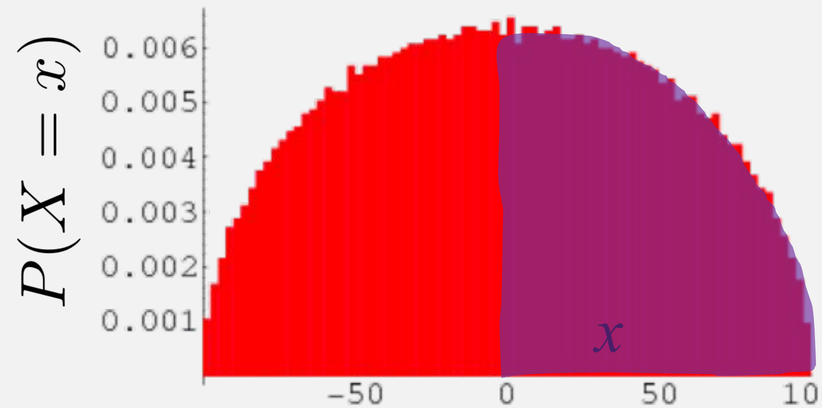
**Theory**

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




**Practice**

From simulations




Approach #2: Discrete Approximation

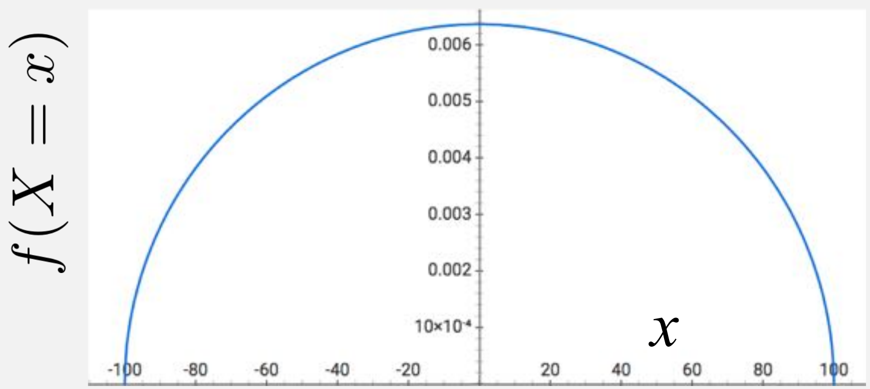
$$P(X > 0) \approx \sum_{i=0}^{100} P(X = i)$$

# Simple Example from Quantum Physics

Let X be a continuous random variable:

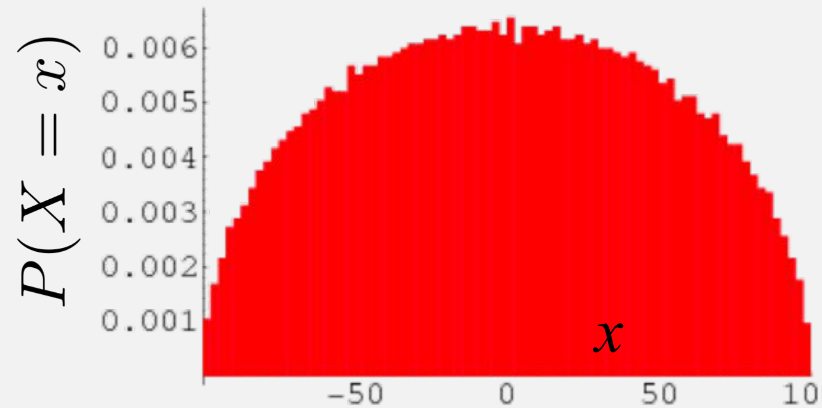
Theory

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




Practice

From simulations




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!

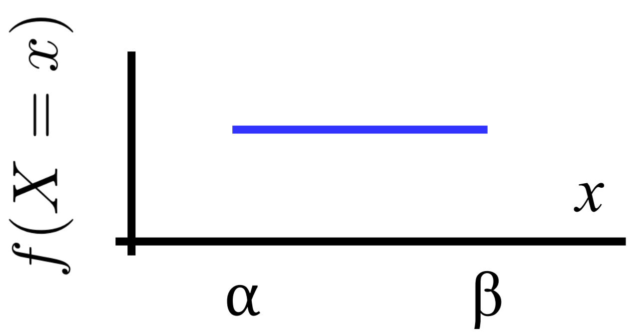
# Uniform Random Variable

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

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

## Probability Density

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




## Properties

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

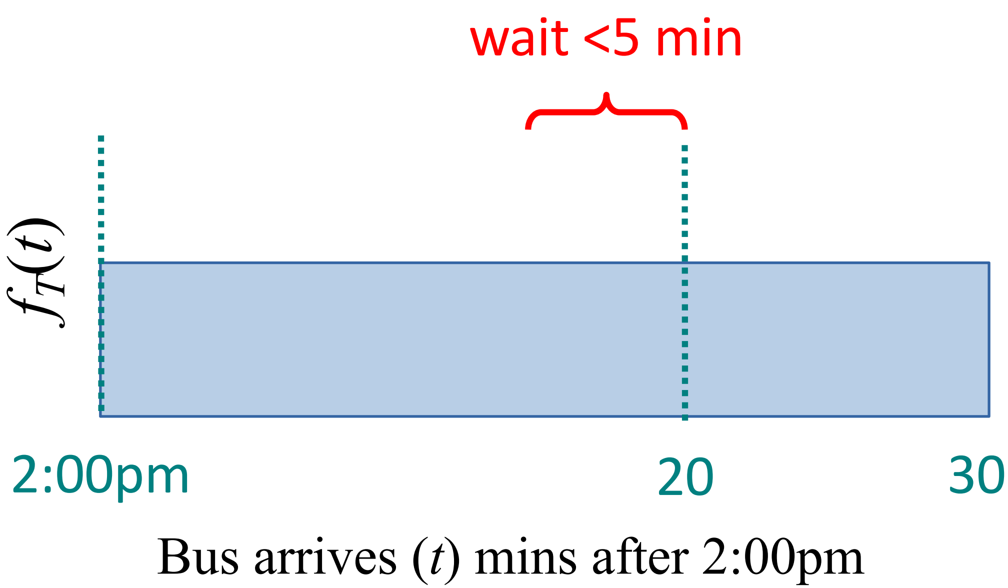
$$\text{Var}(X) = \frac{(\beta - \alpha)^2}{12}$$

# 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 2:15pm. What is P(wait < 5 minutes)?

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




$$P(\text{Wait} < 5) = \int_{15}^{20} \frac{1}{\beta - \alpha}\, dx$$

$$= \left. \frac{x}{\beta - \alpha} \right|_{15}^{20}$$

$$= \left. \frac{x}{30 - 0} \right|_{15}^{20} = \frac{5}{30}$$

# Expectation and Variance

For <u>discrete</u> RV $X$:

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

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

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

For <u>continuous</u> RV $X$:

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

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

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

For <u>both</u> discrete and continuous RVs:

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

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

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

# Expectation of Uniform

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

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

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

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

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

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

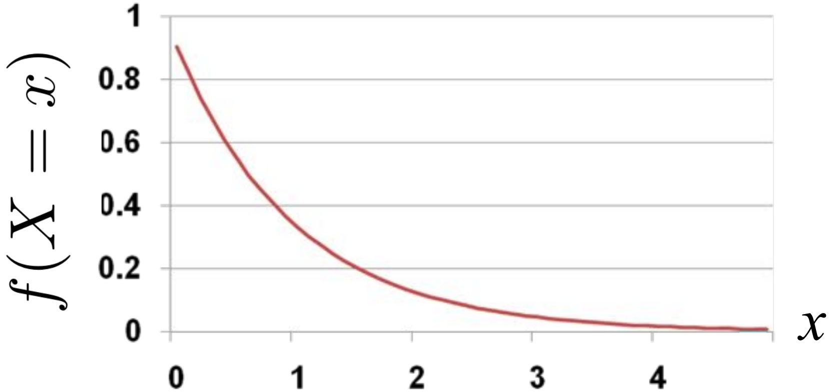
just average the start and end!

