Reg. No.

B.Tech/ M.Tech (Integrated) DEGREE EXAMINATION, MAY 2024

Fourth Semester

21ECE270T – STATISTICS FOR DATA SCIENCE

(For the candidates admitted from the academic year 2022-2023 onwards)

Note:

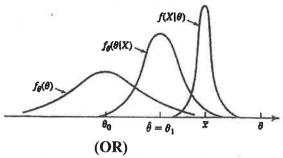
- Part A should be answered in OMR sheet within first 40 minutes and OMR sheet should be handed over to (i) hall invigilator at the end of 40th minute.

_	- C should be answered in	answer booklet.				
Time: 3 Hours			Max. I	Marl	ks: 7	15
	$PART - A (20 \times 1 = 20)$	Marks)	Marks	BL	CO	PO
1. Which of the follow	Answer ALL Quest wing is false?	tions	1	1	1	1,2,12
(A) probability P(a non-negativ	(A) of an event A is (B) re number) probability of the certain even equals 1				
(C) For two ever always write: P(B)	nts, A, B we can (D) $P(A \cup B) = P(A) +$	or If the events A, B are mutually exclusive, then P(A U B) = P(A) + P(B)	-			
2. For a continuous distribution of X in	random variable(X,Y) (X,Y) is obtained,	with Density $f(x,y)$, the marginal	1	2	1	1,2,12
(A) with the densi from f(x,y) by(C) with the densi	ity fl of X obtained (B) integration over x ity fl of X obtained (D)	from f(x,y) by integration over y with the density f1 of X obtained				
from $f(x,y)$ with x	by differentiation	from $f(x,y)$ by differentiation with y	ь			
 For a continuous ra P: probability a: sca which of the follows 	alar constant>0, z: rando	The $F(Z)$ = cumulative density function, m number	1	2	1	1,2,12
(A) $F(Z)=P(Z=z)$ (C) $F(z)=P(Z\leq z)$		$F(Z)=P(Z\geq z)$ $F(z)=P(Z=az)$				
 A call occurs at tin probability P(6 ≤ t ≤ 		om point in the interval (0,10). Find	1	2	1	.1,2,12
(A) 2/5 (C) 3/5	(B) (D)	2/10 3/10				
5. The motion of micr a:	oscopic particles in coll	ision with the molecules in a fluid is	1	2	2	2,12
stochastic proc		stochastic process				
(C) a stochastic p brownian moti	process but not a (D)	Neither a brownian motion nor a stochastic process				
	ring is true: cess is a countable (B) dom variables, one	A stochastic process is a noncountable infinity of random variables, one for each t.	1	2	2	2,12
		A stochastic process is a noncountable infinity of deterministic variables, one for each t			ě	

7	Whio	h of the following is true:			1	2	2	2,12
7.		h of the following is true: The mean of x(t) is the expected	(B)	The mean of $x(t)$ is the expected				
	(2.1)	value of the product $x(t1)$ $x(t2)$.	(-)	value of its autocorrelation				
	(C)		(D)	The mean of $x(t)$ is the expected				
		value of its autocovariance		value of the random variable x(t).		•	2	2.12
8.			t) as a	a stochastic process, the mean of the	1	2	2	2,12
	proce		(B)	f(t)				
	(A) (C)	f(-1) f2(t)	(B)	f3(t)				
9	` ′	sample standard deviation is	(-)	(+)	1	2	3	2,12
2.*			(B)	the square root of the sample mean				
	(C)	the square root of the sample autocorrelation	(D)	the square root of the sample autocovariance				
10.	The 1	runs scored by a batsman in 5 ODI	s are	31,97,112, 63, and 12. The standard	1	2	3	2,12
		ation is						
	` '	24.79	` '	23.79				
	` '	25.79	(D)	26.79		•	2	2.12
11.		sample covariance matrix of a data			1	2	3	2,12
	` '	the pairwise sample covariance		the pairwise sample correlation the pairwise sample autocovariance				
	(C)	the pairwise sample mean	(D)	the pan wise sample autocovariance	1	2	3	2,12
12.		ch of the statement is true?	(D)	The eigen-decomposition of the	1	2	,	2,12
	(A)	The eigen-decomposition of the covariance matrix is not possible	(D)	covariance matrix is possible				
	(C)	The Sample covariance matrix is	(D)	Sample covariance matrix is upper				
	` '	antisymmetric	, ,	triangular in nature				
13.	Whi	ch of the statements is true for freq	uency	y statistics?	1	1	4	2,12
	(A)	If the model is parametric, the	(B)	If the model is parametric, the				
		parameters are deterministic		parameters are random variables				
	(C)	quantities The data is modelled as	(D)	The data is modelled as				
	(C)	realizations from a distribution	` '	realizations from a distribution that				
		that is random		is infinite				
14.		ch of the statements is false for Ba			1	2	4	2,3,12
	(A)	There is a dual interpretation of		There is a single interpretation of				
		the physical meaning of probability		the physical meaning of probability				
	(C)	The probability $p(A)$ of an event	(D)	The probability p(A) is a				
	(-)	A is an objective" measure of		"subjective" measure of our state				
		the relative frequency of the		of knowledge concerning the				
		occurrence of A		occurrence of A	1	2	4	2,3,12
15.		ch of the statements is true for Bay				_	•	2,2,12
	(A)	Posterior distribution is intended to reflect our knowledge of the		to reflect our knowledge of the				
		parameter before we gather data		parameter after we gather data				
	(C)	•	(D)					
		related to Bayesian estimation		discrete case				
16.	Whi	ch of the statements is true for me			1	2	4	2,3,12
	(A)	-	(B)					
	(C)	the MSE The posterior variance	· (D)	MSE The posterior variance maximizes				
	(C)	minimizes the MSE	(1)	the MSE				
					10114	2150	годот	

