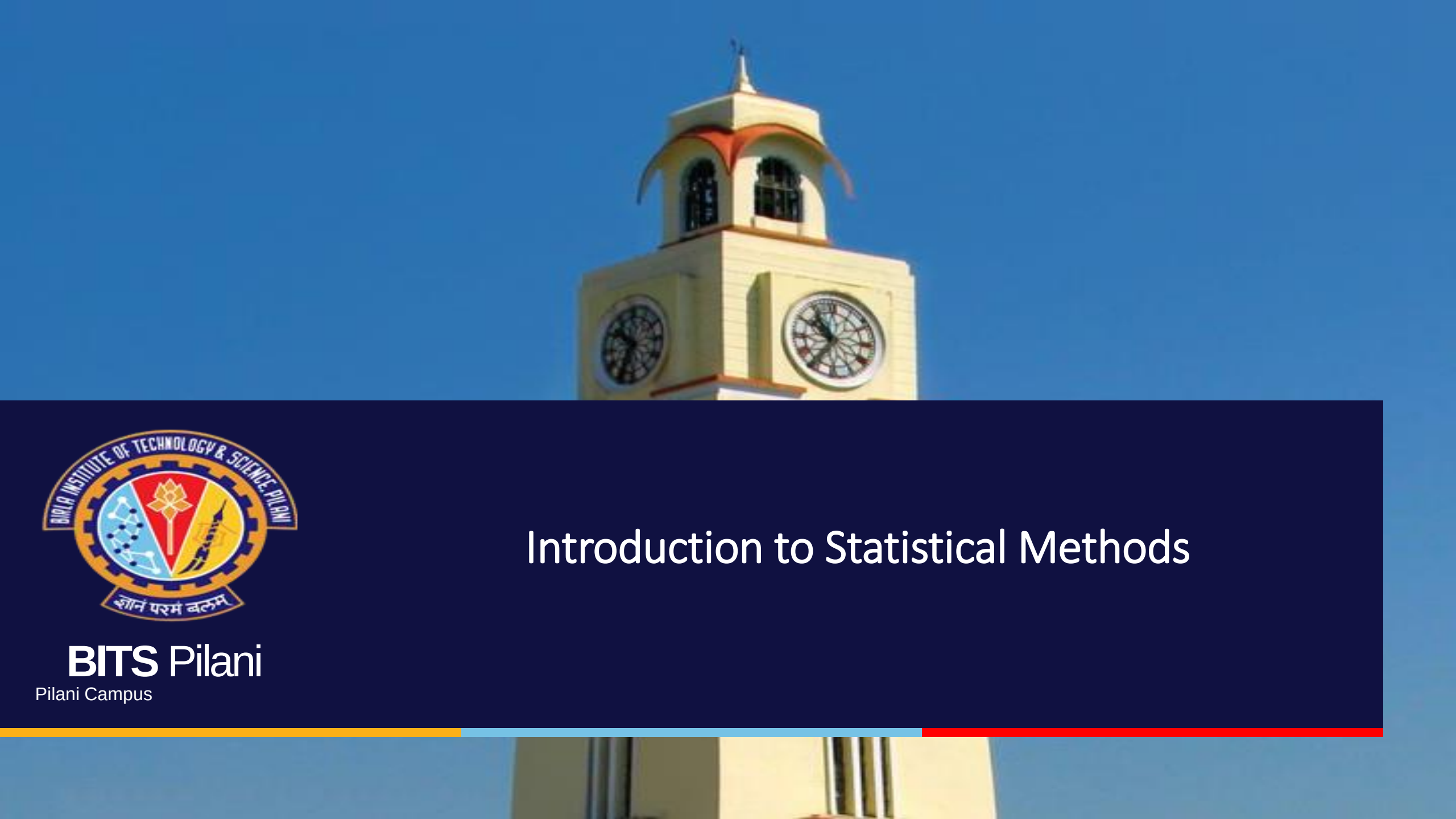

Introduction to Statistical Methods

BITS Pilani

Pilani Campus

innovate achieve lead

BITS Pilani

Pilani Campus

Course No: DSECL ZC413 / AIML ZC418
Course Title: ISM
WEBINAR 2 : 26/12/2024

Topics:

- Bayes' Theorem
- Naïve Bayes Classifier
- Random Variables

BAYES' THEOREM:

Suppose that $E_1, E_2, \text{---------} E_n$ are mutually exclusive and exhaustive events of a sample space " S "

such that $P(E_i) > 0$ for $i = 1,2,3, \text{---------} n$ and A is any arbitrary event of " S "

such that $P(A) > 0$ and $A \subseteq \bigcup_{i=1}^{n} E_i$ then the conditional probability of E_i given A

for $i = 1,2,3 \ldots\ldots\ldots n$ is given by $P(E_i/A) = \dfrac{P(E_i)P(A/E_i)}{\sum_{i}^{n} P(E_i)P(A/E_i)}$

Bayes' theorem is also known as the formula for the **probability of "causes"**. **As we know, the E_i's are a partition of the sample space S, and at any given time only one of the events E_i occurs.** Thus, we conclude that the Bayes' theorem formula gives the probability of a particular E_i, given the event A has occurred.

Terms related to Bayes' Theorem

- **Hypotheses:** Events happening in the sample space **E_1, E_2,… E_n is called the hypotheses**

- **Priori Probability:** Priori Probability is the initial probability of an event occurring before any new data is taken into account. $P(E_i)$ is the priori probability of hypothesis E_i.

- **Posterior Probability:** Posterior Probability is the updated probability of an event after considering new information. Probability $P(E_i|A)$ is considered as the posterior probability of hypothesis E_i.

Applications of Bayes' Theorem

Bayesian inference is very important and has found application in various activities, including medicine, science, philosophy, engineering, sports, law, etc., and Bayesian inference is directly derived from Bayes' theorem.

Example: Bayes' theorem defines the accuracy of the medical test by taking into account how likely a person is to have a disease and what is the overall accuracy of the test.

Problem 1

A hospital has 4 nurses handling respectively 20%, 40%, 25% and 15% of the patients of all the doctors coming to the hospital. The probability that they misguide the patients are respectively 0.05, 0.03, 0.04 and 0.05. Find the probability that the misguided incident can be blamed on the third nurse.

Problem 2

Three machines A, B and C produce respectively 50%, 30% and 20% of the total number of items of a factory. The percentage of defective outputs of these machines are 3%, 5% and 2%. An item is selected at random and is found to be defective.

(i) Find the probability that the item was produced by machine C?

(ii) What is the probability that the item was produced by machine C or B?

(iii) What is the probability that an item selected at random is found to be non-defective and also the probability that the item was produced by machine A given that it is non-defective.

Problem 3

EC2R
JULY 2024

innovate achieve lead

Hardik Pandya, Rishabh Pant and Surya Kumar Yadav are in the race of leading Indian cricket team in the next world cup with probabilities 0.2, 0.5 and 0.3 respectively. The probabilities of getting an increase in the match fee by Hardik, Rishabh and Surya are 0.3, 0.6 and 0.5 respectively if they become the Captain.

If there is an increase in match fee then find the probability

I. that it is because of Hardhik
II. that it is because of Rishabh
III. that it is because of Surya Kumar

Problem 4

EC2R
JAN 2024

innovate achieve lead

In a manufacturing company Machine 1 produces 40% of the items and Machine 2 and Machine 3 produces 25% and 35% of the items respectively. But from the past records it is found that 15%, 20% and 25% of the items they produce are defective. Then find

i) Total percentage of defective produced

ii) If a defective item is selected randomly, then find the probability that it is produced by Machine 1

iii) If a defective item is selected randomly, then find the probability that it is not produced by Machine 2

