The University Of Sheffield.

SCHOOL OF MATHEMATICS AND STATISTICS **Spring Semester 2012–2013**

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

*Attempt **ALL** questions. The allocation of marks is shown in brackets. Total marks 60.*

Please leave this exam paper on your desk
Do not remove it from the hall

Registration number from U-Card (9 digits)
to be completed by student

Blank

1 Let S be the set of outcomes that could result from rolling a six-sided die:

$$S = \{1,2,3,4,5,6\},$$

and consider the set W_1 of subsets of S:

$$W_1 = \{\phi, \{1\}, S\}.$$

(i) Explain why W_1 is not a Boolean algebra. ***(1 mark)***

(ii) Suppose one subset E of S is added to W_1, to give a new set

$$W_2 = \{\phi, \{1\}, E, S\}.$$

What would E have to be in order for W_2 to be a Boolean algebra? ***(1 mark)***

(iii) For your choice of E in part (ii), give the value of the counting measure applied to each element of W_2. ***(1 mark)***

2 A series of five cricket matches is to be played between England and Australia. At the start of each match, a coin is tossed, and the Australian captain calls heads or tails. If he calls correctly, he decides whether his team bats or fields first. A bookmaker judges that, for every match, if the Australian captain calls correctly, the probabilities of the different outcomes are as follows

$$P(\text{Australia win}) = 0.3, \qquad P(\text{England win}) = 0.3, \qquad P(\text{match drawn}) = 0.4,$$

but if the Australian captain calls incorrectly, the probabilities of the different outcomes are as follows

$$P(\text{Australia win}) = 0.2, \qquad P(\text{England win}) = 0.4, \qquad P(\text{match drawn}) = 0.4,$$

Important note: to get full marks on this question, where appropriate, you must define suitable notation, and state clearly any probability distributions that you have assumed to derive your answers, including the parameter values.

(i) Out of the five matches, what is the probability that the Australian captain only calls correctly once? ***(2 marks)***

(ii) What is the bookmaker's probability of each outcome (Australia win, England win, match drawn) in the first match? ***(2 marks)***

(iii) What is the bookmaker's probability that, over the five match series, Australia will win one match, England will win two matches, and there will be two draws? ***(3 marks)***

3 A simple financial model is proposed to describe changes in two stock prices from time t to time $t+1$. At time t, the price of one share in the first stock is 180p, and the price of one share in the second stock is 230p. Let X be the price (in pence) of one share in the first stock at time $t+1$, and Y be the price (in pence) of one share in the second stock at time $t+1$. The joint probability mass function of X and Y is tabulated below.

	$p_{X,Y}(x,y)$	x = 170	180	190
y	220	0.02	0.1	0.18
	230	0.1	0.2	0.1
	240	0.13	0.1	0.07

(i) Write out, in a suitable table, the marginal probability mass function of Y ***(1 mark)***

(ii) Find the variance of Y. State the unit of measurement of the variance. ***(3 marks)***

(iii) If it is known that the first stock has increased in price, what is the probability that the second stock has also increased in price? ***(2 marks)***

(iv) Without doing any further calculation, are X and Y independent? Briefly justify your answer. ***(1 mark)***

(v) By inspecting the joint probability mass function, but without doing any calculations, state whether the covariance between X and Y will be positive or negative, briefly explaining your answer. ***(2 marks)***

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

$$f_X(x) = \frac{e^{\left(1-\frac{x}{4}\right)}}{4(e-1)} \text{ for } 0 \leq x \leq 4,$$

and 0 for $x < 0$ and $x > 4$. Find

(i) the cumulative distribution function $F_X(x)$, tabulated for the three cases $x < 0$, $0 \leq x \leq 4$, and $x > 4$; ***(4 marks)***

(ii) $P(X > 2)$; ***(1 mark)***

(iii) $P(X < 3|X > 2)$. ***(2 marks)***

5 A continuous random variable Y has the cumulative distribution function given by

$$F_Y(y) = \begin{cases} 0 & y < 10, \\ 0.1y & 10 \leq y \leq 20, \\ 1 & y > 20. \end{cases}$$

Calculate the expected value of Y^2.

(4 marks)

6 Each of the independent random variables $X_1, \ldots, X_9$ are normally distributed with mean 2 and variance 4, i.e. $X_i \sim N(2, 4)$, for $i = 1, \ldots, 9$.

(i) Show that the estimator $\hat{\mu} = \frac{1}{9}\sum_{i=1}^{9} X_i$ is unbiased for the mean $\mu = 2$.

(1 mark)

(ii) Given the R output

```
> pnorm(1)
[1] 0.8413447
> pnorm(3/2)
[1] 0.9331928
```

calculate the probabilities

(a) $P(-1 < X_i \leq 4)$; ***(1 mark)***

(b) $P(X_i < 100)$; ***(1 mark)***

(c) $P(Y > 12)$, where the random variable Y is defined as

$$Y = \sum_{i=1}^{9} X_i.$$

(2 marks)

7 The simple random sample $X_1, \ldots, X_n$ is used to estimate an unknown mean μ. With $X_i \sim N(\mu, \sigma^2)$, for some known variance σ^2, the estimator

$$\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} X_i + \frac{\alpha}{n}$$

is considered, where α is a constant with $|\alpha| < \infty$.

(i) Find the distribution of $\hat{\mu}_n$. ***(1 mark)***

(ii) Show that $\hat{\mu}_n$ is a consistent estimator. ***(6 marks)***

8 An experiment is conducted on the heat resistance of some chemical material. 9 measurements of the resistance are collected and given below:

$$58, \quad 59, \quad 58, \quad 57, \quad 57, \quad 60, \quad 61, \quad 60, \quad 59$$

Measurements are assumed to be independent and each of them to follow a normal distribution with some unknown mean μ and with standard deviation 2. An initial analysis in R gave the output

```
> qnorm(0.95)
[1] 1.644854
> pnorm(1.644854-3)
[1] 0.08768552
> pnorm(-1.644854-3)
[1] 1.701585e-06
```

(i) Find a 90% confidence interval for the mean resistance μ. ***(2 marks)***

(ii) If we test the null hypothesis $H_0 : \mu = 58$ against the alternative $H_A : \mu \neq 58$ and assuming that H_A is true with $\mu = 60$, find the power of this test. ***(4 marks)***

9 Each of a random sample of 6 college freshmen takes a mathematics aptitude test both before and after undergoing an intensive training course designed to improve such test scores. The test scores of the 6 students before and after the course are tabulated below:

Student	1	2	3	4	5	6
Before	60	73	42	88	66	77
After	70	80	40	94	79	86

An initial analysis in R gave

```
> pt(df=5, 3.431)
[1] 0.9906923
```

Use a suitable test to assess the evidence of difference in the mean score effects before and after the course. Is the course worthwhile? ***(5 marks)***

10 A random sample of 100 hotels, each having 3 rooms is taken and the number of luxurious rooms in each hotel is recorded, as shown in the table below:

Number of luxurious rooms	0	1	2	3	Total
Frequency	19	40	29	12	100

A suggested model is that each room has probability 0.5 to be luxurious and that basic and luxurious rooms within a hotel are independent.

An initial analysis in R gave the following output

```
> pchisq(df=3, 5.493)
[1] 0.8609421
```

Use an appropriate test to assess the evidence of how consistent is the data above with the suggested model. *(7 marks)*

End of Question Paper