

OCR Maths S2

Past Paper Pack

2006-2014

Jan 2006

- 1 In a study of urban foxes it is found that on average there are 2 foxes in every 3 acres.
- (i) Use a Poisson distribution to find the probability that, at a given moment,
- (a) in a randomly chosen area of 3 acres there are at least 4 foxes, [2]
- (b) in a randomly chosen area of 1 acre there are exactly 2 foxes. [3]
- (ii) Explain briefly why a Poisson distribution might not be a suitable model. [2]
- 2 The random variable W has the distribution $B(40, \frac{2}{7})$. Use an appropriate approximation to find $P(W > 13)$. [7]
- 3 The manufacturers of a brand of chocolates claim that, on average, 30% of their chocolates have hard centres. In a random sample of 8 chocolates from this manufacturer, 5 had hard centres. Test, at the 5% significance level, whether there is evidence that the population proportion of chocolates with hard centres is not 30%, stating your hypotheses clearly. Show the values of any relevant probabilities. [7]
- 4 DVD players are tested after manufacture. The probability that a randomly chosen DVD player is defective is 0.02. The number of defective players in a random sample of size 80 is denoted by R .
- (i) Use an appropriate approximation to find $P(R \geq 2)$. [4]
- (ii) Find the smallest value of r for which $P(R \geq r) < 0.01$. [3]
- 5 In an investment model the increase, $Y\%$, in the value of an investment in one year is modelled as a continuous random variable with the distribution $N(\mu, \frac{1}{4}\mu^2)$. The value of μ depends on the type of investment chosen.
- (i) Find $P(Y < 0)$, showing that it is independent of the value of μ . [4]
- (ii) Given that $\mu = 6$, find the probability that $Y < 9$ in each of three randomly chosen years. [4]
- (iii) Explain why the calculation in part (ii) might not be valid if applied to three consecutive years. [1]
- 6 Alex obtained the actual waist measurements, w inches, of a random sample of 50 pairs of jeans, each of which was labelled as having a 32-inch waist. The results are summarised by
- $$n = 50, \quad \Sigma w = 1615.0, \quad \Sigma w^2 = 52\,214.50.$$
- Test, at the 0.1% significance level, whether this sample provides evidence that the mean waist measurement of jeans labelled as having 32-inch waists is in fact greater than 32 inches. State your hypotheses clearly. [10]

Jan 2006

- 7 The random variable X has the distribution $N(\mu, 8^2)$. The mean of a random sample of 12 observations of X is denoted by \bar{X} . A test is carried out at the 1% significance level of the null hypothesis $H_0: \mu = 80$ against the alternative hypothesis $H_1: \mu < 80$. The test is summarised as follows: 'Reject H_0 if $\bar{X} < c$; otherwise do not reject H_0 '.

(i) Calculate the value of c . [4]

(ii) Assuming that $\mu = 80$, state whether the conclusion of the test is correct, results in a Type I error, or results in a Type II error if:

(a) $\bar{X} = 74.0$, [1]

(b) $\bar{X} = 75.0$. [1]

(iii) Independent repetitions of the above test, using the value of c found in part (i), suggest that in fact the probability of rejecting the null hypothesis is 0.06. Use this information to calculate the value of μ . [4]

- 8 A continuous random variable X has probability density function given by

$$f(x) = \begin{cases} kx^n & 0 \leq x \leq 1, \\ 0 & \text{otherwise,} \end{cases}$$

where n and k are positive constants.

(i) Find k in terms of n . [3]

(ii) Show that $E(X) = \frac{n+1}{n+2}$. [3]

It is given that $n = 3$.

(iii) Find the variance of X . [3]

(iv) One hundred observations of X are taken, and the mean of the observations is denoted by \bar{X} . Write down the approximate distribution of \bar{X} , giving the values of any parameters. [3]

(v) Write down the mean and the variance of the random variable Y with probability density function given by

$$g(y) = \begin{cases} 4\left(y + \frac{4}{5}\right)^3 & -\frac{4}{5} \leq y \leq \frac{1}{5}, \\ 0 & \text{otherwise.} \end{cases} \quad [3]$$

- 1 Calculate the variance of the continuous random variable with probability density function given by

$$f(x) = \begin{cases} \frac{3}{37}x^2 & 3 \leq x \leq 4, \\ 0 & \text{otherwise.} \end{cases} \quad [6]$$

- 2 (i) The random variable R has the distribution $B(6, p)$. A random observation of R is found to be 6. Carry out a 5% significance test of the null hypothesis $H_0: p = 0.45$ against the alternative hypothesis $H_1: p \neq 0.45$, showing all necessary details of your calculation. [4]

- (ii) The random variable S has the distribution $B(n, p)$. H_0 and H_1 are as in part (i). A random observation of S is found to be 1. Use tables to find the largest value of n for which H_0 is not rejected. Show the values of any relevant probabilities. [3]

- 3 The continuous random variable T has mean μ and standard deviation σ . It is known that $P(T < 140) = 0.01$ and $P(T < 300) = 0.8$.

- (i) Assuming that T is normally distributed, calculate the values of μ and σ . [6]

In fact, T represents the time, in minutes, taken by a randomly chosen runner in a public marathon, in which about 10% of runners took longer than 400 minutes.

- (ii) State with a reason whether the mean of T would be higher than, equal to, or lower than the value calculated in part (i). [2]

- 4 (i) Explain briefly what is meant by a random sample. [1]

Random numbers are used to select, with replacement, a sample of size n from a population numbered 000, 001, 002, ..., 799.

- (ii) If $n = 6$, find the probability that exactly 4 of the selected sample have numbers less than 500. [3]

- (iii) If $n = 60$, use a suitable approximation to calculate the probability that at least 40 of the selected sample have numbers less than 500. [6]

- 5 An airline has 300 seats available on a flight to Australia. It is known from experience that on average only 99% of those who have booked seats actually arrive to take the flight, the remaining 1% being called 'no-shows'. The airline therefore sells more than 300 seats. If more than 300 passengers then arrive, the flight is over-booked. Assume that the number of no-show passengers can be modelled by a binomial distribution.

- (i) If the airline sells 303 seats, state a suitable distribution for the number of no-show passengers, and state a suitable approximation to this distribution, giving the values of any parameters. [2]

Using the distribution and approximation in part (i),

- (ii) show that the probability that the flight is over-booked is 0.4165, correct to 4 decimal places, [2]

- (iii) find the largest number of seats that can be sold for the probability that the flight is over-booked to be less than 0.2. [5]

June 2006

- 6 Customers arrive at a post office at a constant average rate of 0.4 per minute.
- (i) State an assumption needed to model the number of customers arriving in a given time interval by a Poisson distribution. [1]

Assuming that the use of a Poisson distribution is justified,

- (ii) find the probability that more than 2 customers arrive in a randomly chosen 1-minute interval, [2]
- (iii) use a suitable approximation to calculate the probability that more than 55 customers arrive in a given two-hour interval, [6]
- (iv) calculate the smallest time for which the probability that no customers arrive in that time is less than 0.02, giving your answer to the nearest second. [5]
- 7 Three independent researchers, *A*, *B* and *C*, carry out significance tests on the power consumption of a manufacturer's domestic heaters. The power consumption, X watts, is a normally distributed random variable with mean μ and standard deviation 60. Each researcher tests the null hypothesis $H_0: \mu = 4000$ against the alternative hypothesis $H_1: \mu > 4000$.

Researcher *A* uses a sample of size 50 and a significance level of 5%.

- (i) Find the critical region for this test, giving your answer correct to 4 significant figures. [6]

In fact the value of μ is 4020.

