

NANYANG JUNIOR COLLEGE

JC2 Preliminary Examination

Higher 2

FURTHER MATHEMATICS

9649/02

Paper 2

18 September 2026

3 hours

Additional Materials: Printed Answer Booklet
List of Formulae and Results (MF27)

READ THESE INSTRUCTIONS FIRST

Answer **all** the questions.

Write your answers on the Printed Answer Booklet. Follow the instructions on the front cover of the answer booklet.

Give non-exact numerical answers correct to 3 significant figures, or 1 decimal place in the case of angles in degrees, unless a different level of accuracy is specified in the question.

You are expected to use an approved graphing calculator.

Unsupported answers from a graphing calculator are allowed unless a question specifically states otherwise.

Where unsupported answers from a graphing calculator are **not** allowed in a question, you must present the mathematical steps using mathematical notations and not calculator commands.

You must show all necessary working clearly.

The number of marks is given in brackets [] at the end of each question or part question.

This document consists of **6** printed pages and **2** blank pages.



NANYANG JUNIOR COLLEGE
Internal Examinations

Section A: Pure Mathematics [50 marks]

- 1 The linear transformation $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$ is defined by

$$T: \begin{pmatrix} x \\ y \\ z \end{pmatrix} \mapsto \mathbf{M} \begin{pmatrix} x \\ y \\ z \end{pmatrix},$$

$$\text{where } \mathbf{M} = \begin{pmatrix} 3 & -1 & 1 \\ -1 & 3 & -1 \\ 1 & -1 & 3 \end{pmatrix}.$$

- (a) Find the dimension of the null space of T . [2]

- (b) It is given that

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \mathbf{v}_3 = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}.$$

Determine with a reason, whether

$$T(\mathbf{v}_1), T(\mathbf{v}_2), T(\mathbf{v}_3)$$

form a basis for $\mathbb{R}^3$. [2]

- (c) For matrix $\mathbf{M}$, state the eigenvalue corresponding to the eigenvector $\mathbf{u} = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$, and determine an eigenvector corresponding to another eigenvalue. [3]

- 2 Anand and Beng set up fish farms on two identical, adjacent plots of land such that the fish populations in both farms may be modelled by the same differential equation

$$\frac{dP}{dt} = 2P \left(1 - \frac{P}{25} \right) - H,$$

where P is the population of fish (in thousands) and H is a constant (in thousands) representing the change in fish population due to harvesting. One unit of t represents one harvest season.

Anand started his farm with 4000 fish, but the fish population in his farm died out after several years.

Beng started his farm with 16500 fish, and the fish population in his farm grew and stabilised at 20000 after many years.

- (a) Find the number of fish harvested each season. [2]
- (b) On the same graph, illustrate how the populations of fish in Anand's and Beng's farm vary over time. [2]
- (c) Find the smallest number of fish Beng could have started his farm with to reach the same population size after many years. [2]

- 3 Amy wishes to find the root of the equation

$$x^5 - x - 1.0093 = 0 \quad (*)$$

First, she tries using a Newton-Raphson iteration method with $x_0 = 0$.

- (a) Find the values of x_1, x_2, x_3 and x_4 , and explain whether Amy will be able to find a solution for the equation. [3]
- (b) Suggest what Amy should have done before choosing an appropriate initial value for her iterative process. [1]
- (c) Without using a graphing calculator, sketch two suitable curves to show that (*) has exactly one real positive root, justifying your answer. [1]

Next, Amy decides to find the root of (*) using a fixed-point iteration formula

$$x_{n+1} = \frac{x_n + 1.0093}{x_n^4}$$

with $x_0 = 1$.

- (d) Sketch a graph showing how x_1 and x_2 are obtained using this method, and explain why Amy will not be able to find a solution to this equation. [2]
- (e) Rewrite the equation (*) in a suitable form and hence use the iterative formula $x_{n+1} = g(x_n)$ to find the solution of (*) correct to 3 significant figures using $x_0 = 1$. [3]

- 4 The polar equations of two curves are given by

$$C_1 : r = 1 + \sin \theta \text{ and } C_2 : r = 1 + \cos 2\theta \text{ for } -\pi \leq \theta \leq \pi.$$

- (a) Find the exact polar coordinates of all point(s) of intersection of C_1 and C_2 . [3]
- (b) Find the exact area of the region of C_1 that lies inside C_2 . [4]

Let P be the point of intersection of C_1 and C_2 in the first quadrant. A straight line parallel to the initial line passes through P and intersects C_2 at another point, Q , in the first quadrant.

- (c) Show that at Q , $\sin \theta = \frac{\sqrt{13} - 1}{4}$. [4]
- (d) The region bounded by the line segments OP , OQ and the arc of the curve C_2 between P and Q forms a polar sector. Find the total perimeter of this sector, giving your answer correct to 4 decimal places. [2]

- 5 A mathematical model is given by the differential equation

$$\frac{dy}{dx} = \frac{e^{-y} - 2x}{1 + x^2}$$

where $y = 0$ when $x = 1$.

- (a) Using a step size of 0.5, obtain an approximation for y when $x = 2$ using
- (i) Euler's method, [2]
 - (ii) the Improved Euler's method. [3]
- Show all your steps clearly.
- (b) It is later found that the actual value of y is -0.5108256 , correct to 7 significant figures, when $x = 2$. Calculate the percentage error of the estimates obtained by both methods. [2]
- (c) Give a reason why the Improved Euler's method estimate gives a larger error than the Euler's method estimate in this case. [1]
- (d) Suggest a scenario when the Euler method and Improved Euler method may give estimates that are very close to each other. [1]
- (e) Using the substitution $y = \ln u$, find y in terms of x and sketch the solution curve for $0 \leq x \leq 3$. [5]

Section B: Probability and Statistics [50 marks]

- 6 Under usual operating conditions, a help desk receives on average one call every 2.5 minutes. It may be assumed that calls are received at random and that the probability of two or more calls being received simultaneously is negligible.
- (a) State the two further assumptions needed for the number of calls received in a fixed period of time to be modelled by a Poisson distribution. [2]
- Assume now that a Poisson model is valid, and that usual operating conditions apply.
- (b) Find the probability that no call is received in a period of 5 minutes. [1]
- Let T minutes denote the waiting time until the next call is received.
- (c) State fully the distribution of T . [1]
- (d) Find the minimum value of t , correct to 1 decimal place, in order to be 99% certain that the next call is received within t minutes. [2]
- (e) A help desk operator has already waited 5 minutes without receiving a call. Find the expected additional waiting time until the next call is received. [2]

- 7 Benford's Law states that, in many tables containing large numbers of numerical values, the probability distribution of the leading non-zero digit D is given by

$$P(D = d) = \log_{10} \left(\frac{d+1}{d} \right), \quad d = 1, 2, \dots, 9.$$

Fabricated financial data often deviate from Benford's Law and auditors use it as an initial analytical procedure to identify datasets that may require closer examination. An auditor takes a random sample of 200 invoice amounts from a branch that is being reviewed. The data are as follows.