	Which of the statements is false for Regression analysis If random variable X is the predictor, and random variable Y is the response? (A) The curve of regression of Y on (B) The curve of regression of X on Y X is used to describe the statistical relation between X, Y relation between X, Y (C) It is possible that the curve is a (D) if X and Y have a bivariate normal distribution, then the curve is a straight line Which of the following is true for the coefficient of correlation? (A) The coefficient of correlation is (B) The coefficient of correlation is not dependent on the change of scale (C) The coefficient of correlation is (D) The coefficient of correlation is a constant quantity	1	2	5	1,2
19.	origin Which of the following techniques is an analysis of the relationship between two variables to help provide the prediction mechanism? (A) Standard error (B) Correlation (C) Regression (D) Covariance	1	2	5	1,2
20.	Which of the following statements is true about the arithmetic mean of two regression coefficients? (A) It is less than the correlation (B) It is equal to the correlation coefficient (C) It is greater than or equal to the (D) It is greater than the correlation coefficient (D) It is greater than the correlation coefficient	1	2	5	1,2
	$PART - B (5 \times 8 = 40 Marks)$	Marks	BL	со	PO
21. a.	$PART - B \ (5 \times 8 = 40 \ Marks)$ Answer ALL Questions Independent trials, consisting of the flipping of a coin having probability p of coming up heads, are continually performed until either a head occurs or a total of n flips is made. Find an expression of the probability for 1 to n th flip. Also show that the sum of all probabilities is equal to 1.		BL 2	co 1	PO 1,2,12
	Answer ALL Questions Independent trials, consisting of the flipping of a coin having probability p of coming up heads, are continually performed until either a head occurs or a total of n flips is made. Find an expression of the probability for 1 to n th flip. Also show that the sum of all probabilities is equal to 1. (OR)				
	Answer ALL Questions Independent trials, consisting of the flipping of a coin having probability p of coming up heads, are continually performed until either a head occurs or a total of n flips is made. Find an expression of the probability for 1 to n th flip. Also show that the sum of all probabilities is equal to 1.	8	2	1	1,2,12
b.i.	Answer ALL Questions Independent trials, consisting of the flipping of a coin having probability p of coming up heads, are continually performed until either a head occurs or a total of n flips is made. Find an expression of the probability for 1 to n th flip. Also show that the sum of all probabilities is equal to 1. (OR) A fair coin is tossed 10000 times. What is the probability that the number of	8	2	1	1,2,12
b.i. ii.	Answer ALL Questions Independent trials, consisting of the flipping of a coin having probability p of coming up heads, are continually performed until either a head occurs or a total of n flips is made. Find an expression of the probability for 1 to nth flip. Also show that the sum of all probabilities is equal to 1. (OR) A fair coin is tossed 10000 times. What is the probability that the number of heads is between 4900 and 5100? Over a period of 12 hours, 180 calls are made at random. What is the probability	4	2	1	1,2,12
b.i. ii. 22. a.	Answer ALL Questions Independent trials, consisting of the flipping of a coin having probability p of coming up heads, are continually performed until either a head occurs or a total of n flips is made. Find an expression of the probability for 1 to n th flip. Also show that the sum of all probabilities is equal to 1. (OR) A fair coin is tossed 10000 times. What is the probability that the number of heads is between 4900 and 5100? Over a period of 12 hours, 180 calls are made at random. What is the probability that in a four-hour interval the number of calls is between 50 and 70? In the fair-coin experiment, we define the process $X(t)$ as follows: $X(t) = \sin \pi t$ if heads shows, $X(t) = 2t$ if tails shows. (i) Find mean $E[x(t)]$ (ii) Find $F(x, t)$ for $t=0.25$, $t=0.5$, $t=1$	4	2 2	1 1	1,2,12 1,2,12 1,2,12
b.i. ii. 22. a. b.	Answer ALL Questions Independent trials, consisting of the flipping of a coin having probability p of coming up heads, are continually performed until either a head occurs or a total of n flips is made. Find an expression of the probability for 1 to n th flip. Also show that the sum of all probabilities is equal to 1. (OR) A fair coin is tossed 10000 times. What is the probability that the number of heads is between 4900 and 5100? Over a period of 12 hours, 180 calls are made at random. What is the probability that in a four-hour interval the number of calls is between 50 and 70? In the fair-coin experiment, we define the process $X(t)$ as follows: $X(t) = \sin \pi t$ if heads shows, $X(t) = 2t$ if tails shows. (i) Find mean $E[x(t)]$ (ii) Find $F(x, t)$ for $t=0.25$, $t=0.5$, $t=1$	4 4 8	2 2 2 2	1 1 2	1,2,12 1,2,12 1,2,12 2,12

