S2 QUESTIONS

TAKEN FROM JANUARY 2006, JANUARY 2007, JANUARY 2008, JANUARY 2009

SECTION 1

<u>The binomial and Poisson distributions.</u> Students will be expected to use these distributions to model a real-world situation and to comment critically on their appropriateness. Cumulative probabilities by calculation or by reference to tables. Students will be expected to use the additive property of the Poisson distribution — eg if the number of events per minute $\sim \text{Po}(\lambda)$ then the number of events per 5 minutes $\sim \text{Po}(5\lambda)$. <u>The mean and variance of the binomial and Poisson distributions.</u> No derivations will be required.

1. A fair coin is tossed 4 times.
Find the probability that
(*a*) an equal number of head and tails occur
(*b*) all the outcomes are the same,
(*c*) the first tail occurs on the third throw.

2. Accidents on a particular stretch of motorway occur at an average rate of 1.5 per week.
(*a*) Write down a suitable model to represent the number of accidents per week on this stretch of motorway.
Find the probability that
(*b*) there will be 2 accidents in the same week,
(*c*) there is at least one accident per week for 3 consecutive weeks,
(*d*) there are more than 4 accidents in a 2 week period.

3. The random variable J has a Poisson distribution with mean 4.
(*a*) Find $\text{P}(J \geq 10)$.
The random variable K has a binomial distribution with parameters $n = 25$, $p = 0.27$.
(*b*) Find $\text{P}(K \leq 1)$.

4. The probability of a bolt being faulty is 0.3. Find the probability that in a random sample of 20 bolts there are
(*a*) exactly 2 faulty bolts,
(*b*) more than 3 faulty bolts.
These bolts are sold in bags of 20. John buys 10 bags.
(*c*) Find the probability that exactly 6 of these bags contain more than 3 faulty bolts.

5. (*a*) State two conditions under which a Poisson distribution is a suitable model to use in statistical work.

The number of cars passing an observation point in a 10 minute interval is modelled by a Poisson distribution with mean 1.

(*b*) Find the probability that in a randomly chosen 60 minute period there will be

(i) exactly 4 cars passing the observation point,

(ii) at least 5 cars passing the observation point.

The number of other vehicles, other than cars, passing the observation point in a 60 minute interval is modelled by a Poisson distribution with mean 12.

(*c*) Find the probability that exactly 1 vehicle, of any type, passes the observation point in a 10 minute period.

6. A botanist is studying the distribution of daisies in a field. The field is divided into a number of equal sized squares. The mean number of daisies per square is assumed to be 3. The daisies are distributed randomly throughout the field. Find the probability that, in a randomly chosen square there will be

(*a*) more than 2 daisies,

(*b*) either 5 or 6 daisies.

The botanist decides to count the number of daisies, *x*, in each of 80 randomly selected squares within the field. The results are summarised below

$$\sum x = 295 \qquad \sum x^2 = 1386$$

(*c*) Calculate the mean and the variance of the number of daisies per square for the 80 squares. Give your answers to 2 decimal places.

(*d*) Explain how the answers from part (*c*) support the choice of a Poisson distribution as a model.

(*e*) Using your mean from part (*c*), estimate the probability that exactly 4 daisies will be found in a randomly selected square.

Answers

1 a) 0.375 b) 0.125 c) 0.125

2 a) Po(1.5) b) 0.2510 c) 0.4689 d) 0.1847

3 a) 0.0081 b) 0.00392

4 a) 0.0279 b) 0.8929 c) 0.0140

5 a) Any two of: i) Events occur at a constant rate ii) Events occur independently or randomly iii) Events occur singly. b) i) 0.1339 ii) 0.7149 c) 0.149

6. a) 0.5768 b) 0.1512 c) 3.69, 3.72 d) For a Poisson model , Mean = Variance ; For these data 3.69 $\approx$ 3.73 $\Rightarrow$ Poisson model e) 0.193

SECTION 2

<u>The use of the Poisson distribution as an approximation to the binomial distribution.</u>

<u>Use of the Normal distribution as an approximation to the binomial distribution and the Poisson distribution, with the application of the continuity correction.</u>

7. For a particular type of plant 45% have white flowers and the remainder have coloured flowers. Gardenmania sells plants in batches of 12. A batch is selected at random.
Calculate the probability this batch contains
(*a*) exactly 5 plants with white flowers,
(*b*) more plants with white flowers than coloured ones.
Gardenmania takes a random sample of 10 batched of plants.
(*c*) Find the probability that exactly 3 of these batches contain more plants with white flowers than coloured ones.
Due to an increasing demand for these plants by large companies, Gardenmania decides to sell them in batches of 50.
(*d*) Use a suitable approximation to calculate the probability that a batch of 50 plants contains more than 25 plants with white flowers.