Leading digit	1	2	3	4	5	≥ 6
Observed frequency	52	34	22	16	38	38

- (a) Show that the theoretical probability for the pooled category $D \geq 6$ is $\log_{10} \left(\frac{5}{3} \right) \approx 0.22185$. [1]
- (b) Carry out an appropriate test to determine if there is very strong evidence that the leading non-zero digit distribution does not follow Benford's Law. [5]
- (c) Calculate the percentage of the total test statistic represented by the contribution of the cell corresponding to $D = 5$. Discuss why the result of the test in part (b) may not be sufficient to conclude that the invoice amounts from the branch is fabricated. [3]
- 8 The continuous random variable X has the distribution $U(0, n)$, where n is an integer larger than 8.

- (a) A success is defined as the event ' $\left| X - \frac{n}{2} \right| < 2$ '.

Independent values of X are generated until the first success occurs. Given that the probability that the third value generated is the first success is 0.128, find the value of n . [5]

A success is now defined as the event ' $X(n - X) > \frac{n^2}{16}$ '.

Independent values of X are generated until the first success occurs. The random variable Y denotes the number of values generated.

- (b) State the distribution of Y , showing clearly that the parameter of the distribution is $\frac{\sqrt{3}}{2}$. [4]
- (c) Find the variance of the number of failures before the first success. [2]

- 9 A medical researcher studies whether a sleep-improvement programme increases the average number of hours slept each night. The average hours of sleep for nine randomly selected participants are recorded before and after the programme.

Participant	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	<i>F</i>	<i>G</i>	<i>H</i>	<i>I</i>
Before	6.2	5.8	6.5	5.9	6.1	5.7	6.4	6.0	5.9
After	6.9	6.3	6.2	6.8	6.7	5.9	7.2	6.4	6.9

- (a) Use an appropriate t -test to examine whether, overall, the sleep improvement programme has increased the average number of sleeping hours, at the 1% significance level. [5]
- (b) Provide an alternative analysis using Wilcoxon signed-rank test, at the 1% significance level. [4]
- (c) What distributional assumption is required in part (a) but not in part (b)? [1]
- (d) Another researcher claims: "Since both tests are carried out on the same data, they must always lead to the same conclusion." Comment on this statement. [2]
- 10 An autonomous drone uses two independent radar sensors, A and B, to measure its distance (in metres) to a landing platform. The distance readings from both sensors are normally distributed with an unknown common population mean μ m. Sensor A's readings have a known standard deviation of 0.4 m while Sensor B's readings have a known standard deviation of 0.6 m.

A random sample of 16 independent readings is taken from Sensor A, yielding a sample mean of 24.20 m. A random sample of 9 independent readings is taken from Sensor B, yielding a sample mean of 24.35 m.

- (a) Find a 95% confidence interval for μ using only the data from Sensor A, giving the end-points of the interval correct to 2 decimal places. [2]

An engineer proposes combining the data streams using a combined sample mean estimator:

$$\bar{X}_C = \alpha \bar{X}_A + (1 - \alpha) \bar{X}_B,$$

where $\bar{X}_A$ and $\bar{X}_B$ are sample mean readings from sensors A and B respectively, and α is a constant such that $0 < \alpha < 1$.

- (b) Show that $\bar{X}_C$ is an unbiased estimator for μ , and find the value of α that minimizes the variance of this combined estimator. [4]
- (c) Using your value of α from part (b), calculate a 95% confidence interval for μ based on the combined data from both sensors. [3]
- (d) Without further computation, state, with a clear statistical reason, whether the confidence interval found in part (c) is narrower than the 95% confidence interval for μ obtained using only data from sensor B. [1]

BLANK PAGE

BLANK PAGE

2026 Preliminary Examination
Further Mathematics Paper 2
Solutions and Marking Scheme

1 The linear transformation $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$ is defined by

$$T: \begin{pmatrix} x \\ y \\ z \end{pmatrix} \mapsto \mathbf{M} \begin{pmatrix} x \\ y \\ z \end{pmatrix},$$

where $\mathbf{M} = \begin{pmatrix} 3 & -1 & 1 \\ -1 & 3 & -1 \\ 1 & -1 & 3 \end{pmatrix}$.

(a) Find the dimension of the null space of T .

[2]

(b) Given that

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \mathbf{v}_3 = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}.$$

Determine with a reason, whether

$$T(\mathbf{v}_1), T(\mathbf{v}_2), T(\mathbf{v}_3)$$

form a basis for $\mathbb{R}^3$.

[2]

(c) For matrix $\mathbf{M}$, state the eigenvalue corresponding to the eigen vector $\mathbf{u} = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$, and

determine an eigenvector corresponding to another eigenvalue.

[3]

By GC, $\begin{pmatrix} 3 & -1 & 1 \\ -1 & 3 & -1 \\ 1 & -1 & 3 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & -1 & 3 \\ 0 & 1 & 1 \\ 0 & 0 & -10 \end{pmatrix}$ Rank($\mathbf{M}$) = 3 Dim(K) = dim($\mathbb{R}^3$) – rank($\mathbf{M}$) = 3–3=0	
Since $\mathbf{v}_1 - \mathbf{v}_2 - \mathbf{v}_3 = 0$ $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ are linearly dependent. By linearity of T, $T(\mathbf{v}_1) + T(\mathbf{v}_2) + T(\mathbf{v}_3) = 0$ $T(\mathbf{v}_1), T(\mathbf{v}_2), T(\mathbf{v}_3)$ are linearly dependent and cannot span $\mathbb{R}^3$. $T(\mathbf{v}_1), T(\mathbf{v}_2), T(\mathbf{v}_3)$ do not form a basis for $\mathbb{R}^3$.	

