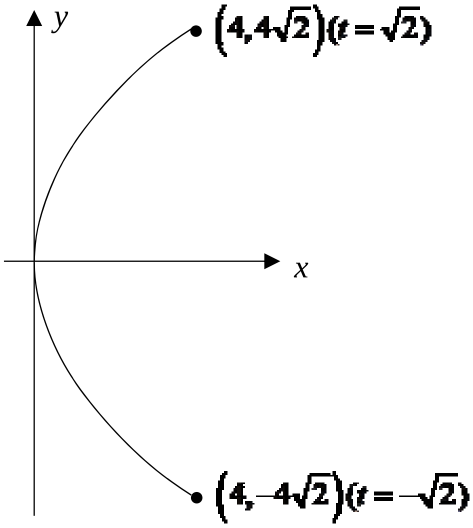
Section A: Pure Mathematics (50 marks)

1 The outer layer of a contact lens is obtained by rotating the curve with parametric equations $x = 2t^2$, $y = 4t$, $-\sqrt{2} \le t \le \sqrt{2}$ through π radians about the x-axis. Find the exact outer surface area of the contact lens. [4]

[Solution]



$$\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 = (4t)^2 + 4^2 = 16(t^2+1)$$

$$\text{Surface area} = 2\pi\int_0^{\sqrt{2}} y\sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}\, dt$$

$$= 2\pi\int_0^{\sqrt{2}} 4t\sqrt{16(t^2+1)}\, dt$$

$$= 16\pi\int_0^{\sqrt{2}} 2t(t^2+1)^{\frac{1}{2}}\, dt$$

$$= 16\pi\left[\frac{(t^2+1)^{\frac{3}{2}}}{\frac{3}{2}}\right]_0^{\sqrt{2}}$$

$$= \frac{32\pi}{3}\left(3\sqrt{3}-1\right)$$

2 Show that the equation $\tan^{-1}\left(e^{x}\right)-\frac{1}{x}=0$ has exactly one positive root α in the interval $(0,1)$. [2]

To find the root α, two methods are proposed.
In method 1, Newton-Raphson method is used to estimate this positive root using 1 as an initial value, giving your answer correct to 4 decimal places. [2]
Explain, by illustrating the first two iterations of Newton-Raphson method on the graph of $y=\tan^{-1}\left(e^{x}\right)-\frac{1}{x}$, why Newton-Raphson method fails when an initial approximation of 2 is being used. [3]

In method 2, the equation $\tan^{-1}\left(e^{x}\right)-\frac{1}{x}=0$ is to be written in the form $x=F(x)$ and the recurrence relation $x_n=F\left(x_{n-1}\right)$ where $n\geq 1$ is being used.
Propose a suitable $F(x)$ that generates a sequence of numbers that converges to α for any initial approximation $x_{\circ}>0$. Justify your answer. [3]

[Solution]

To show that equation $\tan^{-1}\left(e^{x}\right)-\frac{1}{x}=0$ has exactly one positive root α in the interval $(0,1)$:

Method 1:

Let $f(x)=\tan^{-1}\left(e^{x}\right)-\frac{1}{x}$

$$f'(x)=\frac{e^{x}}{e^{2x}+1}+\frac{1}{x^{2}}$$

Since for all real *x*, $\frac{e^{x}}{e^{2x}+1}>0$ and $\frac{1}{x^{2}}>0$, therefore $f'>0$. f is a strictly increasing function for all *x*.

And since f(1) = 0.218 > 0 and

As $x\to 0^{+}, \frac{1}{x}\to\infty, \tan^{-1}\left(e^{x}\right)\to\frac{\pi}{4}\Rightarrow f(x)\to -\infty$

Since f changes sign and is continuous in (0, 1), and f is strictly increasing, therefore f has exactly one positive root in (0,1).

Alternative method:

Check that $\tan^{-1}\left(e^{x}\right)$ is increasing and $\frac{1}{x}$ is decreasing function for all $x > 0$, and so there is exactly one positive root (besides checking change in sign in (0,1)).

