

Mathematical Methods — Bound Reference

March 30, 2024

Contents

I	Theory	1
1	Functions and relations	1
1.1	Set notation	1
1.2	Sets of numbers	1
1.3	Interval notation	2
1.4	Functions VS relations	2
1.4.1	Vertical line test	2
1.5	Function notation	2
1.6	Types of functions (many/one-to-one)	3
1.6.1	Horizontal line test	3
1.7	Parity of functions	3
1.8	Implied/maximal domain	3
1.9	Sum and product of functions	4
1.9.1	Addition of ordinates (sketching $y = (f + g)(x)$)	4
1.10	Composite functions	4
1.11	Increasing and decreasing functions	5
1.12	Inverse functions	6
2	Coordinate geometry	7
2.1	Solving simultaneous equations	7
2.2	Linear coordinate geometry	7
2.2.1	Distance between two points	7
2.2.2	Midpoint of a line	7
2.2.3	Gradient of a line	7
2.2.4	Equation of a line	8
2.2.5	Tangent of the angle of slope	8
2.2.6	Perpendicular and parallel lines	8
2.3	The geometry of simultaneous linear equations	9
2.4	Transformations	10
2.4.1	Dilations	10
2.4.2	Table of transformations	11
2.4.3	Applying transformations	11
3	Polynomial functions	12
3.1	Quadratics	12
3.2	Remainder theorem	12
3.3	Factor theorem	13
3.4	Rational root theorem	13
3.5	Polynomials of degree n	13
3.6	Difference and sum of two variables of the same degree	13
4	Exponential functions	14
4.1	Exponential function characteristics	14
4.2	Euler's number — e	14
4.3	Index laws	14
4.4	Logarithms	15
4.4.1	Log laws	15

4.5	Exponential growth and decay	15
5	Circular functions	16
5.1	Radians and degrees	16
5.2	Unit circle	16
5.3	Trigonometric functions as triangles	17
5.4	Properties of trigonometric functions	17
5.5	Trigonometric identities	18
5.5.1	Pythagorean identities	18
5.5.2	Double-angle identities	18
5.5.3	Sum/Difference identities	18
5.5.4	Product-to-sum identities	18
5.5.5	Triple-angle identities	18
5.6	General solutions	19
5.6.1	General solutions for $\sin(x)$	19
5.6.2	General solutions for $\cos(x)$	19
5.6.3	General solutions for $\tan(x)$	19
5.7	Period of two trigonometric functions' sum/difference	19
6	Differentiation	20
6.1	Average rate of change	20
6.2	Differentiation from first principles	20
6.3	Derivative rules	20
6.3.1	Differentiation results	20
6.4	Chain rule	20
6.5	Limits	21
6.5.1	Algebra of limits	21
6.5.2	Left and right limits	21
6.6	Continuity of a function	21
6.7	Differentiability of a function	21
6.8	Tangent line	21
6.9	Normal line	22
6.10	Second derivative of a function (concavity)	22
6.11	Newton's method	23
7	Integration	24
7.1	Estimating the area under a graph	24
7.1.1	Left-endpoint estimate	24
7.1.2	Right-endpoint estimate	24
7.1.3	Trapezium estimate	24
7.2	The fundamental theorem of calculus	25
7.3	Antidifferentiation rules	25
7.3.1	Antidifferentiation results	25
7.3.2	Properties of the definite integral	25
7.4	Signed area	26
7.5	Average value of a function	26
8	Probability	27
8.1	Basic laws of probability	27
8.2	Mutually exclusive events	27
8.3	Probabilities from data	27

8.4	Probability tables (Karnaugh maps)	27
8.5	Conditional probability	27
8.6	Law of total probability	28
8.7	Independent events	28
8.8	Discrete probability functions	28
8.9	Population parameters	29
8.9.1	Expected value	29
8.9.2	Variance	29
8.9.3	Standard deviation	30
8.10	Bernoulli sequence	30
8.11	Binomial probability distribution	30
8.11.1	Population parameters for the binomial distribution	30
8.12	Probability density functions	31
8.12.1	Visualising a probability density function	31
8.13	Computing improper integrals	32
8.14	Properties for a continuous probability distribution	32
8.14.1	Expected value/mean	32
8.14.2	Percentiles	33
8.14.3	The median	33
8.14.4	Interquartile range	33
8.14.5	The variance of a continuous probability distribution	33
8.14.6	The standard deviation of a continuous probability distribution	34
8.15	The probability density function of $aX + b$	34
8.16	The standard normal distribution	34
8.16.1	Transformations of normal distributions	34
8.16.2	Symmetry properties of the standard normal distribution	35
8.17	Empirical formulas	35
8.18	Normal approximation of a binomial distribution	35

9 Sampling 36

9.1	Sample	36
9.2	Population and sample proportions	36
9.3	Hypergeometric distribution	36
9.4	Types of distributions for calculating \hat{p}	37
9.5	Population parameters for the sample	37
9.6	Normal approximation of the sample distribution	37
9.7	Inference of the population	38
9.7.1	Point estimates	38
9.7.2	Interval estimates (confidence intervals)	38
9.8	Finding confidence intervals	38
9.8.1	k values for confidence intervals	39
9.8.2	Margin of error	39

II Extension 40

1 Angle relationships 40

1.1	Complementary and supplementary angles	40
1.1.1	Complimentary angles	40
1.1.2	Supplementary angles	40
1.2	Angles formed by intersecting lines	40

1.2.1	Vertically opposite angles	40
1.2.2	Angles formed by a transversal	41
2	Pseudocode	42
2.1	What is pseudocode	42
2.2	Concepts of pseudocode	42
2.2.1	Comments	42
2.2.2	Conditional statements	42
2.2.3	For loops	43
2.2.4	While loops	43
2.2.5	Functions	43
2.2.6	Variables	44
2.2.7	All concepts used in conjunction	44
3	Counting methods	45
3.1	Pascal's triangle	45
4	Base functions	46

Part I

Theory

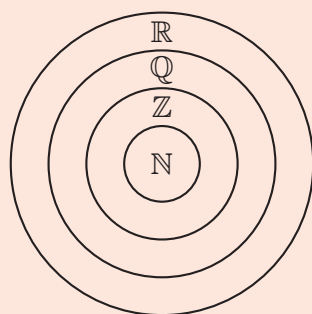
1 Functions and relations

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 $B \subseteq A$.
- \emptyset 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 ($A \cap B = \emptyset$).
- 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

- $\mathbb{N} = \{1, 2, 3, \dots\}$ = Counting numbers.
- $\mathbb{Z} = \{\dots, -3, -2, -1, 0, 1, 2, 3, \dots\}$ = 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 $a < b$.

- $(a, b) = \{x : a < x < b\}$
- $[a, b] = \{x : a \leq x \leq b\}$
- $(a, b] = \{x : a < x \leq b\}$
- $[a, b) = \{x : a \leq x < b\}$
- $(a, \infty) = \{x : a < x\}$
- $[a, \infty) = \{x : a \leq x\}$
- $(-\infty, b) = \{x : x < b\}$
- $(-\infty, b] = \{x : x \leq b\}$

When using number lines to represent intervals,

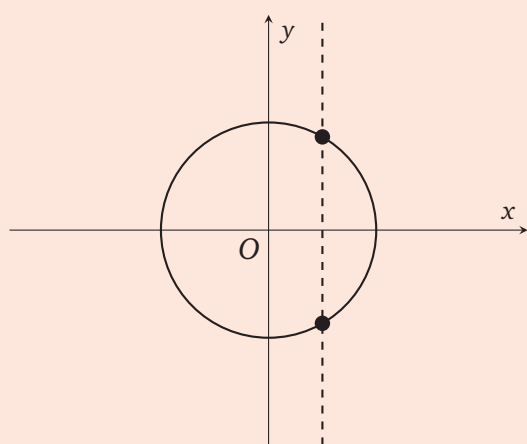
- The 'closed' circle (\bullet) indicates that the number is included.
- The 'open' circle (\circ) indicates that the number is **not** included.

