09: Continuous RVs

David Varodayan

January 27, 2020

Adapted from slides by Lisa Yan

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)	Today!	Interval of time to first success

Kickboxing with RVs

Choose from:

A. $\text{Ber}(p)$
B. $\text{Bin}(n, p)$
C. $\text{Poi}(\lambda)$
D. $\text{Geo}(p)$
E. $\text{NegBin}(r, p)$

How would you model the following?

1. # of snapchats you receive in a day
2. # of children until the first one with brown eyes
3. Whether stock went up or down
4. # of probability problems you try until you get 5 correct (if you are randomly correct)
5. # of years in some decade with more than 6 Atlantic hurricanes

Equity in the Courts

Berghuis v. Smith (2010)

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

- Article by Erin Miller – Friday, 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 red, and nine hundred forty are black, and then you select them at random… twelve at a time."** According to Justice Breyer and the binomial theorem, if the red balls were black jurors then **"you would expect… something like a third to a half of juries would have at least one black person"** on them.

Justice Scalia's rejoinder: "We don't have any urns here."

Justice Breyer Meets CS109

Approximation using the Binomial distribution (even though it's not)

- Urn with 1000 balls, 60 red, 940 black.
- Assume each draw happens with replacement: P(draw red) = 60/1000
- 12 independent trials

What is $P(\geq 1 \text{ red ball drawn})$?

1. Define events/ RVs & state goal

$X \sim \text{Bin}(12, 60/1000)$

Want: $P(X \geq 1)$

2. Solve

$$P(X \geq 1) = 1 - P(X = 0)$$
$$\approx 1 - 0.4769 = 0.5240$$

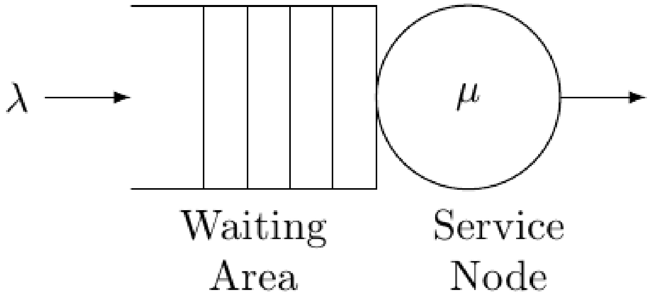
In Breyer's description, you should actually expect just over half of juries to have ≥ 1 non-white person in them.

demo

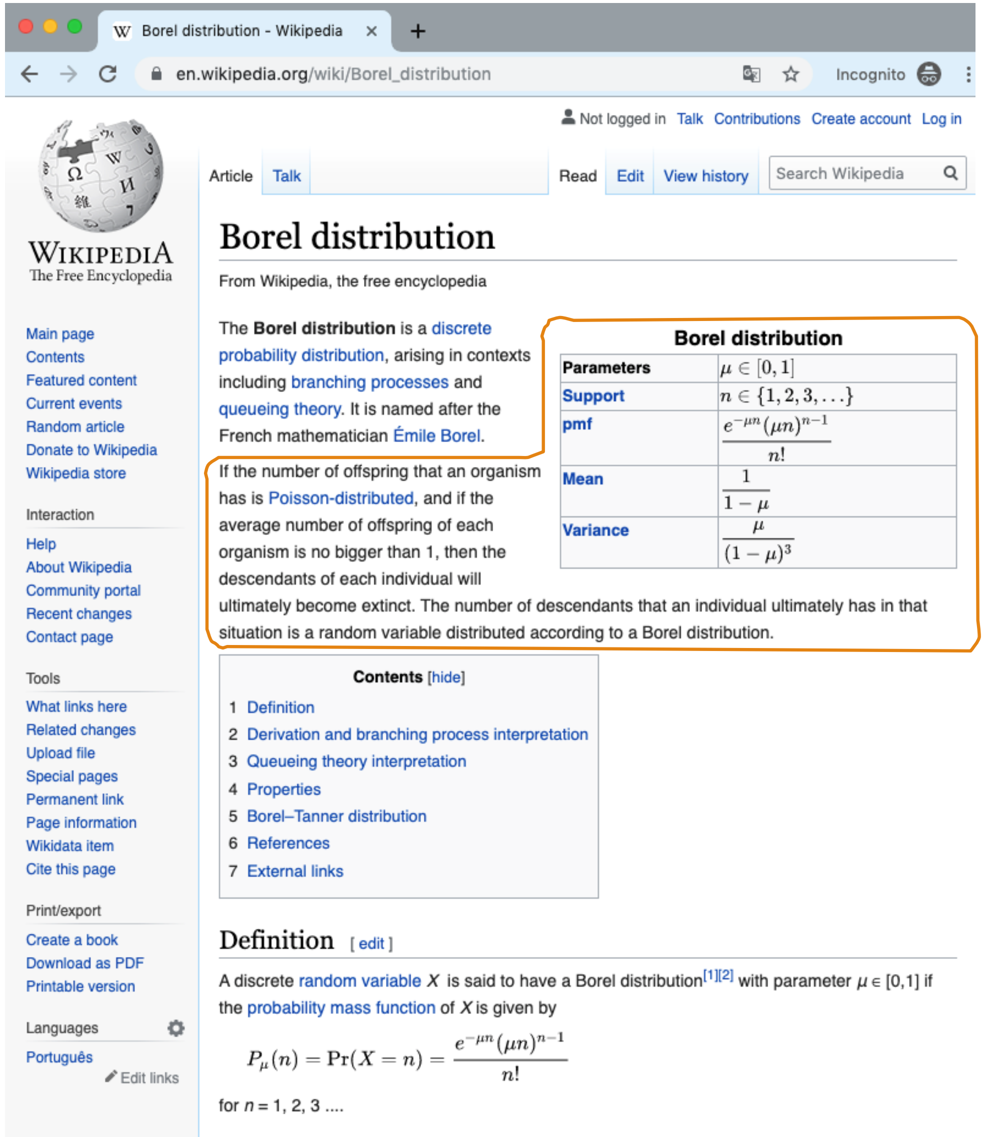
CS109 Learning Goal: Use new RVs

Let's say you are learning about servers/networks.

You read about the M/D/1 queue:



"The service time busy period is distributed as a Borel with parameter $\mu = 0.2$."

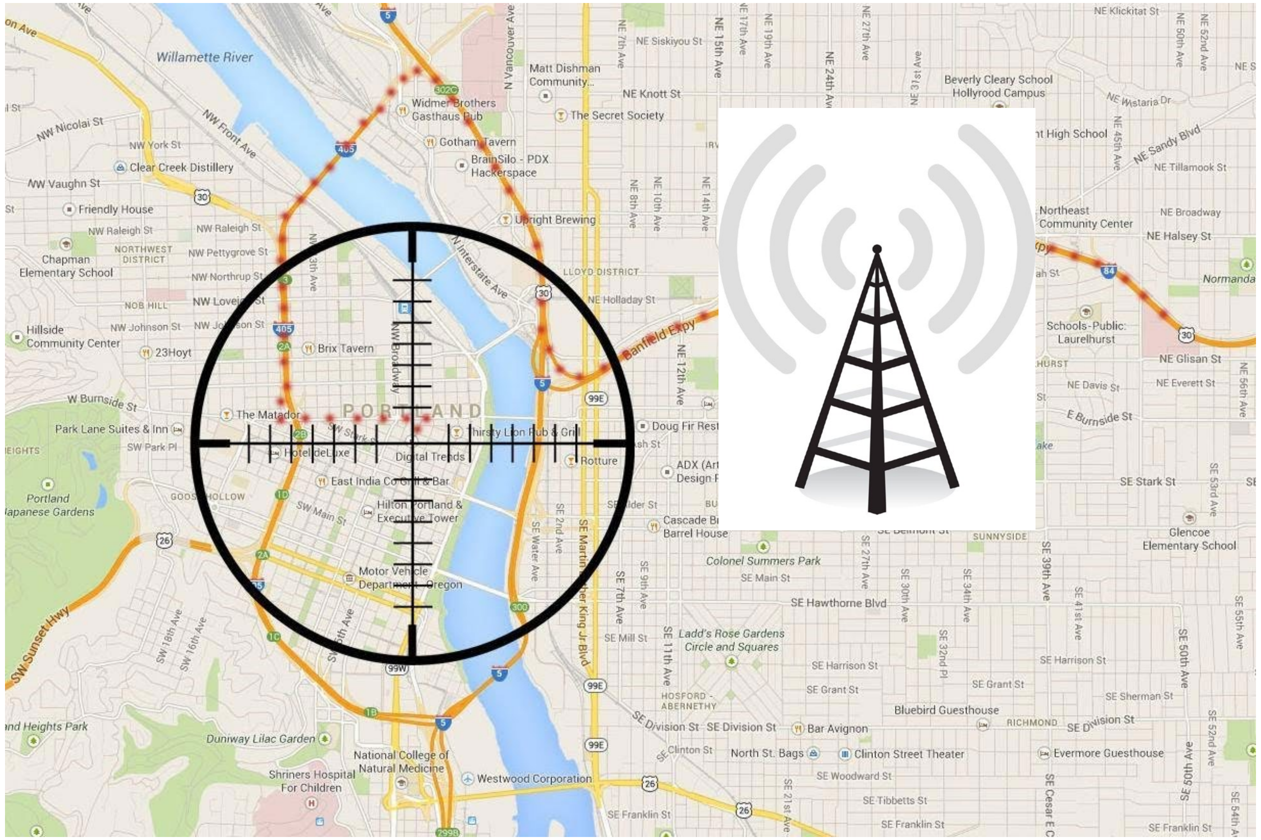


Goal: You can recognize terminology and understand experiment setup.

