# CSCI 3022-002 Intro to Data Science
# Continuous Rvs

**Example:**

A factory makes parts for a medical device company. 6% of those parts are defective. For each one of the problems below:

(i.) Define an appropriate random variable for the experiment.

(ii.) Give the values that the random variable can take on.

(ii.) Find the probability that the random variable equals 2.

(iv.) State any assumptions you need to make.

Problems:

1. Out of 10 parts, X are defective.
2. Upon observing an assembly line, X non-defective parts are observed before finding a defective part.
3. X is the number of defective parts made per day, where the rate of defective parts per day is 10.

# Announcements and Reminders

- Homework ~~5~~ 4 posted! Shorter one!

# Last Time...: the blocks of discrete probability

1. Bernoulli: *one* binary outcome experiment.
2. Binomial: binary outcome experiment success *count* in $n$ tries. → fixing
3. Geometric: Total trials *until a success* of a binary outcome experiment.
4. Negative Binomial: Trials until $\overset{\text{fixed}}{r}$ binary outcome experiment *successes*.
5. Poisson: *counting* outcomes with a fixed rate $\lambda$.

# Last Time...: the blocks of discrete probability

1. Bernoulli: *one* binary outcome experiment.
   $f(x) = p^x(1-p)^{1-x}$
2. Binomial: binary outcome experiment success *count* in $n$ tries.
   $f(x) = \binom{n}{x}p^x(1-p)^{(n-x)}$
3. Geometric: Total trials *until a success* of a binary outcome experiment.
   $f(x) = (1-p)^{x-1}p$
4. Negative Binomial: Trials until *r* binary outcome experiment *successes*.
   $f(x) = \binom{x-1}{r-1}p^r(1-p)^{(x-r)}$
5. Poisson: *counting* outcomes with a fixed rate $\lambda$.
   $f(x) = \frac{e^{-\lambda}\lambda^x}{x!}$

# Last Time...: the blocks of discrete probability

The underlying pieces of discrete RVs:

1. The random variable $X$ takes inputs/events in the (discrete) sample space $\Omega$ and maps them to a (discrete) finite or infinite set of probability values $a_1, a_2, a_3, \dots$.
2. We find probabilities in the probability mass function or probability density function
$$f(x) = P(X = x).$$
3. We can find cumulative probabilities or probability on ranges of outcomes in the cumulative density function
$$F(x) = P(X \leq x) = \sum_{X \leq x} f(x).$$

# Poisson and... binomial?

One way to generate the Poisson is to take limits of a binomial: suppose you get texts during class (☹) at a rate of 29 texts per hour. What is the probability that you get 29 texts in an hour? 12 texts in an hour? 107 texts in an hour?

$\lambda$ is the *rate* of the Poisson.

$$P(X=29) \text{ or } P(X=12) \text{ or } P(X=107)$$

$$\text{if } X \sim \text{Pois}(\text{rate } \lambda = \underline{29})$$

# Poisson and... binomial?

One way to generate the Poisson is to take limits of a binomial: suppose you get texts during class $(\ddot\frown)$ at a rate of 29 texts per hour. What is the probability that you get 29 texts in an hour? 12 texts in an hour? 107 texts in an hour?

$\lambda$ is the *rate* of the Poisson.

Think about a Bernoulli that represents your friends asking ”should I text...?” then flipping a coin with probability $p$. Then:

# Poisson and... binomial?

One way to generate the Poisson is to take limits of a binomial: suppose you get texts during class (☹) at a rate of 29 texts per hour. What is the probability that you get 29 texts in an hour? 12 texts in an hour? 107 texts in an hour?

$\lambda$ is the *rate* of the Poisson.

Think about a Bernoulli that represents your friends asking "should I text...?" then flipping a coin with probability $p$. Then:

$\lambda = \frac{texts}{hour} \approx \frac{flips}{hour} \cdot \frac{texts}{flip} = np$ for the same $n$ and $p$ as a *binomial*.

# Poisson and... binomial?

One way to generate the Poisson is to take limits of a binomial: suppose you get texts during class $(\ddot\frown)$ at a rate of 29 texts per hour. What is the probability that you get 29 texts in an hour? 12 texts in an hour? 107 texts in an hour?

$\lambda$ is the *rate* of the Poisson.

