The University Of Sheffield.

SCHOOL OF MATHEMATICS AND STATISTICS **Spring Semester 2011–2012**

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

*Attempt **ALL** questions. The allocation of marks is shown in brackets. Total marks 90.*

1 Let S be the set {red, green, blue}, and consider the set of subsets of S

$$A = \{\phi, \{\text{red}\}, \{\text{green}\}, \{\text{blue}\}, S\}$$

(i) Explain why A is not a Boolean algebra. ***(2 marks)***

(ii) Let $\mathcal{P}(S)$ be the power set of S, and define the set function

$$m(E) = \begin{cases} 0.5 & \text{if } \{\text{red}\} \subseteq E, \text{ or } \{\text{green}\} \subseteq E, \text{ or } \{\text{red, green}\} \subseteq E, \\ 0 & \text{otherwise.} \end{cases}$$

Demonstrate that m is not a measure on $\mathcal{P}(S)$, explaining what requirement the function m fails to satisfy. ***(2 marks)***

2 In the semi-finals of a football tournament, the four remaining teams are Spain, Brazil, Argentina and Ghana, with Spain playing Brazil. Suppose that this combination of four teams in the semi-finals has not been seen before. You are offered the following bet. If Spain win the tournament, you gain £4, but if Spain fail to win, you must pay £1. Let $P(S)$ be the probability that Spain wins.

(i) Express, in terms of $P(S)$, your expected winnings if you take the bet. ***(1 mark)***

(ii) You decide you will take the bet if your expected winning are positive. Explain briefly (in one or two sentences in each case) whether it would be appropriate to assign a value to $P(S)$ using

(a) classical probability based on symmetry;

(b) frequency probability;

(c) subjective probability. ***(3 marks)***

3 Two random variables W and V have a joint probability mass function, as tabulated below.

$p_{W,V}(w,v)$		$w=0$	$w=1$	$w=2$
v	0	0.2	0.3	0.1
	1	0.15	0.15	0.1

(i) Find the marginal probability mass functions of W and V. ***(5 marks)***

(ii) If it is known that $W = 0$, what is the probability that $V = 1$? ***(2 marks)***

(iii) Calculate the means and the standard deviations of W and V. ***(6 marks)***

(iv) Calculate the covariance between W and V. ***(2 marks)***

(v) Are W and V independent? Briefly justify your answer. ***(1 mark)***

4 In a multiple choice test, there are 10 questions, with four answers per question. For each question, only one out of the four answers is correct. If you were to pick one answer at random for each question, calculate the probability of getting exactly 6 out of 10 answers correct. ***(3 marks)***

5 A continuous random variable X has probability density function $f_X(x)$ given by

$$f_X(x) = \begin{cases} k(x+x^2) & 0 \le x \le 2, \\ 0 & \text{otherwise.} \end{cases}$$

(i) Leaving your answers in terms of k, find

(a) $E(X)$; ***(3 marks)***

(b) $Var(X)$ ***(3 marks)***

(c) The cumulative distribution function $F_X(x)$, tabulated for the three cases $x < 0$, $0 \le x \le 2$, and $x > 2$; ***(4 marks)***

(d) $P(X = 0.5)$; ***(1 mark)***

(e) $P(X > 1)$; ***(2 marks)***

(f) $P(X > 1.5 | X > 1)$. ***(3 marks)***

(ii) Deduce the value of k. ***(2 marks)***

6 A random variable X follows the normal distribution with mean 5 and variance 4.

(i) Find the mean and variance of the random variable $W = 3X - 2$.
(2 marks)

(ii) Given the R output

```
> pnorm(2)
[1] 0.977
> pnorm(7)
[1] 1
```

calculate the probabilities

(a) $P(X > 1)$; *(3 marks)*

(b) $P(|X| \leq 9)$; *(2 marks)*

(c) $P(X < 10)$. *(2 marks)*

7 The proportion of answering a question in an opinion poll is p. In order to estimate p, a simple random sample $X_1, \ldots, X_n$ is selected, each of X_i following a Bernoulli distribution, i.e.

$$X_i = \begin{cases} 1, & \text{with probability } p \\ 0, & \text{with probability } 1-p \end{cases}$$

(i) Find the mean and variance of X_i. *(2 marks)*

(ii) Show that $\hat{p}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ is an unbiased estimator for p. *(3 marks)*

(iii) Show that $\hat{p}_n$ is a consistent estimator. *(4 marks)*

8 A scientist is interested in the mean height of adults in a particular population. A survey will produce a sample of 15 heights $X_1, \ldots, X_{15}$ (in meters).

The actual values obtained can be summarised as $\sum_{i=1}^{15} x_i = 26.7$ and $\sum_{i=1}^{15} x_i^2 = 173.526$.

(i) Given the R output

```
> qt(0.975, 14)
[1] 2.145
```

calculate a 95% confidence interval for the population mean height.
(6 marks)

(ii) Explain what is possibly wrong with this interval and state two possible ways to improve it. *(2 marks)*

9 In a study of a drug development, 6 patients were offered a treatment and results of the effect of the drug (units not given), were recorded before and after treatment, as shown in the following table.

Patient	1	2	3	4	5	6
Before treatment	5	6	6	7	9	4
After treatment	4	3	4	8	7	5

Given the R output

```
> pt(1.46385, 5)
[1] 0.898
```

use an appropriate test, to assess the evidence of difference before and after treatment. ***(8 marks)***

10 A simple genetic theory implies that 25% of a particular population should have the gene AA, 50% should have the gene Aa, and 25% should have the gene aa.

(i) If the theory is true, and the genotypes of 12 randomly selected individuals are observed, what is the probability that

(a) 3 have gene AA, 6 have Aa, and 3 have aa; ***(2 marks)***

(b) 4 have each of the three genes? ***(1 mark)***

(ii) An actual sample of 31 individuals is taken, and their genotypes observed, giving the following results.

Gene	AA	Aa	aa
Number observed	9	16	6

Given the R output

```
> pchisq(0.6129, 2)
[1] 0.2639447
```

carry out a suitable test of the genetic theory, and comment briefly on your results. ***(8 marks)***

End of Question Paper