ONLINE MASTERS IN **DATA SCIENCE**

DSC 215 - PROBABILITY AND STATISTICS FOR DATA SCIENCE

RANDOM VARIABLES

UC San Diego

COMPUTER SCIENCE & ENGINEERING

HALICIOĞLU DATA SCIENCE INSTITUTE

Random Variables

- In the context of statistics, **random variables (rv)** allow us to apply a mathematical framework for better understanding and predicting outcomes in the real world.

- Mathematically, **a random variable** is a function from Ω to another set, such as the real line $\mathbb{R}$, or the natural numbers (positive integers) $\mathbb{N}$.

- If the range is, say, $\mathbb{R}$, we have the random variable $X : \Omega \to \mathbb{R}$, which means that for $\omega \in \Omega,$ we may write $X(\omega) = x \in \mathbb{R}$.

Example: Random Variables

Example:

- Consider Ω to be the set of all length 3 sequences of Heads and Tails so that, for example

$$\underbrace{(H, H, T)}_{\omega} \in \Omega$$

- Suppose the random variable X is the function that counts how many heads are in the sequence. So, in our example

$$\omega = (H, H, T) \implies X(\omega) = 2$$

Discrete VS Continuous Random Variables

- **Discrete random variables:** In our previous example $X : \Omega \to \{0,1,2,3\}$ so we call it a **discrete random variable** (because the range is a discrete set). Note that the discrete set can be infinite, like $\mathbb{N}$.

- We can expand our notation to more conveniently write the probability that there are k heads in our sequence as

$$\mathbb{P}(X = k) := \mathbb{P}(\{\omega : X(\omega) = k\}).$$

- **Continuous random variables**: If, for example $X : \Omega \to \mathbb{R}$, we call it a continuous random variable. Here we can introduce the convenient notation

$$\mathbb{P}(a \leq X \leq b) := \mathbb{P}(\{\omega : a \leq X(\omega) \leq b\})\,.$$

Probability Mass Functions (Discrete Random Variables)

- For a ***discrete*** random variable the **probability mass function (PMF)** is the function

$$p_X : \mathbb{R} \to [0,1] \quad \text{with} \quad p_X(a) = \mathbb{P}(X = a).$$

- It gives the probability that a discrete random variable is exactly equal to some value.
- PMFs satisfy
  - $p_X(x) \geq 0$ (all probabilities are non-negative)
  - $\sum_x p_X(x) = 1$ (because probabilities of all disjoint events sum to 1)

Probability Mass Functions

Example:

- Let X be a random variable associated with a coin-tossing experiment, so $\Omega = \{H, T\}$, and suppose that $X(H) = 1$ and $X(T) = 0$.

- If the coin is fair: $$p_X(x) = \begin{cases} 1/2 & \text{if} \quad x = 0 \\ 1/2 & \text{if} \quad x = 1 \\ 0 & \text{otherwise.} \end{cases}$$

Cumulative Distribution Functions (for Cont. RVs)

- For a real-valued random variable (i.e., when $X(\omega)$ is a number), a **cumulative distribution function (CDF)** is a function

$$F_X : \mathbb{R} \mapsto [0,1] \quad \text{with} \quad F_X(x) = \mathbb{P}(X \leq x)\,.$$

- The CDF allows us to compute the probability of any event in $\mathscr{F}$.

