The University Of Sheffield.

SCHOOL OF MATHEMATICS AND STATISTICS **September 2017**

MAS6002 Assignment 1

To be submitted before 12:00 noon on Tuesday 10th October via MOLE

Answer all questions, giving enough explanation to justify your answers. Where needed for the explanation, you may refer to notes and handouts supplied, rather than write material out. You must prepare your answers yourself, without help. Conferring with others in preparing your answers is forbidden — doing so would be cheating, and the University takes such offences very seriously. Most of the marks will be for correct answers accompanied by adequate explanations. You may use R to obtain probabilities from standard distributions

1 A test for a disease will certainly be positive if the patient has the disease, but will be positive with probability 0.01 if the patient does not have the disease. Based on a patient's symptoms, but without using the test, a doctor believes that the probability the patient has the disease is 0.1.

(a) Based on the doctor's beliefs, find the probability that the patient has the disease if the test is positive. You should define your notation carefully and explain your method. ***(3 marks)***

(b) What is the probability that the patient has the disease if the test is not positive? ***(1 mark)***

2 A continuous random variable X has probability density function given by

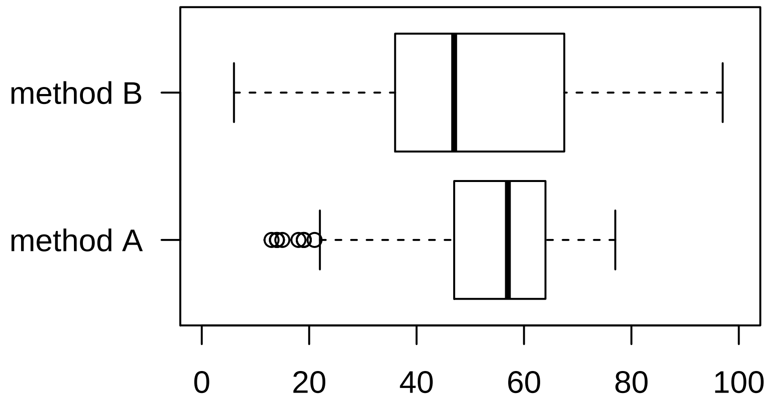
$$f_X(x) = \begin{cases} \frac{3}{4}(1-x^2) & -1 \leq x \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

(a) Find $P\left(X > \frac{1}{2}\right)$. ***(1 mark)***

(b) Find the cumulative distribution function, $F_X(x)$, for $-1 \leq x \leq 1$. ***(2 marks)***

(c) Find $E(X)$ and $\text{Var}(X)$. ***(3 marks)***

3 Two methods of teaching arithmetic to primary school children (aged 7 years) have been compared in an experiment. Children in two schools are taught with method A, and children in another two schools are taught with method B. A total of 200 children were taught using each method. After one month of teaching, the children took a test. Box plots of the children's test scores are plotted below.



(a) Describe the shape of the distribution of test scores for method A.

(1 mark)

(b) What do the box plots suggest about the effectiveness of each method?

(2 marks)

(c) Let a_i be the observed score of child i taught with method A, and let b_i be the observed score of child i taught with method B. Sample means and variances for the two sets of marks are as follows.

$$\bar{a} = \frac{1}{200}\sum_{i=1}^{200} a_i = 53.42, \qquad \bar{b} = \frac{1}{200}\sum_{i=1}^{200} b_i = 50.18,$$

$$s_a^2 = \frac{1}{199}\sum_{i=1}^{200}(a_i - \bar{a})^2 = 214.506,$$

$$s_b^2 = \frac{1}{199}\sum_{i=1}^{200}(b_i - \bar{b})^2 = 472.751.$$

Conduct a hypothesis test that the population mean test scores for each method are equal. Specify your null and alternative hypotheses, defining your notation clearly. Give a range for the p-value, using the following R output, and state clearly the interpretation of your result.

```
qnorm(c(0.9, 0.95, 0.975), mean = 0, sd = 1)
## [1] 1.28 1.64 1.96
```

