Queen Mary
University of London

Main Examination period 2023 – January – Semester A

MTH5129: Probability & Statistics II

Duration: 2 hours

The exam is intended to be completed within **2 hours**. However, you will have a period of **4 hours** to complete the exam and submit your solutions.

For actuarial students only: This module also counts towards IFoA exemptions. For your submission to be eligible, **you must submit within the first 3 hours.**

You should attempt ALL questions. Marks available are shown next to the questions.

All work should be **handwritten** and should **include your student number**. Only one attempt is allowed – **once you have submitted your work, it is final**.

In completing this assessment:

- You may use books and notes.
- You may use calculators and computers, but you must show your working for any calculations you do.
- You may use the Internet as a resource, but not to ask for the solution to an exam question or to copy any solution you find.
- You must not seek or obtain help from anyone else.

When you have finished:

- scan your work, convert it to a **single PDF file**, and submit this file using the tool below the link to the exam;
- e-mail a copy to **maths@qmul.ac.uk** with your student number and the module code in the subject line;

Examiners: C. Beck, D. Kalogiros

 Continue to next page

Question 1 [27 marks]. Suppose that X and Y have a joint probability density function given by

$$f_{X,Y}(x,y) = \begin{cases} 7e^{-x-7y} & \text{if } x, y \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

(a) Verify that $f_{X,Y}$ is indeed a probability density function. [**5**]

(b) Find the marginal probability density function f_X and state the name of the distribution of X. [**6**]

(c) Find the conditional probability density function $f_{Y|X=x}$. [**5**]

(d) Are the random variables X and Y independent? Justify your answer. [**3**]

(e) Let another probabilty density $\widetilde{f}_{X,Y}$ be given by

$$\widetilde{f}_{X,Y}(x,y) = \begin{cases} ce^{-(\sqrt{x}+\sqrt{y})^2+2\sqrt{xy}} & \text{if } x \geq y \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Determine the normalization constant c in this case. Are X and Y statistically independent? Justify your answer. [**8**]

Question 2 [11 marks]. Consider a standard Normal random variable $Z \sim N(0,1)$. Use the method of cumulative distribution functions, or any other method, to find the probability density function $f_U(u)$ of

$$U = Z^5 \,. \qquad [\mathbf{11}]$$

Question 3 [12 marks]. Suppose that $Z_1, Z_2, \ldots, Z_n$ are statistically independent random variables. Define Y as the sum of squares of these random variables:

$$Y = \sum_{i=1}^{n} Z_i^2 \qquad (n \geq 2)$$

(a) Express the moment generating function $M_Y(t)$ of the random variable Y in terms of moment generating functions involving the random variables Z_i^2, $i = 1, \ldots, n$. [**4**]

(b) Determine $M_Y(t)$ for the special case that $Z_i \sim N(0,1)$. [**4**]

(c) For the above special case, calculate $E[Y]$ by using the moment generating function. [**4**]

 Continue to next page

Question 4 [25 marks]. A family software company operating in 4 cities in the UK would like to automate customer service and reduce repeat calls in their call centres. Hence, they would like to fully implement AI-powered virtual agents in order to reduce costs as well as enhance the customer service performance. In order to fully understand consumer acceptance of this new technology, they collected data from a pilot survey during the last 90 days on the number of complaints received after contacting customer service and being served either by a virtual agent or not (e.g. served by a human agent or directed to the FAQs (Frequently Asked Questions) online page, among others). The relevant data are presented in the following tables below:

Number of complaints received when being served by a Virtual Agent	Observed Frequency
0	35
1	12
2	13
3	11
4	5
5	4
6	3
7	2
8	2
9	2
10	1
11 or more than 11	0

Number of complaints received when <u>not</u> being served by a Virtual Agent	Observed Frequency
0	19
1	14
2	15
3	9
4	14
5	6
6	4
7	2
8	2
9	3
10	2
11 or more than 11	0

(a) Estimate the average number of complaints received per day in the two different types of customer service where a customer is served by a Virtual Agent or not served by a Virtual Agent. [**6**]

(b) Test the hypothesis that the average number of complaints received per day when served by a Virtual agent is the same as the number of complaints received per day after not being served by a Virtual Agent at the 5% significance level. [**15**]

(c) Find an approximation to the 95% confidence interval for the difference of the average number of complaints received in the two types of customer service (with or without Virtual Agent). [**4**]

 Continue to next page

NB. In your answers to all the questions above, report the numerical computations to three decimal places.

Question 5 [25 marks]. Lecturers at a university intend to see the effect of stress relief strategies such as deep breathing and relaxation techniques to the student performance in the final exam. For this reason, they assigned students of the same year in two different cohorts. More precisely, students in Cohort 1 did not do any relaxation techniques before the exam, while students in Cohort 2 performed a full range of the relaxation techniques approved by the British Psychological Society. The average mark of a random sample of 17 students from Cohort 1 was 61, with a standard deviation of 17, while the average mark of a random sample of 11 students from Cohort 2 was 67 with a standard deviation of 19.

(a) Test whether the variances of marks for the two cohorts are the same at the 5% significance level. [**7**]

(b) Test the hypothesis that the average mark in the exam for Cohort 1 students is the same as that for students of Cohort 2 at the 5% significance level. For this test, what is the p-value? [**9**]

(c) What does the p-value that you calculated in (b) indicate about the null hypothesis? [**2**]

(d) Calculate a 95% confidence interval for the difference in average marks in the exam for the two cohorts of students. Explain the meaning of the 95% confidence interval that you have calculated. [**7**]

NB. In your answers to all the questions above, report the numerical computations to three decimal places.

End of Paper.