



## Lecture 9: Continuous Random Variables and Their Distributions

Dan Larremore ... again.

daniel.larremore @

codata@colorado.edu.

Tues 5 PM



Statisticians Fall asleep faster by taking a random sample of sheep.

## Announcements and reminders

---

- HW 2 due Friday at 5 PM
- Quizlet 04 due ~~Monday~~ at 10 AM  
Weds.
- Practicum #1 posted Fri.

## Previously, on CSCI 3022...

**Definition:** A discrete random variable (r.v.)  $X$  is a function that maps the elements of the sample space  $\Omega$  to a finite number of values  $a_1, a_2, \dots, a_n$  or an infinite number of values  $a_1, a_2, \dots$

**Definition:** A probability mass function (pmf) is the map between the random variable's values and the probabilities of those values.

$$f(a) = P(X=a)$$

**Definition:** A cumulative distribution function (cdf) is a function whose value at a point  $a$  is the cumulative sum of probability masses up until  $a$

$$F(a) = P(X \leq a) = \sum_{x \leq a} f(x)$$

## Continuous random variables

---

Many real-life random processes must be modeled by random variables that can take on continuous (i.e., not discrete) values. Some examples:

- Peoples' heights:  $X \in \underline{(0, \infty)}$
- Final grades in a class:  $X \in \underline{[0, 100]}$
- Time between people checking out in a line at the store:  $X \in \underline{(0, \infty)}$

Other examples?

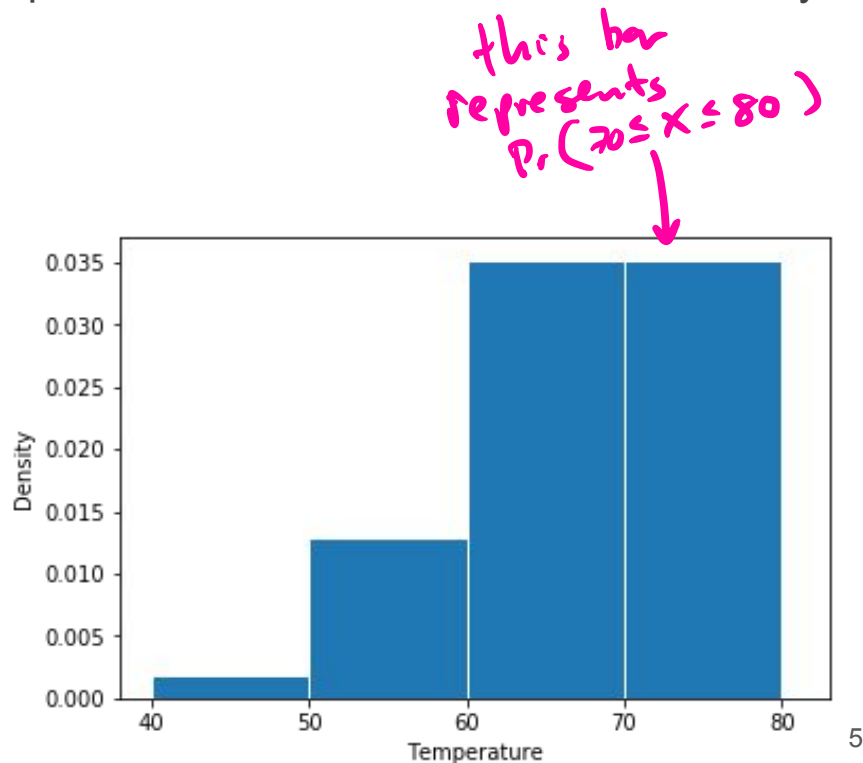
Age.  $(0, \infty)$

Temp (max daily)  $[0, \infty)$

## Continuous from discrete

**Example:** S'pose your friend asks you what the temperature will be like today. Specifically, they want to know what the probability that the temperature is between 70 and 80 °F, so they can decide whether or not to wear shorts.

How would you calculate your response?

