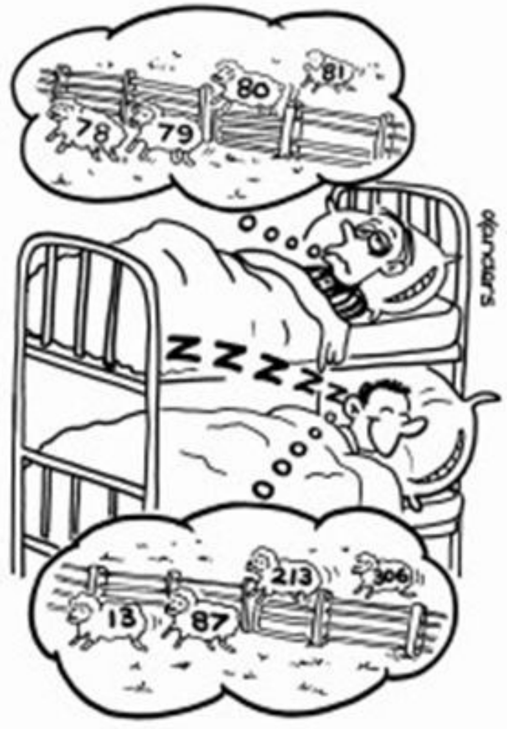
Lecture 9: Continuous Random Variables and Their Distributions

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~~ Weds. at 10 AM
- 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 Ω to a finite number of values $a_1, a_2, \ldots, a_n$ or an infinite number of values $a_1, a_2, \ldots$

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

Other examples?

Age. $(0, \infty)$

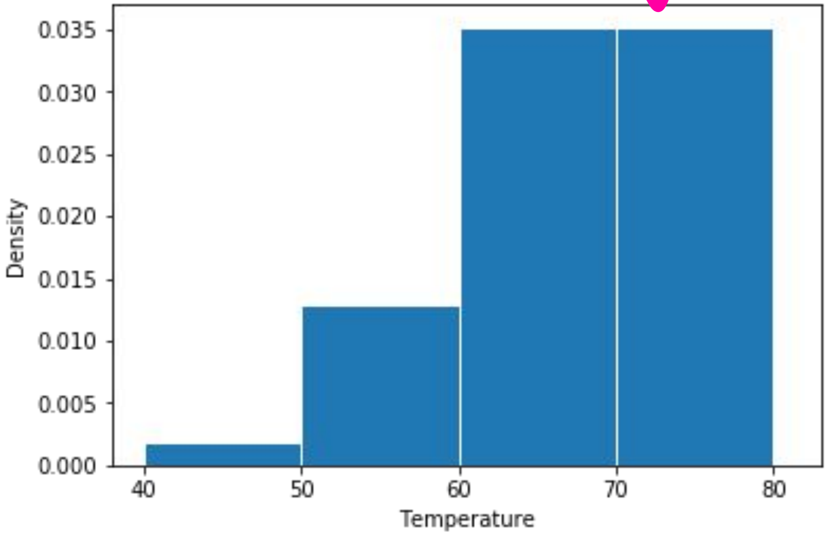
Temp (max daily) $[0, \infty)$

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 ^oF, so they can decide whether or not to wear shorts.

this bar represents $Pr(70 \le X \le 80)$

How would you calculate your response?



