Mathematical Methods — Bound Reference

January 9, 2024

Contents

Part I

Theory

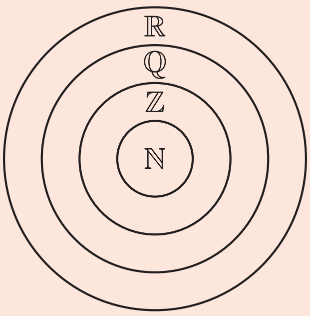
1 Functions and relations

Set notation

- A **set** is a collection of objects called **elements**.
- $x \in A$ means that element x is a member of set A (and its counterpart is $x \notin A$).
- B, another set, is a **subset** of set A if *every element* of B is also in A. We write this as $B \subseteq A$.
- $\varnothing$ is known as the **empty set**.
- The set of elements that are common to two sets A and B is called the **intersection** of A and B, and is denoted by $A \cap B$. Thus, $x \in A \cap B \iff x \in A$ and $x \in B$.
- Sets A and B are **disjoint** if they have no elements in common ($A \cap B = \varnothing$).
- The set of elements that are in A or in B (or in *both*) is called the **union** of sets A and B, and is denoted by $A \cup B$.
- The **set difference** of two sets A and B is given by $A \setminus B = \{x : x \in A, x \notin B\}$.

Sets of numbers

- $\mathbb{N} = \{1, 2, 3, \ldots\}$ = Counting numbers.
- $\mathbb{Z} = \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}$ = Whole numbers.
- $\mathbb{Q} = \left\{p, q \in \mathbb{Z} : \frac{p}{q}\right\}$ = Rational numbers.
- The set of *all the numbers which cannot be represented by ratios of two integers* is called the set of **real numbers**, and is denoted by $\mathbb{R}$.
  - Positive real numbers: $\mathbb{R}^+ = \{x : x > 0\}$
  - Negative real numbers: $\mathbb{R}^- = \{x : x < 0\}$
  - Real numbers excluding zero: $\mathbb{R} \setminus \{0\}$



Interval notation

- Suppose that a and b are real numbers, with $a < b$.
  - $(a, b) = \{x : a < x < b\}$
  - $[a, b] = \{x : a \leq x \leq b\}$
  - $(a, b] = \{x : a < x \leq b\}$
  - $[a, b) = \{x : a \leq x < b\}$
  - $(a, \infty) = \{x : a < x\}$
  - $[a, \infty) = \{x : a \leq x\}$
  - $(-\infty, b) = \{x : x < b\}$
  - $(-\infty, b] = \{x : a < x \leq b\}$

When using number lines to represent intervals,

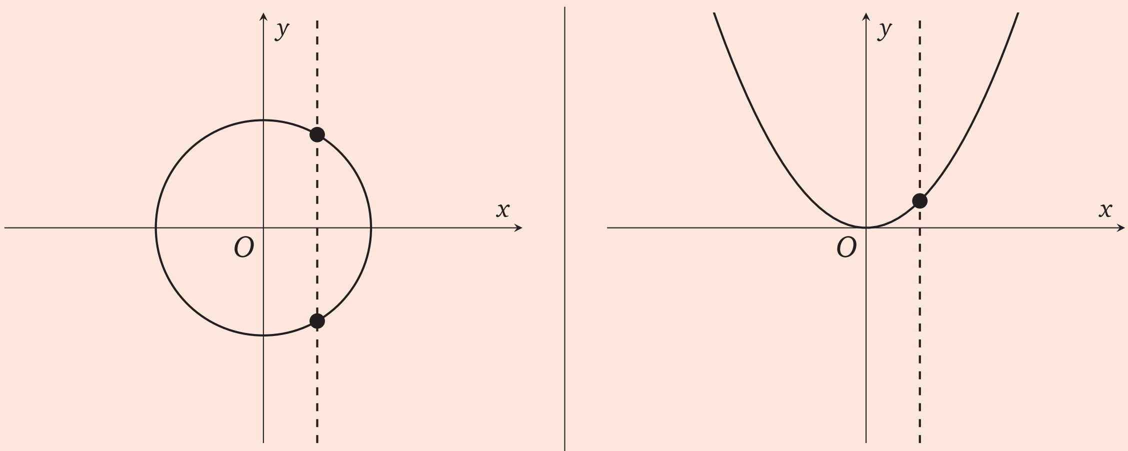
- The 'closed' circle (•) indicates that the number is included.
- The 'open' circle (∘) indicates that the number is **not** included.

Functions VS relations

- A **function** is a relation such that for each x-value there is only one corresponding y-value. This means that, if (a, b) and (a, c) are ordered pairs of a function, then $b = c$.
- In other words, a function cannot contain two different ordered pairs with the same first coordinate.

Vertical line test

- If a vertical line can be drawn anywhere on the graph and it only ever intersects the graph at a maximum of once, then the relation is a function.



$x^2 + y^2 = 1$ is not a function

$y = x^2$ is a function

Function notation

$$f : X \to Y, f(x) = \ldots$$

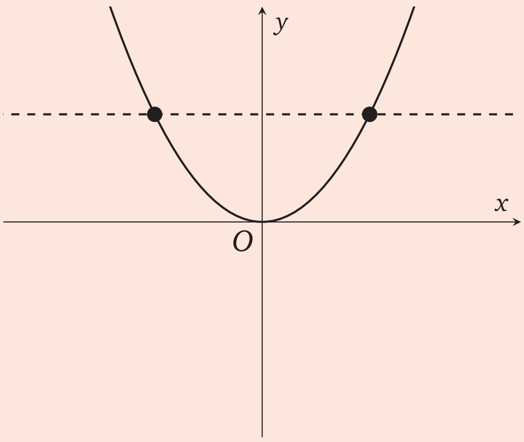
- f is the function name
- X is the **domain** of the function, *i.e.,* the set of values for which the function is defined.
- Y is the **codomain** of the function, *i.e.,* the set of values which the **range** (the range is the set of outputs of the function) of the function falls into.

Types of functions (many/one-to-one)

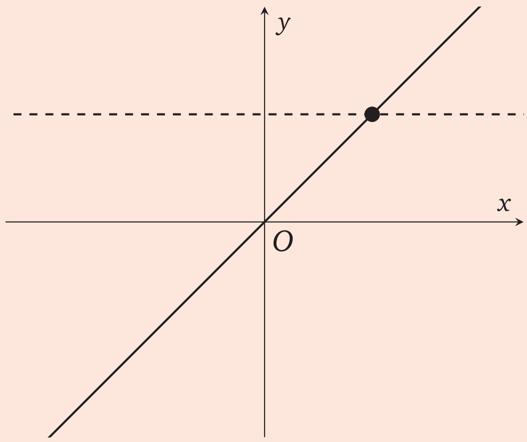
- If $\forall a, b \in \text{dom}(f) : f(a) = f(b) \iff a = b$, or, to put it another way, $\forall a, b \in \text{dom}(f) : f(a) \neq f(b) \iff a \neq b$, then a function is a **one-to-one** function.
- A function that does not satisfy the above condition(s) is a **many-to-one** function.

Horizontal line test

- If a horizontal line can be drawn anywhere on the graph of a function and it only ever intersects the graph a maximum of once, then the function is a **one-to-one**.



$f(x) = x^2$ is a many-to-one function



$f(x) = x$ is a one-to-one function

Parity of functions

- A function is **even** if $\forall x \in \text{dom}(f) : f(x) = f(-x)$.
- A function is **odd** if $\forall x \in \text{dom}(f) : f(-x) = -f(x)$.
- A function can be **neither odd nor even** (if both of the above statements do not apply).

Implied/maximal domain

- The **implied** domain (also referred to as the **maximal** domain) of a function is the largest subset of $\mathbb{R}$ for which the rule for the function is defined.

Sum and product of functions

- $(f+g)(x) = f(x) + g(x)$ for $\text{dom}(f) \cap \text{dom}(g) \neq \varnothing$
  - $\text{dom}(f+g) = \text{dom}(f) \cap \text{dom}(g)$
- $(f-g)(x) = f(x) - g(x)$ for $\text{dom}(f) \cap \text{dom}(g) \neq \varnothing$
  - $\text{dom}(f-g) = \text{dom}(f) \cap \text{dom}(g)$
- $(f \cdot g)(x) = f(x) \cdot g(x)$ for $\text{dom}(f) \cap \text{dom}(g) \neq \varnothing$
  - $\text{dom}(f \cdot g) = \text{dom}(f) \cap \text{dom}(g)$

Addition of ordinates (sketching $y = (f+g)(x)$)

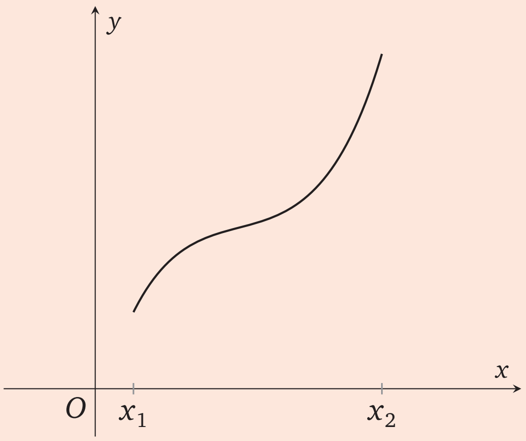
- When $f(x) = 0$, $(f+g)(x) = g(x)$.
- When $g(x) = 0$, $(f+g)(x) = f(x)$.
- If $f(x)$ and $g(x)$ are **both** positive, then $(f+g)(x) > f(x)$ **and** $(f+g)(x) > g(x)$.
- If $f(x)$ and $g(x)$ are **both** negative, then $(f+g)(x) < f(x)$ **and** $(f+g)(x) < g(x)$.
- If $f(x)$ is positive and $g(x)$ is negative, then $g(x) < (f+g)(x) < f(x)$.
- Look for values of x for which $f(x) + g(x) = 0$.

Composite functions

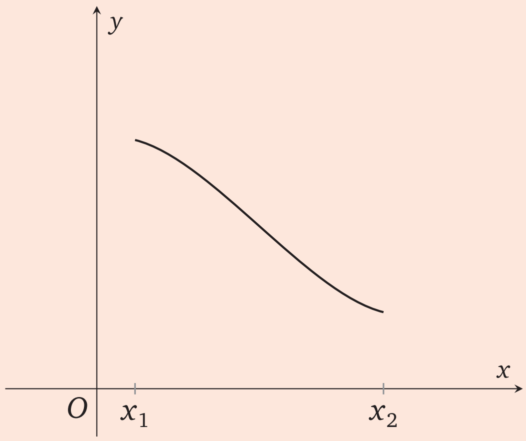
- Given that $\boxed{\text{ran}(g) \subseteq \text{dom}(f)}$, we can define the new function h as a **composition** of f with g.
- This is written $h = f \circ g$ (read 'composition of g followed by f') and the rule for h is given by $h(x) = f(g(x))$.
  - Generally, $\boxed{\text{dom}(h) = \text{dom}(g)}$ (but **do** check any restrictions the final function imposes).
- Generally, $\boxed{f \circ g \neq g \circ f}$

Increasing and decreasing functions

- If $\forall x \in [x_1, x_2] : \boxed{f(x_2) > f(x_1)} \mid x_2 > x_1$, then f is **strictly increasing** over the interval $[x_1, x_2]$.
- If $\forall x \in [x_1, x_2] : \boxed{f(x_2) < f(x_1)} \mid x_2 > x_1$, then f is **strictly decreasing** over the interval $[x_1, x_2]$.
- These intervals include the values of x for which $\dfrac{\mathrm{d}f}{\mathrm{d}x} = 0$, but the gradient never changes sign as such.
  - An example would be that the function of $f : [0, \infty) \to \mathbb{R}, f(x) = x^2$ is strictly increasing.
- A function that is strictly increasing or decreasing is also a one-to-one function.



