Continuous Variables

Chris Piech
CS109, Stanford University

Announcements

1. Freya was born on 1.23.21. Thank you!
2. Quiz #1 on Feb 3rd, take home
3. Chris is back at office hours today
4. AAAI 2021 starts today. Using Radio Archives for Speech Recognition of Low Resource Languages.

   https://www.scientificamerican.com/article/why-ai-needs-to-be-able-to-understand-all-the-worlds-languages/

Quick slide reference

Fun challenge: time to delivery tool

A clue: There must be a better way!
(a) Information is being ignored.
(b) Real world impact is huge

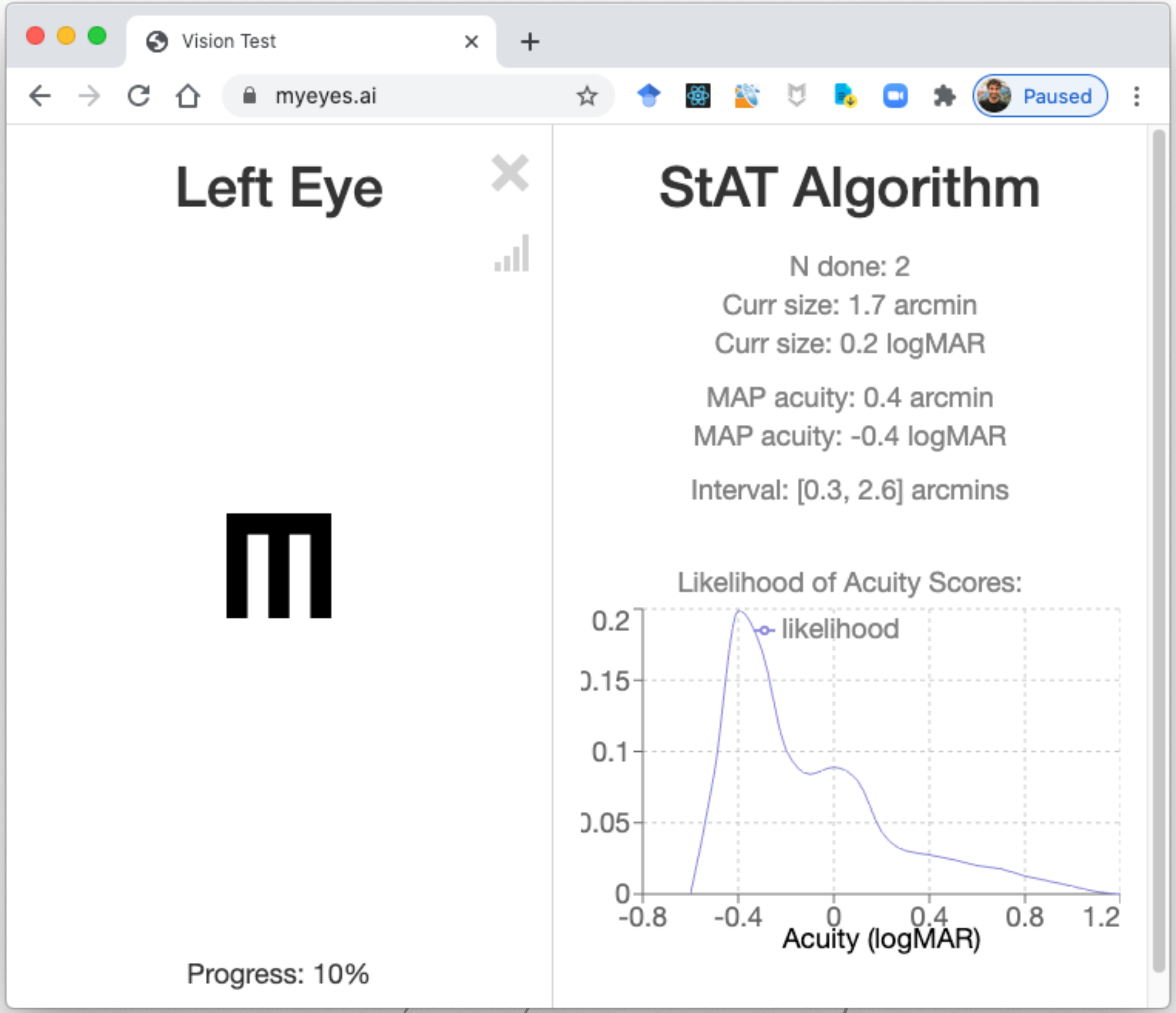
This is often what the first step of a great probability project looks like.

Next steps: Does the tool already exist? Can we find (or ask for) a dataset? Could we build a model based off research?

1906 Earthquake
Magnitude 7.8

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

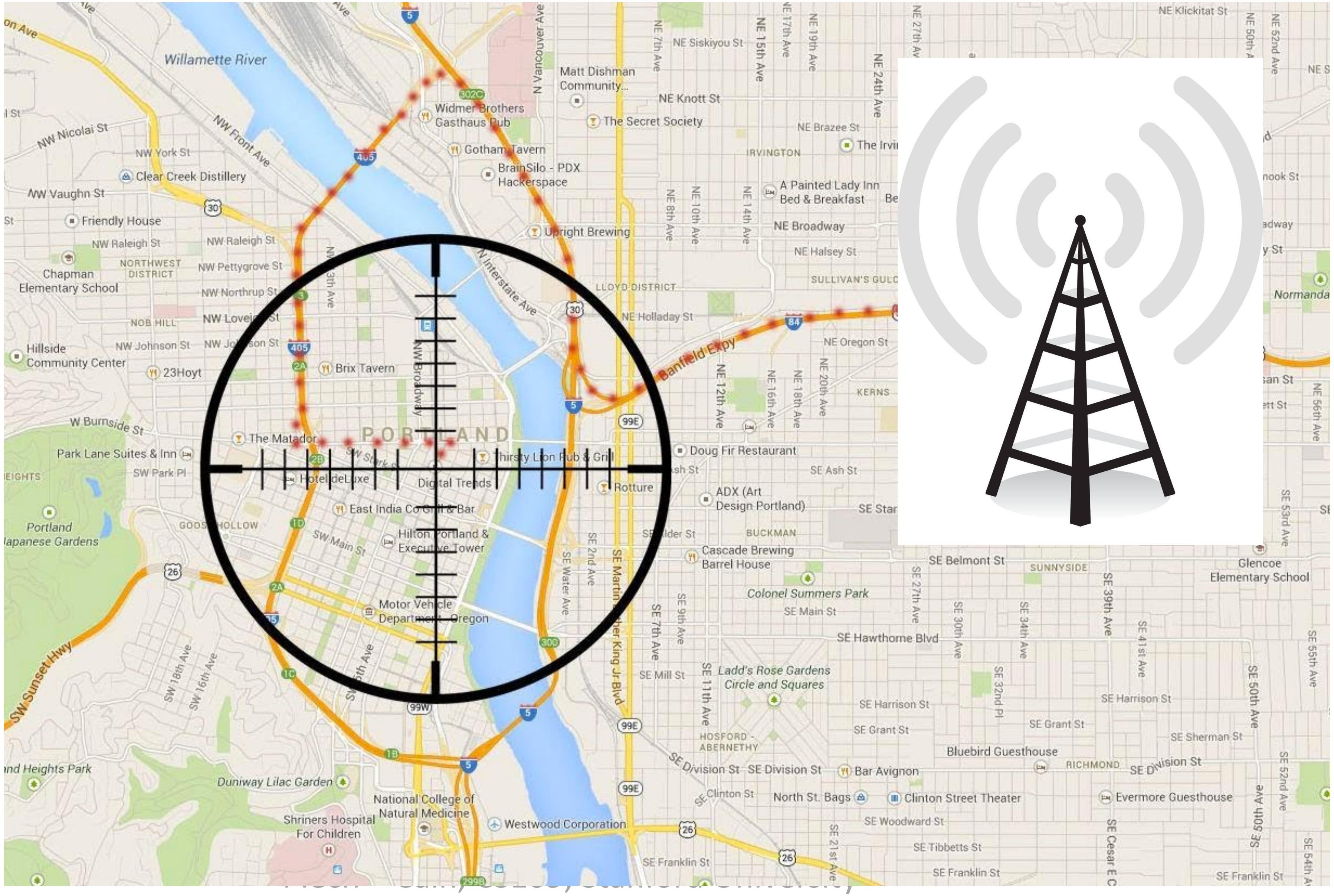
How can you represent belief in "ability to see"



Quick slide reference

Big hole in our knowledge

Not all variables 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)?

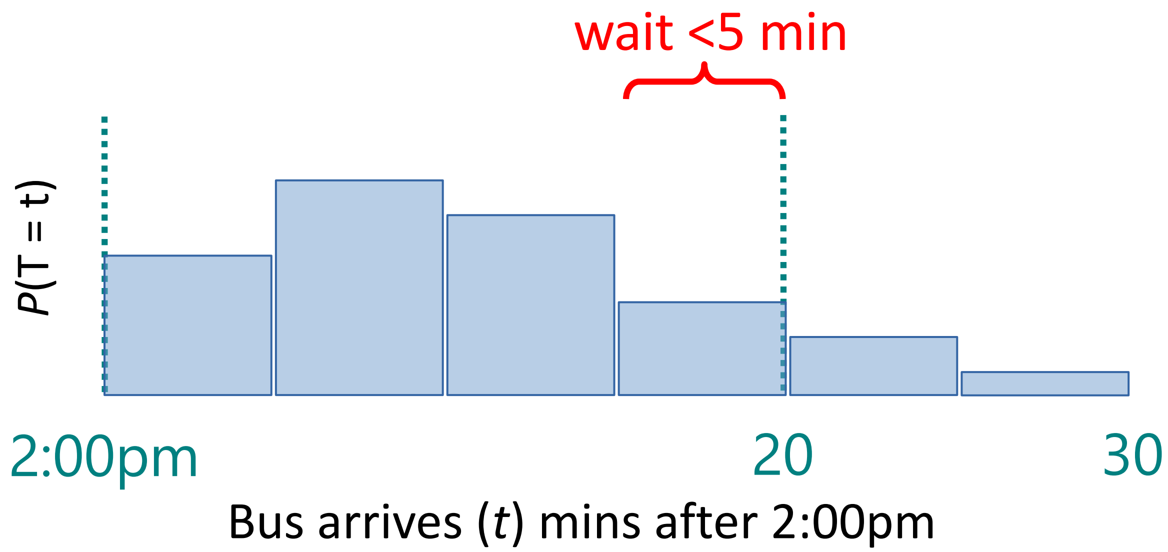
What is the probability that the bus arrives at:
2:17pm and 12.12333911102389234 seconds?

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)?

