OCR

Tuesday 10 June 2014 – Morning

A2 GCE MATHEMATICS

4733/01 Probability & Statistics 2

QUESTION PAPER

Candidates answer on the Printed Answer Book.

OCR supplied materials:
- Printed Answer Book 4733/01
- 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 found 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 The random variable F has the distribution B(50, 0.7). Use a suitable approximation to find P($F > 40$). **[5]**

2 The events organiser of a school sends out invitations to 150 people to attend its prize day. From past experience the organiser knows that the number of those who will come to the prize day can be modelled by the distribution B(150, 0.98).

(i) Explain why this distribution cannot be well approximated by either a normal or a Poisson distribution. **[3]**

(ii) By considering the number of those who do not attend, use a suitable approximation to find the probability that fewer than 146 people attend. **[4]**

3 The random variable G has the distribution $N(\mu, \sigma^2)$. One hundred observations of G are taken. The results are summarised in the following table.

Interval	$G < 40.0$	$40.0 \leqslant G < 60.0$	$G \geqslant 60.0$
Frequency	17	58	25

(i) By considering P($G < 40.0$), write down an equation involving μ and σ. **[2]**

(ii) Find a second equation involving μ and σ. Hence calculate values for μ and σ. **[4]**

(iii) Explain why your answers are only estimates. **[1]**

4 A zoologist investigates the number of snakes found in a given region of land. The zoologist intends to use a Poisson distribution to model the number of snakes.

(i) One condition for a Poisson distribution to be valid is that snakes must occur at constant average rate. State another condition needed for a Poisson distribution to be valid. **[1]**

Assume now that the number of snakes found in 1 acre of a region can be modelled by the distribution Po(4).

(ii) Find the probability that, in 1 acre of the region, at least 6 snakes are found. **[2]**

(iii) Find the probability that, in 0.77 acres of the region, the number of snakes found is either 2 or 3. **[4]**

5 A continuous random variable X has probability density function

$$f(x) = \begin{cases} \frac{1}{2}\pi \sin(\pi x) & 0 \leqslant x \leqslant 1, \\ 0 & \text{otherwise.} \end{cases}$$

(i) Show that this is a valid probability density function. **[4]**

(ii) Sketch the curve $y = f(x)$ and write down the value of $E(X)$. **[3]**

(iii) Find the value q such that $P(X > q) = 0.75$. **[3]**

(iv) Write down an expression, including an integral, for $\text{Var}(X)$. (Do not attempt to evaluate the integral.) **[2]**

(v) A student states that "X is more likely to occur when x is close to $E(X)$." Give an improved version of this statement. **[1]**

6 In a city the proportion of inhabitants from ethnic group Z is known to be 0.4. A sample of 12 employees of a large company in this city is obtained and it is found that 2 of them are from ethnic group Z. A test is carried out, at the 5% significance level, of whether the proportion of employees in this company from ethnic group Z is less than in the city as a whole.

(i) State an assumption that must be made about the sample for a significance test to be valid. **[1]**

(ii) Describe briefly an appropriate way of obtaining the sample. **[2]**

(iii) Carry out the test. **[7]**

(iv) A manager believes that the company discriminates against ethnic group Z. Explain whether carrying out the test at the 10% significance level would be more supportive or less supportive of the manager's belief. **[2]**

7 An examination board is developing a new syllabus and wants to know if the question papers are the right length. A random sample of 50 candidates was given a pre-test on a dummy paper. The times, t minutes, taken by these candidates to complete the paper can be summarised by

$$n = 50, \qquad \sum t = 4050, \qquad \sum t^2 = 329\,800.$$

Assume that times are normally distributed.

(i) Estimate the proportion of candidates that could not complete the paper within 90 minutes. **[6]**

(ii) Test, at the 10% significance level, whether the mean time for all candidates to complete this paper is 80 minutes. Use a two-tail test. **[7]**

(iii) Explain whether the assumption that times are normally distributed is necessary in answering

(a) part **(i)**,

(b) part **(ii)**. **[2]**

Question 8 begins on page 4.

8 The random variable W has the distribution Po(λ). A significance test is carried out of the null hypothesis H_0: $\lambda = 3.60$, against the alternative hypothesis H_1: $\lambda < 3.60$. The test is based on a single observation of W. The critical region is $W = 0$.

