Reg. No.									

MANIPAL INSTITUTE OF TECHNOLOGY

MANIPAL

(A constituent unit of MAHE, Manipal)

FOURTH SEMESTER B.TECH. (Data Science)

END SEMESTER EXAMINATIONS, April 2024

MATHEMATICAL FOUNDATION FOR DATA SCIENCE II [MAT-2239]

REVISED CREDIT SYSTEM

Time: 9:30-12:30pm **Date: 30/04/2024** **Max. Marks: 50**

Instructions to Candidates:

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

Q.NO	Questions	Marks	CO	PO	BTL
1A.	An urn initially contains 5 black balls and 5 white balls. The following experiment is repeated indefinitely. A ball is draw form the urn, if the ball is white it is put back in the urn, otherwise it is left out. Let X_n be the number of black balls remaining in the urn after n draws from the urn. Find the one step transition probability matrix and draw the transition diagram.	3	1	1,2,8, 12	4,5
1B.	Find the limiting- state probability vector associated with the following transition probability matrix $\begin{bmatrix} 2/5 & 2/5 & 1/5 \\ 1/5 & 3/5 & 1/5 \\ 1/5 & 2/5 & 2/5 \end{bmatrix}$.	3	1	1,2,8, 12	4,5
1C.	For a Markov chain with state space S = {1, 2, 3, 4, 5, 6, 7, 8} 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. $P = \begin{bmatrix} 0.4 & 0.3 & 0.3 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.6 & 0.4 & 0 & 0 & 0 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.8 & 0.2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.4 & 0.6 & 0 \\ 0.4 & 0.4 & 0 & 0 & 0 & 0 & 0 & 0.2 \\ 0.1 & 0 & 0.3 & 0 & 0 & 0.6 & 0 & 0 \end{bmatrix}$	4	1	1,2,8, 12	4,5

2A.	Let X be a Markov chain with the state space E = {1,2}, with initial probability vector $\pi[1/3 \quad 2/3]$ and one step transition probability matrix $P = \begin{bmatrix} 0.5 & 0.5 \\ 0.3 & 0.7 \end{bmatrix}$. Find i) $P[X_1 = 2, X_4 = 1, X_6 = 1 \mid X_0 = 1]$ ii) $P[X_2 = 1, X_7 = 2, X_9 = 2 \mid X_0 = 1]$ iii) $P[X_2 = 1 \mid X_7 = 2]$.	3	2	1,2,8, 12	4,5
2B.	Consider a Gambler's ruin problem with $p = 0.4$ and the state space S = {0,1,2,3,4,5,6}, where states 0 and 6 are absorbing states. The following matrix $Q = \begin{bmatrix} 0 & 0.4 & 0 & 0 & 0 \\ 0.6 & 0 & 0.4 & 0 & 0 \\ 0 & 0.6 & 0 & 0.4 & 0 \\ 0 & 0 & 0.6 & 0 & 0.4 \\ 0 & 0 & 0 & 0.6 & 0 \end{bmatrix}$ represents the TPM of transient states {1,2,3,4,5}. Starting in state 3, determine (a) the expected amount of time spent in state 3 and the expected amount of time spent in state 2 (b) the probability of ever visiting state 4.	3	2	1,2,8, 12	4,5
2C.	Let X be $N_3(\mu, \Sigma)$ with $\mu = \begin{bmatrix} 5 \\ 3 \\ 7 \end{bmatrix}$ and $\Sigma = \begin{bmatrix} 4 & -1 & 0 \\ -1 & 4 & 2 \\ 0 & 2 & 9 \end{bmatrix}$. (i) Find $P\{X_1 > 6\}$ (ii) Find $P\{5X_2 + 4X_3 > 70\}$ iii) Find $P\{4X_1 - 3X_2 + 5X_3 < 80\}$.	4	2	1,2,8, 12	4,5
3A.	The following are the three measurements on the variables X_1 and X_2. Find the mean vector, covariance matrix S and X_1: 10, 12, 11 X_2: 100, 110, 105 correlation matrix for the same	3	3	1,2,8, 12	4,5
3B.	Let $X \sim N_3(\mu, \Sigma)$ with $\mu = (-3 \ 1 \ 4)^T$ and $\Sigma = \begin{bmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{bmatrix}$. Find the conditional distribution of X_2 given that $X_1 = x_1$ and $X_3 = x_3$.	3	3	1,2,8, 12	4,5
3C.	Consider the observation vector $X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, the mean vector and the sample covariance matrix of three treatments are given as follows. Let $\overline{X_1} = \begin{bmatrix} 8 \\ 4 \end{bmatrix}$ and $S_1 = \begin{bmatrix} 3 & 3 \\ 3 & 7 \end{bmatrix}$ with $n_1 =$ 3,	4	4	1,2,8, 12	4,5