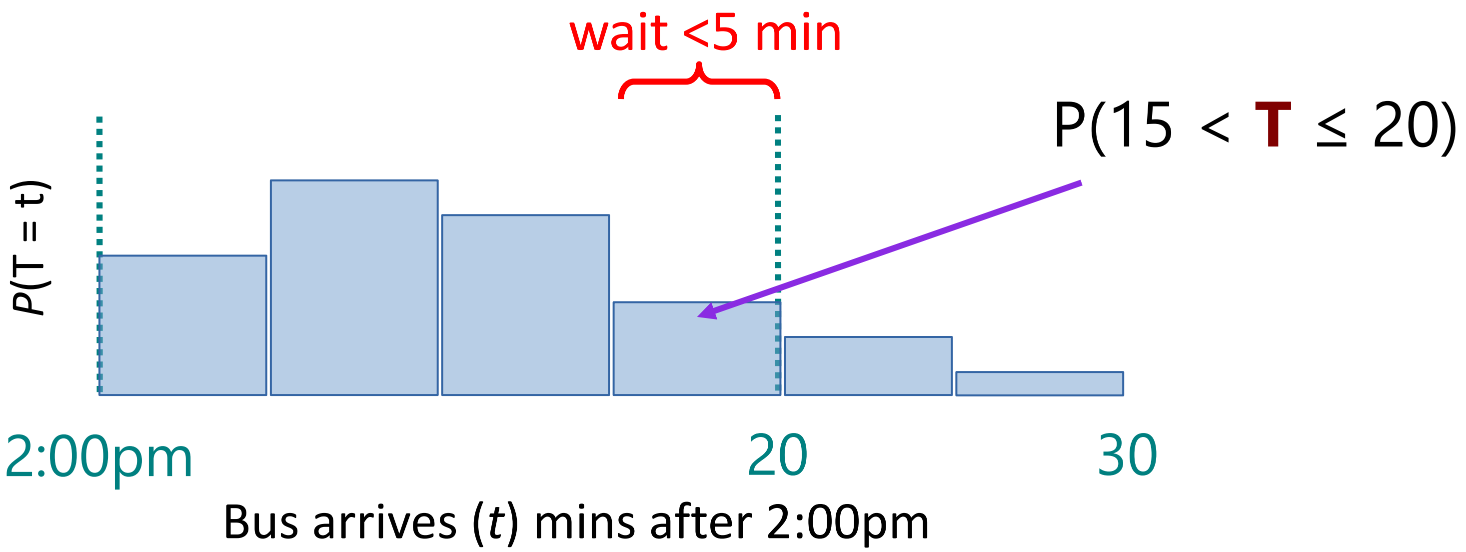


image: Haha169

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)?

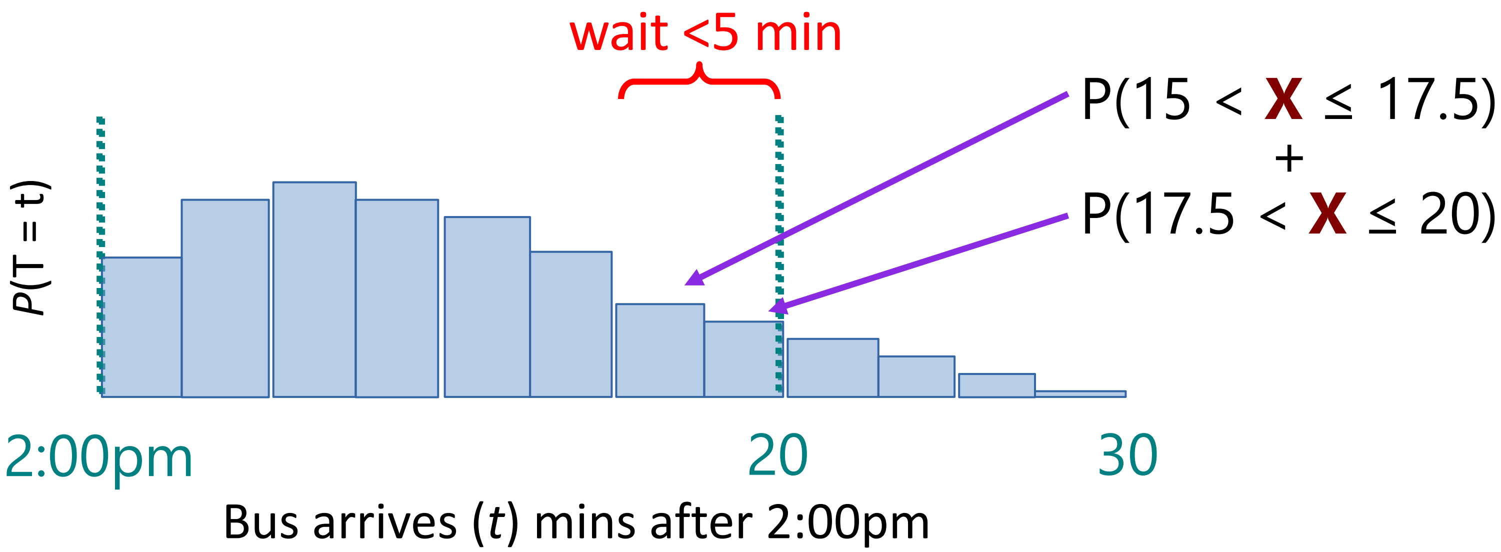


image: [Haha169]

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)?

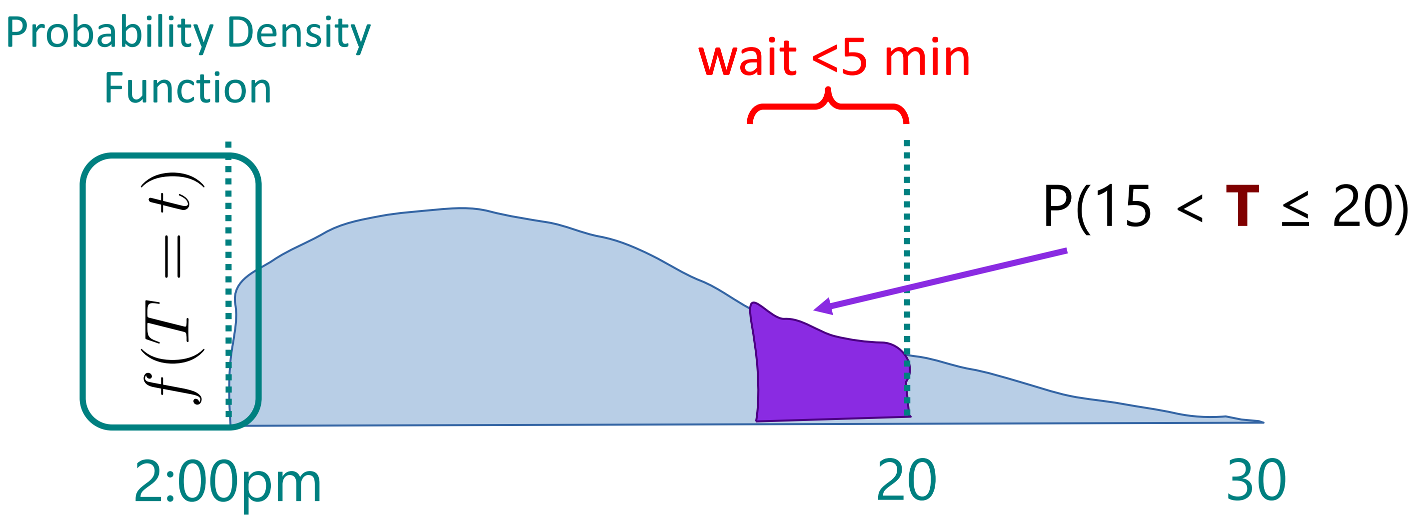


image: Haha169

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!

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

This is another way to write the PDF

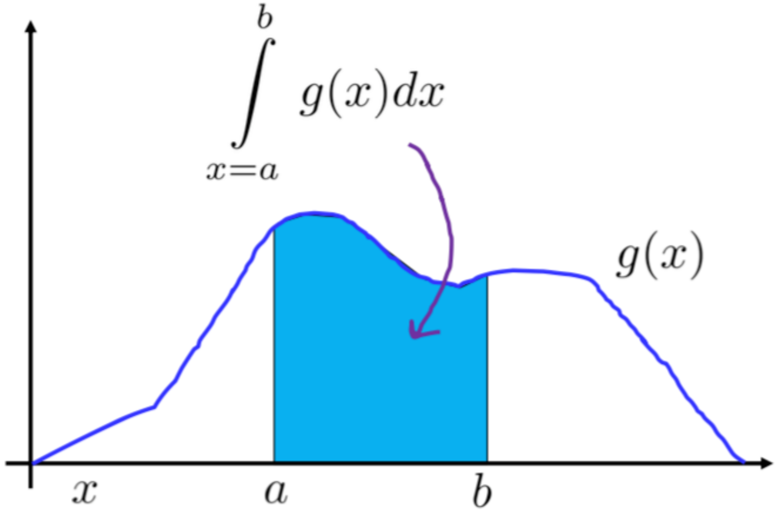
Probability Density Function

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



Integral = area under a curve

Loving, not scary

Continuous RV definition

A random variable X is continuous if there is a **probability density function** $f(x) \geq 0$ such that for $-\infty < x < \infty$:

$$P(a \leq X \leq b) = \int_a^b f(x)\,dx$$

Integrating a PDF must always yield valid probabilities, and therefore the PDF must also satisfy

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

Also written as: $f_X(x)$

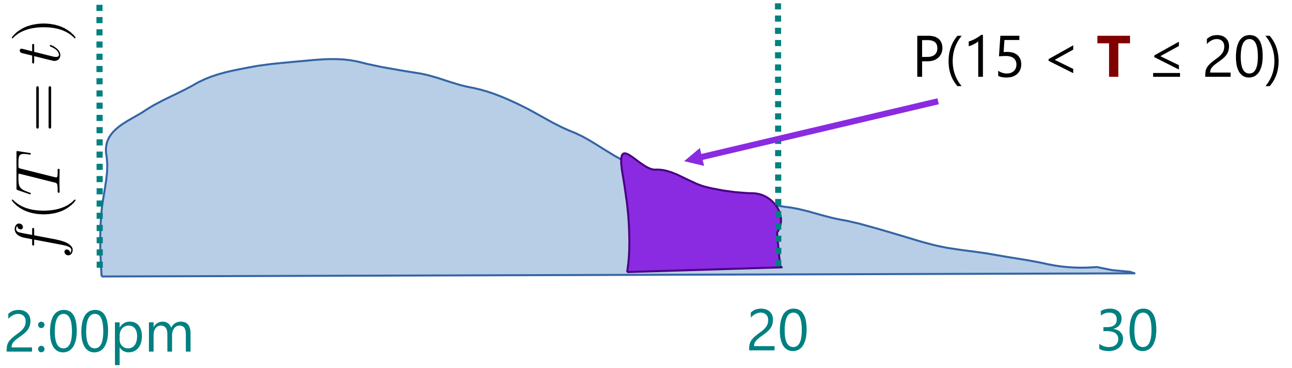
Riding the Marguerite

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



Bus arrives (t) mins after 2:00pm

image: [Haha169]

Properties of PDFs

The integral of a PDF gives a probability. Thus:

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

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

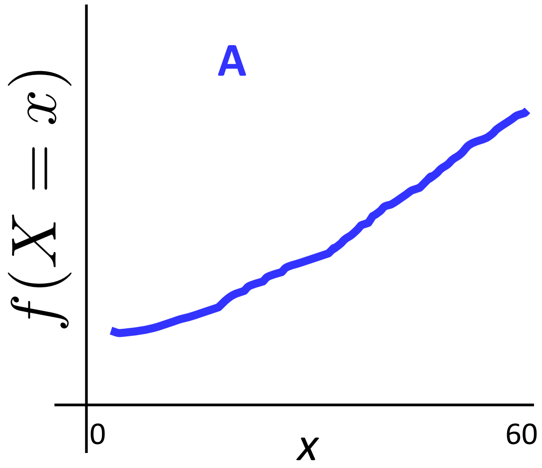
What do you get if you integrate over a probability *density* function?

