Chapter 7
Continuous Random Variables

Continuous Random Variable

Defn: A **continuous random variable (r.v.)** has a continuous range of values that it can take on. This might be an interval or set of intervals.
Thus a continuous r.v. can take on an **uncountable** set of possible values.

Examples:

- ❑ Time of an event
- ❑ Response time of a job
- ❑ Speed of a device
- ❑ Location of a satellite
- ❑ Distance between people's eyeballs

Probability for Continuous Random Variable

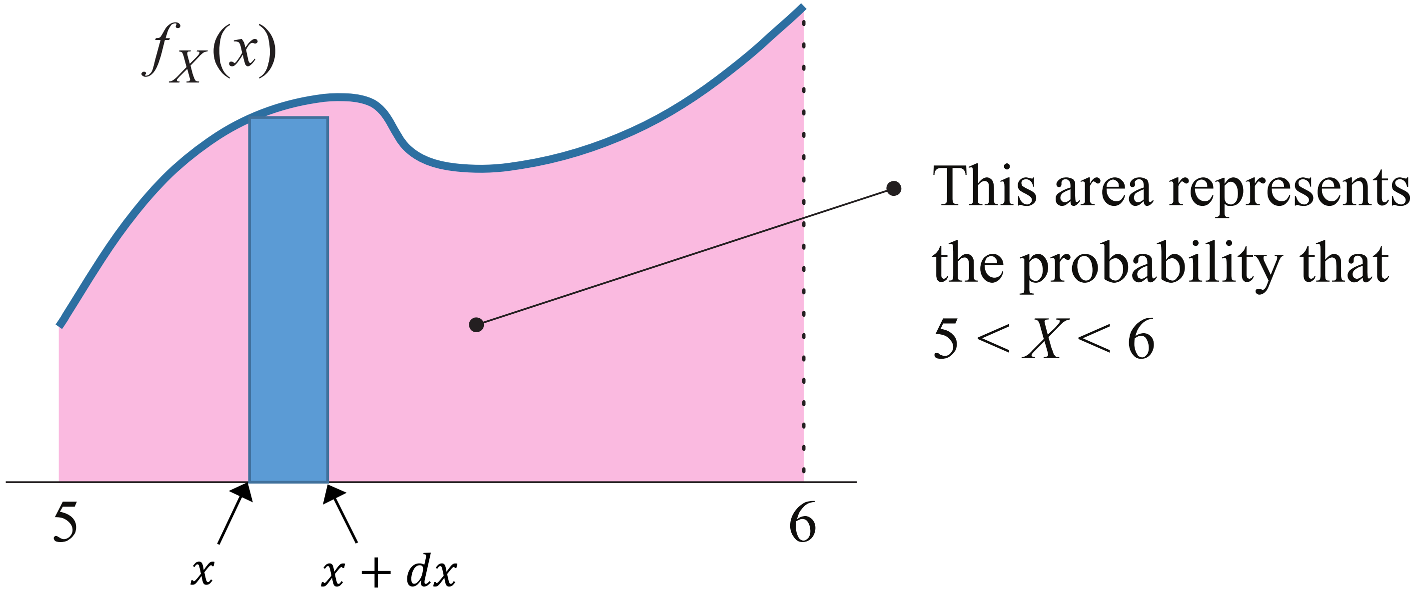
The probability that a continuous r.v. is equal to any particular value is defined to be 0.

Probability for a continuous r.v. is defined via a density function.

Defn 7.2: The **probability density function (p.d.f.)** of a continuous r.v. X is a non-negative function $f_X(\cdot)$, where

$$P\{a \leq X \leq b\} = \int_a^b f_X(x)dx \quad \text{and} \quad \int_{-\infty}^{\infty} f_X(x)dx = 1$$

Probability for Continuous Random Variable



$$f(x)dx \approx \mathbf{P}\{x \le X \le x + dx\}$$

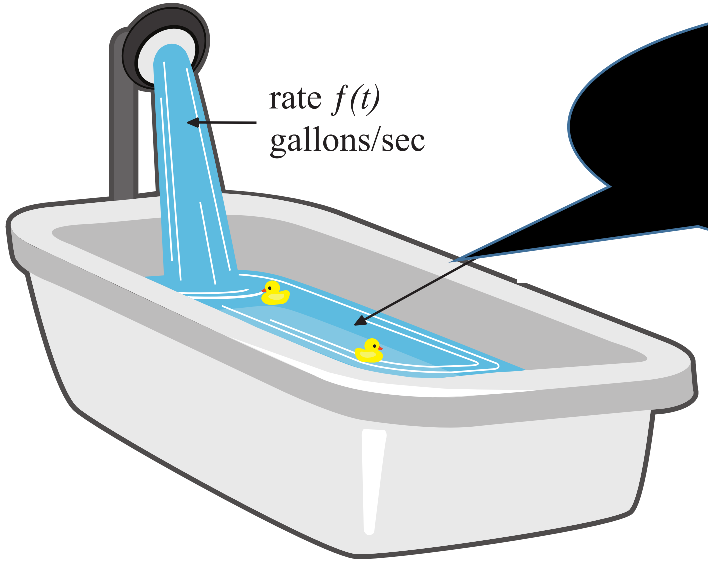
How do $\mathbf{P}\{5 \le X \le 6\}$ and $\mathbf{P}\{5 < X < 6\}$ compare?

Can $f_X(x)$ be larger than 1?

Density as a rate

Density functions are not necessarily related to probability.

<u>Example</u>: Filling a bathtub at rate $f_X(t) = t^2$ gallons/sec, where $t \geq 0$



Q: How much water after 4 seconds?

