Title Slide Title: Probability Concepts and Similarity Measures

---

Discrete Random Variable

- A discrete random variable takes countable values (e.g., whole numbers).
- Example: Number of heads in 3-coin tosses.
- **Significance:** Used in scenarios where outcomes are distinct (e.g., number of defective products).

Continuous Random Variable

- A continuous random variable takes an infinite number of values.
- Example: Height of students in a class.
- **Significance:** Used for measuring real-world phenomena like temperature, time, and weight.

Joint Probability

- The probability of two or more events occurring simultaneously.
- Example: The probability that a student passes both Math and Science exams.
- **Formula:**
  $F(a \le X \le b, c \le Y \le d) = \int_{b}^{a} \int_{d}^{c} fX,Y(x,y)dydx$
- **Significance:**
- Helps in understanding the relationships between multiple random variables.

Marginal Probability

- The probability of a single event occurring, regardless of other events.
- Example: The probability that a student passes Math, regardless of their Science score.
- **Formula:**
  $f_X(x) = \int f(x, y)\, dy$

  $f_Y(y) = \int f(x, y)\, dx$

- **Significance:** Helps in simplifying complex probability distributions and understanding individual event likelihoods.

Conditional Probability

- The probability of an event occurring given that another event has already occurred.
- Example: The probability of a student passing Science given that they passed Math.
- **Formula:**
  $P(A|B)=\frac{P(A\cap B)}{P(B)}$
- **Significance:** Essential for decision-making and real-world applications like spam detection and medical diagnosis.

Probability Density Function (PDF)

- A PDF describes the probability distribution of a continuous random variable.
- It tells us how likely a value is within a given range.
- Properties:

- - Always non-negative ($f(x) \geq 0$)
- **Formula:**

  F(a≤X≤b)=$\int_{b}^{a} fX(x)dx$

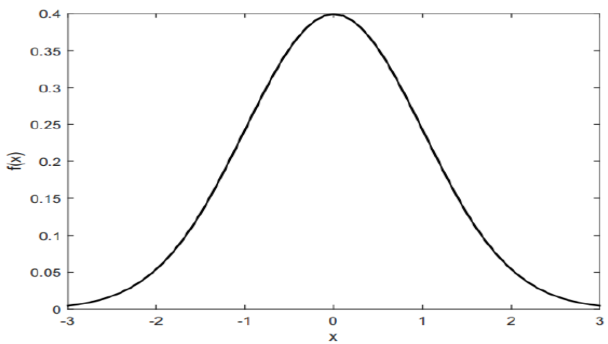
- **Significance:**
- Describes the probability distribution of continuous variables.

Probability Mass Function (PMF)

- A function that gives the probability of a discrete random variable taking a specific value.
- Example: The probability of rolling a 3 on a six-sided die.
- **Formula:**
  P(X=x)P(X = x)
- **Significance:** Crucial for discrete probability distributions such as Binomial and Poisson distributions.

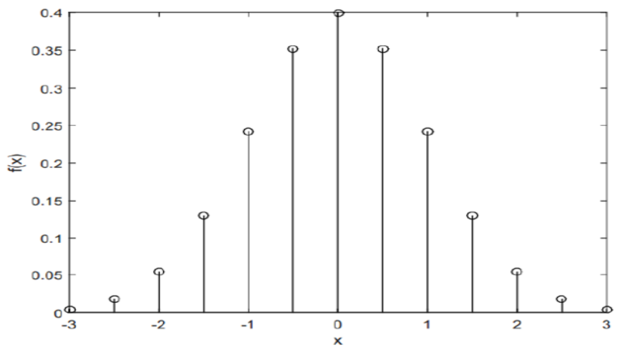
Cumulative Density Function (CDF)

- A function that provides the probability up to a certain point.
- Example: The probability of a student's height being less than 170 cm.
- **Formula:**
  $F(x)=P(X \le x)$
- **Significance:** Used to determine probabilities over intervals.