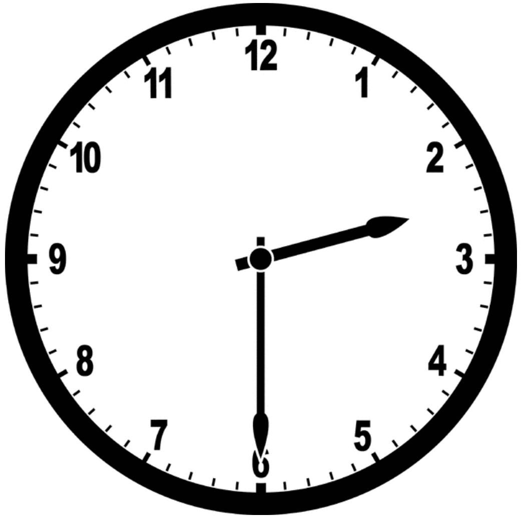
Today's plan

- Continuous RVs
- Uniform RV
- Exponential RV
- CDFs in detail

Not all values are discrete




```python
import numpy as np
np.random.random()
```



People heights

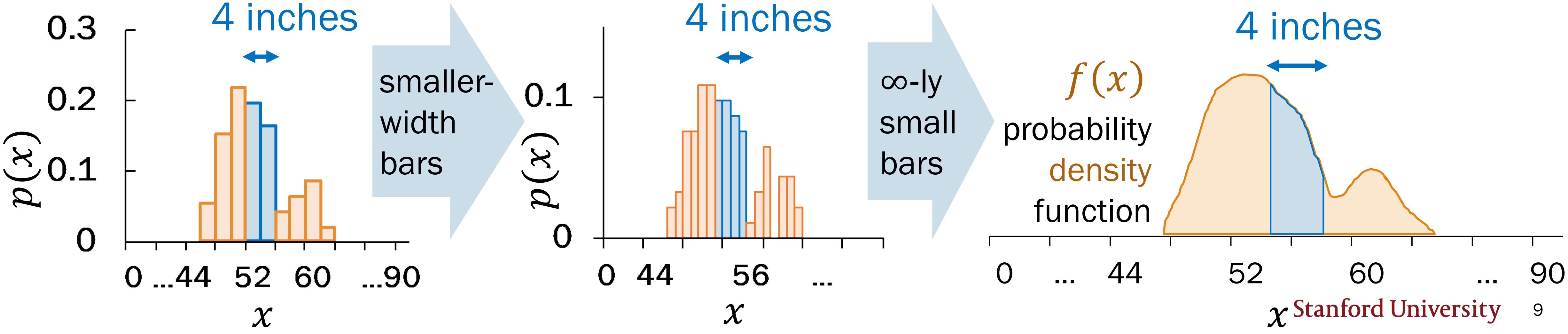
You are volunteering at the local elementary school.

- You are excited to get the perfect Halloween costume for your new 6th grade buddy, Jordan.
- However, you don't know exactly how tall Jordan is.

1. What is the probability that your buddy is 54.0923857234 inches tall?

   Essentially 0

2. What is the probability that your buddy is between 52–56 inches tall?



Continuous RV definition

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

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

The function f is a probability density function (PDF) if:

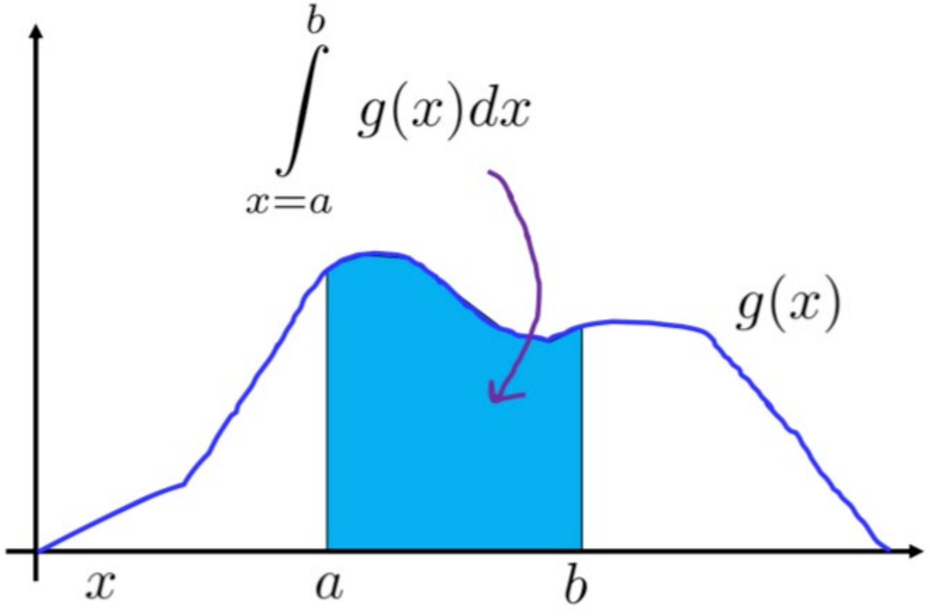
$$P(-\infty \leq X \leq \infty) = \int_{-\infty}^{\infty} f(x)\,dx = 1$$

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

- Often written as: $f_X(x)$
- Units: probability per units of X
- $f(x)$ is not a probability. Integrate to get probabilities.

Today's main takeaway

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

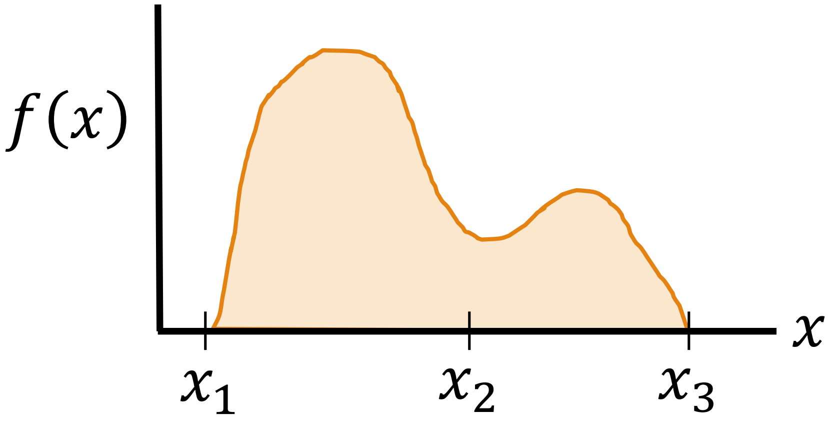


A probability!