$$\mathrm{f}(x_n) = \tan^{-1}(\mathrm{e}^{x_n}) - \frac{1}{x_n}$$

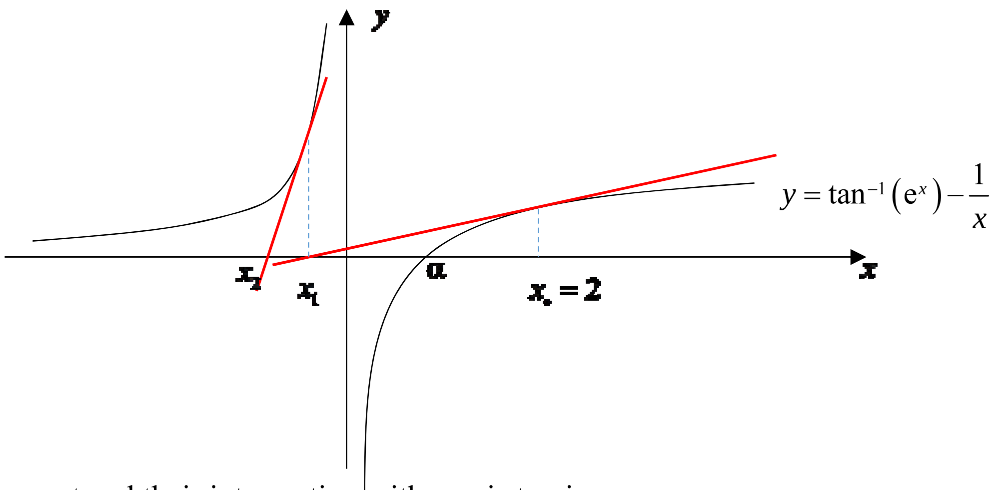
$$\mathrm{f}'(x_n) = \frac{\mathrm{e}^{x_n}}{\mathrm{e}^{2x_n}+1} + \frac{1}{{x_n}^2}$$

By Newton-Raphson method, using $x_\circ = 1$ and $x_{n+1} = x_n - \dfrac{\mathrm{f}(x_n)}{\mathrm{f}'(x_n)}$,

$x_1 \approx 0.835137 = 0.8351(4 \text{ d.p.})$
$x_2 \approx 0.855108 = 0.8551(4 \text{ d.p.})$
$x_3 \approx 0.855524 = 0.8555(4 \text{ d.p.})$
$x_4 \approx 0.855524 = 0.8555(\text{round off 4 d.p. same as } x_3)$

$\therefore \alpha \approx 0.8555 \ (4 \text{ d.p.})$

To explain why N-R method fails when $x_\circ = 2$ is used:



Show each tangent and their intersection with x-axis to give x_1, x_2.

From the graph above, we see that due to the small gradient of the graph when $x_\circ = 2$, the tangent at $x_\circ = 2$ intersects the x-axis far away at a point where $x_1 < 0$ and this results in subsequent $x_n < 0$ and hence Newton-Raphson iterations fail to converge to α when $x_\circ = 2$.

Consider $\mathrm{F}(x) = \dfrac{1}{\tan^{-1}(\mathrm{e}^x)}$.

$$\mathrm{F}'(x) = -\frac{\mathrm{e}^x}{(\mathrm{e}^{2x}+1)\left(\tan^{-1}(\mathrm{e}^x)\right)^2} < 0 \text{ for all } x > 0.$$

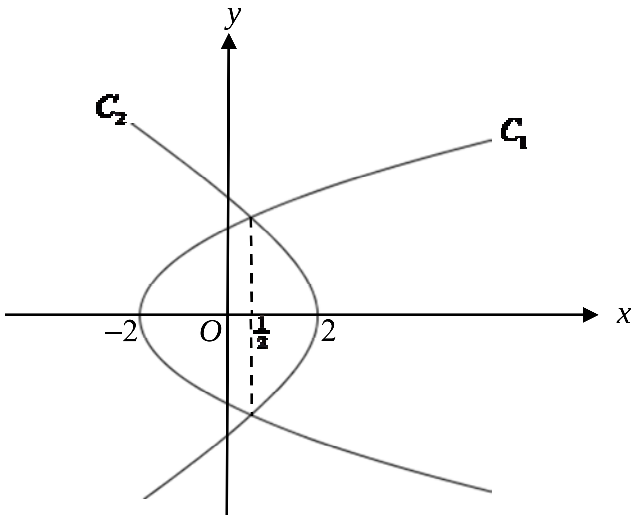
$\mathrm{F}'(0) \approx -0.811$

and $F''(x) = \dfrac{2\left(e^{4x} - e^{2x}\right)}{\left(1+e^{2x}\right)^3 \left(\tan^{-1}\left(e^x\right)\right)^3} > 0$ for all $x > 0$

$\Rightarrow F'(x)$ is an increasing function in x and hence $\left|F'(x)\right| < 1$ for all $x > 0$.

Thus for any $x_\circ > 0$, $x_n = F\left(x_{n-1}\right)$ will always generate a sequence that converges to α.

3



The parabola C_1 and the hyperbola C_2 share the same focus at the origin O. C_1 and C_2 cut the x-axis at (–2, 0) and (2, 0) respectively. Given that C_1 and C_2 intersect at $x = \frac{1}{2}$, find the Cartesian equations of C_1 and C_2. [7]

Hence, show that the distance between the directrix of C_1 and the directrix of C_2 corresponding to the focus at O is $\frac{36}{5}$ units. [2]

[Solution]:

Equation of parabola C_1 is of the form $y^2 = 4k(x+2)$.

