A-Levels Math Notes

Grass

Contents

Part 1

FMA

Chapter 1

Inequalities and Equations

1.1 Solving Inequalities

General Methods

1. Quadratic formula for factorisation / finding roots (of polynomial).
2. Completing the square.
3. Discriminant/Completing the square to eliminate factors which are *always* positive or negative (e.g. removing $x^2 - 3x + 4$). *Note to include coefficient of x^2 in the argument.*
4. GC (include sketch).
5. *Rational Functions*: Move everything to one side by adding or subtracting, then use a number line.

1.2 Modulus Inequalities

Fact

Given $x \in \mathbb{R}$, we have that

- $|x| \geq 0$,
- $|x^2| = |x|^2 = x^2$,
- $\sqrt{x^2} = |x|$.

And as long as $x \in \mathbb{R}^+$,

- $\sqrt{x}^2 = |x|$.

Useful Properties

For every $x, k \in \mathbb{R}$:

(a) $|x| < k$ iff $-k < x < k$.

(b) $|x| > k$ iff $x < -k$ or $x > k$.

1.3 System of Linear Equations

General Information

- For more complicated real-world-context qns, try playing around with the values (e.g. use simultaneous equations) first. It may work out nicer than expected.

1.4 Summary

G.C. Skills

1. Plotting curves $y = f(x)$ in G.C.
2. How to use simultaneous equation solver.

Important Notes

- Eliminating Factors — *only* works for $c = 0$ in $f(x) \geq c$ or $f(x) \leq c$.

  Counterexample: It is false that $P(x) = x(3x^2 - 9x + 10) \leq 2$ iff $x \leq 2$. Notice that $P(1.8) = 6.336 \not\leq 2$.

- Discriminant — include coefficient of x^2 in argument.
- When using factor elimination to remove some $f(x)$, we only need to say that "$f(x)$ is negative".
- Rational functions — exclude the values that causes division by zero to occur.
- With inequalities, be really careful about multiplication! If $x > y$ *and* $z > 0$, then $xz > yz$.
- Cross multiplication preserves order for $\frac{x}{y} < \frac{x'}{y'}$ iff y and y' are *both* positive or negative.

  Note the counterexample $\frac{1}{2} < \frac{1}{-3}$.

- Squaring preserves/reverses order for $x < y$ iff x and y are *both* positive or negative.
- Can't necessarily use differentiation to solve if qns asks for algebraic method.
- Safer to graph out the two functions separately!
- Be careful about whether to include equality! Don't forget to account for it!
- Note that when solving for $|x| = y$, $|x| < y$, etc, y must be greater than or equal to 0. In other words, there may be solutions we will need to reject.
- Carelessness: Look at the question carefully! If they ask for a *set* of values, then rmb to give it as a *set*!
- Exponentiation and Logarithms: Simply use ln and avoid $\log_c$ for $c < 1$.

  Order is *Preserved* under exponentiation/logarithms if the base is *larger than* one. Otherwise, when it is *less than* one, the order is *reversed*. `https://www.desmos.com/calculator/gd8z5fa0bg`

- For more complicated real-world-context qns, try playing around with the values (e.g. use simultaneous equations) first. It may work out nicer than expected.

Chapter 2

Sequences and Series

2.1 Binomial Theorem and Series

Theorem 2.1: The Binomial Theorem

$$(a+b)^n = \sum_{r=0}^{n} \binom{n}{r} a^{n-r} b^r,$$

where $n \in \mathbb{Z}^+$.

Theorem 2.2: The Binomial Series

$$(1+x)^p = \sum_{r=0}^{\infty} \binom{p}{r} x^r,$$

where $p \in \mathbb{Q}$, $|x| < 1$, and

$$\binom{p}{r} := \frac{p(p-1)\cdots(p-r+1)}{r!}.$$

Corollary 2.3

Clearly,

$$(a+x)^p = a^p \left(1 + \frac{x}{a}\right)^p = a^p \sum_{r=0}^{\infty} \binom{p}{r} \frac{x^r}{a^r},$$

under the same conditions.

Fact

We can expand $(a+x)^p$ in descending powers of x by using $(a+x)^p = x^p \left(1 + \frac{a}{x}\right)^p$.

Note

Sometimes computing a couple terms can be useful in finding a pattern. For example, to get the coefficient of x^k explicitly.

2.2 APGP

Basics

	AP	GP
u_n	$u_n = S_n - S_{n-1}$	
	$u_n = a + (n-1)d$	$u_n = ar^{n-1}$
S_n	$S_n = \frac{n}{2}[2a + (n-1)d]$ $= \frac{n}{2}(a + \ell)$	$S_n = \frac{a(1-r^n)}{1-r}$ $= \frac{a(r^n - 1)}{r-1}$
S_∞	Always diverges	$S_\infty = \frac{a}{1-r}$ when $\lvert r \rvert < 1$
Prove AP/GP	I Show $u_n - u_{n-1}$ is a constant / independent of n. II Show $u_n = a + (n-1)d$ explicitly.	I Show $\frac{u_n}{u_{n-1}}$ is constant / independent of n. II Show $u_n = ar^{n-1}$ explicitly
Mean	$u_{n+1} = \frac{u_n + u_{n+2}}{2}$. (Arithmetic Mean)	$\frac{u_{n+1}}{u_n} = \frac{u_{n+2}}{u_{n+1}}$ $u_{n+1}^2 = u_n \cdot u_{n+2}$ (Geometric Mean)

Important Notes

Applications: Write out a few terms in a table and observe the trend. (You can literally say "By observing a trend, ...")

G.C. Skills

Table function

1. Enter eqn into GC.
2. 2nd graph to show table
3. 2nd tblset for setup options

2.3 Summation

Fact

$$\begin{aligned}
\sum_{i=m}^{n} f(i) + g(i) &= \sum_{i=m}^{n} f(i) + \sum_{i=m}^{n} g(i) \\
\sum_{i=m}^{n} af(i) &= a\sum_{i=m}^{n} f(i) \\
\sum_{i=m}^{n} a &= (n-m+1)a, \text{ for any constant a} \\
\sum_{i=m}^{n} f(i) &= \sum_{i=1}^{n} f(i) - \sum_{i=1}^{m-1} f(i)
\end{aligned}$$

Note

- Look out for sums being AP and GPs.
- Results to be provided:

$$\sum_{i=1}^{n} i^2 = \frac{n}{6}(n+1)(2n+1)$$
$$\sum_{i=1}^{n} i^3 = \frac{1}{4}n^2(n+1)^2$$

2.4 Method of Differences

General Information

$$\sum_{i=1}^{n} u_i = \sum_{r=1}^{n} f(r) - f(r-1) = f(n) - f(0).$$

- Explain convergence of a function $h(x) = f(x) + g(x)$: As $n \to \infty$, $f(x) \to 0$ and $g(x) \to 0$. Hence, $h(x)$ converges to...

G.C. Skills

Know how to generate sequences using the two methods:

1. Table method — Present by showing two consecutive values of n so that the values of the sequence are of opposite signs. E.g.:

n	S_n
182	$561.28 < 0$
183	$-1935.91 < 0$

2. 2nd stat seq (& we can use operations on seq, e.g. sum)

Chapter 3

Recurrence Relations

General Information

1. Recurrence relation is *homogenous* if constant (b below) is zero.
2. First order linear recurrence relation: $u_n = au_{n-1} + b$, with $a \neq 0$.
3. Second order *homogenous* linear recurrence relation: $u_n = a_1u_{n-1} + a_2u_{n-2}$, $a_2 \neq 0$.
4. Solving RRs in general:
   (a) Continually expand u_n in terms of u_{n-1}, then in terms of u_{n-2}, $\ldots$, till an explicit formula is obtained.
   (b) Use a_1 to generate $a_2, a_3, \ldots, a_n$.
5. Solving 1st order RRs, $u_{n+1} = au_n + b$:
   (a) Iteration — Essentially technique 4(a). Will need to use G.P. formula at the end.
   (b) Rewriting RR + Using G.P. Formulas ((c) is better)
      i. Write RR as $u_n - k = a(u_{n-1} - k)$, where $k = \frac{b}{1-a}$. Let $v_n = u_n - k$.
      ii. $\frac{v_n}{v_{n-1}} = a$, a constant and $\{v_n\}$ is a G.P. with first term $v_1 - k$ and common ratio a.
      iii. So, $v_n = (u_1 - k)a^{n-1}$, and accordingly, $u_n = v_n + k = (u_1 - k)a^{n-1} + k$.
   (c) ★ Let $u_n = Aa^n + \frac{b}{1-a}$. Then solve for the constant A with info provided.
6. Solving 2nd order (homogenous) RRs, $u_{n+2} = au_{n+1} + bu_n$:
   Assume $u_n = m^n$, then $m^2 - am - b = 0$ (is the *characteristic/auxillary equation* of the RR). Solve for the roots, say m_1 and m_2. Then, the general solution for u_n is

$$u_n = \begin{cases} Am_1^n + Bm_2^n & \text{if } m_1, m_2 \in \mathbb{R}, \text{ and } m_1 \neq m_2, \\ (C + Dn)m_1^n & \text{if } m_1, m_2 \in \mathbb{R}, \text{ and } m_1 = m_2, \\ r^n[A\cos(n\theta) + B\sin(n\theta)] & \text{if } m_1, m_2 \in \mathbb{C}, \ m_1 = re^{i\theta}, \text{ and } m_2 = re^{-i\theta}. \end{cases}$$

Note

Let $x_{n+1} = f(x_n)$ and $L \coloneqq \lim x_n$. To find the possible values of L, we can compare the graph of $y = f(x)$ against the identity function $y = x$. This is done by seeing if $f(x) < x$, $f(x) = x$, or $f(x) > x$.

Note

We should remember Vieta's Formulas. Consider a complex polynomial $a_2z^2 + a_1z + a_0$ with roots r_1 and r_2. Then, the sum $r_1 + r_2 = -a_1/a_2$ and the product $r_1r_2 = a_0/a_2$.

Example 3.1

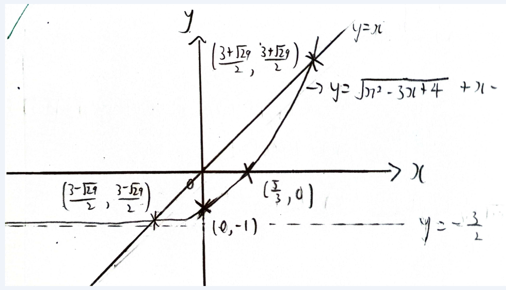


Figure 3.1: The RR $x_{n+1} = \sqrt{x_n^2 - 3x_n + 4} + x_n - 3$.

Let $f(x) = \sqrt{x^2 - 3x + 4} + x - 3$.

1. Suppose $x_1 \le \frac{3+\sqrt{29}}{2}$. For $x_1 < \frac{3-\sqrt{29}}{2}$, we see that $f(x) > x$. So x_n increases till $\frac{3-\sqrt{29}}{2}$. While for $\frac{3-\sqrt{29}}{2} < x_1 < \frac{3+\sqrt{29}}{2}$, we have $f(x) < x$. Thus x_n decreases till $\frac{3-\sqrt{29}}{2}$. Notice the graphs intersects at $x = \frac{3-\sqrt{29}}{2}$. So, when $x_n = \frac{3-\sqrt{29}}{2}$, if ever, then $x_{n+1} = x_n$. That is, $L = \frac{3-\sqrt{29}}{2}$.

2. Similarly, if $x_1 = \frac{3+\sqrt{29}}{2}$, then $x_n = \frac{3+\sqrt{29}}{2}$ is a constant function; $L = \frac{3+\sqrt{29}}{2}$.

3. Presume that $x_n > \frac{3+\sqrt{29}}{2}$. Then, $f(x) > x$ tells us x_n is an increasing sequence that is unbounded. In other words, L does not exist.

Chapter 4

Induction

General Information

Let $P(x)$ be the statement that "...".

When $n = 1$, ...

$\implies P(1)$ is true.

Assume $P(k)$ is true for some $k \in \mathbb{Z}^+$.

Then, ...

$\implies P(k+1)$ is true.

Therefore, since $P(1)$ is true and $P(k)$ true $\implies P(k+1)$ true, $P(n)$ is true for all $n \in \mathbb{Z}^+$.

Chapter 5

Differentiation

Definition

1. A function f is called (strictly) increasing on an interval I iff $f'(x) > 0$ for all $x \in I$.
2. A function f is called monotonically increasing on an interval I iff $f'(x) \geq 0$ for any $x \in I$.

General Information

1. How to sketch the graph of the integral or derivative of a function f.
2. Relationship btw. a function f and its derivative, f':

$y = f(x)$	$y = f'(x)$
Vertical asymptote at $x = a$	Vertical asymptote at $x = a$.
Horizontal asymptote at $y = a$	Horizontal asymptote $y = 0$.

3. Recap:

$f(x)$	$f'(x)$
$\sin^{-1}\left(\dfrac{x}{a}\right)$	$\dfrac{1}{\sqrt{a^2 - x^2}}$, $\lvert x\rvert < a$
$\cos^{-1}\left(\dfrac{x}{a}\right)$	$-\dfrac{1}{\sqrt{a^2 - x^2}}$, $\lvert x\rvert < a$
$\tan^{-1}\left(\dfrac{x}{a}\right)$	$\dfrac{a}{a + x^2}$, $x \in \mathbb{R}$
$\log_a(f(x))$	$\dfrac{1}{x\ln(a)}$
a^x	$a^x \ln(a)$

4. Implicit differentiation: $\dfrac{dz}{dx} = \dfrac{dz}{dy} \cdot \dfrac{dy}{dx}$. ★ Makes life much easier (e.g. finding $f^{(n)}(x)$).
5. Parametric Differentiation: $\dfrac{dy}{dx} = \dfrac{dy}{dt} \cdot \dfrac{dt}{dx}$.
6. Small angle approximation:
   (a) $\sin(x) \approx x$,
   (b) $\cos(x) \approx 1 - \dfrac{x^2}{2}$,

   (c) $\tan(x) \approx x$.

7. Maclaurin Series:

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!} x.$$

Chapter 6

Integration Techniques

6.1 Basic Integration (IBS, IBP, etc)

General Information

1. Factor Formulae ★ (must *rmb*):

   (a) $\sin(mx)\cos(nx) = \frac{1}{2}[\sin((m+n)x) + \sin((m-n)x)]$,

   (b) $\cos(mx)\cos(nx) = \frac{1}{2}[\cos((m+n)x) + \cos(m-n)x]$,

   (c) $\sin(mx)\sin(nx) = -\frac{1}{2}[\cos((m+n)x) - \cos((m-n)x)]$.

2. Common classes of integrals:

   (a) Apply partial fractions:
   $$\int \frac{f(x)}{g(x)}\,dx.$$

   (b) Split $px+q$, then complete the square:
   $$\int \frac{px+1}{\sqrt{ax^2+bx+c}}\,dx \quad \text{or} \quad \int \frac{px+1}{ax^2+bx+c}\,dx$$

3. Integration by Substitution:
   $$\int f(x)\,dx = \int f(x)\frac{dx}{du}\,du.$$

4. Use Pythagoras' Theorem / draw a right-angled triangle to help with trig conversions, e.g.:
   $$\tan(\theta) \qquad \text{to} \qquad \frac{x+1}{\sqrt{2-(x+1)^2}}.$$

5. Integration by Parts:

   Let $u = g(x)$, $\dfrac{dv}{dx} = h(x)$,

   $\dfrac{du}{dx} = g'(x)$, $v = \displaystyle\int h(x)\,dx$.

   $$\int u\left(\frac{dv}{dx}\right)dx = uv - \int v\left(\frac{du}{dx}\right)dx.$$

6.2 Areas & Volumes

General Information

1. Volume of revolution when rotated about x-axis:

   (a) The disc method
   $$\int_{x_1}^{x_2} \pi y^2\, dx = \int_{x=x_1}^{x=x_1} \pi y^2 \frac{dx}{dt}\, dt.$$

   (b) The shell method:
   $$\int_{x_1}^{x_2} 2\pi yx\, dy.$$

2. Arc length:
$$\int_{x_1}^{x_2} \sqrt{\left(\frac{dy}{dx}\right)^2 + 1}\, dx = \int_{y_1}^{y_2} \sqrt{\left(\frac{dx}{dy}\right)^2 + 1}\, dy = \int_{t_1}^{t_2} \sqrt{\left(\frac{dy}{dt}\right)^2 + \left(\frac{dx}{dt}\right)^2}\, dt = \int_{\alpha}^{\beta} \sqrt{r^2 + \left(\frac{dr}{d\theta}\right)^2}\, d\theta.$$

3. Surface area of revolution when rotated about x-axis:
$$\int_{x_1}^{x_2} 2\pi y \sqrt{\left(\frac{dy}{dx}\right)^2 + 1}\, dx = \int_{y_1}^{y_2} 2\pi y \sqrt{\left(\frac{dx}{dy}\right)^2 + 1}\, dy = \int_{t_1}^{t_2} 2\pi y \sqrt{\left(\frac{dy}{dt}\right)^2 + \left(\frac{dx}{dt}\right)^2}\, dx = \int_{\alpha}^{\beta} \frac{1}{2} r^2\, d\theta.$$

★ Rotating about x-axis $\implies$ y in integrand
Rotating about y-axis $\implies$ x in integrand.

6.3 Numerical Methods

6.3.1 Trapezium Rule

General Information

1. Formula for n intervals, or (n+1)ordinates, of width $h := (b-a)/n$:

$$\int_a^b y\,dx = \frac{h}{2}(y_0 + 2y_1 + 2y_2 + \cdots + 2y_{n-1} + y_n).$$

2. Illustration

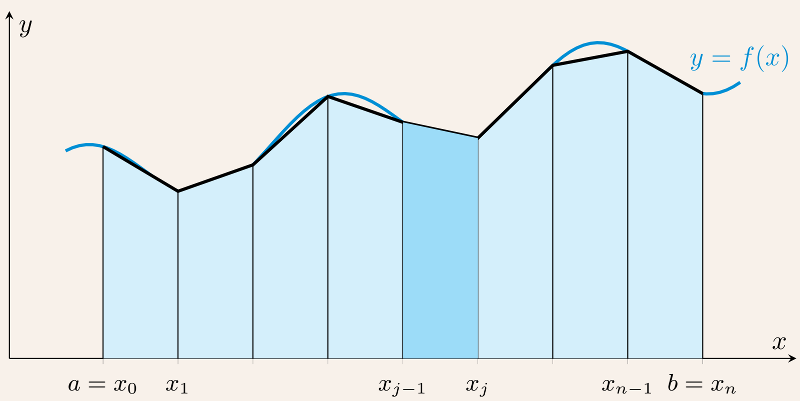


Figure 6.1: Trapezium rule

3. Error:

   (a) Concave upwards, i.e. ($f'(x)$ is increasing / $f''(x) > 0$) $\implies$ overestimation.

   (b) Concave downwards, i.e. ($f'(x)$ is decreasing / $f''(x) < 0$) $\implies$ underestimation.

6.3.2 Simpson's Rule

General Information

1. Formula for n intervals, or (n+1)ordinates, of width $h := (b-a)/n$:

$$\int_a^b y\,dx = \frac{h}{3}(y_0 + 4y_1 + 2y_2 + 4y_3 + \cdots + 2y_{n-2} + 4y_{n-1} + y_n).$$

   Note that the number of intervals n should be *even*, that of ordinates *odd*.

2. Illustration

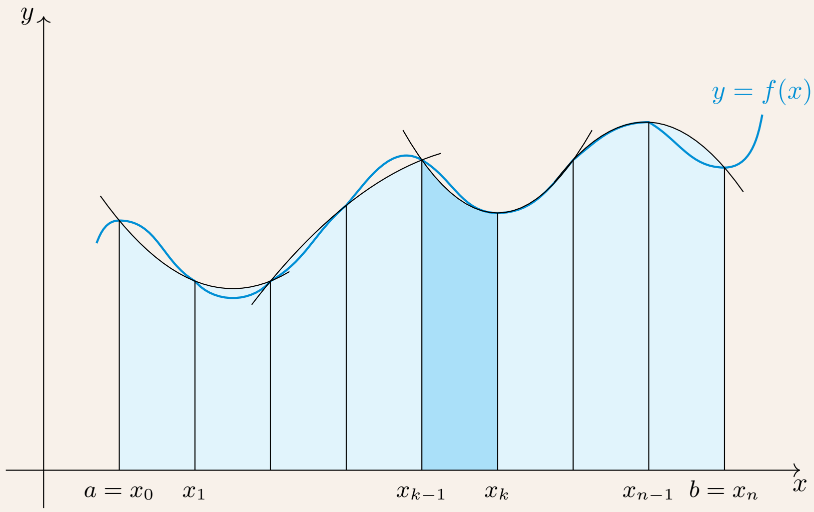


Figure 6.2: Simpson's rule

Note

Accuracy of the Trapezium rule vs Simpson's Rule: "Simpson's Rule uses *quadratic curves* to interpolate the points on the curve so it usually *gives a better approximation* to the actual curve than the trapezium rule which uses *straight lines* to interpolate the ordinates."

