

## Paper 2 Solution

### Section A: Pure Mathematics [40 marks]

- 1 Relative to the origin  $O$ , the position vectors of two points  $A$  and  $B$  are  $\mathbf{a}$  and  $\mathbf{b}$  respectively. The length of  $\mathbf{a}$  is  $k$  units and  $\mathbf{b}$  is a unit vector. The angle between  $\mathbf{a}$  and  $\mathbf{b}$  is  $\frac{\pi}{6}$  radians.

It is given that the point  $M$  lies on the line segment  $AB$  such that  $AM:AB$  is 3:4, find exact area of  $\triangle OAM$ , giving your answer in terms of  $k$ . [4]

Solution

$$\overrightarrow{OM} = \frac{3\mathbf{b} + \mathbf{a}}{4}$$

$$\text{Area of } \triangle OAM = \frac{1}{2} |\overrightarrow{OA} \times \overrightarrow{OM}|$$

$$= \frac{1}{2} \left| \mathbf{a} \times \left( \frac{3\mathbf{b} + \mathbf{a}}{4} \right) \right|$$

$$= \frac{1}{8} |\mathbf{a} \times 3\mathbf{b} + \mathbf{a} \times \mathbf{a}|$$

$$= \frac{3}{8} |\mathbf{a} \times \mathbf{b}| \quad \text{Since } \mathbf{a} \times \mathbf{a} = \mathbf{0}$$

$$= \frac{3}{8} |\mathbf{a}| |\mathbf{b}| \sin \frac{\pi}{6} \hat{\mathbf{n}} \quad \text{where } \hat{\mathbf{n}} \text{ is a unit vector } \perp \mathbf{a} \& \mathbf{b}$$

$$= \frac{3}{8} k \left( \frac{1}{2} \right)$$

$$= \frac{3}{16} k \text{ units}^2$$

**2 (a)** Without the use of a graphing calculator, solve the inequality  $\frac{x+4}{x+1} \geq \frac{x+2}{2}$ . [3]

**(b)** Hence solve  $1 + \frac{3}{e^x + 1} \geq \frac{e^x + 2}{2}$ . [2]

Solution

**(a)**  $\frac{x+4}{x+1} \geq \frac{x+2}{2}$

$$\frac{x+2}{2} - \frac{x+4}{x+1} \leq 0$$

$$\frac{(x+2)(x+1) - 2(x+4)}{2(x+1)} \leq 0$$

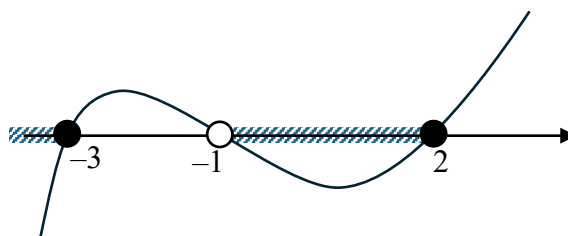
$$\frac{x^2 + x - 6}{2(x+1)} \leq 0$$

$$\frac{(x-2)(x+3)}{(x+1)} \leq 0$$

Multiplying both sides by  $(x+1)^2$ :

$$(x-2)(x+3)(x+1) \leq 0, \quad x \neq -1$$

$$x \leq -3 \quad \text{or} \quad -1 < x \leq 2$$



**(b)**  $1 + \frac{3}{e^x + 1} \geq \frac{e^x + 2}{2}$

$$\frac{e^x + 4}{e^x + 1} \geq \frac{e^x + 2}{2}$$

Replace  $x$  with  $e^x$

$$e^x \leq -3 \quad (\text{NA})$$

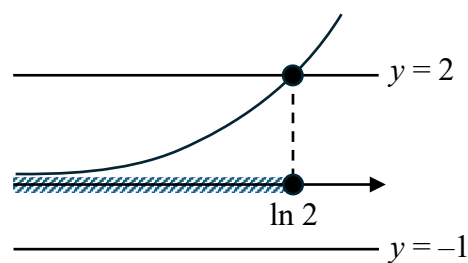
or

$$-1 < e^x \leq 2$$

(No solution  $\because e^x > 0$ )

$$0 < e^x \leq 2$$

$$x \leq \ln 2$$



**3 (a)** Given that  $y = \tan^{-1} \frac{1}{5} x$ , show that  $(25 + x^2) \frac{d^2 y}{dx^2} + 2x \frac{dy}{dx} = 0$ . [2]

**(b)** By further differentiation of this result, find the Maclaurin's series of  $y$  up to and including the term in  $x^3$ . [3]

[Solution]

Method 1

$$\begin{aligned} \text{Differentiating wrt } x: \frac{dy}{dx} &= \frac{1}{1 + \left(\frac{1}{5}x\right)^2} \left(\frac{1}{5}\right) \\ \left(1 + \frac{1}{25}x^2\right) \frac{dy}{dx} &= \frac{1}{5} \end{aligned}$$