Problem 5

EC2M
JAN 2024

innovate achieve lead

A product is manufactured by four companies A, B,C and D. They produce 20%, 30%, 40% and 10% of the products released in the market. But it is observed that 30%, 20%, 10% and 25% are defective items manufactured by them respectively. Then find

i)Total percentage of defective items manufactured

ii)The percentage that the selected defective item is produced by company A

iii)The percentage that the selected defective item is not produced by company B

iv)The percentage that the selected defective item is not produced by company D

Naïve Bayes Classifier

- Naïve Bayes classifiers are a collection of algorithms based on Bayes' Theorem. It is not a single algorithm but a family of algorithms where all of them share a common principle, i.e., every pair of features being classified as independent of each other.
- Naïve Bayes classifier is one of the most simple and effective classification algorithms, that aids in the rapid development of machine learning models with rapid prediction capabilities.
- It is highly used in text classification tasks where data contains high dimension (as each word represent one feature in the data). It is fast and making prediction easy with high dimension data.
- It is a probabilistic classifier, that predicts the probability of an instance belongs to a class with a given set of feature value. It uses Bayes theorem in the algorithm for training and prediction.

Why is it called Naïve Bayes?

- The “Naïve” part of the name indicates the simplifying assumption made by the Naïve Bayes Classifier. The classifier assumes that the features used to describe an observation are conditionally independent, given the class label.

Assumptions of Naïve Bayes’

The fundamental Naive Bayes assumption is that each feature makes an:

- **Feature independence:** The features of the data are conditionally independent of each other, given the class label.
- **Continuous features are normally distributed:** If a feature is continuous, then it is assumed to be normally distributed within each class.
- **Discrete features have multinomial distributions:** If a feature is discrete, then it is assumed to have a multinomial distribution within each class.
- **Features are equally important:** All features are assumed to contribute equally to the prediction of the class label.
- **No missing data:** The data should not contain any missing values.

Types of Naïve Bayes Model

- **Gaussian Naïve Bayes Classifier :** The Gaussian Naive Bayes classifier assumes that the attributes of a dataset have a normal distribution. Here, if the attributes have continuous values, the classification model assumes that the values are sampled from a Gaussian distribution.
- **Multinomial Naïve Bayes:** When the input data is multinomially distributed, we use the multinomial naive Bayes classifier. This algorithm is primarily used for document classification problems like sentiment analysis.
- **Bernoulli Naïve Bayes:** The Bernoulli Naive Bayes classification works in a similar manner to the multinomial classification. The difference is that the attributes of the dataset contain boolean values representing the presence or absence of a particular attribute in a data point.

Advantages of Naïve Bayes Classifier

- Easy to implement and computationally efficient.
- Effective in cases with a large number of features.
- Performs well even with limited training data.
- It performs well in the presence of categorical features.
- For numerical features data is assumed to come from normal distributions

Disadvantages of Naive Bayes Classifier

- Assumes that features are independent, which may not always hold in real-world data.
- Can be influenced by irrelevant attributes.
- May assign zero probability to unseen events, leading to poor generalization.

Implementation of Naïve Bayes classifier

To implement a Naive Bayes classifier, we perform three steps.

- First, we calculate the probability of each class label in the training dataset.
- Next, we calculate the conditional probability of each attribute of the training data for each class label given in the training data.
- Finally, we use the Bayes theorem and the calculated probabilities to predict class labels for new data points. For this, we will calculate the probability of the new data point belonging to each class. The class with which we get the maximum probability is assigned to the new data point.

Problem 6

Consider the following data set:

Find out whether the object with attribute **Confident = Yes, Sick = No** will Fail or Pass using Bayesian Classification.

