SNS COLLEGE OF TECHNOLOGY

An Autonomous Institution
Coimbatore-35

DEPARTMENT OF ARTIFICIAL INTELLIGENCE & DATA SCIENCE

23ADT202 – FUNDAMENTALS OF DATA SCIENCE AND ANALYTICS

II YEAR IV SEM

UNIT II – INTERPRETING DISTRIBUTIONS

EMPATHY:

What is a Probability Distribution, a Discrete Random Variable, and a Probability Distribution Table:

Probability Distribution for a Discrete Random Variable:

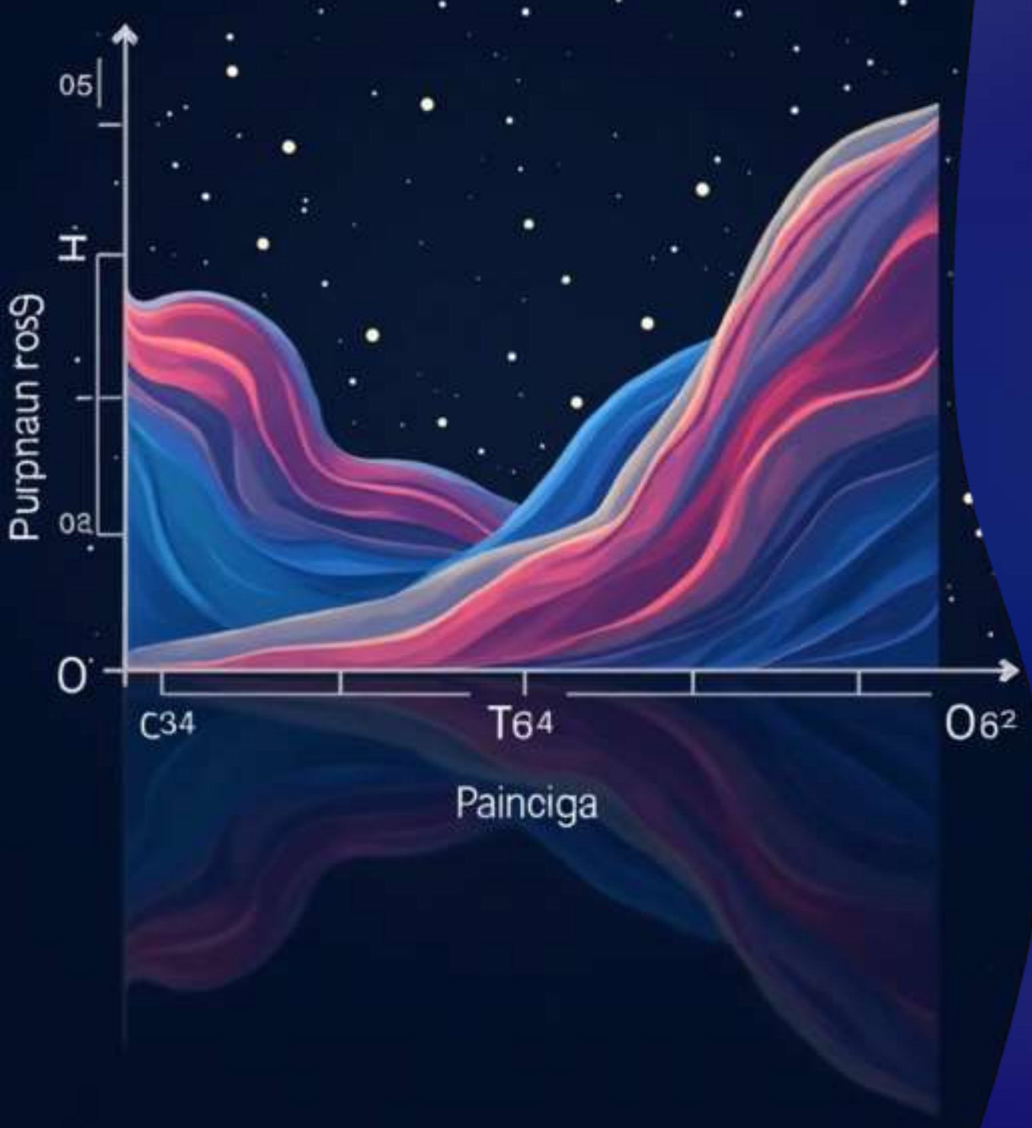
A probability distribution shows the probability of different values for a discrete random variable or, put another way, the probability of different outcomes occurring.

