2018 GCE A Level H2 Further Maths 9649 Paper 2 Solutions

Section A: Pure Mathematics

Question 1

1 In this question, V denotes the set of vectors of the form $\begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix}$, where a, b, c and d are real numbers.

You may assume that V forms a linear space under the usual operations of vector addition and multiplication by a scalar.

(i) Show that the subset of V for which $a + b + c + d = 0$ forms a linear space. [3]

(ii) Show that the subset of V for which $a + b + c + d = 1$ does not form a linear space. [1]

(iii) Determine whether or not the subset of V for which $a + b = c + d$ and $a + 2b = c + 3d$ forms a linear space. [2]

(iv) State the dimension of the linear space defined in part **(i)** and provide a basis for this linear space. [3]

(i) Let $W \subset V$.

Let $\mathbf{u} = \begin{pmatrix} a_1 \\ b_1 \\ c_1 \\ d_1 \end{pmatrix} \in W$ and $\mathbf{v} = \begin{pmatrix} a_2 \\ b_2 \\ c_2 \\ d_2 \end{pmatrix} \in W$. Then $a_1 + b_1 + c_1 + d_1 = 0$ and $a_2 + b_2 + c_2 + d_2 = 0$

$$\mathbf{u} + \mathbf{v} = \begin{pmatrix} a_1 \\ b_1 \\ c_1 \\ d_1 \end{pmatrix} + \begin{pmatrix} a_2 \\ b_2 \\ c_2 \\ d_2 \end{pmatrix} = \begin{pmatrix} a_1 + a_2 \\ b_1 + b_2 \\ c_1 + c_2 \\ d_1 + d_2 \end{pmatrix} \in W$$

since $a_1 + a_2 + b_1 + b_2 + c_1 + c_2 + d_1 + d_2 = (a_1 + b_1 + c_1 + d_1) + (a_2 + b_2 + c_2 + d_2) = 0$

Let $\alpha \in \mathbb{R}$. Then $\alpha\mathbf{u} = \begin{pmatrix} \alpha a_1 \\ \alpha b_1 \\ \alpha c_1 \\ \alpha d_1 \end{pmatrix} \in W$

since $\alpha a_1 + \alpha b_1 + \alpha c_1 + \alpha d_1 = \alpha(a_1 + b_1 + c_1 + d_1) = 0$

Also $\mathbf{0} \in W$ and Since W is closed under vector addition and scalar multiplication, W forms a linear space.

(ii) Let $T \subset V$ such that $a+b+c+d=1$

$\begin{pmatrix}1\\0\\0\\0\end{pmatrix} \in T$ but $2\begin{pmatrix}1\\0\\0\\0\end{pmatrix} \notin T$. Thus T does not form a linear space.

(iii) Let $S \subset V$ such that $a + b = c + d$ and $a + 2b = c + 3d$. Then $b = 2d$ and $c = a + d$

Let $\mathbf{u} = \begin{pmatrix}a_1\\2d_1\\a_1+d_1\\d_1\end{pmatrix} \in S$ and $\mathbf{v} = \begin{pmatrix}a_2\\2d_2\\a_2+d_2\\d_2\end{pmatrix} \in S$

$$\mathbf{u} + \mathbf{v} = \begin{pmatrix}a_1\\2d_1\\a_1+d_1\\d_1\end{pmatrix} + \begin{pmatrix}a_2\\2d_2\\a_2+d_2\\d_2\end{pmatrix} = \begin{pmatrix}a_1+a_2\\2(d_1+d_2)\\a_1+a_2+d_1+d_2\\d_1+d_2\end{pmatrix} \in S$$

Let $\alpha \in \mathbb{R}$. Then $\alpha\mathbf{u} = \begin{pmatrix}\alpha a_1\\2\alpha b_1\\\alpha(a_1+d_1)\\\alpha d_1\end{pmatrix} \in S$

Also, $\mathbf{0} \in S$ and since S is closed under vector addition and scalar multiplication, S forms a linear space.

(iv) $\dim W = 3$

$$\begin{pmatrix}a\\b\\c\\d\end{pmatrix} = \begin{pmatrix}a\\b\\c\\-a-b-c\end{pmatrix} = a\begin{pmatrix}1\\0\\0\\-1\end{pmatrix} + b\begin{pmatrix}0\\1\\0\\-1\end{pmatrix} + c\begin{pmatrix}0\\0\\1\\-1\end{pmatrix}$$

$\left\{\begin{pmatrix}1\\0\\0\\-1\end{pmatrix}, \begin{pmatrix}0\\1\\0\\-1\end{pmatrix}, \begin{pmatrix}0\\0\\1\\-1\end{pmatrix}\right\}$ is a basis for W.

