



The
University
Of
Sheffield.

MAS113

SCHOOL OF MATHEMATICS AND STATISTICS

**Spring Semester
2013–2014**

MAS113 Introduction to Probability and Statistics

2 hours

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

- 1 Let S be the number of rain-free days in Sheffield next week:

$$S = \{0, 1, 2, 3, 4, 5, 6, 7\}.$$

Let $\mathcal{P}(S)$ be the power set of S . Give a counter-example to show that the following set function g is not a valid probability measure:

$$g(A) = \frac{\text{the largest element in } A}{7},$$

for any $A \in \mathcal{P}(S)$.

(2 marks)

- 2 Let X be a discrete random variable with probability mass function $p_X(x)$ and range R_X . For any two constants a and b , prove that

(i) $E(aX + b) = aE(X) + b;$ *(3 marks)*

(ii) $Var(aX + b) = a^2Var(X).$ *(3 marks)*

You may state, without proof, that $\sum_{x \in R_X} p_X(x) = 1$, and you may use the result in (i) to prove (ii).

- 3 Humans each carry 0, 1 or 2 copies of a particular gene. Let X be the number of copies of the gene carried by a child, and Y the number of copies carried by its mother. Standard genetic theory, plus background information about this particular gene, suggest the following joint distribution

		y		
		0	1	2
x	0	6/25	1/25	0
	1	6/25	7/25	2/25
	2	0	1/25	2/25

- (i) One mother is selected at random. Calculate the expected number of copies of the gene that she is carrying. *(2 marks)*
- (ii) A test determines that a woman is carrying two copies of the gene. She asks to know the probability that, if she has a child, the child will also have two copies of the gene. She is told:
- “The probability you want is given in row three, column three of the table: 2/25.”
- In one sentence, explain why the above statement is wrong, and calculate the probability that she has asked for. *(2 marks)*
- (iii) By inspecting the table only (do not give any calculations), explain why X and Y are not independent. *(1 mark)*
- (iv) For the following two parts, you must define your notation carefully and state the probability distributions you are using, to get full marks. Suppose ten children, each with different mothers, are selected at random.
- (a) Calculate the probability that at least two of them are carrying two copies of the gene. *(2 marks)*
- (b) If two of them are known to be carrying two copies of the gene, calculate the probability that four children out of the remaining eight are carrying no copies of the gene. *(3 marks)*

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

$$f_X(x) = \begin{cases} e^{2x} & \text{for } 0 \leq x \leq b, \\ 0 & \text{otherwise,} \end{cases}$$

where $b > 0$.

- (i) Derive the cumulative distribution function $F_X(x)$, tabulated for the three cases $x < 0$, $0 \leq x \leq b$, and $x > b$; *(2 marks)*
- (ii) Give a check that demonstrates that your cumulative distribution function in part (i) is correct. *(1 mark)*
- (iii) Deduce the value of b . *(1 mark)*
- (iv) For the following parts, you may leave your answers in terms of b .
 - (a) Calculate $P\left(X < \frac{b}{4} \mid X < \frac{b}{2}\right)$. *(2 marks)*
 - (b) Using integration by parts, find $E(X)$. *(3 marks)*

- 5 The time T (in days) to failure of a machine is modelled as an exponential random variable with rate 0.001, and density function

$$f_T(t) = \begin{cases} 0.001e^{-0.001t} & \text{for } 0 \leq t \\ 0 & \text{otherwise.} \end{cases}$$

Let $T_{0.95}$ be the 95th percentile of the distribution of T .

- (i) Is it correct to say $P(T = T_{0.95}) = 0.95$? Justify your answer. *(1 mark)*
 - (ii) Find the value of $T_{0.95}$. *(2 marks)*
- 6 (i) Let Z be a random variable that follows the standard normal distribution $N(0, 1)$. If $\Phi(z)$ denotes the c.d.f. of Z at point z , show that

$$\Phi(0) = \frac{1}{2}.$$

(2 marks)

- (ii) Let X be a random variable that follows the normal distribution with mean 1 and variance 4. Given the R output

```
> pnorm(1/2)
[1] 0.6914625
```

calculate the following probabilities

- (a) $P(1 \leq X < 2)$; *(2 marks)*
- (b) $P\left(|X| < \frac{1}{2}\right)$. *(2 marks)*

- 7 In a survey about how many hours per week do grandfathers spend sending and answering email, the results of 8 grandfathers age at least 75 were

0, 1, 2, 2, 7, 10, 14, 15.

Given the R output

```
> qt(0.995, 7)
[1] 3.499483
> qchisq(0.05, 7)
[1] 2.16735
> qchisq(0.95, 7)
[1] 14.06714
```

- (i) find a 99% confidence interval for the mean hours grandfathers spend sending and answering email; *(3 marks)*
- (ii) find a 90% confidence interval for the standard deviation of hours grandfathers spend sending and answering email. *(3 marks)*

- 8 When Vincenzo Baranello's blood pressure is in control, the systolic blood pressure reading has a mean of 130. For the last six times he has monitored his blood pressure, the recordings of which are shown in the table below

140, 150, 135, 155, 155, 140.

Given the R output

```
> pt(4.503, 5)
[1] 0.9968091
```

assess the strength of evidence this data provides on whether or not his blood pressure has changed. *(6 marks)*

- 9 A financial analyst trading regularly shares in the stock market, is interested in monitoring the number of shares, or trading volume, which are bought and sold each day. In the absence of any events that trigger heavy trading, volume tends to be the highest early in the week and lightest late in the week. To assess this claim data received on the trading volume of General Electric shares on Mondays (variable X) and Fridays (variable Y). The data are summarised in the table below

Sample size	Mean	Standard deviation
$n_X = 11$	$\bar{x} = 51.82$	$s_X = 17.39$
$n_Y = 12$	$\bar{y} = 50.00$	$s_Y = 11.34$

Given the R output

```
> pt(0.294, 10)
[1] 0.6126151
```

assess the claim that the trading volume of Monday is the same as that of Friday. *(5 marks)*

- 10 For murderers in the United States in 2009, the table below cross-tabulates the sex of the victim by the sex of the offender.

	Female Victim	Male victim
Female offender	182	484
Male offender	1719	4078

Given the R output

```
> pchisq(1.558, 1)
[1] 0.7880417
```

use a suitable test to test whether there is a difference between the gender of offender and the gender of victim. *(7 marks)*

End of Question Paper