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

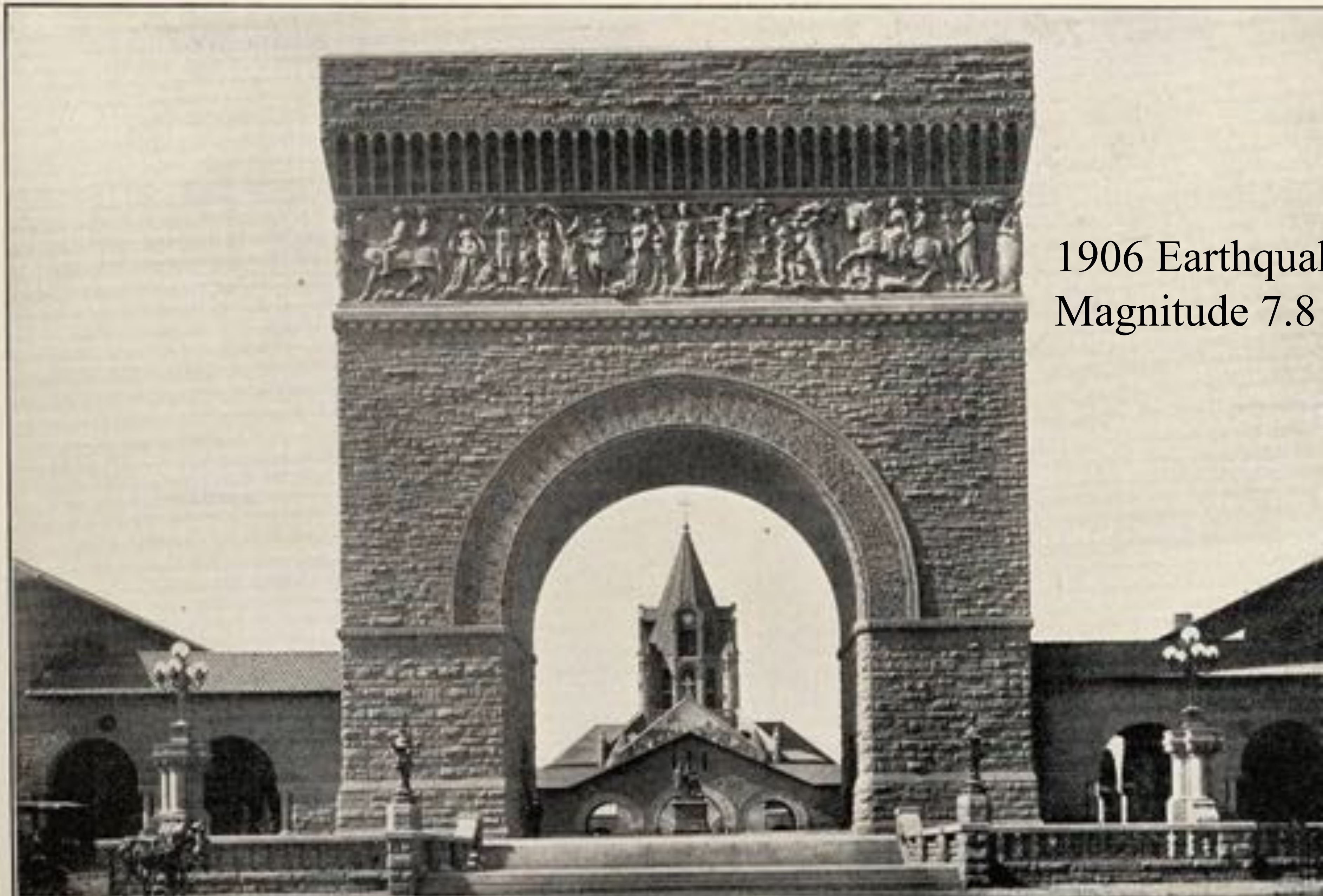
# Exponential Random Variable

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

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




  - $E[X] = \dfrac{1}{\lambda}$
  - $\text{Var}(X) = \dfrac{1}{\lambda^2}$
  - Support: $0 \leq x \leq \infty$
  - Represents time until some event
    - Earthquake, request to web server, end cell phone contract, etc.




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

# How Many Earthquakes

Based on historical data, major earthquakes (magnitude 8.0+) happen at a **rate of 0.002** per year*. What is the probability of **zero major earthquakes magnitude next year?**

$X$ = Number of major earthquakes next year

---

$$X \sim \text{Poi}(\lambda = 0.002)$$

$$P(X = 0) = \frac{\lambda^0 e^{-\lambda}}{0!} = \frac{0.002^0 e^{-0.002}}{0!} \approx 0.998$$

*In California, according to the USGS, 2015

# How Long Until Next Earthquake

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

$Y$ = Years until the next earthquake of magnitude 8.0+

---

$$Y \sim \text{Exp}(\lambda = 0.002)$$

$$f_Y(y) = \lambda e^{-\lambda y}$$
$$= 0.002^{-0.002y}$$

$$P(Y < 30) = \int_0^{30} 0.002 e^{-0.002y} \, dy$$

*In California, according to the USGS, 2015

# Integral Review

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

# How Long Until Next Earthquake

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

$Y$ = Years until the next earthquake of magnitude 8.0+

---

$$Y \sim \text{Exp}(\lambda = 0.002)$$

$$f_Y(y) = \lambda e^{-\lambda y}$$
$$= 0.002^{-0.002y}$$

$$P(Y < 30) = \int_0^{30} 0.002 e^{-0.002y}\, dy$$

$$= 0.002\Big[-500e^{-0.002y}\Big]_0^{30}$$

$$= \frac{500}{500}(-e^{-0.06} + e^{0}) \approx 0.06$$

*In California, according to the USGS, 2015

# How Long Until Next Earthquake

Based on historical data, major earthquakes (magnitude 8.0+) happen at a **rate of 0.002** per year*. What is the **expected number of years until the next earthquake?**

$Y$ = Years until the next earthquake of magnitude 8.0+

---

$$Y \sim \text{Exp}(\lambda = 0.002)$$

$$E[Y] = \frac{1}{\lambda} = \frac{1}{0.002} = 500$$

*In California, according to the USGS, 2015

# How Long Until Next Earthquake

Based on historical data, major earthquakes (magnitude 8.0+) happen at a **rate of 0.002** per year*. What is the **standard deviation of years until the next earthquake?**

