ADVANCED GCE

MATHEMATICS

4735

Probability & Statistics 4

Candidates answer on the answer booklet.

OCR supplied materials:
- 8 page answer booklet (sent with general stationery)
- List of Formulae (MF1)

Other materials required:
- Scientific or graphical calculator

Thursday 23 June 2011
Morning

Duration: 1 hour 30 minutes

INSTRUCTIONS TO CANDIDATES

- Write your name, centre number and candidate number in the spaces provided on the answer booklet. Please write clearly and in capital letters.
- Use black ink. Pencil may be used for graphs and diagrams only.
- Read each question carefully. Make sure 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 scientific or 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.

Turn over

1 The random variable X has the distribution $\mathrm{B}(n, p)$.

(i) Show, from the definition, that the probability generating function of X is $(q + pt)^n$, where $q = 1 - p$. **[2]**

(ii) The independent random variable Y has the distribution $\mathrm{B}(2n, p)$ and $T = X + Y$. Use probability generating functions to determine the distribution of T, giving its parameters. **[4]**

2 A botanist believes that some species of plants produce more flowers at high altitudes than at low altitudes. In order to investigate this belief the botanist randomly samples 11 species of plants each of which occurs at both altitudes. The numbers of flowers on the plants are shown in the table.

Species	1	2	3	4	5	6	7	8	9	10	11
Number of flowers at low altitude	5	3	4	7	2	9	6	5	4	11	2
Number of flowers at high altitude	1	6	10	8	14	16	20	21	15	2	12

(i) Use the Wilcoxon signed rank test at the 5% significance level to test the botanist's belief. **[7]**

(ii) Explain why the Wilcoxon rank sum test should not be used for this test. **[1]**

3 For the events A and B, $\mathrm{P}(A) = \mathrm{P}(B) = \frac{3}{4}$ and $\mathrm{P}(A \mid B') = \frac{1}{2}$.

(i) Find $\mathrm{P}(A \cap B)$. **[4]**

For a third event C, $\mathrm{P}(C) = \frac{1}{4}$ and C is independent of the event $A \cap B$.

(ii) Find $\mathrm{P}(A \cap B \cap C)$. **[1]**

(iii) Given that $\mathrm{P}(C \mid A) = \lambda$ and $\mathrm{P}(B \mid C) = 3\lambda$, and that no event occurs outside $A \cup B \cup C$, find the value of λ. **[5]**

4 The discrete random variable X has moment generating function $\left(\frac{1}{4} + \frac{3}{4}\mathrm{e}^t\right)^3$.

(i) Find $\mathrm{E}(X)$. **[3]**

(ii) Find $\mathrm{P}(X = 2)$. **[3]**

(iii) Show that X can be expressed as a sum of 3 independent observations of a random variable Y. Obtain the probability distribution of Y, and the variance of Y. **[4]**

5 A test was carried out to compare the breaking strengths of two brands of elastic band, A and B, of the same size. Random samples of 6 were selected from each brand and the breaking strengths were measured. The results, in suitable units and arranged in ascending order for each brand, are as follows.

Brand A:	5.6	8.7	9.2	10.7	11.2	12.6
Brand B:	10.1	11.6	12.0	12.2	12.9	13.5

(i) Give one advantage that a non-parametric test might have over a parametric test in this context. **[1]**

(ii) Carry out a suitable Wilcoxon test at the 5% significance level of whether there is a difference between the average breaking strengths of the two brands. **[7]**

(iii) An extra elastic band of brand B was tested and found to have a breaking strength exceeding all of the other 12 bands. Determine whether this information alters the conclusion of your test. **[3]**

6 A City Council comprises 16 Labour members, 14 Conservative members and 6 members of Other parties. A sample of two members was chosen at random to represent the Council at an event. The number of Labour members and the number of Conservative members in this sample are denoted by L and C respectively. The joint probability distribution of L and C is given in the following table.