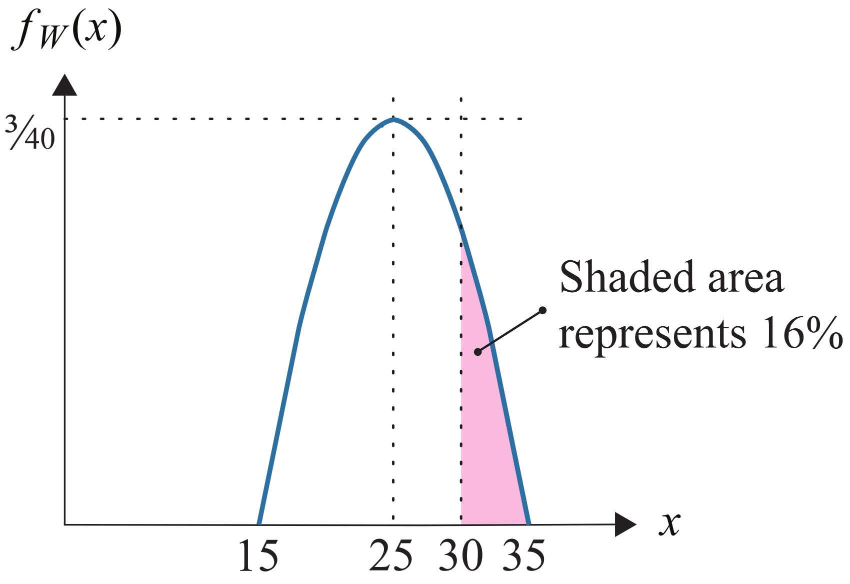
A: $$\int_0^4 f_X(t)dt = \int_0^4 t^2 dt = \frac{64}{3}$$ gallons

Q: Is $f_X(t)$ a p.d.f.?

Example: Computing probability from p.d.f.

Weight of two-year-olds ranges between 15 and 35 pounds with p.d.f. $f_W(x)$:

$$f_W(x) = \begin{cases} \frac{3}{40} - \frac{3}{4000}(x-25)^2 & \text{if } 15 \leq x \leq 35 \\ 0 & \text{otherwise} \end{cases}$$



Q: What is the fraction of two-year-olds that weigh > 30 pounds?

A:

$$\int_{30}^{\infty} f_W(x)dx = \int_{30}^{35} \frac{3}{40} - \frac{3}{4000}(x-25)^2 dx \approx 16\%$$

Cumulative distribution function

<u>Defn</u>: The **cumulative distribution function (c.d.f.)** of a continuous r.v. X is given by:

$$F_X(a) = \boldsymbol{P}\{-\infty < X \leq a\} = \int_{-\infty}^{a} f_X(x)dx$$

The **tail** of X is given by:

$$\bar{F}_X(a) = 1 - F_X(a) = \boldsymbol{P}\{X > a\}$$

Q: How do we get $f_X(x)$ from $F_X(x)$?

F.T.C.

A: $f_X(x) = \frac{d}{dx}\int_{-\infty}^{x} f_X(t)dt = \frac{d}{dx}F_X(x)$

(See Section 1.3 of your book)

Uniform distribution

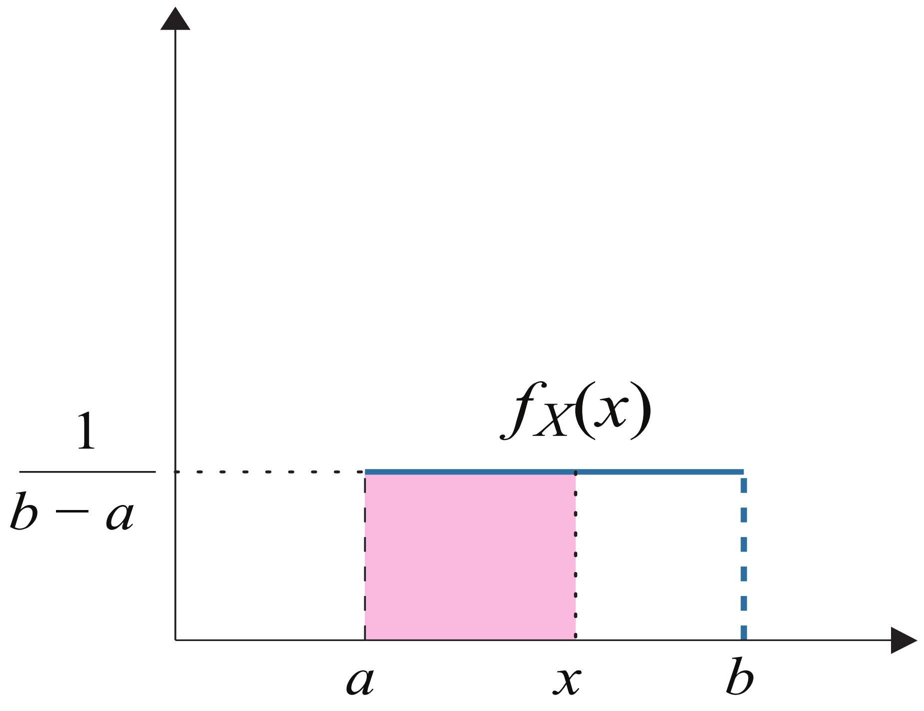
<u>Defn</u>: **Uniform$(\boldsymbol{a}, \boldsymbol{b})$**, often written $U(a, b)$, models the fact that any interval of length δ between a and b is equally likely. Specifically, if $X \sim U(a, b)$, then