$Y$ = Years until the next earthquake of magnitude 8.0+

$$Y \sim \text{Exp}(\lambda = 0.002)$$

$$\text{Var}(Y) = \frac{1}{\lambda^2} = \frac{1}{0.002^2} = 250,000 \text{ years}^2$$

$$\text{Std}(Y) = \sqrt{\text{Var}(X)} = 500 \text{ years}$$

*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

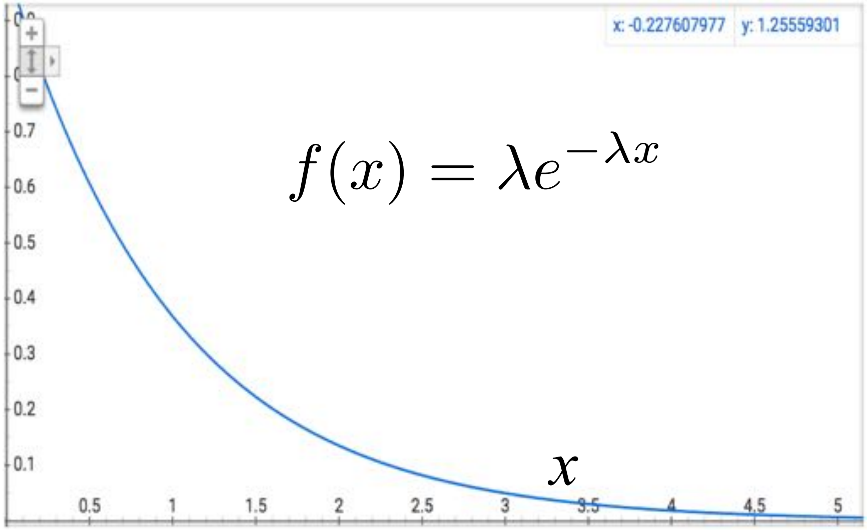
$$F_X(x) = 1 - e^{-\lambda x}$$

---

$$\begin{aligned}
P(X < x) &= \int_{y=-\infty}^{x} f(y)\ dy \\
&= \int_{y=0}^{x} \lambda e^{-\lambda y}\ dy \\
&= \frac{\lambda}{\lambda}\Big[-e^{-\lambda y}\Big]_0^x \\
&= [-e^{-\lambda x}] - [-e^{\lambda 0}] \\
&= 1 - e^{-\lambda x}
\end{aligned}$$

# CDF: X ~ Exp($\lambda$ = 1)

*Probability* density function

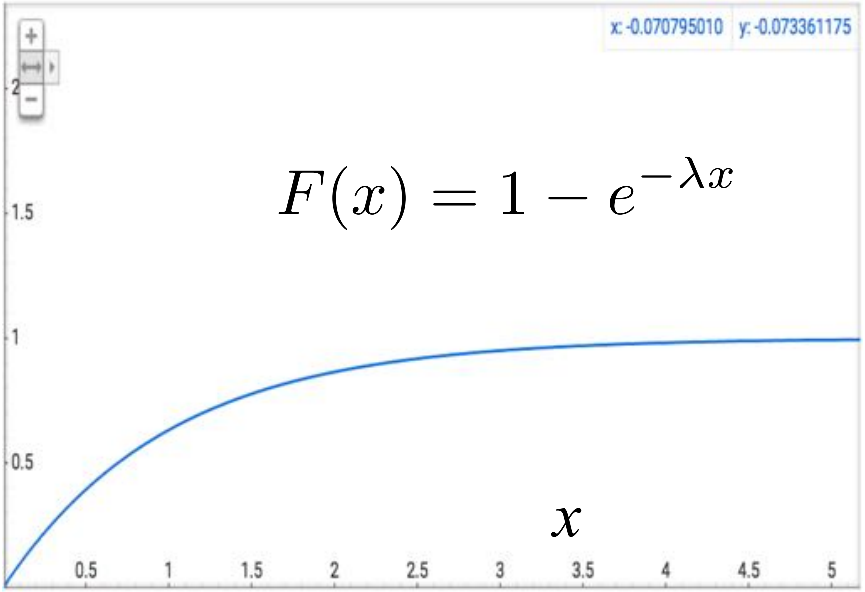



$$f(x) = \lambda e^{-\lambda x}$$

*Cumulative* density function

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

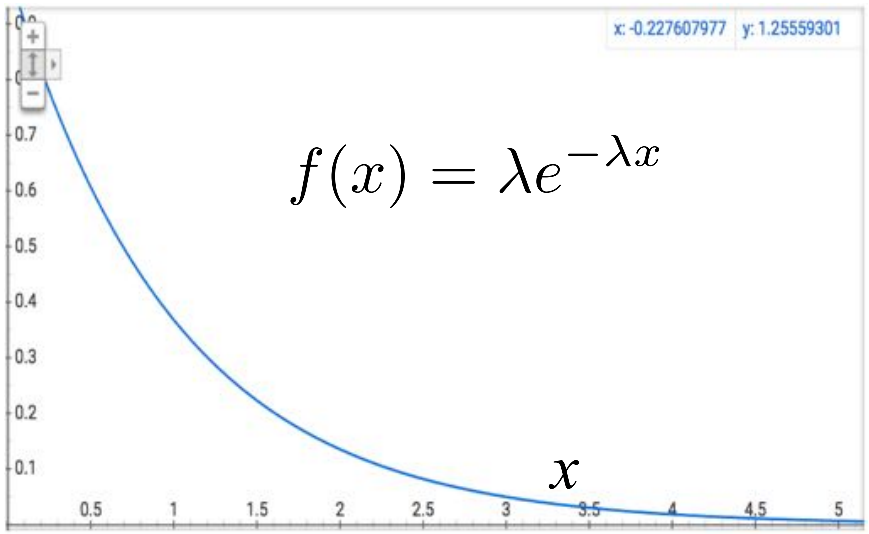
$$= \int_{y=-\infty}^{x} f(y)\, dy$$




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

# CDF: X ~ Exp($\lambda$ = 1)

*Probability* density function




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

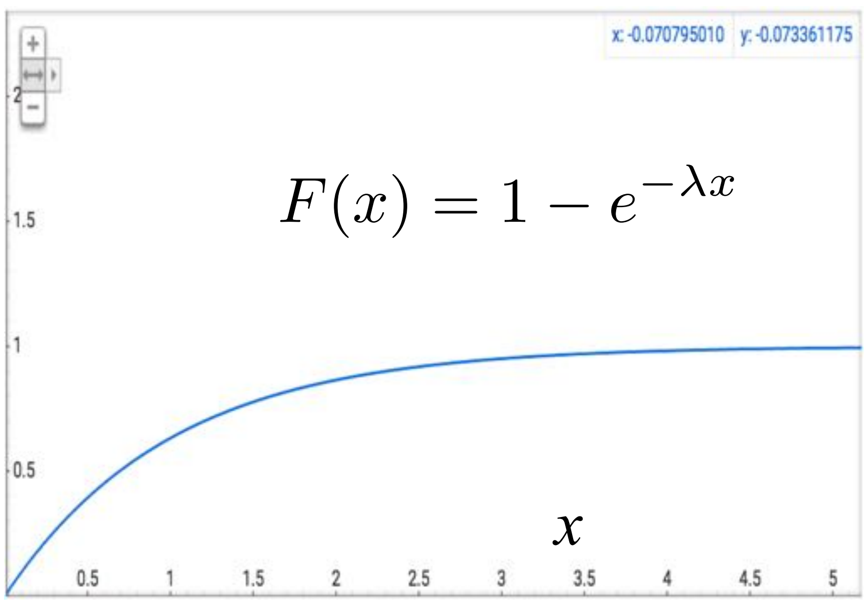
P(X < 2)