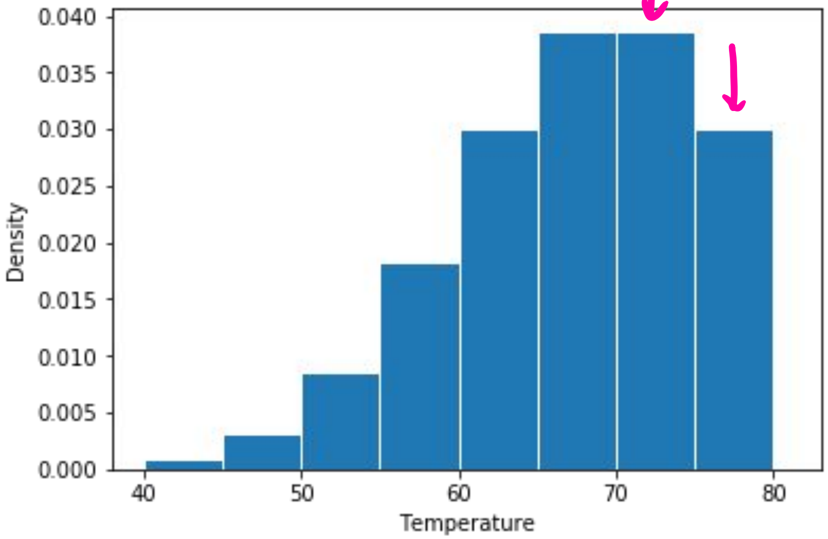
Continuous from discrete

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 ^oF, so they can decide whether or not to wear shorts.

$Pr(70 \le X \le 80)$

How would you calculate your response?

+

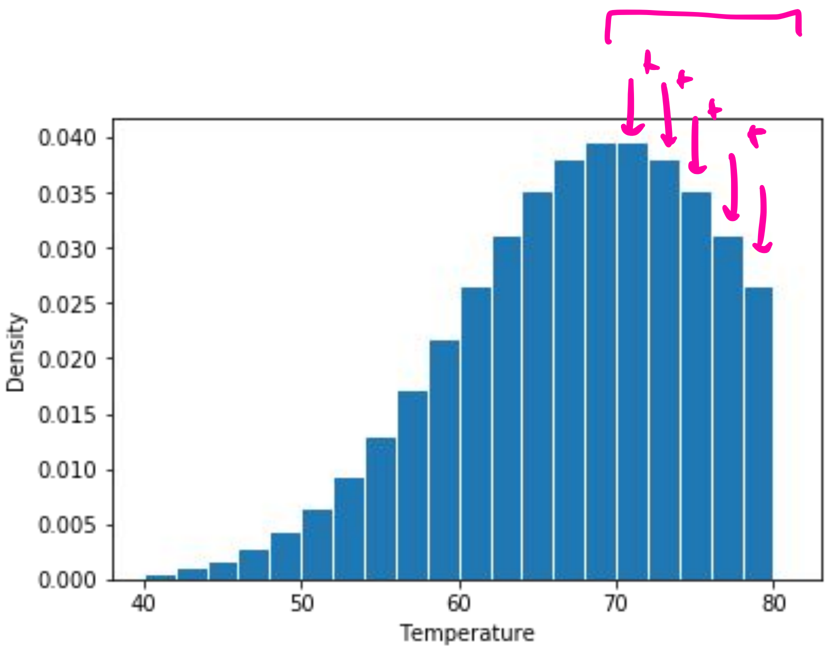


Continuous from discrete

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 ^oF, so they can decide whether or not to wear shorts.

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

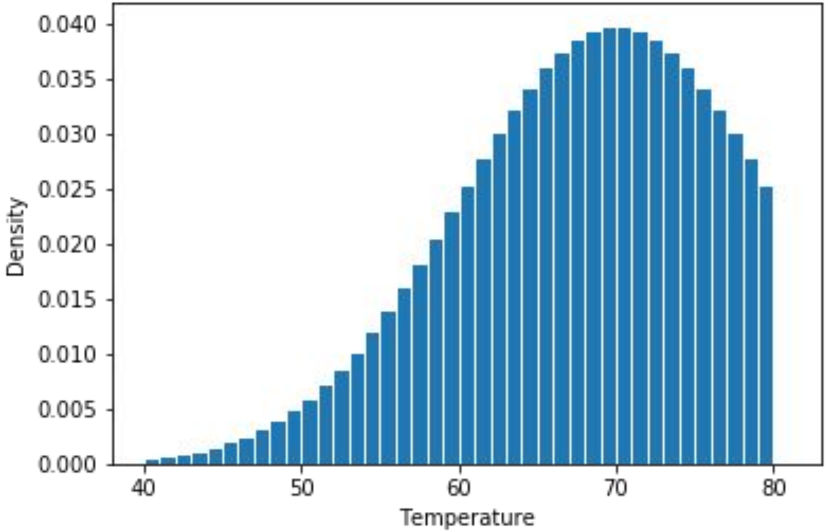
How would you calculate your response?