Method 2

$$y = \tan^{-1} \frac{1}{5} x \Rightarrow \tan y = \frac{1}{5} x$$

$$\begin{aligned} \text{Differentiating wrt } x: \sec^2 y \frac{dy}{dx} &= \frac{1}{5} \\ (1 + \tan^2 y) \frac{dy}{dx} &= \frac{1}{5} \\ \left(1 + \frac{1}{25}x^2\right) \frac{dy}{dx} &= \frac{1}{5} \\ (25 + x^2) \frac{dy}{dx} &= 5 \quad (\text{Shown}) \end{aligned}$$

$$\text{Differentiating wrt } x: (25 + x^2) \frac{d^2 y}{dx^2} + 2x \frac{dy}{dx} = 0$$

$$\text{Differentiating wrt } x: (25 + x^2) \frac{d^3 y}{dx^3} + (2x) \frac{d^2 y}{dx^2} + 2x \frac{d^2 y}{dx^2} + 2 \frac{dy}{dx} = 0$$

$$\text{When } x = 0, y = 0, \frac{dy}{dx} = \frac{1}{5}, \frac{d^2 y}{dx^2} = 0, \frac{d^2 y}{dx^2} = -\frac{2}{125}$$

$$\begin{aligned} \text{Hence the Maclaurin expansion required is } y &= 0 + \frac{1}{5}x + 0 + \frac{-2}{3!}x^3 + \dots \\ &= \frac{1}{5}x - \frac{1}{375}x^3 + \dots \end{aligned}$$

- 4 The curve  $C$  is defined by the parametric equations

$$x = \cos t + \frac{1}{2} \cos 5t, \quad y = \sin t + \frac{1}{2} \sin 5t \quad \text{for } 0 \leq t \leq \frac{\pi}{2}.$$

- (a) Find the coordinates of the point on the curve  $C$  corresponding to  $t = 0$ . [1]

Another curve  $D$  has the following parametric equations

$$x = 2 + \sqrt{k} \cos \theta, \quad y = \frac{3}{2} \sin \theta \quad \text{for } 0 \leq \theta \leq 2\pi \text{ and } k > 0.$$

- (b) Find the cartesian equation for curve  $D$ . [2]

- (c) Sketch the curves  $C$  and  $D$  on the same diagram. [2]

- (d) Hence determine the range of values of  $k$  such that the equation

$$\frac{(\cos t + 0.5 \cos 5t - 2)^2}{k} + \frac{4(\sin t + 0.5 \sin 5t)^2}{9} = 1$$

has no real solutions.

[1]

[Solution]

- (a) When  $t = 0$ ,  $x = 1 + \frac{1}{2} = \frac{3}{2}$ ,  $y = 0$ .

Hence the coordinates of the points on the curve corresponding to  $t = 0$  are  $\left(\frac{3}{2}, 0\right)$

- (b)  $x = 2 + \sqrt{k} \cos \theta \quad \Rightarrow \quad \cos \theta = \frac{1}{\sqrt{k}}(x - 2)$

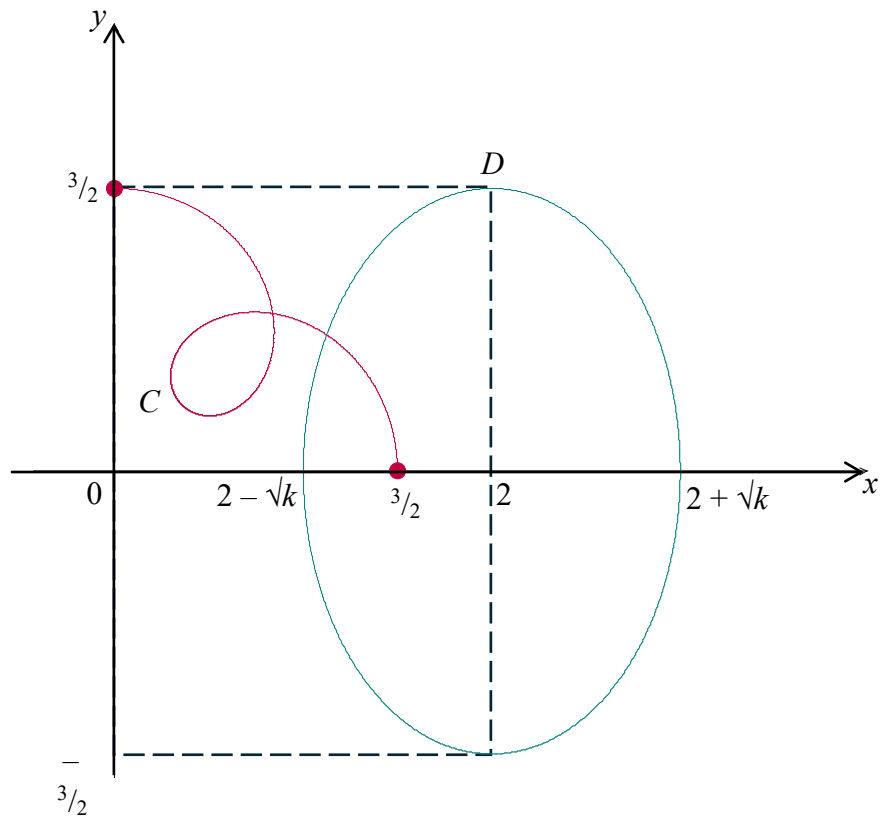
$$y = \frac{3}{2} \sin \theta \quad \Rightarrow \quad \sin \theta = \frac{2}{3} y$$

Since  $\cos^2 \theta + \sin^2 \theta = 1$ ,

$$\left[ \frac{1}{\sqrt{k}}(x - 2) \right]^2 + \left( \frac{2}{3} y \right)^2 = 1$$

$$\frac{(x - 2)^2}{k} + \frac{4y^2}{9} = 1$$

(c)



(d) For  $\frac{(\cos t + 0.5 \cos 5t - 2)^2}{k} + \frac{4(\sin t + 0.5 \sin 5t)^2}{9} = 1$  to have no solutions, there must not be any intersections b/w the 2 curves.

$$\text{Hence } 2 - \sqrt{k} > \frac{3}{2} \Rightarrow \sqrt{k} < \frac{1}{2}$$

Since  $k$  is positive, we have  $0 < k < \frac{1}{4}$

5 (a) Solve the following integral.

