Course Summary

Topic 1 - Number and Algebra

APPROXIMATION AND ESTIMATION

A **measurement** is accurate to $\pm\frac{1}{2}$ of the smallest division on the scale.

An **approximation** is a value given to a number which is close to, but not equal to, its true value.

An **estimation** is a value which is found by judgement or prediction instead of carrying out a more accurate measurement.

If the exact value is V_E and the approximate value is V_A then:

- **absolute error** $= |V_A - V_E|$
- **percentage error** $= \frac{|V_A - V_E|}{V_E} \times 100\%$

SCIENTIFIC NOTATION (STANDARD FORM)

A number is in **scientific notation** if it is written in the form $a \times 10^k$ where $1 \leqslant a < 10$ and $k \in \mathbb{Z}$.

SEQUENCES AND SERIES

A **number sequence** is a set of numbers defined by a rule. Often, the rule is a formula for the **general term** or ***n*th term** of the sequence.

A sequence which continues forever is called an **infinite sequence**. A sequence which terminates is called a **finite sequence**.

Arithmetic sequences

In an **arithmetic sequence**, each term differs from the previous one by the same fixed number.

$u_{n+1} - u_n = d$ for all $n \in \mathbb{Z}^+$, where d is a constant called the **common difference**.

For an arithmetic sequence with first term u_1 and common difference d, the nth term is $u_n = u_1 + (n-1)d$.

Geometric sequences

In a **geometric sequence**, each term is obtained from the previous one by multiplying by the same non-zero constant, called the **common ratio** r.

$u_{n+1} = ru_n$, so we can find $r = \frac{u_{n+1}}{u_n}$ for all $n \in \mathbb{Z}^+$.

For a geometric sequence with first term u_1 and common ratio r, the nth term is $u_n = u_1 r^{n-1}$.

Series

A **series** is the sum of the terms of a sequence.

For a finite sequence with n terms, the corresponding series is $S_n = u_1 + u_2 + ... + u_n$.

For an infinite sequence, the corresponding series $u_1 + u_2 + ... + u_n + ...$ can only be calculated in some cases.

Using **sigma notation** or **summation notation** we write $u_1 + u_2 + u_3 + ... + u_n$ as $\sum_{k=1}^{n} u_k$.

For a **finite arithmetic series**, $S_n = \frac{n}{2}(u_1 + u_n)$ or $S_n = \frac{n}{2}(2u_1 + (n-1)d)$.

For a **finite geometric series** with $r \neq 1$, $S_n = \frac{u_1(r^n - 1)}{r - 1}$.

The sum of an **infinite geometric series** is $S = \frac{u_1}{1-r}$ provided $|r| < 1$.

If $|r| > 1$ the series is **divergent**.

Compound interest

The value of a compound interest investment after n time periods is

$$u_n = u_0(1+i)^n$$

where u_0 is the initial value of the investment
and i is the interest rate per compounding period.

To find the **real value** of the investment, we divide by the inflation multiplier each year.

You should be able to use the TVM solver on your calculator to solve problems involving compound interest investments and loans.

Depreciation

Depreciation is the loss in value of an item over time.

The value of an item after n years is $u_n = u_0(1-d)^n$
where u_0 is the initial value of the item
and d is the rate of depreciation per year.

POLYNOMIAL EQUATIONS

The highest power of x in a polynomial equation is called its **degree**.

If a polynomial equation has degree n then it may have up to n real solutions.

You should be able to use your graphics calculator to solve polynomial equations.

EXPONENTIALS AND LOGARITHMS

Laws of exponents	
$a^m \times a^n = a^{m+n}$	$a^0 = 1,\ a \neq 0$
$\dfrac{a^m}{a^n} = a^{m-n}$	$a^{-n} = \dfrac{1}{a^n}$ and $\dfrac{1}{a^{-n}} = a^n$
$(a^m)^n = a^{mn}$	$a^{\frac{1}{n}} = \sqrt[n]{a}$
$(ab)^n = a^n b^n$	$a^{\frac{m}{n}} = \sqrt[n]{a^m} = (\sqrt[n]{a})^m$
$\left(\dfrac{a}{b}\right)^n = \dfrac{a^n}{b^n}$	

The **logarithm in base 10** of a positive number is the power that 10 must be raised to in order to obtain that number.

If $10^x = b$ for $b > 0$, we say that x is the logarithm of b in base 10, and write $x = \log b$.

$\log 10^x = x$ and $10^{\log x} = x$ for any $x > 0$.

The **natural logarithm** is the logarithm in base e. The natural logarithm of x is written as $\ln x$ or $\log_e x$.

$\ln e^x = x$ and $e^{\ln x} = x$ for all $x > 0$.

Laws of logarithms	
Base 10	**Base** e
$\log xy = \log x + \log y$	$\ln xy = \ln x + \ln y$
$\log\left(\dfrac{x}{y}\right) = \log x - \log y$	$\ln\left(\dfrac{x}{y}\right) = \ln x - \ln y$
$\log(x^m) = m \log x$	$\ln(x^m) = m \ln x$
$\log 1 = 0$	$\ln 1 = 0$

COMPLEX NUMBERS

Any number of the form $a + bi$ where a and b are real and $i = \sqrt{-1}$ is called a **complex number**.

Complex numbers allow us to obtain solutions to quadratic equations of the form $ax^2 + bx + c = 0$ with $b^2 - 4ac < 0$.

If $z = a + bi$ where a and b are real then:

- a is the **real part** of z, written $\mathcal{Re}(z)$
- b is the **imaginary part** of z, written $\mathcal{Im}(z)$.

You should be able to calculate sums, differences, products, quotients, and integer powers of complex numbers.

Two complex numbers are equal if their real parts are equal *and* their imaginary parts are equal

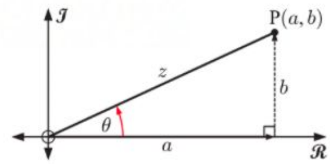
$$a + bi = c + di \Leftrightarrow a = c \text{ and } b = d.$$

The **complex conjugate** of $z = a + bi$ is $z^* = a - bi$.

The complex plane (Argand plane)

On the complex plane, the x-axis is called the **real axis** and the y-axis is called the **imaginary axis**.

The complex number $z = a + bi$ is represented by the vector $\overrightarrow{OP} = \begin{pmatrix} a \\ b \end{pmatrix}$.



Modulus and argument

- The **modulus** of the complex number $z = a + bi$ is the length of the vector $\begin{pmatrix} a \\ b \end{pmatrix}$, which is $|z| = \sqrt{a^2 + b^2}$.
- If z is represented by P(a, b) on the Cartesian plane, the **argument** of z is the angle θ, where $-\pi < \theta \leqslant \pi$ is measured anticlockwise between the positive real axis and $\overrightarrow{OP}$.

Properties of modulus and argument:

- $|wz| = |w||z|$ and $\arg(wz) = \arg w + \arg z$
- $\left|\frac{w}{z}\right| = \frac{|w|}{|z|}$ and $\arg\left(\frac{w}{z}\right) = \arg w - \arg z$ provided $z \neq 0$
- $|z^*| = |z|$ and $\arg(z^*) = -\arg z$
- $zz^* = |z|^2$

For points P_1 and P_2 on the complex plane defined by $z_1 \equiv \overrightarrow{OP_1}$ and $z_2 \equiv \overrightarrow{OP_2}$, the distance between P_1 and P_2 is $|z_1 - z_2|$.

Polar and exponential form

The complex number z can be written in **polar form** $z = |z| \operatorname{cis} \theta = |z|(\cos\theta + i \sin\theta)$ or **exponential form** $z = |z|e^{i\theta}$, where θ is the argument of z.

$\operatorname{cis}\theta \times \operatorname{cis}\phi = \operatorname{cis}(\theta + \phi)$

$\frac{\operatorname{cis}\theta}{\operatorname{cis}\phi} = \operatorname{cis}(\theta - \phi)$

$\operatorname{cis}(\theta + k2\pi) = \operatorname{cis}\theta$ for all $k \in \mathbb{Z}$.

If a complex number is multiplied by $r \operatorname{cis}\theta$ then its modulus is *multiplied* by r, and its argument is *increased* by θ.

You should be able to use the exponential form to add trigonometric functions with the same frequency but different phase, such as $\sin(3t - \frac{\pi}{3}) + \sin(3t + \frac{\pi}{6})$.

MATRICES

A matrix with m rows and n columns has order $m \times n$.

Two matrices are **equal** if they have the **same order** *and* the elements in corresponding positions are equal.

Operations with matrices

To **add** two matrices, they must be of the **same order**, and we **add corresponding elements**.

To **subtract** matrices, they must be of the **same order**, and we **subtract corresponding elements**.

To **multiply** a matrix by a scalar k, we multiply each element of the matrix by k.

The **product** of an $m \times n$ matrix **A** with an $n \times p$ matrix **B**, is the $m \times p$ matrix **AB** in which the element in the rth row and cth column is the product of the rth row of **A** (as a row matrix) and the cth column of **B** (as a column matrix).

In general, $\mathbf{AB} \neq \mathbf{BA}$.

Zero and identity matrices

A **zero matrix** is a matrix in which all the elements are zero.

If **A** is a matrix of any order and **O** is the corresponding zero matrix, then $\mathbf{A} + \mathbf{O} = \mathbf{O} + \mathbf{A} = \mathbf{A}$.

An **identity matrix** is a square matrix with 1s in the leading diagonal and 0s everywhere else.

If **A** is a square matrix and **I** is the corresponding identity matrix then $\mathbf{AI} = \mathbf{IA} = \mathbf{A}$.

Inverse of a matrix

The **multiplicative inverse** of $\mathbf{A}$, denoted $\mathbf{A}^{-1}$, satisfies $\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$.

For the 2×2 matrix $\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$:

- The value $ad - bc$ is called the **determinant** of matrix $\mathbf{A}$, denoted $\det \mathbf{A}$ or $|\mathbf{A}|$.
- If $\det \mathbf{A} \neq 0$, then $\mathbf{A}$ is **invertible** or **non-singular**, and $\mathbf{A}^{-1} = \frac{1}{\det \mathbf{A}} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$.
- If $\det \mathbf{A} = 0$, then $\mathbf{A}$ is **singular**, and $\mathbf{A}^{-1}$ does not exist.

You should be able to use technology to find determinants and inverses of larger matrices.

Simultaneous linear equations

A system of linear equations can be written in the form $\mathbf{A}\mathbf{x} = \mathbf{b}$ where $\mathbf{A}$ is the matrix of coefficients, $\mathbf{x}$ is the vector of unknowns, and $\mathbf{b}$ is a vector of constants.

Provided the inverse matrix $\mathbf{A}^{-1}$ exists, we can solve the matrix equation $\mathbf{A}\mathbf{x} = \mathbf{b}$ for $\mathbf{x}$ by finding $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$.

EIGENVALUES AND EIGENVECTORS

Suppose $\mathbf{A}$ is a square matrix.

If $\mathbf{x}$ is a non-zero vector and λ is a constant such that $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$, then λ is an **eigenvalue** of $\mathbf{A}$ and $\mathbf{x}$ is its corresponding **eigenvector**.

The **characteristic polynomial** of $\mathbf{A}$ is $p(\lambda) = \det(\lambda\mathbf{I} - \mathbf{A})$.

The eigenvalues of $\mathbf{A}$ are the solutions to $p(\lambda) = 0$.

For a given eigenvalue λ, the corresponding eigenvectors are the solutions $\mathbf{x}$ to $(\lambda\mathbf{I} - \mathbf{A})\mathbf{x} = \mathbf{0}$.

Matrix diagonalisation

A non-zero square matrix is said to be **diagonal** if the elements *not* on its leading diagonal are zero.

A square matrix $\mathbf{A}$ is **diagonalisable** if there exists a matrix $\mathbf{P}$ such that $\mathbf{P}^{-1}\mathbf{A}\mathbf{P}$ is a diagonal matrix. We say that $\mathbf{P}$ diagonalises $\mathbf{A}$.

If $\mathbf{A}$ is a 2×2 matrix with *distinct* eigenvalues λ_1, λ_2 and corresponding eigenvectors $\mathbf{x}_1$, $\mathbf{x}_2$, then $\mathbf{P} = (\mathbf{x}_1 \mid \mathbf{x}_2)$ diagonalises $\mathbf{A}$, and $\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$.

In this case, we can calculate powers of $\mathbf{A}$ using $\mathbf{A}^n = \mathbf{P} \begin{pmatrix} \lambda_1^n & 0 \\ 0 & \lambda_2^n \end{pmatrix} \mathbf{P}^{-1}$.