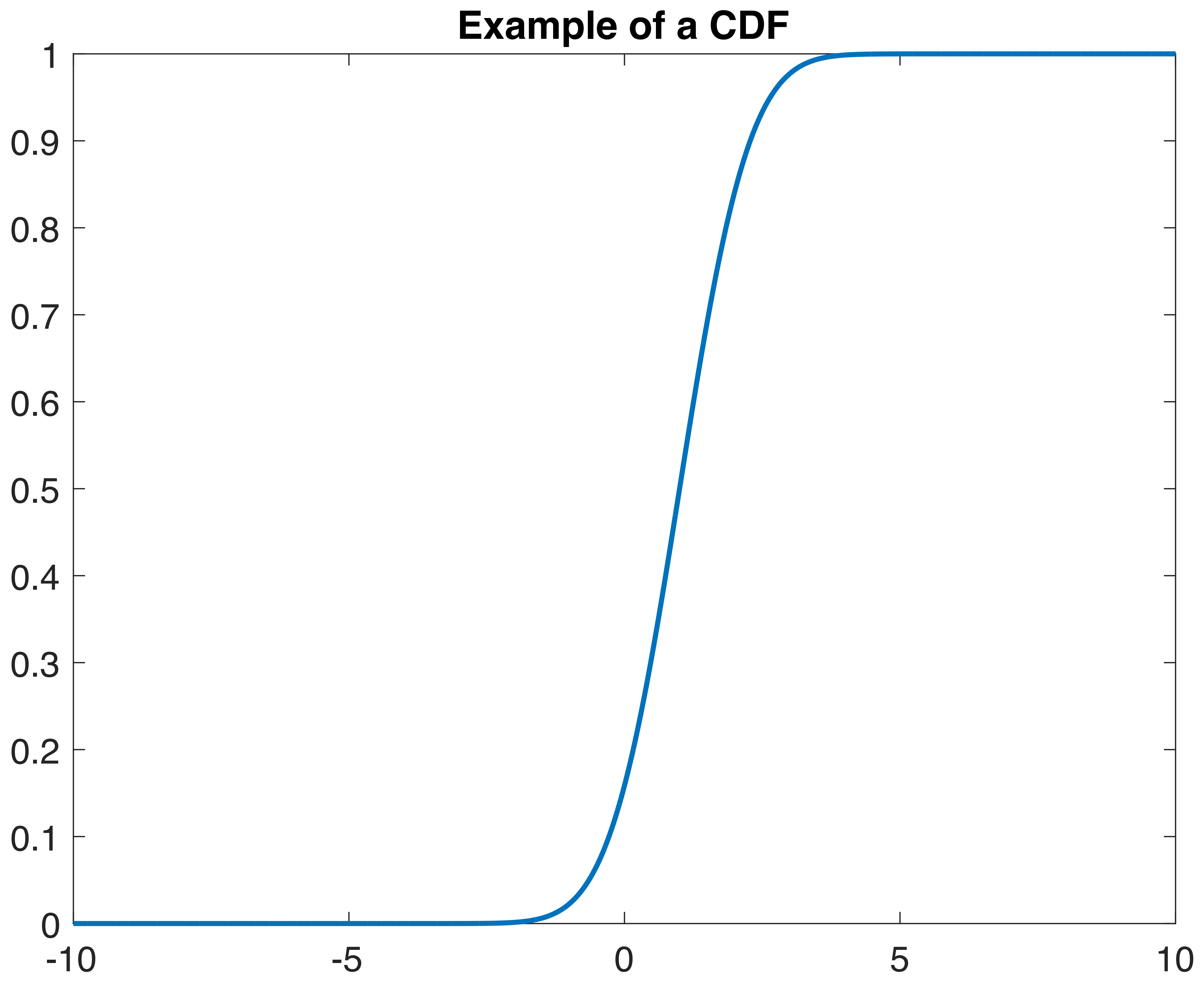
- For example, if we know the CDF, we can calculate

$$\mathbb{P}(1 \leq X \leq 2.7) = F_X(2.7) - F_X(1)$$

Properties of CDFs

- CDFs satisfy important properties (as can be deduced from the definition and the axioms of probability):

  - $0 \leq F_X(x) \leq 1$.
  - $\lim_{x \to -\infty} F_X(x) = 0$.
  - $\lim_{x \to +\infty} F_X(x) = 1$.
  - $x \leq y \implies F_X(x) \leq F_X(y)$.



Probability Density Functions (for Cont. RVs)

- For a ***continuous*** random variable, with a differentiable CDF, the **probability density function (PDF)** is the function

$$f_X(x) = \frac{d}{dx} F_X(x)\,.$$

- So, it is the derivative of the CDF.

- Like the CDF, it can help us calculate probabilities: $\mathbb{P}(a \leq X \leq b) = \int_a^b f_X(x)dx$

- The PDF and CDF also satisfy $F_X(x) = \int_{-\infty}^{x} f_X(z)dz$

Properties of PDFs

- PDFs satisfy important properties:
  - $f_X(x) \geq 0.$
  - $\int_{-\infty}^{\infty} f_X(x)dx = 1$
  - For a set A, $\int_A f_X(x)dx = \mathbb{P}(X \in A)$
- **Important:** In general $f_X(x) \neq \mathbb{P}(X = x)$.
- **Exercise:** Construct an example for which $f_X(x) \neq \mathbb{P}(X = x)$