(i)  $\int \frac{1-2x}{\sqrt{3+2x-x^2}} dx$ . [2]

(ii)  $\int x \ln(2-x^2) dx$ , where  $-\sqrt{2} < x < \sqrt{2}$ . [3]

Solution

(a)(i) 
$$\int \frac{1-2x}{\sqrt{3+2x-x^2}} dx = \int \frac{2-2x}{\sqrt{3+2x-x^2}} - \frac{1}{\sqrt{2^2-(x-1)^2}} dx$$
  

$$= 2\sqrt{3+2x-x^2} - \sin^{-1}\left(\frac{x-1}{2}\right) + c$$

(a)(ii) 
$$\int x \ln(2-x^2) dx = \frac{x^2}{2} \ln(2-x^2) - \int \left(\frac{x^2}{2}\right) \left(\frac{-2x}{2-x^2}\right) dx$$

Integration by parts:

$$\int u dv = uv - \int v du$$

$$= \frac{x^2}{2} \ln(2-x^2) + \int \frac{x^3}{2-x^2} dx$$

$$= \frac{x^2}{2} \ln(2-x^2) + \int -x + \frac{2x}{2-x^2} dx$$

$$= \frac{x^2}{2} \ln(2-x^2) + \int -x - \frac{-2x}{2-x^2} dx$$

$$= \frac{x^2}{2} \ln(2-x^2) - \frac{x^2}{2} - \ln(2-x^2) + c$$

Reverse Chain Rule:

$$\int \frac{f'(x)}{f(x)} dx = \ln|f(x)| + c$$

$$\underline{\text{OR}} \left(\frac{x^2}{2} - 1\right) \ln(2-x^2) - \frac{x^2}{2} + c$$

$$\underline{\text{OR}} \frac{x^2}{2} (\ln(2-x^2) - 1) - \ln(2-x^2) + c$$

Reverse Chain Rule:

$$\int f'(x)[f(x)]^n dx = \frac{[f(x)]^{n+1}}{n+1} + c$$

Integrating  $\frac{1}{\sqrt{\text{Quadratic}}}$ :

#1 Complete the square

#2 Choose correct formula from MF 27:

$$\int \frac{1}{\sqrt{a^2-x^2}} dx = \sin^{-1}\left(\frac{x}{a}\right) + c$$

For integration by parts:

$$u = \ln(2-x^2) \quad \frac{du}{dx} = x$$

$$\frac{du}{dx} = \frac{-2x}{2-x^2} \quad v = \frac{x^2}{2}$$

For long division:

$$\begin{array}{r} -x^2+2 \overline{) \frac{-x}{x^3}} \\ \underline{-(x^3-2x)} \\ 2x \end{array}$$

(b) A function  $f$  is defined by  $f(x) = e^x - 2$ .

(i) Sketch the graph of  $y = |f(x)|$ , indicating clearly the equation(s) of asymptote(s), if any. [2]

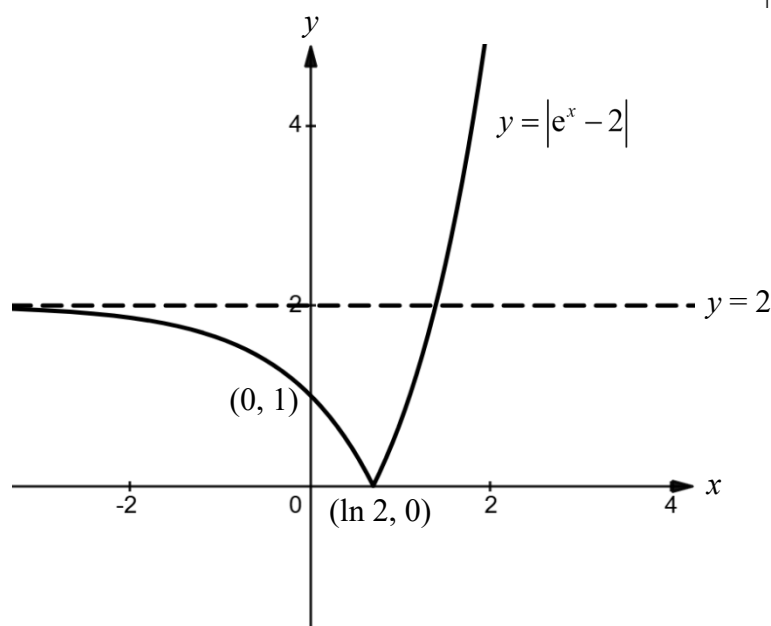
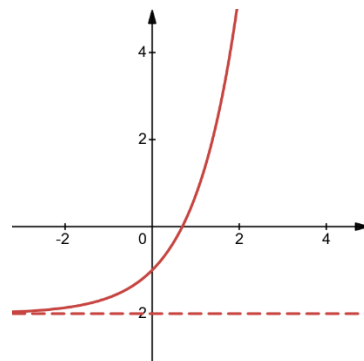
(ii) Hence find  $\int_0^{\ln 3} |f(x)| dx$ , giving your answer in exact form. [3]

Solution

(b)(i) For  $y = e^x - 2$ :

When  $x = 0$ ,  
 $y = e^0 - 2 = -1$

When  $y = 0$ ,  
 $e^x - 2 = 0$   
 $x = \ln 2$



(b)(ii) From (i),  $y = |f(x)| = \begin{cases} e^x - 2, & x > \ln 2 \\ -e^x + 2, & x \leq \ln 2 \end{cases}$

