Topic 7 Part 1 *[328 marks]*

The random variable X has probability distribution Po(8).

1a. (i)   Find
$\text{P}(X = 6)$.
(ii)   Find
$\text{P}(X = 6|5 \leqslant X \leqslant 8)$.

[5 marks]

1b. $\bar{X}$ denotes the sample mean of
$n > 1$ independent observations from
X.
(i)   Write down
$\text{E}(\bar{X})$ and
$\text{Var}(\bar{X})$.
(ii)   Hence, give a reason why
$\bar{X}$ is not a Poisson distribution.

[3 marks]

1c. A random sample of
40 observations is taken from the distribution for
X.
(i)   Find
$\text{P}(7.1 < \bar{X} < 8.5)$.
(ii)   Given that
$\text{P}\left(\left|\bar{X} - 8\right| \leqslant k\right) = 0.95$, find the value of
k.

[6 marks]

2.   The following table gives the average yield of olives per tree, in kg, and the rainfall, in cm, for nine separate regions of Greece. You may assume that these data are a random sample from a bivariate normal distribution, with correlation coefficient ρ.

[16 marks]

Rainfall (x)	11	10	15	13	7	18	22	20	28
Yield (y)	56	53	67	61	54	78	86	88	78

A scientist wishes to use these data to determine whether there is a positive correlation between rainfall and yield.

(a)   State suitable hypotheses.

(b)   Determine the product moment correlation coefficient for these data.

(c)   Determine the associated p-value and comment on this value in the context of the question.

(d)   Find the equation of the regression line of y on x.

(e)   Hence, estimate the yield per tree in a tenth region where the rainfall was 19 cm.

(f)   Determine the angle between the regression line of y on x and that of x on y . Give your answer to the nearest degree.