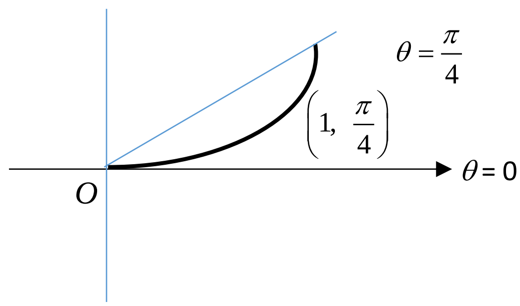
Question 2

2 (i) Sketch the curve with polar equation $r = \tan\theta$ for $0 \leqslant \theta \leqslant \frac{1}{4}\pi$. [1]

(ii) Find, in an exact form, the area of the region enclosed by the curve and the line $\theta = \frac{1}{4}\pi$. [3]

(iii) Find a cartesian equation for the curve, expressing y in terms of x and giving the domain for x. [4]

(i)



(ii) Area $= \frac{1}{2}\int_0^{\frac{\pi}{4}} \tan^2\theta \, \mathrm{d}\theta$

$= \frac{1}{2}\int_0^{\frac{\pi}{4}} \sec^2\theta - 1 \, \mathrm{d}\theta$

$= \frac{1}{2}\left[\tan\theta - \theta\right]_0^{\frac{\pi}{4}}$

$= \frac{1}{2}\left(1 - \frac{\pi}{4}\right)$ units^2

(iii) **Method 1**

By observation, there is a right angled isosceles triangle with a hypotenuse of length 1.

Hence, $x^2 + x^2 = 1 \Rightarrow x = \frac{1}{\sqrt{2}}$

From the graph, the domain is $[0, \frac{1}{\sqrt{2}}]$.

Method 2

$x = r\cos\theta = \tan\theta\cos\theta$

When $\theta = \frac{\pi}{4}$, $x = \tan\left(\frac{\pi}{4}\right)\cos\left(\frac{\pi}{4}\right) = \frac{1}{\sqrt{2}}$

From the graph, the domain is $[0, \frac{1}{\sqrt{2}}]$.

Common Mistake

$r = \tan\theta$

$$\sqrt{x^2+y^2}=\frac{\sin\theta}{\cos\theta}=\frac{r\sin\theta}{r\cos\theta}=\frac{y}{x}$$
$$y^2=x^2\left(x^2+y^2\right)$$
$$y^2=\frac{x^4}{1-x^2}$$

Domain for x:

$$1-x^2>0$$
$$(x+1)(x-1)<0$$
$$-1<x<1$$

Question 3

3 **(i)** Obtain the solution of the differential equation

$$\frac{\mathrm{d}y}{\mathrm{d}x} - y = x$$

given that $y = 0$ when $x = 0$.

Obtain, correct to 3 significant figures, the values of y when $x = 0.1$ and $x = 0.2$. [4]

Now consider the differential equation

$$\frac{\mathrm{d}y}{\mathrm{d}x} - \sin y = x,$$

where $y = 0$ when $x = 0$.

(ii) Use Euler's method with step length 0.1 to estimate the values of y when $x = 0.1$ and $x = 0.2$. [3]

(iii) Use the improved Euler method with step length 0.1 to estimate the values of y when $x = 0.1$ and $x = 0.2$. [3]

(iv) Comment on your numerical answers from parts **(i)**, **(ii)** and **(iii)**. [2]

(i) $\frac{\mathrm{d}y}{\mathrm{d}x} - y = x$

Integrating factor is $\mathrm{e}^{\int -1\,\mathrm{d}x} = \mathrm{e}^{-x}$

$$y\mathrm{e}^{-x} = \int x\mathrm{e}^{-x}\ \mathrm{d}x$$

$$= -x\mathrm{e}^{-x} + \int \mathrm{e}^{-x}\ \mathrm{d}x$$

$$= -x\mathrm{e}^{-x} - \mathrm{e}^{-x} + c$$

$$y = c\mathrm{e}^{x} - x - 1$$

When $x = 0$, $y = 0$, $c = 1$

The solution is $y = \mathrm{e}^{x} - x - 1$

When $x = 0.1$, $y = 0.00517$

When $x = 0.2$, $y = 0.0214$

(ii) $\frac{\mathrm{d}y}{\mathrm{d}x} = \sin y + x$

Using Euler's method, with $x_0 = 0$, $y_0 = 0$,

$y_1 = 0 + 0.1(\sin 0 + 0) = 0$

$y_2 = 0 + 0.1(\sin 0 + 0.1) = 0.01$

(iii) Using Improved Euler's method, with $x_0 = 0$, $y_0 = 0$

$u_1 = 0$

$y_1 = 0 + \frac{1}{2}(0.1)[(\sin 0 + 0) + (\sin 0 + 0.1)] = 0.005$

$u_2 = 0.005 + 0.1(\sin 0.005 + 0.1) = 0.0154999979$

$y_2 = 0.005 + \frac{1}{2}(0.1)[(\sin 0.005 + 0.1) + (\sin 0.01549999979 + 0.2)] = 0.0210$

