OCR
RECOGNISING ACHIEVEMENT

ADVANCED GCE UNIT **4733/01**

MATHEMATICS

Probability & Statistics 2

MONDAY 18 JUNE 2007 Morning

Time: 1 hour 30 minutes

Additional Materials: Answer Booklet (8 pages)
List of Formulae (MF1)

INSTRUCTIONS TO CANDIDATES

- Write your name, centre number and candidate number in the spaces provided on the answer booklet.
- Answer **all** the questions.
- 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.
- The total number of marks for this paper is 72.

ADVICE TO CANDIDATES

- Read each question carefully and make sure you know what you have to do before starting your answer.
- **You are reminded of the need for clear presentation in your answers.**

This document consists of **4** printed pages.

1 A random sample of observations of a random variable X is summarised by

$$n = 100, \quad \Sigma x = 4830.0, \quad \Sigma x^2 = 249\,509.16.$$

(i) Obtain unbiased estimates of the mean and variance of X. [4]

(ii) The sample mean of 100 observations of X is denoted by $\overline{X}$. Explain whether you would need any further information about the distribution of X in order to estimate $\mathrm{P}(\overline{X} > 60)$. [You should not attempt to carry out the calculation.] [2]

2 It is given that on average one car in forty is yellow. Using a suitable approximation, find the probability that, in a random sample of 130 cars, exactly 4 are yellow. [5]

3 The proportion of adults in a large village who support a proposal to build a bypass is denoted by p. A random sample of size 20 is selected from the adults in the village, and the members of the sample are asked whether or not they support the proposal.

(i) Name the probability distribution that would be used in a hypothesis test for the value of p. [1]

(ii) State the properties of a random sample that explain why the distribution in part **(i)** is likely to be a good model. [2]

4 X is a continuous random variable.

(i) State two conditions needed for X to be well modelled by a normal distribution. [2]

(ii) It is given that $X \sim \mathrm{N}(50.0, 8^2)$. The mean of 20 random observations of X is denoted by $\overline{X}$. Find $\mathrm{P}(\overline{X} > 47.0)$. [4]

5 The number of system failures per month in a large network is a random variable with the distribution $\mathrm{Po}(\lambda)$. A significance test of the null hypothesis $\mathrm{H}_0 : \lambda = 2.5$ is carried out by counting R, the number of system failures in a period of 6 months. The result of the test is that H_0 is rejected if $R > 23$ but is not rejected if $R \leqslant 23$.

(i) State the alternative hypothesis. [1]

(ii) Find the significance level of the test. [3]

(iii) Given that $\mathrm{P}(R > 23) < 0.1$, use tables to find the largest possible actual value of λ. You should show the values of any relevant probabilities. [3]

6 In a rearrangement code, the letters of a message are rearranged so that the frequency with which any particular letter appears is the same as in the original message. In ordinary German the letter e appears 19% of the time. A certain encoded message of 20 letters contains one letter e.

(i) Using an exact binomial distribution, test at the 10% significance level whether there is evidence that the proportion of the letter e in the language from which this message is a sample is less than in German, i.e., less than 19%. [8]

(ii) Give a reason why a binomial distribution might not be an appropriate model in this context. [1]

7 Two continuous random variables S and T have probability density functions as follows.

$$S: \quad \mathrm{f}(x) = \begin{cases} \frac{1}{2} & -1 \leqslant x \leqslant 1 \\ 0 & \text{otherwise} \end{cases}$$

$$T: \quad \mathrm{g}(x) = \begin{cases} \frac{3}{2}x^2 & -1 \leqslant x \leqslant 1 \\ 0 & \text{otherwise} \end{cases}$$

(i) *Sketch* on the same axes the graphs of $y = \mathrm{f}(x)$ and $y = \mathrm{g}(x)$. [You should not use graph paper or attempt to plot points exactly.] [3]

(ii) Explain in everyday terms the difference between the two random variables. [2]

(iii) Find the value of t such that $\mathrm{P}(T > t) = 0.2$. [5]

8 A random variable Y is normally distributed with mean μ and variance 12.25. Two statisticians carry out significance tests of the hypotheses $\mathrm{H}_0 : \mu = 63.0$, $\mathrm{H}_1 : \mu > 63.0$.

(i) Statistician A uses the mean $\overline{Y}$ of a sample of size 23, and the critical region for his test is $\overline{Y} > 64.20$. Find the significance level for A's test. [4]

(ii) Statistician B uses the mean of a sample of size 50 and a significance level of 5%.