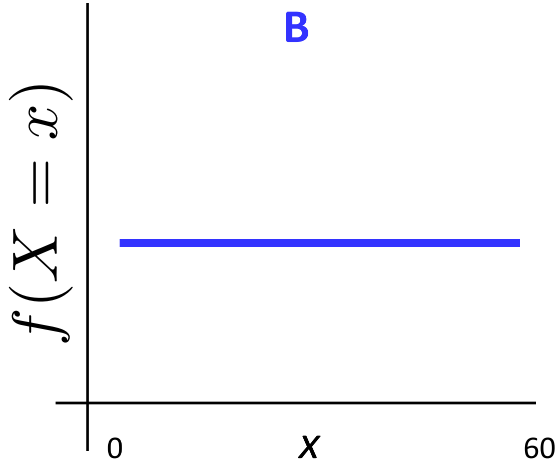
A probability!

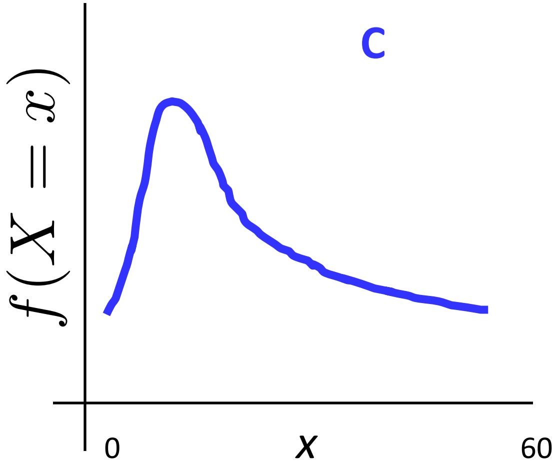
Probability Density Function

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:





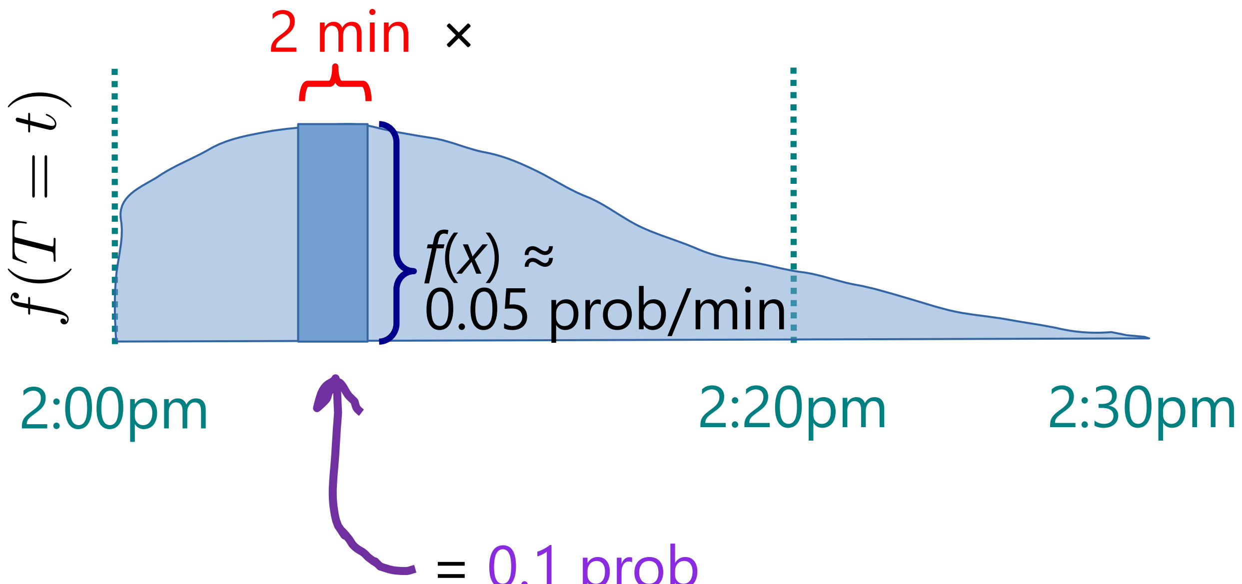


Which of these represent that you think the arrival is more likely to be close to 3:00pm

The ratio of probability densities is meaningful

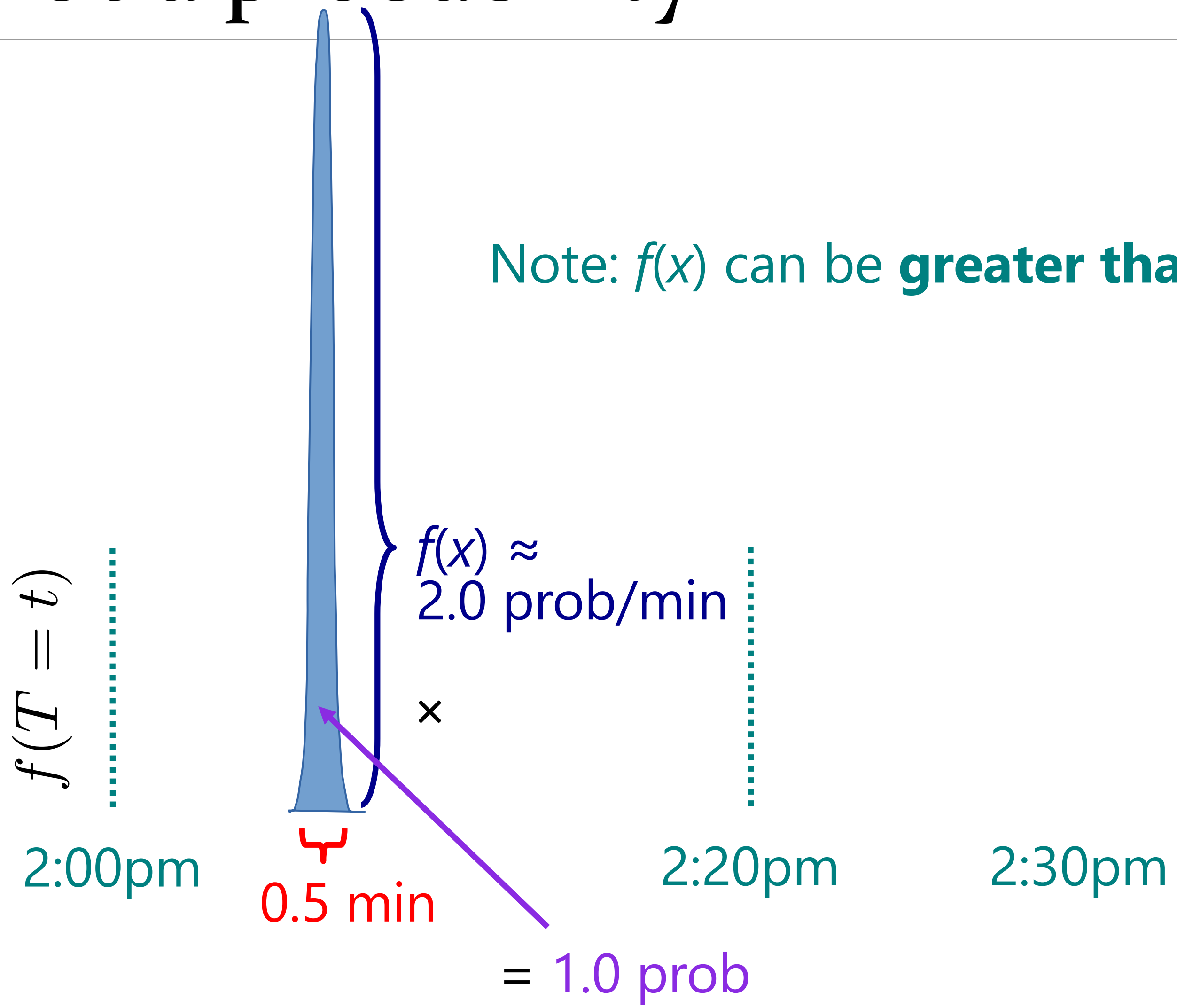
$f(X = x)$ is **not** a probability

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 $\mathrm{P}(X = x)$

"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?*

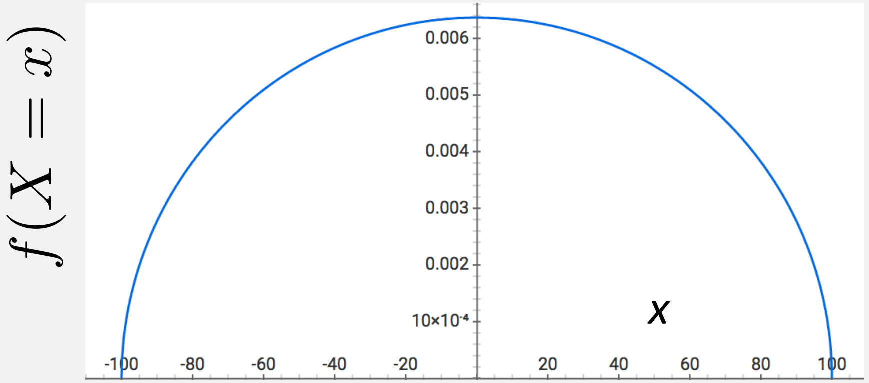
* With help from Wigner, Chris is going to rephrase this problem

Simple Example from Quantum Physics

Let X be a continuous random variable[1]

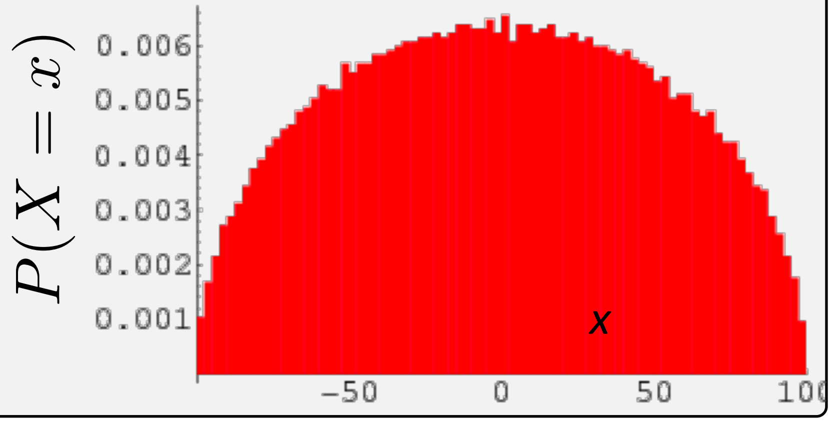
Theory

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



Practice

From simulations



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

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

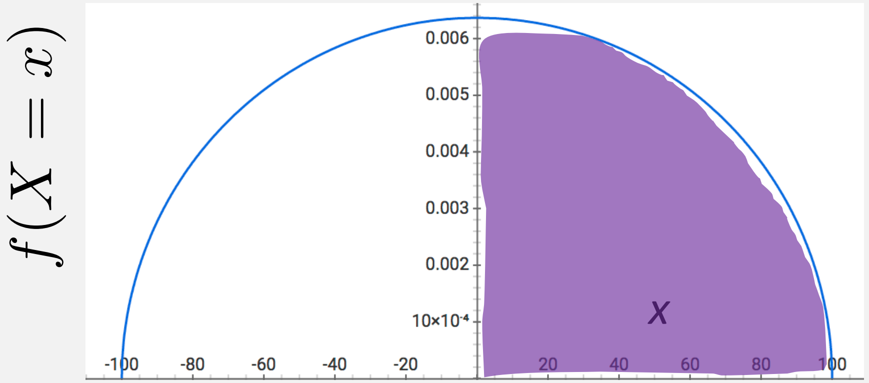
https://en.wikipedia.org/wiki/Wigner_semicircle_distribution

