2021 Further Mathematics Paper 2 (9649/2)

Section A: Pure Mathematics [50 marks]

1 **(a)** By considering $\int_1^{e} \frac{1}{x}\,dx$, use Simpson's rule for two strips to show that e is approximately a root of

$$x^3 + 3x^2 - 15x - 1 = 0. \qquad [5]$$

(b) Hence find an approximation to e, giving your answer to 4 decimal places. [1]

(a) Let $f(x) = \frac{1}{x}$ Using Simpson's rule, $\int_1^{e} \frac{1}{x}\,dx \approx \frac{e-1}{6}\left[f(1) + 4f\left(\frac{e+1}{2}\right) + f(1)\right]$ $[\ln x]_1^{e} \approx \frac{e-1}{6}\left[1 + \frac{4}{\frac{e+1}{2}} + \frac{1}{e}\right]$ $1 \approx \frac{e-1}{6}\left[1 + \frac{8}{e+1} + \frac{1}{e}\right]$ $1 \approx \frac{e-1}{6}\left[\frac{e(e+1) + 8e + (e+1)}{e(e+1)}\right]$ $6e(e+1) \approx (e^3 - e) + 8e^2 - 8e + (e^2 - 1)$ $e^3 + 3e^2 - 15e - 1 \approx 0$ Hence e is approximately a root of $x^3 + 3x^2 - 15x - 1 = 0$. (shown)
(b) $x^3 + 3x^2 - 15x - 1 = 0$ Using GC, $x = -0.06582, x = -5.6319, x = 2.6977$ Since e > 0, e $\approx$ 2.6977 (to 4 d.p)

2 S is the set of complex numbers z for which $|z-6-7\mathrm{i}| \leqslant 5$ and $\arg(z) \leqslant \frac{1}{4}\pi$.

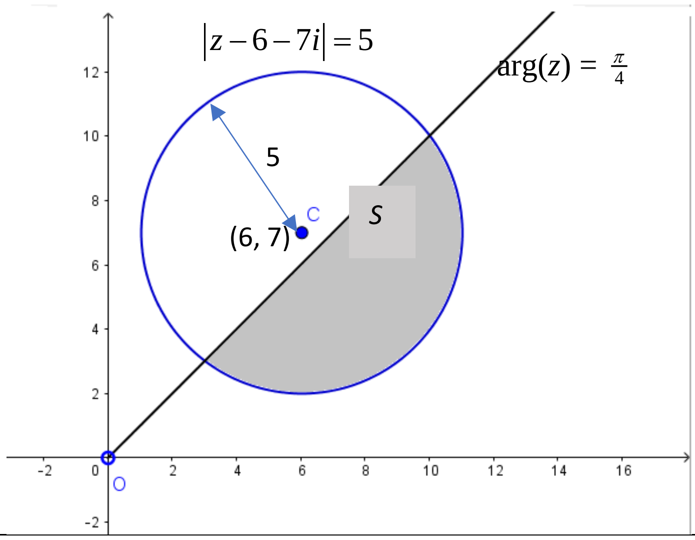
(a) On a single Argand diagram, shade the region corresponding to S. [3]

(b) The complex number z_0 is the element of S which has the smallest argument. Determine the modulus and argument of z_0, giving the argument to 3 significant figures. [5]

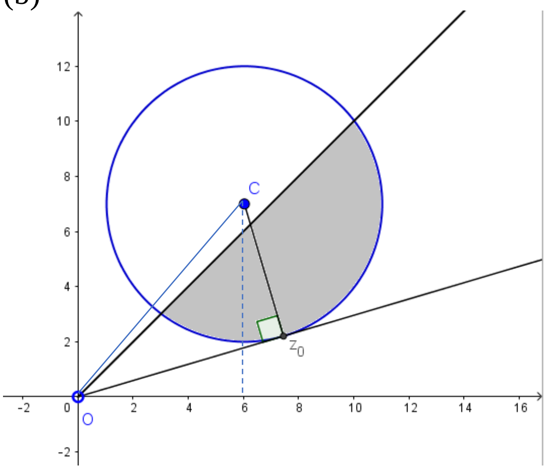
[Solution]

(a)

Note that the centre of circle lies above the half-line.



