RECOGNISING ACHIEVEMENT

ADVANCED GCE
MATHEMATICS
4735

Probability & Statistics 4

Candidates answer on the Answer Booklet

OCR Supplied Materials:
- 8 page Answer Booklet
- List of Formulae (MF1)

Other Materials Required:
None

Friday 19 June 2009
Afternoon

Duration: 1 hour 30 minutes

INSTRUCTIONS TO CANDIDATES

- Write your name clearly in capital letters, your Centre Number and Candidate Number in the spaces provided on the Answer Booklet.
- Use black ink. Pencil may be used for graphs and diagrams only.
- Read each question carefully and make sure that you know what you have to do before starting your answer.
- Answer **all** the questions.
- Do **not** write in the bar codes.
- 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.
- **You are reminded of the need for clear presentation in your answers.**
- The total number of marks for this paper is **72**.
- This document consists of **4** pages. Any blank pages are indicated.

1 X_1 and X_2 are independent random variables with distributions $\mathrm{N}(\mu_1, \sigma_1^2)$ and $\mathrm{N}(\mu_2, \sigma_2^2)$ respectively. Assuming that the moment generating function of a normal variable with mean μ and variance σ^2 is $\mathrm{e}^{\mu t+\frac{1}{2}\sigma^2 t^2}$, find the moment generating function of $X_1 + X_2$. Hence identify the distribution of $X_1 + X_2$, stating the value(s) of any parameter(s). **[5]**

2 A company wishes to buy a new lathe for making chair legs. Two models of lathe, 'Allegro' and 'Vivace', were trialled. The company asked 12 randomly selected employees to make a particular type of chair leg on each machine. The times, in seconds, for each employee are shown in the table.

Employee	1	2	3	4	5	6	7	8	9	10	11	12
Time on Allegro	162	111	194	159	202	210	183	168	165	150	185	160
Time on Vivace	182	130	193	181	192	205	186	184	192	180	178	189

The company wishes to test whether there is any difference in average times for the two machines.

(i) State the circumstances under which a non-parametric test should be used. **[1]**

(ii) Use two different non-parametric tests and show that they lead to different conclusions at the 5% significance level. **[9]**

(iii) State, with a reason, which conclusion is to be preferred. **[1]**

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

$$\mathrm{f}(x) = \begin{cases} \mathrm{e}^{2x} & x < 0, \\ \mathrm{e}^{-2x} & x \geqslant 0. \end{cases}$$

(i) Show that the moment generating function of X is $\dfrac{4}{4 - t^2}$, where $|t| < 2$, and explain why the condition $|t| < 2$ is necessary. **[5]**

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

4 The probability generating function of the discrete random variable Y is given by

$$\mathrm{G}_Y(t) = \frac{a + bt^3}{t},$$

where a and b are constants.

(i) Given that $\mathrm{E}(Y) = -0.7$, find the values of a and b. **[4]**

(ii) Find $\mathrm{Var}(Y)$. **[2]**

(iii) Find the probability that the sum of 10 random observations of Y is -7. **[4]**

5 Alana and Ben work for an estate agent. The joint probability distribution of the number of houses they sell in a randomly chosen week, X_A and X_B respectively, is shown in the table.

		X_A			
		0	1	2	3
X_B	0	0.02	0.13	0.07	0.03
	1	0.16	0.22	0.03	0.04
	2	0.09	0.06	0.03	0.02
	3	0.03	0.04	0.02	0.01

(i) Find $\mathrm{E}(X_A)$ and $\mathrm{Var}(X_A)$. **[3]**

(ii) Determine whether X_A and X_B are independent. **[2]**

(iii) Given that $\mathrm{E}(X_B) = 1.15$, $\mathrm{Var}(X_B) = 0.8275$ and $\mathrm{E}(X_A X_B) = 1.09$, find $\mathrm{Cov}(X_A, X_B)$ and $\mathrm{Var}(X_A - X_B)$. **[4]**

(iv) During a particular week only one house was sold by Alana and Ben. Find the probability that it was sold by Alana. **[4]**

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

$$\mathrm{f}(x) = \begin{cases} 0 & x < a, \\ \mathrm{e}^{-(x-a)} & x \geqslant a, \end{cases}$$

where a is a constant. $X_1, X_2, \ldots, X_n$ are n independent observations of X, where $n \geqslant 4$.

(i) Show that $\mathrm{E}(X) = a + 1$. **[3]**

T_1 and T_2 are proposed estimators of a, where

$$T_1 = X_1 + 2X_2 - X_3 - X_4 - 1 \qquad \text{and} \qquad T_2 = \frac{X_1 + X_2}{4} + \frac{X_3 + X_4 + \ldots + X_n}{2(n-2)} - 1.$$

(ii) Show that T_1 and T_2 are unbiased estimators of a. **[4]**

(iii) Determine which is the more efficient estimator. **[4]**

(iv) Suggest another unbiased estimator of a using all of the n observations. **[2]**

[Question 7 is printed overleaf.]

