University of Technology Sydney
School of Mathematical and Physical Sciences

Mathematical Statistics (37262) –
Quiz 2
Time allowed: 75 minutes
Question paper MUST NOT BE REMOVED from the room
This is an open-book assessment.
You may use textbooks, subject notes and online resources as required. You are not permitted to communicate with other students or to post requests for solutions to online sources

Question 1.

A negative binomial variable with range $\{r, r+1, r+2, ...\}$ can be considered as the sum of r independent identically distributed geometric random variable. It describes how many independent identically distributed Bernoulli ($Bern(p)$) variables must be observed until r successes (1s) have been observed.

For $X \sim NegBin(r,p)$, the probability mass function is

$$P(X=k) = \begin{cases} \dfrac{(k-1)!}{(r-1)!(k-r)!} p^r (1-p)^{k-r} & k \in \{r, r+1, r+2, ...\} \\ 0 & \text{otherwise} \end{cases}.$$

Assume that $r \in \mathbb{Z}^+$ is known but $p \in (0,1)$ is not and that we have a sample $\boldsymbol{X} = \{x_1, x_2, ..., x_k\}$ of independent realisations of $X \sim NegBin(r,p)$.

i) Justify that $E(X) = \dfrac{r}{p}$. Your justification may be a calculation or a verbal explanation of why this is the case.

Hint: You may wish to consider the expectation of a geometric random variable in your answer.

ii) Hence show that using the method of moments with the first sample moment estimates p by $\hat{p}_{MM} = \dfrac{nr}{\sum_{i=1}^{n} x_i}$.

iii) Given the sample $\boldsymbol{X} = \{x_1, x_2, ..., x_k\}$ calculate the associated likelihood function $L(\boldsymbol{X} \mid p)$.

iv) Showing all of your working, calculate the maximum likelihood estimate of p. Comment on your answer.

(10 Marks)

Question 2.

Consider the problem of estimating the value of the definite integral $\int_{-\infty}^{\infty} e^{-x^2}\cos(x)dx$ by the Monte Carlo method, given 5 independent realisations $\{u_1,u_2,u_3,u_4,u_5\} = \{0.310,\ 0.444,\ 0.791,\ 0.588,\ 0.237\}$ of a $U[0,1]$ variable.

i) Using the change of variable $y = \dfrac{e^x}{1+e^x}$, obtain an estimate of the value of the definite integral. Show all stages in your working.

ii) Clearly explain how you could instead use the transformation $z = e^x$

(8 Marks)

Question 3.

Consider the problem of fitting the linear regression model $y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \varepsilon_i$. where, for example, x_{2i} represents the ith observation of the predictor variable x_2.

i) Given the observations

y	7.6	7.3	7.7	8.2	8.9
x_1	1	2	3	4	5
x_2	0	0	1	1	2
x_3	3	2	2	2	1

write the regression model in the form $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$, clearly defining the matrices $\boldsymbol{X}$, $\boldsymbol{Y}$ and $\boldsymbol{\beta}$.
The entries in some of your matrices should be numerical values oand in others, the entries should be unknown parameters which may be estimated. Note, you do not need to estimate the values of these, just to define the matrix form of the model.

ii) If we instead consider the dataset

y	7.6	7.3	7.7	8.2	8.9
x_1	1	2	3	4	5
x_2	0	0	1	1	2
x_3	2	4	6	8	10

explain why we could not obtain a unique solution for the model parameter estimates in this case.

(6 Marks)

Question 4.

Consider the problem of fitting the linear regression model $y_i = \alpha + \beta x_i + \varepsilon_i$.

i) For the dataset

y	0	0	1	1
x	0	1	0	1

write down the regression model obtained by ordinary least squares. Justify your answer.

ii) Write down the R-squared value associated with this model. Justify your answer.

Note: You may answer the above parts without any calculation if you can explain your answer clearly in your own words.

iii) Construct a dataset consisting of at least three (x, y) observations such that the ordinary least squares regression line would have an associated R-squared value of 1.

For example, your answer may be something like $(1,2)$, $(3,0)$, $(2,1)$ and $(3,3)$.

(6 Marks)