The University Of Sheffield.

SCHOOL OF MATHEMATICS AND STATISTICS **Spring Semester 2021–2022**

MAS113 Introduction to Probability and Statistics **2 hours 30 minutes**

This is an open book exam.

Answer ALL questions. Total marks: 30.

You can work on the exam during the 24 hour period starting from 10am (BST), and you must submit your work within two hours and thirty minutes of accessing the exam paper or by the end of the 24 hour period (whichever is earlier).

Late submission will not be considered without extenuating circumstances. *Calculations should be performed by hand. A university-approved calculator may be used. The use of any other calculational device, software or service is not permitted. To gain full marks, you will need to show your working.*

By uploading your solutions you declare that your submission consists entirely of your own work, that any use of sources or tools other than material provided for this module is cited and acknowledged, and that no unfair means have been used.

1 Let S be the set $\{1, 2, 3\}$. A set function P on S has $P(\{1, 2\}) = \dfrac{2}{3}$ and $P(\{2, 3\}) = \dfrac{1}{3}$. If P is a probability measure, give the values of

(i) $P(\{2\})$; ***(1 mark)***

(ii) $P(\{1\})$. ***(1 mark)***

2 An item of jewellery is believed to be possibly fake, with the prior probability of it being fake being 0.1. It is decided to have it inspected by an expert. If the item is fake, the expert will declare it to be fake with probability 0.9, while if it is in fact genuine the expert will declare it to be fake with probability 0.05; otherwise the expert will declare it to be genuine.

The expert declares the item to be fake. Find the posterior probability that it is fake. You should define your notation carefully and explain your method.

(4 marks)

3 Let X be a discrete random variable taking values in $R_X = \{1, 2, 3, 4\}$ with probability mass function given by $p_X(x) = x/10$ for $x \in R_X$.

(i) Find $E(X)$. ***(1 mark)***

(ii) Find $\mathrm{Var}(X)$. ***(1 mark)***

(iii) What is $\mathrm{Var}(4X)$? ***(1 mark)***

4 Xander and Yvonne play a game. Xander first throws a 3-sided die, we call the resulting random variable X (the possible outcomes are 1, 2, or 3, and each outcome is equally likely). Yvonne then throws X fair coins, we call Y the number of heads (which is a number between 0 and 3).

(i) Draw a table representing the joint distribution of X and Y. ***(2 marks)***

(ii) Find $P(X = 3 \mid Y = 1)$. ***(1 mark)***

5 Let X be a random variable which has a uniform distribution on $[0, 1]$, and let Y be an independent Bernoulli random variable with parameter $\frac{1}{2}$. Define $Z = X + Y$. Find the cumulative distribution function of Z: $F_Z(x) = P(Z \leq x)$ for $x \in \mathbb{R}$. (Hint: you might want to reason separately on the following cases: $x < 0$, $0 \leq x < 1$, $1 \leq x < 2$, $x \geq 2$.) ***(2 marks)***

6 Let X be a continuous random variable with probability density function defined by

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

Find the moment generating function $M_X(t) = E(e^{tX})$ for $t \neq 1$. ***(2 marks)***

7 Three independent random variables X, Y and Z have the same expectation:

$$E(X) = E(Y) = E(Z) = \theta,$$

but have different variances:

$$\mathrm{Var}(X) = \sigma^2, \ \mathrm{Var}(Y) = 2\sigma^2, \ \mathrm{Var}(Z) = 3\sigma^2.$$

Two different estimators are proposed for θ:

$$T_1 = \frac{X + Y + Z}{3},$$
$$T_2 = \frac{3X + 2Y + Z}{6}.$$

(i) Show that both T_1 and T_2 are unbiased estimators of θ. ***(2 marks)***

(ii) By considering the standard error of each estimator, explain which of T_1 and T_2 is a better choice of estimator for θ. ***(2 marks)***

8 A manufacturing process makes batteries for the same make and model of laptop. A sample of 20 batteries is obtained from this process. Each battery is tested under identical conditions to see how long it lasts following a full charge. In R, the 20 observed times (in hours) are stored in the vector `times`. Some R output is given below.

```
> mean(times)
[1] 9.7
> var(times)
[1] 0.286
> qt(0.975, df = c(19, 20))
[1] 2.093 2.086
```

(where 9.7 hours equates to 9 hours and 42 minutes).

(i) Assume battery lives (following a full charge) are normally distributed. Calculate a 95% confidence interval for the population mean battery life for all batteries produced by the process. To get the marks, you must make clear what formula you have used to obtain your interval. ***(2 marks)***

(ii) The target mean battery life for the manufacturing process is 10 hours. Briefly explain what your confidence interval from part (a) tells you, regarding whether this target has been achieved. ***(1 mark)***

9 A health food company sells a vitamin supplement (a pill), which they claim reduces the chances of catching a cold. An experiment is conducted to test this. Over a two month period, 50 volunteers take the vitamin supplement each day, and another 50 volunteers take a placebo pill (a dummy pill with no known effects). Let θ and ϕ be defined as follows.

- For any individual volunteer given the vitamin supplement, θ is the probability that they will catch a cold during the experiment.
- For any individual volunteer given the placebo pill, ϕ is the probability that they will catch a cold during the experiment.

Let X and Y be defined as

- X: the number of volunteers, out of the 50 taking the vitamin supplement, who will catch a cold during the experiment;
- Y: the number of volunteers, out of the 50 taking the placebo pill, who will catch a cold during the experiment.

(i) By choosing appropriate probability distributions, write down a statistical model that relates X and Y to θ and ϕ. ***(1 mark)***

(ii) At the end of the experiment, the observed values of X and Y are 28 and 32 respectively. Test the hypothesis that the vitamin supplement has no effect on the chances of catching a cold, using the Neyman-Pearson approach with size 0.05. State your null hypothesis, your alternative hypothesis, and explain clearly the conclusion of your test regarding the company's claim. Some of the following R output will help you.

```
> qnorm(c(0.95, 0.975), mean = 0, sd = 1)
[1] 1.645 1.960
```

(4 marks)

(iii) Draw a sketch to indicate the p-value, and state the R command you would use to find the p-value. ***(2 marks)***

End of Question Paper