OCR RECOGNISING ACHIEVEMENT

ADVANCED GCE **4735/01**

MATHEMATICS

Probability & Statistics 4

WEDNESDAY 18 JUNE 2008 Morning

Time: 1 hour 30 minutes

Additional materials (enclosed): None

Additional materials (required):
Answer Booklet (8 pages)
List of Formulae (MF1)

INSTRUCTIONS TO CANDIDATES

- Write your name in capital letters, your Centre Number and Candidate Number in the spaces provided on the Answer Booklet.
- Read each question carefully and make sure you know what you have to do before starting your answer.
- 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 graphical 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**.
- **You are reminded of the need for clear presentation in your answers.**

This document consists of **4** printed pages.

[Turn over

1 For the mutually exclusive events A and B, $P(A) = P(B) = x$, where $x \neq 0$.

(i) Show that $x \leqslant \frac{1}{2}$. [1]

(ii) Show that A and B are not independent. [2]

The event C is independent of A and also independent of B, and $P(C) = 2x$.

(iii) Show that $P(A \cup B \cup C) = 4x(1 - x)$. [4]

2 Part of Helen's psychology dissertation involved the reaction times to a certain stimulus. She measured the reaction times of 30 randomly selected students, in seconds correct to 2 decimal places. The results are shown in the following stem-and-leaf diagram.

14	1 2
15	2 4
16	0 3 6
17	1 5 7
18	3 4 5 7 9
19	2 4 6 7 8 9
20	0 1 3 4 5 7 8 9
21	7

Key: 18 | 3 means 1.83 seconds

Helen wishes to test whether the population median time exceeds 1.80 seconds.

(i) Give a reason why the Wilcoxon signed-rank test should not be used. [1]

(ii) Carry out a suitable non-parametric test at the 5% significance level. [7]

3 From the records of Mulcaster United Football Club the following distribution was suggested as a probability model for future matches. X and Y denoted the numbers of goals scored by the home team and the away team respectively.

		X = 0	1	2	3
Y	0	0.11	0.04	0.06	0.08
	1	0.08	0.05	0.12	0.05
	2	0.05	0.08	0.07	0.03
	3	0.03	0.06	0.07	0.02

Use the model to find

(i) $\mathrm{E}(X)$, [3]

(ii) the probability that the away team wins a randomly chosen match, [2]

(iii) the probability that the away team wins a randomly chosen match, given that the home team scores. [4]

One of the directors, an amateur statistician, finds that $\mathrm{Cov}(X, Y) = 0.007$. He states that, as this value is very close to zero, X and Y may be considered to be independent.

(iv) Comment on the director's statement. [2]

4 William takes a bus regularly on the same journey, sometimes in the morning and sometimes in the afternoon. He wishes to compare morning and afternoon journey times. He records the journey times on 7 randomly chosen mornings and 8 randomly chosen afternoons. The results, each correct to the nearest minute, are as follows, where M denotes a morning time and A denotes an afternoon time.

M	A	A	M	M	M	M	M	M	A	A	A	A	A	A
19	20	22	24	25	26	28	30	31	33	35	37	38	39	42

William wishes to test for a difference between the average times of morning and afternoon journeys.

(i) Given that $s_M^2 = 16.5$ and $s_A^2 = 64.5$, with the usual notation, explain why a t-test is not appropriate in this case. [1]

(ii) William chooses a non-parametric test at the 5% significance level. Carry out the test, stating the rejection region. [6]

5 The discrete random variable X has moment generating function $\frac{1}{4}e^{2t} + ae^{3t} + be^{4t}$, where a and b are constants. It is given that $\mathrm{E}(X) = 3\frac{3}{8}$.

(i) Show that $a = \frac{1}{8}$, and find the value of b. [6]

(ii) Find $\mathrm{Var}(X)$. [4]

(iii) State the possible values of X. [1]

6 The continuous random variable Y has cumulative distribution function given by

$$\mathrm{F}(y) = \begin{cases} 0 & y < a, \\ 1 - \dfrac{a^3}{y^3} & y \geqslant a, \end{cases}$$

where a is a positive constant. A random sample of 3 observations, Y_1, Y_2, Y_3, is taken, and the smallest is denoted by S.