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 ^{o}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.

Probability **Density** Function

cont. function $f(x)$

How would you calculate your response?

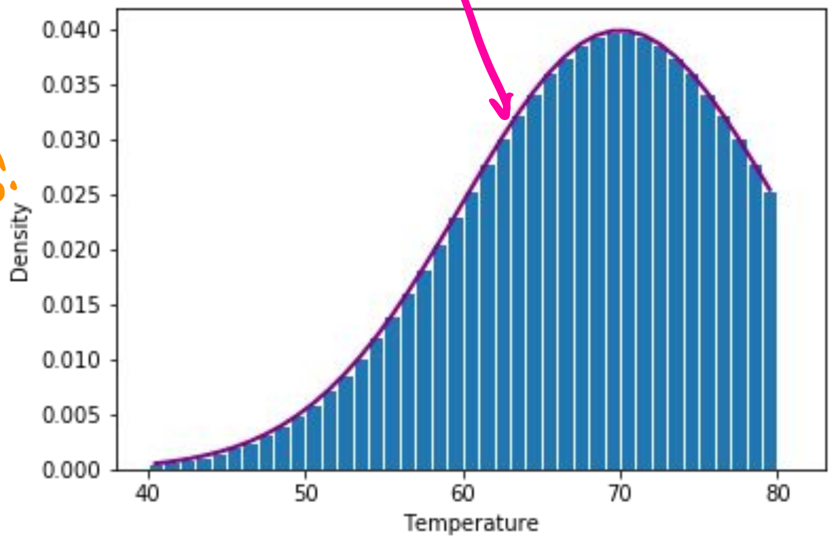
$$P(70 \le X \le 80) = \int_{70}^{80} f(x)\,dx$$

calculus!

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

height rect

width



Continuous random variables

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

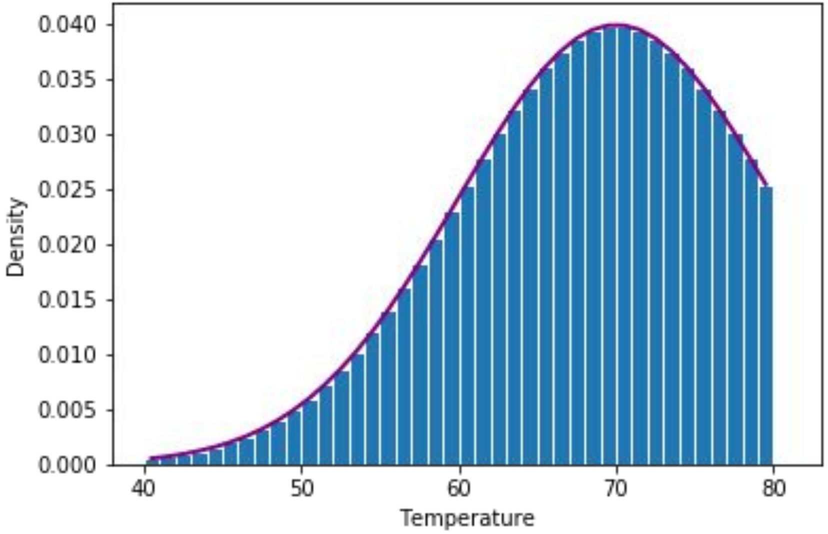
$$P(a \leq X \leq b) = \int_a^b f(x)\ dx$$

The function f must satisfy:

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

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

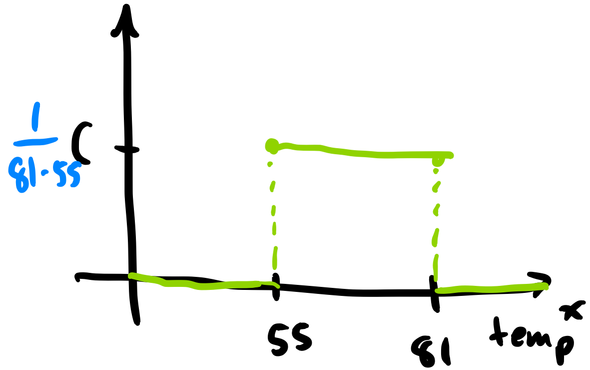
We call f the **probability density function** (pdf) of X.



