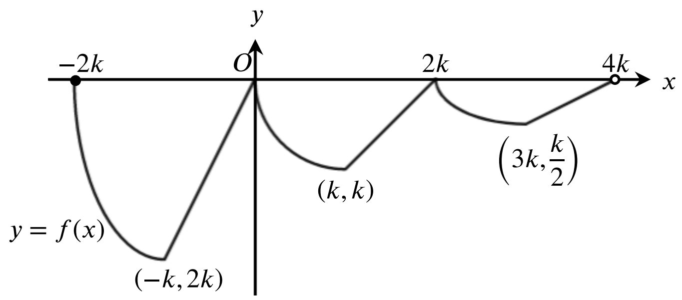
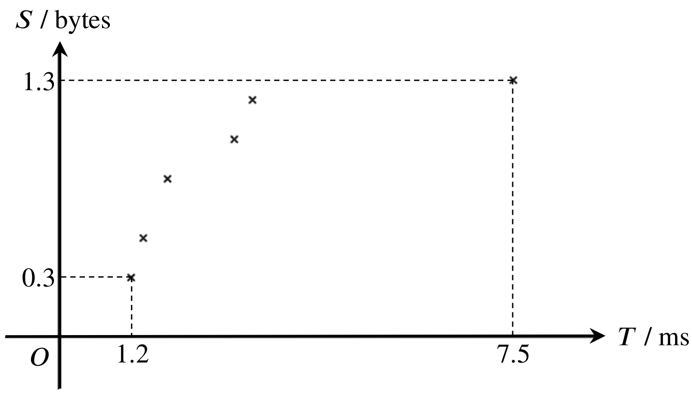
MATHEMATICS

Paper 9758/02
Set II – Paper 2

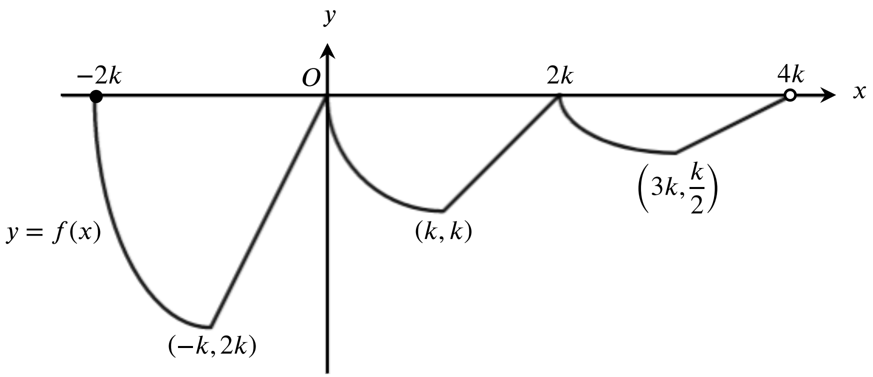
Topic identification and short answers

Qn	Topic(s)	Part	Answers
Section A: Pure Mathematics [40 marks]			
1	Graph (sketching, transformation); Sequences and series (geometric series)	**(i)**	
		(ii)	$\left(\frac{\pi}{4}+\frac{1}{2}\right)k^2$ units^2
		(iii)	$-(\pi+2)k^2$
2	Sequences and series (geometric series, method of differences)		$a_n = 2^n - 1 - \frac{(n-1)n}{2}$
3	Differentiation (maxima and minima)		$\theta = 109.5°$ gives maximum volume. Maximum volume $= \frac{64}{81}r^3$ units^3
4	Complex numbers (cartesian form, conjugate)	**(a)**	[shown]
		(b)	$\gamma = \pm\frac{2}{5}q\text{i}$
5	Vectors (two dimensions, modulus);	**(i)**	[shown]
		(ii)	$\lvert\mathbf{b}-\mathbf{a}\rvert^2 = \lvert\mathbf{a}\rvert^2 + \lvert\mathbf{b}\rvert^2 - 2\lvert\mathbf{a}\rvert\lvert\mathbf{b}\rvert\cos\theta$ $\theta = 72°$
		(iii)	**Regular** pentagon

<table>
<tr><th colspan="4">Section B: Probability and Statistics [60 marks]</th></tr>
<tr><td>6</td><td>Probability (permutations and combinations)</td><td></td><td>Probability $= 1.83 \times 10^{-37}$</td></tr>
<tr><td rowspan="3">7</td><td rowspan="3">Discrete random variables</td><td>(i)</td><td>[shown]</td></tr>
<tr><td>(ii)</td><td>
<table>
<tr><th>x</th><th>$P(X = x)$</th></tr>
<tr><td>3</td><td>$1 - 3p + 3p^2$</td></tr>
<tr><td>4</td><td>$3p - 9p^2 + 12p^3 - 6p^4$</td></tr>
<tr><td>5</td><td>$6p^2 - 12p^3 + 6p^4$</td></tr>
</table>
</td></tr>
<tr><td>(iii)</td><td>$0.354 \le p \le 0.646$</td></tr>
<tr><td rowspan="3">8</td><td rowspan="3">Normal distribution</td><td>(i)</td><td>[shown]</td></tr>
<tr><td>(ii)</td><td>$\sigma \approx 50.01044 \approx 50.0$</td></tr>
<tr><td>(iii)</td><td>$k = \frac{1}{2}$
$m = 325$
Probability $= 1.53 \times 10^{-6}$</td></tr>
<tr><td rowspan="4">9</td><td rowspan="4">Binomial distribution</td><td>(i)</td><td>[shown]</td></tr>
<tr><td>(ii)</td><td>[shown]
Most probable value of $X = 100$</td></tr>
<tr><td>(iii)</td><td>Required probability ≈ 0.0781</td></tr>
<tr><td>(iv)</td><td>$k_{max} = 10$</td></tr>
<tr><td rowspan="4">10</td><td rowspan="4">Sampling; Hypothesis testing</td><td>(i)</td><td>[shown]</td></tr>
<tr><td>(ii)</td><td>$\mathrm{E}(x) = 95.02$
$\mathrm{Var}(x) = 2.5796$</td></tr>
<tr><td>(iii)</td><td>Since n is a multiple of 50, $n > 30$, and thus the sample size is large enough so that, by Central Limit Theorem, the sample mean approximately follows a normal distribution.</td></tr>
<tr><td>(iv)</td><td>$n_{min} = 17450$</td></tr>
</table>

11	Linear regression	**(i)**	A is Quick Sort. B is Merge Sort. C is Selection Sort.
		(ii)	[shown]
		(iii)	Algorithm A, Quick Sort; Best-case $T = -0.248 + \left(2.46 \times 10^{-4}\right) n \log_2 n$ $n \approx 1277$ bytes.
		(iv)	[shown]
		(v)	 Suggestion: $S = 0.368 + 0.540 \ln T$ (Any suitable suggestion acceptable.)

