THE HONG KONG POLYTECHNIC UNIVERSITY
Department of Applied Mathematics

Subject Code: AMA563   **Subject Title:** Principles of Data Science

Programmes: Master of Science in Data Science and Analytics (63027)

Session: Semester 1, 2023/2024

Time: From: 19:00pm, 12 December 2023   **Time Allowed:** 3 hours

To: 22:00pm, 12 December 2023

This question paper has ___**3**___ pages (attachment included).

Instructions to Students:

- This is closed-book exam and the paper contains **6** questions.

- Please attempt all the **6** questions. **Please show all the steps.**

- Please note that you should follow the Regulations on Academic Integrity in Student Handbook and shall not give nor receive any unauthorized aid to/from any person or persons.

Attachment:

NA

Subject Lecturer: Dr. Ting Li

DO NOT TURN OVER THE PAGE UNTIL YOU ARE TOLD TO DO SO.

1. Let $(X_1, \ldots, X_n)$ be a random sample from the uniform distribution on the interval $(\theta, \theta + |\theta|)$. Find the MME and MLE of θ when

   (a) $\theta \in (0, \infty)$;   [5 marks]

   (b) $\theta \in (-\infty, 0)$;   [5 marks]

   (c) $\theta \in \mathcal{R}, \theta \neq 0$ [5 marks]

2. Suppose $X_1, \ldots, X_n$ are iid with geometric distribution. Their common probability function is
$$P(X_i = k) = \theta(1 - \theta)^k, \quad k = 0, 1, 2, \ldots$$

   (a) Find the sufficient and complete statistic and why it is.   [5 marks]

   (b) Write out the probability function of $\sum_{i=1}^{n} X_i$.   [5 marks]

   (c) Find the UMVUE of θ.   [10 marks]

3. Let X have the pdf $f(x; \theta) = \theta^x (1 - \theta)^{1-x}, x = 0, 1$, zero elsewhere. We test $H_0 : \theta = \frac{1}{2}$ and $H_1 : \theta < \frac{1}{2}$ by taking a random sample $X_1, \ldots, X_5$ of size $n = 5$.

   (a) What's the hypothesis, and what is the type of the statistical hypothesis? What is the Type I Error and what is the Type II Error?   [5 marks]

   (b) Develop a uniformly most powerful (UMP) test.   [5 marks]

   (c) What is the power function of the UMP test.   [5 marks]

   (d) What is the significance level of the UMP test.   [5 marks]

   (e) Develop the likelihood ratio test. Is it the same as the UMP test?   [5 marks]

4. Suppose $X_1, \ldots, X_n$ are iid $\sim N(\mu, 1)$.
   For hypotheses $H_0 : \mu = 0$ vs. $H_a : \mu \neq 0$, show that the test that rejects H_0 when $\sqrt{n}|\bar{X}| > z(\alpha/2)$ is NOT UMP at level α, here $z(\alpha/2)$ suits that $P(Z > z(\alpha/2)) = \alpha/2$ with Z from a standard normal distribution.   [15 marks]

5. Suppose $X_1, \ldots, X_n$ are iid $\sim N(\mu, \sigma^2)$ where μ is known.

   (a) Find a UMP at significance level α for $H_0 : \sigma^2 = 1$ v.s. $H_a : \sigma^2 > 1$. [10 marks]

   (b) Derive the generalized likelihood ratio test assuming μ is unknown.   [5 marks]

6. Answer following questions about network data analysis.

   (a) What is the definition of clustering coefficient? Use a real word example to illustrate its meaning.   [5 marks]

(b) Suppose we have a network generated from SBM. Given the adjacency matrix A and the community label that $z_i = 1$ for $i \in \{1, 2, 3, 4, 5\}$ and $z_j = 2$ for $j \in \{6, 7, 8, 9, 10\}$, find the MLE for all the other parameters, i.e., the connection matrix. [5 marks]

— END —