OCR
Oxford Cambridge and RSA

Friday 12 June 2015 – Morning

A2 GCE MATHEMATICS

4735/01 Probability & Statistics 4

QUESTION PAPER

Candidates answer on the Printed Answer Book.

OCR supplied materials:
- Printed Answer Book 4735/01
- List of Formulae (MF1)

Other materials required:
- Scientific or graphical calculator

Duration: 1 hour 30 minutes

INSTRUCTIONS TO CANDIDATES

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

- The Question Paper will be inside 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 For the events A and B it is given that

$$P(A) = 0.6,\ P(B) = 0.3 \text{ and } P(A \text{ or } B \text{ but not both}) = 0.4.$$

(i) Find $P(A \cap B)$. **[3]**

(ii) Find $P(A' \cap B)$. **[1]**

(iii) State, giving a reason, whether A and B are independent. **[1]**

2 The manufacturer of a painkiller, designed to relieve headaches, claims that people taking the painkiller feel relief in at most 30 minutes, on average. A random sample of eight users of the painkiller recorded the times it took for them to feel relief from their headaches. These times, in minutes, were as follows:

33 39 29 35 40 32 26 37

Use a Wilcoxon single-sample signed-rank test at the 5% significance level to test the manufacturer's claim, stating a necessary assumption. **[8]**

3 The manufacturer of electronic components uses the following process to test the proportion of defective items produced.

A random sample of 20 is taken from a large batch of components.

- If no defective item is found, the batch is accepted.
- If two or more defective items are found, the batch is rejected.
- If one defective item is found, a second random sample of 20 is taken. If two or more defective items are found in this second sample, the batch is rejected, otherwise the batch is accepted.

The proportion of defective items in the batch is denoted by p, and $q = 1 - p$.

(i) Show that the probability that a batch is accepted is $q^{20} + 20pq^{38}(q + 20p)$. **[3]**

For a particular component, $p = 0.01$.

(ii) Given that a batch is accepted, find the probability that it is accepted as a result of the first sample. **[3]**

4 The discrete random variable Y has probability generating function

$$G_Y(t) = 0.09t^2 + 0.24t^3 + 0.34t^4 + 0.24t^5 + 0.09t^6.$$

(i) Find the mean and variance of Y. **[5]**

Y is the sum of two independent observations of a random variable X.

(ii) Find the probability generating function of X, expressing your answer as a cubic polynomial in t. **[3]**

(iii) Write down the value of $P(X = 2)$. **[1]**

5 The random variable X has a Poisson distribution with mean λ. It is given that the moment generating function of X is $e^{\lambda(e^t - 1)}$.

(i) Use the moment generating function to verify that the mean of X is λ, and to show that the variance of X is also λ. **[7]**

(ii) Five independent observations of X are added to produce a new variable Y.
Find the moment generating function of Y, simplifying your answer. **[2]**

6 In a two-tail Wilcoxon rank-sum test, the sample sizes are 13 and 15. The sum of the ranks for the sample of size 13 is 135. Carry out the test at the 5% level of significance. **[9]**

7 The discrete random variable X can take the values 0, 1 and 2 with equal probabilities.
The random variables X_1 and X_2 are independent observations of X, and the random variables Y and Z are defined as follows:

Y is the smaller of X_1 and X_2, or their common value if they are equal; $Z = |X_1 - X_2|$.

(i) Draw up a table giving the joint distribution of Y and Z. **[5]**

(ii) Find $P(Y = 0 \mid Z = 0)$. **[2]**

(iii) Find $\text{Cov}(Y, Z)$. **[7]**

8 The independent random variables X_1 and X_2 have the distributions $\text{B}(n_1, \theta)$ and $\text{B}(n_2, \theta)$ respectively.