Suggested solutions and post-mortem

Qn	Suggested Solutions	Comments
Section A: Pure Mathematics [40 marks]		
1 (i) [2]		To begin, it must first be understood that the given piecewise function gives rise to vertically scaled sketches of the same graph in different periods. Be mindful of the domain of the sketch and indicate inclusion of the point $(-2k, 0)$ by means of a solid point, and exclusion of the point $(4k, 0)$ by means of a hollow point.
1 (ii) [1]	Required area = Area of a quarter circle + Area of a right triangle $= \frac{1}{4}\pi k^2 + \frac{1}{2}k^2$ $= \left(\frac{\pi}{4} + \frac{1}{2}\right) k^2$ units^2	There may be a tendency to resort to integration to find the required area due to keywords such as finding "the exact area bounded". Considering the mark for this part, candidates may want to resort to simpler ways to find the area required. As far as presentation is concerned, take care to write mensuration values (such as length, area and/or volume) with appropriate units.
1 (iii) [2]	Since $y = \mathrm{f}(x)$ is entirely blow the x–axis, $\int_{-2k}^{\infty} f(x)\,\mathrm{d}x = -\left(\frac{\pi}{4} + \frac{1}{2}\right) k^2 \left(2 + 1 + \frac{1}{2} + \frac{1}{4} + \cdots\right)$ $= -\left(\frac{\pi}{4} + \frac{1}{2}\right) k^2 \left(\frac{2}{1 - \frac{1}{2}}\right)$ $= -\left(\frac{\pi}{4} + \frac{1}{2}\right) k^2(4) = -(\pi + 2)k^2$	Any successful response to this part would recognise from **(i)** that the integrated result must be **negative**, and from **(ii)** that the area in one period is **subsequently half** of that from the previous period, which calls for geometric sum to infinity.

2 **[6]**	$u_{n+2} = 2u_{n+1} - u_n + 2^n - 1$ $\left(u_{n+2} - u_{n+1}\right) - \left(u_{n+1} - u_n\right) = 2^n - 1$ $v_{n+1} - v_n = 2^n - 1$ $\sum_{r=1}^{n}\left[v_{r+1} - v_r\right] = \sum_{r=1}^{n}(2^r - 1)$ $\begin{bmatrix} v_2 & - & v_1 \\ +v_3 & - & v_2 \\ +v_4 & - & v_3 \\ \vdots & & \\ +v_{n-1} & - & v_{n-2} \\ +v_n & - & v_{n-1} \\ +v_{n+1} & - & v_n \end{bmatrix} = 2\left(\frac{2^n - 1}{2 - 1}\right) - n$ $v_{n+1} - v_1 = 2^{n+1} - 2 - n$ $v_n - v_1 = 2^n - 2 - (n - 1)$ $v_n = 2^n - 2 - (n - 1) + v_1$ $u_{n+1} - u_n = 2^n - 2 - n + 1 + (2 - 1)$ $u_{n+1} - u_n = 2^n - n$ $\sum_{r=1}^{n}\left[u_{r+1} - u_r\right] = \sum_{r=1}^{n}(2^r - r)$ $\begin{bmatrix} u_2 & - & u_1 \\ +u_3 & - & u_2 \\ +u_4 & - & u_3 \\ \vdots & & \\ +u_{n-1} & - & u_{n-2} \\ +u_n & - & u_{n-1} \\ +u_{n+1} & - & u_n \end{bmatrix} = 2\left(\frac{2^n - 1}{2 - 1}\right) - \frac{n(n+1)}{2}$ $u_{n+1} - u_1 = 2^{n+1} - 2 - \frac{n(n+1)}{2}$ $u_n - 1 = 2^n - 2 - \frac{(n-1)n}{2}$ $u_n = 2^n - 1 - \frac{(n-1)n}{2}$	This question mainly deals with method of differences. After obtaining an expression for $v_{n+1} - v_n$, method of differences can then be employed to yield an expression for $u_{n+1} - u_n$. At this point, a second method of differences follows to obtain expression for u_n in terms of n. Along the way, appropriate base modification is required, i.e., changing $u_{n+1} \to u_n$.

3 **[8]**	$OF = r\cos(180° - \theta) = -r\cos\theta$ $EF = r\sin(180° - \theta) = r\sin\theta$ $\therefore$ Volume V of the pyramid $= \frac{1}{3}\left(\frac{1}{2}\cdot 2EF \cdot 2EF\right)(OA + OF)$ $= \frac{1}{3}\left(2r^2\sin^2\theta\right)(r - r\cos\theta)$ $= \frac{2r^3}{3}\sin^2\theta\,(1-\cos\theta)$ For maximum volume, $\frac{\mathrm{d}V}{\mathrm{d}\theta} = \frac{2r^3}{3}\left[2\sin\theta\cos\theta\,(1-\cos\theta) + \sin^2\theta\,(\sin\theta)\right] = 0$ $2\sin\theta\cos\theta\,(1-\cos\theta) + \sin^2\theta\,(\sin\theta) = 0$ $\sin\theta\left(2\cos\theta - 2\cos^2\theta + 1 - \cos^2\theta\right) = 0$ $\sin\theta\left(-3\cos^2\theta + 2\cos\theta + 1\right) = 0$ $\sin\theta\,(3\cos\theta + 1)(1-\cos\theta) = 0$ Since $\theta \neq 0°, 180°$, reject $\sin\theta = 0$ and $\cos\theta = 1$ $\therefore \theta_{max} = \cos^{-1}\left(-\frac{1}{3}\right) = 90° + \cos^{-1}\left(\frac{1}{3}\right) \approx 109.5°$ **<u>Using second derivative</u>** When $\theta = \theta_{max}$, $\frac{\mathrm{d}^2V}{\mathrm{d}\theta^2} = \frac{2r^3}{3}\left[\cos\theta\left(-3\cos^2\theta + 2\cos\theta + 1\right) + \sin\theta\,(6\cos\theta\sin\theta - 2\sin\theta)\right]$ $= \frac{2r^3}{3}\left[\cos\theta\left(-3\cos^2\theta + 2\cos\theta + 1\right) + \left(1-\cos^2\theta\right)(6\cos\theta - 2)\right]$ $= \frac{2r^3}{3}\left[\left(-\frac{1}{3}\right)\left(-3\left(\frac{1}{9}\right) + 2\left(-\frac{1}{3}\right) + 1\right) + \left(1 - \left(\frac{1}{9}\right)\right)\left(6\left(-\frac{1}{3}\right) - 2\right)\right]$ $= \frac{2r^3}{3}\left[-\frac{32}{9}\right] = -\frac{64}{27}r^3 < 0$, since r is positive. $\therefore \theta = 109.5°$ gives maximum volume.	When demanded an expression of mensuration values – such as perimeter, area, or volume – in terms of length (in this case, r) and angle (in this case, θ), candidates may expect that trigonometry will be handy. Upon discovering the stationary values of V, there must be some evidence of appreciation towards edge cases, i.e., making use of $\sin\theta \neq 0$ and $\cos\theta \neq 1$, or the corollary that $\theta \neq 0°$ and $\theta \neq 180°$ (which may also be apparent from the range $0° < \theta < 180°$) to give reasoned rejections and arrive at a feasible value. Be reminded to provide non-exact angles in degrees up to one decimal point as appropriate. Current trend suggests that A–Levels tend to relieve candidates of the need to proof the nature of stationary points. Nonetheless, it is good to be familiar with different methods of proving such points – suggestions provided lists down 3 methods and when to best use it, along with common pitfalls to avoid: **<u>Using second derivative for simplified expressions</u>** Given a simplified expression for a first-order derivative, differentiating a second time proves more timesaving than sign test. There must be **evident engagement with unknown constants**, if present, to arrive at the conclusion. This can be achieved by considering the boundary values of these constants, as contextually implied, and thereafter explaining briefly how they affect the value of the second derivative (in this case, the phrase "since r is positive").

[continued]

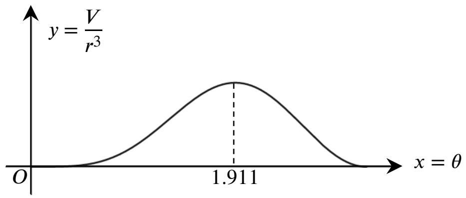
Using sign test

θ	θ_{max}^{-}	θ_{max}	θ_{max}^{+}
$\sin\theta$	–	–	–
$3\cos\theta+1$	–	0	+
$1-\cos\theta$	+	+	+
$\frac{dV}{d\theta}=\frac{2r^3}{3}\sin\theta(3\cos\theta+1)(1-\cos\theta)$	+	0	–
Slope	/	—	\

$\therefore \theta = 109.5°$ gives maximum volume.