8. (*a*) State the condition under which the normal distribution may be used as an approximation to the Poisson distribution.
(*b*) Explain why a continuity correction must be incorporated when using the normal distribution as an approximation to the Poisson distribution.
A company has yachts that can only be hired for a week at a time. All hiring starts on a Saturday. During the winter the mean number of yachts hired per week is 5.
(*c*) Calculate the probability that fewer than 3 yachts are hired on a particular Saturday in winter.
During the summer the mean number of yachts hired per week increases to 25. The company has only 30 yachts for hire.
(*d*) Using a suitable approximation find the probability that the demand for yachts cannot be met on a particular Saturday in summer.
In the summer there are 16 Saturdays on which a yacht can be hired.
(*e*) Estimate the number of Saturdays in the summer that the company will not be able to meet the demand for yachts.

9. The random variable $X \sim \mathrm{B}(150, 0.02)$.
Use a suitable approximation to estimate $\mathrm{P}(X > 7)$.

10. The probability that a sunflower plant grows over 1.5 metres high is 0.25. A random sample of 40 sunflower plants is taken and each sunflower plant is measured and its height recorded.

(*a*) Find the probability that the number of sunflower plants over 1.5 m high is between 8 and 13 (inclusive) using

(i) a Poisson approximation,

(ii) a Normal approximation.

(*b*) Write down which of the approximations used in part (*a*) is the most accurate estimate of the probability. You must give a reason for your answer.

11. A factory produces components of which 1% are defective. The components are packed in boxes of 10. A box is selected at random.

(*a*) Find the probability that the box contains exactly one defective component.

(*b*) Find the probability that there are at least 2 defective components in the box.

(*c*) Using a suitable approximation, find the probability that a batch of 250 components contains between 1 and 4 (inclusive) defective components.

Answers

7. a) 0.2225 b) 0.2607 c) 0.257 d) 0.1977

8. a) $\lambda > 10$ or large b) The Poisson is discrete and the normal is continuous c) 0.1247
d) 0.1357 e) 2.17 or 2 or 3

9. Po(3), 0.0119

10. a) i) 0.6443 ii) 0.7183 b) Normal approx /not Poisson since (n is large) and p close to half.
or (np = 10 npq = 7.5) mean ≠ variance **or** np (= 10) and nq (= 30) both >5 **or** exact binomial
= 0.7148

11. a) 0.0914 b) 0.0043 c) 0.809

SECTION 3

The concept of a continuous random variable.

The probability density function and the cumulative distribution function for a continuous random variable.

Use of the probability density function f(x), where P($a < X \le b$) = $\int_a^b f(x)dx$

Use of the cumulative distribution function F(x_0) = P($X \le x_0$) = $\int_{-\infty}^{xo} f(x)dx$

The formulae used in defining f(x) will be restricted to simple polynomials which may be expressed piecewise.

Relationship between density and distribution functions.

f(x) =$\frac{d(F(x)}{dx}$

Mean and variance of continuous random variables.

Mode, median and quartiles of continuous random variables.

The continuous uniform (rectangular) distribution.

Including the derivation of the mean, variance and cumulative distribution function.

12. The random variable X is uniformly distributed over the interval [–1, 5].

(a) Sketch the probability density function f(x) of X.

Find

(b) E(X),

(c) Var(X),

(d) P(–0.3 < X < 3.3).

13. A continuous random variable X has probability density function f(x) where

$$f(x) = \begin{cases} kx(x-2), & 2 \le x \le 3, \\ 0, & \text{otherwise,} \end{cases}$$

where k is a positive constant.

(a) Show that $k = \frac{3}{4}$.

Find

(b) E(X),

(c) the cumulative distribution function F(x).

(d) Show that the median value of X lies between 2.70 and 2.75.

14. The continuous random variable X has cumulative distribution function

$$\mathrm{F}(x) = \begin{cases} 0, & x < 0, \\ 2x^2 - x^3, & 0 \le x \le 1, \\ 1, & x > 1. \end{cases}$$

(*a*) Find $\mathrm{P}(X > 0.3)$.
(*b*) Verify that the median value of X lies between $x = 0.59$ and $x = 0.60$.
(*c*) Find the probability density function $\mathrm{f}(x)$.
(*d*) Evaluate $\mathrm{E}(X)$.
(*e*) Find the mode of X.
(*f*) Comment on the skewness of X. Justify your answer.