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:

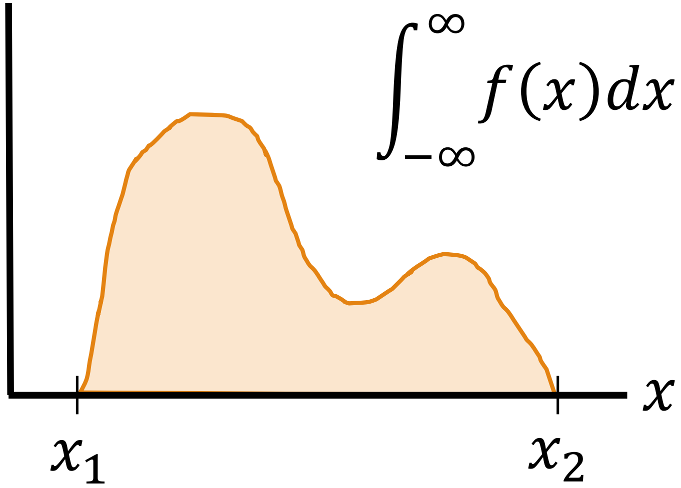
In the graphed PDF above,

1. $P(x_1 \leq X \leq x_2) > P(x_2 \leq X \leq x_3)$
2. $P(X = c) = 0$
3. $P(a \leq X \leq b) = P(a < X < b)$
4. $f(x)$ is a probability

$f(x)$ is NOT a probability

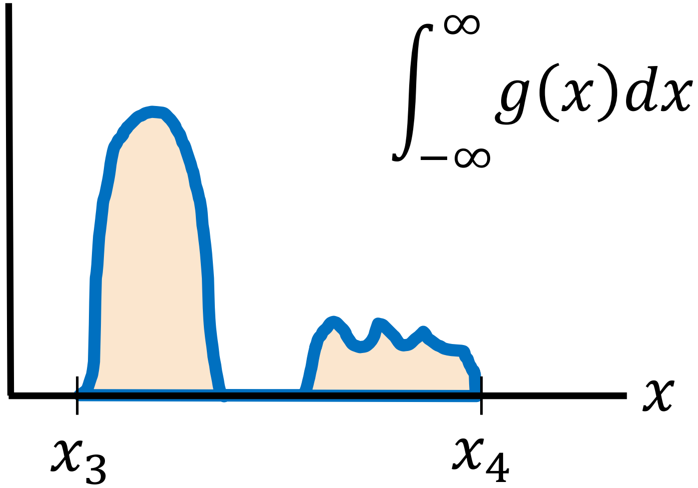
Which of the following functions are valid PDFs?

1. $f(x)$



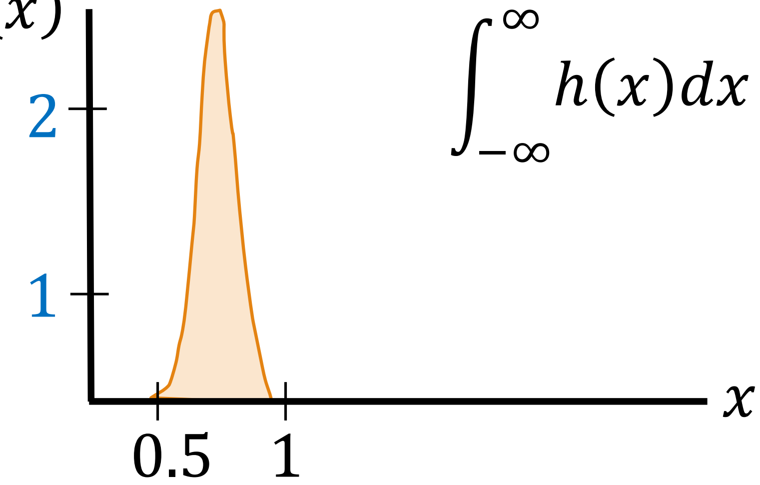
$$\int_{-\infty}^{\infty} f(x)dx = 0.5$$

2. $g(x)$



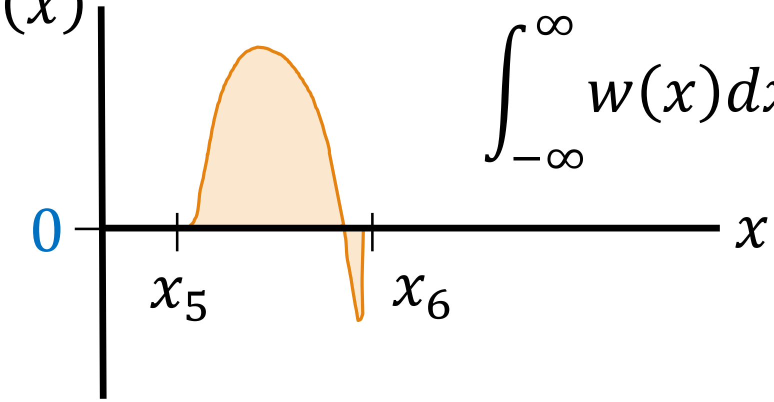
$$\int_{-\infty}^{\infty} g(x)dx = 1$$

3. $h(x)$



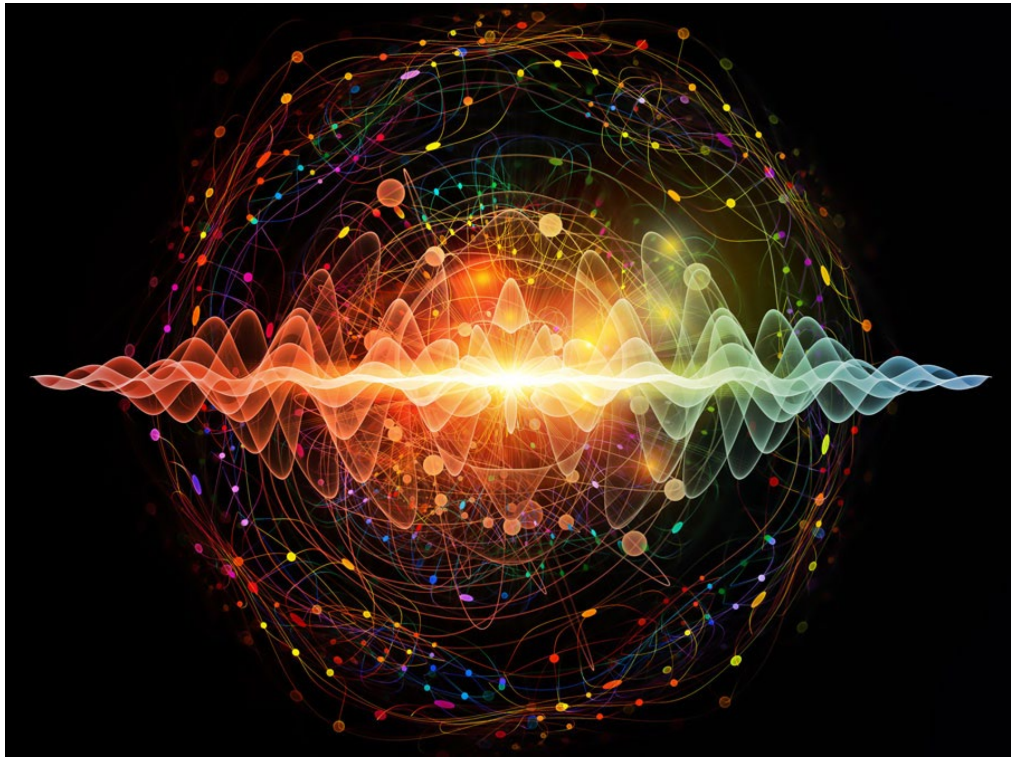
$$\int_{-\infty}^{\infty} h(x)dx = 1$$

4. $w(x)$



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

Simple example from Quantum Physics



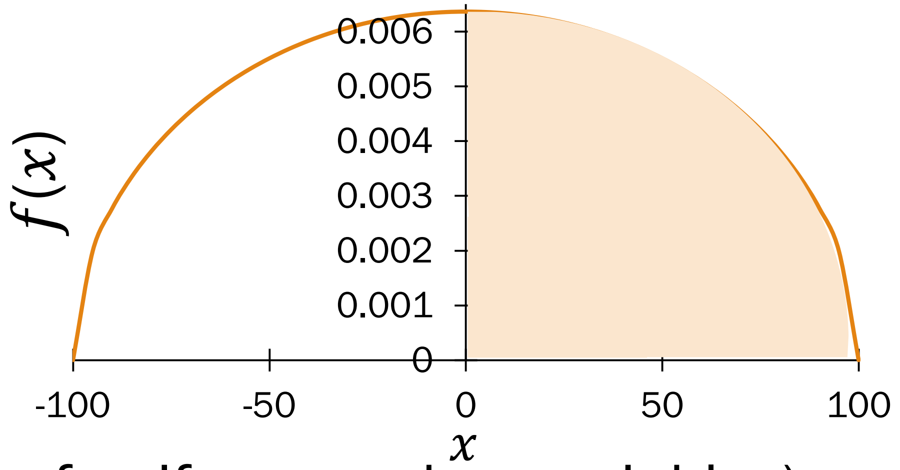
Consider a random 5000x5000 matrix, where each element in the matrix is Uniform(0,1). What is the probability that a selected eigenvalue (λ) of the matrix is greater than 0?*

* With help from Wigner's Semicircle Law, David is going to rephrase this problem.

Simple example from Quantum Physics

Let X = a continuous RV defined as follows:

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



(represents the eigenvalue of a 5000×5000 matrix of uniform random variables)

What is $P(X > 0)$?