Discrete Random Variable:

A discrete random variable is a variable that can take on a countable set of numbers or values. Often, you'll see integers as the variable values.

Probability Distribution Table:

A probability distribution table is a table that shows the different probabilities of a random discrete variable having different outcomes.



DEFINE:

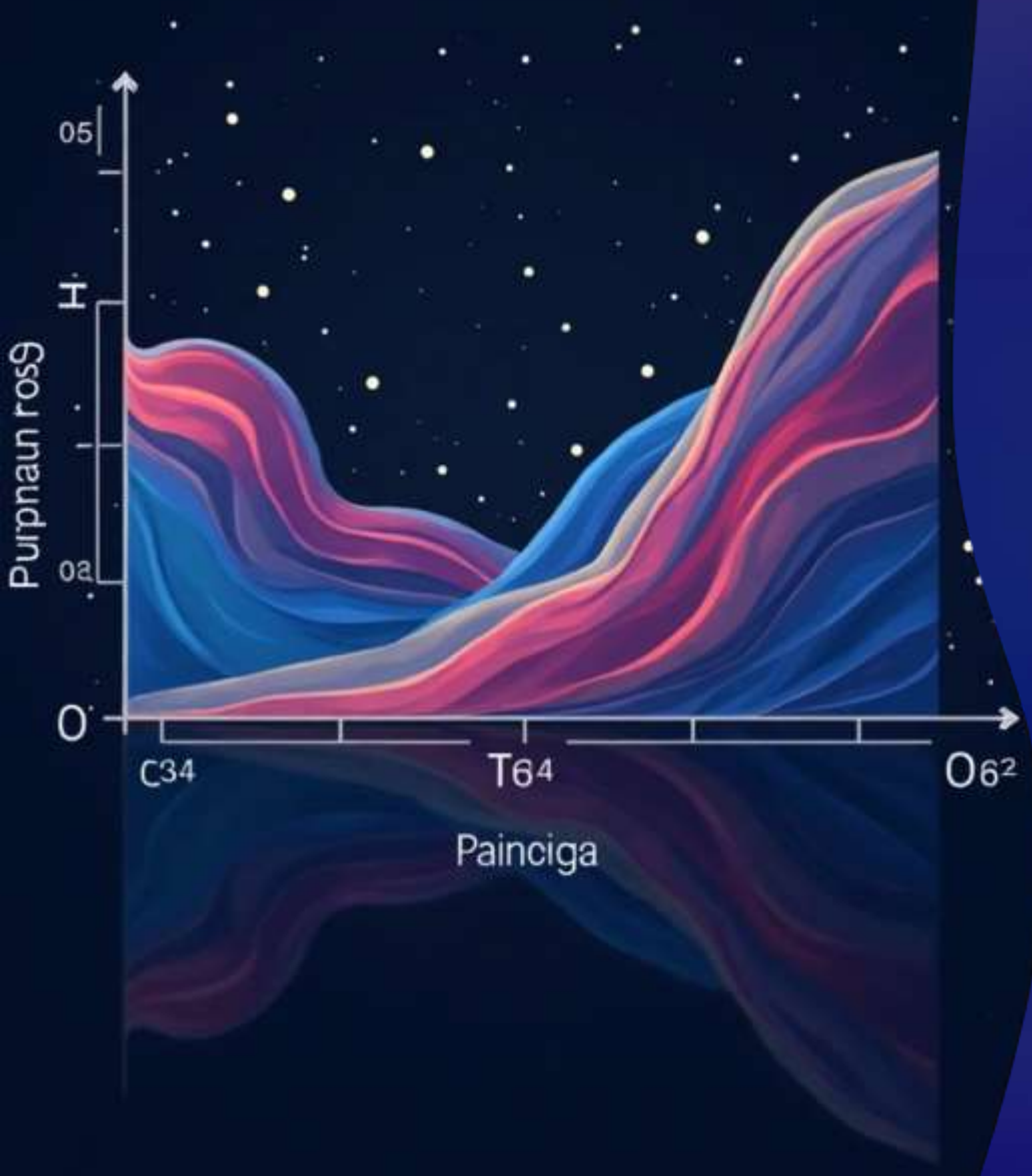
How to Interpret a Probability Distribution

Step 1: A probability distribution table for a discrete random variable has a few properties that can help us interpret it.