Functions VS relations

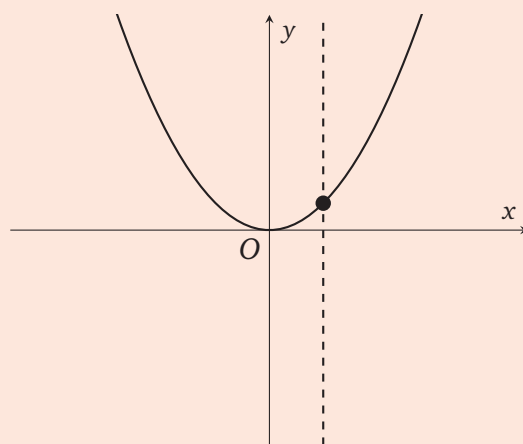
- A **function** is a relation such that for each x -value there is only one corresponding y -value. This means that, if (a, b) and (a, c) are ordered pairs of a function, then $b = c$.
- In other words, a function cannot contain two different ordered pairs with the same first coordinate.

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

$$\underline{f : X \rightarrow Y, f(x) = \dots}$$

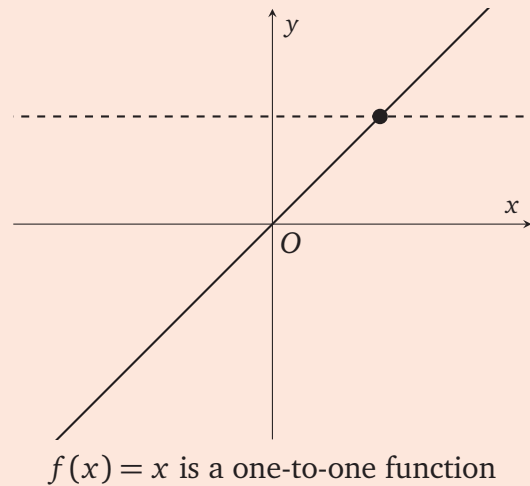
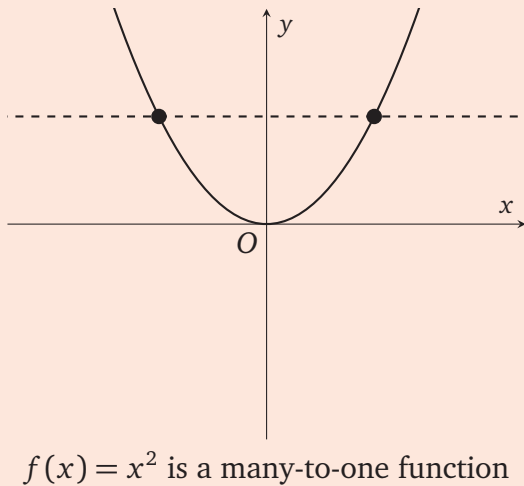
- f is the function name
- X is the **domain** of the function, *i.e.*, the set of values for which the function is defined.
- Y is the **codomain** of the function, *i.e.*, the set of values which the **range** (the range is the set of outputs of the function) of the function falls into.

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**.



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

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

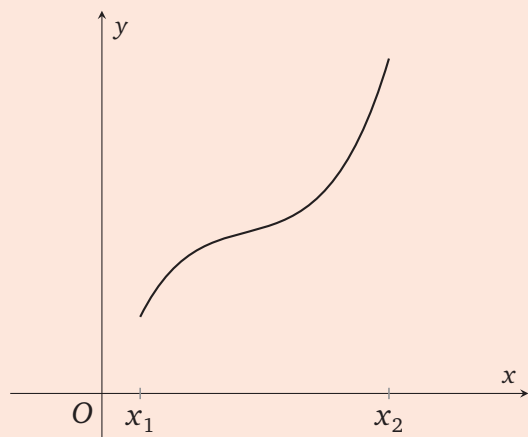
- When $f(x) = 0$, $(f + g)(x) = g(x)$.
- When $g(x) = 0$, $(f + g)(x) = f(x)$.
- If $f(x)$ and $g(x)$ are **both** positive, then $(f + g)(x) > f(x)$ **and** $(f + g)(x) > g(x)$.
- If $f(x)$ and $g(x)$ are **both** negative, then $(f + g)(x) < f(x)$ **and** $(f + g)(x) < g(x)$.
- If $f(x)$ is positive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

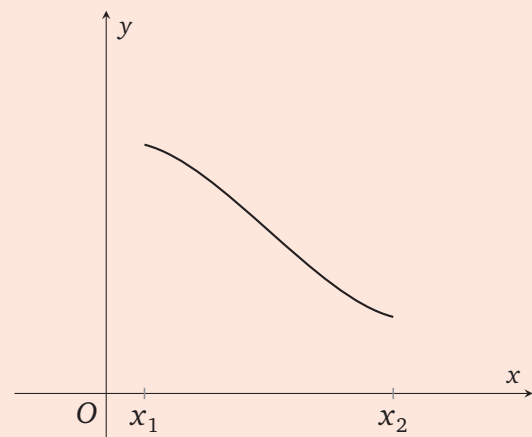
- Given that $\text{ran}(g) \subseteq \text{dom}(f)$, we can define the new function h as a **composition** of f with g .
- This is written $h = f \circ g$ (read ‘composition of g followed by f ’) and the rule for h is given by $h(x) = f(g(x))$.
 - $\text{dom}(h) = \{x \in \text{dom}(g) : g(x) \in \text{dom}(f)\}$. In other words, the domain of h is the set of x such that $x \in \text{dom}(g)$, and using those x -values, the range of $g(x)$ has to fit within the domain of f (because otherwise, f would be undefined for values that didn’t belong in its domain).
 - $\text{ran}(h) = f(\text{ran}(g) \cap \text{dom}(f))$. This is another way of saying that the range of h is the range of $f(x)$ where $x \in \text{dom}(h)$.
- Generally, $f \circ g \neq g \circ f$

Increasing and decreasing functions

- If $\forall x \in [x_1, x_2] : \underline{f(x_2) > f(x_1)} \mid x_2 > x_1$, then f is **strictly increasing** over the interval $[x_1, x_2]$.
- If $\forall x \in [x_1, x_2] : \underline{f(x_2) < f(x_1)} \mid x_2 > x_1$, then f is **strictly decreasing** over the interval $[x_1, x_2]$.
- These intervals include the values of x for which $\frac{df}{dx} = 0$, but the gradient never changes sign as such.
 - An example would be that the function of $f : [0, \infty) \rightarrow \mathbb{R}, f(x) = x^2$ 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 $[x_1, x_2]$

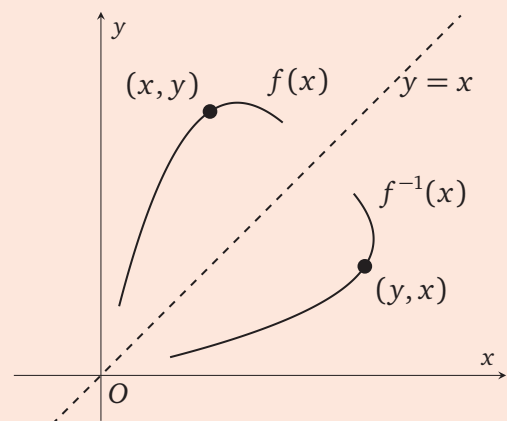
Inverse functions

- 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 $x \in \text{ran}(f)$, $y \in \text{dom}(f)$.