(iv) When y is small, $\sin y \approx y$. Thus the solution to $\frac{dy}{dx} - \sin y = x$ is an approximation to the solution of $\frac{dy}{dx} - y = x$. The answers obtained in part (iii) are closer to those obtained in (i). Hence Improved Euler's method gives a better approximation than the Euler's method.

Question 4

4 The 3×3 matrix $\mathbf{M}$ has characteristic equation

$$\lambda^3 + 2\lambda^2 - 9\lambda - 18 = 0.$$

(i) Show that $\mathbf{M}$ has only one positive eigenvalue and state its value. [2]

It is now given that $\begin{pmatrix} 3 \\ 2 \\ -1 \end{pmatrix}$ is an eigenvector corresponding to the positive eigenvalue of $\mathbf{M}$.

(ii) Write down the vectors given by $\mathbf{M}\begin{pmatrix} -3 \\ -2 \\ 1 \end{pmatrix}$ and $\mathbf{M}^2\begin{pmatrix} 1 \\ \frac{2}{3} \\ -\frac{1}{3} \end{pmatrix}$.

Solve the equation $\mathbf{M}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 18 \\ 12 \\ -6 \end{pmatrix}$. [5]

(iii) Find an expression for $\mathbf{M}^4$ in the form $a\mathbf{M}^2 + b\mathbf{M} + c\mathbf{I}$, where a, b and c are constants to be found and $\mathbf{I}$ is the 3×3 identity matrix. [2]

(i) $\lambda^3 + 2\lambda^2 - 9\lambda - 18 = 0$

Using GC, $\lambda = -3, -2, 3$

Thus M has only one positive eigenvalue 3.

(ii) Given $\mathbf{M}\begin{pmatrix} 3 \\ 2 \\ -1 \end{pmatrix} = 3\begin{pmatrix} 3 \\ 2 \\ -1 \end{pmatrix}$

$$\mathbf{M}\begin{pmatrix} -3 \\ -2 \\ 1 \end{pmatrix} = -3\begin{pmatrix} 3 \\ 2 \\ -1 \end{pmatrix} = \begin{pmatrix} -9 \\ -6 \\ 3 \end{pmatrix}$$

$$\mathbf{M}^2\begin{pmatrix} 1 \\ \frac{2}{3} \\ -\frac{1}{3} \end{pmatrix} = \frac{1}{3}\mathbf{M}\left(\mathbf{M}\begin{pmatrix} 3 \\ 2 \\ -1 \end{pmatrix}\right) = \frac{1}{3}\mathbf{M}\left(3\begin{pmatrix} 3 \\ 2 \\ -1 \end{pmatrix}\right) = \mathbf{M}\begin{pmatrix} 3 \\ 2 \\ -1 \end{pmatrix} = \begin{pmatrix} 9 \\ 6 \\ -3 \end{pmatrix}$$

Since $\mathbf{M}\begin{pmatrix} 3 \\ 2 \\ -1 \end{pmatrix} = 3\begin{pmatrix} 3 \\ 2 \\ -1 \end{pmatrix}$, $\mathbf{M}\begin{pmatrix} 6 \\ 4 \\ -2 \end{pmatrix} = 6\begin{pmatrix} 3 \\ 2 \\ -1 \end{pmatrix} = \begin{pmatrix} 18 \\ 12 \\ -6 \end{pmatrix}$

Thus the solution is $\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 6 \\ 4 \\ -2 \end{pmatrix}$

(iii) **Method 1**: By Cayley-Hamilton Theorem

Note: The Cayley-Hamilton theorem states that ***every matrix satisfies its own characteristic equation.***

Characteristic eqn: $\lambda^3 + 2\lambda^2 - 9\lambda - 18 = 0$

By Cayley-Hamilton Theorm,

$\mathbf{M}^3 + 2\mathbf{M}^2 - 9\mathbf{M} - 18\mathbf{I} = 0$

$\mathbf{M}^3 = -2\mathbf{M}^2 + 9\mathbf{M} + 18\mathbf{I}$ ------- (*)

$$\begin{aligned}\mathbf{M}^4 &= -2\mathbf{M}^3 + 9\mathbf{M}^2 + 18\mathbf{M} \\ &= -2\left(-2\mathbf{M}^2 + 9\mathbf{M} + 18\mathbf{I}\right) + 9\mathbf{M}^2 + 18\mathbf{M} \quad \text{(from (*))} \\ &= 4\mathbf{M}^2 - 18\mathbf{M} - 36\mathbf{I} + 9\mathbf{M}^2 + 18\mathbf{M} \\ &= 13\mathbf{M}^2 - 36\mathbf{I}\end{aligned}$$