7 A particular disease occurs in a proportion p of the population of a town. A diagnostic test has been developed, in which a positive result indicates the presence of the disease. It has a probability 0.98 of giving a true positive result, i.e. of indicating the presence of the disease when it is actually present. The test will give a false positive result with probability 0.08 when the disease is not present. A randomly chosen person is given the test.

(i) Find, in terms of p, the probability that

(a) the person has the disease when the result is positive, **[3]**

(b) the test will lead to a wrong conclusion. **[2]**

It is decided that if the result of the test on someone is positive, that person is tested again. The result of the second test is independent of the result of the first test.

(ii) Find the probability that the person has the disease when the result of the second test is positive. **[2]**

(iii) The town has 24 000 children and plans to test all of them at a cost of £5 per test. Assuming that $p = 0.001$, calculate the expected total cost of carrying out these tests. **[4]**

4735 Probability & Statistics 4

1	$M_{X_1+X_2}(t) = (e^{\mu_1 t+\frac{1}{2}\sigma_1^2 t^2})(e^{\mu_2 t+\frac{1}{2}\sigma_2^2})$	M1	MGF of sum of independent RVs
	$= e^{(\mu_1+\mu_2)t+\frac{1}{2}(\sigma_1^2+\sigma_2^2)t^2}$ *oe*	A1	
	$X_1 + X_2 \sim$ Normal distribution	A1	
	with mean $\mu_1 + \mu_2$, variance $\sigma_1^2 + \sigma_2^2$	A1A1 **5** **{5}**	No suffices:- Allow M1A0A1A0A0
2 (i)	Non-parametric test used when the distribution of the variable in question is unknown	B1 **1**	
(ii)	H_0: $m_{V-A} = 0$, H_1: $m_{V-A} \neq 0$ where m_{V-A} is the median of the population differences	B1	Allow $m_V = m_A$ etc
	Difference and rank, bottom up	M1	
	$P = 65$ $Q = 13$	A1	Allow $P > Q$ stated
	$T = 13$	B1	
	Critical region: $T \leq 13$	M1	
	13 is inside the CR so reject H_0 and accept that there is sufficient evidence at the 5% significance level that the medians differ	A1	Penalise over-assertive conclusions once only.
	Use B(12, 0.5)	M1	
	P($\leq$ 4) = 0.1938 or CR = {0,1,2,10,11,12}	A1	
	> 0.025, accept that there is insufficient evidence, etc CWO	A1 **9**	Or 4 not in CR
(iii)	Wilcoxon test is more powerful than the sign test	B1 **1** **{11}**	Use more information, more likely to reject NH
3(i)	$A + B$		
	$= \int_{-\infty}^{0} e^{2x} e^{xt}\,dx + \int_{0}^{\infty} e^{-2x} e^{xt}\,dx$	M1	Added, correct limits
	$= \left[\frac{1}{2+t} e^{(2+t)x}\right]_{-\infty}^{0} + \left[-\frac{1}{2-t} e^{-(2-t)x}\right]_{0}^{\infty}$	B1 B1	Correct integrals
	$= 1/(2 + t) + 1/(2 - t)$		
	$= 4/(4 - t^2)$ AG	A1	Allow sensible comments about denom of M(t)
	$t < -2$, A infinite; $t > 2$, B infinite	B1 **5**	
(ii)	**Either**: $4/(4 - t^2) = (1 - \frac{1}{4}t^2)^{-1}$	M1	Expand
	$= 1 + \frac{1}{4}t^2 + \ldots$	A1	
	Or: $M'(t) = 8t/(4 - t^2)^2$		M1
	$M''(t) = 8/(4 - t^2)^2 + t \times \ldots$		A1
	E(X) = 0	M1	
	Var(X) = 2×¼ – 0 = ½	A1 **4** **{9}**	For $M''(0) - [M'(0)]^2$ or equivalent 0.5 – 0 = 0.5