15. The continuous random variable Y has cumulative distribution function $\mathrm{F}(y)$ given by

$$\mathrm{F}(y) = \begin{cases} 0 & y < 1 \\ k(y^4 + y^2 - 2) & 1 \le y \le 2 \\ 1 & y > 2 \end{cases}$$

(*a*) Show that $k = \frac{1}{18}$.
(*b*) Find $\mathrm{P}(Y > 1.5)$.
(*c*) Specify fully the probability density function $\mathrm{f}(y)$.

16. The continuous random variable X is uniformly distributed over the interval $\alpha < x < \beta$.
(*a*) Write down the probability density function of X, for all x.
(*b*) Given that $\mathrm{E}(X) = 2$ and $\mathrm{P}(X < 3) = \frac{5}{8}$, find the value of α and the value of β.

A gardener has wire cutters and a piece of wire 150 cm long which has a ring attached at one end. The gardener cuts the wire, at a randomly chosen point, into 2 pieces. The length, in cm, of the piece of wire with the ring on it is represented by the random variable X. Find
(*c*) $\mathrm{E}(X)$,
(*d*) the standard deviation of X,
(*e*) the probability that the shorter piece of wire is at most 30 cm long.

17. The continuous random variable X has probability density function $\mathrm{f}(x)$ given by

$$\mathrm{f}(x) = \begin{cases} 2(x-2) & 2 \le x \le 3 \\ 0 & \text{otherwise} \end{cases}$$

(*a*) Sketch $\mathrm{f}(x)$ for all values of x.
(*b*) Write down the mode of X.
Find
(*c*) $\mathrm{E}(X)$,
(*d*) the median of X.
(*e*) Comment on the skewness of this distribution. Give a reason for your answer.

18. The continuous random variable X is uniformly distributed over the interval [–2, 7].

(a) Write down fully the probability density function f(x) of X.

(b) Sketch the probability density function f(x) of X.

Find

(c) E(X^2),

(d) P(– 0.2 < X < 0.6).

19. A random variable X has probability density function given by

$$f(x) = \begin{cases} -\frac{2}{9}x + \frac{8}{9}, & 1 \le x \le 4 \\ 0, & \text{otherwise.} \end{cases}$$

(a) Show that the cumulative distribution function F(x) can be written in the form $ax^2 + bx + c$, for $1 \le x \le 4$ where a, b and c are constants.

(b) Define fully the cumulative distribution function F(x).

(c) Show that the upper quartile of X is 2.5 and find the lower quartile.

Given that the median of X is 1.88,

(d) describe the skewness of the distribution. Give a reason for your answer.

20. The length of a telephone call made to a company is denoted by the continuous random variable T. It is modelled by the probability density function

$$f(x) = \begin{cases} kt, & 0 \le t \le 10 \\ 0, & \text{otherwise.} \end{cases}$$

(a) Show that the value of k is $\dfrac{1}{50}$.

(b) Find P(T > 6).

(c) Calculate an exact value for E(T) and for Var(T).

(d) Write down the mode of the distribution of T.

It is suggested that the probability density function, f(t), is not a good model for T.

(e) Sketch the graph of a more suitable probability density function for T.

Answers

12. b) 2 c) 3 d) 0.6

13 a) Proof b) 2.69 c) F(x) = 0 ($x \leq 2$), F(x) = $\frac{1}{4}(x^3-3x^2+4)$ ($2<x<3$), F(x) = 1 ($x\geq3$) d) Proof

14 a) 0.847 b) Proof c) f(x) = $-3x^2 + 4x$ ($0\leq x\leq1$), f(x) 0 otherwise d) $^7/_{12}$ e)$^2/_3$ f) mean < median < mode, therefore negative skew.

15 a) $^1/_{18}$ b) 0.705 c) f(y) = $^1/_9(2y^3 + 2y)$, $1 \leq y \leq 2$; f(y) = 0 otherwise

16 a) f(x) = $\frac{1}{\beta-\alpha}$ $\alpha < x < \beta$, f(x) = 0 otherwise b) α=-2, β=6 c) 75 d) 43.3 e) 0.4

17 b) 3 c) $2^2/_3$ d)2.71 e) Negative skew because mean < median < mode

18. a) f(x) = $^1/_9$ ($-2 \leq x \leq 7$), f(x) = 0 otherwise c) 6.75 c) 13 d) 0.0889

19. a) a = $^{-1}/_9$, b = $^8/_9$, c = $^{-7}/_9$ b) F(x) = 0 ($x < 1$), F(x) = $^{-x^2}/_9$ + $^{8x}/_9$ $^{-7}/_9$ ($1 \leq x \leq 4$), F(x) = 1 ($x>4$) c) 1.40 d) mode < median < mode; positive skew

20. b) $^{16}/_{25}$ c) 6 $^2/_3$, 5 $^5/_9$ d) 10

SECTION 4

<u>Population, census and sample. Sampling unit, sampling frame.</u>
Students will be expected to know the advantages and disadvantages associated with a census and a sample survey.