(a)

PDF



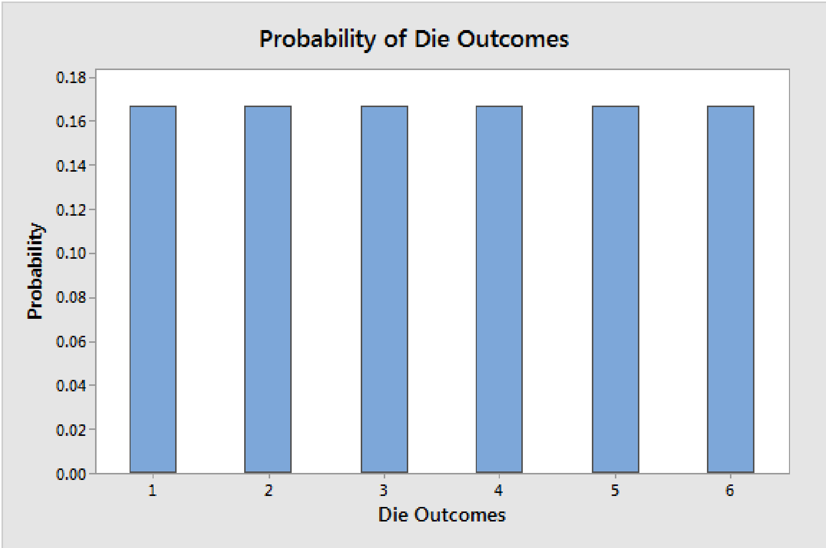
(b)

PMF

Uniform Distribution

- A probability distribution where all outcomes are equally likely.
- Example: Rolling a fair die.
- **Formula:**

  $f(x)=\frac{1}{b-a}$, $a \leq x \leq b$



Binomial Distribution

- A discrete distribution for the number of successes in *n* independent trials.
- Example: The number of heads in 10 coin tosses.
- **Formula:**

$$P_x = \binom{n}{x} p^x q^{n-x}$$

Poisson Distribution

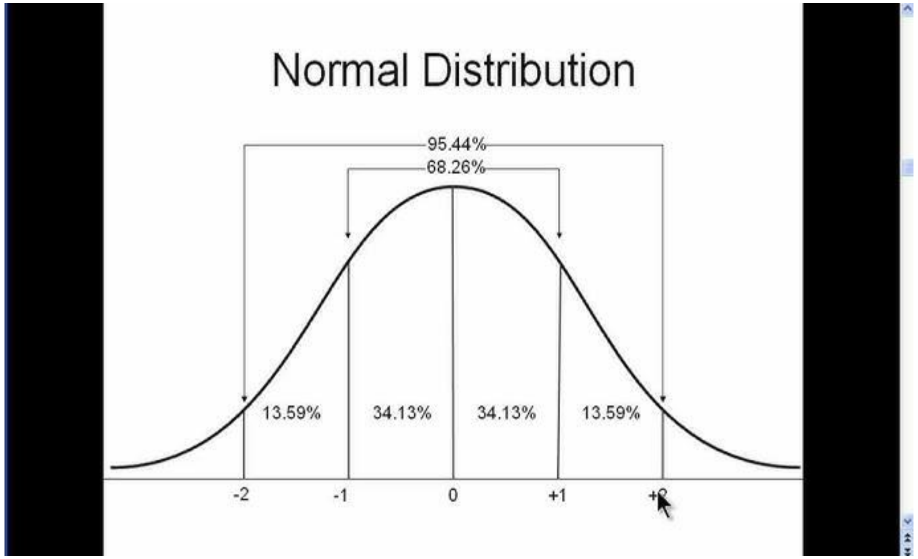
- Models the number of events occurring in a fixed interval.
- Example: The number of customer arrivals at a store per hour.
- **Formula:**

$$P(X=k) = \frac{e^{-\lambda}\lambda^k}{k!}$$

Normal Distribution

- It is a type of probability distribution that is symmetric and bell-shaped, commonly known as the Gaussian distribution.
- It describes how values of a continuous variable are distributed.
- Example: The distribution of student’s heights.
- **Formula:**

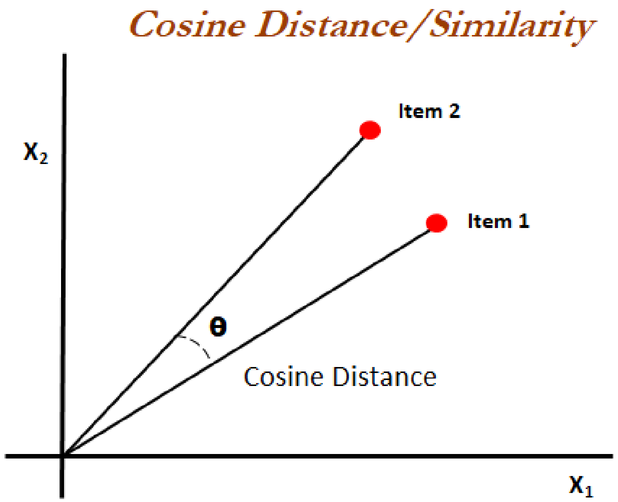
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$



