Reg. No. : | | | | | | | | | | | | |

Question Paper Code : 10803

M.E./M.Tech. DEGREE EXAMINATIONS, APRIL/MAY 2023.

First Semester

Computer Science and Engineering

MA 4151 — APPLIED PROBABILITY AND STATISTICS FOR COMPUTER SCIENCE ENGINEERS

(Common to : M.E. Computer Science and Engineering (with Specialization in Artificial Intelligence and Machine Learning)/M.E. Computer Science and Engineering (with specialization in Networks)/M.E. Multimedia Technology/ Master of Computer Applications (2 years))

(Regulations 2021)

Time : Three hours Maximum : 100 marks

(Tables to be permitted : Normal table, t-table, Chi square table, F-table)

Answer ALL questions.

PART A — ($10 \times 2 = 20$ marks)

1. Find the inner product of the two vectors $u = (1, 2, -1, 3)$ and $v = (-1, 3, 0, 1)$.
2. What is the purpose of psedudo inverse in linear algebra?
3. What are the axioms of Probability?
4. Give the mean, variance and moment generating function of Poisson distribution.
5. For the joint probability density function $f(x,y) = \frac{4xy}{81}$, $0 < x < 3$, $0 < y < 3$, what is $P(X < 0, Y < 4)$?
6. Suppose if X and Y are independent continuous random variables, then show that $CoV(X, Y) = 0$.
7. A major automobile manufacturer has to recall several models from its 1993 line due to quality control problems that were not discovered with its random final inspection procedures. What type of error can be assumed under this situation?
8. In a large sample test, if a null hypothesis Ho is accepted at 10% level of significance, then will it be accepted at both 5% and 1% level of significance? Give reasons.
9. Mention few properties of Multivariate Normal distribution.
10. What do you mean by Principal Component Analysis?

PART B — (5 × 13 = 65 marks)

11. (a) Does the set of all pairs of real numbers (x, y) with the operations $(x,y)+(x', y') = (xx', yy')$ and $k(x,y) = (kx, ky)$ form a vector space. If not list the axioms that fails. (13)

Or

(b) Find the QR factorization of Matrix $A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$. (13)

12. (a) (i) In an semiconductor manufacturing process, the product failure due to high contamination is 0.10, due to medium contamination is 0.01 and due to low contamination is 0.001. In a particular production 20% of the chips are subjected to high contamination level, 30 % are subjected to medium contamination and 50% are subjected to low contamination level. What is the probability that a product using one of the chips fails? If the semiconductor chip in the product fails, what is the probability that the chip was exposed to high level of contamination? (6)

(ii) Assume that the random variable X denotes the number of bits in error in the next four transmitted bits. It is known that the probability of bits transmitted though digital transmission channel is 0.1. Assume that the transmission trials are independent. What is the probability that 2 bits are in error? What is the probability that atleast two bits are in error? What is the mean of the error occurring in the transmitted bits? (7)

Or

(b) (i) In a large corporate computer network, the user log-ons can be modeled as exponential distributions with $\lambda = 25$ log-ons per hour, What is the probability that there are no log-ons in the interval of six minutes? What is the probability that the time until the next log-on is between two and three minutes? Also determine the interval of time that probability of no log-ons occurs in the interval is 0.90. (6)

(ii) The number of telephone calls that arrive at a phone exchange is often modeled as a Poisson random variable. Assume that on the average there are 10 calls per hour. What is the probability that there are exactly 5 calls in one hour? What is the probability that there are 3 or less calls in one hour? What is the probability that there are exactly 15 calls in two hours? (7)

13. (a) Let the response time be considered as the speed of page downloads in mobile webpages, as it is crucial for customer satisfaction. Let X denotes the number of bars of service, and Y denotes the response time for a particular user and site. The joint probability mass function of the random variables is in the table.

	$X=1$	$X=2$	$X=3$
$Y=1$	0.15	0.1	0.05
$Y=2$	0.02	0.1	0.05
$Y=3$	0.02	0.03	0.2
$Y=4$	0.01	0.02	0.25

Find $P(X>1, Y=2)$, $P(X<2, Y>2)$, Marginal Probability distribution of X and Y, $E(X)$, $E(Y)$, $Var(X)$, $Var(Y)$, $E(XY)$, $Cov(X,Y)$. (13)

Or

(b) Grams of solids removed from a material (y) are thought to be related to the drying time (x). Six observations from an experiment are

Time (x)	2.5	3.0	3.5	4.0	4.5	5.0
Solids (y)	4.3	1.5	1.8	4.9	4.2	4.8

Estimate the best fit linear regression model for the given data and what is the amount of solid removed from the material for a time of 4.75 mm. (13)

14. (a) (i) It is claimed that automobiles are driven on average 20,000 kilometers per year. To test this claim, 100 randomly selected automobile owners are asked to keep a record of the kilometers they travel. At 5% level of significance, test this claim if the random sample showed an average of 23,500 kilometers; and a standard deviation of 3900 kilometers. (6)

(ii) In a shop study, a set of data was collected to determine whether or not the proportion of defectives produced was the same for workers on the day, evening and night shifts. Data collected are shown as

	Day	Evening	Night
Defectives	45	55	70
Non Defectives	905	890	870

Determine whether the defectives are independent of shifts. Use 5% Level of significance. (7)

Or

(b) (i) Two different lighting techniques are compared by measuring the intensity of light at selected locations in areas lighted by two methods. If 15 measurements in the first area had a standard deviation of 2.7 foot-candles, and 21 measurements in the second area has standard deviation of 4.2 foot candles, can it be concluded that the lighting in the second area is less uniform? Use 0.05 level of significance. (6)

(ii) A random sample of 60 printed boards has been collected, and the following number of defects observed

Number of defects	Observed frequency
0	32
1	15
2	9
3	4

Does the assumption of a Poisson distribution seem appropriate as a probability model for this data? Use 0.05 level of significance. (7)

15. (a) Compute the covariance matrix for the multiple variable data set

$$X = \begin{bmatrix} 2 & 5 & 10 \\ 4 & 2 & 12 \\ 7 & 6 & 18 \end{bmatrix}.$$ (13)

Or

(b) Calculate the population principal components for the random variables X_1, X_2 and X_3, if its covariance matrix is given by $\Sigma = \begin{bmatrix} 2 & 2 & -2 \\ 1 & 3 & -2 \\ -1 & 1 & 1 \end{bmatrix}$. (13)

PART C — (1 × 15 = 15 marks)

16. (a) A study on the deflection (mm) of particle board from stress levels of relative humidity gave the result in the table.

X : stress level (%)	54	56	61	63	68	70
Y: deflection (mm)	18.4	16.6	15.3	14.1	13.5	11.6

Fit a simple linear regression equation, for the given data. Estimate the value of defection when stress level is 61%, using the fitted equation. What is the estimate of σ^2 ? Estimate the correlation coefficient of the given data. (15)

Or

(b) The line width of semiconductor manufacturing is assumed to be normally distributed with a mean of 0.5 micrometer and a standard deviation of 0.05 micrometer. (15)

(i) What is the probability that a line width is greater than 0.62 micrometer?

(ii) What is the probability that a line width is between 0.47 and 0.63 micrometer?

(iii) What is the probability that a line width is smaller than 0.62 micrometer?

(iv) The line width of 90% of samples is below what value?

———————