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 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 $P(1.8) = 6.336 \nleq 2$.
- 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_∞	Diverges to $\pm\infty$ iff $a \neq 0$ or $d \neq 0$	$S_\infty = \frac{a}{1-r}$ when $\lvert r\rvert < 1$
Prove AP/GP	I Show $u_n - u_{n-1}$ is a constant. II Show $u_n = a + (n-1)d$ explicitly.	I Show $\frac{u_n}{u_{n-1}}$ is constant. 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. Solving RRs in general:

   (a) Continually expand u_n in terms of u_{n-1}, then in terms of u_{n-2}, and so on, till an explicit formula is obtained.

   (b) Alternatively, start from a_1 and iteratively find a_2, a_3, $\ldots$, a_n.

2. Solving $u_{n+1} = au_n + b$:

   (a) Iteration — Apply 4(a) and use the geometric sum formula.

   (b) i. Rewrite the RR as as $u_n - k = a(u_{n-1} - k)$, where $k = \frac{b}{1-a}$. Let $v_n = u_n - k$.

   ii. Then, show that $a = \frac{v_n}{v_{n-1}}$ is a constant, so that $\{v_n\}$ is a geometric progression with first term $v_1 - k$ and common ratio a.

   iii. Now, $v_n = (u_1 - k)a^{n-1}$ and $u_n = v_n + k = (u_1 - k)a^{n-1} + k$.

   (c) ★ Let $u_n = Aa^n + \frac{b}{1-a}$ and solve for the constant A using the initial conditions provided.

3. Solving $u_{n+2} = au_{n+1} + bu_n$:

   i. The characteristic equation is $\lambda^2 - a\lambda - b = 0$, which has roots λ_1 and λ_2.

   ii. The general solution is

$$u_n = \begin{cases} (An + B)\lambda^n & \text{if } \lambda = \lambda_1 = \lambda_2, \\ A\lambda_1^n + B\lambda_2^n & \text{if } \lambda_1 \neq \lambda_2, \\ r^n[A\cos(n\theta) + B\sin(n\theta)] & \text{if } \lambda_1 = re^{i\theta} \text{ and } \lambda_2 = re^{-i\theta} \text{ are not real,} \end{cases}$$

   where A and B are real constants. *Note.* Even if λ_1 and λ_2 are complex, $u_n = A\lambda_1^n + B\lambda_2^n$ is valid.

   iii. Solve for the constants A and B using the initial conditions provided.

Note

We should remember Vieta's Formula. 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 $r_1r_2 = a_0/a_2$.

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 $f(x) > x$.

Example 3.1

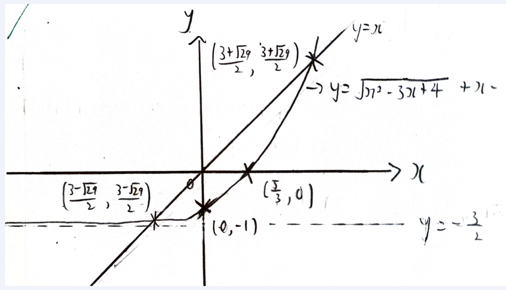


Figure 3.1: The RR $x_{n+1} = \sqrt{x_n^2 - 3x_n + 4} + x_n - 3$.

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

1. Suppose $x_1 \leq \frac{3+\sqrt{29}}{2}$. For $x_1 < \frac{3-\sqrt{29}}{2}$, we see that $f(x) > x$. So x_n increases till $\frac{3-\sqrt{29}}{2}$. In contrast, 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 $x = \frac{3-\sqrt{29}}{2}$. This suggests that $L = \frac{3-\sqrt{29}}{2}$.

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.

Note

When we are asked to comment on the proposed model, it is usually about how suitable it is in modelling the actual situation.

Example 3.2

Investigate what the model, with these values of a and b, predict about the population when $M_0 = 9679$ and $M_0 = 9681$, respectively.

- ✓ The model predicts that the population oscillates between approximately 2420 and 9680 for the first few values of n, but eventually becomes negative at $n = 9$ and $n = 10$. Hence the proposed model is valid only for the first few years and not in the long run.

- × The proposed model is very sensitive to small changes in the initial population of the mammal.

Chapter 4

Induction

General Information

Let $P(n)$ be the statement that "…", where $n \geq n_0$ is an integer.

When $n = n_0$, …

$\implies P(n_0)$ is true.

Assume $P(k)$ is true, for some integer $k \geq n_0$.

Then, …

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

Therefore, since $P(n_0)$ is true and "$P(k)$ is true $\implies P(k+1)$ is true", by Mathematical Induction $P(n)$ is true for all integers $n \geq n_0$.

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:

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

4. Implicit differentiation: $\dfrac{dz}{dx} = \dfrac{dz}{dy} \cdot \dfrac{dy}{dx}$.

Note

Be careful when differentiating implicitly/using the chain rule. Namely, note the power two in the following:

$$\left(f(y)\frac{dy}{dx}\right)' = f'(y)\left(\frac{dy}{dx}\right)^2 + f(y)\frac{d^2y}{dx^2}.$$

More generally, remember to increase the exponent of dy/dx by one with each differentiation.

1. Small angle approximation:
   (a) $\sin(x) \approx x$,
   (b) $\cos(x) \approx 1 - \dfrac{x^2}{2}$,
   (c) $\tan(x) \approx x$.
2. Maclaurin Series:
$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!} x.$$
3. When possible, use the series — of e^x, $\sin(x)$, $\cos(x)$, et cetera — provided in MF26 to find the required series expansion, instead of manually differentiating and applying the Maclaurin series. (Unless otherwise stated, of course.)

Example

Find the series expansion of e^{x+bx^2}. Then, simply write
$$e^{x+bx^2} \approx 1 + (x + bx^2) + \frac{(x+bx^2)^2}{2} \approx 1 + x + \left(b + \frac{1}{2}\right) x^2.$$

Chapter 6

Integration Techniques

6.1 Basic Integration (IBS, IBP, etc)

General Information

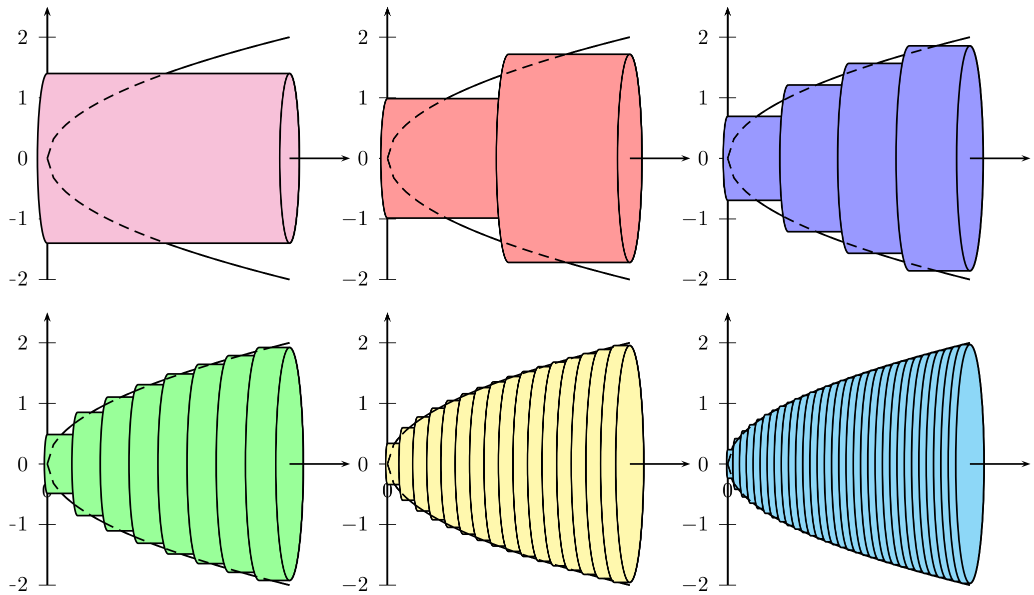
1. Factor Formulae (remember these):
   (a) $\sin(mx)\cos(nx) = \frac{1}{2}[\sin((m+n)x) + \sin((m-n)x)]$,
   (b) $\cos(mx)\cos(nx) = \frac{1}{2}[\cos((m+n)x) + \cos((m-n)x)]$,
   (c) $\sin(mx)\sin(nx) = -\frac{1}{2}[\cos((m+n)x) - \cos((m-n)x)]$.
2. Common classes of integrals:
   (a) Apply partial fractions:
   $$\int \frac{f(x)}{g(x)}\,dx.$$
   (b) Split $px+q$, then complete the square:
   $$\int \frac{px+q}{\sqrt{ax^2+bx+c}}\,dx \quad \text{or} \quad \int \frac{px+q}{ax^2+bx+c}\,dx$$
3. Partial fractions. First ensure the numerator has a smaller degree than that of the denominator, otherwise do long division first. Then,
   $$\text{(i)}\quad \frac{P(x)}{(ax+b)(cx+d)} = \frac{A}{ax+b} + \frac{B}{cx+d}$$
   $$\text{(ii)}\quad \frac{P(x)}{(ax+b)^2} = \frac{A}{ax+b} + \frac{B}{(ax+b)^2}$$
   $$\text{(iii)}\quad \frac{P(x)}{(ax+b)(cx^2+dx+e)} = \frac{A}{ax+b} + \frac{Bx+C}{cx^2+de+e}$$
   where cx^2+de+e is irreducible, i.e. has non-real roots.
4. Integration by Substitution:
   $$\int f(x)\,dx = \int f(x)\frac{dx}{du}\,du.$$
5. Use Pythagoras' Theorem / draw a right-angled triangle to help with trig conversions, e.g.:
   $$\tan(\theta) \qquad \text{to} \qquad \frac{x+1}{\sqrt{2-(x+1)^2}}.$$

6. Integration by Parts:

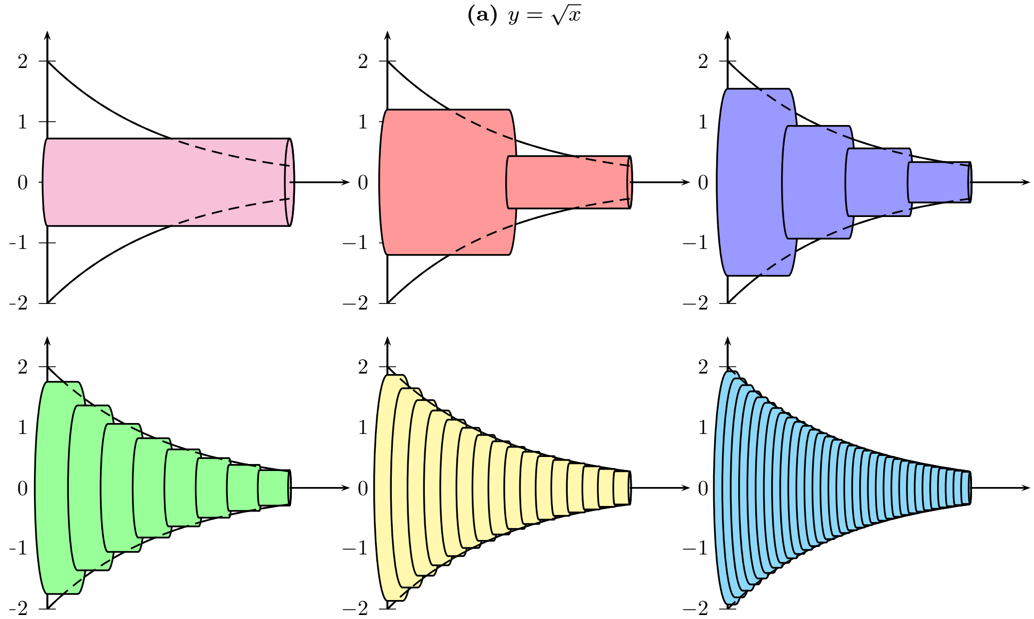
$$\text{Let}\quad u = \_\_\_ \qquad v = \_\_\_ \qquad\qquad \int u\left(\frac{dv}{dx}\right)dx = uv - \int v\left(\frac{du}{dx}\right)dx.$$

$$\frac{du}{dx} = \_\_\_ \qquad \frac{dv}{dx} = \_\_\_$$

6.2 Areas & Volumes



(a) $y = \sqrt{x}$



(b) $y = 2e^{-x/2}$

Figure 6.1: The disc method.

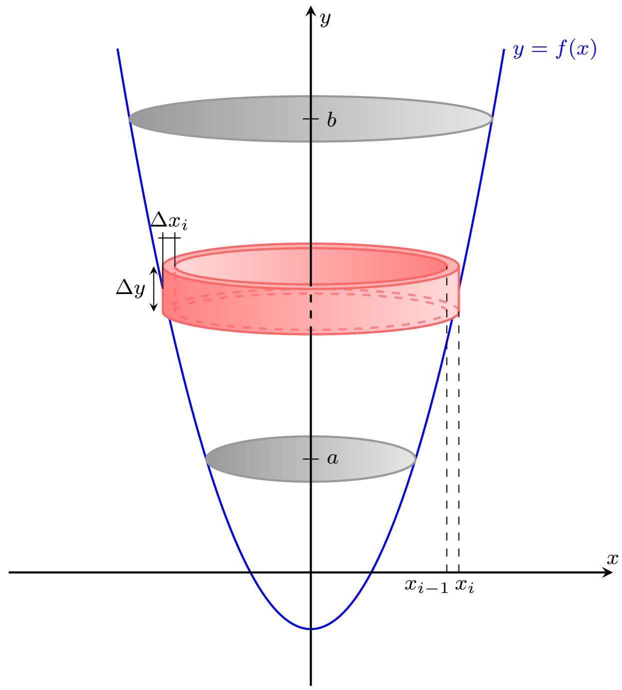


Figure 6.2: The shell method.

General Information

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

   (a) The disc method (Figure 6.1):

$$\int_{x_1}^{x_2} \pi y^2 \, dx = \int_{x=x_1}^{x=x_2} \pi y^2 \frac{dx}{dt} \, dt.$$

   (b) The shell method (Figure 6.2):

$$\int_{y_1}^{y_2} 2\pi y x \, dy.$$

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:

$$\int_{x_1}^{x_2} 2\pi y \sqrt{\left(\frac{dy}{dx}\right)^2 + 1} \, dx = \int_{y_1}^{y_2} 2\pi y \sqrt{\left(\frac{dx}{dy}\right)^2 + 1} \, dy = \int_{t_1}^{t_2} 2\pi y \sqrt{\left(\frac{dy}{dt}\right)^2 + \left(\frac{dx}{dt}\right)^2} \, dt.$$

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

4. Area enclosed by a polar curve:

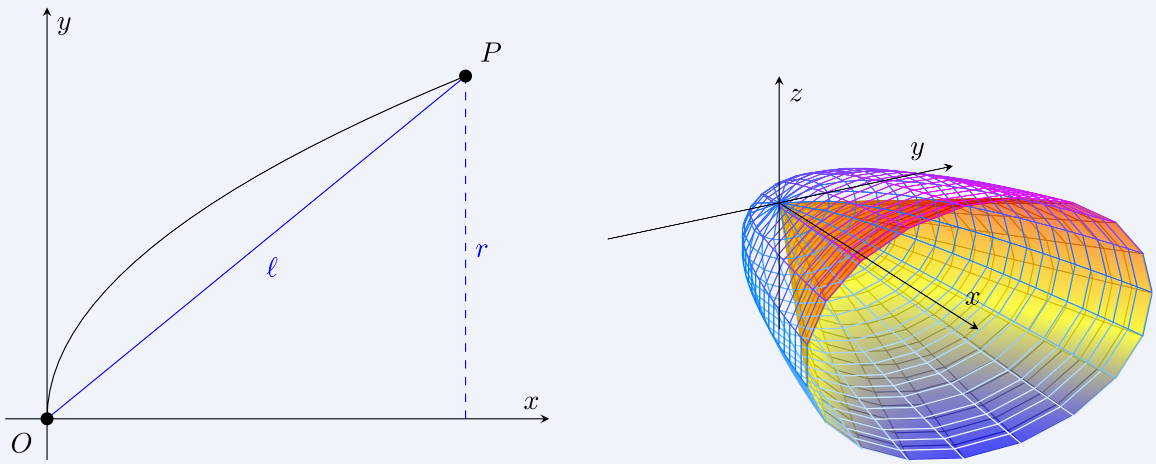
$$\int_{\alpha}^{\beta} \frac{1}{2} r^2 \, d\theta.$$

Example 6.1: Geometrical interpretation of an approximate area

An parabolic arc has equation $x = at^2$, $y = 2at$ for $0 \le t \le p$. When rotated through 2π radians about the x-axis, the surface area of revolution formed is $A = 8\pi a^2/3\left[(p^2+1)^{3/2} - 1\right]$ units^2. Consider the approximation to A when p is sufficiently small, such that powers above p^2 can be ignored. Interpret this approximate area geometrically.

$$A \approx 8/3\pi a^2[1 + 3/2p^2 - 1] = 4\pi a^2 p^2 \text{ units}^2 \ (= \pi(2ap)(2ap)).$$

Geometric interpretation: The approximate area is the surface area of a cone with radius $2ap$ and approximate slant height $2ap$.



(a) 2d illustration. **(b)** 3d illustration.

Figure 6.3

Note. The slant height $\ell = \sqrt{(ap^2)^2 + (2ap)^2} = \sqrt{a^2p^4 + 4a^2p^2} \approx \sqrt{a^2 \cdot 0 + 4a^2p^2} = 2ap$.

6.3 Numerical Methods

6.3.1 Trapezium Rule

General Information

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

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

2. Illustration

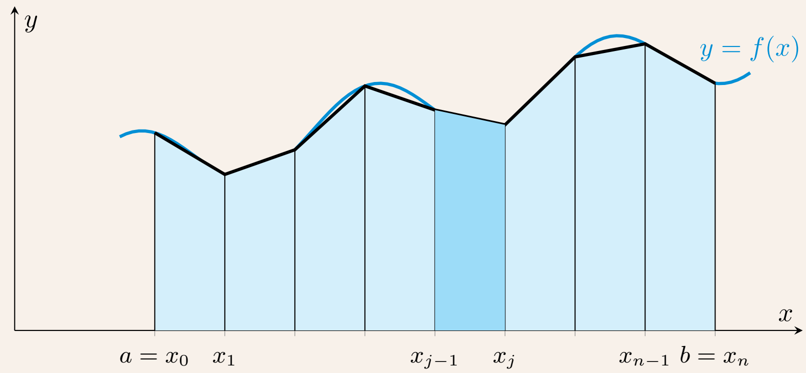


Figure 6.4: Trapezium rule.

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.

Note

State and explain whether the trapezium rule gives an underestimate or an overestimate of the area of under the curve $y = f(x)$ from $x = x_0$ to $x = x_1$.

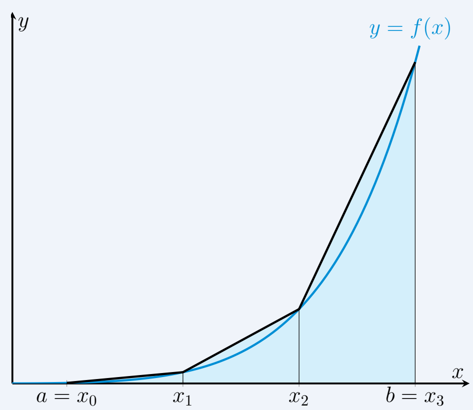


Figure 6.5: The trapezium rule results in an overestimate for the area under a curve that is concave upwards.

From the diagram and the fact that $\frac{d^2y}{dx^2} =$ ___ > 0 for all $x \in [a, b]$, the curve is concave upwards. As such, the trapeziums contain the graph of $y = f(x)$, for $x \in [a, b]$. So, the trapezium rule gives a overestimate.

