Section A: Pure Mathematics [40 marks]

1 A sequence u_1, u_2, u_3, … is such that $u_1 = 2$ and

$$u_n = u_{n-1} - \frac{1}{n(n-1)}, \quad \text{for all integers } n \geq 2.$$

(i) By expressing $\frac{1}{r(r-1)}$ in partial fractions or otherwise, find $\sum_{r=2}^{n} \frac{1}{r(r-1)}$ in terms of n. [4]

(ii) By considering $\sum_{r=2}^{n}(u_r - u_{r-1})$ and using the result in part **(i)**, show that for all integers $n \geq 2$, u_n can be expressed in the form $a + \frac{b}{n}$, where a and b are constants to be determined. [3]

1i	Let $\frac{1}{r(r-1)} = \frac{A}{r} + \frac{B}{r-1}$ $A(r-1) + Br = 1$ Sub $r = 0 \Rightarrow A = -1$ Sub $r = 1 \Rightarrow B = 1$ So $\frac{1}{r(r-1)} = \frac{1}{r-1} - \frac{1}{r}$ $\sum_{r=2}^{n} \frac{1}{r(r-1)} = \sum_{r=2}^{n}\left(\frac{1}{r-1} - \frac{1}{r}\right)$ $= \frac{1}{1} - \cancel{\frac{1}{2}}$ $+ \cancel{\frac{1}{2}} - \cancel{\frac{1}{3}}$ $\vdots$ $\vdots$ $+ \cancel{\frac{1}{n-2}} - \cancel{\frac{1}{n-1}}$ $+ \cancel{\frac{1}{n-1}} - \frac{1}{n}$ $= 1 - \frac{1}{n}$

1ii	$\sum_{r=2}^{n}(u_r - u_{r-1}) = -\sum_{r=1}^{n}\frac{1}{r(r-1)}$ $\begin{array}{ll} \cancel{u_2} & -u_1 \\ +\cancel{u_3} & -\cancel{u_2} \\ \quad\vdots & \\ +\cancel{u_{n-1}} & -\cancel{u_{n-2}} \\ +u_n & -\cancel{u_{n-1}} \end{array} = -\left(1-\frac{1}{n}\right)$ $u_n = u_1 - 1 + \frac{1}{n} = 1 + \frac{1}{n}$

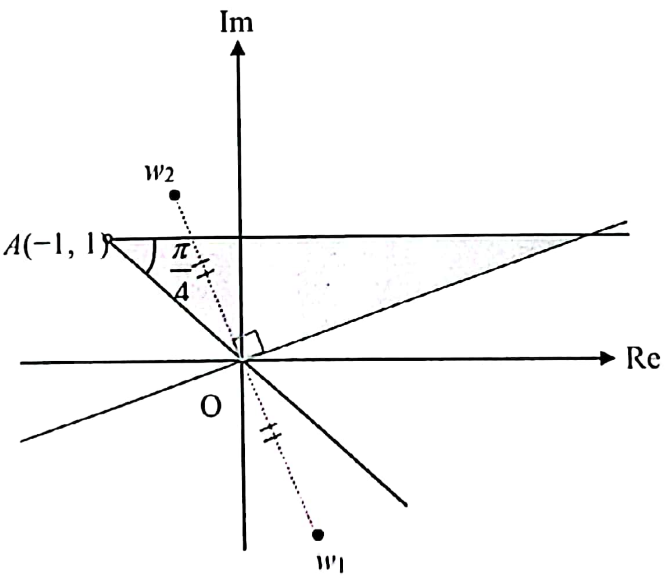
2 Solve the equation $w^4+2-2\sqrt{3}\,\mathrm{i}=0$, giving the roots in the form $re^{\mathrm{i}\theta}$, where $-\pi<\theta\le\pi$ and $r>0$. [4]

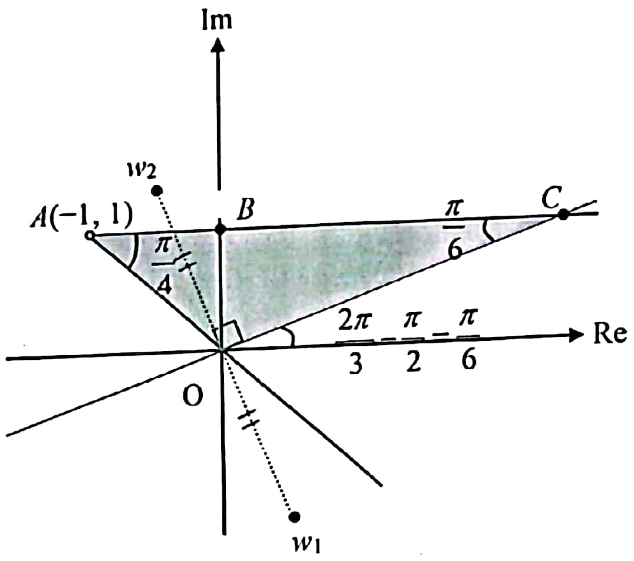
The roots represented by w_1 and w_2 are such that $-\frac{\pi}{2}<\arg(w_1)<0$ and $\frac{\pi}{2}<\arg(w_2)<\pi$.

