Section A: Pure Mathematics [40 marks]

1 A curve C is defined by the parametric equations

$$x = 2\sqrt{3}\cos\theta + \sqrt{3} \text{ and } y = \left(2\sqrt{3}+1\right)\sin\theta, \text{ where } 0 \le \theta \le 2\pi.$$

(i) Find $\frac{\mathrm{d}y}{\mathrm{d}x}$ in terms of θ. [2]

(ii) Show that the Cartesian equation of the curve C is $\left(\frac{x-\sqrt{3}}{2\sqrt{3}}\right)^2 + \left(\frac{y}{2\sqrt{3}+1}\right)^2 = 1$. Hence or otherwise, sketch the curve of C, indicating clearly the x-intercepts in exact form. [3]

(iii) The point $P\left(2\sqrt{3}, 3+\frac{\sqrt{3}}{2}\right)$ lies on curve C. The region R is bounded by the curve C for $x \ge \sqrt{3}$, the x-axis and the line segment joining the points P and $\left(\sqrt{3}, 0\right)$.

Show that the area of R is $\frac{3}{4}\left(2\sqrt{3}+1\right) + 2\left(6+\sqrt{3}\right)\int_0^{\frac{\pi}{3}} \sin^2\theta \,\mathrm{d}\theta$. [4]

[Solution]

(i) $\frac{\mathrm{d}x}{\mathrm{d}\theta} = -2\sqrt{3}\sin\theta$ and $\frac{\mathrm{d}y}{\mathrm{d}\theta} = \left(2\sqrt{3}+1\right)\cos\theta$

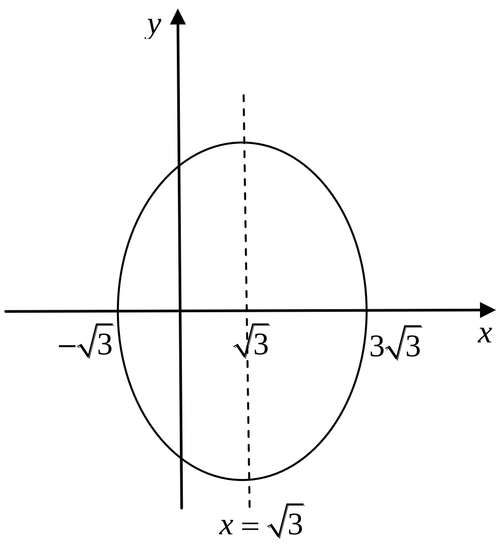
$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{\left(2\sqrt{3}+1\right)\cos\theta}{-2\sqrt{3}\sin\theta} = -\left(1+\frac{1}{2\sqrt{3}}\right)\cot\theta$$

(ii) $x = 2\sqrt{3}\cos\theta + \sqrt{3} \Rightarrow \cos\theta = \frac{x-\sqrt{3}}{2\sqrt{3}}$

$y = \left(2\sqrt{3}+1\right)\sin\theta \Rightarrow \sin\theta = \frac{y}{2\sqrt{3}+1}$

So $\sin^2\theta + \cos^2\theta = 1$

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



(iii) At P, $x = 2\sqrt{3}$

$\therefore 2\sqrt{3}\cos\theta + \sqrt{3} = 2\sqrt{3}$

$\Rightarrow \cos\theta = \frac{1}{2} \therefore \theta = \frac{\pi}{3}$

When $x = 3\sqrt{3}$,

$2\sqrt{3}\cos\theta + \sqrt{3} = 3\sqrt{3}$

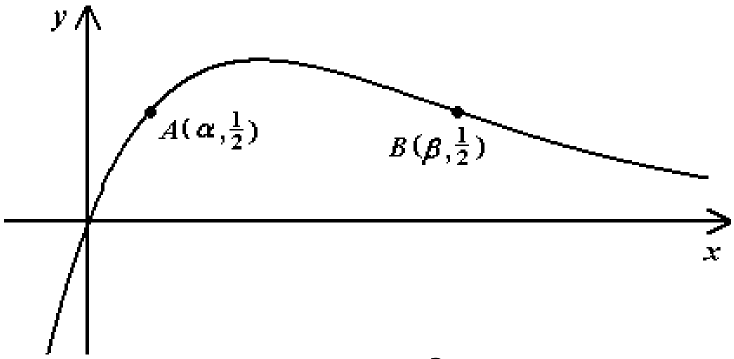
$\Rightarrow \cos\theta = 1, \therefore \theta = 0$

Area of $R =$ area of $\Delta + \int_{2\sqrt{3}}^{3\sqrt{3}} y \, \mathrm{d}x$

$$= \frac{1}{2}\left(\sqrt{3}\right)\left(\frac{\sqrt{3}}{2}\left(2\sqrt{3}+1\right)\right) + \int_{\frac{\pi}{3}}^{0}\left(2\sqrt{3}+1\right)\sin\theta\left(-2\sqrt{3}\sin\theta\right)\,\mathrm{d}\theta$$

$\therefore$ Area of $R = \frac{3}{4}\left(2\sqrt{3}+1\right) + 2\left(6+\sqrt{3}\right)\int_{0}^{\frac{\pi}{3}}\sin^2\theta\,\mathrm{d}\theta$

2



The diagram above shows the curve of $y = \dfrac{2x}{e^x}$. Two points A and B on the curve have coordinates (α, $\frac{1}{2}$) and (β, $\frac{1}{2}$) respectively.

