Continuous Random Variables and Their Distributions

Previously on CSCI 3022

Def: a discrete random variable 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 $a_1, a_2, \ldots$

Def: a probability mass function is the map between the random variable's values and the probabilities of those values

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

Def: a **cumulative distribution function** (CDF) is **a** function whose value at a point **a** is the cumulative sum of probability masses up until **a**.

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

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 include:

- people's heights: $X \in (0, \infty)$
- final grades in a course: $Y \in [0, 100]$
- the time between people arriving in a queue: $Z \in (0, \infty)$

Can you think of other examples?

- congestion on highway
- rain
- pollution

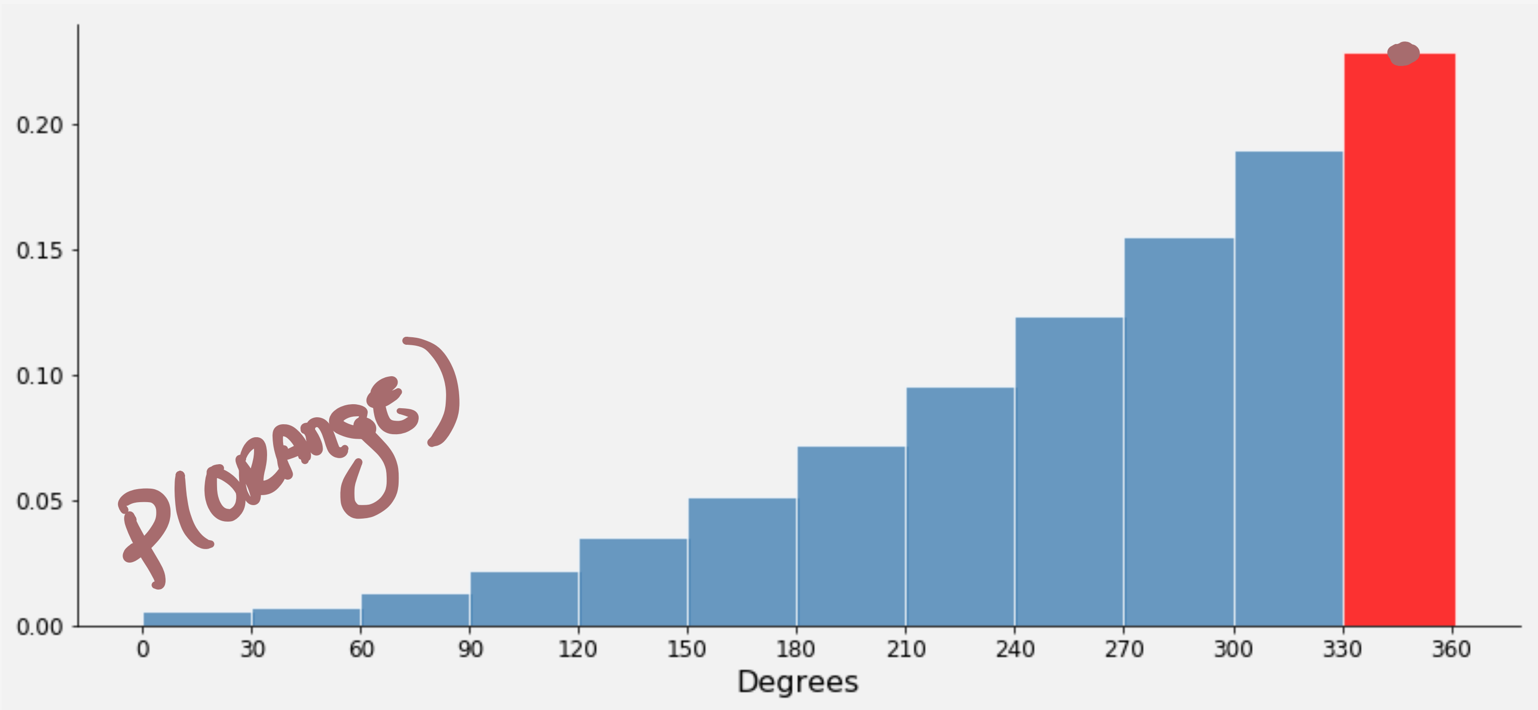
Continuous From Discrete

Example: Suppose you spin the wheel on a game show. Let X be the random variable describing the angle in Degrees at which the wheel stops. Further suppose that the wheel is in disrepair and the closer it gets to 360 Degrees the more likely it is to stop.

Continuous From Discrete

Example: Suppose you spin the wheel on a game show. Let X be the random variable describing the angle in Degrees at which the wheel stops. Further suppose that the wheel is in disrepair and the closer it gets to 360 Degrees the more likely it is to stop.