A random variable X has probability density function

$$f(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{2} & 0 \leq x < 1 \\ \frac{1}{4} & 1 \leq x < 3 \\ 0 & x \geq 3 \end{cases}$$
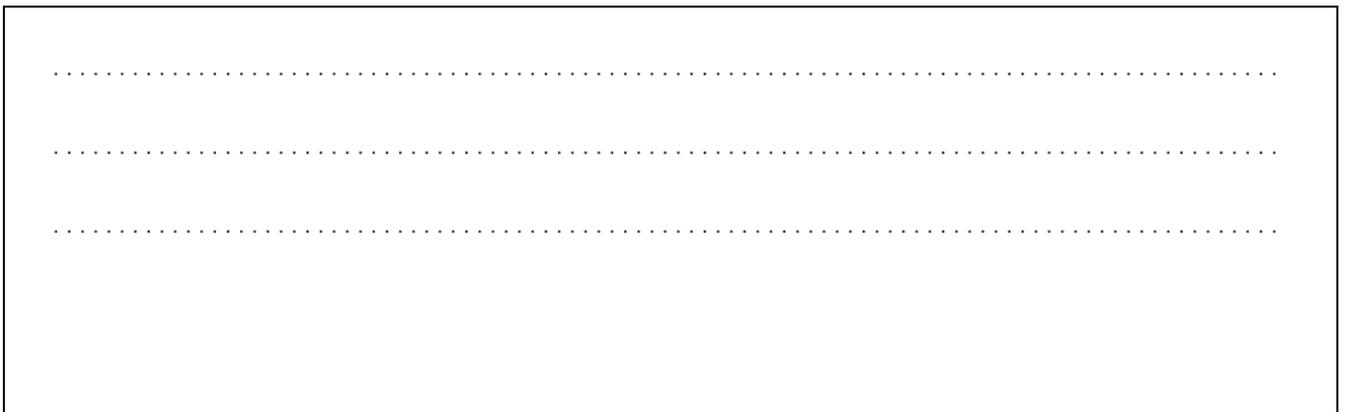
3a. Sketch the graph of $y = f(x)$.   *[1 mark]*
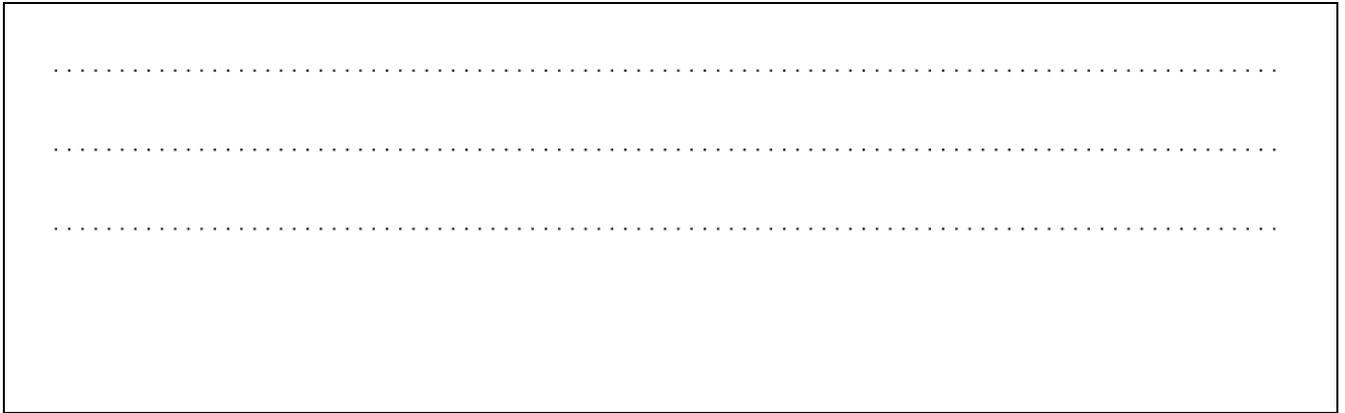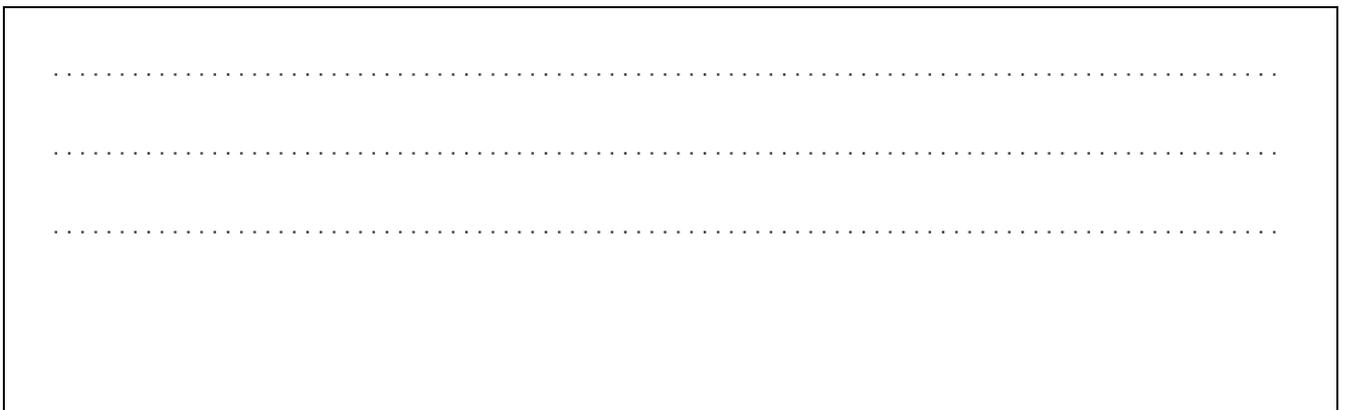
3b. Find the cumulative distribution function for X.   *[5 marks]*

3c. Find the interquartile range for X.   *[3 marks]*

Eric plays a game at a fairground in which he throws darts at a target. Each time he throws a dart, the probability of hitting the target is 0.2. He is allowed to throw as many darts as he likes, but it costs him $1 a throw. If he hits the target a total of three times he wins $10.

