OXFORD CAMBRIDGE AND RSA EXAMINATIONS

Advanced Subsidiary General Certificate of Education
Advanced General Certificate of Education

MATHEMATICS 2641

Probability & Statistics 1

Tuesday **12 JUNE 2001** Afternoon 1 hour 20 minutes

Additional materials:
- Answer booklet
- Graph paper
- List of Formulae (MF8)

TIME 1 hour 20 minutes

INSTRUCTIONS TO CANDIDATES

- Write your Name, Centre Number and Candidate Number in the spaces provided on the answer booklet.
- Answer **all** the questions.
- Give non-exact numerical answers correct to 3 significant figures unless a different degree of accuracy is specified in the question or is clearly appropriate.
- You are permitted to use a graphic calculator in this paper.

INFORMATION FOR CANDIDATES

- The number of marks is given in brackets [ ] at the end of each question or part question.
- The total number of marks for this paper is 60.
- Questions carrying smaller numbers of marks are printed earlier in the paper, and questions carrying larger numbers of marks later in the paper.
- **You are reminded of the need for clear presentation in your answers.**

This question paper consists of 5 printed pages and 3 blank pages.

1 Every time a player throws a dart at a dartboard, the probability that she scores a 'double' is $\frac{1}{8}$ independently of the result of any other throw. She keeps throwing until she scores a double. Let T be the number of throws taken by the player up to and including the throw with which she first scores a double.

(i) Name the distribution of T, and find $\mathrm{E}(T)$. [3]

(ii) Find $\mathrm{P}(T = 3)$. [3]

2 In a game show a contestant was asked to identify the years in which each of 10 events occurred. The table below shows the year in which each event actually occurred and the year given by the contestant for that event.

Actual year of occurrence	1963	1968	1971	1973	1983	1984	1986	1990	1991	1997
Year given by contestant	1970	1983	1964	1977	1969	1992	1981	1986	1994	1997

(i) Calculate Spearman's rank correlation coefficient between the actual year of occurrence and the year given by the contestant. [4]

(ii) Does the value of this correlation coefficient suggest that the contestant is good at remembering the dates of these events? Give a reason for your answer. [2]

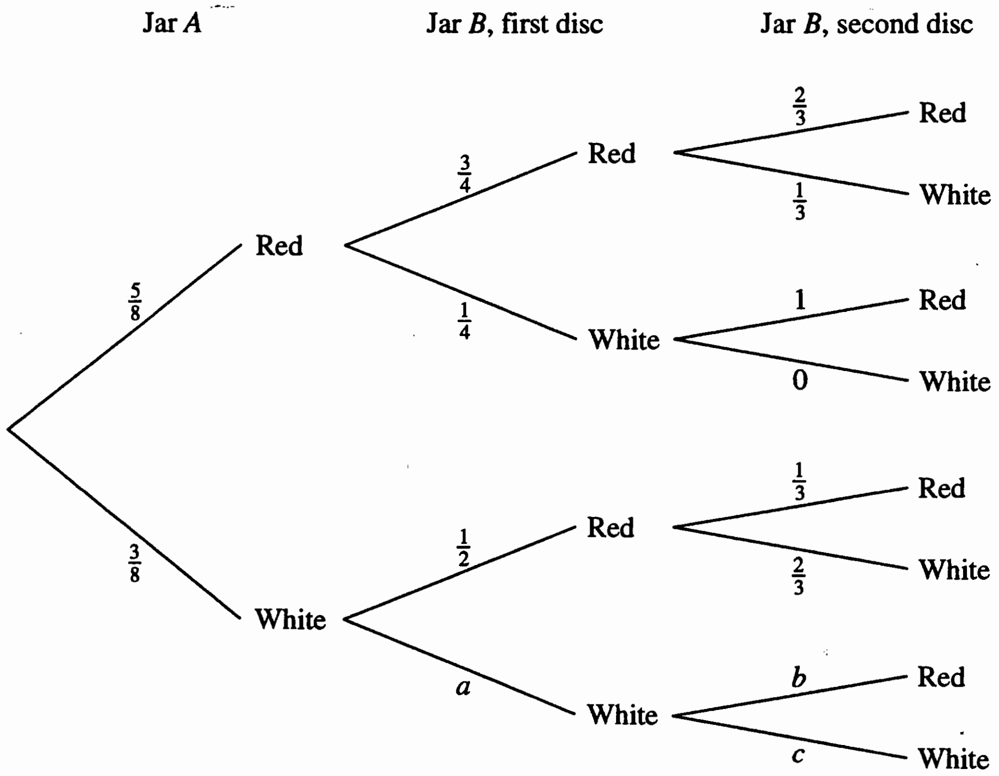
3 A bag contains 30 plastic tiles which are used in a word game. Each tile has a single letter written on it. 12 of the tiles have vowels written on them and the remaining 18 tiles have consonants written on them. A contestant in the game picks 7 tiles at random, without replacement.