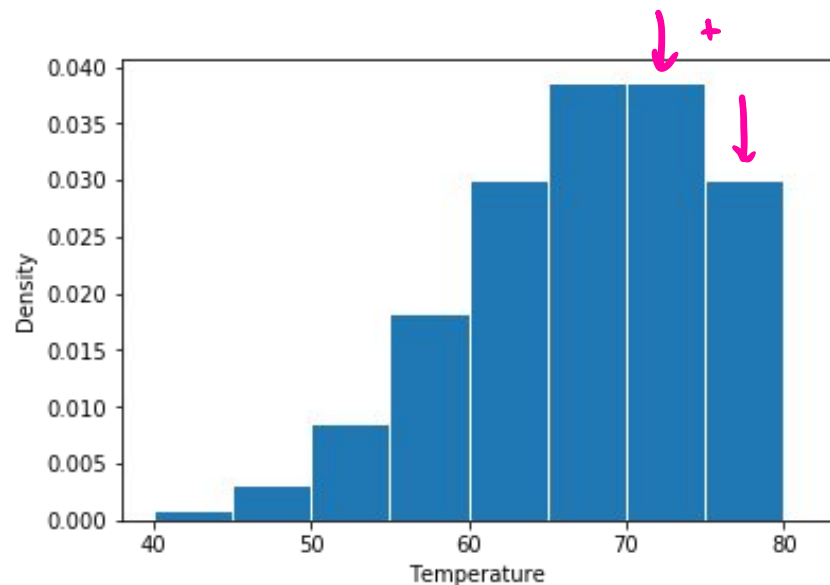


## Continuous from discrete

**Example:** S'pose your friend asks you what the temperature will be like today. Specifically, they want to know what the probability that the temperature is between 70 and 80 °F, so they can decide whether or not to wear shorts.

$$P_r(70 \leq X \leq 80)$$

How would you calculate your response?

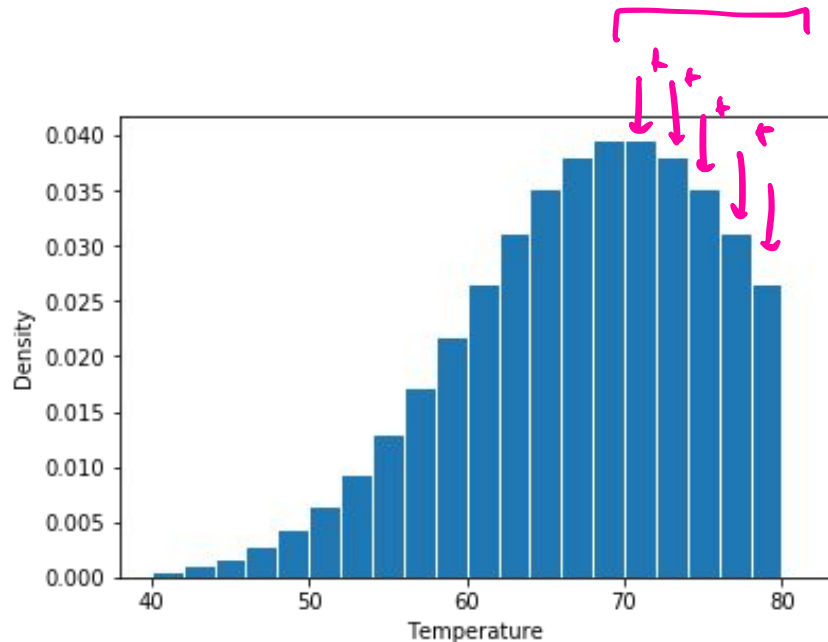


## Continuous from discrete

**Example:** S'pose your friend asks you what the temperature will be like today. Specifically, they want to know what the probability that the temperature is between 70 and 80 °F, so they can decide whether or not to wear shorts.

$$P_r(70 \leq x \leq 80)$$

How would you calculate your response?

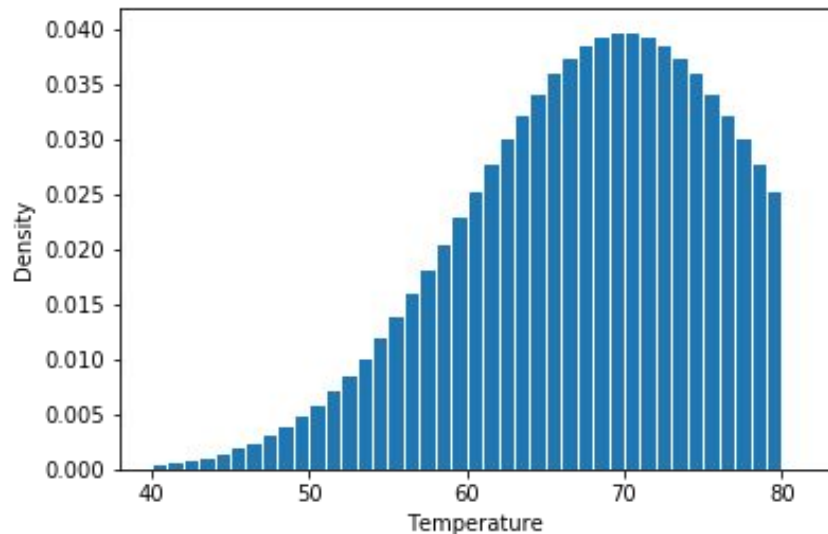


## Continuous from discrete

---

**Example:** S'pose your friend asks you what the temperature will be like today. Specifically, they want to know what the probability that the temperature is between 70 and 80 °F, so they can decide whether or not to wear shorts.