---

*Cumulative* density function

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

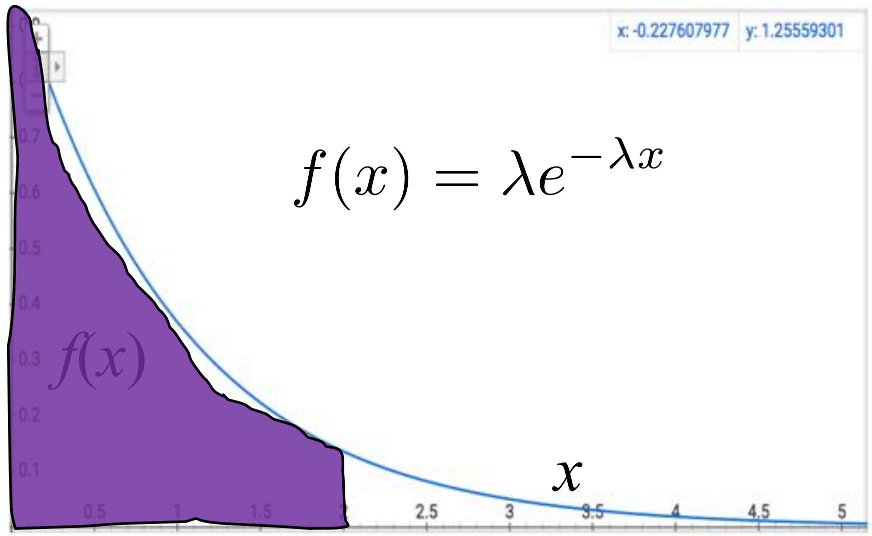
$$= \int_{y=-\infty}^{x} f(y)\ dy$$




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

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

*Probability* density function

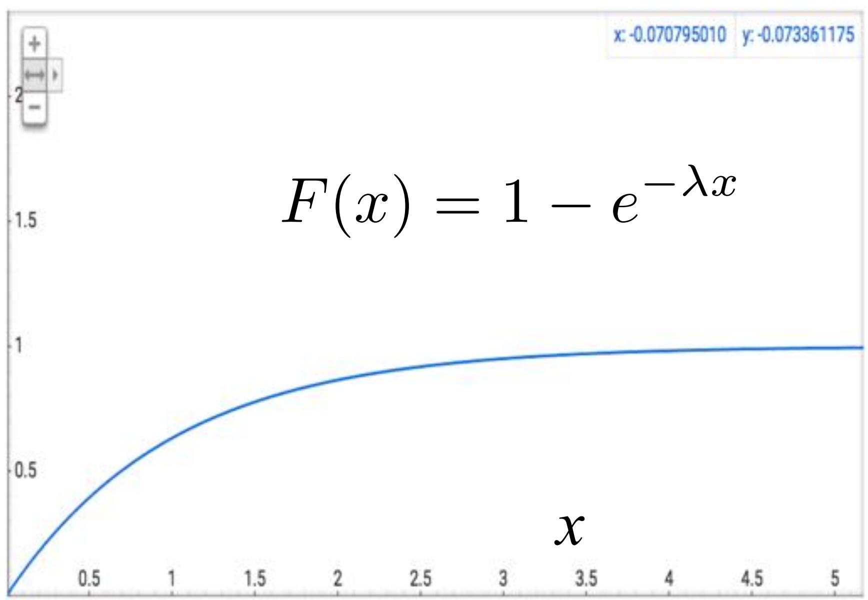



$$\frac{P(X < 2)}{= \int_{x=-\infty}^{2} f(x)\, dx}$$

*Cumulative* density function

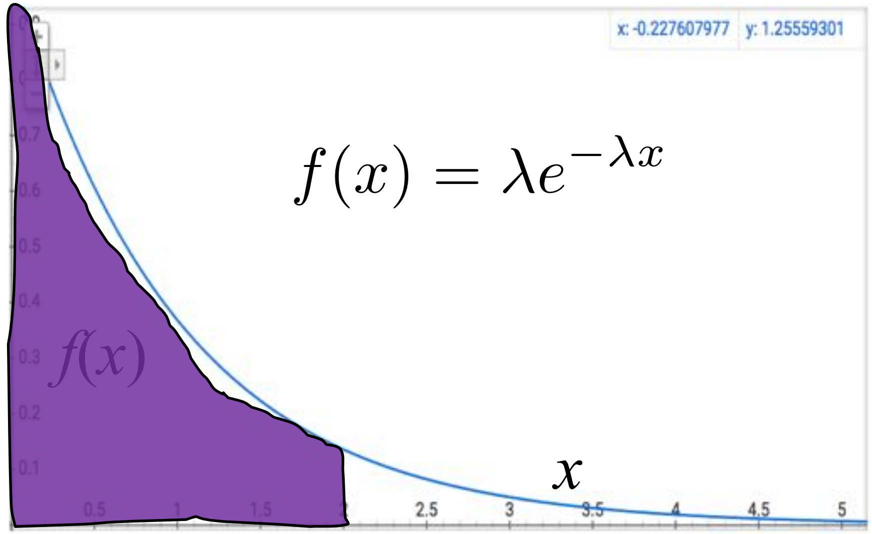
$$F_X(x) = P(X < x)$$

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



# CDF: X ~ Exp($\lambda$ = 1)

*Probability* density function



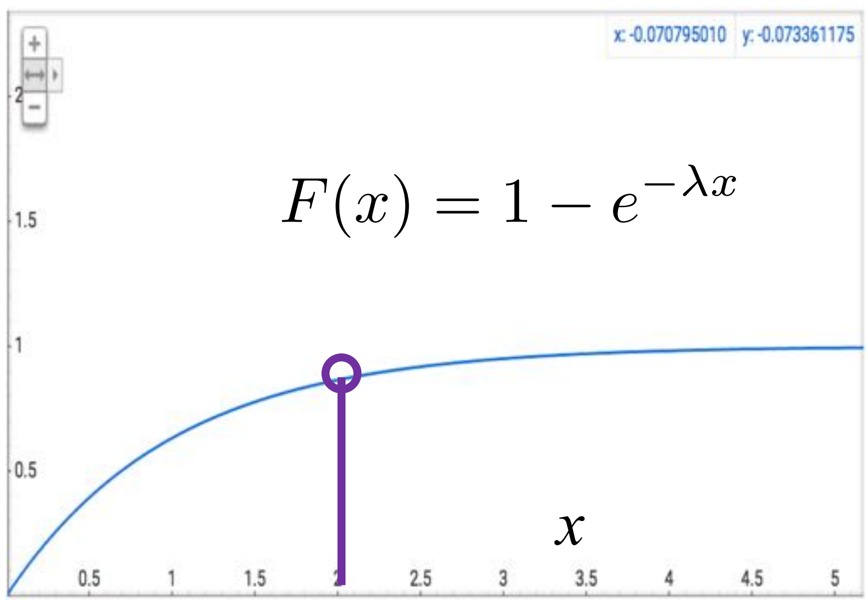

P(X < 2)