24. a. Suppose that $x = \theta + \nu$, where ν is an N $(0, \sigma)$ random variable and θ is the value of an N (θ_0, σ_0) random variable θ (Fig below). Find the bayesian estimate of $\hat{\theta}$ of θ .



- b. Let (X_1, \ldots, X_n) be the random sample of a Bernoulli random variable. X with pmf (probability mass function) given by $f(x; P) = p^x (1-p)^{(1-x)} x=0,1$ where $p, 0 \le p \le 1$ is unknown. Assume that p is a uniform random variable over (0, 1). Find the Bayes' estimator of p.
- 25. a. We consider the problem of building a linear model to predict the Gross Domestic Product (GDP) of a state (Tennessee) in the US from its population and unemployment rate. We have available the following data:

Danulation Ilmanulament

	GDP	Population	Unemployment
	(USD millions)		rate (%)
North Dakota	52 089	757 952	2.4
Alabama	204 861	4 863 300	3.8
Mississippi	107 680	2 988 726	5.2
Arkansas	120 689	2 988 248	3.5
Kansas	153 258	2 907 289	3.8
Georgia	525 360	10 310 371	4.5
Iowa	178 766	3 134 693	3.2
West Virginia	73 374	1 831 102	5.1
Kentucky	197 043	4 436 974	5.2
Tennessee	999	6 651 194	3.0

Build a linear model to predict the GDP.

(OR)

b. For a simple linear regression model, explain the estimation of parameters with 8 3 5 2,3,12 necessary mathematical expressions.

PART - C (1 × 15 = 15 Marks) Answer ANY ONE Ouestion Marks BL CO PO

- 26. Let x represent a binomial random variable with parameters n and p. Show that $\frac{15}{2}$ $\frac{2}{1}$ $\frac{1}{1,2,12}$
 - (i) E(x) = np;
 - (ii) $E[x(x-1)] = n(n-1)p^2$;
 - (iii) $E[x(x-1)(x-2)] = n\{n-1\}(n-2)p^3;$
 - (iv) Compute $E(x^2)$ and $E(x^3)$.
- 27. Explain the concept of Random Walks with necessary mathematical expressions, ¹⁵ ² ² ^{1,2,12} and also examples. Explain Wiener process.

* * * * *

2, 3, 12

2,3,12

2,3,12

3