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

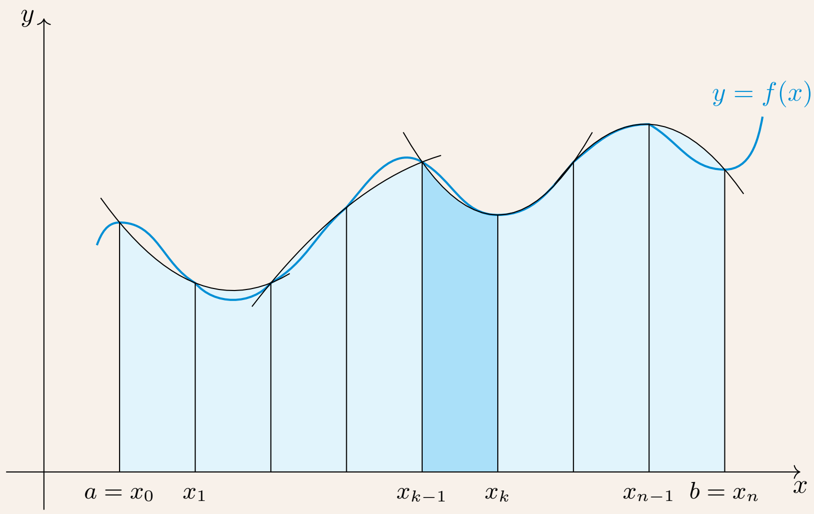


Figure 6.6: Simpson's rule.

3. Simpson's Rule is exact for all polynomials of degree three and below.

Example 6.2

Explain, with the aid of a suitable diagram, why Simpson's Rule does not give a good approximation of the value of $I = \int_a^b f(x)\,dx$ and suggest how the method can be improved to give a better approximation.

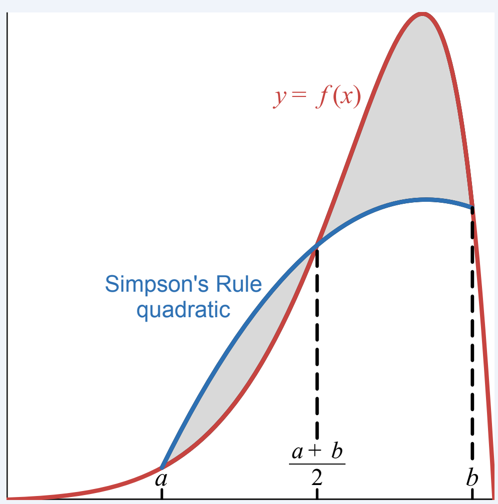


Figure 6.7: (Desmos)

We see that there is a large discrepancy between the area under the actual curve $y = f(x)$ and the approximated quadratic curve for the interval $a \leq x \leq b$, which is caused by [the existence of both a turning point and a point of inflection]. Hence, Simpson's Rule using n ordinates will not provide a good approximation in this case.

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."

6.4 Reduction formulas

General Information

Let $I_n := \int_a^b f(x)\,dx$. To find a reduction formula for I_n, we use integration by parts on I_n.
Note. Integrating I_{n-1}, I_{n+1}, etc (by parts) typically does not produce nice results.

Sometimes, some algebraic tricks are needed to reduce an integral into the desired form. Try to spot these, before attempting another round of integration by parts.

Example 6.3: Rewriting x^2 as $(4-x^2)-4$

Using

$$\begin{aligned} u &= \frac{1}{(4-x^2)^n} & v &= 1 \\ \frac{du}{dx} &= \frac{2nx}{(4-x^2)^{n+1}} & \frac{dv}{dx} &= x \end{aligned}$$

we evaluate $A_n := \int_0^{1/2} \frac{1}{(4-x^2)^n}\,dx$:

$$\begin{aligned} A_n &= \frac{1}{2}\left(\frac{4}{15}\right)^n - 2n\int_0^{1/2} \frac{x^2}{(4-x^2)^{n+1}}\,dx \\ &= \frac{1}{2}\left(\frac{4}{15}\right)^n + 2n\int_0^{1/2} \frac{(4-x^2)-4}{(4-x^2)^{n+1}}\,dx \\ &= \frac{1}{2}\left(\frac{4}{15}\right)^n + 2n\int_0^{1/2} \frac{4-x^2}{(4-x^2)^n} - 4\cdot\frac{1}{(4-x^2)^{n+1}}\,dx \\ &= \frac{1}{2}\left(\frac{4}{15}\right)^n + 2n(A_n - 4A_{n+1}). \end{aligned}$$

Therefore, $(2n-1)A_n + 1/2(4/15)^n = 8nA_{n+1}$.

Chapter 7

Complex Numbers

7.1 Complex Number I

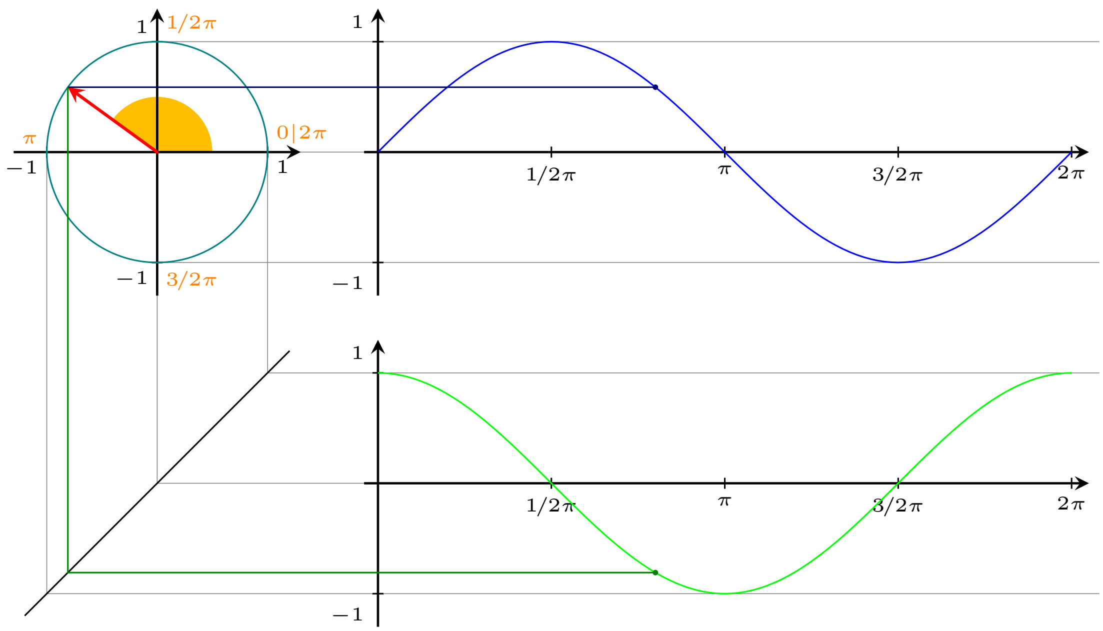


Figure 7.1: Argand diagram.

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,

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

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. To represent roots of unity on an argand diagram, we should annotate
   (a) The points representing the roots, as dots.
   (b) The (dotted) circle which these points lie on.
   (c) The radius of the circle.
   (d) The angular separation between the roots.

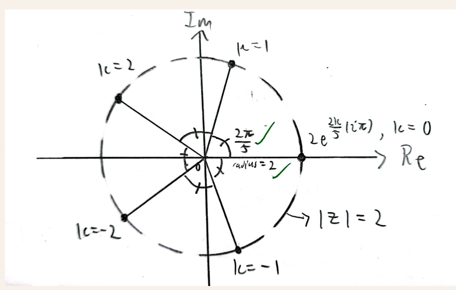


Figure 7.2: Roots of unity on an argand diagram.

6. 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 $a := x + iy$).

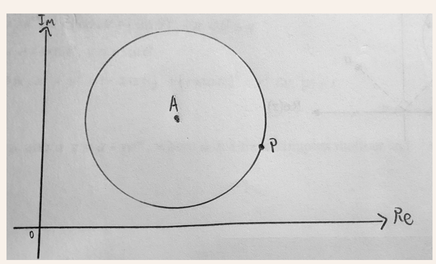


Figure 7.3: The locus of $|z - a| = r$.

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

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

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

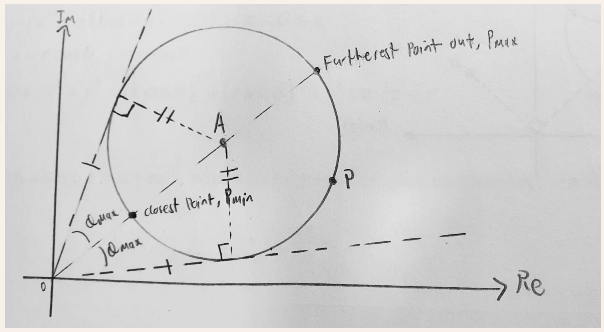


Figure 7.4: Maxmium distance and angle of a point from a circle

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

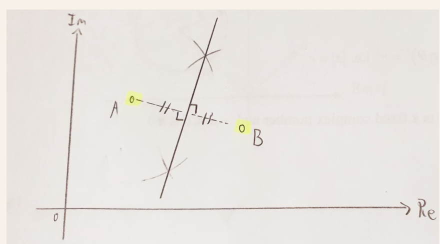


Figure 7.5: The locus of $|z-a| = |z-b|$, a perpendicular bisector

(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.

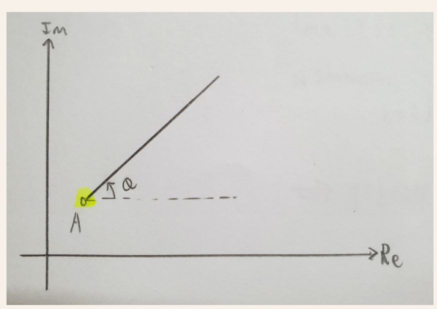


Figure 7.6: The locus of $\arg(z-a) = \theta$, a half-line.

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

8. 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 $P(z-a)$.

9. 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.

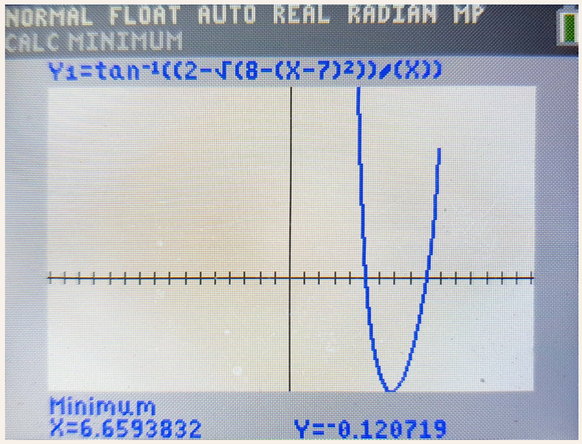


Figure 7.7: Brute force technique for finding maximum/minimum angles.

Note

To show that a *half*-line $\arg(z - a - bi) = \theta$ meets another loci L, we need to show that

(a) The line $y - b = \tan(\theta)(x - a)$ meets the loci L at some (u, v).

(b) $\begin{cases} u > a & \text{if } \theta \in (-\pi/2, \pi/2), \\ u = a & \text{if } \theta = \pm\pi/2, \\ u < a & \text{if } \theta \in (\pi/2, \pi) \cup (-\pi, -\pi/2). \end{cases}$

(c) $\begin{cases} v > b & \text{if } \theta \in (0, \pi), \\ v = b & \text{if } \theta = 0 \text{ or } \theta = \pi, \\ v < b & \text{if } \theta \in (-\pi, 0). \end{cases}$

The latter two conditions must be shown to be clearly satisfied, to illustrate our understanding of *half*-lines. For example, we may write

$$x = \frac{2 - \sqrt{3}}{4} > -2 \qquad y = \frac{1 + 10\sqrt{3}}{4} > 1.$$

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 $5x^2 - 10x + 1 = 0$.

Example 7.4: RV FM 2023 J2 CT

Show, by using De Moivre's theorem, that provided $\cos(\theta) \neq 0$,

$$\sum_{k=1}^{12} (-1)^{k-1} \cos((2k-1)\theta) = \frac{\sin^2(P\theta)}{\cos(\theta)}$$

where P is a constant to be determined.

Let $C = \sum_{k=1}^{12} (-1)^{k-1} \cos((2k-1)\theta)$ and $S = \sum_{k=1}^{12} (-1)^{k-1} \sin((2k-1)\theta)$. Then, for $z = e^{i\theta}$,

$$\begin{aligned} C + iS &= \sum_{k=1}^{12} (-1)^{k-1}[\cos((2k-1)\theta) + i\sin((2k-1)\theta)] \\ &= \sum_{k=1}^{12} z(-z^2)^{k-1} \\ &= \frac{z\left(1 - (-z^2)^{12}\right)}{1 - (-z^2)} \\ &\vdots \\ &= \frac{-ie^{i(12\theta)}\sin(12\theta)}{\cos(\theta)}. \end{aligned}$$

So, comparing real parts,

$$C = \frac{(-i) \cdot i\sin(12\theta) \cdot \sin(12\theta)}{\cos(\theta)} = \frac{\sin^2(12\theta)}{\cos(\theta)}.$$

Note

Algebraic tricks to know.

1. Factoring out $e^{i\theta/2}$, given an expression involving $e^{i\theta}$, can help in simplifying expressions.

2. Let $r_1, \ldots, r_n$ be the roots of the polynomial $p(z) := \sum a_i z^i$. To find the sum or product of the roots, consider (a) Vieta's Formula, or (b) comparing the coefficients of $p(z) = a_n(z - r_1)(z - r_2)\cdots(z - r_n)$.

★ Take extra caution to note whether a geometric progression is present.

3. Let $n \in \mathbb{Z}^+$. Suppose that n is odd, and we want to express $\sin^n(\theta)$ as a linear combination of $\sin(m\theta)$, for $m \in \mathbb{Z}^+$. First notice that $z^k - \frac{1}{z^k} = 2i\sin(k\theta)$, for $z = e^{i\theta}$. Then, use this fact in conjunction with the binomial theorem:

$$\begin{aligned} \left(z - \frac{1}{z}\right)^n &= \sum_{k=0}^{n} \binom{n}{k} (-1)^k z^{n-2k} \\ [2i\sin(\theta)]^n &= \sum_{k=0}^{\lfloor n/2 \rfloor} \binom{n}{n-2k} \left(z^{n-2k} - \frac{1}{z^{n-2k}}\right). \end{aligned}$$

Comparing imaginary parts, we get $\sin^n(\theta)$ in the desired form:

$$\sin^n(\theta) = \frac{1}{2^{n-1}} \sum_{k=0}^{\lfloor n/2 \rfloor} \binom{n}{n-2k} \sin((n-2k)\theta).$$

(Similarly, we can express $\cos^n(\theta)$ as a linear combination of $\cos(m\theta)$.)

4. Now, consider even n, instead. Then, recall that $z^k + \frac{1}{z^k} = 2\cos(k\theta)$. As such,

$$\begin{aligned}[2i\sin(\theta)]^n &= \sum_{k=0}^{n/2} \binom{n}{n-2k} \left(z^{n-2k} - \frac{1}{z^{n-2k}}\right) \\ \sin^n(\theta) &= \frac{1}{2^{n-1}} \sum_{k=0}^{n/2} \binom{n}{n-2k} \cos((n-2k)\theta).\end{aligned}$$

(The case for $\cos^n(\theta)$ is again similar.)

5. Let $n \in \mathbb{Z}^+$. Similarly, to find $\sin(n\theta)$ in terms of powers of $\sin(\theta)$, first apply De Moivre's theorem:

$$\cos(n\theta) + i\sin(n\theta) = (c + is)^n = \sum_{k=0}^{n} \binom{n}{k} c^{n-k} i^k s^k.$$

Then, we compare imaginary parts. (Or real parts, for $\cos(n\theta)$.)

6. Trigonometric identities:

$\pi - \theta$	$\pi/2 - \theta$
$\sin(\pi - \theta) = \sin(\theta)$	$\sin(\pi/2 - \theta) = \cos(\theta)$
$\cos(\pi - \theta) = -\cos(\theta)$	$\cos(\pi/2 - \theta) = \sin(\theta)$
$\tan(\pi - \theta) = -\tan(\theta)$	$\tan(\pi/2 - \theta) = \cot(\theta)$

Table 7.1

7. Let $z^n = 1$. Then, $\sum z^i = z^n \sum z^i = \sum z^{n+i}$.

8. When asked to prove an identity involving binomial coefficients $\binom{n}{k}$, try to factor the given form using the binomial theorem.

Example 7.5: Algebraic tricks to know

1(a). Simplifying a complex number w to a given form.

$$\begin{aligned}w &= \frac{-ie^{2ki\pi/5}}{1 - e^{2ki\pi/5}} = \frac{-ie^{2ki\pi/5}}{e^{ki\pi/5}\left[e^{-ki\pi/5} - e^{ki\pi/5}\right]} = \frac{-ie^{ki\pi/5}}{-2i\sin(k\pi/5)} \\ &= \frac{\cos(k\pi/5) + i\sin(k\pi/5)}{2\sin(k\pi/5)} = \frac{1}{2}[\cot(k\pi/5) + i]\end{aligned}$$

1(b).

$$z = 2\left(1 + e^{2ki\pi/3}\right) = 2e^{ki\pi/3}\left(e^{-ki\pi/3} + e^{ki\pi/3}\right) = 4\cos(k\pi/3)e^{ki\pi/3}.$$

1(c). Let $z_k = e^{2ki\pi/n}$ for all $1 \le k \le n$. In the case where n is odd, i.e. $n = 2m + 1$, evaluate the series $\sum_{k=1}^{n} 2(1+z_k)^{-1}$.

$$\begin{aligned}\sum_{k=1}^{n} \frac{2}{1+z_k} &= \sum_{k=1}^{n} \frac{2}{1+e^{2ki\pi/n}} = \sum_{k=1}^{n} \frac{2}{e^{ki\pi/n}\left(e^{-ki\pi/n}+e^{ki\pi/n}\right)} = \sum_{k=1}^{n} \frac{2e^{-ki\pi/n}}{2\cos(k\pi/n)} \\ &= \sum k = 1^n \frac{\cos(k\pi/n) - i\sin(k\pi/n)}{\cos(k\pi/n)} = \sum_{k=1}^{n} 1 - i\tan(k\pi/n) \\ &= n - i\Big[\big[\tan(\pi/n) + \tan((n-1)\pi/n)\big] + \big[\tan(2\pi/n) + \tan((n-2)\pi/n)\big] + ... \\ &\quad + \big[\tan(m\pi/n) + \tan((n-m)\pi/n)\big] + \tan(\pi)\Big] \\ &= n + i(0+0+\cdots+0) = n\end{aligned}$$

because $\tan(\pi-\theta) = -\tan(\theta)$.

2. We found that the roots of $(1+z)^4 + (1-z)^4 = 0$ has roots $i\tan(k\pi/8)$, where $k = 1,3,4,7$. Then, we were tasked to find the value of $\tan^2(\pi/8) + \tan^2(3\pi/8)$ and $\tan^2(5\pi/8)\tan^2(\pi/8)\tan^2(3\pi/8)$.

   Since $\tan(\pi/8) = -\tan(7\pi/8)$ and $\tan(3\pi/8) = -\tan(5\pi/8)$,

$$\begin{aligned}(1+z)^4 + (1-z)^4 &= 2[z - i\tan(\pi/8)][z + i\tan(\pi/8)][z - i\tan(3\pi/8)][z + i\tan(3\pi/8)] \\ &= 2[z^2 + \tan^2(\pi/8)][z^2 + \tan^2(3\pi/8)].\end{aligned}$$

   Comparing constants/coefficients of z^2, we obtain

$$\tan^2(\pi/8)\tan^2(3\pi/8) = 1 \qquad \text{and} \qquad \tan^2(\pi/8) + \tan^2(3\pi/8) = 6,$$

   respectively.

3. The case of $n = 5$: expressing $\sin^5(\theta)$, in terms of $\sin(m\theta)$.

$$\begin{aligned}\left(z - \frac{1}{z}\right)^5 &= z^5 - 5z^3 + 10z - \frac{10}{z} + \frac{5}{z^3} - \frac{1}{z^5} \\ [2i\sin(\theta)]^5 &= \left(z^5 - \frac{1}{z^5}\right) - 5\left(z^3 - \frac{1}{z^3}\right) + 10\left(z - \frac{1}{z}\right) \\ 32i\sin^5(\theta) &= 2\sin(5\theta) - 5(2)i\sin(3\theta) + 10(2)i\sin(\theta) \\ \sin^5(\theta) &= \frac{1}{16}\sin(5\theta) - \frac{5}{16}\sin(3\theta) + \frac{5}{8}\sin(\theta).\end{aligned}$$

7. Let $\omega = e^{2i\pi/11}$. We define

$$\alpha = \omega + \omega^3 + \omega^4 + \omega^5 + \omega^9 \qquad \text{and} \qquad \beta = \omega^{-1} + \omega^{-3} + \omega^{-4} + \omega^{-5} + \omega^{-9}.$$

   Find $(x-\alpha)(x-\beta)$ in its simplest form. Notice that $\beta = \omega^{11}\beta = \omega^{10} + \omega^8 + \omega^7 + \omega^6 + \omega^2$. As such,

$$\alpha + \beta = \sum_{k=1}^{10} \omega^k = -1 \qquad \text{and} \qquad \alpha\beta = \underbrace{5 + 2\left(\sum_{k=1}^{10} \omega^k\right)}_{\text{the usual expansion}} = 5 - 2 = 3.$$

   Now, $(x-\alpha)(x-\beta) = x^2 - (\alpha+\beta)x + \alpha\beta = x^2 + x + 3$.