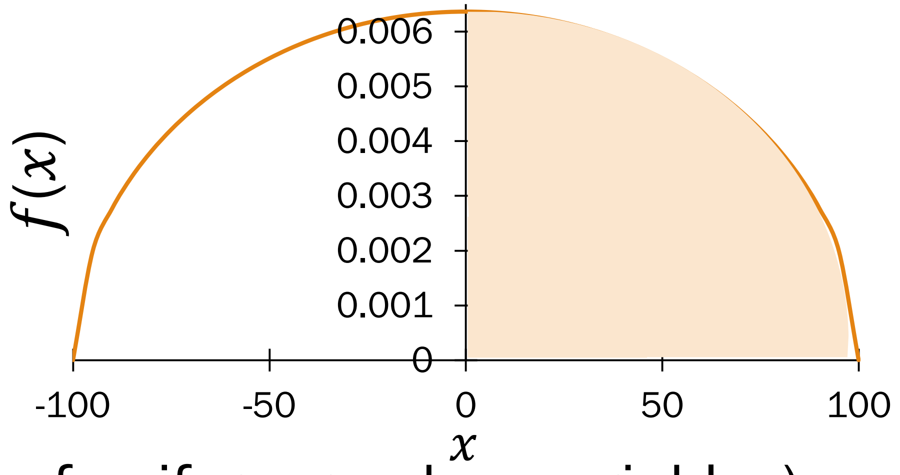
1. Approach 1
   Integrate over PDF
2. Approach 2
3. Approach 3

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

Simple example from Quantum Physics

Let X = a continuous RV defined as follows:

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



(represents the eigenvalue of a 5000×5000 matrix of uniform random variables)

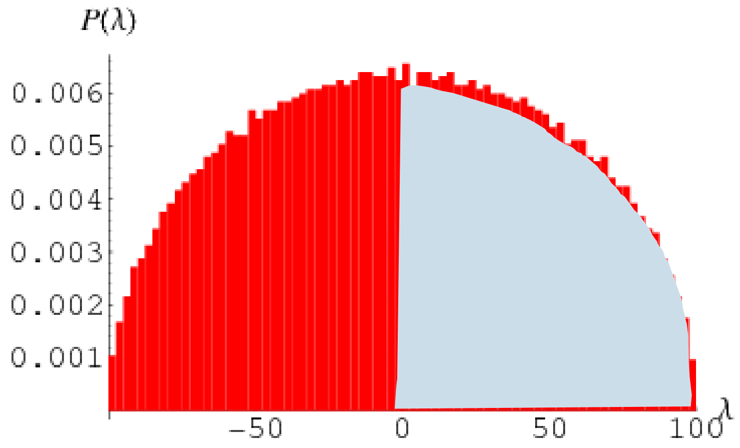
What is $P(X > 0)$?

1. Approach 1
   Integrate over PDF

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

2. Approach 2
   Simulate + discrete approximation

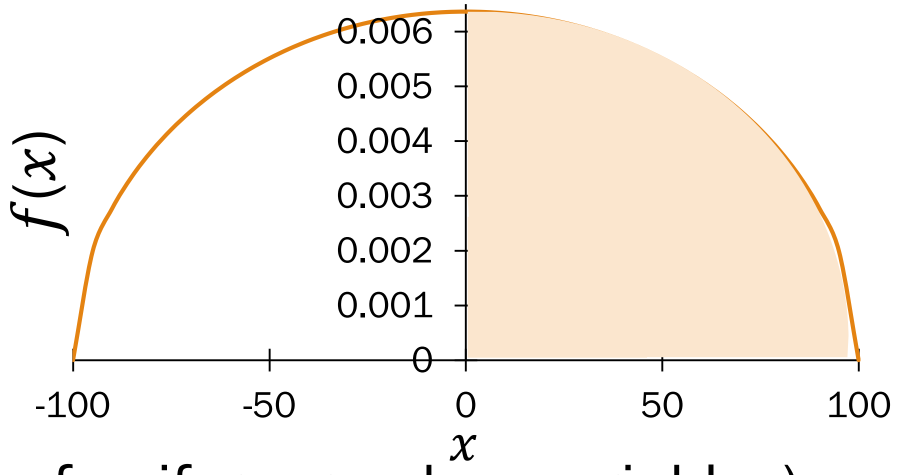
$$P(X > 0) \approx \sum_{k=0}^{100} P(X = k)$$



Simple example from Quantum Physics

Let X = a continuous RV defined as follows:

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



(represents the eigenvalue of a 5000×5000 matrix of uniform random variables)

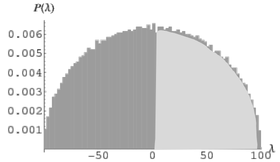
What is $P(X > 0)$?

1. Approach 1
   Integrate over PDF

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

2. Approach 2
   Simulate + discrete approximation

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



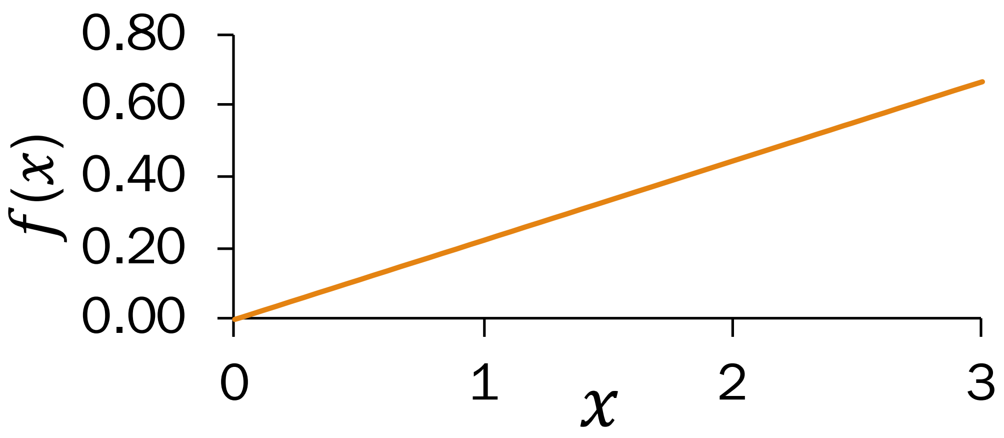
3. Approach 3
   Know semi-circles

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

Another example

Let X be a continuous RV with PDF:

$$f_X(x) = \begin{cases} Cx & \text{if } 0 \le x \le 3 \\ 0 & \text{otherwise} \end{cases}$$



What is the constant C that makes f a valid PDF?

A. Know triangles and $y = mx + b$

B. Solve for C: $\int_{-\infty}^{\infty} f(x)\, dx = C \int_0^3 x\, dx = 0$

C. Solve for C: $\int_{-\infty}^{\infty} f(x)\, dx = C \int_0^3 x\, dx = 1$

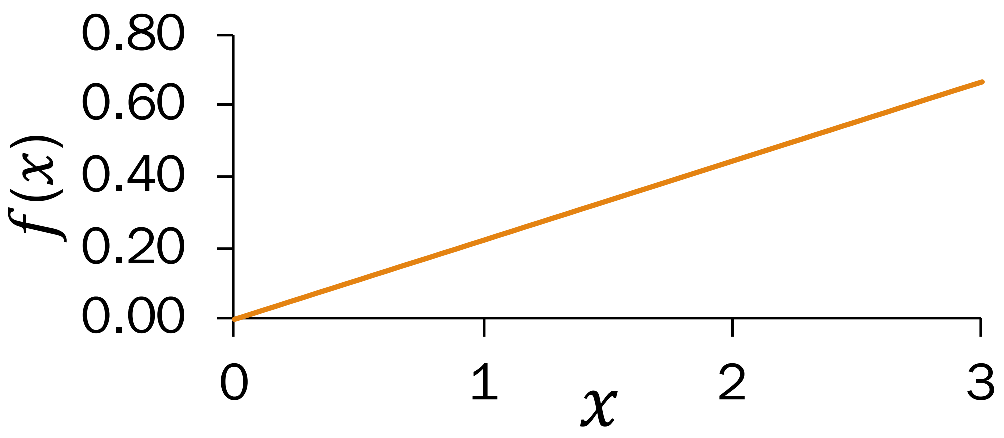
D. $f(x)$ is not a probability

E. None/other

Another example

Let X be a continuous RV with PDF:

$$f_X(x) = \begin{cases} Cx & \text{if } 0 \le x \le 3 \\ 0 & \text{otherwise} \end{cases}$$



What is the constant C that makes f a valid PDF?