A sequence of real numbers $x_1, x_2, x_3, \ldots$ satisfies the recurrence relation

$$x_{n+1} = \frac{1}{4}e^{x_n} \quad \text{for } n \geq 1.$$

(i) Show algebraically that if the sequence converges, then it converges to either α or β. [3]

(ii) Show that $x_{n+1} < \alpha$ if $x_n < \alpha$. [2]

(iii) Show that $x_{n+1} > x_n$ if $x_n < \alpha$. [2]

(iv) Explain briefly how the results in **(ii)** and **(iii)** may be used to deduce that the sequence converges to α when $x_1 = 0$. [2]

[Solution]

(i) If the sequence converges to, say l,
then $x_n \to l$ and $x_{n+1} \to l$ as $n \to \infty$.

i.e. $x_{n+1} = \frac{1}{4}e^{x_n} \Rightarrow \quad l = \frac{1}{4}e^{l}$ --- (A)

$\Rightarrow \dfrac{2l}{e^l} = \dfrac{1}{2}$

$\Rightarrow l = \alpha$ or β from the diagram.

(ii) If $x_n < \alpha$,

$e^{x_n} < e^{\alpha}$ since e^x is an increasing function.

$\Rightarrow \frac{1}{4}e^{x_n} < \frac{1}{4}e^{\alpha}$

$\Rightarrow x_{n+1} < \frac{1}{4}e^{\alpha}$

$\Rightarrow x_{n+1} < \alpha$ [using eqn (A) in (i)]

(iii) Method 1A: Using given graph

Step 1:

Consider $x_{n+1} - x_n = \dfrac{1}{4}\mathrm{e}^{x_n} - x_n$

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

Step 2:

If $x_n < \alpha$, $\quad \dfrac{2x_n}{\mathrm{e}^{x_n}} < \dfrac{1}{2}$ $\quad$ [as seen in given graph $y = \dfrac{2x}{\mathrm{e}^x}$]

$\Rightarrow \quad \dfrac{1}{2}\mathrm{e}^{x_n}\left(\dfrac{1}{2} - \dfrac{2x_n}{\mathrm{e}^{x_n}}\right) > 0$ $\quad$ [$\because$ $\mathrm{e}^{x_n} > 0$ for all $x_n \in \mathbb{R}$]

$\Rightarrow \quad x_{n+1} - x_n > 0$

$\Rightarrow \quad x_{n+1} > x_n$ (shown)

Method 1B: Using given graph

If $x_n < \alpha$, $\quad \dfrac{2x_n}{\mathrm{e}^{x_n}} < \dfrac{1}{2}$ $\quad$ [as seen in given graph $y = \dfrac{2x}{\mathrm{e}^x}$]

$\Rightarrow \quad x_n < \dfrac{1}{4}\mathrm{e}^{x_n}$

$\Rightarrow \quad x_n < x_{n+1}$

$\Rightarrow \quad x_{n+1} > x_n$ (shown)

Method 2: Sketching you own graph

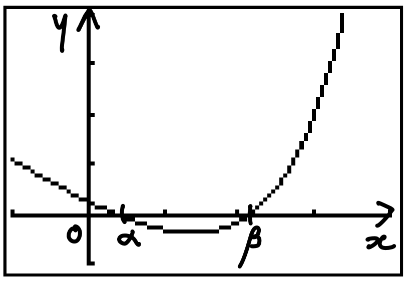
Step 1:

Consider $x_{n+1} - x_n = \dfrac{1}{4}\mathrm{e}^{x_n} - x_n$,

Step 2:

Sketch the graph of $y = \dfrac{1}{4}\mathrm{e}^x - x$.

(We usually use a GC to sketch this graph)



[**Note:** The x-intercepts satisfy $\dfrac{1}{4}\mathrm{e}^x = x \Rightarrow \dfrac{2x}{\mathrm{e}^x} = \dfrac{1}{2} \Rightarrow x = \alpha$ or β]

Hence,

if $x_n < \alpha$, $\quad \frac{1}{4}e^{x_n} - x_n > 0$ $\quad$ [as seen in the graph $y = \frac{1}{4}e^{x} - x$]

$\Rightarrow \quad x_{n+1} - x_n > 0$

$\Rightarrow \quad x_{n+1} > x_n$ (shown)

(iv) When $x_1 = 0 < \alpha$, $\quad x_1 < x_2 < \alpha$ using results in (ii) and (iii)

When $x_2 < \alpha$, $\quad x_2 < x_3 < \alpha$ using results in (ii) and (iii)

Hence $x_1 < x_2 < x_3 < ... < x_n < x_{n+1} < ... < \alpha$

So the sequence is **strictly increasing** and **converges to α.**

3 The function f is defined by

$$\mathrm{f}: x \mapsto x + \frac{1}{x}, \quad x \in \mathbb{R}, x \neq 0$$

(i) Sketch the graph of $y = \mathrm{f}(x)$, showing clearly the coordinates of the stationary points and the equations of asymptotes, if any. [2]