**How would you calculate your response?**





# Continuous from discrete

**Example:** S'pose your friend asks you what the temperature will be like today. Specifically, they want to know what the probability that the temperature is between 70 and 80 °F, so they can decide whether or not to wear shorts.

How would you calculate your response?

$$P(70 \leq X \leq 80) = \int_{70}^{80} f(x) dx$$

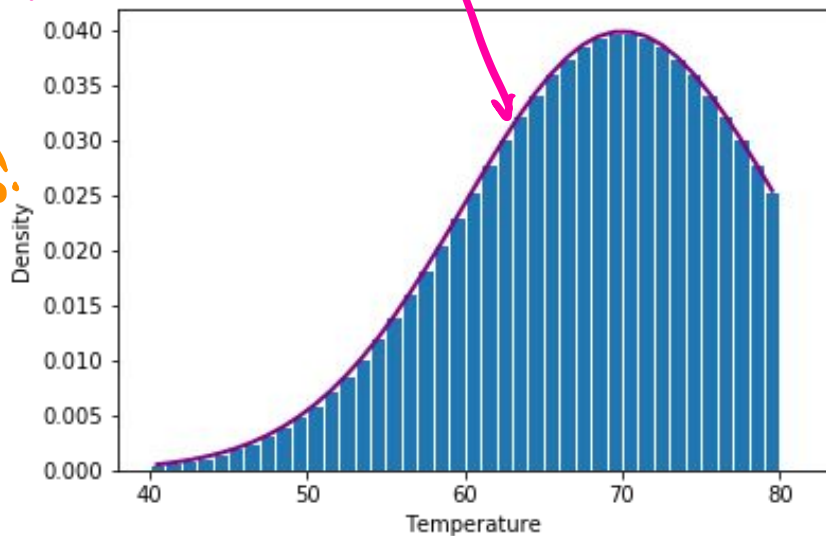
think  $\lim_{\Delta x \rightarrow 0} \sum_{x \text{ between } 70, 80} f(x) \Delta x$

height rect  $\uparrow$   $f(x)$   $\uparrow$  width  $\Delta x$

calculus!

Probability  
Density  
Function

cont. function  
 $f(x)$



# Continuous random variables

**Definition:** A random variable  $X$  is **continuous** if for some function  $f: \mathbb{R} \rightarrow \mathbb{R}$  and for any numbers  $a$  and  $b$  with  $a \leq b$ ,

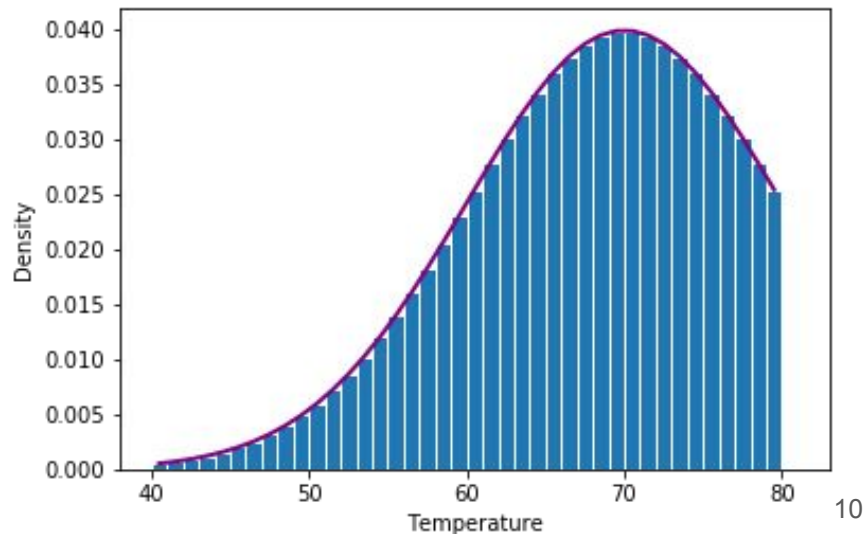
$$P(a \leq X \leq b) = \int_a^b f(x) dx$$

