Mathematics HL

1 Algebra

1.1 Sequences and Series

Arithmetic progressions

- $T_n = U_n = a + (n-1)d$.
- A sequence is an A.P if $T_n - T_{n-1} = d =$ constant.
- $S_n = \frac{n}{2}(a+l) = \frac{n}{2}(2a + (n-1)d)$.
- $T_n = S_n - S_{n-1}$.

Geometric progressions

- $T_n = ar^{n-1}$.
- A sequence is a G.P if $\frac{T_n}{T_{n-1}} = r =$ constant.
- $S_n = \frac{a(1-r^n)}{1-r}$.
- $|r| < 1 \implies S_\infty = \frac{a}{1-r}$.
- $|r| > 1 \implies$ divergent.

1.2 Summation

For $\sum\limits_{r=m}^{n} u_r$, the number of terms is $(n-m+1)$.

$$\sum_{r=1}^{n}(x_r \pm y_r) = \sum_{r=1}^{n} x_r \pm \sum_{r=1}^{n} y_r$$

$$\sum_{r=1}^{n} ku_r = k\sum_{r=1}^{n} u_r$$

$$\sum_{r=m}^{n} u_r = \sum_{r=1}^{n} u_r - \sum_{r=1}^{m-1} u_r$$

Useful sums:

$$\sum_{r=1}^{n} r = \frac{1}{2}n(n+1)$$

$$\sum_{r=1}^{n} r^2 = \frac{1}{6}n(n+1)(2n+1)$$

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

1.3 Permutations and combinations

$$^nC_r = \frac{n!}{r!(n-r)!} \qquad ^nP_r =^n C_r \cdot r!$$

If m objects are identical and the remaining are distinct (a total of n objects), permutations $= \frac{n!}{m!}$

1.4 The Binomial Thoerem

$$(1+x)^n = 1 + nx + \frac{n(n-1)}{2!}x^2 + \frac{n(n-1)(n-2)}{3!}x^3 + \cdots$$

1.5 Mathematical induction

1. Let P_n be the statement: *ello* for all $n \in \mathbb{Z}^+$.
2. For $n = 1$: LHS = *something*. RHS = *something* $\implies P_1$ is true.
3. Assume P_k is true for some $k \in \mathbb{Z}^+$.
4. Showing that P_{k+1} is true: *it is true!*
5. Since P_1 is true, and P_k is true $\implies P_{k+1}$ is true, by Mathematical Induction, P_n is true for all $n \in \mathbb{Z}^+$.

To do the inductive step:

- $\displaystyle\sum_{r=1}^{k+1} u_r = u_{k+1} + \sum_{r=1}^{k} u_r$
- $\displaystyle\frac{d^{k+1}y}{dx^{k+1}} = \frac{d}{dx}(\frac{d^k y}{dx^k})$
- For divisibility, let the expression = a multiple of m. You can always rearrange the inductive hypothesis.

2 Functions and equations

- A function is a to-one relationship.
- If the vertical line $x = a$ cuts the graph at one point only, then f is a function. If it cuts more than once, give an example.
- If a function passes the horizontal line test, it will have an inverse.
- The inverse is just a reflection of the graph in the line $y = x$.
- For inverse functions, $R_f = D_{f^{-1}}$ and $D_f = R_{f^{-1}}$.
- For gf to exist, $R_f \subseteq D_g$.
- $D_{gf} = D_f$.
- $R_{gf} = R_g|(D_g = R_f)$.
- $(g \circ f)^{-1}(x) = (f^{-1} \circ g^{-1})(x)$.
- $ff^{-1}(x)$ may not necessarily intersect with $f^{-1}f(x)$, it depends on the domain.
- For a periodic function, $f(x) = f(x+c)$.

2.1 Graphs

- To transform, TSST. (translate and stretch)x then (translate and stretch)y.
- For $y = |f(x)|$, retain $y \geq 0$, then reflect $y < 0$.
- For $y = f(|x|)$, retain $x \geq 0$, then reflect $x \geq 0$ to the left of the x-axis.
- For each transformation, you're allowed to replace x by something else.

2.2 Polynomials

- For a polynomial of degree n:
  - The sum of individual roots $= -\frac{a_{n-1}}{a_n}$
  - The sum of (choose 2) roots $= \frac{a_{n-2}}{a_n}$
  - The sum of (choose 3) roots $= -\frac{a_{n-3}}{a_n}$
  - The product of roots, i.e the sum of (choose n) roots $= (-1)^n \frac{a_o}{a_n}$
- For the special case of a quadratic: $\alpha + \beta = -\frac{b}{a}$, $\alpha\beta = \frac{c}{a}$
- A polynomial of degree n has a maximum of n roots, but some of these may be complex.