The complex number z satisfies the relations $|z-w_1|\ge|z-w_2|$ and $-\frac{\pi}{4}\le\arg[z-(-1+\mathrm{i})]\le 0$.

On an Argand diagram, sketch the region R in which the points representing z can lie. [3]

Find the exact area of R. [3]

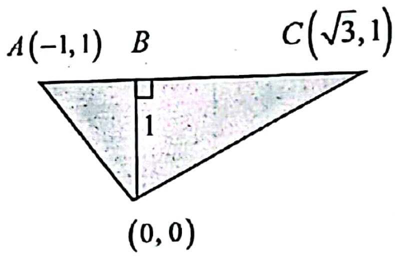
2	$w^4+2-2\sqrt{3}\,\mathrm{i}=0 \Rightarrow w^4=-2+2\sqrt{3}\,\mathrm{i}$ $=4e^{\mathrm{i}\frac{2\pi}{3}}$ $=4e^{\mathrm{i}\left(\frac{2\pi}{3}+2k\pi\right)}$ $\Rightarrow w=(4)^{\frac{1}{4}}\left[e^{\mathrm{i}\left(\frac{2\pi}{3}+2k\pi\right)}\right]^{\frac{1}{4}}=\sqrt{2}e^{\mathrm{i}\left(\frac{\pi}{6}+\frac{k\pi}{2}\right)}$, where $k=0,\pm1,-2$ $\therefore w=\sqrt{2}e^{\mathrm{i}\frac{\pi}{6}},\ \sqrt{2}e^{\mathrm{i}\frac{2\pi}{3}},\ \sqrt{2}e^{\mathrm{i}\left(-\frac{5\pi}{6}\right)},\ \sqrt{2}e^{\mathrm{i}\left(-\frac{\pi}{3}\right)}$
	$\therefore w_1=\sqrt{2}e^{\mathrm{i}\left(-\frac{\pi}{3}\right)},\ w_2=\sqrt{2}e^{\mathrm{i}\frac{2\pi}{3}}$ 



$$\tan\frac{\pi}{6} = \frac{1}{BC}$$
$$BC = \sqrt{3}$$

Area of shaded region

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



3 The line l has equation $\dfrac{x+1}{-1}=\dfrac{z+6}{2}$, $y=4$ and the point A has coordinates $(-1, 3, -5)$.

(i) Find the position vector of the foot of the perpendicular from A to l. [3]

(ii) Plane p_1 contains l and A. Show that the equation of p_1 is $2x+y+z=-4$. [3]

Given that the plane p_2 has equation $x + 2y + cz = -5$ where c is a negative constant, and that the acute angle between p_1 and p_2 is 60°, find the value of c. [3]

(iii) Find the equation of the line of intersection, m, between p_1 and p_2. [1]

(iv) The plane p_3 has equation $3x + \alpha y + 7z = \beta$ where α and β are constants. Given that the planes p_1, p_2 and p_3 have no common point, what can be said about the values of α and β? [2]

3i

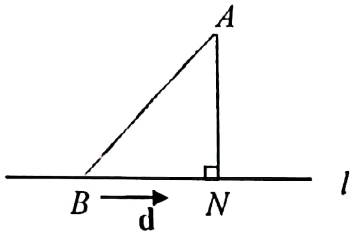
Let $\lambda=\dfrac{x+1}{-1}=\dfrac{z+6}{2}$, $\quad y=4$

$x=-1-\lambda,\quad z=-6+2\lambda,\quad y=4$

$$l:\ \mathbf{r}=\begin{pmatrix}-1\\4\\-6\end{pmatrix}+\lambda\begin{pmatrix}-1\\0\\2\end{pmatrix},\ \lambda\in\square$$

Let N be the foot of the perpendicular from A to l.

$$\overrightarrow{ON}=\begin{pmatrix}-1\\4\\-6\end{pmatrix}+\lambda\begin{pmatrix}-1\\0\\2\end{pmatrix}=\begin{pmatrix}-1+\lambda\\4\\-6+2\lambda\end{pmatrix}\quad\text{for some }\lambda\in\mathbb{R}$$



$$\Rightarrow\ \overrightarrow{AN}=\begin{pmatrix}-\lambda\\1\\-1+2\lambda\end{pmatrix}$$

$$\overrightarrow{AN}\perp l\ \text{ i.e. }\begin{pmatrix}-\lambda\\1\\-1+2\lambda\end{pmatrix}\square\begin{pmatrix}-1\\0\\2\end{pmatrix}=0\ \Rightarrow\lambda=\frac{2}{5}$$

$$\text{Thus }\overrightarrow{ON}=\frac{1}{5}\begin{pmatrix}-7\\20\\-26\end{pmatrix}$$

3ii

Let B be the point on l with coordinates $(-1, 4, 6)$

$$\overrightarrow{BA}=\begin{pmatrix}-1\\3\\-5\end{pmatrix}-\begin{pmatrix}-1\\4\\-6\end{pmatrix}=\begin{pmatrix}0\\-1\\1\end{pmatrix}\quad\text{or}\quad\overrightarrow{AN}=-\frac{1}{5}\begin{pmatrix}2\\-5\\1\end{pmatrix}$$