- Domain and range:
 - $\text{dom}(f^{-1}) = \text{ran}(f)$
 - $\text{ran}(f^{-1}) = \text{dom}(f)$
- Compositions:
 - $\forall x \in \text{dom}(f^{-1}) : (f \circ f^{-1})(x) = x$
 - $\forall x \in \text{dom}(f) : (f^{-1} \circ f)(x) = x$
- 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**.

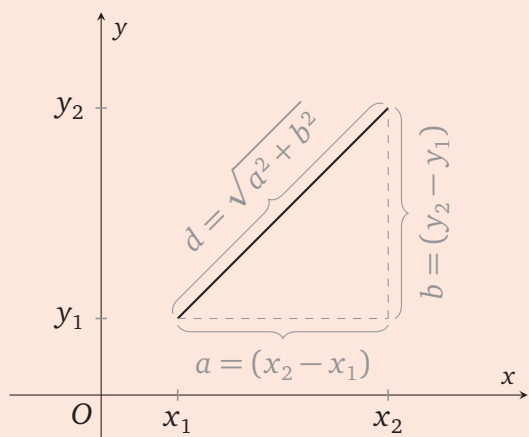
Linear coordinate geometry

- 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:

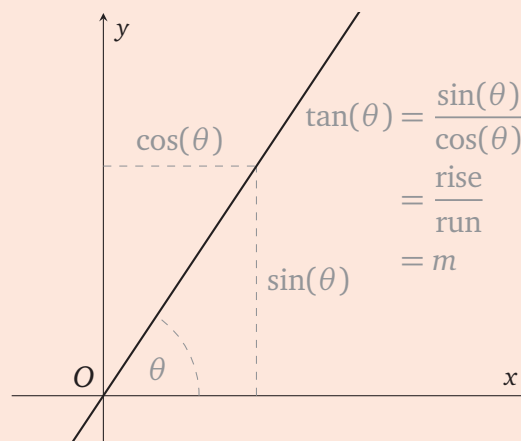
$$\frac{x}{a} + \frac{y}{b} = 1$$

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
<i>Case 1</i>	$2x + y = 5$ $x - y = 4$	Unique solution: $x = 3, y = -1$	Two lines meeting at a point
<i>Case 2</i>	$2x + y = 5$ $2x + y = 7$	No solutions	Distinct parallel lines
<i>Case 3</i>	$2x + y = 5$ $4x + 2y = 10$	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: $y = f\left(\frac{x}{a}\right)$.

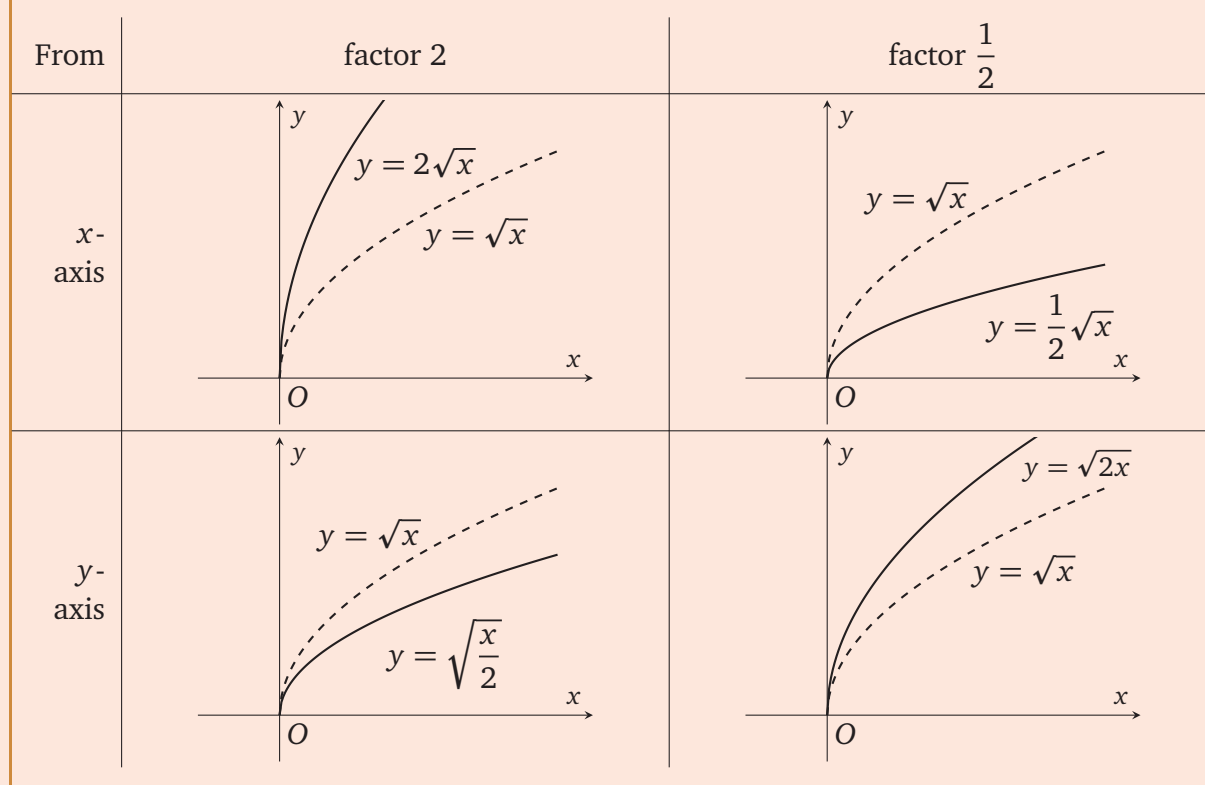


Table of transformations

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 (x', y') .
3. $(x', y') = T(x, y)$
 $\therefore (x', y') = (ax + h, by + k), \quad a \neq 0, b \neq 0$
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 $(ax^2 + bx + c)$,
 - 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 $h = -\frac{b}{2a}$ and $k = \frac{4ac - b^2}{4a}$.
 - the **axis of symmetry** is $x = h$, where $h = -\frac{b}{2a}$.
 - 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 $P\left(-\frac{\alpha}{\beta}\right)$.

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:

$$P(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_2 x^2 + a_1 x + a_0$$

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

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

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:

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

Index laws

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

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

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

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

- 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 $k < 0$.