Since focus is at O and vertex is at (–2, 0), therefore $k = 2$.

Therefore, C_1 has equation $y^2 = 8(x+2)$.

When $x = \frac{1}{2}$, $y^2 = 8\left(\frac{1}{2}+2\right) \Rightarrow \quad y = \pm\sqrt{20}$

Therefore intersection points of C_1 and C_2 are $\left(\frac{1}{2}, \sqrt{20}\right)$ and $\left(\frac{1}{2}, -\sqrt{20}\right)$.

Equation of hyperbola C_2 is of the form $\frac{(x-c)^2}{a^2} - \frac{y^2}{c^2-a^2} = 1$

Since C_2 passes through (2, 0) and $\left(\frac{1}{2}, \sqrt{20}\right)$, we have

$a^2 = (2-c)^2$ ---(1)

$\frac{\left(\frac{1}{2}-c\right)^2}{a^2} - \frac{20}{c^2-a^2} = 1$ ---(2)

Sub (1) into (2), we have $\frac{\left(\frac{1}{2}-c\right)^2}{(2-c)^2} - \frac{20}{c^2-(2-c)^2} = 1$

Since distance between vertex and focus is 2 units, therefore $c > 2$.

Using GC, we have $c = 5$. $\therefore a = 3$.

Therefore equation of hyperbola C_2 is $\dfrac{(x-5)^2}{9}-\dfrac{y^2}{16}=1$

Directrix of C_1 is $x=-4$

Directrix of C_2 is $x=-\dfrac{a^2}{c}+c$, i.e. $x=\dfrac{16}{5}$

Therefore, distance between the 2 directrices is $4+\dfrac{16}{5}=\dfrac{36}{5}$

4 The matrix **A** is given by $\mathbf{A}=\begin{pmatrix}3&2&1\\0&1&2\\0&0&4\end{pmatrix}$.

Write down the eigenvalues of **A** and find the corresponding eigenvectors of **A**. Hence find a non-singular matrix **P** and a diagonal matrix **D** such that $\mathbf{A}=\mathbf{PDP}^{-1}$. [7]

The matrix **B** is such that $\mathbf{B}=\mathbf{QAQ}^{-1}$, where $\mathbf{Q}=\begin{pmatrix}-1&0&1\\1&-1&2\\0&1&-2\end{pmatrix}$.

By using the expression $\mathbf{PDP}^{-1}$ for **A**, find the eigenvalues and a set of corresponding eigenvectors of **B**. [4]

[Solution]:
The eigenvalues of A are 1, 3 and 4.
For $\lambda=1$,

$$\mathbf{A}-\mathbf{I}=\begin{pmatrix}2&2&1\\0&0&2\\0&0&3\end{pmatrix}$$

A corresponding eigenvector is $\begin{pmatrix}2\\2\\1\end{pmatrix}\times\begin{pmatrix}0\\0\\1\end{pmatrix}=2\begin{pmatrix}1\\-1\\0\end{pmatrix}$

For $\lambda=3$,

$$\mathbf{A}-3\mathbf{I}=\begin{pmatrix}0&2&1\\0&-2&2\\0&0&1\end{pmatrix}$$

A corresponding eigenvector is $\begin{pmatrix}0\\2\\1\end{pmatrix}\times\begin{pmatrix}0\\0\\1\end{pmatrix}=2\begin{pmatrix}1\\0\\0\end{pmatrix}$

For $\lambda=4$,

$$\mathbf{A}-4\mathbf{I}=\begin{pmatrix}-1 & 2 & 1\\ 0 & -3 & 2\\ 0 & 0 & 0\end{pmatrix}$$

A corresponding eigenvector is $\begin{pmatrix}-1\\2\\1\end{pmatrix}\times\begin{pmatrix}0\\-3\\2\end{pmatrix}=\begin{pmatrix}7\\2\\3\end{pmatrix}$

i.e. the corresponding eigenvectors of $\lambda=1,3,4$ are $\begin{pmatrix}-1\\-1\\0\end{pmatrix},\begin{pmatrix}1\\0\\0\end{pmatrix},\begin{pmatrix}7\\2\\3\end{pmatrix}$ respectively.