Think about a Bernoulli that represents your friends asking ”should I text...?” then flipping a coin with probability $p$. Then:

$\lambda = \frac{texts}{hour} \approx \frac{flips}{hour} \cdot \frac{texts}{flip} = np$ for the same $n$ and $p$ as a *binomial*.

...but $n$ might vary a bit from hour to hour, so these are only equivalent *in the limit* ($n$ large, $p$ small)!

# Discrete Distributions Example

**Example:**
A factory makes parts for a medical device company. 6% of those parts are defective. For each one of the problems below:

(i.) Define an appropriate random variable for the experiment.

(ii.) Give the values that the random variable can take on.

(iii.) Find the probability that the random variable equals 2.

(iv.) State any assumptions you need to make.

Problems:

1. Out of 10 parts, X are defective.
2. Upon observing an assembly line, X non-defective parts are observed before finding a defective part.
3. X is the number of defective parts made per day, where the rate of defective parts per day is 10.

# Discrete Distributions Example

6% of those parts are defective. trial is this part defective

1. Out of 10 parts, X are defective.

fixed # of trials

(i.) r.v.: Binomial, $n=10$, $p=.06$

(ii.) Values of r.v.: $X: \{0,1,2,\cdots,10\}$

(iii.) $P(X=2)$: $P(X=2)=f(2)=\binom{10}{2}\cdot(.06)^2\,(.94)^8$

$$\binom{n}{x}\,p^x\,(1-p)^{n-x}.$$

(iv.) Assumptions:

iid: trials (parts) are independent & identically distributed

# Discrete Distributions Example

6% of those parts are defective.

1. Out of 10 parts, X are defective.

(i.) r.v.: $X \sim \underline{bin(10, .06)}$

(ii.) Values of r.v.: $X \in \{0, 1, 2, \ldots, 10\}$

(iii.) $P(X = 2)$: $\binom{10}{2} .06^2 .94^8$

(iv.) Assumptions: Parts are $\underline{\textit{i.i.d.}}$

# Discrete Distributions Example

6% of those parts are defective.

2. Upon observing an assembly line, X non-defective parts are observed before finding a defective part.

(i.) r.v.: either geom, $p=.06$ OR NB, $r=1$ $p=.06$

(ii.) Values of r.v.: $X=\{0, 1, 2, \ldots$

(iii.) $P(X=2)$: $P(N\,N\,D) = (.94)^2(.06)^1$

(iv.) Assumptions:

i.i.d.

# Discrete Distributions Example

non-
→ X: # of defects).

6% of those parts are defective.

2. Upon observing an assembly line, X non-defective parts are observed before finding a defective part.

↙ Geom that can be 0

(i.) r.v.: $(X + 1) \sim Geom(.06)$

↪ # of parts observed

(ii.) Values of r.v.: $X \in \{0, 1, 2, \ldots, \infty\}$

(iii.) $P(X = 2)$: $.94^2 .06^1$ $P(N) \cdot P(N) \cdot P(D)$

(iv.) Assumptions: Parts are *i.i.d.*

↗

N.D:

"before finding 2 defects"

# Discrete Distributions Example

6% of those parts are defective.

3. X is the number of defective parts made per day, where the rate of defective parts per day is 10.

count!

(i.) r.v.: $\text{Pois}\left(\text{rate} = \lambda = 10 \,\frac{\text{parts}}{\text{day}}\right) = \text{Pois}(10)$

(ii.) Values of r.v.: $X = \{0, 1, 2, 3, \ldots\ldots\}$

(iii.) $P(X = 2)$: $f(2) = e^{-\lambda}\dfrac{\lambda^2}{2!} = \#) = e^{-10}\dfrac{10^2}{2!} = \dfrac{1}{e^{10}} \cdot 50$

$\approx \dfrac{50}{7^5} \approx \dfrac{7^2}{7^5} \approx \dfrac{1}{7^3}$

(iv.) Assumptions:

# Discrete Distributions Example

6% of those parts are defective.

3. X is the number of defective parts made per day, where the rate of defective parts per day is 10.

(i.) r.v.: $X \sim Pois(10)$

(ii.) Values of r.v.: $X \in \{0, 1, 2, \ldots, \infty\}$

(iii.) $P(X = 2)$: $\frac{e^{-10} \cdot 10^2}{2!}$

(iv.) Assumptions: Parts are... *Poisson*?