8. Let

$$C = 1 - \binom{2n}{1}\cos(\theta) + \binom{2n}{2}\cos(2\theta) - \binom{2n}{3}\cos(3\theta) + \cdots + \cos(2n\theta)$$
$$S = -\binom{2n}{1}\sin(\theta) + \binom{2n}{2}\sin(2\theta) - \binom{2n}{3}\sin(3\theta) + \cdots + \sin(2n\theta),$$

where n is a positive integer. Then,

$$\begin{aligned} C + iS &= \sum_{r=0}^{2n} (-1)^r \binom{2n}{r} [cos(r\theta) + i\sin(r\theta)] \\ &= \sum_{r=0}^{2n} \binom{2n}{r} (-1)^r e^{i(r\theta)} = \sum_{r=0}^{2n} \binom{2n}{r} (1)^{2n-r} \left(-e^{i\theta}\right)^r \\ &= \left(1 - e^{i\theta}\right)^{2n} = \left(e^{i\theta/2}\right)^{2n} \left(e^{-i\theta/2} - e^{i\theta/2}\right)^{2n} \\ &= e^{in\theta}[-2i\sin(\theta/2)]^{2n} \\ &= (-4)[\cos(n\theta) + i\sin(n\theta)]\sin^{2n}(\theta/2). \end{aligned}$$

Comparing real and imaginary parts,

$$C = (-4)^n \cos(n\theta)\sin^{2n}(\theta/2) \qquad \text{and} \qquad S = (-4)^n \sin(n\theta)\sin^{2n}(\theta/2).$$

Note

Geometrical tricks to know.

1. Where clarity may otherwise be lacking, consider making a line segment *thicker* and *label* it with "$\rightarrow$ required points" — to indicate that it is the locus being requested for.
2. Be aware of any triangles, congruent triangles, common angles, and sides of common length. In particular, when a triangle has an angle of $\pi/4 = 45°$, it is an isosceles triangle. These observations are especially helpful in finding circle-line intersections.

Example 7.6

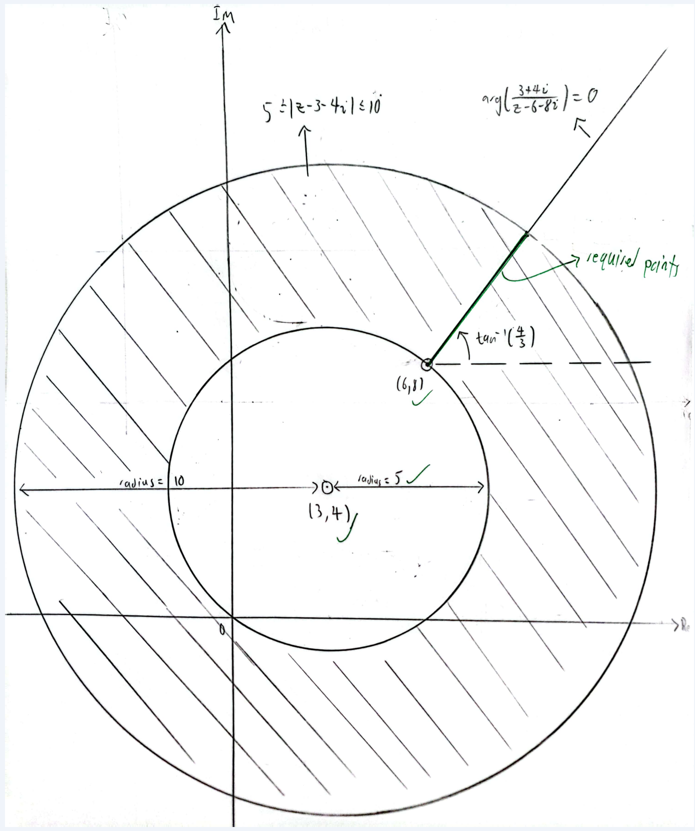
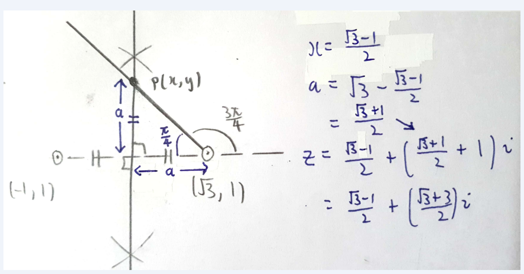
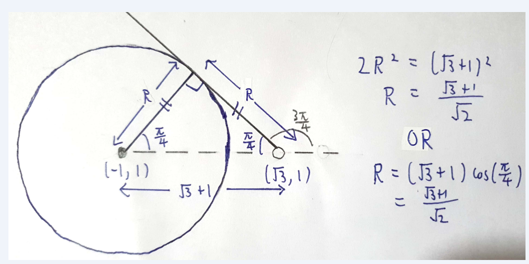


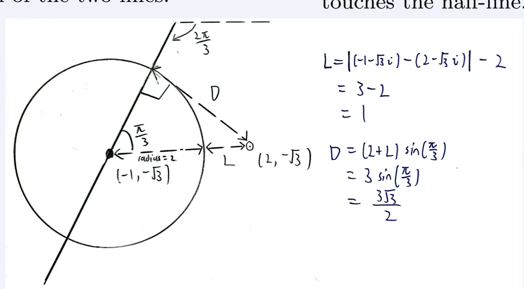
Figure 7.8: Annotations that improve clarity.



(a) Finding the intersection of the two lines.



(b) Finding the radius R for which the circle just touches the half-line.



(c) Finding the shortest distance L and D of $(2, -\sqrt{3})$, from the circle and line, respectively.

Figure 7.9: The writing in blue denote the deductions we should make.

Chapter 8

Linear Algebra

8.1 Vector spaces, subspaces, linear combinations, bases, linear transformations

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 $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$ in S and scalars $a_1, a_2, \ldots, a_n$ in $\mathbb{F}$ such that

$$v = \sum_{i=1}^{n} a_i \mathbf{u}_i.$$

In this case we also say that v is a linear combination of $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$ 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 $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$in S and scalars $a_1, a_2, \ldots, a_n$ not all zero, such that

$$a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + a_n\mathbf{u}_n = \mathbf{0}.$$

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.

Definition 8.7

Let V and W be vector spaces. We call a function $T\colon V \to W$ a *linear transformation from V to W* iff $T(c\mathbf{x} + \mathbf{y}) = cT(\mathbf{x}) + T(\mathbf{y})$ for all $\mathbf{x}, \mathbf{y} \in V$ and $c \in \mathbb{F}$.

Theorem 8.8: The Rank-Nullity Theorem.

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

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

8.2 Matrices and systems of linear equations

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

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

Definition 8.9

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

Theorem 8.10

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

Theorem 8.11

A system $\mathbf{Ax} = \mathbf{0}$ of m linear equations in n unknowns has a solution space of dimension $n - \text{rank}(A)$.

Definition 8.12

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

Theorem 8.13: The Rouché-Capelli Theorem.

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

Definition 8.14

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.

General Information

- 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 reduces a matrix to its rref form.
- Gaussian elimination always reduces $(\mathbf{A} \,|\, \mathbf{I}_n) \to (\mathbf{I}_n \,|\, \mathbf{A}^{-1})$.
- 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, $\{\mathbf{a}_{k_1}, \mathbf{a}_{k_2}, \dots, \mathbf{a}_{k_m}\}$ is linearly independent iff $\{\mathbf{a}'_{k_1}, \mathbf{a}'_{k_2}, \dots, \mathbf{a}'_{k_m}\}$ is. Moreover, the row space of $\mathbf{A}$ and $\mathbf{A}'$ are clearly identical.
- (Example) To find a basis for the intersection of the column spaces of $\mathbf{A}, \mathbf{B} \in \text{M}_{n\times n}(\mathbb{F})$, we 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 columns of $\begin{pmatrix} \mathbf{A} & \mathbf{B} \end{pmatrix}$ and $\begin{pmatrix} \mathbf{A}' & \mathbf{B}' \end{pmatrix}$, respectively. We compare the columns of $\mathbf{A}'$ and $\mathbf{B}'$ to find a basis $\beta' := \{\mathbf{c}'_{i_1}, \mathbf{c}'_{i_2}, \dots, \mathbf{c}'_{i_r}\}$ for the intersection of the column spaces of $\mathbf{A}'$ and $\mathbf{B}'$. Then, $\beta := \{\mathbf{c}_{i_1}, \mathbf{c}_{i_2}, \dots, \mathbf{c}_{i_r}\}$ is a basis for the intersection of the column spaces of $\mathbf{A}$ and $\mathbf{B}$.

8.3 Determinants

Definition 8.15

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.

Note

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

Theorem 8.16

The determinant $\det\colon \mathrm{M}_{n\times n}(\mathbb{F}) \to \mathbb{F}$ is an alternating n-linear function.

(a) Alternating: 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 $\mathbf{A}$,

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

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

Corollary 8.17

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 8.18

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 8.19

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

Theorem 8.20

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 8.21

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 8.22

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

8.4 Diagonalisation

Definition 8.23

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 8.24

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 8.25

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 $\mathbf{A}$ — 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 reducing $\mathbf{A} - \lambda_j\mathbf{I}$.

3. Let $\mathbf{Q} = (\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3)$. Then,
$$\mathbf{D} := \mathbf{Q}^{-1}\mathbf{A}\mathbf{Q}$$
is a diagonal matrix.

Note

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

Theorem 8.26: 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 8.27: The Cayley-Hamilton Theorem for Matrices.

Let $\mathbf{A}$ be an $n \times n$ matrix, and let $f(t)$ be the characteristic polynomial of $\mathbf{A}$. Then, $f(\mathbf{A}) = \mathbf{O}$, 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 matrices $\mathbf{A}$ and $\mathbf{I}_3$ into `[A]` and `[I]`, respectively.
2. Plot $\texttt{Y}_1 = \det(\texttt{[A]})$.
3. `2nd` $\Longrightarrow$ `trace` $\Longrightarrow$ `2:zero` $\Longrightarrow$ Find the roots.

8.5 Miscellaneous

An asterisk denotes the conjugate transpose.

Theorem 8.28

Let $\mathbf{M} \in \mathrm{M}_{n\times n}(\mathbb{K})$ be Hermitian (i.e. $\mathbf{M}^* = \mathbf{M}$), with eigenvectors $\mathbf{u}$ and $\mathbf{v}$ that correspond to the eigenvalues λ and μ. Then, $\mathbf{u}$ and $\mathbf{v}$ are orthogonal with respect to the standard inner product, if $\lambda \neq \mu^*$.

Proof. Let $\mathbf{u}$ and $\mathbf{v}$ be eigenvectors of $\mathbf{M}$. Then,
$$\langle \mathbf{u}, \mu\mathbf{v}\rangle = (\mathbf{M}\mathbf{v})^*\mathbf{u} = (\mathbf{v}^*\mathbf{M}^*)\mathbf{u} = \mathbf{v}^*(\mathbf{M}^*\mathbf{u}) = \mathbf{v}^*(\lambda\mathbf{u}) = \langle \lambda\mathbf{u}, \mathbf{v}\rangle.$$
As such, $(\lambda - \mu^*)\langle \mathbf{u}, \mathbf{v}\rangle = 0$. Hence, $\langle \mathbf{u}, \mathbf{v}\rangle = 0$. $\square$

Example 8.1

Consider a computer that rounds each calculated value to n decimal places, which is then used in later calculations as if it were exact. Perform, for $n = 3$ and $n = 4$, this procedure to find the solution $\mathbf{x} = \begin{pmatrix} x_1 & x_2 & x_3 \end{pmatrix}^\top$ to $\mathbf{A}\mathbf{x} = \mathbf{b}$. Then, find $\sum_{i=1}^{3} \delta_i^2$ where δ_i is the difference between the exact value of x_i and the one found by the computer: this gives a measure for the accuracy of the calculated values. Comment on the difference in results.

One extra decimal place of accuracy in the working (a factor of 10) had led to a significant increase in the measure of accuracy (by a factor of around 250).

8.6 Conics

Given a conic section defined by $Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0$, we may wish to find its lines of symmetry, center, radii, etc. The core idea is simple: complete the square, to express $Ax^2 + Bxy + Cy^2$ in the form $a(x')^2 + b(y')^2$, for some linear combinations x' and y' of x and y. Then, the initial equation becomes $a(x')^2 + b(y')^2 + d(x') + e(y') + F = 0$, which is easily reduced to the standard form for conics. Before that, we need to develop some machinery.

Definition 8.29

Let $\mathbb{F}$ be a field not of characteristic two. A function $K\colon \mathbb{F}^n \to \mathbb{F}$ is called a *quadratic form on* $\mathbb{F}^n$ if there exists a symmetric matrix $A \in \mathrm{M}_{n\times n}(\mathbb{F})$, such that $K(\mathbf{x}) = \mathbf{x}^\top \mathbf{A}\mathbf{x}$ for all $\mathbf{x} \in \mathbb{F}^n$.

Note

Let $\mathbb{F}$ be a field not of characteristic two and scalars $a_i \in \mathbb{F}$. The polynomial $f\colon \mathbb{F}^n \to \mathbb{F}$ given by $f(t_1, t_2, \ldots, t_n) = \sum_{i\leq j} a_{ij} t_i t_j$ is a quadratic form. In fact, the matrix $\mathbf{A}$ with

$$\mathbf{A}_{ij} = \begin{cases} a_{ii} & \text{if } i = j \\ a_{ij}/2 & \text{if } i \neq j \end{cases}$$

gives us our desired quadratic form $K(\mathbf{x}) = \mathbf{x}^\top \mathbf{A}\mathbf{x}$.

Theorem 8.30

Let $K(\mathbf{x}) = \mathbf{x}^\top \mathbf{A}\mathbf{x}$ be a quadratic form on a finite-dimensional real inner product space V. There exists an orthonormal eigenbasis $\beta \coloneqq \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ for $\mathbf{A}$ and eigenvalues λ_i, such that $K\left(\sum_{i=1}^n t_i \mathbf{v}_i\right) = \sum_{i=1}^n \lambda_i t_i^2$ for all $t_i \in \mathbb{R}$.

We now return to our initial problem on conics. We first diagonalise the matrix

$$\begin{pmatrix} A & B/2 \\ B/2 & C \end{pmatrix}$$

to find its eigenvalues λ and μ, then the corresponding unit eigenvectors $\mathbf{u} = \begin{pmatrix} \alpha & \beta \end{pmatrix}^\top$ and $\mathbf{v} = \begin{pmatrix} \gamma & \delta \end{pmatrix}^\top$. Then,

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \mathbf{u} & \mathbf{v} \end{pmatrix} \begin{pmatrix} t_1 \\ t_2 \end{pmatrix} = \begin{pmatrix} \alpha t_1 + \gamma t_2 \\ \beta t_1 + \delta t_2 \end{pmatrix}.$$

Furthermore, $Ax^2 + Bxy + Cy^2 = \lambda t_1^2 + \mu t_2^2$ by 1.28. Therefore,

$$\lambda t_1^2 + \mu t_2^2 + D(\alpha t_1 + \gamma t_2) + E(\beta t_1 + \delta t_2) + F = 0$$

$$\lambda\left(t_1 + \frac{D\alpha + E\beta}{2\lambda}\right)^2 + \mu\left(t_2 + \frac{D\gamma + E\delta}{2\mu}\right)^2 - \frac{(D\alpha + E\beta)^2}{4\lambda} - \frac{(D\gamma + E\delta)^2}{4\mu} + F = 0 \qquad (\text{if } \lambda, \mu \neq 0)$$

gives an equivalent form for our conic section.

Since our basis $\{\mathbf{u}, \mathbf{v}\}$ is orthonormal, the above change of basis from the standard ordered basis to $\{\mathbf{u}, \mathbf{v}\}$ is an isometry (that maps $\mathbf{0}$ to itself). It is obtained via a rotation and/or reflection. Hence, the radii are $\sqrt{\lambda/k}$ and $\sqrt{\mu/k}$, for $4k = (D\alpha + E\beta)^2/\lambda + (D\gamma + E\delta)^2/\mu - F$; the center is $\left(-\frac{D\alpha+E\beta}{2\lambda}, -\frac{D\gamma+E\delta}{2\mu}\right)$. Notice that $t_1 = 0$ and $t_2 = 0$ are parallel to the axes of symmetry. i.e. $\mathbf{u}$ and $\mathbf{v}$ are parallel to the axes of symmetry of our conic. As such, the lines of symmetry are

$$y + \frac{D\gamma + E\delta}{2\mu} = \frac{\beta}{\alpha}\left(x + \frac{D\alpha + E\beta}{2\lambda}\right) \qquad \text{and} \qquad y + \frac{D\gamma + E\delta}{2\mu} = \frac{\delta}{\gamma}\left(x + \frac{D\alpha + E\beta}{2\lambda}\right).$$

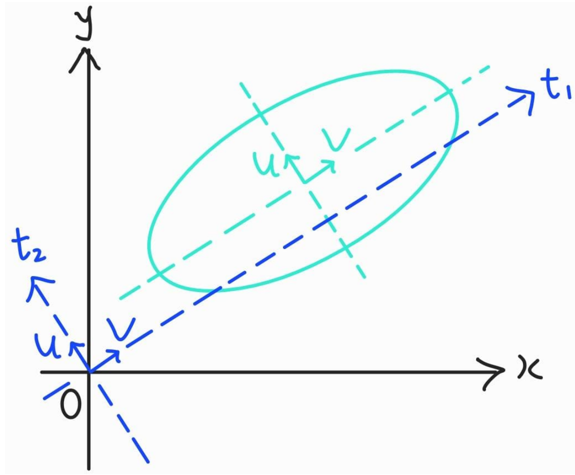


Figure 8.1: Rotated and translated conic.

Note

Without orthonormality, i.e. an isometric change of coordinates, even if we were successful in reducing our conic to the form $a(x')^2 + b(y')^2 = f$, it may not prove to be a useful form. Consider the ellipse $2x^2 + 2xy + y^2 = 1$. Clearly, $x^2 + (x + y)^2 = 1$. But, this gives a circle in $(x, x + y)$ coordinates; we can't deduce much about our initial conic. So, little meaning is found in such a factorisation.

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.
- To show that there a continuous function f attains a root in an interval $[a, b]$, we find two values $x < y$ in the interval (e.g. $a < b$) such that $f(a)f(b) < 0$. i.e. show that f changes sign in $[a, b]$. Then, *by continuity*, a root of f must lie in $[a, b]$.
- To further show that the root is *unique* in $[a, b]$, it suffices to prove that f is *strictly* monotone on $[a, b]$.
- Suppose we have some function $f : \mathbb{R} \to \mathbb{R}$ 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. Our approximation is
  $$\frac{a|f(b)| + b|f(a)|}{|f(a)| + |f(b)|}.$$
     - 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 $\bigcup$	underestimation	overestimation
Downwards $\bigcap$	overestimation	underestimation

Table 9.1: Approximation errors when using linear interpolation.

     - See Figure 9.2 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.

Note

It is important to show which interval we are interpolating on, not just the iteratively obtained values. We can present our working using the table below.

a	$f(a)$	b	$f(b)$	$\dfrac{a\lvert f(b)\rvert + b\lvert f(a)\rvert}{\lvert f(a)\rvert + \lvert f(b)\rvert}$
a	$f(a) > 0$	b	$f(b) < 0$	x_1
x_1	$f(x_1) > 0$	b	$f(b) < 0$	x_2
x_1	$f(x_1) > 0$	x_2	$f(x_2) < 0$	x_3
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

Table 9.2: Required working for linear interpolation.

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 $x_{n+1} \coloneqq F(x_n)$. We want $x_n \to \alpha$.
   - Convergence behavior

Behavior 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 9.3: Convergence behavior of fixed-point iterations.

   - See Figure 9.3 for an illustration.