(ii) Given that $\mathrm{g}(x) = \mathrm{f}(x-b) + 2$ where $b > 2$, state a sequence of transformations which transforms $y = \mathrm{f}(x)$ to $y = \mathrm{g}(x)$. [2]

Sketch the graph of $y = \mathrm{g}(x)$, showing clearly the coordinates of the stationary points and the equations of asymptotes in terms of *b*, if any. [2]

On a separate diagram, sketch $y = \mathrm{g}'(x)$ and solve in terms of *b*, the inequality

$$\mathrm{g}'(x) < (x - b + 1)(b + 1 - x).$$ [5]

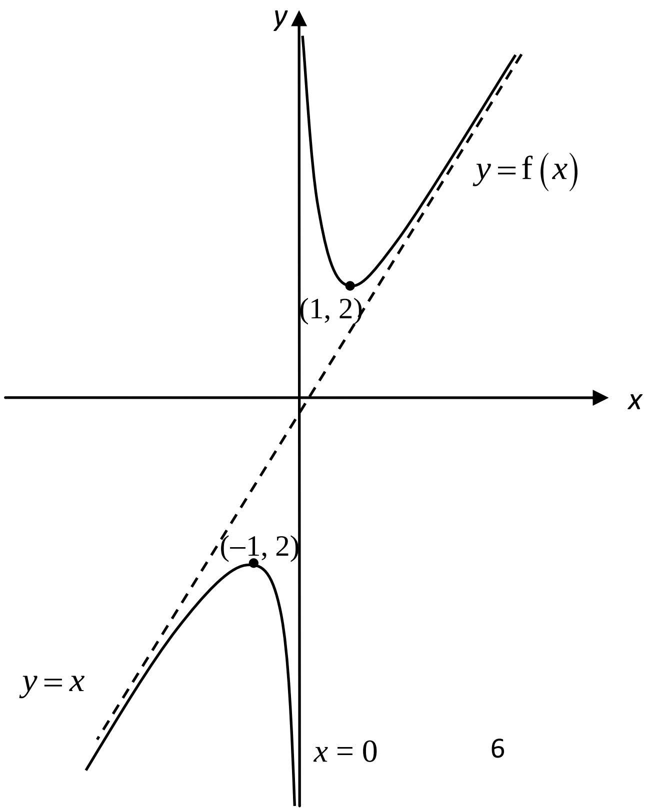
[Solution]

(i) $y = \mathrm{f}(x) = x + \dfrac{1}{x}$

At stationary points, $\mathrm{f}'(x) = 1 - \dfrac{1}{x^2} = \dfrac{x^2 - 1}{x^2} = 0 \quad \therefore x = \pm 1$

For $x = 1, y = 2$ and $x = -1, y = -2$

[Students are allowed to sketch graph using GC, however they have to indicate clearly the coordinates of stationary points and the equations of asymptotes. The details are necessary for next part of the question involving transformations]