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 ºF. Find the probability density function (pdf) for X.

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

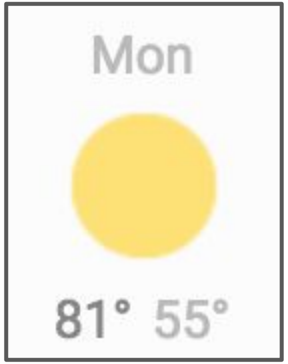


Find C, by using $\int_{-\infty}^{\infty} f(x)\,dx = 1$

set to

$$\int_{-\infty}^{\infty} f(x)\,dx = \underbrace{\cancel{\int_{-\infty}^{55} f(x)\,dx}}_{0} + \int_{55}^{81} \overset{C}{\cancel{f(x)}}\,dx + \underbrace{\cancel{\int_{81}^{\infty} f(x)\,dx}}_{0} = Cx\Big|_{x=55}^{x=81} = C(81-55) = 1$$

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



Continuous random variables

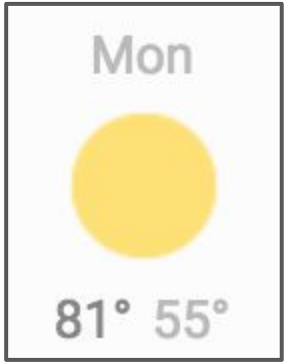
Example: S'pose you have some reason to believe that the temperature is equally likely to be anywhere between 55 and 81 ºF. Find the probability density function (pdf) for X.

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

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

$\Rightarrow \quad C = \dfrac{1}{26}$

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



Continuous random variables

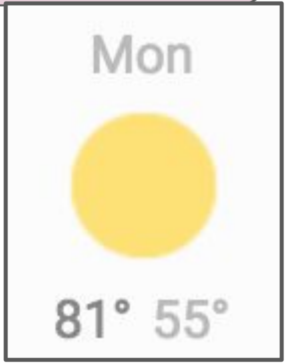
Example: S'pose you have some reason to believe that the temperature is equally likely to be anywhere between 55 and 81 ºF. Find the probability density function (pdf) for X.

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

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

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

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



Continuous random variables

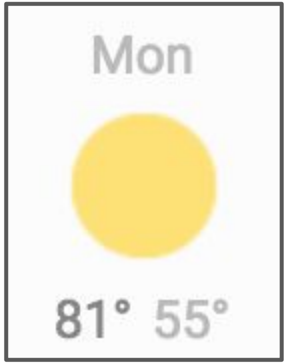
Example: S'pose you have some reason to believe that the temperature is equally likely to be anywhere between 55 and 81 ºF. Find the probability density function (pdf) for X.

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 ºF?

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



Continuous random variables

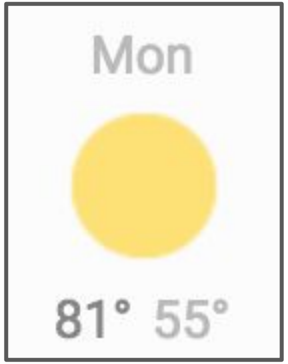
Example: S'pose you have some reason to believe that the temperature is equally likely to be anywhere between 55 and 81 ºF. Find the probability density function (pdf) for X.

Question: What distribution does temperature follow in the example?

X ~ 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 ºF?

$$P(75 \leq X) = P(75 \leq X \leq 81) \qquad \text{b/c } U(55, 81)$$
$$= \int_{75}^{81} \frac{1}{26}\, dx$$
$$= \frac{81 - 75}{26} = \frac{6}{26} \approx 0.23$$



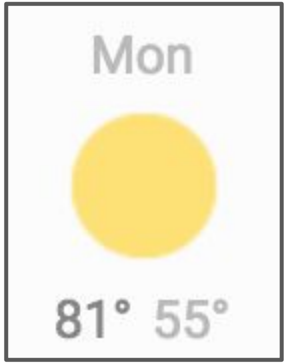
Continuous random variables

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

Zero

$$\Pr(75-\varepsilon \le X \le 75+\varepsilon) = \int_{75-\varepsilon}^{75+\varepsilon} f(x)\,dx$$

If $\varepsilon \to 0$, then $\int_{75}^{75} f(x)\,dx = 0$



Continuous random variables

What if we want to compute things like P(X ≤ *a*)?