- (ii) Calculate the probability that Researcher *A* makes a Type II error. [6]
- (iii) Researcher *B* uses a sample bigger than 50 and a significance level of 5%. Explain whether the probability that Researcher *B* makes a Type II error is less than, equal to, or greater than your answer to part (ii). [2]
- (iv) Researcher *C* uses a sample of size 50 and a significance level bigger than 5%. Explain whether the probability that Researcher *C* makes a Type II error is less than, equal to, or greater than your answer to part (ii). [2]
- (v) State with a reason whether it is necessary to use the Central Limit Theorem at any point in this question. [2]

- 1 The random variable H has the distribution $N(\mu, 5^2)$. It is given that $P(H < 22) = 0.242$. Find the value of μ . [4]
- 2 A school has 900 pupils. For a survey, Jan obtains a list of all the pupils, numbered 1 to 900 in alphabetical order. She then selects a sample by the following method. Two fair dice, one red and one green, are thrown, and the number in the list of the first pupil in the sample is determined by the following table.

		Score on green dice					
		1	2	3	4	5	6
Score on red dice	1, 2 or 3	1	2	3	4	5	6
	4, 5 or 6	7	8	9	10	11	12

For example, if the scores on the red and green dice are 5 and 2 respectively, then the first member of the sample is the pupil numbered 8 in the list.

Starting with this first number, every 12th number on the list is then used, so that if the first pupil selected is numbered 8, the others will be numbered 20, 32, 44,

- (i) State the size of the sample. [1]
- (ii) Explain briefly whether the following statements are true.
- (a) Each pupil in the school has an equal probability of being in the sample. [1]
- (b) The pupils in the sample are selected independently of one another. [1]
- (iii) Give a reason why the number of the first pupil in the sample should not be obtained simply by adding together the scores on the two dice. Justify your answer. [2]
- 3 A fair dice is thrown 90 times. Use an appropriate approximation to find the probability that the number 1 is obtained 14 or more times. [6]

- 4 A set of observations of a random variable W can be summarised as follows:

$$n = 14, \quad \Sigma w = 100.8, \quad \Sigma w^2 = 938.70.$$

- (i) Calculate an unbiased estimate of the variance of W . [4]
- (ii) The mean of 70 observations of W is denoted by \bar{W} . State the approximate distribution of \bar{W} , including unbiased estimate(s) of any parameter(s). [3]

Jan 2007

5 On a particular night, the number of shooting stars seen per minute can be modelled by the distribution $Po(0.2)$.

(i) Find the probability that, in a given 6-minute period, fewer than 2 shooting stars are seen. [3]

(ii) Find the probability that, in 20 periods of 6 minutes each, the number of periods in which fewer than 2 shooting stars are seen is exactly 13. [3]

(iii) Use a suitable approximation to find the probability that, in a given 2-hour period, fewer than 30 shooting stars are seen. [6]

6 The continuous random variable X has the following probability density function:

$$f(x) = \begin{cases} a + bx & 0 \leq x \leq 2, \\ 0 & \text{otherwise,} \end{cases}$$

where a and b are constants.

(i) Show that $2a + 2b = 1$. [3]

(ii) It is given that $E(X) = \frac{11}{9}$. Use this information to find a second equation connecting a and b , and hence find the values of a and b . [6]

(iii) Determine whether the median of X is greater than, less than, or equal to $E(X)$. [4]

7 A television company believes that the proportion of households that can receive Channel C is 0.35.

(i) In a random sample of 14 households it is found that 2 can receive Channel C. Test, at the 2.5% significance level, whether there is evidence that the proportion of households that can receive Channel C is less than 0.35. [7]

(ii) On another occasion the test is carried out again, with the same hypotheses and significance level as in part (i), but using a new sample, of size n . It is found that no members of the sample can receive Channel C. Find the largest value of n for which the null hypothesis is not rejected. Show all relevant working. [4]

[Question 8 is printed overleaf.]

Jan 2007

8 The quantity, X milligrams per litre, of silicon dioxide in a certain brand of mineral water is a random variable with distribution $N(\mu, 5.6^2)$.

(i) A random sample of 80 observations of X has sample mean 100.7. Test, at the 1% significance level, the null hypothesis $H_0 : \mu = 102$ against the alternative hypothesis $H_1 : \mu \neq 102$. [5]

(ii) The test is redesigned so as to meet the following conditions.

- The hypotheses are $H_0 : \mu = 102$ and $H_1 : \mu < 102$.
- The significance level is 1%.
- The probability of making a Type II error when $\mu = 100$ is to be (approximately) 0.05.

The sample size is n , and the critical region is $\bar{X} < c$, where \bar{X} denotes the sample mean.

(a) Show that n and c satisfy (approximately) the equation $102 - c = \frac{13.0256}{\sqrt{n}}$. [3]

(b) Find another equation satisfied by n and c . [2]

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

Permission to reproduce items where third-party owned material protected by copyright is included has been sought and cleared where possible. Every reasonable effort has been made by the publisher (UCLES) to trace copyright holders, but if any items requiring clearance have unwittingly been included, the publisher will be pleased to make amends at the earliest possible opportunity.

OCR is part of the Cambridge Assessment Group. Cambridge Assessment is the brand name of University of Cambridge Local Examinations Syndicate (UCLES), which is itself a department of the University of Cambridge.

June 2007

- 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 \bar{X} . Explain whether you would need any further information about the distribution of X in order to estimate $P(\bar{X} > 60)$. [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 $X \sim N(50.0, 8^2)$. The mean of 20 random observations of X is denoted by \bar{X} . Find $P(\bar{X} > 47.0)$. [4]
- 5 The number of system failures per month in a large network is a random variable with the distribution $Po(\lambda)$. A significance test of the null hypothesis $H_0 : \lambda = 2.5$ 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 $R \leq 23$.
- (i) State the alternative hypothesis. [1]
- (ii) Find the significance level of the test. [3]
- (iii) Given that $P(R > 23) < 0.1$, 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]

June 2007

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

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

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

- (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 \bar{Y} of a sample of size 23, and the critical region for his test is $\bar{Y} > 64.20$. 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, $P(H \geq 71) = 0.0401$ and $P(H \leq 46) = 0.0122$.
- (i) Find the mean and standard deviation of the approximating normal distribution. [6]
- (ii) Hence find the values of n and p . [4]

Jan 2008

- 1 The random variable T is normally distributed with mean μ and standard deviation σ . It is given that $P(T > 80) = 0.05$ and $P(T > 50) = 0.75$. Find the values of μ and σ . [6]
- 2 A village has a population of 600 people. A sample of 12 people is obtained as follows. A list of all 600 people is obtained and a three-digit number, between 001 and 600 inclusive, is allocated to each name in alphabetical order. Twelve three-digit random numbers, between 001 and 600 inclusive, are obtained and the people whose names correspond to those numbers are chosen.
- (i) Find the probability that all 12 of the numbers chosen are 500 or less. [3]
- (ii) When the selection has been made, it is found that all of the numbers chosen are 500 or less. One of the people in the village says, "The sampling method must have been biased." Comment on this statement. [2]
- 3 The random variable G has the distribution $Po(\lambda)$. A test is carried out of the null hypothesis $H_0 : \lambda = 4.5$ against the alternative hypothesis $H_1 : \lambda \neq 4.5$, based on a single observation of G . The critical region for the test is $G \leq 1$ and $G \geq 9$.
- (i) Find the significance level of the test. [5]
- (ii) Given that $\lambda = 5.5$, calculate the probability that the test results in a Type II error. [3]
- 4 The random variable Y has the distribution $N(\mu, \sigma^2)$. The results of 40 independent observations of Y are summarised by
- $$\Sigma y = 3296.0, \quad \Sigma y^2 = 286\,800.40.$$
- (i) Calculate unbiased estimates of μ and σ^2 . [4]
- (ii) Use your answers to part (i) to estimate the probability that a single random observation of Y will be less than 60.0. [2]
- (iii) Explain whether it is necessary to know that Y is normally distributed in answering part (i) of this question. [1]
- 5 Over a long period the number of visitors per week to a stately home was known to have the distribution $N(500, 100^2)$. After higher car parking charges were introduced, a sample of four randomly chosen weeks gave a mean number of visitors per week of 435. You should assume that the number of visitors per week is still normally distributed with variance 100^2 .
- (i) Test, at the 10% significance level, whether there is evidence that the mean number of visitors per week has fallen. [7]
- (ii) Explain why it is necessary to assume that the distribution of the number of visitors per week (after the introduction of higher charges) is normal in order to carry out the test. [2]