Using a graphical method,

$$V=\frac{2r^3}{3}\sin^2\theta(1-\cos\theta) \rightarrow \frac{V}{r^3}=\frac{2}{3}\sin^2\theta(1-\cos\theta)$$



$\frac{V}{r^3}$ is maximum at $\theta \approx 1.911$

$\therefore V$ is maximum at $\theta \approx 1.911 \times \left(\frac{180°}{\pi}\right) \approx 109.5°$

Maximum volume

$$=\frac{2r^3}{3}\left(1-\cos^2\theta\right)(1-\cos\theta)=\frac{2r^3}{3}\left(1-\frac{1}{9}\right)\left(1+\frac{1}{3}\right)=\frac{64}{81}r^3 \text{ units}^3$$

Using sign test for factorised expressions

Given a factorised expression for a first-order derivative, sign test (or first derivative test) will save the hassle of applying tedious chain rules in further differentiation. **To show the nature of the stationary value, a sign test table must appreciate the factors that are significant to the sign change**.

Avoid presenting substandard sign test tables that merely show the signs of the derivative values and the corresponding slopes. There must be some form of appreciation that only specific factors contribute significantly to the sign.

Alternatively, it is acceptable to show derivative values around the stationary point as quoted from a calculator.

Using a graphical method under time constraint

When racing against time, resorting to graphic calculator to prove stationary nature may be acceptable. **Given a contextual question, candidates must recognise any acceptable range(s) of values of the independent variable which is implied by the context** (in this case, $0° < \theta < 180°$). There is risk of mistaking extraneous stationary values outside the implied range as possible answers, which may appear mathematically sensible but contextually inappropriate.

For final presentation, be reminded to append any mensuration results with appropriate units.

4 (a) [5]	**Method 1** Let $z = a + ib, \quad a, b \in \mathbb{R}, \quad b \neq 0$ $\dfrac{z}{1+z^2} = \dfrac{a+ib}{1+(a+ib)^2}$ $= \dfrac{a+ib}{(1+a^2-b^2)+(2ab)i} \times \dfrac{\left(1+a^2-b^2\right)-(2ab)i}{(1+a^2-b^2)-(2ab)i}$ $= \dfrac{a\left(1+a^2-b^2\right)+2ab^2+i\left[b\left(1+a^2-b^2\right)-2a^2b\right]}{(1+a^2-b^2)^2-(2abi)^2}$ $= \dfrac{a\left(1+a^2-b^2\right)+2ab^2}{(1+a^2-b^2)^2+(2ab)^2} + i\left[\dfrac{b\left(1+a^2-b^2\right)-2a^2b}{(1+a^2-b^2)^2+(2ab)^2}\right]$ $\dfrac{z}{1+z^2} \in \mathbb{R}$ $\rightarrow \dfrac{b\left(1+a^2-b^2\right)-2a^2b}{(1+a^2-b^2)^2+(2ab)^2} = 0$ $\rightarrow b\left(1+a^2-b^2-2a^2\right) = 0$ Since $b \neq 0, \ \therefore 1+a^2-b^2-2a^2 = 0$ $\rightarrow a^2+b^2 = 1$ $\rightarrow \left\|\sqrt{a^2+b^2}\right\| = \|z\| = 1$	The most intuitive approach to this question is one that considers an arbitrary Cartesian expression of z, which leads to some juggling of unknowns and conjugate multiplication, before eventually arriving at the result to be shown. Successful responses would not miss the opportunity to make use of the fact that $\mathrm{Im}(z) \neq 0$ in the final deduction.

[continued]

Method 2

Let $z = re^{i\theta}, \quad r > 0, -\pi < \theta \leq \pi$
Since z is not real, $\theta \neq 0$ and $\theta \neq \pi$

$$\frac{z}{1+z^2} = \frac{re^{i\theta}}{1+r^2e^{i2\theta}}$$
$$= \frac{1}{\frac{1}{r}e^{-i\theta} + re^{i\theta}}$$
$$= \frac{1}{\frac{1}{r}(\cos(-\theta) + i\sin(-\theta)) + r(\cos\theta + i\sin\theta)}$$
$$= \frac{1}{\left(\frac{\cos\theta}{r} + r\cos\theta\right) + i\left(-\frac{\sin\theta}{r} + r\sin\theta\right)}$$

$\frac{z}{1+z^2} \in \mathbb{R}$,
$$\rightarrow -\frac{\sin\theta}{r} + r\sin\theta = \left(\frac{r^2-1}{r}\right)\sin\theta = 0$$

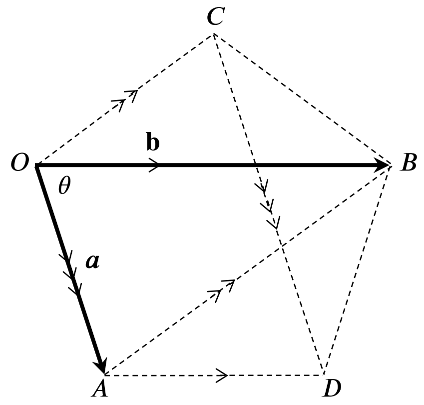
$\theta \neq 0, \theta \neq \pi$ and $r > 0$
$\rightarrow r^2 = 1$
$\rightarrow r = |z| = 1$

Aside from considering an arbitrary Cartesian expression of z, candidates may choose to make do with the more convenient polar form instead.

Unlike the previous approach, in which the imaginary component is in both the nominator and the denominator, the polar form can yield an expression with imaginary part only on one side of the fraction, which eliminates the need for conjugate multiplication. Just like the previous approach, the condition that z is not real is just as relevant for the final deduction, as it implies important conditions on the argument of z.

	[continued] **<u>Method 3</u>** Let $\frac{z}{1+z^2}=m, m\in\mathbb{R}$. $mz^2-z+m=0$ Using the quadratic formula, $z=\frac{1\pm\sqrt{1-4m^2}}{2m}$ Since there are no real solutions for z, $b^2-4ac=1-4m^2<0$ $\therefore z=\frac{1}{2m}\pm i\frac{\sqrt{4m^2-1}}{2m}$ $\therefore \lvert z\rvert=\sqrt{\left(\frac{1}{2m}\right)^2+\left(\frac{\sqrt{4m^2-1}}{2m}\right)^2}=\sqrt{\frac{1+4m^2-1}{4m^2}}=\sqrt{\frac{4m^2}{4m^2}}=1$	A much briefer working would make use of $\frac{z}{1+z^2}$ and express it as some arbitrary real value *m*. By obtaining a quadratic in *z* with no real roots, candidates can make use of the resulting discriminant to directly find *z* in terms of *m* and obtain the modulus value accordingly using a square root. This method is arguably more powerful, especially if candidates are to also find an expression for *z*. Additionally, just like **<u>Method 2</u>**, this working requires no conjugate multiplication, which appeals to time-savers at large.