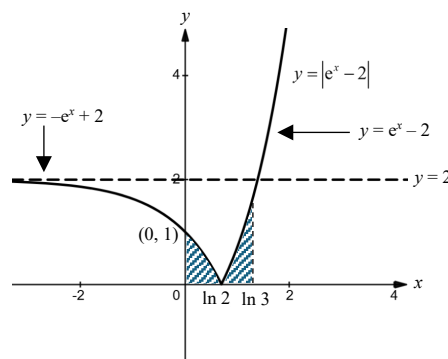
$$\int_0^{\ln 3} |e^x - 2| dx = \int_0^{\ln 2} -e^x + 2 dx + \int_{\ln 2}^{\ln 3} e^x - 2 dx$$

$$= [-e^x + 2x]_0^{\ln 2} + [e^x - 2x]_{\ln 2}^{\ln 3}$$

$$= (2 \ln 2 - 2) - (0 - 1) + (3 - 2 \ln 3) - (2 - 2 \ln 2)$$

$$= 4 \ln 2 - 2 \ln 3$$

Alternative answers:  $\ln \frac{16}{9}$ ;  $2 \ln \frac{4}{3}$



- 6 The points  $P$  and  $Q$  are represented by the complex numbers  $p$  and  $q$  respectively, where

$$p = \sqrt{2} + \sqrt{2}i, \quad \arg(q) = \frac{2}{3}\pi \quad \text{and} \quad |q| = 2.$$

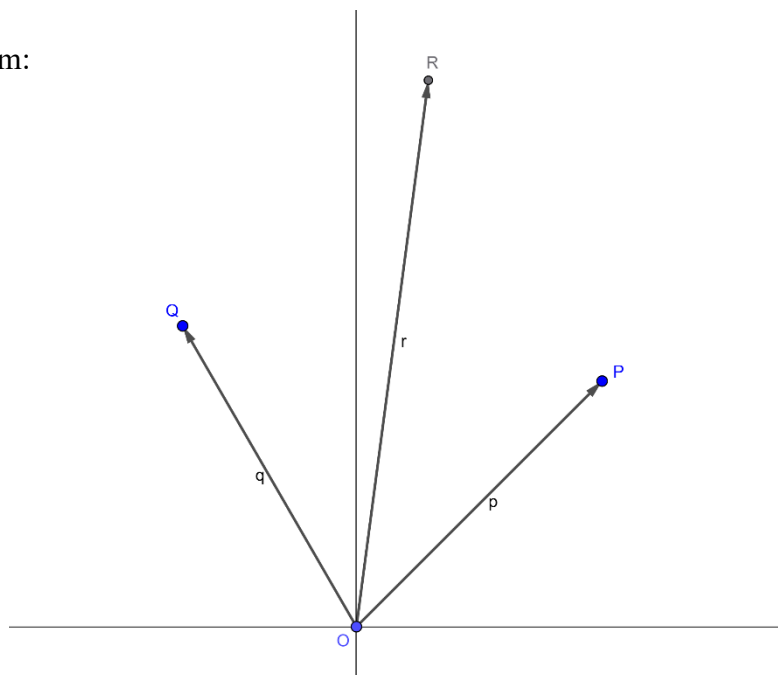
- (a) Find  $|p|$  and  $\arg(p)$  in exact form. [2]
- (b) Sketch the points  $P$  and  $Q$  on an Argand diagram. [2]
- (c) Use the Argand diagram to deduce  $\operatorname{Re}(q)$  and  $\operatorname{Im}(q)$ , giving your answers in exact form. [2]
- (d) The point  $R$  is represented by the complex number  $p + q$ . What can you deduce about the shape of quadrilateral  $OPRQ$ , where  $O$  is the origin? [1]
- (e) By considering  $\arg(p + q)$  or otherwise, show that  $\tan\left(\frac{11}{24}\pi\right) = \sqrt{6} + \sqrt{3} + \sqrt{2} + 2$ . [3]

[Solution]

$$(a) \quad |p| = \sqrt{(\sqrt{2})^2 + (\sqrt{2})^2} = 2$$

$$\arg(p) = \tan^{-1}\left(\frac{\sqrt{2}}{\sqrt{2}}\right) = \frac{\pi}{4}$$

(b) Diagram:





(c) Method 1

Let  $F$  be the foot of perpendicular from  $Q$  to the  $y$ -axis

$$\operatorname{Re}(q) = -QF = -|q| \sin \frac{2\pi}{3} = -1$$

$$\operatorname{Im}(q) = OF = |q| \cos \frac{2\pi}{3} = \sqrt{3}$$

Method 2

$$q = 2 \left( \cos \frac{2\pi}{3} + i \sin \frac{2\pi}{3} \right)$$

$$= 2 \left( -\frac{1}{2} + \frac{\sqrt{3}}{2} i \right)$$

$$= -1 + \sqrt{3}i$$

$$\operatorname{Re}(q) = -1$$

$$\operatorname{Im}(q) = \sqrt{3}$$

(d) Quadrilateral  $OPRQ$  is a rhombus.

$$\begin{aligned} \text{(e) First, } \arg(p+q) &= \frac{1}{2} \left( \frac{\pi}{4} + \frac{2\pi}{3} \right) \\ &= \frac{11\pi}{24} \end{aligned}$$

$$\begin{aligned} \text{Also, } p+q &= \sqrt{2} + \sqrt{2}i - 1 + \sqrt{3}i \\ &= (\sqrt{2} - 1) + (\sqrt{3} + \sqrt{2})i \end{aligned}$$

$$\begin{aligned} \text{Therefore, } \tan \left( \frac{11}{24} \pi \right) &= \frac{\sqrt{3} + \sqrt{2}}{\sqrt{2} - 1} \\ &= \frac{\sqrt{3} + \sqrt{2}}{\sqrt{2} - 1} \times \frac{\sqrt{2} + 1}{\sqrt{2} + 1} \\ &= \sqrt{6} + \sqrt{3} + \sqrt{2} + 2 \end{aligned}$$

### Section B: Probability and Statistics [60 marks]

- 7 Anthony plays a game using a four-sided fair die with the numbers 1, 2, 3 and 4 printed on its sides. Anthony rolls the die and should the outcome be a prime number, he wins an amount equivalent to that number in dollars. Conversely, if the outcome is a non-prime number, Anthony will lose an amount equivalent to that number in dollars. The random variable  $X$  represents the amount of money Anthony wins in a roll of the die.