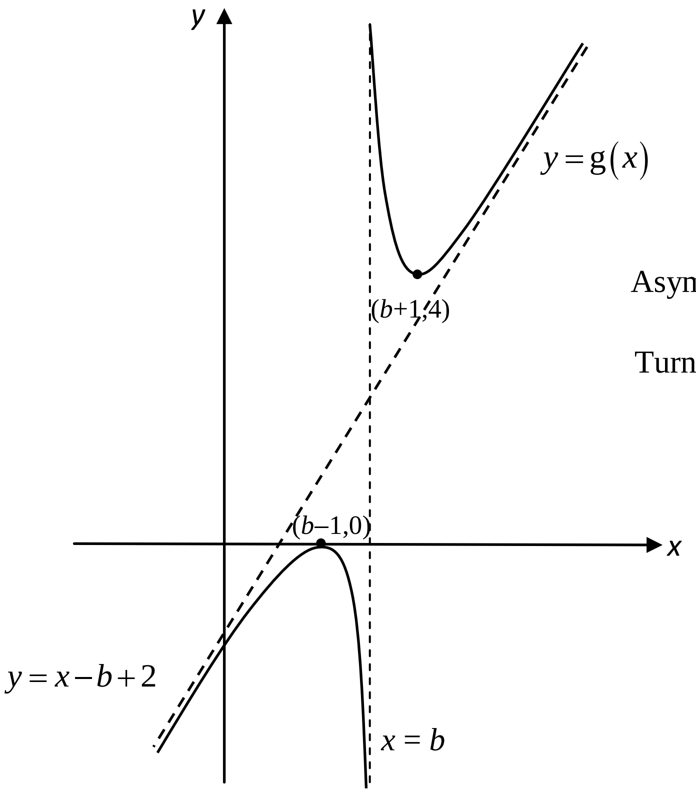
Coordinates of turning points

Shape and equations of asymptotes

(ii) The sequence of transformations is

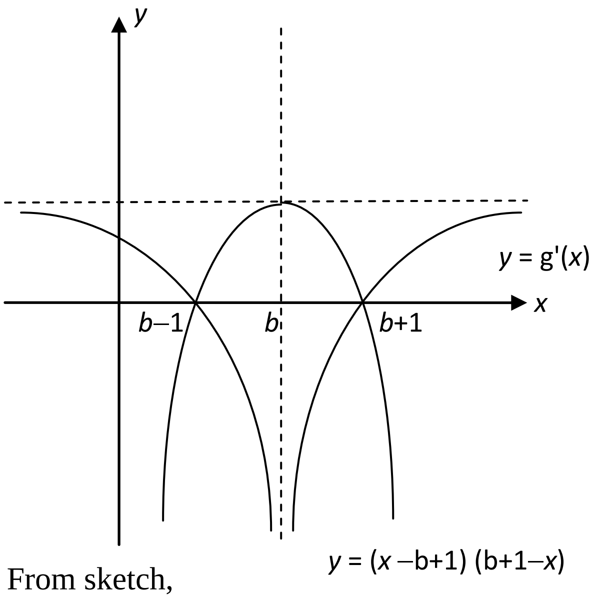
(1) Translate graph +b units along x-axis followed by

(2) Translate graph by +2 units along y-axis



Asymptotes: $y = x - b + 2$; $x = b$

Turning points: $(b-1, 0)$ and $(b+1, 4)$



From sketch,

solution is $b-1 < x < b+1, x \neq b$

4 **(i)** Find the roots of the equation $z^3+8\mathrm{i}=0$, giving them in cartesian form *a* + i*b*, where *a* and *b* are exact real numbers. [3]

(ii) The roots of the equation $\left(z-\sqrt{3}-2\mathrm{i}\right)^3+8\mathrm{i}=0$ are z_1, z_2 and z_3 such that $\mathrm{Re}(z_1)<\mathrm{Re}(z_2)<\mathrm{Re}(z_3)$. Hence find z_1, z_2 and z_3 in cartesian form *a* + i*b*, where *a* and *b* are exact real numbers. [2]

(iii) Show z_1, z_2 and z_3 on an Argand diagram. [1]

(iv) Explain why the locus of all points *z* such that $|z-z_2|=|z-z_3|$ passes through the point representing z_1. Sketch this locus on your Argand diagram and find the minimum value of $|z|$. [5]

[Solution]

(i) $z^3=-8i$

$$=8e^{-i\frac{\pi}{2}}=8e^{i\left(-\frac{\pi}{2}+2k\pi\right)} \quad k=0,\pm 1$$

$$z=2\mathrm{e}^{i\left(-\frac{\pi}{6}+\frac{2k\pi}{3}\right)} \text{ or } z=2\mathrm{e}^{i\left(\frac{4k-1}{6}\right)\pi}$$

$$\therefore z=2\mathrm{e}^{i\left(-\frac{\pi}{6}\right)},2\mathrm{e}^{i\left(\frac{3\pi}{6}\right)},2\mathrm{e}^{i\left(-\frac{5\pi}{6}\right)}$$

