OCR
RECOGNISING ACHIEVEMENT

ADVANCED GCE

MATHEMATICS 4734

Probability & Statistics 3

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

4734

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.

1 The random variables X and Y are independent with $X \sim \mathrm{Po}(5)$ and $Y \sim \mathrm{Po}(4)$. S denotes the sum of 2 observations of X and 3 observations of Y.

(i) Find $\mathrm{E}(S)$ and $\mathrm{Var}(S)$. **[2]**

(ii) The random variable T is defined by $\frac{1}{2}X - \frac{1}{4}Y$. Show that $\mathrm{E}(T) = \mathrm{Var}(T)$. **[3]**

(iii) State which of S and T (if either) does not have a Poisson distribution, giving a reason for your answer. **[2]**

2 The population proportion of all men with red-green colour blindness is denoted by p. Each of a random sample of 80 men was tested and it was found that 6 had red-green colour blindness.

(i) Calculate an approximate 95% confidence interval for p. **[4]**

(ii) For a different random sample of men, the proportion with red-green colour blindness is denoted by p_s. Estimate the sample size required in order that $|p_s - p| \leqslant 0.05$ with probability 95%. **[3]**

(iii) Give one reason why the calculated sample size is an estimate. **[1]**

3 The monthly demand for a product, X thousand units, is modelled by the random variable X with probability density function given by

$$\mathrm{f}(x) = \begin{cases} ax & 0 \leqslant x \leqslant 1, \\ a(x-2)^2 & 1 < x \leqslant 2, \\ 0 & \text{otherwise,} \end{cases}$$

where a is a positive constant. Find

(i) the value of a, **[4]**

(ii) the probability that the monthly demand is at most 1500 units, **[2]**

(iii) the expected monthly demand. **[3]**

4 An experiment by Lord Rutherford at Cambridge in 1909 involved measuring the numbers of α-particles emitted during radioactive decay. The following table shows emissions during 2608 intervals of 7.5 seconds.

Number of particles emitted, x	0	1	2	3	4	5	6	7	8	9	10	$\geqslant 11$
Frequency	57	203	383	525	532	408	273	139	45	27	10	6

It is given that the mean number of particles emitted per interval, calculated from the data, is 3.87, correct to 3 significant figures.

(i) Find the contribution to the χ^2 value of the frequency of 273 corresponding to $x = 6$ in a goodness of fit test for a Poisson distribution. **[4]**

(ii) Given that no cells need to be combined, state why the number of degrees of freedom is 10. **[1]**

(iii) Given also that the calculated value of χ^2 is 13.0, correct to 3 significant figures, carry out the test at the 10% significance level. **[4]**

5 The continuous random variable X has (cumulative) distribution function given by

$$F(x) = \begin{cases} 0 & x < 1, \\ \frac{4}{3}\left(1 - \frac{1}{x^2}\right) & 1 \leqslant x \leqslant 2, \\ 1 & x > 2. \end{cases}$$

(i) Find the median value of X. **[2]**

(ii) Find the (cumulative) distribution function of Y, where $Y = \frac{1}{X^2}$, and hence find the probability density function of Y. **[6]**

(iii) Evaluate $E\left(2 - \frac{2}{X^2}\right)$. **[3]**

6 The Body Mass Index (BMI) of each of a random sample of 100 army recruits from a large intake in 2008 was measured. The results are summarised by

$$\Sigma x = 2605.0, \qquad \Sigma x^2 = 68\,636.41.$$

It may be assumed that BMI has a normal distribution.

(i) Find a 98% confidence interval for the mean BMI of all recruits in 2008. **[5]**

(ii) Estimate the percentage of the intake with a BMI greater than 30.0. **[3]**

(iii) The BMIs of two randomly chosen recruits are denoted by B_1 and B_2. Estimate $P(B_1 - B_2 < 5)$. **[4]**

(iv) State, giving a reason, for which of the above calculations the normality assumption is unnecessary. **[1]**

7 In order to improve their mathematics results 10 students attended an intensive Summer School course. Each student took a test at the start of the course and a similar test at the end of the course. The table shows the scores achieved in each test.