$\mathbf{Mu} = \lambda \mathbf{u}$ $\begin{pmatrix} 3 & -1 & 1 \\ -1 & 3 & -1 \\ 1 & -1 & 3 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \\ -2 \end{pmatrix} = 2 \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$ The corresponding eigen value is 2 From GC, other eigenvalue is 5. Alternative: Solving for another eigen value, $ \mathbf{M} - \lambda \mathbf{I} = 0, \text{ i.e. } \begin{vmatrix} 3-\lambda & -1 & 1 \\ -1 & 3-\lambda & -1 \\ 1 & -1 & 3-\lambda \end{vmatrix} = 0$ $(2 - \lambda)^2 (5 - \lambda) = 0$ $\lambda = 2 \text{ or } \lambda = 5$ For eigenvector, we solve $(\mathbf{M} - 5\mathbf{I})\mathbf{w} = 0$ $\begin{pmatrix} -2 & -1 & 1 \\ -1 & -2 & -1 \\ 1 & -1 & -2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$ $\begin{aligned} -2x - y + z &= 0 \\ -x - 2y - z &= 0 \\ x - y - 2z &= 0 \end{aligned}$ Eigenvector is $\begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}$.	
--	--

- 2 Anand and Beng set up fish farms on two identical, adjacent plots of land such that the fish populations in both farms may be modelled by the same differential equation

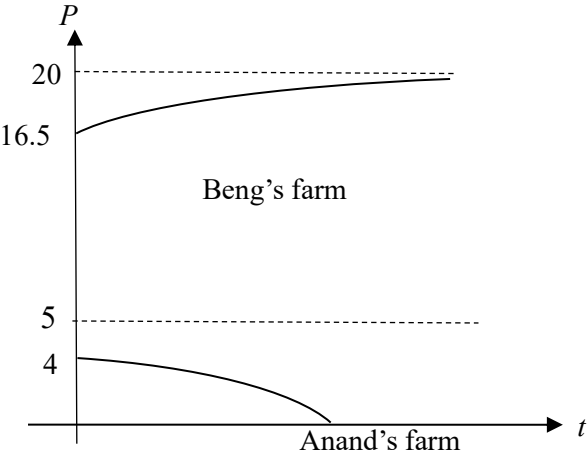
$$\frac{dP}{dt} = 2P \left(1 - \frac{P}{25} \right) - H,$$

where P is the population of fish (in thousands) and H is a constant (in thousands) representing the change in fish population due to harvesting. One unit of t represents one harvest season.

Anand started his farm with 4000 fish, but the fish population in his farm died out after several years.

Beng started his farm with 16500 fish, and the fish population in his farm grew and stabilised at 20000 after many years.

- (a) Find the number of fish harvested each season. [2]
- (b) On the same graph, illustrate how the populations of fish in Anand's and Beng's farm vary over time. [2]
- (c) Find the smallest number of fish Beng could have started his farm with to reach the same population size after many years. [2]

2(a)	Given that $\frac{dP}{dt} = 0$ when $P = 20$, $0 = 2(20)\left(1 - \frac{20}{25}\right) - H \Rightarrow H = 8$. Hence 8000 fish are harvested each season.
(b)	
(c)	$\frac{dP}{dt} = 2P\left(1 - \frac{P}{25}\right) - 8 \stackrel{\text{set}}{=} 0 \Rightarrow P = 5 \text{ or } 20$. If Beng starts with 5001 (least positive integer > 5000) fish, the model predicts that the fish population will increase to 20000 in the long run.

3 Amy wishes to find the root of the equation

$$x^5 - x - 1.0093 = 0 \quad \text{-----} (*)$$

First, she tries using a Newton-Raphson iteration method with $x_0 = 0$.

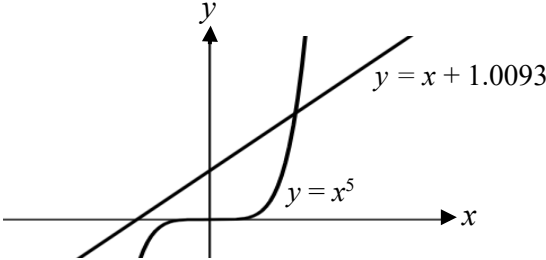
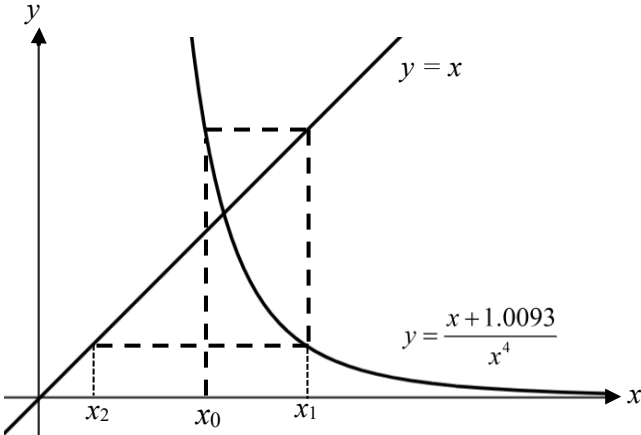
- (a) Find the values of x_1, x_2, x_3 and x_4 , and explain whether Amy will be able to find a solution for the equation. [3]
- (b) Suggest what Amy should have done before choosing an appropriate initial value for her iterative process. [1]
- (c) Without using a graphing calculator, sketch two suitable curves to show that (*) has exactly one real positive root, justifying your answer. [1]

Next, Amy decides to find the root of (*) using a fixed-point iteration formula

$$x_{n+1} = \frac{x_n + 1.0093}{x_n^4}$$

with $x_0 = 1$.

