STRATHMORE INSTITUTE OF MATHEMATICAL SCIENCES
MASTER OF SCIENCE IN DATA SCIENCE AND ANALYTICS
END OF SEMESTER EXAMINATION
DSA 8302 COMPUTATIONAL TECHNIQUES IN DATA SCIENCE

DATE: 25th July 2023 Time: 2 Hours

Instructions

1. This examination consists of **FOUR** questions and an appendix to one of the questions.
2. Answer **Question ONE (COMPULSORY)** and any other **TWO** questions.

Question 1 (20 Marks)

a. Distinguish between numerical and analytical methods in the solution of mathematical problems.

(3 Marks)

b. Describe the bisection method and explain how it differs from the Newton-Raphson method.

(4 Marks)

c. Consider a random sample of size n from the $Bin(1, \pi)$ distribution, with the probability density function

$$f(x;\pi) = \pi^x(1-\pi)^{1-x},\ x = 0,1$$

i) Derive expressions for the likelihood, log-likelihood, score, and information based on a random sample of size n obtained from this population.

(6 Marks)

ii) Derive expressions for the maximum likelihood estimator (MLE) and standard error of π,

(4 Marks)

iii) If the sample estimate of π based on a random sample of 36 is $\hat{\pi} = 0.22$, determine a 95% confidence interval for π.

(3 Marks)

Question 2 (20 Marks)

The probability density function of a continuous random variable Y from the Cauchy distribution is given as

$$g(y) = \frac{1}{\pi[1 + (y - \theta)^2]}$$

a) Derive expressions for the likelihood, log-likelihood, score, and information based on a random sample of size n obtained from this population.

(12 Marks)

b) Write a program or computational algorithm that would be used to determine the maximum likelihood estimate of θ. Ensure that the algorithm has a starting value and a stopping rule and also provides an output of the hessian and log-likelihood.

(8 Marks)

Question 3 (20 Marks)

a) Consider the following data:

X	1	2	4
Y	1	3	1

Use quadratic spline interpolation to find the approximate value of y at x=3.

Hint:

$$\begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 0 \\ 4 & 0 & 2 & 0 & 1 & 0 \\ 0 & 4 & 0 & 2 & 0 & 1 \\ 0 & 16 & 0 & 4 & 0 & 1 \\ 4 & -4 & 1 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}^{-1} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 1 \\ 0.5 & -0.5 & -0.25 & 0.25 & 0.5 & -0.5 \\ -1.0 & 1.0 & 0.00 & 0.00 & 0.0 & -3.0 \\ -3.0 & 3.0 & 1.00 & -1.00 & -3.0 & 3.0 \\ 2.0 & -1.0 & 0.00 & 0.00 & 0.0 & 2.0 \\ 4.0 & -4.0 & 0.00 & 1.00 & 4.0 & -4.0 \end{pmatrix}$$

(10 Marks)

b) Consider the following data

$$(x_1, y_1) = (300{,}0.616);\ (x_2, y_2) = (400{,}0.525);\ (x_3, y_3) = (500{,}0.457);$$

Use polynomial interpolation to determine the value of the function at x=420.

Hint:

$$\begin{pmatrix} 90000 & 300 & 1 \\ 160000 & 400 & 1 \\ 250000 & 500 & 1 \end{pmatrix}^{-1} = 10^5 \begin{pmatrix} 5 & -10 & 5 \\ -4500 & 8000 & -3500 \\ 10^6 & -1.5 \times 10^6 & 0.6 \times 10^6 \end{pmatrix}$$

(10 Marks)

Question 4 (20 Marks)

a) Show that the probability density function of the Bernoulli distribution,

$$f(y) = \pi^y(1-\pi)^{1-y}, \mathrm{y} = 0, 1,$$

belongs to the exponential dispersion family

$$f(y;\theta) = \exp\left[\frac{y\theta - b(\theta)}{\phi}\right] + c(y;\phi).$$

(3 Marks)

b) Also show that the mean and the variance of the Binomial distribution are equal to $b'(\theta)$ and $b''(\theta)$, respectively.

(3 Marks)

c) Consider the generalized linear model

$$g(\mu_i) = \boldsymbol{X}_i\boldsymbol{\beta}$$

where $\mathbf{Y} = (Y_1, \dots, Y_n)'$ is an $(n \times 1)$ vector of response values belonging to the Bernoulli distribution, $\mathbf{X}$ is an $(n \times k)$ design matrix corresponding to the explanatory variables $X_1, \dots, X_k$, $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_k)'$ is the vector of parameters.

Derive an expression for the estimating equation and Hessian that would be used to estimate the vector of parameters $\boldsymbol{\beta}$.

(14 Marks)