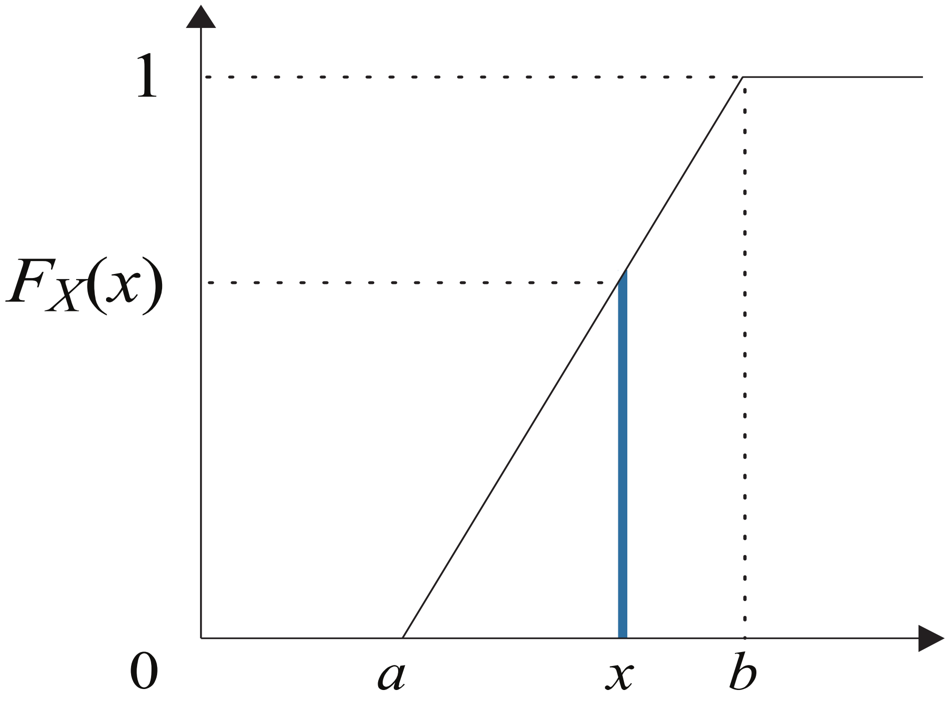
$$f_X(x) = \begin{cases} \dfrac{1}{b-a} & \text{if } a \le x \le b \\ 0 & \text{otherwise} \end{cases}$$

Q: If $X \sim U(a, b)$, what is $F_X(x)$?

A: $F_X(x) = \int_a^x \frac{1}{b-a} dt = \frac{x-a}{b-a}$ if $a \le x \le b$

Graphical depiction of Uniform distribution



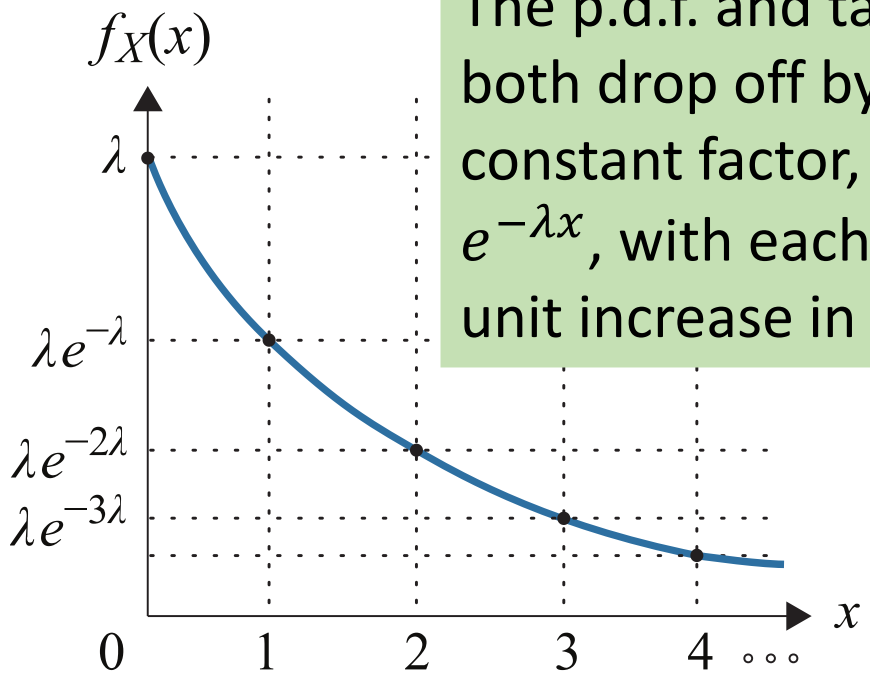


Area of shaded pink region on left $=$ Height of blue line on right

Exponential distribution

<u>Defn</u>: **Exp(λ)** denotes the Exponential distribution with **rate λ**.

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

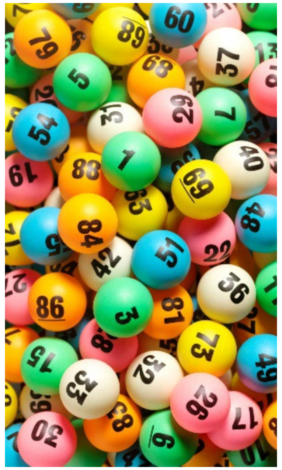


The p.d.f. and tail both drop off by a constant factor, $e^{-\lambda x}$, with each unit increase in x.

