Notations and Abbreviations

The following are used throughout the question paper.

$\mathbb{R}$	Set of real numbers
$\mathbb{R}^n$	n - dimensional Euclidean space
$\mathbb{Q}$	Set of rational numbers
$\mathbb{E}$	Expectation
$\mathbb{V}$	Variance
$\mathbb{P}$	Probability
$\mathbb{C}$ov	Covariance
A^T	Transpose of the matrix A
i.i.d.	independent and identically distributed
p.d.f.	probability density function

---

1. Let $m \neq n$ and let $A = ((a_{ij}))$ be an $m \times n$ real matrix such that $A^T A = I$. Let $b_1, \ldots, b_m \in \mathbb{R}$. Prove that the system of linear equations in the unknowns $x_1, \ldots, x_n$

$$\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = & b_1, \\ a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2n}x_n & = & b_2, \\ \vdots & & \vdots & & & & \vdots & & \vdots \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \cdots & + & a_{mn}x_n & = & b_m \end{array}$$

has at most one solution.

2. Use the identity $n^2 + (2n+1) = (n+1)^2$ to prove that the set $S = \{(x, y) \in \mathbb{Q} \times \mathbb{Q} : x^2 + y^2 = 1\}$ has infinitely many elements.

3. A box contains 7 indistinguishable green balls, 5 indistinguishable white balls and 6 indistinguishable black balls. Suppose balls are drawn using simple random sampling with replacement until balls of all colours are obtained. Let N denote the minimum number of draws needed to achieve this.

   (a) For each nonnegative integer n, calculate $\mathbb{P}(N > n)$.

   (b) Using (a) or otherwise, compute $\mathbb{E}(N)$.

4. Suppose (X, Y) is uniformly distributed over the region

$$\{(x, y) \in \mathbb{R}^2 : \ 0 < y < 1, \ |x| < 1 - y \}.$$

   Calculate the p.d.f. of $|X| + Y$.

5. Let X_1, X_2, X_3 be i.i.d. $N(0, 1)$ random variables. Define, for $1 \leq i \neq j \leq 3$,

$$W_{ij} = \begin{cases} 1 & \text{if } X_i > X_j, \\ 0 & \text{otherwise.} \end{cases}$$

   (a) Calculate $\mathbb{E}(W_{ij})$ and $\mathbb{V}(W_{ij})$ for all $1 \leq i \neq j \leq 3$.

   (b) Calculate $\mathbb{C}\text{ov}(W_{12}, W_{ij})$ for all $1 \leq i \neq j \leq 3$.

6. Assume that the length, in minutes, of a phone-call of an individual follows an exponential distribution with an unknown parameter $\lambda > 0$ with density function $f(x) = \lambda e^{-\lambda x}$, $x > 0$. However, when the phone company calculates the length of a phone-call, it always considers the nearest integer greater than or equal to the actual length. For example, a 22.09 minutes long phone-call will have a call-length of 23 minutes in the phone company records. Suppose you have the data on the lengths of n independent phone-calls $T_1, T_2, \ldots, T_n$ of that individual as reported by the phone company. Based on this data, compute the maximum likelihood estimator of λ.

7. Suppose that X_1, X_2 are i.i.d. $U(0,\theta)$ random variables. We want to test $H_0 : \theta = 1$ against $H_1 : \theta = \theta_1$ where $\theta_1 > 1$ is fixed. For this, we adopt two testing strategies.

$$\begin{array}{ll} \text{Test 1:} & \text{Reject } H_0 \text{ if } X_1 > 0.95. \\ \text{Test 2:} & \text{Reject } H_0 \text{ if } \max\{X_1, X_2\} > \kappa. \end{array}$$

(a) Find κ such that both the tests have the same size.

(b) Find the powers of the two tests. For Test 2, use the value of κ obtained in part (a).

(c) Which of the two tests would you prefer and why?

8. Suppose we have paired observations $(X_1, Y_1), \ldots, (X_n, Y_n)$. A statistician thinks that the following parametric model may be appropriate for this data:
For unknown odd integers α, β and unknown $\rho \in (-1, 1)$, the pairs (X_i^α, Y_i^β) are i.i.d. bivariate normal with parameters $(0, 0, 1, 1, \rho)$.
Write down the likelihood function for this model.

9. Let $Y_1, \ldots, Y_{2n}$ be i.i.d. Bernoulli(p) random variables, where $p \in (0, 1)$ is unknown. Consider the estimators

$$T_{1,n} = \frac{1}{2n-1} \sum_{i=1}^{2n-1} Y_i Y_{i+1} \quad \text{and} \quad T_{2,n} = \frac{1}{n} \sum_{i=1}^{n} Y_{2i-1} Y_{2i}.$$

(a) Show that both estimators are consistent for p^2.

(b) Find $\lim_{n\to\infty} (\mathbb{V}(T_{1,n})/\mathbb{V}(T_{2,n}))$.

(c) For large n, which estimator would you prefer and why?

10. A city has N households numbered $\{1, \ldots, N\}$. Let y_i denote the size of the i-th household. A simple random sample s_1 of size m is drawn without replacement from households $\{1, \ldots, M\}$. Another independent simple random sample s_2, also of size m, is drawn without replacement from households $\{N-M+1, \ldots, N\}$. Assume $m < M$ and $M > N/2$.

   (a) Compute $\pi_i = \mathbb{P}(i \in s_1 \cup s_2)$, the probability that the i-th household is included in either of the two samples, for $i = 1, \ldots, N$.

   (b) Show that T is an unbiased estimator of the average household size in the city, where

$$T = \frac{1}{N} \sum_{i \in s_1 \cup s_2} \left(\frac{y_i}{\pi_i} \right).$$