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

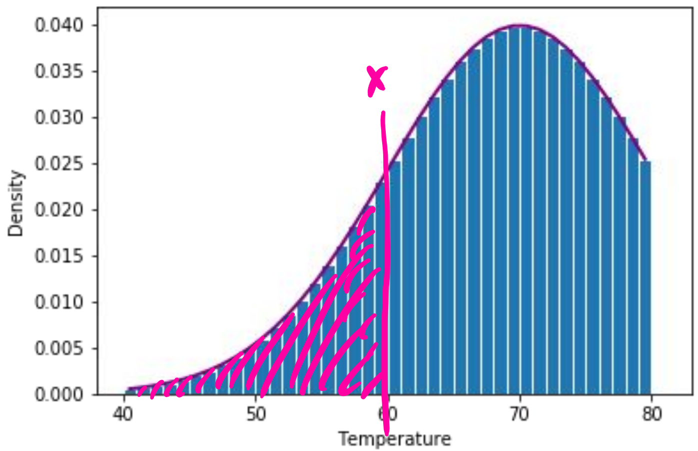
$$F(x) = P(X \leq x) = \sum_{k \leq x} f(k)$$

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 ≤ *a*)?

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

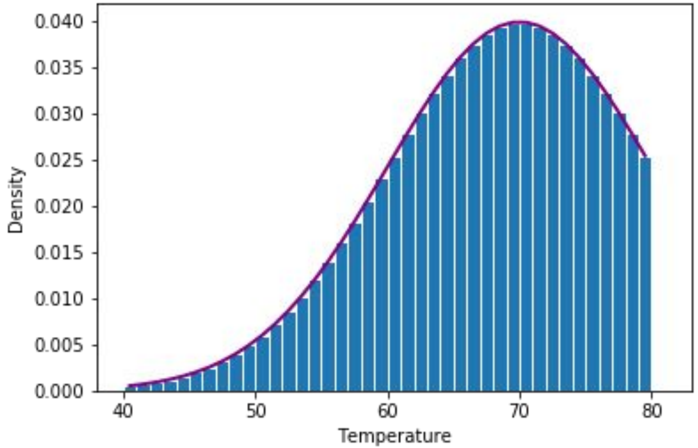
$$F(x) = P(X \leq x) = \sum_{k \leq x} f(k)$$

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

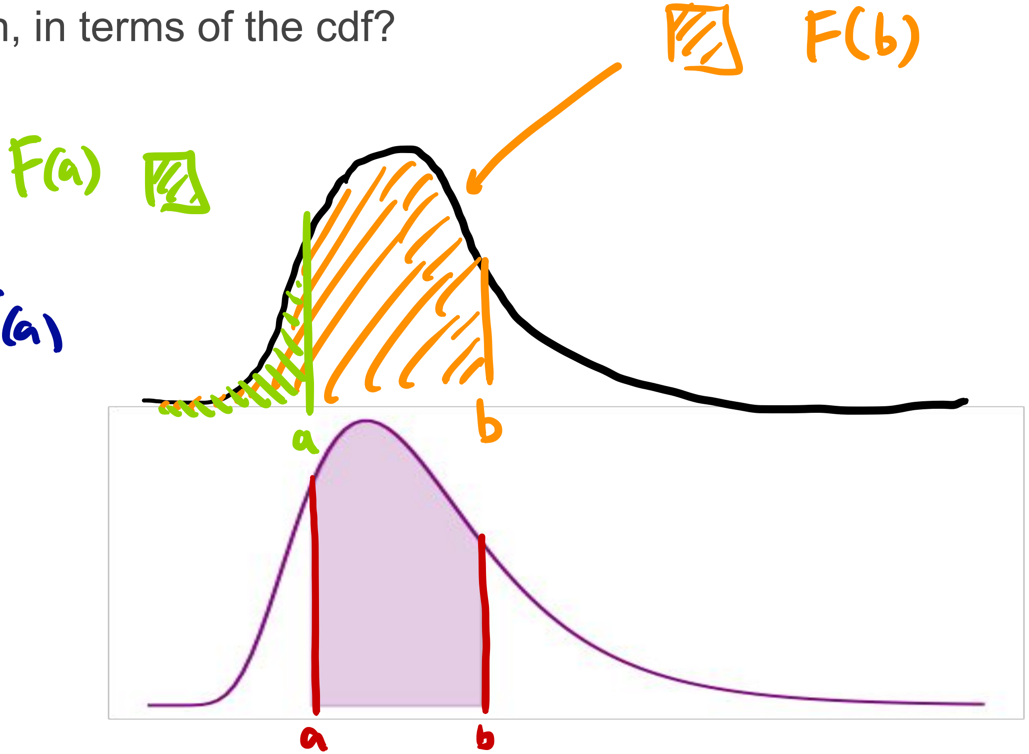
Can we use the cdf to compute things like $P(a \le X \le b)$?