Topic 2 - Functions

PROPERTIES OF LINES

The **gradient** of the line passing through $A(x_1, y_1)$ and $B(x_2, y_2)$ is $m = \frac{y\text{-step}}{x\text{-step}} = \frac{y_2 - y_1}{x_2 - x_1}$.

The gradient of any horizontal line is zero. The gradient of any vertical line is undefined.

The **y-intercept** of a line is the value of y where the line cuts the y-axis.

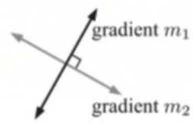
The **x-intercept** of a line is the value of x where the line cuts the x-axis.

PARALLEL AND PERPENDICULAR LINES

The gradients of parallel lines are equal.

The gradients of perpendicular lines are negative reciprocals.

$m_1 = -\frac{1}{m_2}$



EQUATION OF A LINE

The equation of a line can be presented in:

- **gradient-intercept form** $y = mx + c$ where m is the gradient and c is the y-intercept.
- **general form** $ax + by = d$
- **point-gradient form** $y - y_1 = m(x - x_1)$

You should be able to find the equation of a line given:

- its gradient and the coordinates of any point on the line
- the coordinates of two distinct points on the line.

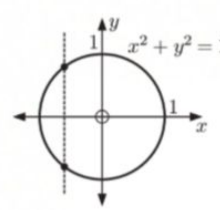
FUNCTIONS $f : x \mapsto f(x)$ OR $y = f(x)$

A **relation** between variables x and y is any set of points in the (x, y) plane.

A **function** is a relation in which no two different ordered pairs have the same x-coordinate or first component. For each value of x there is at most one value of y or $f(x)$. We sometimes refer to y or $f(x)$ as the **image** of x.

We test for functions using the **vertical line test**. A graph is a function if no vertical line intersects the graph more than once.

For example, the graph of the circle $x^2 + y^2 = 1$ shows that this relation is not a function.



The **domain** of a relation is the set of values that x can take.

To find the domain of a function, remember that we cannot:

- divide by zero
- take the square root of a negative number
- take the logarithm of a non-positive number.

The **range** of a relation is the set of values that y or $f(x)$ can take.

Given $f : x \mapsto f(x)$ and $g : x \mapsto g(x)$, the **composite function** of f and g is $f \circ g : x \mapsto f(g(x))$.

In general, $f(g(x)) \neq g(f(x))$, so $f \circ g \neq g \circ f$.

The **identity function** is $f(x) = x$.

INVERSE FUNCTIONS

A function is:

- **one-to-one** if there is only one value of x for each value of y
- **many-to-one** if there is more than one value of x with the same value of y.

The function $y = f(x)$ has an **inverse function** $y = f^{-1}(x)$ if and only if it is one-to-one.

Many-to-one functions do not have an inverse function. However, we can often restrict the domain of a many-to-one function to make it a one-to-one function. This restricted function will have an inverse function.

If $y = f(x)$ has an inverse function $y = f^{-1}(x)$, then the inverse function:

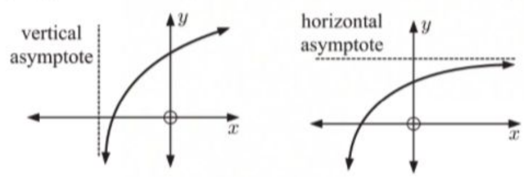
- must satisfy the vertical line test
- is a reflection of $y = f(x)$ in the line $y = x$
- satisfies $(f \circ f^{-1})(x) = (f^{-1} \circ f)(x) = x$
- has range equal to the domain of $f(x)$
- has domain equal to the range of $f(x)$.

GRAPHS OF FUNCTIONS

The **x-intercepts** of a function are the values of x for which $y = 0$. They are the **zeros** of the function.

The **y-intercept** of a function is the value of y when $x = 0$.

An **asymptote** is a line that the graph *approaches* or begins to look like as it tends to infinity in a particular direction.



To find vertical asymptotes, look for values of x for which the function is undefined:

- If $y = \frac{f(x)}{g(x)}$, find where $g(x) = 0$.
- If $y = \log_a(f(x))$, find where $f(x) = 0$.

To find horizontal asymptotes, consider the behaviour as $x \to \pm\infty$.

You should be able to use technology to:

- graph a function
- find turning points
- find the domain and range
- find asymptotes
- find axes intercepts
- find where functions meet.

MODELLING

Mathematical models are developed using a **modelling cycle**:

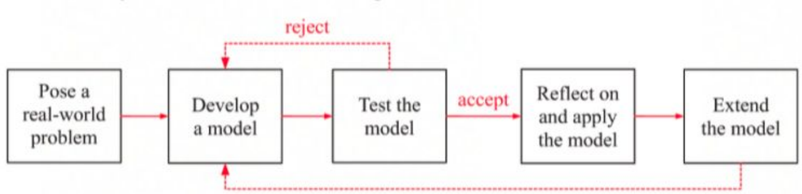
Step 1: **Pose** a real-world problem. Make **assumptions** which simplify the problem without missing key features.

Step 2: **Develop** a model which represents the problem with mathematics. This may involve a formula or an equation.

Step 3: **Test** the model by comparing its predictions with known data. If the model is unsatisfactory, return to *Step 2*.

Step 4: **Reflect** on your model and **apply** it to your original problem, interpreting the solution in its real-world context.

Step 5: If appropriate, **extend** your model to make it more general or accurate as needed.



You should be able to solve systems of equations using technology to find unknown parameters in models.

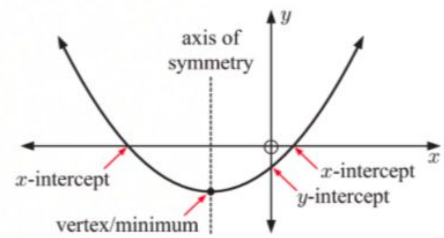
QUADRATICS

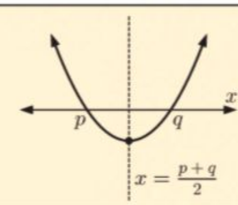
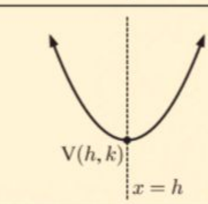
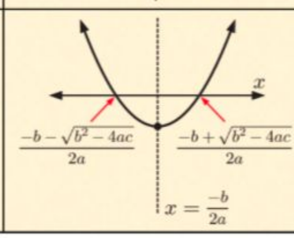
Quadratic functions

A **quadratic function** has the form $y = ax^2 + bx + c,\ a \neq 0$.

The graph is a parabola with the following properties:

- It is *concave up* if $a > 0$  and *concave down* if $a < 0$. 
- Its axis of symmetry is $x = \frac{-b}{2a}$.
- Its vertex has x-coordinate $\frac{-b}{2a}$. The y-coordinate of its vertex is found by substituting $x = \frac{-b}{2a}$ into the function.
  - If $a > 0$ the vertex is a minimum turning point.
  - If $a < 0$ the vertex is a maximum turning point.



$\boldsymbol{y = a(x-p)(x-q)}$ x-intercepts p, q axis of symmetry $x = \frac{p+q}{2}$	
$\boldsymbol{y = a(x-h)^2 + k}$ vertex (h, k) axis of symmetry $x = h$	
$\boldsymbol{y = ax^2 + bx + c}$ axis of symmetry $x = \frac{-b}{2a}$ x-intercepts $\frac{-b \pm \sqrt{b^2-4ac}}{2a}$ where $b^2 - 4ac \geqslant 0$	

Quadratic equations

A quadratic equation of the form $ax^2 + bx + c = 0,\ a \neq 0$, can be solved by:

- factorisation
- completing the square
- the quadratic formula $x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$.

Quadratic inequalities

A **quadratic inequality** can be written in either the form $ax^2 + bx + c \geqslant 0$ or $ax^2 + bx + c > 0$ where $a \neq 0$.

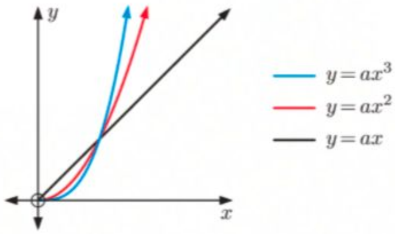
You should be able to use sign diagrams to solve quadratic inequalities.

VARIATION MODELS

Variation models have the form $y = ax^n,\ n \in \mathbb{Z},\ n \neq 0$.

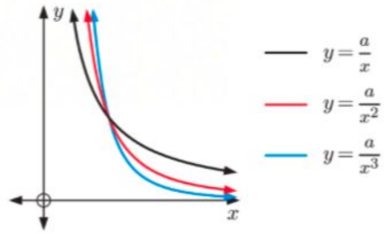
- If $n > 0$ we have **direct variation**.

  The graph passes through the origin $(0, 0)$.



- If $n < 0$ we have **inverse variation**.

  The graph is asymptotic to both the x and y axes.



You should be able to:

- use a point which lies on the graph of a variation model to find the exact equation of the variation model
- use technology to find the variation model which best fits a set of data.

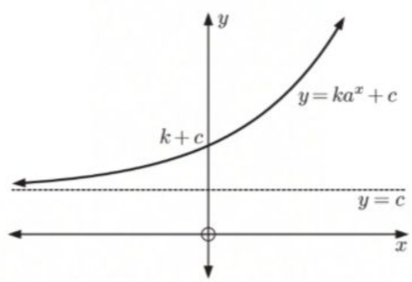
EXPONENTIAL FUNCTIONS

In this course you need to deal with exponential functions of the form:

- $y = ka^x + c$
- $y = ka^{-x} + c$

In each case:

- a and k control the steepness of the curve
- $y = c$ is the equation of the **horizontal asymptote**.

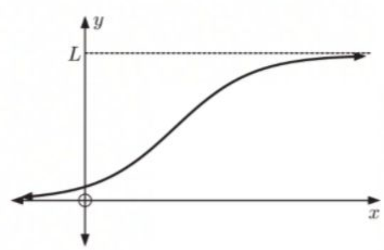


You will also need to deal with natural exponential functions of the form $y = ke^{rx} + c$.

Exponential functions are commonly used to model **growth** and **decay** problems.

Exponential equations are equations where the variable appears in an index or exponent. You should be able to solve exponential equations using technology.

The **logistic model** has the form $\boldsymbol{y = \dfrac{L}{1 + Ce^{-kx}}}$ where $L > 0$ is the limiting value, and C and k are positive constants.

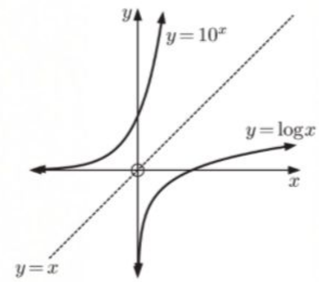


LOGARITHMIC FUNCTIONS

The **logarithmic function** $y = \log x,\ x > 0$ is the inverse function of $y = 10^x$.

The graph of $y = \log x$ has the vertical asymptote $x = 0$.

The natural logarithmic function $y = \ln x,\ x > 0$ is the inverse function of $y = e^x$.



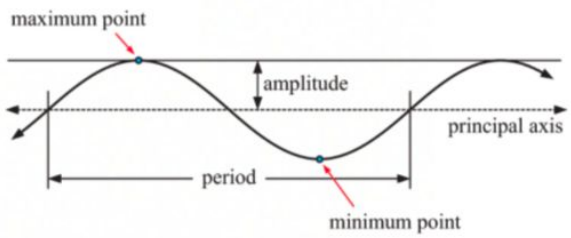
TRANSFORMATIONS OF FUNCTIONS

- $y = f(x) + b$ **translates** $y = f(x)$ vertically b units.
- $y = f(x - a)$ **translates** $y = f(x)$ horizontally a units.
- $y = f(x - a) + b$ **translates** $y = f(x)$ by the vector $\begin{pmatrix} a \\ b \end{pmatrix}$.
- $y = pf(x),\ p > 0$ is a **vertical stretch** of $y = f(x)$ with scale factor p.
- $y = f(qx),\ q > 0$ is a **horizontal stretch** of $y = f(x)$ with scale factor $\frac{1}{q}$.
- $y = -f(x)$ is a **reflection** of $y = f(x)$ in the x-axis.
- $y = f(-x)$ is a **reflection** of $y = f(x)$ in the y-axis.
- If $f^{-1}(x)$ exists, $y = f^{-1}(x)$ is a **reflection** of $y = f(x)$ in the line $y = x$.

PERIODIC FUNCTIONS

A **periodic function** is one which repeats itself over and over in a horizontal direction.

For example, a **wave** is a periodic function which oscillates about a horizontal line called the **principal axis**.



