OCR
RECOGNISING ACHIEVEMENT

Monday 25 June 2012 – Afternoon

A2 GCE MATHEMATICS

4735 Probability & Statistics 4

QUESTION PAPER

Candidates answer on the Printed Answer Book.

OCR supplied materials:
- Printed Answer Book 4735
- List of Formulae (MF1)

Other materials required:
- Scientific or graphical calculator

Duration: 1 hour 30 minutes

4732650612

INSTRUCTIONS TO CANDIDATES

These instructions are the same on the Printed Answer Book and the Question Paper.

- The Question Paper will be found in the centre of the Printed Answer Book.
- Write your name, centre number and candidate number in the spaces provided on the Printed Answer Book. Please write clearly and in capital letters.
- **Write your answer to each question in the space provided in the Printed Answer Book.** Additional paper may be used if necessary but you must clearly show your candidate number, centre number and question number(s).
- Use black ink. HB pencil may be used for graphs and diagrams only.
- Answer **all** the questions.
- Read each question carefully. Make sure you know what you have to do before starting your answer.
- Do **not** write in the bar codes.
- You are permitted to use a scientific or graphical calculator in this paper.
- 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.

INFORMATION FOR CANDIDATES

This information is the same on the Printed Answer Book and the Question Paper.

- The number of marks is given in brackets **[ ]** at the end of each question or part question on the Question Paper.
- **You are reminded of the need for clear presentation in your answers.**
- The total number of marks for this paper is **72**.
- The Printed Answer Book consists of **12** pages. The Question Paper consists of **4** pages. Any blank pages are indicated.

INSTRUCTION TO EXAMS OFFICER / INVIGILATOR

- Do not send this Question Paper for marking; it should be retained in the centre or recycled. Please contact OCR Copyright should you wish to re-use this document.

1 Independent random variables X and Y have distributions $B(7, p)$ and $B(8, p)$ respectively.

(i) Explain why $X+Y \sim B(15, p)$. **[1]**

(ii) Find $P(X=2 \mid X+Y=5)$. **[4]**

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

$$f(x) = \begin{cases} 4xe^{-2x} & x \geqslant 0, \\ 0 & \text{otherwise.} \end{cases}$$

(i) Show that the moment generating function (mgf) of X is

$$\frac{4}{(2-t)^2}, \text{ where } |t| < 2.$$ **[4]**

(ii) Explain why the mgf of $-X$ is $\frac{4}{(2+t)^2}$. **[1]**

(iii) Two random observations of X are denoted by X_1 and X_2. What is the mgf of $X_1 - X_2$? **[1]**

3 Because of the large number of students enrolled for a university geography course and the limited accommodation in the lecture theatre, the department provides a filmed lecture. Students are randomly assigned to two groups, one to attend the lecture theatre and the other the film. At the end of term the two groups are given the same examination. The geography professor wishes to test whether there is a difference in the performance of the two groups and selects the marks of two random samples of students, 6 from the group attending the lecture theatre and 7 from the group attending the films. The marks for the two samples, ordered for convenience, are shown below.

Lecture theatre:	30	36	48	51	59	62	
Filmed lecture:	40	49	52	56	63	64	68

(i) Stating a necessary assumption, carry out a suitable non-parametric test, at the 10% significance level, for a difference between the median marks of the two groups. **[7]**

(ii) State conditions under which a two-sample t-test could have been used. **[1]**

(iii) Assuming that the tests in parts **(i)** and **(ii)** are both valid, state with a reason which test would be preferable. **[1]**

4 The random variable U has the distribution $\text{Geo}(p)$.

(i) Show, from the definition, that the probability generating function (pgf) of U is given by

$$G_U(t) = \frac{pt}{1-qt}, \text{ for } |t| < \frac{1}{q},$$

where $q = 1 - p$. **[3]**

(ii) Explain why the condition $|t| < \frac{1}{q}$ is necessary. **[1]**

(iii) Use the pgf to obtain $\text{E}(U)$. **[3]**

Each packet of Corn Crisp cereal contains a voucher and 20% of the vouchers have a gold star. When 4 gold stars have been collected a gift can be claimed. Let X denote the number of packets bought by a family up to and including the one from which the 4^{th} gold star is obtained.

(iv) Obtain the pgf of X. **[2]**

(v) Find $\text{P}(X = 6)$. **[3]**

5 A one-tail sign test of a population median is to be carried out at the 5% significance level using a sample of size n.

(i) Show by calculation that the test can never result in rejection of the null hypothesis when $n = 4$. **[2]**

The coach of a college swimming team expects Elena, the best 50 m freestyle swimmer, to have a median time less than 30 seconds. Elena found from records of her previous 72 swims that 44 were less than 30 seconds and 28 were greater than 30 seconds.