$$=\sqrt{3}-i,\ 2i,-\sqrt{3}-i$$

(ii) $\left(z-\sqrt{3}-2i\right)^3+8i=0$

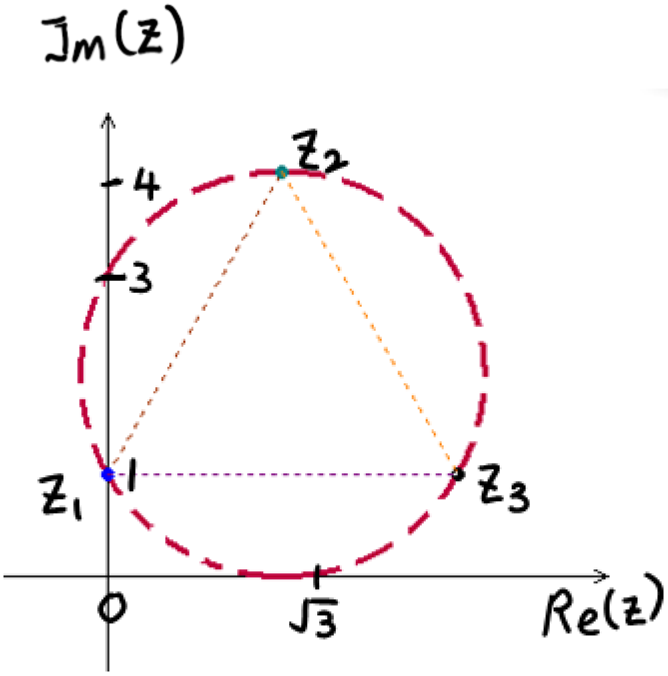
$\left(z-\sqrt{3}-2i\right)^3=-8i$

$\Rightarrow z-\sqrt{3}-2i=\sqrt{3}-i,\ 2i,-\sqrt{3}-i$

$\Rightarrow z=2\sqrt{3}+i,\sqrt{3}+4i,\ i$

So $z_1=i,\ z_2=\sqrt{3}+4i,\ z_3=2\sqrt{3}+i$

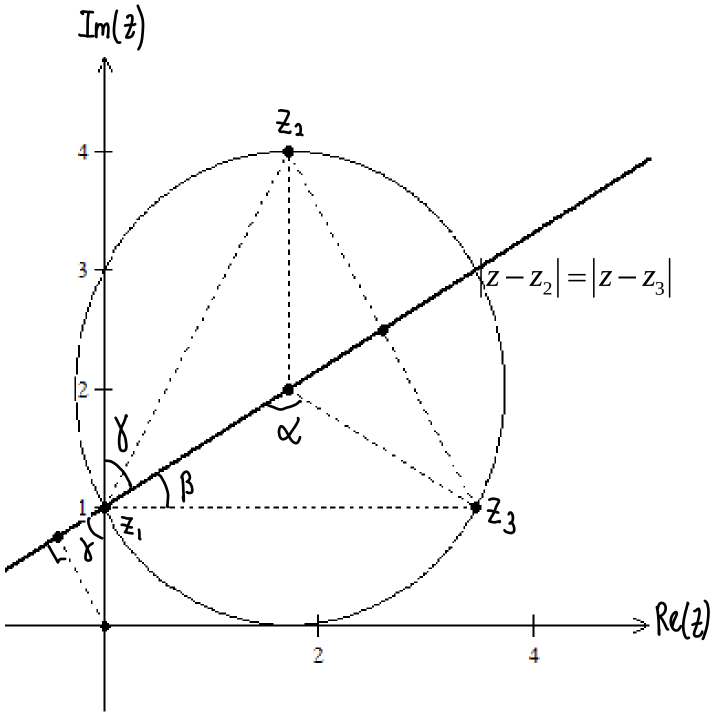
(iii)



(iv) Now $|z_1 - z_2| = |-\sqrt{3} - 3i| = 2\sqrt{3}$

And $|z_1 - z_3| = |-2\sqrt{3}| = 2\sqrt{3}$

Since $|z_1 - z_2| = |z_1 - z_3|$, z_1 lies on the locus.



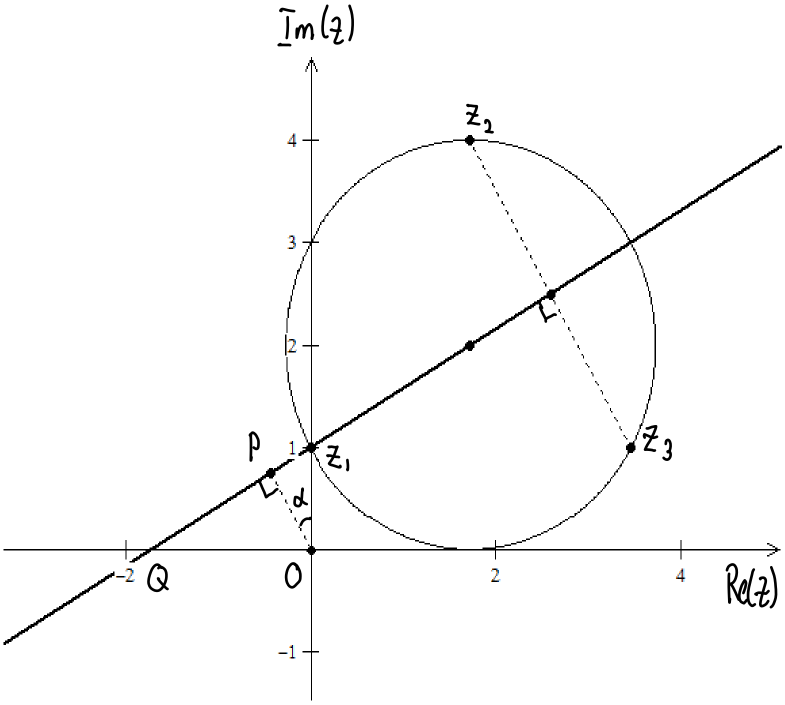
<u>Method 1</u>

$\angle\alpha = \dfrac{2\pi}{3}$

$\therefore \angle\beta = \dfrac{\pi - \angle\alpha}{2} = \dfrac{\pi}{6}$

Hence $\angle\gamma = \dfrac{\pi}{2} - \dfrac{\pi}{6} = \dfrac{\pi}{3}$

$|OP| = |z_1| \sin\angle\gamma = \dfrac{\sqrt{3}}{2}$



Method 2

min $|z| = |OP| = |z_1|\cos\angle\alpha$

where $\angle\alpha = \dfrac{\pi}{2} - \angle POQ$

$\angle POQ = \angle z_2 z_3 z_1$

then $\tan\angle POQ = \tan\angle z_2 z_3 z_1 = \dfrac{3}{\sqrt{3}}$

$\Rightarrow \angle POQ = \dfrac{\pi}{3}\ \therefore \angle\alpha = \dfrac{\pi}{2} - \dfrac{\pi}{3} = \dfrac{\pi}{6}$

$|z - z_2| = |z - z_3|$

Therefore, min $|z| = |OP| = |z_1|\cos\angle\alpha = \cos\dfrac{\pi}{6} = \dfrac{\sqrt{3}}{2}$

Section B: Statistics [60 marks]

5 2 men and 5 women go to a restaurant. They choose an outdoor round table with 7 seats for their meal. Find the number of ways the group can be seated if

(i) the two men are not seated next to each other. [2]

(ii) one of the women, Mary, is to be seated between the two men. [2]

Before their orders arrive, they request to shift to a table in the 'non-smoking' section of the restaurant. They are then given a round table with 10 seats.