The function  $f$  must satisfy:

1)  $f(x) \geq 0$  for all  $x$ , and

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

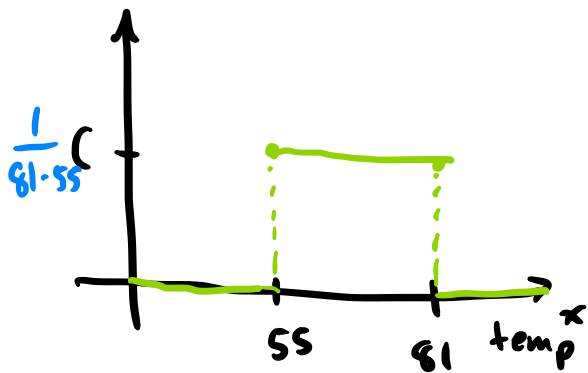
We call  $f$  the **probability density function** (pdf) of  $X$ .



## Continuous random variables

**Example:** S'pose you have some reason to believe that the temperature is equally likely to be anywhere between 55 and 81 °F. Find the probability density function (pdf) for X.

Equally Likely  $\rightarrow f(x) = \begin{cases} C & \text{when } x \text{ between } 55, 81 \\ 0 & \text{otherwise ( } x < 55 \text{ or } x > 81 \text{)} \end{cases}$



Find  $C$ , by using  $\int_{-\infty}^{\infty} f(x) dx = 1$

$$\int_{-\infty}^{\infty} f(x) dx = \int_{-\infty}^{55} f(x) dx + \int_{55}^{81} f(x) dx + \int_{81}^{\infty} f(x) dx = Cx \Big|_{x=55}^{x=81} = C(81-55) = 1$$

so  $C = \frac{1}{81-55}$



## Continuous random variables

---

**Example:** S'pose you have some reason to believe that the temperature is equally likely to be anywhere between 55 and 81 °F. Find the probability density function (pdf) for  $X$ .

$$\text{Equally likely} \Rightarrow f(x) = \begin{cases} C & 55 \leq x \leq 81 \\ 0 & \text{otherwise} \end{cases}$$

But the “sum to 1” condition means:  $1 \stackrel{\heartsuit}{=} \int_{-\infty}^{\infty} f(x) dx = \int_{55}^{81} C dx = (81 - 55)C = 26C$

$$\Rightarrow C = \frac{1}{26}$$

$$\text{and: } f(x) = \begin{cases} \frac{1}{26} & 55 \leq x \leq 81 \\ 0 & \text{otherwise} \end{cases}$$



## Continuous random variables

**Example:** S'pose you have some reason to believe that the temperature is equally likely to be anywhere between 55 and 81 °F. Find the probability density function (pdf) for  $X$ .

**Definition:** A continuous random variable has a uniform distribution on the interval  $[\alpha, \beta]$  if its probability density function  $f$  is given by  $f(x) = 0$  if  $x$  is not in  $[\alpha, \beta]$  and

$$f(x) = \frac{1}{\beta - \alpha} \quad \text{for } \alpha \leq x \leq \beta$$

We say  $X \sim U(\alpha, \beta)$

$$X \sim U(55, 81)$$

Mon



81° 55°

## Continuous random variables

---

**Example:** S'pose you have some reason to believe that the temperature is equally likely to be anywhere between 55 and 81 °F. Find the probability density function (pdf) for  $X$ .

**Question:** What distribution does temperature follow in the example?

$$X \sim U(55, 81)$$

**Follow-up question, getting to what we might actually care about:**

What is the probability that temperature is between 75 and 81 °F?

$$\int_{75}^{81} \frac{1}{81-55} dx = \frac{1}{81-55} x \Big|_{75}^{81} = \frac{81-75}{81-55} = \frac{6}{26} = \frac{3}{13} = \dots$$



## Continuous random variables

---

**Example:** S'pose you have some reason to believe that the temperature is equally likely to be anywhere between 55 and 81 °F. Find the probability density function (pdf) for  $X$ .


**Question:** What distribution does temperature follow in the example?

$$X \sim U(55, 81)$$

**Follow-up question, getting to what we might actually care about:**

What is the probability that temperature is between 75 and 81 °F?

$$P(75 \leq X) = P(75 \leq X \leq 81) \quad \text{b/c } U(55, 81)$$

$$\begin{aligned} &= \int_{75}^{81} \frac{1}{26} dx \\ &= \frac{81 - 75}{26} = \frac{6}{26} \approx 0.23 \end{aligned}$$




## Continuous random variables

Example, follow-up: What is the probability that it is exactly 75 °F?