4a. Find the probability he has his third success of hitting the target on his sixth throw. [3 marks]

4b. (i) Find the expected number of throws required for Eric to hit the target three times. [3 marks]

(ii) Write down his expected profit or loss if he plays until he wins the $10.

4c. If he has just $8, find the probability he will lose all his money before he hits the target three times. [3 marks]

5a. If X and Y are two random variables such that $\mathrm{E}(X) = \mu_X$ and $\mathrm{E}(Y) = \mu_Y$ then [3 marks]
$\mathrm{Cov}(X,\ Y) = \mathrm{E}\left((X - \mu_X)(Y - \mu_Y)\right)$.

Prove that if X and Y are independent then $\mathrm{Cov}(X,\ Y) = 0$.

....

....

....

5b. In a particular company, it is claimed that the distance travelled by employees to work is independent of their [8 marks] salary. To test this, 20 randomly selected employees are asked about the distance they travel to work and the size of their salaries. It is found that the product moment correlation coefficient, r, for the sample is -0.35.

You may assume that both salary and distance travelled to work follow normal distributions.

Perform a one-tailed test at the 5% significance level to test whether or not the distance travelled to work and the salaries of the employees are independent.

....

....

....

If X is a random variable that follows a Poisson distribution with mean $\lambda > 0$ then the probability generating function of X is $G(t) = e^{\lambda(t-1)}$.

6a. (i) Prove that $\mathrm{E}(X) = \lambda$. [6 marks]

(ii) Prove that $\mathrm{Var}(X) = \lambda$.

....

....

....

6b. Y is a random variable, independent of X, that also follows a Poisson distribution with mean λ. *[3 marks]*

If $S = 2X - Y$ find

(i) $\mathrm{E}(S)$;

(ii) $\mathrm{Var}(S)$.

....

....

....

6c. Let $T = \frac{Y}{2} + \frac{Y}{2}$. *[3 marks]*

(i) Show that T is an unbiased estimator for λ.

(ii) Show that T is a more efficient unbiased estimator of λ than S.

....

....

....

6d. Could either S or T model a Poisson distribution? Justify your answer. *[1 mark]*

....

....

....

6e. By consideration of the probability generating function, $G_{X+Y}(t)$, of $X + Y$, prove that $X + Y$ follows a Poisson [3 marks] distribution with mean 2λ.

6f. Find [2 marks]

(i) $G_{X+Y}(1)$;

(ii) $G_{X+Y}(-1)$.

6g. Hence find the probability that $X + Y$ is an even number. [3 marks]

Two species of plant, A and B, are identical in appearance though it is known that the mean length of leaves from a plant of species A is 5.2 cm, whereas the mean length of leaves from a plant of species B is 4.6 cm. Both lengths can be modelled by normal distributions with standard deviation 1.2 cm.

In order to test whether a particular plant is from species A or species B, 16 leaves are collected at random from the plant. The length, x, of each leaf is measured and the mean length evaluated. A one-tailed test of the sample mean, $\bar{X}$, is then performed at the 5% level, with the hypotheses: $H_0 : \mu = 5.2$ and $H_1 : \mu < 5.2$.

7a. Find the critical region for this test.   *[3 marks]*

7b. It is now known that in the area in which the plant was found 90% of all the plants are of species A and 10% are   *[2 marks]*
of species B.

Find the probability that $\bar{X}$ will fall within the critical region of the test.

7c. If, having done the test, the sample mean is found to lie within the critical region, find the probability that the   *[3 marks]*
leaves came from a plant of species A.

Engine oil is sold in cans of two capacities, large and small. The amount, in millilitres, in each can, is normally distributed according to Large $\sim \mathrm{N}(5000,\ 40)$ and Small $\sim \mathrm{N}(1000,\ 25)$.