A normal to p_1 is $\begin{pmatrix}-1\\0\\2\end{pmatrix}\times\begin{pmatrix}0\\-1\\1\end{pmatrix}=\begin{pmatrix}2\\1\\1\end{pmatrix}$ or $\begin{pmatrix}-1\\0\\2\end{pmatrix}\times\begin{pmatrix}2\\-5\\1\end{pmatrix}=5\begin{pmatrix}2\\1\\1\end{pmatrix}$

Equation of p_1 is $\mathbf{r}\bullet\begin{pmatrix}2\\1\\1\end{pmatrix}=\begin{pmatrix}-1\\3\\-5\end{pmatrix}\bullet\begin{pmatrix}2\\1\\1\end{pmatrix}=-4$ i.e $2x+y+z=-4$

Equation of p_2 is $\mathbf{r}\bullet\begin{pmatrix}1\\2\\a\end{pmatrix}=-5$

$$\cos 60^\circ=\frac{\left|\begin{pmatrix}2\\1\\1\end{pmatrix}\bullet\begin{pmatrix}1\\2\\c\end{pmatrix}\right|}{\sqrt{6}\sqrt{5+c^2}}$$

$$\frac{1}{2}=\frac{|4+c|}{\sqrt{6}\sqrt{5+c^2}}$$

$30+6c^2=4\left(c^2+8c+16\right)$

$c^2-16c-17=0$

$(c-17)(c+1)=0$

$c=17$ (rejected since $c<0$) or $c=-1$

3iii

p_1: $2x+y+z=-4$

p_2: $x+2y+cz=-5$

Using GC, $x=-1-t$

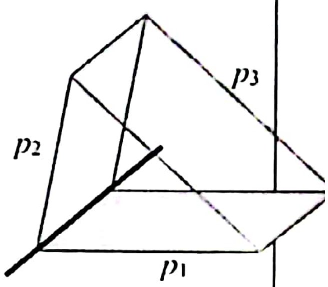
$y=-2+t$

$z=t$

Equation of m is $\mathbf{r}=\begin{pmatrix}-1\\-2\\0\end{pmatrix}+t\begin{pmatrix}-1\\1\\1\end{pmatrix}$, $t\in\mathbb{R}$

3iv

Since the <u>3 planes are not parallel</u> and <u>they have no common point</u>, m is parallel to p_3 but not contained in p_3.



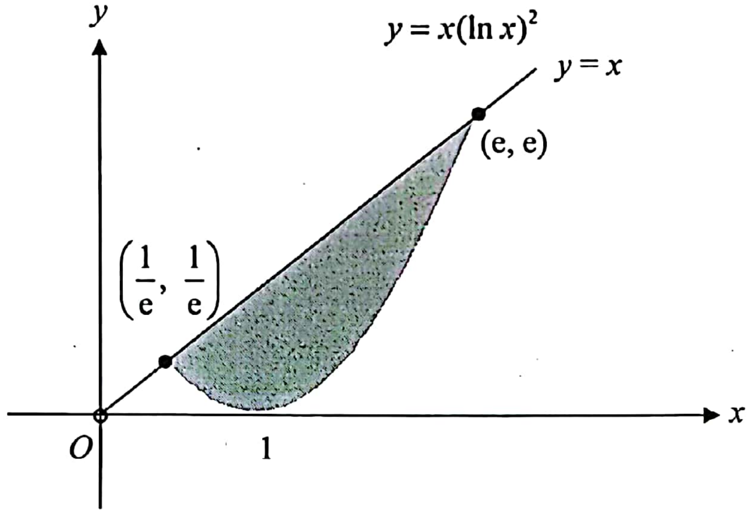
m is perpendicular to $\mathbf{n}_3$: $\begin{pmatrix}3\\\alpha\\7\end{pmatrix}\bullet\begin{pmatrix}-1\\1\\1\end{pmatrix}=0\Rightarrow\alpha=-4$

Also $(-1,-2,0)$ on m does not lie in p_3: $3x+\alpha y+7z=\beta$

Thus $3(-1)+(-4)(-2)+7(0)\neq\beta\Rightarrow\beta\neq5$

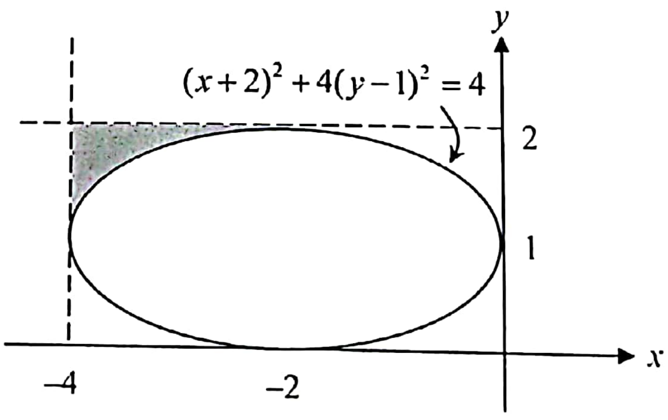
4 (a) The diagram below shows the graphs of $y = x(\ln x)^2$ and $y = x$. The two graphs intersect at the points $\left(\frac{1}{e}, \frac{1}{e}\right)$ and (e, e).