Hence $a = 13,\ b = 0,\ c = -36$

Method 2

$\mathbf{M}^4 = a\mathbf{M}^2 + b\mathbf{M} + c\mathbf{I}$

$\left(\mathbf{PDP}^{-1}\right)^4 = a\left(\mathbf{PDP}^{-1}\right)^2 + b\mathbf{PDP}^{-1} + c\mathbf{PP}^{-1}$

$\mathbf{PD}^4\mathbf{P}^{-1} = \mathbf{P}\left[a\mathbf{D}^2 + b\mathbf{D} + c\mathbf{I}\right]\mathbf{P}^{-1}$

$\mathbf{D}^4 = a\mathbf{D}^2 + b\mathbf{D} + c\mathbf{I}$

By comparing the diagonal entries,

$81 = 9a - 3b + c$

$16 = 4a - 2b + c$

$81 = 9a + 3b + c$

Hence, by GC, $a = 13,\ b = 0,\ c = -36$.

Method 3

$\mathbf{M}^4 = a\mathbf{M}^2 + b\mathbf{M} + c\mathbf{I}$

$\mathbf{M}^4\mathbf{x} = a\mathbf{M}^2\mathbf{x} + b\mathbf{M}\mathbf{x} + c\mathbf{x}$

$\lambda^4\mathbf{x} = a\lambda^2\mathbf{x} + b\lambda\mathbf{x} + c\mathbf{x}$

$\lambda^4 = a\lambda^2 + b\lambda + c$

Using the eigenvalues found in (i),

$81 = 9a - 3b + c$

$16 = 4a - 2b + c$

$81 = 9a + 3b + c$

Hence, by GC, $a = 13,\ b = 0,\ c = -36$.

Question 5

5 (a) In an Argand diagram, the point P represents the complex number $1 + 2i$ and the point Q represents the complex number $4 + 7i$. When the line PQ is rotated anticlockwise about P through θ radians, the image of Q is the point R.

(i) Obtain the complex number represented by R in the case when $\theta = \frac{1}{4}\pi$. [4]

(ii) Find, correct to 3 significant figures, the smallest positive value of θ for which R lies on the imaginary axis. [3]

(b) Illustrate, in separate Argand diagrams, the sets of points z for which

(i) $\text{Re}(z^2) < 0$, [3]

(ii) $\text{Im}(z^3) > 0$. [2]

(a)(i) $PQ = \sqrt{3^2+5^2} = \sqrt{34}$

$\tan\theta = \dfrac{5}{3} \quad \Rightarrow \quad \sin\theta = \dfrac{5}{\sqrt{34}}$ and $\cos\theta = \dfrac{3}{\sqrt{34}}$

Let $z = \sqrt{34}e^{i\theta}$

and $z' = \sqrt{34}e^{i\left(\theta+\frac{\pi}{4}\right)} = \sqrt{34}\left(\cos\left(\theta+\dfrac{\pi}{4}\right)+i\sin\left(\theta+\dfrac{\pi}{4}\right)\right)$

$$= \sqrt{34}\left(\frac{\sqrt{2}}{2}\cos\theta-\frac{\sqrt{2}}{2}\sin\theta+i\left(\frac{\sqrt{2}}{2}\sin\theta+\frac{\sqrt{2}}{2}\cos\theta\right)\right)$$

$$= \frac{\sqrt{2}}{2}(-2+8i)$$

$$= -\sqrt{2}+4\sqrt{2}i$$

Complex number representing R is $1+2i+\left(-\sqrt{2}+4\sqrt{2}i\right) = 1-\sqrt{2}+\left(2+4\sqrt{2}\right)i$

(ii) $z' = \sqrt{34}e^{i(\theta+\alpha)}$

For R to lie on the imaginary axis, the real part of z' is zero.

