The University Of Sheffield.

SCHOOL OF MATHEMATICS AND STATISTICS **Spring Semester 2015–2016**

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

*Attempt **ALL** questions. The allocation of marks is shown in brackets. Total marks 60. Give all numerical solutions correct to two decimal places of accuracy.*

1 Let S be the set $\{1, 2, 3, 4, 5\}$.

(i) Define a set function M by saying that for A a subset of S, $M(A)$ is the product of all the elements of A. (For the empty set, define $M(\varnothing) = 1$.) Give a counterexample to show that M is not a measure. ***(2 marks)***

(ii) Define a set function P by saying that for A a subset of S, $P(A)$ is the number of elements in A divided by 5.

(a) Explain why P is a probability measure. ***(3 marks)***

(b) Under the probability measure P, are the events $A_1 = \{2, 3, 5\}$ and $A_2 = \{2, 4\}$ independent? Give a reason for your answer. ***(2 marks)***

2 In a United States election, a political analyst believes that if the Democrats win Florida their probability of winning the election overall is 0.9 and that if the Democrats lose Florida their probability of winning the election overall is 0.4. The analyst believes that the probability of the Democrats winning Florida is 0.3. Based on the analyst's beliefs, and introducing suitable notation to explain your derivations:

(i) What is the probability of the Democrats winning the election overall? ***(2 marks)***

(ii) What is the probability that the Democrats have won Florida if it is known that they have won the election overall? ***(2 marks)***

3 Six patients are to be tested for whether or not they have a disease. For each individual patient, the doctor believes that the probability of a positive test is 0.3.

(i) Explain why the number of positive tests might be assumed to have a binomial distribution, and state the parameters you would use. ***(3 marks)***

(ii) Under the assumption of a binomial distribution,

(a) what would the probability that there is exactly one positive test be? ***(1 mark)***

(b) Write down the expectation of the number of positive tests. ***(1 mark)***

(c) Write down the variance of the number of positive tests. ***(1 mark)***

4 Andy and Novak are playing a tennis match. I think Novak is definitely going to win, but my friend is not so sure. I offer to give my friend £10 if Andy wins, as long as my friend gives me £20 if Novak wins. The actual probability that Novak wins is 0.8. What is the expectation of my winnings? What is the standard deviation? ***(5 marks)***

5 A random variable Y has probability density function f_Y given by

$$f_Y(y) = \begin{cases} k(y^2+y+1) & -1 \leq y \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

(i) Find the value of k. ***(2 marks)***

(ii) Find the cumulative distribution function of Y, tabulated for the three cases $y < -1$, $-1 \leq y \leq 1$ and $y > 1$. ***(3 marks)***

(iii) Find $P(0 < Y \leq 1/2)$ and $P(Y > 0|Y \leq 1/2)$. ***(2 marks)***

(iv) Find $E(Y)$. ***(1 mark)***

6 (i) Three random variables X, Y and Z are related as follows:

$$Z = 2X - 3Y.$$

If X and Y are independent, with $X \sim N(1, 2)$ and $Z \sim N(0, 17)$, find the mean and variance of Y. ***(3 marks)***

(ii) State Chebychev's inequality for a random variable X which has mean μ and variance σ^2, and use it to show that

$$P(X \geq 4) \leq \frac{1}{4},$$

where X is a random variable with mean 2 and variance 1. ***(4 marks)***

7 A zoologist seeks to gain information about the weights of a population of sparrows, which you may assume to be normally distributed. She is able to catch and weigh 120 birds, and the measurements $x_1, x_2, \ldots, x_{120}$ of these (in grammes) may be summarised by $\sum_{i=1}^{120} x_i = 2007$ and $\sum_{i=1}^{120} x_i^2 = 39632$.

(i) Calculate both the sample mean and sample variance. ***(2 marks)***

(ii) Find a 95% confidence interval for the mean weight of sparrows. ***(2 marks)***

(iii) Find a 95% confidence interval for the variance of the weights of sparrows. ***(2 marks)***

You may use the following R output:

```
> qnorm(0.975,0,1)
[1] 1.96.
> qchisq(0.975,119)
[1] 151.08
> qchisq(0.025,119)
[1] 90.67
```

8 (i) Explain what are meant by *type I* and *type II* errors in the context of hypothesis testing. ***(2 marks)***

(ii) Suppose that an experiment is carried out to test whether or not a new chemical additive, which is designed to extend the shelf–life of a breakfast cereal, is in fact a cause of bowel cancer. What are the practical implications of type I and type II errors in this context? Here you should take the null hypothesis to be that the additive does not cause bowel cancer. ***(2 marks)***

9 A group of 332 teenagers were asked to complete a survey to determine their usage of social media. The survey scored each teenager between -15 and 15, with 15 signifying exceptionally heavy usage and -15 none. The survey found that 150 females had an average score of 2.8, with a standard deviation of 3.6, while 182 males averaged 1.6 with a standard deviation of 2.9. Assuming that the scores in the underlying populations are normally distributed, write down suitable null and alternative hypotheses, and carry out an appropriate test to investigate if gender has any effect on survey score. Your solution should include a calculation of the p-value, and a brief justification of the number selected to be the degrees of freedom. ***(7 marks)***

You are given the following R output:

```
> pt(3.295301,149)
[1] 0.9993856
```

10 In an experiment on chlorophyll inheritance in corn, for 1103 seedlings of self–fertilized heterozygous green plants, 854 seedlings were green and 249 were yellow. The theory predicts that 75% of the seedlings should be green. Are the experimental results far enough away from the predicted values to give evidence against the theory? ***(6 marks)***

You may use the R output:

```
> pchisq(3.459958,1)
[1] 0.937128
```

End of Question Paper