09: Continuous RVs

Jerry Cain
April 15, 2022

Table of Contents

2 Continuous RVs
14 Uniform RV
21 Exponential RV
30 CDFs
38 Exercises

Continuous RVs

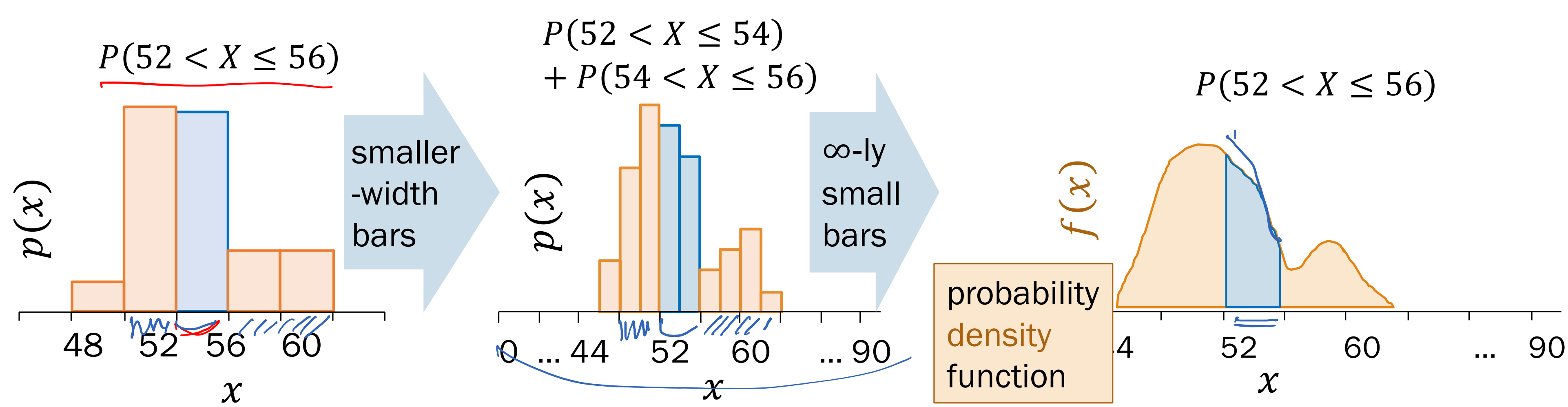
People heights

You are volunteering at the local elementary school.

- To choose a t-shirt for your friend Mia, you need to know their height.

1. What is the probability that your buddy is 54.0923857234 inches tall? — Essentially 0

2. What is the probability Mia is between 52–56 inches tall?



Continuous RV definition

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

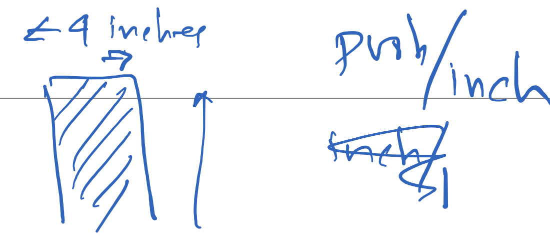
$$P(a \leq X \leq b) = \int_a^b f(x)\,dx$$

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

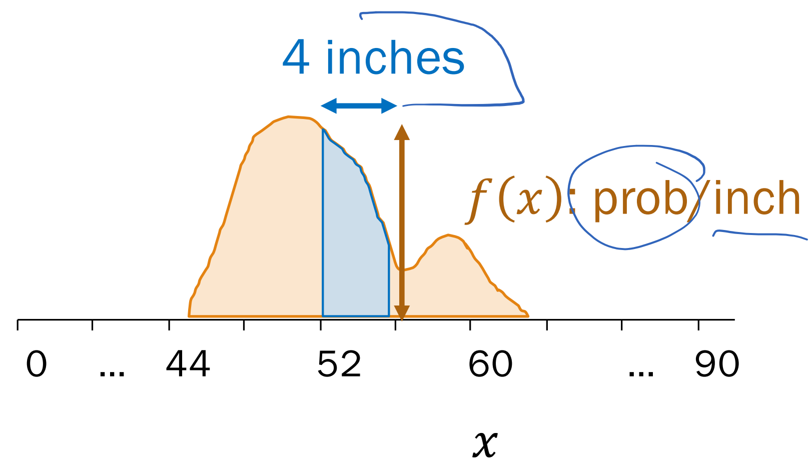
$$\int_{-\infty}^{\infty} f(x)\,dx = P(-\infty < X < \infty) = 1$$

Also written as: $f_X(x)$

Today's main takeaway, #1



Integrate $f(x)$ to get probabilities.



PDF Units: probability per units of X

$$P(52 \leq X \leq 56) = \int_{52}^{56} f(x)\, dx$$

PMF vs PDF

Discrete random variable X

Probability mass function (PMF):

$$p(x)$$

To get probability:

$$P(X = x) = p(x)$$

Continuous random variable X

Probability density function (PDF):

$$f(x)$$

To get probability:

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

Both are measures of how likely X is to take on a value.