(a) Find the critical region for B's test. [3]

(b) Given that $\mu = 65.0$, find the probability that B's test results in a Type II error. [4]

(iii) Given that, when $\mu = 65.0$, the probability that A's test results in a Type II error is 0.1365, state with a reason which test is better. [2]

9 **(a)** The random variable G has the distribution $\mathrm{B}(n, 0.75)$. Find the set of values of n for which the distribution of G can be well approximated by a normal distribution. [3]

(b) The random variable H has the distribution $\mathrm{B}(n, p)$. It is given that, using a normal approximation, $\mathrm{P}(H \geqslant 71) = 0.0401$ and $\mathrm{P}(H \leqslant 46) = 0.0122$.

(i) Find the mean and standard deviation of the approximating normal distribution. [6]

(ii) Hence find the values of n and p. [4]

1	(i)	$\hat{\mu}$ = 4830.0/100 = 48.3 249509.16/100 – (their $\bar{x}^2$) × 100/99 = 163.84	B1 M1 M1 A1	**4**	48.3 seen Biased estimate: 162.2016: can get B1M1M0 Multiply by $n/(n-1)$ Answer, 164 or 163.8 or 163.84
	(ii)	No, Central Limit theorem applies, so can assume distribution is normal	B2	**2**	"No" with statement showing CLT is understood (though CLT does not need to be mentioned) [SR: No with reason that is not wrong: B1]
2		B(130, 1/40) ≈ Po(3.25) $e^{-\lambda}\frac{\lambda^3}{4!}$ = 0.180	B1 M1 A1√ M1 A1	**5**	B(130, 1/40) stated or implied Poisson, *or* correct N on their B(n, p) Parameter their np, *or* correct parameter(s)√ Correct formula, or interpolation Answer, 0.18 or a.r.t. 0.180 [SR: N(3.25, 3.17) or N(3.25, 3.25): B1M1A1]
3	(i)	Binomial	B1	**1**	Binomial stated or implied
	(ii)	Each element equally likely Choices independent	B1 B1	**2**	All elements, or selections, equally likely stated Choices independent [not just "independent"] [can get B2 even if (i) is wrong]
4	(i)	*Two of:* Distribution symmetric No substantial truncation Unimodal/Increasingly unlikely further from μ, etc	B1 B1	**2**	One property Another definitely different property Don't give both marks for just these two "Bell-shaped": B1 only unless "no truncation"
	(ii)	Variance $8^2/20$ $z = \frac{47.0-50.0}{\sqrt{8^2/20}} = -1.677$ Φ(1.677) = 0.9532	M1 A1 A1 A1	**4**	Standardise, allow cc, don't need n Denominator (8 or 8^2 or √8) ÷ (20 or √20 or 20^2) z-value, a.r.t. –1.68 or +1.68 Answer, a.r.t. 0.953
5	(i)	H_1: $\lambda > 2.5$ or 15	B1	**1**	λ > 2.5 or 15, allow μ, don't need "H_1"
	(ii)	Use parameter 15 P(> 23) 1 – 0.9805 = 0.0195 or 1.95%	M1 M1 A1	**3**	λ = 15 used [N(15, 15) gets this mark only] Find P(> 23 or ≥ 23), final answer < 0.5 eg 0.0327 or 0.0122 Answer, 1.95% or 2% or 0.0195 or 0.02 [SR: 2-tailed, 3.9% gets 3/3 here]
	(iii)	P(≤ 23 | λ = 17) = 0.9367 P(≤ 23 | λ = 18) = 0.8989 Parameter = 17 λ = 17/6 or 2.83	M1 A1 M1	**3**	One of these, or their complement: .9367, .8989, 0.9047, 0.8551, .9317, .8933, .9907, .9805 Parameter 17 *[17.1076]*, needs P(≤ 23), cwo [SR: if insufficient evidence can give B1 for 17] Their parameter ÷ 6 *[2.85]* [SR: Solve (23.5 – λ)/√λ = 1.282 M1; 18.05 A0]
6	(i)	H_0: $p = 0.19$, H_1: $p < 0.19$ where p is population proportion $0.81^{20} + 20 \times 0.81^{19} \times 0.19$ = 0.0841 Compare 0.1	B2 M1 A1 A1 B1		Correct, B2. One error, B1, but x or $\bar{x}$ or r: B0 Binomial probabilities, allow 1 term only Correct expression [0.0148 + 0.0693] Probability, a.r.t. 0.084 Explicit comparison of "like with like"
	or	Add binomial probs until ans > 0.1 Critical region ≤ 1	A1 B1		[P(≤ 2) = 0.239]
		Reject H_0 Significant evidence that proportion of *e*'s in language is less than 0.19	M1 A1√	**8**	Correct deduction and method [needs P(≤ 1)] Correct conclusion in context [SR: N(3.8, 3.078): B2M1A0B1M0]
	(ii)	Letters not independent	B1	**1**	Correct modelling assumption, stated in context Allow "random", "depends on message", etc

