

VISVESVARAYA TECHNOLOGICAL UNIVERSITY

Jnana Sangama, Belagavi- 590018



Report on

“MATHEMATICS 3 FOR CSE STREAM-BCS301”

Submitted in partial fulfilment of the requirements for the award of degree of

BACHELOR OF ENGINEERING

In

**COMPUTER SCIENCE AND
ENGINEERING**

SUBMITTED BY: DARSHAN J P

1MV23CS047

Under the Guidance of

LATHA Y L

Assistant Professor
Department of CSE



DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

SIR M VISVESVARAYA INSTITUTE OF TECHNOLOGY

BANGALORE-562157

1 : STOCHASTIC MATRIX

What is Stochastic Matrix?

A **stochastic matrix** is a square matrix used in probability theory where each entry represents a probability. The rows of the matrix sum to 1, indicating a probability distribution. It's commonly used in Markov chains to describe transitions between states.

Key Features:

- **Non-Negative Entries:** All elements in a stochastic matrix are non-negative, as they represent probabilities.

Row Sums Equal 1: Each row of the matrix sums to 1, ensuring a valid probability distribution.

Square Matrix: It is typically a square matrix, meaning the number of rows equals the number of columns, representing transitions between a finite set of states.

ANDREY MARKOV

Stochastic matrices were not explicitly "invented" by a single individual but are a key mathematical concept that emerged as part of probability theory and Markov processes. **Andrey Markov**, a Russian mathematician, is closely associated with their development due to his work on **Markov chains** in the early 20th century. He introduced the foundational ideas of state transitions and probabilities, which naturally led to the use of stochastic matrices to represent these systems .

REPRESENTATION OF STOCHASTIC MATRIX

Key Features in this Example:

1. **Non-Negative Entries:** All elements are ≥ 0 .
2. **Rows Sum to 1:**
 - Row 1: $0.5+0.3+0.2=10.5 + 0.3 + 0.2 = 10.5+0.3+0.2=1$
 - Row 2: $0.1+0.7+0.2=10.1 + 0.7 + 0.2 = 10.1+0.7+0.2=1$

Row 3: $0.4+0.4+0.2=10.4 + 0.4 + 0.2 = 10.4+0.4+0.2=1$

This matrix can represent transitions in a Markov chain, where P_{ij} is the probability of moving from state i to state j .

Applications of Stochastic Matrix

Stochastic Matrix is widely used in:

- Markov Chains
- Google PageRank Algorithm
- Biological and Medical Studies- Cryptography and Network Security

Example Problem:

If the weather has a 70% chance of staying sunny and a 30% chance of switching to rain, represent this as a stochastic matrix and find the probability of it being sunny tomorrow if it is sunny today.

Solution:

1. Stochastic Matrix:

$$P = \begin{bmatrix} 0.7 & 0.3 \end{bmatrix}$$

1. Initial State (Sunny Today):

$$x_0 = \begin{bmatrix} 1 & 0 \end{bmatrix}$$

$$1. \text{ Next Day's Probability: } x_1 = x_0 \cdot P = \begin{bmatrix} 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} 0.7 & 0.3 \end{bmatrix} = \begin{bmatrix} 0.7 & 0.3 \end{bmatrix}$$

Answer:

There is a **70% chance of sunny** and **30% chance of rain** tomorrow.

2 : REGULAR STOCHASTIC MATRIX

WHAT IS REGULAR STOCHASTIC MATRIX ?

A **regular stochastic matrix** is a square stochastic matrix where, after a sufficient number of steps, all entries in the matrix become strictly positive. This means that there exists some power k such that P^k (the matrix raised to the k -th power) has no zero entries.

ANDREY MARKOV

The concept of a **regular stochastic matrix** is rooted in the broader theory of **Markov chains**, which was developed by **Andrey Markov** in the early 20th century. However, the specific classification and terminology, including "regular stochastic matrices," were formalized later as part of the development of linear algebra and probability theory.

Contributions from mathematicians working on Markov processes and matrix theory, such as **David George Kendall** and others in the mid-20th century, helped refine and popularize these concepts.

No single individual is credited with coining the term "regular stochastic matrix," but its development is tied to the broader evolution of Markov chain theory.

Key Points

1. **Strictly Positive Powers:** A stochastic matrix P is regular if there exists some power k such that all entries of P^k are strictly positive (i.e., >0).
2. **Steady-State Convergence:** Regular stochastic matrices guarantee convergence to a unique steady-state probability distribution, regardless of the initial state.
3. **Irreducibility and Aperiodicity:** Regular stochastic matrices are irreducible (every state can reach every other state) and aperiodic (no cycles restrict transitions)

Example Problem

Convert the following **stochastic matrix** with some zero entries into a **regular stochastic matrix**: $P =$

Solution:

$$P^2 = P \times P =$$

$$P^3 = P^2 \times P =$$

Answer: The original matrix P is not regular because it contains a zero entry. By modifying it to P^3 , we obtained a regular stochastic matrix, as P^3 contains no zero entries.