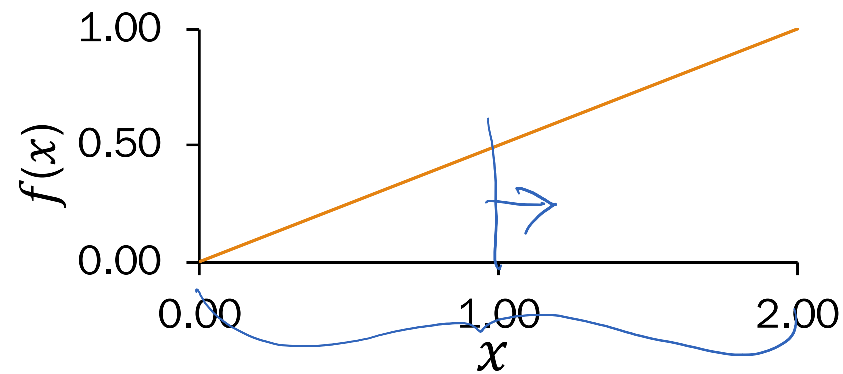
Computing probability

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

Let X be a continuous RV with PDF:

$$f(x) = \begin{cases} \frac{1}{2}x & \text{if } 0 \le x \le 2 \\ 0 & \text{otherwise} \end{cases}$$

What is $P(X \ge 1)$?

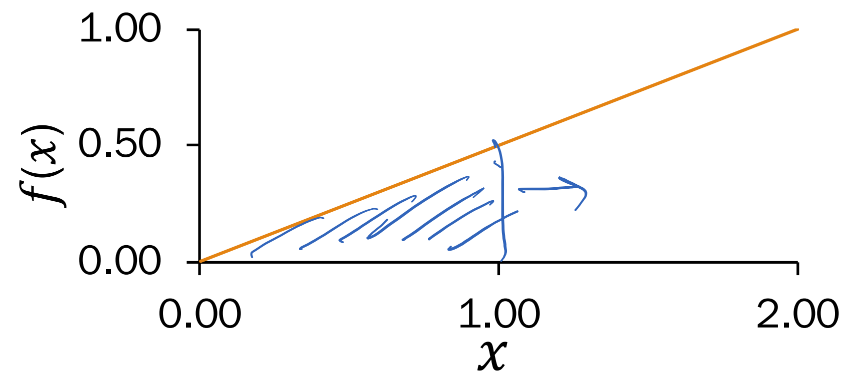


Computing probability

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

Let X be a continuous RV with PDF:

$$f(x) = \begin{cases} \frac{1}{2}x & \text{if } 0 \le x \le 2 \\ 0 & \text{otherwise} \end{cases}$$



What is $P(X \ge 1)$?

Strategy 1: Integrate

$$P(1 \le X < \infty) = \int_1^\infty f(x)dx = \int_1^2 \frac{1}{2}x dx$$

$$= \frac{1}{2}\left(\frac{1}{2}x^2\right)\Big|_1^2 = \frac{1}{2}\left[2 - \frac{1}{2}\right] = \frac{3}{4}$$

Strategy 2: Know triangles

$$1 - \frac{1}{2}\left(\frac{1}{2}\right) = \frac{3}{4}$$

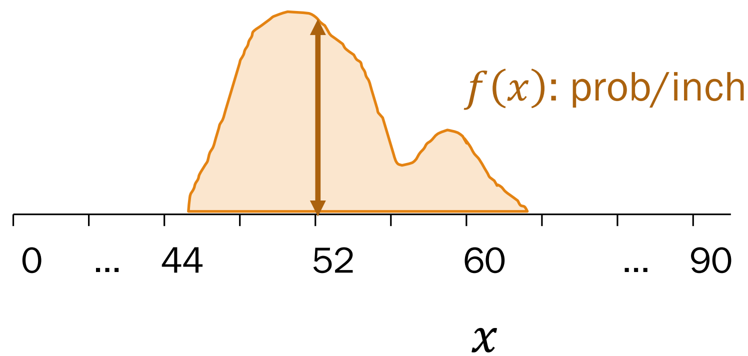
Wait...is this even legal?

