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: July 22, 2024. Time: 3 Hours

Instructions

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

Question 1

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.

(3 Marks)

c) Given an initial guess $x=3$, find an approximate value of the root of the function $f(x) = x^2 - 2$ using 3 iterations. Provide a sample R code that you would use to solve this problem, stopping after 40 iterations.

(6 Marks)

d) Using the inverse-transform approach to explain (mathematically) how you would generate random numbers from the exponential distribution, $f(y) = \theta \exp - (\theta y)$, and further provide an R code that will be used to generate random numbers from this distribution.

(8 Marks)

Question 2

Starting with the Newton-Raphson formula

$$x^{(n+1)} = x^{(n)} - \frac{f(x^{(n)})}{f'(x^{(n)})}$$

shown that the order of convergence of the Newton-Raphson method is

$$|\varepsilon^{(n+1)}| = k|\varepsilon^{(n)}|^2$$

as $n \to \infty$, with $k = \frac{1}{2}\left|\frac{f''(r)}{f'(r)}\right|$ provided $|f'(r)| = 0$.

(20 Marks)

Question 3

a) The probability density function of the logistic distribution is given by

$$f(x) = \frac{1}{\beta}\frac{e^{-(x-\mu)/\beta}}{[1 + e^{-(x-\mu)/\beta}]^2}$$

Derive expressions for its cumulative distribution and inverse cumulative distribution functions and hence explain how random numbers from this distribution would be generated.

(9 Marks)

b) Suppose that $Y_1, \dots, Y_n$ random numbers from a $\text{Exp}(1)$, and that

$$Z = 2\sum_{i=1}^{n} Y_i \sim \chi^2(2n)$$

Using the moment generating function techniques, show that $Z \sim \chi^2(2n)$ and hence explain how you would generate random numbers from the exponential distribution.

(11 Marks)

Question 4

Consider the general linear model

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}, \boldsymbol{\varepsilon} \sim MVN(\boldsymbol{0}, \sigma^2\boldsymbol{I}),$$

where $\mathbf{Y} = (Y_1, \dots, Y_n)'$ is an $(n \times 1)$ vector of response values, $\mathbf{X}$ 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, and σ^2 is the variance.

Using a maximum likelihood approach, clearly showing the likelihood function, the log-likelihood function, and the score-vector, derive the maximum likelihood estimators of $\boldsymbol{\beta}$ and σ^2.

(20 Marks)

Question 5

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 Bernoulli 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 $\mathbf{Y} = (Y_1, \ldots, Y_n)'$ 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 $X_1, \ldots, X_k$, $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_k)'$ 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)