4 (b) [5]	$(6+p\mathrm{i})\omega^2+(-3m+q\mathrm{i})\omega+2m=\left(6\omega^2-3m\omega+2m\right)+\mathrm{i}\left(p\omega^2+q\omega\right)$ Since values are purely imaginary, $6\omega^2-3m\omega+2m=0$ Since values are distinct, b^2-4ac $=(-3m)^2-4(6)(2m)$ $=9m^2-48m$ $=m(9m-48)>0$ $m<0 \text{ or } m>\dfrac{48}{9}\approx 5.333$ $\therefore$ smallest positive integer $m=6$ $6\omega^2-3(6)\omega+2(6)=0$ $\omega^2-3\omega+2=0$ $(\omega-2)(\omega-1)=0$ $\therefore \omega=2$ or $\omega=1$ $\therefore \gamma=(4p+2q)\mathrm{i}$ or $\gamma=(p+q)\mathrm{i}$ The two γ values are conjugate pairs, $4p+2q=-p-q$ $5p=-3q$ $p=-\dfrac{3}{5}q$ $\therefore \gamma=\pm\dfrac{2}{5}q\mathrm{i}$	Successful responses begin by compartmentalising $\mathrm{Re}(\gamma)$ and $\mathrm{Im}(\gamma)$, and letting $\mathrm{Re}(\gamma)=0$ to yield discriminant restrictions and hence the required value of m thereafter. Since the value of m is not yet known at the start, there is no need to initially assume ω can be a complex number and thus must be expressed arbitrarily in some cartesian form. As seen in the answer, the value of m works out such that the quadratic in ω has real roots. Finding the final pairwise-conjugate values for γ follows.

5 (i) [5]	 From the diagram, $\overrightarrow{OC} \parallel \overrightarrow{AB} \quad \rightarrow \overrightarrow{OC} = s(\mathbf{b}-\mathbf{a}), \quad s \in \mathbb{R}$ $\overrightarrow{AD} \parallel \overrightarrow{OB} \quad \rightarrow \overrightarrow{AD} = t\mathbf{b}, \quad t \in \mathbb{R}$ $\rightarrow \overrightarrow{OD} = t\mathbf{b}-\mathbf{a}$ $\therefore \overrightarrow{CD} = (t\mathbf{b}-\mathbf{a})-(s\mathbf{b}-s\mathbf{a}) = (s+1)\mathbf{a}+(t-s)\mathbf{b}$ $\overrightarrow{CD} \parallel \mathbf{a}$ $\rightarrow t-s=0$ $\rightarrow t=s$ $\therefore \overrightarrow{CD} = (s+1)\mathbf{a}$ $\left\lvert\overrightarrow{CD}\right\rvert = \lvert\mathbf{b}\rvert$ $\rightarrow (s+1)\lvert\mathbf{a}\rvert = \lvert\mathbf{b}\rvert$ $\rightarrow s = \dfrac{\lvert\mathbf{b}\rvert}{\lvert\mathbf{a}\rvert}-1$ Since $s=t$, $\overrightarrow{OC} = \left(\dfrac{\lvert\mathbf{b}\rvert}{\lvert\mathbf{a}\rvert}-1\right)(\mathbf{b}-\mathbf{a})$ and $\overrightarrow{AD} = \left(\dfrac{\lvert\mathbf{b}\rvert}{\lvert\mathbf{a}\rvert}-1\right)\mathbf{b}$.	Diagrams would prove to be a boon when approaching this question. It would be convenient to visually refer to the diagram and deduce using the geometric properties of trapeziums to find $\overrightarrow{OC}$ and $\overrightarrow{AD}$ in terms of **a** and **b**, and subsequently use the given information about $\overrightarrow{CD}$ to arrive at the required results.

<table>
<tr>
<td rowspan="2">5
(ii)
[5]</td>
<td><u>Method 1</u>
$|\mathbf{b}-\mathbf{a}|^2$
$= (\mathbf{b}-\mathbf{a})\cdot(\mathbf{b}-\mathbf{a})$
$= \mathbf{b}\cdot\mathbf{b} - 2\mathbf{a}\cdot\mathbf{b} + \mathbf{a}\cdot\mathbf{a}$
$= |\mathbf{b}|^2 - 2|\mathbf{a}||\mathbf{b}|\cos\theta + |\mathbf{a}|^2$</td>
<td><u>Method 2</u>
Applying cosine rule to triangle OAB,
$AB^2 = OA^2 + OB^2 - 2(OA)(OB)\cos\angle AOB$
$|\mathbf{b}-\mathbf{a}|^2 = |\mathbf{a}|^2 + |\mathbf{b}|^2 - 2|\mathbf{a}||\mathbf{b}|\cos\theta$</td>
<td rowspan="2">Finding an expression for $|\mathbf{b}-\mathbf{a}|^2$ at the start can be done by dot product or by cosine rule. Both will yield the same result, as expected.

With the given condition, two key deductions can be made: $|\mathbf{b}-\mathbf{a}| = |\mathbf{b}|$ and $\left(\frac{|\mathbf{b}|}{|\mathbf{a}|}-1\right)|\mathbf{b}| = |\mathbf{a}|$. These will further lead to a quadric equation in terms of $\frac{|\mathbf{a}|}{|\mathbf{b}|}$, which yields an exact value for $\cos\theta$.

Note that the final angle is exact, and therefore no rounding to one decimal place is required.</td>
</tr>
<tr>
<td colspan="2">Given that $|\overrightarrow{OC}| = |\overrightarrow{AD}|$, $\left(\frac{|\mathbf{b}|}{|\mathbf{a}|}-1\right)|\mathbf{b}-\mathbf{a}| = \left(\frac{|\mathbf{b}|}{|\mathbf{a}|}-1\right)|\mathbf{b}|$

Since $|\mathbf{a}| < |\mathbf{b}|$, $|\mathbf{b}-\mathbf{a}| = |\mathbf{b}|$
$\rightarrow |\mathbf{a}|^2 + |\mathbf{b}|^2 - 2|\mathbf{a}||\mathbf{b}|\cos\theta = |\mathbf{b}|^2$
$\rightarrow |\mathbf{a}|^2 = 2|\mathbf{a}||\mathbf{b}|\cos\theta \rightarrow \cos\theta = \frac{|\mathbf{a}|}{2|\mathbf{b}|}$

Given that $|\overrightarrow{AD}| = |\mathbf{a}|$, $\left(\frac{|\mathbf{b}|}{|\mathbf{a}|}-1\right)|\mathbf{b}| = |\mathbf{a}|$
$\rightarrow \frac{|\mathbf{b}|}{|\mathbf{a}|} - 1 = \frac{|\mathbf{a}|}{|\mathbf{b}|}$
$\rightarrow 1 - \frac{|\mathbf{a}|}{|\mathbf{b}|} = \left(\frac{|\mathbf{a}|}{|\mathbf{b}|}\right)^2$
$\rightarrow \left(\frac{|\mathbf{a}|}{|\mathbf{b}|}\right)^2 + \frac{|\mathbf{a}|}{|\mathbf{b}|} - 1 = 0$
$\rightarrow \frac{|\mathbf{a}|}{|\mathbf{b}|} = \frac{-1 \pm \sqrt{1-4(1)(-1)}}{2(1)} = \frac{-1 \pm \sqrt{5}}{2}$

Since $|\mathbf{a}| < |\mathbf{b}|$, $\frac{|\mathbf{a}|}{|\mathbf{b}|} = \frac{\sqrt{5}-1}{2}$
$\rightarrow \cos\theta = \frac{\sqrt{5}-1}{4} \rightarrow \theta = 72°$</td>
</tr>
<tr>
<td>5
(iii)
[1]</td>
<td colspan="2">$OADBC$ is a regular pentagon.</td>
<td>The keyword “regular” is important. It is insufficient to name the shape merely as pentagon, as it is rather obvious that a five-vertex shape would have five sides.</td>
</tr>
</table>