$$P(0 \le X < 1) = \int_0^1 f(x)dx\,??$$

Today's main takeaway, #2

For a continuous random variable X with PDF $f(x)$,

$$P(X = c) = \int_c^c f(x)dx = 0.$$

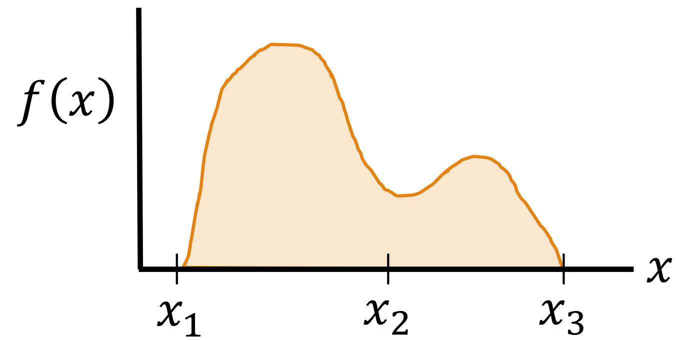


Contrast with PMF in discrete case: $P(X = c) = p(c)$

PDF Properties

For a continuous RV X with PDF f,

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



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

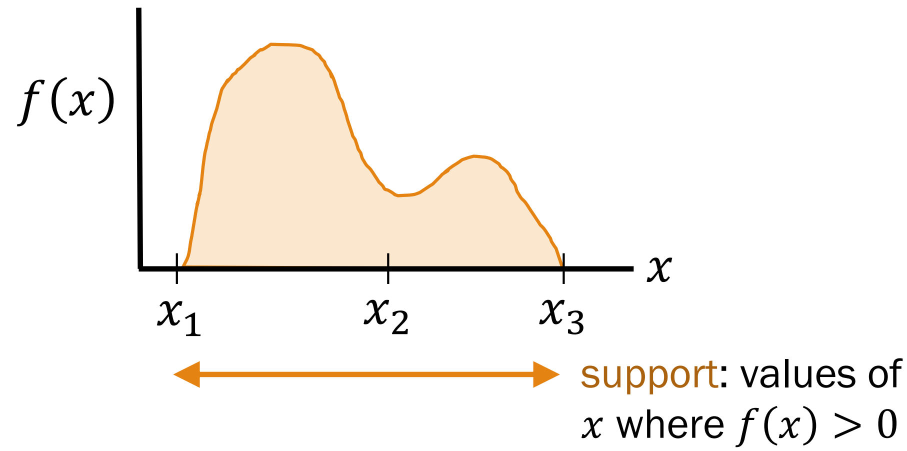
True/False:

1. $P(X = c) = 0$
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, $P(x_1 \leq X \leq x_2) > P(x_2 \leq X \leq x_3)$

PDF Properties

For a continuous RV X with PDF f,

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



True/False:

1. $P(X = c) = 0$
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,
   $P(x_1 \leq X \leq x_2) > P(x_2 \leq X \leq x_3)$

Interval width $dx \to 0$

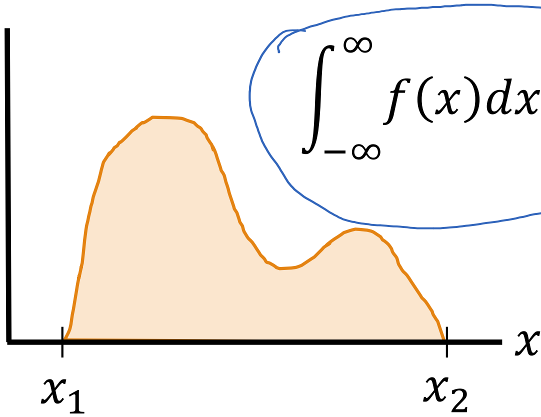
Compare area under the curve f

Determining valid PDFs

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

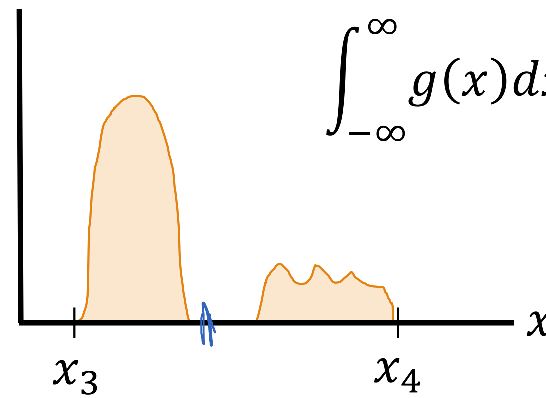
Which of the following functions are valid PDFs?

1. $f(x)$



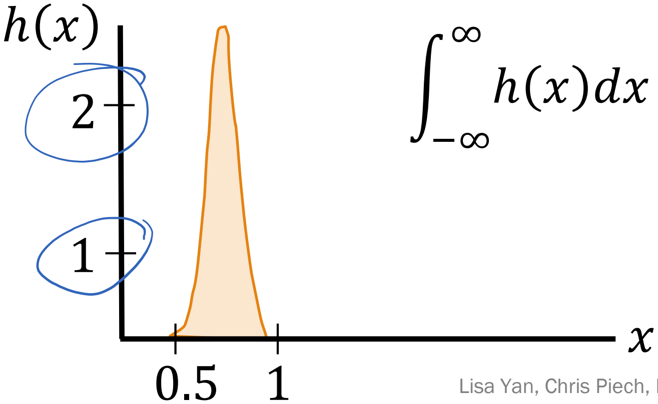
$$\int_{-\infty}^{\infty} f(x)dx = 0.5$$

2. $g(x)$



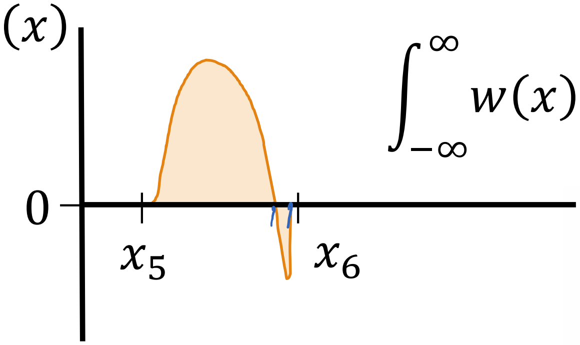
$$\int_{-\infty}^{\infty} g(x)dx = 1$$

3. $h(x)$

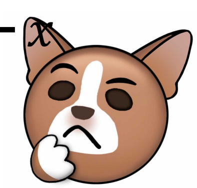


$$\int_{-\infty}^{\infty} h(x)dx = 1$$

4. $w(x)$



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



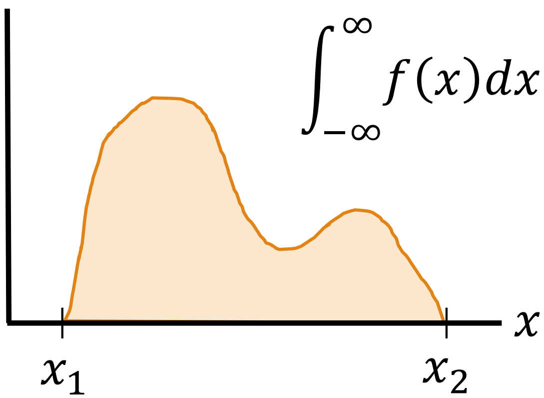
Determining valid PDFs

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

Which of the following functions are valid PDFs?