4 (i)	G(1)=1 [$a+b=1$] G′(1) = – 0.7 [$–a + 2b = – 0.7$] Solve to obtain $\boldsymbol{a = 0.9}$ **,** $\boldsymbol{b = 0.1}$	M1 M1 M1 A1 **4**	
(ii)	$G''(t)$ [$=1.8/t^3 + 0.2$] and $G''(1) + G'(1) – [G'(1)^2]$ used Var = $2 – 0.7 – 0.7^2$ = **0.81**	M1 A1 **2**	
(iii)	$[(0.9 + 0.1t^3)/t]^{10}$ Method to obtain coefficient of t^{-7} $10\times0.9^9\times0.1$ = **0.387** to 3SF	M1 M1 A1 ft A1 **4** **{10}**	$[(a + bt^3)/t]^{10}$ For both Use of MGF. $10a^9b$
5 (i)	Marginal dist of X_A: 0.30 0.45 0.15 0.10 E = 0.45 + 0.3 + 0.3 = 1.05 Var = $0.45 + 0.6 + 0.9 – 1.05^2$ = **0.8475**	B1 B1 B1 **3**	
(ii)	Consider a particular case to show $P(X_A \text{ and } X_B) \neq P(X_A)P(X_B)$ So X_A and X_B are not independent	M1 A1 **2**	Or $E(X_A)$, $E(X_B)$ and $E(X_AX_B)$ 1.05, 1.15, 1.09; $E(X_A)E(X_B)$ = 1.0275, ft on wrong $E(X_A)$
(iii)	Cov = $E(X_AX_B) – E(X_A)E(X_B)$ = 1.09 – 1.15×1.05 = –0.1175 $\text{Var}(X_A – X_B) = \text{Var}(X_A) + \text{Var}(X_B) – 2\text{Cov}(X_A, X_B)$ = **1.91**	M1 A1ft M1 A1 **4**	Or from distribution of X_A - X_B Wrong $E(X_A)$
(iv)	Requires $P(X_A, X_B)/P(X_A+X_B = 1)$ = 0.13/(0.16 + 0.13) = 13/29 = **0.448**	M1 A1A1 A1 **4** **{13}**	

6 (i)	$\int_a^\infty x\mathrm{e}^{-(x-a)}\mathrm{d}x = \left[-x\mathrm{e}^{-(x-a)}\right]_a^\infty + \int_a^\infty \mathrm{e}^{-(x-a)}\mathrm{d}x$ $= a + [-\mathrm{e}^{-(x-a)}]$ $= a + 1$ AG	M1B1 A1 **3**	Correct limits needed for M1; no, or incorrect, limits allowed for B1
(ii)	$\mathrm{E}(T_1) = (a+1) + 2(a+1) - 2(a+1) - 1$ $= a$ $\mathrm{E}(T_2) = ¼(a+1+a+1) + (n-2)(a+1)/[2(n-2)] - 1$ $= a$ (So both are unbiased estimators of a)	M1 A1 M1 A1 **4**	
(iii)	$\sigma^2 = \mathrm{Var}(X)$ $\mathrm{Var}(T_1) = (1 + 4 + 1 + 1)\sigma^2 = 7\sigma^2$ $\mathrm{Var}(T_2) = 2\sigma^2/16 + (n-2)\sigma^2/[2(n-2)^2]$ $= n\sigma^2/[8(n-2)]$ oe This is clearly $< 7\sigma^2$, so T_2 is more efficient	M1 A1 B1 A1 **4**	
(iv)	eg $^1/_n(X_1 + X_2 + \ldots + X_n) - 1$	B2 **2** **{13}**	B1 for sample mean
7 (i)	D denotes "The person has the disease" (a) $\mathrm{P}(D) = p$, $\mathrm{P}(D') = 1 - p$, $\mathrm{P}(+\|D) = 0.98$, $\mathrm{P}(+\|D') = 0.08$ $\mathrm{P}(+) = p\times 0.98 + 0.08\times(1-p)$ $= 0.08 + 0.9p$ $\mathrm{P}(D\|+) = \mathrm{P}(+\|D)(\mathrm{P}(D)/\mathrm{P}(+)$ $= 0.98p/(0.08 + 0.9p)$ (b) $\mathrm{P}(D')\times\mathrm{P}(+\|D') + \mathrm{P}(D)\times\mathrm{P}(-\|D)$ $= 0.08 - 0.06p$	 M1 M1 A1 M1 A1 **5**	 Use conditional probability
(ii)	$\mathrm{P}(++) = 0.98^2\times p + 0.08^2\times(1-p)$ $\mathrm{P}(\mathrm{D}\|++) = 0.9604p/(0.954p + 0.0064)$	M1 A1 **2**	
(iii)	Expected number with 2 tests: $24000\times 0.0809 = a$ Expected number with 1 test: $24000 \times 0.9191 = b$ Expected total cost = £$(10a + 5b)$ = **£129 708**	 M1 M1 M1 A1 **4** **{11}**	 Or: $0.08 + 0.9\times 0.001$ oe $\times 5\times 24000$ $+5\times 24000$ (dep 1st M1) Or £130 000