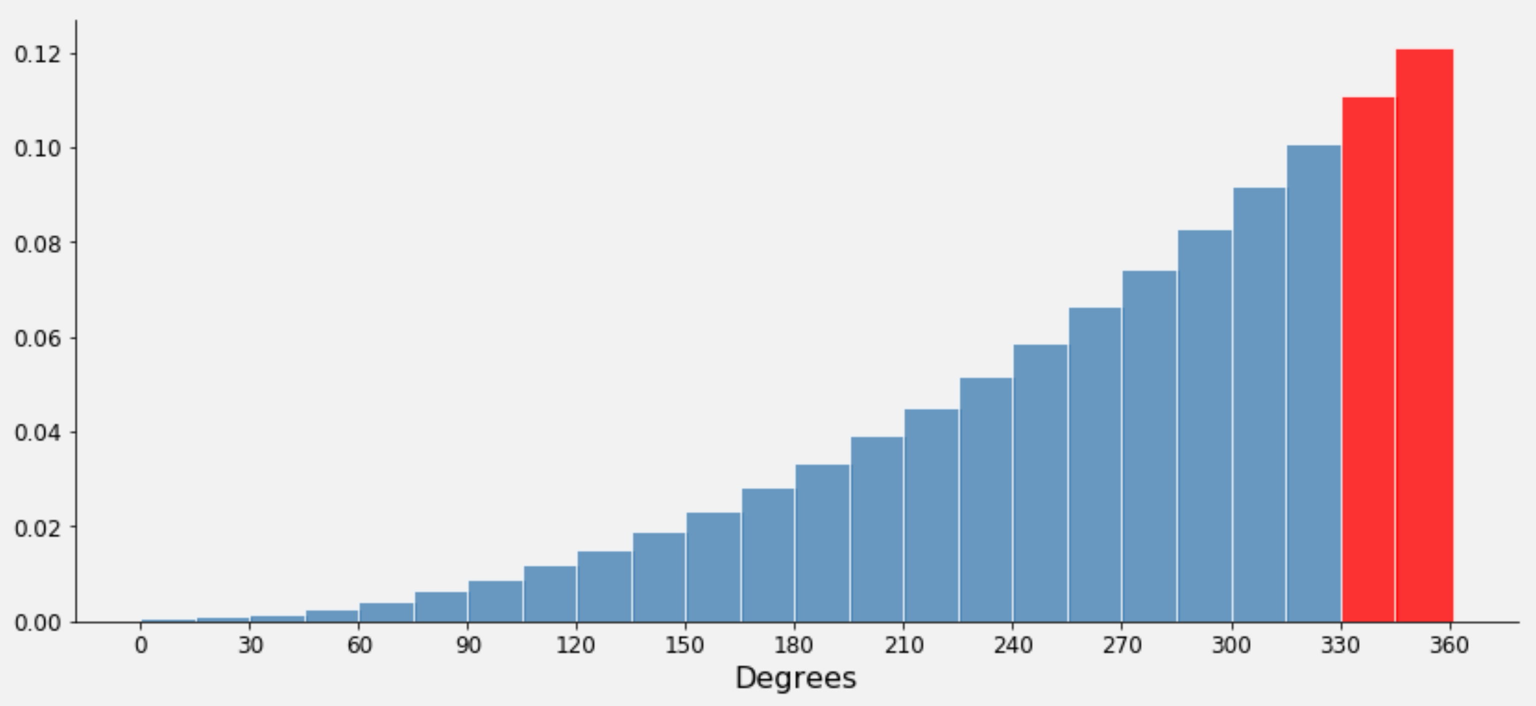
P(RED) = 0.25



Continuous From Discrete

Example: Suppose you spin the wheel on a game show. Let X be the random variable describing the angle in Degrees at which the wheel stops. Further suppose that the wheel is in disrepair and the closer it gets to 360 Degrees the more likely it is to stop.

$P(r_1) + P(r_2) = 0.25$



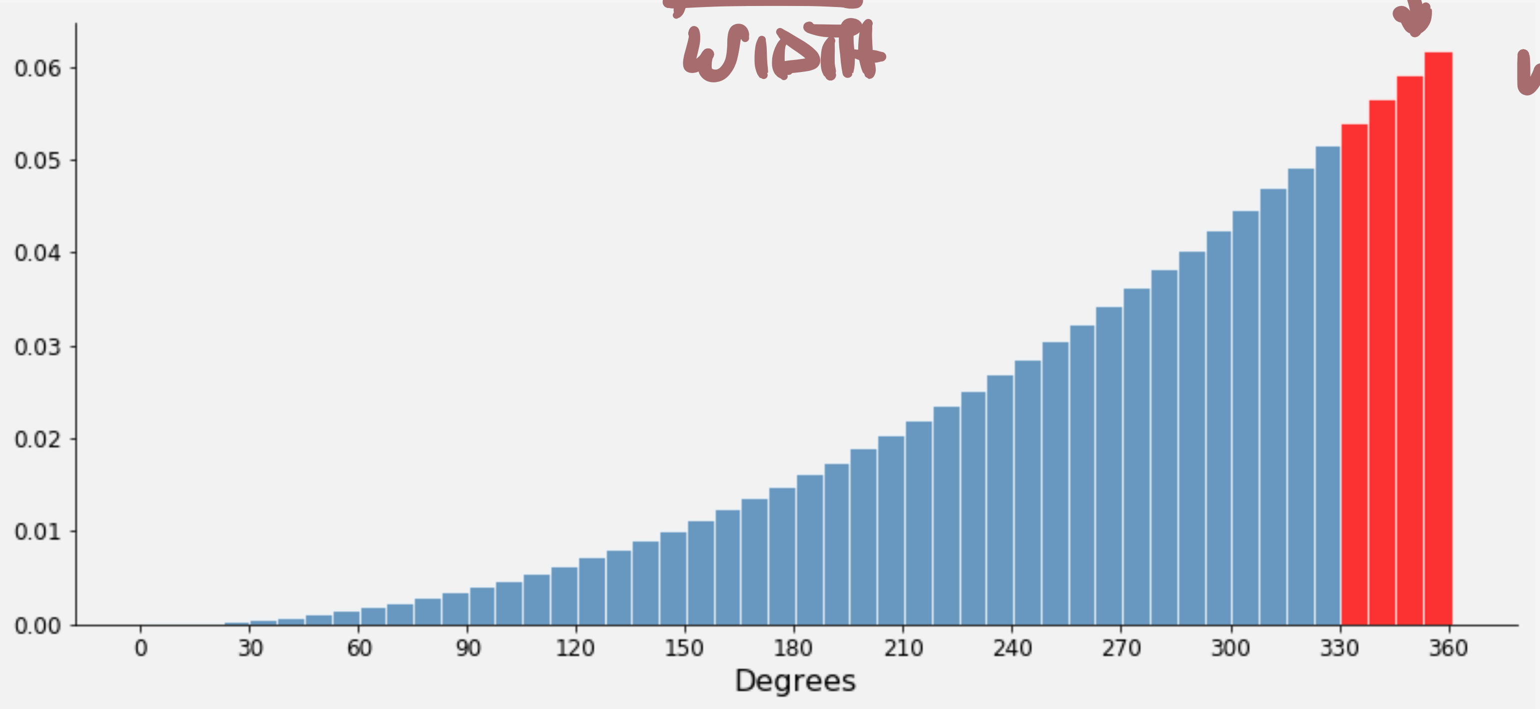
Continuous From Discrete

Example: Suppose you spin the wheel on a game show. Let X be the random variable describing the angle in Degrees at which the wheel stops. Further suppose that the wheel is in disrepair and the closer it gets to 360 Degrees the more likely it is to stop.