2.3 Circular functions and Trigonometry

- The ambiguous case of the sine rule occurs when the angle you are trying to find is opposite the longest side.
- $\sin(-\theta) = -\sin\theta \qquad \tan(-\theta) = -\tan\theta \qquad$ (odd functions).
- $\cos(-\theta) = \cos\theta \qquad$ (even function).
- For $\pi \pm \theta$ or $2\pi \pm \theta$: sin-sin, cos-cos, tan-tan.
- For $\frac{\pi}{2} \pm \theta$ or $\frac{3\pi}{2} \pm \theta$: sin-cos, cos-sin, tan-cot.
- $\tan x = \cot(\frac{\pi}{2} - x)$.
- $\sec x = \csc(\frac{\pi}{2} - x)$.
- The domain of $\arcsin x$ and $\arccos x$ are $[-1,1]$.
- $\cos(\arcsin x) = \sin(\arccos x) = \sqrt{1-x^2}$.
- A circle with centre (h, k) and radius r is described by:

$$(x-h)^2 + (y-k)^2 = r^2$$

- To simplify an expression with trig, it may help to use the half angle formula.

$$\frac{\sin\theta}{1+\cos\theta} = \frac{2\sin\frac{\theta}{2}\cos\frac{\theta}{2}}{1+2\cos^2\frac{\theta}{2}-1} = \tan\frac{\theta}{2}$$

2.4 Systems of equations

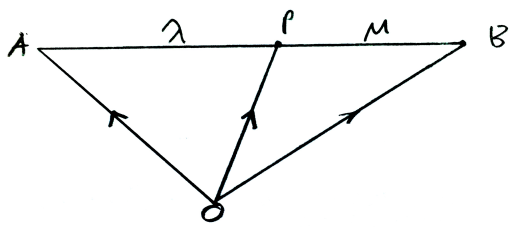
- A system of equations can be written as an augmented matrix:

$$\begin{aligned} 2x + 3y + 4z &= 2 \\ 3x - 2y + z &= -3 \\ x + 4y - z &= 5 \end{aligned} \rightarrow \left(\begin{array}{ccc|c} 2 & 3 & 4 & 2 \\ 3 & -2 & 1 & -3 \\ 1 & 4 & -1 & 5 \end{array}\right)$$

- A system is **consistent** if it has solutions.
- A system is **inconsistent** if one of the rows reduces to $0 = a$.
- If the last row reduces to $0 = 0$, there are infinitely many solutions and the general solution can be found by setting $z = \lambda$ where λ is a parameter.
- If the determinant of the 3×3 matrix is zero, then there is no unique solution (i.e either no solutions or infinite solutions).
- This links to planes, since the Cartesian equation of a plane is $ax + by + cz = d$.

3 Vectors

- A vector $\overrightarrow{AB}$ can be represented by a straight line, with an arrow, joining A and B.
- A vector can also be denoted with a lower case letter, e.g **a**, which is written with a tilde below it.
- A position vector defines the position of a point relative to the origin. $\mathbf{a} = \overrightarrow{OA}$.
- The Cartesian form of a vector: $\mathbf{r} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$, or $\mathbf{r} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$
- $|\mathbf{r}| = \sqrt{x^2 + y^2 + z^2}$
- A unit vector: $\hat{\mathbf{r}} = \frac{\mathbf{r}}{|\mathbf{r}|}$
- The Ratio Thoerem: $\overrightarrow{OP} = \frac{\mu\overrightarrow{OA} + \lambda\overrightarrow{OB}}{\mu + \lambda}$



3.1 Scalar products

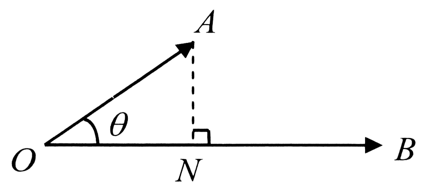
- The scalar product of two vectors is defined as $\mathbf{a} \cdot \mathbf{b} = |\mathbf{a}||\mathbf{b}|\cos\theta$.
- The vectors must both converge or diverge from one point.
- $\mathbf{a} \cdot \mathbf{a} = |\mathbf{a}|^2$.
- Most algebra works, except for cancellation and division.

- $\mathbf{a} \perp \mathbf{b} \iff \mathbf{a} \cdot \mathbf{b} = 0$.
- $\begin{pmatrix} a_1 \\ a_2 \\ b_2 \end{pmatrix} \cdot \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = a_1 b_1 + a_2 b_2 + a_3 b_3$.