Find the number of ways they can be seated if

(iii) the empty seats are adjacent to each other. [2]

(iv) none of the empty seats are adjacent to each other and there must be more than 1 person between any two empty seats. [2]

[Solution]

(i) Number of ways $= 6! - (5!)2 = 480$ or $(5-1)! \times {}^5P_2 = 480$

(ii) Group Mary and the 2 men as one unit, number of ways $= (5-1)! \times 2 = 48$

(iii) Group 3 empty seats as one unit, number of ways $= (8-1)! = 5040$

(iv) Number of ways $= 7!$ or ${}^7P_3 \times {}^4P_2 \times {}^2P_2 = 7! = 5040$

6 Eighteen numbers are arranged in three groups of six as follows:

Group A: 0, 2, 2, 2, 2, 9

Group B: 3, 3, 3, 7, 7, 16

Group C: 1, 1, 1, 1, 6, 6

One number is drawn at random from each group. Let a, b and c denote the number drawn from groups A, B and C respectively.

Event X is defined as "b is greater than the sum of a and c".

Event Y is defined as "b is greater than both a and c".

(i) Show that $\mathrm{P}(X)=\dfrac{23}{54}$. [3]

(ii) Find $\mathrm{P}(X\,|\,Y)$. [4]

A game is played with a biased coin where the probability of getting a Head is p. A player first flips a coin. If the coin shows a Head, the player draws a number from Group B and the score is the number drawn. If the coin shows a Tail, the player draws a number from Group A and C each and the score is the sum of the numbers drawn. If the probability of obtaining a score of 3 is $\dfrac{13}{27}$, find the value of p. [2]

[Solution]

Event X ="b greater than the sum of a and c" $=\left\{\begin{array}{l}(b=3, a=0, c=1)\\ \text{or}\,(b=7, a=0, c=1 \text{ or } 6)\\ \text{or}\,(b=7, a=2, c=1)\\ \text{or}\,(b=16, a=0,\ 2 \text{ or } 9, c=1 \text{ or } 6)\end{array}\right\}$

(i) $\mathrm{P}(X)=\frac{3}{6}\times\frac{1}{6}\times\frac{4}{6}+\frac{2}{6}\times\frac{1}{6}\times 1+\frac{2}{6}\times\frac{4}{6}\times\frac{4}{6}+\frac{1}{6}$ [M2 - 2 cases with correct probability]

[M1- Sum of mutually exclusive cases or otherwise]

$=\frac{92}{216}=\frac{23}{54}$

(ii) Now $\mathrm{P}(X \mid Y) = \dfrac{\mathrm{P}(X \cap Y)}{\mathrm{P}(Y)} = \dfrac{\mathrm{P}(X)}{\mathrm{P}(Y)}$ [M1 for realising $X \cap Y = X$]

Since Event Y = "b greater than both a and c" = $\left\{\begin{array}{l}\left(b=3, a=0 \text{ or } 2, c=1\right) \\ \text{or } \left(b=7, a=0 \text{ or } 2, c=1 \text{ or } 6\right) \\ \text{or } \left(b=16, a=0, 2 \text{ or } 9, c=1 \text{ or } 6\right)\end{array}\right\}$

$\Rightarrow \mathrm{P}(Y) = \dfrac{3}{6}\times\dfrac{5}{6}\times\dfrac{4}{6}+\dfrac{2}{6}\times\dfrac{5}{6}\times 1+\dfrac{1}{6}\times 1\times 1$ [M1 for one correct, M2 for all correct]

$= \dfrac{156}{216} = \dfrac{13}{18}$

$\therefore \mathrm{P}(X \mid Y) = \dfrac{\frac{23}{54}}{\frac{13}{18}} = \dfrac{23}{39}$

$\mathrm{P}(\text{score} = 3) = \dfrac{13}{27} \Rightarrow p\left(\dfrac{3}{6}\right) + (1-p)\left(\dfrac{4}{6}\times\dfrac{4}{6}\right) = \dfrac{13}{27}$ (use of Tree Diagram if applicable)

[M1 - Sum of 2 mutually exclusive cases]

$\therefore p = \dfrac{2}{3}$

7 Two friends, Bob and Patrick, meet up each week at a swimming complex for a 200m freestyle friendly match. The time (in seconds, s) taken by Bob to complete a 200m freestyle swim follows a normal distribution with mean 152 s and standard deviation 2.2 s while the corresponding time taken by Patrick is also normally distributed with mean 156 s and standard deviation 3.0 s.