Simple Example from Quantum Physics

Let X be a continuous random variable[1]

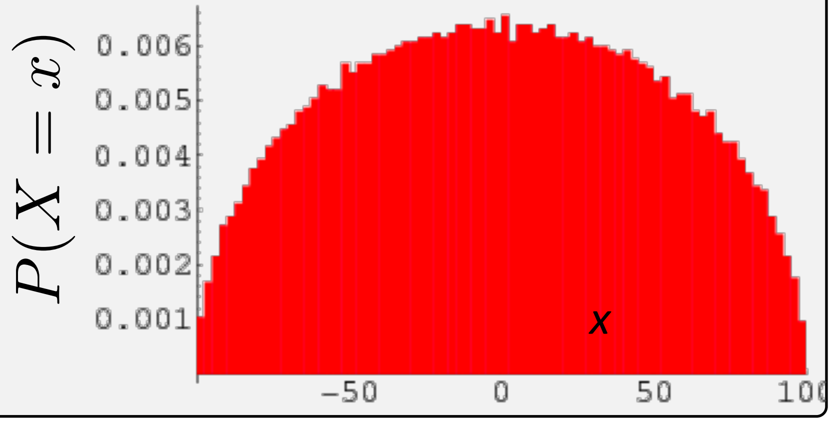
Theory

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



Practice

From simulations



Approach #1: Integrate over the PDF

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

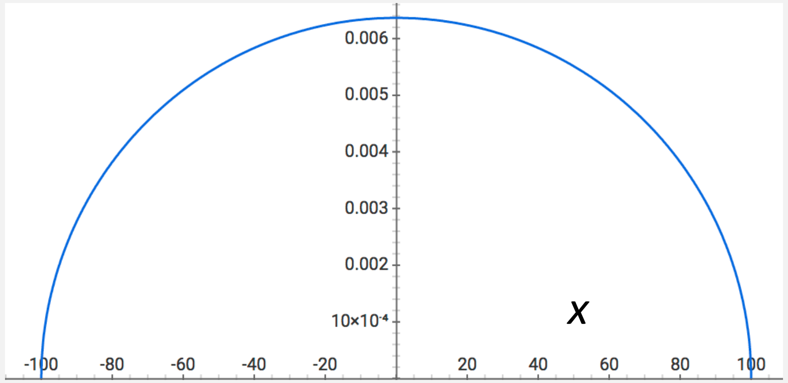
Simple Example from Quantum Physics

Let X be a continuous random variable[1]

Theory

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

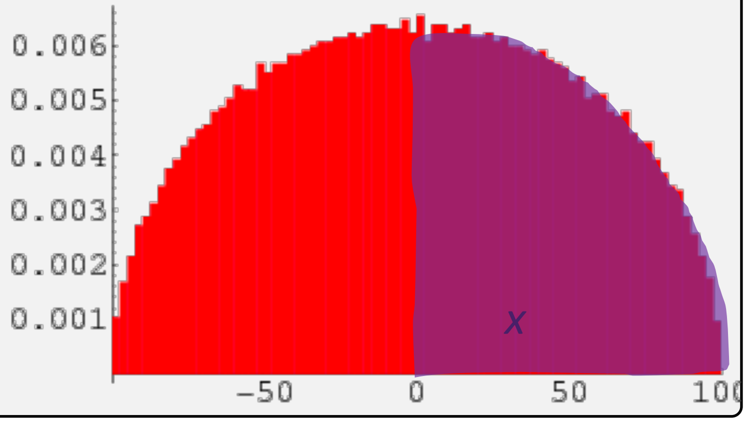
$f(X = x)$



Practice

From simulations

$P(X = x)$



Approach #2: Discrete Approximation

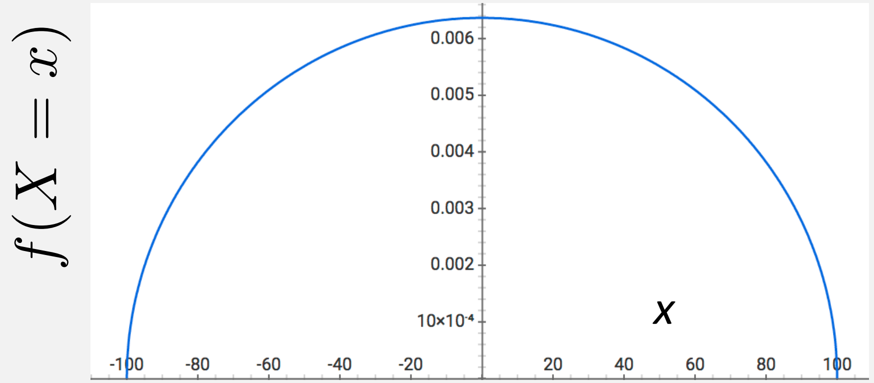
$$P(X > 0) \approx \sum_{i=0}^{100} P(X = i)$$

Simple Example from Quantum Physics

Let X be a continuous random variable[1]

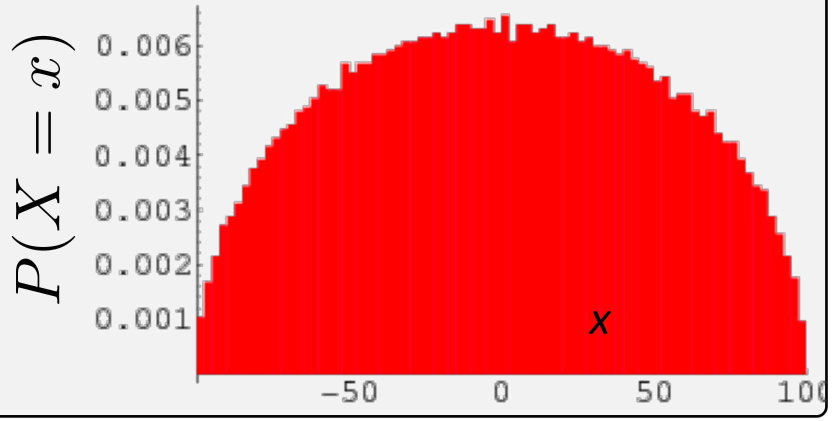
Theory

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



Practice

From simulations



Approach #3: Know Semi-Circles

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

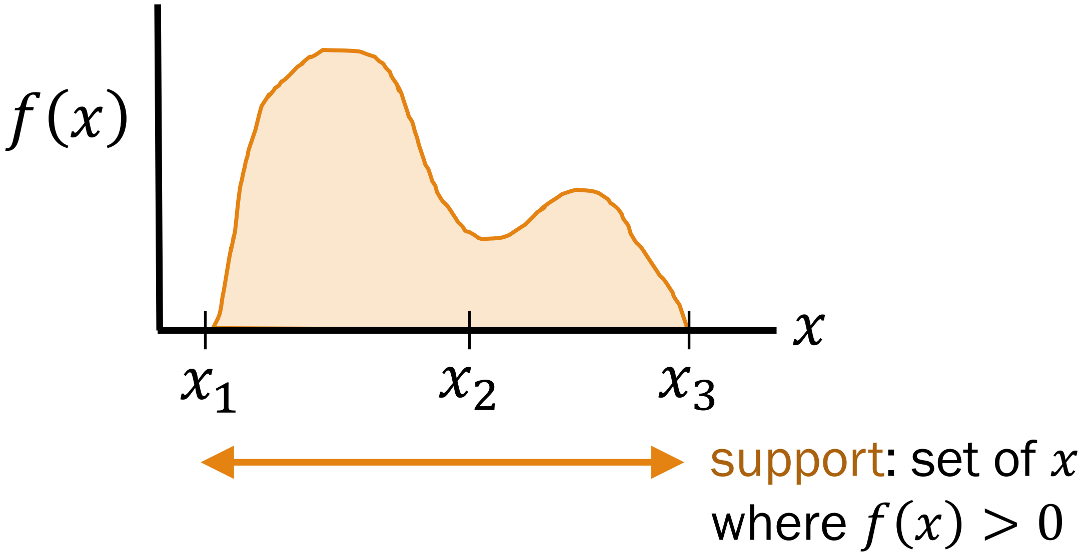
What do you get if you integrate over a probability *density* function?

A probability!

PDF Properties

For a continuous RV X with PDF f,

$$P(a \le X \le b) = \int_a^b f(x)\,dx$$



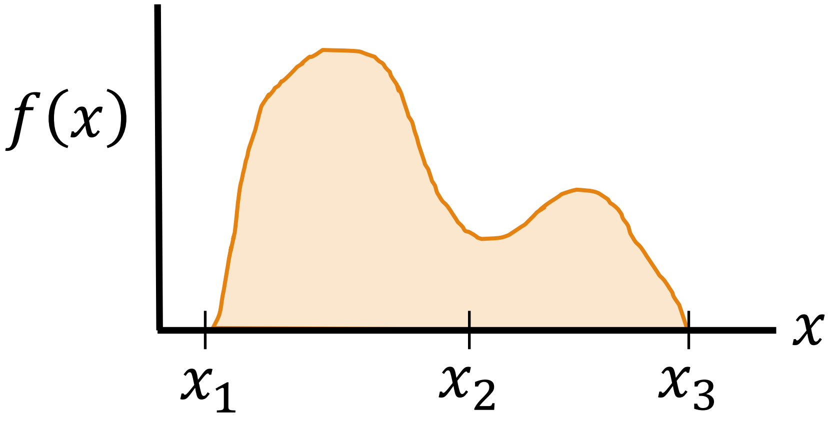
True/False:

1. $P(X = c) = 0$
2. $P(a \le X \le b) = P(a < X < b) = P(a \le X < b) = P(a < X \le b)$
3. $f(x)$ is a probability
4. In the graphed PDF above,
   $P(x_1 \le X \le x_2) > P(x_2 \le X \le x_3)$

PDF Properties

For a continuous RV X with PDF f,

$$P(a \leq X \leq b) = \int_a^b f(x)\,dx$$



support: set of x where $f(x) > 0$

True/False:

1. $P(X = c) = 0$

   Interval width $dx \to 0$

2. $P(a \leq X \leq b) = P(a < X < b) = P(a \leq X < b) = P(a < X \leq b)$

3. $f(x)$ is a probability

4. In the graphed PDF above, $P(x_1 \leq X \leq x_2) > P(x_2 \leq X \leq x_3)$

   Compare area under the curve f

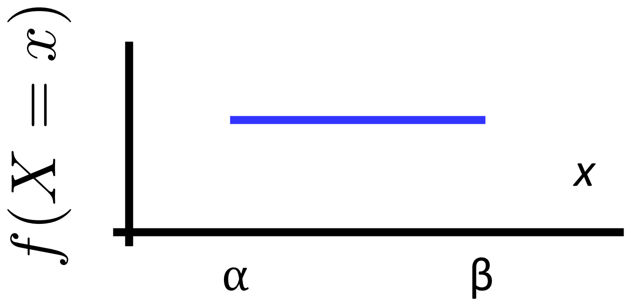
Uniform Random Variable

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

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

Probability Density

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



Properties

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

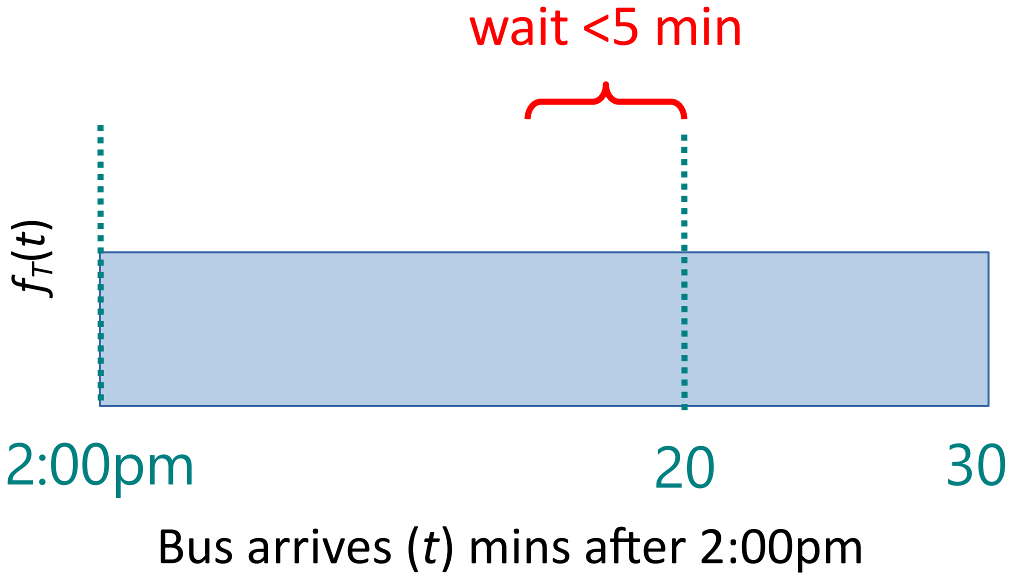
$$\mathrm{Var}(X) = \frac{(\beta - \alpha)^2}{12}$$

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

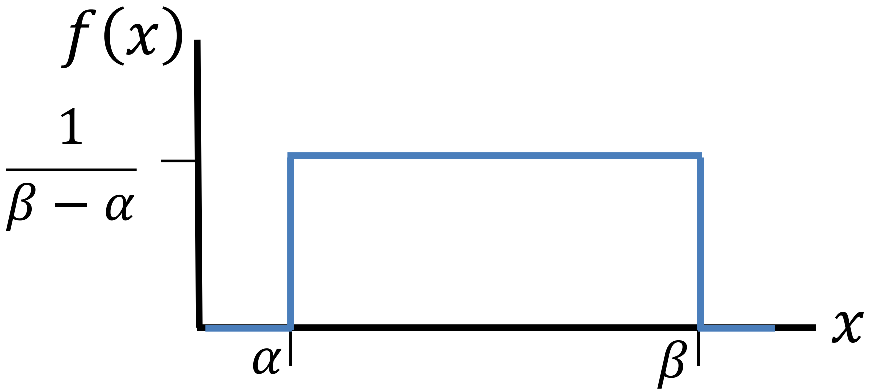
$$E[X] = \int_{-\infty}^{\infty} x \cdot f(x)\ dx$$

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

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

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

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



just average the start and end!

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

Grid of random variables

	Number of successes	Time until success	
One trial	Ber(p)	Geo(p)	One success
	$n = 1$	$r = 1$	
Several trials	Bin(n, p)	NegBin(r, p)	Several successes
Interval of time	Poi(λ)	Exp(λ)	Interval of time to first success

Exponential Random Variable

Consider an experiment that lasts a duration of time until success occurs.

def An **Exponential** random variable X is the amount of time until success.

$$X \sim \text{Exp}(\lambda)$$

Support: $[0, \infty)$

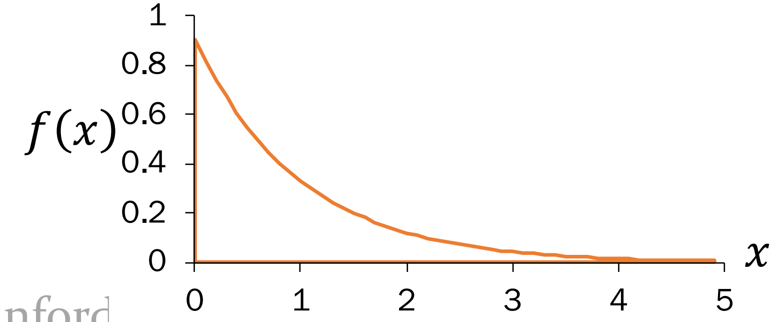
PDF $\quad f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$

Expectation $\quad E[X] = \frac{1}{\lambda}$ (in extra slides)

Variance $\quad \text{Var}(X) = \frac{1}{\lambda^2}$ (on your own)

Examples:

- Time until next earthquake
- Time for request to reach web server
- Time until end of cell phone contract



1906 Earthquake
Magnitude 7.8

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

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

$$\begin{aligned} f_Y(y) &= \lambda e^{-\lambda y} \\ &= 0.002^{-0.002y} \end{aligned}$$

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

*In California, according to the USGS, 2015

Integral Review

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

How Long Until the 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+

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

$$\begin{aligned} f_Y(y) &= \lambda e^{-\lambda y} \\ &= 0.002^{-0.002y} \end{aligned}$$

$$\begin{aligned} P(Y < 30) &= \int_0^{30} 0.002 e^{-0.002y} \, dy \\ &= 0.002 \Big[-500 e^{-0.002y} \Big]_0^{30} \\ &= \frac{500}{500} (-e^{-0.06} + e^0) \approx 0.06 \end{aligned}$$

*In California, according to the USGS, 2015

How Long Until the 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 the 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 CDF

Cumulative Density Function

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

$$x = 2$$

$$0.03125$$

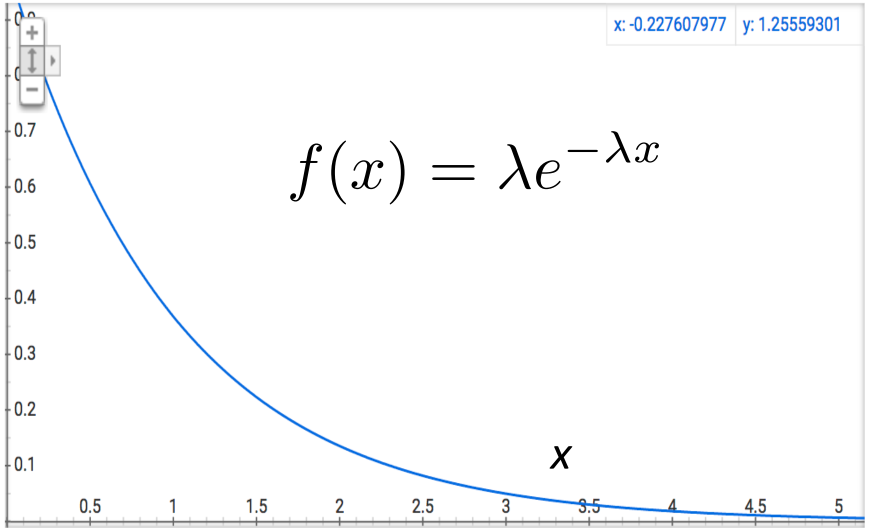
CDF of an Exponential

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

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

Using CDF Example. X is Exp(λ = 1)

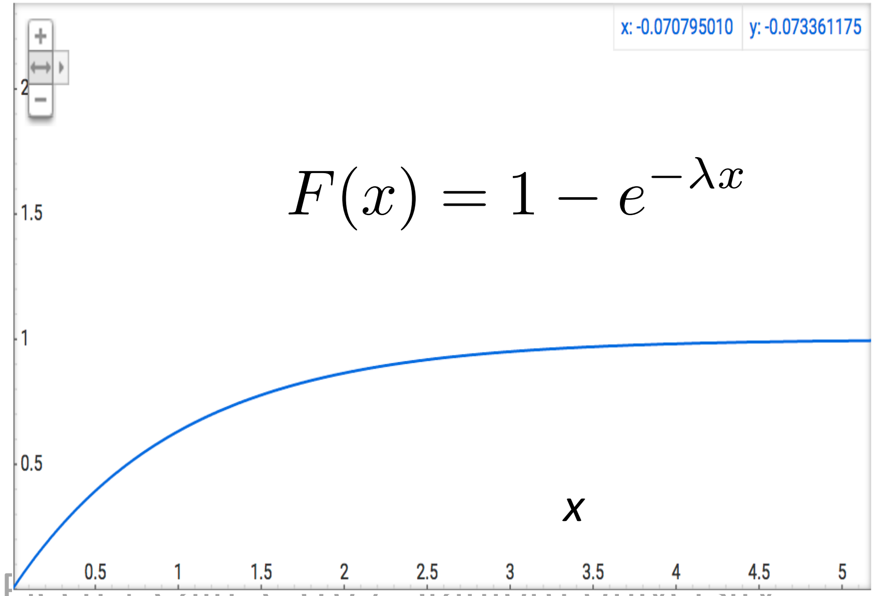
Probability density function



Cumulative density function

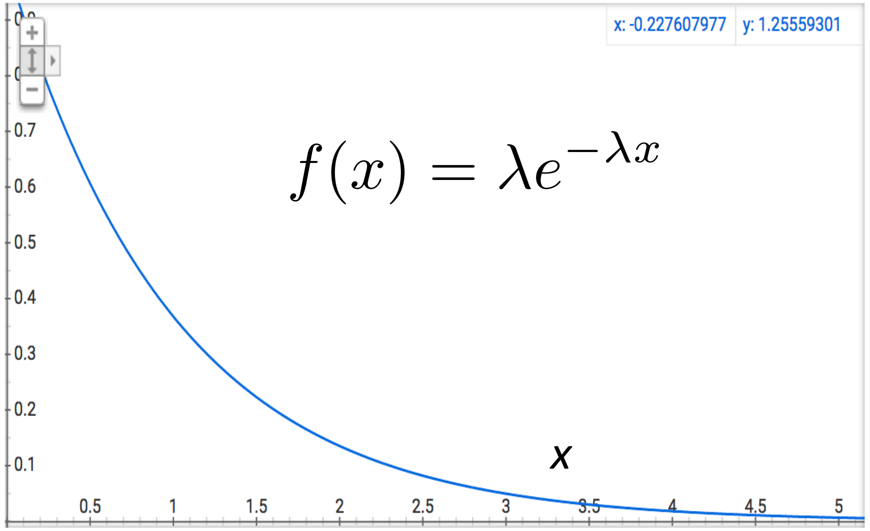
$$F_X(x) = P(X < x)$$

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



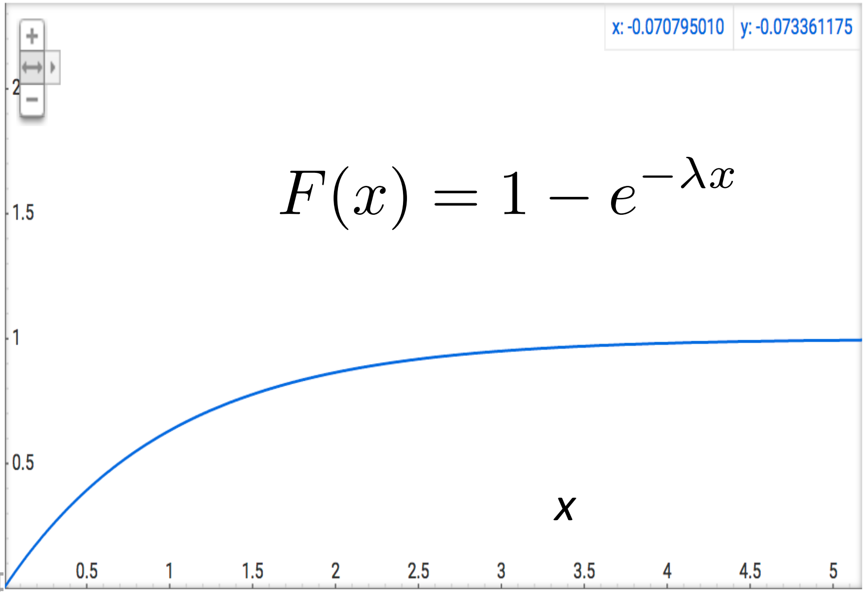
Using CDF Example. X is Exp(λ = 1)

