OXFORD CAMBRIDGE AND RSA EXAMINATIONS

Advanced Subsidiary General Certificate of Education
Advanced General Certificate of Education

MATHEMATICS 4735

Probability & Statistics 4

Specimen Paper

Additional materials:
Answer booklet
Graph paper
List of Formulae (MF 1)

TIME 1 hour 30 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 72.
- 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 4 printed pages.

[Turn over

1 A continuous random variable X has moment generating function given by

$$M_X(t) = \frac{9}{(3-t)^2}.$$

Find the mean and variance of X. [5]

2 The events A and B are independent, and $P(A) = P(B) = p$, where $0 < p < 1$.

(i) Express $P(A \cup B)$ in terms of p. [3]

(ii) Given that $P\big((A \cap B) \mid (A \cup B)\big) = \frac{1}{2}$, find the value of $P\big((A \cap B') \cup (A' \cap B)\big)$. [5]

3 A University's Department of Computing is interested in whether students who have passed A level Mathematics perform better in Computing examinations that those who have not.

A random sample of 19 students was taken from those students who took a particular first year Computing examination. This sample included 12 students who have passed A level Mathematics and 7 students who have not. The marks gained in the Computing examination were as follows:

Students who have passed A level Mathematics: 27, 34, 39, 41, 45, 47, 55, 59, 66, 75, 78, 86.

Students who have not passed A level Mathematics: 17, 21, 28, 35, 37, 54, 64.

Use a suitable non-parametric test to determine if there is evidence, at the 5% significance level, that students who have passed A level Mathematics gain a higher average mark than students who have not passed A level Mathematics. (A normal approximation may be used.) [10]

4 The continuous random variable X has probability density function given by

$$f(x) = \begin{cases} kx & 0 \leqslant x \leqslant a, \\ 0 & \text{otherwise,} \end{cases}$$

where k is a constant and the value of the parameter a is unknown.

(i) Show that $k = \dfrac{2}{a^2}$. [2]

The random variable U is defined by $U = \frac{3}{2}X$.

(ii) Show that U is an unbiased estimator of a. [3]

(iii) Find, in terms of a, the variance of U. [4]

The random variable λX^n, where n is a positive integer and λ is a constant, is an unbiased estimator of a^n.

(iv) Express λ in terms of n. [2]

5 **(i)** Explain briefly the circumstances under which a non-parametric test of significance should be used in preference to a parametric test. [1]

The acidity of soil can be measured by its pH value. As a part of a Geography project a student measured the pH values of 14 randomly chosen samples of soil in a certain area, with the following results.

5.67 5.73 6.64 6.76 6.10 5.41 5.80 6.52 5.16 5.10 6.71 5.89 5.68 5.37

(ii) Use a Wilcoxon signed-rank test to test whether the average pH value for soil in this area is 6.24. Use a 10% level of significance. [5]

Some time later, the pH values of soil samples taken at exactly the same locations as before were again measured. It was found that, for 3 of the 14 locations, the new pH value was higher than the previous value, while for the other 11 locations the new value was lower.

(iii) Test, at the 5% significance level, whether there is evidence that the average pH value of soil in this area is lower than previously. [5]

6 The joint probability distribution of the discrete random variables X and Y is shown in the following table.

		x = −1	x = 0
y	2	$\frac{1}{6}$	$\frac{2}{9}$
	3	$\frac{5}{18}$	$\frac{1}{3}$

(i) Show that $\mathrm{E}(X) = -\frac{4}{9}$ and find $\mathrm{Var}(X)$. [4]

(ii) Write down the distributions of X conditional on $Y = 2$ and X conditional on $Y = 3$. Find the means of these conditional distributions, and hence verify that

$$\mathrm{E}(X) = \mathrm{E}(X \mid Y = 2) \times \mathrm{P}(Y = 2) + \mathrm{E}(X \mid Y = 3) \times \mathrm{P}(Y = 3).$$ [3]

It is given that $\mathrm{E}(Y) = \frac{47}{18}$ and $\mathrm{Var}(Y) = \frac{77}{324}$.

