The University Of Sheffield.

SCHOOL OF MATHEMATICS AND STATISTICS **Spring Semester 2014–2015**

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

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

1 A bookmaker is offering odds of 5 to 2 against Nico Rosberg winning the 2015 Formula 1 World Championship. This means that for every £1 that you bet, if Rosberg wins, you get your £1 back, plus an additional £2.50, but if Rosberg does not win, you lose your £1. If your subjective probability that Rosbserg wins is 0.3, calculate your expectation and variance of your profit if you bet £10 at these odds. ***(4 marks)***

2 Two playing cards are drawn at random from a standard 52 card deck (containing four ace cards). The cards are drawn without replacement (so the second card will be drawn from the remaining 51 cards in the deck).

(i) Let A_i be the event that the ith card is an ace. Are A_1 and A_2 independent? Justify your answer with suitable calculations. ***(3 marks)***

(ii) The first card drawn is given to Player 1, and the second card drawn is given to Player 2. Let X be the number of aces that Player 1 receives and Y be the number of aces that Player 2 receives. Calculate the covariance between X and Y. ***(4 marks)***

3 In a multiple choice test, there are five possible answers provided to each question. Suppose you take the test, and decide to pick one answer at random each time.

(i) If there were ten questions in total, what would be an appropriate probability distribution for the number of correct answers? ***(1 mark)***

(ii) If there were ten questions in total, what is the probability you would get at least two answers correct? ***(2 marks)***

(iii) If there were 100 questions in total, suggest an approximate 95% interval for the number of answers you would get correct (i.e. the probability of the number of correct answers lying in this interval would be approximately 0.95). Give a brief justification for you answer. ***(3 marks)***

4 In an infectious disease model, the time T (in days) from when an individual is infected to when the individual first displays symptoms of the disease is modelled with an exponential distribution, with probability density function given by

$$f_T(t) = \frac{1}{10}\exp\left(-\frac{t}{10}\right),$$

for $t > 0$ and 0 otherwise.

(i) Using the probability density function, derive an expression for $P(T > t)$. (Do not quote the formula for the cumulative distribution function without proof). ***(2 marks)***

(ii) Prove that $P(T > 10|T > 7) = P(T > 3)$. ***(2 marks)***

5 A random variable Y has probability density function given by

$$f_Y(y) = \frac{1}{2} - \frac{y}{8},$$

for $0 \leq y \leq 4$ and 0 otherwise.

(i) Copy and complete the following table for the cumulative distribution function $F_Y(y)$, showing your working.

	$F_Y(y)$
$y < 0$	
$0 \leq y \leq 4$	
$y > 4$	

(3 marks)

(ii) For the middle row of the table, corresponding to $F_Y(y)$ for values of y between 0 and 4, give a check to demonstrate that your expression for $F_Y(y)$ is correct. ***(1 mark)***

(iii) Calculate the standard deviation of Y. ***(5 marks)***

6 (i) Suppose that a random variable X has a moment generating function given by

$$M_X(t) = \frac{1}{3} + \frac{1}{2}e^t + \frac{1}{6}e^{2t}.$$

Find $\mathbb{E}(X)$ and $\mathrm{Var}(X)$. ***(5 marks)***

(ii) Now suppose that $X_1, X_2, \ldots, X_n, \ldots$ are i.i.d. copies of the random variable X from part (a), and define the random variables

$$S_n = X_1 + X_2 + \cdots + X_n.$$

Find expressions for the numbers a_n and b_n so that

$$\frac{S_n - a_n}{b_n} \to Z \text{ as } n \to \infty,$$

where $Z \sim N(0, 1)$. ***(2 marks)***

7 Measurements of the blood pressure of 25 elderly women have a mean of 140 mm of mercury. If these data are extracted from a normal population with $\sigma = 10$ mm of mercury,

(i) Find a 95% confidence interval for the population mean μ. ***(2 marks)***

(ii) How many women should be in the sample if the length of the confidence interval were to be at most 2? ***(3 marks)***

The following R output is available for you to use:

```
> qnorm(0.995,0,1)
[1] 2.576.
> qnorm(0.975,0,1)
[1] 1.96
> qnorm(0.95,0,1)
[1] 1.645
```

8 Suppose that for an unknown mean μ, we have two independent random variables $X \sim N(\mu, 2)$ and $Y \sim N(\mu, 4)$. Define $Z_1 = aX + 4Y$ and $Z_2 = 3X - bY$.

(i) Find the values of a and b so that both Z_1 and Z_2 are unbiased estimators for μ. ***(2 marks)***

(ii) Given the values of a and b found in (a), which of Z_1 and Z_2 is the better estimator? Present calculations to support your conclusion. ***(3 marks)***

9 A sample study was made of the number of business lunches that executives claim as deductible business expenses per month. It is found that 40 executives in the insurance industry averaged 9.1 such deductions, with a standard deviation of 1.9 within a given month; while 50 bank executives averaged 8.0 with a standard deviation of 2.1

Write down suitable (two-sided) hypotheses to be tested, and then calculate a p-value which enables you to test these. What is your conclusion? You may use the following R output:

```
> pt(2.13, nu)
[1] 0.98,
```

and your solution should include the missing value of `nu`. You may also assume that both of the underlying populations are normal. ***(6 marks)***

10 Consider the table below, which gives information about gender and happiness:

Gender	Not Happy	Pretty Happy	Very Happy
Female	174	587	328
Male	142	513	271

Test the null hypothesis that level of happiness is independent of gender, against the alternative that these are not independent. You may use the R output:

```
> pchisq(0.434, 2)
[1] 0.195
```

and you should round all calculated expected frequencies to the nearest integer.

(7 marks)

End of Question Paper