(b)



Let Q represents z_o

In the right-angle triangle, OCQ, where C is the centre of the circle

$$|z_o| = OQ = \sqrt{OC^2 - 5^2} = \sqrt{6^2 + 7^2 - 25} = \sqrt{60}$$

$$= 2\sqrt{15}$$

In $\Delta\ OCQ$, $\tan \angle COQ = \dfrac{5}{2\sqrt{15}}$

$\Rightarrow \angle COQ = 0.5732033$

Also $\tan^{-1}(\frac{7}{6}) = 0.862170$

Thus $\arg z_o = 0.862170 - 0.5732033 \approx 0.289$ (correct to 3sf)

Use above geometric method is faster!

3 A curve is given parametrically by $x = t\cos t - \sin t$, $y = t\sin t + \cos t$ for $0 \leqslant t \leqslant \frac{3}{4}\pi$. When the curve is rotated through 2π radians about the x-axis, a surface of revolution is formed with area A.

Determine the exact value of A. [10]

$$x = t\cos t - \sin t \Rightarrow \frac{\mathrm{d}x}{\mathrm{d}t} = \cos t - t\sin t - \cos t = -t\sin t$$

$$y = t\sin t + \cos t \Rightarrow \frac{\mathrm{d}y}{\mathrm{d}t} = \sin t + t\cos t - \sin t = t\cos t$$

$$\sqrt{\left(\frac{\mathrm{d}x}{\mathrm{d}t}\right)^2 + \left(\frac{\mathrm{d}y}{\mathrm{d}t}\right)^2} = \sqrt{t^2\sin^2 t + t^2\cos^2 t}$$

$$= \sqrt{t^2} = t \ \text{(since } 0 \le t \le \frac{3}{4}\pi\text{)}$$

$$A = 2\pi\int_0^{\frac{3}{4}\pi} (t\sin t + \cos t)t\ \mathrm{d}t$$

$$= 2\pi\int_0^{\frac{3}{4}\pi} \left(t^2\sin t + t\cos t\right) \mathrm{d}t$$

$$= 2\pi\left\{\left[t^2(-\cos t)\right]_0^{\frac{3}{4}\pi} + \int_0^{\frac{3}{4}\pi} 2t\cos t\ \mathrm{d}t + \left[t(\sin t)\right]_0^{\frac{3}{4}\pi} - \int_0^{\frac{3}{4}\pi} \sin t\ \mathrm{d}t\right\}$$

$$= 2\pi\left\{\frac{9\pi^2}{16}\left(\frac{\sqrt{2}}{2}\right) + \left[2t\sin t\right]_0^{\frac{3}{4}\pi} - \int_0^{\frac{3}{4}\pi} 2\sin t\ \mathrm{d}t + \left[\frac{3\pi}{4}\left(\frac{\sqrt{2}}{2}\right)\right] + \left[\cos t\right]_0^{\frac{3}{4}\pi}\right\}$$

$$= 2\pi\left\{\frac{9\sqrt{2}\pi^2}{32} + \frac{3\sqrt{2}\pi}{4} + \left[2\cos t\right]_0^{\frac{3}{4}\pi} + \frac{3\sqrt{2}\pi}{8} + \left(-\frac{\sqrt{2}}{2} - 1\right)\right\}$$

$$= 2\pi\left[\frac{9\sqrt{2}\pi^2}{32} + \frac{9\sqrt{2}\pi}{8} + \left(-\sqrt{2} - 2\right) - \frac{\sqrt{2}}{2} - 1\right]$$

$$= 2\pi\left[\frac{9\sqrt{2}\pi^2}{32} + \frac{9\sqrt{2}\pi}{8} - \frac{3\sqrt{2}}{2} - 3\right]$$

$$= \pi\left(\frac{9\sqrt{2}\pi^2}{16} + \frac{9\sqrt{2}\pi}{4} - 3\sqrt{2} - 6\right) \text{units}^2$$

4 Let $\mathbf{A} = \begin{pmatrix} 1 & 3 & 1 & 2 \\ 2 & 5 & 1 & 5 \\ 1 & 2 & 7 & -10 \\ 5 & 3 & 0 & 9 \end{pmatrix}$.

(a) Define the rank of a matrix and, showing all necessary working, use row operations to determine the rank of **A**. [6]

(b) Explain why the four row vectors of **A** do not form a basis for the space of four-dimensional row vectors. [1]

(c) Express the row vector (5 3 0 9) as a linear combination of the first three row vectors of **A**. [4]

[Solution]

(a) Rank of a matrix is the dimension of the column space represented by the matrix.