Find the exact area of the shaded region bounded by the graphs of $y = x(\ln x)^2$ and $y = x$. [5]



Hence, without integrating, find the exact area of the region bounded by the graphs of $y = x(\ln x)^2$ and the lines $y = e$ and $x = \frac{1}{e}$. [2

(b) Find the volume of the solid formed when the shaded region bounded by the line $x = -4$, $y = 2$ and the ellipse $(x+2)^2 + 4(y-1)^2 = 4$ is rotated through 2π radian about the y-axis. Give your answer correct to 1 decimal place. [4



4a

Area of shaded region

$$=\frac{1}{2}\left(e+\frac{1}{e}\right)\left(e-\frac{1}{e}\right)-\int_{\frac{1}{e}}^{e}x(\ln x)^2\,dx$$

$$=\frac{1}{2}\left(e^2-\frac{1}{e^2}\right)-\left(\left[\frac{x^2}{2}(\ln x)^2\right]_{\frac{1}{e}}^{e}-\int_{\frac{1}{e}}^{e}\frac{x^2}{2}\,\frac{2\ln x}{x}\,dx\right)$$

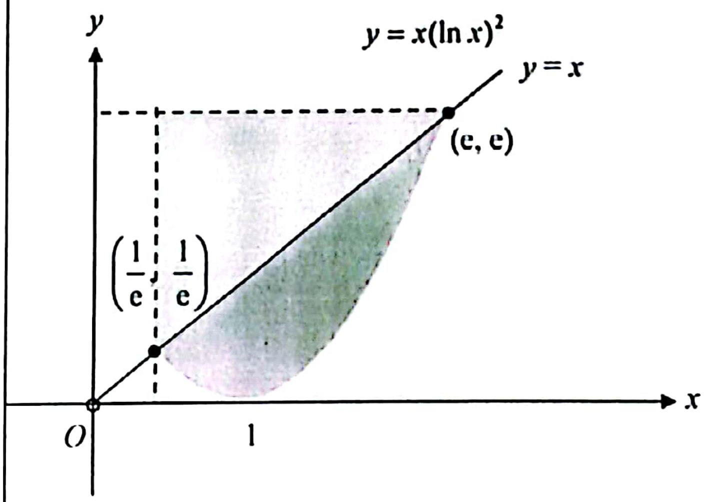
$$=\frac{1}{2}\left(e^2-\frac{1}{e^2}\right)-\left(\left[\frac{e^2}{2}-\frac{1}{2e^2}\right]-\int_{\frac{1}{e}}^{e}x\ln x\,dx\right)$$

$$=\frac{1}{2}\left(e^2-\frac{1}{e^2}\right)-\frac{1}{2}\left(e^2-\frac{1}{e^2}\right)+\left(\left[\frac{x^2}{2}\ln x\right]_{\frac{1}{e}}^{e}-\int_{\frac{1}{e}}^{e}\frac{x^2}{2}\,\frac{1}{x}\,dx\right)$$

$$=\frac{1}{2}\left(e^2+\frac{1}{e^2}\right)-\frac{1}{2}\left[\frac{x^2}{2}\right]_{\frac{1}{e}}^{e}$$

$$=\frac{1}{2}\left(e^2+\frac{1}{e^2}\right)-\frac{1}{2}\left(\frac{e^2}{2}-\frac{1}{2e^2}\right)$$

$$=\frac{1}{4}\left(e^2+\frac{3}{e^2}\right)$$



$$\text{Area}=\frac{1}{4}\left(e^2+\frac{3}{e^2}\right)+\frac{1}{2}\left(e-\frac{1}{e}\right)^2$$

4b

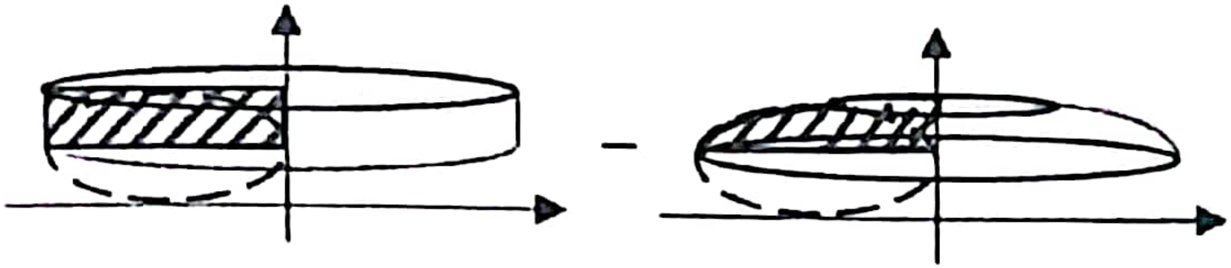
Given $(x+2)^2+4(y-1)^2=4$

$$(x+2)^2=4[1-(y-1)^2]\Rightarrow x+2=\pm2\sqrt{1-(y-1)^2}$$

The shaded region is bounded by the section of the ellipse where $x\le -2$.

Hence $x=-2-2\sqrt{1-(y-1)^2}$

Volume of solid formed



$$= \pi 4^2(2-1) - \pi\int_1^2 x^2 \, dy$$

$$= 16\pi - \pi\int_1^2 \left(-2-2\sqrt{1-(y-1)^2}\right)^2 dy$$

$$= 9.6 \text{ units}^3$$

Section B: Statistics [60 marks]

5 An apartment block has 24 two-bedroom apartments and 88 four-bedroom apartments. A surveyor wishes to conduct interviews on 35 households in this block to learn about their household expenditure. She decides to use stratified sampling across apartment types in the block, assuming that only one household occupies each apartment.