Student	1	2	3	4	5	6	7	8	9	10
First test score	37	27	38	47	54	27	52	39	62	23
Second test score	47	29	50	44	72	37	63	45	76	32

It is desired to test whether there has been an increase in the population mean score.

(i) Explain why a two-sample t-test would not be appropriate. **[1]**

(ii) Stating any necessary assumptions, carry out a suitable t-test at the $\frac{1}{2}$% significance level. **[10]**

(iii) The Summer School director claims that after taking the course the population mean score increases by more than 5. Is there sufficient evidence for this claim? **[4]**

1 (i)	E(S)= 22 Var(S)=E(S)	B1 B1 **2**	
(ii)	E(T) = ½×5 - ¼×4=1.5 Var(T)=¼×5 + $^1/_{16}$×4 = 1.5 =E(T) AG	B1 M1 A1 **3**	Using Var(aX+bY) **CWO**
(iii)	T only does not have a Poisson distribution Some values of T are EITHER negative OR: fractional	B1 B1 **2** **(7)**	Unless wrong reason
2(i)	Use $(^6/_{80})(^{74}/_{80})/80$ $p_s \pm zs$ z= 1.96 (0.0173, 0.1327)	B1 M1 B1 A1 **4**	Or /79 s of the form $\sqrt{(p_s q_s/80)}$ (or 79) or no √ Accept (0.017,0.133)
(ii)	Use $z\sqrt{(p_s q_s/n)}$ ≤ 0.05 $n \ge 106.6$, least is 107	M1 A1 A1 **3**	or no √ and z=1.96 .Or = Allow 110
(iii)	e.g Variance is an estimate OR Distribution of p_s is only approx normal	 B1 **1** **(8)**	Not var unknown Must state distribution of what.
3(i)	$\int_0^1 ax\,dx + \int_1^2 a(x-2)^2\,dx = 1$ $\left[\frac{ax^2}{2}\right]_0^1 + \left[\frac{a(x-2)^3}{3}\right]_1^2$ ½ a + ⅓a = 1 $a = {}^6/_5$	M1 B1 M1 A1 **4**	 With or without limits Correct method for equation with fractions/decimals
(ii)	EITHER: $\int_0^1 ax\,dx + \int_1^{1.5} a(x-2)^2\,dx$ OR $1-\int_{1.5}^2 a(x-2)^2\,dx$ $= {}^{19}/_{20}$	M1 A1 **2**	Any a AEF
(iii)	$\int_0^1 ax^2\,dx + \int_1^2 ax(x-2)^2\,dx$ $=\left[\frac{ax^3}{3}\right]_0^1 + \left[a(\frac{x^4}{4} - \frac{4x^3}{3} + 2x^2\right]_1^2$ =9/10 (Expected monthly demand = 900)	M1 B1 A1 **3** **(9)**	 AEF With or without limits AEF

