GROUP A

1. Let $f(x) = x^2 - 2x + 2$. Let L_1 and L_2 be the tangents to its graph at $x = 0$ and $x = 2$ respectively. Find the area of the region enclosed by the graph of f and the two lines L_1 and L_2.

2. Find the number of 3×3 matrices A such that the entries of A belong to the set $\mathbb{Z}$ of all integers, and such that the trace of A^tA is 6.
(A^t denotes the transpose of the matrix A).

3. Consider n independent and identically distributed positive random variables $X_1, X_2, \dots, X_n$. Suppose S is a fixed subset of $\{1, 2, \dots, n\}$ consisting of k distinct elements where $1 \leq k < n$.

   (a) Compute
   $$\mathbb{E}\left[\frac{\sum_{i \in S} X_i}{\sum_{i=1}^{n} X_i}\right].$$

   (b) Assume that X_i's have mean μ and variance σ^2, $0 < \sigma^2 < \infty$. If $j \notin S$, show that the correlation between $(\sum_{i \in S} X_i)X_j$ and $\sum_{i \in S} X_i$ lies between $-\frac{1}{\sqrt{k+1}}$ and $\frac{1}{\sqrt{k+1}}$.

GROUP B

4. Let $X_1, X_2, \dots, X_n$ be independent and identically distributed random variables. Let $S_n = X_1 + \cdots + X_n$. For each of the following statements, determine whether they are true or false. Give reasons in each case.

   (a) If $S_n \sim Exp$ with mean n, then each $X_i \sim Exp$ with mean 1.

   (b) If $S_n \sim Bin(nk, p)$, then each $X_i \sim Bin(k, p)$.

5. Let $U_1, U_2, \ldots, U_n$ be independent and identically distributed random variables each having a uniform distribution on $(0, 1)$. Let

$$X = \min\{U_1, U_2, \ldots, U_n\}, \qquad Y = \max\{U_1, U_2, \ldots, U_n\}.$$

Evaluate $\mathbb{E}[X|Y = y]$ and $\mathbb{E}[Y|X = x]$.

6. Suppose individuals are classified into three categories C_1, C_2 and C_3. Let $p^2, (1-p)^2$ and $2p(1-p)$ be the respective population proportions, where $p \in (0, 1)$. A random sample of N individuals is selected from the population and the category of each selected individual recorded. For $i = 1, 2, 3$, let X_i denote the number of individuals in the sample belonging to category C_i. Define $U = X_1 + \frac{X_3}{2}$.

(a) Is U sufficient for p? Justify your answer.

(b) Show that the mean squared error of $\frac{U}{N}$ is $\frac{p(1-p)}{2N}$.

7. Consider the following model:

$$y_i = \beta x_i + \varepsilon_i x_i, \;\; i = 1, 2, \ldots, n,$$

where $y_i, i = 1, 2, \ldots, n$ are observed; $x_i, i = 1, 2, \ldots, n$ are known positive constants and β is an unknown parameter. The errors $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ are independent and identically distributed random variables having the probability density function

$$f(u) = \frac{1}{2\lambda} \exp\left(-\frac{|u|}{\lambda}\right), \;\; -\infty < u < \infty,$$

and λ is an unknown parameter.

(a) Find the least squares estimator of β.

(b) Find the maximum likelihood estimator of β.

8. Assume that $X_1, \ldots, X_n$ is a random sample from $N(\mu, 1)$, with $\mu \in \mathbb{R}$. We want to test $H_0 : \mu = 0$ against $H_1 : \mu = 1$. For a fixed integer $m \in \{1, \ldots, n\}$, the following statistics are defined:

$$\begin{aligned} T_1 &= (X_1 + \ldots + X_m)/m, \\ T_2 &= (X_2 + \ldots + X_{m+1})/m, \\ \vdots &= \qquad \vdots \\ T_{n-m+1} &= (X_{n-m+1} + \ldots + X_n)/m. \end{aligned}$$

Fix $\alpha \in (0, 1)$. Consider the test

$$\text{reject } H_0 \quad \text{if} \qquad \max\{T_i : 1 \leq i \leq n - m + 1\} > c_{m,\alpha}.$$

Find a choice of $c_{m,\alpha} \in \mathbb{R}$ in terms of the standard normal distribution function Φ that ensures that the size of the test is at most α.

9. A finite population has N units, with x_i being the value associated with the i^{th} unit, $i = 1, 2, \ldots, N$. Let $\bar{x}_N$ be the population mean. A statistician carries out the following experiment.

- Step 1: Draw a SRSWOR of size n ($< N$) from the population. Call this sample S_1 and denote the sample mean by $\bar{X}_n$.
- Step 2: Draw a SRSWR of size m from S_1. The x-values of the sampled units are denoted by $\{Y_1, \ldots, Y_m\}$.

An estimator of the population mean is defined as,

$$\widehat{T}_m = \frac{1}{m} \sum_{i=1}^{m} Y_i.$$

(a) Show that $\widehat{T}_m$ is an unbiased estimator of the population mean.

(b) Which of the following has lower variance: $\widehat{T}_m$ or $\bar{X}_n$?