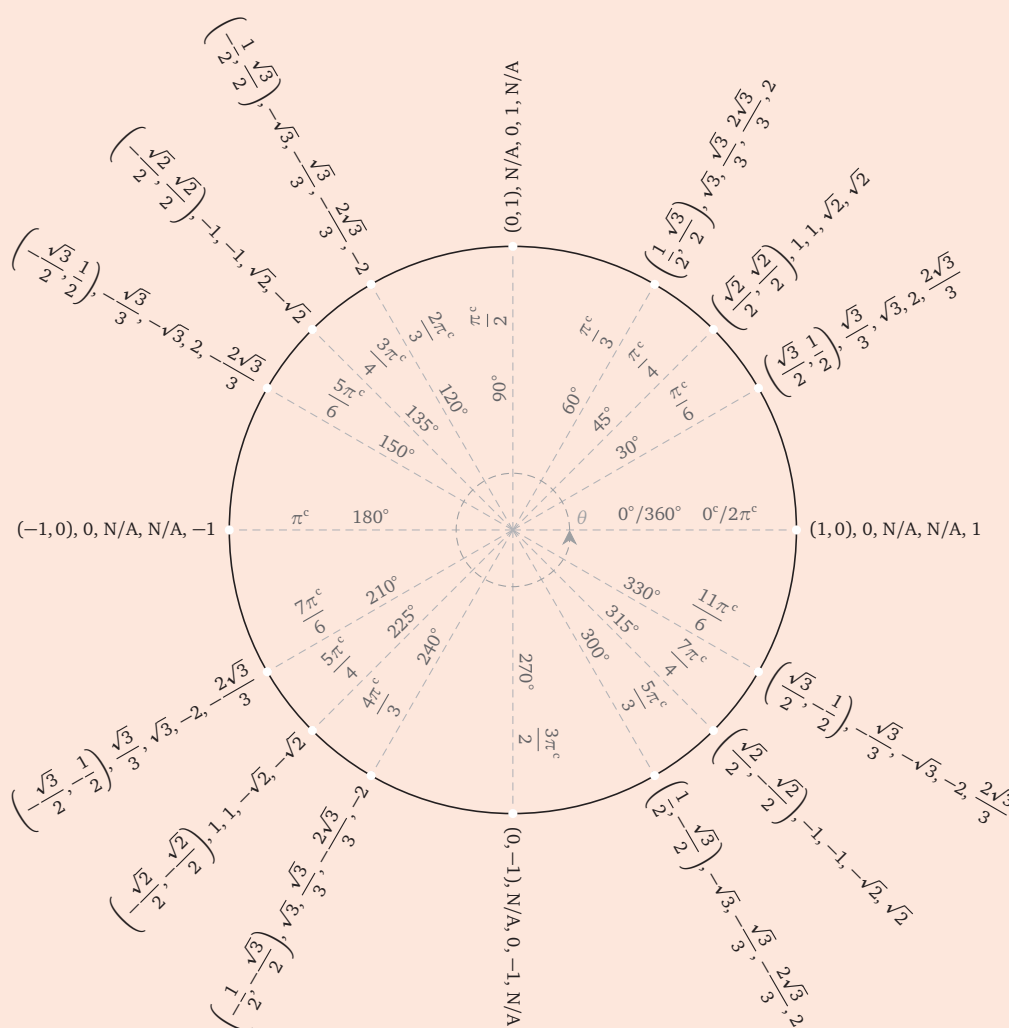
5 Circular functions

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:

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

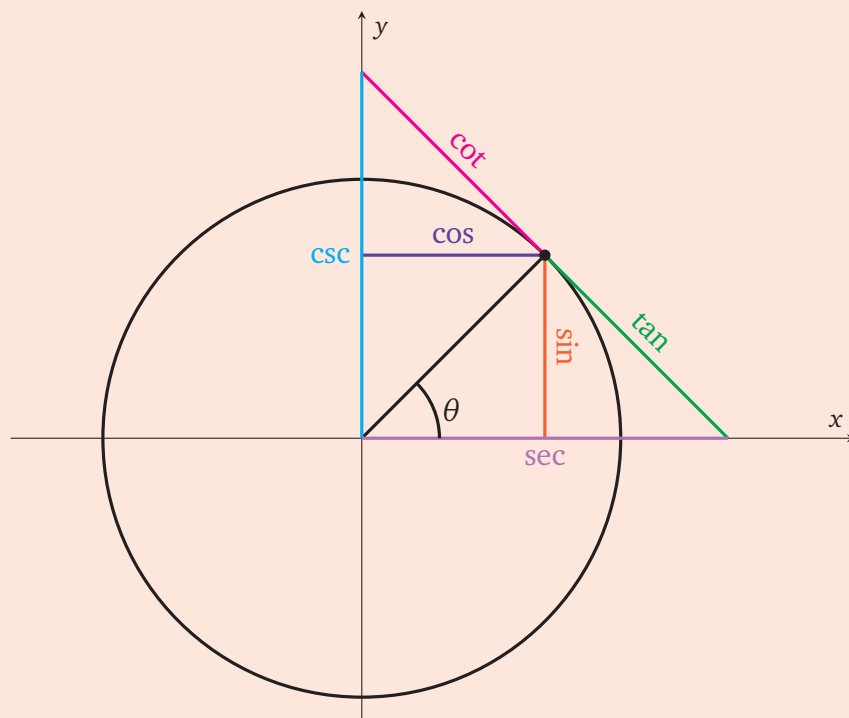
Unit circle



(x, y), z, a, b, c, where:

$x = \cos$
 $y = \sin$
 $z = \tan = \frac{\sin}{\cos}$
 $a = \cot = \frac{1}{\tan} = \frac{\cos}{\sin}$
 $b = \csc = \frac{1}{\sin}$
 $c = \sec = \frac{1}{\cos}$

Trigonometric functions as triangles



Properties of trigonometric functions

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

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

Sum/Difference identities

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

Product-to-sum identities

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

Triple-angle identities

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

General solutions

General solutions for $\sin(x)$

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

General solutions for $\cos(x)$

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

General solutions for $\tan(x)$

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

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 .

$$\text{Let } f(x) = a \sin(bx + c)$$

$$\text{Let } g(x) = k \cos(mx + n)$$

$$\text{Let } h(x) = f(x) + g(x)$$

$$= a \sin(bx + c) + k \cos(mx + n)$$

$$\text{period}(f) = \frac{2\pi}{b}$$

$$\text{period}(g) = \frac{2\pi}{m}$$

$$\therefore \text{period}(h) = \text{lcm}\left(\frac{2\pi}{b}, \frac{2\pi}{m}\right)$$

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:

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

- 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:** $f(x) = g(x) - h(x) \implies f'(x) = g'(x) - h'(x)$

Chain rule

When you have a function, $q(x)$, that can be expressed as the composition of two simpler functions, $u = g(x)$ and $q(x) = f(u)$, which are ‘chained’ together: $x \xrightarrow{g} u \xrightarrow{f} y$, you can use the **chain rule** to differentiate such a function like so:

$$q'(x) = f'(g(x)) \cdot g'(x)$$

or, in Leibniz notation, where $u = g(x)$ and $y = f(u)$:

$$\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx}$$

Limits

Algebra of limits

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

Left and right limits

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

Differentiability of a function

- A function f is said to be differentiable at $x = a$ if $\lim_{h \rightarrow 0} \frac{f(a+h) - f(a)}{h}$ 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:

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

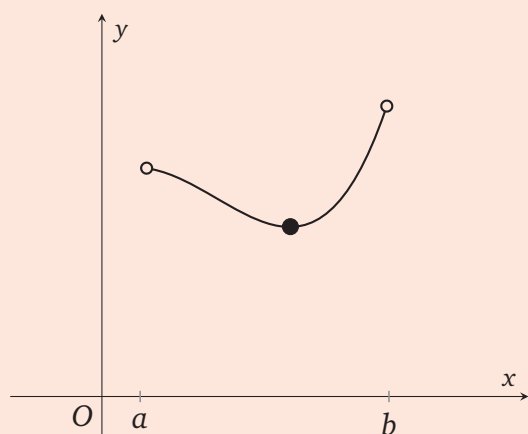
Normal line

- 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:

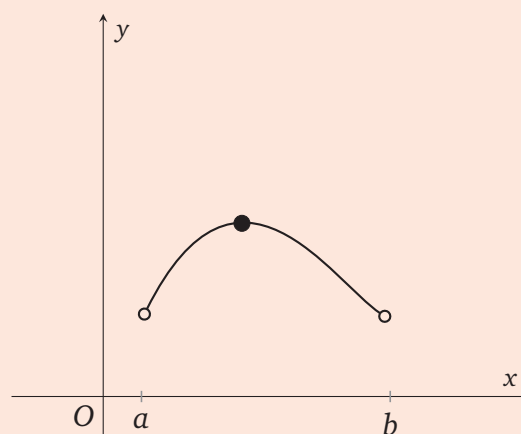
$$y - f(a) = -\frac{1}{f'(a)} \cdot (x - a)$$

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 $\forall x \in (a, b) : f''(x) > 0$, 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 $\forall x \in (a, b) : f''(x) < 0$, 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 $f''(x) = 0$ 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)

Newton's method

To approximate solutions to equations of the form $f(x) = 0$ for x , we can use Newton's method:

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}, \quad \text{where } x_{n+1} \text{ is the next guess}$$

This generates a sequence of x values where eventually it will converge to a root of the function f .

Beware, this method may not work as intended if:

- your initial guess, x_0 , is too close to another root, as it will (likely) converge to that root instead of a “desired” root instead;
- the tangent at your initial guess, x_n , is so steep and so is the next guess' tangent in such a way that it leads back to the previous guess, causing Newton's method to get stuck in an oscillation between x_n and x_{n+1} ;
- x_n , is too close to a stationary point, making the next guess very far away, pushing it further away from the root;
- $f(x_n)$ or $f'(x_n)$ does not exist for x_n ;
- $f'(x_n) = 0$, making the fractional part of the formula undefined.

7 Integration

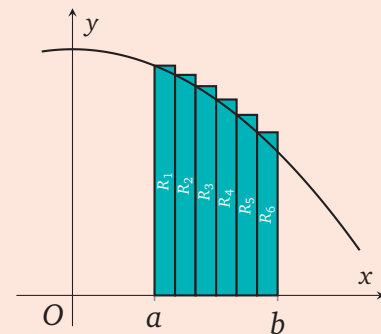
Estimating the area under a graph

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.

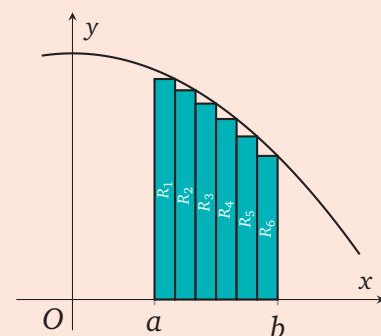


Right-endpoint estimate

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

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

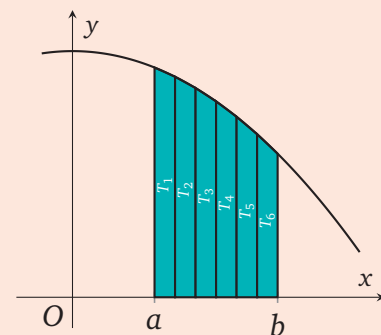
- For a function f that is...
 - strictly increasing in the domain $[a, b]$, the left-endpoint estimate \geq actual area.
 - strictly decreasing in the domain $[a, b]$, the left-endpoint estimate \leq actual area.



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

$$\frac{d}{dx}[F(x)] = f(x) \implies \int_a^b f(x) dx = [F(x)]_a^b = F(b) - F(a)$$

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

Antidifferentiation rules

Antidifferentiation results

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

Properties of the definite integral

- $\int_a^b f(x) dx = \int_a^c f(x) dx + \int_c^b f(x) dx$
- $\int_a^a f(x) dx = 0$
- $\int_a^b [k \cdot f(x)] dx = k \cdot \int_a^b f(x) dx$
- $\int_a^b [f(x) \pm g(x)] dx = \int_a^b f(x) dx \pm \int_a^b g(x) dx$
- $\int_a^b f(x) dx = - \int_b^a f(x) dx$

Signed area

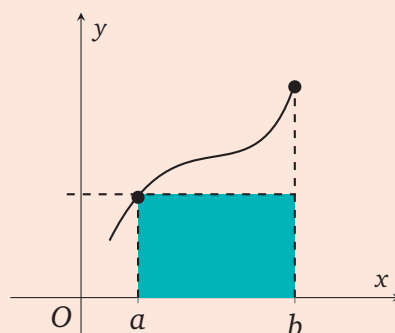
- For any continuous function f on an interval $[a, b]$, the **definite integral** $\int_a^b f(x) \, dx$ gives the **signed area** enclosed by the graph of $y = f(x)$ between $x = a$ and $x = b$.
- To get the **unsigned area**, just take the absolute value of the function like so: $\int_a^b |f(x)| \,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

Average value of a function

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

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

- 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:** $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$
- $\Pr(\emptyset) = 0$
- $\Pr(A') = 1 - \Pr(A)$, 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 rule 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)

	B	B'	
A	$\Pr(A \cap B)$	$\Pr(A \cap B')$	$\Pr(A)$
A'	$\Pr(A' \cap B)$	$\Pr(A' \cap B')$	$\Pr(A')$
	$\Pr(B)$	$\Pr(B')$	1

Conditional probability

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

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

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

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

Law of total probability

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

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

Independent events

- For events A and B with $\Pr(A) \neq 0$ and $\Pr(B) \neq 0$, the following three conditions are all **equivalent conditions** for the independence of A and B :
 - $\Pr(A | B) = \Pr(A)$
 - $\Pr(B | A) = \Pr(B)$
 - $\Pr(A \cap B) = \Pr(A) \cdot \Pr(B)$
- 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,

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

- 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

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

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,

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

- The expected value $E(X)$ 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 .
- $E[g(X)] = \sum_x [g(x) \cdot p(x)]$
- $E(aX + b) = a \cdot E(X) + b$ (for a, b constant)
 - Generally, $E[g(X)] \neq g[E(X)]$, but the linear case is an exception.
- If X and Y are two random variables, then $E(X + Y) = E(X) + E(Y)$

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:

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

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

$$\text{Var}(X) = E(X^2) - [E(X)]^2$$

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

Standard deviation

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

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

- 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 $X \sim \text{Bi}(n, p)$ and the rule is

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

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

- 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

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

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 $\int_c^d f(x) dx = 1$.

- 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:
 - $\Pr(a < X < b)$
 - $\Pr(a \leq X < b)$
 - $\Pr(a < X \leq b)$
 - $\Pr(a \leq X \leq b)$

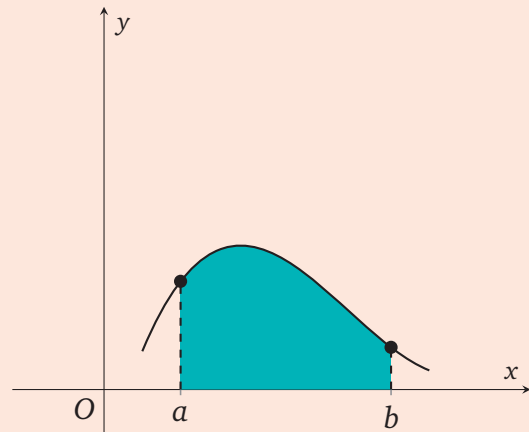
- If f has the domain $[c, d]$ and $a \in [c, d]$, then $\Pr(X < a) = \Pr(X \leq a) = \int_c^a f(x) dx$.