3.2 Vector products

- The vector product of two vectors is defined as $\mathbf{a} \times \mathbf{b} = |\mathbf{a}||\mathbf{b}| \sin\theta \hat{\mathbf{n}}$.
- $\hat{\mathbf{n}}$ is a unit vector perpendicular to $\mathbf{a}$ and $\mathbf{b}$.
- $(\mathbf{a} \times \mathbf{b}) = -(\mathbf{b} \times \mathbf{a})$.
- $(\lambda \mathbf{a}) \times (\mu \mathbf{b}) = (\lambda\mu)(\mathbf{a} \times \mathbf{b})$.
- $|\mathbf{a} \times \mathbf{b}| = |\mathbf{a}||\mathbf{b}| \sin\theta$.
- $\mathbf{a} \parallel \mathbf{b} \iff \mathbf{a} \times \mathbf{b} = 0$, hence $\mathbf{a} \times \mathbf{a} = 0$.
- $\mathbf{a} \perp \mathbf{b} \iff \mathbf{a} \times \mathbf{b} = |\mathbf{a}||\mathbf{b}|$.
- $\mathbf{a} \cdot (\mathbf{a} \times \mathbf{b}) = 0$.
- $\begin{pmatrix} a_1 \\ a_2 \\ b_2 \end{pmatrix} \times \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = \begin{pmatrix} a_2 b_3 - b_2 a_3 \\ -(a_1 b_3 - b_1 a_3) \\ a_1 b_2 - b_1 a_2 \end{pmatrix}$. Cover top find det, cover mid find negative det, cover bot find det.
- Area $\Delta ABC = \frac{1}{2}|\overrightarrow{AB}||\overrightarrow{AC}| \sin\theta = \frac{1}{2}|\overrightarrow{AB} \times \overrightarrow{AC}|$.

3.3 Projections and resolving vectors



- The length of the horizontal projection of $\mathbf{a}$ onto $\mathbf{b} = \overrightarrow{ON} = |\mathbf{a}||\hat{\mathbf{b}}| \cos\theta = \mathbf{a} \cdot \hat{\mathbf{b}}$
- The length of the vertical projection is given by $|AN| = |\mathbf{a} \times \hat{\mathbf{b}}|$
- The horizontal projection vector is then $\mathbf{u} = (\mathbf{a} \cdot \hat{\mathbf{b}})\hat{\mathbf{b}}$, which is the same as the resolved component of $\mathbf{a}$ parallel to $\mathbf{b}$.
- The perpendicular component of $\mathbf{a}$ is $\mathbf{v} = \mathbf{a} - \mathbf{u}$.

3.4 Straight lines

$$l: \quad \mathbf{r} = \mathbf{a} + \lambda \mathbf{d}, \qquad \lambda \in \mathbb{R}.$$

- The vector equation of a line uses a position vector $\mathbf{a}$ of a fixed point on l, and a direction vector $\mathbf{d}$ parallel to l, to find the position vector of any point on the line ($\mathbf{r}$).
- λ is a real parameter, which means that the vector equation of a line is not unique.

- To get the **parametric form**, we write the equation as column vectors then equate components:

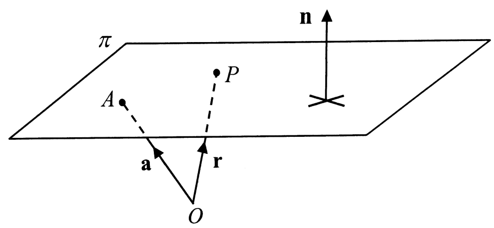
$$\begin{cases} x = a_1 + \lambda d_1, \\ y = a_2 + \lambda d_2, & \lambda \in \mathbb{R} \\ z = a_3 + \lambda d_3, \end{cases}$$

- To get the **Cartesian form**, make λ the subject then eliminate it.

$$\begin{cases} x = \mathbf{a_1} + \lambda \mathbf{d_1}, \\ y = \mathbf{a_2} + \lambda \mathbf{d_2}, \\ z = \mathbf{a_3} + \lambda \mathbf{d_3} \end{cases} \implies \begin{cases} \frac{x-\mathbf{a_1}}{\mathbf{d_1}} = \lambda, \\ \frac{y-\mathbf{a_2}}{\mathbf{d_2}} = \lambda, \\ \frac{z-\mathbf{a_3}}{\mathbf{d_3}} = \lambda \end{cases} \implies \frac{x-\mathbf{a_1}}{\mathbf{d_1}} = \frac{y-\mathbf{a_2}}{\mathbf{d_2}} = \frac{z-\mathbf{a_3}}{\mathbf{d_3}} \quad (= \lambda)$$

- l_1 and l_2 are parallel $\iff$ $\mathbf{d_1}$ and $\mathbf{d_2}$ are parallel $\iff$ $\mathbf{d_1} = k\mathbf{d_2}$, for some $k \in \mathbb{R}$.
- l_1 and l_2 intersect $\iff$
  - $\mathbf{d_1}$ is not parallel to $\mathbf{d_2}$ AND
  - there exist unique values of λ and μ such that $\mathbf{a_1} + \lambda\mathbf{d_1} = \mathbf{a_2} + \mu\mathbf{d_2}$.
- The lines are skew $\iff$ the direction vectors aren't parallel and there aren't unique values of λ and μ.
- The acute angle between two lines is giveb by $\cos^{-1} |\frac{\mathbf{d_1} \cdot \mathbf{d_2}}{|\mathbf{d_1}||\mathbf{d_2}|}|$

3.5 Planes



$$\overrightarrow{AP} \perp \mathbf{n} \implies \overrightarrow{AP} \cdot \mathbf{n} = 0 \implies (\mathbf{r} - \mathbf{a}) \cdot \mathbf{n} = 0.$$