Therefore, $\mathbf{A}=\mathbf{PDP}^{-1}$ where $\mathbf{P}=\begin{pmatrix}1 & 1 & 7\\ -1 & 0 & 2\\ 0 & 0 & 3\end{pmatrix}$, $\mathbf{D}=\begin{pmatrix}1 & 0 & 0\\ 0 & 3 & 0\\ 0 & 0 & 4\end{pmatrix}$

$\mathbf{B}=\mathbf{QAQ}^{-1}$

$=\mathbf{Q}\left(\mathbf{PDP}^{-1}\right)\mathbf{Q}^{-1}$

$=\left(\mathbf{QP}\right)\mathbf{D}\left(\mathbf{QP}\right)^{-1}$

$\Rightarrow$ **B** has the same set of eigenvalues as A and the corresponding eigenvectors are obtained by

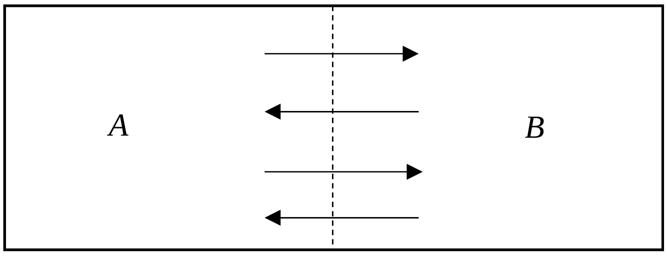
$$\mathbf{QP}=\begin{pmatrix}-1 & 0 & 1\\ 1 & -1 & 2\\ 0 & 1 & -2\end{pmatrix}\begin{pmatrix}1 & 1 & 7\\ -1 & 0 & 2\\ 0 & 0 & 3\end{pmatrix}$$

$$=\begin{pmatrix}-1 & -1 & -4\\ 2 & 1 & 11\\ -1 & 0 & -4\end{pmatrix}$$

i.e. the eigenvalues of B are 1, 3, 4 and the corresponding eigenvectors are

$\begin{pmatrix}-1\\2\\-1\end{pmatrix},\begin{pmatrix}-1\\1\\0\end{pmatrix},\begin{pmatrix}-4\\11\\-4\end{pmatrix}$

5 The diagram is a compartmental representation of the exterior and interior of a cell. The concentration of fluid at any time t in compartment A and B are x units and y units respectively. Nutrients are allowed to permeate through the membrane such that the concentration of fluid in compartment A and B increases at a rate of $6(y-kx)$ and $2\left(y-\dfrac{x}{k}\right)$ respectively where k is a positive constant. Additional nutrients are also constantly supplied so that there is a constant rate of increase of concentration of 13 units to each compartment. The rate of change of x and y are the same at the instant when the concentration of fluid in compartment A is four times that of compartment B.



Membrane

(i) Write down two differential equations, one relating x and t and another relating y and t. Hence show that

$$\frac{\mathrm{d}^2x}{\mathrm{d}t^2}+2\frac{\mathrm{d}x}{\mathrm{d}t}+10x=52.$$ [5]

(ii) Given that when $t = 0$, $x = y = 0$, find the solution of x in terms of t. [7]

(iii) By sketching the solution curve of x against t, describe what happens to the concentration of fluid in compartment A in the long term. [2]

(iv) Given that as $t \to \infty, y \to \dfrac{13}{10}$, explain how these long term values of x and y can be found without solving the second order differential equations. [2]

[Solution]:

(i)

$$\frac{\mathrm{d}x}{\mathrm{d}t}=6(y-kx)+13$$

$$\frac{\mathrm{d}y}{\mathrm{d}t}=2\left(y-\frac{x}{k}\right)+13$$

When $x=4y, \dfrac{\mathrm{d}x}{\mathrm{d}t}=\dfrac{\mathrm{d}y}{\mathrm{d}t}$

$$6(y-4ky)+13=2\left(y-\frac{4y}{k}\right)+13$$

$$6k^2-k-2=0 \text{ when } y\neq 0$$

$$k=\frac{2}{3} \text{ or } k=-\frac{1}{2} \text{ (reject } \because k>0)$$

$$\frac{\mathrm{d}x}{\mathrm{d}t}=6y-4x+13 \text{ ----- (1)}$$

$$\frac{\mathrm{d}y}{\mathrm{d}t}=2y-3x+13 \text{ ------(2)}$$

Differentiate (1) w.r.t. t:

$$\frac{\mathrm{d}^2x}{\mathrm{d}t^2}=6\frac{\mathrm{d}y}{\mathrm{d}t}-4\frac{\mathrm{d}x}{\mathrm{d}t}$$

$$=6(2y-3x+13)-4\frac{\mathrm{d}x}{\mathrm{d}t}$$

From (1), $\frac{\mathrm{d}x}{\mathrm{d}t}=6y-4x+13 \Rightarrow 6y=\frac{\mathrm{d}x}{\mathrm{d}t}+4x-13$

$$\frac{\mathrm{d}^2x}{\mathrm{d}t^2}=6(2y-3x+13)-4\frac{\mathrm{d}x}{\mathrm{d}t}$$

$$=2\left(\frac{\mathrm{d}x}{\mathrm{d}t}+4x-13\right)-18x+78-4\frac{\mathrm{d}x}{\mathrm{d}t}$$

$$\frac{\mathrm{d}^2x}{\mathrm{d}t^2}+2\frac{\mathrm{d}x}{\mathrm{d}t}+10x=52 \text{ (shown)}$$