<table>
<tr><th colspan="3">Section B: Probability and Statistics [60 marks]</th></tr>
<tr>
<td>6
[5]</td>
<td>

<u>Method 1</u>

Each 27 cubes have 27 possible placements and 24 possible orientations
No. of all possible arrangements = $27! \times 24^{27}$

1. <u>8 cubes with 3 painted faces</u>
   Each 8 corner cubes have 8 correct placements and 3 correct orientations
   No. of possible (correct) arrangements = $8! \times 3^8$

2. <u>12 cubes with 2 painted faces</u>
   Each 12 edge cubes have 12 correct placements and 2 correct orientations
   No. of possible (correct) arrangements = $12! \times 2^{12}$

3. <u>6 cubes with 1 painted face</u>
   Each 6 face cubes have 6 correct placements and 4 correct orientations
   No. of possible (correct) arrangements = $6! \times 4^6$

4. <u>1 cube with no painted face</u>
   1 centre cube have 1 correct placement only and 24 correct orientations
   No. of possible (correct) arrangements = 24

∴ Required probability
$$= \frac{(8! \times 3^8)(12! \times 2^{12})(6! \times 4^6)(24)}{27! \times 24^{27}} \approx 1.83 \times 10^{-37}$$

</td>
<td>

This question may prove to be a challenge. Time-pressed candidates would be wise to forego the credits from this question and handle the following questions that are relatively more manageable.

When responding to this question, candidates might find keywords such as “shuffles” and “reorientates” useful: these imply that both the **position** and the **orientation** of each smaller cube are significant when making up the desired cube.

The suggested methods consider the different permutations, combinations, and/or the probability of correctness of each four types of cubes, as per listed.

<u>Method 1</u>

Given that each smaller cubes can be put in 27 possible placements and 24 possible orientations, candidates will need to find the number of cases where each cube satisfies its own condition. The final probability is obtained from dividing the number of correct permutations by the number of all possible permutations.

</td>
</tr>
</table>

	[continued] **<u>Method 2</u>** 1. <u>8 cubes with 3 painted faces</u> No. of ways to place correctly = 8! Probability of each in correct orientation = $\frac{1}{8}$ 2. <u>12 cubes with 2 painted faces</u> No. of ways to place correctly = 12! Probability of each in correct orientation = $\frac{1}{12}$ 3. <u>6 cubes with 1 painted face</u> No. of ways to place correctly = 6! Probability of each in correct orientation = $\frac{1}{6}$ 4. <u>1 cube with no painted face</u> No. of ways to place correctly = 1! Probability of each in correct orientation = 1 No. of ways to arrange into the big cube, without orientation = 27! $\therefore$ Required probability $= \frac{(6!)(8!)(12!)}{27!}\left(\frac{1}{6}\right)^6\left(\frac{1}{8}\right)^8\left(\frac{1}{12}\right)^{12} \approx 1.83\times10^{-37}$	**<u>Method 2</u>** Those who are more likely to stumble in permutations and combinations may instead wish to consider the probability of each cube being in the correct *placement* and *orientation*. To consider the probability of each cube being in the correct *placement*, candidates may have to rely on permutations on the correct slots, which will yield a fraction expressed entirely in factorials. To consider the probability of each cube being in the correct *orientation*, candidates might find it useful to associate the cubes' face(s) relative to a common vertex (for 3 painted faces), edge (for 2 painted faces) or just the face itself, while the unpainted cube at the core will always be correctly orientated.

<table>
<tr>
<td>7
(i)
[2]</td>
<td>Probability of B winning is $1 - p$.

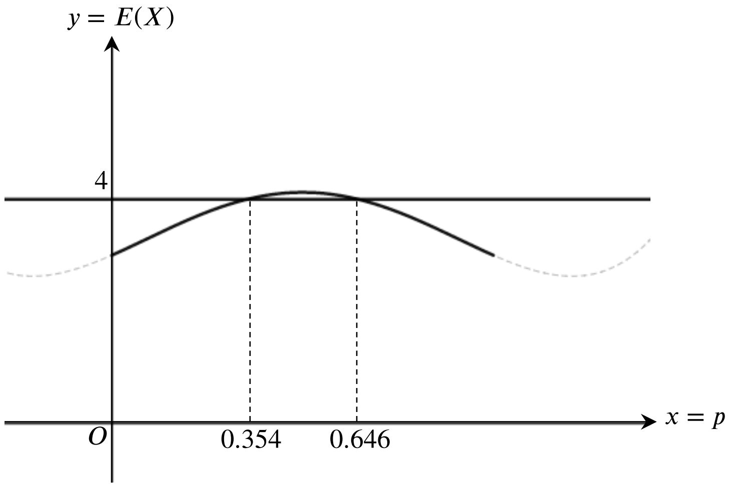
$\therefore \mathrm{P}(X = 3)$
$= \mathrm{P}(A \text{ wins first 3 duels}) + \mathrm{P}(B \text{ wins first 3 duels})$
$= p^3 + (1 - p)^3$
$= p^3 + 1 - 3p + 3p^2 - p^3$
$= 1 - 3p + 3p^2$</td>
<td>In this part, there must be an appreciation that B wins with $1 - p$ chance and for deducing that the required probability is the sum of the probabilities of each player winning 3 duels. Be sure to explicitly write the meaning of these probability.</td>
</tr>
<tr>
<td>7
(ii)
[3]</td>
<td>$\mathrm{P}(X = 4)$
$= \mathrm{P}\left(A \text{ wins in } 4^{th} \text{ duel}\right) + \mathrm{P}\left(B \text{ wins in } 4^{th} \text{ duel}\right)$
$= \left[\left({}^3C_2\right)(p)^2(1 - p)\right] \times p + \left[\left({}^3C_2\right)(1 - p)^2(p)\right] \times (1 - p)$
$= 3p(1 - p)\left[p^2 + (1 - p)^2\right]$
$= \left(3p - 3p^2\right)\left[2p^2 - 2p + 1\right]$
$= 3p - 9p^2 + 12p^3 - 6p^4$

$\mathrm{P}(X = 5)$
$= \mathrm{P}\left(A \text{ wins in } 5^{th} \text{ duel}\right) + \mathrm{P}\left(B \text{ wins in } 5^{th} \text{ duel}\right)$
$= \left[\left({}^4C_2\right)(p)^2(1 - p)^2\right] \times p + \left[\left({}^4C_2\right)(1 - p)^2(p)^2\right] \times (1 - p)$
$= 6p^2(1 - p)^2[p + 1 - p]$
$= 6p^2\left(1 - 2p + p^2\right)$
$= 6p^2 - 12p^3 + 6p^4$

<u>OR</u>

Due to the setup, a round can have $3 \le X \le 5$ duels.
$\therefore \mathrm{P}(X = 5)$
$= 1 - \mathrm{P}(X = 3) - \mathrm{P}(X = 4)$
$= 1 - \left(1 - 3p + 3p^2\right) - \left(3p - 9p^2 + 12p^3 - 6p^4\right)$
$= 6p^2 - 12p^3 + 6p^4$

<table>
<tr><td>x</td><td>3</td><td>4</td><td>5</td></tr>
<tr><td>$\mathrm{P}(X = x)$</td><td>$1 - 3p + 3p^2$</td><td>$3p - 9p^2 + 12p^3 - 6p^4$</td><td>$6p^2 - 12p^3 + 6p^4$</td></tr>
</table>
</td>
<td>To approach this question, consider the different permutations in which the player wins their round. For the latter approach, it could be verified that the sum of probabilities of all possible cases of X is 1.</td>
</tr>
</table>