The **period** of a periodic function is the length of one cycle.

The **amplitude** is the distance between a maximum or minimum point and the principal axis.

THE SINE FUNCTION

If we begin with $y = \sin x$, we can perform transformations to produce the **general sine function** $f(x) = a\sin(b(x - c)) + d$, where $b > 0$.

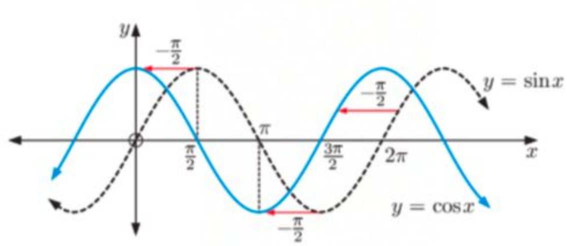
We have a vertical stretch with scale factor $|a|$ and a horizontal stretch with scale factor $\frac{1}{b}$, a reflection in the x-axis if $a < 0$, and a translation through $\begin{pmatrix} c \\ d \end{pmatrix}$.

The general sine function has the following properties:

- the **amplitude** is $|a|$
- the **principal axis** is $y = d$
- the **period** is $\frac{2\pi}{b}$.

THE COSINE FUNCTION

Since $\cos x = \sin\left(x + \frac{\pi}{2}\right)$, the graph of $y = \cos x$ is a horizontal translation of $y = \sin x$, $\frac{\pi}{2}$ units to the left.



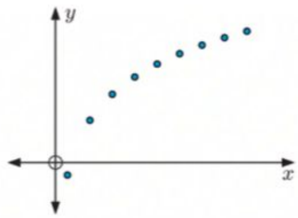
The properties of the **general cosine function** $y = a\cos(b(x - c)) + d$ are the same as those of the general sine function.

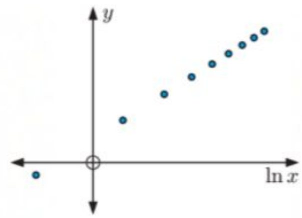
NON-LINEAR MODELLING

Logarithmic models

A **logarithmic model** has the form $\boldsymbol{y = a + b \ln x}$.

If the variables x and y are connected by a logarithmic model, the graph of y against $\ln x$ is linear.



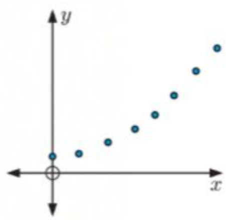


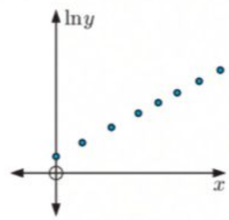
We can use linear regression to find an equation connecting y and $\ln x$.

Exponential models

An **exponential model** has the form $\boldsymbol{y = kb^x}$ where $k > 0$ and $b > 0,\ b \neq 1$ are constants.

If the variables x and y are connected by an exponential model, the graph of $\ln y$ against x is linear.



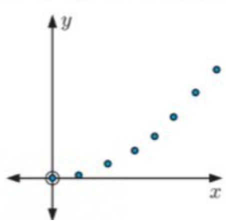


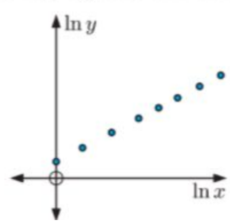
We can use linear regression to find a *linear* equation connecting $\ln y$ and x, which we can rearrange into an *exponential* equation connecting y and x.

Power models

A **power model** has the form $\boldsymbol{y = ax^n}$ where $a > 0$ and $n \neq 0$ are constants.

If the variables x and y are connected by a power model, the graph of $\ln y$ against $\ln x$ is linear.





We can use linear regression to find a *linear* equation connecting $\ln y$ and $\ln x$, which we can rearrange into a *power* equation connecting y and x.

Since $y = ax^n$, all power models pass through the origin. However, since we cannot take the logarithm of 0, we will need to remove data such as $(0, 0)$ before starting our analysis.

Problem solving

To determine whether a linear, logarithmic, exponential, or power model is most appropriate to connect variables x and y:

Step 1: Draw scatter diagrams of y against x, y against $\ln x$, $\ln y$ against x, and $\ln y$ against $\ln x$.

Step 2: Find the regression line for each scatter diagram in which the variables appear linearly related. If more than one model appears appropriate, choose the one with the highest r^2 value.

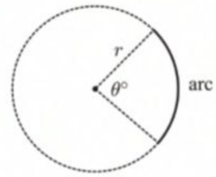
Step 3: Write the equation connecting the variables x and y without logarithms.

Topic 3 - Geometry and Trig

ARCS AND SECTORS

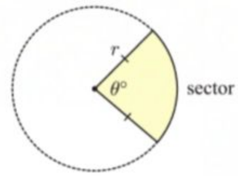
An **arc** is a part of a circle which joins any two different points.

Arc length $= \frac{\theta}{360} \times 2\pi r$



A **sector** is the region between two radii of a circle and the arc between them.

Area $= \frac{\theta}{360} \times \pi r^2$



GEOMETRY OF 3-DIMENSIONAL FIGURES

The **surface area** of a three-dimensional figure with plane faces is the sum of the areas of the faces.

The **volume** of a solid is the amount of space it occupies.

The **capacity** of a container is the quantity of fluid it is capable of holding. You should understand how the units of volume and capacity are related.

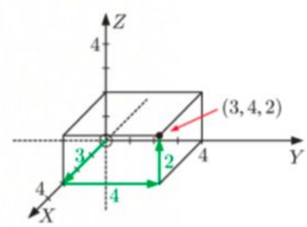
You should be able to calculate the surface area and volume of 3-dimensional figures, including solids of uniform cross-section, pyramids, spheres, and cones.

3-DIMENSIONAL COORDINATE GEOMETRY

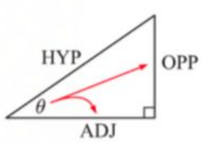
In 3-dimensional coordinate geometry, we specify an origin O, and three mutually perpendicular axes called the X-axis, the Y-axis, and the Z-axis.

For points $A(x_1, y_1, z_1)$ and $B(x_2, y_2, z_2)$:

- the **distance** $AB = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}$
- the **midpoint** of [AB] is $\left(\frac{x_1 + x_2}{2}, \frac{y_1 + y_2}{2}, \frac{z_1 + z_2}{2}\right)$.



RIGHT ANGLED TRIANGLE TRIGONOMETRY



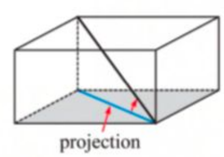
$$\cos\theta = \frac{\text{ADJ}}{\text{HYP}}$$

$$\sin\theta = \frac{\text{OPP}}{\text{HYP}}$$

$$\tan\theta = \frac{\text{OPP}}{\text{ADJ}}$$

Angle between a line and a plane

The **angle between a line and a plane** is the angle between the line and its **projection** on the plane.



True bearings

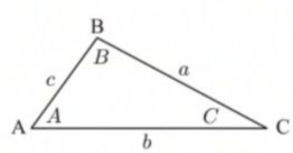
True bearings are used to describe the direction of one object from another. The direction is measured clockwise from true north.

NON-RIGHT ANGLED TRIANGLE TRIGONOMETRY

Area formula: Area $= \frac{1}{2}ab\sin C$

Cosine rule: $a^2 = b^2 + c^2 - 2bc\cos A$

Sine rule: $\frac{\sin A}{a} = \frac{\sin B}{b} = \frac{\sin C}{c}$ or $\frac{a}{\sin A} = \frac{b}{\sin B} = \frac{c}{\sin C}$



RADIAN MEASURE

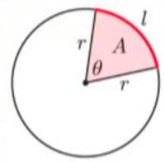
There are $360° \equiv 2\pi$ radians in a circle.

To convert from degrees to radians, multiply by $\frac{\pi}{180}$.

To convert from radians to degrees, multiply by $\frac{180}{\pi}$.

For θ in radians:

- the length of an arc of radius r and angle θ is $l = \theta r$
- the area of a sector of radius r and angle θ is $A = \frac{1}{2}\theta r^2$.



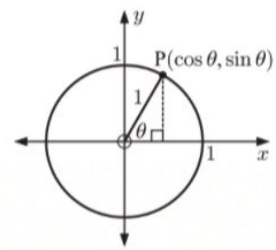
THE UNIT CIRCLE

The **unit circle** is the circle centred at the origin O and with radius 1 unit.

Consider point P on the unit circle where [OP] makes angle θ with the positive x-axis. The coordinates of P are $(\cos\theta, \sin\theta)$.

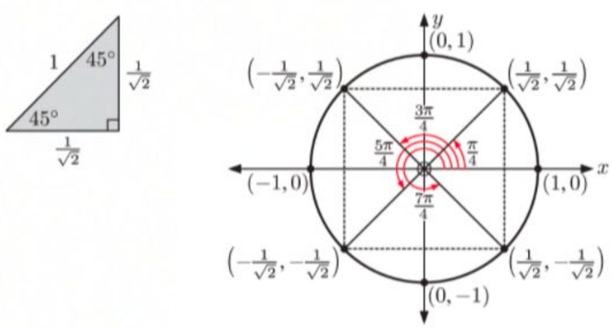
θ is **positive** when measured in an **anticlockwise** direction from the positive x-axis.

$\tan\theta$ is defined as $\frac{\sin\theta}{\cos\theta}$. $\tan\theta$ is the **gradient** of [OP].

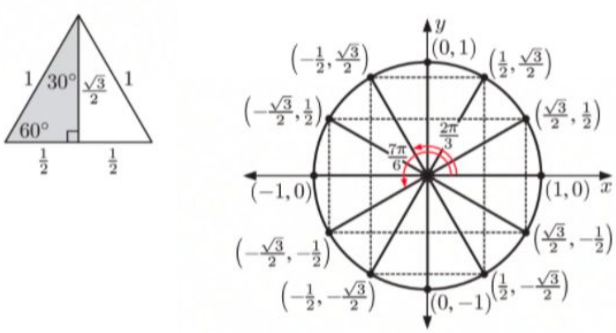


You should memorise or be able to quickly find the values of $\cos\theta$, $\sin\theta$, and $\tan\theta$ that are multiples of $\frac{\pi}{4}$ or $\frac{\pi}{6}$.

Multiples of $\frac{\pi}{4}$ or $45°$



Multiples of $\frac{\pi}{6}$ or $30°$



The **Pythagorean identity** $\cos^2\theta + \sin^2\theta = 1$ can be used to find one trigonometric ratio from another.

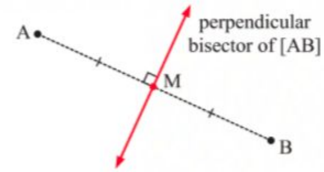
TRIGONOMETRIC EQUATIONS

Trigonometric equations may be solved graphically, using pre-prepared graphs or technology.

PERPENDICULAR BISECTORS

The **perpendicular bisector** of a line segment [AB] is the line perpendicular to [AB] which passes throught its midpoint.

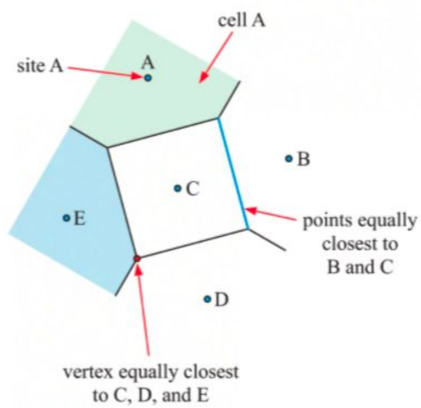
Points on the perpendicular bisector are equidistant from A and B.



VORONOI DIAGRAMS

In a Voronoi diagram:

- Important locations are called **sites**.
- Each site is surrounded by a region or **cell** which contains the points which are closer to that site than to any other site.
- The lines which separate the cells are called **edges**. Each point on an edge is equally closest to the two sites whose cells are adjacent to that edge.
- The points at which the edges meet are called **vertices**. Each vertex is equally closest to the sites whose cells meet at that vertex.



Adding a new site to a Voronoi diagram

To add the cell for a new site X to an existing Voronoi diagram with sites $P_1, P_2, P_3, ..., P_n$, we follow these steps:

Step 1: Identify the site P_i whose cell contains the new site X. Construct the perpendicular bisector of $[P_iX]$, within this cell. At any point where this line meets an existing edge, create a new vertex.

Step 2: For each site P_j whose cell is adjacent to a new vertex, construct the perpendicular bisector of $[P_jX]$ within that cell through the vertex. Continue to create new vertices as in *Step 1*. Repeat this process until no more new vertices are created. At this time cell X is complete.