Wrong: number of non-zero rows in reduced form of the matrix; dimension of the matrix

$$\begin{pmatrix} 1 & 3 & 1 & 2 \\ 2 & 5 & 1 & 5 \\ 1 & 2 & 7 & -10 \\ 5 & 3 & 0 & 9 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 3 & 1 & 2 \\ 0 & -1 & -1 & 1 \\ 0 & -1 & 6 & -12 \\ 0 & -12 & -5 & -1 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 3 & 1 & 2 \\ 0 & 1 & 1 & -1 \\ 0 & 0 & 7 & -13 \\ 0 & 0 & 7 & -13 \end{pmatrix}$$

$$\longrightarrow \begin{pmatrix} 1 & 3 & 1 & 2 \\ 0 & 1 & 1 & -1 \\ 0 & 0 & 7 & -13 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

There are 3 pivot elements (or 3 non-zero row in the REF form of M), so the rank of A is 3

(b) The four row vectors of A are not linearly independent as indicated by an empty row of the REF form of A.

(c) Let $\begin{pmatrix} 5 \\ 3 \\ 0 \\ 9 \end{pmatrix} = c_1 \begin{pmatrix} 1 \\ 3 \\ 1 \\ 2 \end{pmatrix} + c_2 \begin{pmatrix} 2 \\ 5 \\ 1 \\ 5 \end{pmatrix} + c_3 \begin{pmatrix} 1 \\ 2 \\ 7 \\ -10 \end{pmatrix}$

Solving the equations: $c_1 = -18$, $c_2 = 11$ and $c_3 = 1$

5 A conic section S has a focus at the origin O, directrix $x = -1$ and eccentricity e.

(a) On separate diagrams, sketch S in each of the cases

- $e = \frac{1}{2}$,
- $e = 1$,
- $e = 2$.

Illustrate on each diagram the geometrical significance of the value of e. [5]

(b) It is now given that $e = 2$.

(i) Determine, in the form $Ax^2 + By^2 + Cx + Dy + E = 0$ (where A, B, C, D and E are integers), the cartesian equation of S. [2]

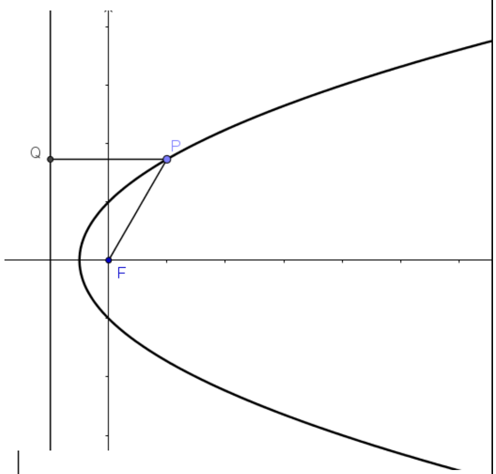
(ii) Write down the cartesian coordinates of O', the second focus of S. [1]

(iii) Let T be a tangent to S. Using the reflective properties of conic sections, or otherwise, prove that the product of the distances from O to T and from O' to T is constant and determine its value. [7]

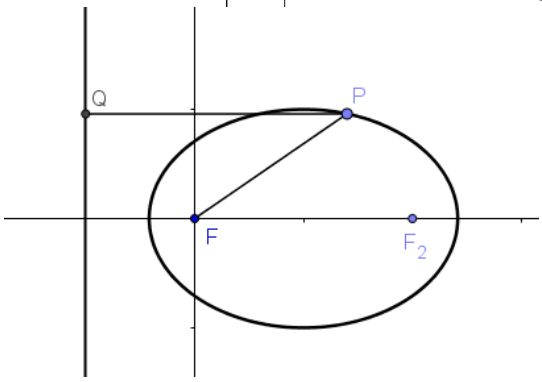
Eccentricity, e, is the ratio between the distance of a point, P, on the conic and a focus, F, to the distance between the point and a directrix d

That is $e = \dfrac{PF}{Pd} = \dfrac{PF}{PQ}$.