7 **(iii)** **[3]**	Since none of the rounds ended at the third duel, $4 \le \mathrm{E}(X) \le 5$ **OR** $\mathrm{E}(X) \ge 4$ $\mathrm{E}(X)$ $= 3\left(1 - 3p + 3p^2\right) + 4\left(3p - 9p^2 + 12p^3 - 6p^4\right) + 5\left(6p^2 - 12p^3 + 6p^4\right)$ $= 3 - 9p + 9p^2 + 12p - 36p^2 + 48p^3 - 24p^4 + 30p^2 - 60p^3 + 30p^4$ $= 3 + 3p + 3p^2 - 12p^3 + 6p^4$, Using GC to draw $y = 3 - 3x + 3x^2 - 12x^3 + 6x^4,\ 0 < x < 1$,  $\therefore 0.354 \le p \le 0.646$	With the keyword "in the long run", this part hints at the use of $\mathrm{E}(X)$ to find the required range. It can be noted that, due to the game setup, if the round is expected to never end at the third duel, then the expected number of duels can only be 4 or 5. With an expression for $\mathrm{E}(X)$ in terms of p, finding the range required can be done easily with graphical methods.

8 (i) [2]	Assume $M \sim \text{N}(250, 150^2)$. We have $\text{P}(X \le 0) = 0.0478$. Under a normal distribution with the given parameters, the number of chocolate bars having "negative masses" is not negligible — in 100 bars, around 4 or 5 have a "negative mass" — which is not supposed to happen in this context. $\therefore M \nsim \text{N}(250, 150^2)$.	This question specifically tests on context awareness. There must be an appreciation towards the fact that the mass of a chocolate bar cannot be negative.
8 (ii) [3]	$M \sim \text{N}\left(450, \sigma^2\right)$ $\text{P}(437.33 < M < 514.09) = \text{P}(M < 450) = 0.5$ $2\text{P}(437.33 < M < 462.67) + 3\text{P}(437.33 < M < 462.67)$ $= 2\text{P}(437.33 < M < 462.67) + 2\text{P}(462.67 < M < 514.09)$ $= 2\text{P}(437.33 < M < 514.09)$ $= 2(0.5) = 1$ $\rightarrow (2 + 3)\text{P}(437.33 < M < 462.67) = 1$ $\rightarrow 5\text{P}(437.33 < M < 462.67) = 1$ $\rightarrow \text{P}\left(-\frac{12.67}{\sigma} < Z < \frac{12.67}{\sigma}\right) = \frac{1}{5}$ Using GC (invNorm), $\frac{12.67}{\sigma} \approx 0.2533471101$ $\rightarrow \sigma \approx 50.01044 \approx 50.0$	This question calls for the manipulation of the ranges, standardisation of the distribution and using GC (invNorm) to deduce the final value of the standard deviation σ, each individually awarded to make up the total credit. Despite the question not asking for a show, be advised to display all relevant steps to prevent any divergence due to careless mistakes, or to otherwise avoid losing marks by leaving out credit-worthy steps.

8 (iii) [5]	Let $kM + m \sim \mathrm{N}\left(550, 25^2\right)$ $\mathrm{E}(kM + m) = k\mathrm{E}(M) + m$ $550 = 450k + m$ $\mathrm{Var}(kM + m) = k^2\mathrm{Var}(M)$ $25^2 = k^2 50^2$ Since k is positive, $\rightarrow k = \sqrt{\frac{50^2}{25^2}} = \frac{1}{2}$ $\rightarrow m = 550 - \frac{450}{2} = 325$ New measurement is twice as heavy as its old measurement: $\rightarrow \frac{1}{2}M + 325 \geq 2M$ $\rightarrow \frac{3}{2}M - 325 \leq 0$ $\mathrm{E}\left(\frac{3}{2}M - 325\right) = \frac{3}{2}\mathrm{E}(M) - 325 = \frac{3}{2}(450) - 325 = 350$ $\mathrm{Var}\left(\frac{3}{2}M - 325\right) = \left(\frac{3}{2}\right)^2 \mathrm{Var}(M) = \left(\frac{3}{2}(50)\right)^2 = 75^2$ $\therefore \mathrm{P}(Y \geq 2M) = \mathrm{P}\left(\frac{3}{2}M - 325 \leq 0\right) = 1.53 \times 10^{-6}$	Deducing the correct values of k and m can be done using arithmetic properties of expectation and variance. The latter part may prove to be tricky. Be advised that the new measurement is not independent of M. As such, avoid denoting $kM + m$ as another random variable, and later forgetting that it is independent of M. **Note that $\mathbf{Var}(Y - 2M) \neq \mathbf{Var}(Y) + 2^2\mathbf{Var}(M)$ when Y depends on M.** As a side, it must be noted that given two random variables X and Y (not necessarily independent), the more general result is $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X, Y)$, where $\mathrm{Cov}(X, Y)$ is the **covariance**, or a measure of the association or dependence, between two random variables X and Y. For two random variables X and Y that are independent of each other, $\mathrm{Cov}(X, Y) = 0$ and therefore $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$.

9 (i) [1]	For X to follow binomial distribution, an infection must happen independently from person to person. In reality, **the event of a person being infected likely depends on the people who are already infected.**	This is a classic "giveaway" question testing on candidates' awareness of independence condition for binomial distribution in contextual questions.
9 (ii) [3]	$X \sim \mathrm{B}(5000, 0.02)$ $\dfrac{\mathrm{P}(X=k+1)}{\mathrm{P}(X=k)} = \dfrac{\binom{5000}{k+1}(0.02)^{k+1}(0.98)^{4999-k}}{\binom{5000}{k}(0.02)^{k}(0.98)^{5000-k}}$ $= \dfrac{\left(\frac{5000!}{(k+1)!\,(4999-k)!}\right)}{\left(\frac{5000!}{k!\,(5000-k)!}\right)}\left(\dfrac{(0.02)^{k+1}(0.98)^{4999-k}}{(0.02)^{k}(0.98)^{5000-k}}\right)$ $= \dfrac{\left(\frac{1}{(k+1)k!\,(4999-k)!}\right)}{\left(\frac{1}{k!\,(5000-k)(4999-k)!}\right)}\left(\dfrac{0.02}{0.98}\right)$ $= \dfrac{\left(\frac{1}{k+1}\right)}{\left(\frac{1}{5000-k}\right)}\left(\dfrac{1}{49}\right) = \dfrac{5000-k}{49(k+1)}$ **<u>Considering $P(X=k+1) > P(X=k)$ to find most probable X</u>** $\dfrac{\mathrm{P}(X=k+1)}{\mathrm{P}(X=k)} = \dfrac{5000-k}{49(k+1)} > 1$ $5000-k > 49k+49$ $k < \dfrac{5000-49}{50} = 99.02$ $\therefore \mathrm{P}(X=1) < \mathrm{P}(X=2) < \mathrm{P}(X=3) < \cdots < \mathrm{P}(X=98) < \mathrm{P}(X=99) < \mathrm{P}(X=100)$ Most probable value of X is 100. **<u>Otherwise</u>** Most probable value of $X = \mathrm{E}(X) = np = 5000(0.2) = 100$	The first part can be shown by making use of relevant expressions from the List of Formulae (MF26) with appropriate substitutions. Subsequently, finding the most probable value of X can be done in at least two ways: **<u>Considering $P(X=k+1) > P(X=k)$</u>** The inequality eventually implies that the probability keeps growing larger until k is at most 99.02. However, it would be a mistake to state that $X=99$ is the most probable value of X. As the inequality holds for at most $k=99$, $\mathrm{P}(X=99) < \mathrm{P}(X=100)$ is true and hence the required value is $X=100$. **<u>Using formula for expected value from MF26</u>** Candidates may choose the hassle-free route by referring to MF26 and using $\mathrm{E}(X)=np$ to obtain similar result.