Probability density function



P(X < 2)

Cumulative density function

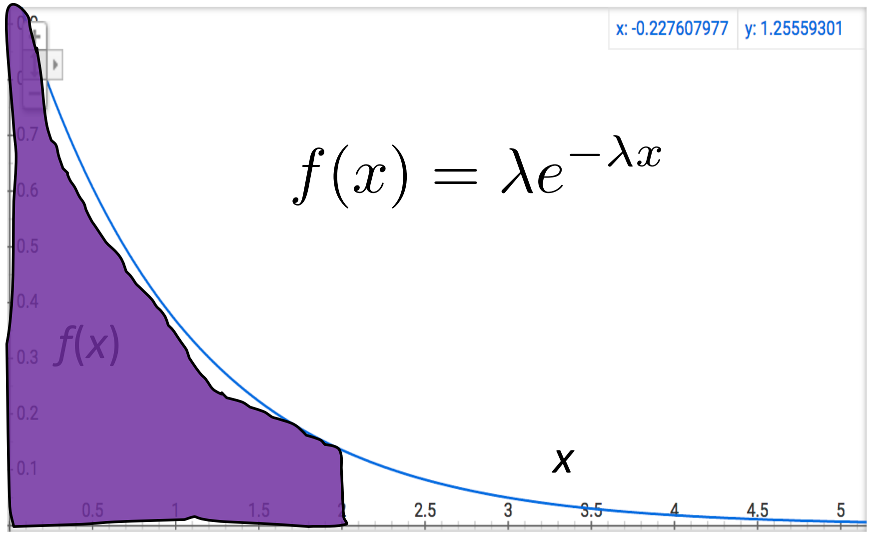


$$F_X(x) = P(X < x)$$

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

Using CDF Example. X is Exp(λ = 1)

Probability density function



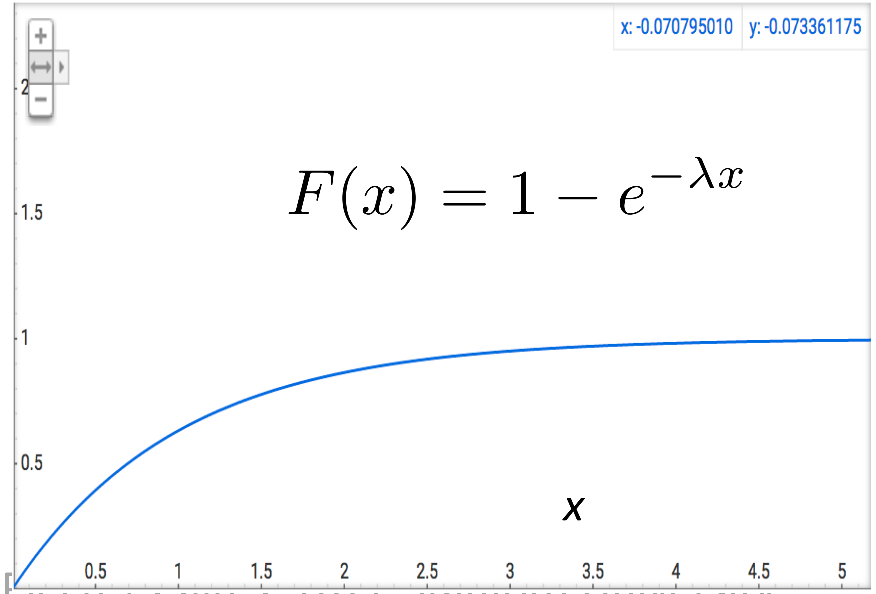
P(X < 2)

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

Cumulative density function

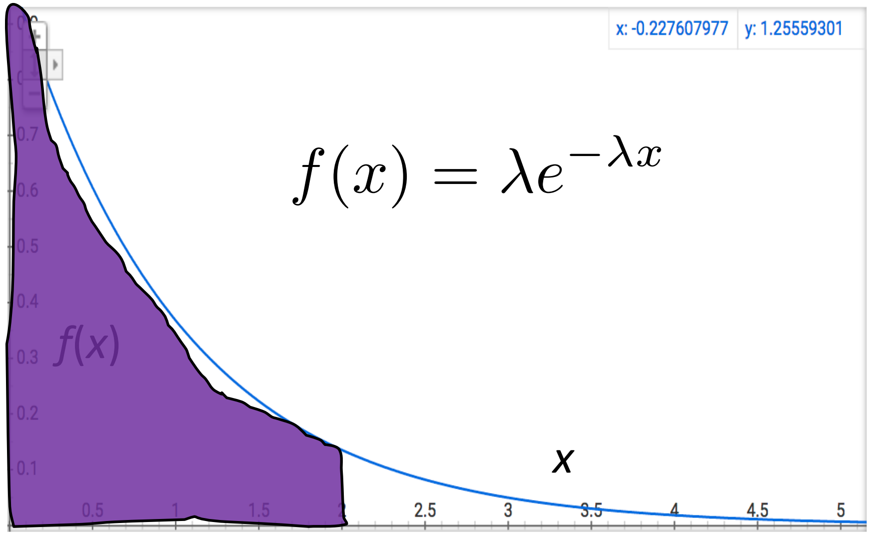
$$F_X(x) = P(X < x)$$

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



Using CDF Example. X is Exp(λ = 1)

Probability
density
function



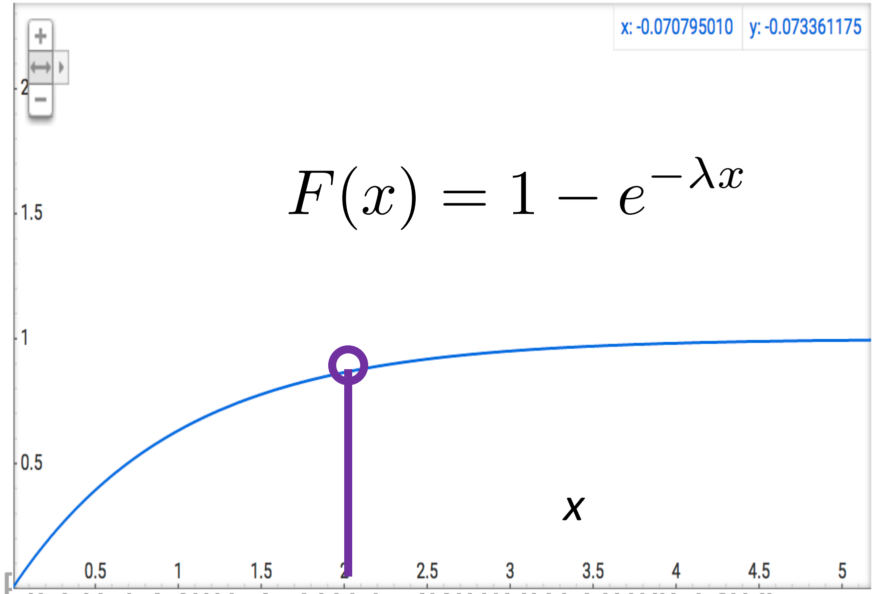
P(X < 2)

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

Cumulative
density function

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

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



or

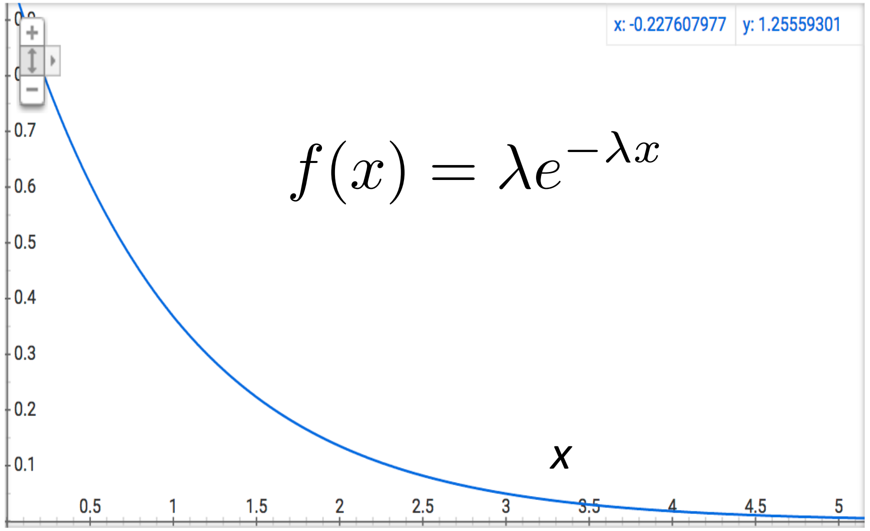
$$= F(2)$$

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

$$\approx 0.84$$

Using CDF Example. X is Exp(λ = 1)

Probability
density
function

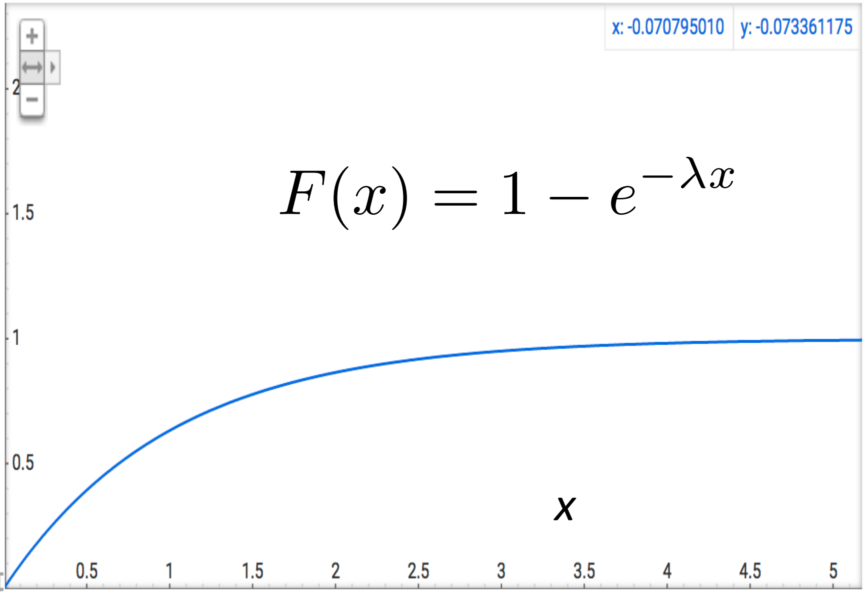


P(X > 1)