Two possible estimators for θ are

$$T_1 = \frac{1}{2}\left(\frac{X_1}{n_1} + \frac{X_2}{n_2}\right) \text{ and } T_2 = \frac{X_1 + X_2}{n_1 + n_2}.$$

(i) Show that T_1 and T_2 are both unbiased estimators, and calculate their variances. **[8]**

(ii) Find $\dfrac{\text{Var}(T_1)}{\text{Var}(T_2)}$. Given that $n_1 \neq n_2$, use the inequality $(n_1 - n_2)^2 > 0$ to find which of T_1 and T_2 is the more efficient estimator. **[4]**

END OF QUESTION PAPER

Question			Answer	Marks	Guidance	
1	(i)		Let P(A∩B)=x , 0.6–x+0.3–x=0.4 x=0.25	M1A1 A1 **[3]**	M1 for attempt to set up equation in x.	x must appear more than once.
	(ii)		0.05	B1 ft **[1]**	0.3-(i). Ans must be≥0.	
	(iii)		No, 0.6×0.3≠0.25	B1 ft **[1]**	Must have an answer to (i)	P(B|A')=0.05÷0.4=0.125≠P(B)
2			H0:m≤(or=)30, H1:m>30 Diffs 3,9,–1,5,10,2, –4,7 Ranks 3,7,1,5,8,2,4,6 Signed ranks 5(–), 31(+) TS=5 5 in CR, reject H0 There is sufficient evidence that the median time for relief is more than 30 mins. Distribution is symmetrical.	B1 M1A1 A1 A1ft M1ft A1 B1 **[8]**	If in words, 'population' needed. M1 for attempting differences AND ranks. Ft TS, not CV CV=5 Cwo, in context, not over-assertive.	
3	(i)		P(batch accepted)= q^{20} …+ $20q^{19}p(q^{20}+20q^{19}p)$ $q^{20}+20q^{38}p(q+20p)$ AG	B1 M1 A1 **[3]**	M1 for attempt at P(1)[P(0)+P(1)]	Allow 20C1
	(ii)		$\frac{0.99^{20}}{0.99^{20}+20\times0.99^{39}\times0.01+400\times0.99^{38}\times0.01^{2}}$ oe 0.834	M1A1 A1 [3]	0.99^{20} / (i) with q=0.99, p=0.01	
4	(i)		G′Y(t)= $0.18t+0.72t^2+1.36t^3+1.2t^4+0.54t^5$ Sub t=1,(0.18+0.72+1.36+1.2+0.54)	 M1	≥3 terms correct. M1 for diffn and sub t=1	y 2 3 4 5 6 p 0.09 0.24 0.34 0.24 0.09 B1 E(Y)=0.18+….=4 B1 E(Y2)=0.36+…..=17.2 M1 Var (Y)=17.2-42=1.2 M1A1

Question			Answer	Marks	Guidance	
			4	A1		
			$G''_Y(t)=0.18+1.44t+4.08t^2+4.8t^3+2.7t^4$	M1	≥3 terms correct ft	
			Sub t=1 and use correct formula for Var (0.18+1.44+4.08+4.8+2.7+"4"-"16")	M1		
			1.2	A1 **[5]**		
	(ii)		Attempt to factorise into $(0.3t+\ldots)(0.3t+\ldots)$	M1M1	Attempt to find $\sqrt{(G_Y(t))}$ seen or implied.	Or $0.3t^3$ twice.
			$0.3t+0.4t^2+0.3t^3$	A1 **[3]**		
	(iii)		0.4	B1ft [1]	Allow non-numerical answer, but coeff of t2 is not enough.	
5	(i)		Attempt differentiation using chain rule, oe.	M1		
			$\lambda e^{\lambda(e^t-1)}e^t$	A1		
			Sub t=0 to correctly obtain λ. AG	A1		
			Attempt differentiation of M', using ch. rule, oe.	M1		
			$\lambda e^{\lambda e^t+t-\lambda}(\lambda e^t+1)$	A1	$\lambda^2e^{2t}e^{\lambda(e^t-1)}+\lambda e^te^{\lambda(e^t-1)}$	from product rule.
			Sub t=0, and use correct formula for Var	M1		
			$\lambda^2+\lambda-\lambda^2=\lambda$ AG	A1 **[7]**		
	(ii)		Attempt to raise MX(t) to power 5.	M1		
			$e^{5\lambda(e^t-1)}$	A1 [2]		
6			H0:The samples are drawn from identical popns. H1:The samples are from different popns.	B1	Allow $m_1=m_2$; $m_1\neq m_2$	Critical region method . First B1B1B1as main scheme $\frac{x+0.5-188.5}{\sqrt{471.25}}$ = or < -1.96 M1A1B1 x<146 A1 135 is in CR, rej H0 M1 Conclusion A1
			Mean=188.5	B1	Allow 13x29/2	
			Var=471.25	B1	Allow 13x15x29/12	