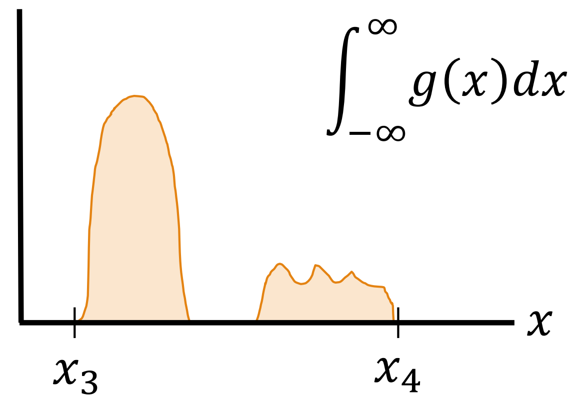
1. $f(x)$

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



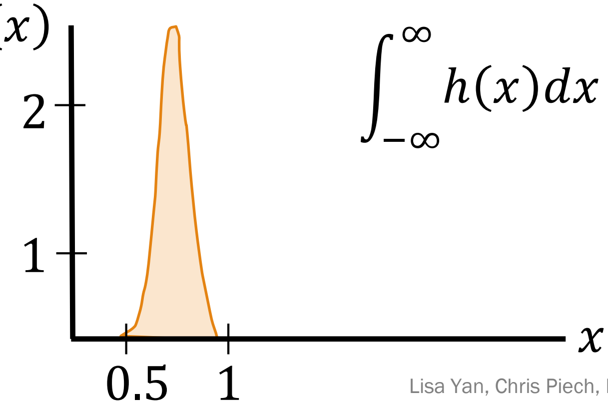
2. $g(x)$

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



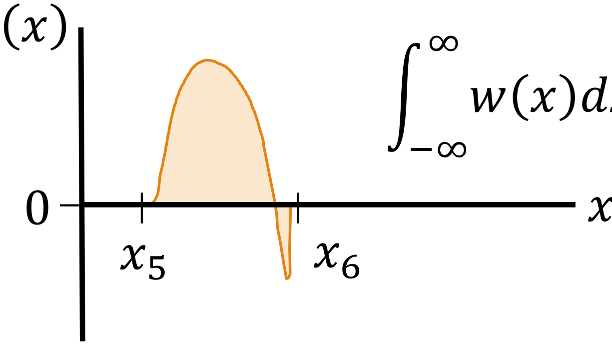
3. $h(x)$

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



4. $w(x)$

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



Uniform RV

Uniform Random Variable

<u>def</u> An **Uniform** random variable X is defined as follows:

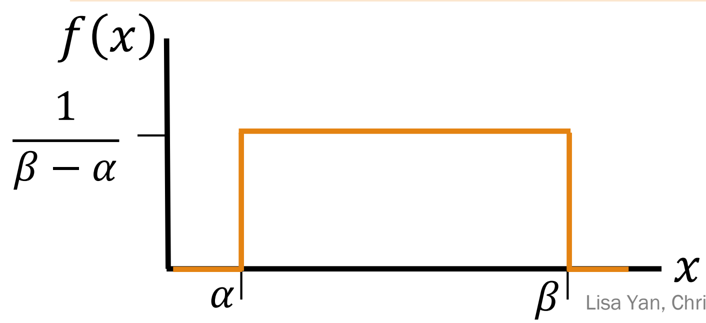
$$X \sim \text{Uni}(\alpha, \beta)$$

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

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

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

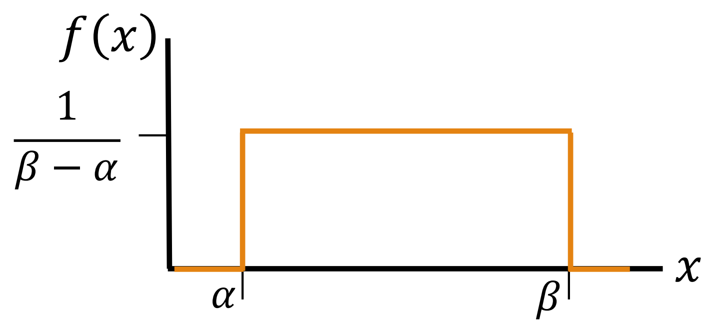
Variance $$\text{Var}(X) = \frac{(\beta - \alpha)^2}{12}$$



Quick check

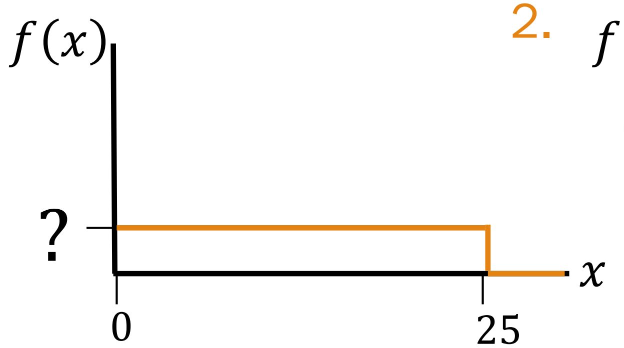
If $X \sim \text{Uni}(\alpha, \beta)$, the PDF of X is:

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

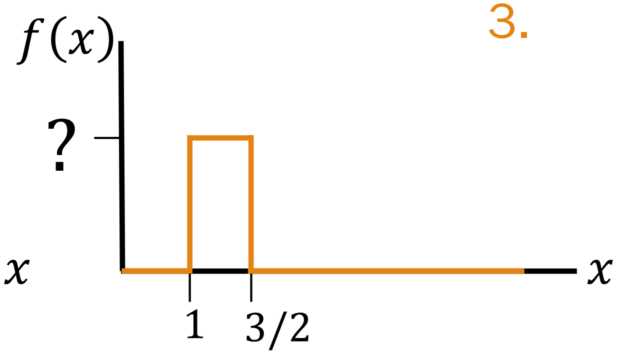


What is $\frac{1}{\beta - \alpha}$ if the following graphs are PDFs of Uniform RVs X?

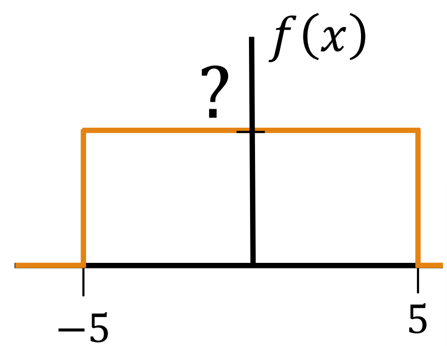
1.

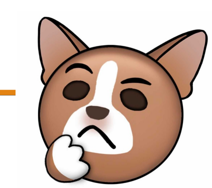


2.



3.

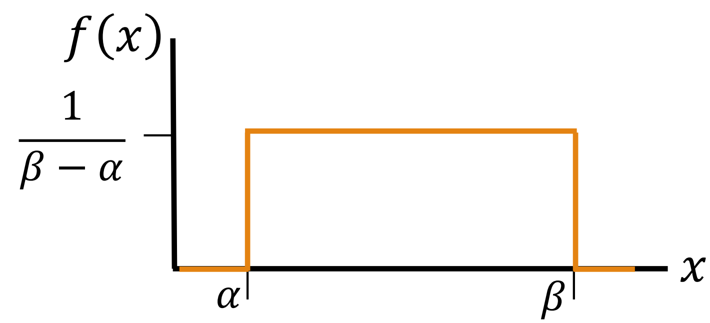




Quick check

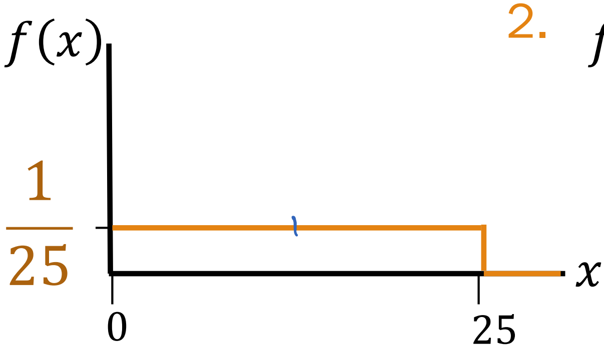
If $X \sim \text{Uni}(\alpha, \beta)$, the PDF of X is:

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

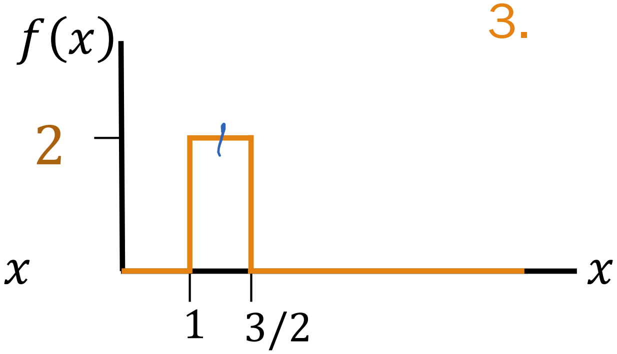


What is $\frac{1}{\beta - \alpha}$ if the following graphs are PDFs of Uniform RVs X?

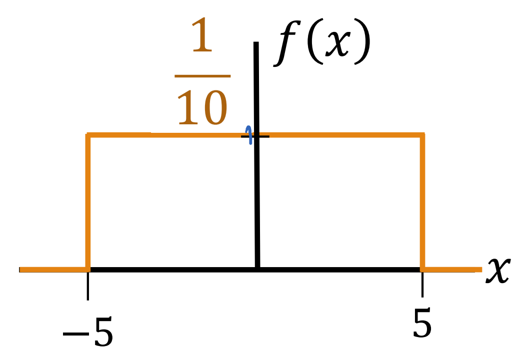
1.



2.



3.



