Course Code				**23STMAL231**			
Title of the Course				**Discrete Probability Distributions**			
Offered to: (Programme/s)				**B.Sc. Hons Statistics**			
L	**4**	**T**	**0**	**P**	**0**	**C**	**3**
Year of Introduction:	**2024-25**			**Semester:**		**3**	
Course Category:	**Major**			**Course Relates to:**	Local, Regional, National, Global		
Year of Revision:	**NA**			**Percentage:**	**NA**		
Type of the Course:				Skill development			
Crosscutting Issues of the Course :				**NA**			
Pre-requisites, if any				**Probability & Random Variable**			

NA : Not Applicable

Course Description:

Discrete probability distributions explore the fundamentals of probability theory with a focus on discrete random variables. Students will learn about key concepts such as probability mass functions, expected values, variance, and common distributions including binomial, Poisson, and geometric. The course emphasizes practical applications and problem-solving techniques, providing a solid foundation for analyzing and interpreting data in various fields. By the end of the course, students will be equipped to model real-world scenarios using discrete probability distributions and apply statistical methods to draw meaningful conclusions from data.

Course Aims and Objectives:

S.NO	COURSE OBJECTIVES
1	**Define and Describe:** Clearly define and describe key concepts related to discrete probability distributions, including probability mass functions and cumulative distribution functions.
2	**Identify and Use Distributions:** Identify and apply common discrete probability distributions, such as binomial, Poisson, and geometric, in various scenarios.
3	**Perform Calculations:** Accurately perform calculations involving probabilities, expected values, and variances for discrete random variables.
4	**Interpret Results:** Interpret the results of probability computations and understand their implications in real-world contexts.
5	**Apply Problem-Solving Techniques:** Utilize discrete probability distributions to solve practical problems, employing appropriate statistical methods and techniques.

Course Outcomes

At the end of the course, the student will be able to…

CO NO	COURSE OUTCOME	BTL	PO	PSO
CO1	explain the key characteristics and parameters of each distribution.	**K2**	**PO1**	**PSO1**
CO2	analyze the properties of each distribution, such as skewness, kurtosis, and mode..	**K4**	**PO1**	**PSO1**
CO3	explain the concepts and techniques of discrete probability distributions clearly and concisely.	**K2**	**PO1**	**PSO1**
CO4	apply discrete probability distributions to real-world problems and scenarios.	**K3**	**PO1**	**PSO1**
CO5	solve problems involving discrete probability distributions using various methods and techniques.	**K3**	**PO1**	**PSO1**

For BTL: K1: Remember; K2: Understand; K3: Apply; K4: Analyze; K5: Evaluate; K6: Create

CO-PO MATRIX

CO NO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PSO1	PSO2
CO1	3							2	
CO2	3							2	
CO3	3							2	
CO4	3							2	
CO5	3							2	

Use the codes 3,2,1 for High, Moderate and Low correlation Between CO-PO-PSO respectively

Course Structure:

Unit – 1: Uniform, Bernoulli and Binomial distributions (12Hrs)

Discrete Uniform distribution – definitions, mean, variance.

Bernoulli distribution – definitions, mean, variance and its moment generating function. Binomial distribution – Definition, Moments of Binomial Distribution, Mode of Binomial Distribution, Recurrence Relation for the moments of Binomial distribution, Mean deviation about Mean of Binomial distribution, Moment generating function, Characteristic function, Cumulants and Probability generating function of Binomial distribution. Additive property of Binomial distribution. Measure of skewness, kurtosis and problems. The first two moments are obtained through moment generating function, Recurrence relation for the probabilities of Binomial distribution.

Examples/Applications/Case Studies:

1. Rolling a fair die. Each number (1-6) has an equal probability of appearing.
2. Random number generation, simulations, load balancing in computer networks.
3. Quality control, medical trials, decision making.
4. Number of heads in 10-coin flips, number of defective items in a sample.

Exercises/Projects:

1. Simulation**:** Simulate the tossing of a fair coin 1000 times and compare the observed frequency of heads to the theoretical probability.
2. **Gambling:** Analyse the probability of winning different types of lottery games using binomial distribution.
3. **Sports Analytics:** Analyse the performance of a basketball player or baseball hitter using binomial distribution and calculate relevant statistics.
4. **Data Analysis:** Collect data on a binary outcome (e.g., success/failure, yes/no) and fit a binomial distribution to the data. Analyse the goodness of fit.

Specific Resources: (web)

Stat Trek: Offers online tutorials and calculators for statistical concepts.