(ii) Consider the auxiliary equation $m^2+2m+10=0$

$$m=\frac{-2\pm\sqrt{-36}}{2}$$

$$=-1\pm 3\mathrm{i}$$

Complementary function is $x=\mathrm{e}^{-t}\left(A\sin 3t+B\cos 3t\right)$

Consider the particular integral, $x=p$ where p is a constant.

$$\frac{\mathrm{d}^2x}{\mathrm{d}t^2}=\frac{\mathrm{d}x}{\mathrm{d}t}=0\Rightarrow 10p=52$$

$$p=\frac{26}{5}$$

General solution is $x=\mathrm{e}^{-t}\left(A\sin 3t+B\cos 3t\right)+\frac{26}{5}$

When $t=0, x=0$

$$B=-\frac{26}{5}$$

Since $\frac{\mathrm{d}x}{\mathrm{d}t}=\mathrm{e}^{-t}\left(3A\cos 3t-3B\sin 3t\right)-\mathrm{e}^{-t}\left(A\sin 3t+B\cos 3t\right)$

and $\frac{\mathrm{d}x}{\mathrm{d}t}=6y-4x+13$, so when $t=0, x=y=0$,

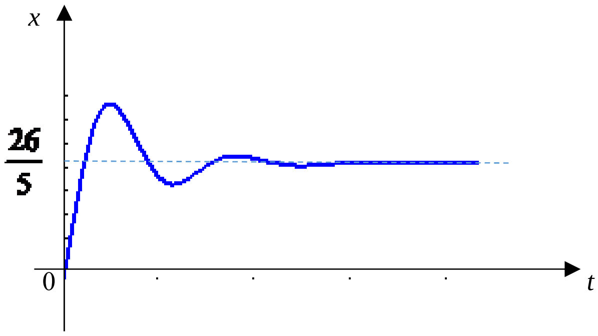
$3A-B=13$

Subst. $B = -\dfrac{26}{5}$, $A = \dfrac{13}{5}$

Solution is

$$x = \frac{e^{-t}}{5}(13\sin 3t - 26\cos 3t) + \frac{26}{5}$$

(iii)



As $t \to \infty, e^{-t} \to 0$,

Since $13\sin 3t - 26\cos 3t = \sqrt{845}\sin\left(3t - \tan^{-1}\dfrac{26}{13}\right)$ is a periodic function with fixed amplitude,

$\therefore \dfrac{e^{-t}}{5}(13\sin 3t - 26\cos 3t) \to 0$ as $t \to \infty$

and $\therefore x \to \dfrac{26}{5}$

(iv) Since both x and y tends to a constant in the long term, we can substitute

$$\frac{dx}{dt} = 0 \text{ and } \frac{dy}{dt} = 0 \text{ into the DE (1) and (2)}$$

$$\frac{dx}{dt} = 6y - 4x + 13 \quad = 0 \text{ ----- (1)}$$

$$\frac{dy}{dt} = 2y - 3x + 13 \quad = 0 \text{ ------(2)}$$

and solve the two simultaneous equations for x and y which will give the limiting values of $\dfrac{26}{5}$ and $\dfrac{13}{10}$ respectively.

Section B: Statistics (50 marks)

6 In a population, a person is either left-handed or right-handed. On average, a proportion, p of the population is left-handed. A random sample of people is chosen, one by one, until a left-handed person is selected. Find in terms of p, the probability that the number of people in the sample is less than 15. [2]

Another random sample is chosen, one by one, until at least one left-handed and at least one right-handed person have been obtained. The number of people in this sample is denoted by N. Show that $\mathrm{P}(N = n) = pq^{n-1} + qp^{n-1}$, for $n \geq 2$, where $q = 1 - p$. Hence find $\mathrm{E}(N)$ in terms of p and q. [5]

[Solution]

Let L be the number of people chosen until a left-handed person is chosen. $L \sim \text{Geo}(p)$

$$\mathrm{P}(L < 15) = \mathrm{P}(L \leq 14)$$
$$= 1 - (1 - p)^{14}$$

Let $q = 1 - p$

Let R be the number of people chosen until a right-handed person is chosen. $R \sim \text{Geo}(q)$

$P(N = r) =$ P($(r-1)$ right-handed people chosen followed by one left-handed person) + P($(r-1)$ left-handed people chosen followed by one right-handed person)

$$= \mathrm{P}(L = r) + \mathrm{P}(R = r)$$
$$= pq^{r-1} + qp^{r-1}$$

$$\mathrm{E}(N) = \sum_{r=2}^{\infty} r\mathrm{P}(N = r)$$
$$= \sum_{r=2}^{\infty} r\left(pq^{r-1} + qp^{r-1}\right)$$
$$= \sum_{r=1}^{\infty}\left(rpq^{r-1}\right) - p + \sum_{r=1}^{\infty}\left(rqp^{r-1}\right) - q$$
$$= \mathrm{E}(L) - p + \mathrm{E}(R) - q$$
$$= \frac{1}{p} - p + \frac{1}{q} - q$$

7 A metalworker makes circular discs of various sizes but with the same thickness of 0.1 cm. The length, X cm, of the radius of a randomly chosen disc has a uniform distribution in the interval $\left[\frac{10}{\pi}, \frac{60}{\pi}\right]$. The volume of a randomly chosen disc is denoted by Y cm^3.