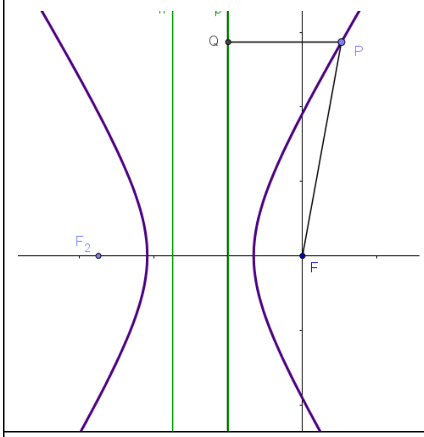
e = 1 , S is a parabola



e = ½, S is an ellipse



e = 2 , S is a hyperbola



-1

(b) (i) e = 2 and F(0, 0) and x = -1

Let P(x, y) any point on the hyperbola. $\dfrac{\sqrt{x^2+y^2}}{\lvert x-(-1)\rvert}=2$ $\Rightarrow$ $x^2 + y^2 = 4(x + 1)^2$ $\Rightarrow x^2 + y^2 = 4x^2 + 8x + 4 \Rightarrow 3x^2 - y^2 + 8x + 4 = 0$
(ii) Equation of conic S is $(x+\frac{4}{3})^2-\dfrac{y^2}{3}=\dfrac{4}{9}\Rightarrow\dfrac{(x+\frac{4}{3})^2}{(\frac{2}{3})^2}-\dfrac{y^2}{\frac{4}{3}}=1$ (hyperbola)   Consider $\dfrac{x^2}{(\frac{2}{3})^2}-\dfrac{y^2}{\left(\frac{2}{\sqrt{3}}\right)^2}=1$ or $3x^2-y^2=\dfrac{4}{3}$ Using $c = ae \Rightarrow c = \frac{2}{3}(2) = \frac{4}{3}$ So the other focus point for conic S is $(-\dfrac{8}{3},0)$
(iii) Method 1 S is a hyperbola with $a=\dfrac{2}{3},b=\dfrac{2}{\sqrt{3}},c=\dfrac{4}{3}$ and $e=2$.   Let P be a point on the hyperbola and line PQ be the tangent to S at P.   WLOG, let $OP > O'P = x$, such that $OP - O'P = 2a = \dfrac{4}{3}$ By the reflective property of hyperbolae, $\angle OPQ=\angle O'PQ$. Denote this angle by θ.   Let $O'P = x$, $OP = x+\dfrac{4}{3}$ and let the distances from O to T and O' to T be d_1 and d_2 respectively.   Then $d_1d_2 = (OP\sin\theta)(O'P\sin\theta) = (x)\left(x+\dfrac{4}{3}\right)\sin^2\theta$ $=\left(x^2+\dfrac{4}{3}x\right)\left(\dfrac{1-\cos 2\theta}{2}\right)=\dfrac{1}{2}\left(x^2+\dfrac{4}{3}x\right)-\dfrac{1}{2}\left(x^2+\dfrac{4}{3}x\right)\cos 2\theta$ --- (*)   Consider triangle OPO': $OO' = 2ae = \dfrac{8}{3}$.   Using the cosine rule, $\left(\dfrac{8}{3}\right)^2=x^2+\left(x+\dfrac{4}{3}\right)^2-2\left(x^2+\dfrac{4}{3}x\right)\cos 2\theta$ $=2x^2+\dfrac{8}{3}x+\dfrac{16}{9}-2\left(x^2+\dfrac{4}{3}x\right)\cos 2\theta$ Hence $\left(\dfrac{8}{3}\right)^2-\dfrac{16}{9}=2\left(x^2+\dfrac{4}{3}x\right)-2\left(x^2+\dfrac{4}{3}x\right)\cos 2\theta$ $\Rightarrow\dfrac{1}{2}\left(x^2+\dfrac{4}{3}x\right)-\dfrac{1}{2}\left(x^2+\dfrac{4}{3}x\right)\cos 2\theta=\dfrac{1}{4}\left(\dfrac{64}{9}-\dfrac{16}{9}\right)=\dfrac{4}{3}$ Substituting into (*), we have $d_1d_2=\dfrac{4}{3}$, a constant value.