- (a) Tabulate the probability distribution of  $X$  and find the value of  $E(X)$ . [2]  
 (b) Find the value of the variance of  $X$ . [1]  
 (c) The random variables  $X_1$  and  $X_2$  are two independent observations of  $X$ . By drawing a table of outcomes for  $|X_1 + X_2|$  or otherwise, find  $E(|X_1 + X_2|)$ . [2]

Assume now that a biased six-sided die with numbers 2, 3, 4, 5, 6 and 7 printed on its sides, is used instead for the same game. It is known that for this biased die, the probability of obtaining a “3” is  $p$ , the probability of obtaining a “5” is  $2p$  and the remaining four numbers each have an equal probability of occurring.

- (d) Find the exact value of  $p$  for the game to be fair using the biased die. [3]

[Solution]

(a)

$x$	-4	-1	2	3
$P(X=x)$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$

$$E(X) = \frac{-4}{4} + \frac{-1}{4} + \frac{2}{4} + \frac{3}{4} = 0$$

$$(b) \quad \text{Var}(X) = \left[ \frac{(-4)^2}{4} + \frac{(-1)^2}{4} + \frac{2^2}{4} + \frac{3^2}{4} \right] - 0^2 = \frac{15}{2}$$

(c) Let  $Y = |X_1 + X_2|$

$X_1 \backslash X_2$	-4	-1	2	3
-4	8	5	2	1
-1	5	2	1	2
2	2	1	4	5
3	1	2	5	6

$y$	1	2	4	5	6	8
$P(Y=y)$	$\frac{4}{16}$	$\frac{5}{16}$	$\frac{1}{16}$	$\frac{4}{16}$	$\frac{1}{16}$	$\frac{1}{16}$

$$\begin{aligned}
 E(Y) &= \frac{4}{16} + 2\left(\frac{5}{16}\right) + 4\left(\frac{1}{16}\right) + 5\left(\frac{4}{16}\right) + 6\left(\frac{1}{16}\right) + 8\left(\frac{1}{16}\right) \\
 &= \frac{13}{4}
 \end{aligned}$$

(d) The new probability distribution for  $X$  is as tabulated below:

$x$	-6	-4	2	3	5	7
$P(X=x)$	$\frac{1-3p}{4}$	$\frac{1-3p}{4}$	$\frac{1-3p}{4}$	$p$	$2p$	$\frac{1-3p}{4}$

$$E(X) = -6\left(\frac{1-3p}{4}\right) - 4\left(\frac{1-3p}{4}\right) + 2\left(\frac{1-3p}{4}\right) + 3p + 10p + 7\left(\frac{1-3p}{4}\right)$$

$$= -\left(\frac{1-3p}{4}\right) + 13p$$

$$= -\frac{1}{4} + \frac{55}{4}p$$

For the game to be fair,  $-\frac{1}{4} + \frac{55}{4}p = 0$

$$\text{Hence } p = \frac{1}{55}$$

- 8 A study was conducted on a group of pre-diabetic seniors to examine whether the number of hours of exercise per week ( $t$ ) has any effect on their fasting blood sugar levels ( $S$ , measured in mg/dL). The results were as follows.

$t$	1.5	1.8	2.1	2.4	2.5	3.1	3.5	4.2
$S$	119	108	100	95	93	88	86	85

- (a) Draw the scatter diagram for these values, labelling the axes. [1]
- (b) Comment on whether a linear model would be appropriate, referring both to the scatter diagram and the context of the question. [2]
- (c) By considering the shape of the following graphs, explain which model is the most appropriate for the data collected, where  $a$  and  $b$  are positive constants.

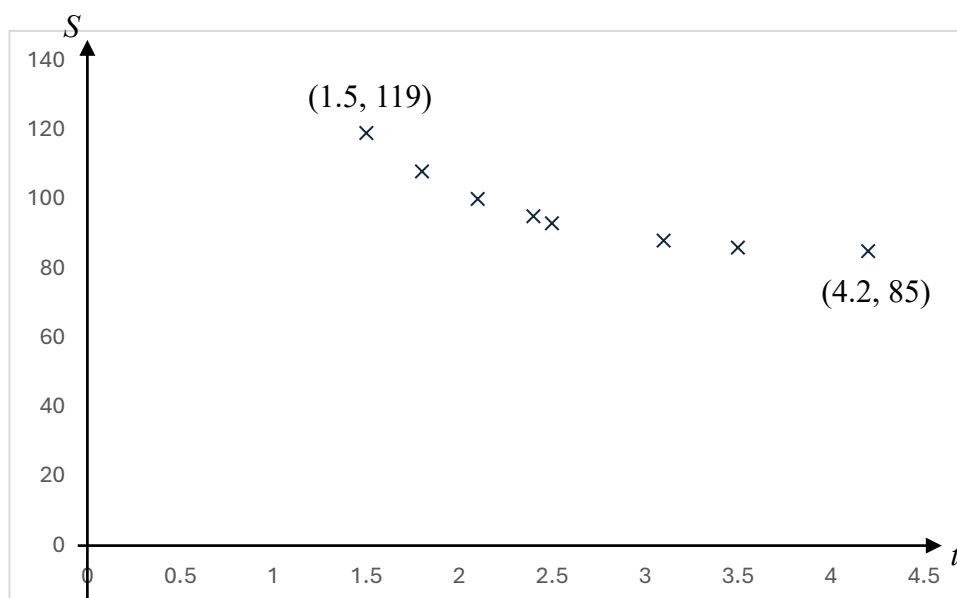
$$\text{A: } S = a + bt^2, \quad \text{B: } S = a + \frac{b}{t}, \quad \text{C: } S = a + b \ln t. \quad [1]$$

