Introduction to Data Science With Probability and Statistics

Lecture 10: Continuous Variables and their Distributions

CSCI 3022 - Summer 2020
Sourav Chakraborty
Dept. of Computer Science
University of Colorado Boulder

What will we learn today?

- ❑ Probability Density Function
- ❑ Continuous Density Function
- ❑ Continuous Uniform Distribution
- ❑ Normal (Gaussian) Distribution
- ❑ Exponential Distribution

- ❑ *A Modern Introduction to Probability and Statistics, Chapter 5*

Review from Last Time

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 $a_1, a_2, \dots, a_n$ or an infinite number of values $a_1, a_2, \dots$

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

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

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

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

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:

People's heights: $X \in$ ________________

Final grades in a class: $X \in$ ______________

Time between people checking out in a line at the store: $X \in$ ________________

Can you think of other examples?

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:

People's heights: $X \in$ (0, 10) ______

Final grades in a class: $X \in$ [0, 100] ______

Time between people checking out in a line at the store: $X \in$ $(0, \infty)$ ______

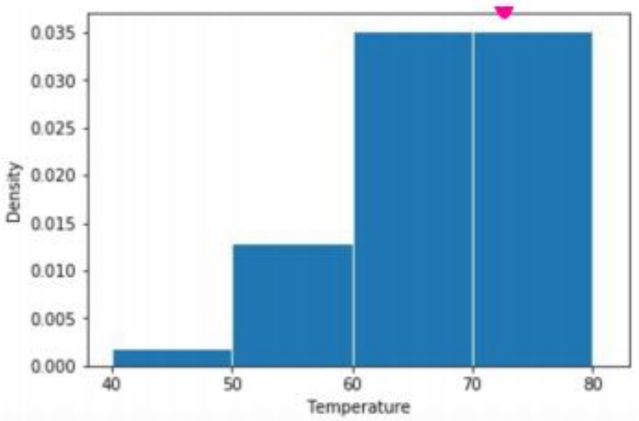
Can you think of other examples?

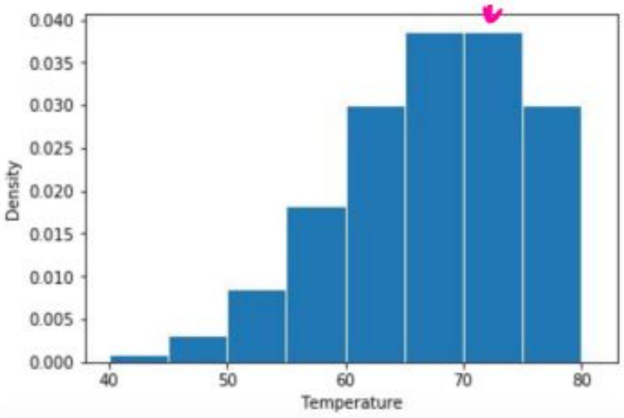
- Temperature
- Speed of car during a race

Continuous Random Variables

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

How would you calculate your response?

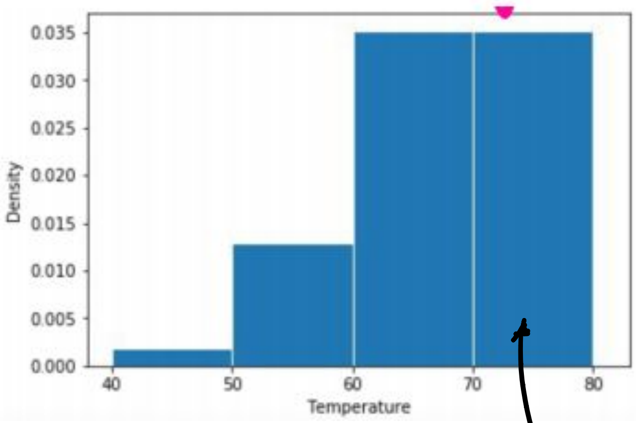




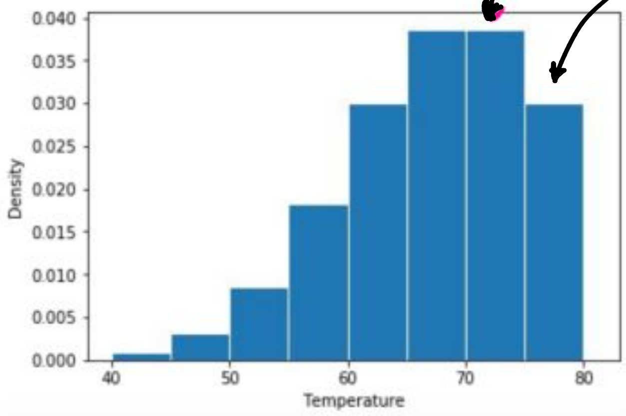
Continuous Random Variables

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

How would you calculate your response?