- The **scalar product form** of the vector equation of the plane is $\mathbf{r} \cdot \mathbf{n} = \mathbf{a} \cdot \mathbf{n}$, where $\mathbf{a}$ is a fixed point on the plane.
- n can be found by taking the cross product of two known vectors parallel to the plane.
- The shortest distance between the origin and the plane: $|d| = |\mathbf{a} \cdot \hat{\mathbf{n}}|$
- The **parametric form** of the vector equation of the plane:

$$\pi : \quad \mathbf{r} = \mathbf{a} + \lambda\mathbf{d_1} + \mu\mathbf{d_2}, \qquad \lambda, \mu \in \mathbb{R}$$

- By expanding the scalar product form, we can arrive at the **Cartesian form**:

$$\mathbf{r} \cdot \mathbf{n} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} \cdot \begin{pmatrix} a \\ b \\ c \end{pmatrix} = ax + by + cz = D$$

- A line will be parallel to a plane if it is perpendicular to $\mathbf{n}$, i.e $\mathbf{n} \cdot \mathbf{d} = 0$ and there is no common point.

- If not parallel, it will intersect at a point, which can be found by substituting the line equation into the plane equation.
- The acute angle between l and π: $\sin\theta = |\frac{\mathbf{d}\cdot\mathbf{n}}{|\mathbf{d}||\mathbf{n}|}|$
- When planes intersect, their Cartesian forms can be combined to form a system of simultaneous equations
  - If there is a unique solution, the planes intersect at a point.
  - If there are infinitely many solutions, the planes intersect in a line.
  - If there are no solutions, the three planes do not intersect.

4 Calculus

4.1 Differentiation

- If the limit of the denominator of a rational function is zero, you cannot substitute to find the limit: either 'juggle' or use l'Hopital's rule, e.g:

$$\lim_{x\to 0}\left(\frac{\sin x}{x}\right) = \lim_{x\to 0}\left(\frac{\cos x}{1}\right) = 1$$

- The definition of the derivative:

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

- Special derivatives:

  $\frac{d}{dx}(\sec x) = \sec x \tan x$

  $\frac{d}{dx}(\csc x) = -\csc x \cot x$

  $\frac{d}{dx}(\cot x) = -\csc^2 x$

  $\frac{d}{dx}(\arcsin x) = \frac{1}{\sqrt{1-x^2}}$

  $\frac{d}{dx}(\arccos x) = -\frac{1}{\sqrt{1-x^2}}$

  $\frac{d}{dx}(\arctan x) = \frac{1}{1+x^2}$

  $\frac{d}{dx}(a^x) = a^x \ln a$

  $\frac{d}{dx}(\log_a x) = \frac{1}{x\ln a}$

- $f(x)$ is an **increasing function** on (a, b) if $\frac{dy}{dx} > 0$ on that interval, or **strictly increasing function** if $\frac{dy}{dx} \geq 0$.
- $f(x)$ is **concave upwards** on (a, b) if $\frac{d^2y}{dx^2} > 0$.
- If the derivative at a point is zero, the function is stationary.
- If the derivative at a point is ∞, there is a vertical line.
- For a point of inflexion, $\frac{d^2y}{dx^2} = 0$ AND the sign of $\frac{d^2y}{dx^2}$ changes, i.e concativity changes.
- Sketching the graph of $f'(x)$ given $f(x)$:

  - Stationary point $\to$ x-intercept.
  - $f(x)$ increasing $\to$ $f'(x)$ above x-axis.
  - Point of inflexion $\to$ turning point.
- The gradient at any point on the curve: $m = \frac{dy}{dx}|_{x=x_0}$.
- The equation of a tangent to the curve at (x_0, y_0): $y - y_0 = m(x - x_0)$.
- If two variables are related, their rates of change are also related:

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

- In kinematics especially:

$$a = \frac{dv}{dt} = \frac{dv}{dx} \cdot \frac{ds}{dt} = v\frac{dv}{ds}$$

4.2 Integration

$$\int (px+q)^n dx = \frac{(px+q)^{n+1}}{p(n+1)} + C$$

$$\int f'(x)(f(x))^n dx = \frac{(f(x))^{n+1}}{n+1} + C$$

$$\int \frac{f'(x)}{f(x)} dx = \ln|f(x)| + C$$

$$\int \frac{1}{x\ln x} dx = \int \frac{x^{-1}}{\ln x} + C = \ln|\ln|x|| + C$$

$$\int \tan x \; dx = \ln|\sec x| + C$$

$$\int \sec x \; dx = \ln|\sec x + \tan x| + C$$

$$\int \csc x \; dx = -\ln|\csc x + \cot x| + C$$

$$\int \frac{1}{(x+k)^2 + a^2} \; dx = \frac{1}{a}\arctan(\frac{x+k}{a}) + C$$

$$\int \frac{1}{\sqrt{a^2 - (x+k)^2}} \; dx = \arcsin(\frac{x+k}{a}) + C$$

- To integrate $\sin^2 x$ or $\cos^2 x$, we expand $\cos(2x)$ and rearrange.
- To integrate $\sin^3 x$, split into $\int \sin x(\sin^2 x)dx$, then use $\sin^2 x + \cos^2 x = 1$.
- If the integral is of the form:

$$\int \frac{px+q}{\sqrt{Ax^2+Bx+C}} dx \quad \text{or} \quad \int \frac{px+q}{Ax^2+Bx+C} dx$$