Using the appropriate model selected above, find

- (d) the value of the product moment correlation coefficient, [1]
- (e) the equation of the regression line of  $S$  on  $t$ , and hence find an estimate of the exercise time required for the fasting blood sugar to be 99 mg/dL. Comment on the reliability of your answer. [3]

**Solution**

- (a) B1 – points are as seen in this diagram with axes labelled with starting and ending points.



(b)

The scatter diagram shows that the relationship between  $s$  and  $t$  is non-linear, decreases at a decreasing rate.

Also, if the linear model is used, the blood sugar level cannot be decreased till it reaches negative values.

(c) From the scatter diagram, we see that  $y$  decreases at a decreasing rate as  $x$  increases, which is the case for model **B** ( $S = a + \frac{b}{t}$ ,  $b > 0$ ).

For models **A** ( $S = a + bt^2$ ) and **C** ( $S = a + b \ln t$ ),  $y$  increases as  $x$  increases, since  $b > 0$ .

Hence model **B** is appropriate. [B1 – selection with explanation]

(d)  $r = 0.988$  (3.s.f.)

NORMAL FLOAT AUTO REAL RADIAN MP					
<b>LinReg</b>					
$y=ax+b$					
$a=81.66051209$					
$b=62.43758456$					
$r^2=0.9762710097$					
$r=0.9880642741$					
L1	L2	L3	L4	L5	o
1.5	119	0.6667	-----	-----	
1.8	108	0.5556			
2.1	100	0.4762			
2.4	95	0.4167			
2.5	93	0.4			
3.1	88	0.3226			
3.5	86	0.2857			
4.2	85	0.2381			
-----	-----	-----			
L3(1)=0.66666666666667					

(e)  $S = \frac{81.6605}{t} + 62.4378$

$$S = \frac{81.7}{t} + 62.4 \text{ (3s.f.)}$$

When  $S = 99$

$$99 = \frac{81.6605}{t} + 62.4378$$

$$t = 2.23$$

Result is reliable as  $r$  is close to 1 and  $S = 99$  is within the data range of [85, 119], hence it is interpolation.

**9 In this question, you should state clearly all the distributions that you use, together with the values of the appropriate parameters.**

A toy designer creates weighted wooden blocks used in balance-based educational toys. There are two types of blocks: cylindrical blocks and cuboidal blocks. The mass, in grams, of each cylindrical block and cuboidal block follow a normal distribution with the following parameters:

	Mean	Standard Deviation
cylindrical block	280	6
cuboidal block	220	5

- (a) Find the probability that the mass of a randomly selected cylindrical block exceeds the mass of a randomly selected cuboidal block by more than 65 grams. [2]
- (b) The probability that the total of twice the mass of a cylindrical block and thrice the mass of a cuboidal block being less than  $M$  g is 0.355. Find the value of  $M$ , giving your answer to the nearest integer. [2]

To build a deluxe seesaw model, the designer uses:

- 2 cylindrical blocks,
- 3 cuboidal blocks,
- 7 screws,
- 1 wooden plank.

The blocks are drilled before assembly to fit screws:

- Each cylindrical block loses 5% of its mass after drilling.
- Each cuboidal block loses 3% of its mass after drilling.

The masses of the remaining components follow a normal distribution with the following parameters:

	Mean	Standard Deviation
screw	8	1.2
wooden plank	540	8

When assembling the deluxe seesaw model, all of its components are chosen at random.

- (c) Find the probability that the total mass of all the modified components in a deluxe seesaw model is more than 1785 g. [4]

[Solution]

- (a) Let  $X$  and  $Y$  be the mass of a randomly selected cylindrical and cuboidal block (in g) respectively.

$$X \sim N(280, 6^2) \text{ and } Y \sim N(220, 5^2)$$

$$X - Y \sim N(280 - 220, 5^2 + 6^2)$$

$$X - Y \sim N(60, 61)$$

$$P(X - Y > 65) = 0.261$$

- (b) Let  $C = 2X + 3Y$

$$C \sim N(2 \times 280 + 3 \times 220, 2^2 \times 6^2 + 3^2 \times 5^2)$$

$$C \sim N(1220, 369)$$

$$\text{Given } P(C < M) = 0.355$$

$$\text{From GC, } M = 1212.8568$$

$$M = 1213$$

- (c) Let  $S$  and  $W$  be the mass of a randomly selected screw and wooden plank (in g) respectively.

$$S \sim N(8, 1.2^2) \text{ and } W \sim N(540, 8^2)$$

$$\text{Let } T = 0.95(X_1 + X_2) + 0.97(Y_1 + Y_2 + Y_3) + S_1 + \dots + S_7 + W$$

$$E(T) = (0.95 \times 2 \times 280) + (0.97 \times 3 \times 220) + (7 \times 8) + 540$$

$$= 532 + 426.8 + 56 + 540$$

$$= 1768.2$$

$$\text{Var}(T) = (0.95^2 \times 2 \times 6^2) + (0.97^2 \times 3 \times 5^2) + (7 \times 1.2^2) + 64$$

$$= 64.98 + 70.5675 + 10.08 + 64$$

$$= 209.6275$$

$$P(T > 1785) = 0.123$$

**10** Green Earth Factory produces solar panels. On average, 2% of the solar panels produced are found to contain defects. The solar panels are packed in batches of 10.