Chapter 7

Complex Numbers

7.1 Complex Number I

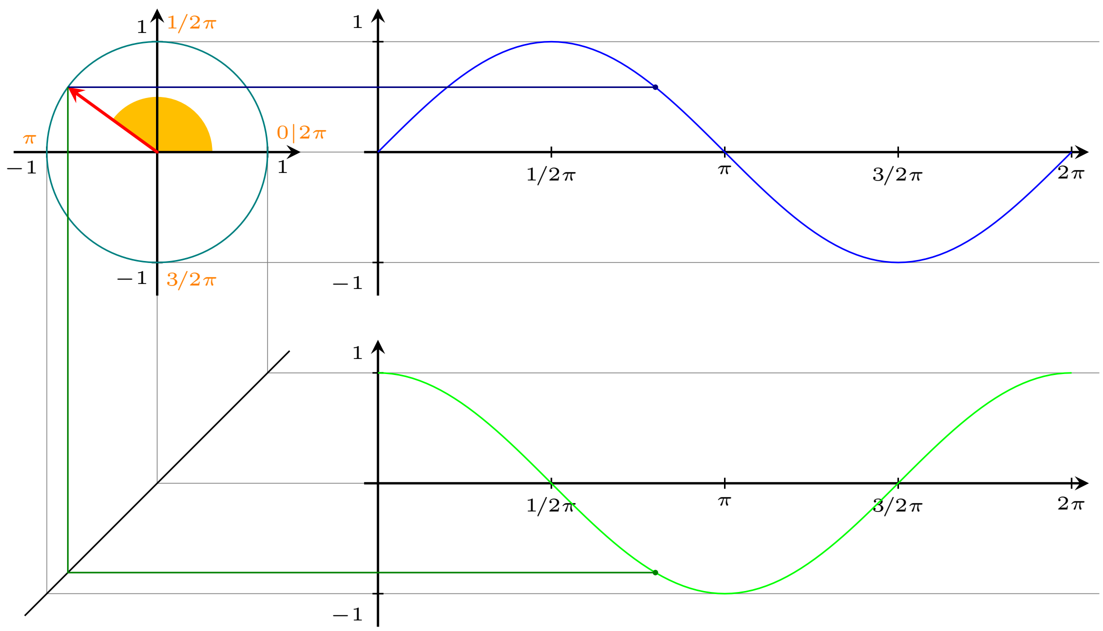


Figure 7.1: Argand diagram.

General Information

1. Find the square root of $x + iy$: Let $\sqrt{x + iy} = a + bi$. Then square both sides & solve.
2. Simplifying fractions: multiply by the denominator's conjugate
$$\frac{a+bi}{c+di} = \frac{a+bi}{c+di} \cdot \frac{c-di}{c-di} = \cdots .$$
3. Polynomials:
   (a) Fundamental Theorem of Algebra: If $p(z) \coloneqq \sum_{i=0}^{n} a_i z^i$ is a polynomial of degree $n \geq 1$ with complex coefficients, then there exists complex numbers c_i for each $1 \leq i \leq n$ such that
$$p(z) = a_n \prod_{i=1}^{n} (z - c_i).$$

(b) If a polynomial in real coefficients only has root $a + bi$, then $a - bi$ is another root.

Example 7.1

Find the roots of $iz^2 + 2z + 3i = 0$.

$$z^2 - 2iz + 3 = 0$$
$$z = \frac{2i \pm \sqrt{(2i)^2 - 4(1)(3)}}{2(1)} = i \pm \frac{\sqrt{-16}}{2} = i \pm 2i$$

So, $z = 3i$ or $z = -i$.

Example 7.2: N2010/2/1

One root of the equation $x^4 + 4x^3 + ax + b = 0$, where a and b are real, is $x = -2 + i$. Find the values of a and b and the other roots.

Substitute $-2 + i$ into the equation:

$$(-2 + i)^4 + 4(-2 + i)^3 + (-2 + i)^2 + a(-2 + i) + b = 0$$
$$-12 + 16i = 2a - b - ai$$
$$a = -16, \quad 2a - b = -12$$

Therefore, $a = -16$, $b = -20$.

Since all the coefficients of the polynomial are real (**explain**), $-2 - i$ is another root. Now, $x^4 + 4x^3 + ax + b = (x - (-2 + i))(x - (-2 - i))(cx + d)$ for some $c, d \in \mathbb{R}$.

Accordingly, substitute $x = 0$, then $x = 2$, and solve. Alternatively, notice $x^4 + 4x^3 + ax + b = (x^2 - 2(-2)x + ((-2)^2 + 1^2))(x^2 + cx + d) = (x^2 + 4x + 5)(x^2 + 4x + 5)$. Either ways, we have $c = 0$ and $d = -4$. As such, the last two roots are $x = -2 \pm i$ and $x = \pm 2$.

(c) Simultaneous equations: Solve as usual.

(d) Properties of modulus: $|z_1^x z_2^y| = |z_1|^x |z_2|^y$, for any $x, y \in \mathbb{R}$.

(e) Properties of arguments (same as log): $\arg(z) \in (-\pi, \pi]$ and $\arg(z_1^x z_2^y) = x \arg(z_1) + y \arg(z_2)$ for any $x, y \in \mathbb{R}$.

(f) Polar form: $z = re^{i\theta}$.

(g) Polar/Trigonometric form: $z = r[\cos(\theta) + i \sin(\theta)]$.

Note

Show that the value of w^n is either 2^n or 2^{-n} for integers n.

Then we **must** show that $w^n = \cdots = \begin{cases} 2^n(1) = 2^n & \text{if } n \text{ even,} \\ 2^n(-1) = -2^n & \text{if } n \text{ odd.} \end{cases}$

Note

Common tricks to know:

1. Replace all occurrences of w in a polynomial $P(x)$ with $-w$.

2. Notice that a geometric series is being used. E.g. $\frac{1}{z^2} - \frac{1}{z} + 1 - z + z^2 = \frac{z^5+1}{z^2(z+1)}$.

7.2 Complex Numbers II

Theorem 7.1: De Moivre's Theorem

Let z be a complex number, n an integer, and θ an angle. Suppose $z = re^{i\theta}$. Then,

$$z^n = e^{i\theta} = r^n[\cos(n\theta) + i \sin n\theta].$$

General Information

1. All nth roots of any complex number are the same distance r from the origin and have the same angular separation, π/n.
2. Note that $1 + e^{i\theta} = e^{i\theta/2}(e^{-i\theta/2} + e^{i\theta/2})$.
3. For $z = re^{i\theta}$, we have $z^n + z^{-n} = 2\cos(n\theta)$ and $z^n - z^{-n} = 2i\sin(n\theta)$.
4. The geometric meaning of multiplying by i is a anti-clockwise rotation by π radians.
5. Loci (Use a *compass*)

   (a) The locus represented by $|z - a| = r$ (or $z = a + re^{i\theta}$) is a *circle* of radius r centered at $A(x, y)$ (where $a \coloneqq x + iy$).

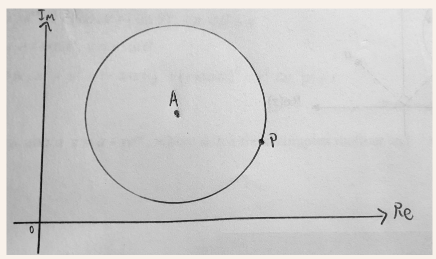


Figure 7.2: The locus of $|z - a| = r$.

   i. Either label the four points to the direct North, South, East, West of the circle, or denote the radius clearly.

   ii. The line segment, representing the furthest distance from a point to a circle, always cuts through the circle's centre. So, the distance

$$\text{OP}_{\text{max}} - \text{OP}_{\text{min}} = 2 \cdot \text{radius}.$$

   iii. The line segments, from a point to a circle that produces the largest angle, are tangents to the circle.

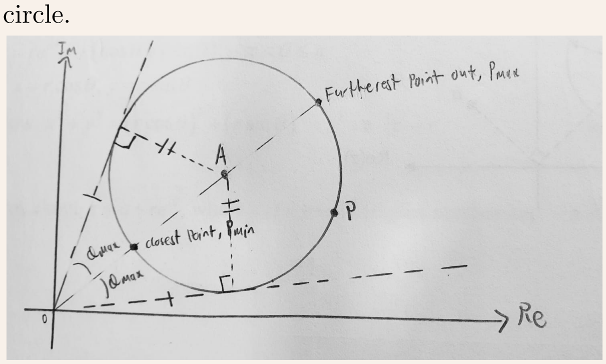


Figure 7.3: Maxmium distance and angle of a point from a circle

   (b) The locus represented by $|z - a| = |z - b|$ is the *perpendicular bisector* of the line segment joining A and B.

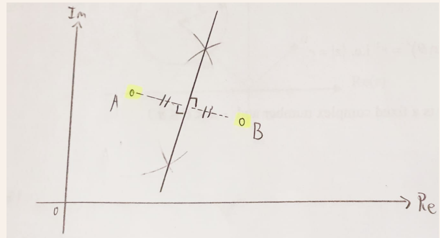


Figure 7.4: The locus of $|z-a| = |z-b|$, a perpendicular bisector

(c) The locus represented by $\arg(z-a) = \theta$ is the *half-line* from A (excluding A) that makes an angle θ with the *positive* real axis.

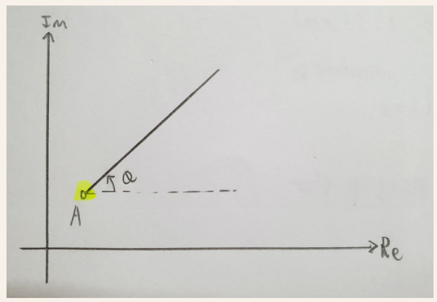


Figure 7.5: The locus of $\arg(z-a) = \theta$, a half-line.

6. There is no need to find the points of intersection between two loci, unless the questions states so.

7. Suppose we have a locus z represented by the predicate $P(z)$. Then, for any $a \in \mathbb{C}$, the locus of $z + a$ is represented by $P(z-a)$.

8. Say we are given a locus z represented by $|z-a| = r$, where $a = \alpha + \beta i$.

   (a) The greatest and least value of $|z|$ are $|a| \pm r$, respectively.

   (b) The greatest and least value of $\arg(z)$ can be obtained geometrically, or by plotting

$$Y_1 = \tan^{-1}\left(\frac{\beta \pm \sqrt{r^2 - (X-\alpha)^2}}{X}\right)$$

   and finding the maximum/minimum point, respectively.

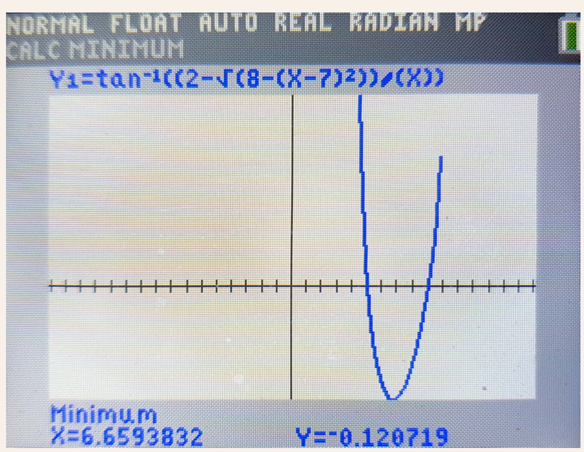


Figure 7.6: Brute Force Technique for Finding Maximum/Minimum angles.

Example 7.3: TQ 10(b)

Show that $\cot^2(2\pi/5)$ is a root of the equation $px^2 + qx + r = 0$, where we are given

$$\cot(4\theta) = \frac{\cot^4(\theta) - 6\cot^2(\theta) + 1}{4\cot^3(\theta) - 4\cot(\theta)}.$$

First notice that $\cot(8\pi/5) = -\cot(2\pi/5)$. So,

$$-\cot(2\pi/5) = \frac{\cot^4(2\pi/5) - 6\cot^2(2\pi/5) + 1}{4\cot^3(2\pi/5) - 4\cot(2\pi/5)}.$$

Simplifying gives

$$5[\cot^2(2\pi/5)]^2 - 10[\cot^2(2\pi/5)] + 1 = 0.$$

Thus, $x = \cot^2(2\pi/5)$ is a root of the equation $5x^2 - 10x + 1 = 0$.

Chapter 8

Linear Algebra

Definition 8.1

A vector space (or linear space) V over a field $\mathbb{F}$ consists of a set on which two operations (called addition and multiplication respectively here) are defined so that;

(A) (V is Closed Under Addition) For all $\mathbf{x}, \mathbf{y} \in V$, there exists a unique element $\mathbf{x} + \mathbf{y} \in V$.

(M) (V is Closed Under Scalar Multiplication) For all elements $a \in \mathbb{F}$ and elements $\mathbf{x} \in V$, there exists a unique element $a\mathbf{x} \in V$.

Such that the following properties hold:

(VS 1) (Commutativity of Addition) For all $\mathbf{x}, \mathbf{y} \in V$, we have $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$.

(VS 2) (Associativity of Addition) For all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in V$, we have $(\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z})$.

(VS 3) (Existance of The Zero/Null Vector) There exists an element in V denoted by $\mathbf{0}$, such that $\mathbf{x} + \mathbf{0} = \mathbf{x}$ for all $\mathbf{x} \in V$.

(VS 4) (Existance of Additive Inverses) For all elements $\mathbf{x} \in V$, there exists an element $\mathbf{y} \in V$ such that $\mathbf{x} + \mathbf{y} = \mathbf{0}$.

(VS 5) (Multiplicative Identity) For all elements $x \in V$, we have $1\mathbf{x} = \mathbf{x}$, where 1 denotes the multiplicative identity in $\mathbb{F}$.

(VS 6) (Compatibility of Scalar Multiplication with Field Multiplication) For all elements $a, b \in \mathbb{F}$ and elements $\mathbf{x} \in V$, we have $(ab)\mathbf{x} = a(b\mathbf{x})$.

(VS 7) (Distributivity of Scalar Multiplication over Vector Addition) For all elements $a \in \mathbb{F}$ and elements $\mathbf{x}, \mathbf{y} \in V$, we have $a(\mathbf{x} + \mathbf{y}) = a\mathbf{x} + a\mathbf{y}$.

(VS 8) (Distributivity of Scalar Multiplication over Field Addition) For all elements $a, b \in \mathbb{F}$, and elements $\mathbf{x} \in V$, we have $(a + b)\mathbf{x} = a\mathbf{x} + b\mathbf{x}$.

Theorem 8.2

Let V be a vector space and W a subset of V. Then W is a subspace of V iff the following 3 conditions hold for the operations defined in V.

(a) $\mathbf{0} \in W$

(b) $\mathbf{x} + \mathbf{y} \in W$ whenever $\mathbf{x} \in W$ and $\mathbf{y} \in W$.

(c) $c\mathbf{x} \in W$ whenever $c \in \mathbb{F}$ and $\mathbf{x} \in W$.

Definition 8.3

A subset S of a vector space V *generates* (or *spans*) V iff $\operatorname{span}(S) = V$. In this case, we also say that the vectors of S generate (or span) V.

Definition 8.4

Let V be a vector space and S a nonempty subset of V. A vector $v \in V$ is called a *linear combination* of vectors of S iff there exists a finite number of vectors $u_1, u_2, \ldots, u_n$ in S and scalars $a_1, a_2, \ldots, a_n$ in $\mathbb{F}$ such that

$$v = \sum_{i=1}^{n} a_i u_i.$$

In this case we also say that v is a linear combination of $u_1, u_2, \ldots, u_n$ and call $a_1, a_2, \ldots, a_n$ the *coefficients* of the linear combination

Definition 8.5

A set subset S of a vector space V is called *linearly dependent* iff there exists a finite number of distinct vectors $u_1, u_2, \ldots, u_n$in S and scalars $a_1, a_2, \ldots, a_n$ not all zero, such that

$$a_1 u_1 + a_2 u_2 + a_n u_n = \mathbf{0}.$$

Definition 8.6

A *basis* β for a vector space V is a linearly independent subset of V that generates V. If β is a basis for V, we also say that the vectors of β form a basis for V.

Theorem 8.7: The Rank-Nullity Theorem.

For any vector spaces V and W, and a linear operator $T\colon V \to W$, it holds that

$$\operatorname{rank}(T) + \operatorname{nullity}(T) = \dim(V).$$

General Information

- Let $\mathbf{A}$ be an $m \times n$ matrix, and $\mathbf{a}_j$ its jth column. For any $\mathbf{x} = \begin{pmatrix} x_1 & x_2 & \cdots & x_n \end{pmatrix}^\top$,

$$\mathbf{A}\mathbf{x} = \sum_{j=1}^{n} x_j \mathbf{a}_j.$$

- Let $\mathbf{A}$ and $\mathbf{B}$ be matrices having n rows. For any matrix $\mathbf{M}$ with n columns, we have

$$(\mathbf{A} \mid \mathbf{B}) = (\mathbf{MA} \mid \mathbf{MB}).$$

Definition 8.8

A system $\mathbf{Ax} = \mathbf{b}$ is *homogeneous* iff $\mathbf{b} = 0$; otherwise it is *nonhomogeneous*.

Theorem 8.9

For any matrix, its row space, column space, and rank are identical.

Theorem 8.10

A system $\mathbf{Ax} = \mathbf{b}$ of m linear equations in n unknowns has a solution space of dimension $n - \operatorname{rank}(A)$.

Definition 8.11

A system $\mathbf{Ax} = \mathbf{b}$ of linear equations is *consistent* iff its solution set is nonempty; otherwise it is *inconsistent*.

Theorem 8.12: The Rouché-Capelli Theorem.

A system $\mathbf{A}\mathbf{x} = \mathbf{b}$ is consistent iff $\operatorname{rank}(\mathbf{A}) = \operatorname{rank}(\mathbf{A}|\mathbf{b})$.

Definition 8.13

A matrix is said to be in *reduced row echelon form* iff

- Any row containing a nonzero entry precedes any row in which all the entries are zero (if any).
- The first nonzero entry in each row is the only nonzero entry in its column.
- The first nonzero entry in each row is 1 and it occurs in a column to the right of the first nonzero entry in the preceding row.

- Gaussian elimination.
  - In the forward pass, the augmented matrix is transformed into an upper triangular matrix in which the first nonzero entry of each row is 1 and it occurs in a column to the right of the first nonzero entry of each preceding row.
  - In the backward pass, the upper triangular matrix is transformed into reduced row echelon form by making the first nonzero entry of each row the only nonzero entry of its column.
- Gaussian elimination always reduecs a matrix to its rref form.
- Let $\mathbf{A}$ be an invertible $n \times n$ matrix. Then, for some elementary row matrices $\mathbf{E}_1$ to $\mathbf{E}_p$,
$$\mathbf{E}_p\mathbf{E}_{p-1}\ldots\mathbf{E}_1(\mathbf{A}\,|\,\mathbf{I}_n) = \mathbf{A}^{-1}(\mathbf{A}\,|\,\mathbf{I}_n) = (\mathbf{I}_n\,|\,\mathbf{A}^{-1}).$$
In other words, we can perform Gaussian elimination, so that $(\mathbf{A}\,|\,\mathbf{I}_n) \to (\mathbf{I}_n\,|\,\mathbf{A}^{-1})$.
- Let $\mathbf{A} := \begin{pmatrix}\mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n\end{pmatrix}$ be $m \times n$ matrix, and $\mathbf{A}' := \begin{pmatrix}\mathbf{a}'_1 & \mathbf{a}'_2 & \cdots & \mathbf{a}'_n\end{pmatrix}$ its rref. Then, $\{\mathbf{a}_{k_1}, \mathbf{a}_{k_2}, \ldots, \mathbf{a}_{k_m}\}$ is linearly independent iff $\{\mathbf{a}'_{k_1}, \mathbf{a}'_{k_2}, \ldots, \mathbf{a}'_{k_m}\}$ is. Moreover, the row space of $\mathbf{A}$ and $\mathbf{A}'$ are clearly identical.
- Finding a basis for an intersection of subspaces. Let V and W be subspaces of $\mathbb{F}^n$ generated by the columns of the $n \times m$ matrix $\mathbf{A}$ and $n \times k$ matrix $\mathbf{B}$, respectively. Find a basis for the subspace $V \cap W$.