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

$$\lim_{k \rightarrow \infty} \int_{-k}^a f(x) \, dx$$

- If $\text{dom}(f) = [a, \infty)$, then $\int_a^{\infty} f(x) \, dx = 1$. This integral is computed as

$$\lim_{k \rightarrow \infty} \int_a^k f(x) \, dx$$

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

$$\lim_{k \rightarrow \infty} \int_{-k}^k f(x) \, dx$$

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

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

provided the integral exists.

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

$$E(X) = \int_c^d f(x) \, dx$$

- 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

$$\int_{-\infty}^p f(x) dx = q$$

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

$$\int_{-\infty}^m f(x) dx = 0.5$$

- 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).

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

where a and b are such that

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

The variance of a continuous probability distribution

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

The standard deviation of a continuous probability distribution

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

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

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

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

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

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:

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

- 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:

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

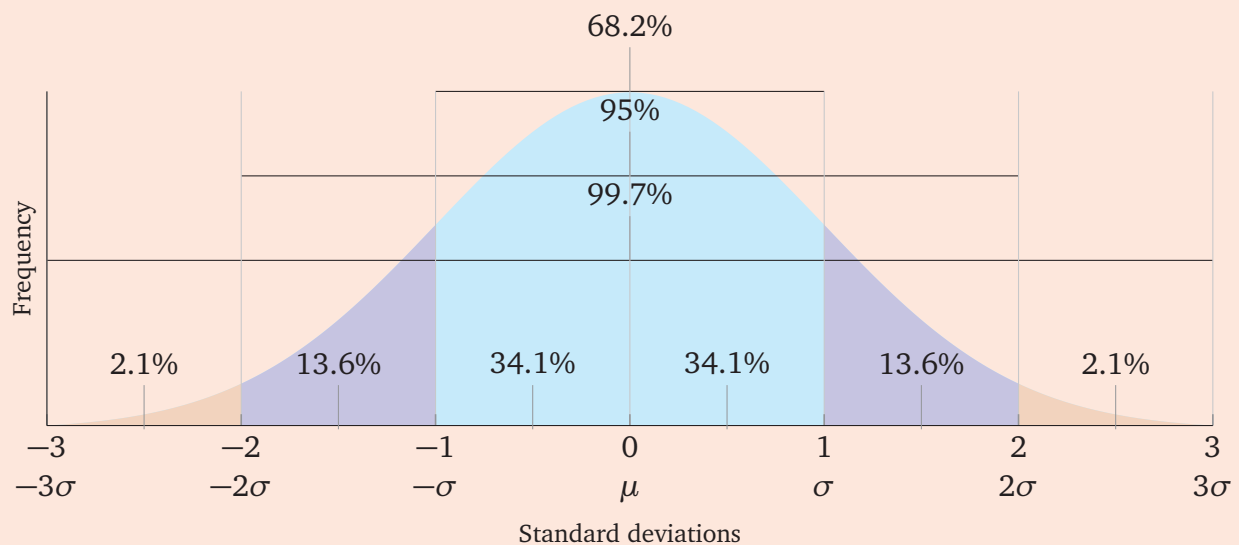
- These transformations are “area preserving”.

Symmetry properties of the standard normal distribution

- $\Pr(Z > a) = 1 - \Pr(Z \leq a)$
- $\Pr(Z < -a) = \Pr(Z > a)$
- $\Pr(-a < Z < a) = 1 - 2\Pr(Z \geq a)$
 $= 1 - 2\Pr(Z \leq -a)$

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 $x = 0, 1, \dots, n$)).
 - 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:

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

- 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 $\text{Pr}(\hat{P} > a)$ (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 $\sigma = \sqrt{\frac{p(1-p)}{n}}$.
- $\therefore Z_{\hat{p}} = \frac{\hat{p} - \mu_{\hat{p}}}{\sigma_{\hat{p}}} = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$, 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 $\Pr(-k < Z_{\hat{p}} < k) = \frac{C}{100}$.

- The **1-prop z interval** function can be used on the CAS to find the un-standardised C.I. (found in Menu \rightarrow Statistics \rightarrow Confidence Intervals \rightarrow 1-Prop z interval).

k values for confidence intervals

- 68.2% C.I.: $k = \text{invNorm}(0.841, 0, 1) = 0.99857627845453 \approx 0.9986$
- 90% C.I.: $k = \text{invNorm}(0.95, 0, 1) = 1.6448536259066 \approx 1.6449$
- 95% C.I.: $k = \text{invNorm}(0.975, 0, 1) = 1.9599639859915 \approx 1.9600$
- 99% C.I.: $k = \text{invNorm}(0.995, 0, 1) = 2.5758293030016 \approx 2.5758$
- 99.7% C.I.: $k = \text{invNorm}(0.9985, 0, 1) = 2.9677379271247 \approx 2.9677$

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

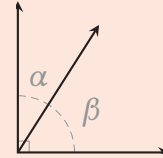
Extension

1 Angle relationships

Complementary and supplementary angles

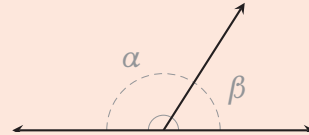
Complimentary angles

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



Supplementary angles

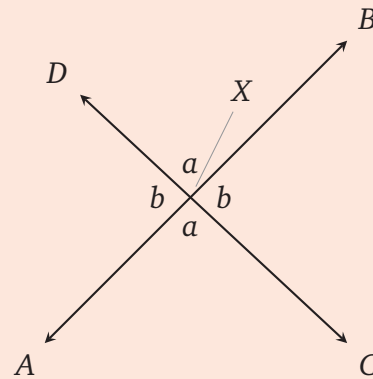
- In this case, the angles α and β are complimentary, as $\alpha + \beta = 180^\circ$.



Angles formed by intersecting lines

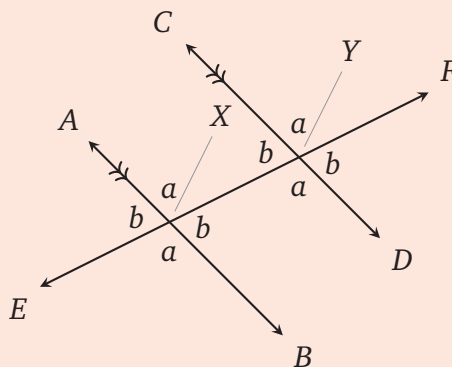
Vertically opposite angles

- 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):**
 - $\angle XYD + \angle YXB = 180^\circ$
 - $\angle AXY + \angle CYX = 180^\circ$



2 Pseudocode

What is pseudocode

We can use computers to perform calculations using algorithms (e.g., Newton's method, the bisection method, etc.). To accomplish this, we need to describe to the computer how to carry out a specific algorithm. We use **pseudocode** to do this. A computer would need to have the algorithm described using a programming language, but pseudocode provides a language-agnostic way to accomplish the same thing using an informal set of tokens and keywords.