Method 2 [Coordinate geometry – straight forward but more tedious]
Consider the equation of tangent to the hyperbola

$3x^2 - y^2 = \frac{4}{3}$ (before S is translated) so its foci are

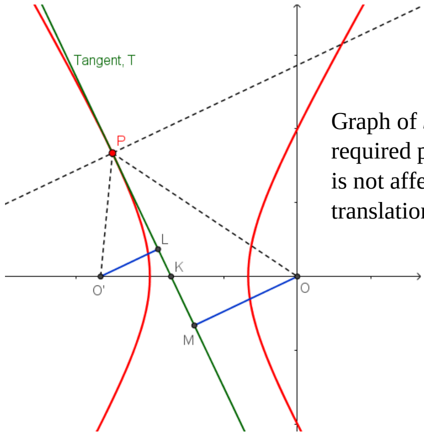
$(\pm\frac{4}{3}, 0)$ and a point $P(x_0, y_0)$ on the hyperbola

$6x - 2y\frac{dy}{dx} = 0 \Rightarrow \frac{dy}{dx} = \frac{3x}{y}$

Equation of tangent to the hyperbola at P is:

$$y - y_o = \frac{3x_0}{y_o}(x - x_o) \Rightarrow y_o(y - y_o) = 3x_0(x - x_o)$$

$$yy_o - 3x_0x + 3x_o^2 - y_0^2 = 0 \quad \text{------ (*)}$$



Graph of S. The required product is not affected by translation

Distance between $O(\frac{4}{3}, 0)$ to the tangent line (*) is :

OM (refer to the diagram) =

$$\frac{|0 - 3x_o(\frac{4}{3}) + 3x_o^2 - y_o^2|}{\sqrt{y_0^2 + 9x_o^2}} = \frac{|-4x +_o 3x_o^2 - y_o^2|}{\sqrt{y_0^2 + 9x_o^2}}$$

Since $P(x_0, y_0)$ lies on the hyperbola $3x^2 - y^2 = \frac{4}{3}$ (before S is translated),

$3x_o^2 - y_o^2 = \frac{4}{3}$, thus $OM = \dfrac{|0 - 3x_o(\frac{4}{3}) + 3x_o^2 - y_o^2|}{\sqrt{y_0^2 + 9x_o^2}} = \dfrac{|\frac{4}{3} - 4x_o|}{\sqrt{y_0^2 + 9x_o^2}}$

Similarly, distance between $F_2(-\frac{4}{3}, 0)$ and the tangent line (*) is

$O'L$ (refer to the diagram)

$$= \frac{|0 - 3x_o(-\frac{4}{3}) + 3x_o^2 - y_o^2|}{\sqrt{y_0^2 + 9x_o^2}} = \frac{|4x +_o 3x_o^2 - y_o^2|}{\sqrt{y_0^2 + 9x_o^2}} = \frac{|\frac{4}{3} + 4x_o|}{\sqrt{y_0^2 + 9x_o^2}}$$

The required product is not affected by translation along the x-axis, thus

The product is = $OM \times O'L = \dfrac{|\frac{4}{3} - 4x_o|}{\sqrt{y_0^2 + 9x_o^2}} \times \dfrac{|\frac{4}{3} + 4x_o|}{\sqrt{y_0^2 + 9x_o^2}}$

$$= \frac{|\frac{16}{9} - 16x_0^2|}{y_0^2 + 9x_o^2} = \frac{|\frac{16}{9} - 16x_0^2|}{3x_0^2 - \frac{4}{3} + 9x_0^2}$$

$$= \frac{16|\frac{1}{9} - x_0^2|}{12(x_0^2 - \frac{1}{9})} = \frac{4}{3} \text{ is a constant}$$

Section B: Probability and Statistics [50 marks]

6 A treatment intended to reduce high levels of anxiety in adults is being trialled. A random sample of 9 adults with high levels of anxiety is selected. Each adult in the sample is assessed for anxiety levels before and after the treatment. The assessment gives a score between 0 and 100 for anxiety level where a higher score indicates a higher level of anxiety. The scores for the 9 adults are given in the table.

Adult	A	B	C	D	E	F	G	H	I
Anxiety before treatment	82	94	69	69	87	76	92	61	78
Anxiety after treatment	63	84	71	76	74	64	91	65	84

(a) Explain why it might not be appropriate to use a test based on the t-distribution in this situation. [1]

(b) Carry out a suitable Wilcoxon test, using a 5% significance level, to investigate whether or not the treatment appears to be effective. [6]

[Solution]

(a) The probability distribution of the difference in level of anxiety in adults before and after treatment may not follow normal distribution. Hence t-test may not be appropriate.

(b)

Let X = Anxiety level before treatment and Y = Anxiety level after treatment.