$1+3\cos\alpha-5\sin\alpha = 0$

$\sqrt{34}\cos\left(\alpha+\tan^{-1}\dfrac{5}{3}\right) = -1$

$\alpha+\tan^{-1}\dfrac{5}{3} = 1.398445737$

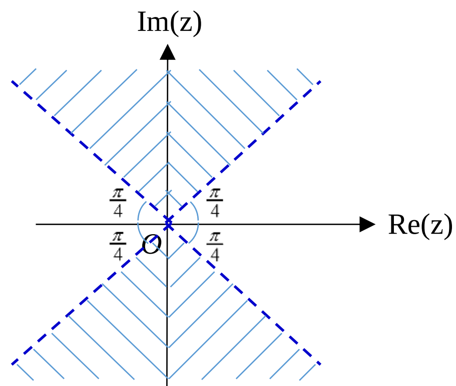
Smallest positive value of $\alpha = 0.713$

(b)(i)

Let $\arg(z) = \theta$, $\theta \in [0, 2\pi]$

Then $\arg z^2 = 2\arg z = 2\theta$, $2\theta \in [0, 4\pi]$

For $\text{Re}(z^2) < 0$,



$$2\theta \in \left(\frac{\pi}{2}, \frac{3\pi}{2}\right] \text{or} \left(\frac{5\pi}{2}, \frac{7\pi}{2}\right]$$

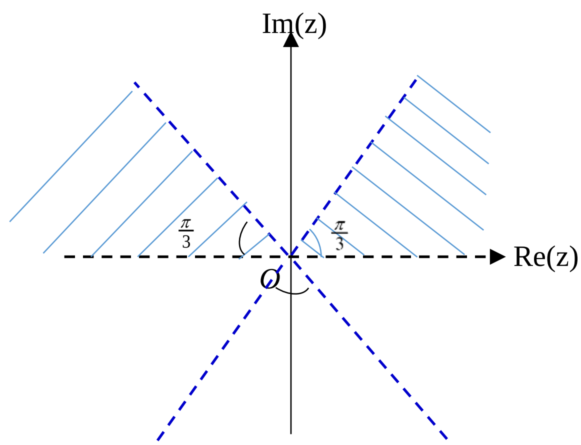
Hence $\theta \in \left(\frac{\pi}{4}, \frac{3\pi}{4}\right] \text{or} \left(\frac{5\pi}{4}, \frac{7\pi}{4}\right]$

(ii)

For $\text{Im}(z^3) > 0$,

$$3\theta \in (0, \pi) \text{or } (2\pi, 3\pi) \text{ or } (4\pi, 5\pi)$$

$$\theta \in \left(0, \frac{\pi}{3}\right) \text{or } \left(\frac{2\pi}{3}, \pi\right) \text{ or } \left(\frac{4\pi}{3}, \frac{5\pi}{3}\right)$$



Section B: Probability and Statistics

Question 6

6 The random variable X has a 'power law' distribution with parameter n if its probability density function (pdf) is of the form

$$f(x) = \begin{cases} kx^{-n} & \text{for } x \geqslant 1, \\ 0 & \text{otherwise,} \end{cases}$$

where k is a constant which depends on n. (Power law distributions occur frequently in the natural sciences and social sciences.)

(i) Show that, for the pdf to exist, n must be greater than 1 and that $k = n - 1$. [3]

(ii) For the case $n = 4$, find $\mathrm{E}(X)$ and $\mathrm{Var}(X)$. [2]

(iii) Show that, if $n = 3$, $\mathrm{Var}(X)$ does not exist. [2]

(iv) Find the set of values of n for which $\mathrm{E}(X)$ does not exist. [1]

$$f(x) = \begin{cases} kx^{-n} & \text{for } x \ge 1, \\ 0 & \text{otherwise} \end{cases}.$$

(i) $\int_{-\infty}^{\infty} f(x)\,dx = 1 \Rightarrow k\int_{1}^{\infty} x^{-n}\,dx = 1$

If $n = 1$, $k\left[\ln x\right]_1^{\infty} = 1$ Impossible as $\ln x \to \infty$ when $x \to \infty$.

If $n \neq 1$, $k\left[\dfrac{x^{1-n}}{1-n}\right]_1^{\infty} = 1 \Rightarrow \dfrac{1}{1-n}\left[\lim_{x\to\infty} x^{1-n} - 1\right] = 1$

x^{1-n} converges when $1-n<0$ i.e. $n>1$ and $\lim_{x\to\infty} x^{1-n} = 0$.

Hence $\dfrac{k}{1-n}(0-1) = 1 \Rightarrow k = n-1$

(ii) $n = 4$, $k = 3$, $f(x) = 3x^{-4}$

$\mathrm{E}(X) = \int_1^{\infty} x\left(3x^{-4}\right)dx = 3\int_1^{\infty} x^{-3}\,dx = 3\left[\dfrac{x^{-2}}{-2}\right]_1^{\infty} = \dfrac{3}{2}(0+1) = \dfrac{3}{2}$ (OR use GC)

$\mathrm{E}(X^2) = \int_1^{\infty} x^2\left(3x^{-4}\right)dx = 3\int_1^{\infty} x^{-2}\,dx = 3\left[-\dfrac{1}{x}\right]_1^{\infty} = 3$

$\mathrm{Var}(X) = \mathrm{E}(X^2) - \left[\mathrm{E}(X)\right]^2 = 3 - \left(\dfrac{3}{2}\right)^2 = \dfrac{3}{4}$

(iii) If $n = 3$, $k = 2$, $f(x) = 2x^{-3}$

$\mathrm{E}(X^2) = \int_1^{\infty} x^2\left(3x^{-3}\right)dx = 2\int_1^{\infty} \dfrac{1}{x}dx = 2\left[\ln x\right]_1^{\infty}$

Since $\ln x \to \infty$ when $x \to \infty$, $\mathrm{E}(X^2)$ is infinite.