Jan 2008

- 6 The number of house sales per week handled by an estate agent is modelled by the distribution $Po(3)$.
- (i) Find the probability that, in one randomly chosen week, the number of sales handled is
- (a) greater than 4, [2]
- (b) exactly 4. [2]
- (ii) Use a suitable approximation to the Poisson distribution to find the probability that, in a year consisting of 50 working weeks, the estate agent handles more than 165 house sales. [5]
- (iii) One of the conditions needed for the use of a Poisson model to be valid is that house sales are independent of one another.
- (a) Explain, in non-technical language, what you understand by this condition. [1]
- (b) State another condition that is needed. [1]

- 7 A continuous random variable X_1 has probability density function given by

$$f(x) = \begin{cases} kx & 0 \leq x \leq 2, \\ 0 & \text{otherwise,} \end{cases}$$

where k is a constant.

- (i) Show that $k = \frac{1}{2}$. [2]
- (ii) Sketch the graph of $y = f(x)$. [2]
- (iii) Find $E(X_1)$ and $\text{Var}(X_1)$. [5]
- (iv) Sketch the graph of $y = f(x - 1)$. [2]
- (v) The continuous random variable X_2 has probability density function $f(x - 1)$ for all x . Write down the values of $E(X_2)$ and $\text{Var}(X_2)$. [2]
- 8 Consultations are taking place as to whether a site currently in use as a car park should be developed as a shopping mall. An agency acting on behalf of a firm of developers claims that at least 65% of the local population are in favour of the development. In a survey of a random sample of 12 members of the local population, 6 are in favour of the development.
- (i) Carry out a test, at the 10% significance level, to determine whether the result of the survey is consistent with the claim of the agency. [7]
- (ii) A local residents' group claims that no more than 35% of the local population are in favour of the development. Without further calculations, state with a reason what can be said about the claim of the local residents' group. [2]
- (iii) A test is carried out, at the 15% significance level, of the agency's claim. The test is based on a random sample of size $2n$, and exactly n of the sample are in favour of the development. Find the smallest possible value of n for which the outcome of the test is to reject the agency's claim. [4]

June 2008

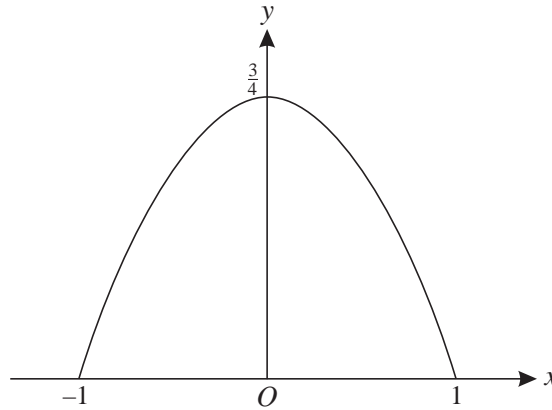
- 1 The head teacher of a school asks for volunteers from among the pupils to take part in a survey on political interests.
- (i) Explain why a sample consisting of all the volunteers is unlikely to give a true picture of the political interests of all pupils in the school. [2]
 - (ii) Describe a better method of obtaining the sample. [3]
- 2 The annual salaries of employees in a company have mean £30 000 and standard deviation £12 000.
- (i) Assuming a normal distribution, calculate the probability that the salary of one randomly chosen employee lies between £20 000 and £24 000. [4]
 - (ii) The salary structure of the company is such that a small number of employees earn much higher salaries than the others. Explain what this suggests about the use of a normal distribution to model the data. [2]
- 3 In a factory the time, T minutes, taken by an employee to make a single item is a normally distributed random variable with mean 28.0. A new ventilation system is installed, after which the times taken to produce a random sample of 40 items are measured. The sample mean is 26.44 minutes and it is given that $\frac{\sum t^2}{40} - 26.44^2 = 37.05$. Test, at the 10% significance level, whether there is evidence of a change in the mean time taken to make an item. [8]
- 4 The random variable U has the distribution $N(\mu, \sigma^2)$, where the value of σ is known. A test is carried out of the null hypothesis $H_0 : \mu = 50$ against the alternative hypothesis $H_1 : \mu > 50$. The test is carried out at the 1% significance level and is based on a random sample of size 10.
- (i) The test is carried out once. The value of the sample mean is 53.0. The outcome of the test is that H_0 is not rejected. Show that $\sigma > 4.08$, correct to 3 significant figures. [4]
 - (ii) The test is carried out repeatedly. In each test the actual value of μ is 50. Find the probability that the first test to result in a Type I error is the fifth to be carried out. Give your answer correct to 2 significant figures. [3]

June 2008

- 5 (i) A continuous random variable X has probability density function given by

$$f(x) = \begin{cases} \frac{3}{4}(1-x^2) & -1 \leq x \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

The graph of $y = f(x)$ is shown in the diagram.



Calculate the value of $\text{Var}(X)$. [4]

- (ii) A continuous random variable W has probability density function given by

$$g(x) = \begin{cases} k(9-x^2) & -3 \leq x \leq 3, \\ 0 & \text{otherwise,} \end{cases}$$

where k is a constant.

- (a) Sketch the graph of $y = g(x)$. [1]
- (b) By comparing the graphs of $y = f(x)$ and $y = g(x)$, explain how you can tell without calculation that $9k < \frac{3}{4}$. [2]
- (c) State with a reason, but without calculation, whether the standard deviation of W is greater than, equal to, or less than that of X . [2]
- 6 (a) On average I receive 19 e-mails per (8-hour) working day. Assuming that a Poisson distribution is a valid model, find the probability that in one randomly chosen hour I receive either 3 or 4 e-mails. [4]
- (b) (i) State the conditions needed to use a Poisson distribution as an approximation to a binomial distribution. [2]
- (ii) 108 people each throw a pair of fair six-sided dice. Use a Poisson approximation to find the probability that at least 4 people obtain a double six. [5]

[Questions 7 and 8 are printed overleaf.]

June 2008

- 7 Wendy analyses the number of ‘dropped catches’ in international cricket matches. She finds that the mean number of dropped catches per day is 2. In a recent 5-day match she found that there was a total of c dropped catches. She tests, at the 5% significance level, whether the mean number of dropped catches per day has increased.
- (i) State conditions needed for the number of dropped catches per day to be well modelled by a Poisson distribution. [2]
- Assume now that these conditions hold.
- (ii) Find the probability that the test results in a Type I error.
- (iii) Given that $c = 14$, carry out the test. [10]
- 8 A company sponsors a series of concerts. Surveys show that on average 40% of audience members know the name of the sponsor. As this figure is thought to be disappointingly low, the publicity material is redesigned.
- (i) After the publicity material has been redesigned, a random sample of 12 audience members is obtained, and it is found that 9 members of this sample know the name of the sponsor. Test, at the 5% significance level, whether there is evidence that the proportion of audience members who know the name of the sponsor has increased. [7]
- (ii) A more detailed 5% hypothesis test is carried out, based on a random sample of size 400. This test produces significant evidence that the proportion of audience members knowing the name of the sponsor has increased. Using an appropriate approximation, calculate the smallest possible number of audience members in the sample of 400 who know the name of the sponsor. [7]