# Continuous RVs

Many real-life random processes must be modeled by random variables that can take on continuous (non-discrete) values. Some example:

1. Peoples' heights: $X \in [3', \cancel{7}.5']$ $[0, 7.5')$
2. Final grades in a class: $X \in [0, 100]$ $+ [0, 105]$
3. Time between people checking out at a store : $t \in [0, 10\,m].$

# Continuous RVs

Many real-life random processes must be modeled by random variables that can take on continuous (non-discrete) values. Some example:

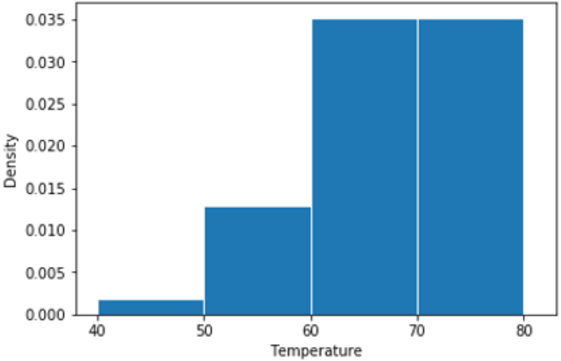
1. Peoples' heights: $X \in \{[0, 7.5ft]\}$
2. Final grades in a class: $X \in \{[0, 100]\}$
3. Time between people checking out at a store : $t \in \{[0, \infty]\}$

Chipotle on 29th St.

# More precision!

Suppose your friend asks you what the high temperature will be today. They want to know the probability it will be between 70F and 80F, so they can wear shorts.

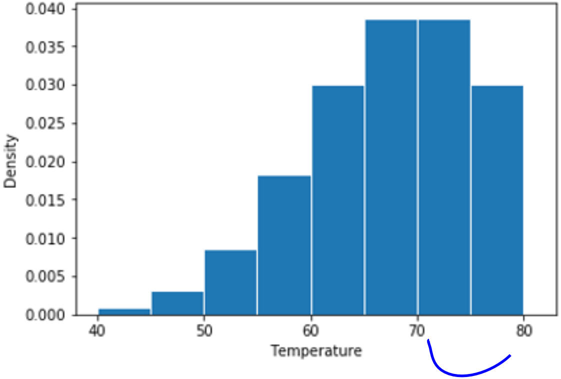
Probability:
Add up the share of outcomes between 70F and 80F!



# More precision!

Suppose your friend asks you what the high temperature will be today. They want to know the probability it will be between 70F and 80F, so they can wear shorts.

Probability:
Add up the share of outcomes between 70F and 80F!

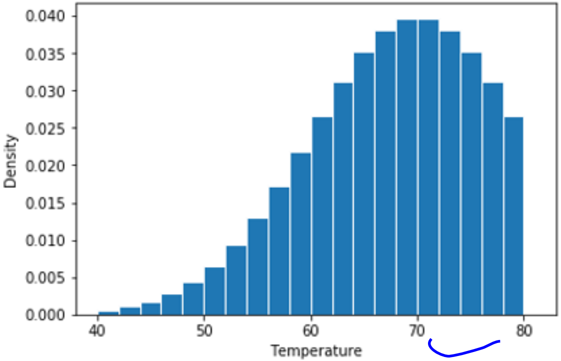


2 bins
smaller

# More precision!

Suppose your friend asks you what the high temperature will be today. They want to know the probability it will be between 70F and 80F, so they can wear shorts.

Probability:
Add up the share of outcomes between 70F and 80F!

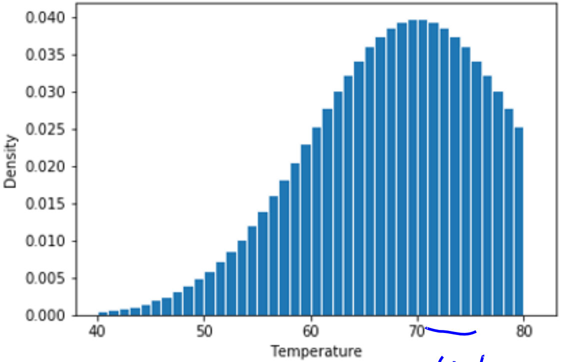


5 bins,
2° each

# More precision!

Suppose your friend asks you what the high temperature will be today. They want to know the probability it will be between 70F and 80F, so they can wear shorts.

Probability:
Add up the share of outcomes between 70F and 80F!