Zero

$$\Pr(75 - \epsilon \leq X \leq 75 + \epsilon) = \int_{75 - \epsilon}^{75 + \epsilon} f(x) dx$$

$$\text{If } \epsilon \rightarrow 0, \text{ then } \int_{75}^{75} f(x) dx = 0$$





# Continuous random variables

What if we want to compute things like  $P(X \leq a)$ ?

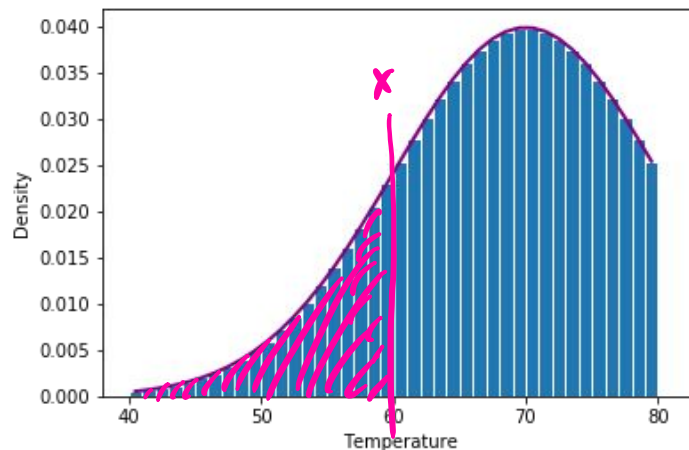
Is there an analog for the **cumulative distribution function** from the discrete case?

$$F(x) = P(X \leq x) = \sum_{k \leq x} f(k)$$

**Question:** What would the continuous analog of this sum be?

$$F(x) = \int_{-\infty}^x f(t) dt$$

↑ CDF                      ↑ PDF



# The cumulative distribution function

Note:

Discrete

Cont.

PMF

PDF

CDF

CDF

What if we want to compute things like  $P(X \leq a)$ ?

Is there an analog for the **cumulative distribution function** from the discrete case?

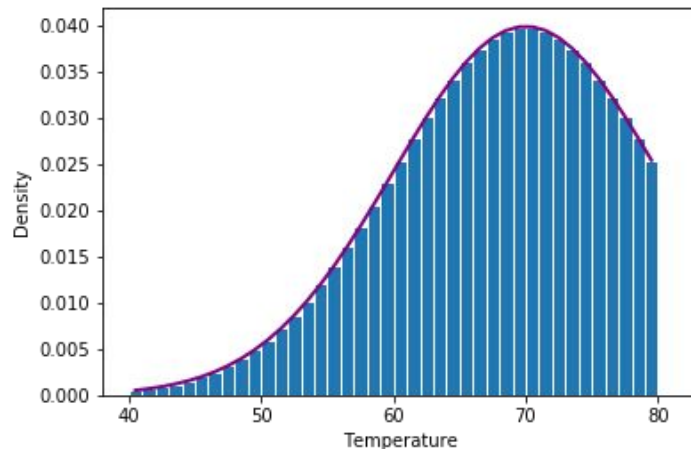
$$F(x) = P(X \leq x) = \sum_{k \leq x} f(k)$$