  1. First notice that $\mathbf{v} \in V \cap W$ iff
  $$\mathbf{v} = \mathbf{A}\mathbf{x}_1 = \mathbf{B}\mathbf{x}_2$$
  for some $\mathbf{x}_2 \in \mathbb{F}^m$ and $\mathbf{x}_2 \in \mathbb{F}^k$. That is,
  $$\begin{pmatrix}\mathbf{A} & \mathbf{B}\end{pmatrix}\begin{pmatrix}\mathbf{x_1} \\ -\mathbf{x_2}\end{pmatrix} = \mathbf{0}.$$
  So, equivalently, we write
  $$\begin{pmatrix}\mathbf{A} & \mathbf{B}\end{pmatrix}\mathbf{y} = \mathbf{0}.$$
  for some $\mathbf{y} \in \mathbb{F}^{m+k}$. As such, by row reducing $\begin{pmatrix}\mathbf{A} & \mathbf{B}\end{pmatrix}$, we find a basis
  $$\beta := \left\{\begin{pmatrix}\mathbf{u_1} \\ \mathbf{u'_1}\end{pmatrix}, \begin{pmatrix}\mathbf{u_2} \\ \mathbf{u'_2}\end{pmatrix}, \ldots, \begin{pmatrix}\mathbf{u_r} \\ \mathbf{u'_r}\end{pmatrix}\right\},$$
  where $\mathbf{u}_i \in \mathbb{F}^m$ and $\mathbf{u}_i \in \mathbb{F}^k$. Now, a generating set for $V \cap W$ is
  $$\Gamma := \{\mathbf{Au_1}, \mathbf{Au_2}, \ldots, \mathbf{Au_r}\}.$$
  Alternatively, another generating set for $V \cap W$ is
  $$\Delta := \{\mathbf{Bu'_1}, \mathbf{Bu'_2}, \ldots, \mathbf{Bu'_r}\}.$$
  From here, it is simple to choose bases $\gamma \subseteq \Gamma$ and $\delta \subseteq \Delta$ for $V \cap W$.
  (Naturally, it holds that $\mathbf{Au_i} + \mathbf{Bu'_i} = 0$.)

2. An alternative method. By row reduction, we can calculate

$$\begin{aligned} r := \dim(V \cap W) &= \dim(U) + \dim(V) - \dim(U+V), \\ &= \operatorname{rank}(\mathbf{A}) + \operatorname{rank}(\mathbf{B}) - \operatorname{rank}\begin{pmatrix}\mathbf{A} & \mathbf{B}\end{pmatrix}, \\ &= \operatorname{rank}\left(\mathbf{A}^\top\right) + \operatorname{rank}\left(\mathbf{B}^\top\right) - \operatorname{rank}\begin{pmatrix}\mathbf{A}^\top \\ \mathbf{B}^\top\end{pmatrix}. \end{aligned}$$

   Then, a basis for $V \cap W$ can be formed by choosing r linearly independent columns of $\begin{pmatrix}\mathbf{A} & \mathbf{B}\end{pmatrix}$, or rows of $\begin{pmatrix}\mathbf{A}^\top \\ \mathbf{B}^\top\end{pmatrix}$.

3. Another alternative, probably the best option! Skip the row reduction of $\mathbf{A}$ and $\mathbf{B}$ in the above method. We just reduce

$$\begin{pmatrix}\mathbf{A} & \mathbf{B}\end{pmatrix} \to \begin{pmatrix}\mathbf{A}' & \mathbf{B}'\end{pmatrix}.$$

   Let $\mathbf{c_i}$ and $\mathbf{c_i'}$ be the ith column of $\begin{pmatrix}\mathbf{A} & \mathbf{B}\end{pmatrix}$ and $\begin{pmatrix}\mathbf{A}' & \mathbf{B}'\end{pmatrix}$, respectively. We compare the columns of A' and B' to find (with relative ease) a basis $\beta' := \left\{\mathbf{c}'_{\mathbf{i_1}}, \mathbf{c}'_{\mathbf{i_2}}, \ldots, \mathbf{c}'_{\mathbf{i_r}}\right\}$ for the intersection of the column spaces of A' and B'. Then, $\beta := \left\{\mathbf{c}_{\mathbf{i_1}}, \mathbf{c}_{\mathbf{i_2}}, \ldots, \mathbf{c}_{\mathbf{i_r}}\right\}$ is a basis for $V \cap W$ (the intersection of the column spaces of A and B).

4. A fourth method for when I learn about orthogonal complements.

Definition 8.14

Let $\mathbf{A} \in \mathrm{M}_{n\times n}(\mathbb{F})$. If $n = 1$, so that $A = (a_11)$, we define $\det(\mathbf{A}) := a_11$. For $n \geq 2$, we define $\det(\mathbf{A})$ recursively as

$$\det(\mathbf{A}) := \sum_{j=1}^{n} (-1)^{1+j} \mathbf{A}_{1j} \cdot \det(\widetilde{\mathbf{A}}_{1j}).$$

THe scalar $\det(\mathbf{A})$ is called the *determinant* of $\mathbf{A}$ and is also denoted by $|\mathbf{A}|$. The scalar

$$(-1)^{i+j} \det(\widetilde{\mathbf{A}}_{1j})$$

is called the cofactor of the entry of $\mathbf{A}$ in row i, column j.

- A matrix $\mathbf{A}$ is invertible iff its determinant is nonzero.

Theorem 8.15

The determinant $\det\colon \mathrm{M}_{n\times n}(\mathbb{F}) \to \mathbb{F}$ is an alternating n-linear function. The former (alternating) means that for $\mathbf{A} \in \mathrm{M}_{n\times n}(\mathbb{F})$ and any $\mathbf{B}$ obtained from $\mathbf{A}$ by interchanging any two rows of $\mathbf{A}$,

$$\det(\mathbf{B}) = -\det(\mathbf{A}).$$

The latter (n-linearity) means that, for any scalar $k \in \mathbb{F}$ and vectors $\mathbf{u}, \mathbf{v}, \mathbf{a}_i \in \mathbb{F}^n$,

$$\det\begin{pmatrix}\mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_{r-1}\mathbf{u} + k\mathbf{v} \\ \mathbf{a}_{r+1} \\ \vdots \\ \mathbf{a}_n\end{pmatrix} = \det\begin{pmatrix}\mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_{r-1}\mathbf{u} \\ \mathbf{a}_{r+1} \\ \vdots \\ \mathbf{a}_n\end{pmatrix} + k\det\begin{pmatrix}\mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_{r-1}\mathbf{v} \\ \mathbf{a}_{r+1} \\ \vdots \\ \mathbf{a}_n\end{pmatrix}.$$

(In fact, it can be shown that $\det\colon \mathrm{M}_{n\times n}(\mathbb{F}) \to \mathbb{F}$ is the *unique* alternating n-linear function, such that $\det(\mathbf{I}) = 1$.)

Corollary 8.16

Let $\mathbf{A} \in \mathrm{M}_{n\times n}(\mathbb{F})$. Then, for any matrix $\mathbf{B}$ obtained by adding a scalar multiple of one row/column of $\mathbf{A}$ to another, $\det(\mathbf{B}) = \det(\mathbf{A})$.

Theorem 8.17

The determinant of a square matrix can be evaluated by cofactor expansion along any row. That is, if $\mathbf{A} \in \mathrm{M}_{n\times n}(\mathbb{F})$, then for any integer $1 \leq i \leq n$,

$$\det(\mathbf{A}) = \sum_{j=1}^{n} (-1)^{i+j} \mathbf{A}_{ij} \cdot \det(\widetilde{\mathbf{A}}_{ij}).$$

Here, $\widetilde{\mathbf{A}}_{ij}$ is the $(n-1) \times (n-1)$ matrix obtained from $\mathbf{A}$ by deleting its ith row and jth column.

Corollary 8.18

The determinant of any triangular matrix is the product of its diagonals.

Theorem 8.19

Let A be an $n \times n$ matrix. Then,

$$\det(\mathbf{A}) = \det(\mathbf{A}^\top).$$

So, the determinant of a square matrix can also be evaluated by cofactor expansion along any column.

Theorem 8.20

Let $\mathbf{A}$ be an invertible $n \times n$ matrix. Then,

$$\mathbf{A}^{-1} = \frac{1}{\det(\mathbf{A})} \operatorname{adj}(A),$$

where $\operatorname{adj}(\mathbf{A})$ is the adjugate/classical adjoint of $\mathbf{A}$. That is, the matrix whose (i,j)th entry is the (j,i)th cofactor $(-1)^{j+i}\det(\widetilde{\mathbf{A}}_{ji})$

Theorem 8.21

For any $\mathbf{A}, \mathbf{B} \in \mathrm{M}_{n\times n}(\mathbb{F})$, we have $\det(\mathbf{AB}) = \det(\mathbf{A}) \cdot \det(\mathbf{B})$.

Definition 8.22

A linear operator T on a finite-dimensional vector space V is called *diagonalisable* iff there is an ordered basis β for V such that $[T]_\beta$ is a diagonal matrix. A square matrix $\mathbf{A}$ is called diagonalisable iff $L_{\mathbf{A}}$ is diagonalisable.

Definition 8.23

Let T be a linear operator on a vector space V. A nonzero vector $\mathbf{v} \in V$ is called an *eigenvector* of T iff there exists a scalar λ such that $T(\mathbf{v}) = \lambda\mathbf{v}$. The scalar λ is called the *eigenvalue* corresponding to the eigenvector $\mathbf{v}$.
Let $\mathbf{A}$ be in $\mathrm{M}_{n\times n}(\mathbb{F})$. A nonzero vector $v \in \mathbb{F}^n$ is called an *eigenvector* of $\mathbf{A}$ iff v is an eigenvector of $L_{\mathbf{A}}$; that is, iff $\mathbf{A}v = \lambda v$ for some scalar λ. The scalar λ is called the eigenvalue of $\mathbf{A}$ corresponding to the eigenvector v.

Definition 8.24

Let $\mathbf{A} \in \mathrm{M}_{n\times n}(\mathbb{F})$. The polynomial $f(t) = \det(\mathbf{A} - \lambda\mathbf{I}_n)$ is called the *characteristic polynomial* of $\mathbf{A}$.

- A matrix $\mathbf{A} \in \mathrm{M}_{n\times n}(\mathbb{F})$ is diagonalizable iff there exists an ordered basis $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ for $\mathbb{F}^n$ consisting of eigenvectors of $\mathbf{A}$, i.e. a eigenbasis. Furthermore, if $\mathbf{Q}$ is the $n \times n$ matrix whose jth column is $\mathbf{v}_j$, then $\mathbf{A} = \mathbf{Q}^{-1}\mathbf{D}\mathbf{Q}$ is a diagonal matrix such that d_{jj} is the eigenvalue of A corresponding to $\mathbf{v}_j$. The matrix $\mathbf{Q}$ is said to *diagonalise* $\mathbf{A}$.
- Hence, we obtain the following procedure to diagonalise a 3×3 matrix $\mathbf{A}$ with three distinct eigenvalues.
  1. Find the eigenvalues λ_1, λ_2, and λ_3 of $\mathbf{A}$. They are just the roots of the characteristic polynomial of $\mathbf{A}$. This can be done using the GC.
  2. Find an eigenvector $\mathbf{v}_j$ corresponding to each eigenvalue λ_j by finding the nullspace of $\mathbf{A} - \lambda_j\mathbf{I}$.
  3. Let $\mathbf{Q} = (\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3)$. Then,
  $$\mathbf{D} := \mathbf{Q}^{-1}A\mathbf{Q}$$
  is a diagonal matrix.

Note

Let $\mathbf{A}$ be a 3×3 real matrix with the eigenvalue λ. Then, the cross product of two nonzero rows/columns of $\mathbf{A} - \lambda\mathbf{I}$ is an eigenvector of $\mathbf{A}$.

Theorem 8.25: The Cayley-Hamilton Theorem.

Let T be a linear operator on a finite dimensional vector space V, and let $f(t)$ be the characteristic polynomial of T. Then $f(T) = T_0$, the zero transformation. That is, T "satisfies" its characteristic equation.

Corollary 8.26: The Cayley-Hamilton Theorem for Matrices.

Let A be an $n \times n$ matrix, and let $f(t)$ be the characteristic polynomial of A. Then, $f(A) = O$, the $n \times n$ zero matrix.

G.C. Skills

Finding eigenvalues of a matrix $\mathbf{A}$ using the GC.

1. `2nd` $\Longrightarrow$ `x`$^{-1}$ (`matrix`) $\Longrightarrow$ Key in the matrix $\mathbf{A} - t\mathbf{I}$, e.g. into `[A]`.
2. Plot `Y`$_1$ = det (`[A]`).
3. `2nd` $\Longrightarrow$ `trace` $\Longrightarrow$ `2:zero` $\Longrightarrow$ Find the roots.

Chapter 9

Numerical Methods

General Information

- The parity of the degree of a real polynomial is the same as that of its number of real roots.
- Let the real polynomial p given by $p(x) = a_{2n}x^{2n} + a_{2n-1}x^{2n-1} + \cdots + a_0$ have coefficients $a_n > 0$ and $a_0 < 0$. Then, it has at least one positive and one negative root.
- Suppose we have some function $f\colon \mathbb{R} \to \mathbb{R}$ with a root α, whose value we want to approximate. There are three ways to obtain this approximation.

  1. Linear interpolation on an interval $[a, b]$ containing α. We let $x_0 \coloneqq b$ and

  $$x_{n+1} \coloneqq \frac{a|f(x_n)| + x_n|f(a)|}{|f(x_n)| + |f(a)|}.$$

     - The sequence $\{x_n\}$ of approximations *always* converges to α.
     - The smaller $f''(x)$ is (i.e. the slower the gradient $f'(x)$ changes) near α, the faster the rate of convergence.
     - Error:

Concave/Gradient	Positive	Negative
Upwards $\bigcap$	underestimation	overestimation
Downwards $\bigcup$	overestimation	underestimation

Table 9.1: Approximation errors when using linear interpolation.

     - See Figure 9.1 for an illustration.

  2. Fixed-point Iteration. First select a function $F\colon \mathbb{R} \to \mathbb{R}$, such that $F(\alpha) = \alpha$, and choose some initial approximation x_0 to α. Then, we recursively define $x_{n+1} \coloneqq F(x_n)$. The desired convergence behavior is for x_n to approach α.

     - Convergence behavior

Behvaior of $\lvert F'(x)\rvert$	Converges?	Rate of convergence
$\lvert F'(x)\rvert < 1$ and is small near α	✓	fast
$\lvert F'(x)\rvert < 1$ but is close to 1 near α	✓	slow
$\lvert F'(x)\rvert \geq 1$ near α	×	-

Table 9.2: Convergence behavior of fixed-point iterations.

     - See Figure 9.2 for an illustration.

3. The Newton-Raphson Method. For a function $f\colon \mathbb{R} \to \mathbb{R}$ with a root α, the Newton-Raphson formula is
$$x_{n+1} \coloneqq x_n - \frac{f(x_n)}{f'(x_n)}.$$
   - The Newton-Raphson method fails in the following cases.
     (a) The gradient at x_0 is too gentle.
     (b) The gradient changes too rapidly.
     (c) The initial approximation x_0 is too far from the root α.
     (d) There is a turning point between the initial approximation x_0 and the root α.
     (e) There is a point of inflection.
   - Error:

Concave/Gradient	Positive	Negative
Upwards $\bigcap$	overestimation	underestimation
Downwards $\bigcup$	underestimation	overestimation

Table 9.3: Approximation errors when using the Newton-Raphson method.

   - See Figure 9.3 for an illustration.

Screw trying to make nice diagonal cells. Pain. Suffering.

Note

At every iteration of linear interpolation, we must ensure that $\alpha \in [a, x_n]$. Otherwise x_n may not approximate α. If $\alpha \notin [a, x_n]$, simply consider $\alpha \in [x_n, b]$ (or any other suitable interval) instead.

G.C. Skills

Linear interpolation: finding an approximation to a root in $[a, b]$ up to n decimal places.

1. $Y_1 = f(x)$,
2. $a \to A$ and $b \to B$,
3. $\dfrac{B|Y_1(A)| + A|Y_1(B)|}{|Y_1(A)| + |Y_1(B)|}$,
4. Ans $\to A$ or B (choose the one that has the opposite sign to Ans),
5. Repeat steps 4 to 5,
6. Terminate this process when the approximations are consistent up to n decimal places.

You can freely enter any function and shift the initial values in the Desmos graphs below!

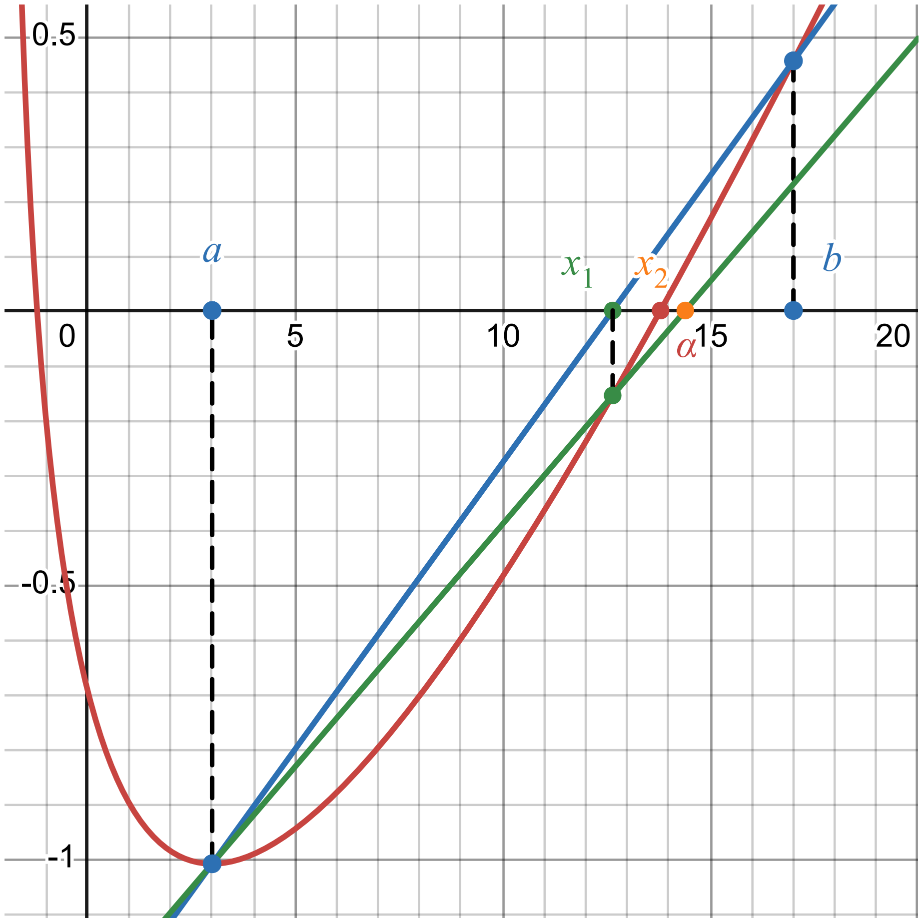


Figure 9.1: An illustration of linear interpolation (Desmos).