Show that $\mathrm{E}(Y) = \frac{430}{3\pi}$ and $\mathrm{Var}(Y) = \frac{c}{9\pi^2}$, where c is a constant to be determined. [6]

A random sample of 200 discs made by the metalworker has mean volume 40 cm^3. Obtain a 90% confidence interval for the mean volume of the discs produced by the metalworker. [1]

Give a possible reason why the interval you have found may not contain the population mean. [1]

[Solution]:

$$\mathrm{f}(x) = \begin{cases} \frac{\pi}{50}, & \frac{10}{\pi} \le x \le \frac{60}{\pi} \\ 0, & \text{otherwise} \end{cases}$$

$$Y = \frac{\pi x^2}{10}$$

$$\mathrm{E}(Y) = \mathrm{E}\left(\frac{\pi x^2}{10}\right)$$
$$= \int_{\frac{10}{\pi}}^{\frac{60}{\pi}} \frac{\pi x^2}{10}\left(\frac{\pi}{50}\right) \mathrm{d}x$$
$$= \frac{\pi^2}{500}\int_{\frac{10}{\pi}}^{\frac{60}{\pi}} x^2 \ \mathrm{d}x$$
$$= \frac{\pi^2}{500}\left[\frac{x^3}{3}\right]_{\frac{10}{\pi}}^{\frac{60}{\pi}}$$
$$= \frac{\pi^2}{1500}\left[\left(\frac{60}{\pi}\right)^3 - \left(\frac{10}{\pi}\right)^3\right]$$
$$= \frac{430}{3\pi}$$

$$\mathrm{E}(Y^2) = \mathrm{E}\left(\frac{\pi^2 x^4}{100}\right)$$
$$= \int_{\frac{10}{\pi}}^{\frac{60}{\pi}} \frac{\pi^2 x^4}{100}\left(\frac{\pi}{50}\right) \mathrm{d}x$$
$$= \frac{\pi^3}{5000}\int_{\frac{10}{\pi}}^{\frac{60}{\pi}} x^4 \ \mathrm{d}x$$
$$= \frac{\pi^3}{5000}\left[\frac{x^5}{5}\right]_{\frac{10}{\pi}}^{\frac{60}{\pi}}$$
$$= \frac{\pi^3}{25000}\left[\left(\frac{60}{\pi}\right)^5 - \left(\frac{10}{\pi}\right)^5\right]$$
$$= \frac{31100}{\pi^2}$$

$$\mathrm{Var}(Y) = \mathrm{E}(Y^2) - \left[\mathrm{E}(Y)\right]^2$$
$$= \frac{31100}{\pi^2} - \left(\frac{430}{3\pi}\right)^2$$
$$= \frac{95000}{9\pi^2}$$
$c = 95000$

Using GC, a 90% confidence interval for the mean volume of the discs is (36.2, 42.8)

There is a 10% chance that the population mean lie outside the 90% confidence interval.

8 Given that the random variable X has the Poisson distribution with mean a, show that

$$\sum_{r=0}^{n} r\mathrm{P}(X=r) = a\mathrm{P}(X \le n-1)\,. \qquad [3]$$

A garage receives delivery of new cars at the beginning of each month and accepts as many new cars as is necessary to bring its stock of new cars to 10. The monthly demand for new cars at the garage has a Poisson distribution with mean 8.

(i) Find the probability that in a particular month there will be insufficient cars to sell. [1]

(ii) Calculate the minimum number of stock of new cars that are to be increased to have at least 90% chance of meeting the demand. [3]

(iii) Find to two significant figures, the mean of the number of new cars sold per month by the garage. [3]

[Solution]:

$X \sim \mathrm{P_o}(a)$, $\quad \mathrm{P}(X=r) = \dfrac{\mathrm{e}^{-a}a^r}{r!}$, $r = 0, 1, 2, \ldots$

$$\begin{aligned}
\sum_{r=0}^{n} r\mathrm{P}(X=r) &= \sum_{r=0}^{n} r\left(\frac{\mathrm{e}^{-a}a^r}{r!}\right) = \sum_{r=1}^{n} r\left(\frac{\mathrm{e}^{-a}a^r}{r!}\right) \\
&= a\sum_{r=1}^{n} \frac{\mathrm{e}^{-a}a^{r-1}}{(r-1)!} \\
&= a\left(\frac{\mathrm{e}^{-a}}{0!} + \frac{\mathrm{e}^{-a}a}{1!} + \frac{\mathrm{e}^{-a}a^2}{2!} + \ldots + \frac{\mathrm{e}^{-a}a^{n-1}}{(n-1)!}\right) \\
&= aP(X \le n-1) \qquad \text{(Shown)}
\end{aligned}$$

Let X be the monthly demand for new cars. $X \sim \mathrm{P_o}(8)$.