Hence Var(X) does not exist.

(iv) $\quad \mathrm{E}(X) = (n-1)\int_1^{\infty} x^{1-n}\,\mathrm{d}x = (n-1)\left[\dfrac{x^{2-n}}{2-n}\right]_1^{\infty}$

E(X) does not exist when $\lim\limits_{x\to\infty}\left(\dfrac{x^{2-n}}{2-n}\right)$ is infinite. i.e. when $2-n \geq 0$.

Set of values of n for which E(X) does not exist = $\{n : n \in \mathbb{R}, n \leq 2\}$

Note that *n* is not an integer!

Question 7

7 The long run average number of goals per game in a particular football league is 2.639.

(i) State the assumptions necessary for a Poisson distribution to provide a good model for the rate at which goals are scored.

Comment on how reasonable these assumptions are in practice. [3]

(ii) Assuming that a Poisson model holds, estimate the probability that a randomly chosen game ends with

(a) a score of 0-0,

(b) 4 or more goals having been scored.

[2]

(iii) In fact, 8.2% of games in this league ended with a score of 0-0, and 30.3% of games ended with 4 or more goals having been scored.

Comment on whether or not this information validates the use of the Poisson model. [1]

Football games last, on average, 95 minutes. (That is the scheduled time of 90 minutes plus an average of 5 minutes added time.)

(iv) Assuming that the Poisson model holds, find the average time between goals and hence state the probability density function for the time between goals. [2]

(v) Find the probability that, in a randomly chosen game,

(a) at least one goal is scored in the first 10 minutes,

(b) no goals are scored in the first 45 minutes.

[2]

(i) Assumptions:

- Goals are scored randomly and independently. (Not the goals are independent of each other!)
- The **mean** rate of goals scored is a constant in a single game. (This means that the mean number of goals scored within a fixed length of time remains constant throughout the game.)

The assumptions are unlikely to hold true in practice as the players' performance will likely be influenced by their earlier performance, especially whether their team has already scored any goal. The mean number of goals scored within any fixed length of time is also unlikely to be a constant throughout the game as the players' physical and mental conditions will vary during the game.

Reasons pertaining to "the scoring of goals was not a random process as it depends on the tactics, players' skills and exact game situations" are not acceptable.

(ii) Let X denote the number of goals scored in a game.

$X \sim \text{Po}(2.639)$

(a) $P(X = 0) = 0.0714$

(b) $P(X \geq 4) = 1 - P(X \leq 3) = 0.273$

(iii) Given: $P(X = 0) = 0.082 \quad P(X \geq 4) = 0.303$

Although the probabilities found in part **(ii)** are quite close to the actually probabilities given, the other outcomes are not considered, so the information is not sufficient to validate the use of the Poisson Model.

(iv) Let T be the time in minutes between any two goals.

Average time between goals = $\dfrac{95}{2.639} = 36.0$

The mean number of goals scored per minute is $\lambda = \dfrac{2.639}{95}$

T follows an exponential distribution with parameter λ.

pdf of T is given by: $f(t) = \begin{cases} \lambda e^{-\lambda t}, & t \geq 0 \\ 0, & \text{otherwise} \end{cases}$, where $\lambda = \dfrac{2.639}{95}$

(v) **(a)** Let Y denote the number of goals scored in the first 10 minutes.

$$Y \sim \mathrm{Po}\left(\frac{2.639}{95} \times 10\right) \text{ i.e. } \mathrm{Po}\left(\frac{26.39}{95}\right)$$

$P(Y \geq 1) = 1 - P(Y = 0) = 0.243$

OR Using Exponential Distribution:

$P(T \leq 10) = \int_0^{10} \lambda e^{-\lambda t}\,dt = 0.243$, where $\lambda = \dfrac{2.639}{95}$

(b) Let W denote the number of goals scored in the first 45 minutes.

$$W \sim \mathrm{Po}\left(\frac{2.639}{95} \times 45\right) \text{ i.e. } \mathrm{Po}\left(\frac{23.751}{19}\right)$$

$P(W = 0) = 0.286$

OR Using Exponential Distribution:

$$\begin{aligned} P(T > 45) &= 1 - P(T \leq 45) \\ &= 1 - \int_0^{45} \lambda e^{-\lambda t}\,dt, \ \lambda = \frac{2.639}{95} \\ &= 0.286 \end{aligned}$$

Question 8

8 A random sample of adults were asked to rate their level of income and their level of happiness, using a scale of 'low', 'medium' and 'high' for each. The percentages of people in the various categories were as follows.