Link: https://stattrek.com/

Unit – 2: Poisson Distribution (12Hrs)

Poisson distribution – Definition, Moments of Poisson distribution, Mode of Poisson distribution, Recurrence Relation for the moments of Poisson distribution, Measure of skewness, kurtosis and problems. Moment generating function, Characteristic function, Cumulants and Probability generating function of Poisson distribution. The first two moments are obtained through moment generating function. Additive or reproductive property of Poisson distribution. Recurrence relation for the probabilities of Poisson distribution. Poisson distribution as a limiting case of Binomial distribution.

Examples/Applications/Case Studies:

1. **Number of calls received by a call center in an hour.** The arrival of calls can be considered independent events occurring at a constant rate.
2. **Number of cars arriving at a toll booth in a minute.** Similar to call center example, car arrivals can be modelled as independent events with a constant rate.
3. **Case Study: Call Center Staffing**

   A call center wants to determine the optimal number of agents to handle incoming calls. By modelling the arrival of calls as a Poisson process, the call center can calculate the average number of calls per hour and the probability of different call volumes. This information can be used to optimize staffing levels and minimize customer wait times.

Exercises/Projects:

Project 1: Real-Life Application

Objective: Apply the Poisson distribution to model a real-life scenario.

1. **Choose a Scenario**: For instance, modeling the number of phone calls received by a customer service center per hour.
2. **Data Collection**: Collect data on the number of calls received over several hours.
3. **Model Fitting**: Fit a Poisson model to the data. Estimate the parameter λ\lambda and check if the model is a good fit using goodness-of-fit tests.
4. **Analysis**: Analyze the results. Compare the predicted probabilities (e.g., probability of receiving exactly 10 calls in an hour) with observed frequencies.

Project 2: Simulation and Visualization

Objective: Simulate Poisson random variables and visualize their properties.

1. **Simulation**: Write a program (using Python, R, etc.) to simulate Poisson random variables with different parameters.
2. **Visualization**:
   - Plot histograms of the simulated data for various λ values.
   - Show how the distribution changes with different λ.
   - Plot the theoretical PMF and compare it with the histogram from the simulation.
3. **Analysis**: Calculate and compare the empirical moments and cumulants with theoretical values. Verify the additive and reproductive properties through simulations.

Specific Resources: (web)

https://stattrek.com/probability-distributions/poisson.aspx

Unit – 3: Negative Binomial Distribution (12Hrs)

Negative Binomial distribution – Definition, Moments of Negative Binomial Distribution, Recurrence Relation for the moments of Negative Binomial distribution, Moment generating function, Characteristic function, Cumulants and Probability generating function of Negative Binomial distribution. Additive property of Negative Binomial distribution. Measure of skewness, kurtosis and problems. The first two moments are obtained through moment generating function, Recurrence relation for the probabilities of Negative Binomial distribution. Poisson distribution as a limiting case of Negative Binomial distribution.

Examples/Applications/Case Studies:

1. Modelling the Number of Trials Until a Fixed Number of Successes

Scenario: Consider a factory where each machine has a probability p of producing a defective part. You are interested in the number of machines needed to find a fixed number r of defective parts.

Application: This scenario can be modelled using the Negative Binomial distribution, where r represents the number of defective parts (successes), and the trials represent machines inspected until the rth defective part is found.

2. Modelling the Number of Failures Before a Set Number of Successes

Scenario: In a clinical trial, suppose you are examining the number of patients who must be tested before achieving a certain number of successful treatments. If each treatment has a success probability p, the number of trials needed to achieve r successes follows a Negative Binomial distribution.

Application: This distribution helps in understanding how many patients need to be treated before reaching the desired number of successful outcomes.

3 Case Study: Insurance Claims

Scenario: Insurance companies often use the Negative Binomial distribution to model the number of claims made by policyholders before a certain number of claims is reached. For instance, an insurer might want to model the number of policies that must be written before observing a certain number of large claims.

Application: This helps in risk assessment and setting premium levels based on the number of claims and the probability distribution of these claims.

Exercises/Projects:

1. Coin Tossing Simulation

Objective: Simulate a negative binomial experiment by tossing a coin repeatedly until a specified number of heads is obtained.

- Steps:
  a) Define the desired number of heads (r).
  b) Simulate coin tosses using a random number generator.
  c) Count the number of tails (failures) before obtaining r heads.
  d) Repeat the experiment multiple times to create a dataset.
  e) Calculate the mean, variance, and other statistics of the simulated data.
  f) Compare the results with the theoretical values for a negative binomial distribution with parameters r and p (probability of heads).
  g)

2. Simulation and Application of the Negative Binomial Distribution

Objective: To simulate the Negative Binomial distribution and understand its properties through empirical data and compare with theoretical values.