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt = \begin{cases} 1 - e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$

$$\bar{F}_X(x) = e^{-\lambda x}, \qquad x \geq 0$$

Memorylessness



Defn: Random variable X has the **memoryless property** if

$$\boldsymbol{P}\{X > t + s \mid X > s\} = \boldsymbol{P}\{X > t\} \qquad \forall s, t \geq 0$$

$X =$ Time to win lottery.

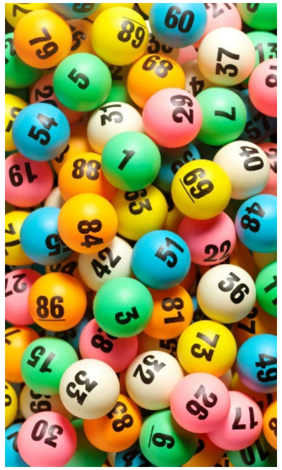
Suppose I haven't won the lottery by time s. Then the probability that I'll need $> t$ more time to win is independent of s.

Equivalently: X has the **memoryless property** if

$$[X \mid X > s] \stackrel{d}{=} s + X \qquad \forall s \geq 0$$

That is, the r.v.s $[X \mid X > s]$ and $s + X$ have the same distribution.

Memorylessness



Defn: Random variable X has the **memoryless property** if

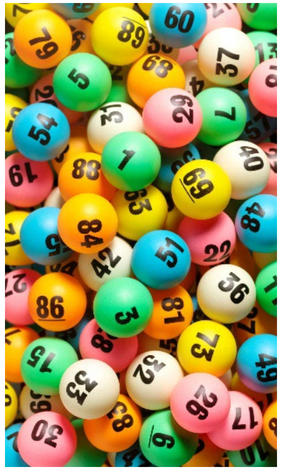
$$\boldsymbol{P}\{X > t + s \mid X > s\} = \boldsymbol{P}\{X > t\} \qquad \forall s, t \geq 0$$

Q: Prove that if $X \sim Exp(\lambda)$, then X has the memoryless property.

A: First recall that: $\bar{F}_X(x) = e^{-\lambda x}, \qquad x \geq 0$

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

Memorylessness



Defn: Random variable X has the **memoryless property** if

$$\boldsymbol{P}\{X > t + s \mid X > s\} = \boldsymbol{P}\{X > t\} \qquad \forall s, t \geq 0$$

Q: What other distribution has the memoryless property?

A: The Geometric distribution

Q: Does $X \sim Uniform(a, b)$ also have the memoryless property?

A: No. If $X \sim Uniform(a, b)$ and we're given that $X > b - \epsilon$, then we know that X will end soon.

Memorylessness Example

Mortality rate normally increases with age.
But not for the naked mole-rat!
Its remaining lifetime is independent of its age.

Q: Let $X \sim Exp(1)$ denote the lifetime of the naked mole-rat in years.
If a naked mole-rate is 4 years old, what is the probability of surviving at least one more year?

A:

$$\boldsymbol{P}\{X > 4 + 1 \mid X > 4\} = \frac{\boldsymbol{P}\{X > 5\}}{\boldsymbol{P}\{X > 4\}} = \frac{e^{-5}}{e^{-4}} = e^{-1} = \boldsymbol{P}\{X > 1\}$$

Post Office Example

A post office has 2 clerks.
When customer A walks in, customer B is being served by one clerk, and customer C is being served by the other.
All service times $\sim Exp(\lambda)$.

B C

Q: What is $\boldsymbol{P}\{$A is last to leave$\}$?

A: $\frac{1}{2}$ One of B or C will leave first. At that point, the remaining customer's lifetime restarts. A will then compete with that remaining customer.

Expectation, Variance, and Higher Moments

Defn: For a continuous r.v. X with p.d.f. $f_X(\cdot)$, we have:

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

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

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

$$\boldsymbol{Var}(X) = \boldsymbol{E}[(X - \boldsymbol{E}[X])^2] = \int_{-\infty}^{\infty} (x - \boldsymbol{E}[X])^2 f_X(x)dx = \boldsymbol{E}[X^2] - \boldsymbol{E}[X]^2$$

Uniform distribution: Mean and Variance

Q: Derive mean and variance of $X \sim U(a,b)$.

$X \sim Uniform(a,b)$

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{if } a \le x \le b \\ 0 & \text{otherwise} \end{cases}$$

A:

$$\boldsymbol{E}[X] = \int_a^b \frac{1}{b-a} \cdot t\,dt = \frac{1}{b-a} \cdot \frac{b^2-a^2}{2} = \frac{a+b}{2}$$

$$\boldsymbol{E}[X^2] = \int_a^b \frac{1}{b-a} \cdot t^2 dt = \frac{1}{b-a} \cdot \frac{b^3-a^3}{3} = \frac{b^2+ab+a^2}{3}$$

$$\boldsymbol{Var}(X) = \boldsymbol{E}[X^2] - \boldsymbol{E}[X]^2 = \frac{(b-a)^2}{12}$$

Exponential distribution: Mean and Variance

Q: Derive mean and variance of $X \sim Exp(\lambda)$.