(i) Give an advantage of this sampling method. [1]

(ii) Describe how a stratified sample can be obtained. [3]

5i	Stratified sampling ensures that households occupying two-bedroom apartments and households occupying four-bedroom apartments are **proportionately represented** in the sample.
5ii	Total number of units = 24 + 88 = 112 Number of 2-bedroom households needed = $\frac{24}{112}\times 35 = 7.5$ Number of 4-bedroom households needed = $\frac{88}{112}\times 35 = 27.5$ She should choose 8 two-bedroom apartments and 27 four-bedroom apartments (or 7 two-bedroom apartments and 28 four-bedroom apartments) Using the list of apartment numbers as sampling frames for the two types of apartments, she would use simple random sampling (or systematic sampling) to select the 2-bedroom apartments and 4-bedroom apartments to be interviewed.

6 A mathematician arranges all eight letters in the word PARALLEL to form different 8-letter code words. Find the number of different code words that can be formed if

(i) the code words start with an L and end with an A, [1]

(ii) the 2 A's are not adjacent to each other, [3]

(iii) there is exactly one letter between the first and second L, and exactly one letter between the second and third L. [3]

6i	Number of ways $=\dfrac{6!}{2!}=360$ P AA R LLL E
6ii	**Method 1** (By Slotting) Number of ways $=\binom{7}{2}\dfrac{6!}{3!}=2520$ **Method 2** (Complement) Number of ways to arrange letters without restriction $=\dfrac{8!}{2!3!}=3360$ Number of ways to arrange letters with both A's together $=\dfrac{7!}{3!}=840$ Total number of ways $=\dfrac{8!}{2!3!}-\dfrac{7!}{3!}=2520$
6iii	Number of ways $=4\times\dfrac{5!}{2!}=240$ L _ L _ L _ _ _ **Method 2** Number of ways $=\binom{5}{2}\times 2!\times\dfrac{4!}{2!}=240$ Number of ways $=\dfrac{{}^5C_1\times{}^4C_1\times 4!}{2!}=240$ **Method 3** Case 1: LALAL_ _ _ Number of ways $=4!=24$ Case 2: LXLAL_ _ _ or LALXL_ _ _ Number of ways $=2\times\binom{3}{1}\times 4!=144$ Case 3: LXLYL_ _ _ or LYLXL_ _ _ Number of ways $=2\times\binom{3}{2}\times\dfrac{4!}{2!}=72$ Total 240

7 Anand, Beng and Charlie are selling cupcakes to raise funds for the charity, Boys And Girls Understand Singapore (BAGUS). Anand will bake 60% of the cupcakes, Beng will bake 40% of the cupcakes and Charlie will spread frosting on all the cupcakes.

20% of Anand's cupcakes will turn out flawed, while 12% of Beng's cupcakes will turn out flawed. Charlie has a 6% chance of spreading the frosting badly on any cupcake, regardless of whether it is flawed or otherwise.

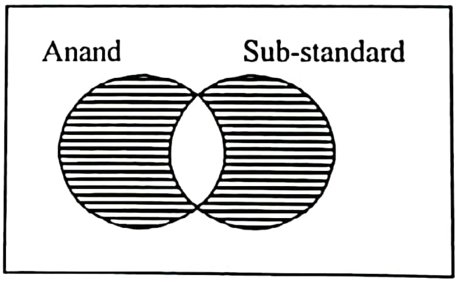
If a cupcake is flawed or has frosting spread badly (or both), it is considered sub-standard. Otherwise, a cupcake is considered "good".

Find the probability that

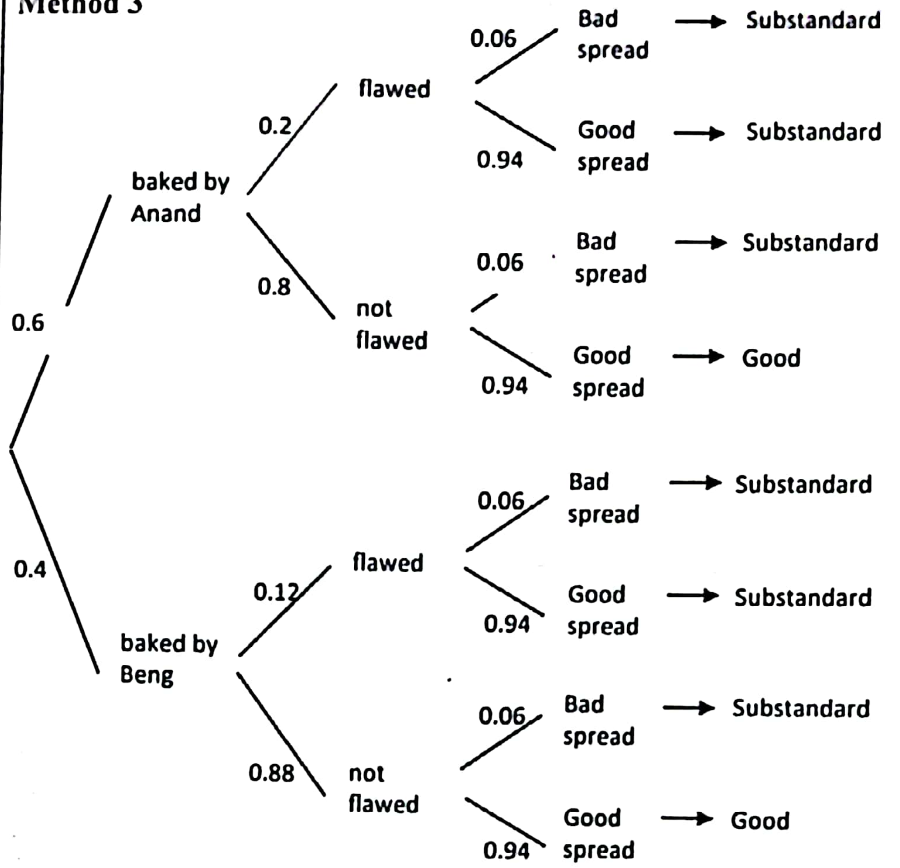
(i) a randomly chosen cupcake is "good", [2]