This function is **strictly increasing** over $[x_1, x_2]$



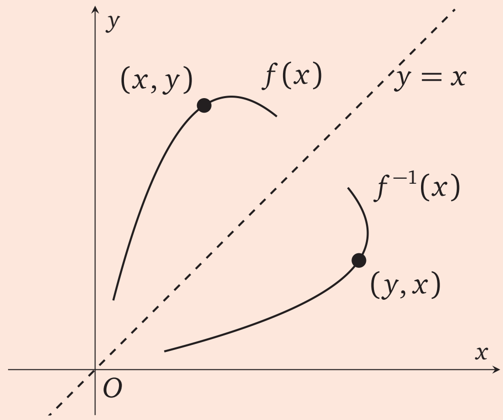
This function is **strictly decreasing** over $[x_1, x_2]$

Inverse functions

- If f is a **one-to-one function**, then a new function f^{-1}, called the **inverse** of f, may be defined by

$$f^{-1}(x) = y \quad \text{if } f(y) = x,$$

  for $x \in \operatorname{ran}(f)$, $y \in \operatorname{dom}(f)$.
- Domain and range:
  - $\operatorname{dom}\left(f^{-1}\right) = \operatorname{ran}(f)$
  - $\operatorname{ran}\left(f^{-1}\right) = \operatorname{dom}(f)$
- Compositions:
  - $\forall x \in \operatorname{dom}(f^{-1}) : \left(f \circ f^{-1}\right)(x) = x$
  - $\forall x \in \operatorname{dom}(f) : \left(f^{-1} \circ f\right)(x) = x$
- The point (x, y) is on the graph of f^{-1} if and only if the point (y, x) is on the graph of f. Thus, the graph of f^{-1} is a **reflection** of the graph of f in the line $y = x$.
- If f is strictly increasing, then f^{-1} is also strictly increasing (and vice versa).
- For a continuous function f (and some non-continuous functions, too), **at least one** of the intersections of f and f^{-1} lie on the line $y = x$ (if the functions intersect at all, that is), so, to solve for this intersection point, either $f(x) = x$ or $f^{-1}(x) = x$ will suffice.
  - If "just one intersection point" is needed, then the line $y = x$ is **tangential** to both functions at that point. This means that the gradients of f, f^{-1}, and $y = x$ are equal at that point (they are all 1, as the gradient of $y = x$ is always 1).



2 Coordinate geometry

Solving simultaneous equations

- There are two ways to solve simultaneous equations: **substitution** and **elimination**.

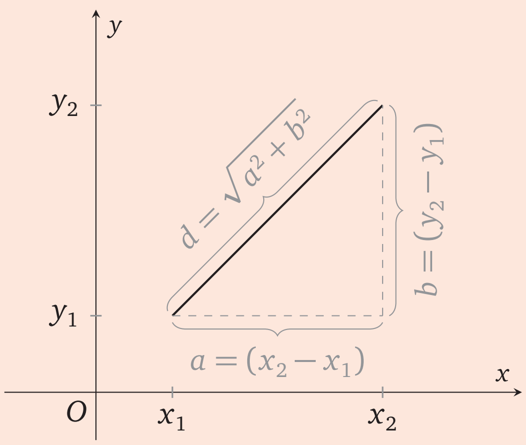
Linear coordinate geometry

- The following is revision of basic concepts of linear coordinate geometry.
- This is **linear** coordinate geometry, meaning these concepts can only apply to straight lines.

Distance between two points

- Let d be the **distance** between two points $A(x_1, y_1)$ and $B(x_2, y_2)$.

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$



The Pythagorean theorem comes into play for the derivation of the distance formula, as it is just the hypotenuse of a right-angled triangle formed by the horizontal and vertical components of the line.

Midpoint of a line

- The **midpoint**, of a line beginning and ending at points $A(x_1, y_1)$ and $B(x_2, y_2)$ respectively is given by the formula:

$$\text{Midpoint} = \left(\frac{x_1 + x_2}{2}, \frac{y_1 + y_2}{2}\right)$$

Gradient of a line

- The **gradient**, m, of a line going through the points $A(x_1, y_1)$ and $B(x_2, y_2)$ is given by the formula:

$$m = \frac{y_2 - y_1}{x_2 - x_1}$$

Equation of a line

- The equation of a line (in slope-intercept form) with a y-intercept at the point $A(0, c)$ is given by the formula:

$$y = mx + c$$

- The equation of a line (in point-slope form) going through the point $A(x_1, y_1)$ is given by the formula:

$$y - y_1 = m(x - x_1)$$

- The equation of a line (in intercept form) going through the two points $A(a, 0)$ and $B(0, b)$ is given by the formula:

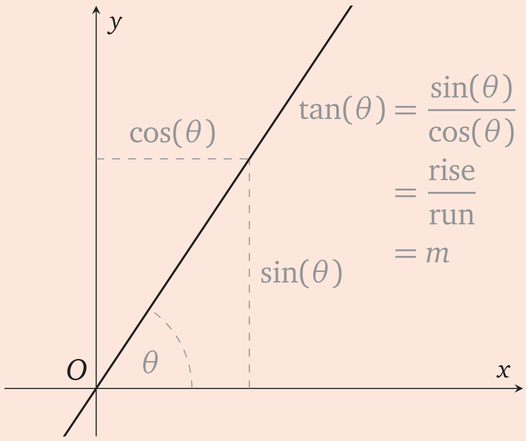
$$\frac{x}{a} + \frac{y}{b} = 1$$

Tangent of the angle of slope

- For a straight line with gradient m, the angle of slope is found using:

$$m = \tan(\theta)$$

where θ is the angle that the line makes with the positive direction of the x-axis.



Perpendicular and parallel lines

- If two straight lines are perpendicular to each other (meet at right angles), the product of their gradients is -1 (unless one is vertical and the other horizontal).

$$m_\perp \times m = -1$$

- Parallel lines have the same gradient.

The geometry of simultaneous linear equations

- There are three cases for a system of two linear equations with two variables.

	Example	Solutions	Geometry
Case 1	$2x + y = 5$ $x - y = 4$	Unique solution: $x = 3,\ y = -1$	Two lines meeting at a point
Case 2	$2x + y = 5$ $2x + y = 7$	No solutions	Distinct parallel lines
Case 3	$2x + y = 5$ $4x + 2y = 10$	Infinitely many solutions	Two copies of the same line

Dilations

- Dilation from the x-axis:
  - For $b \in \mathbb{R}^+$, a dilation of factor b from the x-axis is described by the rule:

$$(x, y) \rightarrow (x, by)$$

  - This means that this dilation can also be applied as such: $y = b \cdot f(x)$.
- Dilation from the y-axis:
  - For $a \in \mathbb{R}^+$, a dilation of factor a from the y-axis is described by the rule:

$$(x, y) \rightarrow (ax, y)$$

  - This means that this dilation can also be applied as such: $y = f\left(\frac{x}{a}\right)$.

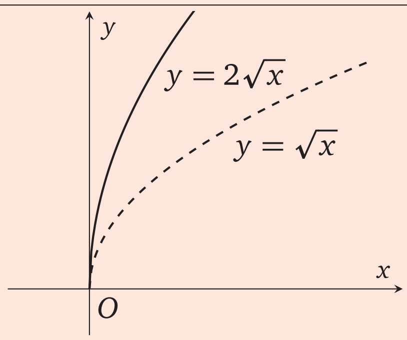
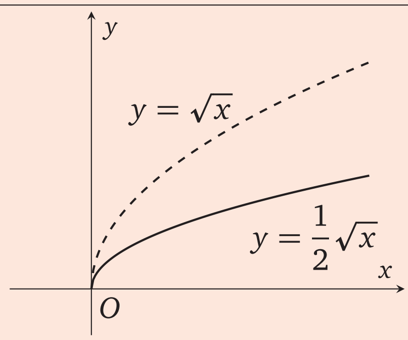
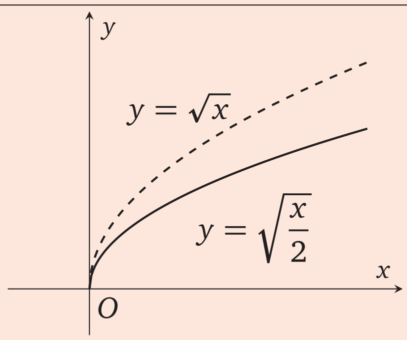
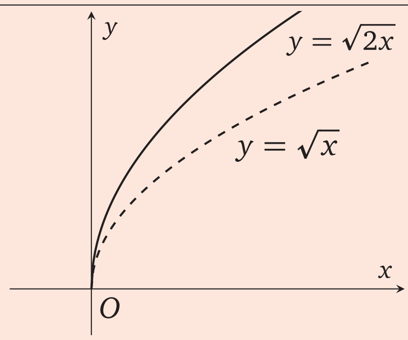
<table>
<tr><th>From</th><th>factor 2</th><th>factor $\frac{1}{2}$</th></tr>
<tr><td>x-axis</td><td>
</td><td>
</td></tr>
<tr><td>y-axis</td><td>
</td><td>
</td></tr>
</table>

Table of transformations

Mapping	$(x,y) \rightarrow$	$y = f(x) \rightarrow$
Reflection in the x-axis	$(x, -y)$	$y = -f(x)$
Reflection in the y-axis	$(-x, y)$	$y = f(-x)$
Dilation of factor a from the y-axis	(ax, y)	$y = f\left(\frac{x}{a}\right)$
Dilation of factor b from the x-axis	(x, by)	$y = bf(x)$
Reflection in the line $y = x$ (inverse function)	(y, x)	$x = f(y)$
Translation of h units in the positive direction of the x-axis	$(x+h, y)$	$y = f(x-h)$
Translation of k units in the positive direction of the y-axis	$(x, y+k)$	$y - k = f(x)$

Applying transformations

1. $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2, T(x,y) = (ax+h, by+k), \quad a \neq 0, b \neq 0$
   - Note that this notation is the same as writing $(x,y) \rightarrow (ax+h, by+k)$.
2. Denote the **transformed** pair of coordinates (the new ones) as $\left(x', y'\right)$.
3. $\left(x', y'\right) = T(x,y)$
   $\therefore \left(x', y'\right) = (ax+h, by+k), \quad a \neq 0, b \neq 0$
4. Solve for the original x and y to be subbed into the function in question.
   $$x' = ax + h$$
   $$\therefore x = \frac{x' - h}{a}$$
   $$y' = bx + k$$
   $$\therefore y = \frac{y' - k}{b}$$
5. Substitute x and y back into the function $y = f(x)$.
   - Remember to solve for y if there is more than one term on that side of the equation.