(i) Show that $\mathrm{P}(S > s) = \left(\dfrac{a}{s}\right)^9$ and hence obtain the probability density function of S. [5]

(ii) Show that S is not an unbiased estimator of a, and construct an unbiased estimator, T_1, based on S. [4]

It is given that T_2, where $T_2 = \frac{2}{9}(Y_1 + Y_2 + Y_3)$, is another unbiased estimator of a.

(iii) Given that $\mathrm{Var}(Y) = \frac{3}{4}a^2$ and $\mathrm{Var}(S) = \frac{9}{448}a^2$, determine which of T_1 and T_2 is the more efficient estimator. [3]

(iv) The values of Y for a particular sample are 12.8, 4.5 and 7.0. Find the values of T_1 and T_2 for this sample, and give a reason, unrelated to efficiency, why T_1 gives a better estimate of a than T_2 in this case. [3]

7 The probability generating function of the random variable X is given by

$$\mathrm{G}(t) = \frac{1 + at}{4 - t},$$

where a is a constant.

(i) Find the value of a. [2]

(ii) Find $\mathrm{P}(X = 3)$. [4]

The sum of 3 independent observations of X is denoted by Y. The probability generating function of Y is denoted by $\mathrm{H}(t)$.

(iii) Use $\mathrm{H}(t)$ to find $\mathrm{E}(Y)$. [5]

(iv) By considering $\mathrm{H}(-1) + \mathrm{H}(1)$, show that $\mathrm{P}(Y \text{ is an even number}) = \frac{62}{125}$. [2]

4735 Statistics 4

1 (i)	Use $P(A) + P(B) - P(A\cap B) \le 1$, $P(A\cap B)=0$	B1	**1**	AEF
(ii)	Use $P(A\|B)=P(A\cap B) / P(B)$ Use $P(A\cap B) = 0$ with argument with $x\neq 0$	M1 A1	 **2**	AEF e.g. Inependent if $(A\cap B)=P(A)P(B)=x^2$, $P(A\cap B)=0$, $x\neq 0$, so A and B are not indep.
(iii)	Use $P(A\cup B\cup C) = P(A)+P(B)+P(C)- P(A\cap B) - P(B \cap C) - P(C\cap A) + P(A \cap B \cap C)$ Use $P(A\cap B)= 0$; $P(A \cap B \cap C)=0$ $P(B\cap C)=2x^2$; $P(C\cap A)=2x^2$ Substitute and obtain required result AG	M1 A1 A1 A1	 4 (7)	Or equivalent. Allow one sign error For both For both
2 (i)	Wilcoxon test requires a symmetric distribution not supported by the diagram	B1	**1**	Or equivalent
(ii)	H_0: $m = 1.80$, H_1: $m > 1.80$ Use sign test Number exceeding 1.8 = 20 Use B(30,0.5), $P(\ge 20)$ Or $P(\le 10)$ 0.0494 Compare with 0.05 correctly Conclude there is significant evidence that the median time exceeds 1.80 sec	 M1 A1 M1 A1 M1 A1√	B1 **7 (8)**	Needs "population median" if words OR: 1.645 if N(15,7.5) , z =1.643, 1.816, 2.008 used; OR CR ($X \ge 20$) ft p or z
3 (i)	Marginal distribution of X x 0 1 2 3 p 0.27 0.23 0.32 0.18 1×0.23+2×0.32+ 3×0.18 =1.41	 B1 M1 A1	 **3**	
(ii)	$P(Y>X)=0.08+0.05+0.03+0.08+0.06+0.07$ $= 0.37$	M1 A1	 **2**	
(iii)	Use $P(Y > X \cap X > 0)/ P(X > 0)$ $P(X > 0)= 0.73$ $P(Y>X \cap X > 0)=0.08+0.06+0.07$ 21/73	M1 A1 A1 A1	 **4**	 From marginal distribution AEF
(iv)	The director cannot conclude independence from cov. So director's conclusion incorrect. OR: Eg $P(X=0 \cap Y=0)=0.11$, $P(X=0)P(Y=0)=0.27\times 0.29\neq P(X=0 \cap Y=0)$	M1 A1 M1 A1	 **2 (11)**	Idea that independence implies cov = 0 but not the reverse