- (a) State, in the context of solar panel production, two assumptions needed for the number of defective solar panels in a batch to be well modelled by a binomial distribution. [2]

It is now given that the number of defective solar panels per batch follows a binomial distribution.

- (b) Find the probability that a batch contains at least 2 defective solar panels. [2]

As part of Green Plan initiative, the Education Department plans to install solar panels on the rooftops of 100 primary schools to reduce carbon footprint. Each school will be fitted with 10 solar panels, and the school will lodge a complaint if there are at least two defective panels in their allocation.

- (c) Find the probability that at most 5 schools lodge complaints. [2]

Green Earth Factory also produces LED lights with a defect rate of 1%. The number of defective LED lights also follows a binomial distribution. As part of the same green initiative, the Education Department orders installations for the 100 primary schools, with each school receiving another 25 LED lights. For quality control purposes, the Education Department will fine the factory if the total number of defective items (combining both solar panels and LED lights) across all 100 schools exceeds 50 units.

- (d) By using an approximation, find the probability that Green Earth Factory receives a fine. [5]

Solution

- (a) Whether a solar panel is defective is independent of whether any other solar panel is faulty.

The probability of a randomly chosen solar panel being defective is constant at 0.02.

- (b) Let  $S$  be the number of defective solar panels out of 10 panels.

$$S \sim B(10, 0.02)$$

$$P(S \geq 2) = 1 - P(S \leq 1)$$

$$= 0.01617764$$

$$= 0.0162 \text{ (3sf)}$$

- (c) Let  $Y$  be the number of schools that lodge complaint out of 100 schools.

$$Y \sim B(100, 0.016177)$$

$$P(X \leq 5) = 0.994$$



(d) Let  $L$  be the number of defective LED lights out of 25 lights.

$$L \sim B(25, 0.01)$$

$$E(L) = 25 \times 0.01 = 0.25 \quad \text{Var}(L) = 25 \times 0.01 \times 0.99 = 0.2475$$

$$E(S) = 10 \times 0.02 = 0.2 \quad \text{Var}(S) = 10 \times 0.02 \times 0.98 = 0.196$$

$$\text{Let } A = L_1 + L_2 + \cdots + L_{100} \text{ and } B = S_1 + S_2 + \cdots + S_{100}$$

$$E(A) = 100 \times 0.25 = 25 \quad \text{Var}(A) = 100 \times 0.2475 = 24.75$$

$$E(B) = 100 \times 0.2 = 20 \quad \text{Var}(B) = 100 \times 0.196 = 19.6$$

Since 100 is large, by Central Limit Theorem,

$$A \sim N(25, 24.75) \text{ and } B \sim N(20, 19.6) \text{ approximately.}$$

Therefore,  $A + B \sim N(45, 44.35)$  approximately.

Probability that Green Earth Factory receives a fine from MOE

$$= P(A + B > 50)$$

$$\approx 0.226$$

- 11 (a)** In a survey conducted by a library, a group of 100 members were asked the number of days in a month they visit the library. The results are shown below:

	1 or less	2 to 4	5 or more
Male	15	25	22
Female	20	$18 - n$	$n$

A member is selected at random from the group. The events are defined as follows:

$A$ : The member visits the library on 4 days or less in a month.

$B$ : The member visits the library on 2 days or more in a month.

$F$ : The member is a female.

Find the following probabilities in terms of  $n$ .

**(i)**  $P(A \cup B')$  [1]

**(ii)**  $P(F | A')$  [2]

**(iii)** Given that  $P(A \cap B) = \frac{3}{10}$ , find the value of  $n$ . Hence determine if  $A$  and  $B$  are independent, justifying your answer. [3]

- (b)** A librarian is organising a large community reading event. To manage registration, he needs to create 4-letter access codes for the participants using the letters from the word BOOKKEEPER.

**(i)** How many access codes can be formed if all 4 letters are distinct from one another? [1]

**(ii)** Find the total number of access codes that can be created given that there is no restriction. [3]

To run an event, the librarian plans to select a team of 12 volunteers from a group of 18. The volunteers comprise three age groups: 5 youths, 6 young adults, and 7 seniors.

**(iii)** In how many ways can the team be formed if at least one volunteer must be chosen from each age group? [3]

[Solution]

**(a)(i)**  $P(A \cup B') = P(A)$

$$= \frac{15 + 25 + 20 + (18 - n)}{100} \text{ or } \frac{100 - 22 - n}{100}$$

$$= \frac{78 - n}{100}$$

$$(ii) \quad P(F | A') = \frac{P(F \cap A')}{1 - P(A)} = \frac{\frac{n}{100}}{1 - \frac{78-n}{100}} = \frac{n}{22+n}$$

$$(iii) \quad \text{Given } P(A \cap B) = \frac{3}{10}$$

$$P(A \cap B) = \frac{25+18-n}{100} = \frac{43-n}{100}$$

$$\frac{43-n}{100} = \frac{3}{10} \Rightarrow n = 13$$

$$P(A)P(B) = \frac{65}{100} \times \frac{65}{100} = \frac{169}{400} \neq \frac{3}{10}$$

Since  $P(A)P(B) \neq P(A \cap B)$ ,  $A$  and  $B$  are **NOT** independent

**Alternatively**, comparing  $P(A|B)$  with  $P(A)$

$$\begin{aligned} P(A|B) &= \frac{P(A \cap B)}{P(B)} \\ &= \frac{\frac{3}{100}}{\frac{65}{100}} = \frac{3}{65} \end{aligned}$$