- (d) Sketch a graph showing how x_1 and x_2 are obtained using this method, and explain why Amy will not be able to find a solution to this equation. [2]
- (e) Rewrite the equation (*) in a suitable form and hence use the iterative formula $x_{n+1} = g(x_n)$ to find the solution of (*) correct to 3 significant figures using $x_0 = 1$. [3]

(a)	Let $f(x) = x^5 - x - 1.0093$. Then $f'(x) = 5x^4 - 1$. Hence the Newton-Raphson iteration is given by $x_{n+1} = x_n - \frac{x_n^5 - x_n - 1.0093}{5x_n^4 - 1}$. From GC, $x_0 = 1, x_1 = -1.0093, x_2 = -0.75925, x_3 = 0.00015346, x_4 = -1.0093$. As the iterations repeat themselves with a period of 3, Amy will not be able to find a solution.
(b)	Amy should have sketched suitable graphs to determine the rough location of the root and choose an initial value nearer to the root that is also away from turning points.
(c)	Sketch $y = x^5$ and $y = x + 1.0093$:  From the sketch it is clear that the graphs intersect exactly once at $x > 0$; therefore the equation (*) has exactly one real positive root.
(d)	 Based on the sketch, successive iterations will get further away from the intersection, so Amy will not be able to find a solution.
(e)	$x^5 - x - 1.0093 = 0 \Rightarrow x = \sqrt[5]{x + 1.0093}$ From GC, $x_0 = 1, x_1 = 1.1498, x_2 = 1.1664 \approx 1.17, x_3 = 1.16821 \approx 1.17$ Check: $f(1.165) = -0.0283 < 0$ and $f(1.175) = 0.0554 > 0$, hence the root is 1.17 (3sf)

4 The polar equations of two curves are given by

$$C_1 : r = 1 + \sin \theta \text{ and } C_2 : r = 1 + \cos 2\theta \text{ for } -\pi \leq \theta \leq \pi.$$

(a) Find the exact polar coordinates of all point(s) of intersection of C_1 and C_2 .

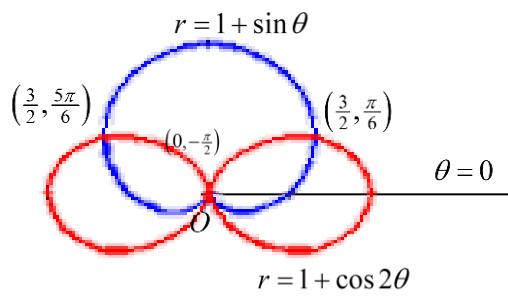
[3]

- (b) Find the exact area of the region of C_1 that lies inside C_2 . [4]

Let P be the point of intersection of C_1 and C_2 in the first quadrant. A straight line parallel to the initial line passes through P and intersects C_2 at another point, Q , in the first quadrant.

- (c) Show that at Q , $\sin \theta = \frac{\sqrt{13}-1}{4}$. [4]

- (d) The region bounded by the line segments OP , OQ and the arc of the curve C_2 between P and Q forms a polar sector. Find the total perimeter of this sector, giving your answer correct to 4 decimal places. [2]

4(a)	$1 + \sin \theta = 1 + \cos 2\theta$ $\sin \theta = \cos 2\theta$ $\sin \theta = 1 - 2\sin^2 \theta$ $2\sin^2 \theta + \sin \theta - 1 = 0$ $(2\sin \theta - 1)(\sin \theta + 1) = 0$ $\sin \theta = \frac{1}{2} \quad \text{or} \quad \sin \theta = -1$ $\theta = \frac{\pi}{6}, \frac{5\pi}{6} \quad \theta = -\frac{\pi}{2}$ $r = 1 + \sin \frac{\pi}{6} = 1 + \sin \frac{5\pi}{6} = \frac{3}{2}$ $r = 1 + \sin \left(-\frac{\pi}{2}\right) = 0$ Polar coordinates of points of intersection are $\left(\frac{3}{2}, \frac{\pi}{6}\right), \left(\frac{3}{2}, \frac{5\pi}{6}\right), \left(0, -\frac{\pi}{2}\right),$	
(b)	 Exact area	