9 (iii) [3]	Each group has $\frac{5000}{100} = 50$ individuals. For a group to be positive, $2\% \times \frac{5000}{100} = 1$ individual at least must be infected. Let Y be the event that a random person in a group is infected. $\text{P}(Y \geq 1) = 1 - \text{P}(Y \leq 0) = 1 - 0.36416968 = 0.63583032$ Identify 2% of the groups as positive $\rightarrow 2\% \times 100 = 2$ groups are identified positive. For 2 groups to be identified as positive just after the 5th test, (1) the first 4 tests must have exactly 1 positive test, and (2) the 5th test must be positive. Let A be the event that a test returns positive in the first 4 tests. $\therefore A \sim \text{B}(4, 0.63583032)$ Required probability $= \text{P}(A = 1) \times \text{P}(\text{group tested positive})$ $= 0.1228322985 \times 0.63583032$ ≈ 0.0781	The first part on finding the probability of a group being tested positive should be straightforward. For the subsequent part, the following pitfalls to avoid may prove noteworthy: • The distribution $X \sim \text{Geo}(0.636)$, or any similar approach fixing a positive group at the 5th test, is not appropriate. This will yield a probability the **first** positive group is found at the 5th test, instead of the **second** positive group. • The distribution $X \sim \text{B}(5, 0.636)$ is more on the right track, but still not appropriate. The probability $\text{P}(X = 2)$ in this case **might not promise a positive group** at the 5th test. An appropriate approach is to allocate one positive group to the first four, and subsequently forcing the last 5th group to be positive.

9 (iv) [4]	Each group has $\frac{5000}{k}$ individuals. For a group to be positive, $2\% \times \frac{5000}{k} = \frac{100}{k}$ individuals at least must be infected. Let Y be the event that a random person in a group is infected. $\therefore \mathrm{P}\left(Y \geq \frac{100}{k}\right) = 1 - \mathrm{P}\left(X \leq \frac{100}{k} - 1\right) = 1 - \textbf{binomcdf}\left(\frac{5000}{k}, 0.02, \frac{100}{k} - 1\right)$ $\rightarrow$ Write $p_1 = \mathrm{P}\left(Y \geq \frac{100}{k}\right)$ Let A be the event that a group is tested positive. $A \sim \mathrm{B}(k, p_1)$ More than 20% probability that exactly half of the groups are tested positive: $\mathrm{P}\left(A = \frac{k}{2}\right) > 0.2$ $\textbf{binompdf}\left(k, p_1, \frac{k}{2}\right) > 0.2$ $\rightarrow$ Write $p_2 =$ $\textbf{binompdf}\left(k, p_1, \frac{k}{2}\right)$ All in all, we use GC to find k such that: $\textbf{binompdf}\left(k, 1 - \textbf{binomcdf}\left(\frac{5000}{k}, 0.02, \frac{100}{k} - 1\right), \frac{k}{2}\right) > 0.2$ Since $\frac{k}{2}$ is the number of positive groups, we only consider finding p_2 values in which k is even. k | p_2 2 | 0.4992 > 0.2 4 | 0.3728 > 0.2 **10** | **0.237** > 0.2 20 | 0.1514 < 0.2 50 | 0.0433 < 0.2 $\therefore k_{max} = 10$	To begin, it would be useful to lay grounds by deducing the required integer numbers, namely the number of groups k and hence the number of infected people in a group that is required to render it positive. Subsequent steps involve working with a nested distribution – the bold parts indicate how some parts are keyed into the graphical calculator. Finding the final value for k can then be done by considering the even division of the groups. When tabulating the probabilities, candidates might wish to adjust the viewing window of the GC so that only relevant integers are displayed.

10 (i) [1]	The probability for $x = 95$ drops in n runs is $0.26 = \frac{13}{50}$. Such occurrences happen $\frac{13}{50}n$ times, which is only an integer when n is a multiple of 50.	A proper justification for *n* to be a multiple of 50, involves a discussion on the particular number of drops *x* whose simplest fractional probability has a nominator that shares no prime factors (or is *coprime*) with 50. Mentioning a suitable greatest common divisor would also suffice.
10 (ii) [2]	$\mathrm{E}(x)$ $= 92(0.06) + 93(0.12) + 94(0.20) + 95(0.26) + 96(0.16) + 97(0.12) + 98(0.08)$ $= 95.02$ $\mathrm{E}\left(x^2\right)$ $= 92^2(0.06) + 93^2(0.12) + 94^2(0.20) + 95^2(0.26) + 96^2(0.16) + 97^2(0.12) + 98^2(0.08)$ $= 9031.38$ $\mathrm{Var}(x) = \mathrm{E}\left(x^2\right) - [\mathrm{E}(x)]^2 = 9031.38 - [95.02]^2 = 2.5796$	This question tests on the expected value and variance of a discrete random variable using relevant formula. Be reminded to treat the final decimal value as **exact**, as such rounding to three significant figures would be inappropriate.
10 (iii) [2]	**Since *n* is a multiple of 50**, $n > 30$, and thus the sample size is large enough so that, by Central Limit Theorem, the sample mean approximately follows a normal distribution.	While this question mainly tests on the understanding on Central Limit Theorem, there must be engagement with the context to earn full marks, in particular the restriction on the value of *n* which allows for the large sample size.

10 (iv) [7]	Unbiased estimate of the population mean = $\mathrm{E}(x) = 95.02$ Unbiased estimate of the population variance $= \dfrac{n}{n-1}\mathrm{Var}(x) = \dfrac{n}{n-1}(2.5796)$ Let X be the random variable for the number of acid drops in a reaction. $H_0: \mu = 95$ $H_1: \mu > 95$ Test at 5% significance level. Assume H_0 is true.	There is many to look out for in a hypothesis testing question. The following are common pitfalls and points to look out for when conducting hypothesis testing: **Unbiased estimate of the population variance, s^2** ▪ Candidates may use the formulae found in MF26 to deduce the required expression accordingly. ▪ The formula for the unbiased estimate of the population variance given a sample variance is in MF26. However, it may not be obvious to some which formula can be used with sample variance: $$s^2 = \frac{n}{n-1} \times \left(\frac{\Sigma(x-\bar{x})^2}{n}\right).$$ $$s^2 = \frac{n}{n-1} \times (\text{sample variance})$$ As such, **be advised to remember this result.** **Null and alternative hypotheses H_0 and H_1** ▪ In H2 Mathematics, the hypotheses involve only the population mean. In particular, H_0 only takes in one value, and H_1 takes the alternative range. ▪ Questions may frame the hypothesis test such that the null hypothesis is a range of values (in this case $\mu \geq 95$). In this case, select the boundary value as the null hypothesis ($\mu = 95$) and test it against the alternative range ($\mu < 95$). ▪ Beware of incorrect notations such as "$H_0 = \ldots$"

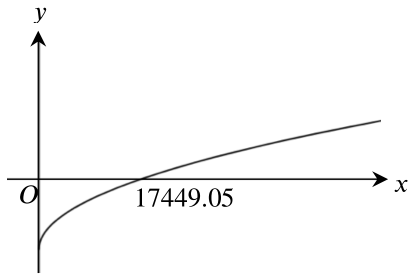
[continued]

Since n is a multiple of 50 and thus sufficiently large, by Central Limit Theorem,
$\bar{X} \sim \mathrm{N}\left(95, \frac{2.5796}{n-1}\right)$ approximately.