Note

We must write out *all* iterations, not just the final two. The working below is sufficient.

Let $x_0 = \_\_\_$ and $x_{n+1} = F(x_n)$, $x \geq 0$.

$$\begin{aligned} x_1 &= \_\_\_ \\ x_2 &= \_\_\_ \\ &\vdots \\ x_{m-1} &= \_\_\_ \\ x_m &= \_\_\_ \end{aligned}$$

Therefore, $\alpha = x_m$ (k d.p.), since $f(x_m - 0.0\cdots05)f(x_m + 0.0\cdots05) = \_\_\_ < 0$.

3. The Newton-Raphson Method. Let α be a root of the function $f\colon \mathbb{R} \to \mathbb{R}$. 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 — where the concavity changes/the sign of $f''(x)$ changes.

– Error:

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

Table 9.4: Approximation errors when using the Newton-Raphson method.

– See Figure 9.4 for an illustration.

Note

We must write out *all* iterations, not just the final two. One way to present our working is as follows.

Let $x_0 = \_\_$ and $x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} = \_\_$, $x \geq 0$.

$$\begin{aligned} x_1 &= \_\_ \\ x_2 &= \_\_ \\ &\vdots \\ x_{m-1} &= \_\_ \\ x_m &= \_\_ \end{aligned}$$

Therefore, $\alpha = x_m$ (k d.p.), since $f(x_m - 0.0\cdots05)f(x_m + 0.0\cdots05) = \_\_ < 0$.

Note

Explain whether $x_0 = \_\_$ is a suitable starting value for using the Newton-Raphson method to find an approximation to α.

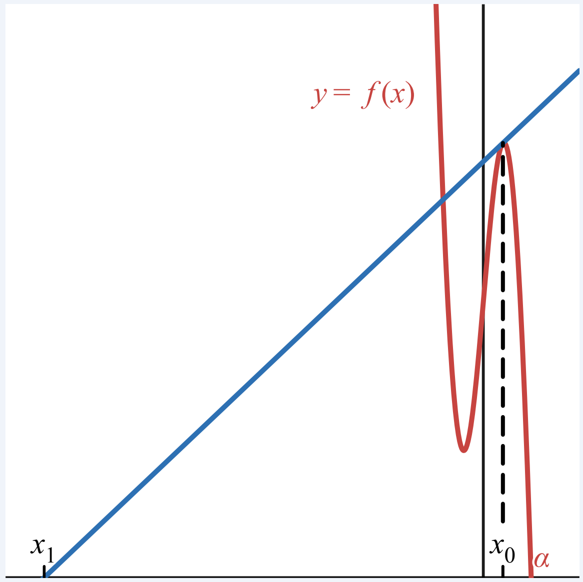


Figure 9.1: (Desmos)

1. Since $x_0 =$ ___ is very close to the stationary point, the tangent to the curve $y = f(x)$ has a very gentle gradient. Thus, it cuts the x-axis far away from the initial approximation.

2. Furthermore, as $x_0 =$ ___ is to the left/right of the minimum/maximum point $x =$ ___, the values of the gradient $f'(x_n)$ will be negative/positive for all $n \geq 0$. Hence, x_n converges to the root β instead α.

(The second point may be omitted if it is irrelevant.)

Note

Suppose a question asks for the approximation of a root to k significant figures/k decimal places. Then:

1. We leave our iterative approximations x_n to at least $k + 2$ significant figures/$k + 2$ decimal places.

2. We continue the iterative process till two consecutive ones agree up to k significant figures/k decimal places.

Note

Perform ________ (e.g. linear interpolation) to obtain an approximation for α, correct to two decimal places. Justify whether this approximation is sufficiently accurate.

Suppose our approximation is some $a = 1.00$, then we note the sign of f at $a \pm 0.005$. (For an arbitrary number of s.f. or d.p., simply adjust the value 0.005 accordingly. E.g. for 3 d.p. we instead use 0.0005). Our working should look similar to the following:

> Since $f(0.995) =$ ___ < 0 and $f(1.005) =$ ___ > 0, we conclude that 1.00 is a sufficiently accurate approximation, at 2 d.p..

Note

The *error obtained* when using an approximation should be the *absolute* difference of the true value and the approximation.

Note

Use the results of part (i) and the differential equation $dy/dx = \sin(xy)$ to estimate the x-coordinate x_P of P.

The maximum point occurs when $dy/dx = \sin(xy) = 0$, i.e. $xy = k\pi$ where $k \in \mathbb{Z}$. From (i), $\frac{dy}{dx}\Big|_{x=5/3} \approx 0.643 > 0$ so $y_P > y(5/3) \approx 2.0468$. Hence, $x_P \approx \pi/2.04679402 = 1.53$ (3.s.f).

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!

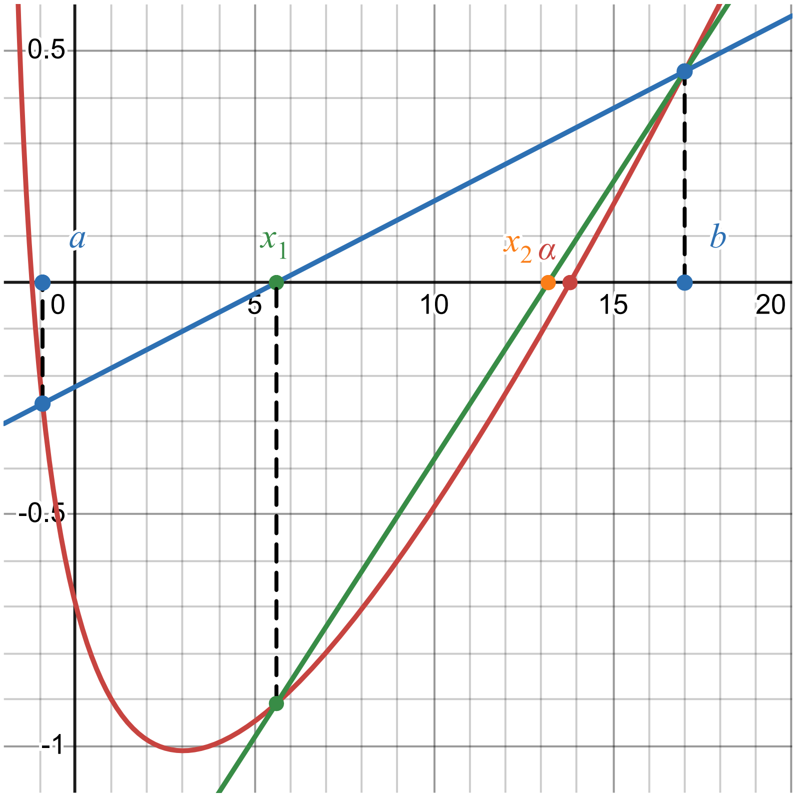


Figure 9.2: An illustration of linear interpolation (Desmos).

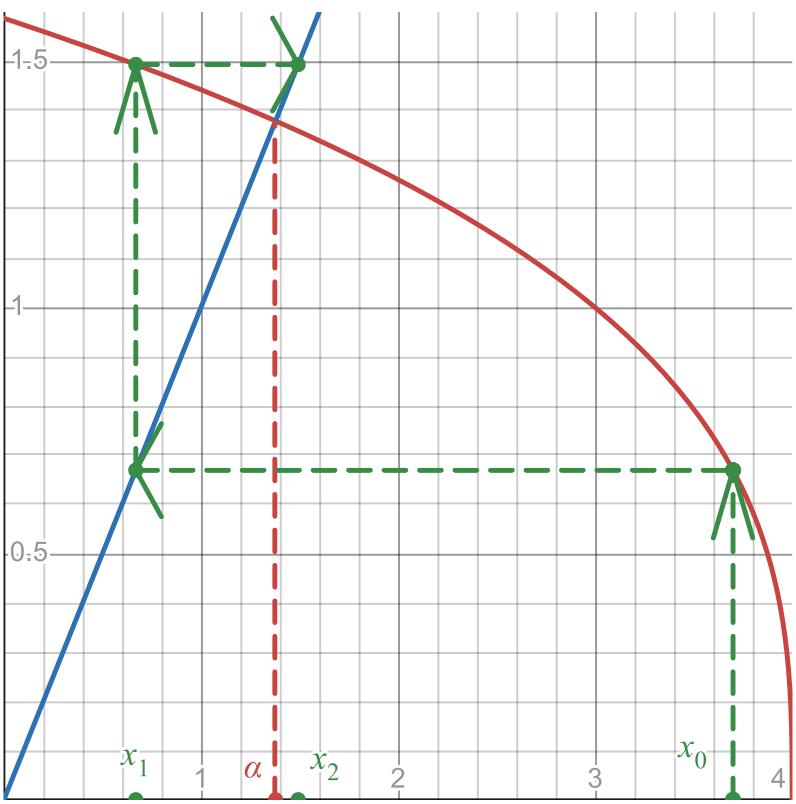


Figure 9.3: An illustration of fixed-point iteration (Desmos).

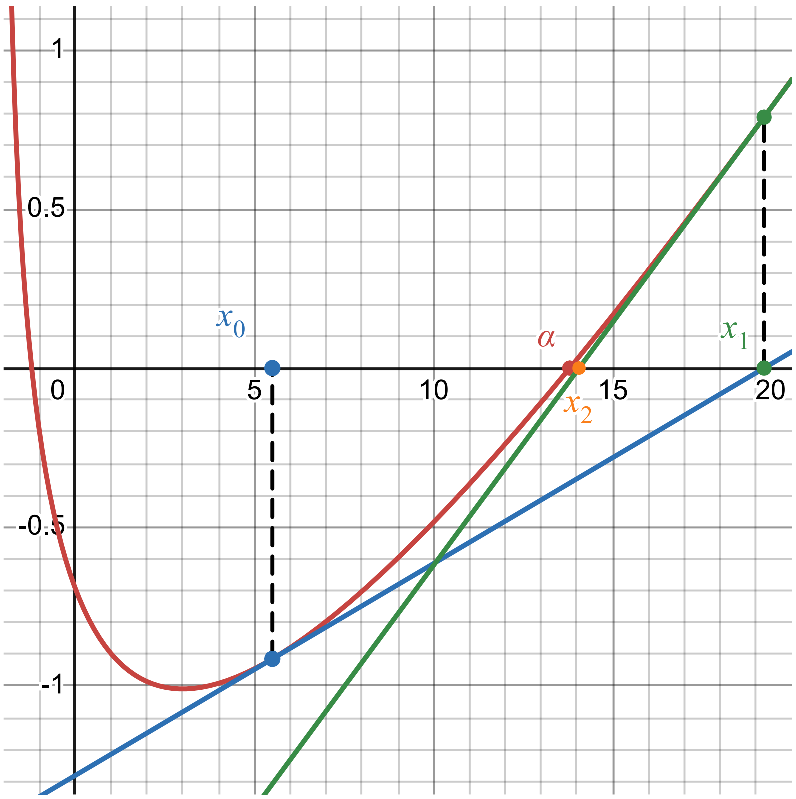


Figure 9.4: An illustration of Newton's Method (Desmos).

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

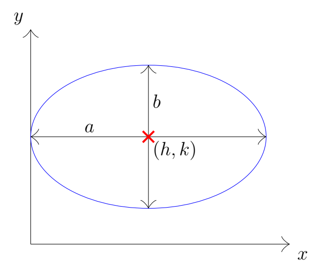
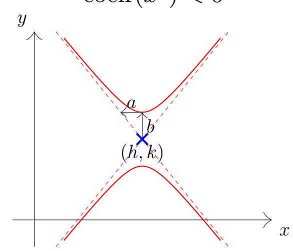
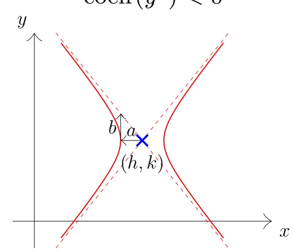
Lines of Symmetry: $y = x + 16$, $y = -x + 24$.

Important Notes

- The discriminant can be very useful.
- Be aware of 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 k that satisfy some condition(s).

10.2 Conics

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

	Ellipses	Hyperbolas
Standard Forms	$\dfrac{(x-h)^2}{a^2}+\dfrac{(y-k)^2}{b^2}=1$	$\dfrac{(x-h)^2}{a^2}-\dfrac{(y-k)^2}{b^2}=1$ $\dfrac{(y-k)^2}{b^2}-\dfrac{(x-h)^2}{a^2}=1$
General Equation	$ax^2+by^2+cx^2+dx+e=0$, where $\operatorname{sgn}(a)=\operatorname{sgn} b$.	$ax^2+by^2+cx^2+dex+e=0$, where $\operatorname{sgn}(a)\neq\operatorname{sgn} b$.
Center	(h,k)	
Vertical 'Radius' (variables here from *standard form*!)	b	
Horizontal 'Radius' (variables here from *standard form*!)	a	
Vertical Vertices (variables here from *standard form*!)	$(h,k\pm b)$	
Horizontal Vertices (variables here from *standard form*!)	$(h\pm a,k)$	
Shape		coeff$(x^2)<0$  coeff$(y^2)<0$ 
Asymptotes (No need to rmb!)	-	$y=k\pm\dfrac{b(x-h)}{a}$
Lines of Symmetry	$x=h,\ y=k$	

General Information

- To find asymptote of hyperbolas, solve
$$\frac{(x-h)^2}{a^2}=\frac{(y-k)^2}{b^2}.$$
- When sketching any conic, label its vertices or radii, together with its center and asymptotes.

10.3 Parametric Equations

Important Notes

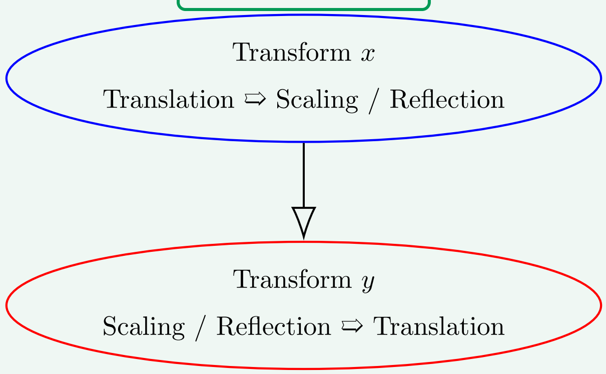
⋆ Check the qns for any *restrictions* on the parameter! And modify that of the G.C.'s accordingly (Tmin & Tmax).

⋆ 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

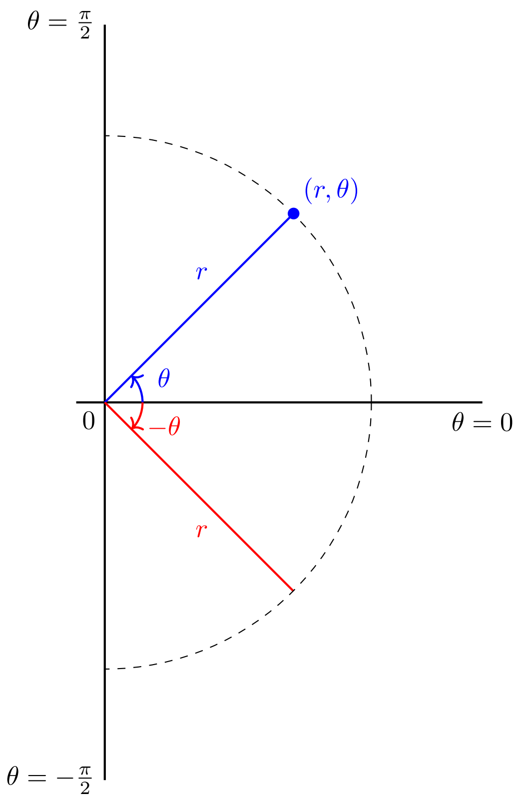
- For $|f(x)|$, simply flip the part of the graph of $f(x)$ that is below the x-axis, to above the x-axis.
- For $f(|x|)$, its graph is symmetric about the x-axis

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

<table>
<tr><th>Behavior of $f(x)$</th><th>Behavior of $1/f(x)$</th></tr>
<tr><td>$f(x) > 0$</td><td>$\dfrac{1}{f(x)} > 0$</td></tr>
<tr><td>$f(x) < 0$</td><td>$\dfrac{1}{f(x)} < 0$</td></tr>
<tr><td>Vertical Asymptote at $x = c$</td><td>$\dfrac{1}{f(x)}$ *tends* to 0
* $\frac{1}{f(x)}$ is undefined at $x = c$</td></tr>
<tr><td colspan="2">$$\frac{df}{dx} = -\frac{d}{dx}\left(\frac{1}{f(x)}\right)$$
i.e. when $f(x)$ increases, $\frac{1}{f(x)}$ decreases.</td></tr>
<tr><td>(a, b) is a *minimum* pt</td><td>$\left(a, \dfrac{1}{b}\right)$ is a *maximum* pt</td></tr>
<tr><td>(a, b) is a *maximum* pt</td><td>$\left(a, \dfrac{1}{b}\right)$ is a *minimum* pt</td></tr>
</table>

Chapter 11

Polar Curves



Definition

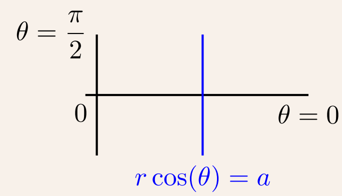
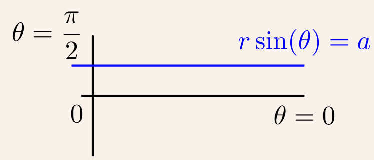
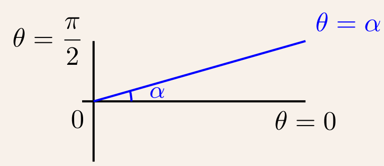
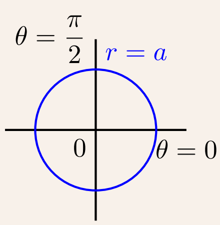
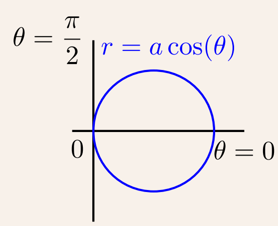
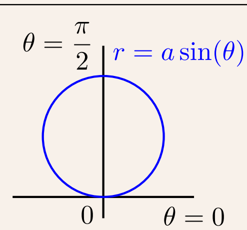
1. The *pole* is the origin.
2. The *initial line/polar axis* is the *half line* $\theta = 0$.

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

<table>
<tr><th>Polar Equation</th><th>Cartesian Equation</th></tr>
<tr><td>
</td><td>$x = a$</td></tr>
<tr><td>
</td><td>$y = a$</td></tr>
<tr><td>
</td><td>$y = x\tan(\alpha)$</td></tr>
<tr><td>
</td><td>$x^2 + y^2 = a^2$</td></tr>
<tr><td>
</td><td>$\left(x - \frac{a}{2}\right)^2 + y^2 = \frac{a^2}{4}$</td></tr>
<tr><td>
</td><td>$x^2 + \left(y - \frac{a}{2}\right)^2 = \frac{a^2}{4}$</td></tr>
</table>

- Tangent lines at the pole are obtained by solving $r = 0$.
- $r = f(\theta)$ is symmetrical about the polar (horizontal) axis iff $f(\theta) = f(-\theta)$ for all θ.
- $r = f(\theta)$ is symmetrical about the vertical line $\theta = \pi/2$ iff $f(\theta) = f(\pi - \theta)$ for all θ.
- Suppose r is a function of $\cos(n\theta)$ *only*. E.g. $r = a\sqrt{(4 + \sin^2(\theta))^2 \cos(\theta)}$. Then, the lines of symmetry are $n\theta = k\pi$, for $k \in \mathbb{Z}$.
- Suppose r is a function of $\sin(n\theta)$ *only*. Then, the lines of symmetry are $n\theta = k\pi/2$, for $k \in \mathbb{Z}$.
- $r = f(\theta)$ is symmetrical about the pole iff (r, θ) is a point on the curve whenever $(-r, \theta)$ is.
- The use of R-formulas may be necessary.
- Area of a sector,
$$A = \frac{1}{2}\int_{\alpha}^{\beta} r^2 \, \mathrm{d}\theta,$$
where $\alpha < \beta$.
- Arc length,
$$\ell = \int_{\theta_1}^{\theta_2} \sqrt{r^2 + \left(\frac{dr}{d\theta}\right)^2} \, \mathrm{d}\theta.$$

Important Notes

1. Normally, $r \geq 0$. But, in some questions, it can be negative.
2. There is no need to fully expand/simplify our final answers. E.g. $(x^2+y^2)^2 = 3y(x^2+y^2)-4y^2$ suffices.
3. The essentials of sketching polar curves:
   (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. When finding dy/dx, try to keep it in polar form if possible instead of converting to cartesian form.
6. As usual, be *careful*, such as about which values need to be rejected.
7. When reflecting/rotating, a diagram may be useful to find the 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) A reflection 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. 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

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

into formulas (e.g. for arc length) without having to manually differentiate it!