<table>
<tr><th colspan="3">Question</th><th>Answer</th><th>Marks</th><th colspan="2">Guidance</th></tr>
<tr><td></td><td></td><td></td><td>$\frac{135+0.5-"188.5"}{\sqrt{471.25}}$
-2.44
CV=-1.96
TS<CV, reject H_0
Sufficient evidence that the samples were drawn from different populations.</td><td>M1A1 ft
A1
B1
M1
A1
[9]</td><td>Allow M1A0 for missing or incorrect c.c.
Allow -2.46 no c.c, -2.49 wrong c.c.

Ft both TS,CV
Not over-assertive. Cwo, allow from(-2.46 or -2.49</td><td>0.0073 (or 0.0069 or 0.0064) B1
pft< 2.5% (allow 5% for M1), rej H_0</td></tr>
<tr><td>7</td><td>(i)</td><td></td><td><table><tr><td>Y\Z</td><td>0</td><td>1</td><td>2</td></tr><tr><td>0</td><td>1/9</td><td>2/9</td><td>2/9</td></tr><tr><td>1</td><td>1/9</td><td>2/9</td><td>0</td></tr><tr><td>2</td><td>1/9</td><td>0</td><td>0</td></tr></table></td><td>M1M1
A3

[5]</td><td>M1 for attempt to allocate all 9 ordered pairs to correct cell or for correct Y,Z tables.
A1 3,4 or 5 correct; A2 6,7 or 8 correct
A3 all correct.</td><td>2nd M1 for attempt to evaluate entries in Y\Z table, even if <9 pairs. NOT from independence assumed.</td></tr>
<tr><td></td><td>(iii)</td><td></td><td>Attempt either Σyp or Σzp
E(Y)=5/9
E(Z)=8/9
<table><tr><td>YZ</td><td>0</td><td>1</td></tr><tr><td>p</td><td>7/9</td><td>2/9</td></tr><tr><td></td><td></td><td></td></tr></table>E(YZ)=2/9
“2/9”-“5/9”x”8/9”
$-\frac{22}{81}$</td><td>M1
A1
A1

M1A1
M1
A1
[7]</td><td>
ft
ft
No marks for table.</td><td></td></tr>
<tr><td></td><td>(ii)</td><td></td><td>P(0,0)/P(Z=0)
⅓</td><td>M1
A1
[2]</td><td>
CWO</td><td></td></tr>
<tr><td>8</td><td>(i)</td><td></td><td>$E(T_1)=E(X_1)/2n_1 + E(X_2)/2n_2$
AND $E(T_2)=\frac{E(X_1)+E(X_2)}{n_1+n_2}$
$E(X_1)=n_1\theta$ and $E(X_2)=n_2\theta$
$E(T_1)=E(T_2)=\theta$, validly shown</td><td>M1
B1
A1</td><td>Allow missed suffices for M marks only.

Allow</td><td></td></tr>
</table>

Question			Answer	Marks	Guidance	
			$Var(X_1)=n_1\theta(1-\theta)$ and $Var(X_2)=n_2\theta(1-\theta)$	B1		
			$Var(T_1)=\left(\frac{1}{2n_1}\right)^2 Var(X_1)+\left(\frac{1}{2n_2}\right)^2 Var(X_2)$	M1		
			$=\frac{\theta(1-\theta)}{4}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)$	A1		
			$Var(T_2)=\frac{1}{(n_1+n_2)^2}\left(Var(X_1)+Var(X_2)\right)$	M1		
			$\frac{\theta(1-\theta)}{n_1+n_2}$	A1		
				[8]		
	(ii)		$\frac{(n_1+n_2)^2}{4n_1n_2}$	B1		
			$(n_1-n_2)^2=(n_1+n_2)^2-4n_1n_2$	M1	Attempt to use $(n_1+n_2)^2$	
			$(n_1+n_2)^2>4n_1n_2$	A1		
			T_2 is more efficient.	A1		
				[4]		