Test statistics is given by
$$Z = \frac{\bar{X} - 95}{\sqrt{\frac{2.5796}{n-1}}} \sim \mathrm{N}(0,1)$$

Since the acid is stronger than the alkali, do not reject $H_0 \rightarrow z$–value lies in the acceptance region
5% significance level → when $P(Z > k) = 0.05, k = 1.644853626$

Substituting $\bar{x} = 95.02$,
$$\frac{0.02}{\sqrt{\frac{2.5796}{n-1}}} > 1.644853626 \rightarrow \frac{0.02}{\sqrt{\frac{2.5796}{n-1}}} - 1.644853626 > 0$$



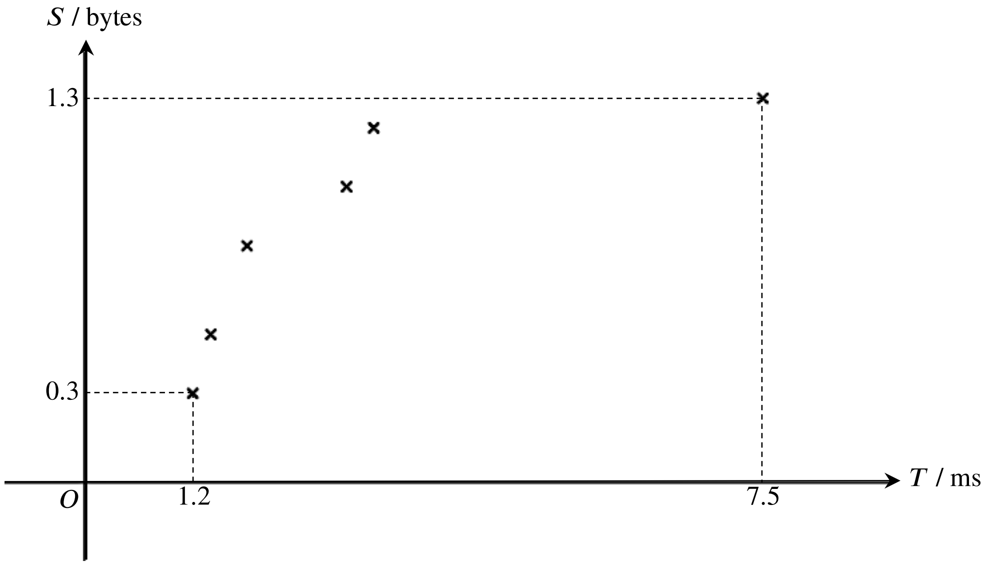
$\therefore n > 17449.05 \rightarrow n_{min} = 17450$

Central Limit Thorem, given a random variable X
- Central Limit Theorem approximates the distribution of $\bar{X}$, the mean of n observations of X, **not on X itself.**
- Central Limit Theorem must be explicitly stated if: (1) the distribution of X is unknown, **and** (2) X is observed for sufficiently large times $n > 30$. On the other hand, **normally distributed observations need not be justified** with this theorem, even if n is small.
- **When a question asks for an assumption which validates the conclusion, Central Limit Theorem is most likely not the answer.** The theorem applies for any distribution of X. Find other assumptions.
- Beware of writing its abbreviation "CLT". Most marking schemes rule against such presentation.
- Remember to divide the variance by n for the approximated distribution of $\bar{X}$.
- Beware of missing out the word "approximately".

Test conclusions
- Avoid writing generic test conclusions such as "reject H_0" or "do not reject H_0". It is good practice to write down its full interpretation in context.
- If some range(s) of values is to be found, be wary of two-tailed tests. For such cases, avoid using the p–value with the **CENTER** qualifier in GC (invNorm). Use $1 - p$ instead, or alternatively use $\frac{p}{2}$ with either **LEFT** or **RIGHT** qualifiers.

Having come this far, beware not to overlook the "multiple of 50" condition established in **(i)** for the least value of n.

11 (i) [3]	Check best-cases correlation with n^2 • T_A and n^2: $r = 0.9858779204$ • T_B and n^2: $r = 0.9819469142$ • T_C and n^2: $r = 0.9999920261$ $\therefore C$ is Selection Sort. Check worst-cases correlation with n^2 • T_A and n^2: $r = 0.9989170205$ • T_B and n^2: $r = 0.9823633575$ $\therefore A$ is Quick Sort. $\therefore B$ is Merge Sort.	The best strategy to tackle this question is to recognize which time complexity is the odd-one out for each best-case and worst-case. Finding and comparing product moment correlation coefficients follow.
11 (ii) [1]	The runtime 3.0ms is out of the range for T_B, **both for best-case and worst-case**. In any case, the data size suggested would thus have been obtained via extrapolation, which is unreliable.	Take care to mention that it is extrapolation for **both** best-case and worst-case in this context.
11 (iii) [3]	Use algorithm A, Quick Sort. Propose a best-case scenario. Since n is the independent variable, use regression line of T against $n \log_2 n$: $T = -0.2479929003 + \left(2.464222741 \times 10^{-4}\right) n \log_2 n$ When $T = 3.0$, $n \log_2 n = 13180.59787$ Using GC, $n = 1277.3246803198 \approx 1277$ bytes.	Upon investigating, only one row will allow for interpolation. After identifying the algorithm from **(i)** and the appropriate case, the regression line follows using GC. Take care to **not** use the regression line of $n \log_2 n$ against T in this case, as this would suggest that n depends on T, which is contextually inaccurate.
11 (iv) [2]	$\bar{T} = \frac{1.2 + 1.4 + 1.8 + 2.9 + 3.2}{5} = \frac{10.5}{5} = 2.1$ $\bar{S} = \frac{0.3 + 0.5 + 0.8 + s + 1.2}{5} = \frac{s + 2.8}{5}$ Regression line of S on T passes through $\left(\bar{T}, \bar{S}\right)$ $\rightarrow \frac{s + 2.8}{5} = \frac{7}{18}(2.1) - \frac{17}{300} = \frac{19}{25}$ $\rightarrow s = \frac{19}{5} - 2.8 = 1.0$	This question tests on the candidates' ability to recall that regression lines pass through the mean values of the bivariate data. Finding s subsequently follows with algebra.

11 (v) [4]	[Any model is acceptable if it shows that **as T increases, S <u>increases at a decreasing rate</u>**, e.g.: $\ln x$]  For $S = \frac{7}{18}T - \frac{17}{300}$, $r = 0.7987181381 \approx 0.799$ Suggest $S = a + b\ln T$ Using GC, $a = 0.3682482527 \approx 0.368$, $b = 0.5403484857 \approx 0.540$, $r = 0.9245838014 \approx 0.925$, $\therefore$ New regression line is $S = 0.368 + 0.540T$, The product moment correlation coefficient of the second model ($r \approx 0.925$) is closer to 1 than that of the first (linear) model ($r \approx 0.799$), suggesting a stronger positive linear correlation between S and T. Hence, the second model is better.	After drawing the scatter diagram which includes the new data, it becomes clear that the bivariate data is no longer suitably modelled as a linear relationship, but instead a curvilinear one. Take care to propose a model that shows the appropriate behaviour. Subsequently, finding the new regression line follows. When providing justification, there must be evidence of awareness that the r–value of the new model is closer to 1 than the old model, or that the diagram shows that as T increases, S increases at a decreasing rate. Vague explanations that simply states "the r–value of the new model is higher" or "S is not linearly related to T anymore" may not be as well-received.