Cumulative
density function

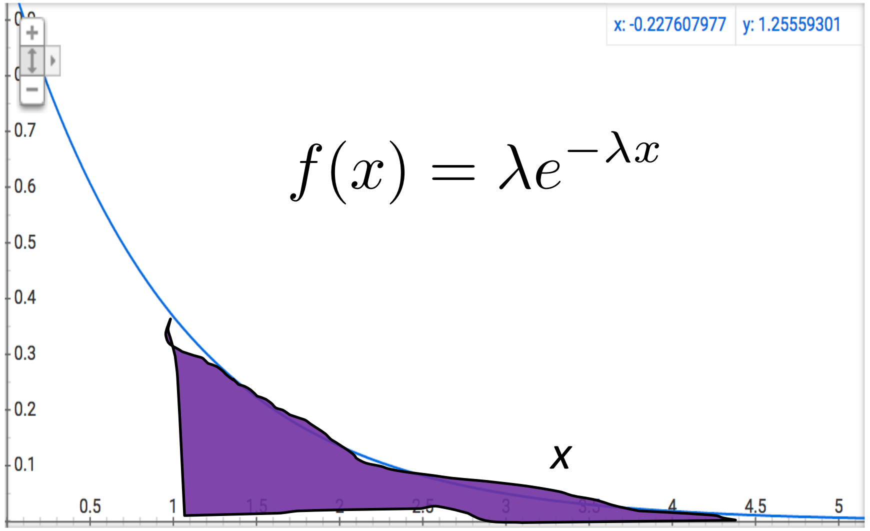
$$F_X(x) = P(X < x)$$

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



Using CDF Example. X is Exp(λ = 1)

Probability density function



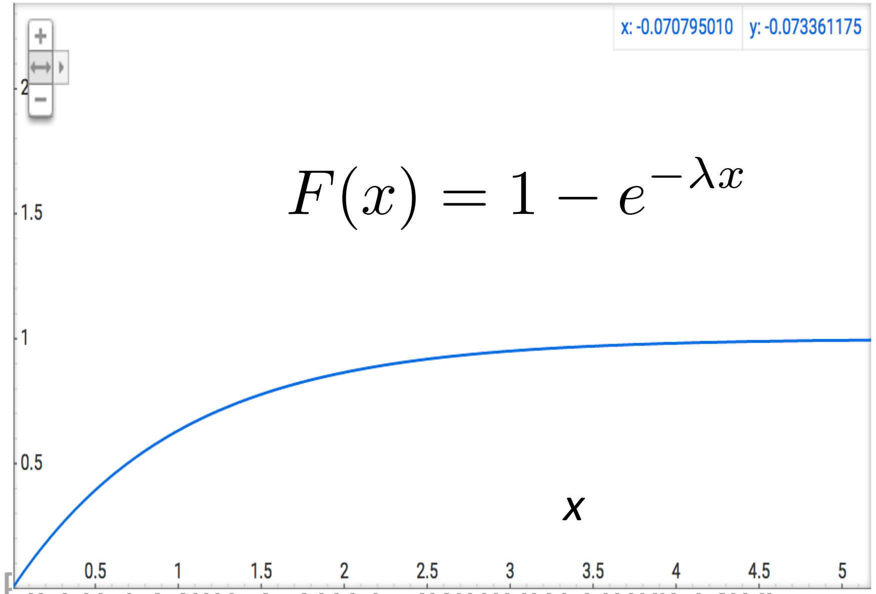
P(X > 1)

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

Cumulative density function

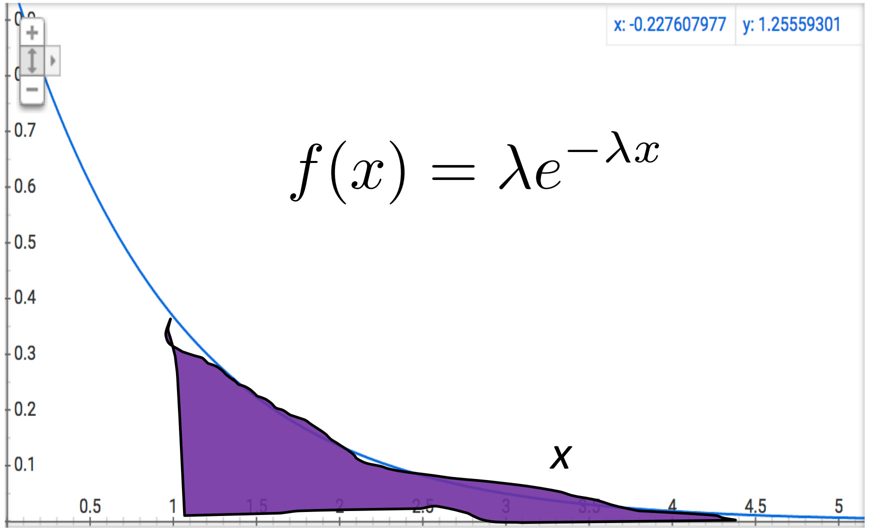
$$F_X(x) = P(X < x)$$

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



Using CDF Example. X is Exp(λ = 1)

Probability density function



P(X > 1)

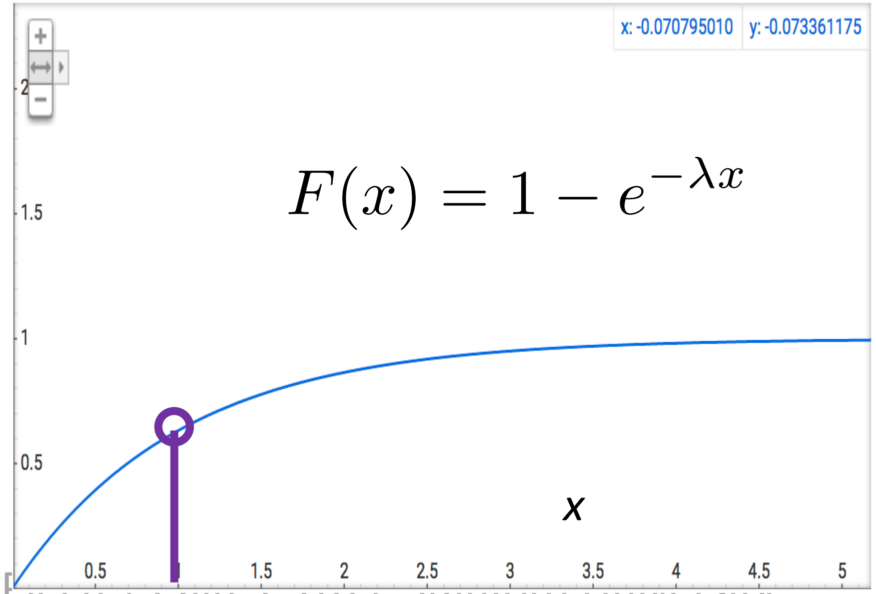
$$= \int_{x=1}^{\infty} f(x)\, dx$$

or

Cumulative density function

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

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



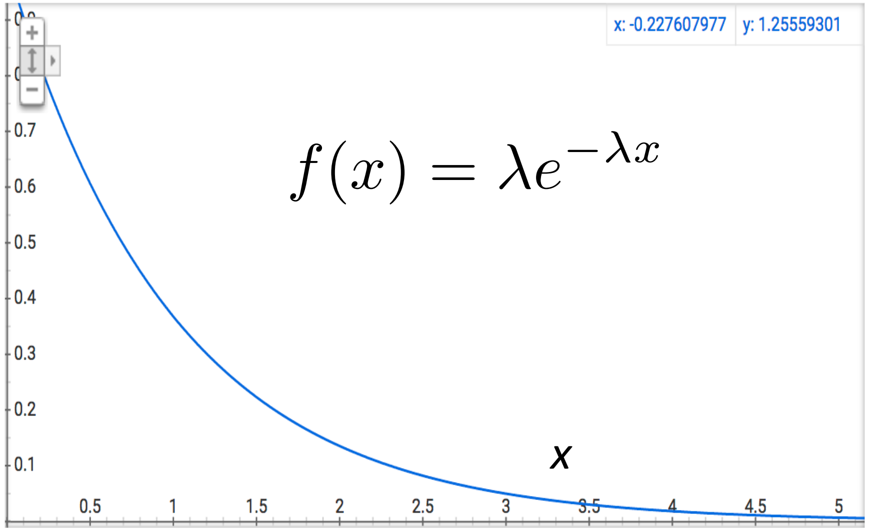
$$= 1 - F(1)$$

$$= e^{-1}$$

$$\approx 0.37$$

Using CDF Example. X is Exp(λ = 1)

Probability density function

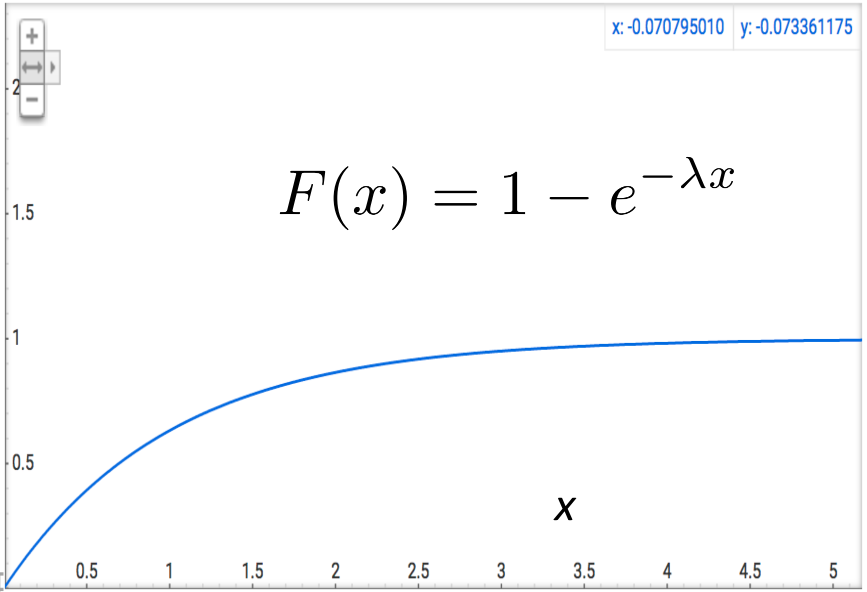


P(1 < X < 2)