Expectation and Variance

Discrete RV X

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

LOTUS

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

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

TL;DR: $\sum_{x=a}^{b} \Rightarrow \int_a^b dx$

Uniform RV expectation

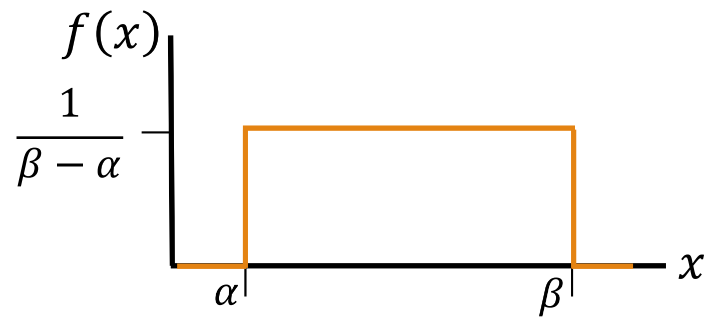
$$E[X] = \int_{-\infty}^{\infty} x \cdot f(x)dx$$

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

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

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

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



Interpretation:
Average the start & end

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))

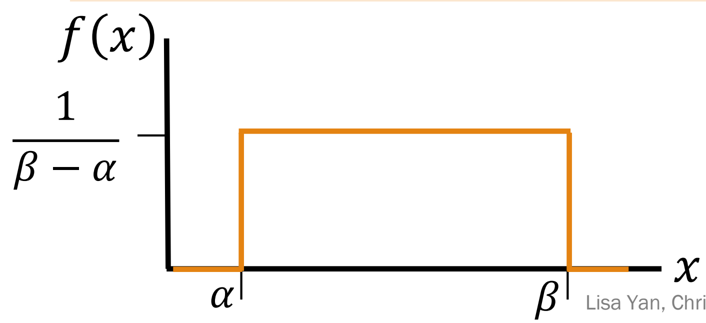
PDF $$f(x) = \begin{cases} \frac{1}{\beta - \alpha} & \text{if } \alpha \le x \le \beta \\ 0 & \text{otherwise} \end{cases}$$

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

Just now

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

On your own time



Exponential RV

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 first success occurs.

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

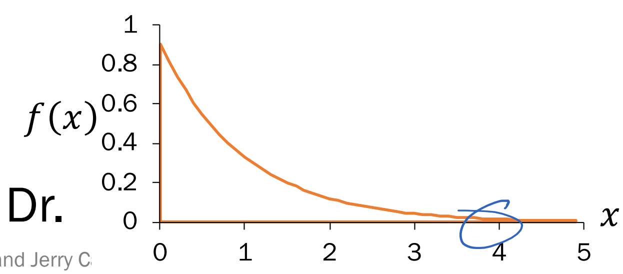
$X \sim \text{Exp}(\lambda)$

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

Examples:

- Time until next earthquake
- Time for request to reach web server
- Time until water main break on Campus Dr.



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?

Interpreting Exp(λ)

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

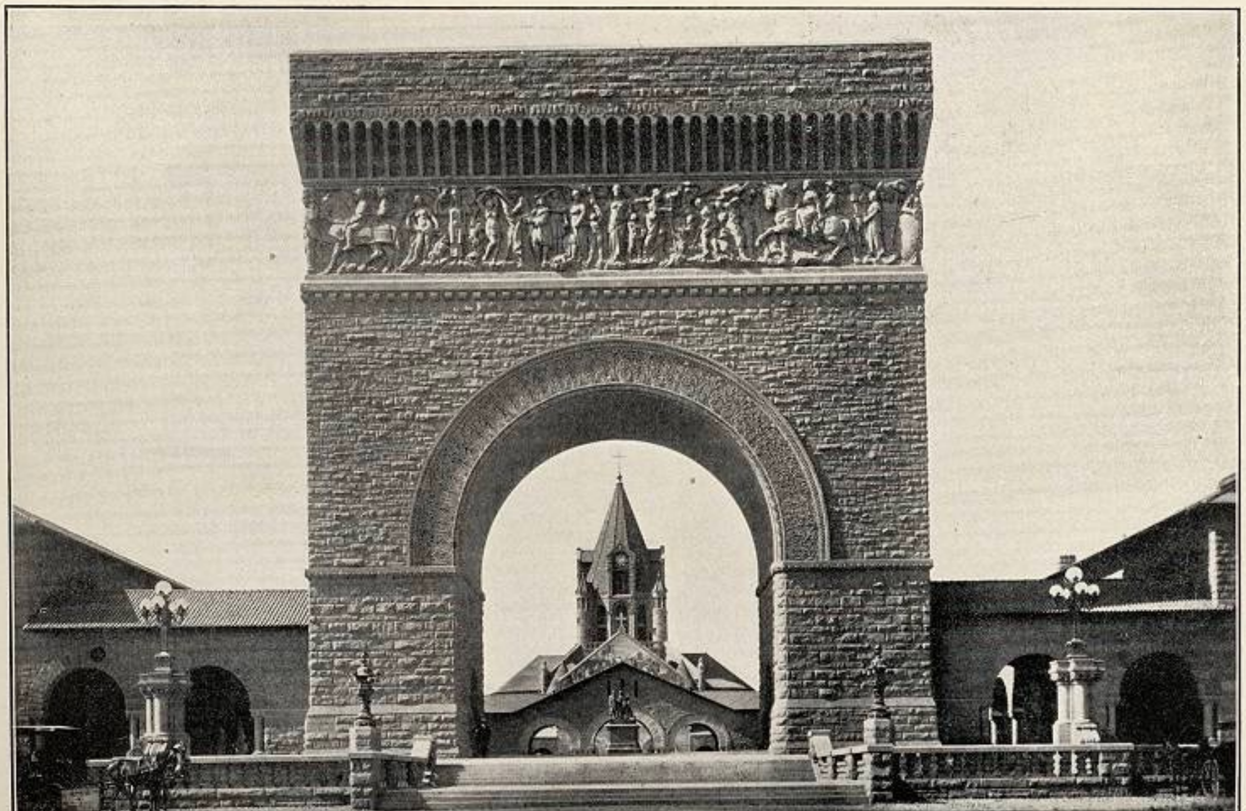
$X \sim \text{Exp}(\lambda)$ Expectation $E[X] = \frac{1}{\lambda}$

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

e.g., average # of successes per second

For both Poisson and Exponential RVs, $\lambda =$ # successes/time.