		Level of happiness		
		Low	Medium	High
Level of income	Low	3%	11.5%	8%
	Medium	5.5%	23.5%	14.5%
	High	8.5%	19%	6.5%

(i) Given that the sample size was 200, carry out a chi-squared test at the 5% level of significance to investigate the independence of the two factors, level of income and level of happiness. Show in your working the contribution to the test statistic in each cell of the table. [7]

(ii) Suppose now that the sample size had been 400. State the new value of the test statistic and explain whether or not the conclusion of the test would be changed. [2]

(iii) A researcher believes that the figures shown in the table are broadly accurate for the population as a whole, and she wishes to have very strong evidence that the two factors are not independent. Assuming she is correct, estimate the sample size required to show, at the 0.1% level of significance, that the two factors are not independent. [3]

(i) sample size $n = 200$

		Level of happiness			
		Low	Medium	High	**Total**
Level of income	Low	6	23	16	45
	Medium	11	47	29	87
	High	17	38	13	68
	Total	34	108	58	200

H_0: Level of income and level of happiness are independent
H_1: Level of income and level of happiness are associated

Check: Do not write

H_0: Level of income is independent of level of happiness

H_1: Level of income is dependent of level of happiness

Level of significance: 5%
Based on H_0, the expected frequencies are computed and shown in brackets below:

		Level of happiness			
		Low	Medium	High	**Total**
Level of income	Low	7.65	24.3	13.05	45
	Medium	14.79	46.98	25.23	87
	High	11.56	36.72	19.72	68
	Total	34	108	58	200

Degree of freedom = (3 – 1) (3 – 1) = 4

Test statistic: $\sum \frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}} \sim \chi_4^2$

The contributions to the test statistic are shown in the table below:

		Level of happiness		
		Low	Medium	High
Level of income	Low	$\frac{(6-7.65)^2}{7.65} = 0.356$	0.0695	0.667
	Medium	0.971	8.51×10^{-6}	0.563
	High	2.56	0.0446	2.29

From GC, $\chi^2_{\text{cal}} = 7.521$, p-value = 0.11077 > 0.05

Since the p-value > level of significance, we do not reject H_0

Hence there is insufficient evidence at the 5% level of significance to conclude that level of income and level of happiness are associated.

(ii) If $n = 400$,

(All entries in the observed frequency table and expected frequency table are doubled)

New $\chi^2_{\text{cal}} = 15.04$, p-value = $P\left(\chi^2 \geq 15.04\right) = 0.00461 < 0.05$

Since the p-value < level of significance, we reject H_0.

The conclusion of the test would be changed.

(iii) Let $n = 200k$. Then $\chi^2_{\text{cal}} = \sum \frac{\left(kO_{ij} - kE_{ij}\right)^2}{kE_{ij}} = 7.521k$.

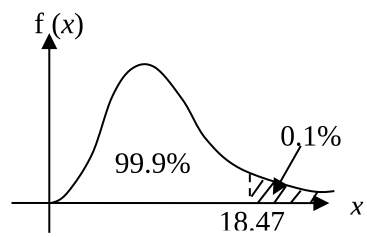
From MF26, $P(\chi_4^2 \geq 18.47) = 0.001$

For H_0 to be rejected at 0.01% level of significance,

$$7.521k \geq 18.47$$

$$200k \geq 491.2$$

Least n = 492



Question 9

9 A wildlife conservationist investigated the masses of wolves in two regions, A and B, in Mexico. In region A, he took a sample of adult male wolves and found their masses, in kg, to be as follows.

39.1 34.1 41.4 31.8 35.9 36.4 30.9 32.3 33.6 29.1

10 values
σ^2 is unknown

(i) Construct a 95% confidence interval for the mean mass of adult male wolves in region A. State the assumptions necessary for your method to be valid. [4]

The conservationist wished to investigate whether or not the mean masses of adult male wolves differed in regions A and B. He found the masses of a sample of adult male wolves in region B. These data are summarised as follows.

$$n = 20 \qquad \Sigma x = 607.7 \qquad \Sigma x^2 = 18\,749$$

(ii) Carry out an appropriate hypothesis test with a 5% significance level. State the conclusion the conservationist should reach, and state the assumptions necessary for the method to be valid. [8]

(i) Let Y denote the masse of a randomly chosen adult male wolf in region A and μ_Y be the population mean of Y.