10 bins

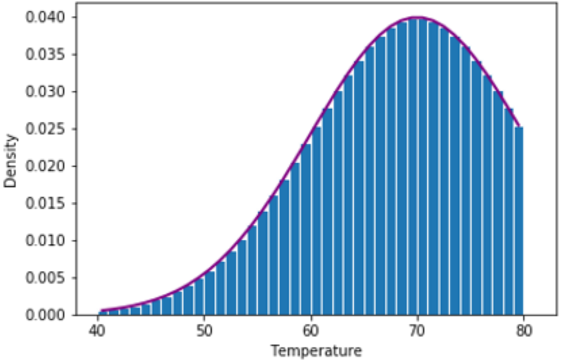
1° ea

# More precision!

Suppose your friend asks you what the high temperature will be today. They want to know the probability it will be between 70F and 80F, so they can wear shorts.

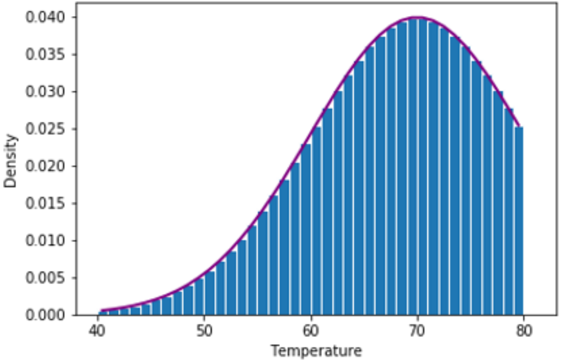
Probability:
Add up the share of outcomes between 70F and 80F!

Integrate!



# More precision!

Suppose your friend asks you what the high temperature will be today. They want to know the probability it will be between 70F and 80F, so they can wear shorts.



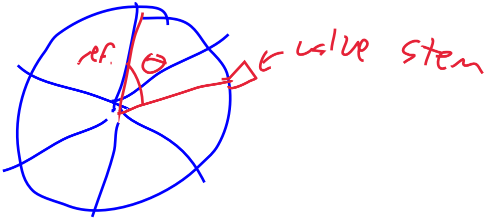
Probability:

*Integrate* up the share of outcomes between 70F and 80F!

# Continuous Distributions

**Example:**

Consider the reference line connecting the valve stem on a tire to the center point.



Let X be the angle measured clockwise to the location of an imperfection. The pdf for X is:

↓

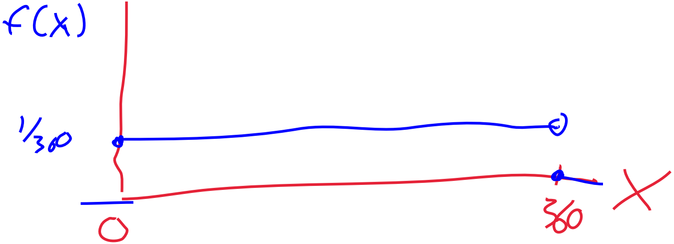
$$f(x) = \begin{cases} \frac{1}{360} & 0 \leq X < 360 \\ 0 & else \end{cases}$$

# Continuous Distributions

**Example, cont'd:**

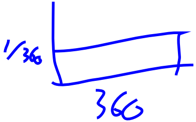
$$f(x) = \begin{cases} \frac{1}{360} & 0 \le X < 360 \\ 0 & else \end{cases}$$

Graphically, the pdf of $X$ is:



## Continuous Distributions

$f(x) = 1/360 \qquad 0 \le x < 360$



**Example, cont'd:** How can we show that:

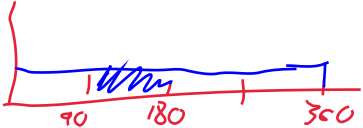
1. the total area of the pdf of $x$ is 1?

   width = 360°

   height : $1/360 = f(x)$

   $A = 1$ ☺

2. How do we calculate $P(90 \le X \le 180)$?

   $$= \int_{90}^{180} f(x)\,dx = \int_{90}^{180} \frac{1}{360}\,dx$$

   $$= \frac{x}{360}\Big|_{90}^{180} = 1/4$$

3. What is the probability that the angle of occurrence is within 90 of the reference line? (The reference line is at 0 degrees.)

   either: $X \in [0, 90]$ OR $X \in [270, 360]$

# Continuous Distributions

**Example, cont'd:** How can we show that:

1. the total area of the pdf of $x$ is 1?

$$\int_0^{360} f(x)\,dx = 1?$$

2. How do we calculate $P(90 \leq X \leq 180)$?

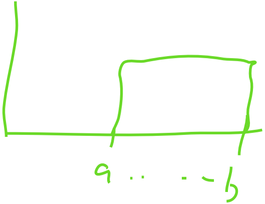
$$\int_{90}^{180} f(x)\,dx = \ldots?$$

3. What is the probability that the angle of occurrence is within 90 of the reference line? (The reference line is at 0 degrees.)