**Question:** What would the continuous analog of this sum be?

**Answer:** For continuous r.v., we **also** have a **cumulative distribution function:**

cdf

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



# The cumulative distribution function

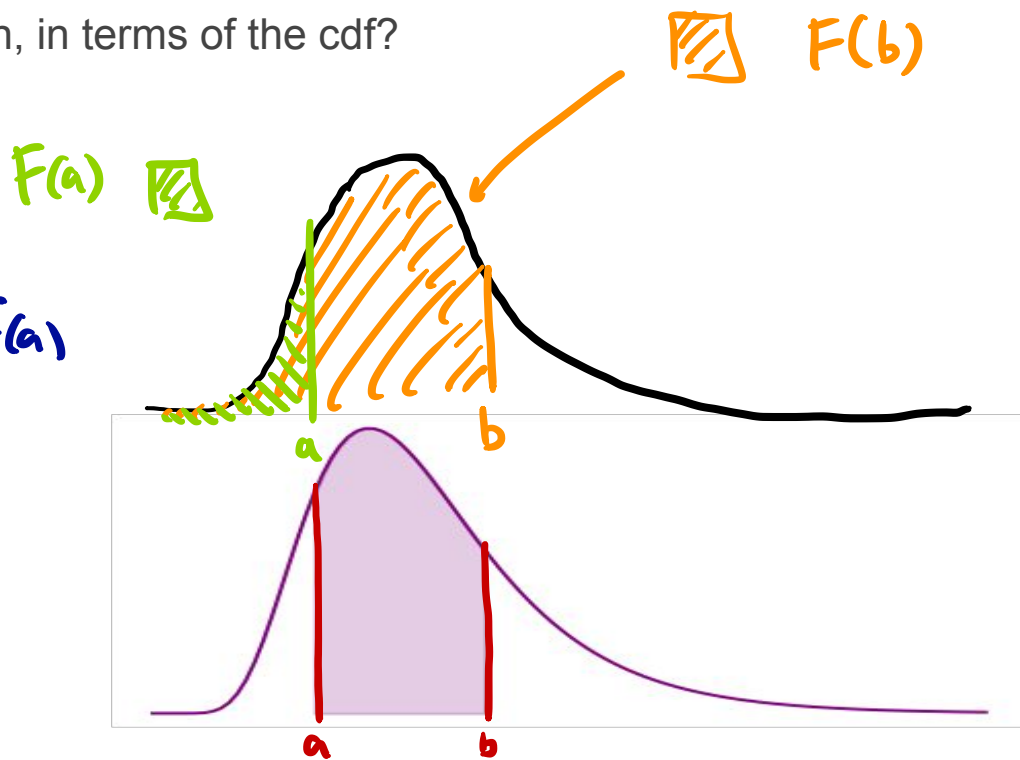
Can we use the cdf to compute things like  $P(a \leq X \leq b)$  ?

**Example:** What is the shaded region, in terms of the cdf?

$$\Pr(a \leq X \leq b) = F(b) - F(a)$$

↓

$$\int_a^b f(x) dx$$



# The cumulative distribution function

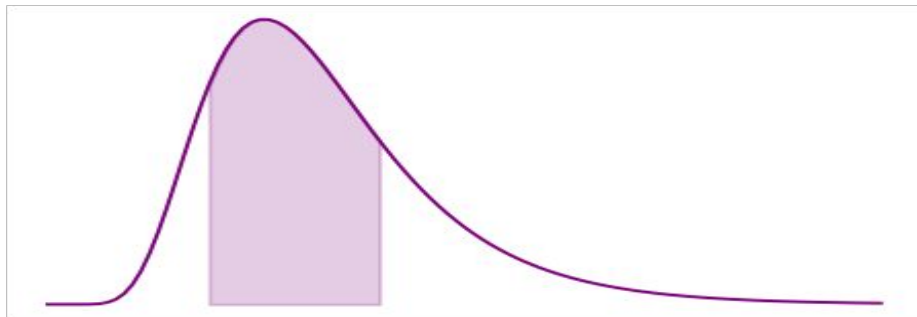
Can we use the cdf to compute things like  $P(a \leq X \leq b)$  ?

**Example:** What is the shaded region, in terms of the cdf?

More generally:  $P(a \leq X \leq b) = \int_a^b f(t) dt = F(b) - F(a)$

*Fundamental Theorem of Calculus!*

*F antiderivative of f*



# The cumulative distribution function

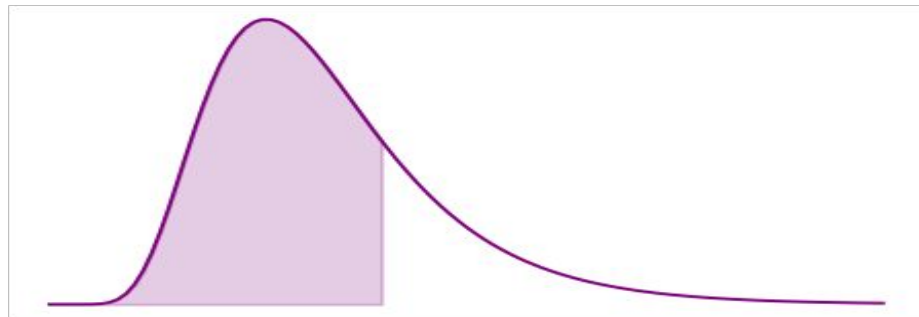
---

Can we use the cdf to compute things like  $P(a \leq X \leq b)$  ?

**Example:** What if we viewed the cdf as a function of  $x$ ?

See annotated slides later!

F.T.C.