Instructions:

Simulation:

Write a program (in Python, R, etc.) to simulate random variables from a Negative Binomial distribution with given parameters r and p

Generate a large sample (e.g., 10,000 samples) for various sets of r and p values.

Empirical Analysis:

Compute the sample mean and variance from the simulated data.

Compare the empirical mean and variance with the theoretical values derived from the distribution's parameters.

Visualization:

Plot histograms of the simulated data for different values of r and p.

Create plots to compare the empirical distribution with the theoretical PMF.

Application:

Use the simulated data to solve a practical problem, such as modelling the number of trials needed to achieve a certain number of successes in a sequence of Bernoulli trials.

Analyse the skewness and kurtosis of the simulated data and compare these with theoretical values.

Specific Resources: (web)

https://www.khanacademy.org/math/statistics-probability

Unit – 4: Geometric Distribution (12Hrs)

Geometric distribution – Definition, Moments of Geometric distribution, Moment generating function, Characteristic function, Cumulants and Probability generating function of geometric distribution. Additive property of geometric distribution. Measure of skewness, kurtosis and problems. The first two moments are obtained through moment generating function, Recurrence relation for the probabilities of geometric distribution. Lack of memory property.

Examples/Applications/Case Studies:

1.Modelling the Number of Trials Until the First Success

Scenario: Consider a quality control test where each unit produced has a probability p of passing the test. You want to know the number of units that need to be tested until the first unit passes the test.

Application: This scenario is modelled using the Geometric distribution, where X represents the number of trials until the first success. The probability mass function (PMF) of the Geometric distribution is given by: $P(X=k)=q^{k-1}p,\ k=1,2,3,...$

where k is the trial number of the first success, and p is the probability of success in each trial.

2. Example: Modelling the Number of Calls Until a Sale

Scenario: In a telemarketing campaign, the probability of making a sale on any given call is p. You want to determine the number of calls needed to achieve the first sale.

Application: This scenario can be modelled using the Geometric distribution to determine the expected number of calls required. For instance, if p=0.1, you can calculate the expected number of calls needed using the mean of the Geometric distribution, which is $\frac{1}{p}$.

Exercises/Projects:

Simulation and Application of the Geometric Distribution

Objective:

To simulate the Geometric distribution and analyse empirical data to understand its properties and applications.

Instructions:

1. **Simulation**:
   - Write a program (using Python, R, etc.) to simulate random variables from a Geometric distribution with parameter p.
   - Generate a large sample (e.g., 10,000 samples) for various values of p.
2. **Empirical Analysis**:
   - Compute the sample mean and variance from the simulated data.
   - Compare the empirical mean and variance with the theoretical values E(X) and V(x)
3. **Visualization**:
   - Plot histograms of the simulated data for different values of p.
   - Create plots to compare the empirical PMF with the theoretical PMF.
4. **Application**:
   - Use the simulated data to solve a practical problem. For example, model the number of trials required to achieve the first success in a sequence of Bernoulli trials and use the results to make predictions or decisions.
   - Analyse and interpret the results in the context of the application. For instance, if you model the number of calls needed before achieving a sale, discuss how the results can inform sales strategies.
5. **Lack of Memory Property**:
   - Demonstrate the lack of memory property of the Geometric distribution through simulation. For example, simulate the number of trials needed after a

certain number of trials have already occurred and verify that the distribution of the remaining trials follows the same Geometric distribution.

Unit – 5: Hyper Geometric Distribution (12Hrs)

Hyper Geometric Distribution – Definition, mean and variance, problems. Recurrence relation for probabilities. Limiting case of Hyper Geometric distribution to Binomial distribution.

Examples/Applications/Case Studies:

1. Example: Quality Control in Manufacturing

Scenario: Suppose a factory produces a batch of 100 items, 15 of which are defective. A quality control inspector randomly selects 10 items from the batch for inspection. You want to find the probability that exactly 3 out of the 10 inspected items are defective.

2. Case Study: Election Polling

Scenario: During an election, a poll is conducted to estimate the proportion of voters who support a particular candidate. Suppose there are 500 registered voters, of whom 200 are known to support the candidate. If a poll of 50 voters is randomly selected, what is the probability that exactly 20 of them support the candidate?

Exercises/Projects:

Simulation and Limiting Case to Binomial Distribution

Objective: To simulate the Hypergeometric distribution and examine its approximation to the Binomial distribution as the population size becomes large.