Chapter 12

Conic Sections

Definition 12.1

Eccentricity e is defined as

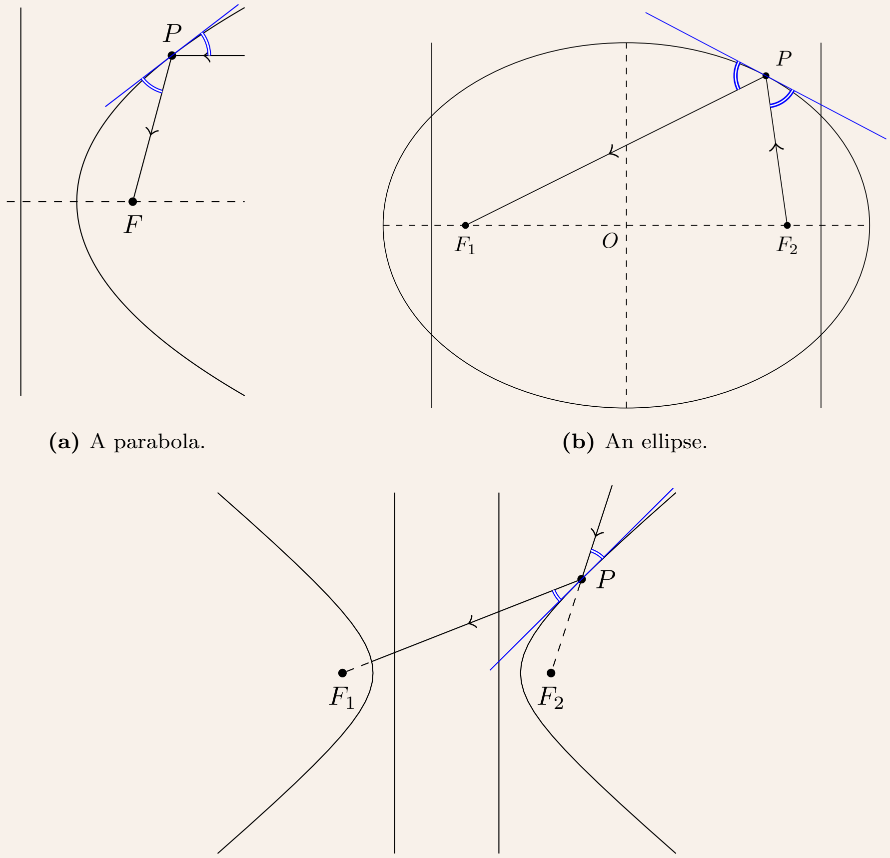
$$\frac{\text{distance of } P \text{ from focus}}{\text{distance of } P \text{ from directrix}}.$$

General Information

Conic section	Circle	Ellipse	Parabola	Hyperbola
Eccentricity e	0	$(0,1)$	1	$(1,\infty)$

Table 12.1: Values of eccentricity e and the associated conic sections.

Conic	Parabolas		Ellipses		Hyperbolas	
Cartesian	$x^2 = 4py$	$y^2 = 4px$	$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1,\ a>b$	$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1,\ b>a$	$\frac{x^2}{a^2}-\frac{y^2}{b^2}=1$	$\frac{y^2}{b^2}-\frac{x^2}{a^2}=1$
Parametric	$x=t$ $y=t^2/4p$	$x=t^2/4p$ $y=t$	$x=a\cos(\theta),\ y=b\sin(\theta)$		$x=a\sec(\theta)$ $y=b\tan(\theta)$	$x=a\tan(\theta)$ $y=b\sec(\theta)$
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$(x^2)>0$, $\lvert PF_1-PF_2\rvert=2a$	



(a) A parabola.

(b) An ellipse.

(c) A hyperbola

Figure 12.1: The reflective property of conics, their directrices, and foci.

- Conics in polar forms. Consider a conic with a focus at the origin, and a directrix

	Top $y = p$	
Left $x = -p$		Right $x = p$
	Bottom $y = -p$	

for some $p > 0$. In each respective case, the polar equation for the conic is given by

	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 $r = \dfrac{ep}{1 - e\sin(\theta)}$	

G.C. Skills

The parametric forms of the various conics can be found within the GC; no memorisation is needed!

$$\texttt{apps} \implies \texttt{2:Conics} \implies \texttt{mode} \implies \texttt{PAR} \implies \texttt{y=} \implies \begin{matrix} \texttt{1:CIRCLE} \\ \texttt{2:ELLIPSE} \\ \texttt{3:HYPERBOLA} \\ \texttt{4:PARABOLA} \end{matrix}.$$

Note

- The area of an ellipse is given by πab.
- The *major/minor axes* of an ellipse refer to the longest/shortest diameter of the ellipse.
- The *semi-major/semi-minor axes* of an ellipse refer to the longest/shortest radius of the ellipse.
- A *focal radius* of a conic is the distance from a point on the conic to a focus of the conic.

Note

Some possible things to try:

- Using $PF_1 + PF_2 = 2a$ to form simultaneous equations.
- Finding the polar form of a given conic (when $e < 1$ so $r \geq 0$) to solve for distances.
- Congruent/Similar triangles.
- Classic use of discriminants.
- Vieta's formula.

Example 12.1

Describe the significance of the reflective property of parabolas in relation to the reception of the satellite TV.

- As the satellite is very far from the parabolic dish, the rays coming towards the dish can be taken to be parallel rays.

- These rays would all be reflected by the dish and converge at the focus. So, the receiver so should be placed at the focus to maximise the reception of the satellite TV.

Chapter 13

Functions

General Information

1. The 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. $D_{gf} = D_f$ & $R_{gf} \subseteq R_g$.

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

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

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

Note

A useful fact about cross products. For any $a_i, b_i \in \mathbb{R}$:

$$\left\| \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \times \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} \right\| = \begin{vmatrix} 1 & a_1 & b_1 \\ 1 & a_2 & b_2 \\ 1 & a_3 & b_3 \end{vmatrix}.$$

Lines	Planes
Equivalent Forms	
1. Vector Equation: $$\ell \colon \mathbf{r} = \mathbf{a} + \lambda \mathbf{m},\ \lambda \in \mathbb{R},$$ 2. Cartesian Equation: $$\frac{x-a_1}{m_1} = \frac{y-a_2}{m_2} = \frac{z-a_3}{m_3}.$$	1. Vector Equation: $$\Pi \colon \mathbf{r} = \mathbf{a} + \lambda \mathbf{m_1} + \mu \mathbf{m_2},\ \ \lambda, \mu \in \mathbb{R}.$$ 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 $\mathbf{n} \coloneqq \begin{pmatrix} n_1 & n_2 & n_3 \end{pmatrix}^\top$.
Foot of Perpendicular	
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) $\overrightarrow{AN} = \left(\overrightarrow{AQ} \cdot \widehat{\mathbf{m}}\right) \widehat{\mathbf{m}}$, (b) $\overrightarrow{ON} = \overrightarrow{OA} + \overrightarrow{AN}$.	M1: (a) $\overrightarrow{ON} = \overrightarrow{OQ} + \lambda \mathbf{n}$, (b) $\overrightarrow{ON} \cdot \mathbf{n} = p$, solve for λ, (c) Substitute λ into (a). M2: (a) $\overrightarrow{QN} = \left(\overrightarrow{QA} \cdot \widehat{\mathbf{n}}\right) \widehat{\mathbf{n}}$, (b) $\overrightarrow{ON} = \overrightarrow{OQ} + \overrightarrow{QN}$.

<table>
<tr><th colspan="2">Shortest Distance of Point To Line, QN</th></tr>
<tr>
<td>
M1: $\left\| \overrightarrow{AQ} \times \widehat{\mathbf{m}} \right\|$.

M2: (a) $AN = \left\| \overrightarrow{AQ} \cdot \widehat{\mathbf{m}} \right\|$,
(b) Pythagoras' Theorem.

M3: Using the foot of perpendicular, find distance QN.
</td>
<td>
M1: $\left\| \overrightarrow{AQ} \cdot \widehat{\mathbf{n}} \right\|$.

M2: Distance of plane to *origin*:
If $\Pi\colon \mathbf{r} \cdot \mathbf{n} = p$, then $\dfrac{p}{\|\mathbf{n}\|}$ is the shortest distance from the origin to the plane Π.
Note:
• If $\dfrac{p}{\|\mathbf{n}\|} > 0$, then Π is 'above' the origin.
• If $\dfrac{p}{\|\mathbf{n}\|} < 0$, then Π is 'below' the origin.

M3: Using the foot of perpendicular, then find distance QN.
</td>
</tr>
<tr><th>The Relationship Between Two Lines</th><th>The Relationship Between Two Planes</th></tr>
<tr>
<td>
1. Parallel, Non-Intersecting
(a) $\mathbf{m_1} \parallel \mathbf{m_2}$,
(b) Solving $\mathbf{r_1} = \mathbf{r_2}$ gives no real solution. Or, show that $\mathbf{a_1}$ does not lie in ℓ_2.

2. Parallel, Coinciding
(a) $\mathbf{m_1} \parallel \mathbf{m_2}$,
(b) $\mathbf{a}$ lies in ℓ_1 and ℓ_2.

3. Non-Parallel, Intersecting
(a) $\mathbf{m_1}$ not $\parallel$ $\mathbf{m_2}$,
(b) Solve $\mathbf{r_1} = \mathbf{r_2}$ to find intersection.

4. Skew Lines
(Non-Parallel, Non-Intersecting)
(a) $\mathbf{m_1}$ not $\parallel$ $\mathbf{m_2}$,
(b) Solving $\mathbf{r_1} = \mathbf{r_2}$ gives no real solution.
</td>
<td>
1. Distinct Parallel Planes:
(a) Show that $\mathbf{n_1} \parallel \mathbf{n_2}$,
(b) Find a vector $\mathbf{b}$ for which
(i) $\mathbf{b} \cdot \mathbf{n_1} = p_1$,
(ii) $\mathbf{b} \cdot \mathbf{n_2} \neq p_2$.

2. Same Plane:
(a) Show that $\mathbf{n_1} \parallel \mathbf{n_2}$,
(b) Find a vector $\mathbf{b}$ for which
(i) $\mathbf{b} \cdot \mathbf{n_1} = p_1$,
(ii) $\mathbf{b} \cdot \mathbf{n_2} = p_2$.

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

The Relationship Between A Line and A Plane		
1. ℓ lies in Π M1: i. Show $\mathbf{m} \cdot \mathbf{n} = 0$ so $\ell \parallel \Pi$. ii. Then $\mathbf{a}_\ell \cdot \mathbf{n} = p$ tells us ℓ lies in Π. M2: Substitute ℓ into Π and show the system (of lin eqns) is consistent for all λ. 2. $\ell \parallel \Pi$ but nonintersecting M1: i. Show $\mathbf{m} \cdot \mathbf{n} = 0$ so $\ell \parallel \Pi$. ii. Then $\mathbf{a}_\ell \cdot \mathbf{n} \neq p$ tells us ℓ and Π are nonintersecting. M2: Substitute ℓ into Π, and show the system (of lin eqns) is inconsistent. 3. Intersect at one point (a) Check that $\mathbf{m} \cdot \mathbf{n} \neq 0$. (b) 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
$\theta = \cos^{-1} \lvert \widehat{\mathbf{m}}_\mathbf{1} \cdot \widehat{\mathbf{m}}_\mathbf{2} \rvert$.	$\theta = \sin^{-1} \lvert \widehat{\mathbf{m}} \cdot \widehat{\mathbf{n}} \rvert$.	$\theta = \cos^{-1} \lvert \widehat{\mathbf{n}}_\mathbf{1} \cdot \widehat{\mathbf{n}}_\mathbf{2} \rvert$.

Note

Describe the line $\ell\colon \mathbf{r} = \mathbf{a} + \lambda\mathbf{m}, \lambda \in \mathbb{R}$ geometrically.

It is the line *passing through the fixed point* with position vector $\mathbf{a}$ and is *parallel to* $\mathbf{m}$.

Note

Let π be a plane containing the line ℓ and P be a point. The foot of perpendicular of P on π is not necessarily the same as the foot of perpendicular of P on ℓ.

Note

When finding the plane π that is 'above' another plane $\Pi\colon \mathbf{r} \cdot \mathbf{n} = p$ by a constant q units, take note of the direction of $\mathbf{n} = \begin{pmatrix} n_1 & n_2 & n_3 \end{pmatrix}^\top$. If $n_3 > 0$, then $\mathbf{n}$ is pointing 'upwards'. So, $\pi\colon \mathbf{r} \cdot \widehat{\mathbf{n}} = +q$. Otherwise $n_3 < 0$, meaning $\mathbf{n}$ is pointing 'downwards'. Hence, $\pi\colon \mathbf{r} \cdot \widehat{\mathbf{n}} = -q$.

I'm slightly skeptical that Cambridge will not make clear which side is considered 'upwards/downwards'. But, just in case, this is a good point to note.

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:

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

$$e^{\int P(x)\,dx}\frac{dy}{dx} + ye^{\int P(x)\,dx}P(x) = Q(x)e^{\int P(x)\,dx},$$

$$ye^{\int P(x)\,dx} = \int Q(x)e^{\int P(x)\,dx}\,dx.$$

Example 17.1: Justification for D.E. models of real world scenarios

Consider the proportion u of susceptible individuals — those who have not yet caught the disease — and the proportion v of infected individuals and the proportion w of recovered individuals. The standard model used in measuring the rate of spread of the disease assumes that $du/dt = -kuv$, where k is a positive constant; the unit of time used is the average length of time for which a person is infected.

(a) Describe the relevance of each bracketed term in the equation $du/dt = (-k)(uv)$ in terms of the spread of the disease.

(b) Justify the assertion that $dw/dt \approx v$.

(a) The uv term measures the interaction between susceptibles and infected. This should be directly proportional to du/dt: When interactions between susceptibles and infected are more common ($uv > 0$ is larger), more susceptibles should become infected (du/dt is more negative). The proportionality constant $-k$ between uv and du/dt is negative to illustrate this relationship; the proportion of susceptibles cannot increase when population size is constant.

(b) Since the unit of time used in the model is the average length of time for which a person is infected, if there are I infected individuals at time t, then after one unit time, roughly I people would have recovered. i.e. $dR/dt \approx I$ so $dw/dt \approx v$, where R is the number of people who have recovered.

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 17.2

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

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

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 17.3

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

Example 17.4

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 17.5

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: It is More accurate as it takes the mean of the initial and next gradient.

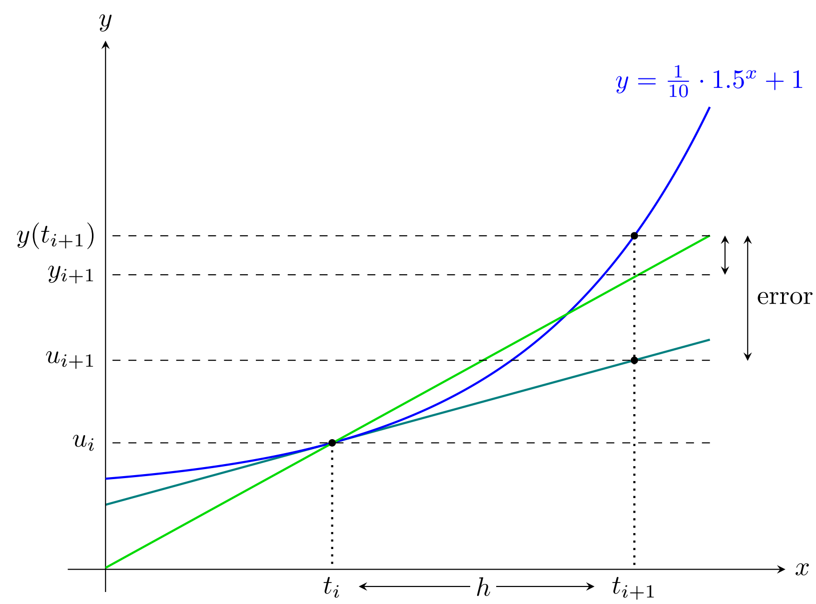


Figure 17.1: An illustration of Euler's Method and the Improved Euler's Method.

Note

To find the percentage difference of the true value and the estimated value, we calculate

$$\frac{\text{true value} - \text{estimated value}}{\text{true value}} \cdot 100\%.$$

Note

Explain the large discrepancy between the approximations yielded by Euler's Method and the Improved Euler's Method.

- There is a large increase in the gradient of the curve from [the value of $y'(a)$] at $x = a$ to [the value of $y'(b)$] at $x = b$.
- So, the average of the two gradients used in the Improved Euler's Method is significantly larger than [the value of $y'(a)$] used in Euler's Method.
- Hence, a large discrepancy between the results of the two methods is observed.

Note

State the condition when the Improved Euler's Method is not an improvement over Euler's Method

When the gradient function is a constant throughout the allocated interval.

Example 17.6: Comparing three sets of approximations.

(i) Obtain the solution of the differential equation

$$\frac{dy}{dx} - y = x$$

given that $y = 0$ when $x = 0$. Obtain, correct to 3 significant figures, the values of y when $x = 0.1$ and $x = 0.2$.

Now consider the differential equation

$$\frac{dy}{dx} - \sin(y) = x,$$

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

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

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

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

(i) $y(0.1) = 0.00517$ and $y(0.2) = 0.0214$.

(ii) $y(0.1) \approx 0$ and $y(0.2) \approx 0.01$.

(iii) $y(0.1) \approx 0.005$ and $y(0.2) \approx 0.210$.

(iv)
- Since the values of y are small, $\sin(y) \approx y$ by small angle approximation. [1]
- Hence, the values of y from (i) can be compared with estimates from (ii) and (iii).
- Since the answers from (iii) are closer to those in (i) than the those in (ii), the answers in (iii) are more accurate estimates of the exact values of y. [1]

Example 17.7

The function $y = y(x)$ satisfies $\frac{dy}{dx} = \frac{1}{10}(\sin(x) - xy)$.

(i) The value of $y(h)$ is to be found, where h is a small positive number, and $y(0) = 0$.

(a) Use two steps of Euler's method to determine an approximation to $y(h)$ in terms of h.

(b) Use one step of the improved Euler formula to find alternative solutions to $y(h)$ in terms of h.

(ii) ★ Show that $y = y(x)$ satisfies $e^{0.05h^2} y(h) = \int_0^h 0.1e^{0.05x^2} \sin(x)\, dx$.

Proof. We see that

$$\begin{aligned} \frac{dy}{dx} &= \frac{1}{10}[\sin(x) - xy] \\ \frac{dy}{dx} + \left(\frac{1}{10}x\right) &= \frac{1}{10}\sin(x) \end{aligned}$$

Let I.F. be $e^{\int \frac{1}{10}x\, dx} = e^{0.05x^2}$:

$$\begin{aligned} e^{0.05x^2} y &= \int \frac{1}{10} e^{0.05x^2} \sin(x)\, dx \\ \left[e^{0.05x^2} y\right]_0^h &= \int_0^h \frac{1}{10} e^{0.05x^2} \sin(x)\, dx \\ e^{0.05h^2} y(h) - e^0 y(0) &= \int_0^h \frac{1}{10} e^{0.05x^2} \sin(x)\, dx \end{aligned}$$

Since $y(0) = 0$,

$$e^{0.05h^2} y(h) = \int_0^h \frac{1}{10} e^{0.05x^2} \sin(x)\, dx.$$

□

(iii) Use the fact that h is small to estimate $\int_0^h 0.1e^{0.05x^2} \sin(x)\, dx$. Hence find another approximation to $y(h)$ in terms of h.

(iv) ★ Discuss the relative merits of the three methods employed to obtain these approximations.

- Two step of Euler's method might be lead to a more accurate approximation than one step of the improved Euler method, because of the larger number of steps.
- The improved Euler method might lead to a more accurate approximation, as it takes the mean of the gradients $y'(0)$ and $y'(h)$, rather than the simplistic tangent line approximation used by Euler's method.
- The method in (iii) might be more accurate than Euler's method and the improved Euler method, because it uses quadratic polynomials/an exponential curve[a] instead of straight line segments.