# The cumulative distribution function

Can we use the cdf to compute things like  $P(a \leq X \leq b)$  ?

**Example:** What if we viewed the cdf as a function of  $x$ ?

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

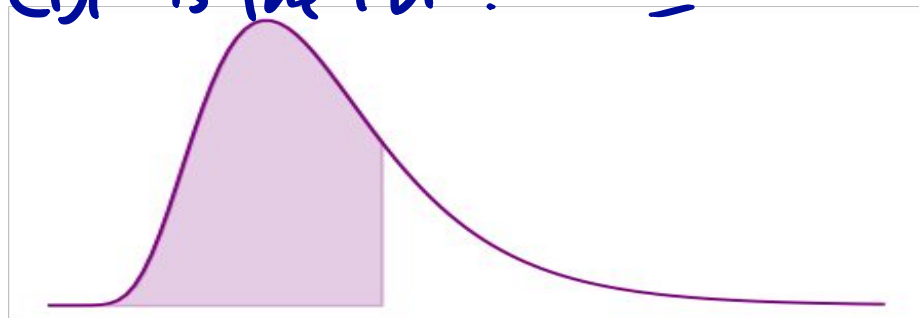
$$\frac{d}{dx}F(x) = \frac{d}{dx} \int_{-\infty}^x f(t) dt = f(x) - "f(-\infty)" = f(x)$$

$$\frac{d}{dx}F(x) = f(x)$$

Derivative of CDF is the PDF! ^\_^



**THIS** is a wildly important and useful relationship between  $F(x)$  and  $f(x)$



# The Normal distribution [List of Common Misconceptions] Gauss.

**Definition:** A continuous random variable  $X$  has a normal (or Gaussian) distribution with parameters  $\mu$  and  $\sigma^2$  if its probability density function is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

We say  $X \sim N(\mu, \sigma^2)$

Let's play around with this distribution: <https://academo.org/demos/gaussian-distribution/>

# The Exponential distribution

texts during class.

Sometimes it's easier to first find the **cdf** and then derive the **pdf** by taking a derivative.

**Example:** Recall that the Poisson distribution describes the number of arrivals (or hits), assuming some constant average rate of arrivals per time period.

A Poisson random variable is discrete because we're counting things  
(e.g., number of cars entering a parking garage)

Discrete Poisson = how many "events"

Cont Exponential = how long between "events"

But s'pose each arrival is someone finishing checking out in a grocery store line?

Waiting times = cont. r.v.  
Exponential Distr.





# The Exponential distribution

---

Sometimes it's easier to first find the **cdf** and then derive the **pdf** by taking a derivative.

**Example:** Recall that the Poisson distribution describes the number of arrivals (or hits), assuming some constant average rate of arrivals per time period.

A Poisson random variable is discrete because we're counting things (e.g., number of cars entering a parking garage)

But s'pose each arrival is someone finishing checking out in a grocery store line?

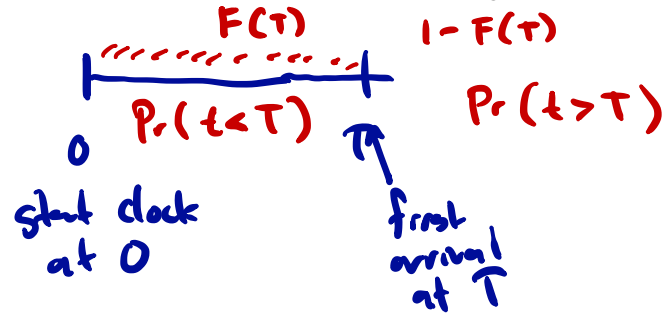
Now, we are interested in the **amount of time between each arrival**  
(so we can choose which line will be the shortest!)



# The Exponential distribution

S'pose the # arrivals follows a Poisson distribution (process) with rate  $\lambda$  arrivals/minute

- Start from  $t = 0$  and let  $T$  be the random variable describing the first arrival
- What is the probability that the first arrival does not occur in the first  $t$  minutes?



$$1 = \Pr(t < T) + \Pr(t > T)$$

$\Pr(T > t) \dots$  Poisson... take deriv.