(iii) Find $\mathrm{Cov}(X, Y)$ and state, with a reason, whether X and Y are independent. [4]

(iv) Find $\mathrm{Var}(X + Y)$. [2]

7 The random variable X has a geometric distribution with parameter p.

(i) Show that the probability generating function $\mathrm{G}_X(t)$ of X is given by

$$\mathrm{G}_X(t) = \frac{pt}{1 - t(1-p)}.$$ [3]

(ii) Hence show that $\mathrm{E}(X) = \frac{1}{p}$ and that $\mathrm{Var}(X) = \frac{1-p}{p^2}$. [5]

A child has 4 fair, six-sided dice, one white, one yellow, one blue and one red.

(iii) The child rolls the white die repeatedly until the die shows a six. The number of rolls up to and including the roll on which the white die first shows a six is denoted by W. Write down an expression for $\mathrm{G}_W(t)$. [1]

(iv) The child then repeats this process with the yellow die, then with the blue die and then with the red die. By finding an appropriate probability generating function, find the probability that the total number of rolls of the four dice, up to and including the roll on which the red die first shows a six, is exactly 24. [4]

1 *EITHER*: $M_X'(t) = \frac{18}{(3-t)^3}$	B1	For correct differentiation of the mgf
Hence $E(X) = M_X'(0) = \frac{2}{3}$	B1√	For correct value for the mean
$M_X''(t) = \frac{54}{(3-t)^4}$	B1	For correct second derivative
Hence $\text{Var}(X) = M_X''(0) - \{E(X)\}^2 = \frac{2}{3} - \frac{4}{9} = \frac{2}{9}$	M1	For correct method for the variance
	A1√	For correct answer
OR: $M_X(t) = 1 + \frac{2}{3}t + \frac{1}{3}t^2 + \ldots$	M1	For attempting binomial expansion of mgf
	A1	For first three terms correct (unsimplified)
Hence $E(X) = \frac{2}{3}$	A1√	For correct value for the mean
$\text{Var}(X) = (2!) \times \frac{1}{3} - \{E(X)\}^2 = \frac{2}{3} - \frac{4}{9} = \frac{2}{9}$	M1	For correct method for the variance
	A1√ **5** **[5]**	For correct answer
2 **(i)** $P(A \cup B) = p + p - p \times p = 2p - p^2$	M1	For use of $P(A) + P(B) - P(A \cap B)$
	B1	For $P(A \cap B) = P(A)P(B)$ since independent
	A1 **3**	For correct expression $2p - p^2$
(ii) $\frac{p^2}{2p - p^2} = \frac{1}{2} \Rightarrow 2p = 2 - p \Rightarrow p = \frac{2}{3}$	B1√	For equation $\frac{P(A \cap B)}{P(A \cup B)} = \frac{1}{2}$
	M1	For solving relevant equation for p
	A1	For correct value
Hence $P\big((A \cap B') \cup (A' \cap B)\big) = 2 \times \frac{2}{3} \times \frac{1}{3} = \frac{4}{9}$	M1	For calculation of $2p(1-p)$ or equivalent
	A1 **5** **[8]**	For correct answer $\frac{4}{9}$
3 H_0 : population medians equal , H_1 : higher median for those who passed Mathematics	B1	For both hypotheses stated correctly
Ranking: Pass: 3, 5, 8, 9, 10, 11, 13, 14, 16, 17, 18, 19 Not pass: 1, 2, 4, 6, 7, 12, 15	M1	For attempt at ranking correctly
Sum of ranks of those not passing is 47	A1	For correct sum of ranks
$R_m \sim N(\frac{1}{2} \times 7 \times 20, \frac{1}{12} \times 7 \times 12 \times 20) = N(70, 140)$	M1	For using the appropriate normal approx
	A1	For both parameters correct
EITHER: Test statistic is $\frac{47.5 - 70}{\sqrt{140}} = -1.902$	M1	For standardising
	A2	For correct value of test statistic (allow A1 if correct apart from missing or wrong c.c.)
This is less than -1.645	M1	For comparison with correct critical value
OR: Critical region is $\frac{X + 0.5 - 70}{\sqrt{140}} < -1.645$	M1	For setting up the appropriate inequality
i.e. $X \leqslant 50$	A2	For correct critical region (allow A1 if correct apart from missing or wrong c.c.)
Sample value 47 lies in the critical region	M1	For comparing 47 with critical region
Hence there is evidence that those passing Mathematics have a higher average score	A1√ **10** **[10]**	For conclusion stated in context