8a. A large can is selected at random. Find the probability that the can contains at least 4995 millilitres of oil.   *[2 marks]*

8b. A large can and a small can are selected at random. Find the probability that the large can contains at least 30 *[6 marks]* milliliters more than five times the amount contained in the small can.

8c. A large can and five small cans are selected at random. Find the probability that the large can contains at least   *[5 marks]* 30 milliliters less than the total amount contained in the small cans.

Eleven students who had under-performed in a philosophy practice examination were given extra tuition before their final examination. The differences between their final examination marks and their practice examination marks were

$$10, \; -1, \; 6, \; 7, \; -5, \; -5, \; 2, \; -3, \; 8, \; 9, \; -2.$$

Assume that these differences form a random sample from a normal distribution with mean μ and variance σ^2.

9a. Determine unbiased estimates of μ and σ^2. *[4 marks]*

. . . .

. . . .

. . . .

9b. (i) State suitable hypotheses to test the claim that extra tuition improves examination marks. *[8 marks]*

(ii) Calculate the p-value of the sample.

(iii) Determine whether or not the above claim is supported at the 5% significance level.

. . . .

. . . .

. . . .

A manufacturer of stopwatches employs a large number of people to time the winner of a 100 metre sprint. It is believed that if the true time of the winner is μ seconds, the times recorded are normally distributed with mean μ seconds and standard deviation 0.03 seconds.

The times, in seconds, recorded by six randomly chosen people are

$$9.765, \ 9.811, \ 9.783, \ 9.797, \ 9.804, \ 9.798.$$

10a. Calculate a 99% confidence interval for μ. Give your answer correct to three decimal places.   *[4 marks]*

10b. Interpret the result found in (a).   *[2 marks]*

10c. Find the confidence level of the interval that corresponds to halving the width of the 99% confidence interval. Give your answer as a percentage to the nearest whole number.   *[3 marks]*

A random variable
X has a population mean μ.

11a. Explain briefly the meaning of [3 marks]

(i) an estimator of μ;

(ii) an unbiased estimator of μ.

....

....

....

11b. A random sample X_1, X_2, X_3 of three independent observations is taken from the distribution of X. [12 marks]

An unbiased estimator of μ, $\mu \neq 0$, is given by $U = \alpha X_1 + \beta X_2 + (\alpha - \beta)X_3$,

where α, $\beta \in \mathbb{R}$.

(i) Find the value of α.

(ii) Show that $\operatorname{Var}(U) = \sigma^2 \left(2\beta^2 - \beta + \frac{1}{2}\right)$ where $\sigma^2 = \operatorname{Var}(X)$.

(iii) Find the value of β which gives the most efficient estimator of μ of this form.

(iv) Write down an expression for this estimator and determine its variance.

(v) Write down a more efficient estimator of μ than the one found in (iv), justifying your answer.

....

....

....

12a. Determine the probability generating function for $X \sim \mathrm{B}(1,\ p)$. [4 marks]

....

....

....

12b. Explain why the probability generating function for $B(n,\ p)$ is a polynomial of degree n. *[2 marks]*

. . . .

. . . .

. . . .

12c. Two independent random variables X_1 and X_2 are such that $X_1 \sim B(1,\ p_1)$ and $X_2 \sim B(1,\ p_2)$. Prove that if $X_1 + X_2$ has a binomial distribution then $p_1 = p_2$. *[5 marks]*

. . . .

. . . .

. . . .

A baker produces loaves of bread that he claims weigh on average 800 g each. Many customers believe the average weight of his loaves is less than this. A food inspector visits the bakery and weighs a random sample of 10 loaves, with the following results, in grams:

$$783, 802, 804, 785, 810, 805, 789, 781, 800, 791.$$

Assume that these results are taken from a normal distribution.

13a. Determine unbiased estimates for the mean and variance of the distribution. *[3 marks]*

13b. In spite of these results the baker insists that his claim is correct. *[7 marks]*
Stating appropriate hypotheses, test the baker's claim at the 10 % level of significance.

