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.

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

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

(i) *Sketch* on the same axes the graphs of $y = f(x)$ and $y = g(x)$. [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 $P(T > t) = 0.2$. [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 $H_0 : \mu = 63.0$, $H_1 : \mu > 63.0$.

(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 $B(n, 0.75)$. 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 $B(n, p)$. It is given that, using a normal approximation, $P(H \geqslant 71) = 0.0401$ and $P(H \leqslant 46) = 0.0122$.

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

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