4	**(i)** $\int_0^a kx\,dx = 1 \Rightarrow \frac{1}{2}ka^2 = 1 \Rightarrow k = \frac{2}{a^2}$	M1		For use of $\int_0^a f(x)\,dx = 1$
		A1	**2**	For showing the given answer correctly
	(ii) $E(U) = \frac{3}{2}\int_0^a kx^2\,dx = \frac{3}{2}\times\frac{1}{3}ka^3 = a$	B1		For stating or implying $E(U) = \frac{3}{2}E(X)$
		M1		For use of $\int_0^a x f(x)\,dx$
	Hence U is an unbiased estimator of a	A1	**3**	For showing the given result correctly
	(iii) $E(U^2) = \int_0^a \left(\frac{3}{2}x\right)^2 kx\,dx = \frac{9}{16}ka^4 = \frac{9}{8}a^2$	M1		For correct process for $E(U^2)$
		A1		For correct value $\frac{9}{8}a^2$
	Hence $\text{Var}(U) = \frac{9}{8}a^2 - a^2 = \frac{1}{8}a^2$	M1		For correct process for $\text{Var}(U)$
		A1√	**4**	For correct answer (Alternatively via $\text{Var}(U) = \frac{9}{4}\text{Var}(X)$.)
	(iv) $\frac{2\lambda}{a^2}\int_0^a x^{n+1}\,dx = a^n \Rightarrow \frac{2\lambda}{a^2}\times\frac{a^{n+2}}{n+2} = a^n$	M1		For using $\lambda E(X^n) = a^n$
	Hence $\lambda = \frac{1}{2}(n+2)$	A1	**2**	For correct answer
			11	
5	**(i)** A non-parametric test is needed when there is no information (or reasonable assumption) available about an underlying distribution	B1	**1**	For a correct statement
	(ii) H_0: population median pH is 6.24, H_1: population median pH is not 6.24	B1		For both hypotheses stated correctly
	Deviations from NH value 6.24 are: –0.57 –0.51 0.40 0.52 –0.14 –0.83 –0.44 0.28 –1.08 –1.14 0.47 –0.35 –0.56 –0.87	M1		For calculating signed differences from 6.24
	Signed ranks are : –10 –7 4 8 –1 –11 –5 2 –13 –14 6 –3 –9 –12	M1		For calculating signed ranks
	Test statistic is $2+4+6+8=20$	A1		For the correct value of the test statistic
	This is less than the critical value of 25, so we conclude that there is evidence to suggest that the average pH value is not 6.24	M1		For comparing with the correct critical value
		A1	**6**	For correct conclusion based on correct work
	(iii) H_0: same average pH as before; H_1: lower value	B1		For both hypotheses stated correctly
	$P(\leqslant 3 \text{ out of } 14 \mid H_0) = 0.0287$	M1		For relevant use of $B(14, \frac{1}{2})$
		A1		For correct value 0.0287
	This is less than 0.05, so we reject H_0 and	M1		For comparing with 0.05
	conclude that the average pH is now lower	A1	**5**	For correct conclusion based on correct work
			12	