- 1 A newspaper article consists of 800 words. For each word, the probability that it is misprinted is 0.005, independently of all other words. Use a suitable approximation to find the probability that the total number of misprinted words in the article is no more than 6. Give a reason to justify your approximation. [4]
- 2 The continuous random variable Y has the distribution $N(23.0, 5.0^2)$. The mean of n observations of Y is denoted by \bar{Y} . It is given that $P(\bar{Y} > 23.625) = 0.0228$. Find the value of n . [4]
- 3 The number of incidents of radio interference per hour experienced by a certain listener is modelled by a random variable with distribution $Po(0.42)$.
- (i) Find the probability that the number of incidents of interference in one randomly chosen hour is
- (a) 0,
(b) exactly 1. [3]
- (ii) Find the probability that the number of incidents in a randomly chosen 5-hour period is greater than 3. [3]
- (iii) One hundred hours of listening are monitored and the numbers of 1-hour periods in which 0, 1, 2, ... incidents of interference are experienced are noted. A bar chart is drawn to represent the results. Without any further calculations, sketch the shape that you would expect for the bar chart. (There is no need to use an exact numerical scale on the frequency axis.) [2]
- 4 A television company believes that the proportion of adults who watched a certain programme is 0.14. Out of a random sample of 22 adults, it is found that 2 watched the programme.
- (i) Carry out a significance test, at the 10% level, to determine, on the basis of this sample, whether the television company is overestimating the proportion of adults who watched the programme. [8]
- (ii) The sample was selected randomly. State what properties of this method of sampling are needed to justify the use of the distribution used in your test. [2]
- 5 The continuous random variables S and T have probability density functions as follows.
- $$S: \quad f(x) = \begin{cases} \frac{1}{4} & -2 \leq x \leq 2 \\ 0 & \text{otherwise} \end{cases}$$
- $$T: \quad g(x) = \begin{cases} \frac{5}{64}x^4 & -2 \leq x \leq 2 \\ 0 & \text{otherwise} \end{cases}$$
- (i) Sketch, on the same axes, the graphs of f and g . [3]
- (ii) Describe in everyday terms the difference between the distributions of the random variables S and T . (Answers that comment only on the shapes of the graphs will receive no credit.) [2]
- (iii) Calculate the variance of T . [4]

Jan 2009

- 6 The weight of a plastic box manufactured by a company is W grams, where $W \sim N(\mu, 20.25)$. A significance test of the null hypothesis $H_0 : \mu = 50.0$, against the alternative hypothesis $H_1 : \mu \neq 50.0$, is carried out at the 5% significance level, based on a sample of size n .

(i) Given that $n = 81$,

(a) find the critical region for the test, in terms of the sample mean \bar{W} , [5]

(b) find the probability that the test results in a Type II error when $\mu = 50.2$. [5]

(ii) State how the probability of this Type II error would change if n were greater than 81. [1]

- 7 A motorist records the time taken, T minutes, to drive a particular stretch of road on each of 64 occasions. Her results are summarised by

$$\Sigma t = 876.8, \quad \Sigma t^2 = 12\,657.28.$$

(i) Test, at the 5% significance level, whether the mean time for the motorist to drive the stretch of road is greater than 13.1 minutes. [11]

(ii) Explain whether it is necessary to use the Central Limit Theorem in your test. [1]

- 8 A sales office employs 21 representatives. Each day, for each representative, the probability that he or she achieves a sale is 0.7, independently of other representatives. The total number of representatives who achieve a sale on any one day is denoted by K .

(i) Using a suitable approximation (which should be justified), find $P(K \geq 16)$. [7]

(ii) Using a suitable approximation (which should be justified), find the probability that the mean of 36 observations of K is less than or equal to 14.0. [7]

June 2009

- 1 The random variable H has the distribution $N(\mu, \sigma^2)$. It is given that $P(H < 105.0) = 0.2420$ and $P(H > 110.0) = 0.6915$. Find the values of μ and σ , giving your answers to a suitable degree of accuracy. [6]
- 2 The random variable D has the distribution $Po(20)$. Using an appropriate approximation, which should be justified, calculate $P(D \geq 25)$. [6]
- 3 An electronics company is developing a new sound system. The company claims that 60% of potential buyers think that the system would be good value for money. In a random sample of 12 potential buyers, 4 thought that it would be good value for money. Test, at the 5% significance level, whether the proportion claimed by the company is too high. [7]
- 4 A survey is to be carried out to draw conclusions about the proportion p of residents of a town who support the building of a new supermarket. It is proposed to carry out the survey by interviewing a large number of people in the high street of the town, which attracts a large number of tourists.
- (i) Give two different reasons why this proposed method is inappropriate. [2]
- (ii) Suggest a good method of carrying out the survey. [3]
- (iii) State two statistical properties of your survey method that would enable reliable conclusions about p to be drawn. [2]
- 5 In a large region of derelict land, bricks are found scattered in the earth.
- (i) State two conditions needed for the number of bricks per cubic metre to be modelled by a Poisson distribution. [2]
- Assume now that the number of bricks in 1 cubic metre of earth can be modelled by the distribution $Po(3)$.
- (ii) Find the probability that the number of bricks in 4 cubic metres of earth is between 8 and 14 inclusive. [3]
- (iii) Find the size of the largest volume of earth for which the probability that no bricks are found is at least 0.4. [4]
- 6 The continuous random variable R has the distribution $N(\mu, \sigma^2)$. The results of 100 observations of R are summarised by
- $$\Sigma r = 3360.0, \quad \Sigma r^2 = 115\,782.84.$$
- (i) Calculate an unbiased estimate of μ and an unbiased estimate of σ^2 . [4]
- (ii) The mean of 9 observations of R is denoted by \bar{R} . Calculate an estimate of $P(\bar{R} > 32.0)$. [4]
- (iii) Explain whether you need to use the Central Limit Theorem in your answer to part (ii). [2]

June 2009

7 The continuous random variable X has probability density function given by

$$f(x) = \begin{cases} \frac{2}{9}x(3-x) & 0 \leq x \leq 3, \\ 0 & \text{otherwise.} \end{cases}$$

(i) Find the variance of X . [5]

(ii) Show that the probability that a single observation of X lies between 0.0 and 0.5 is $\frac{2}{27}$. [2]

(iii) 108 observations of X are obtained. Using a suitable approximation, find the probability that at least 10 of the observations lie between 0.0 and 0.5. [6]

(iv) The mean of 108 observations of X is denoted by \bar{X} . Write down the approximate distribution of \bar{X} , giving the value(s) of any parameter(s). [3]

8 In a large company the time taken for an employee to carry out a certain task is a normally distributed random variable with mean 78.0 s and unknown variance. A new training scheme is introduced and after its introduction the times taken by a random sample of 120 employees are recorded. The mean time for the sample is 76.4 s and an unbiased estimate of the population variance is 68.9 s^2 .

(i) Test, at the 1% significance level, whether the mean time taken for the task has changed. [7]

(ii) It is required to redesign the test so that the probability of making a Type I error is less than 0.01 when the sample mean is 77.0 s. Calculate an estimate of the smallest sample size needed, and explain why your answer is only an estimate. [4]

Jan 2010

- 1 The values of 5 independent observations from a population can be summarised by

$$\Sigma x = 75.8, \quad \Sigma x^2 = 1154.58.$$

Find unbiased estimates of the population mean and variance. [4]

- 2 A college has 400 students. A journalist wants to carry out a survey about food preferences and she obtains a sample of 30 pupils from the college by the following method.

- Obtain a list of all the students.
- Number the students, with numbers running sequentially from 0 to 399.
- Select 30 random integers in the range 000 to 999 inclusive. If a random integer is in the range 0 to 399, then the student with that number is selected. If the number is greater than 399, then 400 is subtracted from the number (if necessary more than once) until an answer in the range 0 to 399 is selected, and the student with that number is selected.