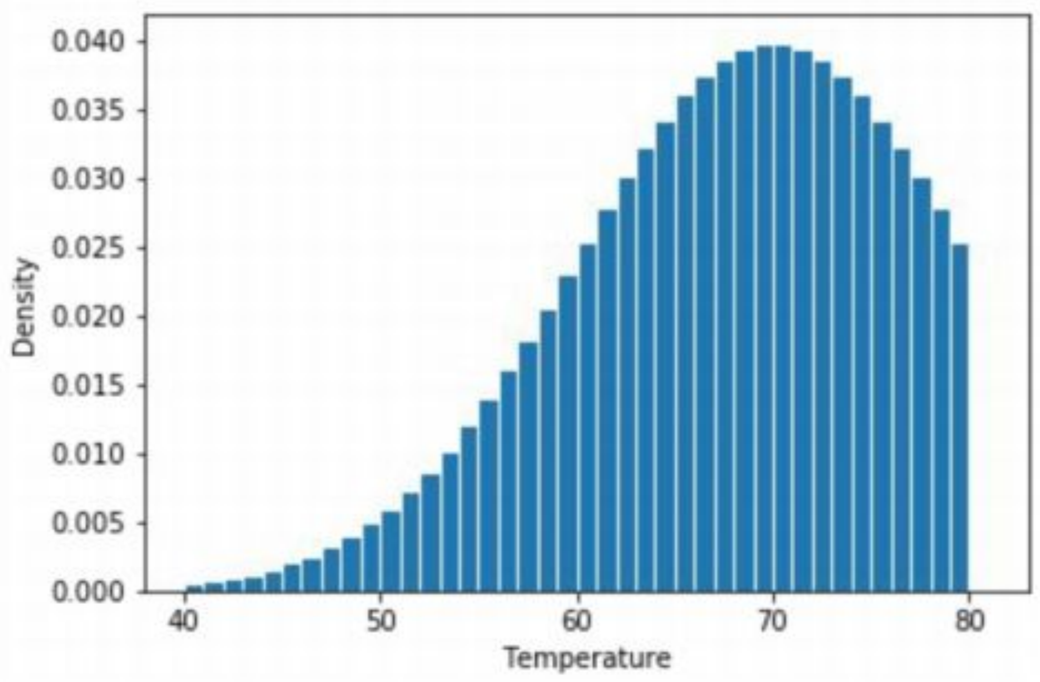


$0.035 \times (80-70) = 0.35$



$5 \times (0.04) + 5 \times (0.03) \approx 0.35$

Continuous Random Variables



How would you calculate your response?

You could use a probability **density** function.

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

Continuous Random Variables

A random variable X is **continuous** if for some function $f\colon \mathbb{R} \to \mathbb{R}$ 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

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

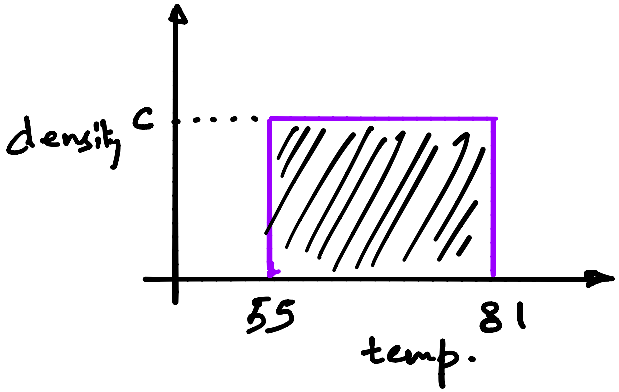
❖ We call f the probability density function (pdf) of X.

Continuous Random Variables

Example: Suppose 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.

Continuous Random Variables

Example: Suppose 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.



$$f(x) = \begin{cases} c, & 55 \le x \le 81 \\ 0, & x < 55 \text{ or } x > 81 \end{cases}$$

what is c?

Continuous Random Variables

Example: Suppose 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.

Properties:

1. $f(x) \geq 0 \;,\; \forall x \in [55, 81]$
2. $\int_{-\infty}^{\infty} f(x)\,dx = 1.$

$$f(x) = \begin{cases} 1/26, & x \in [55, 81] \\ 0, & \text{otherwise} \end{cases}$$

$$= \underbrace{\int_{-\infty}^{55} f(x)\,dx}_{0} + \int_{55}^{81} f(x)\,dx + \underbrace{\int_{81}^{\infty} f(x)\,dx}_{0} = 1.$$

$$\Rightarrow c \int_{55}^{81} dx = 1 \quad \Rightarrow \boxed{c = 1/26}$$

Uniform Distribution

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

What distribution does temperature follow in the example?

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

What is the probability that the temperature is exactly 75°F?

Uniform Distribution

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

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

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

What distribution does temperature follow in the example? $T \sim U(55, 81)$

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

What is the probability that the temperature is exactly 75°F?

Uniform Distribution

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

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

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

What distribution does temperature follow in the example? $T \sim U(55, 81)$

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

$$P(75 \le x \le 81) = \int_{75}^{81} (1/26)\,dx = 3/13$$

What is the probability that the temperature is exactly 75°F?

Uniform Distribution

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

What distribution does temperature follow in the example? $T \sim U(55, 81)$

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

$$P(75 \leq x \leq 81) = \int_{75}^{81} (1/26)\,dx = 3/13$$

What is the probability that the temperature is exactly 75°F?

$$\int_{75}^{75} 1/26 \cdot dx = 0$$

Cumulative Distribution Function

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?