<u>Concepts of a statistic and its sampling distribution.</u>

21. A bag contains a large number of coins. Half of them are 1p coins, one third are 2p coins and the remainder are 5p coins.
(*a*) Find the mean and variance of the value of the coins.
A random sample of 2 coins is chosen from the bag.
(*b*) List all the possible samples that can be drawn.
(*c*) Find the sampling distribution of the mean value of these samples.

22. (*a*) Define a statistic.
A random sample $X_1, X_2, ..., X_n$ is taken from a population with unknown mean μ.
(*b*) For each of the following state whether or not it is a statistic.

(i) $\frac{X_1 + X_4}{2}$,

(ii) $\frac{\sum X^2}{n} - \mu^2$.

23. (*a*) Explain what you understand by a census.
Each cooker produced at GT Engineering is stamped with a unique serial number. GT Engineering produces cookers in batches of 2000. Before selling them, they test a random sample of 5 to see what electric current overload they will take before breaking down.
(*b*) Give one reason, other than to save time and cost, why a sample is taken rather than a census.
(*c*) Suggest a suitable sampling frame from which to obtain this sample.
(*d*) Identify the sampling units.

<u>Answers</u>

21.a) mean = 2, variance = 2 b) (1,1) (1,2) (2,1) (1,5) (5,1) (2,2) (2,5) (5,2) (5,5)
c) p(X= 1) =$^1/_4$, p(X=1.5)=$^1/_3$, p(X=2)=$^1/_9$, p(X=3)= $^1/_6$, p(X=3.5)= $^1/_9$, p(X=5)= $^1/_{36}$

22 a) A random variable; a function of known observations (from a population). b) i) Yes ii) No

23 a) A census is when every member of the population is investigated. b) There would be no cookers left to sell. c) A list of the unique identification numbers of the cookers. d) A cooker

SECTION 5

<u>Concept and interpretation of a hypothesis test. Null and alternative hypotheses.</u>
Use of hypothesis tests for refinement of mathematical models.

<u>Critical region.</u>
Use of a statistic as a test statistic.

<u>One-tailed and two-tailed tests.</u>

<u>Hypothesis tests for the parameter *p* of a binomial distribution and for the mean of a Poisson distribution.</u>
Students are expected to know how to use tables to carry out these tests. Questions may also be set not involving tabular values. Tests on sample proportion involving the normal approximation will not be set.

24. A teacher thinks that 20% of the pupils in a school read the Deano comic regularly.
He chooses 20 pupils at random and finds 9 of them read the Deano.

(*a*) (i) Test, at the 5% level of significance, whether or not there is evidence that the percentage of pupils that read the Deano is different from 20%. State your hypotheses clearly.

(ii) State all the possible numbers of pupils that read the Deano from a sample of size 20 that will make the test in part (*a*)(i) significant at the 5% level.

The teacher takes another 4 random samples of size 20 and they contain 1, 3, 1 and 4 pupils that read the Deano.

(*b*) By combining all 5 samples and using a suitable approximation test, at the 5% level of significance, whether or not this provides evidence that the percentage of pupils in the school that read the Deano is different from 20%.

(*c*) Comment on your results for the tests in part (*a*) and part (*b*).

25. Past records from a large supermarket show that 20% of people who buy chocolate bars buy the family size bar. On one particular day a random sample of 30 people was taken from those that had bought chocolate bars and 2 of them were found to have bought a family size bar.

(*a*) Test, at the 5% significance level, whether or not the proportion *p* of people who bought a family size bar of chocolate that day had decreased. State your hypotheses clearly.

The manager of the supermarket thinks that the probability of a person buying a gigantic chocolate bar is only 0.02. To test whether this hypothesis is true the manager decides to take a random sample of 200 people who bought chocolate bars.

(*b*) Find the critical region that would enable the manager to test whether or not there is evidence that the probability is different from 0.02. The probability of each tail should be as close to 2.5% as possible.

(*c*) Write down the significance level of this test.

26. Dhriti grows tomatoes. Over a period of time, she has found that there is a probability 0.3 of a ripe tomato having a diameter greater than 4 cm. She decides to try a new fertiliser. In a random sample of 40 ripe tomatoes, 18 have a diameter greater than 4 cm. Dhriti claims that the new fertiliser has increased the probability of a ripe tomato being greater than 4 cm in diameter.