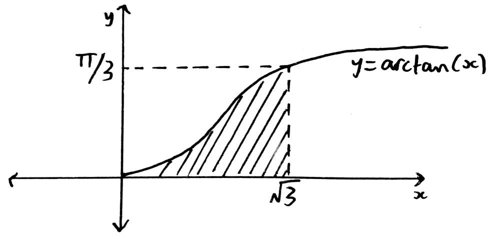
  use sorcery to change it into $\int \frac{f'(x)}{f(x)} dx$ or $\int f'(x)(f(x))^n dx$.

- Integration by substitution:
  1. Replace dx by $\frac{dx}{dt} \cdot dt$.
  2. Substitute by replacing all x with $g(t)$.
     Then: $\int f(x)\ dx = \int f(g(t))\frac{dx}{dt} \cdot dt$
- Integration by parts:

  $\int udv = uv - \int vdu$
- To choose which one to differentiate, use LIATE: **L**ogs, **I**nverse trig, **A**lgebraic, **T**rig, **E**xponentials.

4.3 Definite integrals

- $\int_a^b f(x)dx = -\int_b^a f(x)dx$
- The definite integral $\int_a^b f(x)dx$ can only be found if $f(x)$ is defined for all $x \in (a, b)$.
- The area between a curve and the y-axis: $\int_a^b f(y)dy$
- If a function is difficult to integrate, try integrating its inverse w.r.t y then subtract from a rectangle. e.g:



$$\int_0^{\sqrt{3}} \arctan x\ dx = \frac{\pi\sqrt{3}}{3} - \int_0^{\frac{\pi}{3}} \tan y\ dy$$

- The area between the curve and the axis is always $\int_a^b |f(x)|dx$.
- The area between two curves is always $\int_a^b y_1 - y_2 dx$.
- The volume of revolution:
$$V = \pi \int_a^b y^2 dx$$
- The volume of revolution of the area enclosed by two curves:
$$V = \pi \int_a^b (y_1)^2 dx - \pi \int_a^b (y_2)^2 dx$$

5 Probability and Statistics

5.1 Probability

- Two events A and B are **mutually exclusive** if $P(A \cap B) = 0$.
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
- $P(A|B) = \frac{P(A \cap B)}{P(B)}$.
- A and B are **independent** if $P(A|B) = P(A)$, so if they are independent $P(A \cap B) = P(A)P(B)$.

5.2 Discrete random variables

- $P(X = x)$ is the probability that the r.v X will assume a value of x.
- A discrete r.v can assume a countable number of values.
- For a d.r.v taking values $x_1, x_2, x_3, ..., x_n$, the **probability distribution** is defined as $P(X = x_i)$, such that:
$$0 \leq P(X = x_i) \leq 1 \qquad \text{and} \qquad \sum_{\text{all } i} P(X = x_i) = 1$$
- The expectation of a d.r.v:
$E(X) = \mu = \sum xP(X = x)$
$E(g(X)) = \sum g(x)P(X = x)$
$E(a) = a$
$E(aX \pm B) = aE(X) \pm b$
$E(X \pm Y) = E(X) \pm E(Y)$
- The variance of a d.r.v:
$\text{Var}(X) = \sigma^2 = E((x - \mu)^2) = E(X^2) - [E(X)]^2$
$\text{Var}(a) = 0$
$\text{Var}(ax + b) = a^2\text{Var}(()X)$
$\text{Var}(X \pm Y) = \text{Var}(X) + \text{Var}(Y)$ (only if X and Y are independent)
- Note: never subtract variance.

5.3 Discrete distributions

<u>The Binomial distribution</u>

$$X \sim B(n, p) \qquad P(X = x) = \binom{n}{x} p^x q^{n-x} \qquad E(X) = np \qquad \text{Var}(X) = npq$$

- There are n independent trials, two possible outcomes (either 'success' or 'failure'), with constant probability of success p, X is the number of 'successes'.
- The Binomial distribution is a combination of n Bernoulli trials.
- For $P(X \leq x)$, we find $P(X = 0) + P(X = 1) + P(X = 2) + ... + P(X = x)$.

<u>The Poisson distribution</u>

$$X \sim Po(\lambda) \qquad P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!} \qquad E(X) = \text{Var}(X) = \lambda$$

- For a random variable in time or space, if there is no chance of simultaneous events, the events are independent, and the events have a constant probability of occuring, it is a Poisson process.
- λ is the parameter, and defines the number of events in a given time/space.
- If $X \sim Po(\lambda)$ and $Y \sim Po(\mu)$, then $X + Y \sim Po(\lambda + \mu)$.

