Reg. No.									

MANIPAL INSTITUTE OF TECHNOLOGY

MANIPAL

(A constituent unit of MAHE, Manipal)

FOURTH SEMESTER B.TECH. (Data Science)

MAKEUP EXAMINATIONS, June 2024

MATHEMATICAL FOUNDATION FOR DATA SCIENCE II [MAT-2239]

REVISED CREDIT SYSTEM

Time: 9:30-12:30pm **Date: 14/06/2024** **Max. Marks: 50**

Instructions to Candidates:

- Answer **ALL** the questions.
- Missing data may be suitably assumed.

Q.NO	Questions	Marks	CO	PO	BTL
1A.	Customers tend to exhibit loyalty to product brands but may be persuaded through clever marketing and advertising to switch brands. Consider the case of three brands: A, B and C. Customer "unyielding" loyalty to a given brand is estimated at 75%, giving the competitors only a 25% margin to realize a switch. Competitors launch their advertising campaigns once a year. For brand A customers, the probability of switching to brands B and C are 0.1 and 0.15 respectively. Customers of brand B are likely to switch to A and C with probabilities 0.2 and 0.05 respectively. Brand C customers can switch to brands A and B with equal probabilities. Express the situation as a Markov chain and write the transition probability matrix. Clearly mention the elements of the state space and draw the state transition diagram.	3	1	1,2,8, 12	4,5
1B.	Find the stationary distribution associated with the following transition probability matrix $\begin{bmatrix} 0.5 & 0.2 & 0.3 \\ 0.4 & 0.4 & 0.2 \\ 0.4 & 0.1 & 0.5 \end{bmatrix}$.	3	1	1,2,8, 12	4,5
1C.	For a Markov chain with state space S = {1, 2, 3, 4, 5, 6} having transition probability matrix given below, draw the transition diagram, identify the Markov chain is reducible or not. Identify the states are transient, recurrent and absorbing states, if any. If there exist recurrent states, classify them into null recurrent and non-null states. Also determine the periodicity of the positively recurrent states.	4			

	$P = \begin{bmatrix} 1/2 & 1/2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 1/3 & 0 & 0 & 1/3 & 1/3 & 0 \\ 0 & 0 & 0 & 1/2 & 1/2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$		*1*	*1,2,8, 12*	*4,5*
2A.	Let $\{X_n, n \geq 0\}$ be a Markov Chain with State space $S = \{0, 1, 2\}$ and transition probability matrix $\begin{pmatrix} \frac{3}{4} & \frac{1}{4} & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & \frac{3}{4} & \frac{1}{4} \end{pmatrix}$ and initial probability distribution $P[X_0 = i] = \frac{1}{3}, \ \forall i = 0, 1, 2.$ Find (i) $P[X_1 = 1 \mid X_0 = 2]$ (ii) $P[X_2 = 2, X_1 = 1 \mid X_0 = 2]$ (iii) $P[X_3 = 1, X_2 = 2, X_1 = 1, X_0 = 2]$	*3*	*2*	*1,2,8, 12*	*4,5*
2B.	Explain the four types of stochastic processes with reference with an example for each clearly mentioning the state space and the index set.	*3*	*2*	*1,2,8, 12*	*4,5*
2C.	Let $\boldsymbol{X}$ be $\boldsymbol{N_3(\mu, \Sigma)}$ with $\mu = \begin{bmatrix} 2 \\ 3 \\ 6 \end{bmatrix}$ and $\Sigma = \begin{bmatrix} 16 & -1 & 0 \\ -1 & 4 & 1 \\ 0 & 1 & 9 \end{bmatrix}$. (i) Find $P\{2X_1 > 8\}$ (ii) Find $P\{3X_2 + 2X_3 > 50\}$ iii) Find $P\{3X_1 - 4X_2 + 2X_3 < 70\}$.	*4*	*2*	*1,2,8, 12*	*4,5*
3A.	The following $X = \begin{bmatrix} 4.0 & 2.0 & 0.60 \\ 4.2 & 2.1 & 0.59 \\ 3.9 & 2.0 & 0.58 \\ 4.3 & 2.1 & 0.62 \\ 4.1 & 2.2 & 0.63 \end{bmatrix}$ is the data matrix on the variables X_1, X_2 and X_3. Find the mean vector, covariance matrix S.	*3*	*3*	*1,2,8, 12*	*4,5*
3B.	Let $X \sim N_3(\mu, \Sigma)$ with $\mu = (2 \ -3 \ \ 1)^T$ and $\Sigma = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 3 & 2 \\ 1 & 2 & 2 \end{bmatrix}$. Find the conditional distribution of X_3 given that $X_1 = x_1$ and $X_2 = x_2$.	*3*	*3*	*1,2,8, 12*	*4,5*