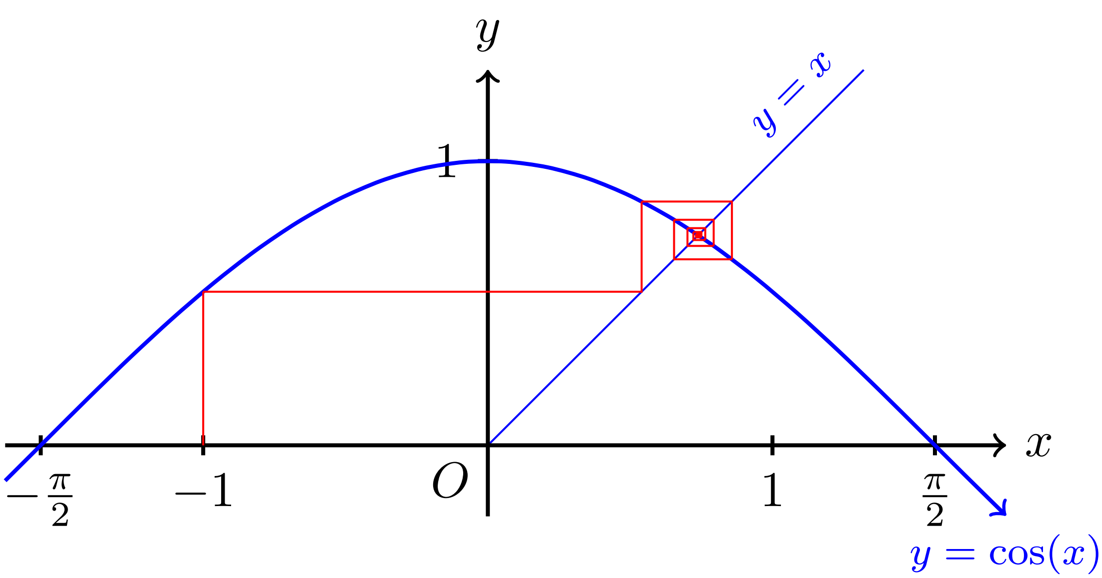


Figure 9.2: An illustration of fixed-point iteration.

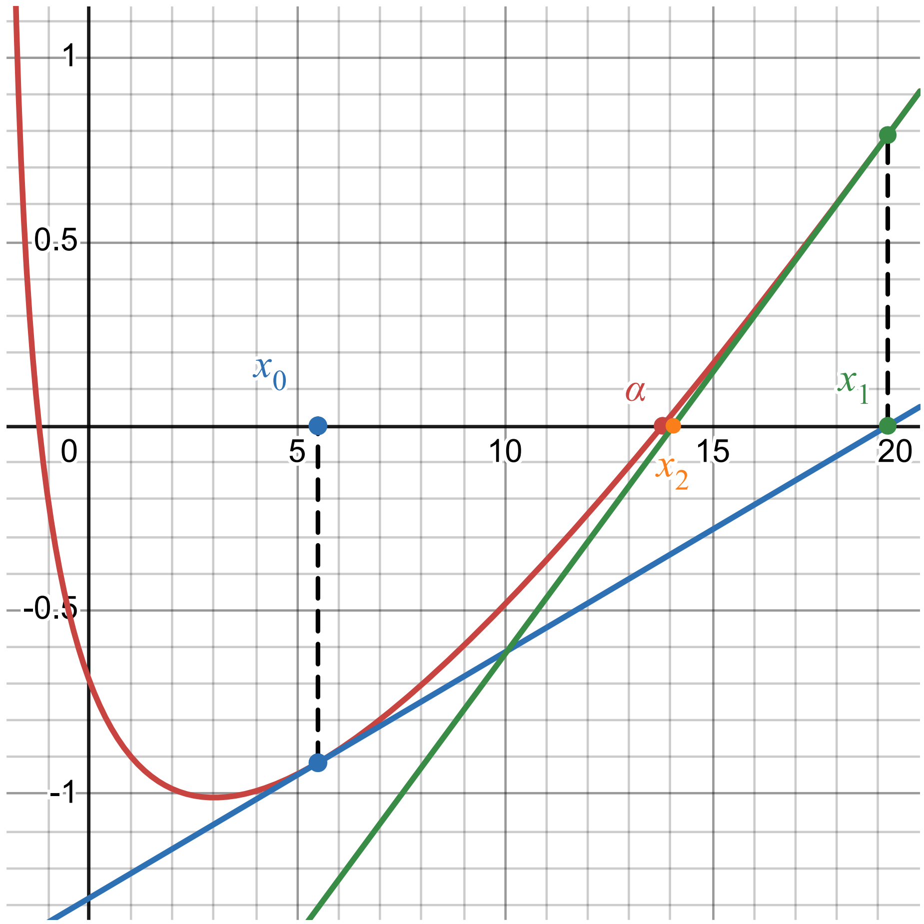


Figure 9.3: An illustration of Newton's Method (Desmos).

Part 2

FMB

Chapter 10

Graphing Techniques

10.1 Graphing 'Familiar' Functions and Asymptotic bois

Definition

1. **Lines of Symmetry**: A *line of symmetry* of a function is a line, such that the function is a reflection of itself about that line.
2. **Horizontal Asymptotes**: A (horizontal) line $g(x) = c$ is the *horizontal asymptote* of the curve $f(x)$ iff $\lim_{x\to\infty} f(x) = c$ (or with $-\infty$ instead of ∞).[a]
3. **Vertical Asymptotes**: A (vertical) line $x = c$ is a *vertical asymptote* of the curve $f(x)$ iff $\lim_{x\to c} f(x) = \infty$ or $-\infty$.
4. **Oblique Asymptotes**: A line $g(x) = mx + c$ — where $m \neq 0$ — is an *oblique asymptote* of the curve $f(x)$ iff $\lim_{x\to\infty}[f(x) - g(x)] = 0$ (or with $-\infty$ instead of ∞).

[a] Otherwise notated by $f(x) \to c$ as $x \to \infty$.

Curve Sketching of Rational Functions

S Stationary points

I Intersection with axes

A Asymptotes

i Know how to sketch the graphs of $y = \dfrac{ax+b}{cx+d}$ and $y = \dfrac{ax^2+bx+c}{dx+e}$.

ii Rectangular Hyperbolas $\left(\text{of the form } y = \dfrac{ax+b}{cx+d}\right)$:

- *Two* asymptotes, namely $x = -\dfrac{d}{c}$ and $y = \dfrac{a}{c}$.
- *Two* lines of symmetry with gradients ± 1 *and* pass through the intersection point of the aforementioned two asymptotes.

iii *If* $n = \deg P = \deg Q$, then

- $y = R(x)$ is the *horizontal* asymptote of $\dfrac{P(x)}{Q(x)} = R(x) + \dfrac{S(x)}{Q(x)}$.
- Equivalently, $y = \dfrac{\text{coeff}_P(x^n)}{\text{coeff}_Q(x^n)}$ is a *horizontal* asymptote.[a]

iv If $\deg P = \deg Q + 1$, then $R(x)$ is an *oblique* asymptote of $\dfrac{P(x)}{Q(x)} = R(x) + \dfrac{S(x)}{Q(x)}$.

v Write down asymptotes and lines of symmetry.[b] If none are present indicate with "No lines of symmetry."

[a] E.g.: $y = \dfrac{1}{15}$ is a horizontal asymptote of $y = \dfrac{1x^2 + 2x - 3}{(5x+1)(3x+2)}$.

[b] E.g.:

Asymptotes: $x = 4$, $y = 20$.

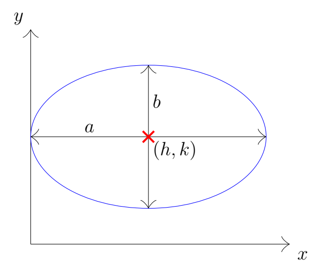
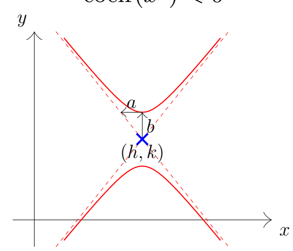
Lines of Symmetry: $y = x + 16$, $y = -x + 24$.

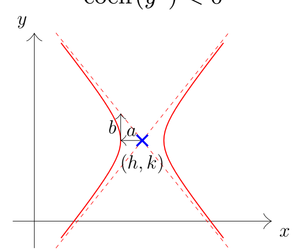
Important Notes

- The discriminant can be very useful.
- Know how to use the G.C. Transfrm app. It allows you to vary the value of some parameter A for a function $f(Ax)$. Use this to graphically find the values of integer k satisfying some conditions.

10.2 Conics

"Tikz is pain, PGFPlots is suffering" — Wise Man.

<table>
<tr><th></th><th>Ellipses</th><th>Hyperbolas</th></tr>
<tr><td>Standard Forms</td><td>$$\frac{(x-h)^2}{a^2}+\frac{(y-k)^2}{b^2}=1$$</td><td>$$\frac{(x-h)^2}{a^2}-\frac{(y-k)^2}{b^2}=1$$ $$\frac{(y-k)^2}{b^2}-\frac{(x-h)^2}{a^2}=1$$</td></tr>
<tr><td>General Equation</td><td>$ax^2+by^2+cx^2+dx+e=0$,
where $\operatorname{sgn}(a)=\operatorname{sgn} b$.</td><td>$ax^2+by^2+cx^2+dex+e=0$,
where $\operatorname{sgn}(a)\neq\operatorname{sgn} b$.</td></tr>
<tr><td>Center</td><td colspan="2">(h,k)</td></tr>
<tr><td>Vertical ‘Radius’
(variables here from standard form!)</td><td colspan="2">b</td></tr>
<tr><td>Horizontal ‘Radius’
(variables here from standard form!)</td><td colspan="2">a</td></tr>
<tr><td>Vertical Vertices
(variables here from standard form!)</td><td colspan="2">$(h,k\pm b)$</td></tr>
<tr><td>Horizontal Vertices
(variables here from standard form!)</td><td colspan="2">$(h\pm a,k)$</td></tr>
<tr><td>Shape</td><td>
</td><td>coeff(x^2) <0


coeff(y^2) <0

</td></tr>
<tr><td>Asymptotes
(No need to rmb!)</td><td>-</td><td>$y=k\pm\frac{b(x-h)}{a}$</td></tr>
<tr><td>Lines of Symmetry</td><td colspan="2">$x=h,\ y=k$</td></tr>
</table>

General Information

- To find asymptote of hyperbolas, just solve
$$\frac{(x-h)^2}{a^2}=\frac{(y-k)^2}{b^2}.$$
- Label vertices or radii, together with the center and asymptotes.

10.3 Parametric Equations

Important Notes

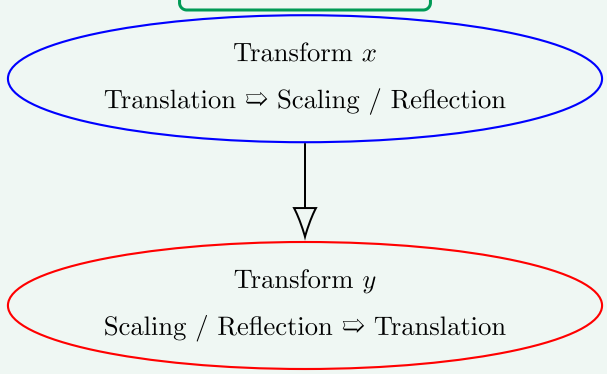
⋆ Check the qns for any *restrictions* on the parameter! And modify that of the G.C.'s accordingly (Tmin & Tmax).

⋆ Vary the t-step or resolution (when using cartesian coordinates) when the graph is oddly jagged.

10.4 Scaling, Translations, and Reflections

<table>
<tr><th colspan="3">Playing With x</th></tr>
<tr><td>Function</td><td>x is replaced with</td><td>(Horizontal) Transformation</td></tr>
<tr><td>$f(x+a)$</td><td>$x+a$</td><td>Translate a units in the positive ($a \leq 0$)
O/R negative x-direction ($a \geq 0$).</td></tr>
<tr><td>$f(-x)$</td><td>$-x$</td><td>Reflect about the y-axis</td></tr>
<tr><td>$f(ax)$</td><td>ax</td><td>Scale parallel to the x-axis by a scale factor of $\frac{1}{a}$ if $a \geq 0$.</td></tr>
<tr><th colspan="3">Playing With $f(x)$</th></tr>
<tr><td colspan="2">Function / Change to $f(x)$</td><td>(Vertical) Transformation</td></tr>
<tr><td colspan="2">$f(x)+a$</td><td>Translate a units in the positive ($a \geq 0$)
O/R negative y-direction ($a \leq 0$).</td></tr>
<tr><td colspan="2">$-f(x)$</td><td>Reflect about the x-axis.</td></tr>
<tr><td colspan="2">$af(x)$</td><td>Scale parallel to the y-axis by scale factor a.</td></tr>
</table>

Important Notes



10.5 $|f(x)|$ and $f(|x|)$

General Information

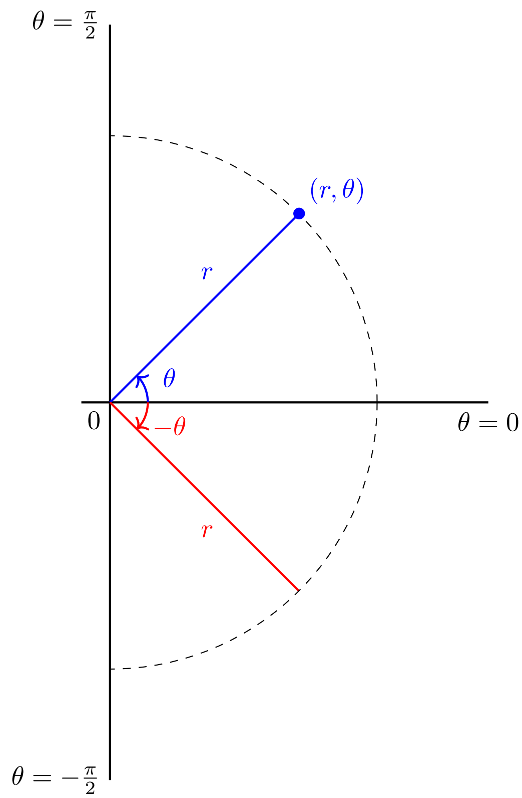
- For $|f(x)|$, simply flip the part of the graph of $f(x)$ that is below the x-axis, to above the x-axis.
- For $f(|x|)$, its graph is symmetric about the x-axis

10.6 $y = \frac{1}{f(x)}$

Behavior of $f(x)$	Behavior of $1/f(x)$
$f(x) > 0$	$\dfrac{1}{f(x)} > 0$
$f(x) < 0$	$\dfrac{1}{f(x)} < 0$
Vertical Asymptote at $x = c$	$\dfrac{1}{f(x)}$ *tends* to 0 * $\frac{1}{f(x)}$ is undefined at $x = c$
$\dfrac{df}{dx} = -\dfrac{d}{dx}\left(\dfrac{1}{f(x)}\right)$ i.e. when $f(x)$ increases, $\frac{1}{f(x)}$ decreases.	
(a, b) is a *minimum* pt	$\left(a, \dfrac{1}{b}\right)$ is a *maximum* pt
(a, b) is a *maximum* pt	$\left(a, \dfrac{1}{b}\right)$ is a *minimum* pt

Chapter 11

Polar Curves



Definition

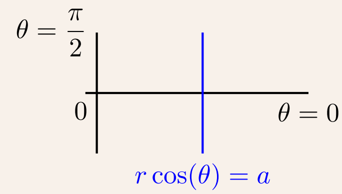
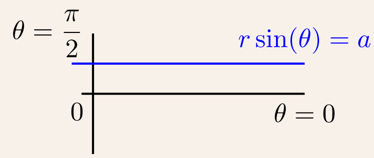
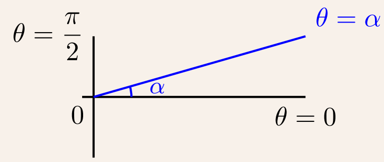
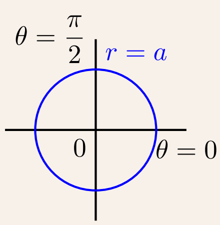
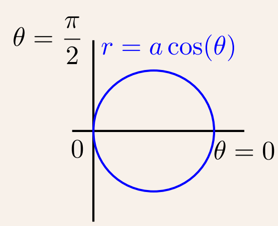
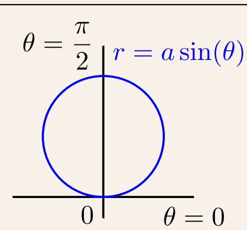
1. The *pole* is the origin, i.e. the point 0.
2. The *initial line / polar axis* is the *half line* $\theta = 0$.

General Information

- Coordinate Conversion

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

- Standard Functions

Polar Equation	Cartesian Equation
	$x = a$
	$y = a$
	$y = x\tan(\alpha)$
	$x^2 + y^2 = a^2$
	$\left(x - \frac{a}{2}\right)^2 + y^2 = \frac{a^2}{4}$
	$x^2 + \left(y - \frac{a}{2}\right)^2 = \frac{a^2}{4}$

- Tangent lines at the pole are obtained by solving $r = 0$.
- Know how to find range of r and θ (given a func/eqn).
- $r = f(\theta)$ is symmetrical about the polar (horizontal) axis iff $f(\theta) = f(-\theta)$.
  - Suppose r is a function of $\cos(n\theta)$[a] *only*. Then, the lines of symmetry are $n\theta = 0, \pi, 2\pi, \ldots$
- $r = f(\theta)$ is symmetrical about the vertical line $\theta = \pi/2$ iff the equation $f(\theta) = f(\pi - \theta)$.
  - Suppose r is a function of $\sin(n\theta)$ *only*. Then, the lines of symmetry are

$$n\theta = \frac{\pi}{2}, \frac{3\pi}{2}, \ldots, \frac{(2m+1)\pi}{2}, \ldots$$

- $r = f(\theta)$ is symmetrical about the pole iff (r, θ) is a point on the curve whenever $(-r, \theta)$ is.
- R-formula may be necessary
- Area of a sector,

$$A = \frac{1}{2}\int_{\alpha}^{\beta} r^2 \, \mathrm{d}\theta,$$

  where $\alpha < \beta$.
- Arc length,

$$\ell = \int_{\theta_1}^{\theta_2} \sqrt{r^2 + \left(\frac{dr}{d\theta}\right)^2} \, \mathrm{d}\theta.$$

[a] E.g.: $r = a\sqrt{(4 + \sin^2(\theta))^2 \cos(\theta)}$

Important Notes

1. r is normally ≥ 0. But, in some questions, it can be negative.
2. No need to fully expand; a final answer such as $(x^2 + y^2)^2 = 3y(x^2 + y^2) - 4y^2$ suffices.
3. Polar curve sketching essentials:
   (a) Shape of curve
   (b) Intersection(s) with ('axial') half lines
   (c) Nothing else *unless* the qns asks for it

   - ☐ Be careful of the sharpness / smoothness of points! Points supposed to be sharp should be sufficiently so, and those supposed to be smooth should be sufficiently rounded / not look sharp.
   - ☐ Check whether the polar axes are tangents at pole. Use this information to draw accurate graphs.
   - ☐ Best to add a small dotted line to show tangentiality at intercepts.
   - ☐ Careful about constants like a in $r = a\sin(\theta)$ for axial intercepts.
   - ☐ No need to state points at the pole unless they are 'axial', i.e. $\theta = 0$, or $\pi/2$, etc.
4. When finding maximum / minimum y values ($dy/dx = 0$), we need to check its nature (1st/2nd deriv test). However, this is unnecessary for max / min r values.
5. For stuff like dy/dx, try to keep it in polar form if possible instead of converting to cartesian form.

6. As usual, be *careful*! E.g. Which values need to be rejected.
7. When reflecting/rotating, a diagram may be useful in finding the necessary angle/expression to replace θ with. E.g.:
   (a) To reflect about $r = \theta$ or $y = x$, we map $(r, \theta) \to (r, \pi/2 - \theta)$.
   (b) Reflect about the half-line $\theta = \pi/2$ is obtained by mapping $(r, \theta) \to (r, \pi - \theta)$.

G.C. Skills

1. To display a nicely scaled polar curve, we use `Zoom fit`, followed by `Zoom square`
2. Simply press alpha trace 1 to get r_1. In fact, this works for the other modes available in the GC as well.
3. We can type

$$\left.\frac{d}{d\theta} r_1\right|_{\theta=\theta}$$

info formulas (like the one for arc length) without having to manually differentiate it!