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

*Cumulative* density function

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

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




or

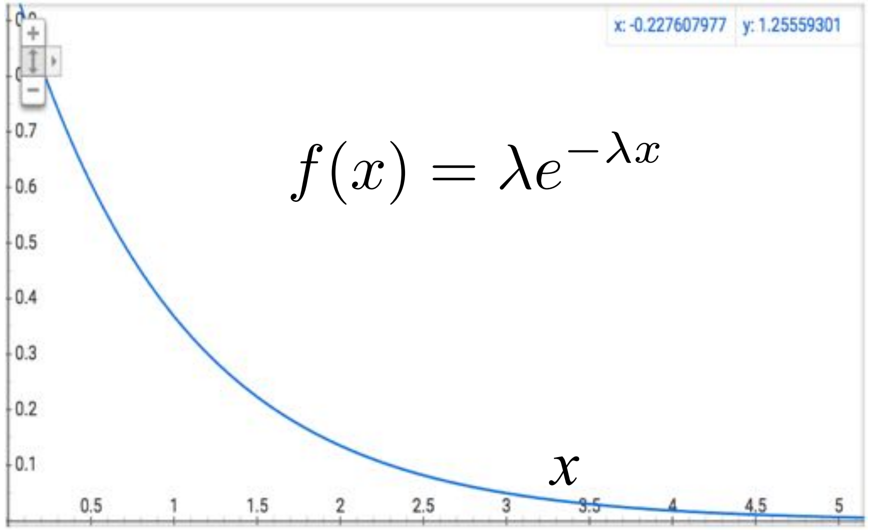
$$= F(2)$$

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

$$\approx 0.84$$

# CDF: X ~ Exp($\lambda$ = 1)

*Probability* density function



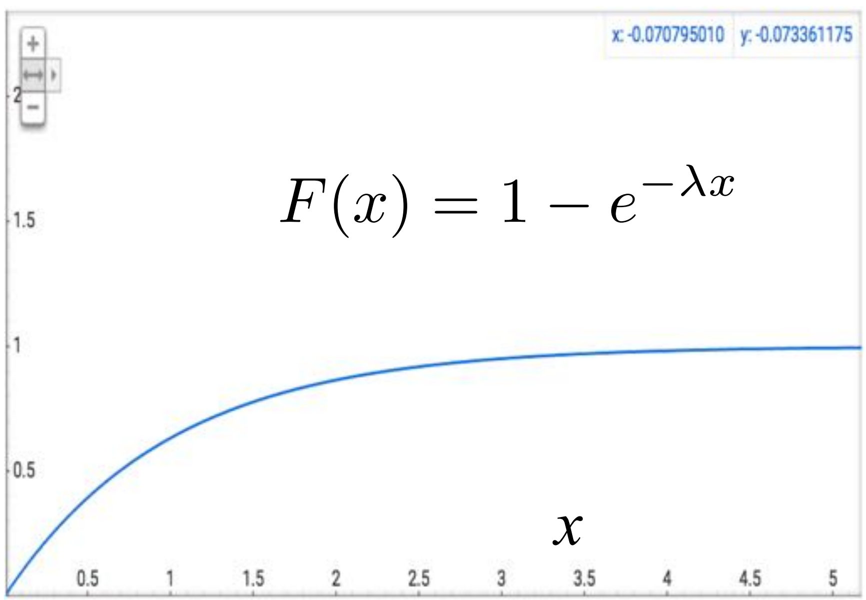

P(X > 1)

______________________

*Cumulative* density function

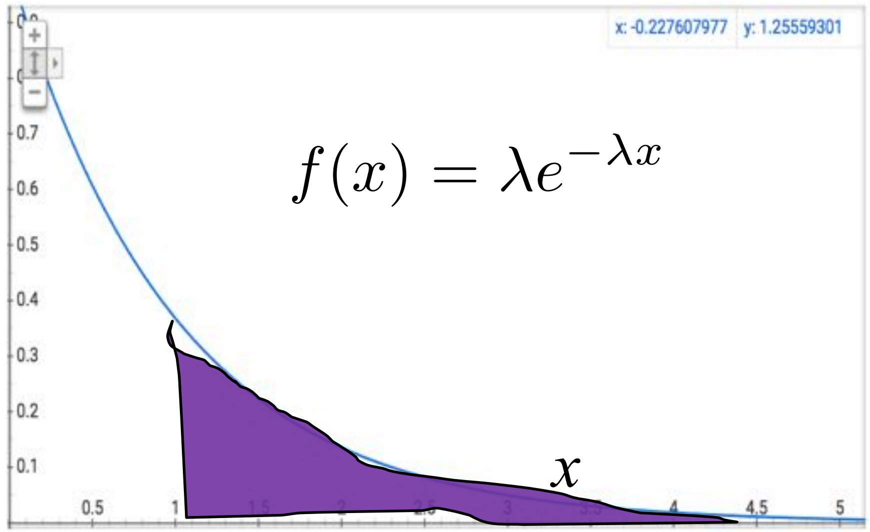
$$F_X(x) = P(X < x)$$

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



# CDF: X ~ Exp(λ = 1)

*Probability* density function



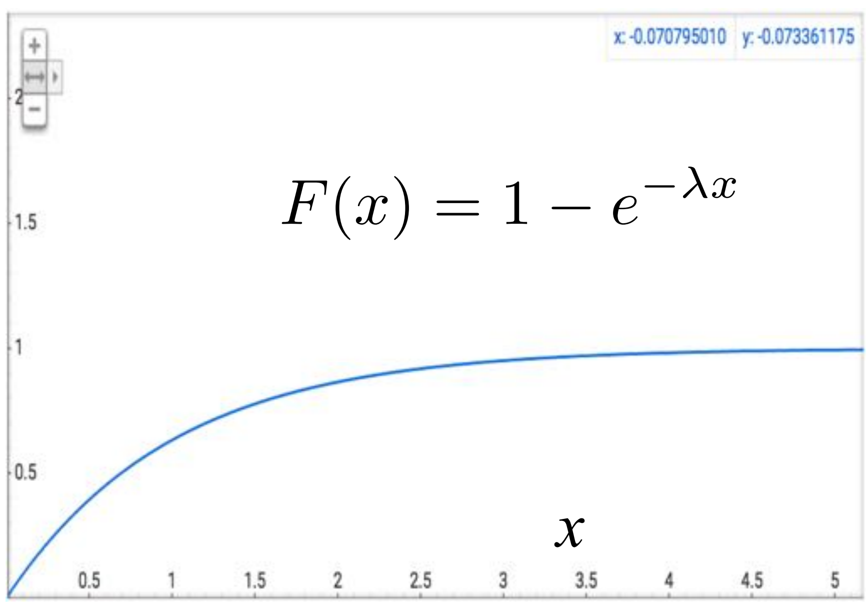

P(X > 1)