$f(x_{200})\Delta x$

$\frac{\text{PROB}}{\text{WIDTH}}$

WIDTH

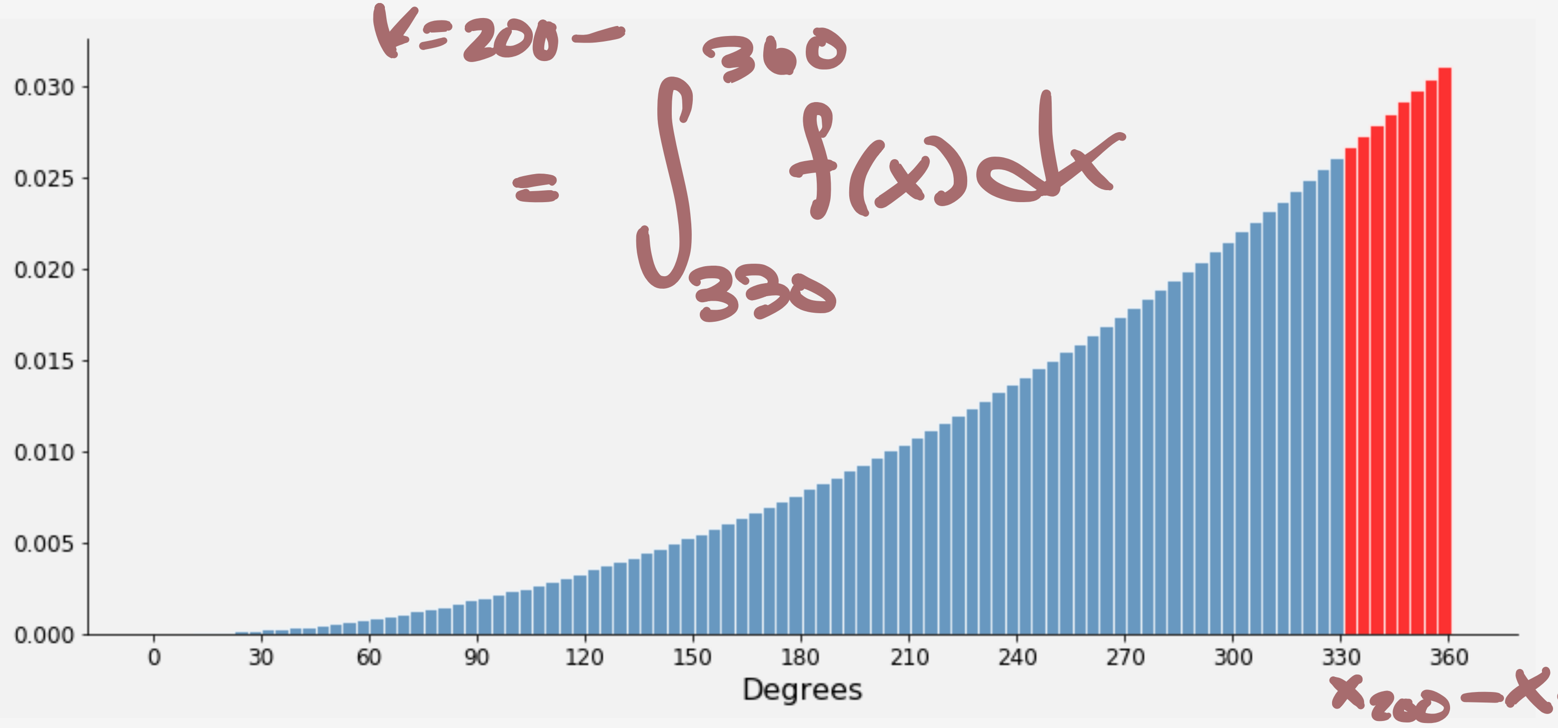


Continuous From Discrete

Example: Suppose you spin the wheel on a game show. Let X be the random variable describing the angle in Degrees at which the wheel stops. Further suppose that the wheel is in disrepair and the closer it gets to 360 Degrees the more likely it is to stop.

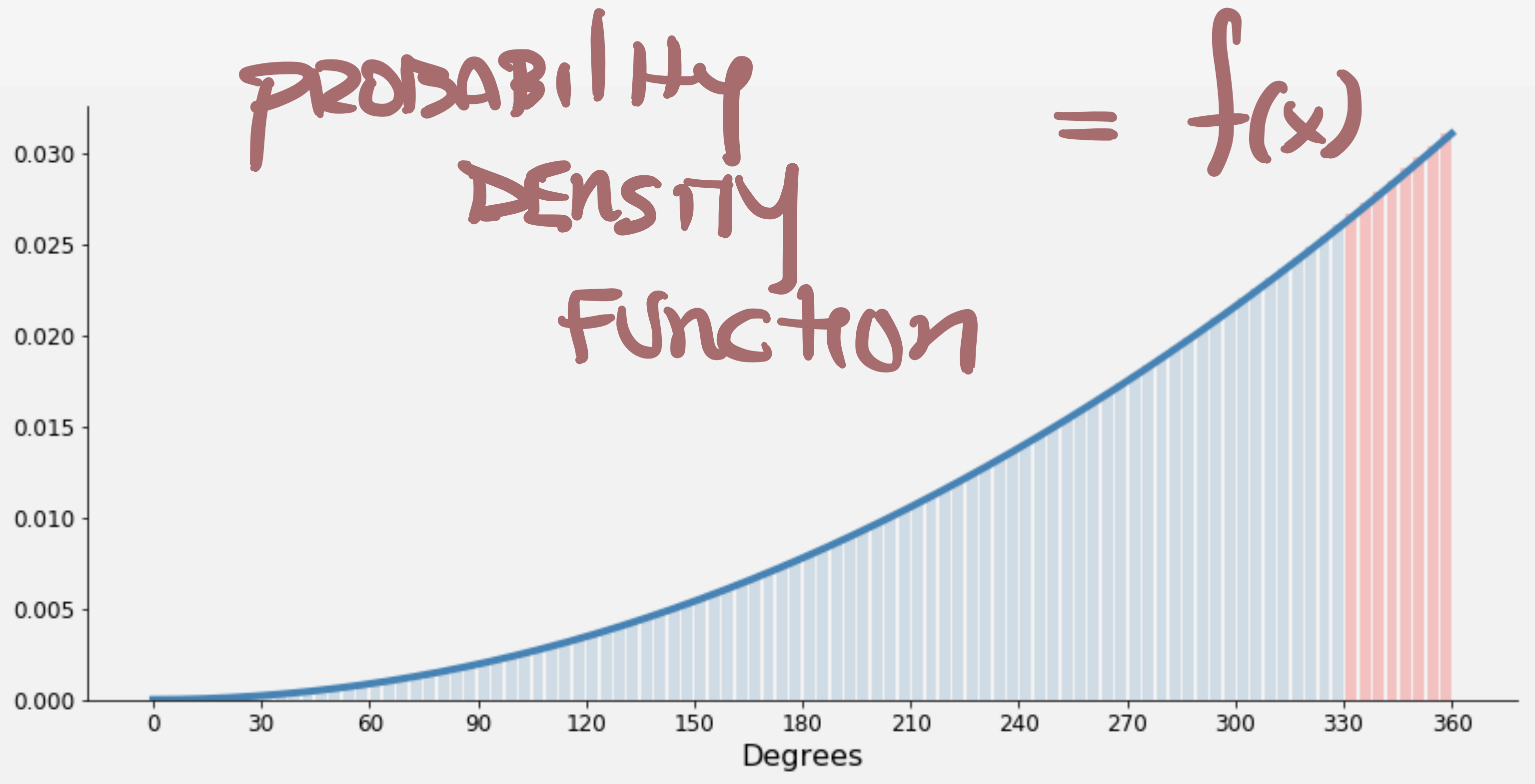
$$\sum_{k=200}^{250} f(x_k)\Delta x = 0.25$$

$$= \int_{330}^{360} f(x)\,dx$$



Continuous From Discrete

Example: Suppose you spin the wheel on a game show. Let X be the random variable describing the angle in Degrees at which the wheel stops. Further suppose that the wheel is in disrepair and the closer it gets to 360 Degrees the more likely it is to stop.