(i) Explain why this method is unsatisfactory. [2]

(ii) Explain how it could be improved. [1]

- 3 In a large town, 35% of the inhabitants have access to television channel *C*. A random sample of 60 inhabitants is obtained. Use a suitable approximation to find the probability that 18 or fewer inhabitants in the sample have access to channel *C*. [6]

- 4 80 randomly chosen people are asked to estimate a time interval of 60 seconds without using a watch or clock. The mean of the 80 estimates is 58.9 seconds. Previous evidence shows that the population standard deviation of such estimates is 5.0 seconds. Test, at the 5% significance level, whether there is evidence that people tend to underestimate the time interval. [7]

- 5 The number of customers arriving at a store between 8.50 am and 9 am on Saturday mornings is a random variable which can be modelled by the distribution $Po(11.0)$. Following a series of price cuts, on one particular Saturday morning 19 customers arrive between 8.50 am and 9 am. The store's management claims, first, that the mean number of customers has increased, and second, that this is due to the price cuts.

(i) Test the first part of the claim, at the 5% significance level. [7]

(ii) Comment on the second part of the claim. [1]

- 6 The continuous random variable X has the distribution $N(\mu, \sigma^2)$.

(i) Each of the three following sets of probabilities is impossible. Give a reason in each case why the probabilities cannot both be correct. (You should not attempt to find μ or σ .)

(a) $P(X > 50) = 0.7$ and $P(X < 50) = 0.2$ [1]

(b) $P(X > 50) = 0.7$ and $P(X > 70) = 0.8$ [1]

(c) $P(X > 50) = 0.3$ and $P(X < 70) = 0.3$ [1]

(ii) Given that $P(X > 50) = 0.7$ and $P(X < 70) = 0.7$, find the values of μ and σ . [4]

Jan 2010

- 7 The continuous random variable T is equally likely to take any value from 5.0 to 11.0 inclusive.
- (i) Sketch the graph of the probability density function of T . [2]
 - (ii) Write down the value of $E(T)$ and find by integration the value of $\text{Var}(T)$. [5]
 - (iii) A random sample of 48 observations of T is obtained. Find the approximate probability that the mean of the sample is greater than 8.3, and explain why the answer is an approximation. [6]
- 8 The random variable R has the distribution $B(10, p)$. The null hypothesis $H_0: p = 0.7$ is to be tested against the alternative hypothesis $H_1: p < 0.7$, at a significance level of 5%.
- (i) Find the critical region for the test and the probability of making a Type I error. [3]
 - (ii) Given that $p = 0.4$, find the probability that the test results in a Type II error. [3]
 - (iii) Given that p is equally likely to take the values 0.4 and 0.7, find the probability that the test results in a Type II error. [2]
- 9 Buttercups in a meadow are distributed independently of one another and at a constant average incidence of 3 buttercups per square metre.
- (i) Find the probability that in 1 square metre there are more than 7 buttercups. [2]
 - (ii) Find the probability that in 4 square metres there are either 13 or 14 buttercups. [3]
 - (iii) Use a suitable approximation to find the probability that there are no more than 69 buttercups in 20 square metres. [5]
 - (iv) (a) Without using an approximation, find an expression for the probability that in m square metres there are at least 2 buttercups. [2]
 - (b) It is given that the probability that there are at least 2 buttercups in m square metres is 0.9. Using your answer to part (a), show numerically that m lies between 1.29 and 1.3. [4]

June 2010

- 1** (i) The number of inhabitants of a village who are selected for jury service in the course of a 10-year period is a random variable with the distribution $Po(4.2)$.
- (a) Find the probability that in the course of a 10-year period, at least 7 inhabitants are selected for jury service. [2]
- (b) Find the probability that in 1 year, exactly 2 inhabitants are selected for jury service. [3]
- (ii) Explain why the number of inhabitants of the village who contract influenza in 1 year can probably not be well modelled by a Poisson distribution. [2]
- 2** A university has a large number of students, of whom 35% are studying science subjects. A sample of 10 students is obtained by listing all the students, giving each a serial number and selecting by using random numbers.
- (i) Find the probability that fewer than 3 of the sample are studying science subjects. [3]
- (ii) It is required that, in selecting the sample, the same student is not selected twice. Explain whether this requirement invalidates your calculation in part (i). [2]
- 3** Tennis balls are dropped from a standard height, and the height of bounce, H cm, is measured. H is a random variable with the distribution $N(40, \sigma^2)$. It is given that $P(H < 32) = 0.2$.
- (i) Find the value of σ . [3]
- (ii) 90 tennis balls are selected at random. Use an appropriate approximation to find the probability that more than 19 have $H < 32$. [6]
- 4** The proportion of commuters in a town who travel to work by train is 0.4. Following the opening of a new station car park, a random sample of 16 commuters is obtained, and 11 of these travel to work by train. Test at the 1% significance level whether there is evidence of an increase in the proportion of commuters in this town who travel to work by train. [7]
- 5** The time T seconds needed for a computer to be ready to use, from the moment it is switched on, is a normally distributed random variable with standard deviation 5 seconds. The specification of the computer says that the population mean time should be not more than 30 seconds.
- (i) A test is carried out, at the 5% significance level, of whether the specification is being met, using the mean \bar{t} of a random sample of 10 times.
- (a) Find the critical region for the test, in terms of \bar{t} . [4]
- (b) Given that the population mean time is in fact 35 seconds, find the probability that the test results in a Type II error. [3]
- (ii) Because of system degradation and memory load, the population mean time μ seconds increases with the number of months of use, m . A formula for μ in terms of m is $\mu = 20 + 0.6m$. Use this formula to find the value of m for which the probability that the test results in rejection of the null hypothesis is 0.5. [4]

June 2010

- 6** (a) The random variable D has the distribution $\text{Po}(24)$. Use a suitable approximation to find $P(D > 30)$. [5]
- (b) An experiment consists of 200 trials. For each trial, the probability that the result is a success is 0.98, independent of all other trials. The total number of successes is denoted by E .
- (i) Explain why the distribution of E cannot be well approximated by a Poisson distribution. [1]
- (ii) By considering the number of failures, use an appropriate Poisson approximation to find $P(E \leq 194)$. [4]

- 7** A machine is designed to make paper with mean thickness 56.80 micrometres. The thicknesses, x micrometres, of a random sample of 300 sheets are summarised by

$$n = 300, \quad \Sigma x = 17\,085.0, \quad \Sigma x^2 = 973\,847.0.$$

Test, at the 10% significance level, whether the machine is producing paper of the designed thickness. [11]

- 8** The continuous random variable X has probability density function given by

$$f(x) = \begin{cases} kx^{-a} & x \geq 1, \\ 0 & \text{otherwise,} \end{cases}$$

where k and a are constants and a is greater than 1.

- (i) Show that $k = a - 1$. [3]
- (ii) Find the variance of X in the case $a = 4$. [5]
- (iii) It is given that $P(X < 2) = 0.9$. Find the value of a , correct to 3 significant figures. [4]

- 1 A random sample of nine observations of a random variable is obtained. The results are summarised as

$$\Sigma x = 468, \quad \Sigma x^2 = 24\,820.$$

Calculate unbiased estimates of the population mean and variance. [4]

- 2 The random variable H has the distribution $N(\mu, 5^2)$. The mean of a sample of n observations of H is denoted by \bar{H} . It is given that $P(\bar{H} > 53.28) = 0.0250$ and $P(\bar{H} < 51.65) = 0.0968$, both correct to 4 decimal places. Find the values of μ and n . [6]

- 3 The probability that a randomly chosen PPhone has a faulty casing is 0.0228. A random sample of 200 PPhones is obtained. Use a suitable approximation to find the probability that the number of PPhones in the sample with a faulty casing is 2 or fewer. Justify your approximation. [6]

- 4 The continuous random variable X has mean μ and standard deviation 45. A significance test is to be carried out of the null hypothesis $H_0: \mu = 230$ against the alternative hypothesis $H_1: \mu \neq 230$, at the 1% significance level. A random sample of size 50 is obtained, and the sample mean is found to be 213.4.