Applications

1. Queueing Theory
2. Population Dynamics in Biology
3. Finance and Economics
4. Reinforcement Learning

3 : Chi-Square Test

Introduction

The **Chi-Square Test** is a statistical method used to determine if there is a significant association between categorical variables. It compares the observed

frequencies of events or outcomes to the frequencies that would be expected under the null hypothesis (no relationship between the variables).

KARL PEARSON

The **Chi-Square Test** was introduced by the British statistician **Karl Pearson** in 1900. Pearson developed the test as a method for assessing the goodness of fit between observed data and a theoretical model, specifically to test whether a given distribution of categorical data fits an expected distribution.

His work on the Chi-Square Test was published in his paper "On the Criterion that a Given System of Deviations from the

Probable in the Case of a Correlated System of Variates is Least Likely to Have Arisen from Random Sampling," in the journal *Philosophical Magazine*.

Pearson's development of the test was a significant contribution to the field of statistics, and it remains one of the most widely used statistical tests today.

There are two common types of Chi-Square tests:

1. Chi-Square Goodness of Fit Test:

- This test assesses how well an observed distribution of a single
- categorical variable fits an expected distribution. Example: Testing whether a die is fair by comparing the observed frequency of each face of the die to the expected frequency.

2. Chi-Square Test of Independence:

- This test evaluates whether two categorical variables are independent or
- related. Example: Testing whether gender is related to the choice of a particular brand

Steps in performing a Chi-Square Test:

1. State the Hypotheses:

- Null Hypothesis (H_0): Assumes no association or difference (e.g., the
- variables are independent). Alternative Hypothesis (H_1): Assumes an association or difference.

2. Calculate the Chi-Square Statistic: The formula for the test statistic is:

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

where: O = Observed frequency

 E = Expected frequency (calculated based on the null hypothesis)

3. Determine the Degrees of Freedom:

- For the Goodness of Fit Test: $df = \text{number of categories} - 1$

For the Test of Independence: $df = (r-1) \times (c-1)$, where r is the number of rows and c is the number of columns in the contingency table.

4. Find the p-value:

- The p-value is compared to the significance level (α , typically 0.05) to determine whether to reject or fail to reject the null hypothesis.

5. Conclusion:

- If the p-value is less than α , reject the null hypothesis, indicating that there is a significant association between the variables.

If the p-value is greater than α , fail to reject the null hypothesis, indicating no significant association.

Assumptions:

- Data should be categorical (nominal or ordinal).
- The observations should be independent.

The expected frequency for each cell should be at least 5 for reliable results.

The Chi-Square test is widely used in various fields such as biology, social sciences, and marketing research.

APPLICATIONS :

- **Goodness of fit:** Testing if an observed distribution matches an expected distribution.

- **Independence:** Testing if two categorical variables are independent or related.

- **Survey data:** Analysing relationships between demographic characteristics and survey responses.

Market research: Understanding customer preferences and behaviour in relation to demographic variables.

Genetics: Testing inheritance patterns or allele frequencies.

Quality control: Assessing product defects or manufacturing consistency.

Social sciences: Exploring relationships between behaviours and demographic factors.

Health: Investigating associations between health conditions and risk factors.

4 : ONE WAY ANOVA

WHAT IS ONE WAY ANOVA ?

One-Way ANOVA (Analysis of Variance) is a statistical test used to compare the means of three or more independent groups to determine if there is a statistically significant difference between them. It helps to assess whether at least one of the group means is different from the others.

RONALD A.FISHER

The **One-Way ANOVA** (Analysis of Variance) was introduced by the British statistician **Ronald A. Fisher** in the early 20th century. Fisher developed ANOVA as a method for analysing experimental data, particularly for agricultural experiments, to determine whether different treatments or factors led to significant differences in outcomes.

He introduced the technique in his 1925 book, *"Statistical Methods for Research Workers,"* and later expanded on it in his 1935 book *"The Design of Experiments."* Fisher's work laid the foundation for modern statistical analysis, and ANOVA became a crucial tool for testing hypotheses in many fields, including agriculture, psychology, biology, and social sciences.

Fisher's development of ANOVA helped researchers analyse and interpret data more effectively, allowing them to test multiple groups simultaneously rather than relying on pairwise comparisons.

Key Points:

- One-Way: Refers to the presence of only one independent variable or factor, which divides the data into different groups.

ANOVA: Refers to the process of analysing the variance within and between groups to assess differences in their means.

Hypotheses:

- Null Hypothesis (H_0): Assumes that all group means are equal (no significant difference).