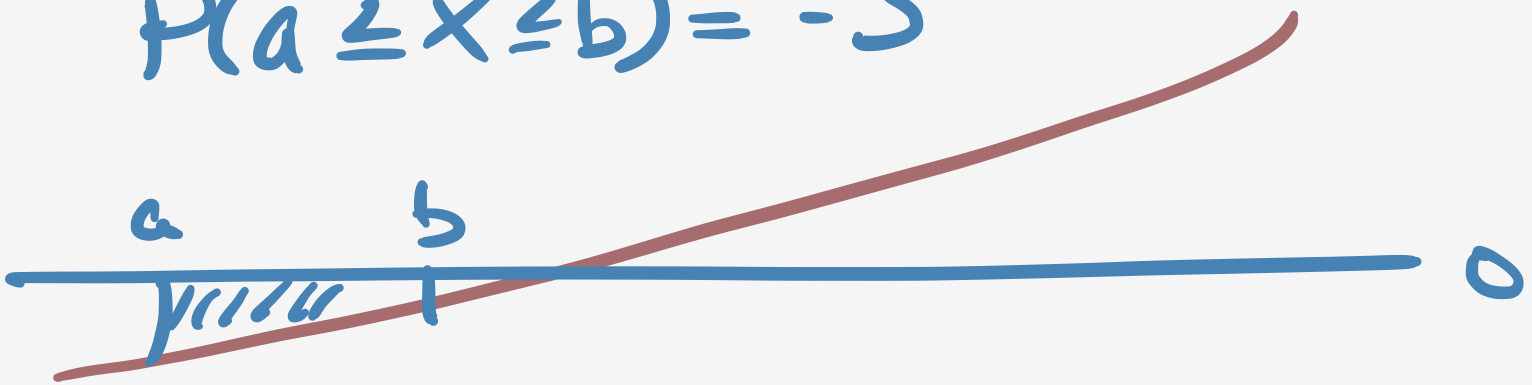
Continuous Random Variables

Def: 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 has to satisfy $f(x) \geq 0$ for all x and $\int_{-\infty}^{\infty} f(x)\, dx = 1$. We call f the probability density function of X.

$P(a \leq X \leq b) = -3$



Continuous Random Variables

Example: Suppose you spin the wheel on a game show. Let X be the random variable describing the angle in Degrees at which the wheel stops. Further suppose that the wheel is in disrepair and the closer it gets to 360 Degrees the more likely it is to stop.

Question: What is the probability the wheel stops on a particular value, say

$$P(X = 355.12345)$$

$$P(355 \leq X \leq 355)$$

$$= \int_{355}^{355} f(x)\,dx = 0$$

Continuous Random Variables

Example: Suppose you fix the wheel so that stopping at any particular angle is equally likely. Find the probability density function for X.

$$f(x) = \begin{cases} c & 0 \le x \le 360 \\ 0 & \text{else} \end{cases}$$

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

$$1 = \int_{-\infty}^{\infty} f(x)\,dx = \int_0^{360} f(x)\,dx = c\int_0^{360} dx$$

$$1 = cx \Big|_0^{360} \Rightarrow c(360-0) = c\,360$$

$$\Rightarrow \boxed{c = \frac{1}{360}}$$

The Uniform Distribution

Example: Suppose you fix the wheel so that stopping at any particular angle is equally likely. Find the probability density function for X.

$X \sim U(0, 360)$

Def: 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 $X \sim U(\alpha, \beta)$.

The Uniform Distribution

Example: Suppose you fix the wheel so that stopping at any particular angle is equally likely. Find the probability density function for X.

Question: What distribution does X follow?

The Uniform Distribution

Example: Suppose you fix the wheel so that stopping at any particular angle is equally likely. Find the probability density function for X.

Question: What distribution does X follow? $X \sim U(0, 360)$

Question: What is the probability that that X falls in red? $\frac{30}{360}$

$$P(330 \leq X \leq 360) = \int_{330}^{360} \frac{1}{360}\, dx$$

$$= \frac{360 - 330}{360} = \frac{30}{360} = \frac{1}{12}$$

The Cumulative Distribution Function

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

For discrete random variables we had the cumulative distribution function

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

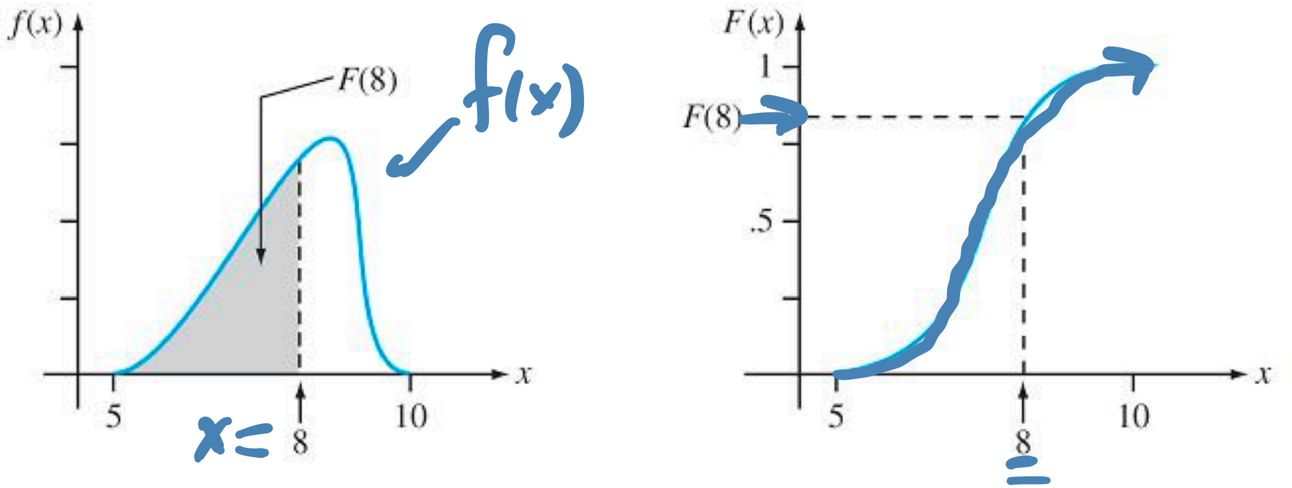
Question: Can we do this for a continuous RV? What would the analogue be?