# The Exponential distribution

S'pose the # arrivals follows a Poisson distribution (process) with rate  $\lambda$  arrivals/minute

- Start from  $t = 0$  and let  $T$  be the random variable describing the first arrival
- What is the probability that the first arrival does not occur in the first  $t$  minutes?

$$\begin{aligned}P(T > t) &= P(\text{no hits in } \leq t) \\&= \frac{(\lambda t)^0 e^{-\lambda t}}{0!} = e^{-\lambda t}\end{aligned}$$

$$\begin{aligned}P(T \leq t) &= 1 - P(T > t) = 1 - e^{-\lambda t} \\F(t) &= 1 - e^{-\lambda t}\end{aligned}$$

$$\begin{aligned}f(t) &= \frac{d}{dt}F(t) = \frac{d}{dt}(1 - e^{-\lambda t}) \\&= \begin{cases} \lambda e^{-\lambda t} & \text{for } t \geq 0 \\ 0 & \text{for } t < 0 \end{cases}\end{aligned}$$

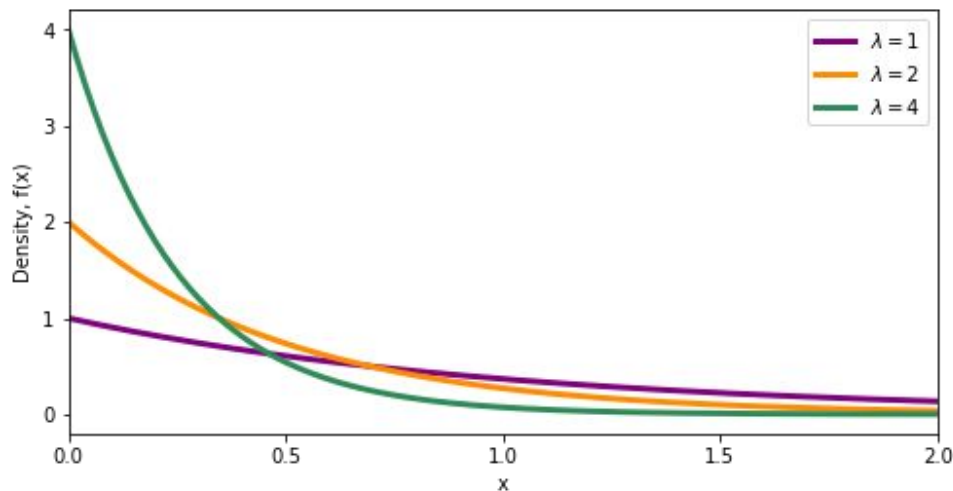


# The Exponential distribution

**Definition:** A continuous random variable  $X$  has an exponential distribution with **rate parameter**  $\lambda > 0$  if its probability density function is given by

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

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



# The Exponential distribution

---

Now hang on a second!

- That works for the **first** arrival. But what about the second? Or third?
- We wanted the distribution of time between all arrivals



# The Exponential distribution

---

Now hang on a second!

- That works for the **first** arrival. But what about the second? Or third?
- We wanted the distribution of time between all arrivals

## Fun fact:

It turns out the Exponential distribution has something called the **memoryless property**

**Theorem:** (memoryless property)

If  $T \sim \text{Exp}(\lambda)$ , then  $P(T > t + t_0 \mid T > t_0) = P(T > t)$

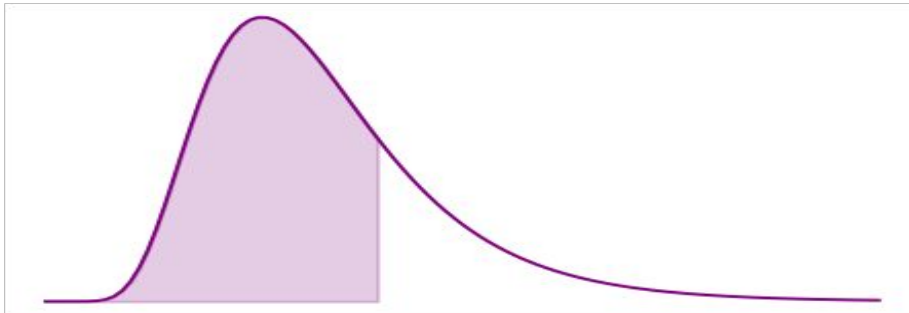


# Quartiles and Percentiles

---

Given everything we have learned since Week 1, consider the following...

**Question:** How can we compute an  $x$  such that, say,  $P(X \leq x)$  75% of the time?



# What just happened?

---

- We learned about **probability density functions** and **continuous density functions**!  
(pdfs) (cdfs)
  - We learned about some important **continuous** distributions!
    - **Normal** distribution: \_\_\_\_\_
    - **Exponential** distribution: \_\_\_\_\_
  - How are the exponential and Poisson distributions related?
- 

Next time: Great expectations!