Chapter 12

Conic Sections

Definition 12.1

Eccentricity, e, is defined as

$$\frac{\text{distance of } P \text{ from focus}}{\text{distance of } P \text{ from directrix}}.$$

General Information

- Shapes associated with the value of e
  - $e = 0$: Circle
  - $0 < e < 1$: Ellipse
  - $e = 1$: Parabola
  - $e > 1$: Hyperbola

Conic	Parabolas		Ellipses		Hyperbolas	
Equation	$x^2 = 4py$	$y^2 = 4px$	$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1,\ a > b$	$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1,\ b > a$	$\frac{x^2}{a^2} - \frac{y^2}{b^2}$	$\frac{y^2}{b^2} - \frac{x^2}{a^2} = 1$
Foci	$(0, p)$	$(p, 0)$	$(\pm c, 0)$	$(0, \pm c)$	$(\pm c, 0)$	$(0, \pm c)$
a, b, c	N.A.		$c^2 = a^2 - b^2$	$c^2 = b^2 - a^2$	$c^2 = a^2 + b^2$	
Directrices	$y = -p$	$x = -p$	$x = \pm\frac{a}{e} = \pm\frac{a^2}{c}$	$y = \pm\frac{b}{e} = \pm\frac{b^2}{c}$	$x = \pm\frac{a}{e} = \pm\frac{a^2}{c}$	$y = \pm\frac{b}{e} = \pm\frac{b^2}{c}$
e	$e = 1$		$0 < e < 1$		$e > 1$	
	N.A.		$e = \frac{c}{a} = \frac{\sqrt{a^2 - b^2}}{a} = \sqrt{1 - \frac{b^2}{a^2}}$	$e = \frac{c}{b} = \frac{\sqrt{b^2 - a^2}}{b} = \sqrt{1 - \frac{a^2}{b^2}}$	$e = \frac{c}{a} = \frac{\sqrt{a^2 + b^2}}{a} = \sqrt{1 + \frac{b^2}{a^2}}$	$e = \frac{c}{b} = \frac{\sqrt{a^2 + b^2}}{b} = \sqrt{1 + \frac{a^2}{b^2}}$
Reflective Property	When light parallel to its axis of symmetry ($x = 0$ or $y = 0$) hits its concave side, the light is reflected to the focus.		For any point P on the ellipse with $a > b$, $PF_1 + PF_2 = 2a$		For any point P on the hyperbola with coeff$(x^2) > 0$, $\lvert PF_1 - PF_2 \rvert = 2a$	

- Polar Form: $x = p$, $x = -p$, $y = p$, or $y = -p$ being the directrix

	Top $r = \dfrac{ep}{1 + e\sin(\theta)}$	
Left $r = \dfrac{ep}{1 - e\cos(\theta)}$		Right $r = \dfrac{ep}{1 + e\cos(\theta)}$
	Bottom $r = \dfrac{ep}{1 - e\sin(\theta)}$	

Definition

- Major / minor axes $\implies$ lengths of longest and shortest diameters respectively.
- Semi-major / semi-minor $\implies$ half of major / minor axes respectively.
- Focal radius $\implies$ distance from point on conic section to focus.

Note

Some possible things to try:

- Using the fact that $PF_1 + PF_2 = 2a$ to do simultaneous equations.
- Converting to polar form (when $e < 1$ so $r \geq 0$) for distances.
- Congruent/Similar triangles.
- Classic use of discriminants.
- Sum and product of roots: Given any polynomial $ax^2 + bx + c$ with the roots α and β,

$$\alpha + \beta = -\frac{b}{a} \qquad \text{and} \qquad \alpha\beta = \frac{c}{a}.$$

Chapter 13

Functions

General Information

1. Horizontal Line Test:
   (a) Fail: Since[a] $y = k$ intersects the graph of $y = f(x)$ more than once, therefore f is not injective.
   (b) Success: Since *any* horizontal line $y = k$ will intersect the graph of $y = g(x)$ *at most once*, so $f(x)$ is one-one.
2. The inverse function, f^{-1}, of a function f exists iff f is one-one.
3. $y = f^{-1}$ is a reflection of $y = f(x)$ about the line $y = x$.
4. The composite function gf exists iff $R_f \subseteq D_g$.
5. $D_{gf} = D_f$ & $R_{gf} = R_g$.
6. Finding the range:
   (a) Graphing method:
   (b) Mapping method, e.g.: $D_f = (0, \infty) \xrightarrow{f} (1, \infty) \xrightarrow{g} (0, \infty) = R_{gf}$

[a] some specific k, e.g. $y = 1/2$

Chapter 14

Permutations and Combinations

Definition 14.1

The terms n *pick* r and n *choose* r respectively denote

$$^nP_r := \frac{n!}{(n-r)!} \qquad \text{and} \qquad \binom{n}{r} = {}^nC_r := \frac{n!}{(n-r)!r!}.$$

General Information

- Addition and multiplication principles
- Know how to 'bundle' objects together so as to calculate the total no. of permutations.
- There are
$$\frac{n!}{n_1!n_2!\cdots n_r!}$$
number of ways to arrange n objects, of which n_i are similar, for each i.

Fact

Intuition: If there are n_1 objects are non-distinct out of n objects, then there are $n_1!$ ways to arrange these objects that results in 'the same' permutation.

- Case-wise considerations/calculations (then summing together the total number of permutations)
- Unordered circular permutations:
There are $n!/n = (n-1)!$ number of ways of arranging n distinct objects in a circle.

Fact

For unordered circular permutations, we do not care if you rotate the seating arrangement, as long as neighbours are preserved for each object. i.e. $(A, B, C, D) \sim (B, C, D, A)$. As a result, each such collection of n permutations reduces down to one. Thus, explaining the division by n.

- Complementary Method, i.e. taking number of arrangements without restriction - number of arrangements with the opposite of that restriction.

Example 14.1

Number of ways two girls *cannot* sit next to each other = number of arrangements *without restriction* − number of arrangements with girls sitting *together*.

- Insertion Method, place down some of your objects and then insert the rest in the gaps.

Example 14.2

Boys sit at table first: 2! ways.

From the 3 gaps, choose 2 for the 2 girls to sit at: 3 ways.

The girls can arrange themselves in 2! ways.

So, total no. of ways is $2! \cdot 3 \cdot 2! = 12$.

- Ordered circular permutations: First calculate the number of unordered permutations, then add the ordering at the end.

Note

Circular arrangements are not the same as row arrangements.
We know that A and B are not considered to be seating together in the row arrangement of (A, C, D, E, B). But, they are seating together in a corresponding row arrangement. The number of row arrangements can be less than, equal to, or more than the number of circular arrangements.

Chapter 15

Vectors

<table>
<tr><th>Lines</th><th>Planes</th></tr>
<tr><td colspan="2">Equivalent Forms</td></tr>
<tr><td>1. Vector Equation:
$$\ell\colon \mathbf{r} = \mathbf{a} + \lambda\mathbf{m},\ \lambda \in \mathbb{R},$$
2. Cartesian Equation:
$$\frac{x - a_1}{m} = \frac{y - a_2}{m_2} = \frac{z - a_3}{m_3.}$$</td><td>1. Vector Equation:
$$\Pi\colon \mathbf{r} = \mathbf{a} + \lambda\mathbf{m_1} + \mu\mathbf{m_2} \text{ where } \lambda, \mu \in \mathbf{R},$$
2. Scalar Product Form:
$$\Pi\colon \mathbf{r} \cdot \mathbf{n} = p$$
where the scalar $p \coloneqq \mathbf{a} \cdot \mathbf{n}$,
3. Cartesian Equation:
$$n_1x + n_2x + n_3z = p$$
where the normal vector $\mathbf{n} \coloneqq \begin{pmatrix} n_1 & n_2 & n_3 \end{pmatrix}^\top$.</td></tr>
<tr><td colspan="2">Foot of Perpendicular</td></tr>
<tr><td>M1: (a) $\overrightarrow{ON} = \mathbf{a} + \lambda\mathbf{m}$,
(b) $\overrightarrow{QN} \cdot \mathbf{m} = 0$, solve for λ,
(c) Substitute λ into (a).
M2: (a) $\overrightarrow{AN} = \left(\overrightarrow{AQ} \cdot \hat{\mathbf{m}}\right)\hat{\mathbf{m}}$,
(b) $\overrightarrow{ON} = \overrightarrow{OA} + \overrightarrow{AN}$.</td><td>(a) $\ell_{NQ}\colon \mathbf{r} = \overrightarrow{OQ} + \lambda n$, where $\lambda \in \mathbb{R}$, and
$$\Pi\colon \mathbf{r} \cdot \mathbf{n} = \mathbf{a} \cdot \mathbf{n},$$
(b) $\left(\overrightarrow{OQ} + \lambda\mathbf{n}\right) \cdot \mathbf{n} = \mathbf{a} \cdot \mathbf{n}$, solve for λ,
(c) $\overrightarrow{ON} + \overrightarrow{OQ} + \lambda\mathbf{n}$.</td></tr>
<tr><td colspan="2">Shortest Distance of Point To Line, QN</td></tr>
</table>

M1: $\left\lVert \overrightarrow{AQ} \times \hat{\mathbf{m}} \right\rVert$. M2: (a) $AN = \left\lVert \overrightarrow{AQ} \cdot \hat{\mathbf{m}} \right\rVert$, (b) Pythagoras' Theorem. M3: Using the foot of perpendicular, find distance QN.	M1: $\left\lVert \overrightarrow{AQ} \cdot \hat{\mathbf{n}} \right\rVert$. M2: Distance of plane to *origin*: If $\Pi\colon \mathbf{r} \cdot \mathbf{n} = p$, then $\dfrac{p}{\lVert \mathbf{n} \rVert}$ is the shortest distance from the origin to the plane Π. *Note:* • If $\dfrac{p}{\lVert \mathbf{n} \rVert} > 0$, then Π 'above' 0. • If $\dfrac{p}{\lVert \mathbf{n} \rVert} < 0$, then Π 'below' 0. M3: Using the foot of perpendicular, then find distance QN.
The Relationship Between Two Lines	The Relationship Between Two Planes
1. Parallel, Non-Intersecting (a) $\mathbf{m_1} \mathbin{/\mkern-5mu/} \mathbf{m_2}$, (b) Solving $\ell_1 = \ell_2$ gives no real solution. 2. Parallel, Coinciding (a) $\mathbf{m_1} \mathbin{/\mkern-5mu/} \mathbf{m_2}$, (b) $\mathbf{a}$ lies in ℓ_1 and ℓ_2. 3. Non-Parallel, Intersecting (a) $\mathbf{m_1}$ not $\mathbin{/\mkern-5mu/}$ $\mathbf{m_2}$, (b) Solve $\ell_1 = \ell_2$ to find intersection. 4. Skew Lines (Non-Parallel, Non-Intersecting) (a) $\mathbf{m_1}$ not $\mathbin{/\mkern-5mu/}$ $\mathbf{m_2}$, (b) Solving $\ell_1 = \ell_2$ gives no real solution.	1. Parallel Planes: Show there exists an $\mathbf{a}$ for which (a) $\mathbf{a} \cdot \mathbf{n_1} = p_1$, (b) $\mathbf{a} \cdot \mathbf{n_2} \neq p_2$. 2. Same Plane: Show there exists an $\mathbf{a}$ for which (a) $\mathbf{a} \cdot \mathbf{n_1} = p_1$, (b) $\mathbf{a} \cdot \mathbf{n_2} = p_2$. 3. Intersect in a line ℓ; To find this line: M1: $\mathbf{n_1} \times \mathbf{n_2}$ gives the direction vector. So find a common point with simultaneous equations. M2: Solving system of linear equations, from the *cartesian* form of the planes, using G.C.
The Relationship Between A Line and A Plane	
1. ℓ lies in Π M1: i. Show $\mathbf{m} \cdot \mathbf{n} = 0$ so $\ell \mathbin{/\mkern-5mu/} \Pi$. ii. Then $\mathbf{a} \cdot \mathbf{n} = p$ tells us ℓ lies in Π. M2: Substitute ℓ into Π and show the system (of lin eqns) is consistent for all λ. 2. $\ell \mathbin{/\mkern-5mu/} \Pi$ but Nonintersecting M1: i. Show $\mathbf{m} \cdot \mathbf{n} = 0$ so $\ell \mathbin{/\mkern-5mu/} \Pi$. ii. Then $\mathbf{a} \cdot \mathbf{n} \neq p$ tells us ℓ and Π are nonintersecting. M2: Substitute ℓ into Π, and show the system (of lin eqns) is inconsistent. 3. Intersect at 1 point M1: Check that $\mathbf{m} \cdot \mathbf{n} \neq 0$. Then, to find the point of intersection of the plane $\Pi\colon \mathbf{r} \cdot \mathbf{n} = p$ with the line $\ell\colon \mathbf{r} = \mathbf{a} + \lambda \mathbf{m}$, we solve for λ using simultaneous equations or G.C.	

The Point of Reflection		
1. Find foot of perpendicular $\overrightarrow{ON}$,	2. Notice $\overrightarrow{OA'} = \overrightarrow{OA} + 2\overrightarrow{AN} = 2\overrightarrow{ON} - \overrightarrow{OA}$.	
Angle Between		
Two Lines	Line and Plane	Two Planes
$\theta = \cos^{-1} \left\| \widehat{\mathbf{m}}_{\mathbf{1}} \cdot \widehat{\mathbf{m}}_{\mathbf{2}} \right\|.$	$\theta = \sin^{-1} \left\| \widehat{\mathbf{m}} \cdot \widehat{\mathbf{n}} \right\|.$	$\theta = \cos^{-1} \left\| \widehat{\mathbf{n}}_{\mathbf{1}} \cdot \widehat{\mathbf{n}}_{\mathbf{2}} \right\|.$

Chapter 16

Probability

General Information

1. Principle of Inclusion and Exclusion for
   (a) Two events:
   $$\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cap B),$$
   (b) Three events:
   $$\mathrm{P}(A \cup B \cup C) = \mathrm{P}(A) + \mathrm{P}(B) + \mathrm{P}(C) - \mathrm{P}(A \cap B) - \mathrm{P}(A \cap C) - \mathrm{P}(B \cap C) + \mathrm{P}(A \cap B \cap C).$$
2. Mutually Exclusive Events:
   $$\begin{aligned} \mathrm{P}(A \cap B) &= 0, \\ \mathrm{P}(A \cup B) &= \mathrm{P}(A) + \mathrm{P}(B). \end{aligned}$$
3. Independent Events:
   $$\begin{aligned} \mathrm{P}(A \mid B) &= \mathrm{P}(A), \\ \mathrm{P}(A \cap B) &= \mathrm{P}(A)\,\mathrm{P}(B). \end{aligned}$$
4. Conditional Probability:
   $$\mathrm{P}(A \mid B) = \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)}.$$
5. Use PnC to help compute stuff faster.
6. When we want to find the greatest and least possible probability (e.g. of $\mathrm{P}(A^{\mathsf{c}} \cap B^{\mathsf{c}} \cap C^{\mathsf{c}})$), it is advisable to draw a Venn diagram and fill in all relevant probabilities.

Example 16.1

There are 6 white balls and 5 black balls. Two are randomly drawn. What is the probability that one is white and the other black?

$$\left(\frac{5}{11}\right)\left(\frac{6}{10}\right) + \left(\frac{6}{11}\right)\left(\frac{5}{10}\right) = \frac{6}{11} \qquad \text{vs} \qquad \frac{\binom{6}{1}\binom{5}{1}}{\binom{11}{2}} = \frac{6}{11}.$$

Chapter 17

Differential Equations

17.1 First Order D.E.s

17.1.1 Elementary Solving Techniques

General Information

- Separable Variables:

$$\frac{dy}{dx} = f(y)g(x), \int \frac{1}{f(y)}\,dy = \int g(x)\,dx.$$

- Integrating Factor:

$$\frac{dy}{dx} + P(x)y = Q(x), \quad \text{let I.F.} = e^{\int P(x)\,dx}$$
$$e^{\int P(x)\,dx}\frac{dy}{dx} + ye^{\int P(x)\,dx}P(x) = Q(x)e^{\int P(x)\,dx},$$
$$ye^{\int P(x)\,dx} = \int Q(x)e^{\int P(x)\,dx}\,dx.$$

17.1.2 Numerical Methods

General Information

- Euler's Method:

$$y_{i+1} = y_i + hf(x_i, y_i), \text{ where } x_n = x_0 + nh.$$

We can present our working directly, as shown in Example 17.1, if there are only one or two iterations. Otherwise, draw the following table.

x	y	$y + hf(x, y)$
x_0	y_0	y_1
$\vdots$	$\vdots$	$\vdots$
x_n	y_n	

Table 17.1: Tabular presentation for Euler's Method.

Example 17.1

Let (step size) $h = 0.25$ and $f(x, y) = \frac{dy}{dx}$:

$$\begin{aligned}\text{By MF26,}\quad y_2 &= \frac{2}{3} + hf\left(0, \frac{2}{3}\right)\\ &= \frac{13}{18}\\ y_3 &= \frac{13}{18} + hf\left(0.25, \frac{13}{18}\right)\\ &= 0.6701865657.\end{aligned}$$

$$\text{Therefore, } y(0.5) \approx 0.670.$$

- Improved Euler's Method:

$$u_{i+1} = y_i + hf(x_i, y_i) \quad \& \quad y_{i+1} = y_i + \frac{h}{2}[f(x_i, y_i) + f(x_{i+1}, u_{i+1})].$$

  Usually only one or two iterations is necessary, so presenting our working directly is sufficient.

- Error:
  1. If $\frac{dy}{dx}$ can be shown to be *increasing* from the calculations of $f(x, y)$, then the curve is *concave upwards*, leading to a *underestimate*.
  2. If $\frac{dy}{dx}$ can be shown to be *decreasing* from the calculations of $f(x, y)$, then the curve is *concave downwards*, leading to a *overestimate*.

Example 17.2

From the computation, *the values of* $\frac{dy}{dx}$ *increases*, i.e. $\frac{d^2y}{dx^2} > 0$, and thus implying the solution curve to be *concave upwards*. Therefore, we have an *underestimation*.

Example 17.3

It is suggested that the estimation in part (ii)[a] can be further improved by reducing the step size. Sketch the solution curve and hence comment on this suggestion.

The solution curve has a *stationary point at* $x = 1.47$, which is between 1 and 2 and also the gradient of the curve is close to zero for x value beyond this stationary point. Thus, when the step size is reduced, *tangent* at point close to this stationary point becomes *almost parallel* to the curve, making *little improvement* to the estimation due to *little difference in* y.

[a]Given the point (1,1), we estimated the value of y(2) using the Improved Euler's Method

Example 17.4

It is found that the approximation obtained in (i) for the y-coordinate where $x = 0.75$ is an underestimation and has a percentage error of 120.633%. Explain why there is such a substantial error.

From gradient values calculated above, we suspect sharp charges in gradient values within the interval (from negative to positive). Yet *Euler's Method*[a] *simply uses a straight line segment* with gradient[b] -4.6409 to estimate the curve for the first iteration, which could have lead to a significant underestimation of the y-value.

[a] We are explaining what it does
[b] Emphasising negative gradient (Show its value)

Note

Compare the relative merits of Euler's Method and the Improved Euler's Method.

Euler's Method: It is computationally simpler.

Improved Euler's Method More accurate as it takes the mean of the initial and next gradient.

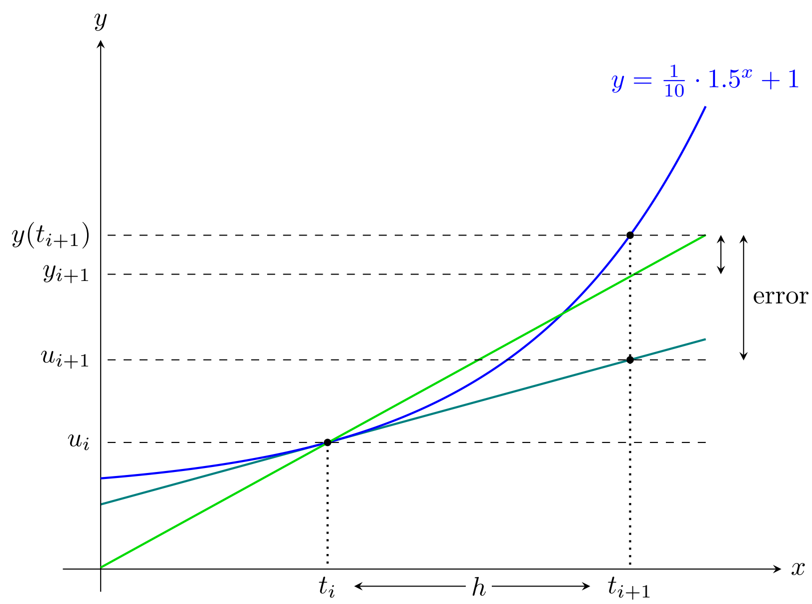


Figure 17.1: An illustration of Euler's Method and the Improved Euler's Method.