Step 3: Remove any segments of edges from the original Voronoi diagram which now lie within cell X.

Nearest neighbour interpolation

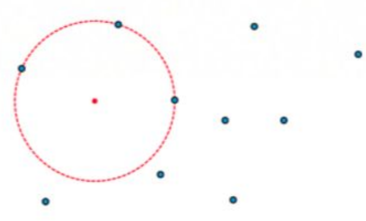
If we are given the values of a variable at a set of known data points, we can *estimate* the value of the variable at some other point. We use the variable's value at the *nearest* known data point.

From a Voronoi diagram with the known data points as sites, we can quickly identify the nearest known data point to any given point.

If the given point lies on an edge or at a vertex, we take the average of the closest known data points.

The Largest Empty Circle problem

The **Largest Empty Circle problem** is the problem of finding the largest circle whose interior does not contain any sites.



In the problems considered in this course, the optimal position for the circle's centre will occur at one of the vertices of the Voronoi diagram. The vertex with the greatest distance from its nearest site is the optimal position for the circle's centre.

VECTORS

A **vector** is a quantity with both **magnitude** and **direction**.

Two vectors are **equal** if and only if they have the same magnitude *and* direction.

In examinations:

- scalars are written in italics a
- vectors are written in bold $\mathbf{a}$.

On paper, you should write vector **a** as $\underset{\sim}{a}$.

The 3-dimensional base unit vectors are $\mathbf{i} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$, $\mathbf{j} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$, and $\mathbf{k} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$.

The 3-dimensional **zero vector 0** is $\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$.

The general 3-dimensional vector $\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = v_1\mathbf{i} + v_2\mathbf{j} + v_3\mathbf{k}$.

You should understand the following for vectors in both algebraic and geometric forms:

- vector addition
- vector subtraction $\mathbf{v} - \mathbf{w} = \mathbf{v} + (-\mathbf{w})$
- multiplication by a scalar k to produce vector $k\mathbf{v}$ which is parallel to **v**
- the magnitude of vector **v**, $|\mathbf{v}| = \sqrt{v_1^2 + v_2^2 + v_3^2}$
- the distance between two points in space is the magnitude of the vector which joins them.

The **position vector** of $\mathrm{A}(x, y, z)$ is $\overrightarrow{\mathrm{OA}}$ or $\mathbf{a} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$.

The **displacement vector of $\mathbf{B}(b_1, b_2, b_3)$ relative to $\mathbf{A}(a_1, a_2, a_3)$** is $\overrightarrow{\mathrm{AB}} = \mathbf{b} - \mathbf{a} = \begin{pmatrix} b_1 - a_1 \\ b_2 - a_2 \\ b_3 - a_3 \end{pmatrix}$.

A, B, and C are **collinear** if $\overrightarrow{\mathrm{AB}} = k\overrightarrow{\mathrm{BC}}$ for some scalar k.

The unit vector in the direction of **a** is $\frac{1}{|\mathbf{a}|}\mathbf{a}$.

THE SCALAR OR DOT PRODUCT OF TWO VECTORS

$\mathbf{v} \bullet \mathbf{w} = v_1w_1 + v_2w_2 + v_3w_3$

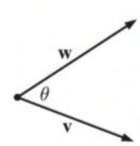
For non-zero vectors **v** and **w**:

- $\mathbf{v} \bullet \mathbf{w} = 0 \Leftrightarrow$ **v** and **w** are perpendicular
- $|\mathbf{v} \bullet \mathbf{w}| = |\mathbf{v}||\mathbf{w}| \Leftrightarrow$ **v** and **w** are parallel

The angle θ between vectors **v** and **w** can be found using $\cos\theta = \frac{\mathbf{v} \bullet \mathbf{w}}{|\mathbf{v}||\mathbf{w}|}$.

If $\mathbf{v} \bullet \mathbf{w} > 0$ then θ is acute.

If $\mathbf{v} \bullet \mathbf{w} < 0$ then θ is obtuse.

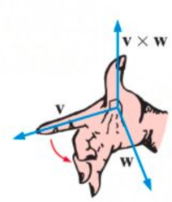


THE VECTOR CROSS PRODUCT OF TWO VECTORS

$$\mathbf{v} \times \mathbf{w} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix}$$

$$= \begin{vmatrix} v_2 & v_3 \\ w_2 & w_3 \end{vmatrix}\mathbf{i} - \begin{vmatrix} v_1 & v_3 \\ w_1 & w_3 \end{vmatrix}\mathbf{j} + \begin{vmatrix} v_1 & v_2 \\ w_1 & w_2 \end{vmatrix}\mathbf{k}$$

$$= (v_2w_3 - v_3w_2)\mathbf{i} - (v_1w_3 - v_3w_1)\mathbf{j} + (v_1w_2 - v_2w_1)\mathbf{k}$$



$\mathbf{v} \times \mathbf{w}$ is perpendicular to both **v** and **w**. Its direction is found using the right hand rule.

Geometric properties of the vector product:

$|\mathbf{v} \times \mathbf{w}| = |\mathbf{v}||\mathbf{w}|\sin\theta$ where θ is the angle between the vectors.

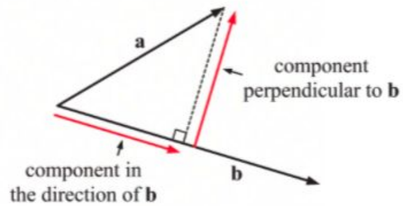
$|\mathbf{v} \times \mathbf{w}| =$ area of parallelogram formed by vectors **v** and **w**.

$\frac{1}{2}|\mathbf{v} \times \mathbf{w}| =$ area of triangle formed by vectors **v** and **w**.

VECTOR COMPONENTS

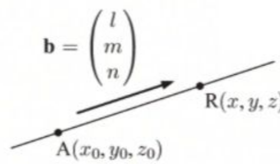
For 3-dimensional vectors **a** and **b**:

- the component of **a** in the direction of **b** is $\frac{\mathbf{a} \bullet \mathbf{b}}{|\mathbf{b}|}$
- the component of **a** perpendicular to **b** is $\frac{|\mathbf{a} \times \mathbf{b}|}{|\mathbf{b}|}$.



LINES

Suppose $\mathrm{R}(x, y, z)$ is any point on the line,
$\mathrm{A}(x_0, y_0, z_0)$ is a known point on the line,
and $\mathbf{b} = \begin{pmatrix} l \\ m \\ n \end{pmatrix}$ is the **direction vector** of the line.



Then:

- The **vector equation** of the line is $\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x_0 \\ y_0 \\ z_0 \end{pmatrix} + \lambda \begin{pmatrix} l \\ m \\ n \end{pmatrix}$, $\lambda \in \mathbb{R}$
- The **parametric equations** of the line are: $\begin{cases} x = x_0 + \lambda l \\ y = y_0 + \lambda m \ , \quad \lambda \in \mathbb{R} \\ z = z_0 + \lambda n \end{cases}$

The **acute angle θ between two lines** is given by $\cos\theta = \frac{|\mathbf{b}_1 \bullet \mathbf{b}_2|}{|\mathbf{b}_1||\mathbf{b}_2|}$ where $\mathbf{b}_1$ and $\mathbf{b}_2$ are the direction vectors of the lines.

The shortest distance from point P to a line with direction vector **b** occurs at the point R on the line such that $\overrightarrow{\mathrm{PR}}$ is perpendicular to **b**.

You should be able to find the shortest distance between two objects moving with constant velocity.

You should also be able to determine whether a pair of lines are **parallel**, **coincident**, **intersecting**, or **skew**.

MOTION WITH VARIABLE VELOCITY

For an object moving with parametric equations $\mathrm{P}(x(t), y(t))$:

- the velocity vector $\mathbf{v} = \begin{pmatrix} x'(t) \\ y'(t) \end{pmatrix}$
- the speed $= |\mathbf{v}| = \sqrt{[x'(t)]^2 + [y'(t)]^2}$.

GEOMETRIC TRANSFORMATIONS

A **linear transformation** is a geometric transformation which can be described by the matrix equation $\mathbf{x}' = \mathbf{A}\mathbf{x}$ where $\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$.

Linear transformations include **stretches**, **rotations**, **reflections**, and any combination or **composition** of these.

An **affine transformation** is a geometric transformation which can be described by the matrix equation $\mathbf{x}' = \mathbf{A}\mathbf{x} + \mathbf{b}$ where $\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ and $\mathbf{b} = \begin{pmatrix} e \\ f \end{pmatrix}$.

In addition to the linear transformations, affine transformations include **translations**.

Translations

Suppose the point (x, y) is **translated** through $\mathbf{b} = \begin{pmatrix} e \\ f \end{pmatrix}$ to the image point (x', y').

The image is given by $\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} e \\ f \end{pmatrix}$.

Rotations

For a **rotation** anticlockwise about $\text{O}(0, 0)$ through θ, the transformation matrix is

$$\mathbf{A} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \text{ with } |\,\mathbf{A}\,| = 1.$$

Reflections

For a **reflection** in the mirror line $y = (\tan\alpha)x$, the transformation matrix is

$$\mathbf{A} = \begin{pmatrix} \cos 2\alpha & \sin 2\alpha \\ \sin 2\alpha & -\cos 2\alpha \end{pmatrix} \text{ with } |\,\mathbf{A}\,| = -1.$$

Stretches and enlargements

For a **horizontal stretch** with scale factor k, the transformation matrix is $\mathbf{A} = \begin{pmatrix} k & 0 \\ 0 & 1 \end{pmatrix}$.

For a **vertical stretch** with scale factor k, the transformation matrix is $\mathbf{A} = \begin{pmatrix} 1 & 0 \\ 0 & k \end{pmatrix}$.

For an **enlargement** with scale factor k, the transformation matrix is $\mathbf{A} = \begin{pmatrix} k & 0 \\ 0 & k \end{pmatrix}$.

Composite transformations

If we perform a linear transformation with transformation matrix **A**, followed by another linear transformation with transformation matrix **B**, then the resulting composite transformation is itself a linear transformation, with transformation matrix **BA**.

The affine transformation $\mathbf{x}' = \mathbf{A}\mathbf{x} + \mathbf{b}$ is the composition of the linear transformation with transformation matrix **A**, followed by a translation through **b**.

Area

For the transformation $\mathbf{x}' = \mathbf{A}\mathbf{x} + \mathbf{b}$, area of image $= |\det \mathbf{A}\,| \times$ area of object.

Topic 4 - Statistics and Probability

SAMPLING

We obtain data from a **sample** of a population when it is impractical to obtain data from the entire population.

You should know the four main categories of **error** that can arise from sampling:

- **Sampling errors** occur when a characteristic of a sample differs from that of the population.
- **Measurement errors** are inaccuracies in measurement during data collection.
- **Coverage errors** occur when a sample does not truly reflect the population.
- **Non-response errors** occur when a large number of people selected for a survey choose not to respond.

SAMPLING METHODS

- In **simple random sampling**:
  - Each member of the population has the same chance of being selected in the sample.
  - Each set of n members of the population has the same chance of being selected as any other set of n members.
- In **systematic sampling**, the sample is created by selecting members of the population at regular intervals.
- In **convenience sampling**, members are chosen for the sample because they are easier to select or more likely to respond.
- In **stratified sampling** or **quota sampling**, the population is divided into subgroups, and the number of members sampled from each subgroup is proportional to the fraction of the population represented by that subgroup. If the members of each subgroup are randomly selected, the sample is a **stratified sample**. If the members are specifically chosen, the sample is a **quota sample**.

TYPES OF DATA AND ITS REPRESENTATION

Categorical data refers to data which describes a particular quality or characteristic.

Discrete data can take any of a set of exact number values $\{x_1, x_2, x_3, ...\}$. It is normally **counted**.

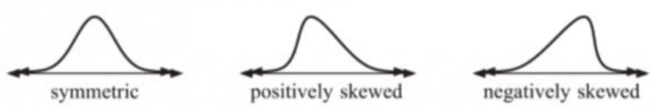
Continuous data can take any numerical value within a certain range. It is normally **measured**.

Grouped data is numerical data which is collected in groups or classes. The **modal class** is the class with the highest frequency.

A **column graph** is used to display discrete data and grouped data. The columns have spaces between them.

A **frequency histogram** is used to display continuous data. The classes are of equal width, and there are no spaces between the columns.

Data may be symmetric, positively skewed, or negatively skewed.



We use a **cumulative frequency graph** to display the cumulative frequency for each data value in a distribution. This enables us to read off the values at each percentile.