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

*Cumulative* density function

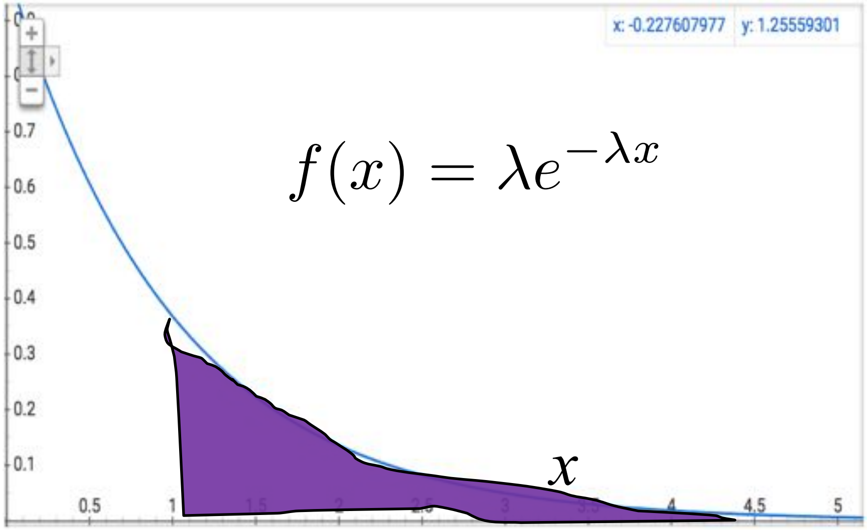
$$F_X(x) = P(X < x)$$

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



# CDF: X ~ Exp($\lambda$ = 1)

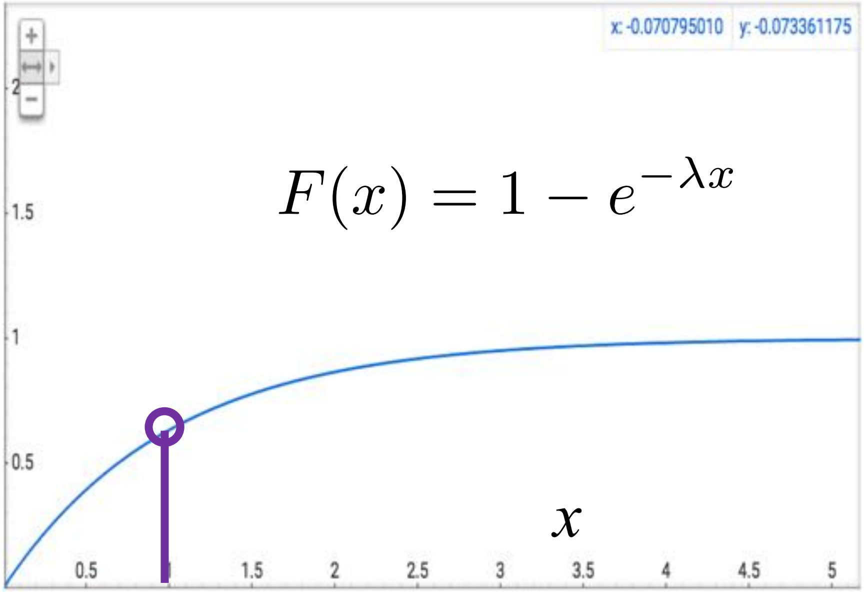
*Probability* density function




*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

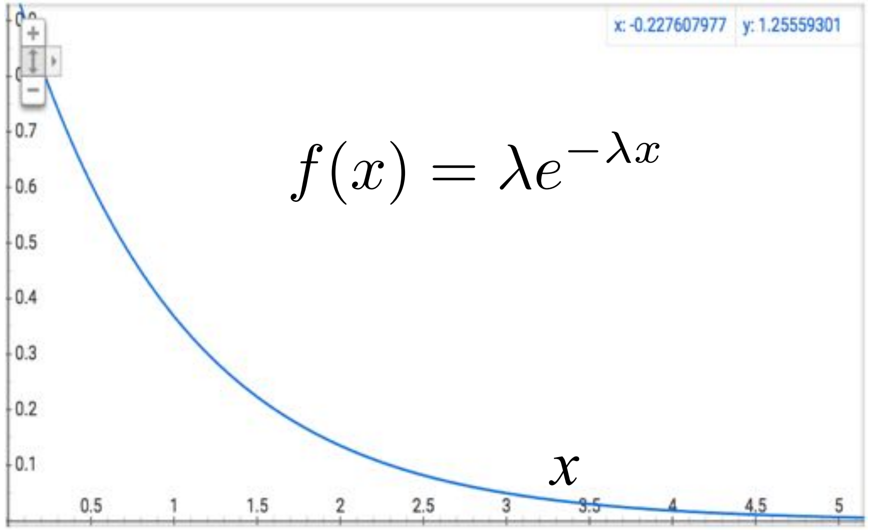
$$= 1 - F(1)$$

$$= e^{-1}$$

$$\approx 0.37$$

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

*Probability* density function

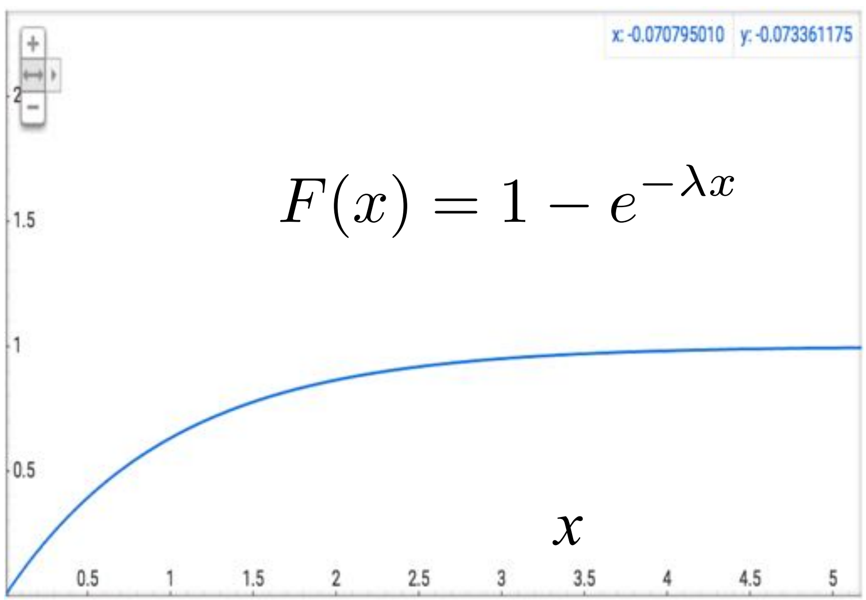