(i) Carry out the test. [5]

(ii) Explain whether it is necessary to use the Central Limit Theorem in your test. [2]

- 5 A temporary job is advertised annually. The number of applicants for the job is a random variable which is known from many years' experience to have a distribution $Po(12)$. In 2010 there were 19 applicants for the job. Test, at the 10% significance level, whether there is evidence of an increase in the mean number of applicants for the job. [7]

- 6 The number of randomly occurring events in a given time interval is denoted by R . In order that R is well modelled by a Poisson distribution, it is necessary that events occur independently.

(i) Let R represent the number of customers dining at a restaurant on a randomly chosen weekday lunchtime. Explain what the condition 'events occur independently' means in this context, and give a reason why it would probably not hold in this context. [2]

Let D represent the number of tables booked at the restaurant on a randomly chosen day. Assume that D can be well modelled by the distribution $Po(7)$.

(ii) Find $P(D < 5)$. [2]

(iii) Use a suitable approximation to find the probability that, in five randomly chosen days, the total number of tables booked is greater than 40. [6]

- 7 Two continuous random variables S and T have probability density functions f_S and f_T given respectively by

$$f_S(x) = \begin{cases} \frac{a}{x^2} & 1 \leq x \leq 3, \\ 0 & \text{otherwise,} \end{cases}$$
$$f_T(x) = \begin{cases} b & 1 \leq x \leq 3, \\ 0 & \text{otherwise,} \end{cases}$$

where a and b are constants.

- (i) Sketch on the same axes the graphs of $y = f_S(x)$ and $y = f_T(x)$. [3]
- (ii) Find the value of a . [3]
- (iii) Find $E(S)$. [3]
- (iv) A student gave the following description of the distribution of T : "The probability that T occurs is constant". Give an improved description, in everyday terms. [1]
- 8 A company has 3600 employees, of whom 22.5% live more than 30 miles from their workplace. A random sample of 40 employees is obtained.
- (i) Use a suitable approximation, which should be justified, to find the probability that more than 5 of the employees in the sample live more than 30 miles from their workplace. [8]
- (ii) Describe how to use random numbers to select a sample of 40 from a population of 3600 employees. [3]
- 9 A pharmaceutical company is developing a new drug to treat a certain disease. The company will continue to develop the drug if the proportion p of those who have the disease and show a substantial improvement after treatment is greater than 0.7. The company carries out a test, at the 5% significance level, on a random sample of 14 patients who suffer from the disease.
- (i) Find the critical region for the test. [3]
- (ii) Given that 12 of the 14 patients in the sample show a substantial improvement, carry out the test. [5]
- (iii) Find the probability that the test results in a Type II error if in fact $p = 0.8$. [3]

- 1 In Fisher Avenue there are 263 houses, numbered 1 to 263. Explain how to obtain a random sample of 20 of these houses. [3]

- 2 The random variable Y has the distribution $N(\mu, \sigma^2)$. It is given that

$$P(Y < 48.0) = P(Y > 57.0) = 0.0668.$$

Find the value y_0 such that $P(Y > y_0) = 0.05$. [7]

- 3 The random variable X has the distribution $N(\mu, 5^2)$. A hypothesis test is carried out of $H_0: \mu = 20.0$ against $H_1: \mu < 20.0$, at the 1% level of significance, based on the mean of a sample of size 16. Given that in fact $\mu = 15.0$, find the probability that the test results in a Type II error. [7]

- 4 A continuous random variable X has probability density function

$$f(x) = \begin{cases} \frac{3}{16}(x-2)^2 & 0 \leq x \leq 4, \\ 0 & \text{otherwise.} \end{cases}$$

(i) Sketch the graph of $y = f(x)$. [2]

(ii) Calculate the variance of X . [5]

(iii) A student writes “ X is more likely to occur when x takes values further away from 2”. Explain whether you agree with this statement. [1]

- 5 A travel company finds from its records that 40% of its customers book with travel agents. The company redesigns its website, and then carries out a survey of 10 randomly chosen customers. The result of the survey is that 1 of these customers booked with a travel agent.

(i) Test at the 5% significance level whether the percentage of customers who book with travel agents has decreased. [7]

(ii) The managing director says that “Our redesigned website has resulted in a decrease in the percentage of our customers who book with travel agents.” Comment on this statement. [1]

- 6 Records show that before the year 1990 the maximum daily temperature $T^\circ\text{C}$ at a seaside resort in August can be modelled by a distribution with mean 24.3. The maximum temperatures of a random sample of 50 August days since 1990 can be summarised by

$$n = 50, \quad \Sigma t = 1314.0, \quad \Sigma t^2 = 36\,602.17.$$

(i) Test, at the 1% significance level, whether there is evidence of a change in the mean maximum daily temperature in August since 1990. [11]

(ii) Give a reason why it is possible to use the Central Limit Theorem in your test. [1]

7 The number of customer complaints received by a company per day is denoted by X . Assume that X has the distribution $Po(2.2)$.

(i) In a week of 5 working days, the probability there are at least n customer complaints is 0.146 correct to 3 significant figures. Use tables to find the value of n . [3]

(ii) Use a suitable approximation to find the probability that in a period of 20 working days there are fewer than 38 customer complaints. [5]

A week of 5 working days in which at least n customer complaints are received, where n is the value found in part (i), is called a 'bad' week.

(iii) Use a suitable approximation to find the probability that, in 40 randomly chosen weeks, more than 7 are bad. [6]

8 (a) A group of students is discussing the conditions that are needed if a Poisson distribution is to be a good model for the number of telephone calls received by a fire brigade on a working day.

(i) Alice says "Events must be independent". Explain why this condition may not hold in this context. [1]

(ii) State a different condition that is needed. Explain whether it is likely to hold in this context. [2]

(b) The random variables R , S and T have independent Poisson distributions with means λ , μ and $\lambda + \mu$ respectively.

(i) In the case $\lambda = 2.74$, find $P(R > 2)$. [3]

(ii) In the case $\lambda = 2$ and $\mu = 3$, find $P(R = 0 \text{ and } S = 1) + P(R = 1 \text{ and } S = 0)$. Give your answer correct to 4 decimal places. [3]

(iii) In the general case, show algebraically that

$$P(R = 0 \text{ and } S = 1) + P(R = 1 \text{ and } S = 0) = P(T = 1). \quad [4]$$

- 1 A random sample of 50 observations of the random variable X is summarised by

$$n = 50, \Sigma x = 182.5, \Sigma x^2 = 739.625.$$

Calculate unbiased estimates of the expectation and variance of X . [4]

- 2 The random variable Y has the distribution $B(140, 0.03)$. Use a suitable approximation to find $P(Y = 5)$. Justify your approximation. [5]

- 3 The random variable G has a normal distribution. It is known that

$$P(G < 56.2) = P(G > 63.8) = 0.1.$$

Find $P(G > 65)$. [6]

- 4 The discrete random variable H takes values 1, 2, 3 and 4. It is given that $E(H) = 2.5$ and $\text{Var}(H) = 1.25$. The mean of a random sample of 50 observations of H is denoted by \bar{H} . Use a suitable approximation to find $P(\bar{H} < 2.6)$. [5]

- 5 (i) Six prizes are allocated, using random numbers, to a group of 12 girls and 8 boys. Calculate the probability that exactly 4 of the prizes are allocated to girls if

(a) the same child may win more than one prize, [2]

(b) no child may win more than one prize. [2]

- (ii) Sixty prizes are allocated, using random numbers, to a group of 1200 girls and 800 boys. Use a suitable approximation to calculate the probability that at least 30 of the prizes are allocated to girls. Does it affect your calculation whether or not the same child may win more than one prize? Justify your answer. [6]

- 6 The number of fruit pips in 1 cubic centimetre of raspberry jam has the distribution $\text{Po}(\lambda)$. Under a traditional jam-making process it is known that $\lambda = 6.3$. A new process is introduced and a random sample of 1 cubic centimetre of jam produced by the new process is found to contain 2 pips. Test, at the 5% significance level, whether this is evidence that under the new process the average number of pips has been reduced. [8]

- 7 (i) The continuous random variable X has the probability density function

$$f(x) = \begin{cases} \frac{1}{2\sqrt{x}} & 1 \leq x \leq 4, \\ 0 & \text{otherwise.} \end{cases}$$

Find (a) $E(X)$, [3]

(b) the median of X . [3]

- (ii) The continuous random variable Y has the probability density function

$$g(y) = \begin{cases} \frac{1.5}{y^{2.5}} & y \geq 1, \\ 0 & \text{otherwise.} \end{cases}$$

Given that $E(Y) = 3$, show that $\text{Var}(Y)$ is not finite. [3]

- 8 In a certain fluid, bacteria are distributed randomly and occur at a constant average rate of 2.5 in every 10 ml of the fluid.