(ii) Stating a necessary assumption, test at the 5% significance level whether Elena's median time for the 50 m freestyle is less than 30 seconds. **[9]**

6 The random variables S and T are independent and have joint probability distribution given in the table.

		S = 0	S = 1	S = 2
T	1	a	0.18	b
	2	0.08	0.12	0.20

(i) Show that $a = 0.12$ and find the value of b. **[6]**

(ii) Find $\text{P}(T - S = 1)$. **[2]**

(iii) Find $\text{Var}(T - S)$. **[4]**

Questions 7 and 8 are printed overleaf.

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

$$f(x) = \begin{cases} \frac{1}{4}(1 + ax) & -2 \leqslant x \leqslant 2, \\ 0 & \text{otherwise,} \end{cases}$$

where a is a constant.

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

(ii) Find $E(X)$ in terms of a. **[2]**

(iii) Construct an unbiased estimator T_1 of a based on one observation X_1 of X. **[2]**

(iv) A second observation X_2 is taken. Show that T_2, where $T_2 = \frac{3}{8}(X_1+X_2)$, is also an unbiased estimator of a. **[2]**

(v) Given that $\text{Var}(X) = \sigma^2$, determine which of T_1 and T_2 is the better estimator. **[4]**

8 Events A and B are such that $P(A) = 0.3$ and $P(A \mid B) = 0.6$.

(i) Show that $P(B) \leqslant 0.5$. **[3]**

(ii) Given also that $P(A \cup B) = x$, find $P(B)$ in terms of x. **[2]**

Question			Answer	Marks	Guidance	
1		(i)	Successes/variables are indep with same p so B(7+8, p)	B1 **[1]**	or use pgf.	
		(ii)	P(X = 2|X + Y = 5) = P(X = 2, Y = 3)/P(X + Y =5) $=\binom{7}{2}p^2q^5\times\binom{8}{3}p^3q^5\ /\ \binom{15}{5}p^5q^{10}$ = 0.392= (1176/3003)=56/143	M1 B1 B1 A1 **[4]**	 Numerator Denominator	
2		(i)	$\int_0^\infty 4x\mathrm{e}^{-(2-t)x}\mathrm{d}x$ oe $=\left[\frac{-4}{2-t}x\mathrm{e}^{-(2-t)x}\right]_0^\infty+\frac{4}{2-t}\int_0^\infty \mathrm{e}^{-(2-t)x}\mathrm{d}x$ oe $=\frac{-4}{(2-t)^2}\left[\mathrm{e}^{-(2-t)x}\right]_0^\infty$ oe $=\frac{4}{(2-t)^2}$ AG	M1 M1 A1 A1 **[4]**	 Using integration by parts (Allow omission of limits for M1M1) Allow $\frac{4}{(t-2)^2}$	
		(ii)	Requires E(e^{-Xt}) = which is E(e^{Xt}) with $-t$ for t	B1 **[1]**	Or from mgfs	or from $\int_0^\infty 4xe^{-x(t+2)}$ dx Must be –x(t+2), not –xt-2x
		(iii)	$16/(4-t^2)^2$	B1 **[1]**	AEF, ISW	

Question			Answer	Marks	Guidance	
3		(i)	Populations have identical/same distributions (apart from location) (H_0: $m_1 = m_2$, H_1: $m_1 \neq m_2$) Ranks 1 2 4 6 9 10 3 5 7 8 11 12 13 $R_m = 32$, $m(m + n + 1) - R_m = 52$ $W = 32$ Critical value = 29 32 > 29, do not reject H There is insufficient evidence at the 10% significance level of a difference between the median marks of the two groups. oe.	B1 M1 A1 A1 B1 M1 A1 **[7]**	Allow 'Data quantitative' Can be implied. M1A0A1 possible Correct first conclusion ft TS and CV ft TS only.	Allow 'No assumption necessary' stated.
		(ii)	Marks should have normal populations with equal variances.	B1 **[1]**	Need 'population'.	NOT populations of students.
		(iii)	2- sample t-test would be better than the Wilcoxon test since it uses more information.	B1 **[1]**	Or is more powerful.	
4		(i)	$E(t^U) = pt + qpt^2 + q^2pt^3 + ….$ $= pt(1 + qt + q^2t^2 + … .$ $= pt/1 - qt)$ AG	M1 A1 A1 **[3]**	or a=pt, r=qt	or $(1\text{-qt})^{-1} = 1\text{+qt+}….$
		(ii)	For convergence of the infinite series	B1 **[1]**	Or G would be ≤ 0 (or probs or denom)	
		(iii)	$G'(t) = [p(1 - qt) + pqt]/(1 - qt)^2$ Mean = $G'(1)$ =…=$1/p$	M1 M1 A1 **[3]**	or product rule. CWO	
		(iv)	$G_U = 0.2t/(1 - 0.8t)$; $G_X = [0.2t/(1 - 0.8t)]^4$	B1 B1 **[2]**		
		(v)	Find the coefficient of t^6 in expansion of G_X $= 0.2^4 \times (4 \times 5/2) \times 0.8^2$ = 0.01024 =32/3125	M1 M1 A1 **[3]**	Or 3 in the first 5 (B(5, 0.2) and 1 in 6^{th}	=0.0512 x0.2

