The University of Nottingham

SCHOOL OF MATHEMATICAL SCIENCES

AUTUMN SEMESTER 2022-2023

MATH4065 - STATISTICAL FOUNDATIONS

Assessed Coursework

Your neat, clearly-legible solutions should be submitted electronically as a pdf file via the MATH4065 Moodle page by the deadline indicated there. A scan of a handwritten solution is acceptable. Since this work is assessed, your submission must be entirely your own work (see the University's policy on Academic Misconduct). Submissions up to five working days late will be subject to a penalty of 5% of the maximum mark per working day.

IMPORTANT: Your answers should include, where appropriate

(i) any R code you used;
(ii) any plots you produce;
(iii) explanations of what you are doing and why.

1. For this question you will need the file "ExperimentData.csv" which can be downloaded from the module Moodle page. This file contains the results of an experiment in which the input variables (i.e. the predictor variables) are $X1$ and $X2$ and the output variable (i.e. the response variable) is Y.

   (a) Read the data file into R and investigate whether or not there is any graphical evidence for relationships between Y and the predictor variables. [5 marks]

   (b) Using R, fit a linear model (model 1) in which Y is the response variable, $X1$ is the predictor variable and there is also an intercept term. Find the least-squares estimates of the model parameters. Is there evidence that $X1$ can explain the variation in Y, or not? [5 marks]

   (c) Using R, fit a new model (model 2) in which Y is the response variable, both $X1$ and $X2$ are predictor variables, and there is an intercept term. Is there evidence that model 2 fits the data significantly better than model 1, or not? [5 marks]

   (d) Explore how well model 2 fits the data by considering the residuals. [5 marks]

2. Consider a random sample $X_1, \ldots, X_n$ from the continuous distribution with probability density function

$$f(x) = \frac{1}{\sigma} \frac{\mathrm{e}^{x/\sigma}}{(1 + \mathrm{e}^{x/\sigma})^2}, \qquad -\infty < x < \infty,$$

where $\sigma > 0$.

(a) Derive the log-likelihood function $l(\sigma)$. [2 marks]

(b) Use R to find the maximum likelihood estimate of σ given observed data

$$X_1 = 0.337, X_2 = 0.507, X_3 = 0.250, X_4 = 3.131, X_5 = 6.908,$$
$$X_6 = 0.195, X_7 = -3.097, X_8 = 1.002, X_9 = 2.011, X_{10} = 1.710$$

[5 marks]

(c) Produce a plot of $l(\sigma)$ over a suitable range of values with the maximum likelihood estimate of σ shown. [3 marks]

3. Let $\{X_n : n = 1, 2, \ldots\}$ be a Markov chain in which $X_1 = 0$ and for $n = 1, 2, \ldots$,

$$X_{n+1} = 2 - X_n + \varepsilon_n,$$

where $\varepsilon_1, \varepsilon_2, \ldots$ are independent random variables such that

$$\varepsilon_n = \begin{cases} +1 & \text{with probability 1/2,} \\ -1 & \text{with probability 1/2.} \end{cases}$$

(a) Write a function in R whose input is a positive integer T and whose output is a single realisation of $\{X_1, X_2, \ldots, X_T\}$. [4 marks]

(b) Produce a plot showing a single realisation of $\{X_1, X_2, \ldots, X_{10}\}$, with the axes suitably labelled. [2 marks]

(c) Use your function to find numerical estimates of the variance of X_{100} and $P[X_{21} = 0]$. You will not be given marks for using any other method. [4 marks]