C. Solve for $\int_{-\infty}^{\infty} f(x)\,dx = C\int_0^3 x\,dx = 1$

$$C\int_0^3 x\,dx = C\left(\frac{1}{2}x^2\right)\Big|_0^3 = C\left(\frac{9}{2} - 0\right) = 1 \quad \Longrightarrow \quad C = 2/9$$

Today's plan

Continuous RVs

Uniform RV

Exponential RV

CDFs in detail

Uniform Random Variable

def A **Uniform** random variable X is defined as follows:

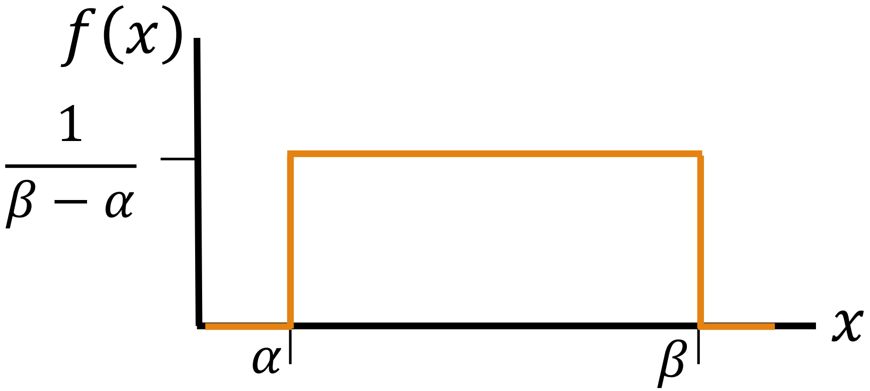
$$X \sim \text{Uni}(\alpha, \beta)$$

Support: $[\alpha, \beta]$
(sometimes defined over (α, β))

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

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

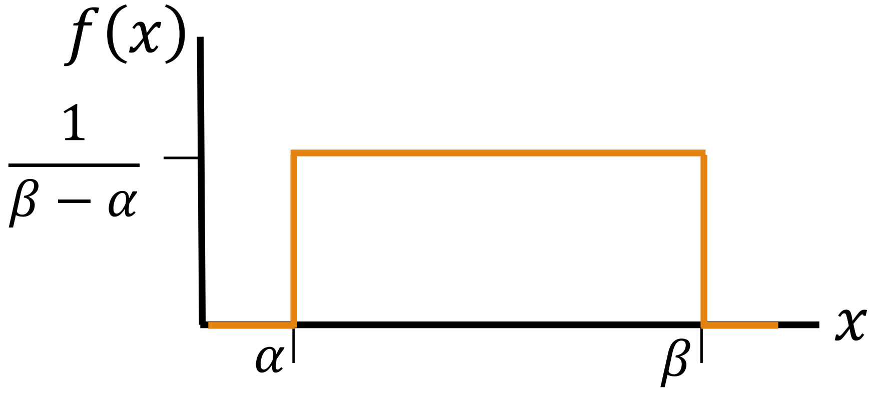
Variance $\text{Var}(X) = \frac{(\beta - \alpha)^2}{12}$



Quick check

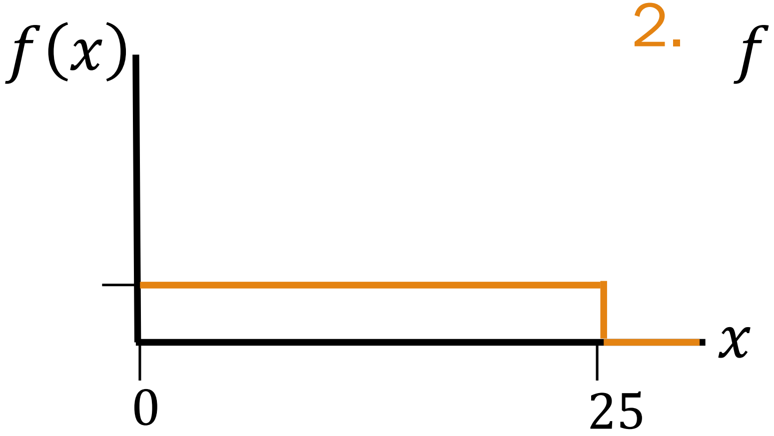
If $X \sim \text{Uni}(\alpha, \beta)$, the PDF of X is:

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

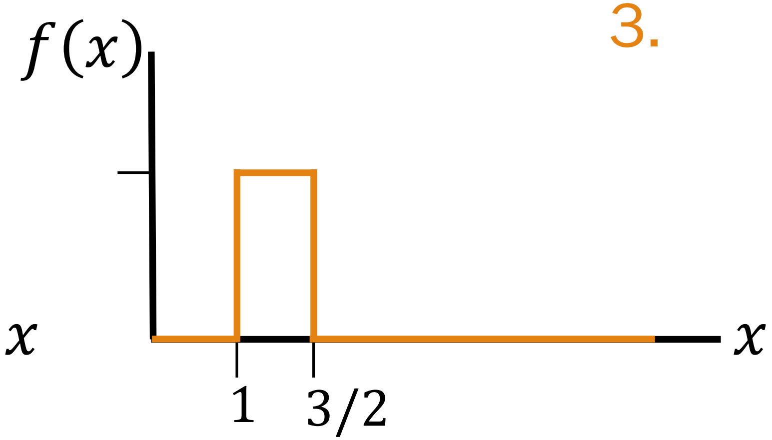


What is $\frac{1}{\beta - \alpha}$ if the following graphs are PDFs of Uniform RVs X?

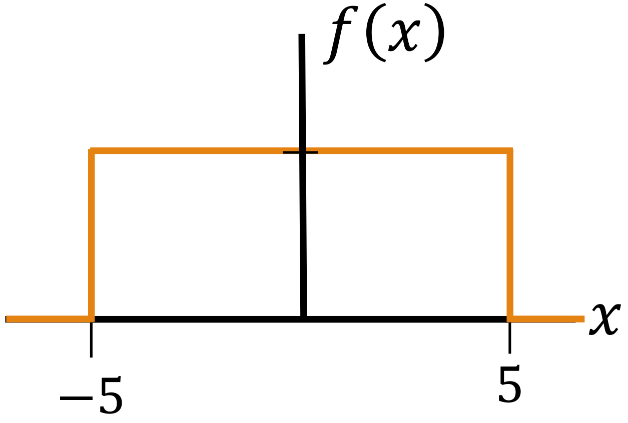
1.



2.



3.



Riding the Marguerite Bus

You want to get on the Marguerite bus.

- The bus stops at the Gates building at 15-minute intervals (2:00, 2:15, etc.).
- You arrive at the stop uniformly b/t 2:00-2:30pm.

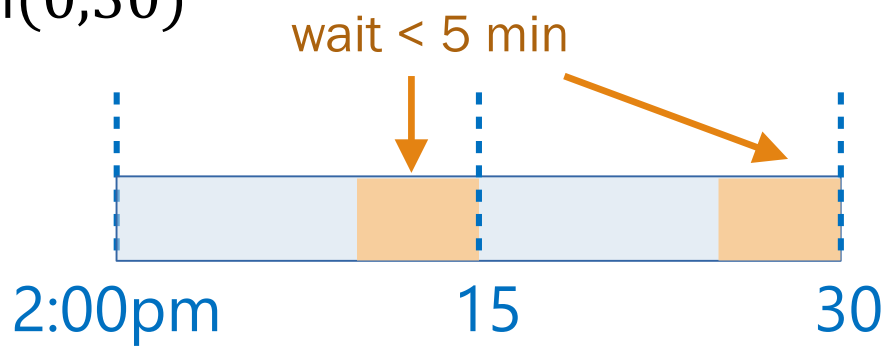
P(you wait < 5 minutes for bus)?

1. Define events/ RVs & state goal

X: time passenger arrives after 2:00

$X \sim \text{Uni}(0,30)$

Want:



2. Solve

Announcements

Problem Set 3

Out: today
Due: Wednesday 2/5
Covers: through this Wednesday

Late days

Free: 2 free class days

No late days after last day of quarter (Fri 3/13)

(note PS#6 due Wed 3/11)

Expectation and Variance

Discrete RV X

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

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

Continuous RV X

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

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

Both continuous and discrete RVs

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

Linearity of Expectation

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

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

Properties of variance

Uniform Random Variable

def An **Uniform** random variable X is defined as follows:

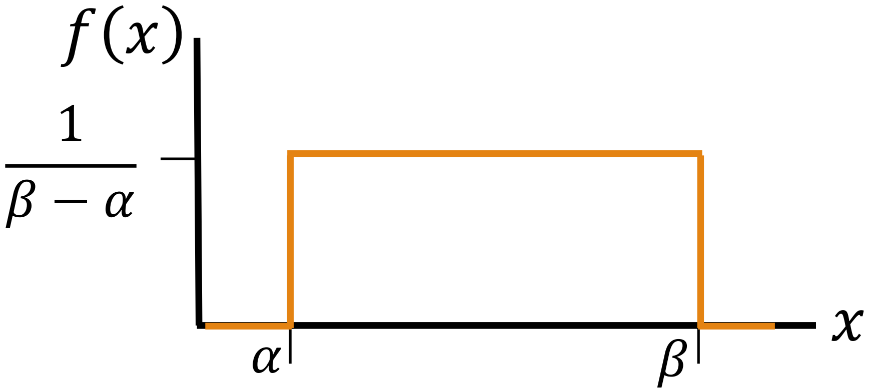
$$X \sim \text{Uni}(\alpha, \beta)$$

Support: $[\alpha, \beta]$
(sometimes defined over (α, β))