Test Dhriti's claim at the 5% level of significance. State your hypotheses clearly.

27. (*a*) Explain what you understand by
(i) a hypothesis test,
(ii) a critical region.

During term time, incoming calls to a school are thought to occur at a rate of 0.45 per minute. To test this, the number of calls during a random 20 minute interval, is recorded.

(*b*) Find the critical region for a two-tailed test of the hypothesis that the number of incoming calls occurs at a rate of 0.45 per 1 minute interval. The probability in each tail should be as close to 2.5% as possible.

(*c*) Write down the actual significance level of the above test.

In the school holidays, 1 call occurs in a 10 minute interval.

(*d*) Test, at the 5% level of significance, whether or not there is evidence that the rate of incoming calls is less during the school holidays than in term time.

28. A single observation x is to be taken from a Binomial distribution B(20, p).
This observation is used to test $H_0 : p = 0.3$ against $H_1 : p \neq 0.3$.

(*a*) Using a 5% level of significance, find the critical region for this test. The probability of rejecting either tail should be as close as possible to 2.5%.

(*b*) State the actual significance level of this test.

The actual value of x obtained is 3.

(*c*) State a conclusion that can be drawn based on this value, giving a reason for your answer.

29. A web server is visited on weekdays, at a rate of 7 visits per minute. In a random one minute on a Saturday the web server is visited 10 times.

(*a*) (i) Test, at the 10% level of significance, whether or not there is evidence that the rate of visits is greater on a Saturday than on weekdays. State your hypotheses clearly.
(ii) State the minimum number of visits required to obtain a significant result.

(*b*) State an assumption that has been made about the visits to the server.

In a random two minute period on a Saturday the web server is visited 20 times.

(*c*) Using a suitable approximation, test at the 10% level of significance, whether or not the rate of visits is greater on a Saturday.

Answers

24. a) $H_0: p = 0.2, H_1: p \neq 0.2$ p(X≥9)=0.01, Evidence that the percentage of pupils that read Deano is not 20% b) 0 or between 9 and 20
b) $W \sim$ Bin (100, 0.2), $W \sim$ N (20, 16), 0.353, Combined numbers of Deano readers suggests 20% of pupils read Deano
c) Either "large sample size gives better result" or "looks as though they are not all drawn from the same population."

25 a) $H_0: p = 0.20$, $H_1: p < 0.20$, $P(X \leq 2) = 0.0442$, 0.0442 < 5%, so significant, there is evidence that the no. of family size bars sold is lower than usual
b) 0 or ≥ 9
c) 0.04

26. $H_0: p = 0.3$; $H_1: p > 0.3$. Let X represent the number of tomatoes greater than 4 cm : X∼ B(40, 0.3). P(X≥ 18) = 0.0320. 0.0320 < 0.05. No evidence to Reject H_0 or it is significant. New fertiliser has increased the probability of a tomato being greater than 4 cm o**r** Dhriti's claim is true.

27. a) i) A hypothesis test is a mathematical procedure to examine a value of a population parameter proposed by the null hypothesis compared with an alternative hypothesis.
ii) The critical region is the range of values **or** a test statistic or region where the test is significant that would lead to the rejection of H_0.
b) x ≤ 3 or x ≥ 16
c) 4.32%
d) P (X ≤ 1) = 0.0611; 0.0611 > 0.05 There is evidence to Accept H_0 or it is not significant. There is no evidence that there are less calls during school holidays.

28. a) x ≤ 2 or x ≥ 11
b) 5.26%
c) Insufficient evidence to reject H_0 **or** sufficient evidence to accept H_0 /not significant. $x = 3$ is not in the critical region **or** 0.1071> 0.025

29. a) i) H0 : $\lambda = 7$ H1 : $\lambda > 7$. X = number of visits. $X \sim$ Po(7). P ($X \geq 10$) = 1 - P($X \leq 9$) = 0.1695; 0.1695 > 0.10 , Critical Region is $X \geq 11$. Not significant or it is not in the critical region or do not reject H_0 The rate of visits on a Saturday is not greater/ is unchanged
ii) 11
b) The visits occur randomly/ independently or singly or constant rate
c) $H_0: \lambda = 7$ $H_1: \lambda > 7$; X ~N;(14,14) ; P (X ≥ 20) = P z ≥ (19.5 - 14)/14 = P (z ≥ 1.47) = 0.0708; 0.0708 < 0.10 therefore significant. The rate of visits is greater on a Saturday.