The first is that, in general, the first column on the left will be the x variable or the different outcomes.

The column to the right will then contain the probability that each of the outcomes will occur.

How to Interpret a Probability Distribution



Step 2: A probability distribution should show all of the probabilities of each outcome.

Therefore, these probabilities should add up to 1.

How to Interpret a Probability Distribution

IDEATE:

Example: Tossing a Fair Coin Twice

X (Number of Heads)	Probability P(X)
0	1/4
1	1/2
2	1/4

How to Interpret a Probability Distribution

PROTOTYPING:

Example: Tossing a Fair Coin Twice

X = 0:

Getting no heads (TT). There is 1 favorable outcome out of 4 possible outcomes, so the probability is 1/4.

X = 1:

Getting exactly one head (HT or TH). There are 2 favorable outcomes out of 4, so the probability is 2/4 = 1/2.

X = 2:

Getting two heads (HH). There is 1 favorable outcome out of 4, so the probability is 1/4.

How to Interpret a Probability Distribution

Example: Tossing a Fair Coin Twice

Important Points:

All probabilities are between 0 and 1.

The sum of all probabilities equals 1, which confirms it is a valid probability distribution: 1/4+1/2+1/4=1

This table clearly shows how likely each possible number of heads is when tossing a fair coin twice.

How to Interpret a Probability Distribution?

Step 3: From a probability distribution table, we can see which outcome has the highest likelihood (i.e. has the highest probability) and which has the lowest likelihood (or lowest probability) of happening.

Step 4: Because we're discussing distributions for discrete random variables, we can calculate the mean, or average, of the distribution.

We can do this by adding up the product of each variable value times their corresponding probability.

The formula of this would be: $\mu = \Sigma x P(x)$

Examples of Interpreting a Probability Distribution:

Look at the following probability distribution table. This hypothetically shows the probability that a random person selected, who's between the ages of 15 and 35, will own a specific number of computers.

Number of Computers Owned (x)	P(x)
0	0.04
1	0.25
2	0.38
3	
4	0.09
5	0.03

TESTING:

Step 1:

We see that x is 0-5, which represents the number of computers owned by a random individual.

The probability of each outcome is then on the right, as we would expect.

Step 2:

We know that the sum of the probabilities should be equal to 1. But, we are missing the probability corresponding to 3 computers owned.

To find this, we can take 1 and subtract all other values from it.

$1 - 0.04 - 0.25 - 0.38 - 0.09 - 0.03 = 0.21$

Examples of Interpreting a Probability Distribution:

So, the table fully filled out would look like this:

Number of Computers Owned (x)	P(x)
0	0.04
1	0.25
2	0.38
3	0.21
4	0.09
5	0.03

Examples of Interpreting a Probability Distribution:

Step 3: From the table, we can see that, if we picked someone random who's between 15 and 35, there is the biggest likelihood that they will own 2 computers because 0.38 is the largest probability.

Similarly, 0.03 is the smallest probability, so there's the least likelihood that they will own 5 computers.

Step 4: To calculate the **average number of computers** owned by this age group, we can use the formula:

$$\mu = \Sigma x P(x) = 0 \cdot 0.04 + 1 \cdot 0.25 + 2 \cdot 0.38 + 3 \cdot 0.21 + 4 \cdot 0.09 + 5 \cdot 0.03 = 0 + 0.25 + 0.76 + 0.63 + 0.39 + 0.15 = 2.18$$

So, the mean or the average number of computers owned by 15-35 years old, when given this example, would be 2.18 computers.

$$\text{Mean (Average)} = \frac{\text{Sum of all observations}}{\text{Number of observations}}$$