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

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

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



Uniform RV expectation

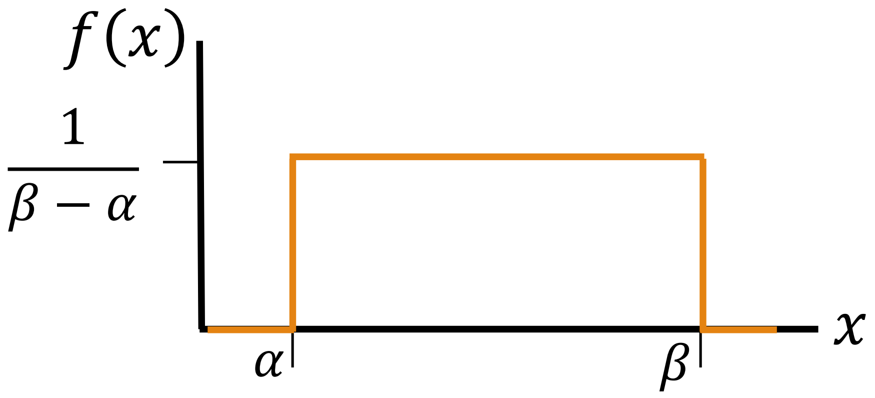
$X \sim \text{Uni}(\alpha, \beta)$ Expectation $E[X] = \dfrac{\alpha + \beta}{2}$

$$E[X] = \int_{-\infty}^{\infty} x \cdot f(x)dx = \int_{\alpha}^{\beta} x \cdot \frac{1}{\beta - \alpha} dx$$

$$= \frac{1}{\beta - \alpha} \cdot \frac{1}{2} x^2 \Big|_{\alpha}^{\beta}$$

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

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



Interpretation:
Average the start & end

Today's plan

Continuous RVs

Uniform RV

Exponential RV

CDFs in detail

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.

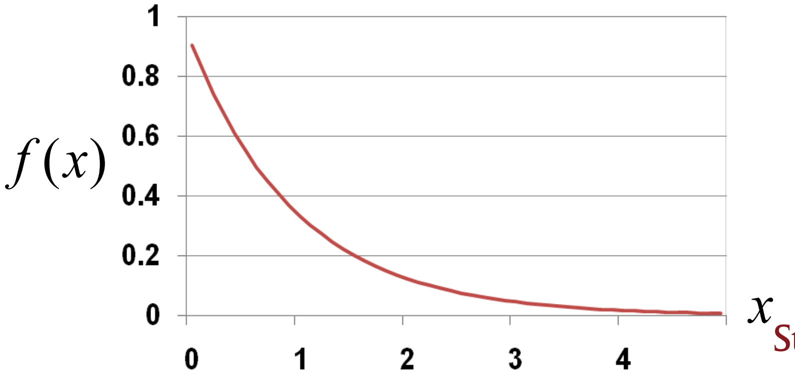
$X \sim \text{Exp}(\lambda)$

Support: $[0, \infty)$

PDF	$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$	
Expectation	$E[X] = \frac{1}{\lambda}$	(in extra slides)
Variance	$\text{Var}(X) = \frac{1}{\lambda^2}$	

Examples:

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



Interpreting Exp(λ)

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

$X \sim \text{Exp}(\lambda)$ Expectation $E[X] = \frac{1}{\lambda}$

Based on the expectation $E[X]$, what are the units of λ?

A. Probability
B. Probability^{-1}
C. Time
D. Time^{-1}
E. Not sure, but $f(x)$ is not a probability

Earthquakes

$X \sim \text{Exp}(\lambda)$ $E[X] = 1/\lambda$
$f(x) = \lambda e^{-\lambda x}$ if $x \geq 0$

Major earthquakes (magnitude 8.0+) occur once every 500 years.*

1. What is the probability of a major earthquake in the next 30 years?

Define events/ RVs & state goal

X: when next earthquake happens

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

λ: year^{-1} $= 1/500$

Want: $P(X < 30)$

Solve

Recall

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

*In California, according to historical data form USGS, 2015

Earthquakes

$X \sim \text{Exp}(\lambda)$ $E[X] = 1/\lambda$
$f(x) = \lambda e^{-\lambda x}$ if $x \geq 0$

Major earthquakes (magnitude 8.0+) occur once every 500 years.*

1. What is the probability of a major earthquake in the next 30 years?
2. What is the standard deviation of years until the next earthquake?

Define events/ RVs & state goal

X: when next earthquake happens

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

λ: year^{-1}

Want: $P(X < 30)$

Solve

*In California, according to historical data form USGS, 2015

Today's plan

Continuous RVs

Uniform RV

Exponential RV

CDFs in detail

Cumulative Distribution Function (CDF)

For a random variable X, the cumulative distribution function (CDF) is defined as

$$F(a) = F_X(a) = P(X \leq a), \text{where } -\infty < a < \infty$$

For a discrete RV X, the CDF is:

$$F(a) = P(X \leq a) = \sum_{\text{all } x \leq a} p(x)$$

For a continuous RV X, the CDF is:

$$F(a) = P(X \leq a) = \int_{-\infty}^{a} f(x)dx$$

If you learn to use CDFs, you can avoid integrals.

CDF of an Exponential RV

$X \sim \text{Exp}(\lambda)$ $f(x) = \lambda e^{-\lambda x}$ if $x \geq 0$

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

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

if $x \geq 0$

Proof:

$$F(x) = P(X \leq x) = \int_{y=-\infty}^{x} f(y)dy = \int_{y=0}^{x} \lambda e^{-\lambda y} dy$$