MEASURING THE CENTRE OF DATA

The **mean** of a set of scores is their arithmetic average.

For a large population, the **population mean** μ is generally unknown. The **sample mean** $\overline{x}$ is used as an approximation for μ.

For ungrouped data, $\overline{x} = \dfrac{\sum_{i=1}^{n} x_i}{n}$

For data in a frequency table, $\overline{x} = \dfrac{\sum xf}{\sum f}$ where f is the frequency of each value.

For grouped data we can only estimate the mean. We use the **mid-interval value** within each group to represent all scores within that group.

The **median** is the middle value of an ordered data set.

- For an **odd number** of data, the median is one of the original data values.
- For an **even number** of data, the median is the average of the two middle values, and may not be in the original data set.

The **mode** is the most frequently occurring score. If there are two modes we say the data is **bimodal**. For continuous data we refer to a **modal class**.

PERCENTILES

The **kth percentile** is the score a such that $k\%$ of the scores are less than a.

The **lower quartile** (Q_1) is the 25th percentile.

The **median** (Q_2) is the 50th percentile.

The **upper quartile** (Q_3) is the 75th percentile.

You should know how to generate a **cumulative frequency graph** and use it to estimate Q_1, Q_2, and Q_3.

MEASURING THE SPREAD OF DATA

The **range** is the difference between the maximum and the minimum data values.

The **interquartile range** $\text{IQR} = Q_3 - Q_1$.

The **variance** σ^2 is the average of the squares of the distances from the mean.

The **standard deviation** σ is the square root of the variance.

You should be able to use technology to calculate standard deviation. If we are given the whole population we use the population standard deviation σ_x. If we are only given a sample from a larger population, we use the sample standard deviation s_x.

OUTLIERS

Outliers are extraordinary data that are separated from the main body of the data. We test for outliers by calculating upper and lower boundaries:

- upper boundary $= Q_3 + 1.5 \times \text{IQR}$
- lower boundary $= Q_1 - 1.5 \times \text{IQR}$

Any data outside of these boundaries is considered an outlier.

BOX AND WHISKER DIAGRAMS

A **box and whisker diagram** or **box plot** illustrates the **five-number summary** of a data set:

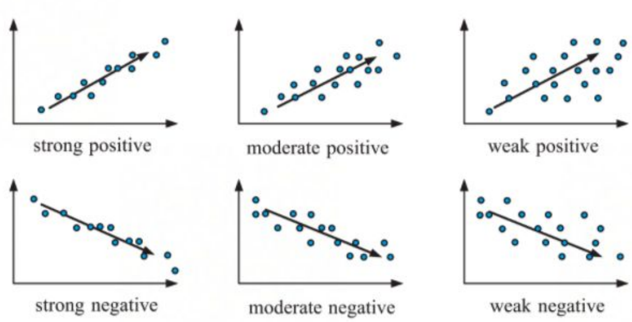
- minimum value
- Q_1
- median
- Q_3
- maximum value

An outlier is indicated by an asterisk $*$.

BIVARIATE STATISTICS

Correlation refers to the relationship between two numerical variables.

We can use a **scatter diagram** to help identify **outliers** and to describe the correlation between variables. We consider **direction**, **strength**, and **linearity**.



If a change in one variable *causes* a change in the other variable then we say there is a **causal relationship** between them.

To measure the strength of the relationship between two variables, we use **Pearson's product-moment correlation coefficient r**.

The correlation coefficient lies in the range $-1 \leqslant r \leqslant 1$.

- The sign of r indicates the direction of correlation.
  - A positive value for r indicates the variables are positively correlated.
  - A negative value for r indicates the variables are negatively correlated.
- The size of r indicates the strength of correlation.
  - A value of r close to $+1$ or -1 indicates strong correlation between the variables.
  - A value of r close to zero indicates weak correlation between the variables.

The coefficient of determination

To help describe the correlation between two variables, we can also calculate the **coefficient of determination**, r^2. This is simply the square of Pearson's product-moment correlation coefficient r, so the direction of correlation is eliminated.

If there is a causal relationship, then r^2 indicates the degree to which change in the independent variable explains change in the dependent variable.

Line of best fit

If two variables are linearly correlated, we can draw a line of best fit to illustrate their relationship.

We can draw a **line of best fit by eye**, which passes through the **mean point** $(\overline{x}, \overline{y})$, and which fits the trend of the data.

To get a more accurate line of best fit, we use a method called **linear regression**. The line obtained is called the **least squares regression line**. You should be able to find this line using your calculator.

When using a line of best fit to estimate values, **interpolation** is usually reliable, whereas **extrapolation** may not be.

Spearman's rank correlation coefficient of a bivariate data set is defined as the Pearson product-moment correlation coefficient of the variables' **ranks**. It is often used when the data is clearly non-linear, but has an upward or downward trend.

Sum of squared residuals

When fitting a model to a set of data points, the **residual** of the ith data point (x_i, y_i) is $\boldsymbol{r_i = y_i - \widehat{y}_i}$ where $\widehat{y}_i$ is the model's predicted value of y at $x = x_i$.

The **sum of squared residuals** is defined as $SS_{\text{res}} = \sum_{i=1}^{n} r_i^2 = \sum_{i=1}^{n} (y_i - \widehat{y}_i)^2$.

Non-linear regression

Non-linear regression is used to fit non-linear models to data by minimising SS_{res}.

When given several fitted models, you should be able to choose the most appropriate one by calculating SS_{res} for each model.

STATISTICAL RELIABILITY AND VALIDITY

- **Statistical reliability** is a measure of how *consistently* a variable can be measured.

  We can consider:
  - **Test-retest reliability**, in which identical tests are performed at different times. We are interested in how consistent the measurement is over *time*.
  - **Parallel forms reliability**, in which similar tests are performed. We are interested in how consistent the measurement is across different *versions*.
- **Statistical validity** considers how *accurately* a variable measures a particular aspect of a population.
  - **Content validity** considers how well the field of study or *content domain* is covered.
  - **Criterion validity** considers how well one variable predicts another valid variable, called the **criterion variable**.

PROBABILITY

A **trial** occurs each time we perform an experiment.

The possible results from each trial of an experiment are called its **outcomes**.

The **sample space** U is the set of all possible outcomes of an experiment.

Experimental probability

In many situations, we can only measure the probability of an event by experimentation.

experimental probability = relative frequency of event

Theoretical probability

If all outcomes are equally likely, the probability of event A is $\text{P}(A) = \frac{n(A)}{n(U)}$.

For any event A, $0 \leqslant \text{P}(A) \leqslant 1$.

For any event A, A' is the event that A does not occur. A and A' are **complementary events**, and $\text{P}(A) + \text{P}(A') = 1$.

The event that both A **and** B occur is written $A \cap B$.

The event that A **or** B **or both** occur is written $A \cup B$.

$\text{P}(A \cup B) = \text{P}(A) + \text{P}(B) - \text{P}(A \cap B)$

For **disjoint** or **mutually exclusive** events, $\text{P}(A \cap B) = 0$.

Making predictions using probability

If there are n trials of an experiment, and an event has probability p of occurring in each of the trials, then the number of times we *expect* the event to occur is np.

Independent events

Two events are **independent** if the occurrence of each of them does not affect the probability that the other occurs. An example of this is sampling **with replacement**.

For independent events A and B, $\text{P}(A \cap B) = \text{P}(A)\,\text{P}(B)$.

Dependent events

Two events are **dependent** if the occurrence of one of them *does* affect the probability that the other occurs. An example of this is sampling **without replacement**.

For dependent events A and B, $\text{P}(A \cap B) = \text{P}(A) \times \text{P}(B \text{ given that } A \text{ has occurred})$.

Conditional probability

For any two events A and B, "$A \mid B$" represents the event "A given that B has occurred", and $\text{P}(A \mid B) = \frac{\text{P}(A \cap B)}{\text{P}(B)}$.

MARKOV CHAINS

A **Markov chain** is a model which describes how a sequence of random events evolves over time.

- Time is measured in **discrete** steps.
- The possible outcomes of each event are the **states** that the system could be in.
- The probabilities for the possible states of the system at the next time step depend *only* on the state at the current time step.

The **state matrix** $\mathbf{s}_n$ shows the state of the system at time n. The state matrix may represent either:

- the probability or proportion of the population that are in each state, or
- the actual number from the population that are in the given state.

The **initial state matrix** is $\mathbf{s}_0$.

In a **transition matrix** $\mathbf{T} = (t_{ij})$:

- The columns represent the current state.
- The rows represent the next state.
- t_{ij} is the probability of moving to state i from state j. $t_{ij} = \text{P}(\text{next state is } i \mid \text{current state is } j)$

The state of the system at time n is given by $\mathbf{s}_n = \mathbf{T}^n \mathbf{s}_0$.

Steady state

When there is very little or no change in the values of the state matrices from one time step to the next, we say the system has reached a **steady state**.

We can find the steady state of a Markov chain with transition matrix **T** by either:

- Calculating $\mathbf{s}_n = \mathbf{T}^n \mathbf{s}_0$ for large values of n.
- Finding **s** such that $\mathbf{Ts} = \mathbf{s}$ algebraically.

DISCRETE RANDOM VARIABLES

A **random variable** represents the possible numerical outcomes of an experiment.

A **discrete random variable** can take any of a set of distinct values.

If X is a discrete random variable with possible values $\{x_1, x_2, ..., x_n\}$ and corresponding probabilities $\{p_1, p_2, ..., p_n\}$, then:

- $0 \leqslant p_i \leqslant 1$ for all $i = 1, ..., n$
- $\sum\limits_{i=1}^{n} p_i = p_1 + p_2 + ... + p_n = 1$
- $\{p_1, p_2, ..., p_n\}$ describes the **probability distribution** of X.

We can also describe the probability distribution of X using a **probability mass function** $P(x) = \mathrm{P}(X = x)$.

The **expectation** of a discrete random variable X is $\mathrm{E}(X) = \mu = \sum\limits_{i=1}^{n} x_i p_i$.

A game where X is the gain to the player is said to be **fair** if $\mathrm{E}(X) = 0$.

The **mode** is the data value x_i whose probability p_i is the highest.

The **variance** is
$$\begin{aligned}\mathrm{Var}(X) &= \sigma^2 \\ &= \mathrm{E}\big[(X-\mu)^2\big] \\ &= \sum (x_i - \mu)^2 p_i \\ &= \sum x_i^2 p_i - \mu^2 \\ &= \mathrm{E}(X^2) - \mu^2 \\ &= \mathrm{E}(X^2) - [\mathrm{E}(X)]^2\end{aligned}$$

The **standard deviation** is $\sigma(X) = \sqrt{\mathrm{Var}(X)}$.

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

THE BINOMIAL DISTRIBUTION

In a **binomial experiment** there are two possible results: success and failure.

Suppose there are n independent trials of the same experiment with the probability of success being a constant p for each trial. If X represents the number of successes in the n trials, then X has a **binomial distribution**, and we write $X \sim \mathrm{B}(n, p)$.

The **binomial probability mass function** is $\mathrm{P}(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$ where $x = 0, 1, 2, ..., n$.

You should be able to use your calculator to find:

- $\mathrm{P}(X = x)$ using the binomial probability distribution function
- $\mathrm{P}(X \leqslant x)$ or $\mathrm{P}(X \geqslant x)$ using the binomial cumulative distribution function.

If $X \sim \mathrm{B}(n, p)$, then:

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

THE POISSON DISTRIBUTION

The **Poisson distribution** arises when considering the number of occurrences within a certain interval (of time or space).

Suppose there are an average of λ occurrences in a given interval. If X represents the number of occurrences in a particular interval, then X has a Poisson distribution, and we write $X \sim \text{Po}(\lambda)$.

The probability mass function of X is $P(x) = \text{P}(X = x) = \dfrac{\lambda^x e^{-\lambda}}{x!}$ for $x = 0, 1, 2, ...$

$\text{E}(X) = \mu = \lambda$ and $\text{Var}(X) = \sigma^2 = \lambda$.

THE NORMAL DISTRIBUTION

If the random variable X has a normal distribution with mean μ and variance σ^2, we write $X \sim \text{N}(\mu, \sigma^2)$.

The probability density function is $f(x) = \dfrac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$ for $x \in \mathbb{R}$.

$f(x)$ is a bell-shaped curve which is symmetric about $x = \mu$.

It has the property that:

- $\approx 68\%$ of all scores lie between $\mu - \sigma$ and $\mu + \sigma$
- $\approx 95\%$ of all scores lie between $\mu - 2\sigma$ and $\mu + 2\sigma$
- $\approx 99.7\%$ of all scores lie between $\mu - 3\sigma$ and $\mu + 3\sigma$.