$D = X - Y$

Let m_d = median of D

Let $d_i = x_i - y_i$ for $i = 1, 2, ..., 9$, then

Adult	A	B	C	D	E	F	G	H	I
d_i	19	10	-2	-7	13	12	1	-4	-6
Rank of $\lvert d_i \rvert$	9	6	2	5	8	7	1	3	4

$H_0 : m_d = 0$ (no difference after treatment)
$H_1 : m_d > 0$ (treatment is effective)
Level of significance: 5%

P = sum of the ranks corresponding to the positive d_i = 1 + 6 + 7 + 8 + 9 = 31
Q = sum of the ranks corresponding to the negative d_i = 2 + 3 + 4 + 5 = 14
$T_{cal} = \min\{P,Q\} = Q = 14$

Since $n = 9$, at 5% level of significance, critical region = $\{t: t \le 8\}$

Since $T_{cal} = 14 > 8$ falls outside the critical region, we do not reject H_0.

Hence, there is insufficient evidence at 5% level of significance to conclude that the treatment is effective.

7 A machine is used to put rice into paper packets. An empty packet has a mass of 25 grams. When the machine is operating correctly, it fills packets with an average of 1.005 kg of rice.

As part of quality control, a random sample of 10 filled packets is taken and each is weighed to find its total mass, x grams. The values of x are as follows.

1028.2	1027.4	1030.8	1032.0	1030.1	1025.4	1027.4	1030.9	1031.3	1024.3

$$\sum x = 10287.8 \qquad \sum x^2 = 10583945.56$$

(a) Assuming that the masses of rice in the packets are normally distributed, carry out a hypothesis test at the 5% level of significance to check whether the machine is operating correctly. [6]

(b) A member of the quality control team points out that the standard deviation for the mass of rice in a packet is known, from past records, to be 2.5 grams. She also says that it would be useful to have a 95% confidence interval for the mean mass of rice in a packet. Calculate this confidence interval, showing your working. [2]

(a)

Method 1

Let μ be the population mean mass in grams of **a filled packet of rice**.

$H_0 : \mu = 1030$ (1005g + 25g) (machine operating correctly)

$H_1 : \mu \neq 1030$ (machine not operating correctly)

Level of significance: 5%

Using GC, unbiased estimate of population variance, $s^2 = 2.63894^2$

Under H_0, Test Statistic: $\dfrac{\overline{X} - 1030}{\frac{2.63894}{\sqrt{10}}} \sim t_9$

Using GC, p-value = 0.178 > 0.05

Since p-value > level of significance, we do not reject H_0. Hence there is insufficient evidence at 5% level of significance to conclude that the machine is not operating correctly.

Wrong to conclude as "There is evidence that the machine is working correctly"

Method 2 (This method may leads to careless mistakes as you need to find the mass of rice by taking each x value to subtract the mass of empty packet)

Let μ be the population mean mass in grams of **rice in a packet**.

Let Y be the mass of rice in a packet

$H_0 : \mu = 1005$

$H_1 : \mu \neq 1005$

Level of significance: 5%

Using GC, unbiased estimate of population variance of Y, $s^2 = 2.63894^2$

Under H_0, Test Statistic: $\dfrac{\overline{Y} - 1005}{\frac{2.63894}{\sqrt{10}}} \sim t_9$

Using GC, p-value = 0.178 > 0.05

Since p-value > level of significance, we do not reject H_0. Hence there is insufficient evidence at 5% level of significance to conclude that the machine is not operating correctly.

(b) sample mean mass of rice in a packet = $\dfrac{10287.8}{10} - 25 = 1003.78$

Confidence limits = $\bar{y} \pm z\dfrac{\sigma}{\sqrt{n}}$

$$= 1003.78 \pm 1.96\frac{2.5}{\sqrt{10}} = 1003.78 \pm 1.55$$ (Must show this working with 1.96)

95% confidence interval = (1002.23, 1005.33)

(End points rounded off to same number of decimal places as 1.55, i.e. 2 decimal places)

Note:

Above is no longer t distribution since the population standard deviation for mass of rice is given as 2.5g.

This part excludes mass of packet.

8 The random variable X has probability density function (pdf)

$$f(x) = \begin{cases} \dfrac{k}{1+x^2} & \text{for } -a \leqslant x \leqslant a, \\ 0 & \text{otherwise,} \end{cases}$$

where k and a are constants.

(a) Show that $k = \dfrac{1}{2\tan^{-1} a}$. [2]

(b) Sketch the pdf. [1]

(c) Write down the value of $E(X)$ and find $\mathrm{Var}(X)$ in terms of a. [4]

The random variable Y is defined by $Y = |X|$.