3 Polynomial functions

Quadratics

- For a quadratic in standard (polynomial) form $\left(ax^2+bx+c\right)$,
  - if $a>0$, then the graph has a **minimum** point.
  - if $a<0$, then the graph has a **maximum** point.
  - the **vertex** (**turning point**) is the point (h,k), where $h=-\dfrac{b}{2a}$ and $k=\dfrac{4ac-b^2}{4a}$.
  - the **axis of symmetry** is $x=h$, where $h=-\dfrac{b}{2a}$.
  - the quadratic can be written in "turning point form" by **completing the square** for x using the formula:

$$ax^2+bx+c=a\left(x+\frac{b}{2a}\right)^2+\frac{4ac-b^2}{4a}$$

  - the solutions to $ax^2+bx+c=0$ can be obtained using the **quadratic formula**:

$$x=\frac{-b\pm\sqrt{b^2-4ac}}{2a},\quad a\neq 0$$

  - the **discriminant** (Δ) for a quadratic polynomial is:

$$\Delta=b^2-4ac$$

    For the equation $ax^2+bx+c=0$,
    ⋆ if $\Delta>0$, there are two solutions.
    ⋆ if $\Delta=0$, there is one solution (tangential).
    ⋆ if $\Delta<0$, there are no solutions.
    For the equation $ax^2+bx+c=0$ where $a,b,c\in\mathbb{Q}$,
    ⋆ if Δ is a perfect square and $\Delta\neq 0$, then the equation has two rational solutions.
    ⋆ if $\Delta=0$, then the equation has one rational solution.
    ⋆ if Δ is not a perfect square and $\Delta>0$, then the equation has two irrational solutions.

Remainder theorem

- When $P(x)$ is divided by $\beta x+\alpha$, the remainder is $P\left(-\dfrac{\alpha}{\beta}\right)$.

Factor theorem

- For the polynomial $P(x)$, if $P(\alpha)=0$, then $x-\alpha$ is a factor of $P(x)$.
- Conversely, if $x-\alpha$ is a factor of $P(x)$, then $P(\alpha)=0$.

More generally:

- For the polynomial $P(x)$, if $\beta x+\alpha$ is a factor of $P(x)$, then $P\left(-\frac{\alpha}{\beta}\right)=0$.
- Conversely, if $P\left(-\frac{\alpha}{\beta}\right)=0$, then $\beta x+\alpha$ is a factor of $P(x)$.

Rational root theorem

- The root of a polynomial function $P(x)$ such that:

  $$P(x)=a_n x^n+a_{n-1}x^{n-1}+\cdots+a_2x^2+a_1x+a_0$$

  where the coefficients are integers is of the form:

  $$\frac{p}{q},\ \text{where } p=\text{a factor of } a_0 \text{ and } q=\text{a factor of } a_n$$

Polynomials of degree n

- For a polynomial $P(x)$ of degree n, there are **at most** n solutions to the equation $P(x)=0$. Therefore, the graph of $P(x)$ has **at most** n x-axis intercepts.
- The graph of a polynomial of even degree may have no x-axis intercepts: for example, $P(x)=x^2+1$. But the graph of a polynomial of odd degree must have at least one x-axis intercept.

Difference and sum of two variables of the same degree

- $x^2-y^2=(x-y)(x+y)$
- $x^3-y^3=(x+y)\left(x^2-xy+y^2\right)$
- If n is odd,
  - $x^n-y^n=(x-y)\left(x^{n-1}+x^{n-2}y+\cdots+xy^{n-2}+y^{n-1}\right)$
  - $x^n+y^n=(x+y)\left(x^{n-1}-x^{n-2}y+\cdots+xy^{n-2}+y^{n-1}\right)$

4 Exponential functions

Exponential function characteristics

- For $a \in \mathbb{R}^+ \setminus \{1\}$, the graph of $y = a^x$ has the following properties:
  - The x-axis is an asymptote.
  - The y-values are always positive.
  - The y-axis intercept is 1.
  - There is no x-axis intercept.

Euler's number — e

- Euler's number is defined as follows:

$$e = \lim_{n \to \infty} \left(1 + \frac{1}{n}\right)^n = 2.718\,281\,828\ldots$$

Index laws

For all positive numbers a and b and all real numbers x and y:

- $a^x \cdot a^y = a^{x+y}$
- $a^x \div a^y = a^{x-y}$
- $(a^x)^y = a^{xy}$
- $(ab)^x = a^x b^x$
- $\left(\dfrac{a}{b}\right)^x = \dfrac{a^x}{b^x}$
- $a^{-x} = \dfrac{1}{a}$
- $a^x = \dfrac{1}{a^{-x}}$
- $a^0 = 1$

Logarithms

- For $a \in \mathbb{R}^+ \setminus \{1\}$, the **logarithm function** with base a is defined as follows:

$$a^x = y \iff \log_a(y) = x$$

- Since a is positive, the expression $\log_a(y)$ is only defined when y is positive ($y > 0$).

Log laws

- $\log_a(1) = 0$
- $\log_a(a) = 1$
- $\log_a\left(x^b\right) = b \cdot \log_a(x)$
- $\log_{a^b}(x) = \frac{1}{b} \cdot \log_a(x)$
- $\log_a\left(\frac{1}{x}\right) = -\log_a(x)$
- $\log_{\frac{1}{a}}(x) = -\log_a(x)$
- $\log_a(b) = \frac{\ln(b)}{\ln(a)}$
- $\log_a\left(a^b\right) = b$
- $\log_a\left[\left(\frac{1}{a}\right)^n\right] = -n$
- $a^{\log_a(b)} = b$
- $\log_a(a) + \log_a(b) = \log_a(ab)$
- $\log_a(a) - \log_a(b) = \log_a\left(\frac{a}{b}\right)$

- The graph of $y = \log_a(x)$ can be obtained from the graph of $y = \log_b(x)$ by a dilation of factor $\frac{1}{\log_b(a)}$ from the x-axis.
- The graph of $y = a^x$ can be obtained from the graph of $y = b^x$ by a dilation of factor $\frac{1}{\log_b(a)}$ from the y-axis.
- When dividing both sides of an inequality by $\log_a(x)$ where $0 < x < 1$, reverse the inequality as the logarithm will evaluate to negative.

Exponential growth and decay

- In a situation where the growth/decay of something is exponential, the amount of that thing can be modelled using a function of the form:

$$A(t) = A_0 \cdot e^{kt}$$

where A_0 is the initial quantity at $t = 0$ (where t is a variable representing a unit of time) and k is a constant.

  - Growth corresponds to $k > 0$.
  - Decay corresponds to $k < 0$.

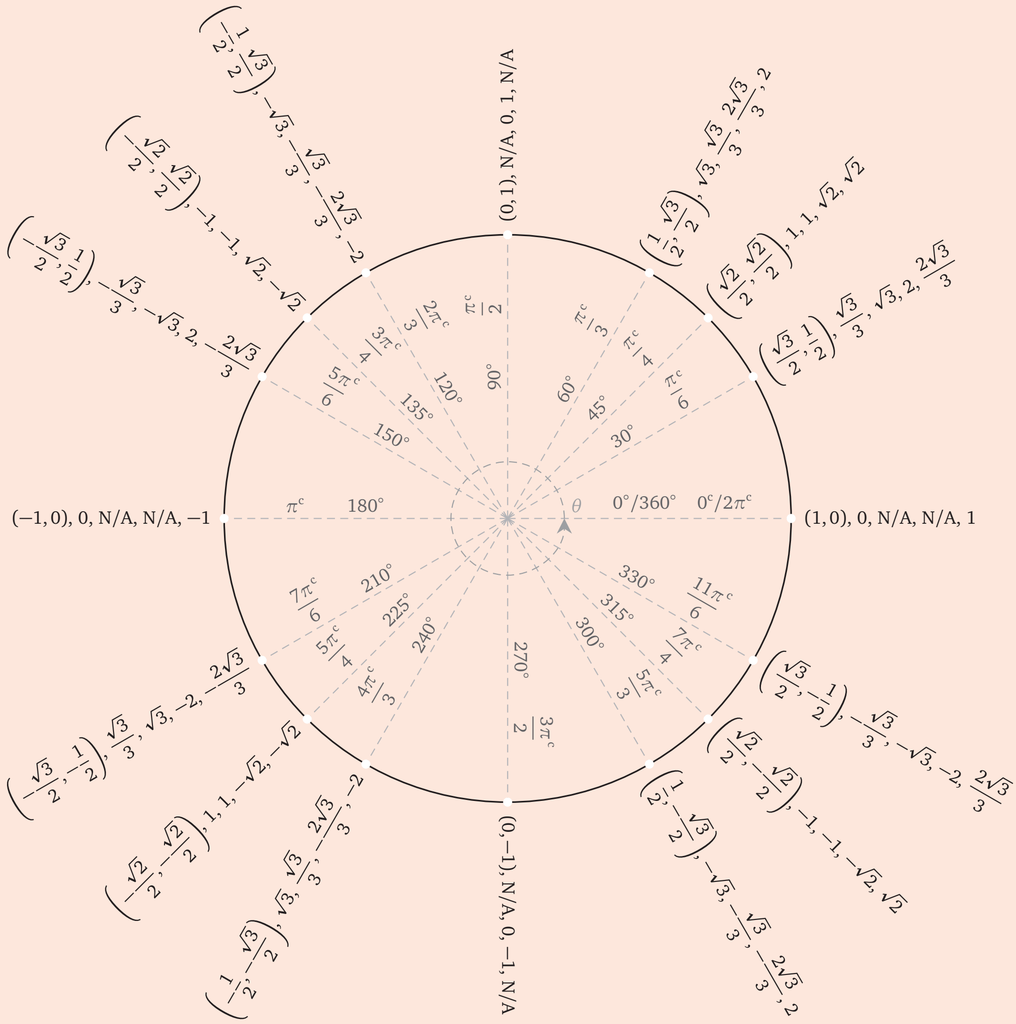
5 Circular functions

Radians and degrees

- One **radian** (written 1^c) is the angle subtended at the centre of the unit circle by an arc of length 1 unit.
- To convert between the two, use the following:

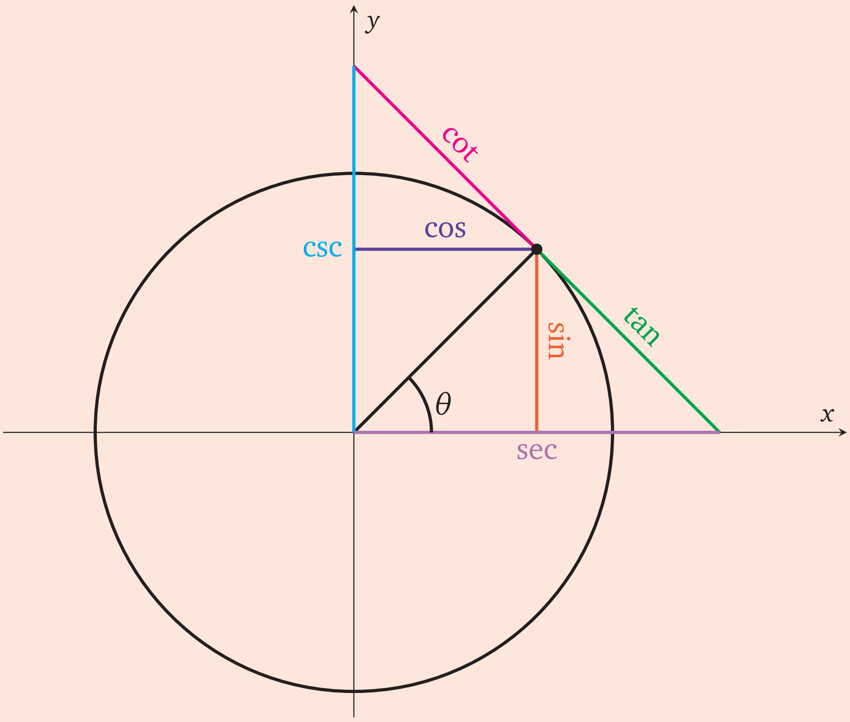
$$1^c = \frac{180^\circ}{\pi} \quad \text{or} \quad 1^\circ = \frac{\pi^c}{180}$$