Concepts of pseudocode

Comments

Comments are lines ignored by the program. They are meant to serve as explanatory sentences or anything else that the reader/programmer should know, but isn't necessarily an instruction for the program.

Comments

```
1 # This is a comment and it will be ignored.  
2 instructions...  
3 # It is here only for the reader/programmer.  
4 more instructions...
```

Conditional statements

Oftentimes, our programs need to make decisions influenced by other things. We can use **If** and **Else** statements.

Conditional statements

```
1 If condition Then  
2     follow these instructions  
3 Else If second condition Then  
4     if the first condition was false, follow these instructions instead  
5 Else  
6     if all else fails, try this instead
```

For loops

When we want to repeat something in a finite sequence of steps, we can use a **For** loop.

For loop

```
1 For variable_name From starting_index To n
2   execute these lines (n - starting_index) times
3 EndFor
```

In a **For** loop, $\text{starting_index} \leq n$.

If $\text{starting_index} = n$, then the **For** loop won't execute.

While loops

When we want to repeat something an infinite amount of time or when we don't know the number of iterations required to complete a task, a **While** loop can be used.

While loop

```
1 While condition Do
2   execute these instructions
3 EndWhile
```

At the end of every iteration, a **While** loop goes back to the start, and checks if *condition* is still **True**.

If *condition* is **False** to begin with, then the **While** loop will never run.

Functions

Functions provide a way of isolating code for readability and repetition. A **Function** takes zero, one, or more arguments and returns a value. A function has to be **defined**, and later **called** to be executed.

Functions

```
1 Define function_name(arg_1, arg_2, ...)
2   execute these instructions
3   Return output_value
4
5 function_name(arg_1, arg_2, ...)
```

Leave the **Function**'s arguments blank if it takes no arguments.

Finally, calling the function is required to execute the code within it.

Variables

Variables, much like in pure math, can be used to store things. Just like we can say $x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$ in math, we can say the same in pseudocode with one caveat: in this specific case, x has two values (owing to the use of the \pm symbol), but in pseudocode, a variable can only have one value at a time. So, to get both our values, we would need to make two variables (one for $+\sqrt{\Delta}$ and one for $-\sqrt{\Delta}$).

Variables

```
1 a ← 42
2 b ← a - 43
```

In pseudocode, variables *store* values, so we typically see the use of \leftarrow instead of the $=$ sign.

All concepts used in conjunction

Sum of the factorial of the first n natural numbers

```
1 Define factorial(x)
2   # Returns x!. This function will not work for negative
   ↳ integers; designed only for  $x \in \mathbb{N} \cup \{0\}$ .
3   result ← 1 # multiplication happens, so result  $\neq 0$ 
4   i ← x
5   While i > 0 Do
6     result ← result * i # new "result" = old
   ↳ "result" × "i"
7     i ← i - 1 # Decrement i with each iteration
8   EndWhile
9   Return result
10
11 Define factorial_sum(n)
12   # Returns the sum of the factorials of every natural
   ↳ number up to n (inclusive upper bound).
13   If n ≤ 0 Then
14     Return -1 # Invalid input of "n"
15
16   result ← 0 # "result" is different the variable in
   ↳ the factorial function.
17   For i From 1 To (n + 1) # "i" is different too
18     result ← result + factorial(i)
19   EndFor
20
21   Return result
22
23 factorial_sum(10) # Gets the sum of the factorials of 1,
   ↳ 2, ..., 10.
```

3 Counting methods

Pascal's triangle

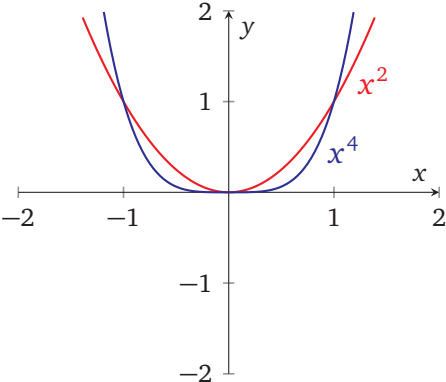
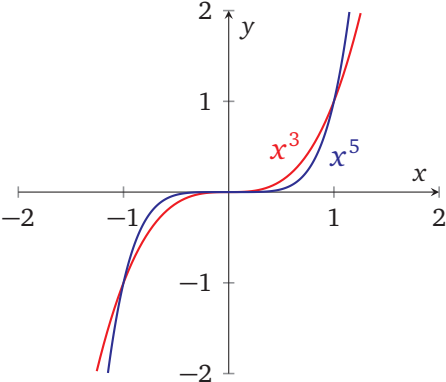
- Featured below is **pascal's triangle**, in which each row n and column k correspond to $\binom{n}{k}$.

- 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

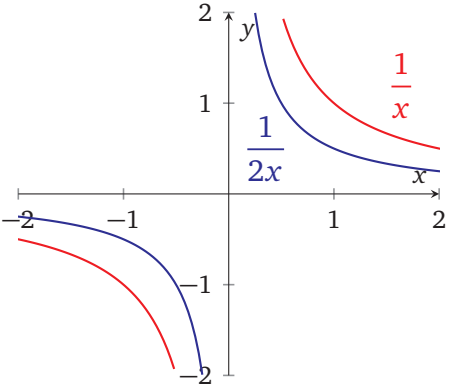
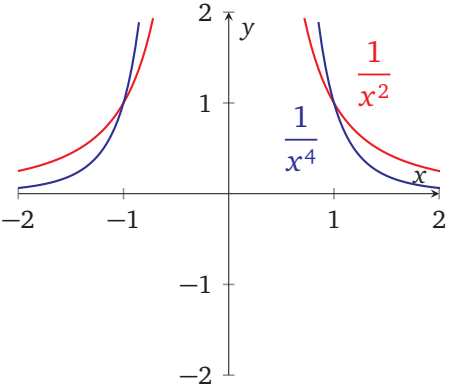
4 Base functions

Table 1: Base graphs

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

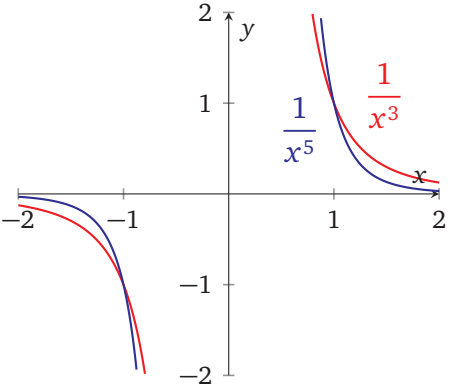
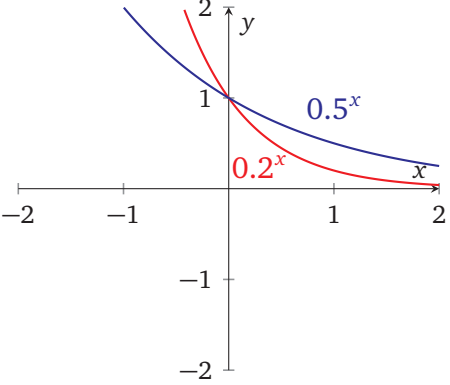
Continued on next page

Table 1: Base graphs (Continued)

