INDIAN STATISTICAL INSTITUTE

POST GRADUATE DIPLOMA IN BUSINESS ANALYTICS - FIRST YEAR
SEMESTER EXAMINATION
FIRST SEMESTER, 2025-26

Date: 10.12.2025

STATISTICAL STRUCTURES IN DATA : PART B

TIME: $1\frac{1}{2}$ HOURS FULL MARKS: 60

Answers should be brief and to the point. If you use any result proved in class, clearly state the result before using it. This question paper carries 64 points. Answer as many as you can. The maximum you can score in 60.

1. (a) Suppose that $X_1, X_2, \ldots, X_n \overset{iid}{\sim} F$, where F is a continuous distribution function. If F_n denotes the corresponding empirical distribution function, for any given x, show that $F_n(x)$ converges to $F(x)$ in probability. [5]

   (b) If (X_1, X_2) follows a trinomial distribution (a multinomial distribution involving two variables) with the dispersion matrix given by

$$\begin{bmatrix} 8 & -4 \\ -4 & 12 \end{bmatrix},$$

   find the parameters of this trinomial distribution. [5]

2. (a) If X follows the normal distribution with the mean μ and variance σ^2, show that $E|X-\mu| = \sqrt{\frac{2}{\pi}}\,\sigma$. [4]

   (b) If (X, Y) follows the bivariate normal distribution $N_2(0, 0, 1, 1, 0.5)$, find the expected value of $\max\{X, Y\}$. [4]

   (c) Suppose that (X_1, X_2) follows a bivariate normal distribution. If $X_1 \sim N(5, 1)$, $X_2 \mid X_1 = 6 \sim N(8, 3)$ and $E(X_2 \mid X_1 = 4) = 10$, then find $E(X_1 \mid X_2 = 5)$. [6]

3. (a) Show that the sub-gradient of a function $f(\mathbf{x}) = \|\mathbf{x}\|$ at $\mathbf{x} = \mathbf{0}$ is $\{\mathbf{x} : \|\mathbf{x}\| \leq 1\}$, where $\|\cdot\|$ denotes the Euclidean norm, [4]

   (b) Consider a linear regression model $\mathbf{Y}_{n\times 1} = \mathbf{X}_{n\times 2}\boldsymbol{\beta}_{2\times 1} + \boldsymbol{\varepsilon}_{n\times 1}$ with two non-stochastic covariates (X_1 and X_2) and no intercept, where $E(\boldsymbol{\varepsilon}) = \mathbf{0}$ and Disp($\boldsymbol{\varepsilon}$)=$\sigma^2\mathbf{I}$. Also, assume that $\mathbf{1}_n^\top\mathbf{Y} = 0$, $\mathbf{Y}^\top\mathbf{Y} = 1$, $\mathbf{1}_n^\top\mathbf{X} = (0, 0)$ and $\mathbf{X}^\top\mathbf{X} = \mathbf{I}_2$, the 2×2 identity matrix. Let $\widehat{\boldsymbol{\beta}} = (\widehat{\beta}_1\ \widehat{\beta}_2)$ be the least squares estimate of $\boldsymbol{\beta} = (\beta_1\ \beta_2)$. If $\widehat{\beta}_1 = 0.8$ and $\widehat{\beta}_2 = -0.6$, find the minimizer of $\Delta(\boldsymbol{\beta}) = \|Y - \mathbf{X}\boldsymbol{\beta}\|^2$ subject to the condition $\beta_1^2 + \beta_2^2 \leq 0.16$. Also find the minimizer of $\Delta(\boldsymbol{\beta})$ given the condition $|\beta_1| + |\beta_2| \leq 0.16$. [3+3]

[Turn Over]

4. The following table gives the frequency distribution of the lifetimes of 200 LED bulbs manufactured by each of the two companies "Shining Sigma" and "Luminous-Lambda".

Liftime (in '000 hours)	0-1	1-2	2-3	3-4	4-5	> 5
Shining Sigma	12	38	29	59	40	22
Luminous Lambda	14	44	35	49	44	14

(a) Construct a suitable test (of 5% level) to check whether the lifetime distributions of the LED bulbs in these companies differ significantly. [6]

(b) If it is known that all bulbs from Shining Sigma stopped working before 10,000 hours, describe how you will test (at 5% level) whether the lifetime distribution of the LED bulbs in that company follows an exponential distribution. (Describe the methodology only, calculations are not needed.) [4]

5. Assignment [20]

Chi-Square Critical Values

df	$\chi^2_{0.10}$	$\chi^2_{0.05}$	$\chi^2_{0.025}$
1	2.706	3.841	5.024
2	4.605	5.991	7.378
3	6.251	7.815	9.348
4	7.779	9.488	11.143
5	9.236	11.070	12.833

df	$\chi^2_{0.10}$	$\chi^2_{0.05}$	$\chi^2_{0.025}$
6	10.645	12.592	14.449
7	12.017	14.067	16.013
8	13.362	15.507	17.535
9	14.684	16.919	19.023
10	15.987	18.307	20.483

df	$\chi^2_{0.10}$	$\chi^2_{0.05}$	$\chi^2_{0.025}$
11	17.275	19.675	21.920
12	18.549	21.026	23.337
13	19.812	22.362	24.736
14	21.064	23.685	26.119
15	22.307	24.996	27.488