The random variable X has a geometric distribution with parameter p.

14a. Show that
$P(X \leqslant n) = 1 - (1-p)^n$, $n \in \mathbb{Z}^+$. *[3 marks]*

14b. Deduce an expression for
$P(m < X \leqslant n)$, $m,\ n \in \mathbb{Z}^+$ and $m < n$. *[1 mark]*

14c. Given that $p = 0.2$, find the least value of n for which
$P(1 < X \leqslant n) > 0.5$, $n \in \mathbb{Z}^+$. *[2 marks]*

The n independent random variables

$X_1, X_2, \ldots, X_n$ all have the distribution

$N(\mu, \sigma^2)$.

15a. Find the mean and the variance of

(i)

$X_1 + X_2$;

(ii)

$3X_1$;

(iii)

$X_1 + X_2 - X_3$;

(iv)

$\bar{X} = \frac{(X_1 + X_2 + \ldots + X_n)}{n}$.

15b. Find

$E(X_1^2)$ in terms of

μ and

σ .

16a.

The random variable X represents the height of a wave on a particular surf beach.

It is known that X is normally distributed with unknown mean

μ (metres) and known variance

$\sigma^2 = \frac{1}{4}$ (metres^2) . Sally wishes to test the claim made in a surf guide that

$\mu = 3$ against the alternative that

$\mu < 3$. She measures the heights of 36 waves and calculates their sample mean

$\bar{x}$. She uses this value to test the claim at the 5 % level.

(i)   Find a simple inequality, of the form

$\bar{x} < A$, where A is a number to be determined to 4 significant figures, so that Sally will reject the null hypothesis, that

$\mu = 3$, if and only if this inequality is satisfied.

(ii)   Define a Type I error.

(iii)   Define a Type II error.

(iv)   Write down the probability that Sally makes a Type I error.

(v)   The true value of

μ is 2.75. Calculate the probability that Sally makes a Type II error.

16b. The random variable Y represents the height of a wave on another surf beach. It is known that Y is normally distributed with unknown mean

μ (metres) and unknown variance

σ^2 (metres^2) . David wishes to test the claim made in a surf guide that

$\mu = 3$ against the alternative that

$\mu < 3$. He is also going to perform this test at the 5 % level. He measures the heights of 36 waves and finds that the sample mean,

$\bar{y} = 2.860$ and the unbiased estimate of the population variance,

$s_{n-1}^2 = 0.25.$

(i)   State the name of the test that David should perform.

(ii)   State the conclusion of David's test, justifying your answer by giving the p-value.

(iii)   Using David's results, calculate the 90 % confidence interval for

μ , giving your answers to 4 significant figures.

Jenny and her Dad frequently play a board game. Before she can start Jenny has to throw a "six" on an ordinary six-sided dice. Let the random variable X denote the number of times Jenny has to throw the dice in total until she obtains her first "six".

17a. If the dice is fair, write down the distribution of X , including the value of any parameter(s).

17b. Write down E(X) for the distribution in part (a).

17c. Before Jenny's Dad can start, he has to throw two "sixes" using a fair, ordinary six-sided dice. Let the random variable Y denote the total number of times Jenny's Dad has to throw the dice until he obtains his second "six".
Write down the distribution of Y , including the value of any parameter(s).

17d. Before Jenny's Dad can start, he has to throw two "sixes" using a fair, ordinary six-sided dice. Let the random variable Y denote the total number of times Jenny's Dad has to throw the dice until he obtains his second "six".
Find the value of y such that
$P(Y = y) = \frac{1}{36}.$

17e. Before Jenny's Dad can start, he has to throw two "sixes" using a fair, ordinary six-sided dice. Let the random variable Y denote the total number of times Jenny's Dad has to throw the dice until he obtains his second "six".
Find
$P(Y \leqslant 6) .$

The random variable X is normally distributed with unknown mean

μ and unknown variance

σ^2. A random sample of 20 observations on X gave the following results.

$$\sum x = 280, \ \sum x^2 = 3977.57$$

18a. Find unbiased estimates of

μ and

σ^2.

