B. Sc. Examination by course unit 2014

MTH5122 Statistical Methods

Duration: 2 hours

Date and time: 8th May 2014, 10.00-12.00

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 awarded are shown next to the questions.

Calculators ARE permitted in this examination. The unauthorized use of material stored in pre-programmable memory constitutes an examination offence. Please state on your answer book the name and type of machine used.

Statistical functions provided by the calculator may be used provided that you state clearly where you have used them.

The New Cambridge Statistical Tables are provided.

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

Important note: the Academic Regulations state that possession of unauthorized material at any time by a student who is under examination conditions is an assessment offence and can lead to expulsion from QMUL.

Please check now to ensure you do not have any notes, mobile phones or unauthorised electronic devices on your person. If you have any, then please raise your hand and give them to an invigilator immediately. Please be aware that 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. Disruption caused by mobile phones is also an examination offence.

Exam papers must not be removed from the examination room.

Examiner(s): L I Pettit, H Grossmann

Question 1 [22 marks] The random variable X has probability density function $f(x) = C(2x^2 + 1)$ for $0 < x < 1$ and zero otherwise. The random variable Y conditional on $X = x$ has probability density function

$$f(y|x) = D\frac{2x^2 + y}{2x^2 + 1} \quad 0 < y < 2$$

and zero otherwise.

(a) Find the values of C and D. [6]

(b) Find the conditional mean of Y given $X = x$. [3]

(c) Hence find the unconditional mean of Y using the result that $\mathrm{E}[Y] = \mathrm{E}[\mathrm{E}[Y|X]]$. [4]

(d) Find the joint probability density function of X and Y. [2]

(e) Find the marginal distribution of Y and hence confirm the value of the mean found in (c). [5]

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

Question 2 [17 marks]

(a) A biased coin has probability p of landing heads. Assuming tosses are independent show that the number of heads, X, before the first tail has probability mass function $P[X = x] = (1 - p)p^x$ for $x = 0, 1, 2, \ldots$ [3]

(b) Show the moment generating function of X is

$$M_X(t) = \frac{1 - p}{1 - pe^t}.$$

[4]

(c) Hence find the mean and variance of X. [6]

(d) The coin tossing is repeated independently until n tails have been seen. Find the moment generating function of the total number of heads. [4]

Question 3 [12 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

Carry out a goodness of fit test of the hypothesis using a significance level of $\alpha = 0.05$. [12]

Question 4 [11 marks] In a study to examine different attitudes to healthy eating, random samples of 647 men and 434 women were selected. Of those sampled 236 men and 195 women said they regularly order a vegetarian meal in a restaurant.

(a) Test the hypothesis that the proportions of men and women who order vegetarian meals regularly are the same against an alternative that women are more likely to order vegetarian meals, use a significance level $\alpha = 0.05$. [8]

(b) Why would using a contingency table approach to this test cause a problem? [3]

Question 5 [8 marks] The random variables $Z_1 \sim N(0,1)$, $Z_2 \sim N(0,1)$, $V_1 \sim \chi^2_n$, $V_2 \sim \chi^2_m$ are mutually independent. Write down the distributions of the following

(a) $Z_1^2 + Z_2^2$, [2]

(b) $\frac{Z_1}{Z_2}$, [2]

(c) $\frac{Z_1}{\sqrt{V_1/n}}$, [2]

(d) $\frac{V_1/n}{V_2/m}$. [2]

Question 6 [18 marks] Eight patients who suffered from severe insomnia took part in a study to determine the effects of two sedatives. Each patient took sedative A for a two week period and the average number of hours sleep were recorded for each patient. This procedure was then repeated for sedative B. The results were as follows.

Patient	1	2	3	4	5	6	7	8
Sedative A	2.1	2.9	5.4	3.8	3.1	4.1	2.4	2.9
Sedative B	1.6	2.0	5.2	4.0	3.3	3.2	1.8	2.3

(a) Test the hypothesis that the effect of the two sedatives are the same by finding the P value. [10]

(b) Find a 95% confidence interval for the difference in average amount of sleep under the two sedatives. [4]

(c) Comment on the design of this trial noting any good features and any improvements which could be made. [4]

Question 7 [12 marks]

(a) Let X be a continuous random variable, g a non-negative function with domain the real line and k a positive real number. Prove Markov's inequality

$$P(g(X) \geq k) \leq \frac{\mathrm{E}[g(X)]}{k}.$$

[6]

(b) Suppose Y is a continuous random variable taking only non-negative values. Let the mean of Y be μ and the median of Y be m. Use Markov's inequality to prove that $m \leq 2\mu$. [6]

End of Paper