	$\overline{X_2} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and $S_2 = \begin{bmatrix} 2 & -4 \\ -4 & 8 \end{bmatrix}$ with $n_2 = 2$ and $\overline{X_3} = \begin{bmatrix} 2 \\ 8 \end{bmatrix}$ and $S_3 = \begin{bmatrix} 1 & -0.5 \\ -0.5 & 1 \end{bmatrix}$ with $n_3 = 3$ respectively. Using the information provided, test for equality of mean vectors using one-way MANOVA statistic at 1% level given the cut off as 7.01.				
4A.	Samples of sizes $n_1 = 45$ and $n_2 = 55$ were taken of Wisconsin homeowners with and without air conditioning, respectively. Two measurements of electrical usage (in kilowatt hours) were considered. The first is a measure of total on-peak consumption (X_1) during July, and the second is a measure of total off-peak consumption (X_2) during July. The resulting summary statistics are $\overline{X_1} = \begin{bmatrix} 204.4 \\ 556.6 \end{bmatrix}, S_1 = \begin{bmatrix} 13825.3 & 23823.4 \\ 23823.4 & 73107.4 \end{bmatrix}$ and $\overline{X_2} = \begin{bmatrix} 130.0 \\ 355.0 \end{bmatrix}, S_1 = \begin{bmatrix} 8632.0 & 19616.7 \\ 19616.7 & 55964.5 \end{bmatrix}$. Test the null hypothesis $\mu_1 - \mu_2 = 0$ for 5% LOS. [Chi- square tabulated value at 5% LOS is 5.99]	3	4	1,2,8, 12	4,5
4B.	Consider the probability distribution function $f(x_1, x_2) = \begin{cases} \frac{x_1}{2} + \frac{3}{2}x_2; 0 \le x_1, x_2 \le 1 \\ 0 \,; otherwise \end{cases}$. Find the mean vector and covariance matrix.	3	3	1,2,8, 12	4,5
4C.	From the following data, obtain the (a) the multiple correlation coefficient, $R_{1.23}$ (b) the partial correlation between $r_{12.3}$ $\begin{array}{\|c\|c\|c\|c\|c\|} \hline X_1 & 2 & 5 & 7 & 11 \\ \hline X_2 & 3 & 6 & 10 & 12 \\ \hline X_3 & 1 & 3 & 6 & 10 \\ \hline \end{array}$	4	4	1,2,8, 12	4,5
5A.	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 $S = \begin{bmatrix} 100 & 50 & -30 \\ 50 & 64 & -20 \\ -30 & -20 & 49 \end{bmatrix}$. The estimated Eigen values of the covariance matrix S are $\hat{Q}_1 = 148.23, \hat{Q}_2 = 36, \hat{Q}_3 = 28.70$ (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.	3	5	1,2,8, 12	4,5
5B.	Given the distance matrix between pairs of five objects {1, 2, 3, 4, 5}, obtain the clusters and the final distance matrix D	3	5	1,2,8, 12	4,5

	using complete linkage. $D = \begin{bmatrix} 0 & . & . & . & . \\ 4 & 0 & . & . & . \\ 7 & 3 & 0 & . & . \\ 9 & 5 & 2 & 0 & . \\ 1 & 3 & 6 & 8 & 0 \end{bmatrix}$ and also draw the Dendrogram.				
5C.	Let $X = \begin{bmatrix} 4 & 11 \\ 8 & 4 \\ 13 & 5 \\ 7 & 14 \end{bmatrix}$ be the data matrix with sample size n=4. Find the first principle component Z_1.	*4*	*5*	*1,2,8, 12*	*4,5*