The Cumulative Distribution Function

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

For continuous random variables we also have a cumulative distribution function

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

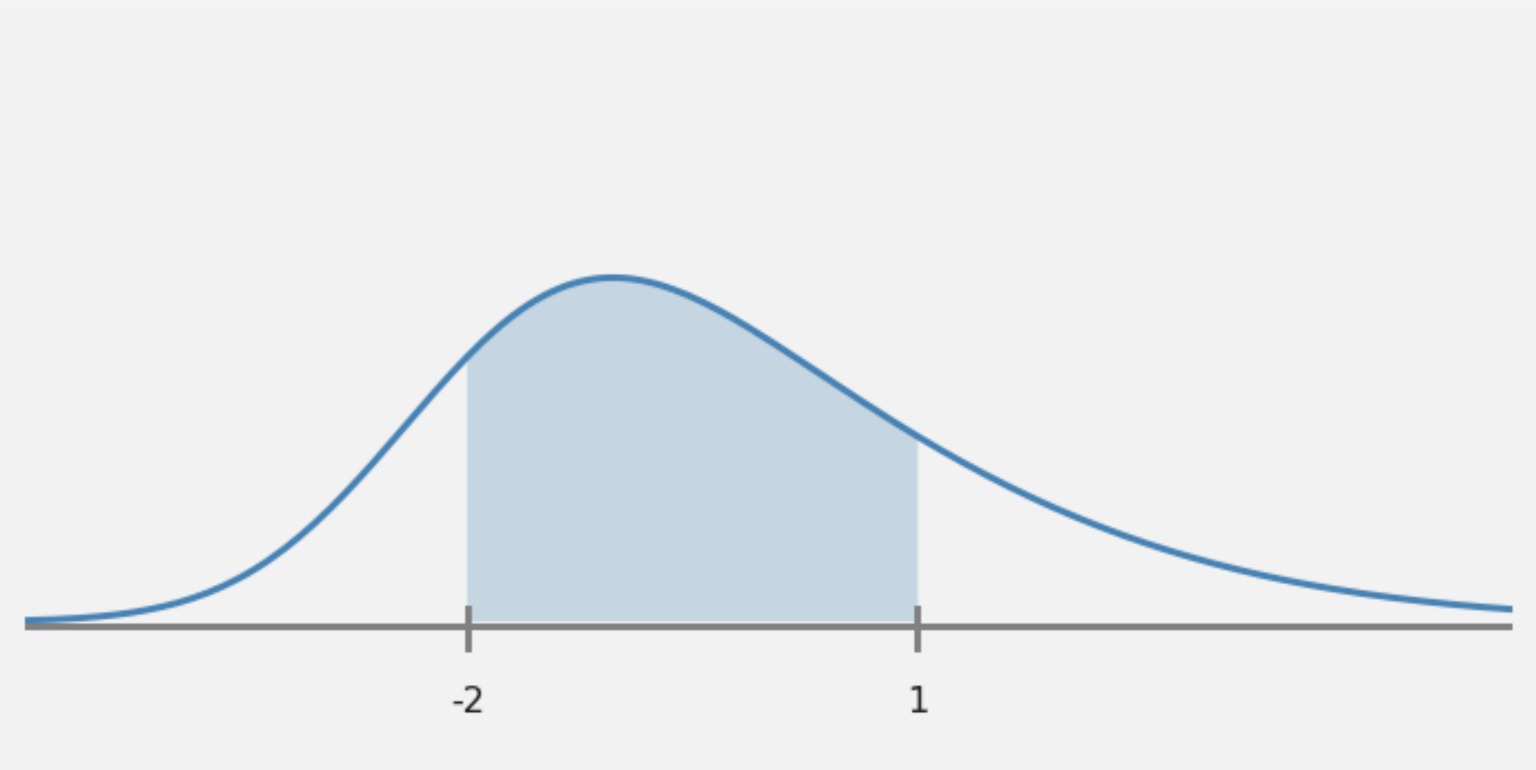


The Cumulative Distribution Function

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

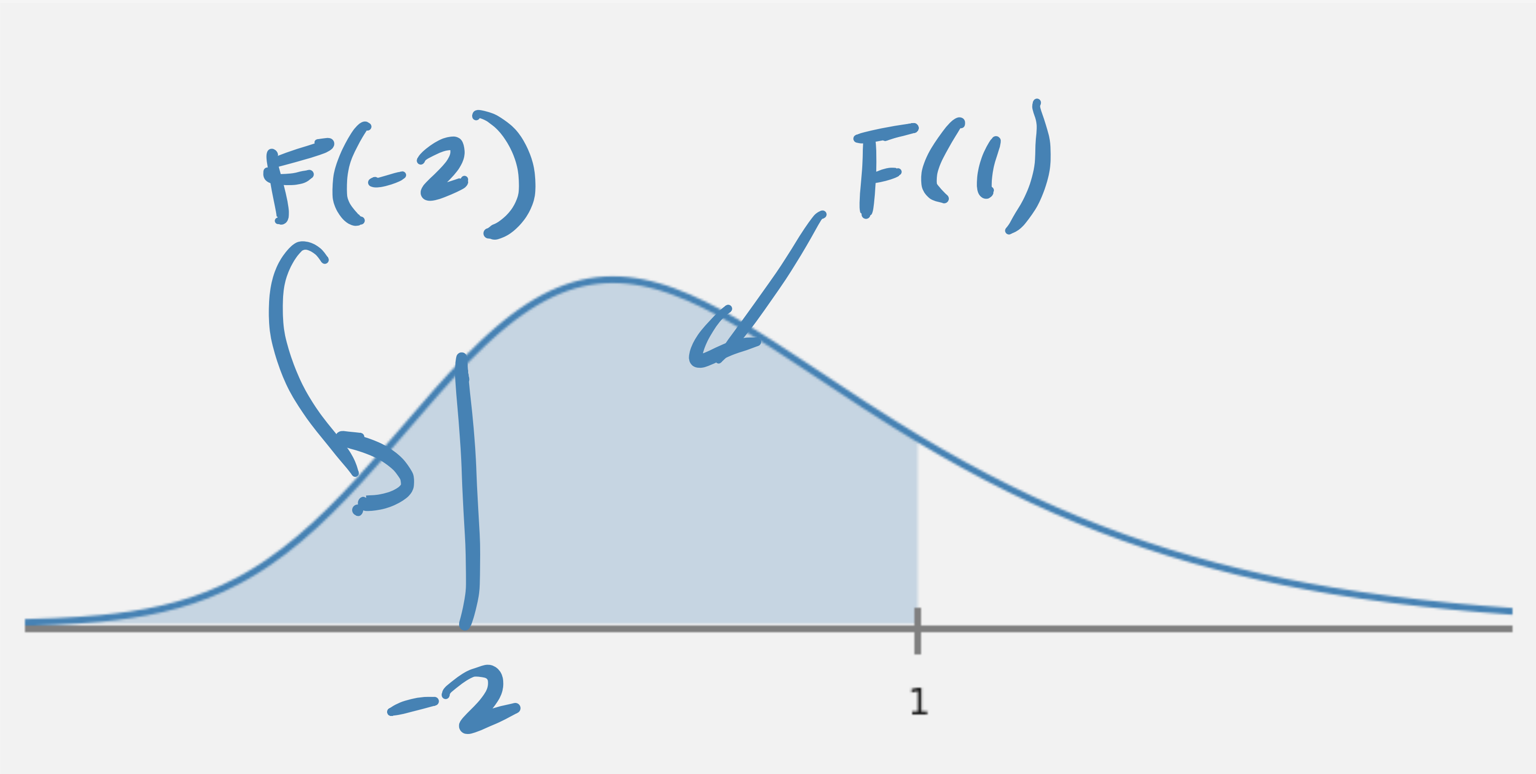
The Cumulative Distribution Function

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



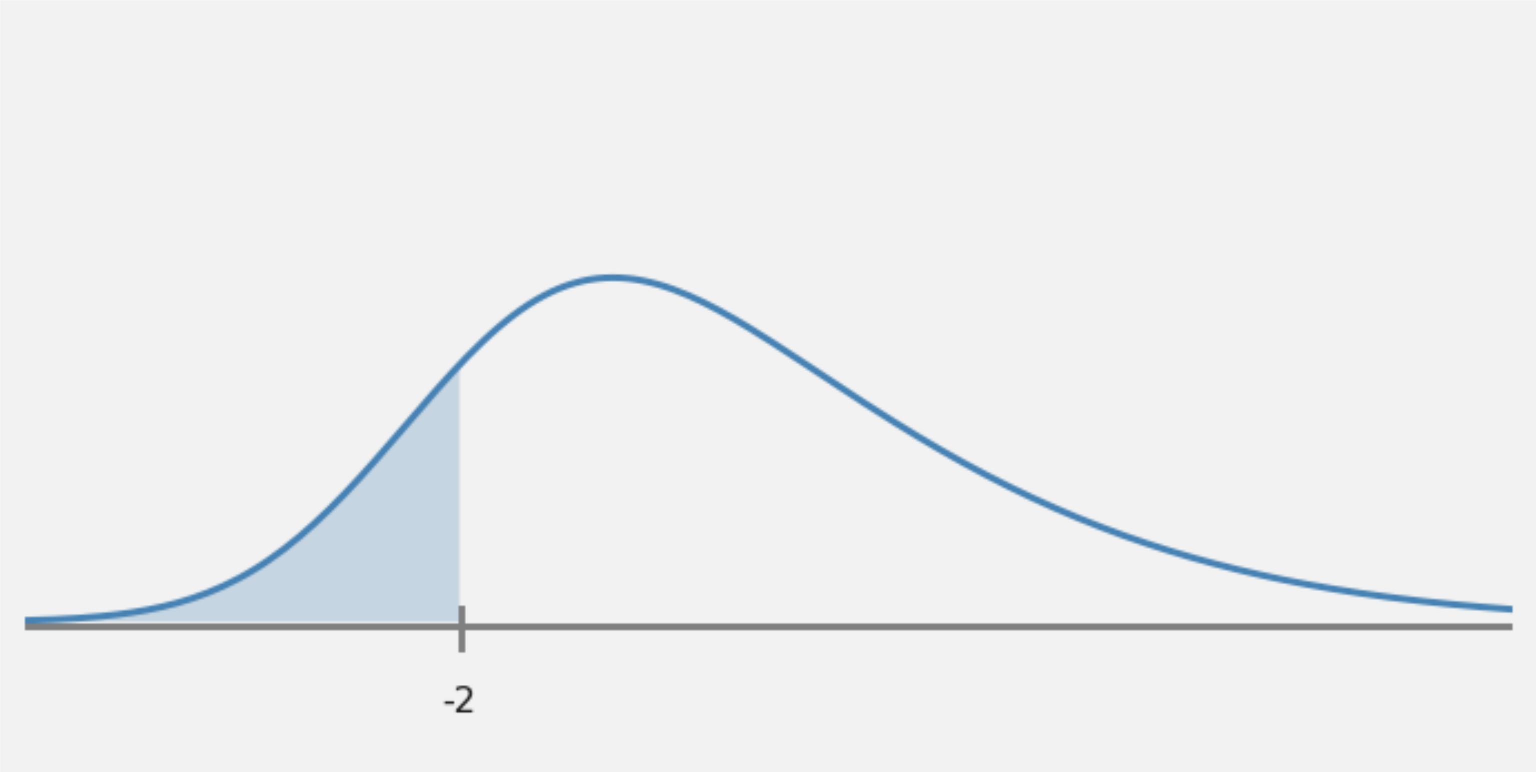
The Cumulative Distribution Function

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



The Cumulative Distribution Function

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

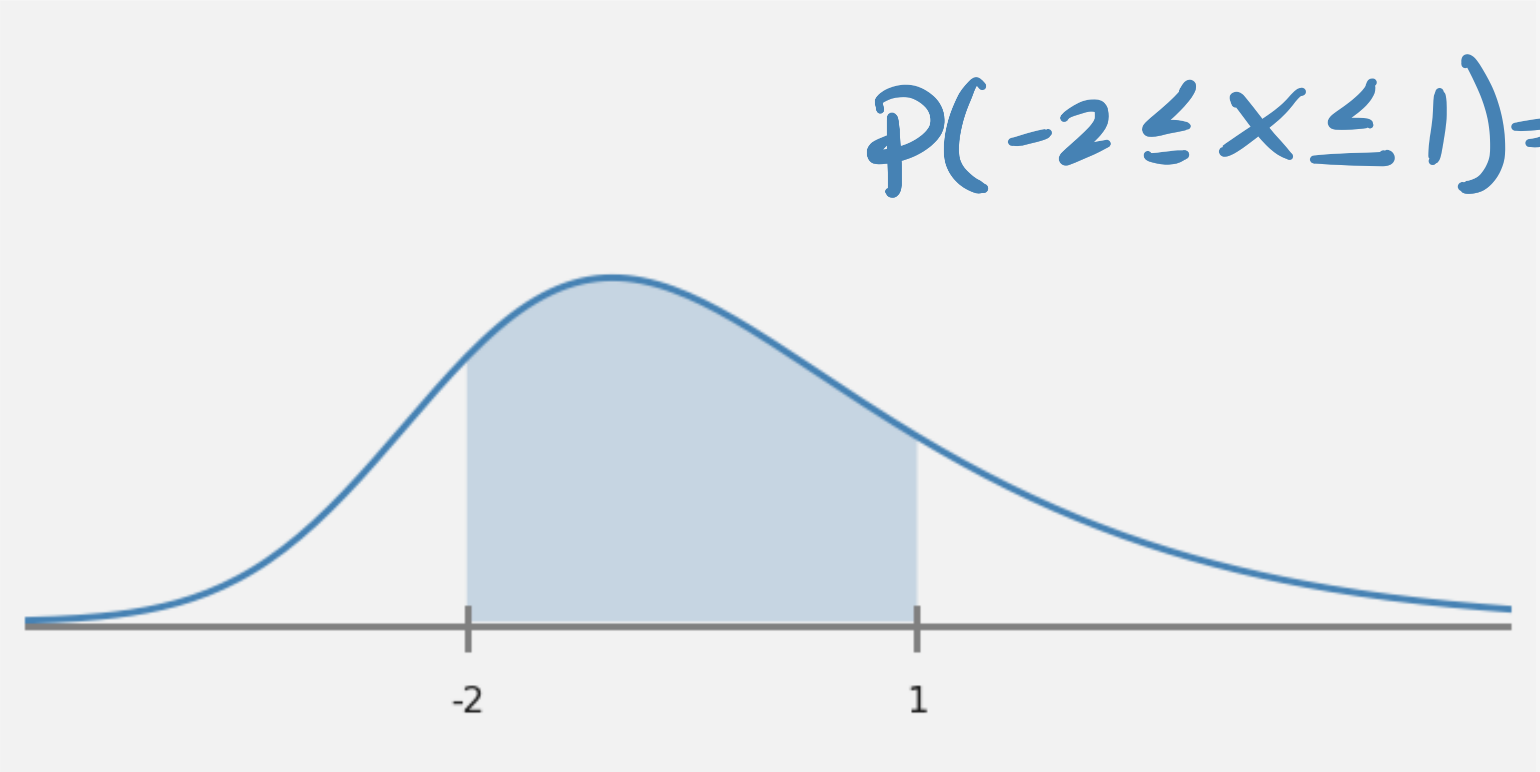


The Cumulative Distribution Function

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

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

$P(-2 \leq X \leq 1) = F(1) - F(-2)$



The Cumulative Distribution Function

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

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

Question: Written this way, does the relationship between F(x) and f(x) seem familiar?

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

The Normal Distribution

Def: a continuous random variable has a normal (or Gaussian) distribution with parameters μ and σ^2 if its probability density function is given by

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

Let's Explore! https://academo.org/demos/gaussian-distribution/