Let Y be the number of new cars sold per month.

(i) P (on a given month, there will be insufficient cars to sell)

$= \mathrm{P}(X > 10) = 1 - \mathrm{P}(X \le 10) = 0.184$

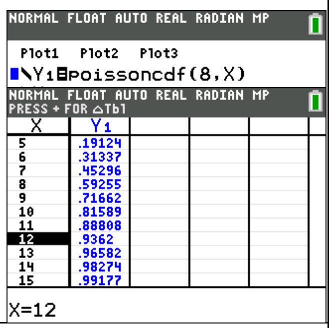
(ii) Require to find the smallest integer n, such that $\mathrm{P}(X \le n) \ge 0.9$

Since from GC, $\mathrm{P}(X \le 11) = 0.8881 < 0.9$

and $\mathrm{P}(N \le 12) = 0.9362 > 0.9$

$\therefore$ Least $n = 12$.

Hence, we need to increase 2 more stock of new cars to have at least 90% chance of meeting the demand.



(iii) $$\mathrm{E}(Y) = \sum_{r=0}^{n} r\mathrm{P}(Y=r) = \sum_{r=0}^{9} r\mathrm{P}(X=r) + \sum_{r=10}^{\infty} 10\mathrm{P}(X=r)$$

$$= 8\mathrm{P}(X \le 9-1) + 10\,\mathrm{P}(X \ge 10)$$

$= 8\mathrm{P}(X \le 8) + 10\{1 - \mathrm{P}(X \le 9) = 7.6$ (to 2 s.f.)

9 **(a)** A sample of 1000 observations of the continuous random variable X was obtained and the results are summarised in the following table, in which s is an unknown integer.

Interval	$0 \le x < 0.2$	$0.2 \le x < 0.4$	$0.4 \le x < 0.6$	$0.6 \le x < 0.8$	$0.8 \le x \le 1.0$
Observed frequency	s	$65 - s$	159	310	466

It is assumed that these results are consistent with a probability distribution having the following probability density function:

$$\mathrm{f}(x) = \begin{cases} 3x^k, & 0 \le x \le 1 \\ 0 \ , & \text{otherwise} \end{cases}$$, where k is a constant to be determined.

A χ^2 -test is to be carried out to test the goodness of fit of this distribution. The expected frequencies are given in the following table where a and b are constants.

Interval	$0 \le x < 0.2$	$0.2 \le x < 0.4$	$0.4 \le x < 0.6$	$0.6 \le x < 0.8$	$0.8 \le x \le 1.0$
Expected frequency	8	56	a	b	488

(i) Show that $a = 152$ and find the value of b. [3]

(ii) Find the largest value of s that would result in the null hypothesis not being rejected at the 5% level of significance. [5]

(b) The following table gives the percentage of the different grades obtained by candidates taking Statistics and Pure Mathematics in a school examination. The two groups of candidates are mutually exclusive.

Grade	A, B or C	D or E	S or U
Statistics	41.5	23.8	34.7
Pure Mathematics	29.2	26.2	44.6

(i) What further information is needed to form a 3×2 expected frequency table from which the independence of subjects and grades can be tested? [1]

(ii) When such a table was formed, the calculated value of

$$\sum \frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}}$$ was 13.16.

Assume that all E_{ij} are more than 5, carry out the test using a 0.5% significance level. [4]

[Solution]

(a) $\int_0^1 (3x^k)\,dx = 1 \quad \Rightarrow \quad \left[\dfrac{3x^{k+1}}{k+1}\right]_0^1 = 1 \quad \Rightarrow \quad \dfrac{3}{k+1} = 1 \quad \Rightarrow \quad k = 2$

(i) H_0: The sample is consistent with the given pdf, f(x).
H_1: The sample is not consistent with the given pdf, f(x).