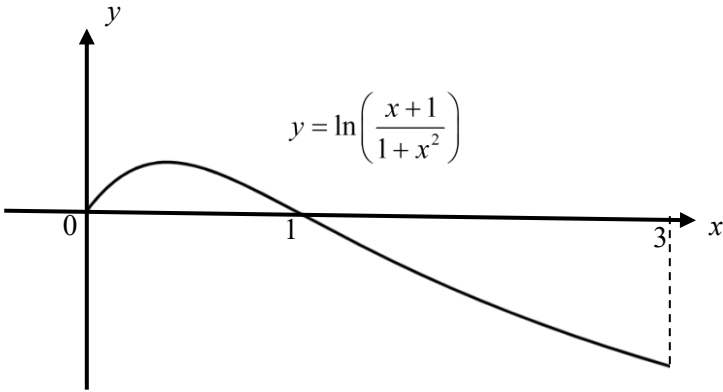
	$ \begin{aligned} &= 2 \times \left[\frac{1}{2} \int_{-\frac{\pi}{2}}^{\frac{\pi}{6}} (1 + \sin \theta)^2 d\theta + \frac{1}{2} \int_{\frac{\pi}{6}}^{\frac{\pi}{2}} (1 + \cos 2\theta)^2 d\theta \right] \\ &= \int_{-\frac{\pi}{2}}^{\frac{\pi}{6}} 1 + 2\sin \theta + \sin^2 \theta d\theta + \int_{\frac{\pi}{6}}^{\frac{\pi}{2}} 1 + 2\cos 2\theta + \cos^2 2\theta d\theta \\ &= \int_{-\frac{\pi}{2}}^{\frac{\pi}{6}} \frac{3}{2} + 2\sin \theta - \frac{\cos 2\theta}{2} d\theta + \int_{\frac{\pi}{6}}^{\frac{\pi}{2}} \frac{3}{2} + 2\cos 2\theta + \frac{\cos 4\theta}{2} d\theta \\ &= \left[\frac{3}{2}\theta - 2\cos \theta - \frac{\sin 2\theta}{4} \right]_{-\frac{\pi}{2}}^{\frac{\pi}{6}} + \left[\frac{3}{2}\theta + \sin 2\theta + \frac{\sin 4\theta}{8} \right]_{\frac{\pi}{6}}^{\frac{\pi}{2}} \\ &= \left[\left(\frac{3}{2} \left(\frac{\pi}{6} \right) - 2 \left(\frac{\sqrt{3}}{2} \right) - \frac{\sqrt{3}}{8} \right) - \left(\frac{3}{2} \left(-\frac{\pi}{2} \right) - 0 - 0 \right) \right] \\ &\quad + \left[\left(\frac{3}{2} \left(\frac{\pi}{2} \right) + 0 + 0 \right) - \left(\frac{3}{2} \left(\frac{\pi}{6} \right) + \frac{\sqrt{3}}{2} + \frac{\sqrt{3}}{16} \right) \right] \\ &= \left(\pi - \frac{9\sqrt{3}}{8} \right) + \left(\frac{\pi}{2} - \frac{9\sqrt{3}}{16} \right) \\ &= \frac{3\pi}{2} - \frac{27}{16}\sqrt{3} \end{aligned} $	
(c)	At point $P \left(\frac{3}{2}, \frac{\pi}{6} \right)$, $y = r \sin \theta = \left[1 + \cos 2 \left(\frac{\pi}{6} \right) \right] \sin \frac{\pi}{6} = \frac{3}{4}$ $y = r \sin \theta = (1 + \cos 2\theta) \sin \theta = \frac{3}{4}$ $(1 + \cos 2\theta) \sin \theta = \frac{3}{4}$ $(2 - 2\sin^2 \theta) \sin \theta = \frac{3}{4}$ $4(2 - 2\sin^2 \theta) \sin \theta = 3$ $8\sin^3 \theta - 8\sin \theta + 3 = 0$ $(2\sin \theta - 1)(4\sin^2 \theta + 2\sin \theta - 3) = 0$ $4\sin^2 \theta + 2\sin \theta - 3 = 0$ $\sin \theta = \frac{-2 \pm \sqrt{2^2 - 4(4)(-3)}}{2(4)} = \frac{-2 \pm 2\sqrt{13}}{8} = \frac{-1 \pm \sqrt{13}}{4}$ Since Q is in first quadrant, $\sin \theta = \frac{\sqrt{13}-1}{4} > 0$ (shown).	
(d)	Total perimeter of this sector $= 1 + \cos \frac{\pi}{3} + 1 + \cos 2 \left(\sin^{-1} \left(\frac{\sqrt{13}-1}{4} \right) \right)$ $+ \int_{\frac{\pi}{6}}^{\sin^{-1} \left(\frac{\sqrt{13}-1}{4} \right)} \sqrt{(1 + \cos 2\theta)^2 + (-2\sin 2\theta)^2} d\theta$ $= 3.0793 \text{ (4dp)}$	

$$\frac{dy}{dx} = \frac{e^{-y} - 2x}{1 + x^2}$$

where $y = 0$ when $x = 1$.

- (a) Using a step size of 0.5, obtain an approximation for y when $x = 2$ using
- (i) Euler's method, [2]
 - (ii) the Improved Euler's method. [3]
- Show all your steps clearly.
- (b) It is later found that the actual value of y is -0.5108256 , correct to 7 significant figures, when $x = 2$.
Calculate the percentage error of the estimates obtained by both methods. [2]
- (c) Give a reason why the Improved Euler's method estimate gives a larger error than the Euler's method estimate in this case. [1]
- (d) Suggest a scenario when the Euler method and Improved Euler method may give estimates that are very close to each other. [1]
- (e) Using the substitution $y = \ln u$, find y in terms of x and sketch the solution curve for $0 \leq x \leq 3$. [5]

5(a) (i)	Let $f(x, y) = \frac{e^{-y} - 2x}{1 + x^2}$ Using Euler's Method with $x_0 = 1, y_0 = 0$ and $h = 0.5$: $y_1 = y_0 + hf(x_0, y_0) = 0 + \frac{1}{2} \left(\frac{e^{-0} - 2}{1 + 1^2} \right) = -\frac{1}{4} = -0.25$ $y_2 = y_1 + hf(x_1, y_1) = -0.25 + \frac{1}{2} \left(\frac{e^{-0.25} - 3}{1 + 1.5^2} \right) = -0.5139961 = -0.514 \text{ (3sf)}$
(ii)	Using the Improved Euler's Method with $x_0 = 1, y_0 = 0$ and $h = 0.5$: $u_1 = y_0 + hf(x_0, y_0) = -0.25$ $y_1 = y_0 + \frac{1}{2}h[f(x_0, y_0) + f(x_1, u_1)]$ $= 0 + \frac{1}{2} \left(\frac{1}{2} \right) [-0.5 - 0.52799]$ $= -0.2569980 \text{ (7sf)}$ $u_2 = y_1 + hf(x_1, y_1) = -0.2569980 + \frac{1}{2}[-0.5252177] = -0.5196069 \text{ (7sf)}$ $y_2 = y_1 + \frac{1}{2}h[f(x_1, y_1) + f(x_2, u_2)]$ $= -0.2569980 + \frac{1}{2} \left(\frac{1}{2} \right) [-0.5252177 - 0.4637267]$ $= -0.5042341 \text{ (7sf)} = -0.504 \text{ (3sf)}.$