17.2 Second Order D.E.

Homogenous	
Roots	Solution y_c
$m_1 \neq m_2$	$y = Ae^{m_1 x} + Be^{m_2 x}$
$m \coloneqq m_1 = m_2$	$y = (Ax + B)e^{mx}$
$m = p \pm qi$	$y = e^{px}(A\cos(qx) + B\sin(qx))$
Non-Homogenous, $c_2 \dfrac{d^2y}{dx^2} + c_1 \dfrac{dy}{dx} + c_0 y = f(x)$	
$y = y_c + y_p$ (C.F. + P.I.)	
$f(x)$	Trial Function for P.I.
Degree n polynomial	$y_p = \sum_{i=0}^{n} a_i x^i$
ke^{ax}	$y_p = ae^{ax}$.
$\alpha\cos(kx) + \beta\sin(kx)$	$y_p = a\cos(kx) + b\sin(kx)$

Note

If y_c and $f(x)$ share some common term, then y_p should be multiplied by x (some least $i \in \mathbb{N}$ times till $x^i y_p$ has no common term with y_c).

Example 17.5

1. If $y_c = A^{-3x}$ and $f(x) = 10e^x$, then $y_p = kxe^x$
2. If $y_c = Ae^x + Be^{-3x}$ and $f(x) = 10e^x$, then $y_p = kxe^x$.
3. If $y_c = Ae^x + Bxe^x + Ce^{-3x}$ and $f(x) = 10e^x$, then $y_p = kx^2e^x$.

17.3 Applications

17.3.1 Exponential Growth

General Information

Let k be the *per-capita growth rate*[a] and $P(t)$ be the population at time t. Then we have the model:

$$\frac{dP}{dt} = kP,$$

with the solution

$$P(t) = P_0e^{kt}.$$

[a]i.e. after accounting for births and deaths.

17.3.2 Logistics Growth

General Information

Let k be the *per-capita growth rate*[a], $P(t)$ be the population at time t, and N be the *carrying capacity* of the system. Then we have the model:

$$\frac{dP}{dt} = kP\left(1 - \frac{P}{N}\right).$$

1. Without solving the logistics equation, we can sketch the solution curve by noting the sign of dP/dt:
   (a) Equilibrium population values occur at $P = 0$ and $P = N$.
   (b) If, for instance $k > 0$,
   $0 < p < N$: $1 - \frac{P}{N} > 0$ so $dP/dt > 0$,
   $P > N$: $1 - \frac{P}{N} < 0$ so $dP/dt < 0$.

   "As t increases, the population of _____ increases to the stable population of _____."

[a]i.e. after accounting for births and deaths.

Example 17.6: Neat trick of letting $A = \pm$constant

$$\begin{aligned} \frac{dP}{dt} &= 3P\left(1 - \frac{P}{200}\right), \\ \int \frac{1}{3P} + \frac{1}{600-3P}\, dP &= \int 1\, dt, \\ \ln\left|\frac{3P}{600-3P}\right| &= 3t + 3c, \\ \frac{3P}{600-3P} &= Ae^{3t}, \text{ where } A = \pm e^{3c}, \\ P &= \frac{200A}{A + e^{-3t}} \end{aligned}$$

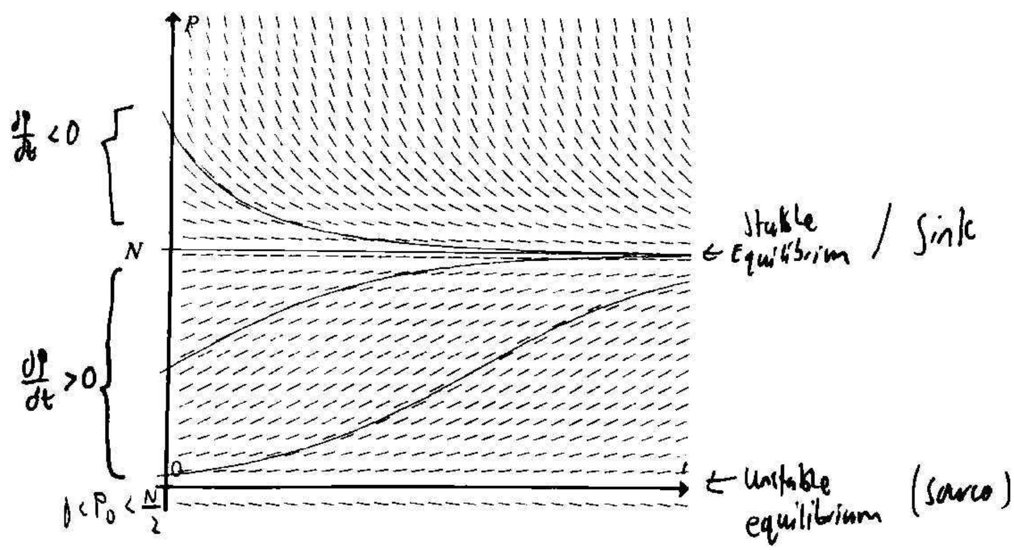


Figure 17.2: Logistics curve

17.3.3 Harvesting

General Information

Let k be the *per-capita growth rate*, $P(t)$ be the population at time t, N be the *carrying capacity* of the system, and H the constant *harvesting rate*. Then we have the model:

$$\frac{dP}{dt} = kP\left(1 - \frac{P}{N}\right) - H.$$

1. Bifurcation Point
   (a) When $0 \le H < \frac{kN}{4}$, there are two equilibrium points, $P = \frac{N}{2} \pm \sqrt{\frac{N^2}{4} - \frac{HN}{k}}$.
   (b) When $H = \frac{kN}{4}$, there is one equilibrium point at $P = \frac{N}{2}$ (the bifurcation point).
   (c) When $H > \frac{kN}{4}$, there is no equilibrium point
2. For Non-Extinction:

$$\frac{N^2}{4} - \frac{HN}{k} \ge 0 \qquad \text{and} \qquad P_0 \ge 450 - \sqrt{\frac{N^2}{4} - \frac{HN}{k}}.$$

17.3.4 Physics

General Information

MUST rmb the forms.

1. Spring System (where $k > 0$ is the spring constant)

$$\ddot{x} + \frac{k}{m}x = 0.$$

   Solution: use R-formula to convert to $A\cos(\omega t + \phi)$ where angular frequency $\omega = \sqrt{k/m}$. Period $T = 2\pi/\omega = 2\pi\sqrt{m/k}$.

2. Simple Pendulum (where ℓ is its length)

$$\ddot{x} + \frac{g}{\ell}x = 0.$$

   Angular frequency $\omega = \sqrt{g/\ell}$ and period $T = 2\pi\sqrt{\ell/g}$.

3. Spring-Mass-Dashpot System (where $c > 0$ is the damping constant)

$$\ddot{x} + \frac{c}{m}\dot{x} + \frac{k}{m}x = 0.$$

   Solution
   (a) Real and Distinct Roots: *Overdamped*
   (b) Identical Real Roots: *Critically Damped*
   (c) Complex Conjugate Roots: *Underdamped*
   "It will oscillate about the equilibrium position with decreasing amplitude."

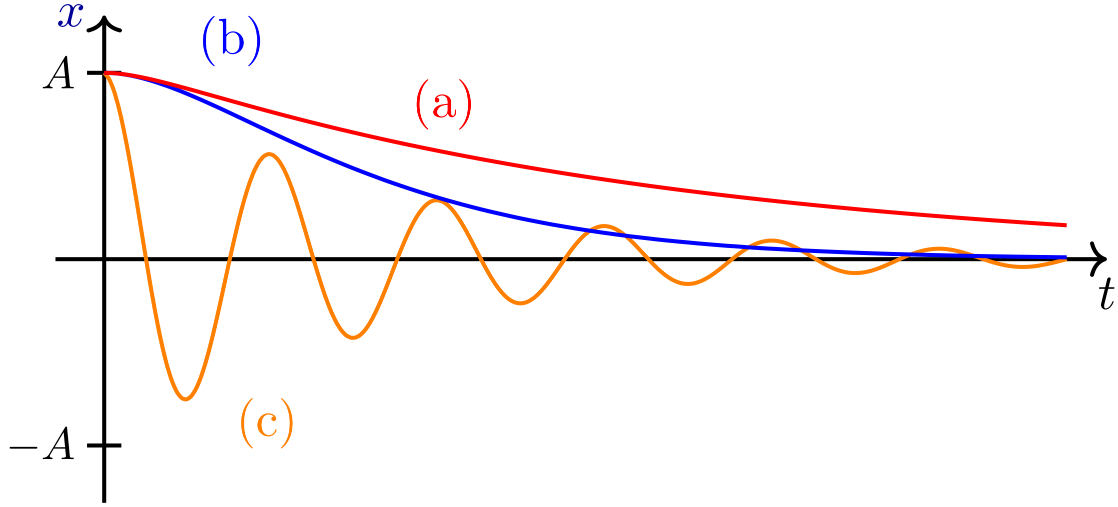


Figure 17.3: Oscillatory behaviors

Chapter 18

Discrete Random Variables

General Information

1. Expectation / Mean,
$$\mathrm{E}(X) \coloneqq \sum_{\text{all } x} x\,\mathrm{P}(X = x).$$
2. Variance
$$\mathrm{Var}(X) \coloneqq E\big(X^2\big) - [E(X)]^2.$$
3. Standard Deviation
$$\sigma \coloneqq \sqrt{\mathrm{Var}(X)}.$$
4. Properties for two *independent* random *variables* X and Y; two *independent observations* X_1 and X_2 of X:
   (a) $\mathrm{E}(aX + bY + c) = a\,\mathrm{E}(X) + b\,\mathrm{E}(Y) + c$,
   (b) $\mathrm{E}(X_1 + X_2) = \mathrm{E}(X_1) + \mathrm{E}(X_2) = 2\,\mathrm{E}(X)$.
   (c) $\mathrm{Var}(aX + bY + c) = a\,\mathrm{Var}(X) + b\,\mathrm{Var}(Y)$,
   (d) $\mathrm{Var}(X_1 + X_2) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) = 2\,\mathrm{Var}(X)$.
5. Probability Distribution Table:

x	1	$\cdots$	n
$\mathrm{P}(X = x)$	$\mathrm{P}(X = 1)$	$\cdots$	$\mathrm{P}(X = n)$

Chapter 19

Special Discrete Random Variables

Definition 19.1

A discrete random variable X which takes all values in $\mathbb{Z}_0^+$ is a *binomial distribution* with probability of success p, denoted by $X \sim \mathrm{B}(n, p)$, iff

$$\mathrm{P}(X = x) = \binom{n}{x} p^x (1-p)^{n-x}.$$

Definition 19.2

A discrete random variable X which takes all values in $\mathbb{Z}^+$ has a *geometric distribution* with probability of success p, denoted by $X \sim \mathrm{Geo}(p)$, iff

$$\mathrm{P}(X = x) = (1-p)^{x-1} p.$$

Note

We can assume $X \sim \mathrm{B}(n, p)$ (or $W \sim \mathrm{Geo}(n, p)$) iff the following three conditions hold

1. The event of a [trial in context] is independent of that of another [trial in context].
2. The probability of each [trial in context] is constant.
3. Each trial has only 2 mutually exclusive outcomes.

Note

Defining random variables:

1. Binomial distribution: Let X be the number of [trial in context], out of [number of trials n in context].
2. Geometric distribution: Let W be the number of [trial in context], up to and including the first [successful trial in context].

Note

Let $W \sim \mathrm{Geo}(p)$, and $q := 1 - p$. Then,

1. $\mathrm{P}(W > m) = q^m$,
2. $\mathrm{P}(X > m + n \,|\, X > n) = \mathrm{P}(X > m) = q^m$,
3. $\mathrm{P}(X < m + n \,|\, X > n) = \mathrm{P}(X < m) = 1 - q^m$.

Definition 19.3

A discrete random variable X which takes all values in $\mathbb{Z}_0^+$ has a *Poisson Distribution* with parameter $\lambda > 0$, denoted by $X \sim \text{Po}(\lambda)$, iff

$$\text{P}(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}.$$

Note

We can assume $Y \sim \text{Po}(\lambda)$ iff the following three conditions hold

1. The event of a [trial in context] is *independent* of that of another [trial in context].
2. The *mean number of occurrences* of [trial in context] is *constant* over an fixed interval of time/space.
3. The *mean number of occurrences* of [trial in context] is *proportional* to the length of the space/time interval.

Note

Additive property of the Poisson distribution: If $U \sim \text{Po}(\mu)$ and $V \sim \text{Po}(\lambda)$ are *independent* variables, then

$$U + V \sim \text{Po}(\mu + \lambda).$$

Note

Defining random variables: Let Y be the number of [event in context], in [space/time interval in context].

General Information

1. Expectation and Mean:

Distribution	Expectation	Variance
$X \sim \text{B}(n,p)$	np	$np(1-p)$
$Y \sim \text{Po}(\lambda)$	λ	
$W \sim \text{Geo}(p)$	p^{-1}	$(1-p)p^{-2}$

2. Use graphing or a table to deal with questions involving inequalities
3. It is helpful to remember the following formulas for when you're asked to derive a formula for mean/mode:
$$\sum_{r=1}^{\infty} r x^{r-1} = (1-x)^{-2} \qquad \text{and} \qquad \sum_{r=1}^{\infty} r^2 x^{r-1} = \frac{1+x}{(1-x)^3}.$$
4. Why is the probability for (b) is smaller than that for (a): The case of (b) is a proper subset of (a).
5. A discrete random variable M can have other probability distributions. In such cases, defining a random variable W having a Binomial/Poisson/Geometric distribution, and then writing M as a function of W may help.

   For example, it may be that $M = W - 1$, or $M = W_1 + W_2$.

Note

When the question asks for the *most likely* number of [event], it is asking for the *mode*.

G.C. Skills

Finding *mode* (e.g. for binomial distributions):

1. Set $Y_1 = \texttt{binompdf}(n, p, X)$.
2. Go to table.
3. Find the value of X for which the highest value of Y_1 occurs.

G.C. Skills

1. `2nd + Vars + 'A'` $\implies$ $\texttt{binompdf}(n, p, x) = \mathrm{P}(X = x)$
2. `2nd + Vars + 'B'` $\implies$ $\texttt{binomcdf}(n, p, x) = \mathrm{P}(X \leq x)$

Note

Let X be the random variable such that $X \sim \mathrm{B}(n, p)$. If $\mathrm{P}(X = n)$ is the *highest probability* that occurs, $X = n$ is the modal value. So, we solve the two inequalities $\mathrm{P}(X = 5) > \mathrm{P}(X = 4)$ and $\mathrm{P}(X = 5) > \mathrm{P}(X = 6)$. This gives the *strictest* range of values that p can take (Fig 17.1).

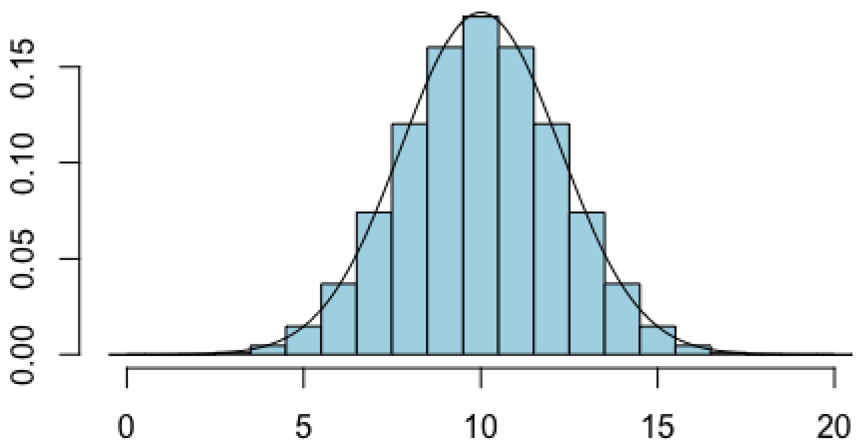


Figure 19.1: In this case, $X = 10$ is the mode.

Example 19.1: 2018 TPJC JC2 H2 MYE P2 8

On average, 3.5% of a certain brand of chocolate turn out misshapen. The chocolates are sold in packets of 25.

(i) State, in context, two assumptions needed for the number of misshapen chocolates in a packet to be well modelled by a binomial distribution.

(ii) Explain why one of the assumptions stated in part (i) may not hold in this context.

Answer:

(i)
1. Each chocolate is *equally likely* (3.) to be misshapen.
2. The event that a chocolate is misshapen is *independent* (2.) of the event that another chocolate is misshapen.

(ii) While on average, the probability that a chocolate is misshapen is 3.5%, it is possible that there are more misshapen chocolates at certain times, possible due to equipment malfunction, which would mean the probability is not constant.

OR

> Misshapen chocolates could be the result of equipment used and as the equipment used would not be the same for the same portion of the chocolate produced, whether a chocolate is misshapen may not be independent of another chocolate being misshapen.

Chapter 20

Continuous Random Variables

General Information

- A function $f\colon \mathbb{R} \to \mathbb{R}$ is a *probability mass function* (pdf) of a continuous random variable X iff f is nonnegative and $\int_{-\infty}^{\infty} f(x)\,dx = 1$.
- For any probability mass function f, we have $\mathrm{P}(a \le X \le b) = \int_a^b f(x)\,dx$. Whether the inequality is strict or nonstrict does not affect the above identity.
- A *mode* of X is any value m such that $f(m)$ is maximum.
- A *cumulative distribution function* (cdf) $F\colon \mathbb{R} \to [0,1]$ of a random variable X is defined by

$$F(x) \coloneqq P(X \le x) = \int_{-\infty}^{x} f(x)\,dx.$$

- When writing out the cdf as a piecewise function, we explicitly write out the range of values for each case. We reserve the use of "otherwise" for pdf's.
- Any cdf is continuous and nondecreasing.
- Let X be a continuous random variable with cdf F. To find the pdf g of any $y(X)$, we first find its cdf, then differentiate. We achieve this by reverse engineering $y(X) \le y$ to find an inequality that relates X with y. E.g. $e^X \le y$ iff $X \le \ln(y)$.
- A *median* of X is any value m such that $\mathrm{P}(X \le m) = F(m) = 1/2$.
- Mean/Expectation:

$$\mu = \mathrm{E}(X) \coloneqq \int_{-\infty}^{\infty} x f(x)\,dx \qquad \text{and} \qquad \mathrm{E}(g(X)) = \int_{-\infty}^{\infty} g(x) f(x)\,dx.$$

- Important property:

$$\mathrm{E}(ag(X) \pm bh(x)) = a\,\mathrm{E}(g(X)) \pm \mathrm{E}(h(X)).$$

- Variance:

$$\mathrm{Var}(X) \coloneqq \mathrm{E}(X^2) - [\mathrm{E}(X)]^2.$$

- Important property:

$$\mathrm{Var}(aX \pm b) = a^2\,\mathrm{Var}(X).$$

Chapter 21

Special Continuous Random Variables

Definition 21.1

A continuous random variable X has a *normal distribution* with mean μ and standard deviation σ, denoted by $X \sim \mathrm{N}(\mu, \sigma^2)$, iff its pdf f is such that

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

General Information

- A normal distribution is symmetrical about the line $x = \mu$. That is

  $$\mathrm{P}(X \leq \mu - \delta) = \mathrm{P}(X \geq \mu + \delta)$$

  for each $\delta > 0$. Note that the mean, median, and mode coincide with μ.

- Properties of the normal distribution. Let X and Y be independent, such that $X \sim \mathrm{N}(\mu, \sigma^2)$ and $Y \sim \mathrm{N}(m, s^2)$. Then, for any $n \in \mathbb{N}$ and x, $y \in \mathbb{R}$,
  - $nX \sim \mathrm{N}(n\mu, n^2\sigma^2)$,
  - $X_1 + X_2 + \cdots + X_n \sim \mathrm{N}(n\mu, n\sigma^2)$,
  - $aX \pm bY \sim \mathrm{N}(a\mu \pm bm, a^2\sigma^2 + b^2 s^2)$.

- At times, the question may be phrased in a misleading manner. Try using some inference to figure out the intended interpretation.