Question	Answer	Marks	Guidance
6 **(i)**	Marginal probabilities for X are $\frac{4}{9}, \frac{5}{9}$	B1	For appropriate addition
	Hence $\mathrm{E}(X) = -1\times\frac{4}{9} + 0\times\frac{5}{9} = -\frac{4}{9}$	B1	For showing the given answer correctly
	$\mathrm{Var}(X) = (-1)^2\times\frac{4}{9} - \left(-\frac{4}{9}\right)^2 = \frac{20}{81}$	M1	For correct process for variance
		A1 **4**	For correct value
(ii)	x: -1, 0; $\mathrm{P}_2(X = x)$: $\frac{3}{7}$, $\frac{4}{7}$ x: -1, 0; $\mathrm{P}_3(X = x)$: $\frac{5}{11}$, $\frac{6}{11}$	B1	For both conditional distributions correct
	Hence $\mathrm{E}(X \mid Y = 2) = -\frac{3}{7}$, $\mathrm{E}(X \mid Y = 3) = -\frac{5}{11}$	B1	For both conditional expectations correct
	$\mathrm{RHS} = -\frac{3}{7}\times\frac{7}{18} - \frac{5}{11}\times\frac{11}{18} = -\frac{4}{9} = \mathrm{E}(X)$	B1 **3**	For correct verification
(iii)	$\mathrm{E}(XY) = -2\times\frac{1}{6} - 3\times\frac{5}{18} = -\frac{7}{6}$	M1	For evaluation of $\mathrm{E}(XY)$
	$\mathrm{Cov}(X, Y) = -\frac{7}{6} - (-\frac{4}{9})\times\frac{47}{18} = -\frac{1}{162}$	M1	For correct method for $\mathrm{Cov}(X, Y)$
		A1	For correct value (fraction or decimal)
	X and Y are not independent, as $\mathrm{Cov}(X, Y) \neq 0$	B1 **4**	For correct conclusion, with correct reason
(iv)	$\mathrm{Var}(X + Y) = \frac{20}{81} + \frac{77}{324} - \frac{2}{162} = \frac{17}{36}$	M1	For use of $\mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X, Y)$
		A1√ **2**	For correct value
		13	
7 **(i)**	$\mathrm{G}_X(t) = \sum_{r=1}^{\infty} q^{r-1}pt^r$, where $q = 1 - p$	B1	For correct statement of the required sum
	$= pt\sum_{r=1}^{\infty}(qt)^{r-1} = \dfrac{pt}{1-qt} = \dfrac{pt}{1-(1-p)t}$	M1	For summing the relevant GP
		A1 **3**	For showing the given answer correctly
(ii)	$\mathrm{G}'_X(t) = \dfrac{p}{(1-qt)^2}$	B1	For correct derivative, in any form
	Hence $\mathrm{E}(X) = \mathrm{G}'_X(1) = \dfrac{p}{p^2} = \dfrac{1}{p}$	B1	For showing the given answer correctly
	$\mathrm{G}''_X(t) = \dfrac{2pq}{(1-qt)^3}$	B1√	For correct second derivative, in any form
	Hence $\mathrm{Var}(X) = \mathrm{G}''_X(1) + \dfrac{1}{p} - \dfrac{1}{p^2}$	M1	For use of $\mathrm{G}''(1) + \mathrm{G}'(1) - \{\mathrm{G}'(1)\}^2$
	$= \dfrac{2pq}{p^3} + \dfrac{1}{p} - \dfrac{1}{p^2} = \dfrac{q}{p^2} = \dfrac{1-p}{p^2}$	A1 **5**	For showing the given answer correctly
(iii)	$\mathrm{G}_W(t) = \dfrac{\frac{1}{6}t}{1-\frac{5}{6}t}$	B1 **1**	For correct expression, in any form
(iv)	Required pgf is $\left(\dfrac{\frac{1}{6}t}{1-\frac{5}{6}t}\right)^4$	B1	For stating fourth power of $\mathrm{G}_W(t)$
	Required probability is the coefficient of t^{24}	B1	For stating or implying the required coeff
	This is $\left(\frac{1}{6}\right)^4 \times \dfrac{(-4)(-5)(-6)...(-23)}{20!} \times \left(\frac{5}{6}\right)^{20}$	M1	For use of appropriate binomial coefficient
	≈ 0.0356	A1 **4**	For correct value
		13	