The Exponential Distribution

Sometimes it's easier to first find the cdf and then derive the pdf by taking derivatives

Example: Recall that the Poisson distribution describes the number of arrivals of customers in a line assuming some constant rate of arrivals per time period

A Poisson RV is a discrete RV because we're counting things (the number of arrivals)

But what if we're interested in the amount of time between each arrival?

The Exponential Distribution

Sometimes it's easier to first find the cdf and then derive the pdf by taking derivatives

Example: Recall that the Poisson distribution describes the number of arrivals of customers in a line assuming some constant rate of arrivals per time period

A Poisson RV is a discrete RV because we're counting things (the number of arrivals)

But what if we're interested in the amount of time between each arrival? That's continuous.

The time period between arrivals is a continuous RV following an exponential distribution

μ = RATE OF ARRIVALS PER SOME TIME UNIT

$$f(k) = \frac{\mu^k e^{-\mu}}{k!}$$

The Exponential Distribution

POISSON RATE μt

Strategy: Let's derive the cdf of the exponential distribution from the pmf of Poisson

Suppose the no. of arrivals are governed by a Poisson distribution with rate λ arrivals/min

Start from t=0 and let T be the RV describing the first arrival

What is the probability that the first arrival does not occur in the first t minutes?

$$P(T>t) = P(\text{no arrivals in } \leq t)$$

$$= \frac{(\mu t)^0 e^{-\mu t}}{0!} = e^{-\mu t}$$