(ii) a randomly chosen cupcake is either baked by Anand or is sub-standard but not both, [3]

(iii) a randomly chosen cupcake is baked by Anand given that it is sub-standard. [3]

7i	P(a randomly chosen cupcake is "good") $= 0.6 \times 0.8 \times 0.94 + 0.4 \times 0.88 \times 0.94$ $= 0.78208 = 0.782$ (3sf)
7ii	P(cupcake is either baked by Anand, is sub-standard but not both)  = P(cupcake is baked by Anand and is good) + P(cupcake is baked by Beng and is sub-standard) $= 0.6 \times 0.8 \times 0.94 + 0.4 \times (0.12 + 0.88 \times 0.06)$ $= 0.520$ (3sf) **Method 2** $= 0.6 \times 0.8 \times 0.94 + 0.4 \times (1 - 0.88 \times 0.94)$ $= 0.520$ (3sf)

Method 3



=P(baked by Anand) + P(sub-standard) – 2 P(baked by Anand and sub-standard)

$= 0.6 + (1 - 0.78208) - 2\ (\ 0.6 \times (1 - 0.8 \times 0.94)\ )$

$= 0.520$ (3sf)

7iii P(cupcake baked by Anand | cupcake is sub-standard)

$$= \frac{\text{P(cupcake is sub-standard and baked by Anand)}}{\text{P(cupcake is sub-standard)}}$$

$$= \frac{0.6 \times (0.2 + 0.8 \times 0.06)}{1 - 0.78208} \quad \text{or} \quad \frac{0.6 \times (1 - 0.8 \times 0.94)}{1 - 0.78208}$$

$= 0.683$ (3sf)

8 The rate of growth, x units per hour, of a particular family of bacteria is believed to depend in some way on the controlled temperature T °C. Experiments were undertaken in the laboratory to investigate this and the results were tabulated as follows:

T	10	20	30	40	50	60	70	80
X	33	37	41	48	55	65	78	94

Draw a scatter diagram for these values, labelling the axes clearly. [1]

State, with a reason, which of the following model is most appropriate for the given data,

(A) $x = a + bT$ (B) $x = ae^{bT}$ (C) $x = a + b\ln T$

where a and b are constants, and $b > 0$. [1]

In addition, when $T = 90$, the rate of growth of bacteria was k units per hour.

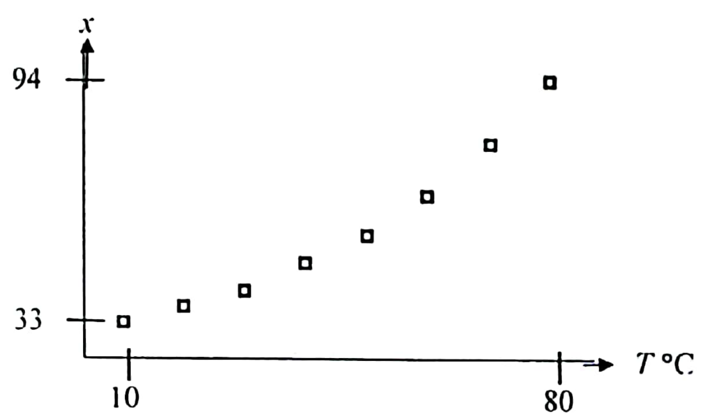
Use the most appropriate model identified for the subsequent parts of this question.

(i) A suitable linear regression line was constructed based on all 9 pairs of transformed data, including the additional pair of data. If the values of a and b were determined to be 27.06 and 0.01497 respectively, find the value of k, correct to the nearest whole number. [4]

(ii) Find the product moment correlation coefficient for all 9 pairs of transformed data. [1]

(iii) Use the regression line in **(i)** to predict the growth rate of the bacteria when the temperature is 105 °C, giving your answer to the nearest whole number. Comment on the reliability of this prediction. [2]

8



8

Model B is the most appropriate as the curve is concave upwards and x increases with T, which is the same as the shape shown in the scatter diagram.