7	(i)		B1 B1 B1 **3**	Horizontal straight line Positive parabola, symmetric about 0 Completely correct, including correct relationship between two Don't need vertical lines or horizontal lines outside range, but don't give last B1 if horizontal line continues past "±1"
	(ii)	*S* is equally likely to take any value in range, *T* is more likely at extremities	B2 **2**	Correct statement about distributions (*not* graphs) *[Partial statement, or correct description for one only: B1]*
	(iii)	$\int_t^1 \frac{3}{2}x^2 dx = \left[\frac{x^3}{2}\right]_t^1$ $\frac{1}{2}(1-t^3) = 0.2$ or $\frac{1}{2}(t^3+1) = 0.8$ $t^3 = 0.6$ $t = 0.8434$	M1 B1 M1 M1 A1 **5**	Integrate f(x) with limits (-1, t) or (t, 1) [recoverable if t used later] Correct indefinite integral Equate to 0.2, or 0.8 if [-1, t] used Solve cubic equation to find t Answer, in range [0.843, 0.844]
8	(i)	$\frac{64.2-63}{\sqrt{12.25/23}} = 1.644$ P(z > 1.644) = 0.05	M1dep A1 dep M1 A1 **4**	Standardise 64.2 with $\sqrt{n}$ z = 1.644 or 1.645, must be + Find $\Phi(z)$, answer < 0.5 Answer, a.r.t. 0.05 or 5.0%
	(ii)	(a) $63 + 1.645 \times \frac{3.5}{\sqrt{50}}$ ≥ 63.81	M1 B1 A1 **3**	$63 + 3.5 \times k / \sqrt{50}$, k from Φ^{-1}, *not* – k = 1.645 (allow 1.64, 1.65) Answer, a.r.t. 63.8, allow >, ≥, =, c.w.o.
		(b) P(< 63.8 | μ = 65) $\frac{63.8-65}{3.5/\sqrt{50}} = -2.3956$ 0.0083	M1 M1 A1 A1 **4**	Use of correct meaning of Type II Standardise their c with $\sqrt{50}$ z = (±) 2.40 [or –2.424 or – 2.404 etc] Answer, a.r.t. 0.008 [eg, 0.00767]
	(iii)	B better: Type II error smaller (and same Type I error)	B2√ **2**	This answer: B2. "B because sample bigger": B1. *[SR: Partial answer: B1]*
9	(a)	$np > 5$ and $nq > 5$ $0.75n > 5$ is relevant $n > 20$	M2 A1 **3**	Use either $nq > 5$ or $npq > 5$ [SR: If M0, use $np > 5$, or "$n = 20$" seen: M1] Final answer $n > 20$ or $n \geq 20$ only
	(b)	(i) $70.5 - \mu = 1.75\sigma$ $\mu - 46.5 = 2.25\sigma$ Solve simultaneously $\mu = 60$ $\sigma = 6$	M1 A1 B1 M1 A1√ A1√ **6**	Standardise once, and equate to Φ^{-1}, ± cc Standardise twice, signs correct, cc correct Both 1.75 and 2.25 Correct solution method to get one variable μ, a.r.t. 60.0 or ± 154.5 σ, a.r.t. 6.00 [Wrong cc (below): A1 both] [SR: σ^2: M1A0B1M1A1A0]
		(ii) $np = 60$, $npq = 36$ $q = 36/60 = 0.6$ $p = 0.4$ $n = 150$	M1dep depM1 A1√ A1√ **4**	$np = 60$ and $npq = 6^2$ or 6 Solve to get q or p or n $p = 0.4$ √ on wrong cc or z $n = 150$ √ on wrong cc or z

		σ	μ	q	p (±0.01)	n
70.5	46.5	6	60	0.6	0.4	150
			60.062			
71	46	6.25	5	0.6504	0.3496	171.8
			60.562			
71.5	46.5	6.25	5	0.6450	0.3550	170.6
			59.562			
70.5	45.5	6.25	5	0.6558	0.3442	173.0
71.5	45.5	6.5	60.125	0.7027	0.2973	202.2
70	46	6	59.5	0.6050	0.3950	150.6