Example 21.1

"The mass of the padding is 30% of the mass of a randomly selected light bulb of mass L. Find the probability that a light bulb with padding has mass c."
Then for any light bulb of mass L_1, the mass of the padding is $0.3L_2$ (and *not* $0.3L_1$). i.e. we are to find $\mathrm{P}(L_1 + 0.3L_2)$.

- A variable $Z \sim \mathrm{N}(0, 1)$ is said to follow the *standard* normal distribution.

  Note: Z is reserved for this purpose.

- Let $X \in \mathrm{N}(\mu, \sigma^2)$. Then, $\frac{X-\mu}{\sigma}$ follows the standard normal distribution.

- What `Tail` do we select for `invNorm`?

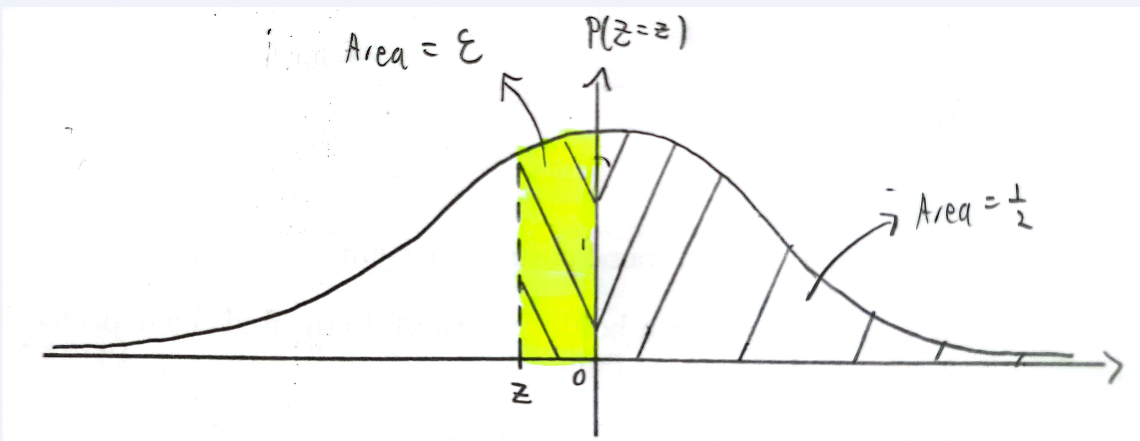
$\mathrm{P}(X < x) = p$	`LEFT`
$\mathrm{P}(-x < X < x) = p$	`CENTER`
$\mathrm{P}(X > x) = p$	`RIGHT`

- When using `invNorm` on an inequality, what should the sign be? For simplicity, we write $\mathscr{L}(p) = \texttt{invNorm}(p, 0, 1, \texttt{RIGHT})$, and $\mathscr{R}(p) = \texttt{invNorm}(p, 0, 1, \texttt{LEFT})$. Then,

$\mathrm{P}(Z > z) \geq p$	$z \leq \mathscr{L}(p)$
$\mathrm{P}(Z > z) \leq p$	$z \geq \mathscr{L}(p)$
$\mathrm{P}(Z < z) \geq p$	$z \geq \mathscr{R}(p)$
$\mathrm{P}(Z < z) \leq p$	$z \leq \mathscr{R}(p)$

Example 21.2

Suppose we want to find the least integer value of m for which $\mathrm{P}(Z > 1 - m) \geq 1/2$.
Then, using `invNorm (RIGHT)`, we infer that $z \leq 0$, *not* $z \geq 0$. An illustration:



Definition 21.2

A continuous random variable X has a *uniform distribution* over the interval (a, b), which is denoted by $X \sim \mathrm{U}(a, b)$, iff its pdf f is such that

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } a < x < b, \\ 0 & \text{otherwise.} \end{cases}$$

Note

Let l and u be the lower and upper quartiles, of a normal distribution $X \sim \mathrm{N}(\mu, \sigma^2)$. i.e. $\mathrm{P}(X < l) = 1/4$ and $\mathrm{P}(X < u) = 3/4$. Then,

$$P\left(\mu - \frac{u-l}{2} < X < \mu + \frac{u-l}{2}\right) = \mathrm{P}(l < X < u) = 1/2.$$

Definition 21.3

A continuous random variable Y has an (negative) exponential distribution, which we denote with $Y \sim \mathrm{Exp}(\lambda)$, iff its pdf g is such that

$$g(Y) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

(An exponential distribution models time between occurrences.)

Note

Let $Y \sim \text{Exp}(\lambda)$, then

$$\text{P}(Y > z + y \,|\, Y > y) = \text{P}(Y > z) \qquad \text{and} \qquad \text{P}(Y < z + y \,|\, Y > y) = \text{P}(Y < z).$$

- Expectation and variance:

Distribution	Expectation	Variance
$X \sim \text{U}(a, b)$	$\dfrac{a+b}{2}$	$\dfrac{(b-a)^2}{12}$
$Y \sim \text{Exp}(\lambda)$	$\dfrac{1}{\lambda}$	$\dfrac{1}{\lambda^2}$

  Note: We need to remember the expectation and variance for the uniform distribution, as it is not provided in the MF26 formula sheet (unlike all other distributions).

- *Warning*: The G.C. tends to incorrectly process an integral if its upper and lower bounds contain ±E99.

- Let T be the time taken between two consecutive arrivals and $\# \sim \text{Po}(\lambda t)$ the number of arrivals in time t. Then,
$$\text{P}(T > t) = \text{P}(\# = 0) = e^{-4t}.$$
  As such, the probability that there is at least one arrival in an interval of time t is
$$\text{P}(T \leq t) = 1 - e^{-4t}.$$

Chapter 22

Sampling and Estimation

Definition 22.1

A sample is a finite subset of the population.

Definition 22.2

A random sample is a sample selected such that each member of the population has an equal probability of being selected into the sample.

Note

State, in context, what it means for the sample to be random.

> It means that every [a member of the population] has an equal probability of being selected into the sample.

Note

Explain why the sample would actually not be random.

> [Contextual reason], so not all the [members of the population] have an equal probability of being selected into the sample.

Definition 22.3

Any statistic T derived from a random sample and used to estimated an unknown population parameter θ is known as an *estimator*. It is an *unbiased* estimator iff $\mathrm{E}(T) = \theta$. If T is unbiased we commonly write $\hat{\theta}$ for T.

General Information

- Either write $\hat{\mu} = \bar{x} = \ldots$ or write out "Unbiased estimate of the population mean μ, $\bar{x} = \ldots$" Same holds for other population parameters θ.
- Estimators you should know:

-

Parameter	Estimator	Unbiased?	Formula
Population Mean μ	Sample Mean $\overline{X}$	✓	$\dfrac{X_1 + X_2 + \cdots + X_n}{n}$
Population Variance σ^2	Sample Variance σ_n^2	×	$\dfrac{\sum (X_i - \overline{X})^2}{n}$ $\dfrac{\sum X_i^2}{n} - \overline{X}^2$
	S^2	✓	$\dfrac{n}{n-1}\sigma_n^2$ $\dfrac{\sum(X_i - \overline{X})^2}{n-1}$ $\dfrac{1}{n-1}\left[\sum X_i^2 - \dfrac{(\sum X_i)^2}{n}\right]$
Population Proportion p	Sample Proportion P_s	✓	$\dfrac{X}{n}$

- Let X be a random variable following *any distribution*, and suppose we have a random sample $X_1, X_2, \ldots, X_n$ of size $n \geq 50$. Then by CLT (Central Limit Theorem), since $n \geq 50$ is large,

$$\overline{X} \sim \mathrm{N}\left(\mu, \frac{\sigma^2}{n}\right) \qquad \text{and} \qquad X_1 + X_2 + \cdots + X_n \sim \mathrm{N}(n\mu, n\sigma^2)$$

  approximately.

- Assumptions when using CLT:
  - The sample is random.
  - Each X_i is independent and identically distributed.

- Suppose $X \sim \mathrm{N}(\mu, \sigma^2)$ is known and we pick a *particular* sample. Then,

Distribution	Is An Approximation?
$\overline{X} \sim \mathrm{N}(\mu, \sigma^2)$	No
$\overline{X} \sim \mathrm{N}(\bar{x}, \sigma^2)$	Yes
$\overline{X} \sim \mathrm{N}(\mu, s^2)$	Yes
$\overline{X} \sim \mathrm{N}(\bar{x}, s^2)$	Yes

  So, if we obtain any of the latter three in solving a question, we must write "$X \sim \mathrm{N}(__, __)$approximately" (even though we knew X *exactly* follows a normal distribution!)

- Pooled estimators. First assume we have two populations, from which we select a random sample of size n_1 and n_2. We let $\overline{X}_1$ and S_1^2 denote the sample mean and unbiased estimator for variance, respectively, for the first sample. Similarly define $\overline{X}_2$ and S_2^2, for the second sample.

Parameter	Unbiased Pooled Estimator
Mean	$\hat{\mu} = \dfrac{n_1\overline{X}_1 + n_2\overline{X}_2}{n_1 + n_2}$
Variance	$S_p^2 = \dfrac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$

The following definition is found in Hogg-McKean-Craig. Similar definitions are also found in Wackerly-Mendenhall-Schaefer and Nitis Mukhopadhyay.

Definition 22.4

Let $X_1, X_2, \ldots, X_n$ be a sample on a random variable X, where X has pdf $f(x;\theta)$, $\theta \in \Omega$. Let $0 < \alpha < 1$ be specified. Let $L = L(X_1, X_2, \ldots, X_n)$ and $U = U((X_1, X_2, \ldots, X_n))$ be two statistics. We say that the interval (L, U) is a $(1-\alpha)100\%$ *confidence interval* for θ iff

$$1 - \alpha = P_\theta[\theta \in (L, U)].$$

That is, the probability that the interval contains θ is $1-\alpha$, which is called the *confidence coefficient* or *confidence level* of the interval.

- We cannot write "a $1-\alpha$ (e.g. 0.95) confidence interval". The $1-\alpha$ must always be expressed as a *percentage*.
- Let $\hat{\theta}$ be a statistic that is normally distributed with mean θ and standard error $\sigma_{\hat{\theta}}$. We see that
  $$\frac{\hat{\theta} - \theta}{\sigma_{\hat{\theta}}} = Z \sim \mathrm{N}(0, 1).$$
  Rewriting $\mathrm{P}(-z_{1-\alpha/2} < Z < z_{1-\alpha/2}) = 1 - \alpha$ gives
  $$\mathrm{P}(\hat{\theta} - z_{1-\alpha/2}\sigma_{\hat{\theta}} < \theta < \hat{\theta} + z_{1-\alpha/2}\sigma_{\hat{\theta}}) = 1 - \alpha.$$
  Hence, a $(1-\alpha)100\%$ confidence interval for θ is
  $$(\hat{\theta} - z_{1-\alpha/2}\sigma_{\hat{\theta}}\,,\ \hat{\theta} + z_{1-\alpha/2}\sigma_{\hat{\theta}}).$$
  (Wackerly-Mendenhall-Schaefer)
- Let $0 < \alpha < 1$ and $X_1, X_2, \ldots, X_n$ be a sample on a random variable X with mean μ, where n is large. Then, an approximate $(1-\alpha)100\%$ confidence interval for μ is
  $$\left(\overline{x} - z_{1-\alpha/2}\frac{s}{\sqrt{n}}\,,\ \overline{x} + z_{1-\alpha/2}\frac{s}{\sqrt{n}}\right).$$
  When the variance σ^2 is known, we can replace s with σ. If the distribution of X is known to be normal, in addition to σ^2 being known exactly, then the confidence interval is exact; it is not just an approximation.

  (Hogg-McKean-Craig)
- Let X be a Bernoulli random variable with probability of success p, where X is 1 or 0 if the outcome is success or failure, respectively. Suppose $X_1, X_2, \ldots, X_n$ is a random sample from the distribution of X, where n is large. Let $\hat{p} = \overline{X}$ be the sample proportion of successes. Then, an approximate $(1-\alpha)100\%$ confidence interval for p is given by
  $$\left(\hat{p} - z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\,,\ \hat{p} + z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right).$$
  (Letting $Y = X_1 + X_2 + \cdots + X_n \sim \mathrm{B}(n, p)$ gives $\hat{p} = Y/n$, which is the presentation used in the school's notes.)

  (Hogg-McKean-Craig)

Note

Standard phrasing for the interpretation of a $(1-\alpha)100\%$ confidence interval (a, b).

The probability that the interval (a, b) contains the true value of the [population mean/proportion in context] is $1 - \alpha$.

Note

Standard phrasing for what is a $(1-\alpha)100\%$ confidence interval for θ?

It is an interval which has probability $1 - \alpha$ of containing the true value of θ.

Note

Standard phrasing for whether mean/proportion in context has likely increased/decreased, when given suitable confidence intervals.

1. There is no conclusive result.

   Since the old and new $(1-\alpha)\%$ confidence intervals overlap, we are unable to conclude whether the [mean/proportion in context] has decreased or not. Hence, it is inconclusive from these figures as to whether the [context (e.g. an awareness campaign)] has been effective.

2. It has likely increased/decreased.

   The old $(1-\alpha)\%$ confidence interval is to the left/right of the new $(1-\alpha)\%$ confidence interval, such that they do not overlap. So, can conclude that the [mean/proportion in context] likely increased/decreased. Hence, these figures suggests that the [context (e.g. an awareness campaign)] has been effective.

Note

Advantage and disadvantage of a $(1-\beta)\%$ confidence interval compared to a $(1-\alpha)\%$ confidence interval, where $\beta < \alpha$.

Advantage:	A $(1-\beta)\%$ CI is more likely to contain the true mean.
Disadvantage:	A $(1-\beta)\%$ CI is less precise (or wider).

Note. Clearly state which is the advantage and disadvantage, as illustrated above.

G.C. Skills

Calculating statistics (i.e. $\bar{x}$, s, etc) by G.C. given data for a sample.

1. Keying in the data: `stat` $\Longrightarrow$ `1:Edit` $\Longrightarrow$ Key in the data into one of the lists $\texttt{L}_i$.

2. Calculating the statistic: `stat` $\Longrightarrow$ `CALC` $\Longrightarrow$ `1-Var Stats` (`List:`$\texttt{L}_i$) $\Longrightarrow$ `Calculate`.

3. Getting the statistic for further calculations: `vars` $\Longrightarrow$ `5:Statistics` $\Longrightarrow$ Select the desired statistic.

G.C. Skills

Calculating the symmetric confidence interval for a normally distributed random variable.

Mean:	`stat` $\Longrightarrow$ `TESTS` $\Longrightarrow$ `7:ZInterval...`
Proportion:	`stat` $\Longrightarrow$ `TESTS` $\Longrightarrow$ `A:1-PropZInt...`

Chapter 23

Statistics: Hypothesis Testing

23.1 General Information

Definition 23.1

The *null hypothesis* H_0 and *alternative hypothesis* H_1 are the hypotheses that we hope to reject and accept, respectively.

General Information

- Without going into details, a *critical region* C is just a set that defines the decision rule / test

$$\text{Reject } H_0 \text{ (Accept } H_1\text{)} \quad \text{if } (X_1, X_2, \ldots, X_n) \in C,$$

  for any random sample $X_1, X_2, \ldots, X_n$ from the distribution of a random variable X.

Definition 23.2

The *significance level* $100\alpha\%$ of a test is the probability of rejecting H_0 when it is in fact true. i.e. $\alpha = \mathrm{P}(H_0 \text{ is rejected} \mid H_0 \text{ is true})$.

Note

Explain, in context, the meaning of 'at the $\alpha\%$ level of significance'.

The probability that [H_1 in context], when actually [H_0 in context], is $\alpha\%$.

Definition 23.3

The *p-value* is the lowest level of significance for which the null hypothesis will be rejected. In other words, for the null hypotheses

$$\text{(a) } \mu < \mu_0, \quad \text{(b) } \mu \neq \mu_0, \quad \text{(c) } \mu > \mu_0,$$

we have

$$\text{(a) } p\text{-value} = \mathrm{P}(Z \leq z_{\text{calc}}), \quad \text{(b) } p\text{-value} = \mathrm{P}(|Z| \geq |z_{\text{calc}}|), \quad \text{(c) } p\text{-value} = \mathrm{P}(Z \geq z_{\text{calc}}).$$

Note

Explain what the p-value means in context.

The p-value is the least level of significance to conclude that [H_1 in context].

- One sample z-test. There are various combination of assumptions for which this test applies. For brevity, we shall avoid restating it, instead directing the reader to table 23.2

  1. Let [X in context] and μ be the population mean.
  2. | | |
     |---|---|
     | Test | $H_0\colon \mu = \mu_0$ |
     | against | $H_1\colon$ (a) $\mu < \mu_0$, (b) $\mu \neq \mu_0$, or (c) $\mu > \mu_0$, |
     | at the $100\alpha\%$ significance level. | |
  3. Under H_0, we have $\overline{X} \sim \mathrm{N}(\mu_0, \hat{\sigma}^2/n)$ approximately. Or, if σ^2 is known exactly, then by CLT $\overline{X} \sim \mathrm{N}(\mu_0, \sigma^2/n)$ approximately.
  4. Test statistic:
  $$Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \sim \mathrm{N}(0,1).$$

  4. Find $z_{1-\alpha}$ or $z_{1-\alpha/2}$, which satisfies
     (a) $\mathrm{P}(Z < z_{1-\alpha}) = \alpha$,
     (b) $\mathrm{P}(-z_{1-\alpha/2} < Z < z_{1-\alpha/2}) = 1 - \alpha$, or
     (c) $\mathrm{P}(Z > z_{1-\alpha})$.
  5. Find the test statistic value
  $$z_{\text{calc}} = \frac{\hat{\mu} - \mu_0}{\sigma/\sqrt{n}}.$$
  6. Reject H_0 iff
     (a) $z_{\text{calc}} < z_{1-\alpha}$,
     (b) $|z_{\text{calc}}| > z_{1-\alpha/2}$, or
     (c) $z_{\text{calc}} > z_{1-\alpha}$.

  4. Find the p-value using GC.
  5. Reject H_0 iff p-value is less than α.

  G.C. Skills

  Calculating the p-value of a sample.

  `stat` $\Longrightarrow$ `TESTS` $\Longrightarrow$ `1:Z-Test...`

  7. Since (a) $z_{\text{calc}} < z_{1-\alpha}$, (b) $|z_{\text{calc}}| > z_{1-\alpha/2}$, (c) $z_{\text{calc}} > z_{1-\alpha}$, or p-value $< \alpha$, we reject H_0. There is sufficient evidence at the significance level $100\alpha\%$ that [H_1 in context].

     Note. For *not* rejecting H_0, simply change to the appropriate inequality (such that z_{calc} is outside the critical region) and write "insufficient" instead of "sufficient".

- If we have a null hypothesis, such as
$$H_0\colon \mu \leq \mu_0 \quad \text{or} \quad H_0\colon \mu \geq \mu_0,$$
we can just use $H_0\colon \mu = \mu_0$ instead.

Note

Explain why there is no need to assume that the distribution of X is normal/know anything about the population distribution of X.

> As the sample size n is large, by the Central Limit Theorem, the sample mean of [random variable X in context] will approximately follow a normal distribution.

Note. Spell "Central Limit Theorem" and "the sample mean" out *in full.* Do not use CLT or $\overline{X}$ for this question.

Definition 23.4

random variable X follows Student's t-distribution with ν degrees of freedom iff its pdf is

$$f(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{t^2}{\nu}\right)^{-\frac{1}{2}(\nu-1)}.$$

This is denoted by $X \sim t(\nu)$.

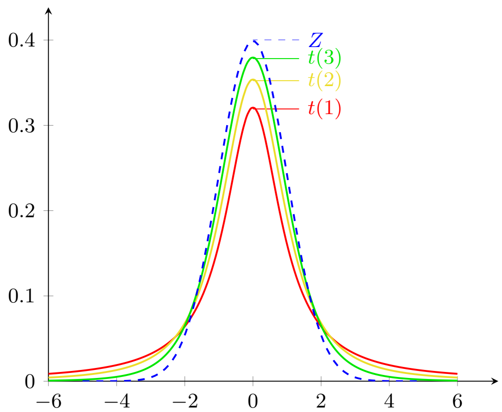


Figure 23.1: Student's t-distribution compared to the standard normal distribution.