$X \sim Exp(\lambda)$

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

A:

$$\boldsymbol{E}[X] = \int_{-\infty}^{\infty} \lambda e^{-\lambda t}\, t\, dt = \frac{1}{\lambda}$$

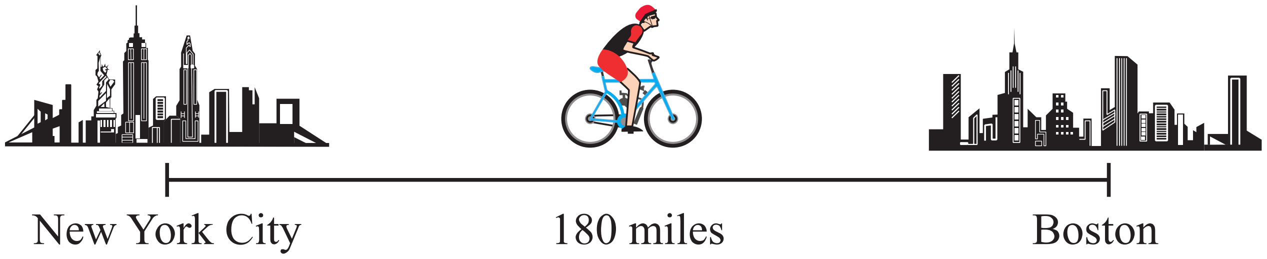
$$\boldsymbol{E}[X^2] = \int_{-\infty}^{\infty} \lambda e^{-\lambda t} \cdot t^2 dt = \frac{2}{\lambda^2}$$

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

The λ parameter is the reciprocal of the mean (hence “the rate”).

Example: Time to get from NYC to Boston

Distance from NYC to Boston is 180 miles.
Motorized bikes have speeds $\sim U(30,60)$.
You buy a random motorized bike.
$T =$ Your time to get from NYC to Boston.



Goal: Derive $\boldsymbol{E}[T]$.

<u>Idea 1</u>: Avg. speed is 45 mph. Thus $\boldsymbol{E}[T] = \frac{180}{45} = 4$ hours.

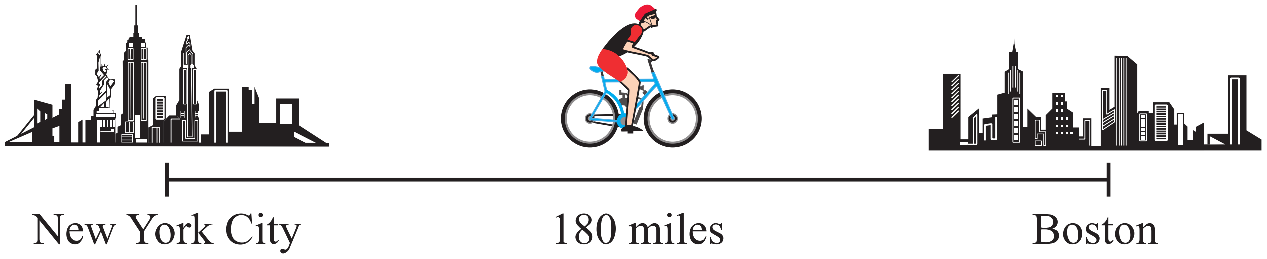
<u>Idea 2</u>: $\boldsymbol{E}[T]$ is the average of $\frac{180}{30} = 6$ and $\frac{180}{60} = 3$. So $\boldsymbol{E}[T] = 4.5$ hours.

Q: Which is correct, Idea 1 or Idea 2?

A: Neither!

Example: Time to get from NYC to Boston

Distance from NYC to Boston is 180 miles.
Motorized bikes have speeds $\sim U(30,60)$.
You buy a random motorized bike.
$T =$ Your time to get from NYC to Boston.



Q: What is $\boldsymbol{E}[T]$?

A: Let $S \sim U(30,60)$ represent the speed of your bike. Then $T = \frac{180}{S}$

$$\boldsymbol{E}[T] = \boldsymbol{E}\left[\frac{180}{S}\right] = \int_{30}^{60} \frac{180}{s} f_S(s) ds = \int_{30}^{60} \frac{180}{s} \cdot \frac{1}{30} ds$$

$$= 6\,(\ln 60 - \ln 30)$$

$$= 6\,(\ln 2) \approx 4.15 \text{ hours}$$

Law of Total Probability for Continuous

Recall the **Law of Total Probability** for event A and **discrete** r.v. X :

$$\boldsymbol{P}\{A\} = \sum_x \boldsymbol{P}\{A \cap (X = x)\} = \sum_x \boldsymbol{P}\{A \mid X = x\} \cdot p_X(x)$$

The same **Law of Total Probability** holds for event A and **continuous** r.v. X :

$$\boldsymbol{P}\{A\} = \int_x f_X(x \cap A)dx = \int_x \boldsymbol{P}\{A \mid X = x\} \cdot f_X(x)dx$$

Here $f_X(x \cap A)$ denotes the density of the intersection of the event A with $X = x$.

Law of Total Probability for Continuous

$$\boldsymbol{P}\{A\} = \int_x f_X(x \cap A)dx = \int_x \boldsymbol{P}\{A \mid X = x\} \cdot f_X(x)dx$$

Here $f_X(x \cap A)$ denotes the density of the intersection of the event A with $X = x$.

Example: Let A be the event $X > 50$.

$$f_X(x \cap A) = \begin{cases} f_X(x) & \text{if } x > 50 \\ 0 & \text{otherwise} \end{cases}$$

$$\boldsymbol{P}\{X > 50\} = \boldsymbol{P}\{A\} = \int_{-\infty}^{\infty} f_X(x \cap A)dx = \int_{50}^{\infty} f_X(x)dx$$

Likewise,

$$\boldsymbol{P}\{X > 50\} = \int_{-\infty}^{\infty} \boldsymbol{P}\{X > 50 \mid X = x\} \cdot f_X(x)dx = \int_{50}^{\infty} 1 \cdot f_X(x)dx$$

Conditioning on a Zero-Probability Event

$$\boldsymbol{P}\{A\} = \int_x f_X(x \cap A)dx = \int_x \boldsymbol{P}\{A \mid X = x\} \cdot f_X(x)dx$$

Here $f_X(x \cap A)$ denotes the density of the intersection of the event A with $X = x$.