(3 marks)

(d) Give one criticism of the design of the experiment to compare the two methods. How might you have done things differently? *(2 marks)*

4 A water company has inspected sections of water pipes for leaks. In each of 50 separate 1km long sections the number of leaks is recorded and tabulated below. A total number of 103 leaks were counted. Test the hypothesis at the 5% level of significance that the number of leaks in a 1km section of pipe follows a Poisson distribution.

number of leaks	0	1	2	3	4	5	6
count	5	13	14	12	5	0	1

You may use the following R output to help you (not all of which is relevant.)

```
dpois(0:3, 103/50)
## [1] 0.127 0.263 0.270 0.186
qchisq(0.95, df = c(3, 4, 5, 6))
## [1]  7.81  9.49 11.07 12.59
```

(5 marks)

5 Let $\mathbf{X} = \begin{pmatrix} X \\ Y \end{pmatrix}$ be a random vector with a bivariate normal distribution, with mean vector $\boldsymbol{\mu} = \begin{pmatrix} 0 \\ 2 \end{pmatrix}$ and covariance matrix $\boldsymbol{\Sigma} = \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}$.

(a) Write down the marginal distribution of X. ***(2 marks)***

(b) Let $U = X - Y$ and $V = X + 2Y + 3$. Find the mean vector and covariance matrix of the random vector $\mathbf{U} = \begin{pmatrix} U \\ V \end{pmatrix}$. ***(5 marks)***

6 Let (X, Y) be a bivariate random variable with probability density function

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

where $k \in \mathbb{R}$ is a deterministic constant.

(a) Show that $k = \frac{6}{5}$. ***(2 marks)***

(b) Calculate $\mathbb{P}[X \leq Y]$. ***(3 marks)***

(c) Find the marginal probability density function $f_Y(y)$ of Y. ***(3 marks)***

7 Recall that the Beta function $B : (0, \infty) \times (0, \infty) \to \mathbb{R}$ is given by

$$B(\alpha, \beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\,dx.$$

(a) Show that $B(\alpha, 1) = \dfrac{1}{\alpha}$. ***(1 mark)***

(b) Let X be a random variable with the Beta distribution, $X \sim Be(8, 1)$, and let $Y = \sqrt[3]{X}$. Find the probability density function of Y and identify its distribution. ***(8 marks)***

8 Let X and Y be a pair of independent and identically distributed random variables. Let $U = X + Y$ and $V = X - Y$.

(a) Show that $\mathrm{Cov}(U, V) = 0$. ***(3 marks)***

(b) Suppose, additionally, that $\lambda \in (0, \infty)$ and that X and Y both have the $Exp(\lambda)$ distribution.

Find the joint probability density function $f_{U,V}(u, v)$ of U and V, stating clearly the range of (u, v) for which it is non-zero. ***(8 marks)***

(c) With X, Y as in (b), are U and V independent? Justify your answer. ***(2 marks)***

9 Let $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ be a set of n independent identically distributed samples from the $Exp(\lambda)$ distribution, where $\lambda \in (0, \infty)$ is an unknown parameter and $n \geq 3$.

(a) Find the likelihood function $L(\lambda; \mathbf{x})$ and the log-likelihood function $\ell(\lambda; \mathbf{x})$. ***(4 marks)***

(b) Derive a formula for the maximum likelihood estimator $\hat{\lambda}$ of λ. ***(5 marks)***

(c) Sketch the log-likelihood function $l(\lambda; \mathbf{x})$, marking the location of $\hat{\lambda}$ clearly on your diagram. ***(2 marks)***

(d) Consider the set

$$R_2 = \left\{ \lambda \in (0, \infty) \ : \ |\ell(\lambda, \mathbf{x}) - \ell(\hat{\lambda}, \mathbf{x})| \leq 2 \right\}$$

Suggest why we might hope that values $\lambda \in R_2$ are good approximations to the true value of λ. ***(2 marks)***

If you wish, you may annotate your diagram from part (c) to help you answer part (d).

End of Question Paper