Alternative Hypothesis (H_1): Assumes that at least one group mean is different from the others.

Procedure:

1. Calculate the overall mean of all the data combined.
2. Calculate the variance within each group (how much the data points within each group vary).
3. Calculate the variance between the group means (how much the means of each group differ from the overall mean).
4. Compute the F-statistic:

$F = \text{Variance Between Groups} / \text{Variance Within Groups}$

The F-statistic compares the variability between group means to the variability within groups.

1. Interpret the F-statistic : If the F-statistic is significantly large, it indicates that there is more variability between groups than within groups, suggesting that at least one group mean is different.

Assumptions:

- Independence: Observations within and between groups should be independent.

Normality: Data within each group should be approximately normally distributed.

Homogeneity of Variance: The variances within each group should be approximately equal.

Example:

A researcher might use One-Way ANOVA to test whether there are significant differences in average test scores between students from three different teaching methods (e.g., traditional, online, and blended learning). The null hypothesis would be that the mean test scores for all three methods are equal, and the alternative hypothesis would suggest that at least one method leads to a different average score.

Summary:

- One-Way ANOVA compares the means of three or more groups.
- It tests if there is a significant difference between the group means.

It is based on the analysis of the variance within and between the groups.

APPLICATIONS :

- **Agriculture:** Comparing the effects of fertilizers, irrigation methods, etc.
- **Medicine:** Testing the effectiveness of different treatments or drugs.
- **Psychology:** Studying different behaviours or conditions based on experimental treatments.
-

Education: Comparing the effectiveness of various teaching methods.

Marketing: Analysing consumer preferences or brand effectiveness.

Manufacturing: Assessing quality control across production methods or machines.

Sports Science: Comparing athletic performance based on training regimens.

Environmental Science: Studying environmental factors and their impact on ecosystems.

5 : TWO WAY ANOVA

WHAT IS TWO WAY ANOVA ?

Two-Way ANOVA (Analysis of Variance) is a statistical test used to examine the influence of two independent categorical variables (factors) on a dependent continuous variable. It is an extension of the One-Way ANOVA and allows for the analysis of interactions between the two factors, as well as their individual effects on the outcome variable.

Key Points of Two-Way ANOVA:

1. Two Independent Variables (Factors):
 - The analysis involves two categorical independent variables (factors),
 - each with multiple levels (groups). For example, one factor might be "Diet" with levels (low-fat, high-fat), and the second factor might be "Exercise Type" with levels (aerobic, resistance).
2. Main Effects:
 - Main Effect of Factor 1: The effect of the first factor (e.g., "Diet") on the dependent variable, ignoring the second factor.
 -

Main Effect of Factor 2: The effect of the second factor (e.g., "Exercise Type") on the dependent variable, ignoring the first factor.

3. Interaction Effect:

- The interaction effect examines whether the effect of one factor depends on the level of the other factor. In other words, it tests if the combined effect of both factors is different from the sum of their individual effects.

Hypotheses:

- Null Hypothesis (H_0):
 - There is no main effect of factor 1 (i.e., no difference between the groups of factor 1).
 - There is no main effect of factor 2 (i.e., no difference between the groups of factor 2). There is no interaction effect between factor 1 and factor 2.
- Alternative Hypothesis (H_1):
 - There is a significant main effect of factor 1.
 - There is a significant main effect of factor 2.
 - There is a significant interaction effect between the two factors.

Procedure:

1. Calculate the means for each combination of the two factors.
2. Compute the variation between groups for each factor (main effects).
3. Compute the variation due to interaction between the two factors.
4. Calculate the F-statistics for each factor and the interaction:
 - F for Factor 1 (Main Effect of Factor 1)
 - F for Factor 2 (Main Effect of Factor 2)
 - F for the interaction between Factor 1 and Factor 2
5. Compare the F-statistics to critical values or calculate p-values to determine if the effects are statistically significant.

Example:

Imagine a study looking at the effects of two factors, "Diet" (low-fat vs. high-fat) and "Exercise Type" (aerobic vs. resistance), on weight loss. The Two-Way ANOVA would allow you to:

- Test whether diet type alone affects weight loss (main effect of "Diet").
- Test whether exercise type alone affects weight loss (main effect of "Exercise Type").

- Test whether the effect of diet type depends on the type of exercise (interaction effect).

Assumptions:

- Independence: Observations should be independent of each other.
- Normality: The dependent variable should be approximately normally distributed within each group. Homogeneity of Variance: The variances in each group should be approximately equal.

Summary:

- Two-Way ANOVA analyses the effects of two independent factors on a continuous dependent variable.
- It assesses main effects (individual effects of each factor) and interaction effects (whether the effect of one factor depends on the level of the other).
- It is commonly used in experimental designs with two categorical factors.