Queen Mary University of London

Main Examination period 2021 – January – Semester A

MTH794P: Probability & Statistics for Data Analytics

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

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.**

All work should be **handwritten** and should **include your student number**.

The exam is available for a period of **24 hours**. Upon accessing the exam, you will have **4 hours** in which to complete and submit this assessment.

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;
- with your e-mail, include a photograph of the first page of your work together with either yourself or your student ID card.

You are expected to spend about **3 hours** to complete the assessment, plus the time taken to scan and upload your work. Please try to upload your work well before the end of the submission window, in case you experience computer problems. **Only one attempt is allowed – once you have submitted your work, it is final**.

Examiners: N. Rodosthenous, W.W. Yoo

 Continue to next page

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.

Question 1 [10 marks]. A player competes at a game involving the rolling of a fair 8-side die.

(a) Suppose that, if it comes up 1, 2 or 3 the player wins, if it comes up 6, 7 or 8 the player loses, if it comes up 4, 5 the player plays again. What is the probability the player wins? **[5]**

(b) Suppose now that, if it comes up 1, 2 or 3 the player wins £3, if it comes up 7 or 8 the player loses £2, if it comes up 4 or 5 the player must roll again and if it comes up 6 the game finishes and the player receives £0. What is the expected reward of the player in this game? **[5]**

Question 2 [22 marks]. Suppose that X and Y have joint probability density function given by

$$f_{X,Y}(x,y) = \begin{cases} 2x^3 + y & \text{if } 0 \le x \le 1,\ 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

(a) Prove that $f_{X,Y}$ is indeed a probability density function. **[4]**

(b) Find $P(0 < X < 0.5\ ,\ 0.5 < Y < 1)$. Give your answer to 4 decimal places. **[4]**

(c) Find $P(X > Y > 0)$. Give your answer to 4 decimal places. **[4]**

(d) Find the marginal probability density function f_X of X. **[4]**

(e) Calculate the conditional density function $f_{Y|X=x}(y)$ of Y given than $X = x$. **[4]**

(f) Are X and Y independent? *Justify your answer.* **[2]**

Question 3 [15 marks]. Let X and Y have the joint probability density function

$$f_{X,Y}(x,y) = \begin{cases} 6x^2 y, & \text{for } 0 < x < 1, 0 < y < 1 \\ 0, & \text{otherwise} \end{cases}$$

(a) Find the joint probability density function $f_{U,V}$ of $U = X + Y$ and $V = X$. **[7]**

(b) Using part (a) or otherwise, find the probability density function of $X + Y$. **[5]**

(c) Describe in a few steps (without doing the calculations) how one can obtain the moment generating function of $X + Y$ by using only the expression $f_{X,Y}(x,y)$ in the above example – **without** using the probability density function of $X + Y$ from part (b). **[3]**

Question 4 [10 marks]. Let $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_m$ be random samples from normal distributions with a common variance σ^2. Assume that $X_1, \ldots, X_n, Y_1, \ldots, Y_m$ are all mutually independent. It can be shown that

$$\mathrm{Var}(S_X^2) = \frac{2\sigma^4}{n-1} \text{ and } \mathrm{Var}(S_Y^2) = \frac{2\sigma^4}{m-1},$$

where S_X^2 is the sample variance for the X's and S_Y^2 is the sample variance for the Y's. Let us denote $T = aS_X^2 + bS_Y^2$ for some constants $a, b \neq 0$.

(a) What condition a and b need to satisfy to guarantee that T is an unbiased estimator for σ^2. **[3]**

(b) Find a that minimises $\mathrm{Var}(aS_X^2 + (1-a)S_Y^2)$. **[7]**

Question 5 [20 marks].

(a) A bottling machine fills wine bottles with amounts that follow a normal distribution $\mathrm{N}(\mu, \sigma^2)$, with $\sigma = 5$ (grams). In a sample of 16 bottles, $\bar{x} = 743$ (grams) was found. Construct a 95% confidence interval for μ. **[3]**

(b) What sample size is needed if we want a 99% confidence interval for μ with length that is at most 1? **[6]**

(c) Now suppose the content of the bottles has to be determined by weighing. It is known that empty wine bottles involved has a weight that follows a normal distribution with mean 250 grams and standard deviation of 15 grams. For a sample of 16 full bottles, an average weight of 998 grams was found. You may assume that the amount of wine is independent of the bottle weight.

(i) Find the distribution of the weight of the i-th full wine bottle, where $i = 1, \ldots, 16$. **[6]**

(ii) Construct a 95% confidence interval for the expected amount of wine per bottle μ. **[5]**

Question 6 [23 marks]. Samples of wood were obtained from the core and periphery of a certain English castle. The date of the wood was determined, giving the following data.

Core		Periphery	
1294	1251	1284	1274
1279	1248	1272	1264
1274	1240	1256	1256
1264	1232	1254	1250
1263	1220	1242	
1254	1218		
1251	1210		

In the following hypotheses testing problems, use **both** critical values and p-values to reach your conclusions. Use significance level 0.01.

(a) Assuming equal variance, test to determine if the mean age of the core is the same as the mean age of the periphery. [**8**]

(b) Assuming unequal variances, test if the mean age of the core is the same as the mean age of the periphery. [**8**]

(c) Test whether the equal variance assumption is reasonable or not. [**7**]

End of Paper.