[a] It depends on how you calculated the approximation to $y(h)$ in (iii).

17.2 Second Order D.E.

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

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 17.8

1. If $y_c = Ae^{-3x}$ and $f(x) = 10e^x$, then $y_p = ke^x$
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 $y_p = kx^2e^x$.

Note

R-formulas. Let $a, b \in \mathbb{R}$. Then, for

$$R := \sqrt{a^2 + b^2} \qquad \text{and} \qquad \tan(\alpha) := \frac{b}{a},$$

we have that

$$a\sin(\theta) + b\cos(\theta) = R\sin(\theta + \alpha) \qquad \text{and} \qquad b\sin(\theta) + a\cos(\theta) = R\cos(\theta - \alpha).$$

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

$$P(t) = P_0 e^{kt}.$$

[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 17.9: Neat trick of letting $A = \pm$constant

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

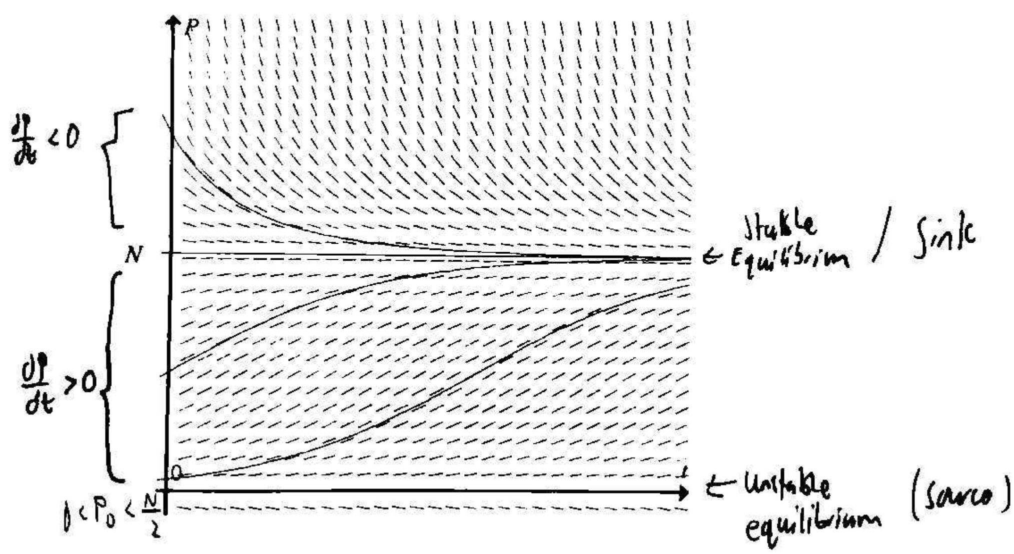


Figure 17.2: Logistics curve

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 $0 \leq H < \frac{kN}{4}$, 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:

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

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."

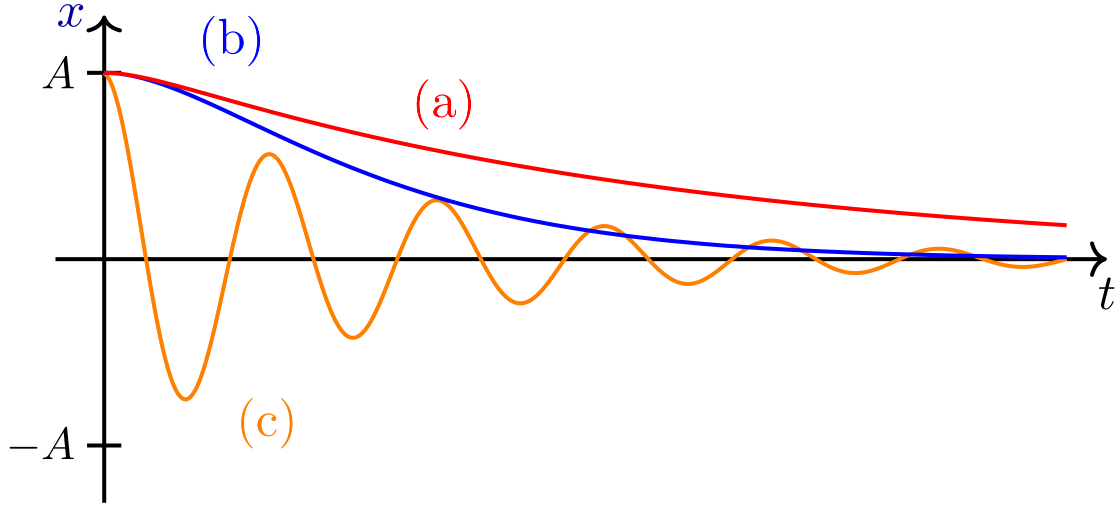


Figure 17.3: Oscillatory behaviors

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
$$\mathrm{Var}(X) \coloneqq E\big(X^2\big) - [E(X)]^2 = \mathrm{E}\left((X-\mu)^2\right).$$

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) $\mathrm{Var}(aX + bY + c) = a^2\,\mathrm{Var}(X) + b^2\,\mathrm{Var}(Y)$.

5. Probability Distribution Table:

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

Table 18.1: A probability distribution table.

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

(The last two represent the memorylessness of the geometric distribution.)

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 $X \sim \mathrm{Po}(\lambda)$, iff

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

Note

We can assume $Y \sim \mathrm{Po}(\lambda)$ 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 $U \sim \mathrm{Po}(\mu)$ and $V \sim \mathrm{Po}(\lambda)$ are *independent* variables, then

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

Note

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

Note

Explain why it might be inappropriate to model X using a Poison distribution.

$$\bar{x} = \underline{\quad\quad} \qquad\qquad s^2 = \underline{\quad\quad}$$

If $X \sim \mathrm{Po}(\lambda)$ is an appropriate model for some λ, then the population mean and population variance of X should coincide. But this may not be true, because the sample mean differs significantly from the unbiased estimate for sample variance. Hence, a Poisson distribution may be an inappropriate model for X.

Example 19.1: Validity of a Poisson model.

Let $\mathscr{X}$ be a discrete random variable and consider the approximation/model $X \sim \mathrm{Po}(\lambda)$. Suppose that $\mathrm{P}(\mathscr{X} = m) \approx \mathrm{P}(X = n)$ and $\mathrm{P}(\mathscr{X} = n) \approx \mathrm{P}(X = n)$ for some integers m and n. Comment on whether this information validates the use of the Poisson model?

No: two specific outcomes are not enough to validate the use of the Poisson model. Instead, we need to consider all possible outcomes.

General Information

1. Expectation and Mean:

Distribution	Expectation	Variance
$X \sim \mathrm{B}(n, p)$	np	$np(1-p)$
$Y \sim \mathrm{Po}(\lambda)$	λ	
$W \sim \mathrm{Geo}(p)$	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 $\mathrm{P}(X = n) > \mathrm{P}(X = n - 1)$ and $\mathrm{P}(X = n) > \mathrm{P}(X = n + 1)$. This gives the *strictest* range of values that p can take (Fig 17.1).

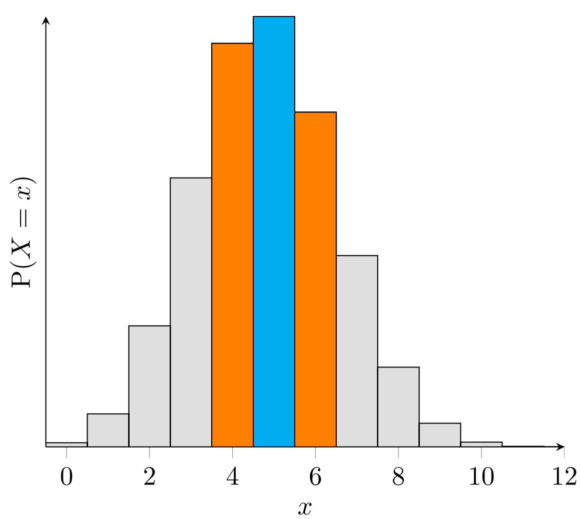


Figure 19.1: The histogram for a binomial distribution $X \sim \mathrm{B}(10, p)$ that has mode $m = 5$.

Example 19.2: 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 density 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 $\mathrm{P}(a \leq X \leq b) = \int_a^b f(x)\,dx$. 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

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

- 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 $Y(X) \leq y$ to find an inequality that relates X with y. E.g. $e^X \leq y$ iff $X \leq \ln(y)$.

Note

The domain conversions for the 'pieces' of the cdf should be shown very clearly.

Example 20.1: Clear workings for domain conversion.

Let $\ln(Y) = -b + bX$, where $1 \leq x \leq 10$. Then,

$$\begin{array}{ccccc} 1 & \leq & x & \leq & 10 \\ 0 & \leq & -b+bx & \leq & 9b \\ 1 & \leq & y & \leq & e^{9b} \end{array}$$

Note

It is important to check whether $Y(X)$ is an increasing or decreasing function.

Example 20.2: An increasing $Y(X)$.

Let $Y = e^X$. Then,

$$P(Y \le y) = P(e^X \le y) = P(X \le \ln(y)).$$

Example 20.3: A decreasing $Y(X)$.

Let $Y = e^{-X}$. Then,

$$P(Y \le y) = P(e^X \le y) = P(X \ge \ln(y)).$$

Example 20.4: Another decreasing $Y(X)$.

Let $X \sim U(0, \pi/2)$ and $Y = \cos(X)$. Then,

$$P(Y \le y) = P(\cos(X) \le y) = P(X \ge \cos^{-1}(y)) \ne P(X \le \cos^{-1}(y)).$$

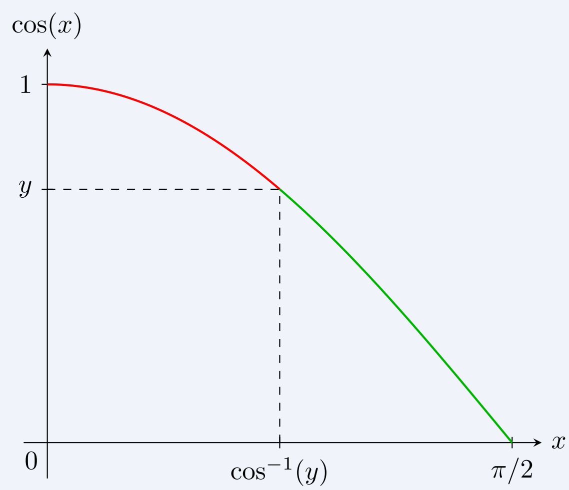


Figure 20.1: The graph of $\cos(x)$ against x. Notice that $[\cos^{-1}(y), \pi/2] \overset{\cos}{\to} [0, y]$, while $[0, \cos^{-1}(y)] \overset{\cos}{\to} [y, 1]$.

- A *median* of X is any value m such that $P(X \le m) = F(m) = 1/2$.
- Mean/Expectation:

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

- Important property:

$$E(aX + bY + c) = a\,E(X) + b\,E(Y) + c.$$

- Variance:

$$\mathrm{Var}(X) := E(X^2) - [E(X)]^2 = E\big((X-\mu)^2\big).$$

- Important property:

$$\mathrm{Var}(aX + bY + c) = a^2\,\mathrm{Var}(X) + b^2\,\mathrm{Var}(Y).$$

Example 20.5: Contextual meaning of a pdf's shape.

Consider the proportion X of cloud cover (i.e. the proportion of the sky that is obscured by the cloud) in one randomly chosen evening. It has the following pdf

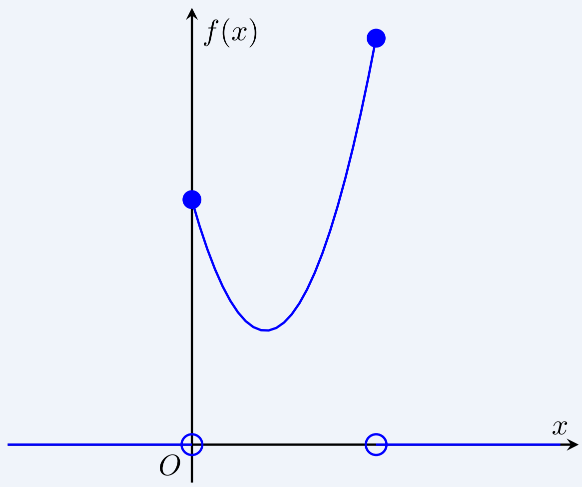


Figure 20.2: Parabolic pdf.

Explain what the shape of the pdf indicates about the cloud cover at noon at the weather station. [2]

- Extremes of cloud cover (i.e. proportion of cloud cover close to 0 or 1) are more common than intermediate values, with [1]
- a high proportion of cloud cover being more common than a low proportion of cloud cover. [1]

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, $y \in \mathbb{R}$,
  - $nX \sim \mathrm{N}(n\mu, n^2\sigma^2)$,
  - $X_1 + X_2 + \cdots + X_n \sim \mathrm{N}(n\mu, n\sigma^2)$,
  - $aX \pm bY \sim \mathrm{N}(a\mu \pm bm, a^2\sigma^2 + b^2s^2)$.

- 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`?

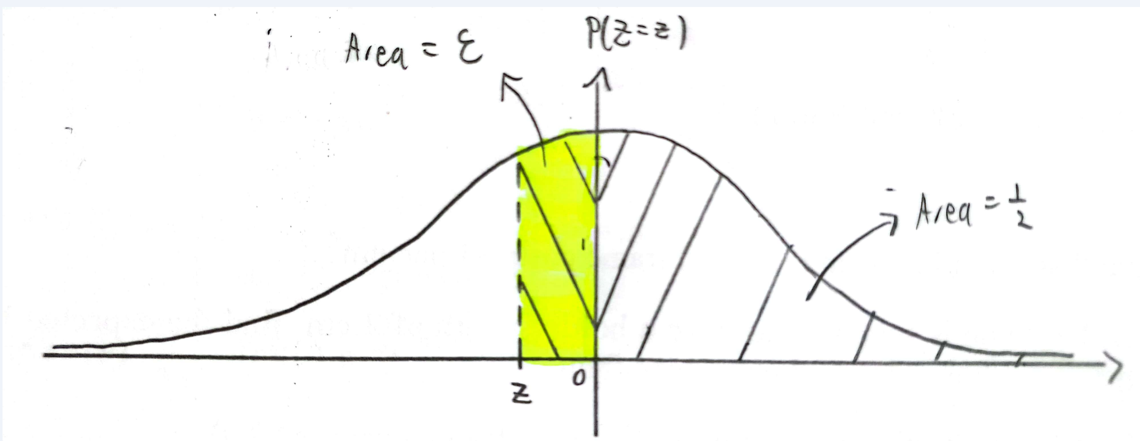
$\mathrm{P}(X < x) = p$	`LEFT`
$\mathrm{P}(-x < X < x) = p$	`CENTER`
$\mathrm{P}(X > x) = p$	`RIGHT`

- 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,

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

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

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

(An exponential distribution models time between occurrences.)

Note

The memorylessness of the exponential distribution. Let $Y \sim \text{Exp}(\lambda)$, then

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

- Expectation and variance:

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

  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,
$$\mathrm{P}(T > t) = \mathrm{P}(\# = 0) = e^{-4t}.$$
  As such, the probability that there is at least one arrival in an interval of time t is
$$\mathrm{P}(T \leq t) = 1 - e^{-4t}.$$

Note

The exponential distribution *begins from zero*. But, contextually, we may begin counting from *one*! Hence, we need to be careful of what bounds to use in probability calculations.

Example 21.3

Find the probability p that the company received the first response in the third hour of the day. (It is implied the that there's no zeroth hour.)

Let $T \sim \text{Exp}(\lambda)$. Then, $p = \mathrm{P}(2 \leq T < 3) \neq \mathrm{P}(3 \leq T < 4)$.

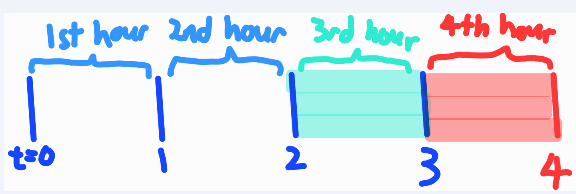


Figure 21.1

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(s)
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 population mean μ, where n is large. Then, an approximate $(1-\alpha)100\%$ confidence interval for μ is
$$\left(\bar{x} - z_{1-\alpha/2}\frac{s}{\sqrt{n}}\,,\ \bar{x} + z_{1-\alpha/2}\frac{s}{\sqrt{n}}\right).$$
When the population 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 a population parameter θ has likely increased/decreased, when given suitable confidence intervals.

1. There is no conclusive result.

   As the old $(1-\alpha)100\%$ confidence interval has a lower/higher central value of $\widehat{\theta}_1 < \widehat{\theta}_2$ than the new $(1-\alpha)100\%$ confidence interval, there is some evidence to suggest that [θ in context] has increased/decreased. However, this evidence is weak: Since the old and new $(1-\alpha)100\%$ confidence intervals overlap, we are unable to conclude whether the [θ 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 $(1-\alpha)100\%$ confidence interval is to the left/right of the new $(1-\alpha)100\%$ confidence interval, such that they do not overlap. So, can conclude that the [θ 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 $(1-\beta)100\%$ confidence interval compared to a $(1-\alpha)100\%$ confidence interval, where $\beta < \alpha$.

Advantage:	A $(1-\beta)100\%$ CI is more likely to contain the true mean. So, any result made on the assumption that the true mean is contained in the $(1-\beta)100\%$ CI is more likely to hold true, than results made assuming the true mean is contained in the $(1-\alpha)100\%$ CI.
Disadvantage:	A $(1-\beta)100\%$ CI is wider. Thus, using the $(1-\beta)100\%$ CI gives us less accuracy about the location of the true value of the mean. Accordingly, stronger results can be concluded by assuming that the true mean is contained in the $(1-\alpha)100\%$ CI, than can be concluded by assuming the true mean is contained in the $(1-\beta)100\%$ CI.

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

Note: combining two samples into one.

Let X be a random variable with population mean μ. We are given two independent random samples A and B, of sizes m and n, respectively. Find a $100(1-\alpha)\%$ confidence interval for μ, based on the combined sample C.

$$\bar{x}_C = \frac{\sum x_A + \sum x_B}{m+n} = \frac{m\bar{x}_A + n\bar{x}_B}{m+n}$$

$$s_C^2 = \frac{1}{m+n-1}\left(\sum x_A^2 + \sum x_B^2 - \frac{(\sum x_A + \sum x_B)^2}{m+n}\right) \neq \frac{(m-1)s_A^2 + (n-1)s_B^2}{m+n-2}.$$

Proceed with the usual calculations to find the $100(1-\alpha)\%$ confidence interval based on C.

Chapter 23

Statistics: Hypothesis Testing

23.1 Preliminary Definitions

Definition 23.1

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

Note

To check we have correctly stated the hypotheses in our hypothesis test, ensure that they are contrasting. For example, we are testing A's claim that $\mu > \mu_0$:

Test	$H_0\colon \mu = \mu_0$	(A's claim is false)
against	$H_1\colon \mu < \mu_0$	(A's claim is false)
at the $100\alpha\%$ significance level.		

Both hypotheses result in A's claim being false. Hence, the hypotheses have been stated erroneously ×.

Test	$H_0\colon \mu = \mu_0$	(A's claim is false)
against	$H_1\colon \mu > \mu_0$	(A's claim is true)
at the $100\alpha\%$ significance level.		

The hypotheses are contrasting. So, they have been correctly stated ✓.

General Information

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

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

  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 we conclude [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].

23.2 One Sample z-Test

General Information

There are various combinations of assumptions for which this test applies, and two variants of the corresponding test statistic. For brevity, we shall avoid restating it, instead directing the reader to table 23.2.

1. Let [X in context] and μ be its 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, since the sample size $n =$ ___ (≥ 50), by CLT we have $\overline{X} \sim \mathrm{N}(\mu_0, s^2/n)$ approximately.
4. Test statistic:
$$Z = \frac{\overline{X} - \mu_0}{s/\sqrt{n}} \sim \mathrm{N}(0,1) \qquad \text{approximately,}$$
where $\bar{x} =$ ___ and $s =$ ___.