The Geometric distribution

$$X \sim Geo(p) \qquad P(X = x) = pq^{x-1},\ x \geq 1 \qquad E(X) = \frac{1}{p} \qquad \text{Var}(X) = \frac{q}{p^2}$$

If we perform a series of independent trials with a probability p of success, X is the number of trials up to and including the first success.

$$\begin{aligned} P(X > x) &= P(X = x+1) + P(X = x+2) + \dots \\ &= pq^x + pq^{x+1} + pq^{x+2} + \dots \\ &= pq^x(1 + q + q^2 + ...) = pq^x(\frac{1}{1-q}) = q^x \end{aligned}$$

$$P(X > a + b | X > a) = P(X > b) = q^b$$

The Negative Binomial distribution

$$X \sim NB(r, p) \qquad P(X = x) = \binom{x-1}{r-1} p^r q^{x-r},\ r \geq 1,\ x \geq 1 \qquad E(X) = \frac{r}{p} \qquad \text{Var}(X) = \frac{rq}{p^2}$$

- X is the number of trials needed to achieve r successes.
- The Negative Binomial distribution is just a combination of r geometric trials.

5.4 Continuous random variables and CDFs

- Instead of probability distributions, we have probability density functions (PDFs), denoted by $f(x)$.
  - $f(x) \geq 0$ for all $x \in \mathbb{R}$
  - $\int_{-\infty}^{\infty} f(x)\ dx = 1$
- Continuous $\implies$ uncountable, so $P(X = x) = 0$. Therefore, $\geq$ or $>$ is irrelevant.

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

$$E(X) = \mu = \int_{-\infty}^{\infty} xf(x)\ dx$$

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

$$P(|X - a| < b) = P(-b < X - a < b)$$

- The mode of a c.r.v is the value of x which gives the maximum probability, i.e the x coordinate of the highest point in the domain.
- The **cumulative distribution function** (CDF):

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\ dt$$

$$\lim_{x \to -\infty} F(x) = 0 \qquad \lim_{x \to \infty} F(x) = 1$$

$$P(a < X < b) = F(b) - F(a)$$

$$\frac{d}{dx} F(x) = f(x)$$

- $F(x)$ is continuous and increasing (since $f(x) > 0$).
- To find the median m, set $F(m) = \frac{1}{2}$ and solve for m, i.e: $\int_{-\infty}^{m} f(t)dt = 0.5$

5.5 The Normal distribution

$$X \sim N(\mu, \sigma^2)$$

- The Normal distribution is a bell curve symmetrical about $x = \mu$.
- The mean = median = mode = μ.
- μ affects the location of the curve, whereas σ^2 affects the spread.
- The standard normal distribution is denoted by $Z \sim N(0, 1)$.
- Any normal distribution can be standardised: $Z = \frac{X-\mu}{\sigma}$
- The Z score represents the number of standard deviations away from the mean.
- To find c given $P(X < c) = p$, use invNorm.
- If $X \sim N(\mu_1, \sigma_1^2)$ and $Y \sim N(\mu_2, \sigma_2^2)$, then $aX + bY$ also has a normal distribution.

$$\begin{aligned} E(aX + bY) &= aE(X) + bE(Y) \\ &= a\mu_1 + b\mu_2 \\ \text{Var}(aX + bY) &= a^2\sigma_1^2 + b^2\sigma_2^2 \end{aligned}$$

$$aX + bY \sim N(a\mu_1 + b\mu_2,\ a^2\sigma_1^2 + b^2\sigma_2^2)$$

5.6 Sampling

- If X is a random variable, $X_1, X_2, X_3, ..., X_n$ are a sample of n independent observations.
- The sample mean:

$$\bar{X} = \frac{X_1 + X_2 + ... + X_n}{n}$$

$$E(\bar{X}) = E\left(\frac{X_1 + X_2 + ... + X_n}{n}\right) = \frac{nE(X)}{n} = E(X) = \mu$$

$$\text{Var}(\bar{X}) = \text{Var}\left(\frac{X_1 + X_2 + ... + X_n}{n}\right) = \frac{1}{n^2}\text{Var}(X_1 + X_2 + ... + X_n) = \frac{n\text{Var}(X)}{n^2} = \frac{\sigma^2}{n}$$

- For the sample sum: $E(S) = n\mu$, $\text{Var}(S) = n\sigma^2$
- Therefore, in a normal population:

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{n}) \qquad \sum_{r=1}^{n} X_r \sim N(n\mu, n\sigma^2)$$

- The **Central Limit Theorem** states that, for a large sample size ($n \geq 50$), the sample mean/sum of a sample from *any* distribution (e.g not normal), will approximately follow the normal distribution.

