MACHAKOS UNIVERSITY

University Examinations for 2024/2025 Academic Year

SCHOOL OF PURE AND APPLIED SCIENCES
DEPARTMENT OF MATHEMATICS, STATISTICS AND ACTUARIAL SCIENCES

FOURTH YEAR FIRST SEMESTER EXAMINATION FOR DEGREE IN BACHELOR OF SCIENCE (STATISTICS AND COMPUTER PROGRAMMING) BACHELOR OF EDUCATION (SCIENCE)

SMA 361: THEORY OF ESTIMATION

DATE: 20TH JANUARY, 2025 **TIME: 2.00 PM – 4.00 PM**

INSTRUCTIONS:
*Answer question **ONE (Compulsory)** and **any other TWO** questions*

QUESTION ONE (30 MARKS), COMPULSORY

(a) Differentiate between point estimation and interval estimation. (2 marks)

(b) Using mathematical symbols define completeness as it is used in statistic. (3 marks)

(c) Suppose that 6.5, 9.2, 9.9, and 12.4 constitute a random sample of size 4 from the pdf

$$f_Y(y;\mu) = \frac{1}{\sqrt{2\pi}(0.8)} e^{-\frac{1}{2}\left(\frac{y-\mu}{0.8}\right)^2} \qquad -\infty < y < \infty$$

That is, the four y_i's come from a normal distribution where σ is equal to 0.8, but the mean, μ, is unknown. What values of μ are believable in light of the four data points? (5 marks)

(d) Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from a population with unknown mean μ. Show that the sample mean, $\bar{x}$,

i. is an unbiased estimator for μ (3 marks)

ii. has variance σ^2 (2 marks)

(e) consider a random sample $X_1, X_2, \ldots, X_n$ of size n from a population with mean μ and variance σ^2. The statistic $T_1 = \frac{1}{3}(2X_1 + X_2)$ is unbiased for μ and the statistic $T_2 = X_3$ is also unbiased for μ. Determine the efficient estimator between T_1 and T_2. (4 marks)

(f) When the mean of a normal population is estimated on the basis of a random sample of size $(2n + 1)$, what is the relative efficiency of the mean with respect to the median? (4 marks)

(g) Show that the statistic $T = \frac{1}{5}(X_1 + 2X_1)$ is not consistent for μ. (3 marks)

(h) Define the r^{th} raw moment about the origin. (2 marks)

(i) State the Cramer-Rao inequality. (2 marks)

QUESTION TWO (20 MARKS)

(a) Let $y_1, y_2, ..., y_n$ be independent variables. Given, $E(y_i) = a_{1i}P_1 + a_{2i}P_2 + ... + a_{mi}P_m$, for $i = 1, 2, ..., n$ and a_i are constants and that $V(y_i) = \sigma^2$ constant.

Show that

i. the least squares estimator of $\underline{P}$ is given by $\underline{\hat{P}} = (AA')^{-1}A\underline{y}$ (7 marks)

ii. $\text{Var}(\underline{P}) = (AA')^{-1}\sigma^2$ (3 marks)

iii. $\underline{\hat{P}}$ is an unbiased estimator of $\underline{P}$ (3 marks)

(b) Y_1, Y_2 and Y_3 are independent random variables with common variance, σ^2, and $E(Y_1) = \theta_1$, $E(Y_2) = \theta_1 + \theta_2$ and $E(Y_3) = \theta_1 + 2\theta_2$. Find the least squares estimators of θ_1 and θ_2 and their means and variances. (7 marks)

QUESTION THREE (20 MARKS)

(a) State two properties of sufficient statistics. (2 marks)

(b) Define the term *minimal sufficient statistics*. (1 mark)

(c) Briefly define the Fisher and Neyman factorization technique of determining sufficient statistics. (3 marks)

(d) Consider $X \sim N(\mu, \sigma^2)$ where σ^2 is known. Using the Fisher and Neyman factorization technique show that $T = \sum_{i=1}^{n} x_i$ is a sufficient statistic for μ. (8 marks)

(e) Let $X_1, X_2, \ldots, X_n$ be iid $N(\mu, \sigma^2)$. Find the minimal sufficient statistic for $\theta = (\mu, \sigma^2)$. (6 marks)

QUESTION FOUR (20 MARKS)

(a) Suppose the random variables $X_1, X_2, \ldots, X_n$ denote the number of successes (0 or 1) in each of *n* independent trials, where $p = P$ (Success occurs at any given trial) is an unknown parameter. Then

$$p_{X_i}(k;\ p) = p^k(1 - p)^{1-k}, \quad k = 0, 1; \quad 0 < p < 1$$

Let $X = X_1 + X_2 + \cdots + X_n = total\ number\ of\ successes$ and define $\hat{p} = \frac{X}{n}$. Clearly, $\hat{p}$ is unbiased for *p*. How does var($\hat{p}$) compare with the Cramer-Rao lower bound for $p_{X_i}(k;\ p)$?

(6 marks)

(b) State three properties of maximum likelihood estimators. (3 marks)

(c) Suppose a random sample of size n is drawn from a Bernoulli distribution given by

$$f(x;\ \theta) = \theta^x(1-\theta)^{1-x} \qquad x = 0, 1 \ \ and\ 0 \le \theta \le 1.$$

Find the maximum likelihood estimate for θ. (5 marks)

(d) Let $X_1, X_2, \ldots, X_n$ be a random sample from a normal distribution, $N(\theta, 1)$, $-\infty < \theta < \infty$. Find the maximum likelihood estimator (MLE) for θ. (6 marks)

QUESTION FIVE (20 MARKS)

(a) Max, a video game pirate (and Bayesian), is trying to decide how many illegal copies of *Zombie Beach Party* to have on hand for the upcoming holiday season. To get a rough idea of what the demand might be, he talks with *n* potential customers and finds that $X = k$ would buy a copy for a present (or for themselves). The obvious choice for a probability model for *X*, of course, would be the binomial pdf. Given *n* potential customers, the probability that *k* would actually buy one of Max's illegal copies is the familiar

$$p_x(k \mid \theta) = \binom{n}{k}\theta^k(1-\theta)^{n-k} \quad \text{for } k = 0, 1, \ldots, n$$

where the maximum likelihood estimate for θ is given by $\theta_e = \frac{k}{n}$

It may very well be the case, though, that Max has some additional insight about the value of θ on the basis of similar video games that he illegally marketed in previous years. Suppose Max suspects, for example, that a reasonable prior distribution for Θ is the *beta pdf*. Written with Θ as the random variable, the (two parameter) beta pdf is given by

$$f_\theta(\theta) = \frac{\Gamma(r+s)}{\Gamma(r)\Gamma(s)}\theta^{r-1}(1 - \theta)^{s-1} \qquad 0 \le \theta \le 1$$

Determine the bayer estimator. (12 marks)

(b) Predicting the annual number of hurricanes that will hit the U.S. mainland is a problem receiving a great deal of public attention, given the disastrous summer of 2004, when four major hurricanes struck Florida causing billions of dollars of damage and several mass evacuations. It has been determined that the obvious pdf for modeling the number of hurricanes reaching the mainland is the Poisson, where the unknown parameter θ would be the expected number in a given year. The table below shows the numbers of hurricanes that actually did come ashore for three fifty-year periods. Use that information to construct a posterior distribution for θ. Assume that the prior distribution is a gamma pdf. (8 marks)

Years	**Number of Hurricanes**
1851–1900	88
1901–1950	92
1951–2000	72