Q: In $\boldsymbol{P}\{A \mid X = x\}$, we're conditioning on a zero-probability event. So we have a zero in the denominator. How is this okay?

$$f_X(x \cap A) = \begin{cases} f_X(x) & \text{if } x \in A \\ 0 & \text{otherwise} \end{cases}$$

A:

$$\boldsymbol{P}\{A \mid X = x\} = \frac{f_X(x \cap A)}{f_X(x)}$$

The ratio is between densities, not probabilities, and the densities are not zero!

Conditioning on a Zero-Probability Event

Example: We have a coin with unknown bias.
Specifically, the coin has probability P of heads where $P \sim Uniform(0,1)$.

Q: What is $\boldsymbol{P}\{\text{Next 10 flips are all heads}\}$?

A:

$$\boldsymbol{P}\{10\ heads\} = \int_0^1 P\{10\ heads \mid P = p\} \cdot f_P(p)dp$$

$$= \int_0^1 P\{10\ heads \mid P = p\} \cdot 1dp$$

$$= \int_0^1 p^{10} \cdot 1dp$$

$$= \frac{1}{11}$$

Conditional p.d.f. and Bayes' Law

<u>Defn</u>: For a continuous r.v. X and an event A, the **conditional p.d.f. of r.v. $\boldsymbol{X}$ given $\boldsymbol{A}$** is:

$$f_{X|A}(x) = \frac{f_X(x \cap A)}{\boldsymbol{P}\{A\}} = \frac{\boldsymbol{P}\{A \mid X = x\} \cdot f_X(x)}{\boldsymbol{P}\{A\}}$$

<u>Comments</u>:

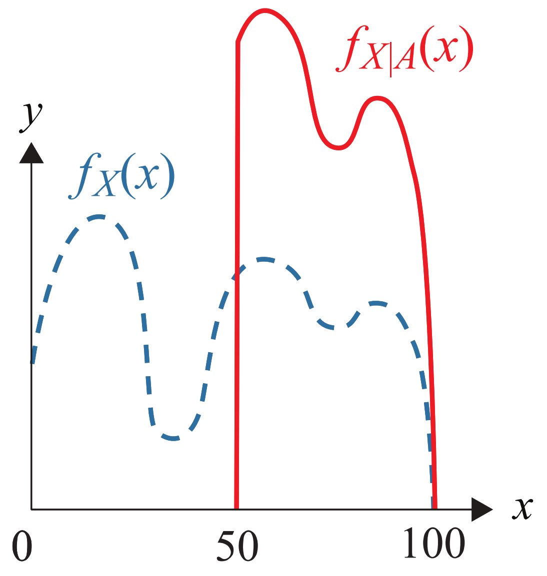
1. Conditional p.d.f $f_{X|A}(x)$ has value 0 outside of A.
2. The conditional p.d.f. is still a proper p.d.f. in that

$$\int_x f_{X|A}(x)dx = 1$$

Conditional p.d.f. and Bayes' Law

Defn: For a continuous r.v. X and an event A, the **conditional p.d.f. of r.v. X given A** is:

$$f_{X|A}(x) = \frac{f_X(x \cap A)}{\boldsymbol{P}\{A\}} = \frac{\boldsymbol{P}\{A \mid X = x\} \cdot f_X(x)}{\boldsymbol{P}\{A\}}$$



Example:
X has p.d.f. $f_X(x)$ defined on $0 < x < 100$.
A is the event $X > 50$.

$f_{X|A}(x)$ is a scaled-up version of $f_X(x)$, allowing it to integrate to 1.

$$f_{X|X>50}(x) = \frac{f_X(x \cap X > 50)}{\boldsymbol{P}\{X > 50\}} = \begin{cases} \dfrac{f_X(x)}{\boldsymbol{P}\{X > 50\}} & \text{if } x > 50 \\ 0 & \text{otherwise} \end{cases}$$

Conditional expectation

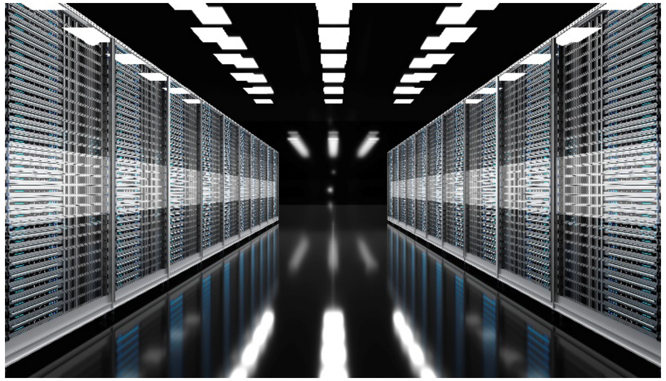
<u>Defn</u>:
For a **discrete** r.v. X and an event A, where $P\{A\} > 0$, the **conditional expectation of $\boldsymbol{X}$ given $\boldsymbol{A}$** is:

$$\boldsymbol{E}[X|A] = \sum_x x \cdot p_{X|A}(x)$$

For a **continuous** r.v. X and an event A, where $P\{A\} > 0$, the **conditional expectation of $\boldsymbol{X}$ given $\boldsymbol{A}$** is:

$$\boldsymbol{E}[X|A] = \int_x x \cdot f_{X|A}(x)dx$$

Conditional expectation example



Example: **Pittsburgh Supercomputing Center (PSC)**
At the PSC, jobs are grouped into different bins based on their size.
Suppose job sizes are Exponentially distributed with
mean 1000 CPU-hours.
Suppose all jobs of size < 500 CPU-hours are sent to bin 1.

a. What is $\boldsymbol{P}\{\text{Job is sent to bin 1}\}$?

b. What is $\boldsymbol{P}\{\text{Job size} < 200 \mid \text{job is in bin 1}\}$?

c. What is $f_{X|A}(x)$, where X is the job size and A is the event that the job is in bin 1?

d. What is $\boldsymbol{E}[\text{Job size} \mid \text{job is in bin 1}]$?

