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, 2022/2023

Time: From: 19:00pm, 07 December 2022 Time Allowed: 3 hours

To: 22:00pm, 07 December 2022

This question paper has 4 pages (attachment included).

Instructions to Students:

- This is closed-book exam and the paper contains 5 questions.
- Please attempt all the 5 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:

1. Some Formulae

Subject Lecturer: Dr. Ting Li

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

1. Let $X_1, \ldots, X_n$ denote a random sample from the distribution with pmf

$$p(x;\theta) = \begin{cases} \theta^x(1-\theta)^{1-x}, & x = 0,1 \\ 0, & elsewhere, \end{cases}$$

where $0 \le \theta \le 1$. Find the MME and MLE of θ. [8 marks]

2. Let $X_1, \ldots, X_n$ be a random sample from a uniform distribution on the interval $[0, \theta]$, with $\theta > 0$.

   a. Find the method of moments estimator $\widehat{\theta}_{\mathsf{MME}}$ of θ. [6 marks]

   b. Find the maximum likelihood estimator $\widehat{\theta^2_{\mathsf{MLE}}}$ of θ^2. [6 marks]

   c. Find the variance of $\widehat{\theta}_{\mathsf{MME}}$. Is $\widehat{\theta}_{\mathsf{MME}}$ a consistent estimator for θ and why? [6 marks]

   d. Write out the pdf of $X_{(n)}$, here $X_{(n)}$ is the largest order statistic among all samples. [6 marks]

   e. Is $X_{(n)}$ a sufficient statistic for θ? Why or why not? [6 marks]

3. Let $X_1, \ldots, X_n$ be a random sample from a Gamma distribution with $\alpha = 5$ and $\beta = \theta > 0$.

   a. Find the moment estimator of θ. [6 marks]

   b. Find the maximum likelihood estimator for θ. [6 marks]

   c. Find the Fisher information $I(\theta)$. [6 marks]

   d. Show that the MLE, $\widehat{\theta}$, of θ is an efficient estimator of θ. [6 marks]

4. Let $X_1, X_2, \ldots, X_n$, $n > 2$, be a random sample from a binomial distribution $b(1, \theta)$.

   a. Find the complete sufficient statistic for θ. [6 marks]

   b. Find the MVUE of θ. [6 marks]

   c. Derive the UMVUE of $\tau(\theta)$, where $\tau(\theta) = e^2(\theta(1-\theta))$. [6 marks]

5. Let X have the pdf $f(x;\theta) = \theta^x(1-\theta)^{1-x}, x = 0, 1$, zero elsewhere. We test $H_0 : \theta = \frac{1}{3}$ and $H_1 : \theta < \frac{1}{3}$ by taking a random sample $X_1, ..., X_4$ of size $n = 4$ and rejecting H_0 if $Y = \sum_{i=1}^n X_i$ is observed to be less than or equal to a constant c.

   a. Show that this is a uniformly most powerful test. [8 marks]

b. Find the significance level when $c = 0$. [6 marks]

c. Find the significance level when $c = 2$. [6 marks]

— END —

Attachment 1: Some Formulae

The gamma distributions, $\underline{\mathsf{Gamma}(\alpha,\beta)}$: $\alpha > 0$ and $\beta > 0$.

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}, \quad x > 0.$$

$$\mu = \alpha\beta, \quad \sigma^2 = \alpha\beta^2, \quad M(t) = (1-\beta t)^{-\alpha}, \quad t < \frac{1}{\beta}.$$

$$X_1, \ldots, X_m \sim \mathsf{Gamma}(\alpha,\beta) \text{ and independent} \Longrightarrow \sum_{i=1}^{m} X_i \sim \mathsf{Gamma}(m\alpha, \beta).$$

The binomial distributions, $\underline{\mathsf{Binomial}(n,p)}$: $0 < p < 1$, and $n = 1, 2, \ldots$,

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, 2, \ldots, n.$$

$$\mu = np, \quad \sigma^2 = np(1-p), \quad M(t) = ((1-p) + pe^t)^n, \quad -\infty < t < \infty.$$

The gamma function, $\Gamma(x)$, defined on $x \in (0, \infty)$, by

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt.$$

$\underline{\text{Properties of } \Gamma(x)\text{:}}$

- $\Gamma(n+1) = n!$ for $n = 0, 1, 2, \ldots$.
- $\Gamma(x+1) = x\Gamma(x)$ for $x > 0$.
- $\Gamma(x)\Gamma(1-x) = \frac{\pi}{\sin(\pi x)}$ for $0 < x < 1$.
- $\Gamma(\frac{1}{2}) = \sqrt{\pi}$.