Cosine Similarity

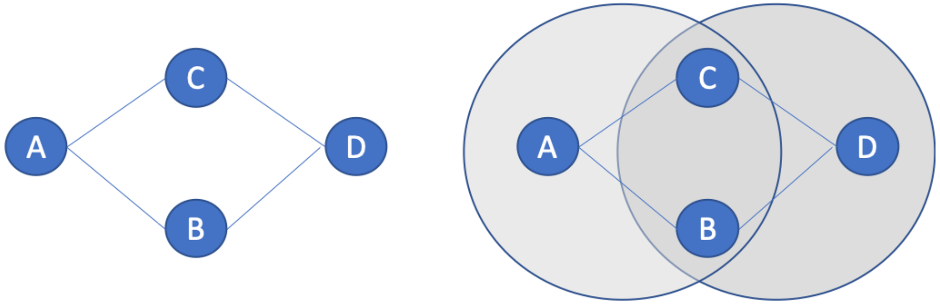
- Measures the cosine of the angle between two vectors, indicating their similarity.
- Example: Used in text analysis to compare document similarity.
- **Formula:**

  $$\cos(\theta)=\frac{A\cdot B}{||A||||B||}$$



Jaccard Similarity

- Measures the similarity between two sets based on their intersection and union.
- Example: Used in clustering and duplicate detection.
- **Formula:**
  J(A,B)= $\frac{|A \cap B|}{|A \cup B|}$



(a) (b)

Pearson Correlation Coefficient

- Measures the strength and direction of a linear relationship between two variables.
- Example: Used to analyze the relationship between study time and exam scores.
- **Formula:**

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

Dot Product

- Measures the similarity between two vectors in terms of their magnitude and direction.
- Example: Used in computer graphics and physics.
- **Formula:**

$$a \cdot b = \sum_{i=1}^{n} a_i b_i$$

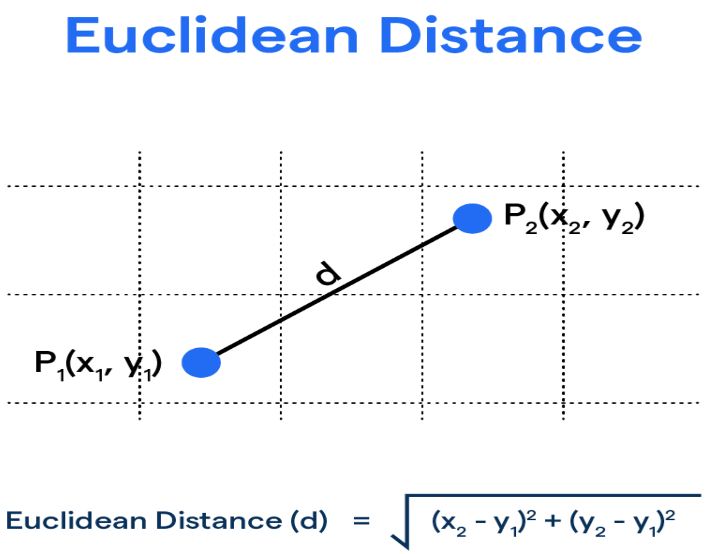
- **If the dot product is 0**, the vectors are orthogonal (perpendicular) to each other.
- **If the dot product is positive**, the vectors point in the same direction.
- **If the dot product is negative**, the vectors point in opposite directions.

Euclidean distance

- It is a measure of the straight-line distance between two points in a multi-dimensional space.
- It is widely used in clustering, classification, and other machine learning algorithms to measure the similarity or dissimilarity between data points.

Formula:

$$d(\mathbf{p}, \mathbf{q}) = \sqrt{\sum_{i=1}^{n} (q_i - p_i)^2}$$



Coin 1 / Coin 2	H (Head)	T (Tail)	Row Total
H (Head)	1	1	**2**
T (Tail)	1	1	**2**
Column Total	**2**	**2**	**4**