- Find $z_{1-\alpha}$ or $z_{1-\alpha/2}$, which satisfies
  1. $\mathrm{P}(Z < z_\alpha) = \alpha$,
  2. $\mathrm{P}(-z_{1-\alpha/2} < Z < z_{1-\alpha/2}) = 1 - \alpha$, or
  3. $\mathrm{P}(Z > z_{1-\alpha}) = \alpha$.
- Find the test statistic value
$$z_{\text{calc}} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}.$$
- Reject H_0 iff
  1. $z_{\text{calc}} < z_\alpha$,
  2. $|z_{\text{calc}}| > z_{1-\alpha/2}$, or
  3. $z_{\text{calc}} > z_{1-\alpha}$.

- Find the p-value using GC.
- 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...`

5. Since (a) $z_{\text{calc}} < z_\alpha$, (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

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.

Note

Explain why a two-tail test should be used instead of a one-tail test.

> We want to see if [μ_X in context] has changed, which could be greater than or less than the original population mean time.

23.3 Student's t-Distribution

Definition 23.4

A 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 $X \sim t(\nu)$.

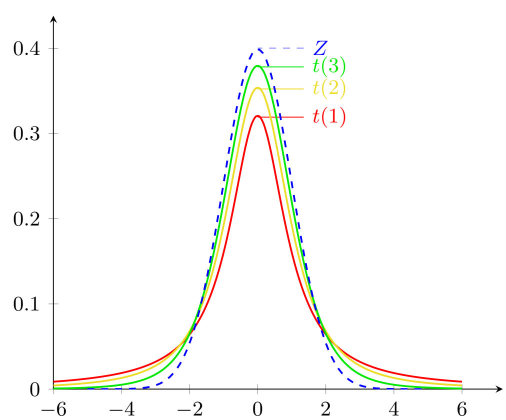


Figure 23.1: Student's t-distribution compared to the standard normal distribution.

General Information

- 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

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

- Suppose we are conducting the following test:

Test	$H_0\colon \mu = \mu_0$
against	$H_1\colon \mu \neq \mu_0$
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...`

23.4 One Sample t-Test

General Information

Again, see table 23.2 for the necessary assumptions.

1. Let [X in context] and μ be its 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
   $$T = \frac{\overline{X} - \mu}{s/\sqrt{n}} \sim t(n-1),$$
   where $\bar{x} =$ ___ and $s =$ ___.
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 $\mathrm{P}(T < t) = A$.

23.5 Two Sample z-Test

General Information

Again, see table 23.3 for the necessary assumptions.

1. Let [X_1, X_2 in context]. Also let μ_1 and μ_2 be the population mean of X_1 and X_2, respectively.

2. | | |
   |---|---|
   | Test | $H_0: \mu_1 - \mu_2 = c$ |
   | against | $H_1:$ (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)
   $$Z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim \mathrm{N}(0, 1)$$
   where $\bar{x}_1 = \_\_\_$ and $\bar{x}_2 = \_\_\_$.

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

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

   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.

Recall

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

23.6 Two Sample t-Test

General Information

Again, see table 23.3 for the necessary assumptions.

1. Let [X_1, X_2 in context]. Also let μ_1 and μ_2 be the population mean of X_1 and X_2, respectively.
2. Test $H_0: \mu_1 - \mu_2 = c$ against $H_1:$ (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
$$T = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2),$$
where $\bar{x}_1 = \_\_\_$, $\bar{x}_2 = \_\_\_$, and $s_p^2 = \_\_\_$.
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`

23.7 Paired Sample t-Test

General Information

Again, see table 23.3 for the necessary assumptions.

1. Let D = [X in context] − [Y in context], and μ_D be the population mean.
2. Test $H_0: \mu_D = \mu_0$ against $H_1:$ (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.
- A two-sample t-test is easier to conduct, because we need not keep track of which [contextual index] is being used for which pair and carefully order the pairs of data each in their own columns.
- A two-sample t-test is faster to implement because we only need to go through a single round of data collection, rather than the two rounds needed for a paired sample t-test.

Note

Suggest how could the data be organised if a paired sample t-test were to be used.

For a paired sample t-test, the data must be paired according to [the contextual indexing]. Thus, [the contextal data pair] must be recorded according to [the contextual indexing].

Note

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

Pairing eliminates the *factor of variability* between different [the contextual indexing, e.g. ages]. Statistically, there is also no need to assume that [X in context] and [Y in context] have exactly the *same population variance*, which is a necessary assumption for a two sample t-test.

Note

If it were required to test whether the [population mean μ_1 of X_1 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_1].

It would be correct to use the estimate of variance based on [sample from the distribution of X_1], since the test statistic

$$T = \frac{\overline{X}_1 - \mu_1}{s/\sqrt{n}} \sim t(n-1).$$

involves only the [sample from the distribution of X_1].

Note

What if

1. the question does not give you a level of significance to conduct your test at?
2. (afterwards) it asks you to explain what it indicates about the strength of the evidence?

Test	$H_0:$
against	$H_1:$
at the $\alpha\%$ significance level.	

1. Since p-value $=$ ___, we reject H_0 if and only if $\overbrace{100 \cdot p\text{-value}}^{\text{the number}} < \alpha \leq 100$. There is sufficient evidence, at the $\alpha\%$ level of significance, to conclude tha [H_1 in context] if and only if $\underbrace{100 \cdot p\text{-value}}_{\text{the number}} < \alpha \leq 100$.
2. Since the p-value is quite small/big, it indicates that there is strong/weak evidence that [H_1 in context].

23.8 Summary

Throughout table 23.2, assume that the sample used for each test is random. Square brackets indicate "and", while round brackets indicate "or".

Assumptions/Reasons	Test (Statistic)
[ii] The population 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 population 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 population 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 population variances σ_1 and σ_2 are known. [ii](1) Sample sizes n_1 and n_2 are large (so CLT applies). [ii](2) At least one 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] Either σ_1^2 or σ_2^2 is unknown. [ii] Sample sizes n_1 and n_2 are large. [iii] The population variances σ_1^2 and σ_2^2 do not coincide. [iv] 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] Either σ_1^2 or σ_2^2 is unknown. [ii] Sample sizes n_1 and n_2 are large. [iii] Both population variances σ_1^2 and σ_2^2 coincide. [iv] 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 $Z = \dfrac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim \mathrm{N}(0,1)$ approximately
[i] Either σ_1^2 or σ_2^2 is unknown. [ii] At least one sample size n_1 or n_2 is small. [iii] Both population variances σ_1^2 and σ_2^2 coincide. [iv] Assume X_1 and X_2 are normally distributed. (Or: Both samples come from normal populations.) Write [iii] and [iv] if the question asks for the necessary assumptions.	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

24.1 The χ^2-Distribution

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

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

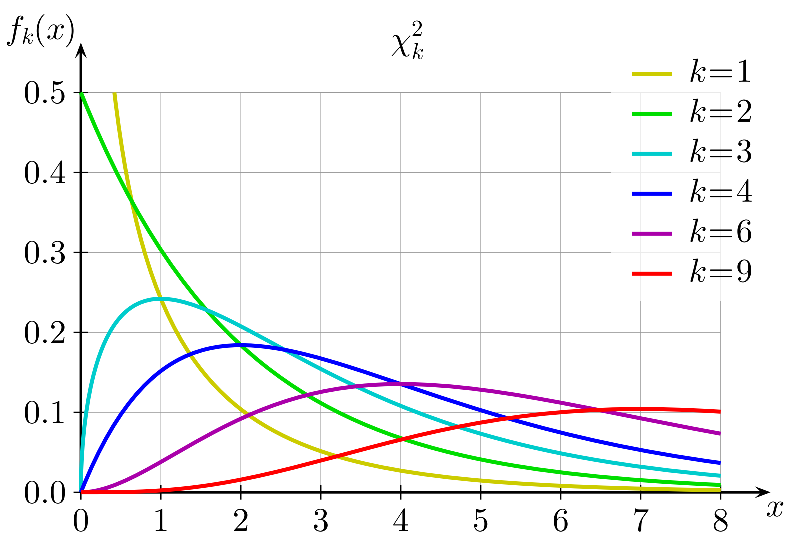


Figure 24.1: Illustration of how the $\chi^2_{(\nu)}$ distribution looks with increasing degree of freedom ν.

General Information

- Properties of chi-squared distributions.
  - $\mathrm{E}(X) = \nu$ 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 $Y \sim \chi^2_{(\upsilon)}$, then $X + Y \sim \chi^2_{(\nu+\upsilon)}$.

24.2 A Goodness-of-Fit Test

General Information

1. Let [X in context].

2. *Note.* Use a pen to draw any necessary tables.

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
f_i	f_1	f_2	$\cdots$	f_n
e_i	e_1	e_2	$\cdots$	e_n
$\frac{(f_i-e_i)^2}{e_i}$	$\frac{(f_1-e_1)^2}{e_1}$	$\frac{(f_2-e_2)^2}{e_2}$	$\cdots$	$\frac{(f_n-e_n)^2}{e_n}$

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

4. Check whether $e_i \geq 5$ for each of the n classes. If it isn't, we need to combine *just enough* adjacent classes, till they do. Working-wise, use some underbraces/overbraces to indicate the combined values.

5. Under H_0, the test statistic is

$$\chi^2 = \sum \frac{(F_i - E_i)^2}{E_i} \sim \chi^2_{(\nu)}.$$

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

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

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:

$$\texttt{stat} \implies \texttt{TESTS} \implies \texttt{D:}\chi^2\texttt{GOF-Test...}$$

Note

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

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 $\nu = n - 1$.

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 values of μ and σ^2 are unknown. In this case, the degree of freedom $\nu = n - 2 - 1 = n - 3$.

Note

Consider when we are testing

Test	$H_0 : X \sim \mathrm{N}(\mu, \sigma^2)$
against	$H_1 : X \nsim \mathrm{N}(\mu, \sigma^2)$
at the $100\alpha\%$ significance level.	

So, we want to fill up the values of e_i below.

x	$a_1 \le x_1 \le a_2$	$a_2 \le x_2 \le a_3$	$\cdots$	$a_n \le x_n \le a_{n+1}$
f_i	f_1	f_2	$\cdots$	f_n
e_i	e_1	e_2	$\cdots$	e_n

Table 24.2: Observed and expected frequencies when testing goodness-of-fit with a normal distribution.

Let the sample size $\sum f_i$ be m. Then, we should calculate $e_1 = m\,\mathrm{P}(-\infty < X \le a_2)$ and $e_n = m\,\mathrm{P}(a_n \le X < \infty)$, instead of $e_1 = m\,\mathrm{P}(a_1 \le X \le a_2)$ or $e_n = m\,\mathrm{P}(a_n \le X \le a_{n+1})$. Similarly, for goodness-of-fit tests with Poisson and Geometric distributions, we must also be careful in ensuring that we account for *all* possible values which X can take on, in calculating e_i.

Note

Suppose we are given a question of the following form.

Some context...

x_i	x_1	x_2	$\cdots$	x_n
f_i	f_1	f_2	$\cdots$	f_n

Table 24.3: Some data.

(i) Show, at the $100\alpha\%$ significance level, that the data does not support the hypothesis of $X \sim \mathrm{Geo}(p)$ with $p = 0.5$.

(ii) State how the test in (i) would have to be amended to test the hypothesis of a geometric distribution for an *unspecified value of p*.

Then, for (ii), two main changes have to be made:

1. Estimate the value of p by computing the sample mean $\bar{x}$ and letting $p = 1/\bar{x}$.
2. Adjust the degree of freedom from 4 to $4 - 1 = 3$, as there is one more restriction, that the mean must agree.

(The phrasing is similar for gof tests for other distributions; simply use the appropriate estimators for the unknown population parameters.)

Note

The χ^2 goodness-of-fit test showed that there is strong evidence for $X \sim \text{Po}(\lambda)$. Suggest a possible reason.

$$s^2 = \_\_\_ \qquad \bar{x} = \_\_\_$$

Since $\bar{x} \approx s^2$, the population mean and population variance of the [X in context] may be approximately equal. This made the data a good fit to the Poisson distribution.

Note

Explain why a test based on a normal distribution might still be valid, despite the χ^2 test-of-independence implying that $X \not\sim \text{N}(\mu, \sigma^2)$.

1. The sample size of ___ is small. Hence, the result of the test may not accurately represent the population.
2. [X in context] might still be normally distributed, but with $\text{E}(X) \neq \mu$ and/or $\text{Var}(X) \neq \sigma^2$.

24.3 Tests of Independence

General Information

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. *Note.* Unless the question asks for it, we do not need to write $\left[\frac{(f_i - e_i)^2}{e_i}\right]$ or its corresponding values, in the following table.

f_i (e_i) $\left[\frac{(f_i-e_i)^2}{e_i}\right]$		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}	$S = \sum t_{r_i} + \sum t_{c_i}$

Table 24.4: *Expected* frequencies for a test of independence.

Remark

The expected frequencies are given by $e_{ij} = \dfrac{\text{row total} \cdot \text{column total}}{\text{total number of observations}} = \dfrac{t_{r_i} t_{c_j}}{S}$.

4. Check whether $e_i \geq 5$ for each of the mn cells. If it isn't, we need to combine *just enough* adjacent classes, till they do. Working-wise, use some underbraces/overbraces/side braces to indicate the combined values.
5. Under H_0, the test statistic is

$$\chi^2 = \sum \frac{(F_i - E_i)^2}{E_i} \sim \chi^2_{(\nu)}.$$

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

6. 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):

`2nd` $\implies$ `x`$^{-1}$ $\implies$ `EDIT` $\implies$ `[A]`.

Then, conduct the test for independence:

`stat` $\implies$ `TESTS` $\implies$ `C:`χ^2`-Test...`

Note

If it's unclear as to what is to be stated as independent/dependent in the hypotheses, consider the expected values and how they relate to the context.

Example 24.4

Consider the following context:

Statement	Independent/Dependent?
There is consistency in the marking of the two T.A.s.	?
There is no consistency in the marking of the two T.A.s.	?

Table 24.5: Two statements on the relationship between the marks awarded and the T.A. marking.

Then, under H_0 — the independence claim — the expected frequencies are as stated below.

	e_{ij}	Grade A	Grade B	Grade C
T.A.	X	a	b	c
	Y	a	b	c

Table 24.6: Expected frequencies.

Since $e_{1j} = e_{2j}$ for all $1 \leq j \leq 3$, we infer the following.

Statement	Independent/Dependent?
There is consistency in the marking of the two T.A.s.	Independent
There is no consistency in the marking of the two T.A.s.	Dependent

Table 24.7: Which statement corresponds to independence and which coresponds to dependence.

Note

If the question says to "use an approximate χ^2-statistic...", then we must use the critical region method. It is incorrect to use the p-value.

Note

Consider when we are asked to state which cells correspond to the highest contributions to the test statistic, and relate that back to the context of the question. Then:

1. State the cells in the form (__, __). E.g. (High, Good) and (Low, Good).

2. In table 24.4, add an asterisk to each of these cells. E.g. $\boxed{1\ (5)\ [10.1]^*}$.
3. Use words that imply correlation and *not* causation. E.g. directly associated, correlates with, etc.

Note

On a similar note, if the question asks "Can it can be concluded that...", but is unclear about whether it's implying correlation or causation, it may be safer to explain both ways. i.e. what correlation is there and why is there no causation.

Note

Explain why we cannot conclude any casual relationships from a test of independence.

> No, the above test does not reflect the actual casual relationship between the two factors, if it exists. Rather, it merely suggests that they are not independent.

Note

Explain why we cannot apply a χ^2-test for independence using the data given.

> The expeceted frequency for (___, ___) is ___ < 5. If we combine the columns, the degree of freedom $\nu = 1 \cdot 0 = 0$. If we combine the rows, $\nu = 0 \cdot 1 = 0$. Thus, we cannot apply a χ^2-test for independence.

Note

Don't be intimidated when a question gives unknown f_i's for multiple cells. By using the formula for e_i, we should be able to rewrite them in terms of one unknown — the observed frequency for one cell. Additional information should be provided, if this is not possible.

Chapter 25

Correlation and Linear Regression

25.1 Scatter Diagrams

Note

Guidelines for drawing a scatter diagram

- 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.

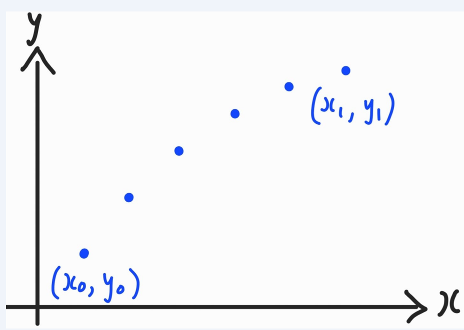


Figure 25.1: An illustration of a scatter plot.

Note. We do not need to start from the origin.

G.C. Skills

To show a scatter plot on the G.C.:

`2nd` $\Longrightarrow$ `y=` $\Longrightarrow$ `1:Plot1...on` $\Longrightarrow$ `enter` $\Longrightarrow$ `on`.

Note. When we no longer need a scatter plot, turn the scatter plot(s) *off* in the G.C., lest it erroneously interferes with other functionalities of the G.C.

Example 25.1

One of the values of t appears to be incorrect. Indicate the corresponding point on your diagram by labelling it P and explain why the scatter diagram for the remaining points may be consistent with a model of the form $y = a + bf(x)$.

With P removed, the remaning points seem to lie, on a curve that [e.g. increases at a decreasing rate], suggesting consistency with the model $y = a + bf(x)$.

25.2 Product Moment Correlation Coefficient r

Definition 25.1

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

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

and takes on a value from 0 to 1. See Figure 25.2 for some scatter plots of various r values.

Note

Explain whether your estimate using the regression line of y on x is reliable.

Since the $|r|$ value of ___ is close to 1, *and* $x =$ ___ is within the data range of ___ $\leq x \leq$ ___, the estimate is reliable.

Note

Explain why the estimate using the regression line y on x is not reliable.

Since $x =$ ___ falls outside of the range of data ___ $\leq x \leq$ ___, we would be extrapolating the observed data points. This makes the estimate of the value of y at $x =$ ___ unreliable.

Note

Explain which dataset would result in a larger absolute value of the product moment correlation coefficient.

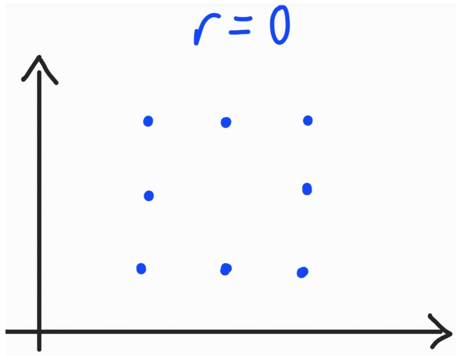
- Set A will have a larger $|r|$ value, because its data points lie relatively *closer to a straight line* (with positive/negative gradient), suggesting a stronger linear correlation.
- Set B's $|r|$ value will be closer to zero, since its data points are *more scattered*, suggesting a weaker linear correlation.

25.3 Regression Lines

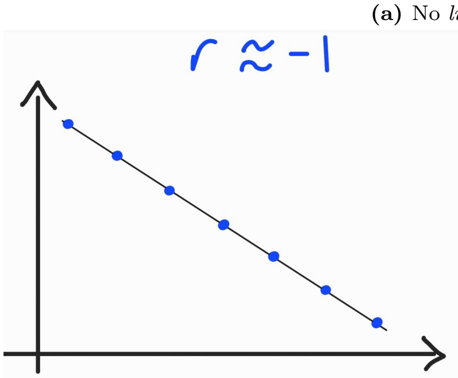
General Information

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

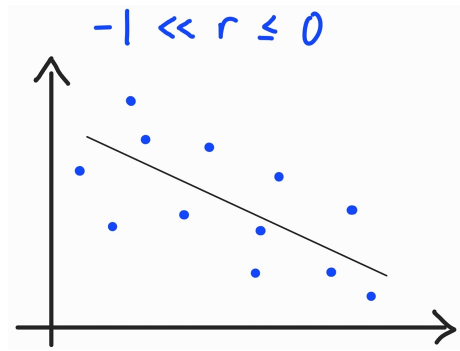
$$y = \bar{y} + b(x-\bar{x}), \qquad \text{where} \qquad b := \frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(x-\bar{x})^2} = \frac{\sum xy - \dfrac{\sum x\sum y}{n}}{\sum x^2 - \dfrac{(\sum x)^2}{n}}.$$



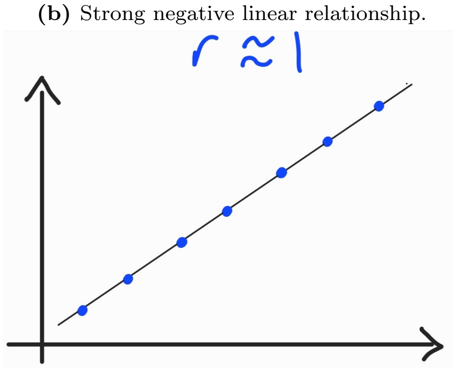
(a) No *linear* relationship.