<table>
<tr><td>3C.</td><td>Age and risky behavior are the two important factors that make difference between accident group (AG) and non-accident group (NAG) of workers. Random samples of 20 individuals from AG and 50 individuals from NAG were collected. The sample mean vector and sample covariance matrix are given. Construct 95% confidence region for the difference between the two population mean vectors.
Sample 1 - $\bar{X}_A = [50\ \ 6]'$ and $S_A = \begin{bmatrix} 16 & -5 \\ -5 & 4 \end{bmatrix}$
Sample 2 - $\bar{X}_B = [40\ \ 8]'$ and $S_B = \begin{bmatrix} 25 & -6 \\ -6 & 9 \end{bmatrix}$</td><td>4</td><td>4</td><td>1,2,8, 12</td><td>4,5</td></tr>
<tr><td>4A.</td><td>The variance covariance matrix of the random vector $X = (X_1, X_2, X_3)$ is given by $\varepsilon = \begin{bmatrix} 25 & -2 & 4 \\ -2 & 4 & 1 \\ 4 & 1 & 9 \end{bmatrix}$. Find the Correlation matrix and also evaluate the correlation between X_1 and $\frac{X_2 + X_3}{2}$.</td><td>3</td><td>4</td><td>1,2,8, 12</td><td>4,5</td></tr>
<tr><td>4B.</td><td>Consider the probability distribution function $f(x,y) = \begin{cases} \frac{x}{2} + \frac{3y}{2}; 0 \le x, y \le 1 \\ 0 ; otherwise \end{cases}$. Find the mean vector and covariance matrix.</td><td>3</td><td>3</td><td>1,2,8, 12</td><td>4,5</td></tr>
<tr><td>4C.</td><td>From the following data, obtain the (a) the multiple correlation coefficient, $R_{2.13}$
(b) the partial correlation between $r_{13.2}$
<table><tr><td>X_1</td><td>2</td><td>5</td><td>7</td><td>11</td></tr><tr><td>X_2</td><td>3</td><td>6</td><td>10</td><td>12</td></tr><tr><td>X_3</td><td>1</td><td>3</td><td>6</td><td>10</td></tr></table></td><td>4</td><td>4</td><td>1,2,8, 12</td><td>4,5</td></tr>
<tr><td>5A.</td><td>In an orthogonal exploratory factor model (assume $m = 1$ factor model), suppose the manifest variables X_1, X_2 and X_3 have the sample covariance matrix $S = \begin{bmatrix} 100 & 40 & 0 \\ 40 & 60 & 0 \\ 0 & 0 & 30 \end{bmatrix}$. The estimated Eigen values of the covariance matrix S are $\hat{Q}_1 = 124.72, \hat{Q}_2 = 35.28, \hat{Q}_3 = 30$
(a)Find the estimated Eigen vector corresponding to the largest Eigen value
(b) Find the estimated loading matrix $\hat{\Lambda}$.
(c) Find the communality $\hat{\Lambda}\hat{\Lambda}^T$ for the variable X_2.</td><td>3</td><td>5</td><td>1,2,8, 12</td><td>4,5</td></tr>
<tr><td>5B.</td><td>Given the distance matrix between pairs of five objects {A, B, C, D, E}, obtain the clusters and the final distance matrix</td><td>3</td><td></td><td></td><td></td></tr>
</table>

	D using complete linkage. $D = \begin{bmatrix} 0 & . & . & . & . \\ 1 & 0 & . & . & . \\ 2 & 2 & 0 & . & . \\ 2 & 5 & 3 & 0 & . \\ 3 & 4 & 6 & 6 & 0 \end{bmatrix}$ and also draw the Dendrogram.		5	1,2,8, 12	4,5
5C.	Let $X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$ be the data matrix and $S = \begin{bmatrix} 14 & -11 \\ -11 & 23 \end{bmatrix}$ be the sample covariance matrix with sample size $n = 50$. (a) Find the first principal component Z_1 (b) Find the variance explained by the first principal component	4	5	1,2,8, 12	4,5