- (i) State a further condition needed for the number of bacteria in a fixed volume of the fluid to be well modelled by a Poisson distribution, explaining what your answer means. [2]

Assume now that a Poisson model is appropriate.

(ii) Find the probability that in 10 ml there are at least 5 bacteria. [2]

(iii) Find the probability that in 3.7 ml there are exactly 2 bacteria. [3]

(iv) Use a suitable approximation to find the probability that in 1000 ml there are fewer than 240 bacteria, justifying your approximation. [7]

- 9 It is desired to test whether the average amount of sleep obtained by school pupils in Year 11 is 8 hours, based on a random sample of size 64. The population standard deviation is 0.87 hours and the sample mean is denoted by \bar{H} . The critical values for the test are $\bar{H} = 7.72$ and $\bar{H} = 8.28$.

(i) State appropriate hypotheses for the test, explaining the meaning of any symbol you use. [3]

(ii) Calculate the significance level of the test. [4]

(iii) Explain what is meant by a Type I error in this context. [1]

(iv) Given that in fact the average amount of sleep obtained by all pupils in Year 11 is 7.9 hours, find the probability that the test results in a Type II error. [3]

- 1 In one day's production, a machine produces 1000 CDs. Explain how to take a random sample of 15 CDs chosen from one day's production. [2]
- 2 (i) For the continuous random variable V , it is known that $E(V) = 72.0$. The mean of a random sample of 40 observations of V is denoted by \bar{V} . Given that $P(\bar{V} < 71.2) = 0.35$, estimate the value of $\text{Var}(V)$. [4]
- (ii) Explain why you need to use the Central Limit Theorem in part (i), and why its use is justified. [2]
- 3 It is known that on average one person in three prefers the colour of a certain object to be blue. In a psychological test, 12 randomly chosen people were seated in a room with blue walls, and asked to state independently which colour they preferred for the object. Seven of the 12 people said that they preferred blue. Carry out a significance test, at the 5% level, of whether the statement "on average one person in three prefers the colour of the object to be blue" is true for people who are seated in a room with blue walls. [7]
- 4 In a rock, small crystal formations occur at a constant average rate of 3.2 per cubic metre.
- (i) State a further assumption needed to model the number of crystal formations in a fixed volume of rock by a Poisson distribution. [1]
- In the remainder of the question, you should assume that a Poisson model is appropriate.
- (ii) Calculate the probability that in one cubic metre of rock there are exactly 5 crystal formations. [2]
- (iii) Calculate the probability that in 0.74 cubic metres of rock there are at least 3 crystal formations. [3]
- (iv) Use a suitable approximation to calculate the probability that in 10 cubic metres of rock there are at least 36 crystal formations. [5]
- 5 The acidity A (measured in pH) of soil of a particular type has a normal distribution. The pH values of a random sample of 80 soil samples from a certain region can be summarised as
- $$\Sigma a = 496, \quad \Sigma a^2 = 3126.$$
- Test, at the 10% significance level, whether in this region the mean pH of soil is 6.1. [11]
- 6 At a tourist car park, a survey is made of the regions from which cars come.
- (i) It is given that 40% of cars come from the London region. Use a suitable approximation to find the probability that, in a random sample of 32 cars, more than 17 come from the London region. Justify your approximation. [7]
- (ii) It is given that 1% of cars come from France. Use a suitable approximation to find the probability that, in a random sample of 90 cars, exactly 3 come from France. [4]

- 7 The continuous random variable X has probability density function

$$f(x) = \begin{cases} kx^2 & 0 \leq x \leq a, \\ 0 & \text{otherwise,} \end{cases}$$

where a and k are constants.

- (i) Sketch the graph of $y = f(x)$ and explain in non-technical language what this tells you about X . [3]
- (ii) Given that $E(X) = 4.5$, find
- (a) the value of a , [6]
- (b) $\text{Var}(X)$. [3]
- 8 The random variable X has the distribution $N(\mu, 8^2)$. A test is carried out, at the 5% significance level, of $H_0: \mu = 30$ against $H_1: \mu > 30$, based on a random sample of size 18.

- (i) Find the critical region for the test. [4]
- (ii) If $\mu = 30$ and the outcome of the test is that H_0 is rejected, state the type of error that is made. [1]

On a particular day this test is carried out independently a total of 20 times, and for 4 of these tests the outcome is that H_0 is rejected. It is known that the value of μ remains the same throughout these 20 tests.

- (iii) Find the probability that H_0 is rejected at least 4 times if $\mu = 30$. Hence state whether you think that $\mu = 30$, giving a reason. [3]
- (iv) Given that the probability of making an error of the type different from that stated in part (ii) is 0.4, calculate the actual value of μ , giving your answer correct to 4 significant figures. [4]

- 1 A random variable has the distribution $B(n, p)$. It is required to test $H_0: p = \frac{2}{3}$ against $H_1: p < \frac{2}{3}$ at a significance level as close to 1% as possible, using a sample of size $n = 8, 9$ or 10 . Use tables to find which value of n gives such a test, stating the critical region for the test and the corresponding significance level. [4]

- 2 A random variable C has the distribution $N(\mu, \sigma^2)$. A random sample of 10 observations of C is obtained, and the results are summarised as

$$n = 10, \sum c = 380, \sum c^2 = 14602.$$

- (i) Calculate unbiased estimates of μ and σ^2 . [4]

- (ii) Hence calculate an estimate of the probability that $C > 40$. [2]

- 3 A factory produces 9000 music DVDs each day. A random sample of 100 such DVDs is obtained.

- (i) Explain how to obtain this sample using random numbers. [3]

- (ii) Given that 24% of the DVDs produced by the factory are classical, use a suitable approximation to find the probability that, in the sample of 100 DVDs, fewer than 20 are classical. [5]

- 4 A continuous random variable X has probability density function

$$f(x) = \begin{cases} kx & 0 \leq x \leq a, \\ 0 & \text{otherwise,} \end{cases}$$

where k and a are constants.

- (i) State what the letter x represents. [1]

- (ii) Find k in terms of a . [2]

- (iii) Find $\text{Var}(X)$ in terms of a . [6]

- 5 In a mine, a deposit of the substance *pitchblende* emits radioactive particles. The number of particles emitted has a Poisson distribution with mean 70 particles per second. The warning level is reached if the total number of particles emitted in one minute is more than 4350.

- (i) A one-minute period is chosen at random. Use a suitable approximation to show that the probability that the warning level is reached during this period is 0.01, correct to 2 decimal places. You should calculate the answer correct to 4 decimal places. [5]

- (ii) Use a suitable approximation to find the probability that in 30 one-minute periods the warning level is reached on at least 4 occasions. (You should use the given rounded value of 0.01 from part (i) in your calculation.) [3]

- 6 Gordon is a cricketer. Over a long period he knows that his population mean score, in number of runs per innings, is 28, and the population standard deviation is 12. In a new season he adopts a different batting style and he finds that in 30 innings using this style his mean score is 28.98.