$$= \lambda \frac{1}{-\lambda} e^{-\lambda y} \Big|_{0}^{x}$$

$$= -1\left(e^{-\lambda x} - e^{-\lambda 0}\right)$$

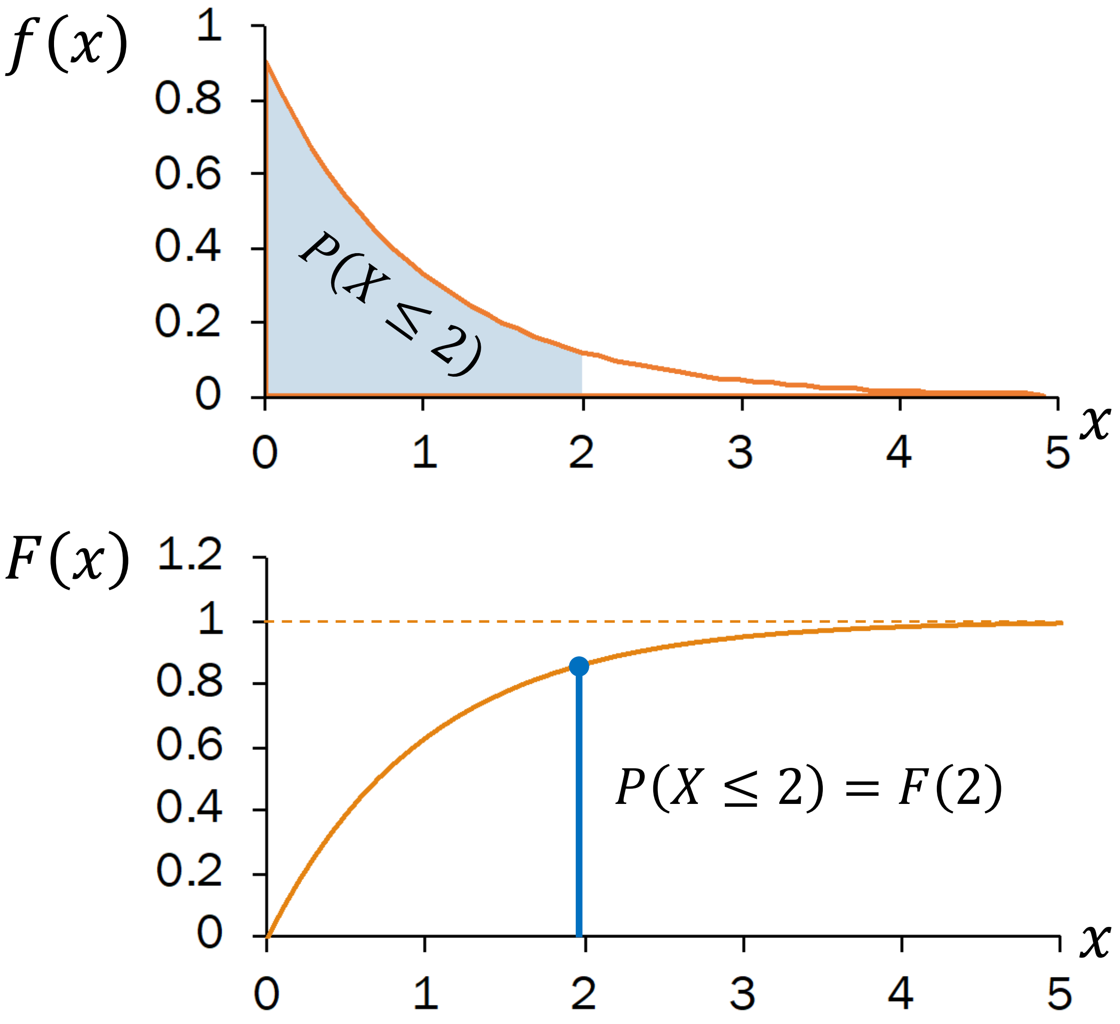
$$= 1 - e^{-\lambda x}$$

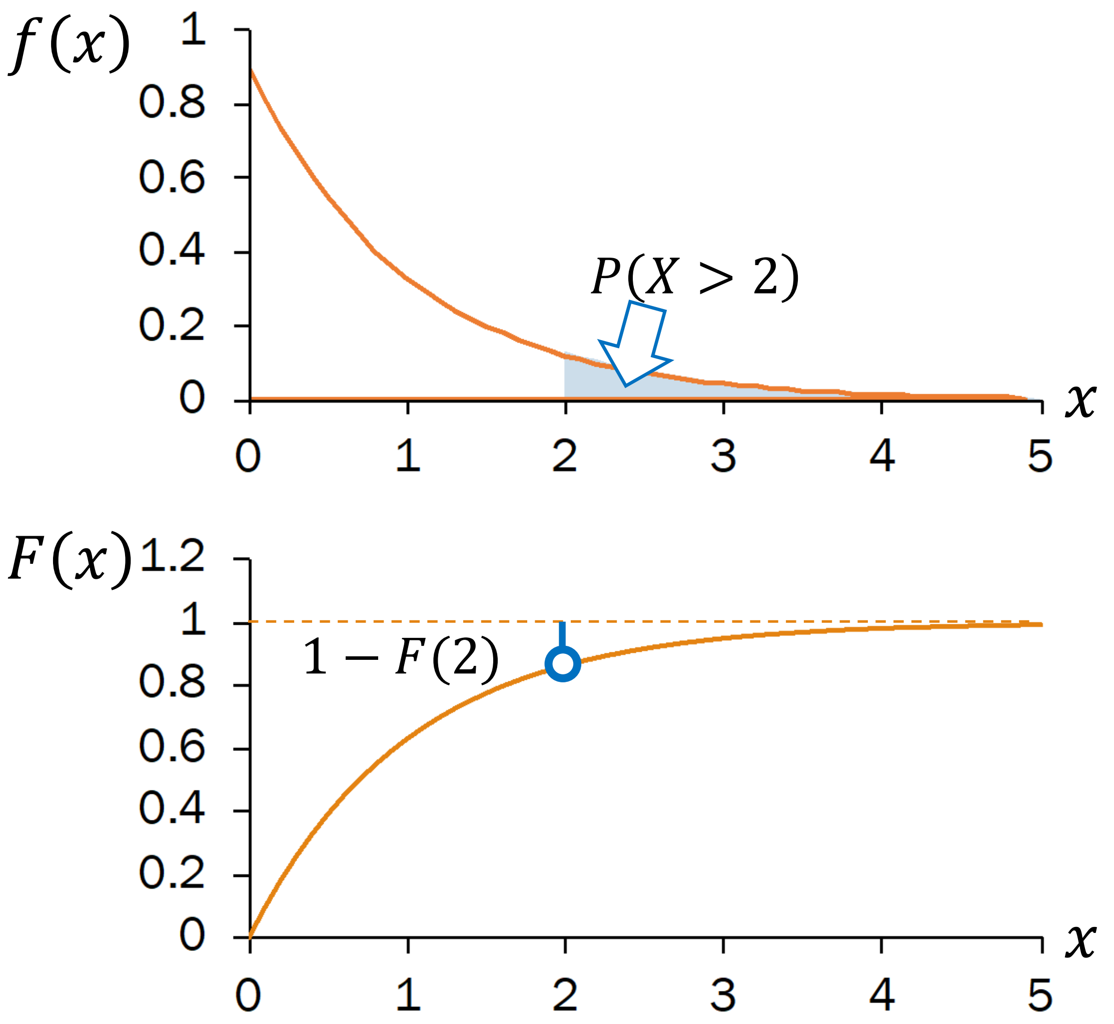
Recall

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

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

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





The CDF $F(x)$ is a probability: $F(x) = P(X \leq x)$

Using the CDF

For a continuous random variable X with PDF $f(x)$, the CDF of X is

$$P(X \leq a) = F(a) = \int_{-\infty}^{a} f(x)dx$$

Matching (choices are used 0/1/2 times)

1. $P(X < a)$
2. $P(X > a)$
3. $P(X \geq a)$
4. $P(a \leq X \leq b)$

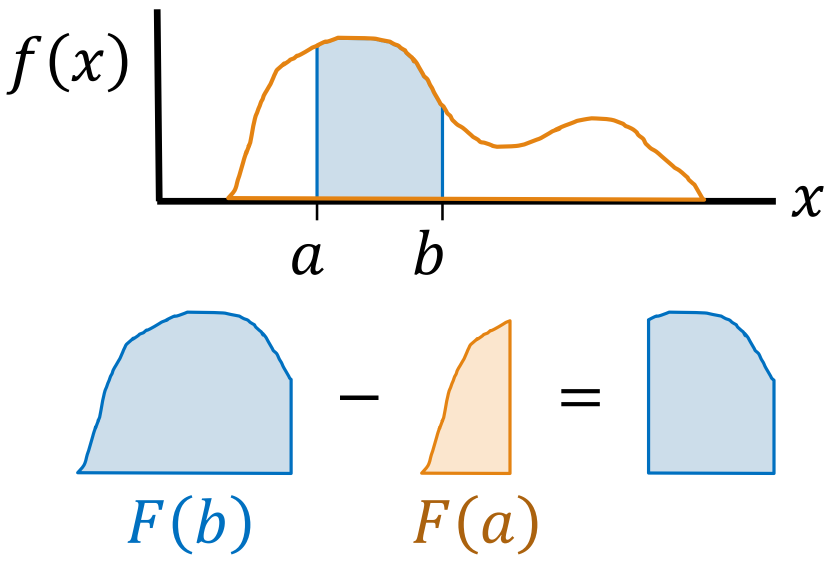
A. $F(a)$
B. $1 - F(a)$
C. $F(a) - F(b)$
D. $F(b) - F(a)$

Using the CDF

For a continuous random variable X with PDF $f(x)$, the CDF of X is

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

4. $P(a \leq X \leq b) = F(b) - F(a)$



Proof:

$$F(b) - F(a) = \int_{-\infty}^{b} f(x)dx - \int_{-\infty}^{a} f(x)dx$$

$$= \left(\int_{-\infty}^{a} f(x)dx + \int_{a}^{b} f(x)dx\right) - \int_{-\infty}^{a} f(x)dx$$

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

Earthquakes with CDFs

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

Major earthquakes (magnitude 8.0+) occur once every 500 years.*

1. What is the probability of a major earthquake in the next 30 years?
2. What is the standard deviation of years until the next earthquake?

Define events/ RVs & state goal

X: when next earthquake happens

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

λ: year^{-1} $= 1/500$

Want: $P(X < 30)$

Solve

$$P(X < 30) = \int_0^{30} 0.002 e^{-0.002x} dx$$

$$P(X < 30) = F(30) = 1 - e^{-\lambda x}$$

$$= 1 - e^{-0.002 \cdot 30}$$

$$\approx 0.058$$

*In California, according to historical data form USGS, 2015

Earthquakes

Major earthquakes (magnitude 8.0+) occur once every 500 years.*

1. What is the probability of a major earthquake in the next 30 years?
2. What is the standard deviation of years until the next earthquake?
3. What is the probability of zero major earthquakes next year?

We know:

$$500 \frac{\text{years}}{\text{earthquake}}$$

$$0.002 \frac{\text{earthquakes}}{\text{year}}$$

$$1 \frac{\text{earthquakes}}{\text{500 years}}$$

Strategy:

A. Bayes' Theorem
B. Total Probability
C. Uniform RV
D. Poisson RV
E. Exponential RV

*In California, according to historical data form USGS, 2015

Earthquakes

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

Major earthquakes (magnitude 8.0+) occur once every 500 years.*

1. What is the probability of a major earthquake in the next 30 years?
2. What is the standard deviation of years until the next earthquake?
3. What is the probability of zero major earthquakes next year?