(i) Show that the probability of Patrick beating Bob in a 200m freestyle match is 0.141, correct to 3 decimal places. [1]

(ii) Find the probability that the total time taken by Bob to complete a 200m freestyle swim on two different occasions is less than twice the time taken by Patrick to complete a 200m freestyle swim on one occasion by less than 5 s. [4]

Bob and Patrick maintained their weekly swimming matches for a total of k weeks, where $k \geq 50$. Use a suitable approximation to find the least value of k such that the probability of Patrick beating Bob on fewer than four occasions is not larger than 5%. [5]

[Solution]

(i) Let the time taken by Bob to complete a 200m swim be X. $X \sim N\left(152, 2.2^2\right)$

Let the time taken by Patrick to complete a 200m swim be Y. $Y \sim N\left(156, 3.0^2\right)$

$Y - X \sim N\left(4, 13.84\right)$

$\text{P}(Y - X < 0) = 0.141$

(ii) $X_1 + X_2 \sim N\left(304, 9.68\right)$ and $2Y \sim N\left(312, 36\right)$

$2Y - \left(X_1 + X_2\right) \sim N\left(8, 45.68\right)$

$\text{P}(0 \leq 2Y - \left(X_1 + X_2\right) < 5) = 0.21029 \approx 0.210$

Let T be the number of weekly swims (out of k) where Patrick beats Bob in the 300m freestyle match. $T \sim B\left(k, 0.14114\right)$

Since n large, $np > 5$, $n(1-p) > 5$, $T \sim N\left(0.14114k, 0.12122k\right)$

Given $\text{P}(T < 4) \leq 0.05 \Rightarrow \text{P}(T < 3.5) \leq 0.05$ (c.c)

$\Rightarrow \text{P}(z < \dfrac{3.5 - 0.14114k}{\sqrt{0.12122k}}) \leq 0.05 \Rightarrow \dfrac{3.5 - 0.14114k}{\sqrt{0.12122k}} \leq -1.64485$

From GC, $k \geq 55$. Least k is 55.

8 An online web survey company wishes to find out the number of hours spent per week, on average, by a typical teenager on Facebook. A survey was conducted on a random sample of 70 teenagers and the time spent per week, x hours, was recorded and summarized as:

$$\sum(x-18)=208, \quad \sum(x-18)^2=8967$$

(i) Find, correct to 1 decimal place, the unbiased estimates of the population mean and variance. [2]

(ii) It is claimed that a typical teenager spends an average of 18 hours a week on Facebook. Test, at the 5% level of significance, whether the population mean weekly time differs from 18 hours. [4]

(iii) Another independent online survey company claims that the population mean weekly time spent on Facebook by teenagers actually exceeds μ_0, where μ_0 is a constant. Use the earlier sample to determine, correct to 1 decimal place, the least value of μ_0 in order for the new claim not to be valid at 5% level of significance. [4]

[Solution]

(i) Let $Y = X - 18$, then $\sum y = 208, \quad \sum y^2 = 8967$.

Then $\bar{x} = 18 + \dfrac{208}{70} = 20.97 \approx 21.0$ (to 1 d.p) .

And $s^2 = \dfrac{1}{69}\left(8967 - \dfrac{208^2}{70}\right) = 120.999 \approx 121.0$ (to 1 d.p)

(ii) $H_0: \mu = 18$

$H_1: \mu \neq 18$

Level of significance = 5%

Since n is large (> 50), σ^2 and distribution of x are unknown,

Test statistics : $Z = \dfrac{\bar{X} - \mu}{s/\sqrt{n}} \sim N(0,1)$ approximately by Central Limit Theorem,

Under H_0, $z_{cal} = 2.26$

p – value = 0.0239 < 5%

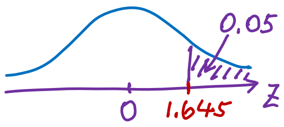
We reject H_0 and conclude that there is sufficient evidence at 5% level of significance to claim that the average weekly time differs from 18 hours.

(iii) $H_0 : \mu = \mu_0$

$H_1 : \mu > \mu_0$

Test statistics : $Z = \dfrac{\bar{X} - \mu}{s / \sqrt{n}} \sim N(0,1)$ approximately by CLT.

Level of significance = 5%, Critical value = 1.645



Under H_0 , $z_{cal} = \dfrac{\bar{x} - \mu_0}{s / \sqrt{n}}$.

For H_0 not to be rejected at 5% level of significance, $z_{cal} < 1.645$

$$\Rightarrow \frac{20.971 - \mu_0}{10.982 / \sqrt{70}} < 1.645$$

$$\Rightarrow \mu_0 > 18.812$$

So the **least value of μ_0 is 18.9**.

9 Records show that one in every 6 students who go to the Career Fair for junior college students will stop by the Police Force's booth to make enquiries.

A class of 25 students visit the Career Fair. Show that the probability that fewer than 10 students make enquiries at the Police Force's booth is 0.99526. [2]

60 classes, each consisting of 25 students, visit the Career Fair.

(i) Find the probability that the average number of students per class who make enquiries at the Police Force's booth is neither fewer than 4 nor greater than 6. [3]

(ii) Using a suitable approximation, find the probability that 58 classes have fewer than 10 students making enquiries at the Police Force's booth. [4]

To gather students' feedback on the Fair, a member of the Career Fair Committee decides to interview 20 male and 30 female students from the student population.