From GC, $\bar{y} = 34.46$ $\quad s_y = 3.7939$

Assume that the sample is a random sample and the mass of adult wolves from region A follows a normal distribution.

$$\frac{\bar{Y} - \mu_Y}{S_y / \sqrt{n}} \sim t_9$$

95% confidence limits for μ_Y are $\bar{y} \pm t_9 \left(\frac{s_y}{\sqrt{n}} \right) = 34.46 \pm 2.2622 \left(\frac{3.793}{\sqrt{10}} \right)$

A 95% confidence interval for μ_Y is (31.75, 37.17).

(ii) Let X denote the masse of a randomly chosen adult male wolf in region $B$and μ_X be the population mean of X.

$$\bar{x} = \frac{\sum x}{n} = \frac{607.7}{20} = 30.385$$

$$s_x^2 = \frac{1}{n-1}\left[\sum x^2 - \frac{\left(\sum x\right)^2}{n} \right] = \frac{1}{19}\left(18749 - \frac{607.7^2}{20} \right) = 14.949$$

Both sample sizes $n_x = 20$, $n_y = 10$ are small.

To carry out a two-sample t-test, we need to assume that

(1) the samples are random samples

(2) X and Y follow two independent normal distribution with a common population variance.

H_0 : $\mu_X = \mu_Y$

H_1 : $\mu_X \neq \mu_Y$

Level of significance: 5%

Test statistic: $\dfrac{(\overline{X}-\overline{Y})-(\mu_X-\mu_Y)}{s_p\sqrt{\dfrac{1}{n_x}+\dfrac{1}{n_y}}} \sim t_{n_x+n_y-2} = t_{28}$

$$s_p^2 = \frac{(n_x-1)s_x^2+(n_y-1)s_y^2}{n_x+n_y-2} = 3.8432^2$$

Based on H_0, $t_{cal} = -2.738$

p-value $= 0.0106 < 0.05$

Since p-value < level of significance, we reject H_o.
Hence there is sufficient evidence at the 5% level of significance to conclude that the mean masses of adult male wolves differed in regions A and B.

Question 10

10 An educational psychologist is working with children who have a specific learning difficulty that affects their short-term memory. The psychologist administers a test of short-term memory when she first meets each child. She administers a similar test after the children have been taught strategies designed to improve short-term memory. She wishes to investigate whether or not the strategies appear to be effective.

The test results for a random sample of 12 of these children are shown in the table. Higher scores indicate better short-term memory.

Child	A	B	C	D	E	F	G	H	I	J	K	L
First test	192	270	204	177	134	291	276	295	196	177	181	255
Second test	251	272	146	343	195	131	355	341	280	167	170	335

(i) Explain why it would not be appropriate, in this case, to carry out a t-test on the data. [2]

(ii) Carry out a suitable Wilcoxon matched-pairs signed rank test for these data, using a 1% significance level. State any assumption(s) required for the Wilcoxon test to be valid. [6]

(i) We do not have information about the population distributions of test scores before and after the children are taught using the strategies. To carry out a t-test, we need the distribution of the difference in the test scores to be normally distributed. Hence it would not be appropriate in this case to carry out a t-test on the data.

(ii) Let X and Y denote the test scores before and after the children are taught using the strategies respectively.

To carry out a Wilcoxon matched-pairs signed rank test:

Let $D = Y - X$ and m_d = population median of D

Assumptions: D has a continuous and symmetric distribution.

Child	A	B	C	D	E	F	G	H	I	J	K	L
1^{st} test x_i	192	270	204	177	134	291	276	295	196	177	181	255
2^{nd} test y_i	251	272	146	343	195	131	355	341	280	167	170	335
$d_i = y_i - x_i$	+59	+2	–58	+166	+61	–160	+79	+46	+84	–10	–11	80
Ranks	6	1	5	12	7	11	8	4	10	2	3	9

$H_0 : m_d = 0$.

$H_1 : m_d > 0$
Level of significance: 1%

P = sum of the ranks corresponding to the positive differences

$= 6 + 1 + 12 + 7 + 8 + 4 + 10 + 9 = 57$

Q = sum of the ranks corresponding to the negative differences

$= 5 + 11 + 2 + 3 = 21$

$T_{\text{cal}} = Q = 21$

Now, $n = 12$, for a 1-tailed test at 1% level of significance, critical region = $\{t: t \le 9\}$
Since T_{cal} does not fall in the critical region, we do not reject H_0.

Hence, there is insufficient evidence at the 1% level of significance to conclude that the strategies are effective in improving short-term memory.