Strategy: D. Poisson RV

Define events/RVs & state goal

X: # earthquakes next year

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

$\lambda: \frac{\text{earthquakes}}{\text{year}}$

Want: $P(X = 0)$

Solve

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

Strategy: E. Exponential RV

Define events/RVs & state goal

X: when first earthquake happens

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

Want: $P(X > 1) = 1 - F(1)$

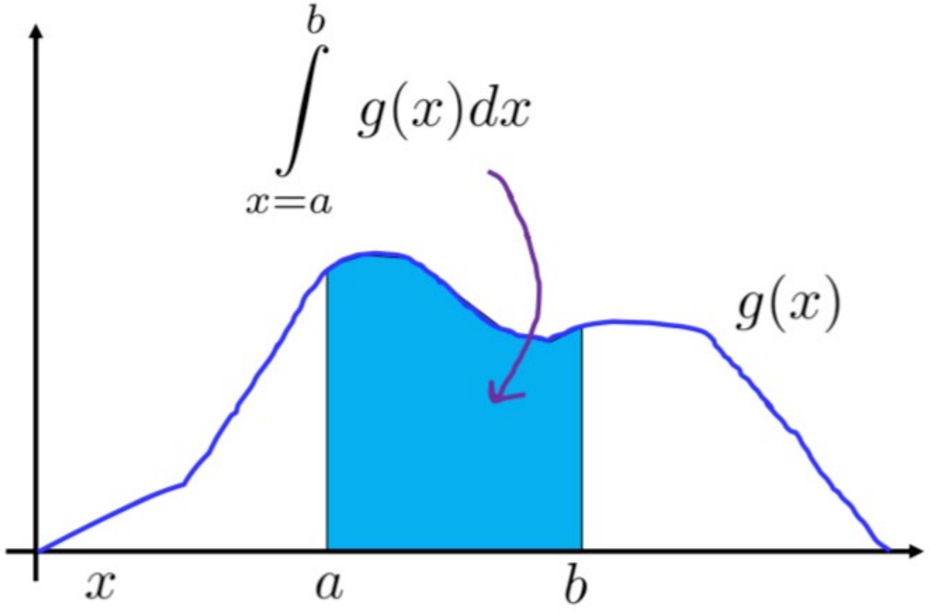
Solve

$$P(X > 1) = 1 - \left(1 - e^{-\lambda \cdot 1}\right) = e^{-\lambda}$$

*In California, according to historical data form USGS, 2015

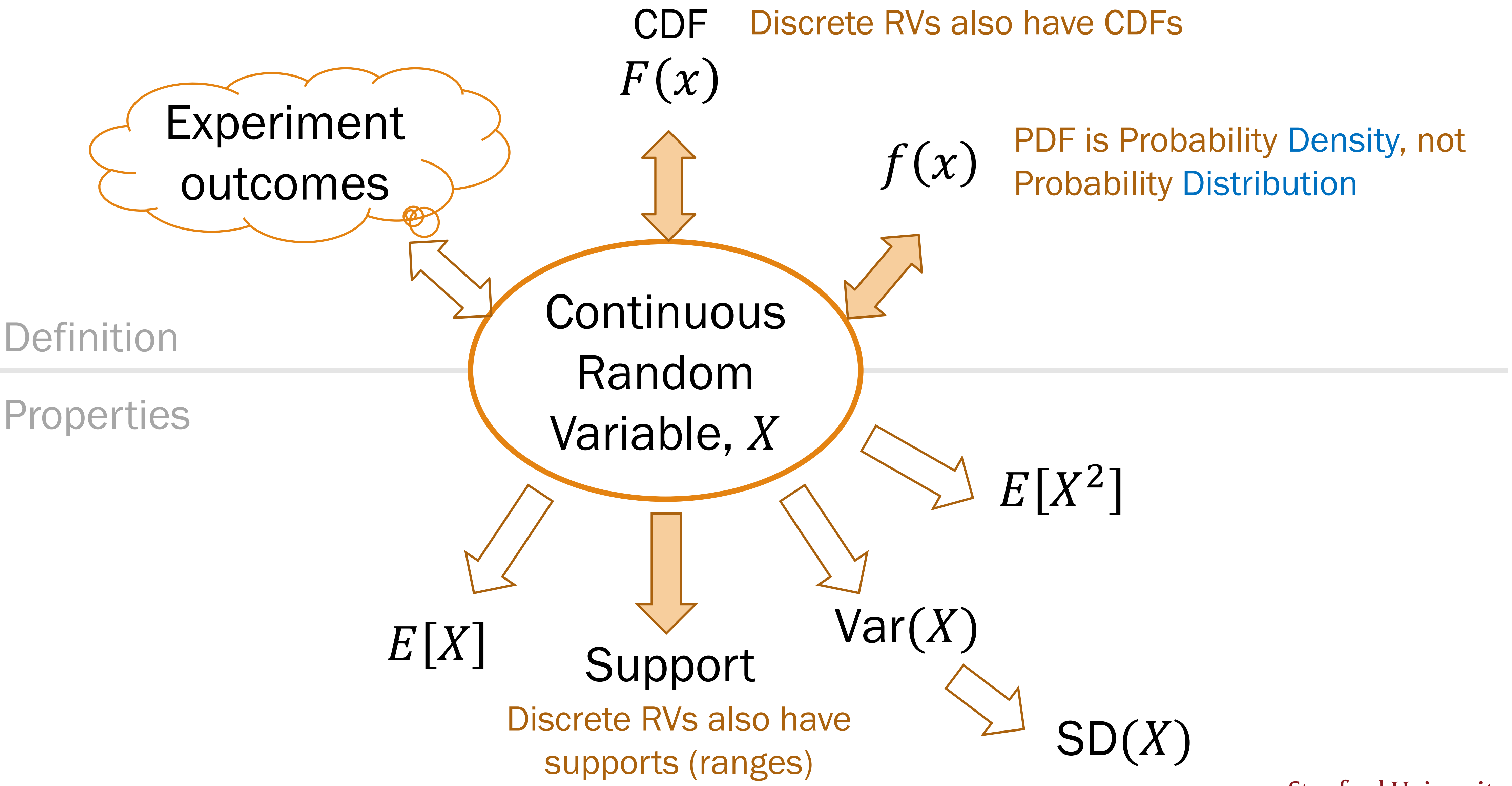
Today's main takeaway

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



A probability!

Continuous random variables



Extra slides

Expectation of the Exponential

Extra problems

Expectation of the Exponential

$X \sim \text{Exp}(\lambda)$ $f(x) = \lambda e^{-\lambda x}$ if $x \geq 0$

$X \sim \text{Exp}(\lambda)$ Expectation $E[X] = \frac{1}{\lambda}$

Proof:

$$E[X] = \int_{-\infty}^{\infty} x f(x) dx = \int_0^{\infty} x\lambda e^{-\lambda x} dx$$

$$= -xe^{-\lambda x}\Big|_0^{\infty} + \int_0^{\infty} e^{-\lambda x} dx$$

$$= -xe^{-\lambda x}\Big|_0^{\infty} - \frac{1}{\lambda} e^{-\lambda x}\Big|_0^{\infty}$$

$$= \left[0 - \frac{1}{\lambda} 0\right] - \left[0 - \frac{1}{\lambda}\right]$$

$$= \frac{1}{\lambda}$$

Integration by parts

$$\int x\lambda e^{-\lambda x} dx = \int u \cdot dv$$

$u = x$ $\quad dv = \lambda e^{-\lambda x} dx$

$du = dx$ $\quad v = -e^{-\lambda x}$

$$\int u \cdot dv = u \cdot v - \int v \cdot du$$

$$-xe^{-\lambda x} - \int -e^{-\lambda x} dx$$

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

Define

X: when visitor leaves

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

Solve

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

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

Replacing your laptop

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

Let X = # hours of use until your laptop dies.

- X is distributed as an Exponential RV, where
- On average, laptops die after 5000 hours of use.
- You use your laptop 5 hours a day.

What is P(your laptop lasts 4 years)?

Define

X: # hours until laptop death

$X \sim \text{Exp}(\lambda = 1/5000)$

Want: $P(X > 5 \cdot 365 \cdot 4)$

Solve

$$P(X > 7300) = 1 - F(7300)$$

$$= 1 - \left(1 - e^{-7300/5000}\right) = e^{-1.46} \approx 0.2322$$

Better plan ahead if you're co-terming!

- 5-year plan: $P(X > 9125) = e^{-1.825} \approx 0.1612$
- 6-year plan: $P(X > 10950) = e^{-2.19} \approx 0.1119$