(i) Find the probability that, of the 7 tiles, 4 have vowels written on them and 3 have consonants written on them. [3]

(ii) Find the probability that, of the 7 tiles, at least 1 has a vowel written on it. [2]

(iii) The letters written on the tiles are A B A E S S U. Calculate the number of different possible arrangements of these letters if the tiles are placed in a straight line. [2]

4 Two jars containing some discs are placed on a table. Jar A contains 5 red discs and 3 white discs, and Jar B contains 2 red discs and 1 white disc. A disc is selected from Jar A at random and placed in Jar B. Two discs are then selected at random from Jar B. The first of these two discs is not replaced in Jar B before the second disc is selected. The possible outcomes are represented by the tree diagram given below.



(i) State the values of the probabilities a, b and c. [3]

(ii) Let D be the total number of red discs selected from Jar B. Show that $P(D = 2) = \frac{3}{8}$. [2]

(iii) The probability distribution of D is given in the table below.

d	0	1	2
$P(D = d)$	$\frac{1}{16}$	$\frac{9}{16}$	$\frac{3}{8}$

Find $\text{Var}(D)$. [4]

5 In a science experiment a student dropped a ball on to the ground from a height of x metres and measured the height, y metres, to which it bounced. The experiment was repeated to give 10 pairs of results. The student chose values of x which increased in steps of exactly 0.25 metres. The results of the experiment are given in the table below.

x	0.25	0.50	0.75	1.00	1.25	1.50	1.75	2.00	2.25	2.50
y	0.11	0.17	0.29	0.38	0.46	0.54	0.63	0.71	0.84	0.90

[$n = 10$, $\Sigma x = 13.75$, $\Sigma y = 5.03$, $\Sigma x^2 = 24.0625$, $\Sigma y^2 = 3.1893$, $\Sigma xy = 8.7575$.]

(i) Calculate the product moment correlation coefficient between x and y. [2]

(ii) By calculating the equation of the appropriate regression line, estimate the height to which the ball will bounce if it is dropped from a height of 2.4 metres. [5]

(iii) Comment on the reliability of

(a) the estimate found in part **(ii)**, [1]

(b) the estimate obtained from the regression line of the height to which the ball will bounce if it is dropped from a height of 10.2 metres. [1]

6 Sheena travels to work by car. From long observation, she has found that she can park in her favourite parking space on 2 days out of 5 on average. Let X be the number of days out of a 5-day working week on which she can park in her favourite parking space.

(i) State two assumptions which need to be made for a binomial model to be valid for the distribution of the random variable X. [2]

(ii) Assuming that $B(n, p)$ is a valid model for the distribution of X,

(a) state the values of the parameters n and p, [2]