4(i)	$2608p$ $p = e^{-3.87}3.87^6/6!$ (×2608=253.82) $(273-253.82)^2/253.82$ =1.449	M1 A1 M1 A1 **4**	p from Poisson From 253.8 or 254 seen Answer between 1.445 and 1.460
(ii)	Number of cells – 1 (estimated mean) – 1(same totals)	B1 **1**	Not 11-1
(iii)	H_0: A Poisson distribution fits the data H_1: A Poisson distribution does not fit the data CV = 15. 99 13.0 < CV and do not reject Ho accept that there is insufficient evidence that a Poisson distribution does not fit data	 B1 B1 M1 A1 **4** **(9)**	For both hypotheses Their CV Sufficient evidence that Poisson distribution fits data , OK
5(i)	Solve $\frac{4}{3}(1-\frac{1}{m^2}) = \frac{1}{2}$ Giving $m = \sqrt{\frac{8}{5}}$	M1 A1 **2**	 Or equivalent. 1.26, 1.265, 2√10/5
(ii)	$G(y) = P(Y \le y)$ or < $= P(X \ge 1/\sqrt{y})$ $= 1 - F(1/\sqrt{y})$ $= 1 - {}^4/_3(1-y)$ or $(4y-1)/3$ $1 \le 1/\sqrt{y} \le 2 \Rightarrow ¼ \le y \le 1$ $g(y) = \begin{cases} 4/3 & 1/4 \le y \le 1, \\ 0 & \text{otherwise.} \end{cases}$	M1 A1 M1 A1 B1 B1√ **6**	Or:$x=1/\sqrt{y}$, $\|dx/dy\|=1/(2y^{3/2})$ B1 $f(x)=8/(3x^3)$; $1\le x\le 2$ M1A1 $g(y)=f(x)\|dx/dy\|$ M1 =4/3 A1 $¼\le y\le 1$ B1 Ft $G(y)$
(iii)	EITHER: $E(2-2Y)$ $= 2 - 2\times{}^5/_8$ $= {}^3/_4$ OR $2 - \int_1^2 16/(3x^5)dx$ OR $\int_1^2 (2-2/x^2))(8/3x^3)dx$ $= 2 + [4/(3x^4)]$ $=[-8/(3x^2) + 4/(3x^4)]$ $= 3/4$ $=3/4$	M1 A1√ A1 M1 A1 A1 **3** **(11)**	 √ $g(y)$ CAO AEF From 2 - $\int xF'(x)dx$ √ $f(x)$ CAO AEF
6(i)	$s^2 = (68636.41 - 2605^2/100)/99$ (=7.84) $\bar{x} = 26.05$ $26.05 \pm zs/10$ $z = 2.326$ or $\Phi^{-1}(0.99)$ ART (25.4, 26.7)	B1 B1 M1 B1 A1 **5**	AEF Allow t(99)=2.365
(ii)	Use N(26.05,7.84) $P(\ge 30) = 1-\Phi([30-26.05]/\sqrt{7.84})$ = 0.0792 = 7.92%	M1 M1 A1 **3**	s^2 from (i) M0 for 7.84/100 No "cc" allow either; ART 0.08 or 8%
(iii)	Use $B_1 - B_2 \sim N(0, 15.68)$ $P(< 5) = \Phi(5/\sigma)$ = 0.897	M1 A1 A1 A1 **4**	With $\mu = 0$ For variance σ^2 Their σ; $\Phi(\pm 5/\sigma) \Rightarrow$ M1
(iv)	(i) only since sample size of 100 is large enough (for CLT to hold)	B1 **1** **(13)**	Must be clear which part and with correct reason.

7(i)	For each student the scores are correlated	B1 **1**	Or equivalent**,** eg paired
(ii)	Increase in score has a normal distribution Sample is considered to be a random sample of all students attending the course H_0: μ_D = 0, H_1: μ_D > 0 where *D*= increase in scores D= 10 2 12 -3 18 10 11 6 14 9 $\bar{D}$ = 8.9 s^2 = 35.88 Test statistic = 8.9/(s/√10) = 4.699 v = 9, CV = 3.25 4.699 > CV Reject H_0 and accept that there is sufficient evidence a at the ½ % significance level of an increase in mean scores. SR 2-sample test: (i)B0(ii)B0B1B1M0 Max 2/11	B1 B1 B1 M1 B1 B1 M1 A1 B1 M1 **10**	Allow pop of differences~ normal Or equivalent , allow independent Or H_0: μ_1=μ_2 H_1: $\mu_1 < \mu_2$: not μ =0 D may be implied Must involve 10 Allow ART 4.70 Or P(t>4.699)=0.00056<0.005 Not OA
(iii)	Test statistic = (8.9-5)/√3.588=2.059 This is significant of an increase at the 5% significance level (CV of 1.833) so director's claim is supported. SR 2-sample t-test. (8.9-5)/s M1 Max1/4 SR: Use of confidence intervals 99% C I 2-sided (2.74,15.1) : 99.5% 1-sided (2.74, ∞) 5 is in this interval so not significant at ½ % level A1 OR 90% CI 2-sided (5.43,12.37) ; 95% 1-sided (5.43, ∞) 5 not in this interval so significant at 5% SL	M1A1 M1 A1 **4** **(15)** M1A1 M1 A1 M1A1 M1 A1	 Or P(t>2.059)=0.035 Any reasonable significance level with corresponding conclusion Allow at ½ %