$P(1 < X < 2)$

*Cumulative* density function

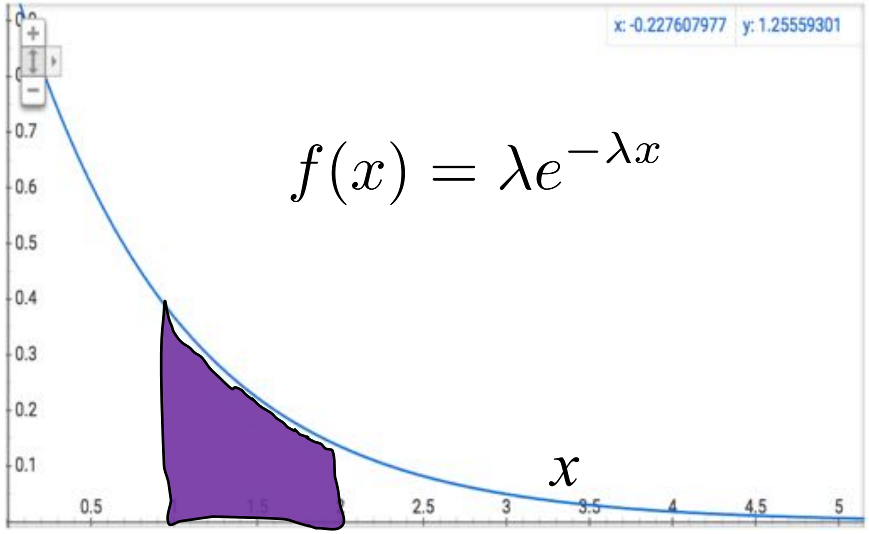
$$F_X(x) = P(X < x)$$

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



# CDF: X ~ Exp($\lambda$ = 1)

*Probability*
density
function



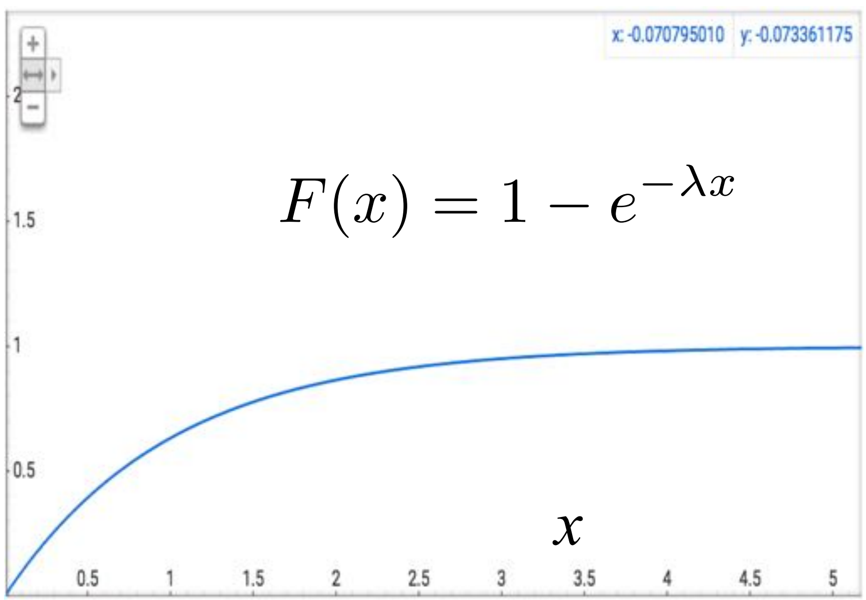

$$P(1 < X < 2)$$

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

*Cumulative*
density
function

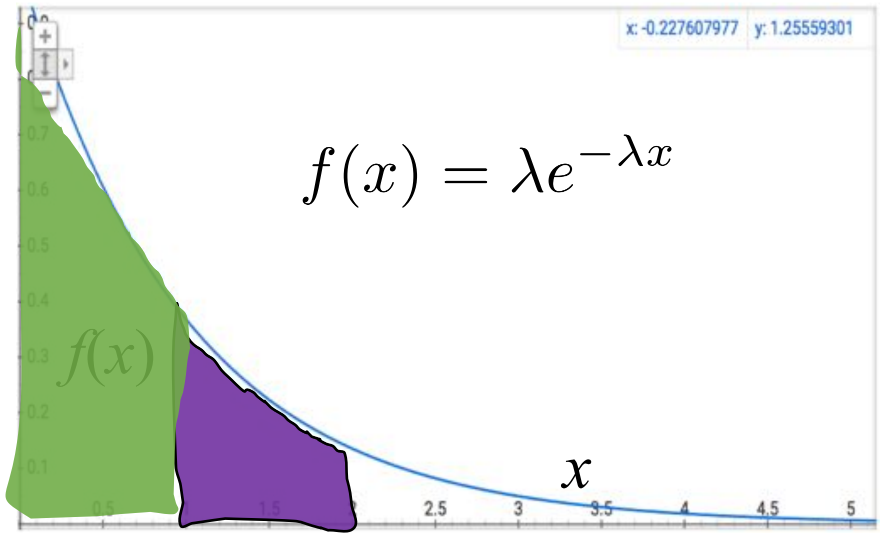
$$F_X(x) = P(X < x)$$

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



# CDF: X ~ Exp($\lambda$ = 1)

*Probability* density function




P(1 < X < 2)