Instructions:

1. Simulation:

Write a program (in Python, R, etc.) to simulate random variables from the Hypergeometric distribution. Use parameters N, K, and n to generate a large sample (e.g., 10,000 samples).

2. Empirical Analysis:

Compute the sample mean and variance from the simulated data. Compare these with the theoretical values derived from the Hypergeometric distribution.

Plot histograms of the simulated data for different parameter values.

3. Limiting Case Analysis:

Simulate the Hypergeometric distribution with large N and fixed n (e.g., N=1,000 and n=50). Compare the Hypergeometric distribution with the corresponding Binomial distribution where $p = \frac{k}{N}$

4. Comparison:

Compare the simulated data of the Hypergeometric distribution with the Binomial distribution using statistical tests or visual plots to evaluate the approximation.

Show how the Binomial distribution approximates the Hypergeometric distribution as N increases while keeping n constant.

5. Application:

Apply the results to a real-world scenario, such as quality control or polling, and discuss how the approximation affects decision-making.

Text Books:

1. Gupta. S.C. &Kapoor,V.K. (2023) . Fundamentals of Mathematical Statistics, Sultan Chand & Sons Pvt. Ltd. New Delhi.

References:

1. Bansilal and Arora (1989). New Mathematical Statistics, Satya Prakashan, New Delhi.
2. Goon A.M., Gupta M.K. and Dasgupta B. (2002): Fundamentals of Statistics, Vol. I & II, 8th Edn. The World Press, Kolkata.
3. Mukhopadhyay, P. (2015). Mathematical Statistics.Publisher: BOOKS AND ALLIED (1 January 2016)

***

PARVATHANENI BRAHMAYYA
SIDDHARTHA COLLEGE OF ARTS & SCIENCE
Autonomous
Siddhartha Nagar, Vijayawada–520010
Re-accredited at 'A+' by the NAAC

Course Code	23STMAL231
Title of the Course	Discrete Probability Distributions
MAX.MARKS: 70	MAX.TIME: 3Hrs

MODEL QUESTION PAPER

SECTION A (5 x 4 = 20 Marks)

(Answer the following questions, internal choice provided for each question. Each question carries 4 marks. Cognitive Level: K2 to K3)

1. **(a)**Define the Discrete Uniform distribution and state its mean and variance.

   OR

   (b) Calculate the mean and variance for the Binomial distribution. (K2)

2. **(a)** Define the geometric distribution. What is the mean and variance of the geometric distribution? **OR**

   (b) Let X ~G(1/2) and Y ~ G(1/2) then find P(X =2|X+Y = 4). (K3)

3. **(a)**What are the applications of Hyper-geometric distribution.

   OR

   (b) In a small pond there are 50 fish, 10 of which have been tagged. A fisherman's catch consists of 7 fish (assume his catch is a random selection done without replacement). What is the probability that exactly 2 tagged fish are caught? (K2)

4. **(a)** Define the Poisson distribution and compute its mean and variance.

   OR

   (b) Show that the Poisson distribution is a limiting case of the Binomial distribution.(K2)

5. **(a)**Write the recurrence relation for the probabilities of the Negative Binomial distribution.

   OR

   (b) Calculate first two moments of the Negative Binomial distribution. (K3)

SECTION B (5 x 10 = 50 Marks)

(Answer the following questions, internal choice provided for each question. Each question carries 10 marks. Cognitive Level: K3 to K4)

6. **(a)** Derive the moment generating function of the Binomial distribution and use it to find the first two moments.(K4)

OR

(b) A coffee connoisseur claims that he can distinguish between a cup of instant coffee and a cup of percolator coffee 75% of the time. It is agreed that his claim will be accepted if he correctly identifies at least 5 of the 6 cups. Find his chances of having the claim (i) accepted, (ii) rejected, when he does have the ability he claims.(K4)

7. **(a)** Explain the recurrence relation for moments of the Poisson distribution and derive the second moment. (K4)

OR

(b) Discuss the additive or reproductive property of the Poisson distribution and its applications. (K4)

8. **(a)** Derive the moment generating function of the Negative Binomial distribution and use it to find its moments. (K3)

OR

(b) Discuss the skewness and kurtosis of the Negative Binomial distribution. (K3)

9. **(a)** Derive the moment generating function and explain the lack of memory property of the Geometric distribution. (K4)

OR

(b) Write the recurrence relation for probabilities of the Geometric distribution and compute its variance. (K4)

10. **(a)** Define the Hypergeometric distribution and derive its mean and variance.(K3)

OR

(b) Show how the Hypergeometric distribution approaches the Binomial distribution as a limiting case.(K3)

***