Course Code				**23STMIL231**			
Title of the Course				**Random Variables & Mathematical Expectations**			
Offered to:				**B.Sc.(Honours) – Data Science**			
L	**4**	**T**	**0**	**P**	**0**	**C**	**3**
Year of Introduction:		**2024-25**		**Semester:**		**3**	
Course Category:		**Minor**		**Course Relates to:**		**Local, Regional, National, Global**	
Year of Revision:		**NA**		**Percentage:**		**NA**	
Type of the Course:				**SKILL DEVELOMENT**			
Crosscutting Issues of the Course :				**NA**			
Pre-requisites, if any				**Basic Mathematics and Probability**			

Course Description:

This course helps the students to familiarize with the ways in which we talk about uncertainty and estimate their situations in which probability arises. Also this course aims at providing basic knowledge about mathematical expectations & generating functions.

Course Aims and Objectives:

S. No	**COURSE OBJECTIVES**
1	Develop a comprehensive understanding of univariate random variables, including their probability distributions, functions, and characteristics, to model and analyze real-world phenomena.
2	Master the concept of mathematical expectation and its properties, enabling the calculation of mean, variance, and other statistical measures for effective decision-making.
3	Acquire proficiency in generating functions, their applications in probability theory, and their role in deriving probability distributions and moments.
4	Explore the Law of Large Numbers and its implications, understanding the convergence of sample means to population means.
5	Grasp the Central Limit Theorem and its applications, recognizing the normal approximation of sample means and its importance.

Course Outcomes

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

NO	COURSE OUTCOME	BTL	PO	PSO
CO1	understand the concept of discrete and continuous random variables with the application of random variables in real time problems.	K2	1	1
CO2	apply the variance and covariance techniques of random variables in terms of expectation	K3	1	1
CO3	understand the definitions of various generating functions and their applications.	K2	1	1
CO4	apply the concepts of Weak Law of Large Numbers (WLLN) and Strong Law of Large Numbers (SLLN)	K3	1	1
CO5	apply the concept of Central limit theorem in real life examples and various inequalities in expectation	K3	1	1

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

CO-PO-PSO MATRIX									
CO NO	**PO1**	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PSO1**	**PSO2**
CO1	**2**							**2**	
CO2	**3**							**3**	
CO3	**2**							**2**	
CO4	**3**							**3**	
CO5	**3**							**3**	

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

Course Structure:

Unit – 1 **Univarate Random Variables** **(12 Hours)**

Definition, Discrete and Continuous random variables -Probability mass function and Probability density function with illustrations. Distribution function and its properties. Calculation of moments, coefficient of skewness and kurtosis to defined probability mass function (PMF) and Probability Density function (PDF)

Examples/Applications/Case Studies:

1. A random experiment of tossing a coin can be modeled using a random variable that represents the outcome of getting a head.
2. Rolling a die - Number of dots on the upper face

Exercises/Project:

1. Simulation

a. Introduce random number generators and simulation tools.
b. Provide a random variable scenario (e.g., flipping a coin, rolling a die, generating random numbers).
c. Have students use technology to simulate the experiment and collect data.

d. Analyze the results and compare them to theoretical probabilities.

2. Probability Distributions

a. Introduce different types of probability distributions (discrete and continuous).
b. Provide examples of random variables.
c. Divide students into groups.
d. Assign each group a random variable.
e. Ask them to determine the possible values and their corresponding probabilities.
f. Create a probability distribution table or graph.
g. Discuss the characteristics of the distribution.

Specific Resources: (web):

https://www.khanacademy.org/math/statistics-probability/random-variables-stats-library

Unit – 2 Mathematical Expectations (12 Hours)

Mathematical expectation of a random variable, Properties of expectations. Moments and covariance using mathematical expectation and their properties. Addition and Multiplication theorems on expectation of two and n variables. Variance, variance of a Linear combination of Random variables

Examples/Applications/Case Studies:

1. **Fair Games:** Used in gambling and game theory to determine the expected payoff or loss.
2. **Decision Making:** Helps in evaluating options by calculating expected values of different choices.

Exercises/Project:

1. Lottery Game Simulation

a. Create a simple lottery game with different prize values and corresponding probabilities.
b. Divide students into groups.
c. Each group should simulate playing the lottery multiple times, recording their winnings.
d. Calculate the average winnings per game (experimental expected value).
e. Compare the experimental expected value with the theoretical expected value calculated using the probability distribution.
f. Discuss the implications of the results in terms of fair games and decision-making.