1/4 + 1/4 = 1/2

$$P(X < 90 \textbf{ OR } X > 270) = \int_0^{90} f(x)\,dx + \int_{270}^{360} f(x)\,dx = \ldots?$$

# Uniform Distribution



The previous problem was an example of the uniform distribution.

**Definition:** *Uniform Distribution*

A continuous rv X is said to have a *uniform distribution* on the interval $[a, b]$ if the pdf of $X$ is:

$f(x) =$ all equally likely . . . ↗ # Area=1 in the interval

$$= \frac{1}{b-a}$$

$$\underbrace{a \leq X \leq b}_{\text{width } b-a}$$

NOTATION: We write $X \sim U(a,b)$ to indicate that X is a uniform rv with lower bound $a$ and upper bound $b$.

# Uniform Distribution

The previous problem was an example of the uniform distribution.

**Definition:** *Uniform Distribution*

A continuous rv X is said to have a *uniform distribution* on the interval $[a, b]$ if the pdf of $X$ is:

$$f(x) = \frac{1}{b-a}; \qquad x \in [a, b]$$

0 else

NOTATION: We write $\underline{X \sim U(a, b)}$ to indicate that X is a uniform rv with lower bound $a$ and upper bound $b$.

# Exponential Distribution

The family of exponential distributions provides probability models that are very widely used in engineering and science disciplines to describe time-to-event data.

It can be thought of as a continuous analogue to the Poisson distribution, but instead of events-per-time, it measure time-per-events.

↳ measures $\frac{\text{events}}{\text{time}}$ "$\lambda$"

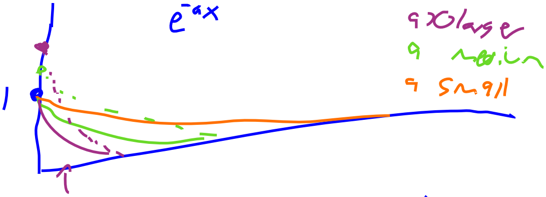
**Examples**:

: time-until-burrito

: time-until-buff bus

: time-until-it-rains

# Exponential Distribution



**Definition:** *Exponential Distribution*
A continuous rv X is said to have an *exponential distribution* with rate parameter $\lambda$ if the pdf of $X$ is:

$$f(x) = \underbrace{\lambda}_{\text{make area} = 1} e^{-\lambda x}$$

Units: $\dfrac{\text{events}}{\text{time}}$

NOTATION: We write $\underline{\exp(\lambda)}$ to indicate that X is an exponential rv with rate $\lambda$.

# Exponential Distribution

**Definition:** *Exponential Distribution*
A continuous rv X is said to have an *exponential distribution* with rate parameter $\lambda$ if the pdf of $X$ is:

$$f(x) = \lambda e^{-\lambda x}; \quad x \geq 0$$

$\underline{\lambda > 0}$

NOTATION: We write $\underline{X \sim exp(\lambda)}$ to indicate that X is an exponential rv with rate $\lambda$.

# Exponential Distribution

$\lambda = \frac{1 \text{ burn out}}{1000 \text{ hours}}$

**Example:**

Suppose a light bulb's lifetime is exponentially distributed with parameter $\lambda = 1/1000$.

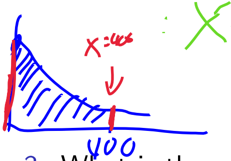
1. What are the units for $\lambda$?

2. What is the probability that the lifetime of the light bulb lasts less than 400 hours?

$X \sim \exp(\lambda)$ $\quad P(X < 400)$ vs. $P(X \leq 400)$

$P(x = 400) = 0$

$$\int_0^{400} \lambda e^{-\lambda x} dx$$



3. What is the probability that the lifetime of the light bulb lasts more than 5 hours?

# Exponential Distribution

**Example:**

Suppose a light bulb's lifetime is exponentially distributed with parameter $\lambda = 1/1000$.