[3 marks]

18b. Determine a 95 % confidence interval for

μ.

[3 marks]

18c. Given the hypotheses

$$H_0 : \mu = 15; \ H_1 : \mu \neq 15,$$

find the p-value of the above results and state your conclusion at the 1 % significance level.

[4 marks]

The number of machine breakdowns occurring in a day in a certain factory may be assumed to follow a Poisson distribution with

mean

μ. The value of

μ is known, from past experience, to be 1.2. In an attempt to reduce the value of

μ, all the machines are fitted with new control units. To investigate whether or not this reduces the value of

μ, the total number of breakdowns, x, occurring during a 30-day period following the installation of these new units is recorded.

19a. State suitable hypotheses for this investigation.

[1 mark]

19b. It is decided to define the critical region by

$x \leqslant 25$.

(i)   Calculate the significance level.

(ii)   Assuming that the value of

μ was actually reduced to 0.75, determine the probability of a Type II error.

[8 marks]

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

$$f(x) = \begin{cases} \frac{3x^2 + 2x}{10}, & \text{for } 1 \leqslant x \leqslant 2 \\ 0, & \text{otherwise.} \end{cases}$$

20a. (i)   Determine an expression for

$F(x)$, valid for

$1 \leqslant x \leqslant 2$, where F denotes the cumulative distribution function of X.

(ii)   Hence, or otherwise, determine the median of X.

[6 marks]

20b. (i)   State the central limit theorem.

(ii)   A random sample of 150 observations is taken from the distribution of X and

$\bar{X}$ denotes the sample mean. Use the central limit theorem to find, approximately, the probability that

$\bar{X}$ is greater than 1.6.

[8 marks]

When Ben shoots an arrow, he hits the target with probability 0.4. Successive shots are independent.

21a. Find the probability that

(i)   he hits the target exactly 4 times in his first 8 shots;

(ii)   he hits the target for the

4th time with his

8th shot.

[6 marks]

21b. Ben hits the target for the

10th time with his

X^{th} shot.

(i)   Determine the expected value of the random variable X.

(ii)   Write down an expression for

$P(X = x)$ and show that

$$\frac{P(X = x)}{P(X = x - 1)} = \frac{3(x - 1)}{5(x - 10)}.$$

(iii)   Hence, or otherwise, find the most likely value of X.

[9 marks]

22. (a) Consider the random variable

X for which

$E(X) = a\lambda + b$, where

a and

b are constants and

λ is a parameter.

Show that

$\frac{X-b}{a}$ is an unbiased estimator for

λ.

(b) The continuous random variable Y has probability density function

$$f(y) = \begin{cases} \frac{2}{9}(3 + y - \lambda), & \text{for } \lambda - 3 \leq y \leq \lambda \\ 0, & \text{otherwise} \end{cases}$$

where

λ is a parameter.

    (i) Verify that

$f(y)$ is a probability density function for all values of

λ.

    (ii) Determine

$E(Y)$.

    (iii) Write down an unbiased estimator for

λ.

23. Consider the random variable

$X \sim \text{Geo}(p)$.

(a) State

$P(X < 4)$.

(b) Show that the probability generating function for X is given by

$G_X(t) = \frac{pt}{1-qt}$, where

$q = 1 - p$.

Let the random variable

$Y = 2X$.

(c) (i) Show that the probability generating function for Y is given by

$G_Y(t) = G_X(t^2)$.

    (ii) By considering

$G_Y'(1)$, show that

$E(Y) = 2E(X)$.

Let the random variable

$W = 2X + 1$.

(d) (i) Find the probability generating function for W in terms of the probability generating function of Y.

    (ii) Hence, show that

$E(W) = 2E(X) + 1$.