(i) Find the significance level of the test. **[2]**

(ii) It is known that, when $\lambda = \lambda_0$, the probability that the test results in a Type II error is 0.8. Find the value of λ_0. **[4]**

END OF QUESTION PAPER

Question		Answer/Indicative content	Marks	Guidance
1		N(35, 10.5) $1 - \Phi\left(\frac{40.5-35}{\sqrt{10.5}}\right) = 1 - \Phi(1.697)$ $= 1 - 0.9552$ $= \mathbf{0.0448}$	M1 A1 M1 A1 A1 **[5]**	Normal, mean 35 Both parameters correct, allow $\sqrt{10.5}$ or 10.5^2 Standardise, their *np*, *npq*, allow no $\sqrt{}$ or 10.5^2, allow wrong or no cc Both 40.5 and $\sqrt{npq}$ [Ans 0.0448 or 0.9552 can imply first 4 marks] Answer, a.r.t. 0.045. [Exact binomial (0.040232): 0/5]
2	(i)	 *np* = 147 > 5 so not Poisson *nq* = 3 < 5 so not normal	M1 A1 A1 **[3]**	Consider any two conditions, out of *np*, *nq* (allow *npq*), size of *n*, size of *p* 147 stated, or "*p* not small", no wrong conditions for Poisson seen 3 [not *just* 2.94] stated, or "*p* not close to ½", no wrong conditions for normal seen (apart from *npq*) *If spurious extra reasons seen ("not independent" etc), max 2/3*
2	(ii)	$A \sim B(150, 0.98)$ so $150 - A \sim B(150, 0.02)$ $\approx$ Po(3) $P(A < 146) = P(150 - A > 4) = 1 - 0.8153$ $= \mathbf{0.1847}$	M1 A1 M1 A1 **[4]**	Clearly consider complement, with $p = 0.02$ Po(3) stated or implied 1 – Po(3) probability, e.g. 0.3528 or 0.0839 0.185 or better [Exact binomial (0.1830): 0/4. N(3, 2.94): M1A0M0A0]
3	(i)	$\frac{\mu - 40}{\sigma} = 0.9544$	M1 B1 **[2]**	Standardise with μ and σ and equate to Φ^{-1}, allow σ^2 but not $\sqrt{n}$, allow 1–, cc, wrong signs. P(…): M0 here. But can recover both marks from part (ii). [0.954, 0.955] seen
3	(ii)	$\frac{60 - \mu}{\sigma} = 0.674(5)$ Solve to get $\sigma = 12.3$ [12.278] $\mu = 51.7(18)$	M1 B1 A1 A1 **[4]**	Standardise as in (i) but do not give if "1 –" or wrong signs in *either* equation [0.674, 0.675] seen. (Other errors lead to loss of A marks.) σ, a.r.t. 12.3, cwo μ, a.r.t. 51.7, cwo [NB: ***CARE!*** either or both can be obtained from wrong equns.] {note for scoris zoning – (i) to be visible in marking (ii)}
3	(iii)	Based on a sample/small sample, etc	B1 **[1]**	Any similar comment, e.g. "frequencies not probabilities" (but not *just* "*n* is small") *and* no wrong comments. Not "because data is grouped". No scattergun.