5.7 Estimators

- An **estimator** is a test statistic T based on observed data that estimates an unknown parameter θ.
- The estimator is **unbiased** if $E(T) = \theta$.
- The sample mean is an unbiased estimator of μ since $E(\bar{X}) = \mu$.
- However, the sample variance is not an unbiased estimator for σ^2 since $E(S_n^2) = \frac{n-1}{n}\sigma^2$.
- An unbiased estimator for σ^2:

$$\begin{aligned} s_{n-1}^2 = \frac{n}{n-1} \times S_n^2 &= \frac{n}{n-1}\left(\frac{1}{n}\sum x^2 - (\bar{x})^2\right) \\ &= \frac{1}{n-1}\left(\sum x^2 - \frac{(\sum x)^2}{n}\right) \end{aligned}$$

- An unbiased estimator is more **efficient** than another if it has a lower variance.

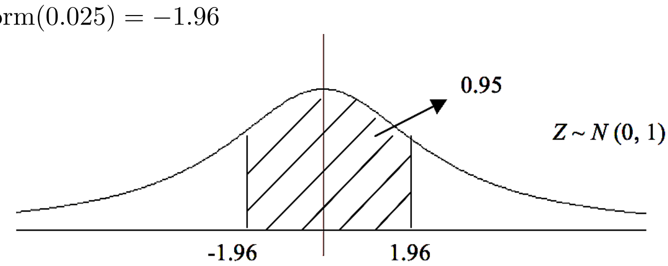
5.8 Confidence intervals

- A 95% confidence interval (CI) means that there is a 95% chance that the interval includes μ.
- For $X \sim N(\mu, \sigma^2)$, if we take a sample: $\bar{X} \sim N(\mu, \sigma^2)$.

$$\text{Confidence limits} = \bar{X} \pm Z_k \frac{\sigma}{\sqrt{n}}$$

$$\text{CI} = \left[\bar{X} - Z_k \frac{\sigma}{\sqrt{n}},\ \bar{X} + Z_k \frac{\sigma}{\sqrt{n}}\right]$$

- Z_k is the **critical value**, and is found using invNorm.
- For a 95% CI: invNorm(0.025) $= -1.96$



- The width of a CI is $2Z_k \frac{\sigma}{\sqrt{n}}$
- If we have a large sample from any population (μ and σ^2 unknown), we can use the CLT.

$$\text{CI} = \left[\bar{x} - Z_k \frac{s_{n-1}}{\sqrt{n}},\ \bar{x} + Z_k \frac{s_{n-1}}{\sqrt{n}}\right]$$

- If the population is normal but we do not know the variance, we use the t-distribution.
  $T = \dfrac{\bar{X} - \mu}{s_{n-1}/\sqrt{n}}$ follows a t-distribution with $n-1$ degrees of freedom.

$$\text{CI} = \left[\bar{x} - t_k \frac{s_{n-1}}{\sqrt{n}},\ \bar{x} + t_k \frac{s_{n-1}}{\sqrt{n}}\right]$$

σ^2	n	Assumptions	Test Statistic
known	large	CLT	$Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$
	small	normal	
unknown	large	CLT	$Z = \frac{\bar{X}-\mu}{s_{n-1}/\sqrt{n}} \sim N(0,1)$
	small	normal	$T = \frac{\bar{X}-\mu}{s_{n-1}/\sqrt{n}} \sim t_{n-1}$

5.9 Hypothesis testing

1. State H_0 and H_1.
2. Test statistic.
3. Level of significance and rejection criteria.
4. Compute p-value (or z-value or t-value).
5. Conclusion in context.

e.g

$H_0 : \mu = 3$
$H_1 : \mu > 3$
Test statistic: $Z = \dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}}$
Sig level $= 5\%$, one tailed.
Reject H_0 if $p < 0.05$
Since p-value $= 0.03 < 0.05,$ we reject H_0 and conclude that there is significant evidence at the 5% level that...

- $P(\text{Type I Error}) = P(H_0 \text{ rejected}|H_0 \text{ true}) = \alpha\%$. i.e $P(\text{Type I Error}) =$ **significance level**.
- $P(\text{Type II Error}) = P(H_0 \text{ accepted}|H_1 \text{ true})$.
- For example, for $H_0 : \mu = \mu_0 \quad H_1 : \mu = \mu_1$,

$$P(\text{Type II Error}) = P(H_0 \text{ accepted}|H_1 \text{ true}) = P(\bar{X} < \text{critical value } |\bar{X} \sim N(\mu_1, \sigma^2))$$

5.10 PGFs

$$G(t) = E(t^X) = \sum t^x P(X = x)$$
$$G(1) = 1$$
$$G'(t) = \sum x t^{x-1} P(X = x) \therefore E(X) = G'(1)$$

$$G''(t) = \sum x(x-1)t^{x-2} P(X = x)$$
$$G''(1) = \sum x^2 P(X = x) - \sum x P(X = x) = E(X^2) - E(X)$$
$$\therefore E(X^2) = G''(1) + G'(1)$$
$$\therefore \text{Var}(X) = G''(1) + G'(1) - [G'(1)]^2$$

$$\text{If } Z = X + Y,\ G_Z(t) = E(t^Z) = E(T^{X+Y}) = E(t^X)E(t^Y) = G_X(t)G_Y(t)$$

- To find $P(X = n)$, we use the Maclaurin series: $P(X = n) = \frac{G^n(0)}{n!}$.
- To prove most things about PGFs, differentiation will be involved (sometimes using the product rule and chain rule).