Conditional expectation example

<u>Example</u>: **Pittsburgh Supercomputing Center (PSC)**
At the PSC, jobs are grouped into bins based on their size.
Suppose job sizes are Exponentially distributed with *mean* 1000 CPU-hours.
Suppose all jobs of size < 500 CPU-hours are sent to bin 1.

$$X \sim Exp\left(\frac{1}{1000}\right)$$

$$f_X(x) = \begin{cases} \frac{1}{1000} e^{-\frac{1}{1000}x} & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$

$$\bar{F}_X(x) = \boldsymbol{P}\{X > x\} = e^{-\frac{1}{1000}x}$$

a. What is $\boldsymbol{P}\{\text{Job is sent to bin 1}\}$?

$$\boldsymbol{P}\{\text{Job is sent to bin 1}\} = F_X(500) = 1 - e^{-\frac{500}{1000}} = 1 - e^{-\frac{1}{2}} \approx 0.39$$

Conditional expectation example

<u>Example</u>: **Pittsburgh Supercomputing Center (PSC)**
At the PSC, jobs are grouped into bins based on their size.
Suppose job sizes are Exponentially distributed with *mean* 1000 CPU-hours.
Suppose all jobs of size < 500 CPU-hours are sent to bin 1.

$$X \sim Exp\left(\frac{1}{1000}\right)$$

$$f_X(x) = \begin{cases} \frac{1}{1000} e^{-\frac{1}{1000}x} & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$

$$\bar{F}_X(x) = \boldsymbol{P}\{X > x\} = e^{-\frac{1}{1000}x}$$

b. What is $\boldsymbol{P}\{\text{Job size} < 200 \mid \text{job is in bin } 1\}$?

$$\boldsymbol{P}\{\text{Job size} < 200 \mid \text{job is in bin } 1\} = \frac{\boldsymbol{P}\{X < 200 \cap \text{bin } 1\}}{\boldsymbol{P}\{\text{bin } 1\}} = \frac{F_X(200)}{F_X(500)} \approx 0.46$$

Conditional expectation example

<u>Example</u>: **Pittsburgh Supercomputing Center (PSC)**
At the PSC, jobs are grouped into bins based on their size.
Suppose job sizes are Exponentially distributed with *mean* 1000 CPU-hours.
Suppose all jobs of size < 500 CPU-hours are sent to bin 1.

$$X \sim Exp\left(\frac{1}{1000}\right)$$

$$f_X(x) = \begin{cases} \frac{1}{1000} e^{-\frac{1}{1000}x} & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$

$$\overline{F}_X(x) = \boldsymbol{P}\{X > x\} = e^{-\frac{1}{1000}x}$$

c. What is $f_{X|A}(x)$, where X is the job size and A is the event that the job is in bin 1?

$$f_{X|A}(x) = \frac{f_X(x \cap A)}{\boldsymbol{P}\{A\}} = \frac{f_X(x \cap A)}{F_X(500)} = \begin{cases} \frac{f_X(x)}{F_X(500)} = \frac{1}{1 - e^{-\frac{1}{2}}} \cdot \frac{1}{1000} e^{-\frac{1}{1000}x} & \text{if } x < 500 \\ 0 & \text{otherwise} \end{cases}$$

Conditional expectation example

<u>Example</u>: **Pittsburgh Supercomputing Center (PSC)**
At the PSC, jobs are grouped into bins based on their size.
Suppose job sizes are Exponentially distributed with *mean* 1000 CPU-hours.
Suppose all jobs of size < 500 CPU-hours are sent to bin 1.

$$X \sim Exp\left(\frac{1}{1000}\right)$$

$$f_X(x) = \begin{cases} \frac{1}{1000} e^{-\frac{1}{1000}x} & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$

$$\bar{F}_X(x) = \boldsymbol{P}\{X > x\} = e^{-\frac{1}{1000}x}$$

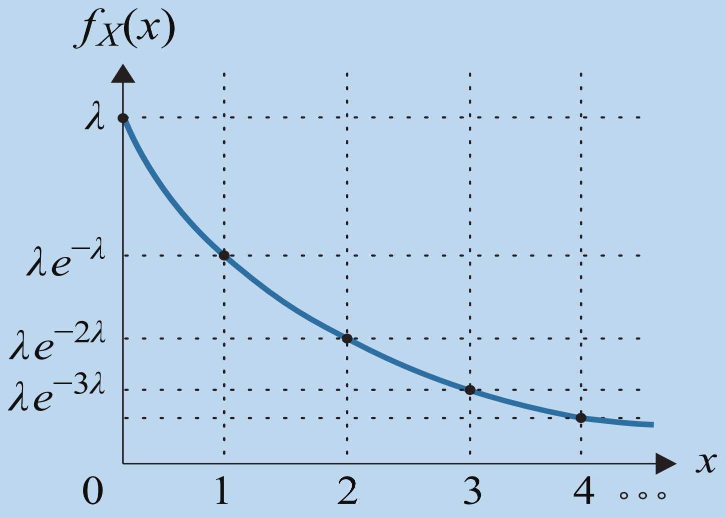
d. What is $\boldsymbol{E}$[Job size | job is in bin 1]?