4	(i)	Snakes must occur independently of one another	B1 **[1]**	Contextualised ("snakes" must be mentioned); not *just* "singly" but allow both independent and singly. Allow explanation, e.g. "Occurrence of one snake doesn't affect occurrences of others". Allow "snakes must occur randomly". Otherwise, more than one condition, "e.g. "randomly, independently, singly and at constant rate": 0.
4	(ii)	$1 - P(\leq 5)$ $= 1 - 0.7851$ $= \mathbf{0.2149}$	M1 A1 **[2]**	Give M1 for 0.3712, 0.1107 or 0.2307. Answer 0.7851 is M0. Answer, a.r.t. 0.215
4	(iii)	Po(3.08) $e^{-3.08}\left(\frac{3.08^2}{2!}+\frac{3.08^3}{3!}\right)$ [= 0.2180 + 0.2238] $= \mathbf{0.4418}$	M1 M1 A1ft A1 **[4]**	Po(3.08) stated or implied. [Just $\lambda = 3.08$ is M0 unless Poisson later.] Correct formula for Po ($r > 0$) used at least once, can be implied Completely correct formula for their λ (not 4), can be implied Final answer, a.r.t. 0.442 No working: last 3 marks either 0 or 3, no "nearly right".
5	(i)	$\int_0^1 \frac{\pi}{2}\sin(\pi x)\,dx = \left[-\frac{1}{2}\cos(\pi x)\right]_0^1 = \frac{1}{2}-\left(-\frac{1}{2}\right)=1$ and function non-negative for all x in range	M1 B1 A1 B1 **[4]**	Attempt to integrate f(x), limits (0, 1) somewhere, evidence e.g. "from calculator" Correctly integrate $\sin(\pi x)$ to $-\frac{1}{2}\cos(\pi x)$ Fully correct, need to see $-\frac{1}{2}\cos(\pi x)$ and final 1, no wrong working seen Non-negative asserted explicitly, allow positive or equivalent. Not just graph drawn. *(Most will not get this mark!)*
5	(ii)	 $E(X) = \frac{1}{2}$	M1 A1 B1 **[3]**	Correct shape, through 0, allow below axis outside range. Allow partial curve if clearly part of sine curve. Fully correct including no extension beyond [0, 1]. Don't worry about grads at ends. Ignore labelling of axes ½ or 0.5, needs to be simplified, no working needed, *no* ft
5	(iii)	$\int_q^1 \frac{1}{2}\pi\sin(\pi x)dx = 0.75$; $\{\left[-\frac{1}{2}\cos(\pi x)\right]_q^1 = 0.75\}$ $\cos(\pi q) = 0.5$ Solve to get $q = \frac{1}{3}$	M1 A1 A1 **[3]**	Equate integral to correct probability, correct limits *somewhere* allow complementary probability (= 0.25) only if limits (0, q) $\cos(\pi q) = 0.5$ or exact equivalent $q = \frac{1}{3}$ or a.r.t. 0.333. SR: Numerical (no working needed): 0.333 B3, 0.33 B2
5	(iv)	$\int_0^1 \frac{\pi}{2}x^2\sin(\pi x)\,dx - \left(\frac{1}{2}\right)^2$	M1 A1ft **[2]**	Integral part correct, allow limits omitted, ignore dx Subtract their $[E(X)]^2$, allow μ in form of integral, correct limits needed, not just "μ^2" {note for scoris zoning – (ii) needs to be visible here}
5	(v)	Values of x in range close to E(X) are more likely than those further away	B1 **[1]**	Need to see "values of x" or equivalent, and probably not "occur" *Not* "the probability of x is greater when x is close to E(X)" etc. *Not* "PDF greater …"

6	(i)	Sample is random	B1 **[1]**	Indicate random sample. Allow "unbiased sample" or "randomly selected" or "all equally likely". Allow "representative" provided it's clearly "of company" (not city) Not just "independent". Withhold if extra wrong bits.
6	(ii)	List population, number sequentially Select using random numbers	B1 B1 **[2]**	List can be implied; must imply employees or people. "Sequential" can be assumed. Not "select numbers randomly", Don't need "ignore outside range" etc. Number randomly *and* select randomly, B1, but "assign random nos & arrange", B2 SC: Put names into hat/lottery machine and take them out: B2 SC: Systematic: B1 for list, can get second B1 if starting-point random
6	(iii)	H_0: $p = 0.4$; H_1: $p < 0.4$ B(12, 0.4) α: P(≤ 2) = **0.0834** > 0.05	B2 M1 A1 A1	Both correct, B2. Allow π. One error, e.g, μ or no symbol, B1, but $\bar{x}$, z etc: B0. B(12, 0.4) stated or implied. Can be implied by N(4.8, 2.88) but no further marks. 0.1673, 0.0398, 0.1513, 0.0421: M1A0(A1M1A1) P(≤ 2) = 0.0834, or P(> 2) = 0.9166. Compare numerical P(≤ 2) with 0.05, or P(> 2) with 0.95
		β: CR is ≤ 1 0.0196 seen and compare 2 with ≤ 1	A1 A1	CR is ≤ 1 stated. Explicitly compare 2 with CR, probability 0.0196 must be seen
		Do not reject H_0. Insufficient evidence that proportion of employees from group *Z* is less.	M1 A1ft **[7]**	Correct first conclusion, needs P(≤ 2 | $p = 0.4$) or fully consistent equivalent In context (mention "employees", "city" etc), acknowledge uncertainty ("evidence") *Not* "there is evidence that the proportion of employees is 0.4" FT on wrong *p*-value or wrong critical value if previous mark gained SC: Normal: B2 M1 max SC: P(= 2) or P(≥ 2) or P(< 2): B2 M1 max SC: two-tailed: can get B1B0 M1A1A0 M1A1 (don't give second A1 for 0.05)
6	(iv)	Yes as H_0 is rejected	M1 A1 **[2]**	Realise this changes conclusion (FT!), or "more likely to reject H_0", "larger CR" More supportive [just "more supportive" without evidence is M0A0]
7	(i)	$\hat{\mu} = \bar{x} = 81$ $\frac{329800}{50} - 81^2$ [= 35] $\times \frac{50}{49}$; = 35.71 $1 - \Phi\left(\frac{90-81}{\sqrt{35.71}}\right) = 1 - \Phi(1.506) = 1 - 0.9339$ **= 6.61%** or **0.0661**	B1 M1 M1 A1 M1 A1 **[6]**	81 only, can be implied Correct formula for biased estimate, their "81", can be implied Multiply by 50/49. SC: single formula: M2, or M1 if wrong but divisor 49 anywhere [can be recovered if correctly done in part (ii)] A.r.t. 35.7 – can't be recovered from part (ii). Can be implied Standardise with their μ and σ, allow σ^2, cc but not $\sqrt{50}$ Answer, a.r.t. 6.6% or 0.066