(b)	Percentage error for Euler Method = $\left \frac{-0.5108256 - (-0.5139961)}{-0.5108256} \right \times 100\% = 0.620\%$ Percentage error for Improved Euler Method = $\left \frac{-0.5108256 - (-0.5042341)}{-0.5108256} \right \times 100\% = 1.29\%$
(c)	The gradient of the solution function of the DE may not be consistently increasing or decreasing within the interval studied, and so the Improved Euler's Method may not always give a better estimate. [May also accept answers suggesting "fluctuating gradients" ?]
(d)	If the gradient of the solution function of the DE is relatively constant within the interval studied, both methods of estimation will give estimates that are very close to one another.
(e)	$y = \ln u \Rightarrow \frac{dy}{dx} = \frac{1}{u} \frac{du}{dx}$ $\frac{dy}{dx} = \frac{e^{-y} - 2x}{1 + x^2}$ $\Rightarrow \frac{1}{u} \frac{du}{dx} = \frac{\frac{1}{u} - 2x}{1 + x^2}$ $\Rightarrow \frac{du}{dx} = \frac{1 - 2xu}{1 + x^2}$ Therefore $\frac{du}{dx} + \frac{2x}{1 + x^2} u = \frac{1}{1 + x^2}.$ Integrating Factor = $e^{\int \frac{2x}{1+x^2} dx} = e^{\ln(1+x^2)} = 1 + x^2.$ Hence $u(1 + x^2) = \int 1 dx = x + C$ Substitute $x = 1, u = e^0 = 1$ into above GS gives $C = 1.$ Therefore $u = \frac{x+1}{1+x^2} \Rightarrow y = \ln\left(\frac{x+1}{1+x^2}\right).$ 

- 6 Under usual operating conditions, a help desk receives on average one call every 2.5 minutes. It may be assumed that calls are received at random and that the probability of two or more calls being received simultaneously is negligible.

- (a) State the two further assumptions needed for the number of calls received in a fixed period of time to be modelled by a Poisson distribution. [2]

Assume now that a Poisson model is valid, and that usual operating conditions apply.

- (b) Find the probability that no call is received in a period of 5 minutes. [1]

Let T minutes denote the waiting time until the next call is received.

- (c) State fully the distribution of T . [1]

- (d) Find the minimum value of t , correct to 1 decimal place, in order to be 99% certain that the next call is received within t minutes. [2]

- (e) A help desk operator has already waited 5 minutes without receiving a call. Find the expected additional waiting time until the next call is received. [2]

	Solution [Poisson & Exponential – 8m]	
6(a)	Calls are received independently of each other. The average rate of calls received remains constant over time.	
(b)	Let C be the number of calls received in a 5-minute period. $C \sim P_o(2)$. Required probability = $P(C = 0) = e^{-2} = 0.135$ (3sf)	
(b)	T follows an exponential distribution with parameter 0.4.	
(c)	Method I: (exponential distribution) Given $P(T \leq t) \geq 0.99$ (or = 0.99) $\int_0^t 0.4e^{-0.4t} dt \geq 0.99$ $1 - e^{-0.4t} \geq 0.99$ $t \geq -\frac{1}{0.4} \ln 0.01 = 11.5129$ Method II (Poisson Distribution) Let C_t be the number of calls received in t minutes. $C_t \sim P_o(0.4t)$ Given $P(C_t \geq 1) \geq 0.99$, $P(C_t = 0) \leq 0.01$ $e^{-0.4t} \leq 0.01$ $t \geq -\frac{1}{0.4} \ln 0.01 = 11.5129$ $\therefore$ minimum value of t is 11.6 (to 1 dp.)	
(d)	By the memoryless property of T ,	

	expected additional waiting time = $E(T) = 2.5$ minutes.	
--	--	--

- 7 Benford's Law states that, in many tables containing large numbers of numerical values, the probability distribution of the leading non-zero digit D is given by

$$P(D = d) = \log_{10} \left(\frac{d+1}{d} \right), \quad d = 1, 2, \dots, 9.$$

Fabricated financial data often deviate from Benford's Law and auditors use it as an initial analytical procedure to identify datasets that may require closer examination. An auditor takes a random sample of 200 invoice amounts from a branch that is being reviewed. The data are as follows.

Leading digit	1	2	3	4	5	≥ 6
Observed frequency	52	34	22	16	38	38

- (a) Show that the theoretical probability for the pooled category $D \geq 6$ is $\log_{10} \left(\frac{5}{3} \right) \approx 0.22185$. [1]
- (b) Carry out an appropriate test to determine if there is very strong evidence that the leading non-zero digit distribution does not follow Benford's Law. [5]
- (c) Calculate the percentage of the total test statistic represented by the contribution of the cell corresponding to $D = 5$. Discuss why the result of the test in part (b) may not be insufficient to conclude that the invoice amounts from the branch is fabricated. [3]

7(a)	$P(D \geq 6) = \log_{10}\left(\frac{7}{6}\right) + \log_{10}\left(\frac{8}{7}\right) + \log_{10}\left(\frac{9}{8}\right) + \log_{10}\left(\frac{10}{9}\right)$ $= \log_{10}\left(\frac{7}{6} \times \frac{8}{7} \times \frac{9}{8} \times \frac{10}{9}\right)$ $= \log_{10}\left(\frac{5}{3}\right) \approx 0.22185 \text{ (5 s.f.)}$																						
(b)	H_0 : The leading non-zero digit distribution follows Benford's Law H_1 : The leading non-zero digit distribution does not follow Benford's Law <table><tr><td>Leading digit</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>≥ 6</td></tr><tr><td>Observed frequency</td><td>52</td><td>34</td><td>22</td><td>16</td><td>38</td><td>38</td></tr><tr><td>Expected frequency</td><td>60.21</td><td>35.22</td><td>24.99</td><td>19.38</td><td>15.84</td><td>44.37</td></tr></table> Degrees of freedom: $6 - 1 = 5$ Test Statistic: $X^2 = \sum_{i=1}^6 \frac{(O_i - E_i)^2}{E_i} \sim \chi^2(5)$ Under H_0 , using GC, $p\text{-value} = 2.3538 \times 10^{-6}$ ($\chi^2_{\text{cal}} = 34.03$)	Leading digit	1	2	3	4	5	≥ 6	Observed frequency	52	34	22	16	38	38	Expected frequency	60.21	35.22	24.99	19.38	15.84	44.37	
Leading digit	1	2	3	4	5	≥ 6																	
Observed frequency	52	34	22	16	38	38																	
Expected frequency	60.21	35.22	24.99	19.38	15.84	44.37																	

	Since p -value is very small, which we will reject H_0 even at a small level of significance. Thus there is very strong evidence that the leading non-zero digit distribution does not follow Benford's Law.	
(c)	$\frac{(38 - 15.84)^2}{34.03}$ Percentage = $\frac{15.84}{34.03} \times 100 = 91.1\%$ The result in (b) only suggests that the distribution differs significantly from that predicted by Benford's Law. Although the largest contribution to the test statistic is the cell corresponding to $D=5$, which is far larger than the rest (still far larger even if the cell $D \geq 6$ is not combined), the branch might have legitimate transactions having similar values which starts with the digit 5, giving rise to much higher observed frequency than expected frequency. Hence further investigation may be required.	