Binomial

If $Y \sim B(n, p)$, we can say that $Y = X_1 + X_2 + X_3 + ... + X_n$ where X is a Bernoulli trial.

x	0	1
$P(X = x)$	q	p

$G_X(t) = \sum t^x P(X = x) = q + pt$

$G_Y(t) = E(t^Y) = E(t^{X_1 + ... + X_n}) = \left[E(t^X)\right]^n = [G_X(t)]^n = (q + pt)^n$

Poisson

If $X \sim Po(\lambda)$, $P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}$.

$$\begin{aligned} G(t) = E(t^X) &= \sum t^x P(X = x) \\ &= \sum t^x \frac{e^{-\lambda}\lambda^x}{x!} \\ &= e^{-\lambda} \sum \frac{(\lambda t)^x}{x!} = e^{-\lambda} e^{\lambda t} = e^{\lambda(t-1)}. \end{aligned}$$

Geometric

If $X \sim Geo(p)$, $P(X = x) = pq^{x-1}$.

$$\begin{aligned} G(t) = E(t^X) &= \sum t^x P(X = x) \\ &= \sum t^x p q^{x-1} \\ &= pt + pt^2 q + pt^3 q^2 + pt^4 q^3 + ... + pt^n q^{n-1} + ... \\ S_\infty &= \frac{a}{1-r} = \frac{pt}{1-qt} \end{aligned}$$

Negative Binomial

If $Y \sim NB(r,p)$, we can say that $Y = X_1 + X_2 + X_3 + ... + X_r$, where $X \sim Geo(p)$.

$$G_Y(t) = E(t^Y) = E(t^{X_1+...+X_r}) = \left[E(t^X)\right]^r = [G_X(t)]^r = \left(\frac{pt}{1-qt}\right)^r$$

5.11 Bivariate data and correlations

- If X and Y are random variables, the joint probability distribution is $P(X = x \cap Y = y)$.
- $\sum\sum p(x,y) = 1$
- $E(XY) = \sum\sum xy\ p(x,y)$
- $\text{Cov}(X,Y) = E(XY) - E(X)E(Y)$. X and Y independent $\implies \text{Cov}(X,Y) = 0$.
- $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) - 2\text{Cov}(X,Y)$.
- The correlation coefficient measures the linear relationship between X and Y
$$\rho = \text{Corr}(X,Y) = \frac{\text{Cov}(X,Y)}{\sigma_X\sigma_Y} = \frac{\sigma_{XY}}{\sigma_X\sigma_Y}$$
- A **bivariate sample** consists of pairs of data (x_1, y_1). For a bivariate sample, the above points do not apply.
$$r = \frac{\sum xy - n\bar{x}\bar{y}}{\sqrt{(\sum x^2 - n\bar{x}^2)(\sum y^2 - n\bar{y}^2)}} = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}, \text{ where } S_{xx} = \sum x^2 - \frac{(\sum x)^2}{n}$$
- If $r = 0$, there is no linear relationship, but it does not imply that X and Y are independent.
- r is independent of the units, and does not show any causality.
- In maths, **controlled variable = independent variable**.
- The y-on-x regression line $y = a + bx$ will always pass through $(\bar{x}, \bar{y})$.
$$y - \bar{y} = b(x - \bar{x}), \text{ where } b = \frac{S_{xy}}{S_{xx}}$$
- The x-on-y regression line is denoted by $x = c + dy$.
$$bd = r^2 \qquad r = \pm\sqrt{bd}, \text{ the sign depends on whether the gradient is positive or negative.}$$
- We can statistically test evidence of a correlation by assuming both variables follow a bivariate normal distribution with correlation coefficient ρ.

  e.g

  $H_0 : \rho = 0$

  $H_1 : \rho \neq 0$

  Test statistic: $T = r\sqrt{\dfrac{n-2}{1-r^2}} \sim t_{n-2}$

  Sig level = 5%, two tailed.

  Reject H_0 if $|T| > \text{invt}(0.975, n-2)$

  Note: $T = r\sqrt{\dfrac{n-2}{1-r^2}}$ (sub in values)

  Since $|T| = 0.08 > 0.05$, we reject H_0 and conclude that there is a significant evidence at the 5% level that there is a correlation between...

6 Miscellaneous

- $(a \pm b)^3 = a^3 \pm 3a^2b + 3ab^2 \pm b^3$
- $a^3 + b^3 = (a + b)(a^2 - ab + b^2)$
- $a^3 - b^3 = (a - b)(a^2 + ab + b^2)$
- $|x + 3||x + 2| = |(x + 3)(x + 2)|$