Question			Answer	Marks	Guidance	
5		(i)	For $n = 4$ P(X = 0) or $P(X = 4) = 2^{-4} = 0.0625$ 0.0625 > 0.05 so H_o cannot be rejected	M1 A1 **[2]**	or 0.9375<0.95	
		(ii)	Sample of times considered random $H_0{:}m = 30$, $H_1{:}m < 30$ Use sign test $X \sim B(72, ½)$ $P(X \le 28) =$ (from N(36,18) Φ (28.5or 43.5 – 36)/$18^{1/2}$ = 0.0385 or 0.0386 Or from B(72, ½) = 0.0382 Compare with 0.05 and reject H_0 There is sufficient evidence to accept that the median time for Elena's swims is less than 30s	B1 B1 M1 M1 M1 M1 A1 A2 M1 A1ft **[9]**	Allow 'data above or below median' Both hypotheses, median or m May be implied =(-)1.767 or CV=(-)1.645 Using calculator procedure or -1.767<-1.645 not over-assertive	 No , or wrong ,CC (27.5 or 44.5) -1.886 or -2.003 M0 Any other CC M0 0.0297 or 0.0227 A1 0.03818457 No, or wrong, CC M1A1ft
6		(i)	Use independence to obtain equation in a and/or b eg 0.4(a+0.08)=0.08, a=(a+b+0.18)(a+0.08) 0.18+2(b+0.12)+0.8=1.4(0.3+2b+0.4) Use independence or Σp=1 or P(T=1)=0.6 to obtain 2nd equation. eg a+0.58+b=1 or above Correct simplified equation eg 0.4a=0.048, a+b=0.42, 0.24=0.8b 2nd correct simplified equation a=0.12 AG b=0.3	M1 M1 A1 A1 A1 A1 **[6]**	P(A ∩ B)=P(A)P(B) or E(TS)=E(T)E(S)	P(S=0)=0.08/0.4 or P(S=2)=0.2/0.4 M1 P(S=0)=0.2 A1 P(S=2)=0.5 A1 (P(T=1)=0.6) a="0.6"x"0.2" or b="0.6"x"0.5" M1 a=0.12 AG A1, b=0.3 A1
		(ii)	$P(T = 2, S = 1) + P(T = 1, S = 0)$ = 0.12 + 0.12 = 0.24	M1 A1 **[2]**		

Question			Answer	Marks	Guidance	
6		(iii)	Var(T) + Var(S) Var(T)= 0.6 + 4 × 0.4 – (0.6 + 0.8)^2 Var(S) = 0.3 + 4 × 0.5 – (0.3 + 1)^2 Var($T – S$) = 0.85	M1 M1 M1 A1 **[4]**	$T – S$: –1 0 1 2 p : 0.3 0.38 0.24 0.08 (M1A1) E($T – S$) = 0.1 E([$T – S$)^2]) = 0.86 Var = 0.86 – 0.01 = 0.85 (M1A1)	(Var(T)=0.24 , Var(S)=0.61) Allow M1 for sum of vars, even if E(T²) and/or E(S²) only used.(2.2,2.3) Allow Var(T)+Var(S)-2Cov(T,S) provided Cov obtained from E(TS)-E(T)E(S) even if incorrectly.
7		(i)	f(– 2) = ¼ (1 – 2a) ≥ 0 =>a≤ ½ f(2) = ¼ (1 + 2a) ≥ 0 => a≥-1/2	M1 A1 **[2]**	Using f(x) ≥ 0 Allow omission of ¼	
		(ii)	$\int_{-2}^{2}\frac{1}{4}(x+\alpha x^2)\mathrm{d}x$ = [¼(x^2/2 + ax^3/3)] = 4a/3	M1 A1 **[2]**		
		(iii)	E(3X/4) = a => T_1 = 3X_1/4	M1 A1 **[2]**		
		(iv)	E(T_2) = $^3/_8$ (E(X_1) + E(X_2)). = $^3/_8$ ($^4/_3a$ + $^4/_3a$)=a, => T_2 unbiased for a	M1 A1 **[2]**		
		(v)	Var(T_1) = $^9/_{16}\sigma^2$ Var(T_2) = $^9/_{64}(\sigma^2+\sigma^2)$ = $^9/_{32}\sigma^2$ Var(T_2) < Var(T_1) => T_2 better	M1 A1 A1 M1 **[4]**	M1 for a²σ ² for either T.	

Question			Answer	Marks	Guidance	
8		(i)	$P(A \cap B)=0.6P(B)$ $P(A \cap B) \le P(A) = 0.3$ $P(B) \le 0.3/0.6 = 0.5$ AG	M1 M1 A1 **[3]**	May be implied.	$0.3P(B \mid A)=0.6P(B)$ M1 ,use $P(B \mid A)\le 1$ M1
		(ii)	$P(A \cup B) = x = 0.3 + P(B) - 0.6P(B)$ $P(B) = (x - 0.3)/0.4$	M1 A1 **[2]**	Use formulae for union and cond prob.	