$$\boldsymbol{E}[\text{Job size} \mid \text{job is in bin 1}] = \int_{-\infty}^{\infty} x\, f_{X|A}(x)dx = \int_0^{500} x \cdot \frac{1}{1 - e^{-\frac{1}{2}}} \cdot \frac{1}{1000} e^{-\frac{1}{1000}x} dx \approx 229$$

Conditional expectation example

<u>Example</u>: **Pittsburgh Supercomputing Center (PSC)**
At the PSC, jobs are grouped into bins based on their size.
Suppose job sizes are Exponentially distributed with *mean* 1000 CPU-hours.
Suppose all jobs of size < 500 CPU-hours are sent to bin 1.

$$X \sim Exp\left(\frac{1}{1000}\right)$$



d. What is $\boldsymbol{E}$[Job size | job is in bin 1]?

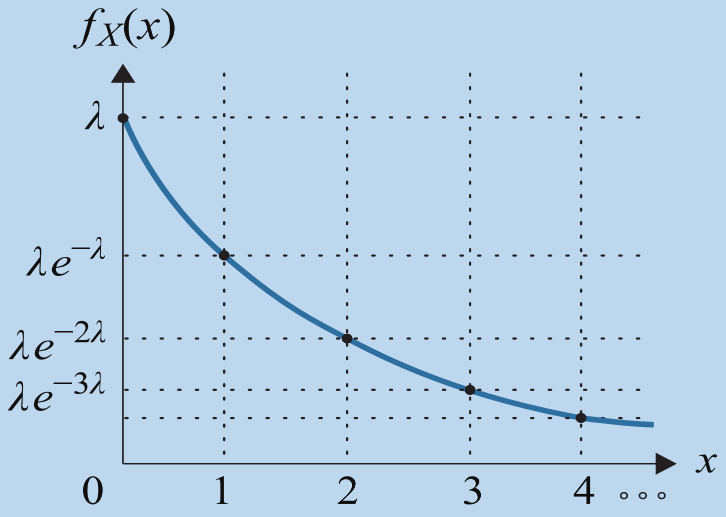
$\boldsymbol{E}$[Job size | job is in bin 1] ≈ 229

Why is the expected job size for bin 1 < 250?

Conditional expectation example

Example: **Pittsburgh Supercomputing Center (PSC)**
At the PSC, jobs are grouped into bins based on their size.
Suppose job sizes are Exponentially distributed with *mean* 1000 CPU-hours.
Suppose all jobs of size < 500 CPU-hours are sent to bin 1.

$$X \sim Exp\left(\frac{1}{1000}\right)$$



d. What is $\boldsymbol{E}$[Job size | job is in bin 1]?

$\boldsymbol{E}$[Job size | job is in bin 1] ≈ 229

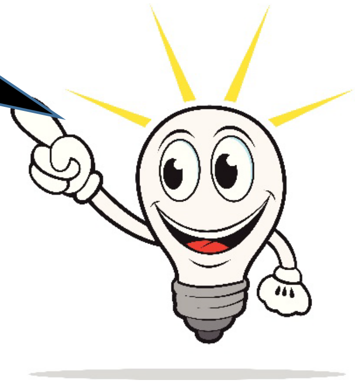
How would the above answer change if $X \sim Uniform(0{,}2000)$?

Learning the bias of a coin, or a human

Example:
We're trying to estimate the likelihood that a human will click on an ad.
We model the human as coin with unknown bias $P \sim Uniform(0,1)$.
The coin has resulted in 10 heads out of the first 10 flips (call this event A).

Q: What is $\boldsymbol{E}[P \mid A]$?

The best estimator of P is the fraction of heads obtained so far!



But this seems shaky …

Learning a person's bias

Example:

We're trying to estimate the likelihood that a human will click on an ad.
We model the human as coin with unknown bias $P \sim Uniform(0,1)$.
The coin has resulted in 10 heads out of the first 10 flips (call this event A).

Q: What is $\boldsymbol{E}[P \mid A]$?

A:

$$\boldsymbol{E}[P \mid A] = \int_0^1 f_{P|A}(p) \cdot p \, dp$$

$$f_{P|A}(p) = \frac{\boldsymbol{P}\{A \mid P=p\} \cdot f_P(p)}{\boldsymbol{P}\{A\}} = \frac{p^{10} \cdot 1}{\boldsymbol{P}\{A\}}$$

So $f_{P|A}(p) = 11p^{10}$

$$\boldsymbol{P}\{A\} = \int_0^1 \boldsymbol{P}\{A \mid P = p\} \cdot f_P(p) dp = \int_0^1 p^{10} \cdot 1 dp = \frac{1}{11}$$

Learning a person's bias

Example:

We're trying to estimate the likelihood that a human will click on an ad.
We model the human as coin with unknown bias $P \sim Uniform(0,1)$.
The coin has resulted in 10 heads out of the first 10 flips (call this event A).

Q: What is $\boldsymbol{E}[P \mid A]$?

A: $$\boldsymbol{E}[P \mid A] = \int_0^1 f_{P|A}(p) \cdot p \, dp = \int_0^1 11p^{10} \cdot p dp = \frac{11}{12}$$

Not 1 but close. The answer depends on the initial assumption that $P \sim Uniform(0,1)$, which is referred to as **the prior** (see Chpt 17).

So $f_{P|A}(p) = 11p^{10}$