Since  $P(A) = \frac{13}{20}$  and  $P(A|B) \neq P(A)$ ,  $A$  and  $B$  are **NOT** independent

(b)(i) 3Es, 2Os, 2Ks, 1P, 1R, 1B

Number of 4-letter code words =  ${}^6C_4 \times 4! = 360$  or  ${}^6P_4 = 360$

(ii) Case 1: All distinct (from b(i))

Number of 4-letter code words =  ${}^6C_4 \times 4! = 360$

Case 2: One pair of repeated letters. (Eg XXYZ)

Number of 4-letter code words =  ${}^3C_1 \times {}^5C_2 \times \frac{4!}{2!} = 360$

Case 3: Two pairs of repeated letters. (Eg XXYY)

$$\text{Number of 4-letter code words} = {}^3C_2 \times \frac{4!}{2!2!} = 18$$

Case 4: One set of 3 repeated letters (Eg EEEX)

$$\text{Number of 4-letter code words} = {}^5C_1 \times \frac{4!}{3!} = 20$$

$$\therefore \text{Total number of 4-letter code word} = 360 + 360 + 18 + 20 = 758$$

(iii)  $\text{Total number of ways} = {}^{18}C_{12} = 18564$

$$\text{No youth} = {}^{13}C_{12} = 13$$

$$\text{No young adult} = {}^{12}C_{12} = 1$$

$$\text{Total possible ways} = 18564 - 13 - 1 = 18550$$

- 12** Flour is packed in bags of 1.5 kg, as stated on the packaging, by two separate machines *A* and *B* in a factory. During a quality control inspection, the production manager selects a random sample of 40 bags filled by machine *A* for checking. The masses,  $x$  kg, of these 40 bags are recorded as follows.

$$\sum x = 60.32 \quad \sum x^2 = 90.993$$

- (a) State what it means for a sample to be random in this context. [1]
- (b) Find unbiased estimates of the population mean and variance of the mass of the filled bags. [2]
- (c) Test, at the 5% level of significance, whether machine *A* produces overweight bags. You should state your hypotheses and define any symbols you use. [5]
- (d) Explain in the context of the question, the meaning of ' $p$ -value'. [1]
- (e) The manager also wants to test whether the mean mass of the filled bags by machine *B* differs from 1.5 kg. It is known that the mass of the filled bags from machine *B* is normally distributed with variance 0.00198 kg<sup>2</sup>. A sample of size  $n$  is randomly selected and its mean is found to be 1.489 kg. Find the range of values that  $n$  can take if there is sufficient evidence to conclude that the mean mass of the filled bags differs from 1.5 kg, at the 5% level of significance. [3]

Solution

- (a) The sample is random when every filled bag of flour has an equal probability of being selected for the sample and the bags for the sample are selected independently.
- (b) Unbiased estimate of the population mean,  $\bar{x} = \frac{60.32}{40} = 1.508$

$$\begin{aligned}
 \text{Unbiased estimate of the population variance, } s^2 &= \frac{1}{39} \left( \sum x^2 - \frac{(\sum x)^2}{40} \right) \\
 &= \frac{1}{39} \left( 90.993 - \frac{60.32^2}{40} \right) \\
 &= 0.0007805128 \\
 &= 0.000781 \text{ (3sf)}
 \end{aligned}$$

- (c) Let  $X$  be the random variable denoting the mass of a randomly chosen bag of flour (in kg) filled by machine  $A$ .

Let  $\mu$  denote the population mean mass of bags of flour filled by machine  $A$  in kg.

$$H_0: \mu = 1.5$$

$$H_1: \mu > 1.5 \quad \text{at 5\% significance level}$$

Under  $H_0$ , since  $n = 40$  is large, by Central Limit Theorem,

$$\bar{X} \sim N\left(1.5, \frac{0.000780513}{40}\right) \text{ approximately or } Z = \frac{\bar{X} - 1.5}{\sqrt{\frac{0.000780513}{n}}} \sim N(0,1)$$

approximately

$$\text{Test statistic: } z = \frac{1.508 - 1.5}{\sqrt{\frac{0.000780513}{40}}} = 1.811$$

$$p\text{-value} = 0.0350667 = 0.0351 \text{ (to 3 s.f.)}$$

Since  $p\text{-value} < 0.05$ , we reject  $H_0$  and conclude there is sufficient evidence that Machine  $A$  produces overweight bags at 5% level of significance.

- (d) It is the probability of obtaining the sample mean mass of the 40 randomly filled bags to be more than 1.508kg when the population mean mass of the bag of flour is 1.5 kg.

- (e) To test  $H_0: \mu = 1.5$

Against  $H_1: \mu \neq 1.5$  at 5% level of significance

Let  $Y$  be the random variable denoting the mass of a randomly chosen bag of flour (in kg) filled by machine  $B$ .

$$\text{Under } H_0, \bar{Y} \sim N\left(1.5, \frac{0.00198}{n}\right)$$

$$\text{Since } H_0 \text{ is rejected, } \frac{1.489 - 1.5}{\sqrt{\frac{0.00198}{n}}} < -1.95996 \quad \text{OR} \quad \frac{1.489 - 1.5}{\sqrt{\frac{0.00198}{n}}} > 1.95996$$

$$\sqrt{n} > 7.9284$$

$$n > 62.8602$$

Solution set is

$$\{n \in \mathbb{Z} : n > 62.9\}$$

No solution since a negative number (LHS) cannot be more than a positive number (RHS)