Assignment 1
Probability & Data (MATH 4043)

Instructions

- This assignment is worth 25% of your final grade and is due - 12pm Friday Week 6.
- Submission is online on the Learnonline website.
- Assignments will be marked and returned online via Learnonline.
- R output will be required to be produced for some of the solutions.
- Your answers should be typed and submitted as a PDF.
- The marks for each question are displayed next to the question.
- The assignment is worth a total of 80 marks.
- **A late submission will attract a penalty of 10% of maximum marks available per day, or part thereof, if the assignment is late. The cut-off time is <u>5pm</u> each day.**

Questions

1. **(22 marks)** Ten small airplanes attempt independently of each other to land on a runway with the ground marked ABC (in order), where B is the "sweet" and aiming point for landing of the line AC. If the plane lands too early (A) then the plane is flying in too low, or if the plane lands too late (C) then the stopping distance becomes unsafe and must abort after position C. Each airplane can land somewhere on the runway with probability 0.9. Otherwise, the airplane ends up landing somewhere else or comes back around. If any one of the airplanes lands on the runway, then it is noticed that it will land in the AB section with probability 0.4, or in the BC section with probability 0.6.

a. **(4 marks)** Define the distribution for X and calculate the probability that any 6 out of the 10 airplanes manage to land somewhere on the runway.

b. **(5 marks)** What is the probability that any 6 out of the 10 airplanes manage to land somewhere on the AB section of the runway? The other ones could have either landed in the BC section or not on the runway at all.

c. **(4 marks)** Use Venn diagrams and probability axioms to prove that
$$(A_1 \cap A_2 \cap A_3^c) \cap (A_1 \cap A_2^c \cap A_3) = \emptyset$$
where A_1, A_2, A_3 are three sets.

d. **(4 marks)** *Flysafe* is one of the airplanes and always lands safely. The airport would like to analyse the scenario that a plane cannot land due to another airplane landing and FlySafe lands. Assume that most commonly six airplanes arrive close to one another, and it has been observed that usually one of the airplanes fails to land when *Flysafe* lands. What is

the probability that *Flysafe* lands and one of the other five fails? *Hint: Use a result similar to part c.*

e. **(5 marks)** What is the probability that six airplanes manage to land somewhere on the runway, given that *Flysafe* is one of them?

2. (10 marks) Consider an individual chosen at random from a city of interest, and let

$$X = 0, \text{ if the individual has neither lung cancer or tuberculosis disease;}$$
$$X = 1, \text{ if the individual has lung cancer;}$$
$$X = 2, \text{ if the individual has tuberculosis disease.}$$

It is known from previous research that

$$P(X = 0) = 0.97, P(X = 1) = 0.01, P(X = 2) = 0.02.$$

These probabilities constitute the probability distribution of X prior to obtaining any information about the individual. We investigate further and take an X-ray of the individual. We can summarise the new information with the variable, Y. If the X-ray is positive then $Y = 1$, and if the X-ray is not "positive" then $Y = 0$.

Suppose the X-ray is positive, and we know

$$P(Y = 1 \mid X = 0) = 0.07,$$
$$P(Y = 1 \mid X = 1) = 0.95,$$
$$P(Y = 1 \mid X = 2) = 0.90.$$

It is critical to understand what the probabilities of no diseases, lung cancer and tuberculosis disease given the X-Ray is positive ($Y = 1$). Find the probabilities $P(X_i \mid Y = 1)$ for $i = 0, 1, 2$ and interpret each probability.

3. (12 marks) A fair coin is tossed 3 times. Let $X = (X_1, X_2)$, where X_1 counts the number of heads in the 3 tosses, and X_2 counts the number of tails in the 3 tosses.

a. **(2 marks)** Write down the sample space S.
b. **(3 marks)** Tabulate the joint probability mass function of X_1 and X_2.
c. **(3 marks)** Calculate the marginal probability mass function of X_1 and X_2. What can you deduce about them?
d. **(4 marks)** Calculate $\rho(X_1, X_2)$, the correlation coefficient. Interpret this value and give an explanation for your answer.

4. (18 marks) Computation Marketing: what is the expected revenue?

- Please download the sales data from Learnonline, ***sales.csv***.

You have started working at a sales company and they have asked you to investigate a particular product, its sales and the expected revenue where the cost of the product is $80. The data which has been given to you has three columns, "Date", "Sales" and "Price ($)".

a. **(2 marks)** Identify the most appropriate distribution that models the variable, *Sales*.
b. **(2 marks)** Use R to compute the mean of the *Price ($)* in the dataset over the time period.
c. **(3 marks)** Use R to calculate the revenue and attach the daily revenue to the right of the sales table in *Sales.csv*. Please provide a screenshot of the daily revenue in your solution.
d. **(2 marks)** Use R to calculate the mean and variance of the *revenue ($)* over the 61 days.
e. **(3 marks)** Use R to compute the probability density function for the number of *Sales* per day and use R to plot a histogram of the number of *Sales* and the *revenue*. Assume that the maximum sold per day is 20 units.
f. **(4 marks)** The company chooses to fix the *Price ($)* per day to be the expected (mean) of *Price ($)* over 61 days. The company states that the lowest revenue amount is $1021 before there is a loss of money. What is the probability that the *revenue* is less than $1021?
g. **(2 marks)** Summarise the results in parts a) – f) and give a conclusion in relation to sales and revenue when the company fixes the price.

5. (18 marks) *Multiple Choice, testing to see if you can count on luck.*

Suppose we conduct an experiment on taking a multiple choice test, let there be 20 questions, each with four possible answers, one of which is correct. We will test this over a large class of 50 students where each student guesses the answers randomly without reading the question. We place the condition that a student passes this test, if they score 8 or more correct answers. Let X be the number of correct answers for a randomly selected student.

a. **(2 marks)** What is the distribution for X?
b. **(3 marks)** What is the expected value for X and the standard deviation for X?
c. **(4 marks)** Use R to plot a probability density function (p.d.f) for X, produce a table for the p.d.f and calculate the probability of a student getting 8 or more correct answers.
d. **(4 marks)** Suppose we randomly select 10 students, what is the probability that at least three students pass just by guessing? From the 10 students, what is the expected number of students who will pass?
e. **(5 marks)** Use R to simulate the number of correct answers someone gets for 50 students to populate a sample for the class. Compute and compare the mean, variance, plot the distribution and probability of passing to the distribution of X.