Unit circle



$(x, y), z, a, b, c$, where:
$x = \cos$
$y = \sin$
$z = \tan = \dfrac{\sin}{\cos}$
$a = \cot = \dfrac{1}{\tan} = \dfrac{\cos}{\sin}$
$b = \csc = \dfrac{1}{\sin}$
$c = \sec = \dfrac{1}{\cos}$

Trigonometric functions as triangles



Properties of trigonometric functions

- $y = \pm a \sin(nt)$
  - The period of $\dfrac{2\pi}{n}$.
  - The amplitude is a.
  - The range is $[-a, a]$.
- $y = \pm a \cos(nt)$
  - The period of $\dfrac{2\pi}{n}$.
  - The amplitude is a.
  - The range is $[-a, a]$.
- $y = a \tan(nt)$
  - The period of $\dfrac{\pi}{n}$.
  - The vertical asymptotes have equations $t = \dfrac{(2k+1)\pi}{2n}$ where $k \in \mathbb{Z}$.
  - The axis intercepts are at $t = \dfrac{k\pi}{n}$ where $k \in \mathbb{Z}$.

Trigonometric identities

Pythagorean identities

- $\cos^2(x) + \sin^2(x) = 1$
- $\sec^2(x) - \tan^2(x) = 1$
- $\csc^2(x) - \cot^2(x) = 1$

Double-angle identities

- $\sin(2x) = 2\sin(x)\cos(x)$
- $\cos(2x) = 2\cos^2(x) - 1$
- $\cos(2x) = 1 - 2\sin^2(x)$
- $\cos(2x) = \cos^2(x) - \sin^2(x)$
- $\tan(2x) = \dfrac{2\tan(x)}{1 - \tan^2(x)}$

Sum/Difference identities

- $\sin(s+t) = \sin(s)\cos(t) + \cos(s)\sin(t)$
- $\sin(s-t) = \sin(s)\cos(t) - \cos(s)\sin(t)$
- $\cos(s+t) = \cos(s)\cos(t) - \sin(s)\sin(t)$
- $\cos(s-t) = \cos(s)\cos(t) + \sin(s)\sin(t)$
- $\tan(s+t) = \dfrac{\tan(s) + \tan(t)}{1 - \tan(s)\tan(t)}$
- $\tan(s-t) = \dfrac{\tan(s) - \tan(t)}{1 + \tan(s)\tan(t)}$

Product-to-sum identities

- $\cos(s)\cos(t) = \dfrac{\cos(s-t) + \cos(s+t)}{2}$
- $\sin(s)\sin(t) = \dfrac{\cos(s-t) - \cos(s+t)}{2}$
- $\sin(s)\cos(t) = \dfrac{\sin(s+t) + \sin(s-t)}{2}$
- $\cos(s)\sin(t) = \dfrac{\sin(s+t) - \sin(s-t)}{2}$

Triple-angle identities

- $\sin(3x) = -\sin^3(x) + 3\cos^2(x)\sin(x)$
- $\sin(3x) = -4\sin^3(x) + 3\sin(x)$
- $\cos(3x) = \cos^3(x) - 3\sin^2(x)\cos(x)$
- $\cos(3x) = 4\cos^3(x) - 3\cos(x)$
- $\tan(3x) = \dfrac{3\tan(x) - \tan^3(x)}{1 - 3\tan^2(x)}$
- $\cot(3x) = \dfrac{3\cot(x) - \cot^3(x)}{1 - 3\cot^2(x)}$

General solutions

General solutions for $\sin(x)$

$$\begin{aligned}
\sin(\theta) &= \alpha, a \in [-1,1] \\
\therefore \theta &= 2n\pi + \sin^{-1}(\alpha), n \in \mathbb{Z} \quad \text{or} \\
&= (2n+1)\pi - \sin^{-1}(\alpha), n \in \mathbb{Z}; \\
&= n\pi + (-1)^n \sin^{-1}(\alpha), n \in \mathbb{Z} \quad \text{(concise)}
\end{aligned}$$

General solutions for $\cos(x)$

$$\begin{aligned}
\cos(\theta) &= \alpha, a \in [-1,1] \\
\therefore \theta &= 2n\pi \pm \cos^{-1}(\alpha), n \in \mathbb{Z}
\end{aligned}$$

General solutions for $\tan(x)$

$$\begin{aligned}
\tan(\theta) &= \alpha, a \in [-1,1] \\
\therefore \theta &= n\pi + \tan^{-1}(\alpha), n \in \mathbb{Z}
\end{aligned}$$

Period of two trigonometric functions' sum/difference

- For two trigonometric functions f and g which are being added to (or subtracted from) each other to produce the function h, the period of h is the LCM (lowest common multiple) of the respective periods of f and g.

$$\begin{aligned}
\text{Let } f(x) &= a\sin(bx+c) \\
\text{Let } g(x) &= k\cos(mx+n)
\end{aligned}$$

$$\begin{aligned}
\text{Let } h(x) &= f(x) + g(x) \\
&= a\sin(bx+c) + k\cos(mx+n)
\end{aligned}$$

$$\begin{aligned}
\text{period}(f) &= \frac{2\pi}{b} \\
\text{period}(g) &= \frac{2\pi}{m} \\
\therefore \text{period}(h) &= \text{lcm}\left(\frac{2\pi}{b}, \frac{2\pi}{m}\right)
\end{aligned}$$

6 Differentiation

Average rate of change

- For any function $y = f(x)$, the **average rate of change** of y with respect to x over the interval $[a, b]$ is the gradient of the line through the two points $A(a, f(a))$ and $B(b, f(b))$.

$$\text{Average rate of change} = \frac{f(b) - f(a)}{b - a}$$

Differentiation from first principles

- The **derivative** of the function f is denoted by f' and is defined by:

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

- The derivative of a function f with respect to x when $x = a$ is also known as the **instantaneous rate of change** of f with respect to x when $x = a$.

Derivative rules

Differentiation results

- **Constant function:** $f(x) = c \implies f'(x) = 0$
- **Multiple:** $f(x) = k \cdot g(x) \implies f'(x) = k \cdot g'(x)$
- **Sum:** $f(x) = g(x) + h(x) \implies f'(x) = g'(x) + h'(x)$
- **Difference:** $f(x) = g(x) - h(x) \implies f'(x) = g'(x) - h'(x)$

Limits

Algebra of limits

- **Sum:** $\lim_{x\to a}(f(x)+g(x)) = \lim_{x\to a}(f(x)) + \lim_{x\to a}(g(x))$
- **Multiple:** $\lim_{x\to a}(k\cdot f(x)) = k\cdot \lim_{x\to a}(f(x)), k\in\mathbb{R}$
- **Product:** $\lim_{x\to a}(f(x)\cdot g(x)) = \lim_{x\to a}(f(x))\cdot \lim_{x\to a}(g(x))$
- **Quotient:** $\lim_{x\to a}\left(\frac{f(x)}{g(x)}\right) = \frac{\lim_{x\to a}(f(x))}{\lim_{x\to a}(g(x))}, \lim_{x\to a}(g(x)) \neq 0$

Left and right limits

- If the value of $f(x)$ approaches the number p as x approaches a from the right-hand side, the it is written as $\lim_{x\to a^+} f(x) = p$.
- If the value of $f(x)$ approaches the number p as x approaches a from the left-hand side, the it is written as $\lim_{x\to a^-} f(x) = p$.
- For $\lim_{x\to a} f(x)$ to exist, $\lim_{x\to a^+} f(x)$ and $\lim_{x\to a^-} f(x)$ must be equal.

Continuity of a function

- A function f is **continuous** at the point $x = a$ if the following conditions are met:
  - $f(a)$ is defined.
  - $\lim_{x\to a} f(x) = f(a)$

Differentiability of a function

- A function f is said to be differentiable at $x = a$ if $\lim_{h\to 0}\frac{f(a+h)-f(a)}{h}$ exists.
- If a function is differentiable at a point, then it is also continuous at that point (the same cannot be said for the converse statement).
- An easy way to remember this is that a function is **not differentiable** at a *sharp corner* or a *cusp* (a sharp point where two points meet).

Tangent line

- The **tangent line** to the graph of the function f at the point $(a, f(a))$ is defined to be the line through $(a, f(a))$ with the gradient $f'(a)$.
- The equation of the tangent line to the graph of $y = f(x)$ at the point $(a, f(a))$ can be found using the formula:

$$y - f(a) = f'(a)\cdot(x-a)$$

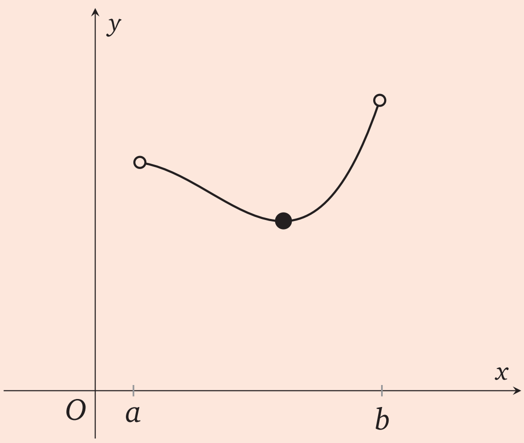
Normal line

- The **normal line** to the graph of the function f at the point $(a, f(a))$ is defined to be the line through $(a, f(a))$ and is perpendicular to the tangent to the function f at that point.
- The equation of the tangent line to the graph of $y = f(x)$ at the point $(a, f(a))$ can be found using the formula:

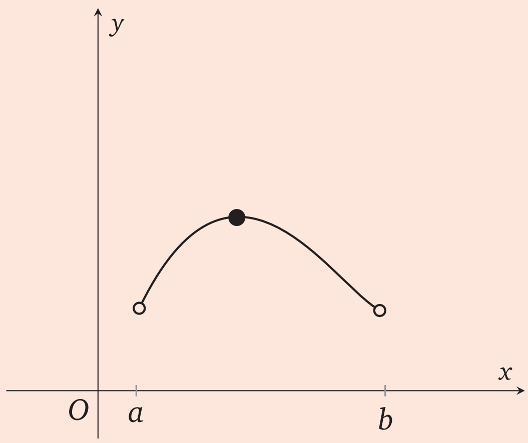
$$y - f(a) = -\frac{1}{f'(a)} \cdot (x - a)$$

Second derivative of a function (concavity)

- Let f be a function defined on an interval (a, b), and assume that both $f'(x)$ and $f''(x)$ exist for all $x \in (a, b)$.
- If $\forall x \in (a, b) : \boxed{f''(x) > 0}$, then the gradient of the curve $y = f(x)$ is increasing in the interval (a, b). The curve is **concave up** (*i.e.,* it has a **local minimum** in the interval (a, b)).
- If $\forall x \in (a, b) : \boxed{f''(x) < 0}$, then the gradient of the curve $y = f(x)$ is decreasing in the interval (a, b). The curve is **concave down** (*i.e.,* it has a **local maximum** in the interval (a, b)).
- If $\boxed{f''(x) = 0}$ for $x = a$, then there is a **stationary point of inflection** in the curve $y = f(x)$ at the point when $x = a$.