You should be able to use your calculator to find normal probabilities for the situations:

- $\text{P}(X \leqslant a)$
- $\text{P}(X \geqslant a)$
- $\text{P}(a \leqslant X \leqslant b)$

You should also be able to use your calculator to find the scores corresponding to particular probabilities. These scores are known as **quantiles**.

ESTIMATION AND CONFIDENCE INTERVALS

Consider a population with distribution X.

A **random sample** of size n $\{X_1, X_2, ..., X_n\}$ is a set of independent observations from the **same** population.

- Each X_i is a random variable that has the same distribution as the population.
- $X_1, X_2, ..., X_n$ are **independent random variables**.

Linear combinations of random variables

A **linear combination** of the random variables $X_1, X_2, ..., X_n$ has the form $a_1X_1 + a_2X_2 + ... + a_nX_n$ where $a_1, a_2, ..., a_n$ are constants.

For n random variables $X_1, X_2, ..., X_n$, and constants $a_1, a_2, ..., a_n$:

1 $\text{E}\left(\sum\limits_{i=1}^{n} a_iX_i\right) = \sum\limits_{i=1}^{n} a_i\,\text{E}(X_i)$

2 If $X_1, X_2, ..., X_n$ are all independent of one another, then $\text{Var}\left(\sum\limits_{i=1}^{n} a_iX_i\right) = \sum\limits_{i=1}^{n} a_i^2\,\text{Var}(X_i)$.

If $X \sim \text{Po}(\lambda_1)$ and $Y \sim \text{Po}(\lambda_2)$ are independent Poisson random variables, then $(X + Y) \sim \text{Po}(\lambda_1 + \lambda_2)$.

Any linear combination of independent normally distributed random variables is itself a normally distributed random variable.

For a random sample $\{X_1, X_2, ..., X_n\}$:

- the sample mean is $\overline{X}_n = \dfrac{1}{n}\sum\limits_{i=1}^{n} X_i$
- the sample variance is $S_{n-1}^2 = \dfrac{1}{n-1}\sum\limits_{i=1}^{n}(X_i - \overline{X}_n)^2$.

Properties of $\overline{X}_n$ and S_{n-1}^2

For a population with mean μ and standard deviation σ:

- $\mathrm{E}(\overline{X}_n) = \mu$, so $\overline{X}_n$ is an unbiased estimator for μ.
- $\mathrm{Var}(\overline{X}_n) = \frac{\sigma^2}{n}$ and $\sigma(\overline{X}_n) = \frac{\sigma}{\sqrt{n}}$
- If the population is normally distributed, then $\overline{X}_n \sim \mathrm{N}\left(\mu, \frac{\sigma^2}{n}\right)$ for all n.
- $\mathrm{E}(S_{n-1}^2) = \sigma^2$, so S_{n-1}^2 is an unbiased estimator for σ^2.

The Central Limit Theorem

For any population, even one not normally distributed, $\overline{X}_n$ is approximately normally distributed for sufficiently large n, with $\overline{X}_n \sim \mathrm{N}\left(\mu, \frac{\sigma^2}{n}\right)$.

You should use the "rule of thumb" of $n \geqslant 30$ being sufficiently large.

Confidence intervals

A **confidence interval** for a population mean μ, is an interval of values between two limits, together with a percentage indicating our conficence that μ lies in that interval.

Confidence intervals for a population mean with known variance

The general confidence interval for μ given a known population variance σ^2 and a data set with sample mean $\overline{x}$ is

$$\overline{x} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leqslant \mu \leqslant \overline{x} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

where $\mathrm{P}(Z \geqslant z_{\frac{\alpha}{2}}) = \frac{\alpha}{2}$ and $Z \sim \mathrm{N}(0, 1^2)$.

The interval contains μ with probability $1 - \alpha$, so the confidence level of the interval is $(1 - \alpha) \times 100\%$.

Confidence intervals for a population mean with unknown variance

The general confidence interval for μ given a data set with sample mean $\overline{x}$ and sample standard deviation s is

$$\overline{x} - t_{n-1,\, \frac{\alpha}{2}} \frac{s}{\sqrt{n}} \leqslant \mu \leqslant \overline{x} + t_{n-1,\, \frac{\alpha}{2}} \frac{s}{\sqrt{n}}$$

where $\mathrm{P}(T \geqslant t_{n-1,\, \frac{\alpha}{2}}) = \frac{\alpha}{2}$ and $T \sim t_{n-1}$ is the t-distribution with $n - 1$ degrees of freedom (df).

The interval contains μ with probability $1 - \alpha$, so the confidence level of the interval is $(1 - \alpha) \times 100\%$.

HYPOTHESIS TESTING

Terminology

- A **statistical hypothesis** is a claim about a population parameter.
- The **null hypothesis** H_0 is a claim that the population parameter is *equal* to a particular value.
- The **alternative hypothesis** H_1 is a claim that the population parameter is *different* to the value specified by H_0. For example, given the null hypothesis H_0: $\mu = \mu_0$, the alternative hypothesis could be:
  - H_1: $\mu > \mu_0$ **(one-tailed hypothesis)**
  - H_1: $\mu < \mu_0$ **(one-tailed hypothesis)**
  - H_1: $\mu \neq \mu_0$ (**two-tailed hypothesis**, as $\mu \neq \mu_0$ could mean $\mu > \mu_0$ or $\mu < \mu_0$).
- A **Type I error** is when we make the mistake of rejecting H_0 when H_0 is in fact true.
- A **Type II error** is when we make the mistake of accepting H_0 when H_0 is in fact false.
- A **test statistic** is a random variable that summarises the information in a sample.
- The distribution of the test statistic under the assumptions of H_0 is called the **null distribution**.

- The **p-value** of a test statistic is the probability of a result that is as or more "extreme" being observed if H_0 is true.
- The **significance level** α of a statistical hypothesis test is the largest p-value that would result in rejecting H_0. Any p-value less than or equal to α results in H_0 being rejected.
  - If a statistical hypothesis test has significance level α, the probability of a Type I error is α.
  - The significance level may be given as a decimal or a percentage.
- The **critical region** $\mathcal{C}$ is the set of all values of the test statistic which result in H_0 being rejected.
- The **acceptance region** $\mathcal{A}$ is the set of all values of the test statistic which result in H_0 being accepted.
- We can make a decision about H_0 using the test statistic directly by comparing it to a **critical value** c which is the value in the critical region which has the largest p-value associated with it.

General procedure

Step 1: Formulate **statistical hypotheses**.

Step 2: Choose a **significance level** for the test. This is a threshold for making a decision, like the confidence levels we saw previously.

Step 3: Use data from a sample to calculate a **test statistic**.

Step 4: Calculate a **p-value** for the test statistic. This is the probability of that test statistic occurring under the assumptions of one of the hypotheses.

Step 5: Make decisions about the hypotheses.

Step 6: Interpret the decision in the context of the problem.

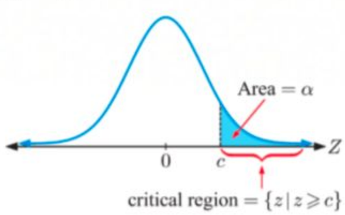
The Z-test

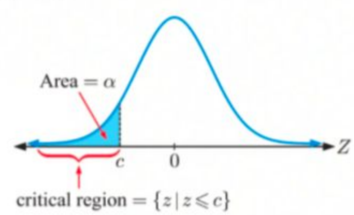
The **Z-test** is used to test hypotheses about a population mean μ when:

- the population is normally distributed
- the population variance σ^2 is **known**.

For a Z-test of $H_0\colon\ \mu = \mu_0$ (where the population standard deviation is σ) using a sample of size n with sample mean $\overline{x}$:

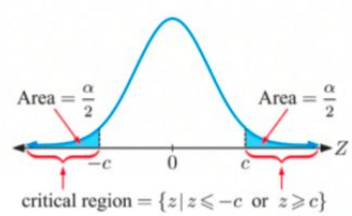
- the **test statistic** is $Z = \dfrac{\overline{X}_n - \mu_0}{\frac{\sigma}{\sqrt{n}}}$ which has **observed value** $z = \dfrac{\overline{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$
- the **null distribution** is $Z \sim \mathrm{N}(0, 1^2)$
- the **p-value** calculation depends on H_1:
  - If $H_1\colon\ \mu > \mu_0$, $p\text{-value} = \mathrm{P}(Z \geqslant z)$.
  - If $H_1\colon\ \mu < \mu_0$, $p\text{-value} = \mathrm{P}(Z \leqslant z)$.
  - If $H_1\colon\ \mu \neq \mu_0$, $p\text{-value} = 2 \times \mathrm{P}(Z \geqslant |\, z\,|)$.
- the **critical value(s)** and **critical region** depend on H_1 and the significance level α:
  - If $H_1\colon\ \mu > \mu_0$, the critical value c satisfies $\mathrm{P}(Z \geqslant c) = \alpha$.
  - If $H_1\colon\ \mu < \mu_0$, the critical value c satisfies $\mathrm{P}(Z \leqslant c) = \alpha$.





  - If $H_1\colon\ \mu \neq \mu_0$, then the critical value c satisfies $2 \times \mathrm{P}(Z \geqslant |\, c\,|) = \alpha$

$$\therefore\ \mathrm{P}(Z \geqslant |\, c\,|) = \frac{\alpha}{2}$$



The one-sample t-test

The **t-test** is used to test hypotheses about a population mean μ when:

- the population is normally distributed
- the population variance is **unknown**.

For a t-test of $H_0\colon\ \mu = \mu_0$ using a sample of size n with sample mean $\overline{x}$ and sample standard deviation s:

- the **test statistic** is $T = \dfrac{\overline{X}_n - \mu_0}{\frac{S_{n-1}}{\sqrt{n}}}$ which has **observed value** $t = \dfrac{\overline{x} - \mu_0}{\frac{s}{\sqrt{n}}}$
- the **null distribution** is $T \sim t_{n-1}$
- the p-value calculation depends on H_1:
  - If $H_1\colon\ \mu > \mu_0,$ p-value $= \mathrm{P}(T \geqslant t)$.
  - If $H_1\colon\ \mu < \mu_0,$ p-value $= \mathrm{P}(T \leqslant t)$.
  - If $H_1\colon\ \mu \neq \mu_0,$ p-value $= 2 \times \mathrm{P}(T \geqslant |\,t\,|)$.

Paired t-tests

A **paired t-test** is used to compare two sets of results for **one** sample. In other words, the data is **matched in pairs**.

To perform a paired t-test we:

- calculate the difference of each pair of data values d_i
- perform a one-sample t-test for the mean of the differences μ_D with null hypothesis $H_0\colon\ \mu_0 = 0$.

The two-sample t-test

The **two-sample t-test** is used to compare the means of **two** samples from different populations.

If the populations have means μ_1 and μ_2, the null hypothesis has the form:

$$H_0\colon\quad \mu_1 = \mu_2 \quad \text{or equivalently}$$
$$H_0\colon\ \mu_1 - \mu_2 = 0$$

You should be able to use technology to calculate the test statistic and p-value.

In this course you are expected to assume **equal variances** and hence use the **pooled two-sample t-test** on your calculator.

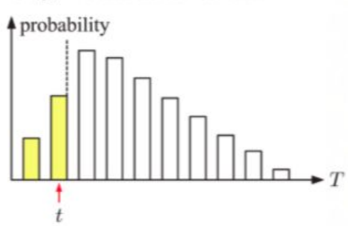
Hypothesis tests for the mean of a Poisson population

Consider a statistical hypothesis test of $H_0\colon\ \lambda = \lambda_0$ for a Poisson distributed population. Given a sample of size n with observed sample mean $\overline{x}$:

- the **test statistic** is $T = n\overline{X}_n$ which has observed value $t = n\overline{x}$
- the **null distribution** is $T \sim \mathrm{Po}(n\lambda_0)$.

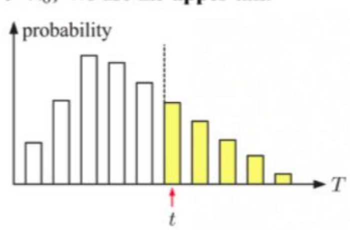
In this course we only consider p-value calculations for Poisson tests with a one-tailed alternative hypothesis H_1.

- If $H_1\colon\ \lambda < \lambda_0,$ we use the **lower tail**.



p-value $= \mathrm{P}(T \leqslant t)$

- If $H_1\colon\ \lambda > \lambda_0,$ we use the **upper tail**.



p-value $= \mathrm{P}(T \geqslant t)$

Hypothesis tests for a population proportion

Consider a statistical hypothesis test of $H_0\colon\ p = p_0,$ where p is the proportion of the population with a particular property. Given a sample of size n:

- the **test statistic** is X, the number of members in the sample with the property of interest
- the **null distribution** is $X \sim \mathrm{B}(n, p_0)$.