OR

Model B is the most appropriate since x increases at an increasing rate as T increases, which is the same as the shape shown in the scatter diagram.

8i	**Regression line for Model B:** $\ln x = \ln a + bT$ $\ln x = \ln(27.06) + 0.01497T$ $\overline{T} = \dfrac{10+20+30+40+50+60+70+80+90}{9} = 50$ and $\dfrac{\sum \ln x}{9}$ $= \dfrac{\ln 33 + \ln 37 + \ln 41 + \ln 48 + \ln 55 + \ln 65 + \ln 78 + \ln 94 + \ln k}{9}$ $= \dfrac{31.77392262 + \ln k}{9}$ Since $\left(\dfrac{\sum \ln x}{9}, \overline{T}\right)$ lies on the regression line, $\dfrac{31.77392262 + \ln k}{9} = \ln(27.06) + 0.01497(50)$ $\Rightarrow k = 104$ (nearest whole number)
8ii	From GC, $r = 0.997$
8iii	When $T = 65$, $\ln x = \ln(27.06) + 0.01497(105)$ $x = 130$ (nearest whole number) The estimated growth rate cannot be taken as reliable as the temperature $105^{0}\mathrm{C}$, from which the value of $x = 130$ is computed from, lies outside the data range of T i.e. [10, 90].

9 Small defects in a twill weave and satin weave occur randomly and independently at a constant mean rate of 1.2 defects and 0.8 defects per square metre respectively.

(i) Find the probability that there are exactly 9 defects in 7 square metres of twill weave. [2]

(ii) A box is made from 2 square metres of twill weave and 3 square metres of satin weave, chosen independently. A box is considered "faulty" if there are more than 10 defects. Show that the probability that a randomly chosen box is "faulty" is 0.0104, correct to 3 significant figures. [2]

(iii) A random batch of 50 boxes is delivered to a customer once every week. The customer can reject the entire batch if there are at least 2 "faulty" boxes in the batch. Using a suitable approximation, find the probability that the customer will reject the entire batch in a randomly selected week. [3]
Hence estimate the probability that the customer will reject 2 batches in a randomly selected period of 4 weeks. [2]

9i	Let T be the number of defects in 7 m^2 of twill weave. $T \sim P_o(1.2\times 7)$, i.e. $T \sim P_o(8.4)$ $P(T=9)=0.129025899=0.129$
9ii	Let W be the number of defects in a box. $W \sim P_o(2\times 1.2+3\times 0.8)$, i.e. $W \sim P_o(4.8)$ $P(\text{box is faulty}) = P(W>10)$ $=1-P(W\le 10)$ $=0.0104$ (shown)
9iii	Let F be the number of "faulty" boxes in a random batch of 50 boxes. $F \sim \text{B}(50,\ 0.0104)$ Since n is large and $np = 0.52 < 5$, $F \sim \text{P}_0(0.52)$ approximately. P(customer rejects the entire batch) $=\text{P}(F\ge 2)$ $=1-\text{P}(F\le 1)$ ≈ 0.096329 $=0.0963$ or 0.0966 (3 sig fig) Required probability $\approx \binom{4}{2}(0.096329)^2(1-0.096329)^2$ $=0.0455$ or 0.0457 (3 sig fig) Alternative Solution: Let X be the number of batches out of 4 that are rejected. $X \sim \text{B}(4,\ 0.096329)$ $\text{P}(X=2)=0.0455$ or 0.0457

10 Newmob is a mobile phone service provider which sells several brands of mobile phones. uPhones and Samseng phones are sold at a subsidy to its subscribers. Each subscriber can either buy one uPhone or one Samseng phone or both one uPhone and one Samseng phone. The probability that a randomly chosen subscriber buys a uPhone is 0.3, and independently, the probability that the subscriber buys a Samseng phone is p.

(i) In a random sample of 50 subscribers, the probability that at most 20 subscribers buy a Samseng phone is twice the probability that exactly 15 subscribers buy a uPhone. Find the value of p. [3]

For the remainder of this question, you may take the value of p to be 0.4.

(ii) In a random sample of 50 subscribers, find the probability that the number of subscribers who buy a uPhone is greater than the expected number of subscribers who buy a Samseng phone. [2]

(iii) Each subsidy for the uPhone costs Newmob \$280 and each subsidy for a Samseng phone costs \$200. Newmob has 1000 subscribers. Using suitable approximations, find the probability that the total subsidy given by Newmob for uPhone purchases exceeds the total subsidy given for Samseng phone purchases. [5]

10i

Let U and S be the number of subscribers who purchased a uPhone and a Samseng phone respectively in a sample of size n.