1. What are the units for $\lambda$?

   Same as Poisson: outcomes per time; so maybe burnouts per hour?

2. What is the probability that the lifetime of the light bulb lasts less than 400 hours?

$$P(X < 400) = \int_0^{400} \lambda e^{-\lambda x} = -e^{-\lambda x}|_0^{400} = 1 - e^{2/5}$$

3. What is the probability that the lifetime of the light bulb lasts more than 5 hours?

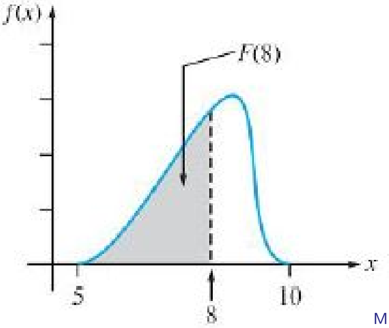
$$P(X > 5) = \int_5^{\infty} \lambda e^{-\lambda x} = -e^{-\lambda x}|_5^{\infty} = 0 - -e^{1/2000} \approx 1$$

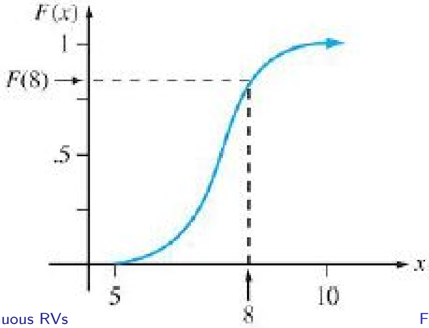
# Cumulative Density Function

**Definition:** *Cumulative Density Function*

The *cumulative distribution function* (cdf) is denoted with $F(x)$. For a continuous r.v. X with pdf $f(x)$, $F(x)$ is defined for every real number x by:

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





# Continuous CDFs

**Example:**

The distribution of the amount of gravel (in tons) sold by a particular construction supply company in a given week is a continuous rv $X$ with pdf

$$f(x) = \begin{cases} \frac{3}{2}(1-x^2) & 0 \leq X < 1 \\ 0 & else \end{cases}$$

1. What is the cdf of sales for any x?
2. Find the probability that $X$ is less than .25?
3. $X$ is greater than .75?
4. $P(.25 < X < .75)$?

# Continuous CDFs

**Example:**

The distribution of the amount of gravel (in tons) sold by a particular construction supply company in a given week is a continuous rv $X$ with pdf

$$f(x) = \begin{cases} \frac{3}{2}(1 - x^2) & 0 \leq X < 1 \\ 0 & else \end{cases}$$

1. What is the cdf of sales for any x?
   $F(x) = P(X \leq x) = \int_0^x \frac{3}{2}(1 - t^2)\,dt$
   $F(x) = \frac{3x}{2} - \frac{x^3}{2}$
2. Find the probability that $X$ is less than .25? $F(.25)$
3. $X$ is greater than .75? $1 - F(.75)$
4. $P(.25 < X < .75)$? $F(.75) - F(.25)$

# Continuous CDFs

Wait, we've seen this before...

**Recall:** *The Fundamental Theorem of Calculus*.

Suppose $F$ is an anti-derivative of $f$. Then:

1. $$\frac{d}{dx}\int_a^x f(t)\,dt = f(x);$$

   a.k.a.

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

2. $$\int_a^b f(x)\,dx = F(B) - F(A).$$

# Percentiles of a Distribution

Definition: The median $\tilde{x}$ of a continuous distribution is the 50th percentile or $.5$ quantile of the distribution.

How can we express this in terms of $f(x), F(x)$?

**Notation**:

**Visually**:

# Percentiles of a Distribution

Definition: The median $\tilde{x}$ of a continuous distribution is the 50th percentile or $.5$ quantile of the distribution.

How can we express this in terms of $f(x), F(x)$?

**Notation**:

$\tilde{x}$ satisfies $F(\tilde{x}) = .5$, or

**Visually**:

$$.5 = \int_{\infty}^{\tilde{x}} f(x)\,dx$$

# Daily Recap

Today we learned

1. Continuous RVs: Uniform, Exponential

Moving forward:

- nb day Friday!

Next time in lecture:

- What to actually do with **all these pdfs**