7	(ii)	H_0: $\mu = 80$ H_1: $\mu \neq 80$	B2	Correct, B2. One error, e.g. wrong or no symbol, >, B1, but x or $\bar{x}$ or t etc, or 81, B0. NB: If both hypotheses involve 81, *can't* get final M1
		α: $z = \dfrac{81-80}{\sqrt{35.71/50}} = 1.183$ [*or* $p = 0.1183$]	M1 A1	Standardise, with $\sqrt{50}$, allow $\sqrt{}$, sign or cc errors, allow from biased variance z, a.r.t. 1.18, or p, a.r.t. 0.118. Allow –1.18.
		< 1.645	B1	Their $z < 1.645$ or $p > 0.05$, *not* if one-tail. Allow –1.18 > –1.645. *Not* just 1.645 seen.
		β: CV $80 + 1.645\sqrt{\dfrac{35.71}{50}} = 81.39$	M1 B1	$80 + zs/\sqrt{50}$, allow $\sqrt{}$ or cc errors, ignore – (no marks for – alone); $z = 1.645$ used in this expression (not just seen), *not* from one-tail
		$81 < 81.39$	A1	Compare CV with 81, allow 81.08 from one-tailed ($z = 1.282$) (but not on their σ) SC: $81 - 1.645\sqrt{\dfrac{35.71}{50}}$: If H_0: $\mu = 80$: (B2) M1B1A0M0A0. If H_0: $\mu = 81$: (B0) M1B1A1 (79.61) M0A0
		Do not reject H_0.	M1	Correct first conclusion, needs $\sqrt{50}$, correct comparison type, μ and $\bar{x}$ not consistently wrong way round (thus H_0: $\mu = 81$ can get B0 M1A1A1 M0A0, max 3/7) In method β, it needs to be clear that comparison involves $\bar{x}$.
		Insufficient evidence that the mean time is not 80 minutes.	A1ft **[7]**	Contextualised (mention "time"), acknowledge uncertainty ("evidence that…") *Not* "significant evidence that mean time is 80" FT on wrong z-value or wrong critical value if previous mark gained SC: One-tailed: can get B1B0 M1A1B0 M1A1, max 5/7 No $\sqrt{50}$: can get B2 M0 B1 M0, max 3/7
7	(iii)	(a) Yes (single observation only) (b) No, CLT applies to large sample	B1 B1 **[2]**	No reason needed, but withhold if wrong reason seen. Allow "yes, no dist^n given" "No" *and* refer to central limit theorem or "large sample" {note for scoris zoning – (a) and (b) to be in single zone}
8	(i)	$P(W = 0 \mid \lambda = 3.6)$ = **0.0273** or 2.73%	M1 A1 **[2]**	Use this conditional probability. *Not* 0.9727, not *just* 2.5% etc Answer a.r.t. 0.0273 or 2.73%. ISW if appropriate (e.g. "0.0273, $\therefore$2.5%")
8	(ii)	$1 - e^{-\lambda_0} = 0.8$ $e^{-\lambda_0} = 0.2$ $\lambda_0 = -\ln(0.2)$ = **1.609**	M1 A1 M1 A1 **[4]**	Use $P(W > 0 \mid \lambda = \lambda_0)$, formula needed but allow if wrong This exact equation, or $e^{\lambda_0} = 5$, or exact equivalent RHS Solve using ln or otherwise [independent of first M1, e.g. –ln(0.8) = 0.223 is M1 here] Final answer, exact or a.r.t. 1.61, cwo SC: No working: 1.60 (tables etc): B0. 1.61 (T&I): SC B4.