Confident	Studied	Sick	Result
Yes	No	No	Fail
Yes	No	Yes	Pass
No	Yes	Yes	Fail
No	Yes	No	Pass
Yes	Yes	Yes	Pass

Data set for classification

Problem 7

Example No.	Color	Type	Origin	Stolen?
1	Red	Sports	Domestic	Yes
2	Red	Sports	Domestic	No
3	Red	Sports	Domestic	Yes
4	Yellow	Sports	Domestic	No
5	Yellow	Sports	Imported	Yes
6	Yellow	SUV	Imported	No
7	Yellow	SUV	Imported	Yes
8	Yellow	SUV	Domestic	No
9	Red	SUV	Imported	No
10	Red	Sports	Imported	Yes

$$New\ Instance = (Red, SUV, Domestic) \rightarrow (Yes\ or\ No)$$

Problem 8

EC2M
JULY 2024

innovate achieve lead

Vehicle	Credit card/Debit card	Travel	Employment
BMW	Credit card	Flight	Business
Scooter	Debit card	Bus	Salaried
Bicycle	None	Local train	Business
Verna	Credit card	Vande Bharat	Salaried
Scooter	Credit card	Flight	Salaried
Auto	None	Bus	Business
BMW	Debit card	Vande Bharat	Salaried
No	No	Bus	Salaried

Based on the data is it possible to find whether the following is salaried or business person. If possible find with the help of suitable theorem / algorithm. Show the working, mention assumptions if any.

"Vehicle = scooter; Credit Card = yes, Travel = Bus"

Random Variables

Motivation:

In most statistical problems we are concerned with one number or a few numbers that are associated with the outcomes of experiments.

When inspecting a manufactured product we may be interested only in the number of defectives; in the analysis of a road test we may be interested only in the average speed and the average fuel consumption; and in the study of the performance of a miniature rechargeable battery we may be interested only in its power and life length. All these numbers are associated with situations involving an element of chance—in other words, they are values of random variables.

Random Variables

A random variable is a real valued function defined on the sample space

A **random variable** consists of an experiment with a probability measure P[·] defined on a sample space S and a function that assigns a real number to each outcome in the sample space of the experiment.

Advantages:

- When the elements of the sample space are non-numeric, they can be quantified by assigning a real number to every event of the sample space. A real number X associated with the outcome of a random experiment.

- Infinite sample space can be converted to finite

Random Variables – Types:

Random variables are usually classified according to the number of values they can assume. Random variable can be classified as

1) discrete random variables, which can take on only a finite number
2) an uncountable infinity of values; called continuous random variables

Examples:

- **Number of accidents on a highway – Discrete**
- **Weights of the students of the class – Continuous**

Discrete Random Variables – Problem 9

EC2M July 2022

A random variable X can assume five values:0,1,2,3,4. A portion of the probability distribution is shown here:

X	0	1	2	3	4
P(x)	0.1	0.25	0.25	?	0.2

i) Find p(3)

ii) Calculate the mean

iii) What is the probability that X is greater than 3?

iv) What is the probability that X is 3 or less?

v) What is the probability that X is in between 1 and 3?

Discrete Random Variables – Problem 10

APR EC2 22

The probabilities that number of times in a day certain computer may malfunction is as follows:

No. of malfunctions in a day	0	1	2	3	4	5	6
Probability	0.17	k	0.27	0.16	0.07	0.03	0.01

Find

(a) the value of k

(b) mean and variance of the malfunction of certain computer

Random Variable – Probability Distribution - Problem 11

A student takes two courses. In each course, the student will earn a B with probability 0.6 or a C with probability 0.4, independent of the other course. To calculate a grade point average (GPA), a B is worth 3 points and a C is worth 2 points. The student's GPA is the sum of the GPA for each course divided by 2. Make a table of the sample space of the experiment and the corresponding values of the student's GPA, G.

The sample space, probabilities and corresponding grades for the experiment are

Outcome	$P[\cdot]$	G
BB	0.36	3.0
BC	0.24	2.5
CB	0.24	2.5
CC	0.16	2


THANK YOU