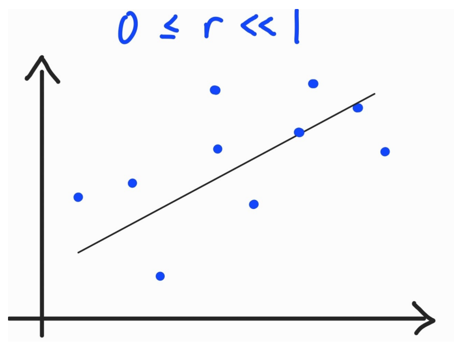
(b) Strong negative linear relationship.



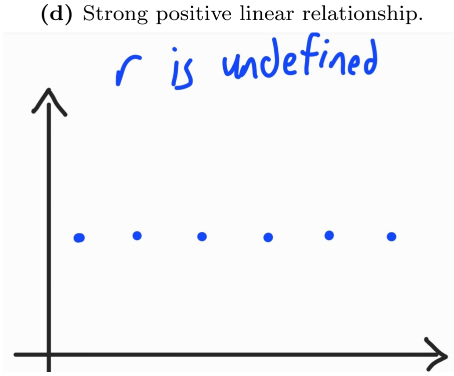
(c) Weak negative linear relationship.



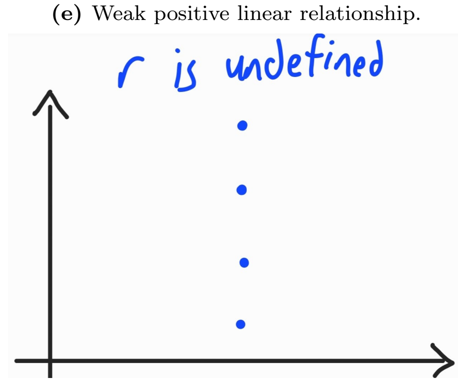
(d) Strong positive linear relationship.



(e) Weak positive linear relationship.



(f) $\sum (y - \bar{y})^2 = 0$.



(g) $\sum (x - \bar{x})^2 = 0$.

Figure 25.2: Example scatter plots with different values of r.

Note. Even though there is no *linear* relationship when $r = 0$, there might be a *nonlinear* relationship present.

G.C. Skills

To find the r-value, or the regression line of y on x, for a given dataset:

`stat` $\Longrightarrow$ `CALC` $\Longrightarrow$ `4:LinReg(ax+b)` or `8:LinReg(a+bx)`

Note

With the aid of diagrams, explain the difference between the least square regression lines of y on x and that of x on y.

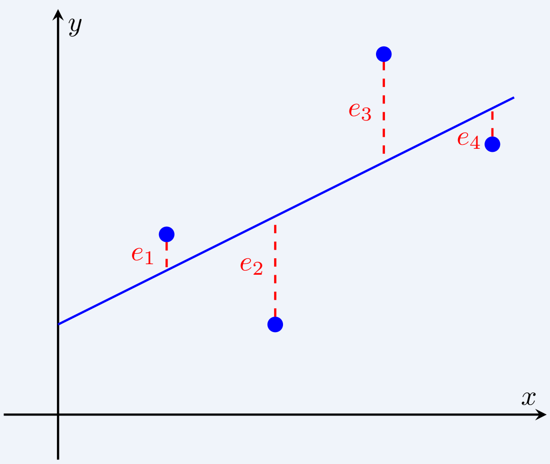


Figure 25.3: Regression line of y on x.

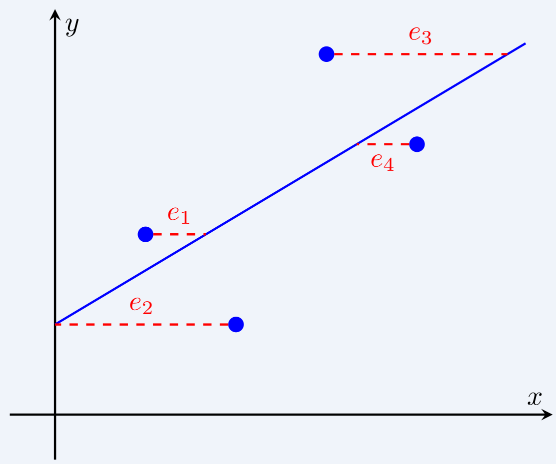


Figure 25.4: Regression line of x on y.

The regression line of y on x assumes that the values of x are known exactly and to perfect accuracy. As such, it minimises the sum of squared distances, $\sum {e_i}^2$ in the y-direction, as shown above.

The regression line of x on y assumes that the values of y are known exactly and to perfect accuracy. As such, it minimises the sum of squared distances, $\sum {e_i}^2$ in the x-direction, as shown above.

Example 25.2

Show on the scatter diagram in part (d) the distances which are used in drawing the least squares regression line of y on x. Explain why these distances are squared, and why this is referred to as the 'method of least squares'.

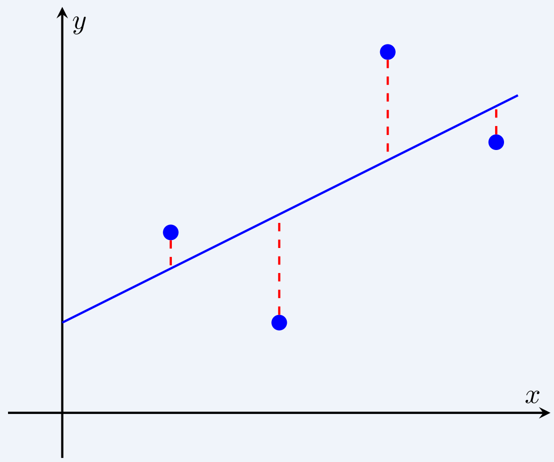


Figure 25.5: Regression line of y on x.

- The distances are signed, i.e. can be positive or negative.
- So, squaring the distances ensures the sum will not become zero (unless the plotted points are collinear) or negative.
- The least squares regression line is the line for which the sum of squared distances is minimised. Hence, this is referred to as the 'method of least squares'

Example 25.3

Suppose that we are given pairs of data for x and y, as shown below:

x	x_1	x_2	$\cdots$	x_n
y	y_1	y_2	$\cdots$	y_n

Table 25.1: A dataset of n pairs of x and y values.

Let Y be the value obtained by substituting a sample value of x into the equation of the regression line of y on x, given by $Y = ax + b$. Consider any $Y' = \alpha x + \beta$. What can you say about the value of $\sum(y - Y')^2$?

Since $\sum(y - Y')^2$ is minimised when $Y' = ax + b$, we see that $\sum(y - Y')^2 \geq \sum(y - Y)^2$ for any $Y' = \alpha x + \beta$.

Note

The regression lines of y on x and x on y intersect at $(\bar{x}, \bar{y})$.

Note

To estimate the value of a variable y, given a the value of another variable x, we always use the regression line of the *dependent* variable on the *independent* variable.

Independent variable	Dependent variable	Regression line
x	y	y on x
y	x	x on y

Table 25.2: Regression line to use for estimations.

Note

Explain why a linear model would not be appropriate. Choose any relevant ones.

- The scatter diagram/data indicates that, as x increases, y [e.g. increases at a decreasing rate], which is *not* a linear relationship.
- A linear model will increase *indefinitely* with more [x in context]. This is contextually *unrealistic*, as [reason in context].
- A linear model would imply that, in the long run, the [e.g. time taken] would be negative, which is impossible.

Note

By calculating the product moment correlation coefficients, explain whether model $y = ax + b$ or model $\mathbf{y} = \mathbf{ax} + \mathbf{b}$ is more appropriate.

> The $|r|$ value for the model $y = ax + b$ is higher at ___, compared to ___ for the model $\mathbf{y} = \mathbf{ax} + \mathbf{b}$. Thus, there is a stronger (positive/negative) correlation between x and y. As such, the model $y = ax + b$ is more appropriate.

Note

Let x be in unit$_x$ and $\mathbf{x}$ be in unit$_\mathbf{x}$. Suppose that the c unit$_x$ $= 1$ unit$_\mathbf{x}$, where c is a constant. Then, if $y = ax + b$, we have $y = ac\mathbf{x} + b$.

25.4 Other Notes

Note

Explain whether it is valid to conclude that a higher value of x will *result in* a lower/higher value of y.

> No. While a higher value of x is *correlated* with a higher value of y, this does not imply any *causal* relationship between x and y.

Note. "result in" tends to refer to a *causal* relationship.

Example 25.4

Suggest an improvement to the data collection process so that the results could provide a fairer gauge of the expected outcome.

> The randomly selected [members of population] might have been of different [category 1; e.g. gender] and [category 2; e.g. age]. To make the results fairer, the data could have been separated based on [category 1] and [category 2].

Chapter 26

Non-Parametric Tests

26.1 Sign Test

General Information

- A *sign test*.

  1. Let m be the population median of $D =$ _________ $-$ _________.
  2. | | | |
     |---|---|---|
     | Test | $H_0: m = m_0$ | |
     | against | $H_1:$ (a) $m < m_0$, (b) $m \neq m_0$, or (c) $m > m_0$, | |
     | at the $100\alpha\%$ significance level. | | |
  3.

[label in context]	1	2	3	$\cdots$	N
Sign	$+$	0	$-$	$\cdots$	$+$

Table 26.1: The signs of $d_1, d_2, \ldots, d_N$, for a sigm test. Instead of $1, 2, \ldots, N$ the labeling/column headers can differ in the given context. E.g. $A, B, \ldots, K$. Similarly, the signs here are mere examples; the ith sign cell should be filled with $+$ $(-)$ $[0]$ if $\operatorname{sgn}(d_i) = 1$ $(= -1)$ $[= 0]$.

  4. Let X_+ be the number of '$+$'. Under H_0, $X_+ \sim \mathrm{B}(n, 1/2)$, $x_+ = 11$. (Alternatively, X_- can also be used.)
  5. Since p-value $=$ ______ $< 100\alpha\%$ $(\geq 100\alpha\%)$, there is sufficient (insufficient) evidence, at the $100\alpha\%$ significance level, to conclude that [H_1 in context].

- *Note.* The p-value for a sign test is given by

H_1	$m < m_0$	$m > m_0$	$m \neq m_0$
X_+	$\mathrm{P}(X_+ \leq x_+)$	$\mathrm{P}(X_+ \geq x_+)$	$2\min\{\mathrm{P}(X_+ \geq x_+), \mathrm{P}(X_+ \leq x_+)\}$
X_-	$\mathrm{P}(X_- \geq x_-)$	$\mathrm{P}(X_- \leq x_-)$	$2\min\{\mathrm{P}(X_- \geq x_-), \mathrm{P}(X_- \leq x_-)\}$

Table 26.2: The p-value for a sign test.

Note

Sign test. Suppose we have $H_1: m \neq m_0$. To find the range of values of x_+ that result in the rejection of H_0, use GC to compute the following tables.

x_+	$\alpha/2 - 2\,\mathrm{P}(X_+ \leq x_+)$
$n-1$	___ > 0
n	___ > 0
$n+1$	___ < 0

x_+	$\alpha/2 - 2\,\mathrm{P}(X_+ \geq x_+)$
$N-1$	___ < 0
N	___ < 0
$N+1$	___ > 0

Then, we conclude that $x_+ \leq n$ or $x_+ \geq N$.

26.2 Wilcoxon Matched-Pairs Signed Rank Test

Note

Assumptions needed for the Wilcoxon Matched-Pairs Signed Rank Test:

1. The data within each pair are dependent on each other, but pairs are independent of each other.
2. The distribution of the differences is continuous and symmetrical.

General Information

- A Wilcoxon matched-pairs signed rank test.

  1. Let m be the population median of $D =$ ________ $-$ ________.
  2. Test $H_0: m = 0$ against H_1: (a) $m < 0$, (b) $m \neq 0$, or (c) $m > 0$, at the $100\alpha\%$ significance level.
  3.

[label in context]	1	2	3	$\cdots$	N
d	d_1	0	d_3	$\cdots$	d_N
Rank	1	0	5	$\cdots$	2

Table 26.3: The value of the differences $d_1, d_2, \ldots, d_N$, which are then ranked according to their absolute size $|d_i|$. For our syllabus, each d_i is always distinct.

  4.
     - $t_- =$ ___ $+$ ___ $+ \cdots +$ ___ $=$ ___
     - $t_+ =$ ___ $+$ ___ $+ \cdots +$ ___ $=$ ___
     - The test statistic is $T := \min\{T_-, T_+\} =$ ___.
     - Reject H_0 if $T =$ ___. (see table 26.4)

Remark

$$\frac{n(n+1)}{2} = t_- + t_+.$$

  5. Since $t =$ ___ $\square$ ___, there is sufficient/insufficient evidence, at the $100\alpha\%$ significance level, to conclude that [H_1 in context].

- The test statistics T_+ and T_- can also be used, depending on our preference.
- The critical regions for a Wilcoxon test, for each alternative hypothesis and test statistic T_- or T_+. The value of c is obtained from MF26*.

  Note. the value of c may differ for a one-tail vs a two-tail test, so look at the table carefully, to obtain the correct value.

H_1	$m < m_0$	$m > m_0$	$m \neq m_0$
T_+	$T_+ \leq c$	$T_+ \geq \dfrac{n(n+1)}{2} - c$	$T_+ \leq c$ or $T_+ \geq \dfrac{n(n+1)}{2} - c$
T_-	$T_- \geq \dfrac{n(n+1)}{2} - c$	$T_- \leq c$	$T_- \leq c$ or $T_- \geq \dfrac{n(n+1)}{2} - c$
T	$T \leq c$[1]		$T \leq c$ or $T \geq \dfrac{n(n+1)}{2} - c$

Table 26.4: The critical regions for Wilcoxon tests.

[1]Assuming $T_- \geq T_+$ for $m < m_0$, and $T_+ \geq T_-$ for $m > m_0$.

- For large sample sizes $n \geq 21$, we use the approximation

$$T \sim \mathrm{N}\left(\frac{n(n+1)}{4}, \frac{n(n+1)(2n+1)}{24}\right)$$

and conduct a one/two-tailed z-test.

G.C. Skills

After calculating our list of differences $\mathtt{L}_3$, we can calculate $\mathtt{L}_4 = |\mathtt{L}_3|$ and use the G.C. to rank this list in ascending order:

$$\texttt{stat} \Longrightarrow \texttt{2:SortA} \Longrightarrow \mathtt{L}_4.$$

This allows us to easily compute the ranks associated with each difference.

Note

The value of n for the test statistic/MF26 critical region in both tests should be the number of columns with nonzero difference d. i.e.

$$n := \#\{i \mid d_i \neq 0\} = \#\text{cols} - \#\{i \mid d_i = 0\}.$$

Note

If we need to use both the sign test and a Wilcoxon test on the same sample, then consider creating just a single table, as shown below.

[label in context]	1	2	3	$\cdots$	n
d	d_1	0	d_3	$\cdots$	d_n
Sign	$+$	0	$-$	$\cdots$	$+$
Rank	1	0	5	$\cdots$	2

Table 26.5: Combined table for both the sign test and Wilcoxon test.

Note

How do you improve the Wilcoxon test used in [the previous part]?

Increase the sample size for the test.

Note

State the circumstances under which a non-parametric test would be used rather than a parametric test.

We use a non-parametric test, rather than a parametric test, when:

1. The population is not known to be normally distributed.
2. The population mean is not the best way to measure central tendency.
3. The measurement scale has no predetermined rank or ordering.

Note

Why is it not appropriate to use a paired-sample t-test?

There is no contextual evidence to support the assumption that $D_1, D_2, \ldots, D_n$ are normally distributed. So, conducting a paired-sample t-test may result in unreliable results, given our small sample size n.

Note

State the precautions that should be taken to avoid (statistical) bias.

Choose any approperate ones.

1. The test should be *'blind'*. [Testers in context] should not know which of the [two variations involved in the test, in context] they are [tasting/wearing/etc, in context]. If the [testers] knew, their preconceptions may affect ____________.
2. Pick a random sample of n [testers].
3. The *order* of the test — whether the [first variation] or [second variation] comes first — should be randomised.
4. The [testers] should not communicate with each other.
5. There should be sufficient rest time between the two runs, so that the running timing of the second run would not be affected due to fatigue.

Note

Explain why it is better to conduct a Wilcoxon test than a sign test.

While a sign test only considers the sign of the differences, a Wilcoxon test takes into account both the sign and *magnitude* of the differences. Therefore, a Wilcoxon test is more reliable, as it incorporates more (relevant) information about the data.

Note

Explain why it is appropriate use a Wilcoxon test in this situation.

✓ Explain why a Wilcoxon test is better than other tests (t-test, sign test).

× Explain why the assumptions of the Wilcoxon test hold true, and hence, the test can be carried out to reach a reliable conclusion.

Note

Explain why a sign test is more suitable/a Wilcoxon test is inappropriate.

Choose any approperiate ones

1. The data here is non-numeric and is not measured on an ordinal scale. Hence, it is inappropriate to conduct a Wilcoxon test. A sign test is better, as the data can still be represented by positive and negative responses — denoting ______ and ______, respectively.
2. The magnitude of the differences is irrelevant because ______. So, a sign test — which only accounts for the sign of the differences — is more appropriate.
3. In this case, the data has too many *tied ranks*. Thus, the conclusion obtained from a Wilcoxon test may not be reliable.
4. An additional assumption that the distribution of the differences $D = \_\_\_ - \_\_\_$ must be continuous and symmetric about the median.

Example 26.1: A tricker question, involving an unknown in the data provided.

Let m be the median of $D\colon X - Y$. For the data in Table 26.6, assume that there are no tied ranks, and $x_i \neq y_i$ for each $1 \leq i \leq 7$. Carry out a Wilcoxon test, at the 5% significance level, to determine if the data supports the alternative hypothesis $H_1: m > 0$.

Index	1	2	3	4	5	6	7
x_i	4	8	7	7	1	9	9
y_i	6	9	3	4	a	1	2

Table 26.6: Data with an unknown variable $a \in \mathbb{Z}^+$.

First, we calculate the differences. Since $x_i \neq y_i$, we have $a \neq 1$. In fact, $a \neq 1, 2, 3, 4, 7, 8$ because $d_i \neq d_j$, for $i \neq j$. Thus, $a = 6, 7$ or $a \geq 10$. The corresponding rank r_5 is hence 5 or 7.

Index	1	2	3	4	5	6	7
d_i	-2	-1	4	3	$1-a$	8	7
$\lvert d_i \rvert$	2	1	4	3	$a-1$	8	7
rank r_i	2	1	4	3	r_5	r_6	r_7

Table 26.7: The values of the differences d_i and the associated ranks. The columns highlighted in grey are those with negative differences d_i.

Now,

$$t_- = 2 + 1 + r_5 = 8, 10 \qquad \text{and} \qquad t_+ = 7(7+1)/2 - t_- = 25 - r_5 = 20, 18.$$

Hence, the test statistic $T \coloneqq \min\{T_-, T_+\} = T_-$, where we reject H_0 if $T \leq 3$. So, since $t_- = 3 + r_5 > 3$, we do not reject H_0.

Chapter 27

Bibliography

1. Figure 6.1 Disc method (Source)
2. Figure 6.2 Shell method (Source)
3. Fig 6.1 Trapezium rule (Source)
4. Fig 6.2 Simpson's rule (Source)
5. Fig 7.1 Argand Diagram (Source)
6. Figure 17.1 An illustration of Euler's Method and the Improved Euler's Method (Source)
7. Oscillatory behavior of DEs modelling physical phenomena Fig 15.2 (Source)
8. Figure 19.1 Mode of a binomial distribution (Source)
9. Used some inspiration from the beautiful preamble by tearfox and det.uwu from Discord, to make my environments look better.
10. Figure 9.2 on linear interpolation by me (Grass) (Source)
11. Figure 9.3 on fixed-point iteration by me (Grass) (Source)
12. Figure 9.4 on the Newton-Raphson method by me (Grass) (Source)
13. Figure 24.1 Chi-squared χ^2 distribution (Source)