This function is **concave up** over (a, b)



This function is **concave down** over (a, b)

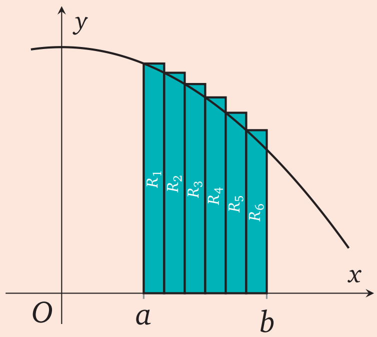
Estimating the area under a graph

Left-endpoint estimate

- The formula for the left-endpoint estimate for a function f over the domain $[a,b]$ with rectangles of width w is as follows:

$$\text{Area}_{\text{est.}} = \sum_{k=1}^{(b-a)/w} w \cdot f(a + w \cdot (k-1))$$

- For a function f that is...
  - strictly increasing in the domain $[a,b]$, the left-endpoint estimate $\leq$ actual area.
  - strictly decreasing in the domain $[a,b]$, the left-endpoint estimate $\geq$ actual area.

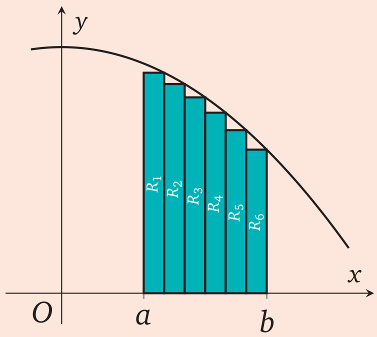


Right-endpoint estimate

- The formula for the right-endpoint estimate for a function f over the domain $[a,b]$ with rectangles of width w is as follows:

$$\text{Area}_{\text{est.}} = \sum_{k=1}^{(b-a)/w} w \cdot f(a + wk)$$

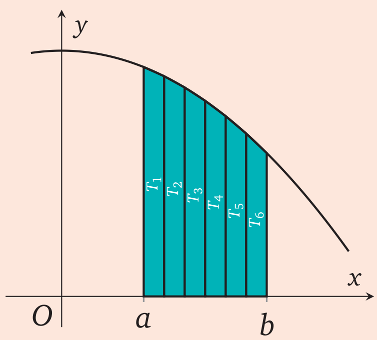
- For a function f that is...
  - strictly increasing in the domain $[a,b]$, the left-endpoint estimate $\geq$ actual area.
  - strictly decreasing in the domain $[a,b]$, the left-endpoint estimate $\leq$ actual area.



Trapezium estimate

- The formula for the right-endpoint estimate for a function f over the domain $[a,b]$ with rectangles of width w is as follows:

$$\text{Area}_{\text{est.}} = \sum_{k=1}^{(b-a)/w} w \cdot [f(a + w \cdot (k-1)) + f(a + wk)]$$



The fundamental theorem of calculus

$$\frac{\mathrm{d}}{\mathrm{d}x}[F(x)] = f(x) \quad \Longrightarrow \quad \int_a^b f(x)\,\mathrm{d}x = [F(x)]_a^b = F(b) - F(a)$$

- As the constant $(+C)$ cancels out, we normally ignore it and take the antiderivative of f with $C = 0$.

Antidifferentiation rules

Antidifferentiation results

- **Sum:** $\int [f(x) + g(x)]\,\mathrm{d}x = \int f(x)\,\mathrm{d}x + \int g(x)\,\mathrm{d}x$
- **Difference:** $\int [f(x) - g(x)]\,\mathrm{d}x = \int f(x)\,\mathrm{d}x - \int g(x)\,\mathrm{d}x$
- **Multiple:** $\int [k \cdot f(x)]\,\mathrm{d}x = k \cdot \int f(x)\,\mathrm{d}x, k \in \mathbb{R}$

Properties of the definite integral

- $\int_a^b f(x)\,\mathrm{d}x = \int_a^c f(x)\,\mathrm{d}x + \int_c^b f(x)\,\mathrm{d}x$
- $\int_a^a f(x)\,\mathrm{d}x = 0$
- $\int_a^b [k \cdot f(x)]\,\mathrm{d}x = k \cdot \int_a^b f(x)\,\mathrm{d}x$
- $\int_a^b [f(x) \pm g(x)]\,\mathrm{d}x = \int_a^b f(x)\,\mathrm{d}x \pm \int_a^b g(x)\,\mathrm{d}x$
- $\int_a^b f(x)\,\mathrm{d}x = -\int_b^a f(x)\,\mathrm{d}x$

Signed area

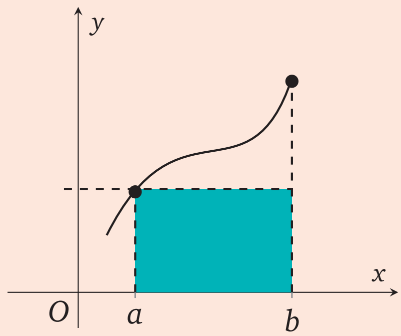
- For any continuous function f on an interval $[a, b]$, the **definite integral** $\int_a^b f(x)\,\mathrm{d}x$ gives the **signed area** enclosed by the graph of $y = f(x)$ between $x = a$ and $x = b$.
- To get the **unsigned area**, just take the absolute value of the function like so: $\int_a^b |f(x)|\,\mathrm{d}x$.

Average value of a function

- The **average value** of a continuous function f over an interval $[a, b]$ is:

$$\frac{1}{b-a} \cdot \int_a^b f(x)\,\mathrm{d}x$$

- In terms of the graph of $y = f(x)$, the average value is the **height of a rectangle** having the same area as the area under the graph for the interval $[a, b]$ (the interval forms the rectangle's base).



8 Probability

Basic laws of probability

- **Total law of probability:** $\forall x \subseteq \mathcal{E} : \Pr(X = x) = 1$
- $\Pr(X = x) \geq 0 \iff x \in \mathcal{E}$
- **Addition rule:** $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$
- $\Pr(\varnothing) = 0$
- $\Pr(A') = 1 - A$, where A' is the complement of A.

Mutually exclusive events

- Two events A and B are mutually exclusive if:

$$\Pr(A \cap B) = 0$$

- For mutually exclusive events, the addition rules becomes:

$$\Pr(A \cup B) = \Pr(A) + \Pr(B)$$

Probabilities from data

- When the number of trials is sufficiently large, the observed relative frequency of an event A becomes close to the probability $\Pr(A)$. That is,

$$\Pr(A) \approx \frac{\text{number of times } A \text{ occurs}}{\text{number of trials}} \quad \text{for a large number of trials}$$

Probability tables (Karnaugh maps)

	$\boldsymbol{B}$	$\boldsymbol{B'}$	
$\boldsymbol{A}$	$\Pr(A \cap B)$	$\Pr(A \cap B')$	$\mathbf{Pr}(\boldsymbol{A})$
$\boldsymbol{A'}$	$\Pr(A' \cap B)$	$\Pr(A' \cap B')$	$\mathbf{Pr}(\boldsymbol{A'})$
	$\mathbf{Pr}(\boldsymbol{B})$	$\mathbf{Pr}(\boldsymbol{B'})$	**1**

Conditional probability

- The **conditional probability** of an event A, given that event B has already occured, is given by:

$$\Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)} \quad \text{if } \Pr(B) \neq 0$$

- This formula may be rearranged to give the **multiplication rule of probability:**

$$\Pr(A \cap B) = \Pr(A \mid B) \cdot \Pr(B)$$

Law of total probability

- The **law of total probability** states that, in the case of two events A and B,

$$\Pr(A) = \Pr(A \mid B) \cdot \Pr(B) + \Pr(A \mid B') \cdot \Pr(B')$$

Independent events

- For events A and B with $\Pr(A) \neq 0$ and $\Pr(A) \neq 0$, the following three conditions are all **equivalent conditions** for the independence of A and B:
  - $\Pr(A \mid B) = \Pr(A)$
  - $\Pr(B \mid A) = \Pr(B)$
  - $\Pr(A \cap B) = \Pr(A) \cdot \Pr(B)$
- In the special case that $\Pr(A) = 0$ or $\Pr(B) = 0$, the third condition ($\Pr(A \cap B) = \Pr(A) \cdot \Pr(B)$) still holds since both sides are zero, so events A and B are still independent.

Discrete probability functions

- The probability distribution of X is a function $p(x) = Pr(X = x)$ that assigns a probability to each value of X. It can be represented by a rule, a table or a graph, and must give a probability $p(x)$ for every value x that X can take.
- For *any* discrete probability function $p(x)$, the following two conditions must hold:
  - Each value of $p(x)$ belongs to the interval $[0, 1]$. That is,

$$\forall x \in \mathrm{dom}(p) : 0 \leq p(x) \leq 1$$

  - The sum of all the values of $p(x)$ must be 1. That is,

$$\sum_x p(x) = 1$$

- The sum of the values of values of $p(x)$ for x between a and b inclusive is written as

$$\sum_{a \leq x \leq b} p(x) = \Pr(a \leq X \leq b)$$

Population parameters

Expected value

- The **expected value** of a discrete random variable X is determined by summing the products of each value of X and the probability that X takes that value. That is,

$$\begin{aligned}\mathrm{E}(X) &= \sum_x [x \cdot \Pr(X = x)] \\ &= \sum_x [x \cdot p(x)]\end{aligned}$$

- The expected value $\mathrm{E}(X)$ may be considered as the long-run average value of X.
- It is generally denoted by μ, and is also called the **mean** of X.
- $\mathrm{E}[g(X)] = \sum_x [g(x) \cdot p(x)]$
- $\mathrm{E}(aX + b) = a \cdot \mathrm{E}(X) + b \quad$ (for a, b constant)
  - Generally, $\mathrm{E}[g(X)] \neq g[\mathrm{E}(X)]$, but the linear case is an exception.
- If X and Y are two random variables, then $\mathrm{E}(X + Y) = \mathrm{E}(X) + \mathrm{E}(Y)$

Variance

- The **variance** of a random variable X is the measure of the spread of the probability distribution about its mean or expected value μ.
- It is defined as:

$$\begin{aligned}\mathrm{Var}(X) &= \mathrm{E}\left[(X - \mu)^2\right] \\ &= \sum_x \left[(x - \mu)^2 \cdot \Pr(X = x)\right] \\ &= \sum_x \left[(x - \mu)^2 \cdot p(x)\right]\end{aligned}$$

- Alternatively, the computational formula for calculating variance is as such:

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

- It may be considered the long-run average value of the square of the distance from X to *mu*.
- The variance is denoted using σ^2.
- $\mathrm{Var}(aX + b) = a^2 \cdot \mathrm{Var}(X) \quad$ (for a, b constant)

Standard deviation

- The **standard deviation** is defined as the square-root of the variance σ^2. That is,

$$\mathrm{sd}(X) = \sqrt{\mathrm{Var}(X)}$$

- It is usually denoted with σ.