Based on H_0,

$$a = 1000\int_{0.4}^{0.6}(3x^2)dx = 1000(0.152) = 152$$

$$b = 1000 - 8 - 56 - 152 - 488 = 296$$

Degrees of freedom = 5 – 1 = 4
The only constraint is: $\sum O_i = \sum E_i$ = 1000,

(ii) Test statistic: $\sum \dfrac{(O_i - E_i)^2}{E_i} \sim \chi_4^2$

$$\chi^2_{\text{cal}} = \frac{(s-8)^2}{8} + \frac{(65-s-56)^2}{56} + \frac{(159-152)^2}{152} + \frac{(310-296)^2}{296} + \frac{(466-488)^2}{488}$$

$$= \frac{s^2 - 16s + 64}{8} + \frac{81 - 18s + s^2}{56} + \frac{49}{152} + \frac{196}{296} + \frac{484}{488}$$

$$= \frac{529 - 130s + 8s^2}{56} + 1.9763$$

At 5% level of significance, critical region = $\{\chi^2 : \chi^2 \ge 9.488\}$
For H_0 not to be rejected, $\chi^2_{\text{cal}} < 9.488$
$8s^2 - 130s + 108.345 < 0 \Rightarrow 0.8812 < s < 15.369$
Largest value of s = 15.

(b) (i) The number of candidates taking Statistics paper and the number of candidates taking Pure Mathematics paper are required.

(iii) H_0: Subjects and grades are independent

H_1: Subjects and grades are not independent

Degrees of freedom = (3 – 1) (2 – 1) = 2

Test statistics: $\sum \dfrac{(O_{ij} - E_{ij})^2}{E_{ij}} \sim \chi_2^2$

Given that χ^2_{cal} = 13.16,
Level of significance: 0.5%

From χ^2 table, we have critical region = $\{ \chi^2 : \chi^2 \ge 10.60 \}$

Since χ^2_{cal} = 13.16 > 10.60, we reject H_0
Hence there is sufficient evidence to conclude that subjects and grades are not independent at the 0.5% level of significance.

10 A study was carried out to investigate the effect of a certain drug on blood cholesterol levels. The levels were measured before and after the drug was administered to a random sample of 9 people. The results, in suitable units, are given in the following table.

Person	A	B	C	D	E	F	G	H	I
Before	170	190	200	188	206	247	191	222	263
After	161	199	190	179	195	235	191	225	252

(i) Perform a suitable t-test at the 5% significance level whether the drug is effective in reducing the blood cholesterol level, stating any necessary conditions for the validity of the test. [5]

(ii) Find the greatest integer value of a for which it could be claimed at the 10% significance level that the drug is effective in reducing the blood cholesterol level by more than a units. [3]

(iii) If the conditions for the test in **(i)** are not met, test whether the drug is effective in reducing the blood cholesterol level at the 5% significance level using the Wilcoxon matched-pairs signed rank test. [4]

[Solution]:

(i) Let X be the cholesterol level of a person before administration of the drug.
Let Y represents the cholesterol level of a person after administration of the drug.
Let $D = X - Y$
Condition: D follows a normal distribution
Using GC,

$\bar{d} = 5.55556$

$s_D^{\ 2} = (7.58471)^2$

$n = 9$

Let $\mu_D = \mu_x - \mu_y$

To test $\quad H_0 : \mu_D = 0$

against $\quad H_1 : \mu_D > 0$

Level of significance: 5% .

Test statistic: $T = \dfrac{\bar{D} - \mu_D}{\frac{S_D}{\sqrt{n}}} \sim t(8)$

If H_0 is true, from GC, p-value = 0.0296.

Since p-value = $0.0296 \le 0.05$, we reject Ho and conclude that there is sufficient evidence at 5% significance level that the drug is effective in reducing the blood cholesterol level.

(ii) $H_0 : \mu_D = a$

$H_1 : \mu_D > a$

Level of significance: 10%

Test statistic: $T = \dfrac{\overline{D} - \mu_D}{\dfrac{S_D}{\sqrt{n}}} \sim t(8)$

If H_0 is true, $t_{cal} = \dfrac{5.55556 - a}{\dfrac{7.58471}{3}}$

For H_0 to be rejected, $t_{cal} = \dfrac{5.55556 - a}{\dfrac{7.58471}{3}} \geq 1.39682$

$a \leq 2.02$

Greatest integer value of a is 2.

(iii) Using the Wilcoxon match-pair signed rank test, we have

$H_0 : M_d = 0$ where M_d represents the median score of the difference $D = Y - X$.
$H_1 : M_d < 0$

Level of significance: 5%

d_i	−9	9	−10	−9	−11	−12	0	3	−11
Ranks	3	3	5	3	6.5	8		1	6.5

P = sum of "+" rank = 3 + 1 = 4
Q = sum of "−" rank = 3 + 5 + 3 + 6.5 + 8 + 6.5 = 32

$T_{cal} = \min(P, Q) = P = 4$

Now, for $n = 8$ (pairs of observations),
1-tailed test at 5% level of significance, critical region = $\{t: t \leq 5\}$

As T_{cal} falls inside the critical region, we reject H_0.
There is sufficient evidence at 5% level of significance that the drug is effective in reducing the blood cholesterol level.

-------- ***End of Paper*** --------