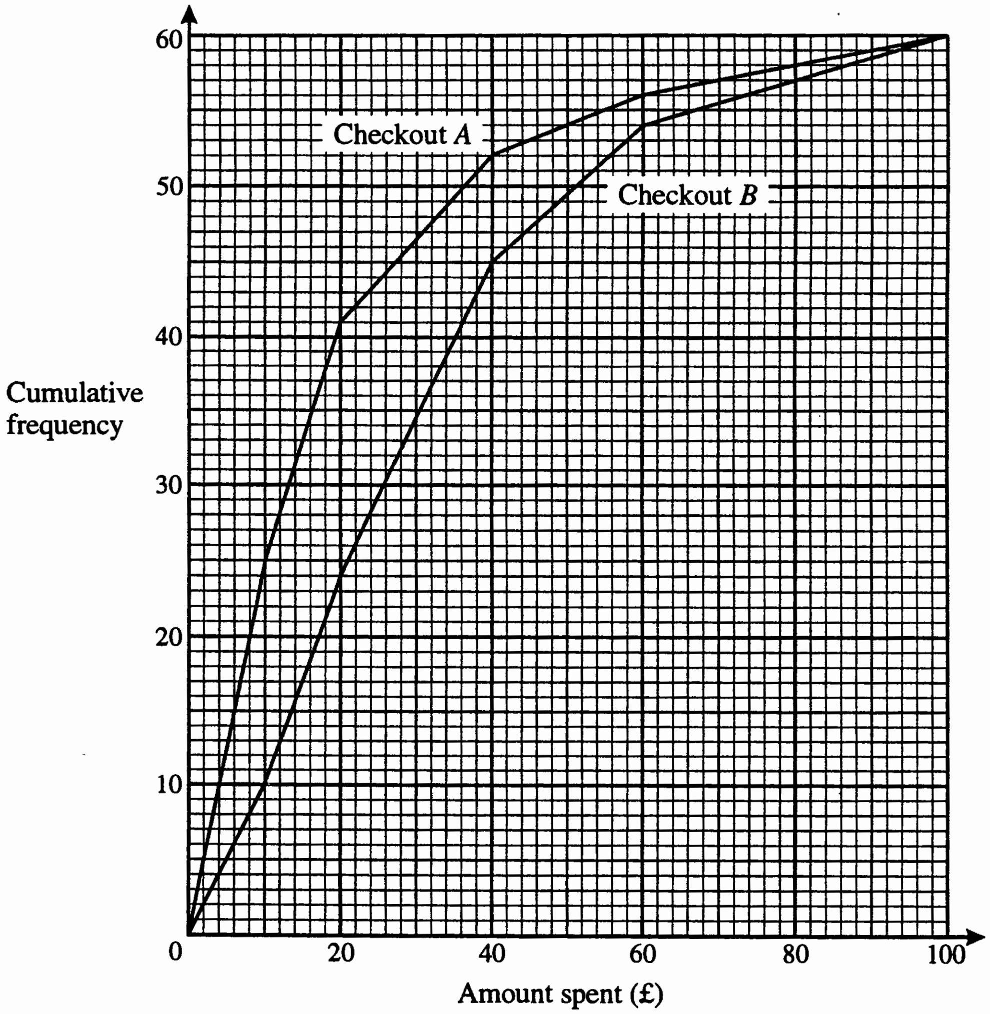
(b) show that $P(X > 3) = 0.0870$ correct to 3 significant figures. [2]

(iii) A 5-day working week in which Sheena can park in her favourite parking space on more than 3 days is a 'good' week. Find the probability that, out of 7 randomly chosen 5-day working weeks, fewer than 2 are good weeks. [4]

7 As part of a statistics project a student recorded the amount of money spent, in £, by each of a random sample of 60 customers at checkout A in a supermarket. She also recorded the amount spent by each of a random sample of 60 customers who used another checkout at checkout B in the same supermarket. The results are given in the table below.

Amount spent	$\leqslant$ £10	$\leqslant$ £20	$\leqslant$ £40	$\leqslant$ £60	$\leqslant$ £100
Cumulative frequency for Checkout A	25	41	52	56	60
Cumulative frequency for Checkout B	10	24	45	54	60

The diagram shows the cumulative frequency graphs for the data.



(i) Use the diagram to estimate the median amount spent at

(a) checkout A,

(b) checkout B. [3]

(ii) Use the diagram to estimate the interquartile range of the amount spent at

(a) checkout A,

(b) checkout B. [4]

(iii) One of the two checkouts was an 'express' checkout. Customers are allowed a maximum of nine items when they pass through an express checkout. State, with a reason, which of the two checkouts, A or B, was more likely to have been the express checkout. [2]

(iv) Calculate an estimate of the mean amount spent at checkout B. [4]

1 Random variable T has a **Geometric Distribution** with $p = \frac{1}{8}$. [3]

$\text{E}[T] = \frac{1}{p} = \mathbf{8}$ $\qquad$ $\text{P}(T = 3) = q^2 p = \mathbf{\frac{49}{512}}$ [3]

2

actual rank	1	2	3	4	5	6	7	8	9	10
contestant's rank	3	6	1	4	2	8	5	7	9	10
d	-2	-4	2	0	3	-2	2	1	0	0

$$r_s = 1 - \frac{6\sum d^2}{n\left(n^2 - 1\right)} = 1 - \frac{6 \times 42}{10 \times 99} = \mathbf{0 \cdot 74\dot{5}\dot{4}5}$$ [3]

The coefficient indicates that the contestant is quite good at *ordering* the events (since the coefficient is positive and quite near $+1$), though not necessarily at remembering the dates precisely. [2]