Bernoulli sequence

- A **Bernoulli sequence** is the name used to describe a sequence of repeated trials with the following properties:
  - Each trial results in one of two outcomes, which are usually designated as either a success, S, or a failure, F.
  - The probability of success on a single trial, p, is constant for all trials (and thus the probability of failure on a single trial is $1-p$).
  - The trials are independent (so that the outcome of any trial is not affected by the outcome of any previous trial).

Binomial probability distribution

- The number of successes in a Bernoulli sequence of n trials is called a **binomial random variable** and is said to have a **binomial probability distribution**.
- If the random variable X is the number of successes in n independent trials, each with probability of success p, then X has a **binomial distribution**, written $X \sim \mathrm{Bi}(n,p)$ and the rule is

$$\Pr(X=x) = \binom{n}{x} \cdot p^x \cdot (1-p)^{n-x} \quad x = 0, 1, \ldots, n$$

where $\binom{n}{x} = \dfrac{n!}{x! \cdot (n-x)!}$

- As the value of p increases, the graph of the binomial distribution is more skewed to the right (negatively skewed). A value of $p = 0.5$ makes the peak of the graph of $y = p(x)$ line up with the midway of the interval $[0, n]$ of the x-axis.

Population parameters for the binomial distribution

- $\mathrm{E}(X) = np$
- $\mathrm{Var}(X) = np(1-p)$

Probability density functions

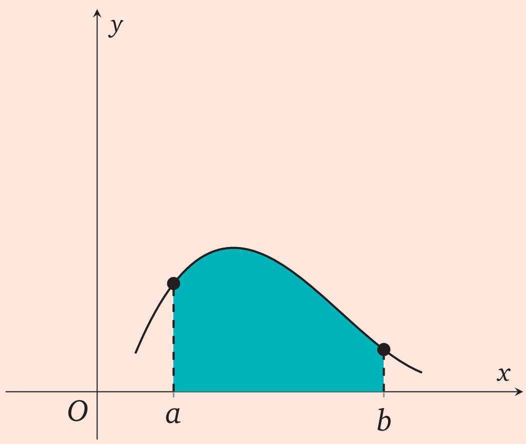
- In general, the probability density function f is a function with domain some interval (*e.g.*, domain $[c,d]$ or $\mathbb{R}$) such that:
  1. $\forall x \in \text{dom}(f) : f(x) \geq 0$
  2. The area under the graph of $y = f(x)$ is equal to 1.
     - If the domain of f is $[c,d]$, then this condition corresponds to $\displaystyle\int_c^d f(x)\,dx = 1$.
- The values of a probability density function f are not probabilities, and $f(x)$ may take values greater than 1.
- The probability of any specific value of X is 0. That is, $\Pr(X = a) = 0$.
- It follows that all of the following expressions have the same numerical value:
  - $\Pr(a < X < b)$
  - $\Pr(a \leq X < b)$
  - $\Pr(a < X \leq b)$
  - $\Pr(a \leq X \leq b)$
- If f has the domain $[c,d]$ and $a \in [c,d]$, then $\Pr(X < a) = \Pr(X \leq a) = \displaystyle\int_c^a f(x)\,dx$.

Visualising a probability density function

- If X is a continuous random variable with density function f, then

$$\Pr(a < X < b) = \int_a^b f(x)\,dx$$

which is the area of the shaded region.



Computing improper integrals

- If $\mathrm{dom}(f) = (-\infty, a]$, then $\displaystyle\int_{-\infty}^{a} f(x)\,\mathrm{d}x = 1$. This integral is computed as

$$\lim_{k\to\infty} \int_{-k}^{a} f(x)\,\mathrm{d}x$$

- If $\mathrm{dom}(f) = [a, \infty)$, then $\displaystyle\int_{a}^{\infty} f(x)\,\mathrm{d}x = 1$. This integral is computed as

$$\lim_{k\to\infty} \int_{a}^{k} f(x)\,\mathrm{d}x$$

- If $\mathrm{dom}(f) = (-\infty, \infty)$, then $\displaystyle\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1$. This integral is computed as

$$\lim_{k\to\infty} \int_{-k}^{k} f(x)\,\mathrm{d}x$$

Properties for a continuous probability distribution

Expected value/mean

- For a continuous random variable X with probability density function f, the **mean** or **expected value** of X is given by

$$\mathrm{E}(X) = \int_{-\infty}^{\infty} f(x)\,\mathrm{d}x$$

provided the integral exists.

- If $f(x) = 0$ for all $x \notin [c, d]$, then

$$\mathrm{E}(X) = \int_{c}^{d} f(x)\,\mathrm{d}x$$

- This definition is consistent with the definition provided in the “Expected Value” section of the “Population parameters” box. Where appropriate, substitute an integral for the summation symbol and f in place of p.

Percentiles

- The value p of X which is the solution of an equation of the form

$$\int_{-\infty}^{p} f(x)\,\mathrm{d}x = q$$

is called a **percentile** of the distribution.

- For example, the 75th percentile is the value p found by taking $q = 75\% = 0.75$.

The median

- The **median** is another measure of centre for a continuous probability distribution.
- The median, m, of a continuous random variable X is the value of X such that

$$\int_{-\infty}^{m} f(x)\,\mathrm{d}x = 0.5$$

- It is also known as the **50th** percentile.

Interquartile range

- The **interquartile range** is the range of the middle 50% of the distribution; it is the difference between the 75th percentile (also known as Q3) and the 25th percentile (also known as Q1).

$$\mathrm{IQR} = b - a$$

where a and b are such that

$$\int_{-\infty}^{a} f(x)\,\mathrm{d}(x) = 0.25 \quad \text{and} \quad \int_{-\infty}^{b} f(x)\,\mathrm{d}(x) = 0.75$$

The variance of a continuous probability distribution

$$\begin{aligned}\mathrm{Var}(X) &= \mathrm{E}\left(X^2\right) - \mu^2 \\ &= \mathrm{E}\left[(X-\mu)^2\right] \\ &= \int_{-\infty}^{\infty} \left[(x-\mu)^2 \cdot f(x)\right] \mathrm{d}x\end{aligned}$$

The standard deviation of a continuous probability distribution

$$\mathrm{sd}(X) = \sqrt{\mathrm{Var}(X)}$$

The probability density function of $aX + b$

- If the probability density function of X has the rule $f(x)$, then the probability density function of $aX + b$ is $\frac{1}{a} \cdot f\left(\frac{x-b}{a}\right)$ and is described by the transformation

$$(x, y) \to \left(ax + b, \frac{y}{a}\right)$$

The standard normal distribution

- A random variable Z with the standard normal distribution has probability density function

$$f(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2} \cdot x^2}$$

- The standard normal distribution has mean $\mu = 0$ and standard deviation $\sigma = 1$.

Transformations of normal distributions

- If X is a **normally distributed random variable** with mean μ and standard deviation σ, then the probability density function of X is given by

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

and

$$\Pr(X \le a) = \Pr\left(Z \le \frac{a-\mu}{\sigma}\right)$$

where Z is the random variable of the standard normal distribution.

- The transformation which maps the graph of a normal distribution with mean μ and standard deviation σ to the graph of the standard normal distribution is as follows:

$$(x, y) \to \left(\frac{x-\mu}{\sigma}, \sigma y\right)$$

- Conversely, the transformation which maps the graph of the standard normal distribution to the graph of a normal distribution with mean μ and standard deviation σ is as follows:

$$(x, y) \to \left(\sigma x + \mu, \frac{y}{\sigma}\right)$$

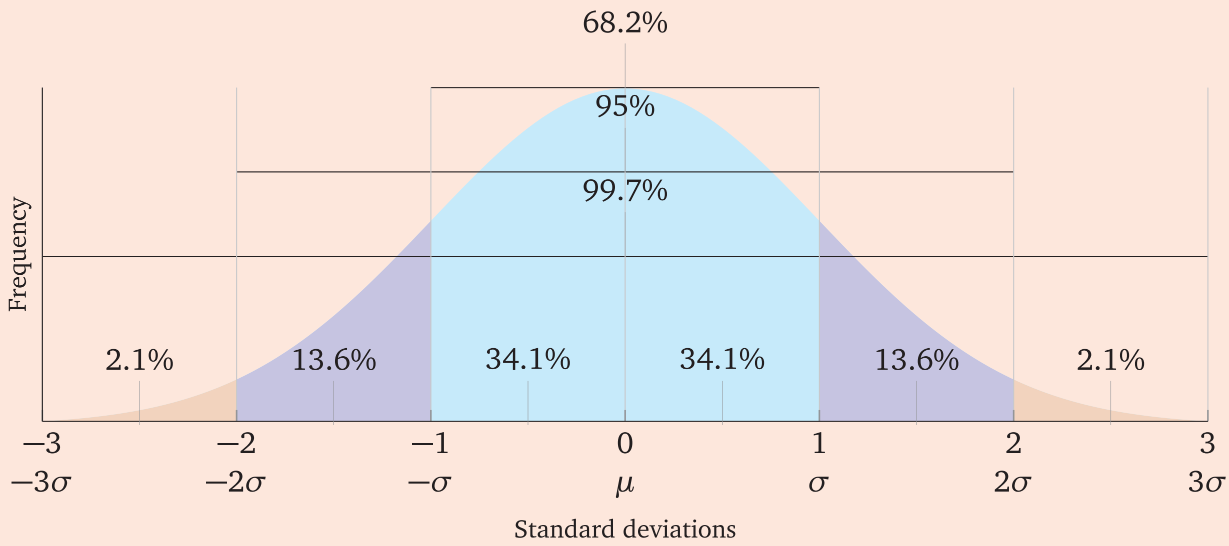
- These transformations are “area preserving”.

Symmetry properties of the standard normal distribution

- $\Pr(Z > a) = 1 - \Pr(Z \le a)$
- $\Pr(Z < -a) = \Pr(Z > a)$
- $\Pr(-a < Z < a) = 1 - 2\Pr(Z \ge a)$
  $= 1 - 2\Pr(Z \le -a)$

Empirical formulas

- For a normally distributed random variable, approximately:
  - 68% of values lie within one standard deviation of the mean, which is the interval $[\mu - \sigma, \mu + \sigma]$.
  - 95% of values lie within two standard deviation of the mean, which is the interval $[\mu - 2\sigma, \mu + 2\sigma]$.
  - 99.7% of values lie within three standard deviation of the mean, which is the interval $[\mu - 3\sigma, \mu + 3\sigma]$.



Normal approximation of a binomial distribution

- If n is sufficiently large, the binomial random variable X will be approximately normally distributed, with a mean of $\mu = np$ and a standard deviation of $\sigma = \sqrt{np(1-p)}$.
- One rule of thumb is that $np > 5$ **and** $n(1-p) > 5$ for a satisfactory approximation.

9 Sampling

Sample

- A sample of size n is called a **simple random sample** if it is selected from the population in such a way that every subset of size n has an equal chance of being chosen as the sample.
- In particular, every member of the population must have an equal chance of being included in the sample.