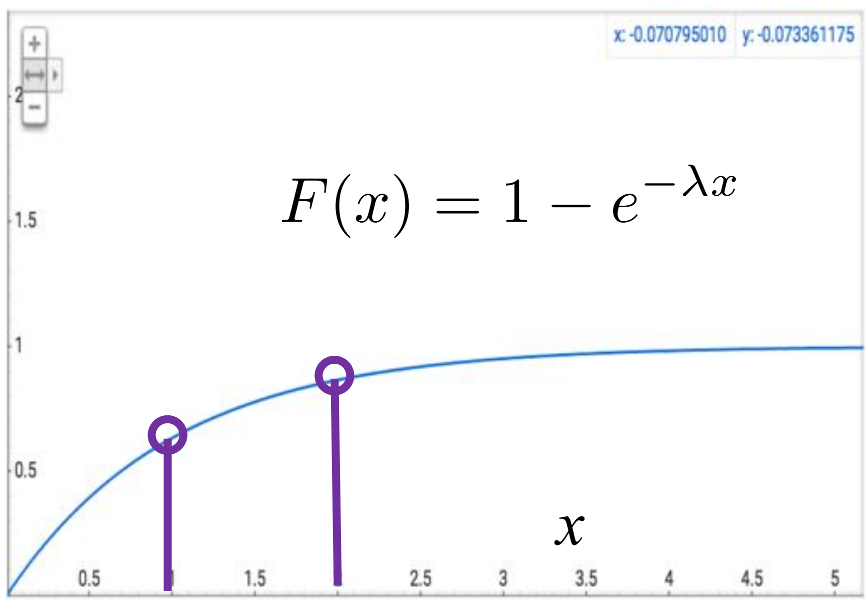
$$= \int_{x=1}^{2} f(x)\, dx$$

or

*Cumulative* density function

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

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




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

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

$$\approx 0.23$$

# 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 major earthquake in the next 4 years?**

$Y$ = Years until the next earthquake of magnitude 8.0+

---

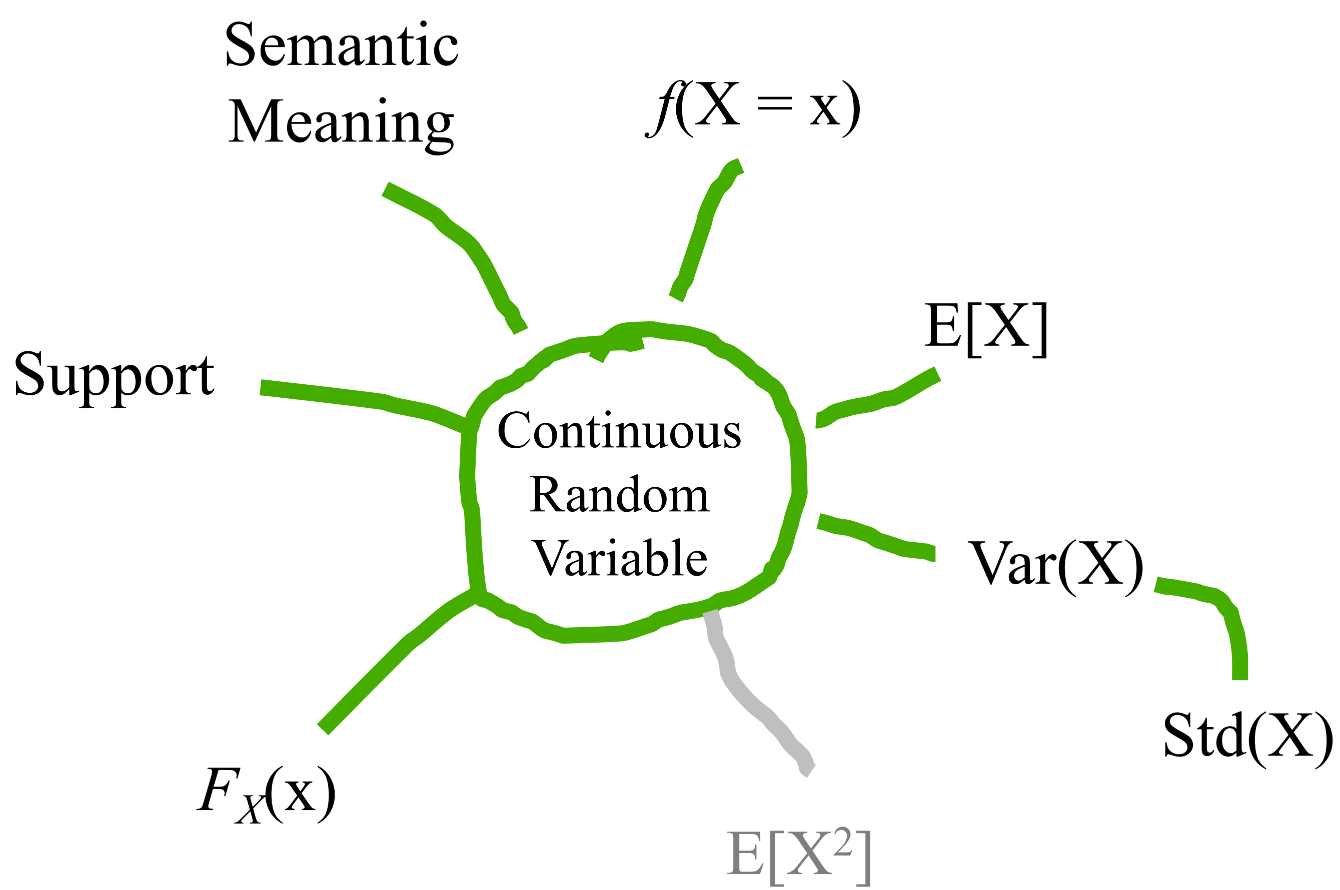
$$Y \sim \text{Exp}(\lambda = 0.002) \qquad F(y) = 1 - e^{-0.002y}$$

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

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 ~ Exp(λ = 1/5),   since E[X] = 1/λ = 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 = # hours of use until your laptop dies
  - On average, laptops die after 5000 hours of use
  - X ~ Exp(λ = 1/5000), since E[X] = 1/λ = 5000
  - You use your laptop 5 hours/day.
  - What is P(your laptop lasts 4 years)?
  - That is: P(X > (5)(365)(4) = 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 coterming:

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

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

# Exponential is Memoryless

- X = time until some event occurs
  - X ~ Exp(λ)
  - What is P(X > s + t | X > s)?

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

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

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

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

# Disk Crashes

- X = 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)?

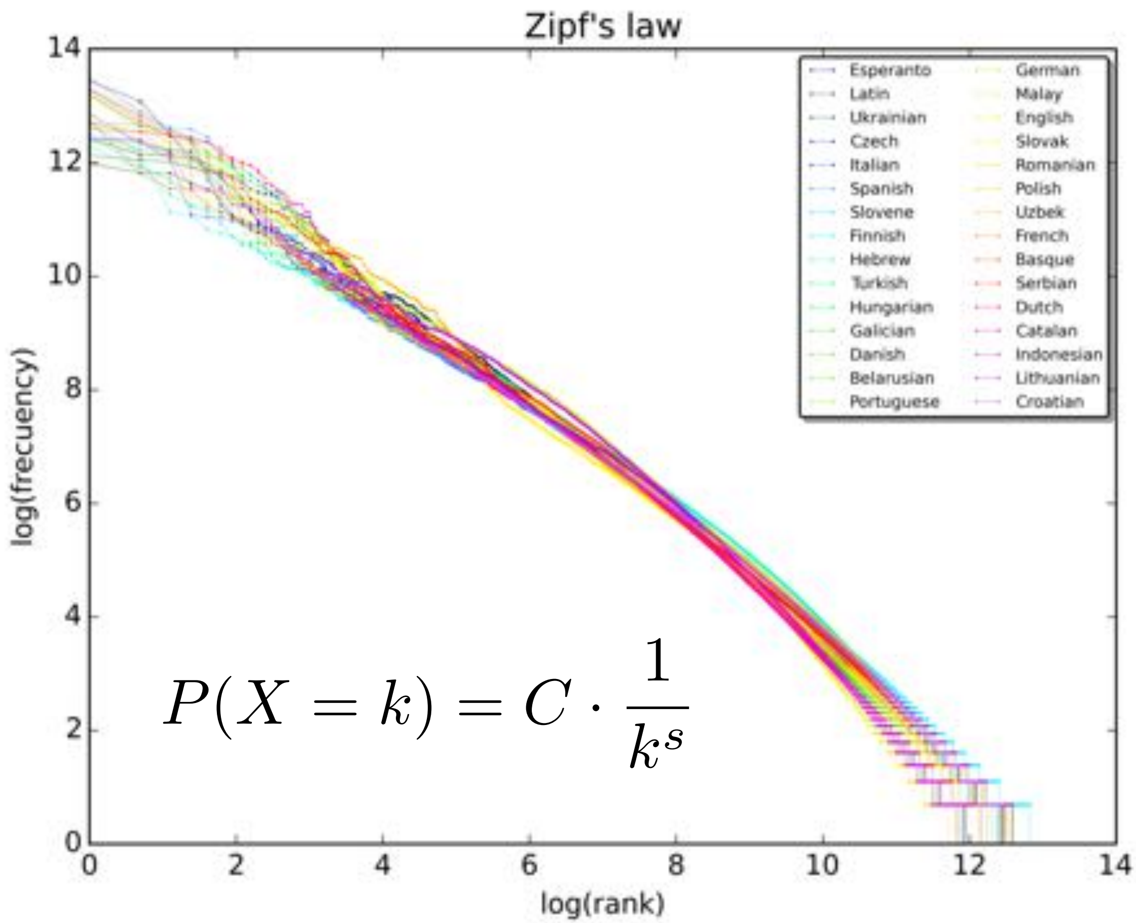
$$F(150) - F(50) = \int_{50}^{150} \frac{1}{100} e^{-x/100} dx = -e^{-x/100} \Big|_{50}^{150} = -e^{-3/2} + e^{-1/2} \approx 0.383$$

  - What is P(X < 10)?

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

# Zipf Random Variable

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




$$P(X = k) = C \cdot \frac{1}{k^s}$$