- Properties of Student's t-distribution.
    1. It is continuous and symmetric about the vertical axis, i.e. $t = 0$.
    2. From Figure 23.1, we see that the t-distribution has a flatter peak and fatter tails, than the standard normal distribution.
    3. As $\nu \to \infty$, we have $t(\nu) \to \mathrm{N}(0,1)$.
- Let $T \sim t(n-1)$ and $t_{(n-1,1-\alpha/2)}$ be such that $\mathrm{P}\left(-t_{(n-1,1-\alpha/2)} < T < t_{(n-1,1-\alpha/2)}\right) = 1-\alpha$. A $(1-\alpha)100\%$ confidence interval, for the population mean μ of T, is

$$\left(\overline{x} - t_{(n-1,1-\alpha/2)}\frac{s}{\sqrt{n}}\,,\ \overline{x} - t_{(n-1,1-\alpha/2)}\frac{s}{\sqrt{n}}\right).$$

- Suppose we are conducting the following test:

Test	$H_0 : \mu = \mu_0$
against	$H_1 : \mu \neq \mu_0$
at a $100\alpha\%$ significance level.	

  Then, we reject H_0 iff the appropriate symmetric interval (z or t-interval) does *not* contain μ_0.

G.C. Skills

Calculating the symmetric t-confidence interval, for the population mean, of a random variable following Student's t-distribution.

`stat` $\Longrightarrow$ `TESTS` $\Longrightarrow$ `8:TInterval...`

- A one sample t-test. Again, see table 23.2 for the necessary assumptions.

1. Let [X in context], *which we assume to be normally distributed*, and μ be the population mean.
2.

Test	$H_0\colon \mu = \mu_0$	
against	$H_1\colon$	(a) $\mu < \mu_0$, (b) $\mu \neq \mu_0$, or (c) $\mu > \mu_0$,
at the $100\alpha\%$ significance level.		

3. Under H_0, the test statistic
$$T = \frac{\overline{X} - \mu}{s/\sqrt{n}} \sim t(n-1).$$
4. Continue as per usual, calculating the critical region or the p-value.

G.C. Skills

Calculating, for a one sample t-test, the

p-value:	`stat` $\Longrightarrow$ `TESTS` $\Longrightarrow$ `2:T-Test...`
critical region:	`2nd` $\Longrightarrow$ `vars` $\Longrightarrow$ `4:invT(`

Note

In the GC, `invT` is always 'to the `LEFT`'. That is, the output t of

```
invT
area:A
df:ν
Paste
```

is such that $\mathrm{P}(T < t) = A$.

- A two-sample z-test. Again, see table 23.3 for the necessary assumptions.

(i) σ_1 and σ_2 are known, in addition to

(1) X_1 and X_2 being normally distributed, or

(2) both sample sizes, n_1 and n_2, being large.

(ii) σ_1 and σ_2 are unknown, but X_1 and X_2 are normally distributed, and both samples are large (so we can use the fact that a t-distribution approximates to a normal distribution with large sample sizes).

1. Let [X_1, X_2 in context], (which we assume to be normally distributed)[a] and μ be the population mean.
2.

Test	$H_0\colon \mu_1 - \mu_2 = c$	
against	$H_1\colon$	(a) $\mu_1 - \mu_2 < c$, (b) $\mu_1 - \mu_2 = c$, or (c) $\mu_1 - \mu_2 > c$,
at the $100\alpha\%$ significance level.		

3. Under H_0, the test statistic

(i)
$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim \mathrm{N}(0,1).$$

(ii)(1)
$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim \mathrm{N}(0,1).$$

(ii)(2)

$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim \mathrm{N}(0,1) \quad \text{where} \quad s_p^2 = \_\_\_.$$

Case (ii)(2) is used when the population variances coincide, i.e. $\sigma_1 = \sigma_2$.

4. Continue as per usual, calculating the critical region or the p-value.

[a] if applicable

Recall

$$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}.$$

- A two-sample t-test. Again, see table 23.3 for the necessary assumptions.

1. Let [X_1, X_2 in context], *which we assume to be normally distributed*, and μ be the population mean.
2. | | |
   |---|---|
   | Test | $H_0\colon \mu_1 - \mu_2 = c$ |
   | against | $H_1\colon$ (a) $\mu_1 - \mu_2 < c$, (b) $\mu_1 - \mu_2 = c$, or (c) $\mu_1 - \mu_2 > c$, |
   | at the $100\alpha\%$ significance level. | |
3. Under H_0, the test statistic

$$T = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1+n_2-2) \quad \text{where} \quad s_p^2 = \_\_\_.$$

4. Continue as per usual, calculating the critical region or the p-value.

G.C. Skills

Calculating the p-value for a

two-sample z-test: `stat` $\Longrightarrow$ `TESTS` $\Longrightarrow$ `3:2-SampZTest...`
two-sample t-test: `stat` $\Longrightarrow$ `TESTS` $\Longrightarrow$ `4:2-SampTTest...` $\Longrightarrow$ `Pooled:Yes`

- A paired sample t-test. Again, see table 23.3 for the necessary assumptions.

1. Let $D = [X \text{ in context}] - [Y \text{ in context}]$, and μ_D be the population mean.
2. | | |
   |---|---|
   | Test | $H_0\colon \mu_D = \mu_0$ |
   | against | $H_1\colon$ (a) $\mu_D < \mu_0$, (b) $\mu_D \neq \mu_0$, or (c) $\mu_D > \mu_0$, |
   | at the $100\alpha\%$ significance level. | |
3. Under H_0, the test statistic

$$T = \frac{\overline{D} - \mu_0}{s_D/\sqrt{n}} \sim t(n-1).$$

4. $d = x_1 - y_1, x_2 - y_2, \ldots, x_n - y_n$ (insert contextual values) so

$$\bar{d} = \_\_\_ \qquad \text{and} \qquad s_d^2 = \frac{1}{n-1}\left(\sum d^2 - \frac{(\sum d)^2}{n}\right) = \_\_\_.$$

5. Continue as per usual, calculating the critical region or the p-value.

Note

How does the question signal the use of a paired sample t-test? It would be done in one of the following ways:

(a) Via a table

Index	1	2	$\cdots$	n
X	x_1	x_2	$\cdots$	x_n
Y	Y_1	Y_2	$\cdots$	Y_n

Table 23.1: Table containing data of two paired samples.

(b) Stated very explicitly. For instance, "The two sets of data are arranged according to respective students."

Note

Explain why a two-sample t-test would be better than a paired sample t-test.

A two-sample t-test would be better since the *samples are independent*, and we do not know if the data is organised such that each pair comes from the same column.

Note

If it were required to test whether the [population mean μ_1 of X in context] is k, give a reason, whether it would be correct to use the [pooled estimate of variance in context] or an estimate based on the [sample from the distribution of X].

It would be correct to use the estimate of variance based on [sample from the distribution of X], since the test statistic

$$T = \frac{\overline{X} - \mu}{s/\sqrt{n}} \sim t(n-1).$$

involves only the [sample from the distribution of X].

Note

23.2 Summary

Throughout the two tables, we *always* assume that the (both) sample(s) independent and random. Square brackets indicate “and”, while round brackets indicate “or”.

Assumptions/Reasons	Test (Statistic)
[ii] The variance σ^2 is known. [ii](1) Sample size n is large (so CLT applies). [ii](2) Sample size n is small, but we assume X is normally distributed.	One-sample z-test $Z = \dfrac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \sim \mathrm{N}(0, 1)$ (approximately if CLT was used)
[i] The variance σ^2 is unknown. [ii] Sample size n is large. [iii](1) X is known to be normally distributed. (FM) So $t(n-1)$ approximates to $\mathrm{N}(0, 1)$. (H2 Math) No specific reason, just write “approximately.”. [iii](2) X is not known to be normally distributed. (H2 Math Handwaving) CLT applies.	One-sample z-test $Z = \dfrac{\overline{X} - \mu_0}{s/\sqrt{n}} \sim \mathrm{N}(0, 1)$ (approximately)
[i] The variance σ^2 is unknown. [ii] Sample size n is small. [iii] Assume X is normally distributed.	One-sample t-test $T = \dfrac{\overline{X} - \mu_0}{s/\sqrt{n}} \sim t(n-1)$

Table 23.2: Summary table for one-sample hypothesis testing.

Assumptions/Reasons	Test (Statistic)
[i] Both variances σ_1 and σ_2 are known. [ii](1) Both sample sizes n_1 and n_2 are large (so CLT applies). [ii](2) Either sample size n_1 or n_2 is small, but we assume X_1 and X_2 are normally distributed.	Two-sample z-test $Z = \dfrac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim \mathrm{N}(0,1)$ (approximately if CLT was used)
[i] One of the variances σ_1 and σ_2 are unknown. [ii] Both sample sizes n_1 and n_2 are large. [iii] Assume X_1 and X_2 are normally distributed. So $t(n_1 + n_2 - 2)$ approximates to $\mathrm{N}(0,1)$.	Two-sample z-test $Z = \dfrac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim \mathrm{N}(0,1)$ approximately
[i] Both variances σ_1^2 and σ_2^2 coincide. [ii] Assume X_1 and X_2 are normally distributed. (Alt: Both samples come from normal populations.)	Two-sample t-test $T = \dfrac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2)$
[i] Assume that $D_1, D_2, \ldots, D_n$ are normally distributed. [ii] Assume that the data within each pair (X_i, Y_i) are dependent on each other, but pairs (X_i, Y_i) and (X_j, Y_j) are independent of each other, for $i \neq j$.	Paired-sample t-test $T = \dfrac{\overline{D} - \mu_D}{s_D/\sqrt{n}} \sim t(n-1).$

Table 23.3: Summary table for two-sample hypothesis testing.

Chapter 24

Chi-Squared χ^2 Tests

Definition 24.1

A random variable X is said to follow a χ^2-distribution, with degree of freedom ν, iff its probability density function is given by

$$f(x) = \begin{cases} \frac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{(\nu/2)-1} e^{-x/2} & \text{if } x > 0, \\ 0 & \text{otherwise.} \end{cases}$$

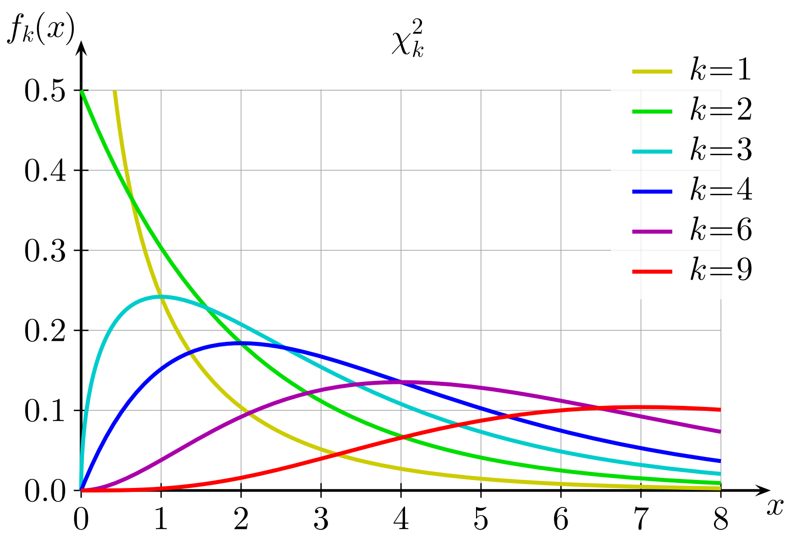


Figure 24.1: Illustration of how the $\chi^2_{(\nu)}$ distribution looks with increasing degree of freedom ν.

General Information

- Properties of chi-squared distributions.
  - $\mathrm{E}(X) = v$ and $\mathrm{Var}(X) = 2\nu$.
  - The $\chi^2_{(\nu)}$ distribution tends to a normal distribution as $\nu \to \infty$.
  - Suppose $Z_i \sim \mathrm{N}(0, 1)$ are independent. Then, $Z_1^2 + \cdots + Z_n^2 \sim \chi^2_{(n)}$.
  - If $X \sim \chi^2_{(\nu)}$ and $Y \sim \chi^2_{(v)}$, then $X + Y \sim \chi^2_{(\nu+v)}$.
- A goodness-of-fit test.
  1. Let [X in context].

2. Test H_0: [X follows the distribution in context]
   against H_1: [X does not follows the distribution in context]
   at the $100\alpha\%$ significance level.

3.

x	x_1	x_2	$\cdots$	x_n
Observed frequency f_i	f_1	f_2	$\cdots$	f_n
Expected frequency e_i	e_1	e_2	$\cdots$	e_n

Table 24.1: Observed and expected frequencies for a goodness-of-fit test

4. Under H_0, the test statistic is

$$\chi^2 = \sum_{i=1}^{n} \frac{(F_i - E_i)^2}{E_i} \sim \chi^2_{(\nu)}.$$

   Here, $n :=$ #classes and $\nu =$ (#classes − #estimated parameters) − 1.

5. Continue as per usual, calculating the critical region $\chi^2_{(\nu)} > \chi^2_{(\nu,1-\alpha)}$ or the p-value.

Note

If X follows a *discrete* normal distribution, we must state it out in words. We cannot write $X \sim \mathrm{N}(\mu, \sigma^2)$ as this would denote that X is a *continuous* random variable.
But if we really have $X \sim \mathrm{N}(n, p)$ (or $X \sim \mathrm{B}(n, p)$, $X \sim \mathrm{Po}(\lambda)$, etc), then we can just denote it as such.

Note

The expected frequency for each of the n classes should be at least 5. If it isn't, we need to combine *just enough* adjacent classes, till they do.

Example 24.1: #estimated parameters = 0

Given $X \sim \mathrm{N}(0, 1)$ (note how the *population parameters* that define the distribution are *known*), the degree of freedom $\nu =$ #estimated parameters $:= n$.

Example 24.2: #estimated parameters = 1

Consider when $X \sim \mathrm{B}(m, p)$, such that the expected frequency for each of the n classes is at least 5, but we do not know the exact value of p. So, we *estimate* it according to the sample given. Then, the degree of freedom is $\nu = n - 1 - 1 = n - 2$.

Example 24.3: #estimated parameters = 2

Similarly, suppose $X \sim \mathrm{N}(\mu, \sigma^2)$, such that the expected frequency of each of the n classes is at least 5, and the true value of μ and σ^2 are unknown. In this case, the degree of freedom $\nu = n - 2 - 1 = n - 3$.

G.C. Skills

- To find the value of $\chi^2_{(\nu,1-\alpha)}$, which satisfies $\mathrm{P}\left(X > \chi^2_{(\nu,1-\alpha)}\right) = \alpha$, we use the table in the MF26 formula sheet (Page 9). Unfortunately, there is no inverse χ^2 function available.
- For the p-value:

`stat` $\Longrightarrow$ `TESTS` $\Longrightarrow$ `D:`χ^2`GOF-Test...`

Tests of independence.

1. Let [X in context].

2. | | |
   |---|---|
   | Test | H_0: [X in context] is independent of [Y in context] |
   | against | H_1: [X in context] is dependent on [Y in context] |
   | at the $100\alpha\%$ significance level. | |

3.

		X				
		x_1	x_2	$\cdots$	x_n	Total
Y	y_1					t_{r_1}
	y_2					t_{r_2}
	$\vdots$					$\vdots$
	y_m					t_{r_m}
	Total	t_{c_1}	t_{c_2}	$\cdots$	t_{c_n}	$\sum t_{r_i} + \sum t_{c_i}$

Table 24.2: *Expected* frequencies for a test of independence.

4. Under H_0, the test statistic is

$$\chi^2 = \sum_{i=1}^{n} \frac{(F_i - E_i)^2}{E_i} \sim \chi^2_{(\nu)}.$$

   Here, $n \coloneqq$ #cols and $\nu = (\text{\#rows} - 1)(\text{\#cols} - 1)$.

5. Continue as per usual, calculating the critical region $\chi^2_{(\nu)} > \chi^2_{(\nu,1-\alpha)}$ or the p-value.

G.C. Skills

Key in the matrix of *observed frequencies* (not Table 1.2 of *expected* frequencies):

$$\texttt{2nd} \Longrightarrow \texttt{x}^{-1} \Longrightarrow \texttt{EDIT} \Longrightarrow \texttt{[A]}.$$

Then, conduct the test for independence:

$$\texttt{stat} \Longrightarrow \texttt{TESTS} \Longrightarrow \texttt{C:}\chi^2\texttt{-Test...}$$

Chapter 25

Correlation and Linear Regression

Note

A good scatter diagram should follow the guidelines below.

- The relative position of each point on the scatter diagram should be clearly shown.
- The range of values for the set of data should be clearly shown by marking out the extreme x and y values on the corresponding axis.
- The axes should be labeled clearly with the variables.

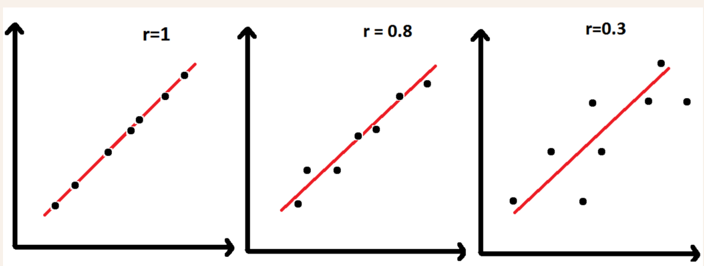
General Information

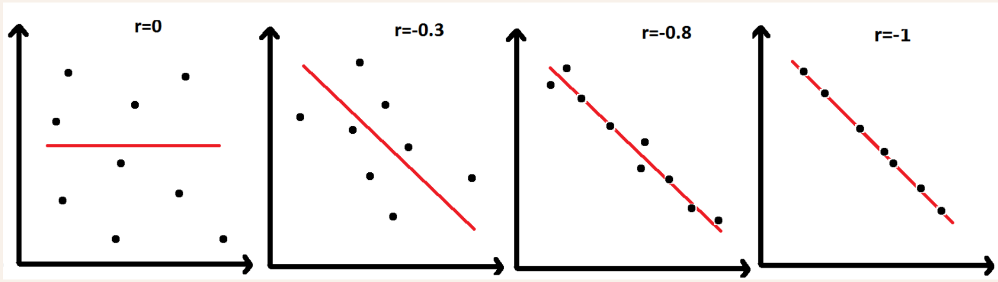
- The Product Moment Correlation Coefficient is a measure of the linear correlation between two variables. It is defined by

$$r = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sqrt{\sum (x-\bar{x})^2 \sum (y-\bar{y})^2}} = \frac{\sum xy - \dfrac{\sum x \sum y}{n}}{\sqrt{\left[\sum x^2 - \dfrac{\left(\sum x\right)^2}{n}\right]\left[\sum y^2 - \dfrac{\left(\sum y\right)^2}{n}\right]}},$$

  which takes on a value from 0 to 1.

- When $r = 0$, there is no linear relationship. But, a nonlinear relationship may be present. Additionally, the regression lines are perpendicular.
- The closer the value of r is to 1 (or -1), the stronger the positive (or negative) linear correlation. Furthermore, the regression lines coincide.





- The regression line of y on x minimises the sum of squares deviation (error) in the y-direction. (i.e. we are assuming x is the independent variable whose values are known exactly.) It is given by

$$y = \bar{y} + b(x - \bar{x}), \qquad \text{where} \qquad b = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sum (x-\bar{x})^2} = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{\left(\sum x\right)^2}{n}}.$$

- The regression lines of y on x and x on y intersect at $(\bar{x}, \bar{y})$.
- Say we are given the value of one variable, and asked to approximate the the value of the other variable. Then, we should always use the line of the *dependent* variable on the *independent*.
- Estimations should not be taken for data outside the range of the sample provided, even if the value of r is close to 1.

Chapter 26

Bibliography

1. Fig 6.1 Trapezium rule (Source)
2. Fig 6.2 Simpson's rule (Source)
3. Fig 7.1 Argand Diagram (Source)
4. Figure 17.1 An illustration of Euler's Method and the Improved Euler's Method `https://tex.stackexchange.com/a/639280`
5. Oscillatory behavior of DEs modelling physical phenomena Fig 15.2 (Source)
6. Fig 17.1 Mode of a binomial distribution (Source)
7. Product moment correlation (Source)
8. Used some inspiration from the beautiful preamble by tearfox and det.uwu from Discord, to make my environments look better.
9. Figure 9.1 on linear interpolation by me (Grass) (Source)
10. Figure 9.2 on fixed-point iteration (Source)
11. Figure 9.3 on the Newton-Raphson method by me (Grass) (Source)
12. Figure 24.1 Chi-squared χ^2 distribution (Source)