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

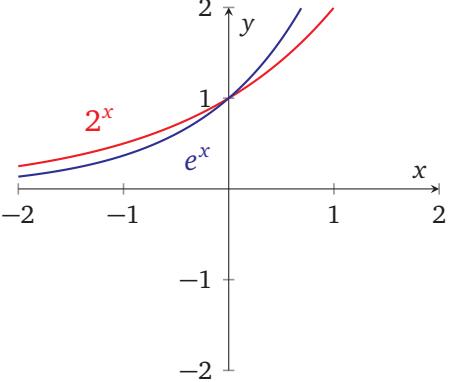
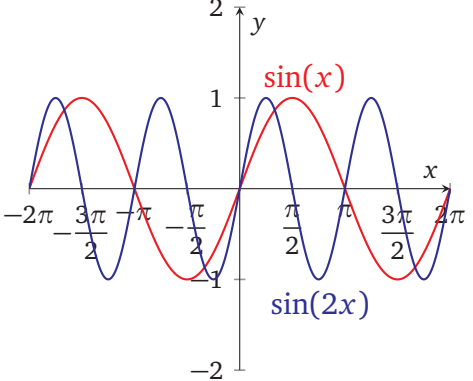
Continued on next page

Table 1: Base graphs (Continued)

Rule	Implied domain	Range	Parity	Graph	Inverse	Asymptote
$\frac{1}{x^n}, n \text{ is odd}$	$\mathbb{R} \setminus \{0\}$	$\mathbb{R} \setminus \{0\}$	Odd	 <p>The graph shows two curves on a Cartesian coordinate system. The x-axis ranges from -2 to 2, and the y-axis ranges from -2 to 2. A blue curve, labeled $\frac{1}{x^5}$, and a red curve, labeled $\frac{1}{x^3}$, both pass through the points (-1, -1) and (1, 1). They both approach the y-axis (x=0) as y goes to positive or negative infinity and the x-axis (y=0) as x goes to positive or negative infinity. The blue curve is steeper than the red curve for x > 1.</p>	$\frac{1}{\sqrt[n]{x}}$	$y = 0$ $x = 0$
$a^x, 0 < a < 1$	\mathbb{R}	\mathbb{R}^+	None	 <p>The graph shows two exponential decay curves on a Cartesian coordinate system. The x-axis ranges from -2 to 2, and the y-axis ranges from -2 to 2. A blue curve, labeled 0.5^x, and a red curve, labeled 0.2^x, both pass through the point (0, 1). They both approach the x-axis (y=0) as x goes to positive infinity. The blue curve is above the red curve for all x. The red curve approaches the y-axis (x=0) as y goes to positive infinity.</p>	$\sqrt[a]{x}$	$y = 0$

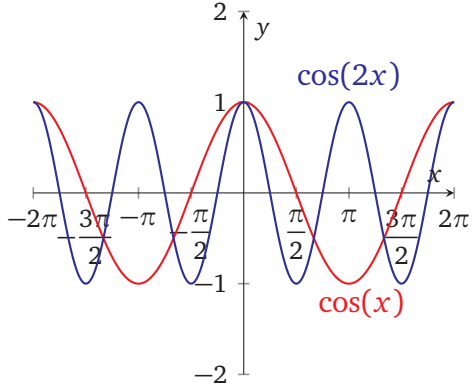
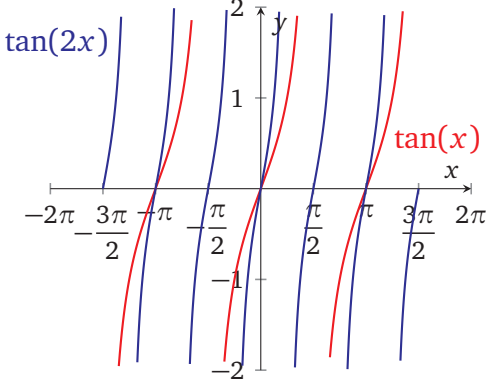
Continued on next page

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

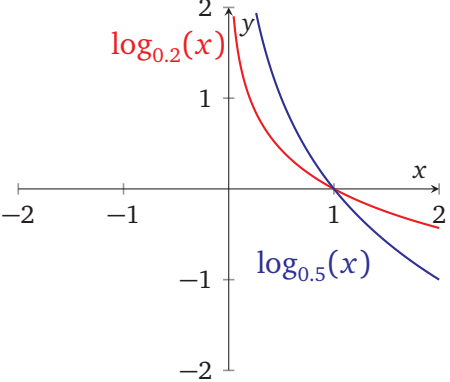
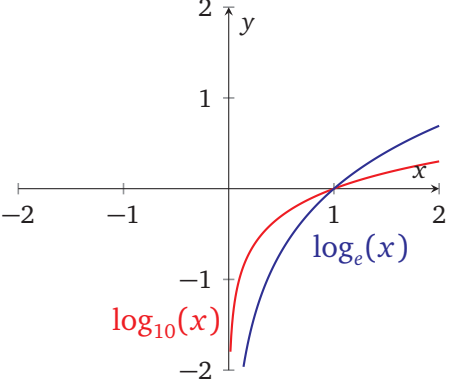
Continued on next page

Table 1: Base graphs (Continued)

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

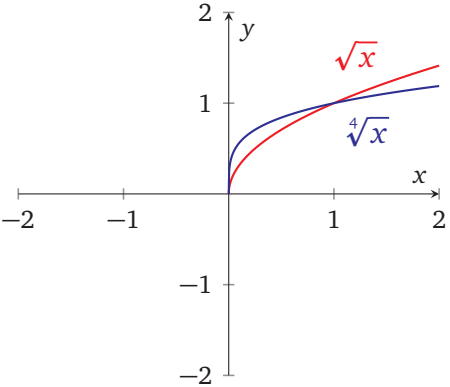
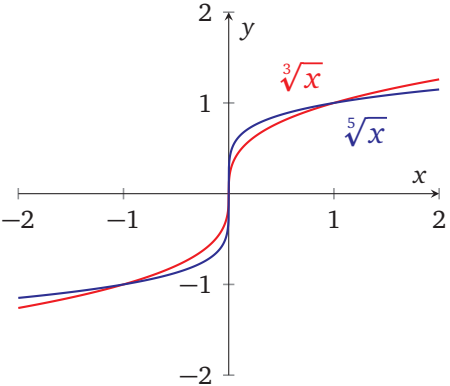
Continued on next page

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
$\sqrt[n]{x}, n \text{ is even}$	$\mathbb{R}^+ \cup \{0\}$	\mathbb{R}^+	None	 <p>The graph shows two curves in the first quadrant of a Cartesian coordinate system. The x-axis ranges from -2 to 2, and the y-axis ranges from -2 to 2. A red curve, labeled \sqrt{x}, and a blue curve, labeled $\sqrt[4]{x}$, both start at the origin (0,0) and pass through the point (1,1). The red curve is above the blue curve for $x > 1$, and below it for $0 < x < 1$.</p>	x^n	None
$\sqrt[n]{x}, n \text{ is odd}$	\mathbb{R}	\mathbb{R}	None	 <p>The graph shows two curves passing through the origin in a Cartesian coordinate system. The x-axis ranges from -2 to 2, and the y-axis ranges from -2 to 2. A red curve, labeled $\sqrt[3]{x}$, and a blue curve, labeled $\sqrt[5]{x}$, both pass through the origin (0,0) and the point (1,1). They also pass through (-1,-1). The red curve is above the blue curve for $x > 1$ and below it for $0 < x < 1$. For $x < -1$, the red curve is below the blue curve, and for $-1 < x < 0$, the red curve is above the blue curve.</p>	x^n	None