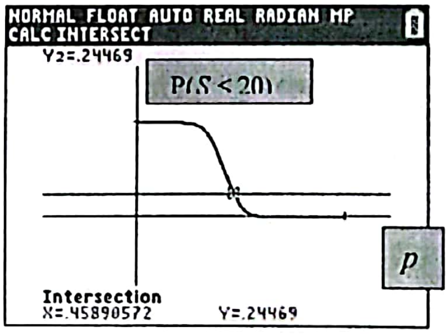
$U \sim \mathrm{B}(n, 0.3) \qquad S \sim \mathrm{B}(n, p)$

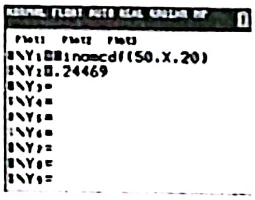
When $n = 50$,

$$\mathrm{P}(S \le 20) = 2\mathrm{P}(U = 15)$$

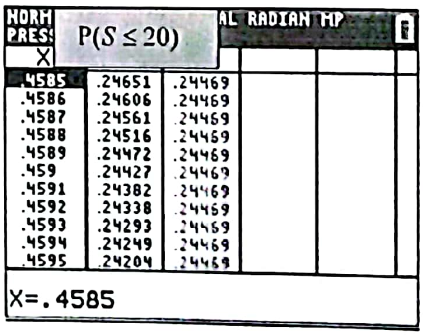
$$\mathrm{P}(S \le 20) = 0.24469$$

From GC,





or



X		
.4585	.24651	.24469
.4586	.24606	.24469
.4587	.24561	.24469
.4588	.24516	.24469
.4589	.24472	.24469
.459	.24427	.24469
.4591	.24382	.24469
.4592	.24338	.24469
.4593	.24293	.24469
.4594	.24249	.24469
.4595	.24204	.24469

X=.4585

$p = 0.459$ (3 sig fig)

10ii	Given $p = 0.4$, $n = 50$, $S \sim B(50, 0.4)$ $E(S) = 50\,(0.4) = 20$ $P(U > 20) = 1 - P(U \le 20)$ $= 0.0478$ (3 sig fig)
10iii	When $n = 1000$ $U \sim B(1000, 0.3)$ Since n is large, $np = 300 > 5$, $n\,(1-p) = 700 > 5$ $U \sim N(300, 210)$ approximately $S \sim B(1000, 0.4)$ Since n is large, $np = 400 > 5$, $n\,(1-p) = 600 > 5$ $S \sim N(400, 240)$ approximately $280U - 200S \sim N\,(\,280(300) - 200(400), 280^2(210) + 200^2(240)\,)$ $280U - 200S \sim N\,(\,4000, 26064000)$ $P\,(\,280U > 200S\,)$ $= P(\,280U - 200S > 0\,)$ $= P(\,280U - 200S > 0.5\,)$ (by continuity correction) $= 0.783$ (3 sig fig)

11 The volume of a packet of soya bean milk is denoted by V ml and the population mean of V is denoted by μ ml. A random sample of 80 packets of soya bean milk is taken and the results are summarised by

$$\sum(v-250)=-217,\ \sum(v-250)^2=30738.$$

Test, at the 4% significance level, whether μ is less than 250 ml. [5]

Explain, in the context of the question, the meaning of 'at the 4% significance level'. [1]

In another test, using the same data, and also at the 4% significance level, the hypotheses are as follows.

Null hypothesis: $\mu=\mu_0$

Alternative hypothesis: $\mu \neq \mu_0$

Given that the null hypothesis is rejected in favour of the alternative hypothesis, find the set of possible values of μ_0. [4]

It is now given that $\mu = 250$ and the population variance is 385. A random sample of 50 packets of soya bean milk is taken and the total volume of the packets is denoted by T ml. By considering the approximate distribution of T and assuming that the volumes of all packets of soya bean milk are independent of one another, find P(T > 12600). [3]

11i	H_0: $\mu=250$ H_1: $\mu<250$ Level of significance: 4% $\bar{v}=\dfrac{-217}{80}+250=247.2875$ $s^2=\dfrac{1}{79}\left(30738-\dfrac{(-217)^2}{80}\right)=381.6378165$ Since n = 80 is large, by Central Limit Theorem, $\bar{V}$ follows a normal distribution approximately. Test statistic: $\dfrac{\bar{V}-\mu}{\left(\dfrac{s}{\sqrt{n}}\right)} \sim \text{N}(0,1)$ approximately If H_0 is true, p-value = 0.107 > 0.04, we do not reject H_0. (Or $z_{calc}=-1.24<-1.751$) There is insufficient evidence at 4% level significance to conclude that $\mu<250$. 'at the 4% significance level' means there is a 4% chance that we wrongly concluded that $\mu<250$. (Or wrongly rejected $\mu=250$ when it is true)

	$H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$ If H_0 is true, $z_{cal} = \dfrac{\bar{v} - \mu_0}{\left(\dfrac{s}{\sqrt{n}}\right)}$ In order to reject H_0, z_{cal} must lie in the critical region. $\dfrac{247.2875 - \mu_0}{\sqrt{\dfrac{381.6378165}{80}}} < -2.053749$ or $\dfrac{247.2875 - \mu_0}{\sqrt{\dfrac{381.6378165}{80}}} > 2.053749$ $\mu_0 > 251.77$ or $\mu_0 < 242.80$ Set of possible values of $\mu_0 = \{\mu_0 : \mu_0 < 243 \text{ or } \mu_0 > 252\}$ (Also accept 242 other than 243) $T = V_1 + V_2 + V_3 + ... + V_{50}$ Given E(V) =250 and Var(V) = 385 Since $n = 50$ is large, by Central Limit Theorem, $T \sim$ N(50×250, 50× 385) = N(12500, 19250) approximately. P(T> 12600) = 0.236 (3 sig fig)