Population and sample proportions

- The **population proportion** p is a **population parameter**; its value is constant. This is also what is used as the value for the probability of success when calculating $\hat{p}$ from a binomial distribution.

$$p = \frac{\text{number in population with attribute}}{\text{population size}}$$

- The **sample proportion** $\hat{p}$ is a **sample statistic**; its value is not constant, but varies from sample to sample.

$$\hat{p} = \frac{\text{number in sample with attribute}}{\text{sample size}} = \frac{X}{n}$$

  where $X \sim \text{Bi}(n, p)$, $p =$ probability of a member of the population having the desired attribute.
- Since $\hat{p}$ varies according to the contents of the random samples, we can consider the sample proportions $\hat{p}$ as being the values of a random variable, which we will denote by $\hat{P}$.

Hypergeometric distribution

- The **hypergeometric distribution** is a *discrete* probability distribution that describes the probability of k successes (random draws for which the object drawn has a specified/desired feature) in n draws (a sample size of n), **without replacement** (the next draw is happening from a population size of $N-1$) from a finite population of size N that contains exactly K objects with that feature, wherein each draw is either a success or failure (a Bernoulli trial).
- The probability density function of such a distribution is as described:

$$p_X(k) = \Pr(X = k) = \frac{\binom{K}{k} \cdot \binom{N-K}{n-k}}{\binom{N}{n}}$$

- This is denoted as $X \sim \text{Hypergeometric}(N, K, n)$.
- This distribution is converse to the binomial distribution, which describes the probability of k successes in n draws *with replacement*.

Types of distributions for calculating $\hat{p}$

- If the sample is being taken **without replacement**, then we can say that $\hat{p} = \frac{X}{n}$, where $X \sim \text{Hypergeometric}(N, K, n)$ (N is the population size, K is the number of members of the population with the desired/specified feature, and n is the sample size).
  - This is typically done with small, countable population sizes (*e.g.,* marbles in a bag, *etc.*).
- If the sample is being taken **with replacement**, $\hat{p} = \frac{X}{n}$, where $X \sim \text{Bi}(n, p)$ (n is the sample size, and p is the probability of selecting x member(s) out of the population which possess the desired/specified feature (*i.e.,* a success) (where $x = 0, 1, \ldots, n$)).
  - This is typically done with large populations consisting of an uncountable number of members (*i.e.,* a country). Normally, this is because you are not given N, the population size, but just p, which can be used to work out $\hat{p}$.
- The distribution of a statistic which is calculated from a sample (such as the sample proportion) has a special name — it is called a **sampling distribution**.

Population parameters for the sample

- If we are selecting a random sample of size n from a *large* population (binomial distribution), then the mean and standard deviation of the sample proportion $\hat{P}$ are given by:

$$\text{E}(\hat{P}) = p \quad \text{and} \quad \text{sd}(\hat{P}) = \sqrt{\frac{p(1-p)}{n}}$$

- The standard deviation of a sample statistic is called the **standard error**.

Normal approximation of the sample distribution

- When the sample size n is *large*, the sample proportion $\hat{P}$ has an approximately normal distribution, with mean $\mu = p$ and standard deviation $\sigma = \sqrt{\frac{p(1-p)}{n}}$.
  - Approximate the sample distribution to a normal distribution when asked to find n, the sample size and when given p, and $\Pr(\hat{P} > a)$ (or anything of the sort). To do this, you may use the `invNorm(Area,`μ`,`σ`)` function on your CAS.

Inference of the population

Point estimates

- The value of the sample proportion $\hat{p}$ can be used to estimate the population proportion p.
- Since this is a single-valued estimate, it is called a **point estimate** of p.

Interval estimates (confidence intervals)

- The value of the sample proportion $\hat{p}$ obtained from a single sample is going to change from sample to sample.
- What is required is an interval that we are reasonably sure contains the parameter value p.
- An **interval estimate** for the population proportion p is called a **confidence interval** for p.

Finding confidence intervals

- When the sample size n is *large* (both np and $n(1-p)$ must be larger than 5), the sample proportion $\hat{P}$ has an approximately normal distribution with $\mu = p$ and $\sigma = \sqrt{\dfrac{p(1-p)}{n}}$.
- $\therefore Z_{\hat{p}} = \dfrac{\hat{P} - \mu_{\hat{P}}}{\sigma_{\hat{P}}} = \dfrac{\hat{P} - p}{\sqrt{\frac{p(1-p)}{n}}}$, where $Z_{\hat{p}}$ is the standard normal variable of the sample distribution $\hat{P}$.
- The **standardised** a% confidence interval can be found using:

$$\begin{aligned}\Pr(-c < Z_{\hat{p}} < c) &= a, 0 < a < 1\\ \implies \Pr(Z_{\hat{p}} < c) &= \frac{1-a}{2} + a, 0 < a < 1\\ &= \frac{a+1}{2}\end{aligned}$$

  This is thanks to the symmetry properties of the approximated normal distribution. The `invNorm(Area,`μ`,`σ`)` function on your CAS can be used to find the value of c.
- Remember, the sample proportion $\hat{p}$ **lies in the middle** of the confidence interval.
- Rearranging this (to the **un-standardised** version), we get the formula given on the formula sheet:

$$C\%\text{ confidence interval} = \left(\hat{p} - k\cdot\sqrt{\frac{\hat{p}\cdot(1-\hat{p})}{n}}, \hat{p} + k\cdot\sqrt{\frac{\hat{p}\cdot(1-\hat{p})}{n}}\right)$$

  where k is such that $\Pr(-k < Z_{\hat{p}} < k) = \dfrac{C}{100}$.
  - The **1-prop z interval** function can be used on the CAS to find the un-standardised C.I. (found in Menu → Statistics → Confidence Intervals → 1-Prop z interval.

k values for confidence intervals

- 68.2% C.I.: $k = \texttt{invNorm(0.841,0,1)} = 0.998\,576\,278\,454\,53 \approx 0.9986$
- 90% C.I.: $k = \texttt{invNorm(0.95,0,1)} = 1.644\,853\,625\,906\,6 \approx 1.6449$
- 95% C.I.: $k = \texttt{invNorm(0.975,0,1)} = 1.959\,963\,985\,991\,5 \approx 1.9600$
- 99% C.I.: $k = \texttt{invNorm(0.995,0,1)} = 2.575\,829\,303\,001\,6 \approx 2.5758$
- 99.7% C.I.: $k = \texttt{invNorm(0.9985,0,1)} = 2.967\,737\,927\,124\,7 \approx 2.9677$

Margin of error

- The **distance** between the sample estimate and the endpoints of the confidence interval is called the **margin of error** (M).
- For a C% confidence interval, the margin of error M is as such:

$$M = k \cdot \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}}$$

  where k is the value corresponding to the confidence interval percentage.
- If p^* is an estimated value for the population proportion p, then
  - a C% confidence interval for a population proportion p will have margin of error approximately equal to a specified value of M when the sample size is:

$$n = \left(\frac{k}{M}\right)^2 \cdot p^* \cdot (1 - p^*)$$

    where M is the margin of error and k is the value associated with the C% confidence interval.

Part II

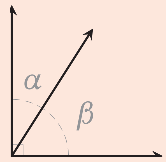
Extension

1 Angle relationships

Complementary and supplementary angles

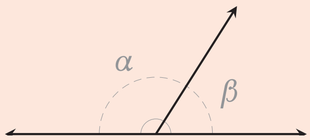
Complimentary angles

- In this case, the angles α and β are complimentary, as $\alpha + \beta = 90^\circ$.



Supplementary angles

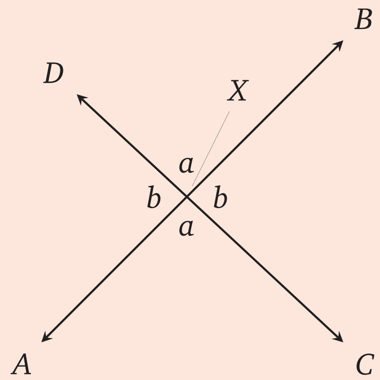
- In this case, the angles α and β are complimentary, as $\alpha + \beta = 180^\circ$.



Angles formed by intersecting lines

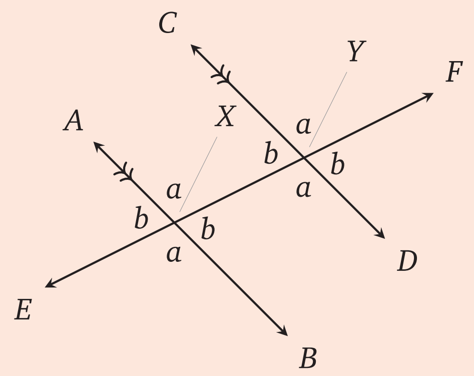
Vertically opposite angles

- In this case, $\angle AXC = \angle DXB$ and $\angle DXA = \angle BXC$.



Angles formed by a transversal

- A **transversal** is a line which crosses two or more lines.
- In this case, the lines AB and CD are **parallel**, which is denoted by $AB \parallel CD$.
- **Vertically opposite angles**:
  - $\angle CYF = \angle XYD$
  - $\angle YFD = \angle CYX$
  - $\angle AXY = \angle EXB$
  - $\angle AXE = \angle YXB$
- **Alternate interior angles**:
  - $\angle AXY = \angle DYX$
  - $\angle CYX = \angle BXY$
- **Alternate exterior angles**:
  - $\angle FYD = \angle AXE$
  - $\angle FYC = \angle EXB$
- **Corresponding angles**:
  - $\angle FYD = \angle YXB$
  - $\angle FYC = \angle YXA$
  - $\angle EXB = \angle XYD$
  - $\angle EXA = \angle XYC$
- **Same side interior angles (supplementary)**:
  - $\angle XYD + \angle YXB = 180°$
  - $\angle AXY + \angle CYX = 180°$



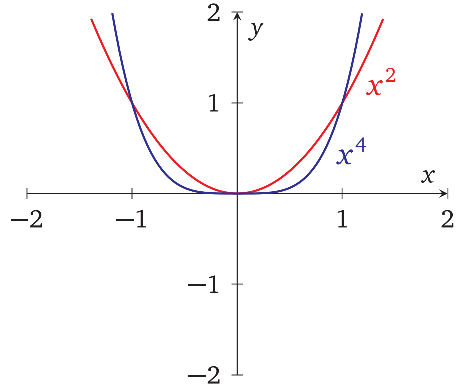
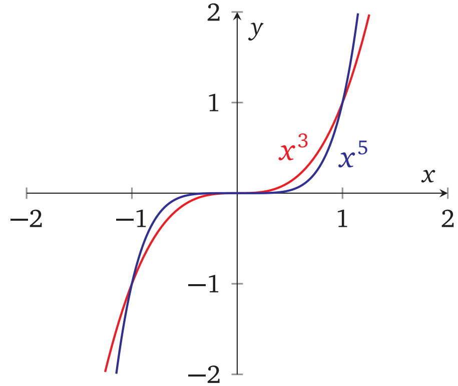
Pascal's triangle