Like the Poisson distribution, the binomial distribution is discrete. We calculate the p-value using the same principles.

Hypothesis tests for a population correlation coefficient

When we talk about the correlation between two variables in a population, we use the **population product-moment correlation coefficient** ρ.

If no relationship or association exists between two variables, then there is **no correlation** between then and $\rho = 0$. In a hypothesis test, this is equivalent to stating a null hypothesis H_0: $\rho = 0$.

If we are interested in detecting:

- **positive** correlation, we use H_1: $\rho > 0$
- **negative** correlation, we use H_1: $\rho < 0$
- **any** correlation, whether positive or negative, we use H_1: $\rho \neq 0$.

You should be able to use your calculator to conduct hypothesis tests about ρ.

Error probabilities and statistical power

significance level $= \alpha =$ P(Type I error) $=$ P(Reject $H_0 \mid H_0$ true)

$\beta =$ P(Type II error) $=$ P(Retain $H_0 \mid H_0$ false)

The **power** of a hypothesis test is defined as $1 - \beta$, the probability of (correctly) rejecting the null hypothesis H_0 when H_0 is false.

For a Z-test, a test for a Poisson population mean, and a test for a population proportion, you should be able to calculate:

- α given the critical region
- β given the critical region and the value that the population parameter takes under the alternative hypothesis H_1.

The χ^2 goodness of fit test

The χ^2 goodness of fit test is used to determine whether a probability distribution fits a set of data.

Consider a scenario with k categories. Let p_i be the population proportion of individuals in category i, where $p_1 + p_2 + ... + p_k = 1$.

The **hypotheses** in a χ^2 goodness of fit test have the form:

$$H_0\text{: } p_1 = p_{01},\ p_2 = p_{02},\ \ldots.,\ \text{and } p_k = p_{0k}$$
$$H_1\text{: at least one of } p_i \neq p_{0i}$$

where p_{0i} is the population proportion of category i under the null hypothesis.

You should also be able to calculate p_{01}, p_{02}, ..., p_{0k} for special probability distributions including the binomial distribution, Poisson distribution, and normal distribution.

The **test statistic** for a χ^2 goodness of fit test is: $\chi^2_{\text{calc}} = \sum \frac{(f_{\text{obs}} - f_{\text{exp}})^2}{f_{\text{exp}}}$

where f_{obs} is an **observed** frequency

f_{exp} is an **expected** frequency.

You should combine similar categories to ensure that no expected frequencies are less than 5.

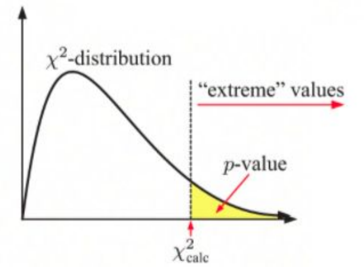
Degrees of freedom (df) refers to the number of values that are "*free to vary*".

For a χ^2 goodness of fit test, **df = number of categories − number of estimated parameters − 1** .

This table gives the degrees of freedom of the goodness of fit test for various distributions when the data is sorted into k categories:

Distribution	*Estimated parameters*	df
Binomial B(n, p)	$p \approx \frac{\overline{x}}{n}$	$k-2$
Poisson Po(λ)	$\lambda \approx \overline{x}$	$k-2$
Normal N(μ, σ^2)	$\sigma^2 \approx s^2$	$k-2$
	$\mu \approx \overline{x},\ \ \sigma^2 \approx s^2$	$k-3$

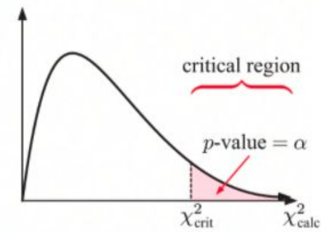
p-value = probability of observing a value greater than or equal to χ^2_{calc}.



For a χ^2 goodness of fit test, we denote the critical value as χ^2_{crit}.

Since we use the upper tail of the null distribution in calculating the p-value, the critical region is the set of values $\geqslant \chi^2_{\text{crit}}$.

The inequality $\chi^2_{\text{calc}} \geqslant \chi^2_{\text{crit}}$ is called the **rejection inequality**.



The χ^2 test for independence

The **χ^2 test for independence** is used to determine if two variables in a **contingency table** are independent or not. It is a special case of the χ^2 goodness of fit test.

The hypotheses for the χ^2 test for independence are H_0: the variables are independent
H_1: the variables are dependent

The test statistic for the χ^2 test for independence is calculated in a similar way to the χ^2 goodness of fit test. The expected frequency of each cell in the contingency table is given by $f_{\text{exp}} = \dfrac{\text{row sum} \times \text{column sum}}{\text{total}}$.

The p-value and critical value χ^2_{crit} for the χ^2 test for independence are calculated in the same way as the χ^2 goodness of fit test.

For a contingency table which has r rows and c columns, $\mathbf{df = (r-1)(c-1)}$.

Topic 5 - Calculus

LIMITS

If $f(x)$ can be made as close as we like to some real number A by making x sufficiently close to a, we say that $f(x)$ has a **limit** of A as x approaches a, and we write $\lim_{x \to a} f(x) = A$.

We say that as x approaches a, $f(x)$ **converges** to A.

RATES OF CHANGE

The **instantaneous rate of change** of a variable at a particular instant is given by the **gradient of the tangent** to the graph at that point.

$\frac{dy}{dx}$ gives the rate of change in y with respect to x.

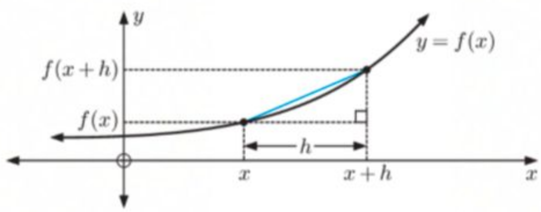
If $\frac{dy}{dx}$ is positive, then as x increases, y also increases.

If $\frac{dy}{dx}$ is negative, then as x increases, y decreases.

DIFFERENTIATION

The **gradient function** or **derivative function** $f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$ provides:

- the rate of change of f with respect to x
- the gradient of the tangent to $y = f(x)$ for any value of x.



RULES OF DIFFERENTIATION

$f(x)$	$f'(x)$	*Name of rule*
c	0	
x^n	nx^{n-1}	
$e^{f(x)}$	$e^{f(x)} f'(x)$	exponentials
$\ln f(x)$	$\frac{f'(x)}{f(x)}$	logarithms
$\sin x$	$\cos x$	
$\cos x$	$-\sin x$	trigonometric functions
$\tan x$	$\frac{1}{\cos^2 x}$	

$f(x)$	$f'(x)$	*Name of rule*
$cu(x)$	$cu'(x)$	
$u(x) + v(x)$	$u'(x) + v'(x)$	addition rule
$u(x)v(x)$	$u'(x)v(x) + u(x)v'(x)$	product rule
$\frac{u(x)}{v(x)}$	$\frac{u'(x)v(x) - u(x)v'(x)}{[v(x)]^2}$	quotient rule

Chain rule

If $y = f(u)$ where $u = u(x)$ then $\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx}$.

SECOND DERIVATIVES

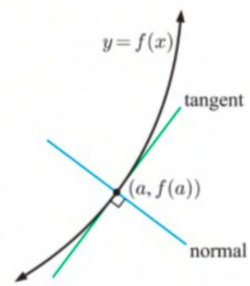
The second derivative of $y = f(x)$ is written $f''(x)$ or $\frac{d^2y}{dx^2}$.

PROPERTIES OF CURVES

Tangents and normals

For the curve $y = f(x)$:

- The gradient of the tangent at $x = a$ is $f'(a)$.
- The equation of the tangent at $x = a$ is $y = f'(a)(x - a) + f(a)$.
- The gradient of the normal at $x = a$ is $-\frac{1}{f'(a)}$.
- The equation of the normal at $x = a$ is $y = -\frac{1}{f'(a)}(x - a) + f(a)$.



Increasing and decreasing functions

$f(x)$ is **increasing** on an interval S $\Leftrightarrow$ $f(a) \leqslant f(b)$ for all $a, b \in S$ such that $a < b$.

$f(x)$ is **decreasing** on S $\Leftrightarrow$ $f(a) \geqslant f(b)$ for all $a, b \in S$ such that $a < b$.

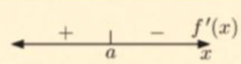
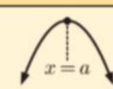
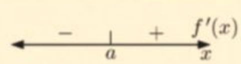
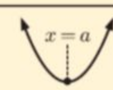
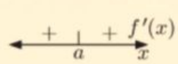
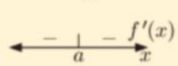
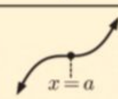
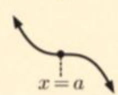
For most functions:

- $f(x)$ is increasing on S $\Leftrightarrow$ $f'(x) \geqslant 0$ for all x in S.
- $f(x)$ is decreasing on S $\Leftrightarrow$ $f'(x) \leqslant 0$ for all x in S.

Stationary points

A **stationary point** of a function is a point such that $f'(x) = 0$.

You should be able to identify and explain the significance of local and global maxima and minima, as well as stationary and non-stationary inflections.

<table>
<tr><th>Stationary point where
$f'(a) = 0$</th><th>Sign diagram of $f'(x)$
near $x = a$</th><th>Shape of curve near
$x = a$</th></tr>
<tr><td>local maximum</td><td></td><td></td></tr>
<tr><td>local minimum</td><td></td><td></td></tr>
<tr><td>stationary inflection</td><td>
or
</td><td>
or
</td></tr>
</table>

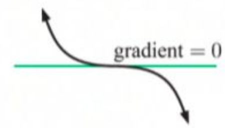
Shape

If $f''(x) \leqslant 0$ for all $x \in S$, the curve is **concave down** on the interval S. 

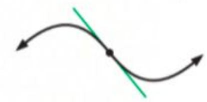
If $f''(x) \geqslant 0$ for all $x \in S$, the curve is **concave up** on the interval S. 

There is a **point of inflection** at $x = a$ if $f''(a) = 0$ **and** the sign of $f''(x)$ changes on either side of $x = a$. It corresponds to a change in shape of the curve.

If $f'(a) = 0$, the point of inflection is a **stationary inflection**: the tangent at $x = a$ is horizontal.



If $f'(a) \neq 0$, the point of inflection is a **non-stationary inflection**: the tangent at $x = a$ is *not* horizontal.



OPTIMISATION PROBLEMS

It is important to remember that a local minimum or maximum does not always give the minimum or maximum value of a function in a particular domain. You must check for other turning points in the domain, and the values of the function at the end points of the domain.

Optimisation problem solving method

Step 1: Draw a large, clear diagram of the situation.

Step 2: Construct a **formula** with the variable to be optimised as the subject. It should be written in terms of one convenient variable, for example x. You should write down what domain restrictions there are on x.

Step 3: Find the **first derivative** and find the value(s) of x which make the first derivative **zero**.

Step 4: For each stationary point, use the **sign diagram test** or **second derivative test** to determine whether you have a local maximum or local minimum.

Step 5: Identify the optimal solution, also considering end points where appropriate.

Step 6: Write your answer in a sentence, making sure you specifically answer the question.

RELATED RATES

If the variables x and y are related, then $\frac{dx}{dt}$ and $\frac{dy}{dt}$ are **related rates**.

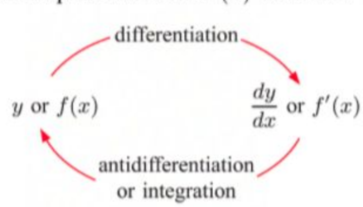
To solve problems involving related rates, we:

- Write an equation connecting the variables.
- Use the chain rule to differentiate the equation with respect to time t.
- Substitute the values for the *particular case* corresponding to some instant in time, and solve to find the required unknown.

INTEGRATION

Antidifferentiation or **integration** is the reverse process of differentiation.

The **antiderivative** or **integral** of $f(x)$ is the simplest function $F(x)$ such that $F'(x) = f(x)$.



Techniques for integration

When integrating, we use the rules for differentiation in reverse. Do not forget to include the **constant of integration**.