Cumulative density function

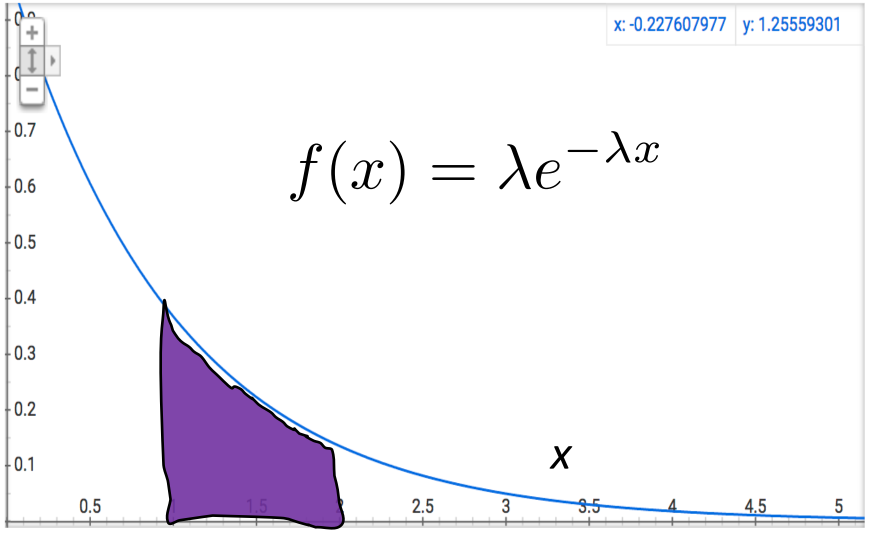
$$F_X(x) = P(X < x)$$

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



Using CDF Example. X is Exp(λ = 1)

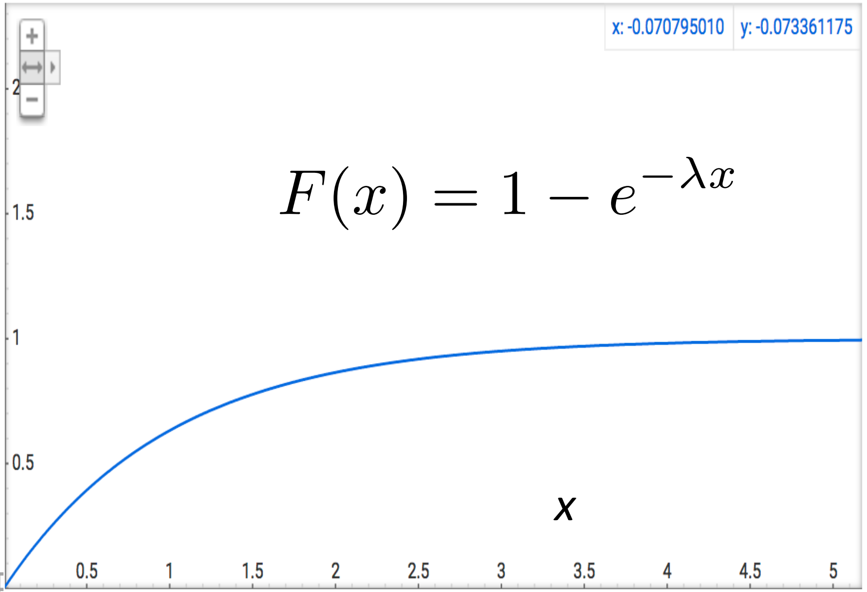
Probability density function



P(1 < X < 2)

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

Cumulative density function

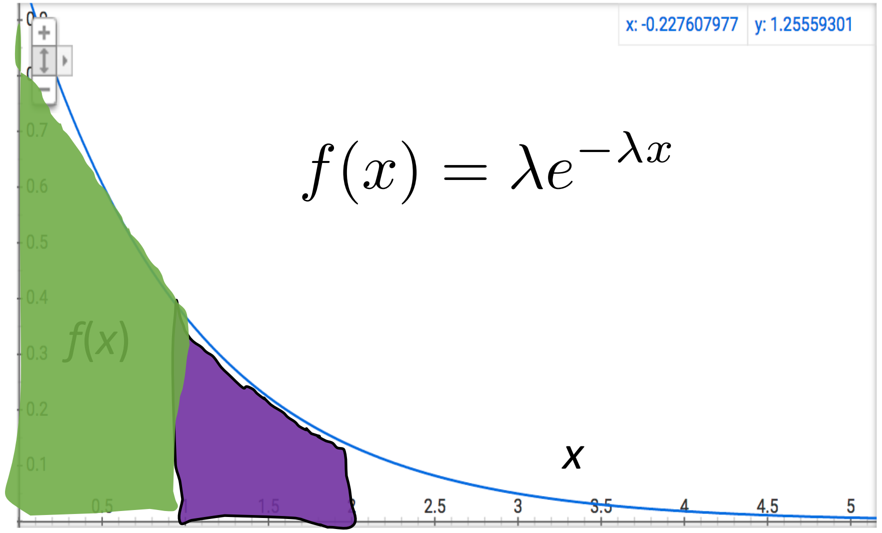


$$F_X(x) = P(X < x)$$

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

Using CDF Example. X is Exp(λ = 1)

Probability
density
function



P(1 < X < 2)

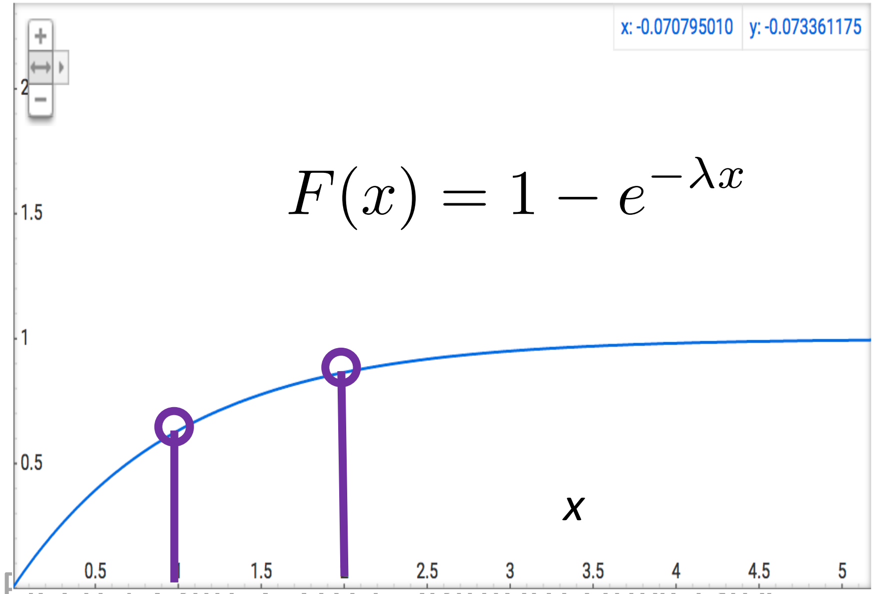
$$= \int_{x=1}^{2} f(x)\, dx$$

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+

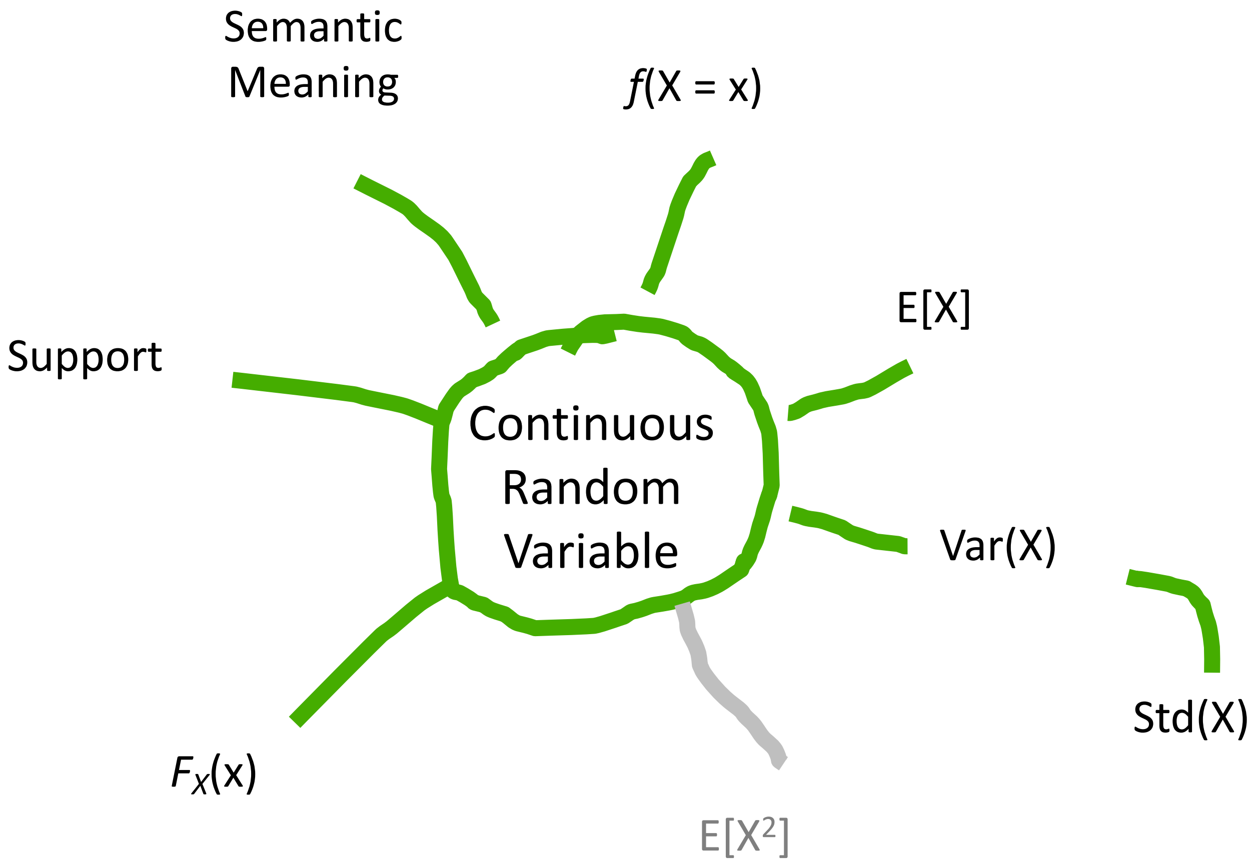
$$Y \sim \text{Exp}(\lambda = 0.002) \qquad F(y) = 1 - e^{-0.002y}$$

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

Feeling lucky?

*According to USGS, 2015

Properties for Continuous Random Variable



Extra Problem

Website visits

$X \sim \text{Exp}(\lambda)$ $\quad E[X] = 1/\lambda$
$\quad F(x) = 1 - e^{-\lambda x}$

Suppose a visitor to your website leaves after X minutes.

- On average, visitors leave the site after 5 minutes.
- The length of stay, X, is exponentially distributed.

1. $P(X > 10)$?

2. $P(10 < X < 20)$?

Website visits

$X \sim \text{Exp}(\lambda)$ $\quad E[X] = 1/\lambda$ $\quad F(x) = 1 - e^{-\lambda x}$

Suppose a visitor to your website leaves after X minutes.

- On average, visitors leave the site after 5 minutes.
- The length of stay, X, is exponentially distributed.

1. $P(X > 10)$?

Define

X: when visitor leaves

$X \sim \text{Exp}(\lambda = 1/5 = 0.2)$

Solve

$$P(X > 10) = 1 - F(10)$$

$$= 1 - \left(1 - e^{-10/5}\right) = e^{-2} \approx 0.1353$$

2. $P(10 < X < 20)$?

Solve

$$P(10 < X < 20) = F(20) - F(10)$$

$$= (1 - e^{-4}) - (1 - e^{-2}) \approx 0.1170$$

Where do $f(X=x)$ equations come from?

[Brand New] How to Represent Visual Ability?