4 (i)	Variances seem not to be equal	B1	**1**	
(ii)	H_0: $m_M = m_A$, H_1: $m_M \neq m_A$	B1		Both hypotheses, AEF. Not "average"
	$R_m = 40$, $m(m+n+1)$- $R_m = 72$	M1		Both found
	$W = 40$	A1		A0 if no or wrong 72
	CR: $W \leq 38$	B1		
	40 not in CR, so do not reject H_0	M1		Or equivalent
	Insufficient evidence that median times differ	A1	**6 (7)**	In context. B1 if no M1 but conclusion correct Allow average here
5 (i)	$a+b = \frac{3}{4}$	B1		From M(0)=1
	$M'(0)= 3^3/_8$	M1		
	$\frac{1}{2} + 3a + 4b = 3\frac{3}{8}$	A1		AEF
	Solve simultaneously	M1		Elimination or substitution
	$a = \frac{1}{8}$ AG	A1		
	$b = \frac{5}{8}$	A1	**6**	
(ii)	$M''(t) = e^{2t} + {}^9/_8 e^{3t} + 10e^{4t}$	B1		
	$M''(0) - (M'(0))^2$	M1		
	${}^{97}/_8 - (3\frac{3}{8})^2$; ${}^{47}/_{64}$	A1A1	**4**	
(iii)	x= 2, 3, 4	B1	**1 (11)**	
6 (i)	$P(Y > y) = 1- F(y)$	M1		Allow any variables
	$= a^3/y^3$	A1		
	$P(S > s)$ =P(all 3 values $>s$)$=(a/s)^9$ AG	A1		
	$\mathbf{f}(s) = d/ds(1 - (a/s)^9)$	M1		
	$= \begin{cases} 9\dfrac{a^9}{s^{10}} & s \geq a, \\ 0 & s < a \end{cases}$	A1	**5**	
(ii)	$\int_a^\infty \frac{a^9}{s^9} ds$	M1		
	$= 9a/8$	A1		
	S not unbiased since this not equal to a	M1		
	$T_1 = 8S/9$	B1√	**4**	Ft E(S)
(iii)	$Var(T_1)= a^2/63$, $Var\ T_2 = a^2/9$	M1 A1 for both		Correct method
	$Var(T_1) < Var(T_2)$, T_1 is more efficient	A1√	**3**	Comparison, completion.. √ one variance correct with same dimensions
(iv)	$t_1 = 4.0$, $t_2 = 5.4$	B1		Both
	From data $a \leq 4.5$ and $t_2 > 4.5$	B1B1	**3 (15)**	AEF

7	**(i)**	G(1) = 1	M1		
		$a = 2$	A1	**2**	
	(ii)	$(1+2t)/(4-t) = c\,(1+2t)(1-\tfrac{1}{4}t)^{-1}$	M1		c= ¼ or 4
		$=\tfrac{1}{4}(1+2t)(1 + \tfrac{1}{4}t + (\tfrac{1}{4}t)^2 + \ldots)$	A1		
		Coefficient of $t^3 = \tfrac{1}{4}[(\tfrac{1}{4})^3 + 2(\tfrac{1}{4})^2]$	M1√		With 2 terms from previous line
		$= {}^9/_{256}$			A1 **4**
	(iii)	$\mathrm{H}(t) = \left(\dfrac{1+2t}{4-t}\right)^3$	B1		
		$\mathrm{H}'(t) = 3\left(\dfrac{1+2t}{4-t}\right)^2\left[\dfrac{2(4-t)+1+2t}{(4-t)^2}\right]$	M1A1		
		E(Y) = H′(1)	M1		
		=3	A1	**5**	
	(iv)	$\mathrm{H}(1)= p_0 + p_1 + p_2 + p_3 + p_4 + \ldots = 1$			
		$\mathrm{H}(-1) = p_0 - p_1 + p_2 - p_3 + p_4 - \ldots = -{}^1/_{125}$			
		Add: $2(p_0+p_2+p_4 +\ldots)= 1 - {}^1/_{125}$	M1		With sufficient detail
		${}^1/_2(1 - {}^1/_{125})$ AG	A1	**2 (13)**	