A traffic radar records the speed,

v kilometres per hour (

$\mathrm{km\,h^{-1}}$), of cars on a section of a road.

The following table shows a summary of the results for a random sample of 1000 cars whose speeds were recorded on a given day.

Speed	Number of cars
$50 \leq v < 60$	5
$60 \leq v < 70$	13
$70 \leq v < 80$	52
$80 \leq v < 90$	68
$90 \leq v < 100$	98
$100 \leq v < 110$	105
$110 \leq v < 120$	289
$120 \leq v < 130$	142
$130 \leq v < 140$	197
$140 \leq v < 150$	31

24a. Using the data in the table,   [4 marks]

   (i)   show that an estimate of the mean speed of the sample is 113.21

$\mathrm{km\,h^{-1}}$;

  (ii)   find an estimate of the variance of the speed of the cars on this section of the road.

24b. Find the 95% confidence interval,   [2 marks]

   I, for the mean speed.

24c. Let   [2 marks]

   J be the 90% confidence interval for the mean speed.

Without calculating

J, explain why

$J \subset I.$

Jenny tosses seven coins simultaneously and counts the number of tails obtained. She repeats the experiment 750 times. The following frequency table shows her results.

Number of tails	Frequency
0	6
1	19
2	141
3	218
4	203
5	117
6	38
7	8

25a. Explain what can be done with this data to decrease the probability of making a type I error.   [2 marks]

25b. (i)   State the meaning of a type II error.   [2 marks]

  (ii)   Write down how to proceed if it is required to decrease the probability of making both a type I and type II error.

26. Francisco and his friends want to test whether performance in running 400 metres improves if they follow a particular training *[10 marks]* schedule. The competitors are tested before and after the training schedule.

The times taken to run 400 metres, in seconds, before and after training are shown in the following table.

Competitor	A	B	C	D	E
Time before training	75	74	60	69	69
Time after training	73	69	55	72	65

Apply an appropriate test at the 1% significance level to decide whether the training schedule improves competitors' times, stating clearly the null and alternative hypotheses. (It may be assumed that the distributions of the times before and after training are normal.)

27. Let   *[14 marks]*

X and

Y be independent random variables with

$X \sim P_o\,(3)$ and

$Y \sim P_o\,(2)$.

Let

$S = 2X + 3Y$.

(a)   Find the mean and variance of

S.

(b)   Hence state with a reason whether or not

S follows a Poisson distribution.

Let

$T = X + Y$.

(c)   Find

$P(T = 3)$.

(d)   Show that

$P(T = t) = \sum_{r=0}^{t} P(X = r)P(Y = t - r)$.

(e)   Hence show that

T follows a Poisson distribution with mean 5.

The discrete random variable X has the following probability distribution, where

$0 < \theta < \frac{1}{3}$.

x	1	2	3
$P(X = x)$	θ	2θ	$1 - 3\theta$

28a. Determine   *[4 marks]*

$E(X)$ and show that

$\text{Var}(X) = 6\theta - 16\theta^2$.

28b. In order to estimate

θ, a random sample of n observations is obtained from the distribution of X.

(i)   Given that

$\bar{X}$ denotes the mean of this sample, show that

$$\hat{\theta}_1 = \frac{3 - \bar{X}}{4}$$

is an unbiased estimator for

θ and write down an expression for the variance of

$\hat{\theta}_1$ in terms of n and

θ.

(ii)   Let Y denote the number of observations that are equal to 1 in the sample. Show that Y has the binomial distribution

$B(n, \theta)$ and deduce that

$\hat{\theta}_2 = \frac{Y}{n}$ is another unbiased estimator for

θ. Obtain an expression for the variance of

$\hat{\theta}_2$.

(iii)   Show that

$\text{Var}(\hat{\theta}_1) < \text{Var}(\hat{\theta}_2)$ and state, with a reason, which is the more efficient estimator,

$\hat{\theta}_1$ or

$\hat{\theta}_2$.