2. Insurance Premium Calculation

a. Present a hypothetical insurance scenario (e.g., car insurance).
b. Provide data on the probability of different types of accidents and corresponding claim amounts.
c. Calculate the expected claim amount per policyholder.

d. Determine the insurance premium that would cover the expected claims and administrative costs.
e. Discuss factors that might affect the actual premium charged (e.g., profit margin, risk assessment).
f. Explore the concept of risk aversion and how it influences insurance purchasing decisions.

Specific Resources: (web):

https://ocw.mit.edu/courses/18-05-introduction-to-probability-and-statistics-spring-2022/resources/lecture-notes/

Unit – 3 Generating Functions (12 Hours)

Definitions of Moment Generating Function (M.G.F.), Cumulant Generating Function (C.G.F), Probability Generating Function (P.G.F), Characteristic Function (c.f.) and their properties with applications.

Examples/Applications/Case Studies:

1. Generating functions are used to solve counting problems, like finding the number of ways to partition a number or the number of solutions to a linear equation.
2. In the study of branching processes, PGFs are used to analyze the size of populations over time.

Exercises/Project:

1. Generating Function Applications in Probability

a. Provide students with probability problems involving sums of independent random variables or finding probabilities of specific outcomes.
b. Guide students in using generating functions to solve these problems.
c. Emphasize the efficiency of using generating functions compared to traditional methods.
d. Discuss the limitations of generating functions and when they might not be applicable.

2. Generating Functions and Combinatorics

a. Introduce the concept of generating functions as a tool for solving counting problems.
b. Provide examples of counting problems (e.g., number of ways to distribute identical objects to distinct recipients, number of non-negative integer solutions to a linear equation).
c. Show how to represent the problem as a generating function.
d. Demonstrate how to extract the desired information from the generating function.
e. Explore different types of generating functions (ordinary, exponential) and their applications.

Specific Resources: (web):

https://ocw.mit.edu/courses/18-05-introduction-to-probability-and-statistics-spring-2022/resources/lecture-notes/

Unit – 4 Law of Large Numbers & Inequalities (12 Hours)

Weak Law of Large Numbers (WLLN) and Strong Law of Large Numbers (SLLN- Statement only) for identically and independently distributed (i.i.d.) random variables with finite variance. Markov's inequality (Statements only), Khinchin's Theorem for WLLN (Statements only).

Examples/Applications/Case Studies:

1. In manufacturing processes, it can be used to assess the long-run average quality of products based on sample data.
2. In finance, it helps in estimating the long-term average return of an investment based on historical data.

Exercises/Project:

1. Coin Tossing Simulation

a. Divide students into groups.
b. Each group tosses a coin a specified number of times (e.g., 10, 20, 50, 100).
c. Calculate the proportion of heads for each trial.
d. Plot the proportion of heads against the number of tosses.
e. Discuss how the proportion of heads converges to 0.5 (the theoretical probability of heads) as the number of tosses increases.
f. Relate the experiment to the WLLN.

2. Markov's Inequality and Gambling

a. Introduce Markov's inequality and its statement.
b. Present a gambling scenario (e.g., a game with a random payoff).
c. Calculate the expected value of the payoff.
d. Use Markov's inequality to estimate the probability of winning a large amount.
e. Discuss the limitations of Markov's inequality and the need for tighter bounds.

Specific Resources: (web):

https://ocw.mit.edu/courses/6-041sc-probabilistic-systems-analysis-and-applied-probability-fall-2013/resources/lecture-19-video-2/

Unit – 5 Central Limit Theorems (CLT) (12 Hours)

Chebychev's and Cauchy - Schwartz inequalities and their applications. Central limit theorem- Statement of De-Movire's Laplace theorem, Lindberg – Levy CLT and its applications, Statement of Liapounoff's CLT, relationship between CLT and WLLN.

Examples/Applications/Case Studies:

1. The CLT can be used to monitor process means and detect shifts in the process.
2. The distribution of portfolio returns is often assumed to be normal, based on the CLT applied to individual asset returns.

Exercises/Project:

1. Dice Rolling Simulation

a. Divide students into groups.
b. Each group rolls a set of dice multiple times (e.g., 5, 10, 20, 30 dice).
c. Calculate the average of the dice rolls for each trial.
d. Create a histogram of the sample means.
e. Observe the shape of the distribution as the number of dice increases.
f. Discuss how the distribution of sample means approaches a normal distribution, illustrating the CLT.