(d) Find the median value of Y in the case $a = \sqrt{3}$. [4]

(a) $\displaystyle\int_{-a}^{a} \frac{k}{1+x^2}\,dx = 1$

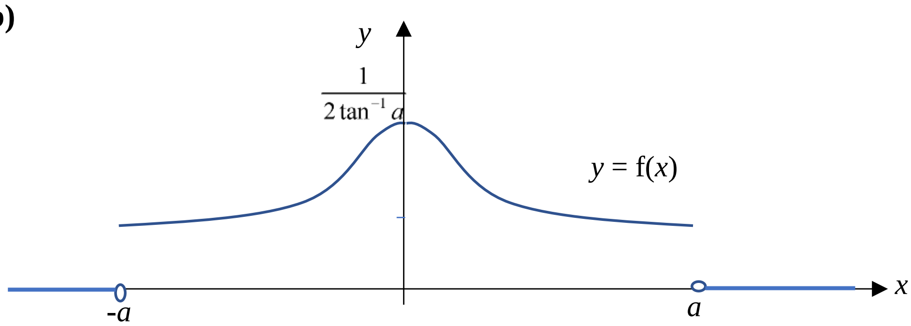
$$k\left[\tan^{-1} x\right]_{-a}^{a} = 1$$

$$k\left[\tan^{-1} a - \tan^{-1}(-a)\right] = 1$$

$$k\left(2\tan^{-1} a\right) = 1$$

$$k = \frac{1}{2\tan^{-1} a}$$ **(shown)**

(b)



(c) Since the pdf is symmetrical about the line $x = 0$, $E(X) = 0$ ("Write down", do not integrate to find $E(X)$)

$$\begin{aligned}
\operatorname{Var}(X) &= E(X^2) - [E(X)]^2 \\
&= \int_{-a}^{a} x^2 \frac{k}{1+x^2}\,dx - 0 \\
&= k\int_{-a}^{a} 1 - \frac{1}{1+x^2}\,dx \\
&= \int_{-a}^{a} k\,dx - \int_{-a}^{a} \frac{k}{1+x^2}\,dx \\
&= [kx]_{-a}^{a} - 1 \\
&= 2ak - 1 \\
&= \frac{a}{\tan^{-1} a} - 1
\end{aligned}$$

(d) $k = \dfrac{1}{2\tan^{-1}\sqrt{3}} = \dfrac{1}{2\left(\frac{\pi}{3}\right)} = \dfrac{3}{2\pi}$

Let m be the median of Y.

$$P(Y \le m) = \frac{1}{2}$$

$$P(|X| \le m) = \frac{1}{2}$$

$$P(-m \le X \le m) = \frac{1}{2}$$

$$\Rightarrow \frac{3}{2\pi}\int_{-m}^{m} \frac{1}{1+x^2}\,dx = \frac{1}{2}$$

$$\left(\frac{3}{2\pi}\right)\left(2\tan^{-1} m\right) = \frac{1}{2}$$

$$\tan^{-1} m = \frac{\pi}{6}$$

$$m = \frac{1}{\sqrt{3}}$$

Do not need to find cdf of Y in this case

Note that:

However if cdf of Y, $G(y)$, is already been found/known, then you can substitute m into $G(y)$ and solve $G(m) = 0.5$

9 A regional health authority is considering a new policy under which all adults over the age of 60 will be invited to attend a health check. In order to assess the likely response rate, the health authority first invited a random sample of adults over the age of 60 to attend. The size of the random sample was 800 and 569 attended.

(a) Find a 95% confidence interval for the proportion of adults over the age of 60 in the region who would attend a health check. Working should be shown. [3]

The health authority then used medical records to categorise those who did and did not attend according to their state of health. The results of this analysis are shown in the table.

State of health	Did attend	Did not attend
Poor	125	83
Average	297	85
Good	147	63

(b) Use a chi-square test of independence to show that there is very strong evidence of an association between state of health and attendance at the health check. [7]

(c) By considering contributions to the test statistic, identify the main issue that the health authority should be concerned about after this analysis. [2]

(a) $p_s = \dfrac{569}{800}$ =0.71125

95% confidence limits = $p_s \pm z\sqrt{\dfrac{p_s(1-p_s)}{n}}$

$$= \frac{569}{800} \pm 1.96\sqrt{\frac{\left(\frac{569}{800}\right)\left(1-\frac{569}{800}\right)}{800}} = \frac{569}{800} \pm 0.0314$$