- (i) Stating a necessary assumption, test at the 5% significance level whether his population mean score has increased. [8]

- (ii) Explain whether it was necessary to use the Central Limit Theorem in part (i). [2]

- 7 The continuous random variable X has the distribution $N(\mu, \sigma^2)$. The mean of a random sample of n observations of X is denoted by \bar{X} . It is given that $P(\bar{X} < 35.0) = 0.9772$ and $P(\bar{X} < 20.0) = 0.1587$.

(i) Obtain a formula for σ in terms of n . [5]

Two students are discussing this question. Aidan says “If you were told another probability, for instance $P(\bar{X} > 32) = 0.1$, you could work out the value of σ .” Binya says, “No, the value of $P(\bar{X} > 32)$ is fixed by the information you know already.”

(ii) State which of Aidan and Binya is right. If you think that Aidan is right, calculate the value of σ given that $P(\bar{X} > 32) = 0.1$. If you think that Binya is right, calculate the value of $P(\bar{X} > 32)$. [4]

- 8 In a large city the number of traffic lights that fail in one day of 24 hours is denoted by Y . It may be assumed that failures occur randomly.

(i) Explain what the statement “failures occur randomly” means. [1]

(ii) State, in context, two different conditions that must be satisfied if Y is to be modelled by a Poisson distribution, and for each condition explain whether you think it is likely to be met in this context. [4]

(iii) For this part you may assume that Y is well modelled by the distribution $Po(\lambda)$. It is given that $P(Y = 7) = P(Y = 8)$. Use an algebraic method to calculate the value of λ and hence calculate the corresponding value of $P(Y = 7)$. [5]

- 9 The random variable A has the distribution $B(30, p)$. A test is carried out of the hypotheses $H_0: p = 0.6$ against $H_1: p < 0.6$. The critical region is $A \leq 13$.

(i) State the probability that H_0 is rejected when $p = 0.6$. [1]

(ii) Find the probability that a Type II error occurs when $p = 0.5$. [2]

(iii) It is known that on average $p = 0.5$ on one day in five, and on other days the value of p is 0.6. On each day two tests are carried out. If the result of the first test is that H_0 is rejected, the value of p is adjusted if necessary, to ensure that $p = 0.6$ for the rest of the day. Otherwise the value of p remains the same as for the first test. Calculate the probability that the result of the second test is to reject H_0 . [5]

June 2013

- 1 It is required to select a random sample of 30 pupils from a school with 853 pupils. A student suggests the following method.

“Give each pupil sequentially a three-digit number from 001 to 853. Use a calculator to generate random three-digit numbers from 0.000 to 0.999 inclusive, multiply the answer by 853, add 1 and round off to the nearest whole number. Select the corresponding pupil, and repeat as necessary”.

- (i) Determine which pupil would be picked for each of the following calculator outputs:

0.103, 0.104, 0.105, 0.106, 0.107. [2]

- (ii) Use your answers to part (i) to show that this method is biased, and suggest an improvement. [2]

- 2 The number of neutrinos that pass through a certain region in one second is a random variable with the distribution $Po(5 \times 10^4)$. Use a suitable approximation to calculate the probability that the number of neutrinos passing through the region in 40 seconds is less than 1.999×10^6 . [4]

- 3 The mean of a sample of 80 independent observations of a continuous random variable Y is denoted by \bar{Y} . It is given that $P(\bar{Y} \leq 157.18) = 0.1$ and $P(\bar{Y} \geq 164.76) = 0.7$.

- (i) Calculate $E(Y)$ and the standard deviation of Y . [6]

- (ii) State

(a) where in your calculations you have used the Central Limit Theorem,

(b) why it was necessary to use the Central Limit Theorem,

(c) why it was possible to use the Central Limit Theorem. [3]

- 4 The number of floods in a certain river plain is known to have a Poisson distribution. It is known that up until 10 years ago the mean number of floods per year was 0.32. During the last 10 years there were 6 floods. Test at the 1% significance level whether there is evidence of an increase in the mean number of floods per year. [7]

- 5 Two random variables S and T have probability density functions given by

$$f_S(x) = \begin{cases} \frac{3}{a^3}(x-a)^2 & 0 \leq x \leq a, \\ 0 & \text{otherwise,} \end{cases}$$

$$f_T(x) = \begin{cases} c & 0 \leq x \leq a, \\ 0 & \text{otherwise,} \end{cases}$$

where a and c are constants.

- (i) On a single diagram sketch both probability density functions. [3]

- (ii) Calculate the mean of S , in terms of a . [5]

- (iii) Use your diagram to explain which of S or T has the bigger variance. (Answers obtained by calculation will score no marks.) [2]

- 6 The random variable X denotes the yield, in kilograms per acre, of a certain crop. Under the standard treatment it is known that $E(X) = 38.4$. Under a new treatment, the yields of 50 randomly chosen regions can be summarised as

$$n = 50, \quad \Sigma x = 1834.0, \quad \Sigma x^2 = 70027.37.$$

Test at the 1% level whether there has been a change in the mean crop yield. [11]

- 7 Past experience shows that 35% of the senior pupils in a large school know the regulations about bringing cars to school. The head teacher addresses this subject in an assembly, and afterwards a random sample of 120 senior pupils is selected. In this sample it is found that 50 of these pupils know the regulations. Use a suitable approximation to test, at the 10% significance level, whether there is evidence that the proportion of senior pupils who know the regulations has increased. Justify your approximation. [11]

- 8 The random variable R has the distribution $B(14, p)$. A test is carried out at the $\alpha\%$ significance level of the null hypothesis $H_0: p = 0.25$, against $H_1: p > 0.25$.

(i) Given that α is as close to 5 as possible, find the probability of a Type II error when the true value of p is 0.4. [4]

(ii) State what happens to the probability of a Type II error as

(a) p increases from 0.4,

(b) α increases, giving a reason. [2]

- 9 The managers of a car breakdown recovery service are discussing whether the number of breakdowns per day can be modelled by a Poisson distribution. They agree that breakdowns occur randomly. Manager A says, "it must be assumed that breakdowns occur at a constant rate throughout the day".

(i) Give an improved version of Manager A 's statement, and explain why the improvement is necessary. [2]

(ii) Explain whether you think your improved statement is likely to hold in this context. [1]

Assume now that the number B of breakdowns per day can be modelled by the distribution $Po(\lambda)$.

(iii) Given that $\lambda = 9.0$ and $P(B > B_0) < 0.1$, use tables to find the smallest possible value of B_0 , and state the corresponding value of $P(B > B_0)$. [2]

(iv) Given that $P(B = 2) = 0.0072$, show that λ satisfies an equation of the form $\lambda = 0.12e^{k\lambda}$, for a value of k to be stated. Evaluate the expression $0.12e^{k\lambda}$ for $\lambda = 8.5$ and $\lambda = 8.6$, giving your answers correct to 4 decimal places. What can be deduced about a possible value of λ ? [5]

- 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	$G < 40.0$	$40.0 \leq G < 60.0$	$G \geq 60.0$
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]

June 2014

- 5 A continuous random variable X has probability density function

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

- (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

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

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.

June 2014

- 8 The random variable W has the distribution $Po(\lambda)$. 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**Copyright Information**

OCR is committed to seeking permission to reproduce all third-party content that it uses in its assessment materials. OCR has attempted to identify and contact all copyright holders whose work is used in this paper. To avoid the issue of disclosure of answer-related information to candidates, all copyright acknowledgements are reproduced in the OCR Copyright Acknowledgements Booklet. This is produced for each series of examinations and is freely available to download from our public website (www.ocr.org.uk) after the live examination series.

If OCR has unwittingly failed to correctly acknowledge or clear any third-party content in this assessment material, OCR will be happy to correct its mistake at the earliest possible opportunity.

For queries or further information please contact the Copyright Team, First Floor, 9 Hills Road, Cambridge CB2 1GE.

OCR is part of the Cambridge Assessment Group; Cambridge Assessment is the brand name of University of Cambridge Local Examinations Syndicate (UCLES), which is itself a department of the University of Cambridge.