2. Simulation of Sampling Distributions

a. Simulate a population with a known distribution (e.g., normal, uniform, exponential).
b. Draw random samples of different sizes from the population.
c. Calculate the sample mean for each sample.
d. Create a histogram of the sample means.
e. Observe the shape of the distribution and compare it to a normal distribution.
f. Explore the impact of sample size on the shape of the sampling distribution.

Specific Resources: (web):

https://ocw.mit.edu/courses/6-041-probabilistic-systems-analysis-and-applied-probability-fall-2010/resources/lecture-20-the-central-limit-theorem/

Text Book

1. S. C. Gupta and V. K. Kapoor,Fundamentals of Mathematical Statistics, 12th Edition, 10th September 2020, Sultan Chand & Sons, New Delhi.

Recommended References books:

1. D. Biswas, Probability and Statistics,Volume I, New central book Agency (P) Ltd, New Delhi.
2. A.M.Goon,M.K. Gupta, B.Dasgupta , An outline of Statistical theory, Volume Two,3rd Edition,2010(with corrections) The World Press Pvt.Ltd., Kolakota.

***

23STMIL231: Random Variables & Mathematical Expectations

B.Sc.(Honours) – Data Science

Max. Marks : 70 **Semester III** **Max. Time : 3Hrs**

Section - A

Answer the following questions **5 X 4M = 20M**

1. a. Define the Random variables and state its types. (Co-1, K-1)

   (OR)

   b. Define distribution function and state its properties (Co-1, K-1)

2. a. Show that the mathematical expectation of the sum of two random variables is the sum of their individual expectation. (Co-2, K-1)

   (OR)

   b. State and prove multiplication theorem on Mathematical expectationof two events. (Co-2, K-1)

3. a. Define Probability Generating Function(PGF) and write its properties. (Co-3, K-1)

   (OR)

   b. Define Characteristic Function(CF) and write its properties. (Co-3, K-1)

4. a. Explain the concept of Weak law of large numbers(WLLN). (Co-4, K-2)

   (OR)

   b. Explain the concept of Strong Law of Large Numbers (SLLN). (Co-4, K-2)

5. a. State the Liapounoff'scentral limit theorem. (Co-5, K-1)

   (OR)

   b. State the Lindberg – Levy's theorem and its assumptions. (Co-5, K-1)

Section - B

Answer the following questions **5 X 10M = 50M**

6. a. A random variable has the following probability distribution

x	0	1	2	3	4	5	6	7	8
P(X= x)	a	3a	5a	7a	9a	11a	13a	15a	17a

   (i) Determine 'a'

   (ii) Find $P(X<3)$, $P(X \geq 3)$ and $P(0<X<5)$

   (iii) Find the distribution function of X. (Co-1, K-3)

   (OR)

b. The diameter of an electric cable, say X, is assumed to be a continuous random variable with p.d.f. $f(x) = 6x(1-x),\ 0 \le x \le 1$.

i) Check that f(x) is p.d.f.,

ii) Determine a number **b** such that P(X <**b**) = P(X >**b**). (Co-1, K-3)

7. a. State and prove the Linear combination of Random variables (Co-2, K-5)

(OR)

b. Given the following table:

X = x	-3	-2	-1	0	1	2	3
P(X = x)	0.05	0.1	0.3	0	0.3	0.15	0.1

Compute (i) E(X), (ii) E(2X+3), (iii) V(X) and (iv) V(2X+3) (Co-2, K-5)

8. a. Derive the relation between cumulants in terms central moments. (Co-3, K-3)

(OR)

b. Prove that $\mu_r^1 = \left[\frac{d^r M_X(t)}{dt^r}\right]_{t=0}$ (Co-3, K-3)

9. a. Examine whether the weak law of large numbers holds for the sequence {X_k} of independent random variable defined as follows: $P\left(X_k = \pm 2^k\right) = 2^{-(2k+1)},\ P\left(X_k = 0\right) = 1 - 2^{-2k}$ (Co-4, K-3)

(OR)

b. Write the statements of W.L.L.N and S.L.L.N. for the sequence ofi.i.d. random variables (Co-4, K-3)

10. a. Use Chebychev's inequality to determine how many times a fair coin must be tossed in order that the probability will be atleast 0.90 that the ratio of the observed number of heads to the number of tosses will lie between 0.4 and 0.6. (Co-5, K-3)

(OR)

b. State and prove Cauchy - Schwartz inequalities (Co-5, K-3)

***