		L		
		0	1	2
C	0	$\frac{1}{42}$	$\frac{16}{105}$	$\frac{4}{21}$
	1	$\frac{2}{15}$	$\frac{16}{45}$	0
	2	$\frac{13}{90}$	0	0

(i) Verify the two non-zero probabilities in the table for which $C = 1$. **[4]**

(ii) Find the expected number of Conservatives in the sample. **[3]**

(iii) Find the expected number of Other members in the sample. **[3]**

(iv) Explain why L and C are not independent, and state what can be deduced about $\text{Cov}(L, C)$. **[3]**

[Question 7 is printed overleaf.]

7 The continuous random variable U has unknown mean μ and known variance σ^2. In order to estimate μ, two random samples, one of 4 observations of U and the other of 6 observations of U, are taken. The sample means are denoted by $\overline{U}_4$ and $\overline{U}_6$ respectively. One estimator S, given by $S = \frac{1}{2}(\overline{U}_4 + \overline{U}_6)$, is proposed.

(i) Show that S is unbiased and find $\text{Var}(S)$ in terms of σ^2. **[4]**

A second estimator T of the form $a\overline{U}_4 + b\overline{U}_6$ is proposed, where a and b are chosen such that T is an unbiased estimator for μ with the smallest possible variance.

(ii) Find the values of a and b and the corresponding variance of T. **[7]**

(iii) State, giving a reason, which of S and T is the better estimator. **[1]**

(iv) Compare the efficiencies of this preferred estimator and the mean of all 10 observations. **[2]**

1 (i)	$\sum_{x=0}^{n}\binom{n}{x}p^x q^{n-x}t^x$ $=\sum_{x=0}^{n}\binom{n}{x}(pt)^x q^{n-x}$	M1 A1 **2**	From E(t^X) M1A0 ∑ without limits $G_X(t)$=q+pt M1 then argument A0
(ii)	$G_T(t) = (q+pt)^n(q+pt)^{2n}$ $= (q+pt)^{3n}$ So $T \sim B(3n, p)$	M1 A1 M1 A1 **4** **[6]**	Multiplying pgfs For B For parameters
2 (i)	H_0: $m_d = 0$, H_1: $m_d > 0$, (where d = high – low) D: -4 3 6 1 12 7 14 16 11 -9 10 Rank -3 2 4 1 9 5 10 11 8 -6 7 $P = 57$, $Q = 9$ $T = 9$ CV = 13 9 < CV so reject H_0 There is sufficient evidence at the 5% significance level to support the botanist's belief	B1 M1 A1 B1 B1 M1 A1 ft **7**	Or H_o:m_H=m_L , etc. Medians Ranking top down, -9,-10,8, ..M1A0 T=15 B0 [SR last 3 marks: z=-2;09 B1 <-1.96 etc M1A1] Or equivalent ft T
(ii)	The rank sum test is for independent samples, the H and L values are correlated	B1 **1** **[8]**	Accept data paired
3 (i)	$P(A\|B') = P(A\cap B')/P(B')$ => $P(A\cap B')$ = 1/8 AEF Use $P(A\cap B) = P(A) - P(A\cap B')$ To give $P(A\cap B)$ = 5/8 AEF	M1 A1 M1 A1 **4**	 May be implied Or equivalent
(ii)	$P(A\cap B\cap C) = \frac{5}{8}\times\frac{1}{4}=\frac{5}{32}$ AEF	B1 √ **1**	Ft 5/8
(iii)	$P(B\cap C) = 3\lambda/4$ and $P(C\cap A) = 3\lambda/4$ Use formula for $P(A\cup B\cup C)$ And $P(A\cup B\cup C)$=1 Sub into formula for $P(A\cup B\cup C)$ and solve for λ giving λ = 3/16 AEF	M1 M1 B1 M1 A1 **5** **[10]**	For use of both conditional probs Allow one sign error
4 (i)	$M'(t) = 3(\frac{1}{4}+\frac{3}{4}e^t)^2\times\frac{3}{4}e^t$ $E(X) = M'(0) = \frac{9}{4}$	M1 A1 A1 **3**	Allow one error
(ii)	mgf $(\frac{1}{64}+\frac{9}{64}e^t+)\ \frac{27}{64}e^{2t}\ (+\frac{27}{64}e^{3t})$ $P(X = 2)$ = coefficient of $e^{2t} = \frac{27}{64}$	M1 A1 A1 **3**	Or PGF=$(\frac{1}{4}+\frac{3}{4}z)^3$ expand find coefficient of z^2 27/64
(iii)	Sum of 3 obs of Y with mgf ¼ + ¾e^t has mgf of X y : 0 1 p: ¼ ¾ Var(Y) = ¾ - (¾)^2 = $\frac{3}{16}$	M1*dep A1 *M1A1 **4** **[10]**	 OR $B(1,\frac{3}{4})$ Using $E(Y^2) - (E(Y))^2$ OR $1\times\frac{3}{4}\times\frac{1}{4}$ M0 if integration used