8 The continuous random variable X has the distribution $U(0, n)$, where n is an integer larger than 8.

(a) A success is defined as the event ' $\left|X - \frac{n}{2}\right| < 2$ '.

Independent values of X are generated until the first success occurs. Given that the probability that the third value generated is the first success is 0.128, find the value of n .

[5]

A success is now defined as the event ' $X(n - X) > \frac{n^2}{16}$ '.

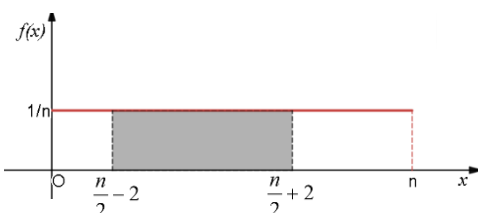
Independent values of X are generated until the first success occurs. The random variable Y denotes the number of values generated.

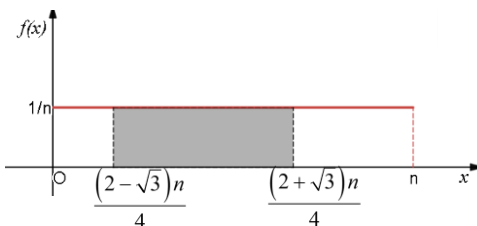
(b) State the distribution of Y , showing clearly that the parameter of the distribution is $\frac{\sqrt{3}}{2}$.

[4]

(c) Find the variance of the number of failures before the first success.

[2]

8(a)	$\left X - \frac{n}{2}\right < 2$ $\Rightarrow \frac{n}{2} - 2 < X < \frac{n}{2} + 2$ $P(\text{success}) = P\left(\frac{n}{2} - 2 < X < \frac{n}{2} + 2\right)$ $= \frac{1}{n} \left[\frac{n}{2} + 2 - \left(\frac{n}{2} - 2\right) \right]$ $= \frac{4}{n}$  $P(3^{\text{rd}} \text{ value generated is the 1st success}) = \left(1 - \frac{4}{n}\right)^2 \frac{4}{n} = 0.128$ Solve $(1 - p)^2 p = 0.128$, where $p = \frac{4}{n}$ $p^3 - 2p^2 + p - 0.128 = 0$	
------	---	--

	By GC, $p = 0.2, 0.48769$ or 1.3123 $\Rightarrow n = \frac{4}{p} = 20$ (the others are non-integer values)	
(b)	Y follows a geometric distribution. $X(n - X) > \frac{n^2}{16} \Rightarrow X^2 - nX + \frac{n^2}{16} < 0$ Consider $X^2 - nX + \frac{n^2}{16} = 0$. $X = \frac{n \pm \sqrt{n^2 - \frac{n^2}{4}}}{2} = \frac{(2 \pm \sqrt{3})n}{4}$ $P(\text{success}) = P\left(\frac{(2 - \sqrt{3})n}{4} < X < \frac{(2 + \sqrt{3})n}{4}\right)$ $= \frac{1}{n} \left[\frac{2\sqrt{3}n}{4} \right]$ $= \frac{\sqrt{3}}{2}$  $\therefore Y \sim \text{Geo}\left(\frac{\sqrt{3}}{2}\right)$.	
(c)	Number of failures before first success, $W = Y - 1$ $\therefore \text{Var}(W) = \text{Var}(Y)$ $= \frac{1-p}{p^2}$ $= \frac{1 - \sqrt{3}/2}{3/4}$ $= \frac{4 - 2\sqrt{3}}{3} [= 0.179 \text{ (3sf)}]$	

- 9 A medical researcher studies whether a sleep-improvement programme increases the average number of hours slept each night.
- The average hours of sleep for nine randomly selected participants are recorded before and after the programme.

Participant	A	B	C	D	E	F	G	H	I
Before	6.2	5.8	6.5	5.9	6.1	5.7	6.4	6.0	5.9
After	6.9	6.3	6.2	6.8	6.7	5.9	7.2	6.4	6.9

- (a) Use an appropriate t -test to examine whether, overall, the sleep improvement programme has increased the average number of sleeping hours, at the 1% significance level. [5]
- (b) Provide an alternative analysis using Wilcoxon signed-rank test, at the 1% significance level. [4]
- (c) What distributional assumption is required in part (a) but not in part (b)? [1]

- (d) Another researcher claims: "Since both tests are carried out on the same data, they must always lead to the same conclusion." Comment on this statement.

[2]

9(a)	Let X and Y be the average number of sleeping hours for the participants before and after the sleep programme respectively. Let $D_i = Y_i - X_i, i = 1, 2, \dots, 9$. $H_0: \mu_D = 0$ $H_1: \mu_D > 0$ where μ_D is the mean of D_i Assumption: $D \sim N(\mu_D, \sigma_D^2)$ Level of significance: 1% From the given data, we have d_i: <table><tr><td>0.7</td><td>0.5</td><td>-0.3</td><td>0.9</td><td>0.6</td><td>0.2</td><td>0.8</td><td>0.4</td><td>1.0</td></tr></table> Thus using GC, $\bar{d} = \frac{8}{15}$ and $s_D^2 = 0.4^2$ Under H_0, Test statistic: $\frac{\bar{D} - 0}{0.4/\sqrt{9}} \sim t_8$ Using GC, $t_{\text{cal}} = \frac{\frac{8}{15} - 0}{0.4/\sqrt{9}} = 4$ $p\text{-value} = 0.00197 < 0.01$ Since $p\text{-value} < \text{level of significance}$, we reject H_0. There is sufficient evidence at 1% level of significance to conclude that the sleep improvement programme has increased the average number of sleeping hours.	0.7	0.5	-0.3	0.9	0.6	0.2	0.8	0.4	1.0												
0.7	0.5	-0.3	0.9	0.6	0.2	0.8	0.4	1.0														
(b)	Let X and Y be the average number of sleeping hours for the participants before and after the sleep programme respectively. $D = Y - X$ $m_d = \text{median of } D$ Let $d_i = y_i - x_i$ for $i = 1, 2, \dots, 9$, then <table><tr><td>d_i</td><td>0.7</td><td>0.5</td><td>-0.3</td><td>0.9</td><td>0.6</td><td>0.2</td><td>0.8</td><td>0.4</td><td>1.0</td></tr><tr><td>Ranks</td><td>6</td><td>4</td><td>2</td><td>8</td><td>5</td><td>1</td><td>7</td><td>3</td><td>9</td></tr></table> $H_0: m_d = 0$ $H_1: m_d > 0$	d_i	0.7	0.5	-0.3	0.9	0.6	0.2	0.8	0.4	1.0	Ranks	6	4	2	8	5	1	7	3	9	
d_i	0.7	0.5	-0.3	0.9	0.6	0.2	0.8	0.4	1.0													
Ranks	6	4	2	8	5	1	7	3	9													