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

Cumulative Distribution Function

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?

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

Continuos : $F(x) = \int_{-\infty}^{x} f(x)\,dx$

pdf

Cumulative Distribution Function

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

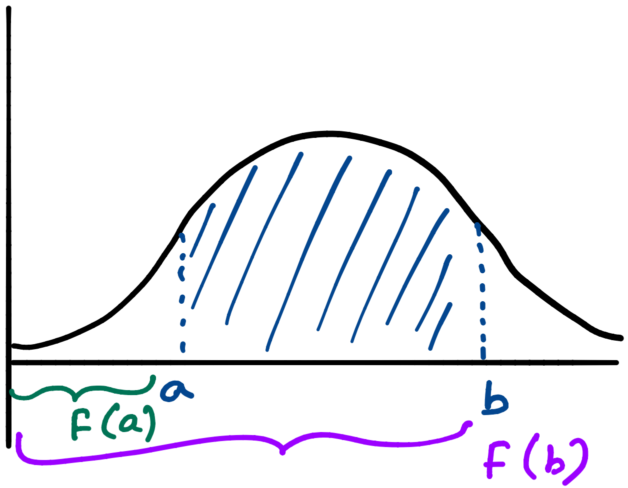
Cumulative Distribution Function

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

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

$$= \int_{-\infty}^{b} f(x)\,dx - \int_{-\infty}^{a} f(x)\,dx$$

$$= \boxed{F(b) - F(a)}$$



Cumulative Distribution Function

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

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

Cumulative Distribution Function

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

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

$$\frac{d}{dx} F(x) = \frac{d}{dx} \int_{-\infty}^{x} f(t)\,dt$$; taking derivatives both sides

$$= f(x)$$

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

Normal Distribution

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/

μ : mean.

σ : std.

Exponential Distribution

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 $X \sim Exp(\lambda)$

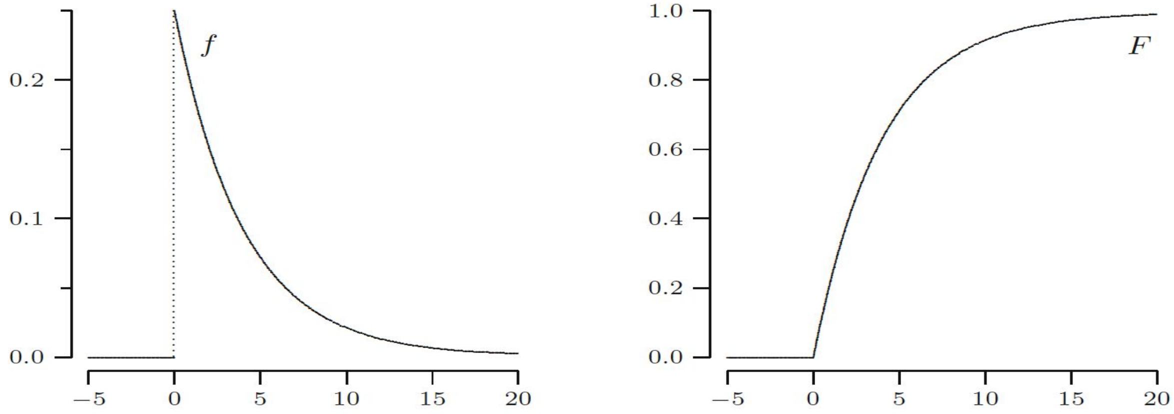


Fig. left: PDF for X~ Exp(0.25)

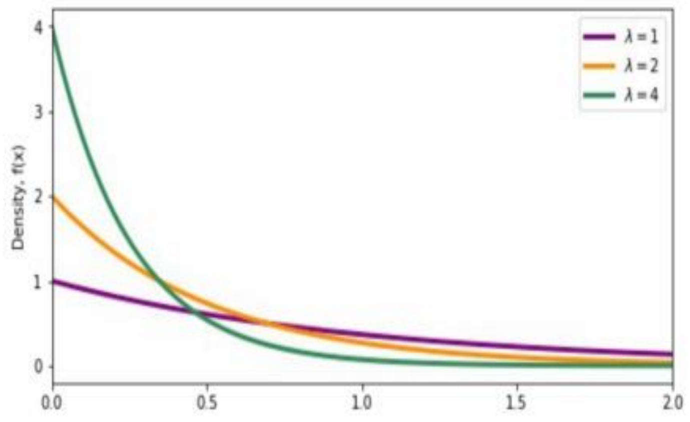
Fig. Right: CDF for X~Exp(0.25)

Exponential Distribution

The exponential distribution may be viewed as a continuous counterpart of the geometric distribution, but the exponential distribution describes the time for a continuous process to change state.

Few Situations where this is used:

- The time until a radioactive particle decays, or the time between clicks of a Geiger counter
- The time it takes before your next telephone call
- For situations where certain events occur with a constant probability per unit length, such as the distance between mutations on a DNA strand, or between roadkills on a given road.



Exponential Distribution

Theorem: (memoryless property)

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

Next Time

❖ Expectation