95% confidence interval = (0.6799, 0.7427)

(b) H_0 : State of heath and attendance at the health check are independent

H_1 : State of heath and attendance at the health check are associated

Under H_0, the expected frequencies, E_{ij} are as follows:

State of health, O_{ij} (E_{ij})	Did attend	Did not attend	**Total**
Poor	125 (147.94)	83 (60.06)	**208**
Average	297 (271.6975)	85 (110.3025)	**382**
Good	147 (149.3625)	63 (60.6375)	**210**
Total	**569**	**231**	**800**

Degree of freedom = $(3-1)(2-1) = 2$

Test Statistic: $\sum \dfrac{(O_{ij}-E_{ij})^2}{E_{ij}} \sim \chi^2_2$

Using GC, $\chi^2_{\text{cal}} = 20.609$, p-value = 3.35×10^{-5}

The small p-value suggests that we need a very small level of significance of less than 0.00335% in order for H_0 not to be rejected. Hence there is very strong evidence of an association between state of health and attendance at the health check.

(c) The contributions to the test statistic are as follows:

State of health, O_{ij} (E_{ij})	Did attend	Did not attend	**Total**
Poor	3.5571	8.762	**208**
Average	2.3564	5.8042	**382**
Good	0.00374	0.092	**210**
Total	**569**	**231**	**800**

The largest contribution comes from those in poor state of health and did not attend the health check where the observed frequency (83) are higher than expected (60.06). The large contribution to the test statistic from this category suggests that more people who are in poor state of heath are not attending health check as expected which is the main issue that the health authority should be concerned.

10 A digital message is made up of binary digits (bits). When a digital message is transmitted from sender to receiver, the 'bit error rate' measures the average proportion of bits of data that have an error when received.

In a particular type of data channel, errors occur randomly with a bit error rate of 2 per million.

(a) State the assumptions needed for the number of errors in this type of data channel to be well modelled by a Poisson distribution. [2]

You are now given that these assumptions hold.

(b) A message of length 1 000 000 bits is transmitted. Calculate the probability that this message is received with no errors. [1]

Messages with exactly one error can be corrected automatically by the receiver. Messages with more than one error have to be re-transmitted.

(c) Find the probability that a message of length 1 000 000 bits does *not* need to be re-transmitted. [1]

(d) Find the probability that a message of length 1 000 000 bits has to be transmitted

(i) exactly 3 times, [2]

(ii) at most 3 times. [2]

(e) Two separate messages, each of length 500 000 bits, are transmitted. Calculate the probability that neither message needs to be re-transmitted. [2]

(f) Find the expected number of times that a message of length 500 000 bits has to be transmitted. [2]

(a) The error in the data channel occurs independently.
Errors must occur at constant average rate or mean rate of occurrence of errors is a constant

Note: It is wrong to say "errors must occur at constant rate";
Must not just give a standard list such as "errors occur randomly, independently and singly". Errors occur randomly is already mentioned in question. As usual in such

questions, "singly" is either part of the independence condition or is automatically satisfied by the scenario. Students must always consider which of the standard conditions do not need to be assumed in a particular context.
(b) Let X be the number of errors received in a message of length 1 million bits. $X \sim \mathrm{P}_0(2)$ $\mathrm{P}(X=0) = 0.13533 \approx 0.135$ (to 3 s.f)
(c) $\mathrm{P}(X \le 1) = 0.406005 \approx 0.406$ (to 3 s.f.)
(d) (i) Let Y be the number of times a message of length 1 million bits needs to be transmitted. $Y\sim\mathrm{Geo}(0.406005)$ $\mathrm{P}(Y=3) = 0.143$ (to 3 s.f.)
(ii) $\mathrm{P}(Y \le 3) = 0.790$ (to 3 s.f.)
(e) Let W be the number of errors received in a message of length 500 000 bits. $W \sim \mathrm{P}_0(1)$ Required probability = $\mathrm{P}(W_1 \le 1) \times \mathrm{P}(W_2 \le 1) = 0.54134 \approx 0.541$
(f) Let T be the number of times a message of length 500 000 bits needs to be transmitted. $T \sim \mathrm{Geo}(\mathrm{P}(W \le 1))$ i.e. $T \sim \mathrm{Geo}(0.73576)$ $\mathrm{E}(T) = \dfrac{1}{0.73576} \approx 1.36$ (to 3 s.f.)