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

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

$$\downarrow$$

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



The cumulative distribution function

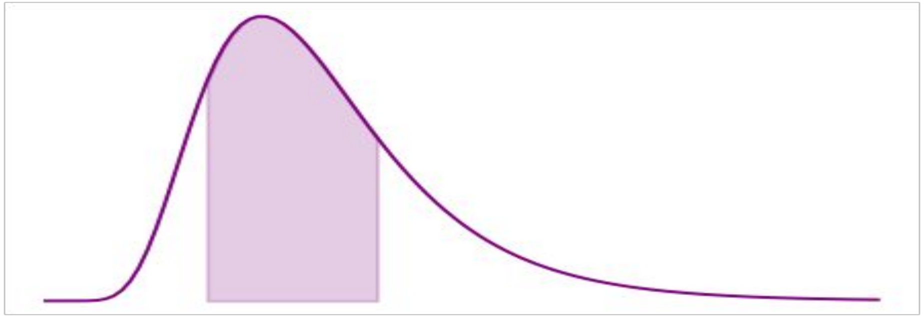
Can we use the cdf to compute things like $P(a \leq X \leq b)$?

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

Fundamental Theorem of Calculus!

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

F antiderivative of f



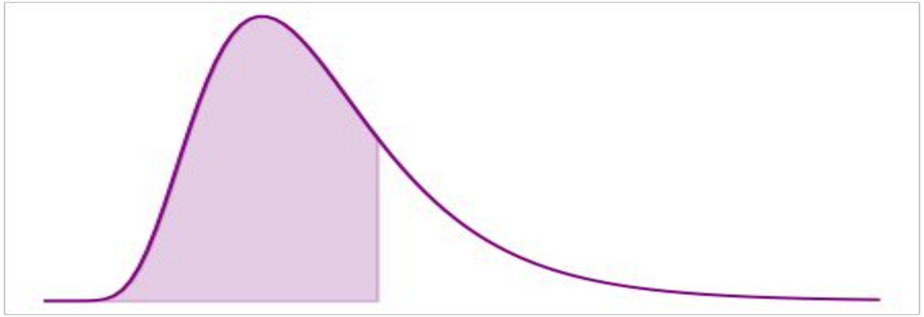
The cumulative distribution function

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 $P(a \le X \le b)$?

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

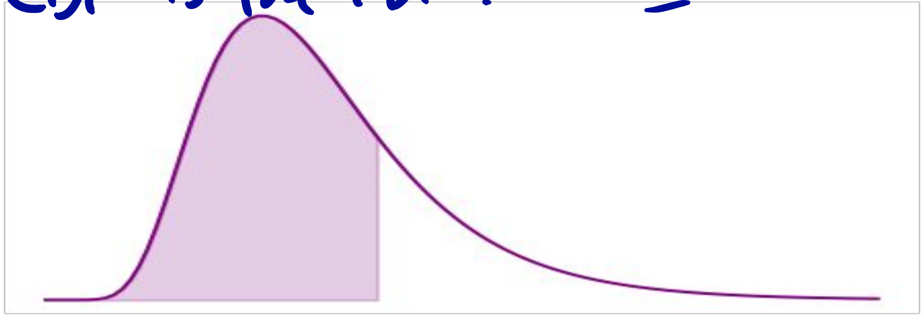
$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\ dt$$

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

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

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 μ and σ^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.

