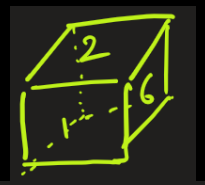
Probability Distribution Functions (PDF)

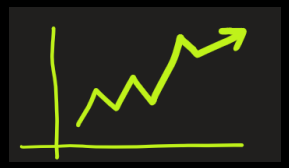
PDF are used for

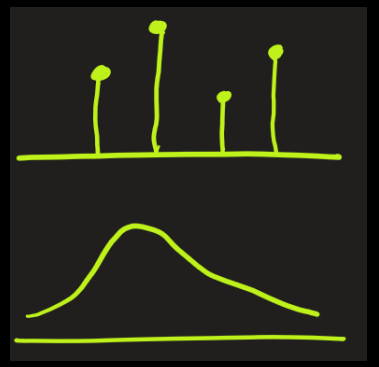
- Predicting number when rolling a dice



- Predicting tomorrow's weather

- Predicting stock market fluctuations





In AI and ML, PDFs appear whenever we predict discrete categories

- like email is spam vs not spam, or picture of cat vs dog.

- SoftMax output of a neural network is actually a PDF that shows the probability of each class.

Here we will understand what a probability distribution function is, why it's important, its types, and some real-life applications.

Probability Distribution Functions (PDF)

A Probability Distribution Function is a **function that describes how the probabilities of a random variable are distributed**. It tells us how likely each outcome is.

Few terms that are used in this domain.

Random variable could something whose outcome is uncertain (like the roll of a dice, test scores, or temperature). It is generally denoted by capital letter X.

Probability is the chance of a particular outcome happening

A simple example:

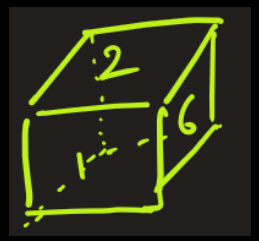
Suppose you roll a fair 6-sided dice.

Here the random variable is 1, 2, 3, 4, 5, or 6

The probability of each face (1 to 6) is 1/6.

So, the PDF is

P(X = x) = 1/6 , where x can be 1, 2, 3, 4, 5, or 6



Outcome (X)	Probability
1	1/6
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6
Total	1

2 Types of Probability Distribution Functions: Discrete and Continuous

1. Discrete Probability Distribution

Another name for Discrete PD is Probability Mass Function (PMF). A **PMF** is a function that gives the **probability of each possible discrete outcome** of a random variable. It is used for **discrete random variables** where outcomes are **countable**: like dice, coins, or number of students in a class.

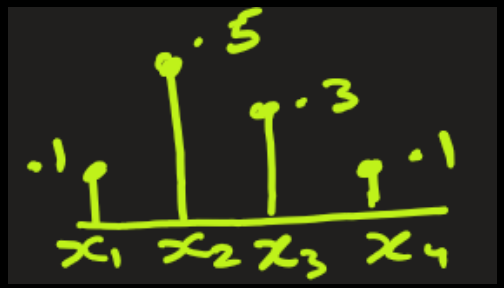
PMF is often written as $P(X = x)$,which means "the probability that random variable X takes the value x."

The 2 key points of PMF are

1) **Non-negative:** $P(X = x) \geq 0$ for all x

2) **Sum of probabilities = 1:**

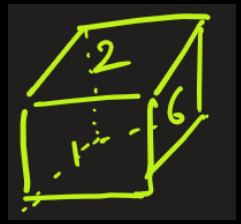
$$\sum_{x} P\,(X = x) = 1$$



2 Types of Probability Distribution Functions: Discrete and Continuous

1. Discrete Probability Distribution

Example: Let's take an experiment of rolling a **fair** dice with the random variable X being the face with odd number that showing up. What is PDF of X ?

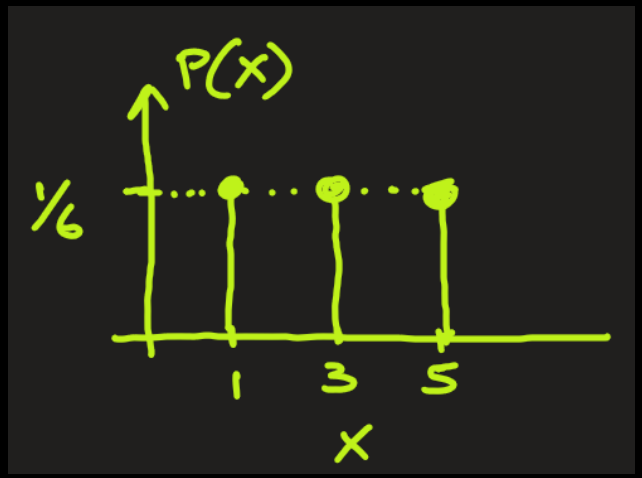


Ans:
Here X = {1, 3, 5}. So,

P(X=1) = 1/6

P(X=3) = 1/6

P(X=5) = 1/6



2 Types of Probability Distribution Functions: Discrete and Continuous

1. Discrete Probability Distribution

Example: A dataset containing the number of emails received each day for a week is provided. Let X be a random variable representing spam email received on any given day. Assuming the data allows for the calculation of the probability of determining whether an email is spam, what is the probability mass function (PMF) of X ?

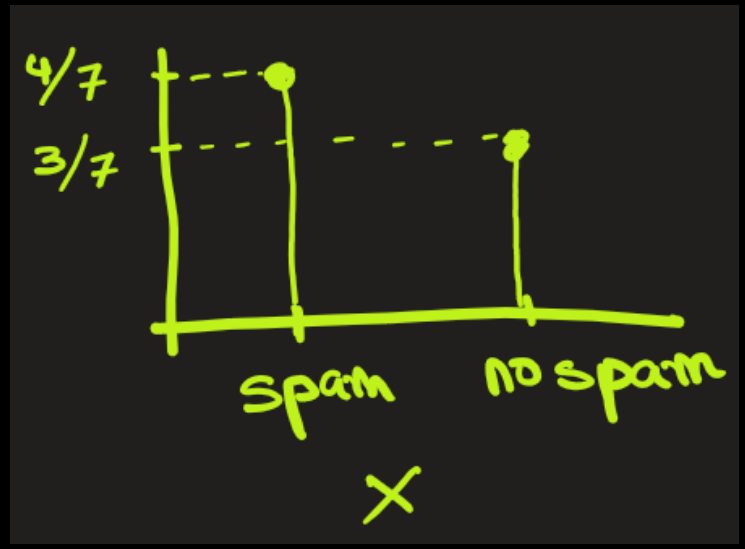
email	Is_spam
E1	Y
E2	Y
E3	Y
E4	N
E5	N
E6	Y
E7	N

Ans:
Here X is a random variable indicating whether new email is spam or not.
X = {spam, no-spam} .

By looking at data, we see

P(X=spam) = 4/7

P(X = no-spam) = 3/7



2 Types of Probability Distribution Functions: Discrete and Continuous

1. Discrete Probability Distribution

Example: Suppose a product has been rated by customers using discrete rating values, and let X be the random variable that denotes the rating given to the product. What is the PMF of X?

Ratings	Count
5-star	15 times
4-star	10 times
3-star	10 times
2-star	25 times
1-star	40 times

Ans:
Here X is a random variable indicating the ratings. X = { 1, 2, 3, 4, 5} .
By looking at data, we see that probability that you randomly pick a 1-star rating is
P(X=1) = 40/100 = 0.40

P(X =2) = 0.25

P(X =3) = 0.10

P(X =4) = 0.10

P(X =5) = 0.15