- Featured below is **pascal's triangle**, in which each row n and column k correspond to $\binom{n}{k}$.
  - Binomial expansion: $(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} \cdot a^{n-k} \cdot b^k$ and $(qa+b)^n = \sum_{k=0}^{n} \binom{n}{k} \cdot (q \cdot a)^{n-k} \cdot b^k$

n																
0	1															
1	1	1														
2	1	2	1													
3	1	3	3	1												
4	1	4	6	4	1											
5	1	5	10	10	5	1										
6	1	6	15	20	15	6	1									
7	1	7	21	35	35	21	7	1								
8	1	8	28	56	70	56	28	8	1							
9	1	9	36	84	126	126	84	36	9	1						
10	1	10	45	120	210	252	210	120	45	10	1					
11	1	11	55	165	330	462	462	330	165	55	11	1				
12	1	12	66	220	495	792	924	792	495	220	66	12	1			
13	1	13	78	286	715	1287	1716	1716	1287	715	286	78	13	1		
14	1	14	91	364	1001	2002	3003	3432	3003	2002	1001	364	91	14	1	
15	1	15	105	455	1365	3003	5005	6435	6435	5005	3003	1365	455	105	15	1
k	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

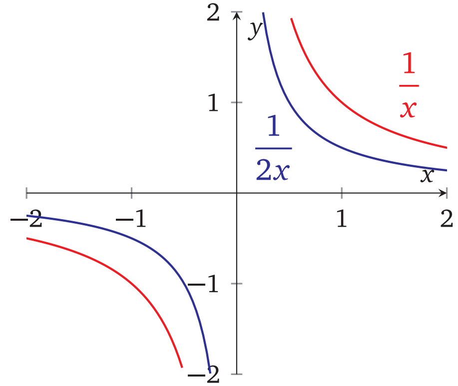
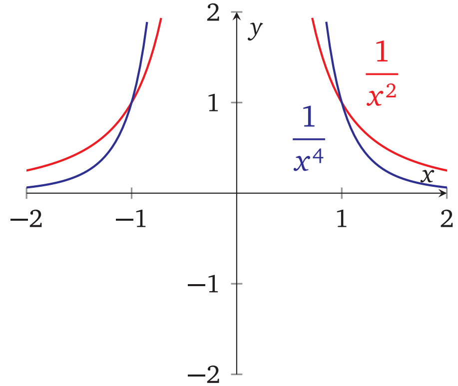
3 Base functions

Table 1: Base graphs

<table>
<tr><th>Rule</th><th>Implied domain</th><th>Range</th><th>Parity</th><th>Graph</th><th>Inverse</th><th>Asymptote</th></tr>
<tr><td>x^n, n is even</td><td>$\mathbb{R}$</td><td>$[0, \infty)$</td><td>Even</td><td>
</td><td>$\sqrt[n]{x}, n$ is even</td><td>None</td></tr>
<tr><td>x^n, n is odd</td><td>$\mathbb{R}$</td><td>$\mathbb{R}$</td><td>Odd</td><td>
</td><td>$\sqrt[n]{x}, n$ is odd</td><td>None</td></tr>
</table>

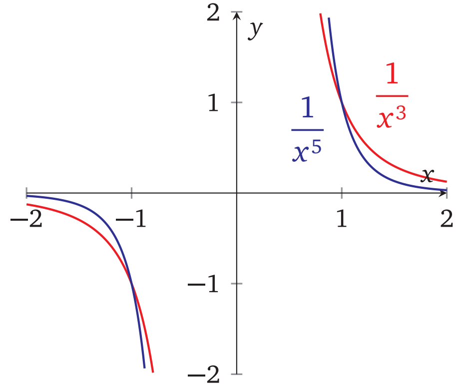
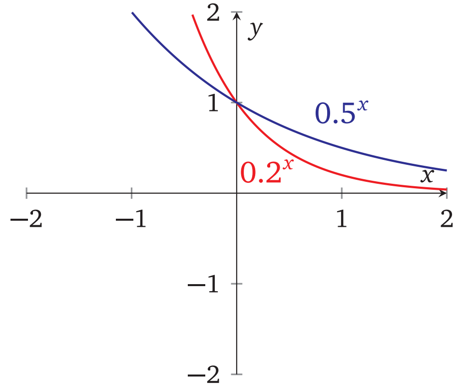
Continued on next page

Table 1: Base graphs (Continued)

Rule	Implied domain	Range	Parity	Graph	Inverse	Asymptote
$\frac{1}{x}$	$\mathbb{R} \setminus \{0\}$	$\mathbb{R}$	Odd		$\frac{1}{x}$	$y = 0$ $x = 0$
$\frac{1}{x^n}$, n is even	$\mathbb{R} \setminus \{0\}$	$\mathbb{R}^+$	Even		$\pm\frac{1}{\sqrt[n]{x}}$	$y = 0$ $x = 0$

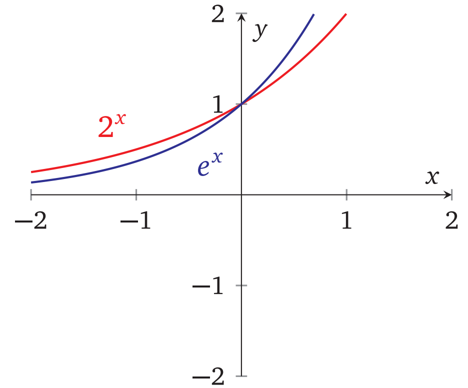
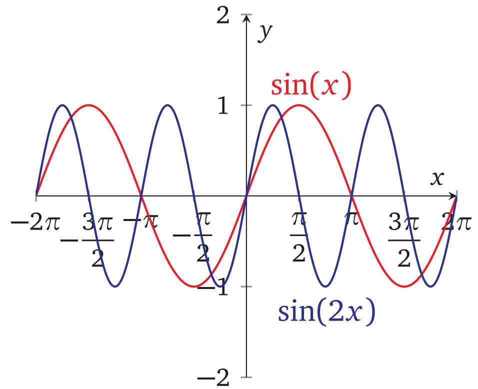
Continued on next page

Table 1: Base graphs (Continued)

Rule	Implied domain	Range	Parity	Graph	Inverse	Asymptote
$\frac{1}{x^n}, n$ is odd	$\mathbb{R} \setminus \{0\}$	$\mathbb{R} \setminus \{0\}$	Odd		$\frac{1}{\sqrt[n]{x}}$	$y = 0$ $x = 0$
$a^x, 0 < a < 1$	$\mathbb{R}$	$\mathbb{R}^+$	None		$\sqrt[a]{x}$	$y = 0$

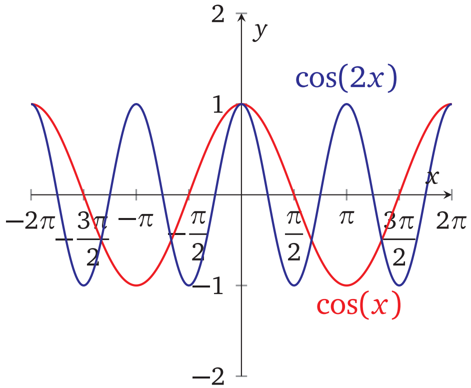
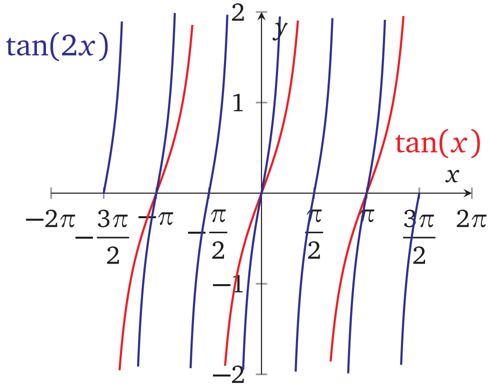
Continued on next page

Table 1: Base graphs (Continued)

Rule	Implied domain	Range	Parity	Graph	Inverse	Asymptote
$a^x, a > 1$	$\mathbb{R}$	$\mathbb{R}^+$	None		$\sqrt[a]{x}$	$y = 0$
$\sin(x)$	$\mathbb{R}$	$[-1, 1]$	Odd		$\sin^{-1}(x)$	None

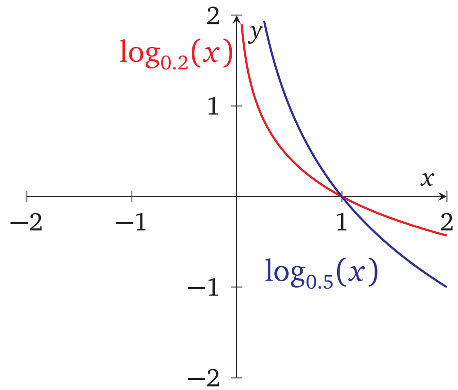
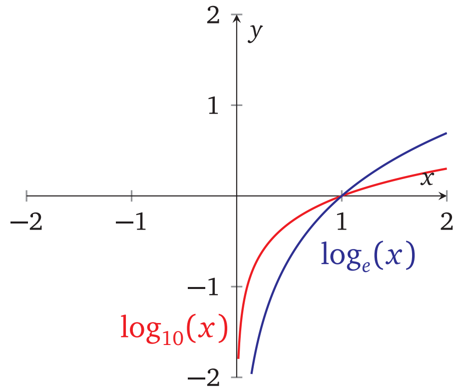
Continued on next page

Table 1: Base graphs (Continued)

Rule	Implied domain	Range	Parity	Graph	Inverse	Asymptote
$\cos(x)$	$\mathbb{R}$	$[-1,1]$	Even		$\cos^{-1}(x)$	None
$\tan(x)$	$x \neq \dfrac{(2n-1)\pi}{2}, n \in \mathbb{Z}$	$\mathbb{R}$	Odd		$\tan^{-1}(x)$	$x = \dfrac{(2n-1)\pi}{2}$

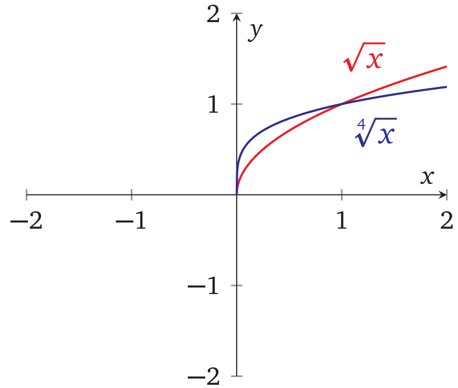
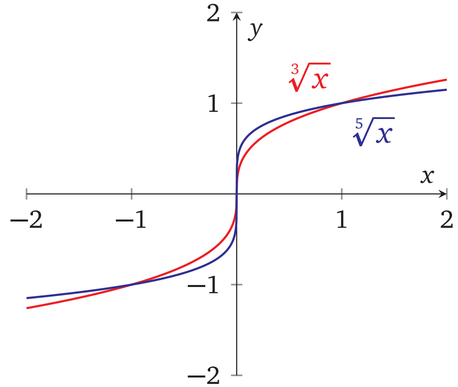
Continued on next page

Table 1: Base graphs (Continued)

Rule	Implied domain	Range	Parity	Graph	Inverse	Asymptote
$\log_a(x)$, $0 < a < 1$	$\mathbb{R}^+$	$\mathbb{R}$	None		$a^x, 0 < a < 1$	$y = 0$
$\log_a(x), a > 1$	$\mathbb{R}^+$	$\mathbb{R}$	None		$a^x, a > 1$	$y = 0$

Continued on next page

Table 1: Base graphs (Continued)

Rule	Implied domain	Range	Parity	Graph	Inverse	Asymptote
$\sqrt[n]{x}, n$ is even	$\mathbb{R}^+ \cup \{0\}$	$\mathbb{R}^+$	None		x^n	None
$\sqrt[n]{x}, n$ is odd	$\mathbb{R}$	$\mathbb{R}$	None		x^n	None