5(i)	Does not require a known probability distribution	B1 **1**	Or equivalent
(ii)	H_0: $m_A = m_B$, H_1: $m_A \neq m_B$ Ranks: A 1 2 3 5 6 10 B 4 7 8 9 11 12 R_A = 27, 78 – 27 = 51, so W = 27 OR: R_B = 51, 78 – 51 = 27 5% CV = 26 27 > CV so do not reject H_0 there is insufficient evidence at the 5% SL to indicate a difference in breaking strengths	B1 M1 M1 A1 B1 M1 A1 **7**	Medians Use N(39,39) with cc B1 P(W≤27.5), Z=-1.84 or equivalent M1 Not in CR etc A1
(iii)	B would have an extra rank 13 W still 27 but CV now 27 H_0 is now rejected	M1 B1 A1 **3** **[11]**	 P(W≤27.5)=-2.07 M1A1 In CR Ho rejected A1
6(i)	L=0, C = 1, choose 1C from 14 and 1 from 6 Others $14\times6 / {}^{36}C_2 = 2/15$ AG L = 1, C = 1, choose 1 from 16, 1 from 14 $16\times14 / {}^{36}C_2 = 16/45$ AG	M1 A1 M1 A1 **4**	Or $\frac{14}{36}\times\frac{6}{35}\times2$ Or $\frac{14}{36}\times\frac{16}{35}\times2$
(ii)	Marginal C probs: 11/30 22/45 13/90 E(C) = 22/45 + 26/90 = 35/45 = 7/9	B1 M1 A1 **3**	 AEF
(iii)	EITHER: 2×1/42 + 2/15 + 16/105 OR: E(L) = 8/9, E(O) = 2 – 15/9 = 1/3	M1 A1 A1 **3**	Other: 0 1 2 M1 p: $\frac{29}{42}$ $\frac{2}{7}$ $\frac{1}{42}$ A1 E(O)=$\frac{2}{7}+\frac{2}{42}=\frac{1}{3}$ A1
(iv)	EITHER: Argument OR: Use idea that for independence $P(L\cap C)= P(L)P(C)$ Conclude that covariance is non-zero	B2 M1A1 B1 **3** **[13]**	e.g The more Ls the fewer Cs OR Use conditional probability OR Cov(L,C)=-136/405 M1A1 L,C not indep B1
7(i)	$E(S) = \frac{1}{2}(E(\bar{U}_4) + E(\bar{U}_6))$ $= \frac{1}{2}(\mu + \mu) = \mu$, so S is unbiased $Var(S) = \frac{1}{4}(\sigma^2/4 + \sigma^2/6)$ $= 5\sigma^2/48$	M1 A1 M1 A1 **4**	 With conclusion
(ii)	$E(T) = (a+b)\mu = \mu$, $a+b=1$ $Var(T) = a^2\sigma^2/4 + b^2\sigma^2/6$ Minimise $y = a^2/4 + b^2/6 = a^2/4 + (1-a)^2/6$ EITHER by differentiation OR, completing square, OR from a sketch graph. Giving a = 2/5, b = 3/5 Justify minimum value Variance = $\sigma^2/10$	M1 B1 M1 M1 A1 B1 A1 **7**	 Allow from completion of square
(iii)	T is better since (both are unbiased and) $Var(T)<Var(S)$	B1 **1**	From calculated variances
(iv)	Sample mean of 10 observations (is also unbiased) with $\sigma^2/10$ They have the same efficiency	M1 A1 **2** **[14]**	Or show that T = mean of 10 observations