Function	*Integral*
k	$kx+c$
x^n	$\frac{x^{n+1}}{n+1}+c, \quad n \neq -1$
e^x	e^x+c
$\frac{1}{x}$	$\ln\lvert x \rvert + c$
e^{ax+b}	$\frac{1}{a}e^{ax+b}+c, \quad a \neq 0$
$(ax+b)^n$	$\frac{(ax+b)^{n+1}}{a(n+1)}+c, \quad a \neq 0, \quad n \neq -1$
$\frac{1}{ax+b}$	$\frac{1}{a}\ln\lvert ax+b \rvert, \quad a \neq 0$
$\cos(ax+b)$	$\frac{1}{a}\sin(ax+b)+c, \quad a \neq 0$
$\sin(ax+b)$	$-\frac{1}{a}\cos(ax+b)+c, \quad a \neq 0$
$\frac{1}{\cos^2(ax+b)}$	$\frac{1}{a}\tan(ax+b)+c, \quad a \neq 0$

Integration by substitution

$$\int f(u)\frac{du}{dx}\,dx = \int f(u)\,du$$

When using substitution to evaluate a definite integral, make sure you change the limits of integration to correspond to the new variable.

DEFINITE INTEGRALS

Fundamental Theorem of Calculus

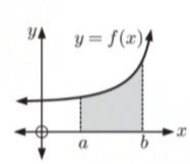
For a continuous function $f(x)$ with antiderivative $F(x)$, $\int_a^b f(x)\,dx = F(b) - F(a)$.

Properties of definite integrals

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

Area under a curve

If $f(x)$ is a continuous *positive* function on the interval $a \leqslant x \leqslant b$, then $\int_a^b f(x)\,dx$ is the area under the curve between $x = a$ and $x = b$.



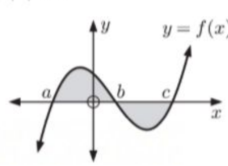
To find the total area enclosed by $y = f(x)$ and the x-axis between $x = a$ and $x = b$, we need to be careful about where $f(x) < 0$.

On an interval $c \leqslant x \leqslant d$ where $f(x) < 0$, the area is $-\int_c^d f(x)\,dx$.

For example:

The total shaded area $= \int_a^b f(x)\,dx - \int_b^c f(x)\,dx$

$\neq \int_a^c f(x)\,dx.$

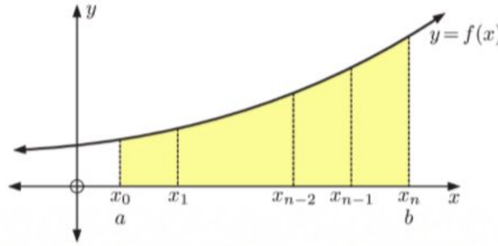


The trapezoidal rule

Suppose we divide the interval $a \leqslant x \leqslant b$ into n subintervals of equal width $h = \frac{b-a}{n}$.

The shaded area $\int_a^b f(x)\,dx$ is approximated by

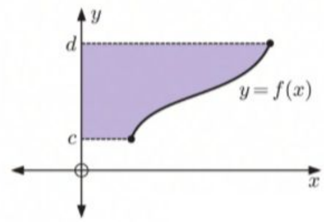
$$\int_a^b f(x)\,dx \approx \frac{h}{2}\left(f(x_0) + 2\sum_{i=1}^{n-1} f(x_i) + f(x_n)\right)$$



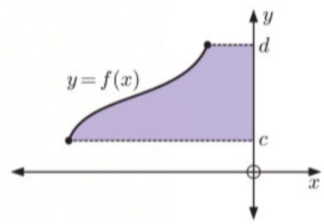
The area between a curve and the y-axis

Consider an invertible function $f(x)$.

- If $x = f^{-1}(y) > 0$ for $c \leqslant y \leqslant d$, shaded area $= \int_c^d f^{-1}(y)\,dy$



- If $x = f^{-1}(y) < 0$ for $c \leqslant y \leqslant d$, shaded area $= -\int_c^d f^{-1}(y)\,dy$

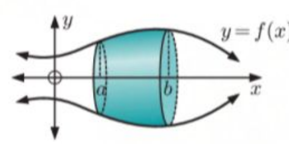


Solids of revolution

- When the region enclosed by $y = f(x)$, the x-axis, and the vertical lines $x = a$ and $x = b$ is revolved through 2π about the x-axis to generate a solid, the volume of the solid is given by

$$V = \pi \int_a^b [f(x)]^2\,dx$$

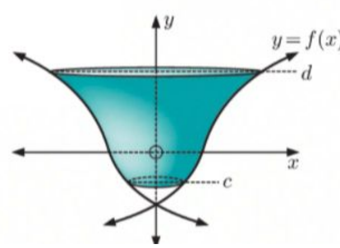
$$\text{or} \quad \pi \int_a^b y^2\,dx$$



- When the region enclosed by the invertible function $y = f(x)$, the y-axis, and the horizontal lines $y = c$ and $y = d$ is revolved through 2π about the y-axis to generate a solid, the volume of the solid is given by

$$V = \pi \int_c^d [f^{-1}(y)]^2\,dy$$

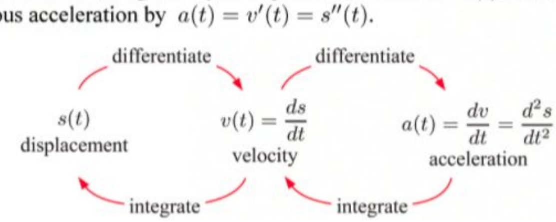
$$\text{or} \quad \pi \int_c^d x^2\,dy.$$



KINEMATICS

Suppose an object moves along a straight line.

Its position relative to the origin at time t is given by a displacement function $s(t)$. Its instantaneous velocity is given by $v(t) = s'(t)$, and its instantaneous acceleration by $a(t) = v'(t) = s''(t)$.



You should understand the physical meaning of the different signs of displacement, velocity, and acceleration.

Displacement:

$s(t)$	*Interpretation*
$= 0$	The object is at O
> 0	The object is to the right of O
< 0	The object is to the left of O

Velocity:

$v(t)$	*Interpretation*
$= 0$	The object is instantaneously at rest
> 0	The object is moving to the right
< 0	The object is moving to the left

Acceleration:

$a(t)$	*Interpretation*
> 0	The velocity of the object is increasing
< 0	The velocity of the object is decreasing
$= 0$	The velocity of the object may be at a maximum or a minimum

Speed

The **speed** at any instant is the magnitude of the object's velocity. If $S(t)$ represents the speed then $S = |\,v\,|$.

If the signs of $v(t)$ and $a(t)$ are the same then the speed of the object is increasing.

If the signs of $v(t)$ and $a(t)$ are different then the speed of the object is decreasing.

Displacement and distance travelled

For the time interval $t_1 \leqslant t \leqslant t_2$:

- displacement $= s(t_2) - s(t_1) = \int_{t_1}^{t_2} v(t)\,dt$
- total distance travelled $= \int_{t_1}^{t_2} |\,v(t)\,|\,dt$.

Velocity and acceleration in terms of displacement

If we are given an object's velocity in terms of its displacement s, we can write its acceleration in terms of s using the formula $a = v\,\dfrac{dv}{ds}$.

DIFFERENTIAL EQUATIONS

A **differential equation** is an equation involving a derivative of a function.

Euler's method allows us to approximate the solution curve to the differential equation $\dfrac{dy}{dx} = f(x, y)$ with particular solution passing through the point (x_0, y_0).

At each stage we perform the iterative procedure $\begin{cases} x_i = x_{i-1} + h \\ y_i = y_{i-1} + hf(x_{i-1}, y_{i-1}) \end{cases}$ where h is the step size.

To approximate $y(x_n)$ where $x_n = x_0 + nh$, we perform the procedure n times.

You should be able to use and interpret **slope fields** for differential equations of the form $\dfrac{dy}{dx} = f(x, y)$.

Separable differential equations are differential equations of the form $\dfrac{dy}{dx} = f(x)g(y)$.

To solve these equations, we rearrange the equation and integrate both sides with respect to x to obtain the form $\int \dfrac{1}{g(y)}\,dy = \int f(x)\,dx$. The variables are now separated, so we can find the two integrals separately and solve the equation.

Coupled differential equations

Differential equations which need to be solved simultaneously are said to be **coupled**.

For the coupled system of differential equations $\begin{cases} \dfrac{dx}{dt} = f_1(x, y) \\ \dfrac{dy}{dt} = f_2(x, y) \end{cases}$,

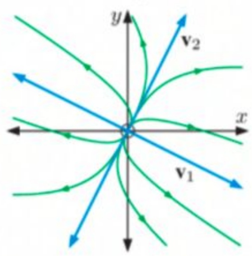
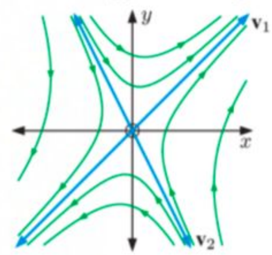
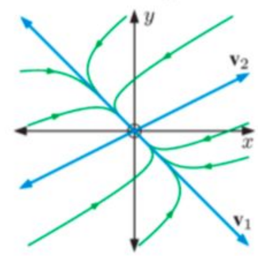
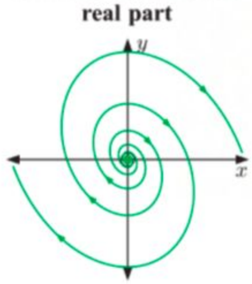
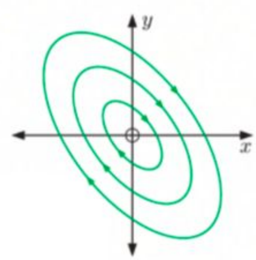
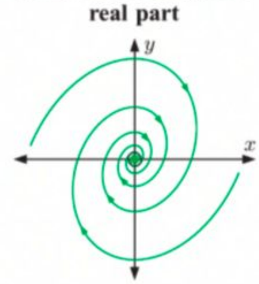
the **trajectory vector** at any point (x, y) is given by $\begin{pmatrix} \dfrac{dx}{dt} \\ \dfrac{dy}{dt} \end{pmatrix}$.

The **phase portrait** for the system is constructed by drawing the trajectory vector at every point on a grid. You should be able to draw a solution curve on a phase portrait from a given initial point.

An **equilibrium point** occurs when $\frac{dx}{dt} = \frac{dy}{dt} = 0$ simultaneously.

Coupled linear differential equations have the form $\begin{cases} \frac{dx}{dt} = ax + by \\ \frac{dy}{dt} = cx + dy \end{cases}$. This can be written more concisely as $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x}$, where $\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is the **matrix of coefficients.**

For coupled linear differential equations, the only equilibrium point is $(0, 0)$. The behaviour of the system around the equilibrium point is determined by the eigenvalues of $\mathbf{A}$.

real and positive	**real and opposite in sign**	**real and negative**
 **unstable fixed** If $\lambda_1 > \lambda_2 > 0$, trajectories become parallel to $\mathbf{v}_1$ as $t \to \infty$.	 **saddle point** If $\lambda_1 > 0 > \lambda_2$, trajectories approach $k\mathbf{v}_1$, $k \in \mathbb{R}$ as $t \to \infty$.	 **stable fixed** If $0 > \lambda_1 > \lambda_2$, trajectories approach O along $k\mathbf{v}_1$, $k \in \mathbb{R}$.
complex with positive real part	**purely imaginary**	**complex with negative real part**
 **unstable spiral**	 **centre**	 **stable spiral**

If $\mathbf{A}$ has real eigenvalues λ_1, λ_2 with corresponding eigenvectors $\mathbf{v}_1$, $\mathbf{v}_2$, the general solution to the system is $\mathbf{x} = Ae^{\lambda_1 t}\mathbf{v}_1 + Be^{\lambda_2 t}\mathbf{v}_2$. If an initial point is given, we can determine A and B and hence find a particular solution.

Euler's method for coupled equations

Euler's method allows us to approximate the solution curve to the coupled equations $\begin{cases} \frac{dx}{dt} = f_1(x, y, t) \\ \frac{dy}{dt} = f_2(x, y, t) \end{cases}$ with initial point (x_0, y_0) at time t_0.

At each stage we perform the iterative procedure $\begin{cases} t_i = t_{i-1} + h \\ x_i = x_{i-1} + hf_1(x_{i-1}, y_{i-1}, t_{i-1}) \\ y_i = y_{i-1} + hf_2(x_{i-1}, y_{i-1}, t_{i-1}) \end{cases}$ where h is the step size.

Second order differential equations

A **second order differential equation** is a differential equation which involves a second derivative, for example

$$\frac{d^2x}{dt^2} - 2\frac{dx}{dt} + 3x = 0.$$

You should be able to transform a second order differential equation into a system of coupled first order differential equations.