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

The Exponential Distribution

Strategy: Let's derive the cdf of the exponential distribution from the pmf of Poisson

Suppose the no. of arrivals are governed by a Poisson distribution with rate λ arrivals/min

Start from t=0 and let T be the RV describing the first arrival

What is the probability that the first arrival does not occur in the first t minutes?

$$F(t) = 1 - e^{-\mu t}$$

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

The Exponential Distribution

Strategy: Let's derive the cdf of the exponential distribution from the pmf of Poisson

Suppose the no. of arrivals are governed by a Poisson distribution with rate λ arrivals/min

Start from t=0 and let T be the RV describing the first arrival

What is the probability that the first arrival does not occur in the first t minutes?

Def: A continuous random variable has an exponential distribution with rate parameter λ if it's probability density function is given by

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

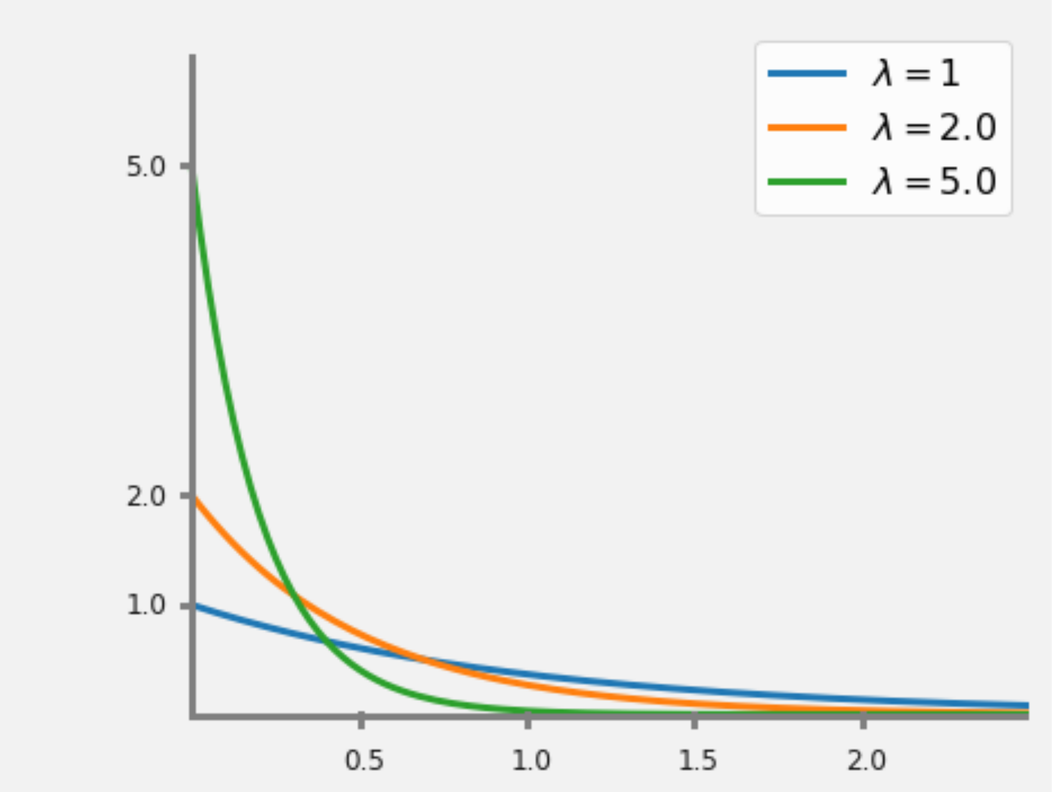
We say $T \sim Exp(\lambda)$

The Exponential Distribution

Def: A continuous random variable has an exponential distribution with rate parameter λ if it's probability density function is given by

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

We say $T \sim Exp(\lambda)$



The Exponential Distribution

OK. That'll do for the first arrival. But we wanted the distribution of time *between* arrivals