	Level of significance: 1% P = sum of the ranks corresponding to the positive differences $= 6 + 4 + 8 + 5 + 1 + 7 + 3 + 9 = 43$ Q = sum of the ranks corresponding to the negative differences $= 2$ $T_{\text{cal}} = \min\{P, Q\} = Q = 2$ Now, $n = 9$ (no. of pairs of observations) For 1-tailed test at 1% level of significance, critical region = $\{T : T \leq 3\}$ (Refer to MF27) Since $T_{\text{cal}}=2$ falls in the critical region, we reject H_0. There is sufficient evidence at 1% level of significance to conclude that the sleep improvement programme has increased the average number of sleeping hours.	
(c)	For the test in (a), we need to assume that the difference of the number of sleeping hours before and after the programme to follow a normal distribution. For (b), we only need to assume that the difference in number of sleeping hours before and after the programme to follow a continuous and symmetric distribution.	
(d)	This statement is not necessarily true. The paired t-test and the Wilcoxon signed-rank test test similar hypotheses but use different information from the same data The paired t-test uses the <u>actual numerical differences and assumes they are Normally distributed</u> whereas the Wilcoxon signed-rank test uses the <u>ranks of the absolute differences together with their signs and does not require Normality.</u>	

- 10** An autonomous drone uses two independent radar sensors, A and B, to measure its distance (in metres) to a landing platform. The distance readings from both sensors are normally distributed with an unknown common population mean μ m. Sensor A's readings have a known standard deviation of 0.4 m while Sensor B's readings have a known standard deviation of 0.6 m.

A random sample of 16 independent readings is taken from Sensor A, yielding a sample mean of 24.20 m. A random sample of 9 independent readings is taken from Sensor B, yielding a sample mean of 24.35 m.

- (a) Find a 95% confidence interval for μ using only the data from Sensor A, giving the end-points of the interval correct to 2 decimal places. [2]

An engineer proposes combining the data streams using a combined sample mean estimator:

$$\bar{X}_C = \alpha \bar{X}_A + (1 - \alpha) \bar{X}_B,$$

where $\bar{X}_A$ and $\bar{X}_B$ are sample mean readings from sensors A and B respectively, and α is a constant such that $0 < \alpha < 1$.

- (b) Show that $\bar{X}_C$ is an unbiased estimator for μ , and find the value of α that minimizes the variance of this combined estimator. [4]
- (c) Using your value of α from part (b), calculate a 95% confidence interval for μ based on the combined data from both sensors. [3]
- (d) Without further computation, state, with a clear statistical reason, whether the confidence interval found in part (c) is narrower than the 95% confidence interval for μ obtained using only data from sensor B. [1]

10(a)	$\bar{X}_A \sim N\left(\mu, \frac{0.4^2}{16}\right)$, i.e. $\bar{X}_A \sim N(\mu, 0.01)$, $\bar{X}_B \sim N\left(\mu, \frac{0.6^2}{9}\right)$, i.e. $\bar{X}_B \sim N(\mu, 0.04)$ A 95% confidence interval for μ is $\left(\bar{x}_A - 1.9600\sqrt{0.01}, \bar{x}_A + 1.9600\sqrt{0.01}\right)$, where $\bar{x}_A = 24.20$ $= (24.004, 24.396)$ $= (24.00, 24.40)$ (to 2 dp.)	
(b)	$\bar{X}_C$ is an unbiased estimator of $\mu \Leftrightarrow E(\bar{X}_C) = \mu$. LHS = $E(\bar{X}_C)$ $= \alpha E(\bar{X}_A) + (1 - \alpha)E(\bar{X}_B)$ $= \alpha\mu + (1 - \alpha)\mu$ ($\because \bar{X}_A$ and $\bar{X}_B$ are unbiased estimators of μ) $= \mu$ $= \text{RHS (shown)}$ $\text{Var}(\bar{X}_C) = \alpha^2 \text{Var}(\bar{X}_A) + (1 - \alpha)^2 \text{Var}(\bar{X}_B)$ ($\because \bar{X}_A$ and $\bar{X}_B$ are independent) $= 0.01\alpha^2 + 0.04(1 - \alpha)^2$ $= 0.05(\alpha - 0.8)^2 + 0.008 \geq 0.008$ $\therefore$ minimum $\text{Var}(\bar{X}_C)$ is 0.008 and occurs when $\alpha = 0.8$.	
(c)	$\bar{X}_C \sim N(\mu, 0.008)$ $\bar{x}_c = 0.8(24.20) + 0.2(24.35) = 24.23$ A 95% confidence interval for μ is $\left(\bar{x}_c - 1.9600\sqrt{0.008}, \bar{x}_c + 1.9600\sqrt{0.008}\right)$ $= (24.0547, 24.40547)$ $= (24.05, 24.41)$ (to 2 dp)	
(d)	Width of a 95% confidence interval is $2(1.9600) \times \text{standard deviation}$. Since $\text{Var}(\bar{X}_C) = 0.008 < \text{Var}(\bar{X}_B) = 0.04$, and width of the 95% confidence interval depends only on the standard deviation, it is hence smaller for the one found using combined data than with only data from sensor B.	