Earthquakes



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

1906 Earthquake
Magnitude 7.8

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?

We know on average:

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

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

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

λ

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

Earthquakes

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

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

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

Recall

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

$$= 0.002 \left[\frac{-1}{0.002} e^{-.002x} \right]_0^{30}$$

$$= -\left(e^{-.06} - e^{0}\right) \approx 0.058$$

*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?

$1/\lambda$

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

Cumulative Distribution Functions (CDFs)

Cumulative Distribution Function (CDF)

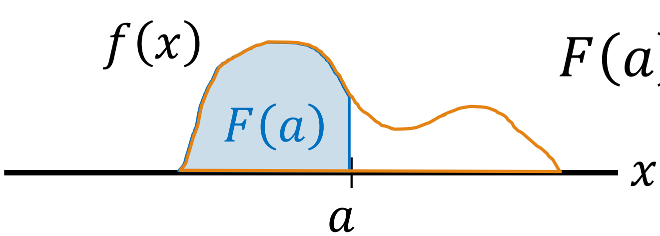
For a random variable X, the cumulative distribution function (CDF) is defined as

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

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:



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

CDF is a probability, though PDF is not.

If you learn to use CDFs, you can avoid integrating the PDF.

Using the CDF for continuous RVs

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

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

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. $F(b) - F(a)$

D. $F(a) - F(b)$

Using the CDF for continuous RVs

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

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

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

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

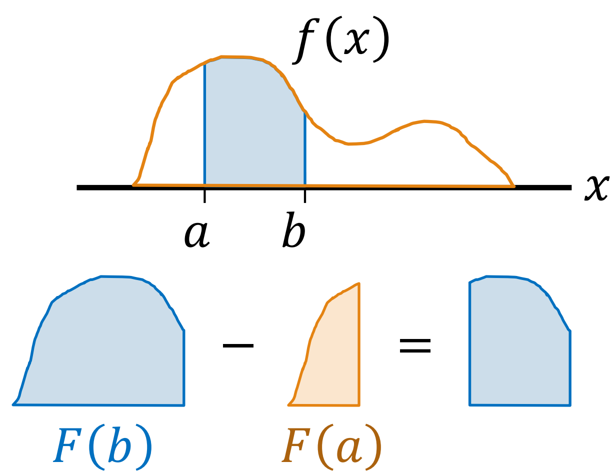
A. $F(a)$
B. $1 - F(a)$
C. $F(b) - F(a)$ (next slide)
D. $F(a) - F(b)$

Using the CDF for continuous RVs

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

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

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



Proof:

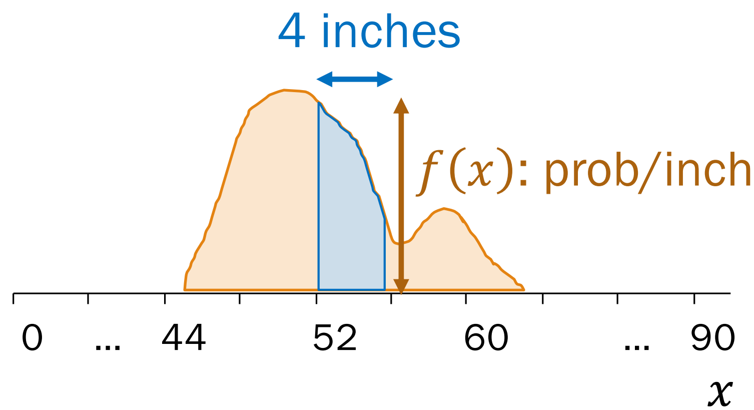
$$F(b) - F(a) = \int_{-\infty}^{b} f(x)dx - \int_{-\infty}^{a} f(x)dx$$

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

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

Addendum to today's main takeaway, #1

Integrate $f(x)$ to get probabilities.*



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

*If you have $F(a)$, you already have probabilities, since $F(a) = \int_{-\infty}^{a} f(x)dx$

CDF of an Exponential RV

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

$$X \sim \text{Exp}(\lambda) \qquad F(x) = 1 - e^{-\lambda x} \text{ if } x \geq 0$$

Proof:

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

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

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

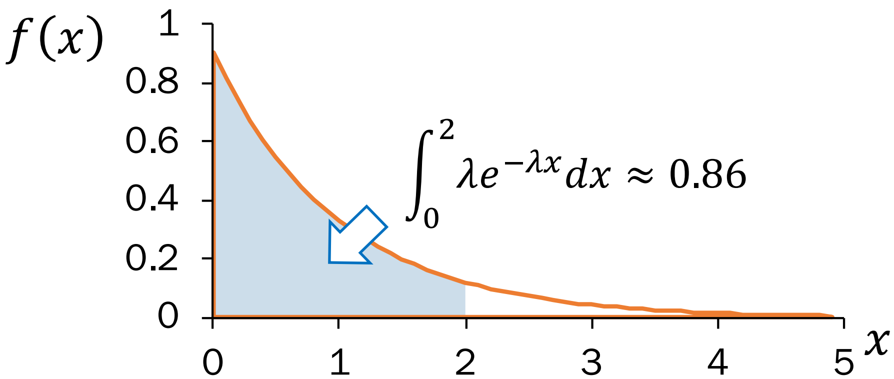
$$= 1 - e^{-\lambda x}$$

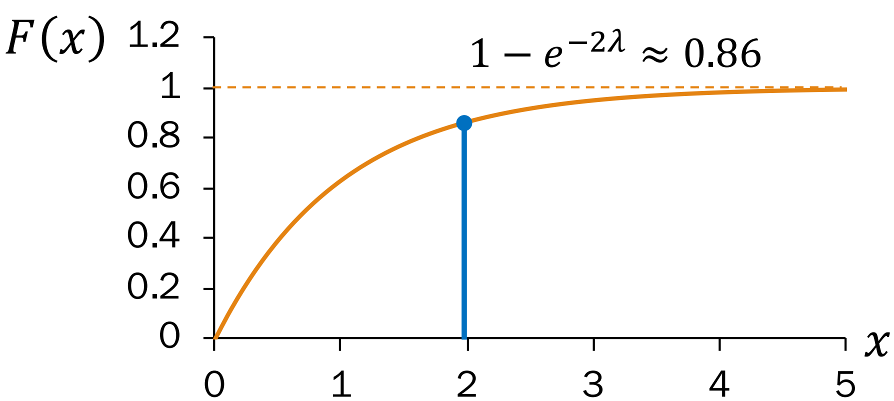
Recall

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

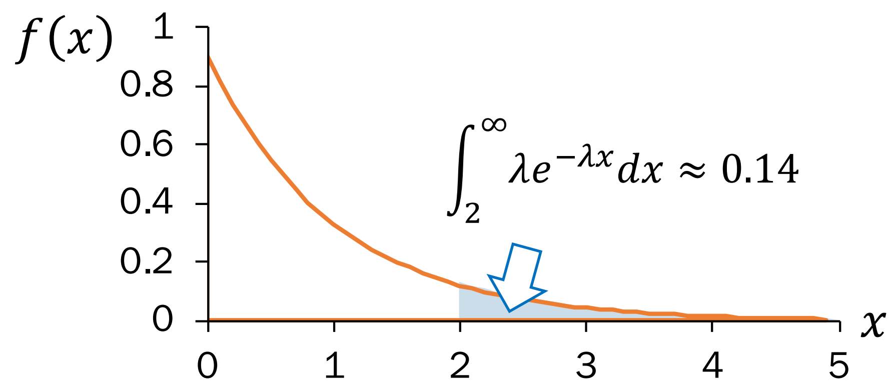
PDF/CDF $X \sim \text{Exp}(\lambda = 1)$

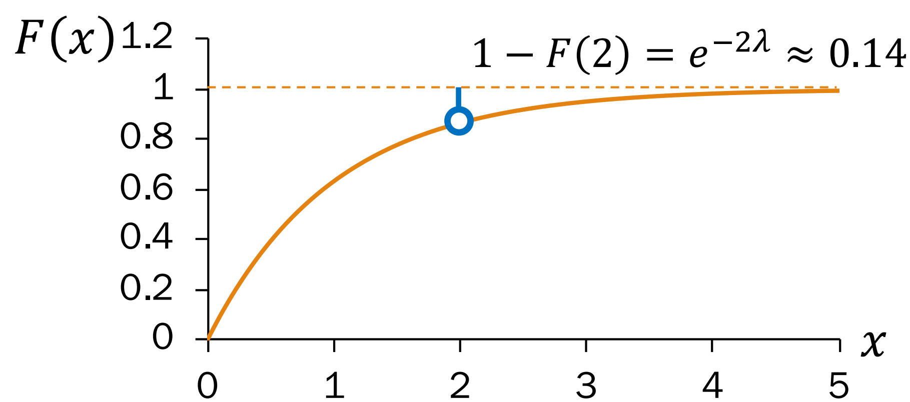
$X \sim \text{Exp}(\lambda)$ $\quad x \geq 0$: $f(x) = \lambda e^{-\lambda x}$, $F(x) = 1 - e^{-\lambda x}$





$P(X \leq 2)$





$P(X > 2)$

Exercises

Earthquakes

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

What is the probability of zero major earthquakes next year?

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

Earthquakes

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

What is the probability of zero major earthquakes next year?

Strategy 1: Exponential RV

Define events/RVs & state goal

T: when first earthquake happens

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

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

Solve

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

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

Earthquakes

$Y \sim \text{Poi}(\lambda) \quad p(k) = e^{-\lambda}\frac{\lambda^k}{k!}$

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

What is the probability of zero major earthquakes next year?

Strategy 1: Exponential RV

Define events/RVs & state goal

T: when first earthquake happens

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

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

Solve

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

Strategy 2: Poisson RV

Define events/RVs & state goal

N: # earthquakes next year

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

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

Want: $P(N = 0)$

Solve

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

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

Website visits

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

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

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

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

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

Extra

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:

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

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

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

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

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

Integration by parts

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

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

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

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

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