Discrete Poisson = how many "events"

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

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 λ 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?

$F(T)$ $\quad 1 - F(T)$

$Pr(t<T)$ $\quad Pr(t>T)$

0

start clock at 0

first arrival at T

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

$Pr(T>t)$ … Poisson … take deriv.

The Exponential distribution

S'pose the # arrivals follows a Poisson distribution (process) with rate λ 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?

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

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

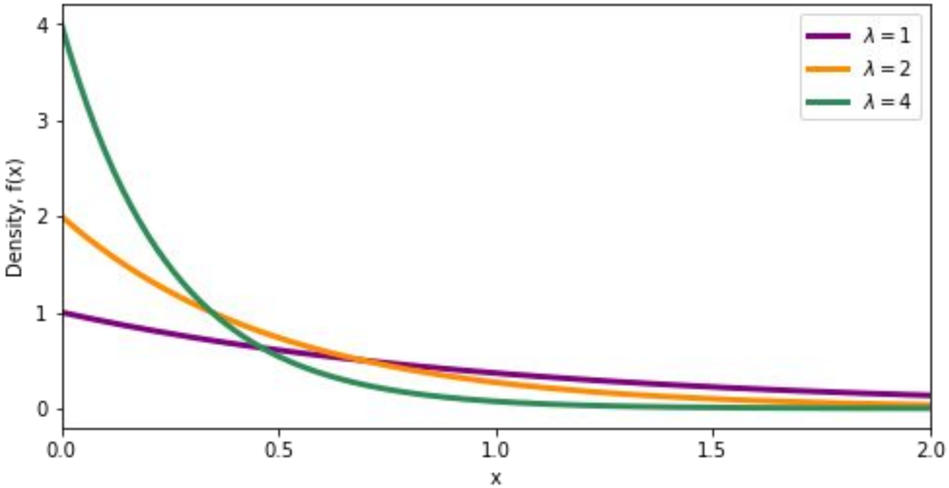
$$f(t) = \frac{d}{dt}F(t) = \frac{d}{dt}\left(1 - e^{-\lambda t}\right)$$
$$= \begin{cases} \lambda e^{-\lambda t} & \text{for } t \geq 0 \\ 0 & \text{for } t < 0 \end{cases}$$

The Exponential distribution

Definition: A continuous random variable X has an **exponential distribution** with **rate parameter** λ > 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 $X \sim Exp(\lambda)$



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**

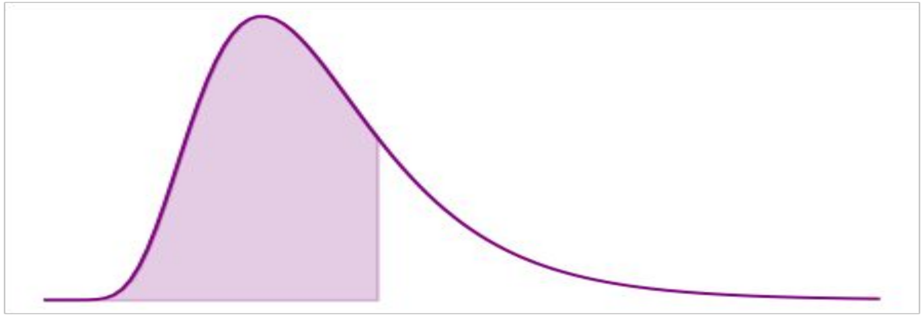
Theorem: (memoryless property)

If $T \sim Exp(\lambda)$, 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** (pdfs) and **continuous density functions** (cdfs)!

- We learned about some important **continuous** distributions!
  - **Normal** distribution: ______________________________
  - **Exponential** distribution: _________________________

- How are the exponential and Poisson distributions related?

  ____

Next time: Great **expectations**!