Queen Mary
University of London

Main Examination period 2020 – January – Semester A

MTH5129: Probability & Statistics II

Duration: 2 hours

Apart from this page, you are not permitted to read the contents of this question paper until instructed to do so by an invigilator.

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

The New Cambridge Statistical Tables are provided.

Only non-programmable calculators that have been approved from the college list of non-programmable calculators are permitted in this examination. Please state on your answer book the name and type of machine used.

Complete all rough work in the answer book and cross through any work that is not to be assessed.

Possession of unauthorised material at any time when under examination conditions is an assessment offence and can lead to expulsion from QMUL. Check now to ensure you do not have any unauthorised notes, mobile phones, smartwatches or unauthorised electronic devices on your person. If you do, raise your hand and give them to an invigilator immediately.

It is also an offence to have any writing of any kind on your person, including on your body. If you are found to have hidden unauthorised material elsewhere, including toilets and cloakrooms, it will be treated as being found in your possession. Unauthorised material found on your mobile phone or other electronic device will be considered the same as being in possession of paper notes. A mobile phone that causes a disruption in the exam is also an assessment offence.

Exam papers must not be removed from the examination room.

Examiners: L. Pettit, N. Rodosthenous

In this exam, $P(\cdot)$ denotes a probability measure defined on a space $(\Omega, \mathcal{F})$ and $E(\cdot)$ denotes the expectation with respect to P.
Please look at the Appendix for probability density functions, quantiles and probabilities of random variables.

Question 1 [24 marks]. Suppose that X and Y have joint density function $f_{X,Y}$ given by

$$f_{X,Y}(x,y) = \begin{cases} ce^{-2(x+y)} & \text{if } x > 0 \text{ and } y > 0 \\ 0 & \text{otherwise} \end{cases}$$

(a) Show that $f_{X,Y}$ is a valid probability density function if and only if $c = 4$. **[6]**

(b) Find the probability $P(X < 2, Y < 3)$. **[4]**

(c) Find the marginal probability density function f_X of X. **[6]**

(d) Calculate the conditional density function $f_{Y|X=x}(y)$. **[5]**

(e) Are X and Y independent? Justify your answer. **[3]**

Question 2 [6 marks]. Let X be a $Bin(4, \frac{1}{2})$ random variable. Find the probability generating function of X. **[6]**

Question 3 [6 marks]. Suppose $X_1, X_2, \ldots, X_n$ are **independent** random variables with moment generating functions $M_{X_i}(t),\ i = 1, \ldots, n$. Find the moment generating function of

$$Y = \sum_{i=1}^{n} X_i$$

and justify all steps in your proof. **[6]**

Question 4 [9 marks]. Suppose that X_1 and X_2 have joint probability density function

$$f(x_1, x_2) = \begin{cases} 8x_1x_2, & 0 < x_1 < x_2 < 2 \\ 0, & \text{otherwise.} \end{cases}$$

What is the joint probability density function of $Y_1 = X_1/X_2$ and $Y_2 = X_2$? **[9]**

Question 5 [24 marks].

(a) A teacher wants to see the effect of changing how reading is taught to primary school children. The children in Year 4 take a reading test at the end of the year. In previous years the score has had a mean of 60 and a variance of 400 and was normally distributed. After the change in teaching method 30 children took the test and the sample mean was $\bar{x} = 66$ with a sample variance $s^2 = 225$. The teacher wants to know if changing the teaching method has affected the score on the test.

Using a 5% significance level:

(i) test the null hypothesis that the variance is $\sigma^2 = 400$ against a two-sided alternative; **[7]**

(ii) test the hypothesis that the mean is $\mu = 60$ against a two-sided alternative. **[7]**

(b) The following year two teachers want to compare two different methods of teaching Mathematics in Year 6. Advise them how to carry out this comparison. Include details on:

- how they could allocate children to the two teaching methods,
- what tests they should use to test their hypotheses,
- what assumptions these are based on,
- how the assumptions can be checked.

[10]

Question 6 [14 marks]. A company's batteries have a mean lifetime of 10 hours. To examine the hypothesis that the distribution of lifetimes has an exponential distribution with mean 10 hours, the lifetimes of one hundred batteries were recorded and are shown below.

Lifetime	0-4	4-8	8-12	12-16	16-20	20+
Number	28	23	16	13	10	10

(a) Assuming the lifetimes have an exponential distribution with mean 10, find the probabilities that the lifetimes lie in the six classes 0-4, 4-8, ..., 20+. **[4]**

(b) Hence find the expected number in each class. **[1]**

(c) Give the formula for a test statistic to test the hypothesis and state its distribution if the hypothesis is true. **[3]**

(d) Calculate the observed value of the test statistic. **[2]**

(e) Give the rejection region for a test with 5% significance level and make a suitable conclusion. **[2]**

(f) Give the R command needed to find the P value of the test. **[2]**

Question 7 [17 marks].

(a) Suppose $\{Z_1, Z_2, Z_3, \ldots\}$ are independent random variables, each having a standard normal distribution, that is with mean 0 and variance 1. Using the random variables $Z_1, Z_2, \ldots$ write down functions of these random variables having

(i) a chi-squared distribution with m degrees of freedom, **[3]**

(ii) a t distribution with n degrees of freedom, **[3]**

(iii) an F distribution with m and n degrees of freedom. **[4]**

(b) Name a method to simulate values from a standard normal distribution. **[2]**

(c) Suppose you wanted to estimate the Expectation of the largest value in a sample of size N from a standard normal distribution. Suggest how you could use simulated values to achieve this. **[5]**

End of Paper – An appendix of 1 page follows.

Appendix: Probability density functions

A random variable $X \sim Bin(n,p)$ follows the binomial distribution with parameters $n \in \mathbb{N}$ and $p \in [0,1]$, if the probability of getting exactly k successes in n trials is given by the probability mass function:

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$$

for $k = 0, 1, 2, ..., n$.

A random variable $X \sim Exp(\lambda)$ has exponential distribution with parameter $\lambda > 0$ if its probability density function is given by:

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geqslant 0 \\ 0 & \text{if } x < 0 \end{cases}$$

Appendix: R studio

You are given the following from R studio

```
> qchisq(0.025,29)
[1] 16.04707
> qchisq(0.975,29)
[1] 45.72229
> qchisq(0.025,30)
[1] 16.79077
> qchisq(0.975,30)
[1] 46.97924
> qchisq(0.95,29)
[1] 42.55697
> qchisq(0.95,30)
[1] 43.77297
> qt(0.975,30)
[1] 2.042272
> qt(0.975,29)
[1] 2.04523
> qt(0.95,30)
[1] 1.697261
> qt(0.95,29)
[1] 1.699127
> qchisq(0.95,4)
[1] 9.487729
> qchisq(0.95,5)
[1] 11.0705
> qchisq(0.95, 6)
[1] 12.59159
```

End of Appendix.