Name the sampling method used and give one disadvantage of this sampling method. [2]

[Solution]

Let X be the number of students who make enquiries at the Police Force's booth (out of 25).

$$X \sim B(25, \frac{1}{6})$$

$$P(X < 10) = P(X \leq 9) = 0.99526$$

(i) **Method 1**

Since $n = 60$ is large, by Central Limit Theorem,

$$\overline{X} \sim \mathrm{N}\left(25\left(\frac{1}{6}\right), \frac{25\left(\frac{1}{6}\right)\left(\frac{5}{6}\right)}{60}\right) \text{ approximately.}$$

i.e. $\overline{X} \sim \mathrm{N}\left(\frac{25}{6}, \frac{125}{2160}\right)$

$$P\left(4 \leq \overline{X} \leq 6\right) = 0.756 \text{ (3s.f.)}$$

Method 2

Let $T = X_1 + X_2 + \ldots + X_{60} \sim B(1500, \frac{1}{6})$

$$P(4 \leq \overline{X} \leq 6) = P(240 \leq T \leq 360) = P(T \leq 360) - P(T \leq 239)$$

$$= 0.76541 \approx 0.765$$

(ii) Let Y be the number of classes (out of 60) having 10 or more students making enquiries at the Police Force's booth.

$Y \sim B(60, 0.0047426)$

Since $n = 60$ large, $np = 0.28455 < 5$, $Y \sim P_0(0.28455)$ approximately

Required Prob $= P(Y = 2) = 0.30459 \approx 0.305$

Quota sampling.

Disadvantage:

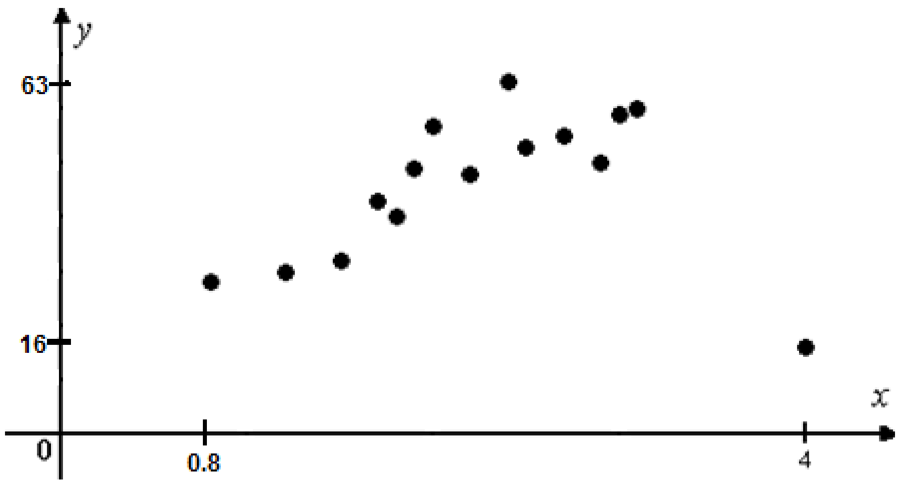
[1] Sample may be biased as the interviewers are allowed to select students who are more approachable to fulfill the quota required.

[2] Sample may not be representative of the student cohort as the male to female ratio may not be 2:3 as stated in the sample.

[3] Quota Sampling method is not random and as a result the sample may be biased as interviewers are allowed to select the students in any manner to fulfill the quota.

10 A study was carried out to investigate possible links between the weights of hens (x kg) and their eggs (y g). A sample of 15 hens was chosen at random and the weights of these hens and their eggs were noted. The scatter diagram and the summarized information for the sample are shown below. The linear product moment coefficient was also computed and found to be 0.200.

By referring to the scatter diagram and the given value of the linear product moment correlation coefficient, comment on the appropriateness of a linear model. [1]



n	$\sum x$	$\sum x^2$	$\sum y$	$\sum y^2$	$\sum xy$
15	33.9	85.99	690	34432	1591.2

One of the points, (4, 16), was identified as an outlier and removed.

(i) For the remaining sample of size 14, recalculate the values in the table above and determine the value of the linear product moment correlation coefficient. Show your workings clearly. [4]

(ii) Use a suitable regression line to estimate the weight of an egg laid by a hen weighing 2.5 kg. [5]

(iii) Comment on the reliability of your answer. [2]

[Solution]

Based on the scatter diagram a linear model is appropriate, the low r value is probably due to the inclusion of the point at $x = 4$.

(i) After removing the point (4, 16), we have

n	$\sum x$	$\sum x^2$	$\sum y$	$\sum y^2$	$\sum xy$
14	29.9	69.99	674	34176	1527.2

$$r_{x,y} = \frac{1527.2 - \frac{(29.9)(674)}{14}}{\sqrt{\left(69.99 - \frac{(29.9)^2}{14}\right)\left(34176 - \frac{(674)^2}{14}\right)}}$$

$= 0.852$ (3sf)

(ii) $\bar{x} = 2.1357$, $\bar{y} = 48.1429$, $b = \dfrac{1527.2 - \frac{(29.9)(674)}{14}}{69.99 - \frac{(29.9)^2}{14}} = 14.306$,

Regression line of y on x: $y - 48.1429 = 14.306\,(x - 2.1357)$

$$\Rightarrow \quad y = 14.306x + 17.589$$

If $x = 2.5$, $y \approx 14.306(2.5) + 17.589 = 53.4$

So the corresponding egg weight is 53.4g.

(iii) Since $r_{x,y} = 0.8523$, there is a strong positive correlation between x and y. Furthermore, the $x = 2.5$ lies within the data range (from 0.8 to roughly 3). Hence it can be concluded that the estimate is a reliable one.