The Exponential Distribution

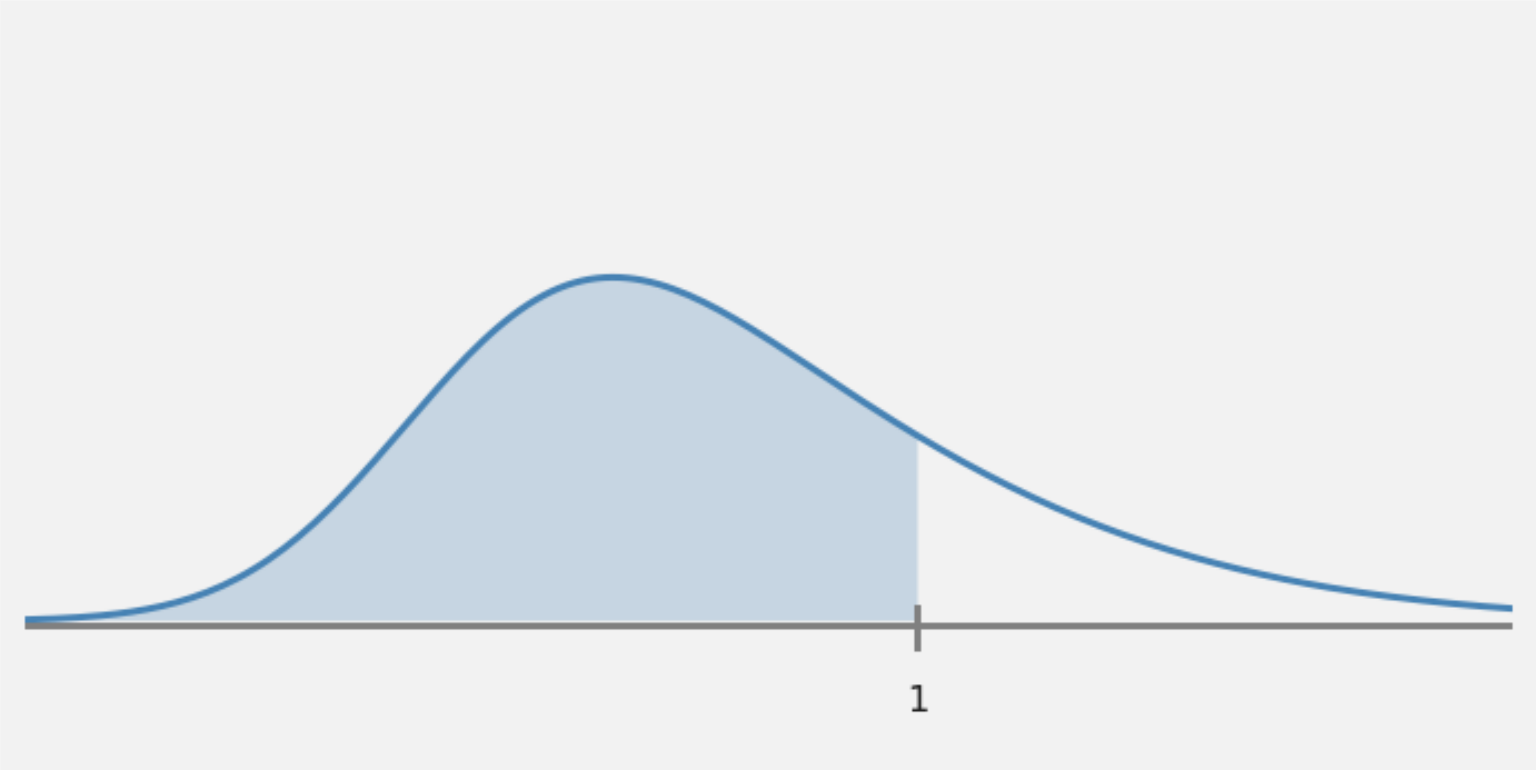
OK. That'll do for the first arrival. But we wanted the distribution of time *between* arrivals

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

Thm: If. $T \sim Exp(\lambda)$ then $P(T > t + t_0 \mid T > t_0) = P(T > t)$

Quartiles and Percentiles

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



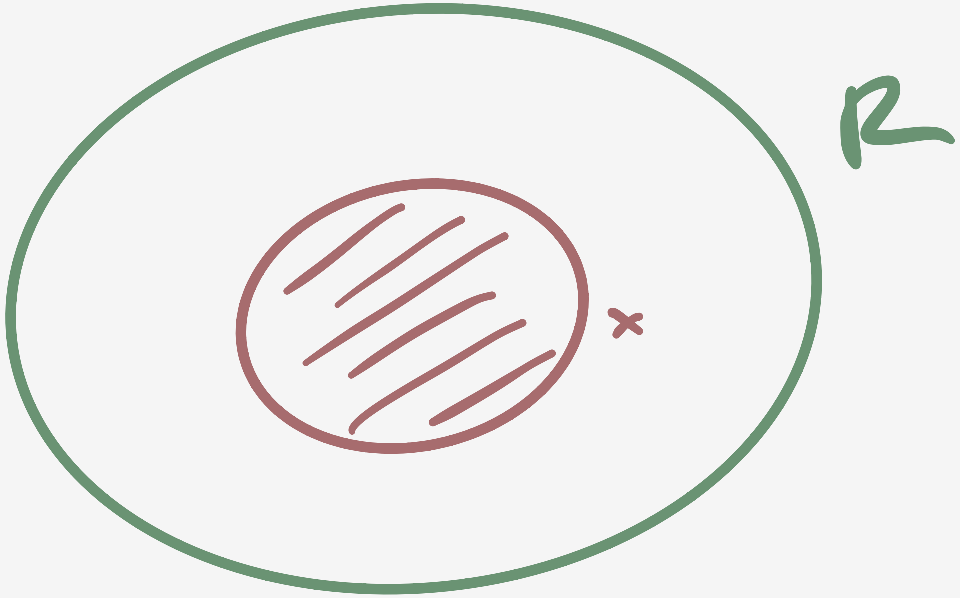
OK! Let's Go to Work!

Get in groups, get out laptop, and open the Lecture 9 In-Class Notebook

Let's:

- Get some more practice with probability density and distribution functions
- See how we can simulate a normal and exponential random variables with Numpy
- See how we can approximate the density function of a continuous RV with histograms

$$F(R) = P(X \le R) = 1$$



$$F(x) = P(X \le x)$$

$$= \frac{\pi x^2}{c} \quad 0 \le x \le R$$

$$1 = F(R) = \frac{\pi R^2}{c} \Rightarrow c = \pi R^2$$

$$F(x) = \frac{\cancel{\pi} x^2}{\cancel{\pi} R^2} \quad 0 \le x \le R$$

$$f(x) = \frac{d}{dx} F(x) = \frac{2\pi x}{\pi R^2} = \frac{2x}{R^2}$$

WHATS THE x S.T. $P(X \leq x) = 0.5$

$= F(x) = 0.5$

$$F(x) = \frac{x^2}{R^2} = \frac{1}{2} \Rightarrow x^2 = \frac{R^2}{2}$$

$$\Rightarrow x = \frac{R}{\sqrt{2}}$$