3

$p\left(4v,\ 3c\right) = \frac{{}^{12}C_4 \times {}^{18}C_3}{{}^{30}C_7} = \frac{495 \times 816}{2035800} = \mathbf{0 \cdot 198}$ $\quad$ $\left(3\ s.f.\right)$ [3]

$p\left(\geq 1v\right) = 1 - p\left(\text{no vowels}\right) = 1 - \frac{{}^{18}C_7}{{}^{30}C_7} = 1 - \frac{31824}{2035800} = \mathbf{0 \cdot 984}$ $\quad$ $\left(3\ s.f.\right)$ [2]

$\text{no. of arrangements} = \frac{7!}{2!2!} = \mathbf{1260}$ [2]

4

$\boldsymbol{a = \frac{1}{2}}$ $\qquad$ $\boldsymbol{b = \frac{2}{3}}$ $\qquad$ $\boldsymbol{c = \frac{1}{3}}$ [3]

$p\left(D = 2\right) = \left(\frac{5}{8} \times \frac{3}{4} \times \frac{2}{3}\right) + \left(\frac{3}{8} \times \frac{1}{2} \times \frac{1}{3}\right) = \frac{5}{16} + \frac{1}{16} = \frac{6}{16} = \mathbf{\frac{3}{8}}$ $\quad$ $(show)$ [2]

$\text{E}\left[D\right] = \sum x_i p_i = 0 + \frac{9}{16} + \frac{6}{8} = \frac{21}{16}$

$\text{Var}\left[D\right] = \sum x_i^{\,2} p_i - \text{E}\left[D\right]^2 = 0 + \frac{9}{16} + \frac{12}{8} - \left(\frac{21}{16}\right)^2 = \frac{87}{256} = \mathbf{0 \cdot 340}$ $\quad$ $\left(3\ s.f.\right)$ [4]

5

$$r = \frac{\frac{1}{n}\sum xy - \overline{x}\,\overline{y}}{\sqrt{\frac{1}{n}\sum x^2 - \overline{x}^2}\sqrt{\frac{1}{n}\sum y^2 - \overline{y}^2}} = \frac{0 \cdot 184125}{0 \cdot 71807... \times 0 \cdot 25675...} = \mathbf{0 \cdot 999} \quad \left(3\ s.f.\right)$$ [4]

regression line of y on x

$y - \overline{y} = \frac{\frac{1}{n}\sum xy - \overline{x}\,\overline{y}}{\frac{1}{n}\sum x^2 - \overline{x}^2}\left(x - \overline{x}\right)$ $\qquad$ $y - 0 \cdot 503 = 0 \cdot 35709...(x - 1.375)$ $\qquad$ $\mathbf{y = 0 \cdot 3571x + 0 \cdot 012}$

estimate of height $\boldsymbol{y}$ **= 0.869 m** (3 s.f.) [5]

This estimate should be reliable because r is very close to 1 and we are not extrapolating beyond the bounds of the data set. [1]

10.2 is far above the maximum x value (2.50) in the data set therefore no confidence should be invested in the prediction from the regression line. [1]

6 **Assumptions:** the prob. of 0.4 applies equally well to the five different days of the week.

success on one day is independent of success on any other day. [2]

$\boldsymbol{n = 5} \quad \boldsymbol{p = \frac{2}{5}} \qquad p(X > 3) = p(4) + p(5) = 5\left(\frac{2}{5}\right)^4\left(\frac{3}{5}\right) + \left(\frac{2}{5}\right)^5 = 0 \cdot 08704 = \mathbf{0 \cdot 0870} \quad (3\ s.f.)$ [2] [2]

No. of 'good' weeks is B(7, 0.08704)

$\therefore\ p\Big(< 2 \text{ good weeks}\Big) = p(0) + p(1) = (0 \cdot 91296)^7 + 7(0 \cdot 08704)(0 \cdot 91296)^6 = \mathbf{0 \cdot 881}$

[4]

7 Estimates

$\text{med}_A = \mathbf{£13}$ $\qquad \text{med}_B = \mathbf{£26}$ [3]

$\text{IQR}_A = 27 - 6 = \mathbf{£21}$ $\qquad \text{IQR}_B = 40 - 14 = \mathbf{£26}$ (*accept* £27) [4]

Checkout A was probably the 'express': since with a limit of five items the median ought to be lower. [2]

$$\text{estimate of mean} = \frac{5 \times 25 + 15 \times 16 + 30 \times 11 + 50 \times 4 + 80 \times 4}{60} = \frac{1215}{60} = \mathbf{£20 \cdot 25}$$

[4]

[Total 60]